

Proposal title: AFS Database Shiny App

Erin Tracy, University of California Davis, etracy1121@gmail.com (or eetracy@ucdavis.edu) (former project lead)

Scott Bonar, University of Arizona, SBonar@ag.arizona.edu, (project lead)

Chad Teal, University of Arizona, chadteal@arizona.edu, (project lead, point of contact)

Brief project summary

Standardization of sampling methods allows fisheries professionals to compare data across large spatial and temporal scales, encourages data sharing, and improves communication. This ultimately enhances their ability to address large-scale threats facing freshwater ecosystems and the fishes they support. The online database allows for the comparison of fish metrics commonly used in management to assess population health, including growth, condition, length-frequency, and catch per unit effort data collected using AFS standard methods.

Approach*Data Acquisition and Compilation*

Data sets were requested from federal, state, and provincial agencies throughout North America. A data set was defined as data from a completed survey of a water body using a standardized gear type conducted once during a season. In some cases, data sets included multi-day surveys of the same waterbody (e.g. large reservoirs) where the effort was summed across all sampling days. We requested 2010-2020 data on any fish species collected with AFS standard gears and methods during routine monitoring programs of an entire fish community in a water body or entire population of specific species, excluding surveys targeting specific size groups or those with other biases (e.g. egg counts, juvenile fish surveys). Following data best practices, both metadata and raw data were requested including contact information, waterbody name and coordinates, type of waterbody, dates of sampling, sampling method, transect or net number if applicable, species name, total length, weight, effort, age, and any additional information necessary to accurately interpret the data. When latitude and longitude were provided, one of the fifteen ecological regions of North America was assigned.

Data analysis and webtool development in R

Data sets went through a quality assurance and control (QA/QC) process which entailed adding missing metadata when possible, formatting to meet the needs of the project (e.g. standard column names, variable names, species names, units, etc.) and performing outlier analysis (removed values outside of 3 interquartile ranges of the median). This process resulted in the exclusion of data that did not meet standard sampling recommendations or was missing essential pieces of information necessary for analysis (e.g. fall electrofishing, non-standard nets, missing metadata). Analysis to produce the averages was conducted in R (R core team, 2021). The code for the data cleaning and analysis is publicly available on GitHub to ensure the science is reproducible (https://github.com/erinetracy/AFS_database_code). An interactive webtool to display those data was also created with R Shiny (Winston et al. 2021) (Figure 15.1). This tool allows users to view, filter, and

download averages at the continent, ecoregional or state level. The code to develop the Shiny decision support tool is also publicly available on Github and can be found at this shiny app site https://erintracy.shinyapps.io/afs_shiny_app/

