Woods Hole Assessment Model: ‘growth’ branch

In this document I only describe the new features implemented in WHAM ('growth’ branch). To see all the base equations, please see Stock and Miller (2021) and references therein.

The new WHAM code can be found here: https://github.com/gmoroncorrea/wham/tree/growth

1. Data

Length compositions

Length compositions are fishery or survey-specific.

$$p_{y,l} = \frac{n_{y,l}}{\sum_l n_{y,l}}$$

$n_{y,l}$ is the number of fish in year y and length bin l.

Conditional age-at-length

Age compositions are conditional on a length bin and can be fishery or survey-specific:

$$p_{y,l,a} = \frac{n_{y,l,a}}{\sum_a n_{y,l,a}}$$

$n_{y,l,a}$ is the number of fish in year y, age a, and length bin l

Input age-length transition matrix

Replaces the age-length transition matrix calculated internally from growth parameters (see below). This matrix should be provided by the user for multiple times within a year (e.g. January 1^{st}, July 1^{st}, and other dates based on fishery and survey timing). This matrix is unique for the entire assessment period.

2. Population model

Mean length-at-age

$\tilde{L}_{y,a}$ is the mean length at age a at the start of the year (January 1st) and can be calculated in two ways.

1. Growth equation (parametric approach): For the first year ($y = 1$), $\tilde{L}_{y,a}$ can be calculated using the classic von Bertalanffy growth equation (Schnute, 1981):

$$\tilde{L}_{y,a} = L_\infty + (L_1 - L_\infty)\exp(-k(a-1))$$

Where L_∞ is the asymptotic length, k is the growth rate, and L_1 is the length at age 1.

For $y > 1$, $\tilde{L}_{y,a}$ is calculated as:

$$\tilde{L}_{y,a} = \begin{cases} L_1 & for\ a = 1 \\ \tilde{L}_{y-1,a-1} + \left(\tilde{L}_{y-1,a-1} - L_\infty\right)(\exp(-k) - 1) & for\ a > 1 \end{cases}$$

Deviations in growth parameters can be estimated as random effects (δ) and are assumed to have a normal distribution with zero mean but autocorrelated over years (y) or cohorts (c):

$$log\left(L_{\infty_t}\right) = \mu_{L_\infty} + \delta_{1,t}$$

$$log(k_t) = \mu_k + \delta_{2,t}$$

$$log\left(L_{1_t}\right) = \mu_{L_1} + \delta_{3,t}$$

$$log\left(CV_{1_t}\right) = \mu_{CV_1} + \delta_{4,t}$$

$$log\left(CV_{A_t}\right) = \mu_{CV_A} + \delta_{5,t}$$

where CV_1 and CV_A are the coefficient of variation of lengths at age 1 and A (age plus group), respectively, and t represents years or cohorts. Then:

$$\delta_t|\delta_{t-1} \sim N(\varphi\delta_{t-1}, \sigma_G^2), \text{ so:}$$

$$Cov(\delta_t, \delta_{\tilde{t}}) = \frac{\sigma_G^2 \varphi^{|t-\tilde{t}|}}{(1-\varphi^2)}$$

where σ_G^2 and φ are the AR(1) variance and correlation coefficient over years or cohorts, respectively.

To calculate the mean length-at-age a at any fraction of year y ($L_{y,a}$), we use:

$$L_{y,a} = \tilde{L}_{y,a} + \left(\tilde{L}_{y,a} - L_\infty\right)\left(\exp\left(-kf_y\right) - 1\right)$$

where f_y represents the fraction of the year [0,1].

2. Mean length-at-age (LAA, non-parametric approach): The user provides a vector with mean length-at-age $\tilde{L}_a$ (parameters) for January 1st. Deviations in $\tilde{L}_a$ as random effects can be estimated:

$$log\left(\tilde{L}_{y,a}\right) = \mu_{\tilde{L}_a} + \delta_{a,y}$$

$$E \sim MVN(0, \Sigma)$$

where $E = \left(\delta_{1,1}, \dots, \delta_{1,Y}, \delta_{2,1}, \dots, \delta_{2,Y}, \dots, \delta_{A,1}, \dots, \delta_{A,Y}\right)'$, Y is the number of years. The covariance matrix Σ is defined by:

$$Cov\left(\delta_{a,y}, \delta_{\tilde{a},\tilde{y}}\right) = \frac{\sigma_L^2 \varphi_{age}^{|a-\tilde{a}|} \varphi_{year}^{|y-\tilde{y}|}}{(1-\varphi_{age}^2)(1-\varphi_{year}^2)}$$

Where σ_L^2, φ_{age}, and φ_{year} are the AR(1) variance and correlation coefficients in age and year, respectively.

