

# Gmacs

A generalized size-structured stock assessment model

The Gmacs Development Team

July 2, 2015

# Outline

- 1 Notation and some definitions
- 2 Growth
- 3 Natural mortality and survival
- 4 Selectivity, retention, fishing
- 5 Recruitment
- 6 Population dynamics
- 7 Data and likelihoods

# Notation

## Generally

- a bold capital symbol  $\mathbf{A}$  refers to a matrix
- a bold lowercase symbol  $\mathbf{a}$  refers to a vector
- an unbolded italic symbol  $a$  refers to a scalar
- $\{a_i\}_{i=1}^n$  is an ordered  $n$ -tuple
- the terms  $p(\cdot)$  or  $\pi(\cdot)$  represent probability distributions
- $a|b$  means event  $a$  conditional on event  $b$  having occurred
- the symbol  $\forall$  means for all values, usually referring to all of the values within an ordered tuple.
- we use red to indicate an estimable parameter
- we use blue to represent covariates and fixed parameters
- we use green to represent data

# Indices

Symbol	Description
$g$	group
$h$	sex
$i$	year
$j$	time step (years)
$k$	gear or fleet
$\ell$	index for size class
$m$	index for maturity state
$o$	index for shell condition

Notice no area index.

# Leading model parameters

Symbol	Support	Description
$M_0$	$0 < M_0 < \infty$	Initial instantaneous natural mortality rate
$R_0$	$0 < R_0 < \infty$	Unfished average recruitment
$\ddot{R}$	$0 < \ddot{R} < \infty$	Initial recruitment
$\bar{R}$	$0 < \bar{R} < \infty$	Average recruitment
$\alpha_r$	$0 < \alpha_r < \infty$	Mode of size-at-recruitment
$\beta_r$	$0 < \beta_r < \infty$	Shape parameter for size-at-recruitment
$\rho$	$-\infty < \rho < \infty$	Recruitment autocorrelation

We group the leading model parameters into the vector

$$\boldsymbol{\theta} = \{M_0, R_0, \ddot{R}, \bar{R}, \alpha_r, \beta_r, \rho\}.$$

# Cubic splines

A spline is a numeric function that is piecewise-defined by polynomial functions, and which possesses a sufficiently high degree of smoothness at the places where the polynomial pieces connect (which are known as knots). A cubic spline is constructed of piecewise third-order polynomials. Cubic splines can be defined by a simple tridiagonal system which can be solved easily to give the coefficients of the polynomials.

# Growth parameters

Symbol	Support	Description
$\alpha_h$	$\alpha_h > 0$	Growth intercept
$\beta_h$	$\beta_h > 0$	Growth slope
$\varphi_h$	$\varphi_h > 0$	Growth scale
$\mu_h$	$\mu_h > 0$	Length at 50% molting probability
$c_h$	$c_h > 0$	Coefficient of variation of molting probability

We group the growth parameters into the vector

$$\boldsymbol{\psi} = \{\alpha_h, \beta_h, \varphi_h, \mu_h, c_h\}.$$

Growth involves:

- average molt increment from size class  $\ell$  to  $\ell'$
- probability of transitioning from size class  $\ell$  to  $\ell'$
- molting probability
- size transition probability

## Growth matrix

The average molt increment from size class  $\ell$  to  $\ell'$  is assumed to be sex-specific and is defined by the linear function

$$a_{h,\ell} = \frac{\alpha_h + \beta_h \ell}{\varphi_h}.$$

The probability of transitioning from size class  $\ell$  to  $\ell'$  assumes that variation in molt increments follows a gamma distribution

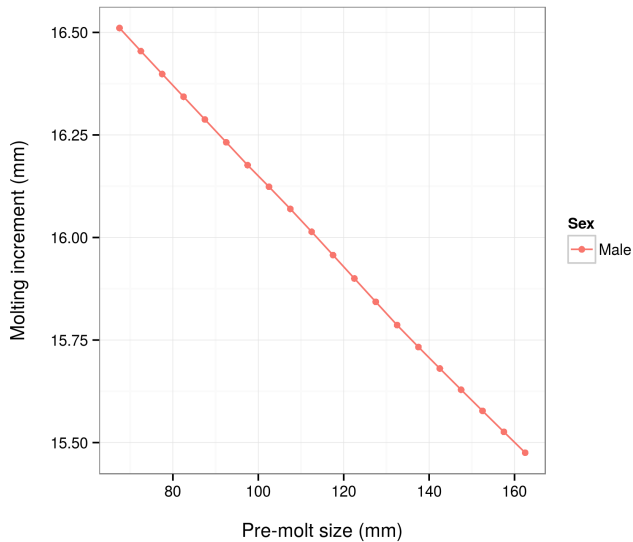
$$p(\ell'|\ell)_h = \mathbf{G}_h = \int_{\ell}^{\ell+\Delta\ell} \frac{\ell^{a_{h,\ell}-1} \exp\left(-\frac{\ell}{\varphi_h}\right)}{\Gamma(a_{h,\ell}) \ell^{a_{h,\ell}}} \quad \text{where} \quad \Delta\ell = \ell' - \ell.$$

