

# Multi-WHAM: A multi-stock, multi-region extension of the Woods Hole Assessment Model

Timothy J. Miller<sup>1</sup>

28 September, 2023

<sup>1</sup>timothy.j.miller@noaa.gov, Northeast Fisheries Science Center, National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543, USA

## Summary

The Woods Hole Assessment Model (WHAM) software package is maintained and developed at the Northeast Fisheries Science Center to enable state-space stock assessments, i.e. where processes such as the annual transitions in numbers-at-age (survival), natural mortality, selectivity, and catchability are treated as time- and/or age-varying random effects. WHAM can also be configured without random effects as a traditional statistical catch-at-age model in order to bridge from current assessments which use Age-Structured Assessment Program (ASAP). Here, we present an extension of WHAM (Multi-WHAM) to model multiple stocks and/or regions and allow movement of stocks among regions.

The new features of Multi-WHAM include

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

Currently, the age and year differences in movement rates are modeled as autoregressive random effects similar to how WHAM models corresponding variation in survival, and  $M$ . Although Multi-WHAM includes several new features, it can still be configured for fitting single-stock, single-region models identically to the current version of WHAM.

## Introduction

The Woods Hole Assessment Model (WHAM) is an R package developed at the Northeast Fisheries Science Center (<https://timjmiller.github.io/wham>, Miller and Stock, 2020; Stock and Miller, 2021). Primary data structures (catch and index observations) are similar to ASAP and can be configured to fit SCAA models nearly identically. There is functionality built into WHAM to migrate ASAP input files to R inputs needed for WHAM, and WHAM uses many of the same types of data inputs, such as empirical weight-at-age, so that existing assessments in the U.S. Northeast can be replicated and tested against models with state-space and environmental effects in a single framework. WHAM primarily differs from ASAP through inclusion of random effects options and implementation via the Template Model Builder package (TMB, Kristensen et al., 2016). In this respect it is similar to the State-space Assessment Model (SAM, Nielsen and Berg, 2014), which is currently used to manage many stocks in the ICES region. WHAM is currently used to manage 5 stocks and likely 4 more after the final report of the Atlantic cod research track assessment. It has been reviewed in the literature and simulation tested (Stock et al., 2021; Stock and Miller, 2021). It has also been used as the operating model in the index-based assessment methods research track assessment (Legault et al., 2023) and is also the main tool for simulation studies in the research track on state-space assessment models.

Here, we describe a multi-stock and multi-region extension of the WHAM package. This extension currently resides in the “lab” branch of the WAM repository (<https://github.com/timjmiller/wham/tree/lab>). The Multi-WHAM package can be installed in R by running

```
devtools::install_github("timjmiller/wham", dependencies=TRUE, ref = "lab",  
  INSTALL_opts=c("--no-multiarch"))
```

## Methods

Many of the equations for the multi-wham version are the same as Stock and Miller (2021) including selectivity options, catch, index and age composition observations, and processes and observations for environmental covariates. Differences from the single-stock version are described below.

### Model description

#### Population

For a single region and stock the numbers at age  $a$  surviving, dying from fishing, and dying from other causes over time interval  $\delta$  are

$$\begin{aligned} N_{L,t+\delta,a} &= N_{t,a} S(\delta, a) = N_{t,a} e^{-Z_{t,a}\delta}, \\ N_{C,t+\delta} &= N_{t,a} H(\delta, a) = N_t \frac{F_{t,a}}{Z_{t,a}} (1 - e^{-Z_{t,a}\delta}), \end{aligned}$$

and

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

In the single-stock version of WHAM, these intervals are  $\delta = 1$  year and  $t$  is January 1 of each year

These outcomes can be defined in vector-matrix form using probability transition matrices. If we define

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

then, for

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

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

Also,

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

When there are  $n_R$  regions and  $n_F$  fleets,

$$\mathbf{N}_{t,a} = (N_{L,t,a,1}, \dots, N_{L,t,a,n_R}, N_{C,t,a,1}, \dots, N_{C,t,a,n_F}, N_{K,t,a})' = (\mathbf{N}'_{L,t,a}, \mathbf{N}'_{C,t,a}, N_{K,t,a})'$$

and

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

Also,

$$\mathbf{N}_{L,t+\delta,a} = \mathbf{O}(\delta, a)' \mathbf{N}_{L,t,a}$$

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

$$N_{K,t+\delta,a} = \mathbf{D}(\delta, a)' \mathbf{N}_{L,t,a}$$

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

is the  $n_R \times n_R$  matrix defining survival and movement from one region to another over the interval.

