A multi-stock, multi-region extension of WHAM

Black sea bass Research Track

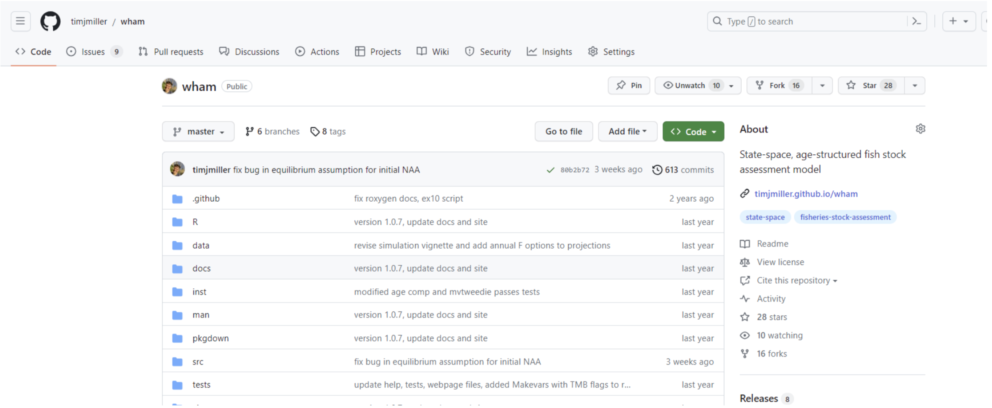
Tim Miller

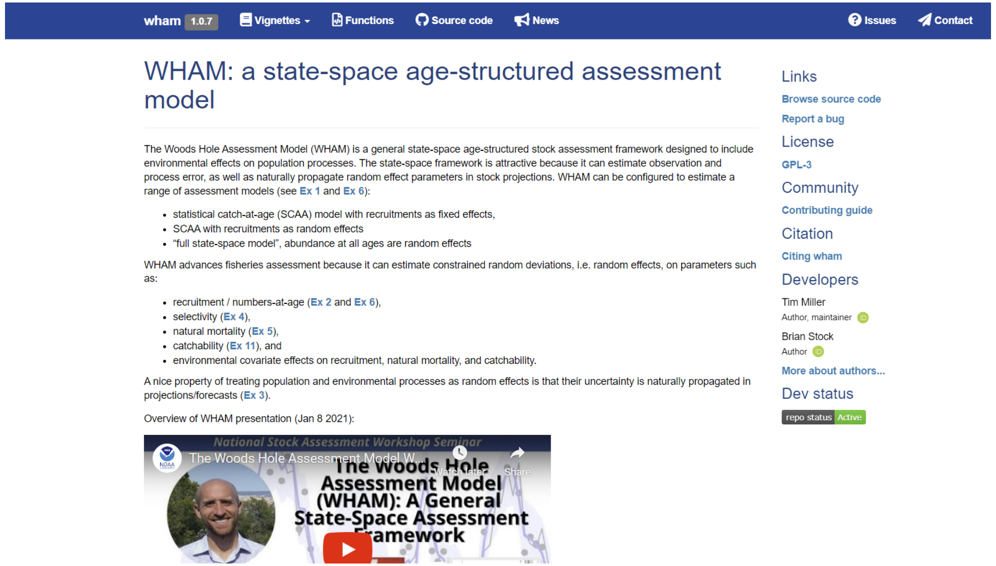
Outline

- Brief description of standard WHAM package
- New features of multi-stock, multi-region extension (Multi-WHAM)
- Generalizing from univariate, to multivariate populations (i.e., survival by region, types of mortality)
- Modeling dynamics of a multi-region populations and predictions for catch and index observations

WHAM

- Woods Hole Assessment Model (WHAM)
- R package
  - `remotes::install_github("timjmiller/wham", dependencies=TRUE)`
- See Stock and Miller 2021 for further details
- "State-space" age-structured assessment model
  - random effects optional in various aspects of the population dynamics
  - Estimation by maximum marginal likelihood
  - Estimation without random effects (traditional SCAA) is default
- Has been recommended for management of 9 stocks in the Northeast US
  - Atlantic butterfish
  - Georges Bank and Eastern GB haddock
  - Atlantic bluefish
  - American plaice
  - 4 Atlantic cod stocks