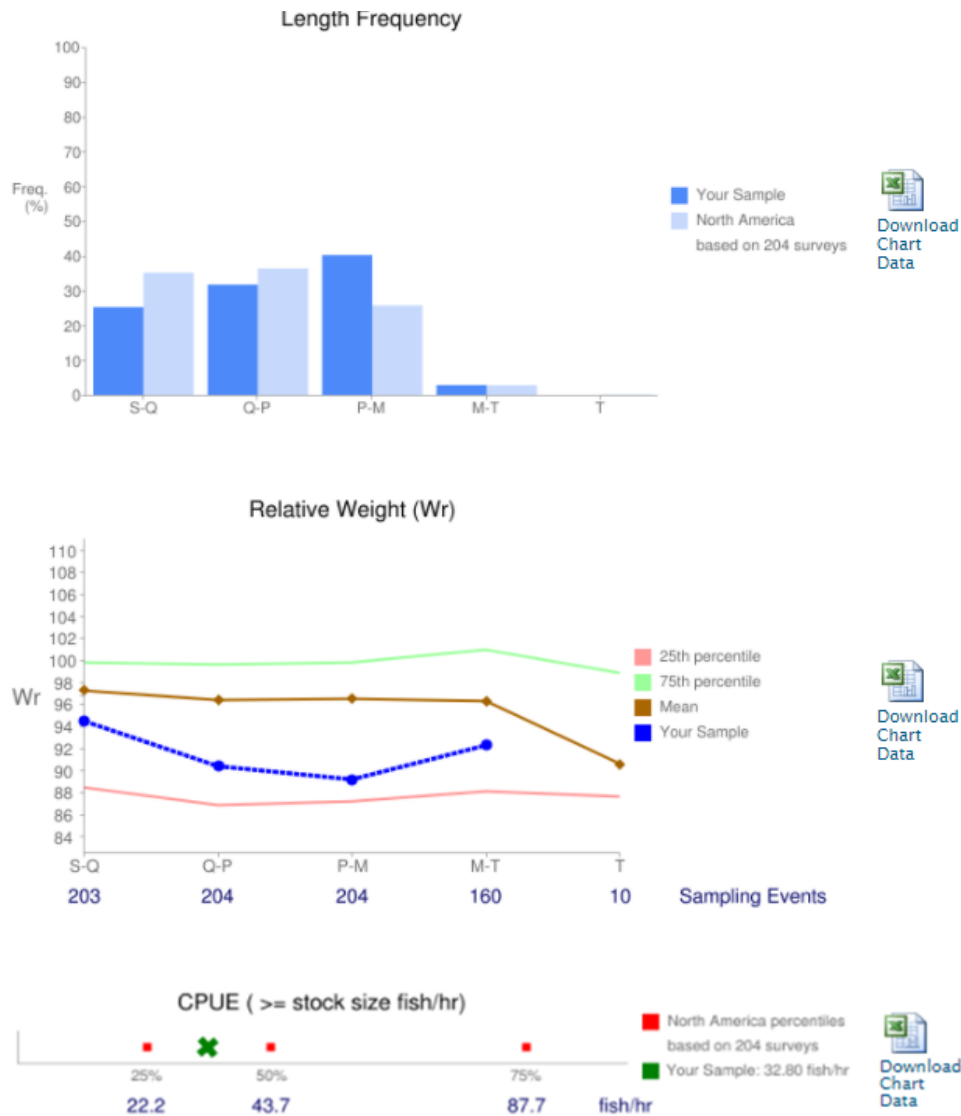
For each species we calculated the following summary data; (1) mean CPUE of stock-length (Gabelhouse 1984) fish, standard error (SE), and 5th, 25th, 50th, 75th, and 95th percentiles; (2) mean relative weight (Wr), SE, and 5th, 25th, 50th, 75th, and 95th percentiles for each proportional size distribution category; and (3) length frequency as mean percentage and SE of each proportional size distribution length. Mean back-calculated total length at age, SE and 5th, 25th, 50th, 75th, and 95th percentiles were obtained from the previous edition due to lack of age data submitted. Though we received data on over 300 fish species, we only included species with proportional size distribution length categories and standard weight equations in the book tables. Trout species were separated into lentic and lotic due to the different standard length and weight equations that exist for these groups. The sample size varies considerably between species, waterbody, sampling method combination. Therefore, users are encouraged to consider effects of small sample size (N) when using some of these data.

