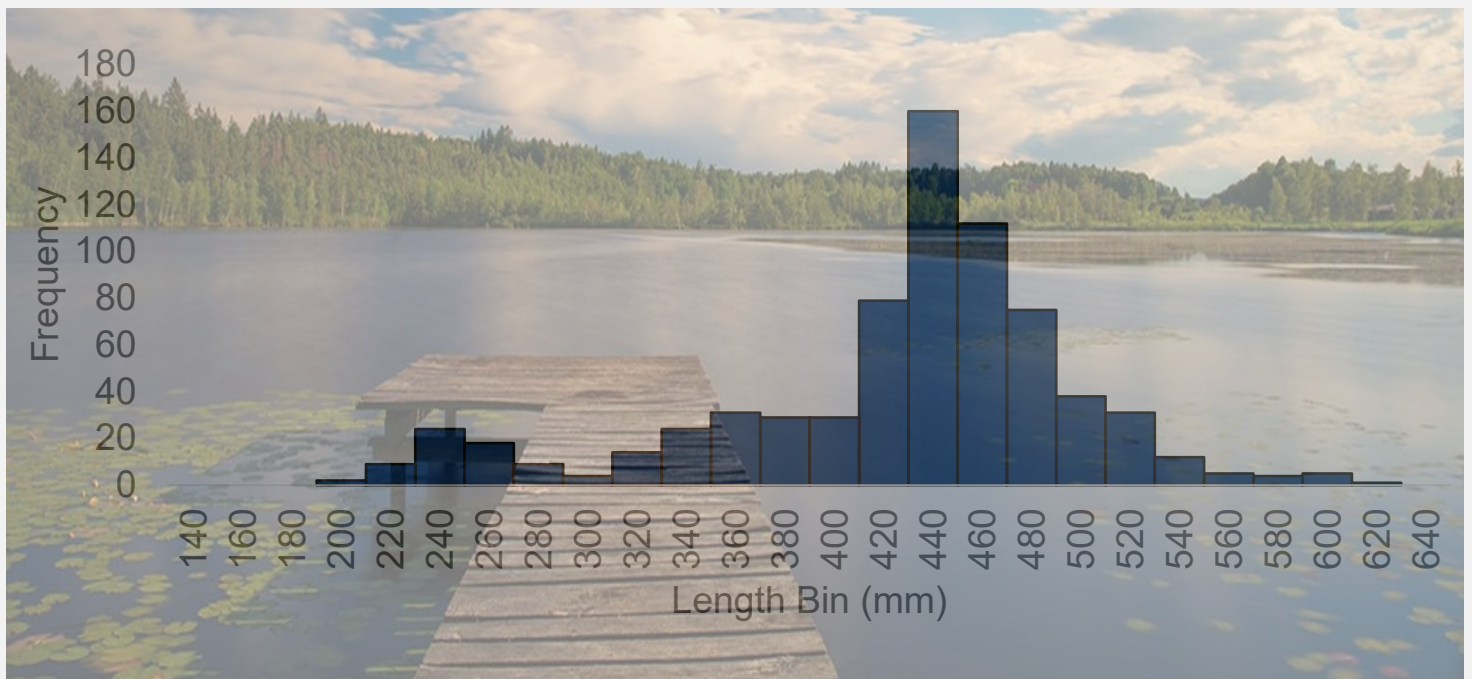# Oklahoma Fishery Analysis Application:

An R-Based Application for Analysis of Standardized Fishery Samples

Oklahoma Department of Wildlife Conservation

Users Guide

Dray D. Carl and Daniel E. Shoup, Oklahoma State University

OFAA v2.0

Manual updated 4/25/2022

Table of Contents

Overview

Our goal in creating the Oklahoma Fishery Analysis Application (hereafter referred to as OFAA or the app) was to develop an application to analyze standardized fishery samples of the Fisheries Division within the Oklahoma Department of Wildlife Conservation (ODWC). Standardization of fishery sampling is essential (see Bonar et al. 2009 for examples), and analysis of samples should also be considered within standardized sampling regimes. Furthermore, we wanted to create an application with an easy, user-friendly interface to give all biologists and technicians the capability to efficiently and accurately analyze fishery samples. We hope this application enables employees to easily explore historical sample data and compare fisheries across the state.

We recommend users read through this Users Guide carefully before using OFAA for analyses of fish populations. Users should be aware of the assumptions regarding each analysis and the method in which each was calculated. We refer users to two textbooks for more detailed explanations of methods used in OFAA: Analysis and Interpretation of Freshwater Fisheries Data (Guy and Brown 2007), Introductory Fisheries Analyses in R (Ogle 2016).

 is an open source programming language and software environment for statistical computing (R Core Team 2017).

Shiny (Chang et al. 2017) is an interactive web application by RStudio and is the user interface for OFAA.

Shiny application and R code written by *Dray D. Carl* and *Daniel E. Shoup*.

Acknowledgements

Using and Navigating OFAA

Online Server and Login

1. The R code governing OFAA is hosted on an online server (Shinyapps.io) paid for by the ODWC. This allows for routine updates to the code and additions to the database in one convenient location, so all users will automatically have the most updated versions of OFAA and the database. Copies of the R code or database can be obtained for personal exploration or use from Dr. Dan Shoup (Oklahoma State University) or Kurt Kuklinski (ODWC Research Supervisor).
2. Once invited to join the Shinyapps.io server, the user will create an account with a username and password. Once an account has been created, the user can access OFAA by simply typing in the web address and logging in.
3. This setup also allows the user to access OFAA through any device (e.g., desktop computers, home laptops, tablets, or even mobile devices). The user interface was developed for typical computer monitor sizes, so the application will look differently on mobile devices. However, the app is still fully functional and may provide employees with useful information while in the field.
4. We recommend users work in Google Chrome while using OFAA. This was the browser the app was created in, so graphics may be slightly different in other web browsers.

