Aurora Villa – Alex Etheridge Internship Plan

Aurora Contact Info

* Cell (alex has in work phone)
* Email: aurorav@pdx.edu
* Preferred notification method: texting
* Work / School Collaboration Operating System: teams – if we need to go to google, we can re-assess.
* Class Schedule:
  + Tues, Weds, Thurs classes
* Other Work Schedule – no work schedule
* Internship Schedule
  + Monday / Friday are open. Mid day is an ideal timing to meet 9 or later and end by 5.

Alex Contact Info

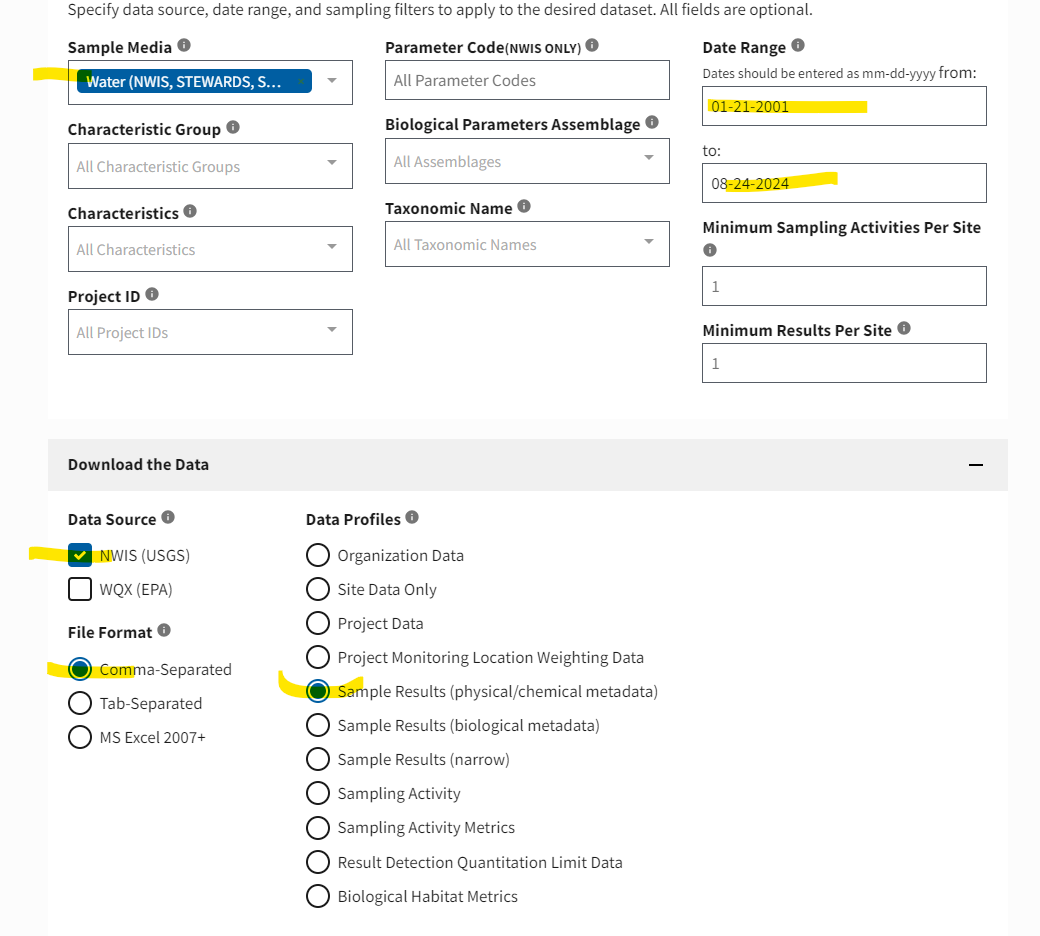
* Work Cell: 503-593-1837
* Work Email: [aetherid@usgs.gov](mailto:aetherid@usgs.gov)
* Personal Cell: 916-208-9639
* Working hours: ~9-6PM Monday through Friday
* Work Collaboration platform: Windows / Office 365, Outlook/Teams
* Optional platform: Google (would be using personal email)
* Preferred method of notification: text/call/teams chat or google chat, scheduled meetings in person or virtual.

Internship Goals:

Draft – subject to change with Aurora’s input.

1. Build and maintain a trusting, open, respectful and collaborative relationship.
2. Collate data for a time-series sites in the Willamette Basin with fluorescent chlorophyll data and chlorophyll-a pigment concentration sample data from a lab.
3. Create a “calibration data set” of sample pairs synced by time for fluorescent chlorophyll (fCHL) and chlorophyll-a pigment concentration data (chl-a).
4. Apply statistical regression to this dataset.
5. Complete the synthesis in R.
6. Utilize open/public access code and data to complete these objectives.
7. Consult and collaborate with other USGS scientists
8. Explore USGS data and consider applying for a longer term position. Openly discuss what is attractive and not attractive about the USGS workplace.