To calculate the mean length-at-age a at any fraction of year y ($L_{y,a}$), we use a linear interpolation between $\tilde{L}_{y,a}$ and $\tilde{L}_{y+1,a+1}$:

$$L_{y,a} = \tilde{L}_{y,a} + (\tilde{L}_{y+1,a+1} - \tilde{L}_{y,a})f_y$$

Age-length transition matrix

Using $L_{y,a}$ information, the proportion in length bin l for age a during year y, $\varphi_{y,l,a}$, is calculated as:

$$\varphi_{y,l,a} = \begin{cases} \Phi\left(\frac{L'_{min} - L_{y,a}}{\sigma_{y,a}}\right) & for\ l = 1 \\ \Phi\left(\frac{L'_{l+1} - L_{y,a}}{\sigma_{y,a}}\right) - \Phi\left(\frac{L'_{l} - L_{y,a}}{\sigma_{y,a}}\right) & for\ 1 < l < n_L \\ 1 - \Phi\left(\frac{L'_{max} - L_{y,a}}{\sigma_{y,a}}\right) & for\ l = n_L \end{cases}$$

where Φ is the standard normal cumulative density function, L'_l is the lower limit of length bin l, L'_{min} is the upper limit of the smallest length bin, L'_{max} is the lower limit of the largest length bin, n_L is the index of largest length bin, and $\sigma_{y,a}$ is the standard deviation of lengths at age a and it is calculated as:

$$\sigma_{y,a} = \left(CV_1 + \frac{(CV_A - CV_1)}{A - 1}.a\right).\tilde{L}_{y,a}$$

The user also has the option to include an input $\varphi_{l,a}$ matrix, so growth parameters or mean length-at-age estimates are not used.

Numbers-at-age and length are calculated:

$$N_{y,l,a} = \varphi_{y,l,a}N_{y,a}$$

Mean weight-at-age

1. Length-weight relationship

If empirical weight-at-age is not provided, the user can opt for a length-weight relationship:

$$w_l = \Omega_1 l^{\Omega_2}$$

Ω_1 and Ω_2 are weight coefficient and exponent, respectively.

Random effects can also be estimated by year or cohorts:

$$log(\Omega_{1_t}) = \mu_{\Omega_1} + \delta_{1,t}$$

$$log(\Omega_{2_t}) = \mu_{\Omega_2} + \delta_{2,t}$$

$$\delta_t | \delta_{t-1} \sim N(\varphi\delta_{t-1}, \sigma_W^2)$$

$$Cov(\delta_t, \delta_{\tilde{t}}) = \frac{\sigma_W^2 \varphi^{|t-\tilde{t}|}}{(1 - \varphi^2)}$$

where σ_W^2 and φ are the AR(1) variance and correlation coefficients over years or cohorts (t), respectively.

2. Mean weight-at-age (WAA, non-parametric approach):

This approach is like the LAA non-parametric approach. The user provides a vector with mean weight-at-age $\widetilde{W}_a$ (parameters) for January 1st. Deviations in $\widetilde{W}_a$ as random effects can be estimated:

$$log(\widetilde{W}_{y,a}) = \mu_{\widetilde{W}_a} + \delta_{a,y}$$

See the LAA non-parametric approach for details on random effects and $\widetilde{W}_{y,a}$ variation within a year.