Select Sample Tab

1. This is the first tab of OFAA and is automatically opened when the app is loaded.
2. The first step is to select a dataset to used based on how much historic data needs to be loaded. Loading just recent data will considerably speed the apps performance. Wait for the "Uploading Data…" (red letters) to disappear before continuing; the app is loading the entire database into memory. If the user later wants to use a different range of years, click on the "change the data being used" button to start the database selection process again.
3. This tab is used for filtering sample data from the entire database. This database will appear as an interactive table below the main selection boxes once it is uploaded and will automatically update as the user fills in the selection boxes. The user can explore the database or selected data with this table (interactive directory buttons on bottom right; sample size on bottom left).
4. Use the boxes provided to select a lake, year, and gear as a minimum to filter the dataset to the sample desired. By deafault, lake and gear are sorted by name (but codes are also present and can be selected by typing these into the box). If you prefer a version sorted by codes rather than names, check the box at the top of this pannel. References to lake and gear codes/names can be found in the

lakinfo and gearinfo .csv's, respectively or in the SSP Manual. The user can type into selection boxes (the options will filter automatically), select options with either a click or the Enter key, and remove selections using the Delete key.

5. Multiple selections are allowed within each selection box (Lake, Year, Month, and Gear). However, the user should be aware of the implications of selecting multiple lakes, years, and gears.

   a. For example, including multiple gears would violate assumptions of CPUE calculations, but may be beneficial for adding to a sample size for length-frequency histograms or length-weight regressions.

6. A minimum of Lake, Year, and Gear fields should be selected before advancing to the next tab.

7. When the user selects criteria in one box, the other boxes will automatically filter to show only available values for these other fields.

8. The user can directly download a .csv of the selected sample dataset for purposes other than use in OFAA using the Download Selected Sample Data button.

9. Alternatively, the user can import a .csv dataset into the app for OFAA to use. This dataset must have gone through the Data Validation App before it is compatible with OFAA. Click the browse button to locate and import the file into the app, and check the box below to use these data instead of data from within the database.

Select Analyses Tab

1. The second tab of OFAA has multiple purposes. First, the user can select which analyses are outputted in the analysis tabs with checkbox inputs. Second, the user can select a single species from the species present in the sample data selected in the Select Sample Tab. Third, the user can match an age dataset to the sample data for calculating population dynamics (e.g., growth, mortality).

2. Catch Analyses Metrics Panel

   a. This panel refers to the third tab, Catch Analyses. The user can select, with checkboxes, which analyses OFAA should calculate when the user navigates to the Catch Analyses tab.

      i. The Catch Analyses tab uses the entire sample dataset (i.e., all species), so the user is not required to select a species before proceeding to this tab.

3. Selected data summaries

   a. Summaries will appear on the left side of the screen for both the selected sample data and selected age data for the benefit of the user.

4. Single Species Analyses Selection Panel

a. This panel refers to the fourth tab, Single Species Analyses.
b. Select species by code or by name
  i. All single species analyses require a species to be selected using the one of the selection boxes, the other (code or name) will populate automatically. Only species codes that are present in the selected sample data will appear as possible selections in these selection boxes.
  ii. The user can download a .csv of the selected data after it has been filtered by the single species for purposes other than OFAA with the Download Selected Species Data button.
    1. However, the user should be aware not to use the Number.of.individuals field in this .csv (see Assumptions and Methods of Calculations – Single Species Analyses Tab for more details).
c. Select Age Dataset
  i. After the user has selected a single species to explore, OFAA will run an algorithm to automatically pair the associated age dataset with the sample data. If a matched age dataset is available and paired (same species, lake, and year), a message will appear "Matched Age Dataset" in green.
  ii. If a matched age dataset is not available, the user can manually choose to select a similar lake or year to "borrow" age data. We recommend not "borrowing" age data more than 2 years before or after the sample data within the same lake. "Borrowing" age data from a different lake is up to the user's discretion (close systems with similar growth). We recommend only using this ability for data exploration (i.e., not for management decisions) and advise careful interpretation of analyses. If the selected age dataset is not a perfect match with the selected sample data, a message "Not a Matched Age Dataset" will appear in red to remind the user to carefully interpret analyses.
  iii. Whether matched or not, once an age dataset is selected the sample size of the age dataset (N = X) will appear.
  iv. All three selection boxes (species, lake, year) are required for selection of an age dataset. Multiple selections within each are accepted.
    1. For example, if age data from 2 consecutive years are available, the user may combine both years of age data to bolster sample size.

v. Unlike the Select Sample Tab selection boxes, these selection boxes are interactive such that only lakes that have age data for the selected species will appear as possible selections, and only years that have age data for the selected species-lake combination will appear as possible selections.

vi. After all selection boxes are filled, the selected age dataset will appear at the bottom of the page in table format. The user can explore these data in the same way as the previous tab and can also download the selected age dataset with the Download Selected Age Data button.

5. Single Species Analyses Metrics Panel

a. Similar to the other panel, the user can select (with checkboxes) which analyses OFAA should calculate when the user navigates to the Single Species Analyses Output tab. These checkboxes are split into two groups (Length and Weight Metrics, Population Dynamics).

b. Alternatively, the user can upload age data in the manner as sample data on the previous tab. Again, the .csv uploaded must have gone through the Data Validation App and be in the proper format.

Common Abbreviations within OFAA

Table 1. Common abbreviations and method of calculation within OFAA.

Abbreviation	Full Meaning	Calculation
Count	Sample Size	
RSE	Relative Standard Error	$\left(\dfrac{SE}{Mean}\right) * 100$
SE	Standard Error	$\dfrac{Std.Dev.}{\sqrt{Count}}$
L 95% CI	Lower 95 Percent Confidence Interval Bound	$Mean - (1.96 * SE)$
U 95% CI	Upper 95 Percent Confidence Interval Bound	$Mean + (1.96 * SE)$
CV	Coefficient of Variation	$\dfrac{Std.Dev.}{Mean}$
N RSE(25)	# samples needed to get RSE= 25 for CPUE (i.e., 50% overall change in CPUE can be detected)	$\left(\dfrac{CV}{25}\right)^2$
N RSE(40)	# samples needed to get RSE= 40 for CPUE (i.e., 80% overall change in CPUE can be detected)	$\left(\dfrac{CV}{40}\right)^2$