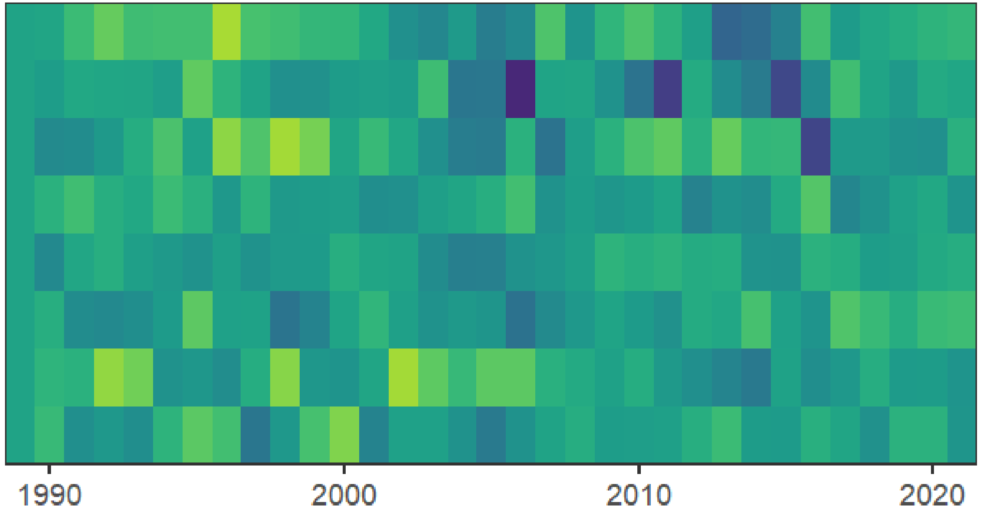




WHAM: Stochastic processes

- Random effects are Gaussian on transformed scale
- Time and/or age varying random effects can be assumed in several parameters:
  - Recruitment (time)
    - log deviations are iid or AR1(year)
  - Transitions in abundance by time step ("survival") (time and age)
    - log deviations from deterministic transitions are iid or 2DAR1
  - Selectivity for indices and fishing fleets (time and/or age)
    - logit deviations are iid, AR1(age), AR1(year) or 2dAR1
  - Natural mortality (time and/or age)
    - log deviations are iid, AR1(age), AR1(year) or 2dAR1
  - Index catchability (time)
    - logit deviations are iid, or AR1(year)

$$Cov\left(\epsilon_{a,y}, \epsilon_{a',y'}\right) = \frac{\rho_{\mathcal{A}}^{|a-a'|}\rho_{\mathcal{Y}}^{|y-y'|}\sigma^2}{\left(1-\rho_{\mathcal{A}}^2\right)\left(1-\rho_{\mathcal{Y}}^2\right)}$$



WHAM: Environmental Covariates

- State-space models for covariates
  - Time series model for latent covariate random effects
  - E.g. first order auto-regressive (AR1):