Selectivity-at-length

In addition to selectivity-at-age functions already available in WHAM, I incorporated some selectivity-at-length functions:

1. Logistic function

$$S_{y,l} = \frac{1}{1 + \exp\left(-\frac{l - \beta_{1_y}}{\beta_{2_y}}\right)}$$

$$\beta_{1_y} = \gamma_{\beta_1} + \frac{u_{\beta_1} - \gamma_{\beta_1}}{1 + \exp\left(-(\upsilon_1 + \zeta_{1,y})\right)}$$

$$\beta_{2_y} = \gamma_{\beta_2} + \frac{u_{\beta_2} - \gamma_{\beta_2}}{1 + \exp\left(-(\upsilon_2 + \zeta_{2,y})\right)}$$

Where β_1 is the length-at-50%-selectivity and β_2 controls the width of the ascending slope. υ_1 is the logit-scale mean β_1 parameter with lower and upper bounds γ_{β_1} and u_{β_1}, respectively. υ_2 is the logit-scale mean β_2 parameter with lower and upper bounds γ_{β_2} and u_{β_2}, respectively. Random effects:

$$Cov(\zeta_{1,y}, \zeta_{2,\tilde{y}}) = \frac{\sigma_s^2 \phi_{par} \phi_{year}^{|y-\tilde{y}|}}{(1 - \phi_{par}^2)(1 - \phi_{year}^2)}$$

where σ_s^2 is the AR(1) variance, and ϕ_{par} and ϕ_{year} are the AR(1) correlation coefficients by parameter and year.

2. Double normal function

The six-parameter version of the double normal selectivity function has three components connected by steep logistic "joiners" to provide overall differentiability (see Methot and Wetzel, 2013), i.e.:

$$S_l = \alpha_l(1 - j_{1,l}) + j_{1,l}((1 - j_{2,l}) + j_{2,l}\beta_l)$$

where α_l, β_l, $j_{1,l}$, and $j_{2,l}$ are the ascending, descending and (two) joiner functions, i.e.:

$$\alpha_l = p_5 + (1-p_5)\left(e^{-(l-p_1)^2/e^{p_3}} - e^{-(L_{min}-p_1)^2/e^{p_3}}\right)/(1-e^{-(L_{min}-p_1)^2/e^{p_3}})$$
$$\beta_l = 1 + (p_6-1)\left(e^{-(l-\gamma)^2/e^{p_4}} - 1\right)/(e^{-(L_{max}-\gamma)^2/e^{p_4}} - 1)$$
$$j_{1,l} = \left(1 + e^{-20(l-p_1)/(1+|l-p_1|)}\right)^{-1}$$
$$j_{2,l} = \left(1 + e^{-20(l-\gamma)/(1+|l-\gamma|)}\right)^{-1}$$

where p_1 is the length at which selectivity=1 starts, γ is the length at which selectivity=1 ends,

$$\gamma = p_1 + L_{bw} + \left(\frac{0.99L_{max} - p_1 - L_{bw}}{1 + e^{-p_2}}\right)$$

L_{bw} is the length bin width. p_2 determines the length at which selectivity=1 ends (the width of the "top", γ is the endpoint), p_3 determines the slope of the ascending section, p_4 determines the slope of the descending section, p_5 is the selectivity at the minimum length (parameterized in logit-space), and p_6 is the selectivity at the maximum length (also parameterized in logit-space). The double normal can mimic an asymptotic selectivity function by setting $p_6 = 1$.

3. Decreasing logistic function

$$S_{y,l} = \frac{1}{1 + \exp\left(\frac{l - \beta_{1_y}}{\beta_{2_y}}\right)}$$

Where β_1 is the decreasing length-at-50%-selectivity and β_2 controls the width of the decreasing slope. β_1, and β_2 are estimated like the logistic function as well as the random effects.

If selectivity-at-length is the primary selectivity function specified by the user, selectivity-at-age is internally calculated by using the transition matrix:

$$S_{y,a} = \sum_l \varphi_{y,l,a} S_{y,l}$$

So $S_{y,a}$ becomes the primary selectivity function internally (e.g. to calculate $N_{y,a}$, $Z_{y,a}$, etc).

3. Environmental effects

An environmental covariate (unobserved values: X_y) can be added and linked to any growth or length-weight parameter like it was already implemented for natural mortality.

There are two options for the process model: a normal random walk and AR(1). For random walk:

$$X_{y+1}|X_y \sim N(X_y, \sigma_X^2)$$

where σ_x^2 is the process variance and X_1 is estimated as a fixed effect parameter. For AR(1):

$$X_1 \sim N(\mu_X, \frac{\sigma_X^2}{1-\phi_X^2})$$

$$X_y \sim N(\mu_X(1-\phi_X) + \phi_X X_{y-1}, \sigma_X^2)$$

where μ_X, σ_X^2, and $|\phi_X| < 1$ are the marginal mean, conditional variance, and autocorrelation parameters.

