**PROBLEM STATEMENT AND IMPLICATIONS**:

*Problem*. Plankton tow sampling for dreissenid mussel veligers is the primary early detection method used in the West because there is a greater probability of detecting veligers as opposed to adults {Carmon, 2014 #1253}. Multiple studies have evaluated how to optimize plankton tow sample analyses (e.g., preservation methods, cross-polarized light microscopy vs. FLOWcam), but for these advances to be realized, plankton tow sampling effort must be high enough to actually encounter and capture a veliger. It is generally recognized that the plankton tow-veliger encounter rate is low since only a small proportion of a waterbody is sampled. In the Pacific Northwest during 2014, 85% of sampled waterbodies had fewer than 10 plankton tows per waterbody and 64% had fewer than three plankton tows. Surprisingly, there has been no direct evaluation of the plankton tow sampling effort required for high probability of early detection. Indirect assessments of other planktonic species, in waters were dreissenids are absent, have concluded that the level of effort expended is not high enough to ensure high probability detection {Counihan, 2017 #1137}. An understanding of the plankton tow effort needed to detect dreissenid veligers when they are rare will allow managers to allocate resources effectively, and when resources are limited, allocate resources strategically by risk.

*How proposed study will address this problem*. We will apply a “single-species, multiple-season” occupancy model framework (*sensu {MacKenzie, 2006 #1000}*) to estimate the plankton tow sampling intensity required for high probability of dreissenid veliger detection. To fit these models, we will use historical plankton tow data provided by federal and state agencies from waterbodies in the West where dreissenids are present. These historical data will allow us to assess how plankton tow detection probabilities are related to a wide range of geographical and physio-chemical covariates (e.g., waterbody size, season), and therefore provide guidance for the sampling effort required in waterbodies where mussels have not yet been detected. Based on these identified relationships, we will then develop an easy-to-use web interface (e.g., Shiny app in R programming language) for conducting plankton tow sample power analyses.

*Geographic area*. Infested and non-infested waters in the West.

**OBJECTIVES:**

*Goal*. Estimate the plankton tow sampling effort needed for high probability detection of dreissenid mussel veligers in the West.

*Objectives*.

1. Collate historical plankton tow data into a common database.
2. Use the ‘unmarked’ statistical package in R to evaluate plankton tow database.
   1. Estimate plankton tow veliger detection probabilities and how detection is associated with covariates (e.g., waterbody size).
3. Create user-friendly, repeatable and reproducible scripts that allow for others to redo our occupancy analyses.
4. Use model output to statistically simulate to determine the effort required for high confidence (e.g., 95%) of detection.
5. Develop easy-to-use web interface (e.g., Shiny app in R programming language) for conducting plankton tow sample power analyses.

**METHODS AND STUDY AREA:**

*Database development*

We will collaborate with the Bureau of Reclamation (BOR) and state of Utah to collate historical plankton tow sample data and metadata. These data will extend across a wide geography in the West, but we will focus efforts on collating data on waterbodies where dreissenid mussels have been documented. Data will be QA/QC’d. We will work with partners to identify their ideal platform for this database (e.g., postgreSQL, MS Access) and their preference for where this database should live (e.g., cloud, USGS ScienceBase, local server).

*Occupancy models*

Occupancy models that account for false-negatives have become the standard for analyzing presence-absence data, since target organisms that are present but are rare or elusive can be easily missed {MacKenzie, 2006 #1000}. We propose to adapt the single-species, multiple-season occupancy model framework that is commonly used in wildlife biology and which PI Sepulveda has used on invasive American bullfrogs {Sepulveda, 2018 #1160}. These models use information from repeated observations within a primary period (e.g., year) to estimate the probability that a site is occupied (i.e., occupancy) by the target species (e.g., dreissenid veliger), given the potential for false-negatives. This modeling approach is especially powerful because it can provide insight on covariates (e.g., waterbody size) that are associated with occupancy and detection probabilities.

For most historical plankton data, multiple sites are sampled in a waterbody multiple times per year, for multiple years. This hierarchical sampling structure (sites within sampling events, sampling events within years) is similar to the structure of multi-season occupancy models, so should be easily adaptable. Metadata associated with the plankton tow data will allow us to assess how detection and occupancy probabilities vary as functions of sampling protocols (e.g., horizontal vs vertical tows) and physio-chemical data (if not associated with metadata, than it is easily retrievable from the [water quality portal](https://www.waterqualitydata.us/) web service). We will provide all analyses scripts via an R Markdown file format as part of the final report.

Derived estimates of occupancywill be used to estimate the number of water samples required to have a 95% probability of detecting a veliger, conditional that veligers are present at the site {Sepulveda, 2019 #1254}. To do this, we will calculate θ\* = , which denotes the cumulative probability of occurrence of veligers in *n* samples taken from a location where veligers are known to occur. Then, we will run post-hoc power analyses to evaluate how sample size (i.e., the number of water samples) influences the precision of estimates. We will use estimates from the most supported covariate models to simulate the veliger detection data. For these simulations, we will vary the number of samplescollected at each sampling event. We will then replicate this process 100 times and assessed whether the 95% credibility interval for each parameter captures the value that generated the data. We will also record the width of the credibility intervals. The sample sizes at which the proportion of the 95% credibility intervals that contain the original parameters stabilize and the average width at which of the credibility intervals stabilize provide insight about the point of diminishing returns, beyond which increasing sample size provides little benefit. These types of post-hoc power analyses will be developed into a Shiny app in the R programming language, which will provide a simple point and click web interface for managers.