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

is the  $n_R \times n_F$  matrix defining proportions captured in each fleet given alive in each region at the beginning of the interval. Multi-WHAM currently assumes each fleet operates in a single region.

Multi-WHAM can assume survival and movement are sequential or simultaneous. When sequential survival and death processes occur over the interval and movement is assumed to happen instantly at the end of the interval:

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

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

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

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

$$\boldsymbol{\mu}(\delta, a) = \begin{bmatrix} 1 - \sum_{r' \neq 1} \mu_{1 \rightarrow r'} & \mu_{1 \rightarrow 2} & \cdots & \mu_{1 \rightarrow R} \\ \mu_{2 \rightarrow 1} & 1 - \sum_{r' \neq 2} \mu_{2 \rightarrow r'} & \cdots & \mu_{2 \rightarrow R} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{R \rightarrow 1} & \cdots & \mu_{R \rightarrow R-1} & 1 - \sum_{r' \neq R} \mu_{R \rightarrow r'} \end{bmatrix}$$

When survival and movement are assumed to occur simultaneously, all movement and mortality parameters are instantaneous rates. We obtain the probability transition matrix over an interval  $\delta$  by exponentiating the instantaneous rate matrix (Miller and Andersen, 2008)

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

The instantaneous rate matrix takes rates of movement between regions and the mortality rates for each fleet and region. Along the diagonal is the - sum of the other rates (the hazard). For two regions and two fleets:

$$\mathbf{A}_{s,t,a} = \begin{bmatrix} -(\mu_{1 \rightarrow 2} + F_{t,a,1} + M_{t,a,1}) & \mu_{1 \rightarrow 2} & F_{t,a,1} & 0 & M_{t,a,1} \\ \mu_{2 \rightarrow 1} & -(\mu_{2 \rightarrow 1} + F_{t,a,2} + M_{t,a,2}) & 0 & F_{t,a,2} & M_{t,a,2} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

For example, when there is one region and  $\delta = 1$ , exponentiating the instantaneous rate matrix

$$\mathbf{A}_{s,t,a} = \begin{bmatrix} -Z_{t,a} & F_{t,a} & M_{t,a} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

results in the Baranov equations

$$e^{\mathbf{A}_{s,t,a}\delta} = \begin{bmatrix} e^{-Z_{t,a}} & \frac{F_{t,a}}{Z_{t,a}}(1 - e^{-Z_{t,a}}) & \frac{M_{t,a}}{Z_{t,a}}(1 - e^{-Z_{t,a}}) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

There are at most  $n_R - 1$  parameters for each stock, season, region for either the sequential or simultaneous configurations. Movement parameters are estimated on a transformed scale. If survival and movement are occur simultaneously, the parameters are estimated on log scale and if they are separable, an additive logit link transformation (like a multinomial regression) is used.

Random effects are normally distributed on the transformed scale

$$f(\mu_{s,r,r',t,y,a}) = \theta_{s,r,r',t} + \epsilon_{s,r,r',t,y,a}$$

$$Cov(\epsilon_{s,r,r',t,y,a}, \epsilon_{s,r,r',t,y',a'}) = \frac{\rho_{s,r,r',t,A}^{|a-a'|} \rho_{s,r,r',t,Y}^{|y-y'|} \sigma_{s,r,r',t}^2}{(1 - \rho_{s,r,r',t,A}^2)(1 - \rho_{s,r,r',t,Y}^2)}$$

for moving from region  $r$  to  $r'$

Covariate effects can be stock-, age-, season-, and/or region-to-region-specific:

$$f(\mu_{s,r,r',t,y,a}) = \theta_{s,r,r',t} + \sum_{k=1}^{n_E} \beta_{s,a,t,r,r',k} E_k$$

and the same orthogonal polynomial options in the standard WHAM package are available. Movement parameters can including both random and environmental effects:

$$f(\mu_{s,r,r',t,y,a}) = \theta_{s,r,r',t} + r_{s,r,r',t,y,a} + \sum_{k=1}^{n_E} \beta_{s,r,r',t,k} E_{k,y}$$

There is currently no likelihood component for tagging data. Therefore, it is likely that movement parameters would either need to be fixed or assumed to have some prior distribution, possibly based on external

