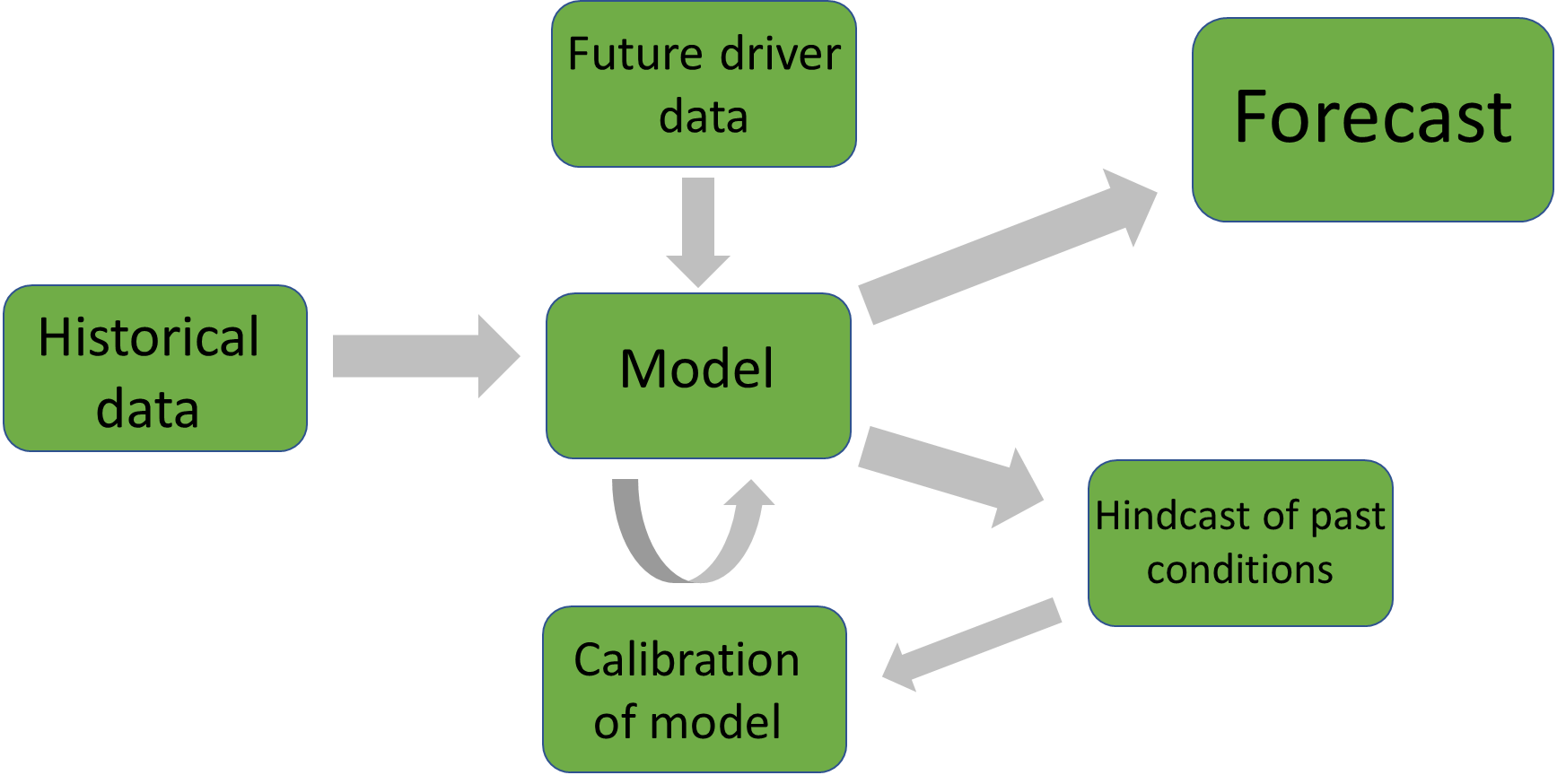
Final Project  
BIOL 5984, Spring 2019  
Whitney Woelmer

Abstract

Lakes and reservoirs are increasingly threatened by eutrophication, a result of rapidly changing land use and climate. Consequently, there is a growing need to assess the current and future state of freshwater ecosystems by adopting iterative, near-term forecasting. Because the field of ecological forecasting is relatively new, there is not yet a consensus as to the best approach for predicting future water quality. We chose to develop an autoregressive integrated moving average (ARIMA) model developed using a suite of physical, chemical, and biological monitoring data to hindcast chlorophyll-a from 2013-2016. The best ARIMA model contained mean daily discharge to the reservoir, mean daily shortwave, and chlorophyll-a from the previous week and hindcasted chlorophyll-a with an R2 = 0.49 and RMSE = 1.71 ug/L. This model will then be used to produce near-term (16-day) forecasts of chlorophyll-a using the Forecasting Lake and Reservoir Ecosystems (FLARE) framework.