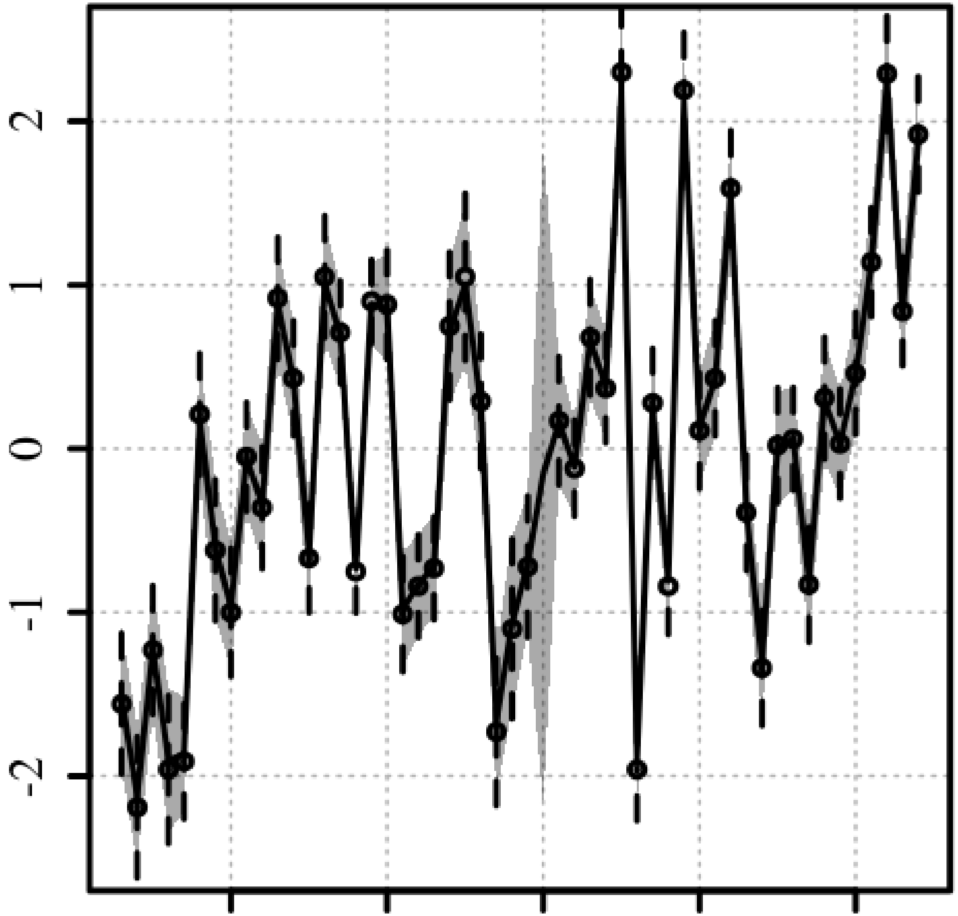
$$X_y | X_{y-1} = \mu(1-\rho) + \rho X_{y-1} + \epsilon_y$$

$$\epsilon_y \sim N(0, \sigma^2_X)$$

  - Random walk $(\rho = 1)$ also an option
  - Observations of covariate have error

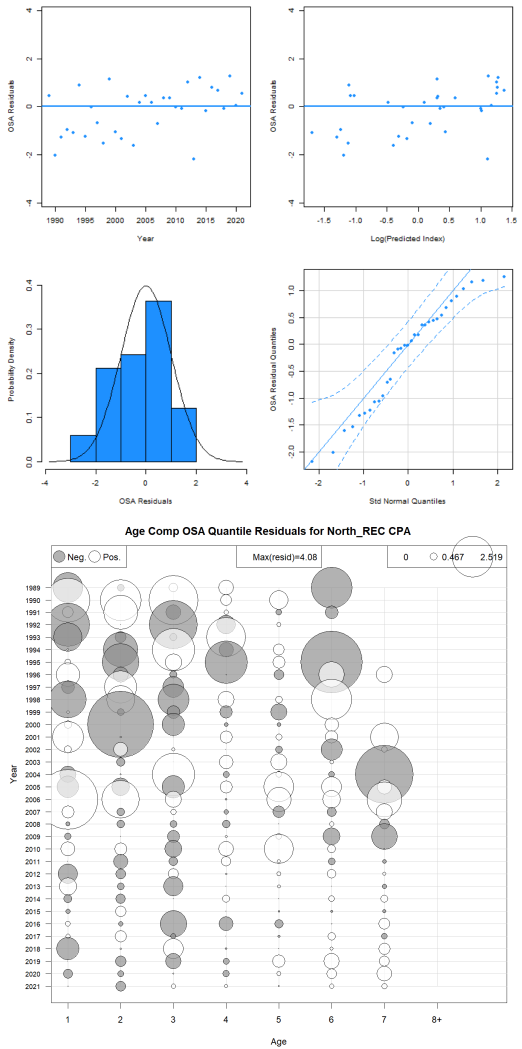
$$x_y | X_y \sim N(X_y, \sigma^2_x)$$

- Latent covariate can affect:
  - Recruitment
  - Natural Mortality
  - Index catchability