Specifically

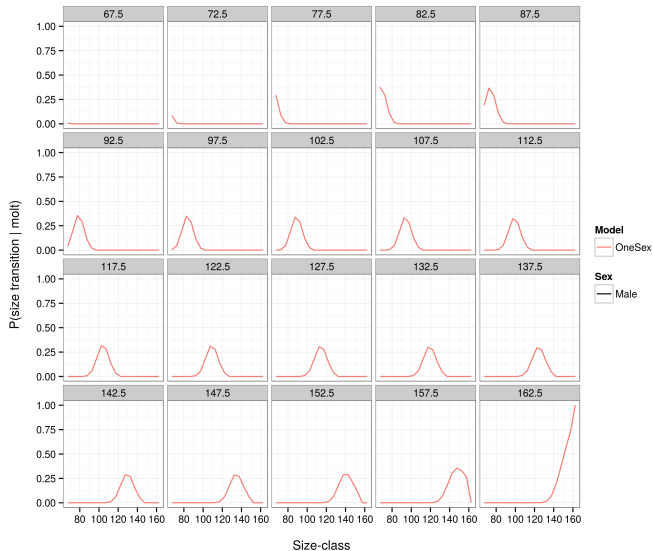
$$\mathbf{G} = G_{\ell,\ell'} = \begin{pmatrix} G_{1,1} & G_{1,2} & \dots & G_{1,n} \\ G_{2,1} & G_{2,2} & \dots & G_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ G_{n,1} & G_{n,2} & \dots & G_{n,n} \end{pmatrix} \quad \text{where} \quad \sum_{\ell'} G_{\ell,\ell'} = 1 \quad \forall \ell.$$



# Growth increments ( $a_{h,\ell}$ )



# Growth transitions ( $G_h$ )



## Molting probability ( $P_h$ )

The standard deviation of molting probability ( $\sigma_h$ ) is calculated from the length at 50% molting probability ( $\mu_h$ ) coefficient of variation of molting probability ( $c_h$ ) as