Catch Analyses Tab

1. The third tab of OFAA begins calculations of catch data. This tab uses all species from the selected sample dataset.
2. Sampling Event Summary Panel
   a. The summary (lake, year, month, and gear) of the Selected Sample Data that the tab is using for analyses.
3. Abiotic Metrics Panel
   a. Mean, Minimum, and Maximum values for Pool Elevation (feet), Surface Temperature (Celsius), Secchi Depth (inches), and Conductivity (microsiemens per centimeter).
4. Sample Size and Effort Panel
   a. Total sample sizes (number of fish) for each species in the selected dataset.
   b. Total amount of effort for each gear type in the selected sample dataset. Units: electrofishing (total minutes fished), trap nets and gillnets (total hours fished), seine (total ft^2 fished).
5. Catch Per Unit Effort Panel
   a. Total CPUE
      i. Mean total CPUE for each species present in the selected sample dataset along with sample size and variance metrics (Table 1).
   b. CPUE by Size Category
      i. Size categories are based on PSD length categories (Gabelhouse 1984) for each species that currently have PSD size categories.
      ii. References are provided for these size categories in both millimeters and inches.
      iii. Mean CPUE for each species within each size category that is present in the selected sample data along with sample size and variance metrics (Table 1).
   c. Both Total CPUE and CPUE by Size Category tables are available for downloading to a .csv using the associated download buttons at the bottom of this panel.

Single Species Analyses Tab

1. The fourth tab of OFAA begins calculations of single species metrics. This tab uses the selected species data.
2. The top-left panel provides a sample summary (lake, year, gear, and species name).
3. Length and Weight Metrics Panel
   a. This panel requires a species to be selected in the Select Analyses Tab.
   b. Length-Frequency Histogram
      i. A length-frequency histogram is provided with the default unit of measure as millimeters. The user can change the length bin increment with the numeric input box. The user can also change the displayed unit of measurement to inches by typing 25.4 into the numeric input box. Y-axis is formatted to proportion for comparisons among samples.
      ii. The user can download a .png of the Length-Frequency Plot or a .csv of the length frequency data using the associated download buttons below the plot.
   c. Proportional Size Distribution
      i. Length references for PSD size categories of the species being analyzed are provided in both millimeters and inches. Both traditional and incremental PSD values are provided (Guy et al. 2007), along with lower and upper 95% confidence interval bounds (see Assumptions and Methods of Calculations for more details).
   d. Relative Weight (W_r)
      i. Provided is detailed information about the standard weight equation used to calculate relative weights for the given species. This information will help the user interpret relative weight results (Reference Percentile and Minimum Total Length) or enable the user to use the standard weight equation away from OFAA (slope and intercept values).
      ii. OFAA calculates overall mean relative weights and mean relative weights grouped into PSD size categories using standard weight equations (most can be found in Neumann et al. 2012). Sample size and variance metrics are also provided.
   e. Length-Weight Regression
      i. A plot is constructed using all length-weight pairings from the selected species data relating log-transformed Total Length (mm) and log-transformed Weight (g). The coefficients for the intercept and a' are provided beneath the plot. Users can use these

coefficients to predict fish weights from known total lengths (Neumann et al. 2012).

- f. Max Length and Weight
    - i. Maximum length (mm and in) and weight (g and lbs) of the species from the selected sample data.
- g. Length-Weight Downloads
    - i. Users can download .png's of plots or .csv's of tables for use in lake reports or other purposes by clicking the associated buttons. Downloadable content includes the PSD Table, Relative Weight Table, and Length-Weight Regression Plot.
4. Population Dynamics Panel
    - a. This panel requires a paired age dataset to calculate associated metrics.
    - b. Selected Age-Length Key
        - i. A bubble plot is constructed for the observed age-length key from the selected age dataset. This is provided so the user can visualize the selected age dataset and any associated biases. The area of each circle reflects the proportion of fish in a length interval that are at a given age.
        - ii. The user can download both the observed age-length key or the smooth age-length key (calculated using multinomial logistic regression) in table format (.csv) as well as the bubble plot of the observed age-length key using provided download buttons underneath the plot. More details about how these two methods of creating age-length keys are provided in Assumptions and Methods of Calculations.
    - c. All metrics from this point forward are based on the entire sample data aged using a smoothed (not observed) age-length key.
    - d. Age Frequency Histogram
        - i. A frequency histogram of all predicted ages in the selected species dataset. Y-axis is formatted as the proportion of the total sample size.
    - e. Growth Metrics Plot and Associated Tables
        - i. The growth metrics plot and associated tables are displayed by default in metric units, but can be displayed in inches and pounds by checking the Display Measurements in English Units checkbox.
        - ii. The growth metrics plot displays each individual fish as an open dot and mean lengths-at-ages by solid triangles. The von Bertalanffy growth curve is also plotted as a solid line. Beneath the plot are tables with mean lengths-at-ages and weights-at-ages along with sample size and variance metrics.

      iii. The age and length data are fit with the von Bertalanffy growth function (Isely and Grabowski 2007). The von Bertalanffy equation table below the plot shows estimates for the three von Bert coefficients (L-infinity, K and t_0) along with upper and lower 95% confidence interval bounds for the fitted model (more details in Assumptions and Methods of Calculations).

  f. Catch Curve (Mortality)

      i. A catch curve is plotted for the user with ln(abundance) on the y-axis and age on the x-axis (Miranda and Bettoli 2007). The descending limb is differentiated with filled circles (open circles for ascending limb), and a linear function is fit to the descending limb. Instantaneous mortality (Z) and annualized mortality (A) are provided in the table below the plot with upper and lower 95% confidence interval bounds (more details in Assumptions and Methods of Calculations).

      ii. Theoretical maximum age is calculated as the X intercept from the catch curve (i.e., the age at which the catch curve line has zero for log abundance). Observed maximum age is simply the oldest fish in the selected age dataset.

  g. Estimates of natural mortality

      i. Natural mortality is usually calculated by measuring fishing mortality (creel survey, reward tag study, etc.) and then subtracting fishing mortality out of the total mortality…the remainder is natural mortality. However, a large body of literature shows natural mortality often correlates with life history attributes (maximum age, max length, max weight, the K parameter of the von Bertalaffy curve, etc.) This table provides two such estimates that were recommended by Then et al. (2015) who reviewed many such estimation methods. Users should appreciate that these approaches are only estimates based on what works for many types of species and any given species or even individual population's natural mortality may not be accurately predicted using these methods.

      ii. Hoenig NLS method – this approach uses theoretical maximum age and is the best such metric for most fish populations according to Then et al. (2015) (but note natural mortality any given population may not work well using this or any other method).

      iii. Pauly NLS -T method – this approach was favored by Then et al. (2015) in cases where theoretical maximum age was unknown and

the Hoenig NLS method could not be used. It is based on the von Bertalanffy parameters.

    h. Population Dynamics Downloads

        i. Users can download .png's of plots or .csv's of tables for use in lake reports or other purposes by clicking the associated buttons.

