**LH\_Sampling**

Description: Functions for simulating fish populations and estimating life history parameters.

URL: https://github.com/evaschemmel/LHsampling.git

Authors: Erin Bohaboy (erin.bohaboy@noaa.gov)

Eva Schemmel (eva.schemmel@noaa.gov)

Updated: 2Aug 2021

References: Schemmel E., Bohaboy E., Kinney M., O'Malley J. (2022) An assessment of sampling strategies for estimating fish growth from fishery-dependent samples. *ICES Journal of Marine Science.*

Introduction: An individual-based model (IBM) incorporating within-population variability in von Bertalanffy growth, size-at-age-dependent natural mortality, and a size-selective fishery to simulate an exploited fish population and catch (harvest). A bootstrap algorithm allows the user to investigate various sampling approaches including sampling strategy (proportional or fixed otolith sampling, POS or FOS, respectively), sample size, supplementation with fishery-independent sampling, and assumptions regarding von Bertalanffy *t0* and the relationship between variance of length at age and age. A function to produce plots of the bootstrap sampling results is also provided.

Required packages: reshape, dplyr, ggplot2, magrittr, assertthat.

simulate\_population\_harvest IBM to simulate a fish population and catches.

**Description**:

An individual-based model (IBM) that simulates fish age and length values for a population and fishery harvest.

**Usage**:

simulate\_population\_harvest(Linf, Linf\_sd, M, Lorenzen, F, mincat, catsd, maxcat, maxcatsd, L0, L0\_sd, k, k\_sd, Amax, age\_max, N)

**Arguments**:

Linf Von Bertalanffy theoretical asymptotic length (cm)

Linf\_sd Population standard deviation of asymptotic length (cm)

M Instantaneous natural mortality rate (yr-1)

Lorenzen TRUE / FALSE specifying whether natural mortality is a function of individual length following Lorenzen (Lorenzen, 2000; Lorenzen, 2005)

F Apical (fully selected) instantaneous fishing mortality rate (yr-1)

mincat Minimum length at 50% fishery selectivity (cm)

catsd Slope of the ascending region of selectivity at length (cm), see details

maxcat Maximum length at 50% fishery selectivity (cm)

maxcatsd Slope of the descending region of selectivity at length (cm), see details

L0 Von Bertalanffy length at age 0 (cm)

L0\_sd Population standard deviation of length at age 0 (cm)

k Von Bertalanffy growth coefficient

k\_sd Population standard deviation of Von Bertalanffy growth coefficient

Amax Maximum longevity (years)

age\_max An arbitrary age selected to represent “old” fish (years)

N The number of age 0 fish in each simulated cohort, typical value = 100,000

**Details**:

The individual-based simulation model begins with a cohort of *N* age 0 fish which progress through ages 0 to *Amax* following the Von Bertalanffy growth function for length at age (*La*) with individual-specific *L∞* and *L0* (length at age 0) parameters drawn from a normal distribution of mean = µ and standard deviation = σ [~N(µ, σ)].

and

= predicted length (cm) of individual *i* at the end of age *a* (years)

*L∞,i* = asymptotic length (cm) of individual *i*

*Ki* = growth coefficient for individual *i*

*a* = age in years

= the theoretical age at which the fish would have zero length

It is important to note that fish enter the population simulation as age 0 fish, meaning they are at the end of their first year of life, about to turn 1 year old. At each age, individual fish are subject to death by natural mortality followed by death from harvest (fishing mortality) in a Bernoulli random process with probability *pM,i* or *pF,i*, for natural death or harvest, respectively. According to the Baranov catch equation (Quinn and Deriso, 1999), the probability of natural mortality at age *a* for fish *i* is

where *Ma,i* is natural mortality at age *a* for fish *i* and is a function of individual length at age (*La,i*, calculated from the VBGF, below), following Lorenzen (Lorenzen, 2000; Lorenzen, 2005):

Parameter *M1* describes the relationship between length and natural mortality in the overall population. *M1* is estimated within the model by simultaneously solving the following set of equations:

where:

*Ma* = population expected natural mortality at age

*La* = population mean length at age

*Moverall* = population overall natural mortality

*NAmax* and *N0* = expected number of fish in the population at ages *Amax* and 0.

*SurvivorshipAmax* = average probability of an individual surviving natural mortality to reach *Amax*.

The probability of fishing mortality (being harvested) at age *a* for fish *i* is

Where *Fa,i* is the fishing mortality at age *a* for fish *i* and is the product of apical (fully selected) fishing mortality (*Fʹ*) and selectivity at age conditioned on length (*SelexL*):

*SelexL* is modeled as the cumulative normal probability density of mean = *mincat* and standard deviation = *mincatsd*.

At each time step, individual fish that experience death by natural mortality are removed from the simulated cohort. Of the fish that survived natural mortality, the individuals that die by fishing mortality are removed from the simulated cohort and set aside as harvested fish. The fish that survived both natural and fishing mortality undergo Von Bertalanffy growth and advance to the next age. *Amax* + 1 cohorts are created and then 1 age (from *a* = 0 to *Amax*) is taken from the survivors and harvest of each cohort to form the simulated population and catch.