WHAM: Observations

- Aggregate catch (by fleet)
    - log-normal
    - $\widehat{C}_{f,a} = N_{y,a}\frac{F_{f,y}s_{f,a}}{Z_{y,a}}\left(1-e^{-Z_{y,a}}\right)W_{f,y,a}$
- Aggregate indices (biomass or abundance)
    - log-normal
    - $\hat{I}_{i,a} = q_i s_{i,a} N_{y,a} e^{-Z_{y,a}\delta_{i,y}} W_{f,y,a}$
- Index or catch age composition
    - Multinomial
    - Dirichlet-multinomial (estimated dispersion)
    - logistic-normal (estimated dispersion)
    - Dirichlet (estimated dispersion)
    - Multivariate-Tweedie (estimated dispersion)
- Environmental covariates
    - normal



Multi-WHAM

New features:

- stock-specific abundance at age and by region
- seasonal intervals within years
- effects of environmental covariates on mortality rates by stock, region, and age
- effects of environmental covariates on recruitment by stock
- variation in movement rates by stock, region-to-region, season, age, and year
- effects of environmental covariates on movement rates by stock, region-to-region, season, and age
- mortality and movement modeled sequentially or simultaneously
- stock-specific stock-recruitment models
- priors for movement rates
- seasonal operation of fleets
- more options for initial abundance at age
- options for weighting of stock-specific SSB/R for global SPR-based reference points

In following slides, everything is stock-specific unless otherwise noted

Abundance transitions: Baranov

- For a single region, the numbers at age a surviving, dying from fishing, and dying from other causes (e.g., M) over a unit time interval (e.g., 1 year) are

$$N_{L,t+1,a} = N_{t,a} S_{t,a} = N_{t,a} e^{-Z_{t,a}}$$

$$N_{C,t+1} = N_{t,a} H_{t,a} = N_t \frac{F_{t,a}}{Z_{t,a}} \left(1 - e^{-Z_{t,a}}\right)$$

$$N_{K,t+1} = N_{t,a} D_{t,a} = N_t \frac{M_{t,a}}{Z_{t,a}} \left(1 - e^{-Z_{t,a}}\right) .$$

Abundance transitions: Baranov

- More generally, for an interval of length δ:

$$N_{L,t+\delta,a} = N_{t,a}S(t,\delta,a) = N_{t,a}e^{-Z_{t,a}\delta}$$

$$N_{C,t+\delta} = N_{t,a}H(t,\delta,a) = N_t\frac{F_{t,a}}{Z_{t,a}}\left(1 - e^{-Z_{t,a}\delta}\right)$$

$$N_{K,t+\delta} = N_{t,a}D(t,\delta,a) = N_t\frac{M_{t,a}}{Z_{t,a}}\left(1 - e^{-Z_{t,a}\delta}\right).$$

Vector-Matrix form

- Define the vector of numbers at age a at time t that are alive, previously captured, and previously dead to to natural mortality:

$$\mathbf{N}_{t,a} = (N'_{L,t,a}, N'_{C,t,a}, N'_{K,t,a})'$$

and the probability transition matrix:

$$\mathbf{P}_{t,\delta,a} = \begin{bmatrix} S(t,\delta,a) & H(t,\delta,a) & D(t,\delta,a) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- The abundance in each category at the end of the interval δ is

$$\mathbf{N}_{t+\delta,a} = \mathbf{P}'_{t,\delta,a}\mathbf{N}_{t,a}$$

- The probability transition matrix over multiple seasons (e.g., a year) is just the product of the seasonal matrices:

$$\mathbf{P}_{t,\delta_1+\delta_2,a} = \mathbf{P}_{t,\delta_1,a}\mathbf{P}_{t,\delta_2,a}$$

- If t is the beginning of the year we typically set $N_{C,t,a} = 0$ to obtain cumulative catch over the yearly interval.