Statewide Percentiles Tab

1. Percentiles provide a useful way to see how a metric from a given sample compares to common values of that metric from other samples in the database. The Xth percentile is the point at which X percent of the values in the data base are below the given point. For example, if the 75th percentile for CPUE is 100 fish/hr, it means that for this given species and gear, 75% of the recorded CPUE values are less than 100 fish/hr and only 25% (100% - 75%) have this value or larger.

2. Some cautionary thoughts. First, percentiles will not be particularly useful if the number of surveys is small (i.e., with 5 surveys, each survey constitutes 1/5 or 20% of all values…so you will not have resolution to accurately estimate percentiles that differ by 5%. Second, poor quality data can set the maximum and minimum values for a metric at unreasonable sizes (catch rates of thousands of fish per hour, maximum lengths that are an order of magnitude larger than possible), and this can strongly effect the resulting percentiles…especially when you get near the 1% or 99% percentiles.

3. Select survey criteria panel

    a. Select at least a gear type to see results. Selecting more than one gear type will allow you to see some metrics that make sense from combined gears (Max length and weight, Wr) but will remove other results where the calculation would be confounded by multiple gears (CPUE, PSD, length/weight at age or mortality)

4. Users can optionally select specific species to simplify the data that result (all species are used by default).

5. Users can also restrict results to a specific range of years, management region(s) or lake(s) using the boxes provided.

6. If desired, users can specify custom percentiles to present. By default, 5%, 25%, 50%, 75%, and 95% are used.

7. Percentiles will not be meaningful if there are too few records included in the calculation. However, including numerous samples with poor quality data (e.g., only 1 or 2 fish caught such that Wr, PSD, etc. is likely skewed by the small number of cases) will heavily skew percentiles and could overwhelm the better data available. By default, the following limitations are in place to restrict

percentile calculation only to metrics with sufficient data quantity. These defaults can be overridden with sliders in the left panel of this tab.

    a. Minimum number of survey required. A survey is composed of all samples using the same lake, year, and gear. The default is 10 surveys. This value is used for all percentile metrics

    b. Minimum # of stock-size fish required is by default 40. This limit is only used with respect to the PSD percentile table. Note a simulation analysis by Gustafson (1988) found PSD precision of $\pm$ 15 or so PSD units requires around 45-55 stock-length fish in the analysis. Obtaining precision of $\pm$ 10 PSD units requires 80-120 sock-length fish. The default of 40 is therefore only providing a precision of 14-19 PSD units and should be considered a minimum acceptable level.

    c. The minimum number of Wr (relative weight) values required is 5 by default. This is only applied to the Wr percentile table.

    d. Minimum number of fish aged is by default 150. Coggins et al. (2013) performed simulations and found reasonable sample sizes require about 300 aged fish with 500-1,000 total fish (i.e., unaged fish to which the key is applied) in the analysis for length-at-age analyses. This value effects length at age and mortality percentile tables.

    e. Minimum number of fish required in an age class before using it for mean length/weight calculations is 5 by default. This value effects length at age and mortality percentile tables.

    f. For catch curve inclusion, a catch-curve R^2 value greater than 0.80 is required by default. This is done because the slope of the line likely no longer represents the mortality trajectory if the data are any messier than this (it is a sign that catch curve assumptions are being violated). Often, quality catch curve data will have R^2 values of 0.9 – 0.98. This only effects the mortality percentile table.

8. Each percentile table will provide output by species when sufficient data exist. If one or more of the above rules protecting data quality are violated, a message indicating this is displayed instead of the results.

9. For metrics that include metric or English units, metric units are used by default and a check box is available to convert to and display the English units.

Stocking Information Tab

1. This tab can be used to explore information about what fish have been stocked in a water body. This feature may complement analyses from other tabs to help aid management decisions.

2. Stocking Information Criteria

a. Three stocking information criteria can be used to filter the stocking database: Year Range, Water Body, and Species.
b. By default, the last 10 years of stocking information is already selected.
c. The stocking information table will show up after the user has selected either a Water Body, Species, or both.
d. Multiple selections are allowed in each selection box (e.g., if the user is interested in stocking information of 2 species within the same water body).

3. Stocking Information Table
a. This table includes six fields: The date stocked, waterbody, species, number of fish stocked, size at stocking (inches), and the hatchery or origin of the fish.
b. The user is able to download the selected stocking information table into a .csv with the associated download button for use in other purposes.

Assumptions and Methods of Calculations

Catch Analyses Tab
1. Abiotic Metrics
a. Mean, maximum, and minimum values are found using the simple mean(), max(), and min() R functions.
2. Sample Size and Effort
a. Sample sizes for each species are found by counting the Number.of.individuals field
b. Total effort is the sum of all individual sampling events grouped by gear code in the selected dataset. Units: electrofishing (total minutes fished), trapnets and gillnets (total hours fished), seine (total ft^2 fished).
3. Total CPUE
a. All CPUE calculations assume only one gear code is in the selected dataset.
b. Calculation Methods
   i. SampleID is created for each unique sampling event (lake code, station, month, day, year, gear code)
   ii. Sums of Number.of.individuals for each species-SampleID combination are calculated
   iii. addZeroCatch() function (FSA package) is used to add zeros to species that were not present in a given sample (SampleID).