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

$$\gamma_{s,r,r',t} \sim N(\theta_{s,r,r',t}, \sigma_{s,r,r',t}^2)$$

When priors are used, the movement is defined instead as

$$f(\mu_{s,r,r',t,y,a}) = \gamma_{s,r,r',t} + r_{s,r,r',t,y,a} + \sum_{k=1}^{n_E} \beta_{s,r,r',t,k} E_{k,y}$$

The default setting in Multi-WHAM is no movement for any stocks even when the number of regions is greater than 1. Also, multiple ASAP \*.dat files can be read in simultaneously:

```
asap <- read_asap3_dat(c("file_1.DAT", "file_2.DAT", "file_3.DAT", "file_4.DAT"))
```

which can be used to quickly structure a model with each stock isolated in its own region. This set up can, for many configurations, provide equivalent results to fitting separate models for each stock.

For each stock, recruitment is assumed to occur on January 1 each year only in the user-defined spawning region for that stock. Abundance at older ages is allowed in all regions on January 1. The user also defines which older ages can be where using the `prepare_wham_input` element `basic_info$NAA_where`. Other elements of `move` argument to `prepare_wham_input`, `move$can_move` and `move$must_move`, define if and where fish can move each season.

## Seasons

Seasonality can be configured to accomodate characteristics of spawning, migration, and fleet-specific behavior. The matrix of transition probabilities over the year  $y$ , is a product of seasonal matrices

$$\mathbf{P}_{s,y,a} = \prod_{t=1}^{n_s} \mathbf{P}_{s,a,y,t}$$

where each of the seasonal matrices can be configured depending on assumptions of migration timing and whether fleets are operating.

Each year can be divided into any number of seasons. For example, `basic_info$fracyr_seasons = rep(0.1, 10)` would create 10 intervals of equal length for the year ( $\delta = 1/10$ ).

## Numbers at age

The numbers at age for a given stock are vector analogs of the equations for numbers at age in the standard WHAM model (Stock and Miller, 2021) and options for configuring recruitment are the same. If recruitment is assumed to be a function of SSB, it is only the SSB in the stock-region at the time of spawning. That is, spawning individuals from the stock may be in other regions at the time of spawning, but only those in the region where the stock recruits to will constitute the spawning stock. Stock-recruit options may differ for each stock.

For ages 2 to A-1, where A is the plus group, The vector of numbers at age  $a$  in year  $y$  by region is

$$\log(\mathbf{N}_{s,y,a}) = \log(\mathbf{O}'_{s,y,a-1} \mathbf{N}_{s,y-1,a-1}) + \epsilon_{s,y,a}$$

and for the plus group

$$\log(\mathbf{N}_{s,y,A}) = \log(\mathbf{O}'_{s,y,A-1} \mathbf{N}_{s,y-1,A-1} + \mathbf{O}'_{s,y,A} \mathbf{N}_{s,y-1,A}) + \epsilon_{s,y,A}.$$

The region-specific errors  $\epsilon_{s,y,a}$  are independent with the same variance  $\sigma_{s,a}$ . The same autoregressive options as the standard WHAM package are available for the numbers at age for each stock and region, but, currently, the variance and correlation parameters are the same across regions for a given stock.

## Initial numbers at age

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

- **age-specific**:  $\log N_{s,r,1,a}$  are fixed effects.
- **equilibrium**:  $\log N_{s,r,1,1}$  and  $\log F_s$  are fixed effects defining equilibrium numbers at age in the first year.
- **iid**:  $\log N_{s,r,1,a} \sim N(\eta_{s,r}, \sigma_{s,r}^2)$  are iid random effects.
- **ar1**:  $\log N_{s,r,1,a} \sim N\left(\eta_{s,r}, \frac{\sigma_{s,r}^2}{(1-\rho_{s,r}^2)}\right)$  are AR1 random effects.

The first two options were previously available, but the **age-specific** option will likely be difficult to estimate when there is movement and individuals can be in multiple regions on January 1 due to the number of parameters for abundances at older ages that may not be well observed in isolation from other stocks. The **equilibrium** option is analogous to the option in the standard WHAM model, but is generalized for vectors of abundances by regions and matrices of survival and movement  $\mathbf{O}$  at a given age. The last two options treat the initial abundances as random effects with or without autocorrelation with age.

## Natural mortality

Natural mortality options have been expanded in the Mult-WHAM version. When not estimated, (mean) mortality rates may be stock-, region-, and age-specific. When random effects are allowed, variance and correlation parameters can be stock and region specific. The options for each stock and 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), as in Cadigan (2016).