Multiple regions

- When there are n_R regions and n_F fleets, the probability transition matrix is generalized to

$$\mathbf{P}_{t,\delta,a} = \begin{bmatrix} \mathbf{O}(t,\delta,a) & \mathbf{H}(t,\delta,a) & \mathbf{D}(t,\delta,a) \\ 0 & \mathbf{I}_H & 0 \\ 0 & 0 & \mathbf{1}_D \end{bmatrix}$$

$\mathbf{O}(t,\delta,a)$ is the $n_R \times n_R$ matrix defining survival and movement from one region (row) to another (column) over the interval δ:

$$\mathbf{O}(t,\delta,a) = \begin{bmatrix} O_{1,1}(t,\delta,a) & \cdots & O_{1,n_R}(t,\delta,a) \\ \vdots & \ddots & \vdots \\ O_{n_R,1}(t,\delta,a) & \cdots & O_{n_R,n_R}(t,\delta,a) \end{bmatrix}$$

$\mathbf{H}(t,\delta,a)$ is the $n_R \times n_F$ matrix defining proportions captured in each fleet (column) given alive in each region (row) at the beginning of the interval:

$$\mathbf{H}(t,\delta,a) = \begin{bmatrix} H_{1,1}(t,\delta,a) & \cdots & H_{1,n_F}(t,\delta,a) \\ \vdots & \ddots & \vdots \\ H_{n_R,1}(t,\delta,a) & \cdots & H_{n_R,n_F}(t,\delta,a) \end{bmatrix}$$

Multiple regions

- Multi-WHAM currently assumes each fleet operates in a single region so each column of $\mathbf{H}(t, \delta, a)$ has at most 1 non-zero value.
- $\mathbf{D}(t, \delta, a)$ is $n_R \times 1$ in Multi-WHAM to minimize dimensions of $\mathbf{P}$, but could theoretically be $n_R \times n_R$ if tracking numbers dead due to natural mortality by region was of interest.
- Multi-WHAM can assume survival and movement processes are sequential or simultaneous within a seasonal interval.
  - When sequential, survival occurs over the interval as usual and movement is assumed to happen instantly at the end of the interval:

$$\mathbf{O}(t, \delta, a) = \mathbf{S}(t, \delta, a)\boldsymbol{\mu}(t, \delta, a).$$

  - When simultaneous, movement parameters are instantaneous rates like mortality. There is an infinitesimal matrix that is a function of the instantaneous rates which is exponentiated for the probability transition matrix:

$$\mathbf{P}_{t,\delta,a} = e^{\mathbf{A}_{t,\delta,a}\delta}$$

Probability transition matrix example

Northern stock, age 5, year 2021

	North	South	North_Commercial	North_Recreational	South_Commercial	South_Recreational	M
North	0.46	0.02	0.10	0.13	0.00	0.00	0.28
South	0.45	0.02	0.07	0.09	0.02	0.07	0.28
North_Commercial	0.00	0.00	1.00	0.00	0.00	0.00	0.00
North_Recreational	0.00	0.00	0.00	1.00	0.00	0.00	0.00
South_Commercial	0.00	0.00	0.00	0.00	1.00	0.00	0.00
South_Recreational	0.00	0.00	0.00	0.00	0.00	1.00	0.00
M	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Sequential survival and movement

- $\mathbf{S}$ is a diagonal matrix of proportions surviving in each region (given they start in that region):

$$\mathbf{S}(t,\delta,a) = \begin{bmatrix} e^{-Z_1(t,\delta,a)} & 0 & \cdots & 0 \\ 0 & e^{-Z_2(t,\delta,a)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & e^{-Z_R(t,\delta,a)} \end{bmatrix}$$

- $\boldsymbol{\mu}$ is matrix of proportions moving from one region to another or staying (given they start in that region)

$$\boldsymbol{\mu}(t,\delta,a) = \begin{bmatrix} 1-\sum_{r'\neq 1}\mu_{1\to r'} & \mu_{1\to 2} & \cdots & \mu_{1\to R} \\ \mu_{2\to 1} & 1-\sum_{r'\neq 2}\mu_{2\to r'} & \cdots & \mu_{2\to R} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{R\to 1} & \cdots & \mu_{R\to R-1} & 1-\sum_{r'\neq R}\mu_{R\to r'} \end{bmatrix}$$

The rows sum to 1 and each of the $\mu_{r\to r'}$ parameters may be year, age, and season-specific.