1. For example, if Black Crappie are present in 4 samples, White Crappie are present in 6 samples, and no fish were caught in 2 samples (species code 98), zeros will be added to 4 samples for Black Crappie and 2 samples for White Crappie.
            iv. CPUE is calculated for each individual sample
                1. Electrofishing: CPUE = (#fish*60)/Gear.Length
                2. Netting (trapnet and gillnet): CPUE = (#fish*24)/Effort
                3. Seine: CPUE = (#fish*1076)/Effort
            v. Mean CPUE is then calculated for each species in the selected sample data.
    c. Metrics of variance and their calculations can be found in Table 1.
    d. Relative Standard Error (RSE) is almost synonymous with Coefficient of Variation (CV) and can be interpreted in much the same way (i.e., detecting a % change in CPUE). Dumont and Schlechte (2004) recommend a RSE of less than 25 to monitor population CPUE's. The values of N RSE(25) and N RSE(40) represent the number of samples (e.g., electrofishing runs) required to achieve a RSE value of 25 and 40, respectively. However, be aware that Dumont and Schlechte (2004) found this method generally underestimates the number of samples required to achieve RSE's of 25.
4. CPUE by Size Category
    a. Calculated with the same methods of Total CPUE with size category as an extra grouping variable.
            i. PSD size category is added to each record (fish) using the psdAdd() function (FSA package) after joining the "Gabelhouse Name" (specific names required for psdAdd() function) to each record.
    b. Species with no existing PSD length categories will not be included in the CPUE by Size Category table. Most species' length categories can be found in (Neumann et al. 2012) or by using the psdVal() function (FSA package). If a species does not have PSD length categories, it is probably a less common species (e.g., Shortnose Gar, Spotted Sucker, etc.) or a small-bodied species (e.g., Inland Silverside, Logperch, etc.) that cannot be easily split up by size categories.
    c. "Trophy" size categories are not included in the CPUE by Size Category table.
    d. Unlike PSD indices (Guy et al. 2007) fish in the "longer" categories do not also belong in the "shorter" categories. Size categories are strictly bound

to the upper and lower bounds of the given size category (i.e., similar to incremental PSD indices).

  e. Metrics of variance and their calculations can be found in Table 1.

Single Species Analyses Tab

1. Before any functions/calculations are run on this tab, the selected species dataset is run through a rep() function. This repeats each record in the dataset according to the value in the Number.of.individuals field (i.e., making an individual record for each individual fish).

2. Length and Weight Metrics Panel

  a. Length-Frequency Histogram

    i. This plot is constructed using the hist() function. The x-axis is set by default to range from 0 to the maximum total length + 30 mm, with the breaks (i.e., length bin size) set by the user in the numeric input box (default 10 mm).

    ii. The Length-Frequency Histogram function uses the hist() function to calculate counts of each length bin, divides by the sum of all fish in the sample, and uses the hist() function again to display the histogram as proportions (enabling comparisons among other samples).

    iii. For more information on interpreting length-frequency histograms, see Neumann et al. (2012).

  b. PSD Table

    i. A PSD table with confidence intervals is produced with the psdCalc() function (FSA package).

      1. This metric assumes only one species is being analyzed.

      2. The sample data is first joined with specific species names required for the psdCalc() function (gabelhousenames.csv).

      3. Only species with established PSD size categories can be used in PSD index calculations. Most species' length categories can be found in (Neumann et al. 2012) or by using the psdVal() function (FSA package).

      4. Both traditional and incremental PSD indices are generated. Traditional indices (PSD-X) are the number of fish greater than or equal to X over the total number of stock-length fish. Incremental PSD indices (PSD X-Y) is the percentage of all stock-length fish that are between two successive Gabelhouse lengths (Neumann et al. 2012).

5. Confidence intervals are calculated using methods of Brenden et al. (2008) assuming a multinomial distribution (Ogle 2016).

c. Relative Weight Table

   i. Relative weight is the most popular measure of condition (Blackwell et al. 2000) and is calculated for each individual fish using the wrAdd() function (FSA package). OFAA provides mean relative weights split among PSD size categories along as well as an overall mean relative weight. Detailed information about the standard weight equation used for the species in question is rendered above the relative weight table. More information about these parameters can be found in Neumann and Allen (2007) or Ogle (2016).

      1. Mean relative weights will only be calculated if 1) both lengths and weights are recorded in the database, 2) PSD size categories are established for the given species, and 3) standard weight equations (for calculating relative weight) are established for the given species.

      2. The sample data is first joined with specific species names required for the psdCalc() function (gabelhousenames.csv) as well as specific species names required for the wrAdd() function (WSnames.csv). Similar to PSD size categories, standard weight equations are species-specific and most can be found in Neumann et al. (2012) or by using the wsVal() function (FSA package).

      3. Most standard weight equations are developed for the 75th percentile of mean weights (see reference for details of standard weight equation); therefore, the user should interpret a relative weight value of 100 as having condition comparable to the 75th percentile.

      4. Metrics of variance and their calculations can be found in Table 1.

d. Length-Weight Regression

   i. Only fish with recorded lengths and weights are used in the length-weight regression

   ii. The length-weight model is a simple linear model [lm()] predicting log10(weight [g]) from log10(total length [mm]). Individual fish are plotted as circles, and a solid line of the fitted linear model is included on the plot using the lines() function.

      iii.   Coefficients of the lm() are provided beneath the plot (a' and intercept). More information about using linear models can be found in Neumann et al. (2012).

   e. Max Length and Weight

      i.   Maximum lengths and weights are found using the max() function. These values (mm and g) are then converted to in and lbs.