**Value**:

A named list:

$population (dataframe: $age, $length): the simulated population

$harvest (dataframe: $age, $length): the simulated harvest (i.e. catch for fisheries-dependent sampling)

$Avg\_age (dataframe: $Ages, $L\_age, $M\_age, and $Selex): average characteristics of the simulated population at age

$parameters named list of 19 elements including all input parameters used in the simulation and the simulated population coefficient of variation of length at age\_max and age\_0

**Example**:

S1\_Auric\_lowF <- simulate\_population\_harvest(Linf = 32.5, Linf\_sd = 2.5, M = 0.18, F = 0.09, Lorenzen = TRUE, mincat = 10, catsd = 2.5, maxcat = 200, maxcatsd = 0, L0 = 10, L0\_sd = 2.5, k = 0.6, k\_sd = 0, Amax = 32, age\_max = 15, N = 100000)

LH\_sample Bootstrap sampling routine to estimate life history parameters

**Description**:

A bootstrap sampling routine to estimate life history parameters from fishery catches simulated by simulate\_population\_harvest().

**Usage**:

LH\_sample <- function(sim\_output, n\_boots, samp\_size, sample\_type, supp\_large = FALSE, supp\_large\_n\_per\_bin = 3, supp\_small = FALSE, supp\_small\_n\_per\_bin = 3, supp\_min\_length = 2, constrained = FALSE, t0 = 0, SD\_L\_const = TRUE, save\_bootstraps = FALSE, Amax = NULL, age\_max = NULL, Lbin\_width = 2)

**Arguments**:

sim\_output Output from simulate\_population\_harvest()

n\_boots Number of bootstrap samples to perform

samp\_size Total sample size for each bootstrap

sample\_type The sampling strategy to be used, either proportional otolith sampling (‘POS’) or fixed otolith sampling (‘FOS’)

supp\_large TRUE / FALSE specifying whether supplemental samples will be collected from large length bins

supp\_large\_n\_per\_bin The number of samples per length bin to be collected from large bins (ignored if *supp\_large* = FALSE)

supp\_small TRUE / FALSE specifying whether supplemental samples will be collected from small length bins

supp\_small\_n\_per\_bin The number of samples per length bin to be collected from small bins (ignored if *supp\_small* = FALSE)

supp\_min\_length The minimum length fish that could be collected from the wild fish population

constrained TRUE / FALSE specifying whether theoretical time at length zero (*t0*) should be estimated

t0 If *constrained* = TRUE, the fixed value for *t0* (typically 0)

SD\_L\_const TRUE / FALSE describing assumptions of population variance in length at age. If TRUE, then standard deviation () of length at age is assumed a linear function of age. If FALSE, then the coefficient of variation of length at age is assumed a linear function of age.

save\_bootstraps TRUE / FALSE specifying whether all bootstrap samples will be included in the function output

Amax Maximum longevity (years). If not specified, this value is taken from sim\_output.

age\_max An arbitrary age selected to represent “old” fish (years). If not specified, this value is taken from *sim\_output*.

Lbin\_width The width of each length bin (cm).

**Details**:

This function will take *n\_boots* samples (without replacement) from the harvested individuals following either a fixed otolith sampling (FOS) or proportional otolith sampling (POS) strategy. The function then parameterizes the von Bertalanffy growth function and estimates the population coefficient of variation of length at age for each bootstrap sample.

To produce each FOS bootstrap, *n* individuals (calculated as total sample size, *samp\_size*, divided by the number of length bins represented in the harvest, rounded up to the next integer) are randomly sampled, without replacement, from the harvest for each length bin. If less than *n* individuals are available in a harvest length bin, then additional samples are randomly drawn from the remaining length bins in the harvest until the total prescribed number of samples (*n* × number of length bins represented in the harvest) are attained. If greater than *samp\_size* individuals have been collected, individuals within fully filled length bins will be randomly discarded from the sample until *samp\_size* is attained. If supplemental sampling is being implemented (*supp\_large* or *supp\_small* = TRUE) then the specified number of samples per large (*supp\_large\_n\_per\_bin*) or small (*supp\_small\_n\_per\_bin*) are taken randomly (without replacement) from the population. For this step, ‘small’ and ‘large’ length bins are defined as any bins containing individuals in the population where fewer than *supp\_large\_n\_per\_bin* or *supp\_small\_n\_per\_bin* individuals in the FOS sample. After supplemental samples are added, individuals from the most populous length bins will be randomly discarded from the sample until *samp\_size* is attained.