Simultaneous survival and movement

- For the scenario where 1 fleet operates in each region the instantaneous rate matrix is

$$\mathbf{A}_{t,\delta,a} = \begin{bmatrix} a_1 & \mu_{1\to 2} & \cdots & \mu_{1\to R} & F_{t,a,1} & 0 & \cdots & 0 & M_{t,a,1} \\ \mu_{2\to 1} & a_2 & \cdots & \mu_{2\to R} & 0 & F_{t,a,2} & \cdots & 0 & M_{t,a,2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & 0 & \vdots \\ \mu_{R\to 1} & \cdots & \cdots & a_R & 0 & \cdots & 0 & F_{t,a,R} & M_{t,a,R} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

  where $a_r = -\left(\sum_{r' \neq r} \mu_{1\to r'} + F_{t,a,r} + M_{t,a,r}\right)$ (rows sum to 0).
- The component matrices $\mathbf{O}$, $\mathbf{H}$, and $\mathbf{D}$ are obtained from exponentiating $\mathbf{A}$.
- This option is still considered experimental because it is currently computationally inefficient.

Movement parameters

- There are at most $n_R - 1$ parameters for each season (i) and region (r) for either the sequential or simultaneous configurations.
- Movement parameters are estimated on a transformed scale.
  - simultaneously: log scale
  - sequential: additive logit link transformation (like a multinomial regression)
- Age- and/or year-specific (1D or 2D AR1/iid) random effects are normally distributed on the transformed scale

$$f(\mu_{r\to r',i,a,y}) = \theta_{r\to r',i} + \epsilon_{r\to r',i,a,y}$$

$$Cov\left(\epsilon_{r\to r',i,a,y}, \epsilon_{r\to r',i,a',y'}\right) = \frac{\rho_{r\to r',i,\mathcal{A}}^{|a-a'|}\rho_{r\to r',i,\mathcal{Y}}^{|y-y'|}\sigma_{r\to r',i}^2}{\left(1-\rho_{r\to r',i,\mathcal{A}}^2\right)\left(1-\rho_{r\to r',i,\mathcal{Y}}^2\right)}$$

- Mean, variance and correlation parameters for the random effects are season- and region-to-region-specfic.
- Covariate effects can be age-, season-, and/or region-to-region-specific:

$$f(\mu_{r\to r',i,a,y}) = \theta_{r\to r',i} + \sum_{k=1}^{n_E} \beta_{a,i,r\to r',k} E_{k,y}$$

and the same orthogonal polynomial options in the current version of WHAM are available.

Movement parameters

- Movement parameters can including both random and environmental effects:

$$f(\mu_{r\to r',i,a,y}) = \theta_{r\to r',i} + \epsilon_{r\to r',i,a,y} + \sum_{k=1}^{n_E} \beta_{r\to r',i,k} E_{k,y}$$

- **Multi-WHAM currently has no likelihood component for tagging observations, but priors can be configured from auxiliary movement parameter estimates.**
  - When prior distributions are used, the (mean) movement parameters are random effects with the mean defined by the user-defined fixed effect counterpart and standard deviation

$$\gamma_{r\to r',i} \sim \mathrm{N}\left(\theta_{r\to r',i}, \sigma^2_{r\to r',i}\right)$$

  and the transformed movement parameter is defined instead as

$$f(\mu_{r\to r',i,a,y}) = \gamma_{r\to r',i} + \epsilon_{r\to r',i,a,y} + \sum_{k=1}^{n_E} \beta_{r\to r',i,k} E_{k,y}$$

.