The environmental covariate observations, x_y, are assumed to be normally distributed with mean X_y and variance $\sigma_{x_y}^2$:

$$x_y | X_y \sim N(X_y, \sigma_{x_y}^2)$$

The environmental covariate affects a parameter through a linear or polynomial equation (see Stock and Miller 2021). The linear formulation is:

$$P_y = P * \exp{(\beta_1 X_y)}$$

where P represents the base parameter (which can be also affected by random effects). Other links (polynomials) are also available.

4. Observation model

Length compositions

Once obtained the catch-at-age (for fisheries) ($C_{y,a}$) or index-at-age (for surveys) ($I_{y,a}$), we estimate the expected catch-at-length:

$$\hat{p}_{y,l} = \frac{\sum_a \varphi_{y,l,a} C_{y,a}}{\sum_l \sum_a \varphi_{y,l,a} C_{y,a}}$$

Or expected index-at-length:

$$\hat{p}_{y,l} = \frac{\sum_a \varphi_{y,l,a} I_{y,a}}{\sum_l \sum_a \varphi_{y,l,a} I_{y,a}}$$

Here, $\varphi_{y,l,a}$ is the age-length transition matrix calculated using mean length-at-age at a specific fraction of a year ($L_{y,a}$).

Conditional age-at-length

For fisheries:

$$\hat{p}_{y,l,a} = \frac{\varphi_{y,l,a} C_{y,a}}{\sum_a \varphi_{y,l,a} C_{y,a}}$$

For indices:

$$\hat{p}_{y,l,a} = \frac{\varphi_{y,l,a} I_{y,a}}{\sum_a \varphi_{y,l,a} I_{y,a}}$$

Weight-at-age

The predicted population weight-at-age is calculated from the length-weight relationship using the age-length transition matrix at any fraction of the year:

$$\hat{w}_{y,a} = \sum_l \varphi_{y,l,a} w_l$$

Or if the WAA non-parametric approach is used, then:

$$\hat{w}_{y,a} = W_{y,a}$$

5. Statistical model

Length compositions

Multinomial:

$$-log\mathcal{L}(\hat{p}|p, Neff) = \Gamma(N_{eff} + 1) - \sum_l \Gamma(p_l N_{eff} + 1) + \sum_l p_l N_{eff} \log(\hat{p}_l)$$

Where Γ represents the gamma function and N_{eff} the effective sample size. Other likelihood functions are also available (e.g. Dirichlet) and are like those for age compositions (see Stock and Miller, 2021).

Conditional age-at-length

I use the same likelihood functions available for age compositions but restricted to a length bin. For example, for length bin l and year:

$$-log\mathcal{L}(\hat{p}_{y,l,a}|p_{y,l,a}, Neff_{y,l}) = \Gamma\left(N_{eff_{y,l}} + 1\right) - \sum_l \Gamma\left(p_{y,l,a} N_{eff_{y,l}} + 1\right) + \sum_l p_{y,l,a} N_{eff_{y,l}} \log(\hat{p}_{y,l,a})$$

$N_{eff_{y,l}} > 0$ is the effective sample size for length bin l during year y. So, the contribution to the total likelihood is the sum over years and length bins.

Weight-at-age

This is used when the empirical weight-at-age is treated as observed weight-at-age ($w_{y,a}$). The likelihood function is:

$$-log\mathcal{L}(\hat{w}_{y,a}|w_{y,a}, n_{y,a}) = \log(CV_{y,a} * \hat{w}_{y,a}) + 0.5\left[\frac{w_{y,a} - \hat{w}_{y,a}}{CV_{y,a} * \hat{w}_{y,a}}\right]^2$$

$n_{y,a} > 0$ is the sample size for age a and year y, and $CV_{y,a}$ can be approximated by $\frac{s_{y,a}}{\sqrt{n_{y,a}}\bar{w}_{y,a}}$. The contribution to the total likelihood is the sum over years and ages.