Each POS bootstrap is produced in much the same way as for FOS. The primary difference is the ‘target’ number *n* of individuals per length bin *L* (*nL*) is calculated based on the proportional distribution of lengths within the entire harvest, where *nL* is rounded to the nearest integer. If then *nL* will be updated by adding or subtracting from the most populous length bin. Making adjustments to the number of samples targeted in the most populous length bin minimizes effects on the POS scheme from adjusting for the desired *samp\_size*. For each POS bootstrap, *nL* individuals are randomly sampled, without replacement, from the harvest for each length bin. If supplemental sampling is being implemented (*supp\_large* or *supp\_small* = TRUE) then the specified number of samples per large (*supp\_large\_n\_per\_bin*) or small (*supp\_small\_n\_per\_bin*) are taken randomly (without replacement) from the population. For this step, ‘small’ and ‘large’ length bins are defined as any bins containing individuals in the population with fewer than *supp\_large\_n\_per\_bin* or *supp\_small\_n\_per\_bin* in the POS sample. After supplemental samples are added, individuals from the non-supplemental length bins will be randomly discarded from the sample until *samp\_size* is attained.

For each bootstrap sample, the coefficient of variation of length at age (*CVa*) is calculated (given ≥ 2 individuals per age). The coefficient of variation of length at age 0 (*CVa0*) and age\_max (*CVage\_max*) are extrapolated assuming a linear relationship between age and *CVa* (SD\_L\_const = FALSE) or standard deviation of length at age (SD\_L\_const = TRUE). The slope and intercept are parameterized using *nls()*. If *nls()* fails to converge, then the average variance in length at age, over all ages, is used.

For each bootstrap sample, the von Bertalanffy growth function is parameterized using *nls()*, with *t0* (where *t0* = *a0* + 1) either estimated or fixed at the user specified value.

and

The total number of bootstrap samples for which *nls()* did not converge is reported out to the R console.

**Value**:

A named list:

$list\_boot\_samps: list containing a dataframe ($age, $length, $binL, $n, $nbin) for each bootstrap sample. Will only contain data if *save\_bootstraps* == TRUE.

$list\_boot\_preds: list containing a dataframes of predicted length at age ($age, $length) calculated from the fitted von Bertalanffy growth model for each bootstrap sample

$list\_boot\_mods: list containing *nls*() fitted model objects for each bootstrap sample

$list\_boot\_CVs: list containing a dataframe of calculated CV of length at age (if defined, $age, $CV\_age) for each bootstrap sample

$list\_boot\_SDs: list containing a dataframe of calculated standard deviation at length (if defined, $age, $SD\_age) for each bootstrap sample

$parameter\_outputs: dataframe ($Linf, $K1, $a0, $CV\_L\_a0, $CV\_L\_age\_max) of *n\_boots* rows of the estimated von Bertalanffy growth parameters

$params\_input\_output: list of 17 elements including all input parameters and $boots\_nls\_fail, which is the total number of bootstrap samples for which *nls*()failed to converge for the von Bertalanffy growth function.

$simulation\_params: list of 19 elements including all input parameters used in simulate\_population\_harvest() and the simulated population coefficient of variation of length at age\_max and age\_0. This element is passed directly from simulate\_population\_harvest().

$parameter\_summary\_all\_boots: 95% confidence intervals, interquartile range, and arithmetic mean von Bertalanffy parameter estimates over all bootstraps (dataframe: $parm\_name, $lower95, $lower50, $avg, $upper50, $upper95)

$list\_some\_boot\_samps: list containing a dataframe for each of 9 randomly selected bootstrap samples ($age, $length, $binL, $n, $nbin)

**Example**:

S1\_A9 <- LH\_sample(sim\_output = S1\_Auric\_lowF, n\_boots = 1000, samp\_size = 300, sample\_type = 'POS', supp\_large = FALSE, supp\_small = FALSE, constrained = FALSE, save\_bootstraps = TRUE, Lbin\_width = 2)

LH\_plot Produce plots from an LH\_sample object

**Description**:

A function that produces plots from the output of LH\_sample()

**Usage**:

LH\_plot <- function(sample\_output, output\_type = 'none')

**Arguments**:

sample\_output Output from LH\_sample()

output\_type How plots are written and saved: ‘none’ displays in R graphics device only, ’pdf’ produces a single .pdf with all plots, and ‘png’ produces a separate .png for each plot.

**Details**:

This function outputs plots to the R graphics device, requiring the user press [enter] in the R console to advance to the next plot. Users are encouraged to examine each plot to confirm the characteristics of the simulated population, fishing fleet, and bootstrap samples match expectations.

Plots:

Selectivity at length as specified in the initial population simulation

Natural mortality at age

Scatterplot of population length at age, overlaid with average length at age as expected from the von Bertalanffy growth parameters passed from simulate\_population\_harvest(). The scatterplot is limited to a random sample of 50,000 individuals from the population.

Histograms of frequency distributions for population length, harvest length, population age, and harvest age.

Histograms of number of individuals per length bin for 9 randomly chosen bootstrap samples.

**Value**:

Files containing plots saved to the working directory [getwd()] if *output\_type* = ‘pdf’ or ‘png’. File names will include system time as [YYYYMMDDhhmmss].

**Example**:

LH\_plot(S1\_A9, output\_type = ‘pdf’)