Natural mortality

When not estimated, (mean) mortality rates may be region-, and age-specific $\mu_{M,r,a}$. When age- and year-specific random effects $m_{r,a,y}$ are allowed, variance and correlation parameters can be region-specific. The options for each region are

- `none`: (default) No random effects by age or year.
- `iid_a`: uncorrelated M by age, constant in time.
- `iid_y`: uncorrelated M by year, constant all ages.
- `ar1_a`: M correlated by age (AR1), constant in time.
- `ar1_y`: M correlated by year (AR1), constant all ages.
- `iid_ay`: M uncorrelated by year and age (2D).
- `ar1_ay`: M correlated by year and age (2D AR1).

Any environmental covariate effects can be region- and age-specific. The general configuration of log natural mortality is

$$\log\left(M_{r,a,y}\right) = \mu_{M,r,a} + m_{r,a,y} + \sum_{k=1}^{n_E} \beta_{r,a,k} E_{k,y}$$

Initial Abundance at age

The options for parameterizing initial numbers at age have been expanded in Multi-WHAM:

- `age-specific`: $\log N_{1,r,a}$ are fixed effects.
- `equilibrium`: $\log N_{1,r,1}$ (initial recruitment) and $\log F$ are fixed effects defining equilibrium numbers at age in the first year.
- `iid`: $\log N_{1,r,a} \sim \mathrm{N}\left(\eta_r, \sigma_r^2\right)$ are iid random effects.
- `ar1`: $\log N_{1,r,a} \sim \mathrm{N}\left(\eta_r, \frac{\sigma_r^2}{(1-\rho_r^2)}\right)$ are AR1 random effects.

Equilibrium assumption

- Natural mortality and selectivity for fleet-specific fishing mortality at age are the same as those that occur during the first year of the model.
- With the assumption that each stock spawns in 1 region, there is only 1 $\log N_{1,r,1}$ parameter.
- The equilibrium calculations are essentially the same as those for SSB/R and Y/R calculations.
- The $n_R \times n_R$ equilibrium probability matrix of survival to age a and being in each region is

$$\widetilde{\mathbf{O}}_a = \begin{cases} \prod_{i=0}^{a-1} \mathbf{O}_i & 1 \leq a < A \\ \left[\prod_{i=0}^{a-1} \mathbf{O}_i\right] \mathbf{O}_+ & a = A \end{cases}$$

  where $\mathbf{O}_0 = \mathbf{I}$.
- $\mathbf{O}_i$ is the upper left survival and movement sub-matrix of the probability transition matrix for age i as defined previously. For the plus group $a = A$, $\mathbf{O}_+ = (\mathbf{I} - \mathbf{O}_A)^{-1}$ is a "fundamental matrix" derived using the matrix generalization of the summation of a convergent geometric series.
- Equilibrium abundance at age is

$$N_{1,r,a} = N_{1,r} \widetilde{\mathbf{O}}_a(s_r, r)$$

  where s_r is the region of spawning and $\mathbf{O}(i,j)$ is the element in row i and column j of the matrix $\mathbf{O}$.

Abundance at age transitions

- The numbers at age are vector analogs of the equations for numbers at age in the standard WHAM model
- Options for configuring recruitment are the same as the standard WHAM package.
- If recruitment is assumed to be a function of SSB, it is only the SSB in the spawning region at the time of spawning.
- For ages $a = 2, \ldots, A-1$, where A is the plus group, The vector of numbers at age a by region in year y is

$$\log\left(\mathbf{N}_{a,y}\right) = \log\left(\mathbf{O}'_{a-1,y-1}\mathbf{N}_{a-1,y-1}\right) + \boldsymbol{\varepsilon}_{a,y}$$

  and for the plus group

$$\log\left(\mathbf{N}_{A,y}\right) = \log\left(\mathbf{O}'_{A-1,y-1}\mathbf{N}_{A-1,y-1} + \mathbf{O}'_{A,y-1}\mathbf{N}_{A,y-1}\right) + \boldsymbol{\varepsilon}_{A,y}.$$