Tools & Software:

1. Data Grapher: [USGS Multi-Site Data Grapher](https://or.water.usgs.gov/cgi-bin/grapher/graph_multisite_setup.pl?parm=chla#step1)
   1. Let’s explore this site together and discuss which sites to select after we look at the data and discuss it.
      1. Tualatin River at Oswego Diversion Dam - USGS ID: 14207200
      2. Willamette R at Portland - USGS ID: 14211720
         1. fCHL: 1/21/2009 – 8/21/2024
   2. Outcome: evaluate patterns in fCHL data and understand fCHL in relation to other parameters, what is it, what does it do, why?
      1. Dissolved oxygen and pH generally go up with increasing “total chlorophyll” fCHL.
   3. Outcome: find a 3-5 long-term sites in the Willamette Basin with potential to explore regression analysis.
      1. Found two sites above to explore further.
      2. Downloaded data from Tualatin at Oswego from Data Grapher here: <https://or.water.usgs.gov/cgi-bin/grapher/tabler.pl?site_id=14207200&chla=on&begin_date=20010426&end_date=20240821>
      3. From Willamette River at Portland here:
      4. <https://or.water.usgs.gov/cgi-bin/grapher/tabler.pl?site_id=14211720&chla=on&begin_date=20010426&end_date=20240821>
2. Water Quality Portal and NWIS Web: [Water Quality Portal Data Sites for USGS-OR (waterqualitydata.us)](https://www.waterqualitydata.us/provider/NWIS/USGS-OR/) and/or [Water Quality Data Home](https://www.waterqualitydata.us/beta/)
   1. We were unable to navigate WQP and find any chl-a data.
   2. 
   3. 
   4. Obtain chl-a samples at sites of interest using the advanced version of the first link.
      1. To Do: clean up data and times and save as csv.
      2. Alex and Aurora finished:
         1. Identifying sites of interest
         2. Downloading fCHL time series from both sites
         3. Downloading chl-a discrete samples from both sites
         4. Formated fCHL timeseries w/ date format as Swedish/Sweden   
            YYYY-MM-DD HH:MM
         5. Review all excel files, filter out Chl-a “corrected for pheophytin” sample results, clean date Time to Swedish/Sweden format.
   5. Note, all of this is a new change to USGS data management systems thus Alex will be learning along with Aurora how to navigate these systems and find what we’re looking for.
   6. We may move to R at this point and begin to simply pull the data into R using dataRetrieval packages.
   7. We may have to consult with our colleague Sean Payne in a scheduled meeting to get some advice. This is common when scientists become stuck. We call people!
3. Move to R – we’ve identified 3-5 sites with sufficient data availability to explore regression analysis.
   1. R & Rstudio will be used.
4. Build R code for data retrieval (dataRetrieval Package)
5. Build R code for graphical exploration of data
   1. Range of values of fCHL – known as a frequency duration curve
   2. Range of samples of chl-a with a concurrent fCHL reading from a sensor.
   3. Subset data into calibration dataset (pairs of fCHL and chl-a synced in time at the same site).
   4. Graph calibration data x vs y; explore in log space.
   5. Define log space mathematically (may need to read up on it together)
   6. Refer to statistical texts including Ryberg and others, 2020. [pubs.usgs.gov/tm/04/a03/tm4a3.pdf](https://pubs.usgs.gov/tm/04/a03/tm4a3.pdf)
   7. Also: [Guidelines and Procedures for Computing Time-Series Suspended-Sediment Concentrations and Loads from In-Stream Turbidity-Sensor and Streamflow Data (usgs.gov)](https://pubs.usgs.gov/tm/tm3c4/pdf/TM3C4.pdf)
6. Build R code for regression model of viable calibration dataset(s).
7. Review model diagnostics.
8. Apply built R code to additional site(s).
9. Assess coefficient and discuss whether a regional coefficient could be considered.
10. To this point, all this work could be done on a non-USGS computer.
11. Engage in discussion with USGS scientists on an approach to publish such models. This meeting is being planned.
12. If needed, write an outline of a potential USGS Open File Report describing the modeling approach. Consult with Amanda Booth of the USGS Florida Caribbean Water Science Center.
13. Produce a model archive summary using a USGS-internal R-shiny app – only works when we meet at Moda Tower on a USGS computer with that network (let’s not ask why… firewalls). [kswsc.cr.usgs.gov](https://kswsc.cr.usgs.gov:3838/peslick/ModelArchiveSummary/) is the website that we would use. Use the output to generate figures in a report. Alternatively, generate these figures using R code.