3.   Population Dynamics Panel

   a. Age-Length Keys

      i.   The app uses an age-length key technique to apply ages to every fish in the sample. This preferred technique has many benefits including 1) greater sample size when applied to entire sample, 2) allows for standardized collection of a subsample of otoliths for aging (i.e., sacrificing less fish), and 3) eliminates bias of otolith subsampling for mean lengths-at-ages.

      ii.   An age-length key provides the probability of a fish being a certain age for each length grouping incorporated in the key. The app calculates two types of age-length keys: an observed age-length key and a smoothed age-length key. The observed age-length key is exactly reflective of the age-length proportions from the otolith subsample. The smoothed age-length key uses a statistical model to predict proportions for every age-length combination (within pre-determined bounds) using data from the subsample of ages and fish lengths. For more information about these methods see Ogle (2016).

      iii.   The observed age-length key is limited because it cannot predict ages of sampled fish not represented in the key. Additionally, highly variable ages and small samples sizes in some length groups may result in pieces of the key that are unreliable (Ogle 2016).

      iv.   The smoothed age-length key is beneficial because it provides the ability to predict the ages of fish that fall in length categories not represented in the aged fish selected data and "smooths" the effects of outliers. However, this key is still limited to the bounds of the selected age data (upper and lower bounds of both total length and age) and will not age fish above or below the lower and upper bounds of total length observed in the selected age data. This key also provides the option to "borrow" an age-length key (different lake or year) to age sample data without associated age data. Though we encourage exploration of data by borrowing age-length keys in cases where age data are lacking, we caution interpretation

of these results and do not recommend using these results for anything other than data exploration (i.e., no management actions).

  v. For these reasons provided above, the observed age-length key is only used for the visual plot provided at the top of the Population Dynamics Panel, and the smoothed age-length key is used for all remaining metrics calculated in the Population Dynamics Panel.

  vi. Observed Age-Length Key
    1. First, the width of the length grouping is determined (either 10, 15, or 20 mm groupings). This is determined by dividing the maximum length observed in the aged sample by 30. If the result (w) is $\geq$ 20, 20 mm length groupings are used, if $15 \geq w > 20$, 15 mm length groupings are used, and if w<15, 10 mm length groupings are used. These length grouping recommendations are from (Coggins et al. 2013).
    2. Length groupings are then assigned to each aged fish using the lencat() function (FSA package), and a frequency table is built with xtabs() and prop.table() functions (base R).

  vii. Smoothed Age-Length Key
    1. This key uses the same methods to generate length groupings as the observed age-length key (Coggins et al. 2013). Each aged fish is assigned a length grouping for use in a multinomial logistic regression model. A model is constructed using the multinom() function (nnet package) to predict a fish's age by length category (maximum iterations = 500).
    2. The app then uses the model to predict a probability for each age-length category combination within the bounds of the aged sample using the predict() function (base R).

  viii. Applying the Smoothed Age-Length Key to the Sample Data
    1. Each fish (within the bounds of the age-length key) is assigned a length category in the same fashion as the associated age-length key.
    2. Each fish from the selected species sample data is assigned an age based on the smoothed age-length key model. This is done using the alkIndivAge() function (FSA package). This function handles the problem of "fractionality" (Isermann and Knight 2005) and random ages are assigned to unaged fish according to the method produced by Isermann and Knight (2005).

b. Selected Age-Length Key Plot

      i. This plot is to visualize the selected age-length key and portrays the observed age-length key. This plot is built using the alkPlot() function (FSA package) with type='bubble.'

  c. Age-Frequency Histogram

      i. The age-frequency histogram provides the user with information of the relative proportion of fish by age in the sample. The y-axis is formatted as proportions to allow comparisons among samples.

      ii. The plot is simply created using by totaling the number of fish in each age class (after the age-length key has been applied to the whole sample) and dividing each of those values by the total number of fish in the sample. The plot is constructed using the barplot() function.

  d. Growth Metrics Plot and Associated Tables

      i. von Bertalanffy Growth Model Fitting

         1. The von Bert growth model is a very common growth function used in fishery science (Ogle 2016). The app fits this model using a nonlinear least squares regression technique with the nls() function (nls package) which uses a maximum of 200 iterations before converging on best-fit coefficients. Starting values for the model coefficients are as follows: Linf = max length of fish in aged sample, K = 0.3, and t_0 = 0.

            a. L infinity is defined as the maximum *mean* length.

            b. K relates how quickly the function approaches Linf.

            c. t_0 is the x-intercept

         2. The function fits a nonlinear regression model through the observed data points of specific lengths at ages.

         3. This model has normal linear regression assumptions including 1) model fits the data (i.e., von Bert function is the best curve for the observed growth trajectory) and 2) measurement errors are independent, normally distributed, and homoscedastic.

      ii. Growth Plot

         1. An original figure is constructed with the plot() function to plot all individual fish from the sample with open dots. Mean lengths-at-ages are added to the plot with the points() function. The app then uses the fitted von Bert growth model to predict ages of 200 sequenced lengths between the minimum and maximum lengths in the sample data with

the predict() function and uses the lines() function to add the curve to the plot.

   iii. Tables
1. Mean length-at-age and weight-at-age tables are simply compiled means, counts, and measures of variance for each cohort of fish from the selected sample. Calculations for measures of variance can be found in Table 1.
2. The von Bertalanffy Growth Equation Table provides the estimated values for the 3 coefficients as well as lower and upper 95% confidence interval bounds for the equation coefficients. These confidence intervals were calculated with a profile likelihood method by using the conf.int() function on the fitted von Bert model.

e. Catch Curve (Mortality)
   i. A catch curve is simply the negative slope of a linear regression model that fits natural log-transformed abundance across age classes (Ogle 2016). The regression model starts with the most abundant age class, termed the "descending limb," which assumes that year class is the first which is fully recruited to the gear used to capture fish for the sample. The resulting negative slope is equal to the instantaneous total mortality (Z) of the population.
   ii. The app also uses a weighted regression to estimate Z which reduces the relative impact of older age classes that generally have lower abundances (details in Ogle 2016).
   iii. The app sums the total number of fish in each age class and converts catch data into natural logs (ln(total +1) to avoid ln(1) = 0). The most abundant age class is identified and all parameters are put into the catchCurve() function (FSA package).
   iv. Total instantaneous mortality rate Z is a measure of how the number of fish in the population declines in an imperceptibly short period of time (Ogle 2016); therefore, it is used in most mathematical models that do not use incremental (e.g., annual) timesteps. Total annual mortality A is the proportion of fish in a closed population that die in one year, and can be converted from Z using the equation $A = 1 - e^{-Z}$. This is convenient because A is much easier to interpret. Both A and Z are provided on the plot and below the plot in a table. Upper and lower 95% confidence interval bounds are also provided and calculated using linear regression theory for slope with the confint() function.
   v. The R^2 value is also calculated to assess the fit of the catch curve.

vi. Catch curve assumptions (Ogle 2016)
        1. All fish from the sample data should be captured with the same gear.
        2. Closed population (no immigration, emigration, births, deaths during sample period)
        3. Constant recruitment across years
        4. Z is constant across ages for ages included in the descending limb of the catch curve
        5. Vulnerability of fish to the gear used in the sample is constant across ages and years (for ages included in the descending limb)
   f. Theoretical maximum age
        i. Is calculated as:

$$Max\ Age = ln(1) - \left(\frac{intercept}{slope}\right)$$

        ii. Slope and intercept in the equation are the slope and intercept of the mortality catch curve.
   g. Estimated natural mortality
        i. Both estimates are fit using the metaM() function from the FSA package ("HoenigNLS" method for the Hoenig approach and "PaulyLNoT" for the Pauly method).