Any environmental covariate effects can be stock-, region-, and age-specific.

## Weight and Maturity at age

Weight at age is treated similarly to the standard WHAM package. Annual weight at age matrices for each fleet are used to calculate total catch and the weight at age is applied to catch numbers at age for any stocks caught by the fleet. Similarly, when indices and/or associated age composition observations are measured in biomass, the weight at age matrices for the index are applied to predicted numbers at age of all stocks observed by the survey. Unique weight at age and maturity at age matrices are allowed for each stock to calculate spawning stock biomass.

## Reference points

Currently a single F reference point is estimated across stocks and regions. Selectivity is determined as before when there are multiple fleets:  $\tilde{F}_{f,a} = \bar{F} \text{sel}_{f,a}$  where  $\text{sel}_{f,a}$  is determined by averaging F at age over a user-defined set of years

$$\text{sel}_{f,a} = \frac{\bar{F}_{f,a}}{\max \bar{F}_a}$$

As in the standard WHAM package, “static” reference points, typically meant to be defined for prevailing conditions, average all of the inputs to the SSB/R or Y/R calculations over the user-specified years (e.g., last 5 years of the model). This same averaging is also applied to possibly time-varying movement parameters.

For  $X\%$  SPR-based reference points, we use a Newton method and iterate

$$\tilde{F}^{(i)} = \tilde{F}^{(i-1)} - \frac{g(\tilde{F}^{(i-1)})}{\tilde{F}^{(i-1)} g'(\tilde{F}^{(i-1)})}$$

where  $g(F)$  is the difference between the weighted spawning biomass per recruit at  $F$  and  $X\%$  of the weighted unfished spawning biomass per recruit:

$$g(F) = \sum_{s=1}^S w_s \left( \text{SSB/R}(F) - \frac{X}{100} \text{SSB/R}(F=0) \right)$$

A similar method is used to find  $F$  that maximizes yield for MSY-based reference points.

$F$  reference points by stock and/or region are not currently implemented because a unique solution is not guaranteed when there is movement.

## Projections

The projection options for Multi-WHAM are the same as those for the standard WHAM package. When there is movement of any stocks, the user has the option to project and use any random effects for time-varying movement or use the average over user specified years, analogous to how natural mortality can be treated in the projection period. The projections of any environmental covariates has been revised to better include error in the estimated latent covariate in any effects on the population.

## References

- Cadigan, N.G., 2016. A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. *Canadian Journal of Fisheries and Aquatic Sciences* 73, 296–308. <https://doi.org/10.1139/cjfas-2015-0047>
- Kristensen, K., Nielsen, A., Berg, C., Skaug, H., Bell, B.M., 2016. TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software* 70, 1–21. <https://doi.org/10.18637/jss.v070.i05>
- Legault, C.M., Wiedenmann, J., Deroba, J.J., Fay, G., Miller, T.J., Brooks, E.N., Bell, R.J., Langan, J., Cournane, J., Jones, A.W., Muffley, B., 2023. Data rich but model resistant: An evaluation of data-limited methods to manage fisheries with failed age-based stock assessments. *Canadian Journal of Fisheries and Aquatic Sciences* 80, 27–42. <https://doi.org/10.1139/cjfas-2022-0045>
- Miller, T.J., Andersen, P.K., 2008. A finite-state continuous-time approach for inferring regional migration and mortality rates from archival tagging and conventional tag-recovery experiments. *Biometrics* 64, 1196–1206. <https://doi.org/10.1111/j.1541-0420.2008.00996.x>
- Miller, T.J., Stock, B.C., 2020. The Woods Hole Assessment Model (WHAM).
- Nielsen, A., Berg, C.W., 2014. Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research* 158, 96–101. <https://doi.org/10.1016/j.fishres.2014.01.014>
- Stock, B.C., Miller, T.J., 2021. The Woods Hole Assessment Model (WHAM): A general state-space assessment framework that incorporates time- and age-varying processes via random effects and links to environmental covariates. *Fisheries Research* 240, 105967. <https://doi.org/10.1016/j.fishres.2021.105967>
- Stock, B.C., Xu, H., Miller, T.J., Thorson, J.T., Nye, J.A., 2021. Implementing two-dimensional autocorrelation in either survival or natural mortality improves a state-space assessment model for Southern New England-Mid Atlantic yellowtail flounder. *Fisheries Research* 237, 105873. <https://doi.org/10.1016/j.fishres.2021.105873>