- The region-specific errors $\boldsymbol{\varepsilon}_{a,y}$ are independent, but within regions, abundance at age have the same autoregressive options as the standard WHAM package. Generally, variance parameters can be region and age specific $\boldsymbol{\sigma}_{r,a}$ and age and/or year AR1 correlation parameters can be region-specific $\left(\rho_{\mathcal{A},r}, \rho_{\mathcal{Y},r}\right)$.
- "Survival" transitions can be deterministic $\left(\boldsymbol{\varepsilon}_{a,y} = 0\right)$ as in traditional statistical catch at age models.

Weight and maturity at age

- Weight and maturity at age is treated similarly to the standard WHAM package.
- Annual weight at age matrices for catches of each fleet are not stock-specific
- Same for weight at age for any surveys measured in biomass
- For SSB, weight at age and maturity at age matrices are stock-specific.

Spawning biomass

- Like the standard version of wham, the spawning biomass is a function of survival through the year up to the time of spawning.
- When spawning for a given stock occurs in one region SSB is

$$\mathrm{SSB}_{a,y} = \sum_{a=1}^{A} w_{a,y} m_{a,y} \sum_{r=1}^{n_R} N_{a,r,y} \mathbf{O}_{a,y}(\delta_s, r, r_s)$$

where $\mathbf{O}_{a,y}(\delta_s, r, r_s)$ is the element of the survival and movement matrix from the beginning of the year to the time of spawning that corresponds to starting in region r and being alive in the spawning region r_s at the time of spawning.

Fishing mortality

- Fishing mortality is treated the same as the standard WHAM package
  - F at age for each fleet is estimated as the product of fully selected F and selectivity at age: $F_{f,a,y} = F_{f,y} s_{f,a,y}$
- In standard WHAM a fully-selected total F is also reported as the maximum of the total F at age summed across fleets

$$F_{\text{total},a,y} = \sum_{i=1}^{n_f} F_{f,a,y}$$

$$F_{\text{total},y} = \arg\max_a F_{\text{total},a,y}$$

- In Multi-WHAM a fully-selected total F by summing across fleets is also reported.
  - **But these fleets may occur in different regions**
  - The magnitude of the total F increases with more regions (that have fishing)
  - There are different ways to average across fleets/regions, but the best way is debatable
  - **Most important is that the representation of F in the model is consistent with that in reference points**

Aggregate catch and index observations

- The likelihood equations for catch and index observations are the same as the standard WHAM package.
- The predicted catch is a function of the abundance at age of each stock and probabilities of capture in each fleet.
- The predicted catch biomass at age for fleet f:

$$\widehat{C}_{f,a,y} = \left(\sum_{s=1}^{n_S} \sum_{r=1}^{n_R} \mathbf{H}_{s,a,y}(r,f) N_{s,r,a,y} \right) c_{f,a,y}.$$

  where $\mathbf{H}_{s,a,y}(r,f)$ element of the matrix representing the probability of capture by fleet f over year y for fish of age a beginning the year in region r and $c_{f,a,y}$ is the weight at age and year in fleet f.
- For index i that occurs at δ_i within each year, the predicted index at age is

$$\hat{I}_{a,y,i} = q_i s_{a,y,i} W_{a,y,i} \sum_{s=1}^{n_S} \sum_{r=1}^{n_R} N_{s,r,a,y} \mathbf{O}_{s,a,y}(\delta_i, r, r_i).$$

  where q_i is the catchability, $s_{a,y,i}$ is the selectivitity at age, $W_{a,y,i}$ is the weight at age (if the index is measured in biomass) and $\mathbf{O}_{s,a,y}(\delta_i, r, r_i)$ is the component of the survival and movement matrix over the interval up to the time of the survey that corresponds to starting in region r and is alive in region r_i where the survey occurs.
- Predicted age composition observations are the same functions of the predicted catch or indices at age as the standard WHAM package.