Expected outcomes

Scientific products

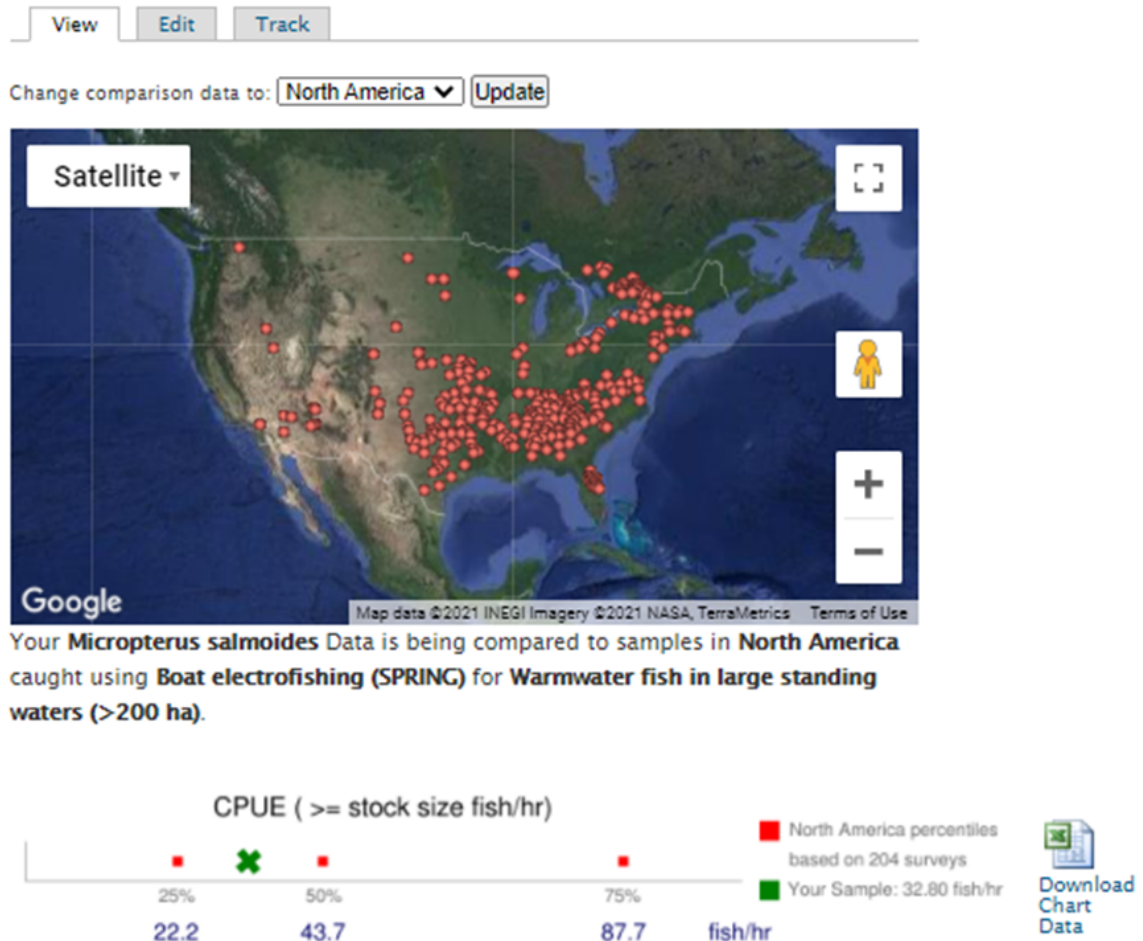
The existing Shiny app will be updated with the following features and migrated to deployment with RStudio Connect.

1. Ability to upload user data and calculate averages including mean **CPUE** of stock-length (Gabelhouse 1984) fish, standard error (SE), and 5th, 25th, 50th, 75th, and 95th percentiles; (2) mean **relative weight** (Wr), SE, and 5th, 25th, 50th, 75th, and 95th percentiles for each proportional size distribution category; and (3) **length frequency** as mean percentage and SE of each proportional size distribution length. The code for data analysis to calculate these metrics can be found here (https://github.com/erintracy/AFS_database_code), however I can't get it to work in the Shiny app code. The code for processing user-contributed data will be refactored using 'dplyr' and tested for compatibility. For user upload to be compatible, project leads may need to codify and communicate the necessary data standards/formats to potential users. A downloadable data template will be provided on Shiny app. Users will upload completed data template and Shiny app will calculate metrics (CPUE, relative weight, and length frequency). Metrics will be displayed per fish species, sampling method, waterbody type, and scale.
2. Plots of the user metrics calculated in step will be displayed as bar graphs or line plots (examples below) and include sample size and standard error bars.



- Development of a map of sample sites to include with pre-existing tables of data in “Database Averages”. Purpose is to display locations of sites in standardized dataset and will not include any user-input data. This does not exist in app yet.

Summary for *Micropterus salmoides*



4. Move app deployment to RStudio Connect (under UA team account)
5. Code improvements: streamline and update analysis code (currently in base R for loops) to generate figures for app with improved performance

Skills development

Erin Tracy will be involved in the development of the Shiny app in the form of biweekly meetings with the UA team to build skills such as data analysis in R, displaying results through plots, graphs, and maps, and user friendly web tool design. When Erin is unavailable, Scott Bonar and Chad Teal, current UA employees, will take over being the primary points of contact for Shiny app development and skills transfer.

Plan for project sustainability

There is a budget to fund CCT Data Science to maintain the app. This would include only a few hours a year keeping the app up-to-date with security patches. Additional work by CCT Data Science, such as

adding features, updating data, or fixing bugs, will require a separate scope-of-work and funding agreement. However, if a Bonar lab member wishes to undertake the additional work, CCT Data Science can provide advice pro bono, e.g., during drop-in hours or individual appointments.