Sample Size and Model Fit

Just because the app calculates it, does not mean the metric is reliable or useful for management decisions. We recommend users follow the recommendations below by evaluating sample sizes and model fits of each sample before applying any of these analyses for among lake/year comparisons or other management decisions. Below are some "baseline" guidelines for evaluating the appropriateness of several of the analyses calculated by the app.

1. Catch-per-unit-Effort
   a. CPUE sample size recommendations are well outlined in the ODWC SSP Manual and are dependent on lake size and catch rate variability.
   b. If catch rates are highly variable, higher sample sizes are required for a reliable index of abundance and vice versa.

i. After the minimum amount of sample effort outlined in the SSP Manual has been reached, sample sizes should be based on observed catch rate variability. Users should try to hit a RSE of either 25 or 40 (indicating that CPUE data are able to detect a 25 or 40% change in population abundance). The number of samples required to achieve this precision are provided in the app output.

2. Length-Frequency
   a. Selection of Length Interval (bin)
      i. Neumann et al. (2012) suggest a general rule of thumb for length frequency histograms. This includes using 10 mm intervals for species which reach about 300 mm, 20 mm intervals for species which reach about 600 mm, and 50 mm intervals for species that reach about 1500 mm. The nature of the app allows the user to explore all options for length intervals, but we recommend following these guidelines for comparisons among samples.
   b. Total Sample Size
      i. Whenever possible, length frequency histograms should include at least 300 individuals for statistical comparisons among populations (Vokoun 2001).
      ii. However, during multi-year trend analyses, 100 total individuals over stock length (Gabelhouse 1984) may be adequate (Anderson and Neumann 1996).

3. Proportional Size Distribution (PSD)
   a. In general, the same recommendations as Length Frequency should be followed.

4. Relative Weight (W_r)
   a. A general rule of thumb should be 30 or more individuals to precisely assess mean condition. This includes means divided into size categories (e.g., still should have at least 30 individuals of quality length to assess mean condition of quality individuals).

5. Length-Weight Regression
   a. Using the same recommendations for selecting length intervals for Length Frequency, five fish per length interval likely achieves acceptable precision for the Length-Weight Regression (Neumann et al. 2012). Individual fish should be measured to the smallest possible increment (e.g., mm) and weighed with appropriate balances. Model fit is generally not a problem in length-weight regressions with adequate sample sizes, as length and weight are usually very predictive with established relationships in fish.

6. Age-Length Key

a. To build a robust age-length key, Coggins et al. 2013 recommended collecting 10 aging structures per length interval (length interval determined by dividing anticipated length infinity by 30).

b. The smooth age-length key method may not be the best method applying ages to sample data for all samples. This method builds a model to predict the probability of a fish being a certain age, given its length. Therefore, in populations with large overlaps in length among age classes (e.g., a stunted population), this model may not be the best option. The user can assess age-length overlap and potential model fit with the age-length key bubble plot.

7. Mean Length and Weight-at-Age
a. We would recommend a sample size of about 15 individuals in an age class before confidently assessing and comparing mean lengths or weights-at-age among samples.

8. von Bertalanffy Growth Equation
a. The app provides upper and lower 95% confidence interval bounds for all three parameters of the von Bert equation. The user can use these to assess model fit as well as the plotted growth curve in relation to observed mean lengths-at-age. If confidence intervals for the three parameters are reasonable (e.g., a 0.2 – 0.8 interval for K is probably not considered reasonable), then the equation likely fits well. Additionally, this would be supported if the fitted line on the plot matches reasonably well with the mean lengths-at-age.

9. Catch Curve
a. Catch curve analyses are sensitive models, and therefore usually require several hundred fish to be present in the sample. The model fit also depends on the general lifespan of the species. For example, longer-lived species (e.g., striper) may require a higher sample size than short-lived species (e.g., bluegill). Assessing the model fit and reliability of the catch curve from app output is similar to the von Bert growth curve. The app provides upper and lower 95% confidence interval bounds for both Z and A when applicable. Thus, if these intervals are reasonable (e.g., +/- 10% annual mortality is probably a good cutoff, as a 20% difference in mortality can be drastic), the model fit and mortality estimates may be adequate and used for management purposes. In addition, the app produces an R^2 value (correlation coefficient) which measures how well the linear model fits the observed data. We recommend a R^2 value greater than 0.5 before relying on calculated mortality rates.

Common Errors

Sometimes while using the app, the user will come across analyses that will not run properly and red errors messages will appear. Below are a few of the more common error types as well as troubleshooting hints or explanations of what might have gone wrong.