Introduction

* Given the unprecedented level of anthropogenic degradation already experienced by freshwater lakes, reservoirs, streams, and wetlands due to land use and climate change, understanding not only the current state of our freshwater ecosystems, but predicting how they will respond tomorrow, next week, and next year is of utmost importance.
* Forecasting as a technique outside of ecology has been developing for decades in many disciplines and has substantial breadth. Across disciplines, forecasts are generally developed by using historical data to set up a model that produces hindcasts of past conditions (Figure 1). The hindcasts are compared with observed data to further calibrate a model, which is then used to produce a forecast when forced with future driver data (Figure 1).
* The development of forecasts of ecosystems and ecosystem services is still in its relative infancy. From a high-level literature review of current ecological forecasting studies (n=15), a majority use empirical approaches (65%) to forecast ecological variables.
* We chose to develop an autoregressive integrated moving average (ARIMA) model developed using a suite of physical, chemical, and biological monitoring data to hindcast chlorophyll-a from 2013-2016 and assessed the model performance using R2 and RMSE.



*Figure 1. Conceptual diagram of the process to develop a forecast.*

Methods

*Study Site*

Falling Creek Reservoir (FCR) is a small (~12 ha), shallow (maximum depth = ~9.3 m) dimictic drinking water reservoir located in southwestern Virginia (37.30333333˚N, -79.83722222˚E). 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 (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 (EDI Portal: Carey et al. 2018, Carey et al. 2019). 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 a staging server and pushed to Github multiple times per day. Sensor data include multiple meteorological (e.g., precipitation, shortwave radiation), physical (e.g., water temperature, dissolved oxygen), chemical (e.g. total and soluble nutrients), and biological variables (e.g., chlorophyll-a concentrations).

*Model Development*

We developed an empirical model to forecast chlorophyll-a, a proxy for phytoplankton, 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 model was 2013-2016 during which we have regularly, weekly coverage of both response and driver data. The autoregressive term in ARIMA was determined by selecting the previous timestep of chlorophyll-a with the highest Pearson’s r correlation coefficient with the current measurement of chlorophyll-a. From a pool of 53 potential meteorological, physical, chemical, and biological driver variables, we first focused on driver variables that have biological significance for phytoplankton growth and which are also predictable using physical models (e.g., temperature, discharge). We excluded variables that were correlated with each other through the use of a Pearson’s correlation analysis (r > 0.5 & r < -0.5). Using these variables, we developed all possible ARIMA combinations with the selected driver variables, and the best model was determined by AICc (corrected Akaike’s Information Criterion). Model fit was then assessed by R2 and RMSE.

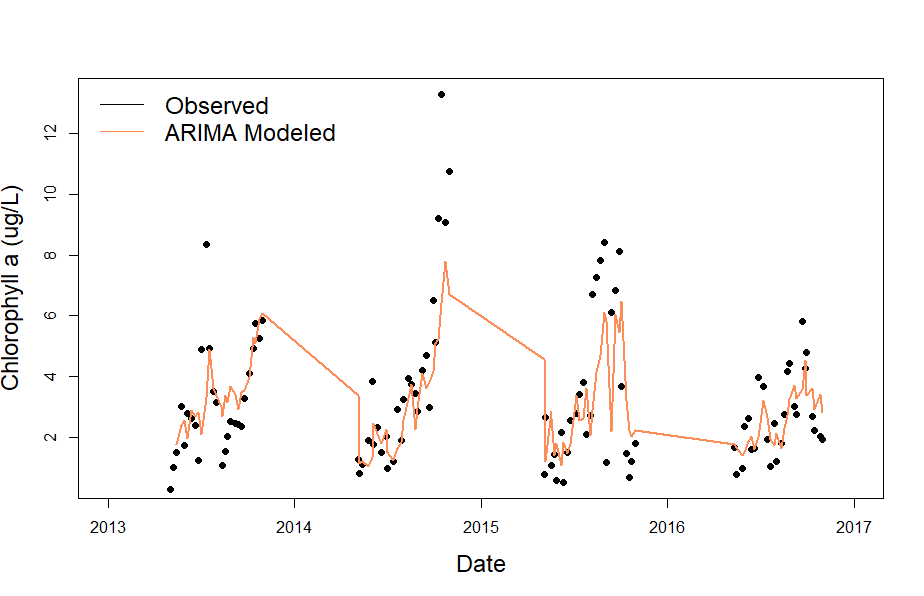
Results

Using the summer period of 2013-2016 as our training period, we developed an empirical model (ARIMA). Our best-fitting ARIMA model over 2013-2016 included discharge to the reservoir and shortwave radiation, where ‘t’ is the current timestep and ‘t-1’ is the previous timestep, with 1 standard error of each parameter term included:

**Chlorophyll-at = 1.65(±0.26) + 0.45(±0.08)Chlorophyll-at-1 – 3.05(±1.39)Discharget**

**– 0.0025(±0.00065)Shortwavet + Ɛ** (eqn. 1)

ARIMA hindcasted chlorophyll-a over 2013-2016 with an R2 = 0.49 and RMSE = 1.71 ug/L. The ARIMA model was able to successfully capture fluctuations at lower chlorophyll-a concentrations (<10 ug/L) (Figure 3). However, when chlorophyll-a reached values above ~10 ug/L, the model the model was unable to recreate these observed dynamics.



*Figure 3. Observed and modeled chlorophyll-a data over 2013-2016.*

Discussion

* Overall, our ARIMA model was able to capture nearly 50% of the variability in chlorophyll-a.
* The inclusion of mean daily discharge to the reservoir in our model highlights the importance of stream inputs to chlorophyll-a dynamics.
* Next steps include using our ARIMA model to produce forecasts of chlorophyll-a for Falling Creek Reservoir using the Forecasting Lakes and Reservoir Ecosystems (FLARE) framework.