1. Overall
    a. Very small sample sizes can cause some metrics to produce errors. More than likely, if one of the functions produces an error, it is usually a data entry error or small sample size problem. In either case, results are likely erroneous in the first place.
    b. Make sure all fields (columns) are filled out in the sample data. Some historical data have numerous missing fields and data entry errors.
2. Catch Analysis Tab
    a. CPUE Panel
        i. If CPUE's are not calculated, it is likely that the Effort (nets) or Gear Length (electrofishing) column was not filled in. CPUE cannot be calculated without a known effort.
        ii. CPUE by Size Category
            1. Look at selected data. Lengths may not have been taken for these fish, lengths are needed to split into size categories.
            2. PSD length categories (Gabelhouse 1984) may not exist for the species. If PSD length categories do not exist, it is probably a less common species (e.g., Shortnose Gar, Spotted Sucker, etc.) or a small-bodied species (e.g., Inland Silverside, Logperch, etc.) that cannot be easily split up by size categories.
3. Single Species Analysis Tab
    a. Make sure to select a single species in the "Select Types of Analyses" tab.
    b. Sample data must, at minimum, have lengths associated with the single species for any of the length-weight or population dynamics metrics.
4. Length and Weight Metrics
    a. PSD Table
        i. Sample may not include fish above stock-size, no PSD proportions can be provided.
5. Population Dynamics Column
    a. Make sure to select an age dataset in the "Select Types of Analyses" tab.

b. Multinomial logistic regression to calculate the smoothed age-length key does not work well with 2 or less age classes and may not converge.
i. If this is the case, use the observed age-length key bubble plot and length-frequency plot to make inferences on population age structure.
6. von Bertalanffy Growth Curve and Coefficients
a. Small sample sizes (most commonly) or abnormal growth trajectories can cause problems with the model convergence or with calculating confidence intervals (confidence intervals can go to infinity)
i. Likely not enough data to use von Bertalanffy growth curve or potentially incorrect growth trajectory model.
7. Mortality Catch Curve
a. Less than 2 data points in the descending limb of the catch-curve.
i. Cannot calculate mortality.

R Packages Used

FSA

Ogle, D.H., J.C. Doll, P. Wheeler, and A. Dinno. 2022. FSA: Fisheries Stock Analysis. R package version 0.9.3

When using output derived from this app, please cite the above FSA Package. Almost every metric used a function from this package.

shiny

Winston Chang, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert and Barbara Borges (2021). shiny: Web Application Framework for R. R package version 1.7.1.

shinyjs

Dean Attali (2021). shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds. R package version 2.1.0.

dplyr

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2022). dplyr: A Grammar of Data Manipulation. R version 1.0.9.

tidyr

Hadley Wickham and Maximilian Girlich (2022). tidyr: Tidy Messy Data. R package version 1.2.0.

nnet

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

nlstools

Florent Baty, Christian Ritz, Sandrine Charles, Martin Brutsche, Jean-Pierre Flandrois, Marie-Laure Delignette-Muller (2015). A Toolbox for Nonlinear Regression in R: The Package nlstools. Journal of Statistical Software, 66(5), 1-21.

DT

Yihui Xie, Joe Cheng and Xianying Tan (2022). DT: A Wrapper of the JavaScript Library 'DataTables'. R package version 0.23.

data.table

Matt Dowle and Arun Srinivasan (2021). data.table: Extension of `data.frame`. R package version 1.14.2.

tibble

Kirill Müller and Hadley Wickham (2021). tibble: Simple Data Frames. R package version 3.1.7**.**

rlang

Lionel Henry and Hadley Wickham (2022). rlang: Functions for Base Types and Core R and 'Tidyverse' Features. R package version 1.0.2

fst

Mark Klik (2022). fst: Lightning Fast Serialization of Data Frames. R package version 0.9.8.

References

Blackwell, B. G., M. L. Brown, and D. W. Willis. 2000. Relative weight (Wr) status and current use in fisheries assessment and management. Reviews in Fisheries Science 8:1-44.

Bonar, S. A., W. A. Hubert, and D. W. Willis. 2009. Standard methods for sampling North American freshwater fishes. American Fisheries Society, Bethesda, MD.

Brenden, T. O., T. Wagner, and B. R. Murphy. 2008. Novel Tools for Analyzing Proportional Size Distribution Index Data. North American Journal of Fisheries Management 28:1233-1242.

Chang, W., J. Cheng, J. Allaire, Y. Xie, and J. McPherson. 2017. Web Application Framework for R. R package version 1.0.5.

Coggins, L. G., D. C. Gwinn, and M. S. Allen. 2013. Evaluation of age–length key sample sizes required to estimate fish total mortality and growth. Transactions of the American Fisheries Society 142:832-840.

Gabelhouse, D. W. 1984. A length-categorization system to assess fish stocks. North American Journal of Fisheries Management 4:273-285.

Gustafson, K. A. 1988. Approximating confidence intervals for indices of fish population size structure. North American Journal of Fisheries Management 8:139-139-141.

Guy, C. S., and M. L. Brown. 2007. Analysis and Interpretation of Freshwater Fisheries Data. American Fisheries Society, Bethesda, MD.

Guy, C. S., R. M. Neumann, D. W. Willis, and R. O. Anderson. 2007. Proportional size distribution (PSD): a further refinement of population size structure index terminology. Fisheries 32:348.

Isely, J. J., and T. B. Grabowski. 2007. Age and Growth. Pages 187-228 *in* C. S. Guy, and M. L. Brown, editors. Analysis and Interpretation of Freshwater Fisheries Data. American Fisheries Society, Bethesda, MD.

Isermann, D. A., and C. T. Knight. 2005. A computer program for age–length keys incorporating age assignment to individual fish. North American Journal of Fisheries Management 25:1153-1160.

Miranda, L. E., and P. W. Bettoli. 2007. Mortality. Pages 229-277 *in* C. S. Guy, and M. L. Brown, editors. Analysis and Interpretation of Freshwater Fisheries Data. American Fisheries Society, Bethesda, MD.

Neumann, R. M., and M. S. Allen. 2007. Size Structure. C. S. Guy, and M. L. Brown, editors. Analysis and Interpretation of Freshwater Fisheries Data. American Fisheries Society, Bethesda, MD.

Neumann, R. M., C. S. Guy, and D. W. Willis. 2012. Length, weight, and associated indices. Pages 637-676 *in* A. V. Zale, D. L. Parrish, and T. M. Sutton, editors. Fisheries Techniques, Third Edition. American Fisheries Society, Bethesda, MD.

Ogle, D. H. 2016. Introductory Fisheries Analyses with R. CRC Press, Boca Raton, FL.

R Core Team. 2017. R: A Language and Environment for Statistical Computing. Pages https://www.R-project.org/ *in*. R Foundation for Statistical Computing, Vienna, Australia.

Then, A. Y., J. M. Hoenig, N. G. Hall, D. A. Hewitt, and H. e. E. Jardim. 2015. Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. ICES Journal of Marine Science 72:82-92.