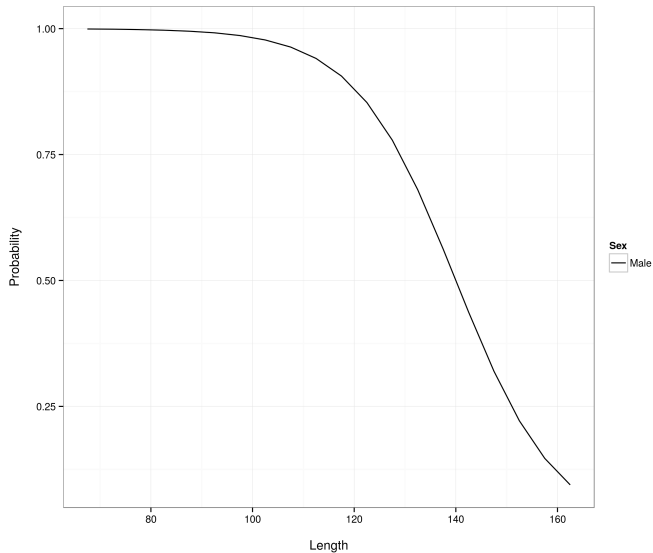
$$\sigma_h = \mu_h c_h.$$

The molting probability ( $P_h$ ) is calculated as

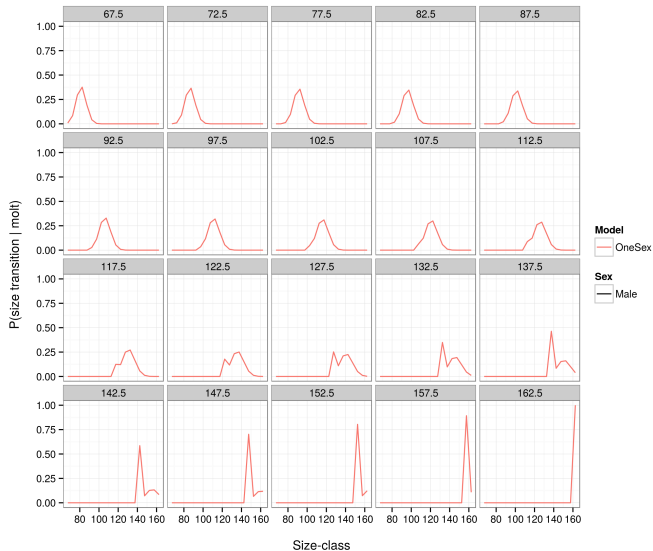
$$P_h = 1 + (-1 - \exp(\mu_h - \ell)/\sigma_h)^{-1}.$$

The molting probability ( $P_h$ ) and the growth probability ( $G_h$ ) are combined to yield the size transition matrix ( $P_h G_h$ ).

# Molting probability ( $P_h$ )



# Size transitions ( $P_h G_h$ )



# Natural mortality variables

Symbol	Description
$M_{0,h}$	Initial instantaneous natural mortality rate
$\sigma_M$	Standard deviation of natural mortality
$\delta_i$	Natural mortality deviate
$M_{h,i}$	Natural mortality by sex $h$ and year $i$

# Natural mortality options

Natural mortality ( $M$ ) is assumed to be sex-specific ( $h$ ), size-independent ( $\ell$ ), and may or may not be constant over time ( $i$ ).

The options currently available in Gmacs include:

- 1 Constant natural mortality ( $M_{h,i} = M_{0,h}$ )
- 2 Random walk (deviates constrained by variance  $\sigma_M^2$ )
- 3 Cubic Spline (deviates constrained by nodes and node placement)
- 4 Blocked changes (deviates constrained by variance in specified blocks  $\iota \in i$ )

## Natural mortality: option 2

If time-varying natural mortality is specified using the **random walk** option, the model constrains  $M_{h,i}$  to be a random-walk process with variance  $\sigma_M^2$

$$M_{h,i+1} = \begin{cases} M_{0,h} & \text{for } i = 1 \\ M_{h,i} e^{\delta_i} & \text{for } i > 1 \end{cases},$$

where

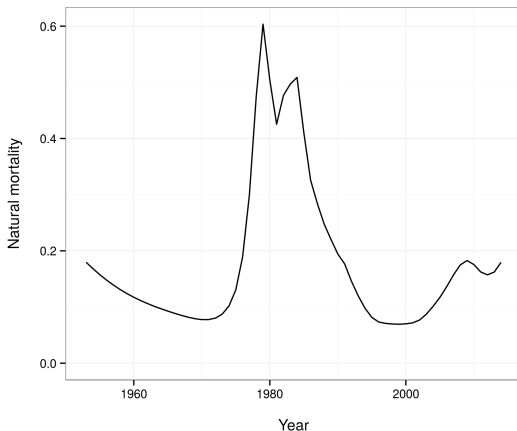
$$\delta_i \sim \mathcal{N}(0, \sigma_M^2).$$

A time-varying natural mortality can be estimated for all years ( $i$ ), or for specified blocks of years ( $\iota \in i$ ).



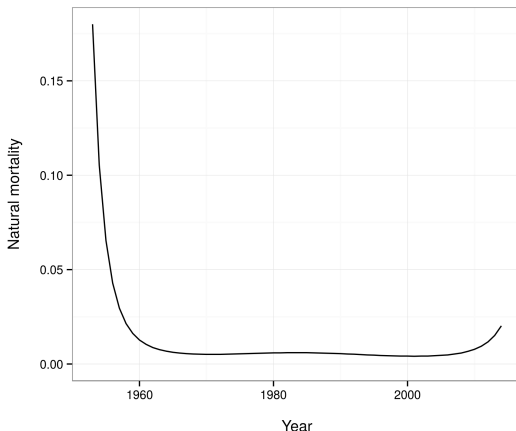
## Natural mortality: option 2

Below we present an example in which time-varying natural mortality is estimated as a **random walk** process for all years (*i*)



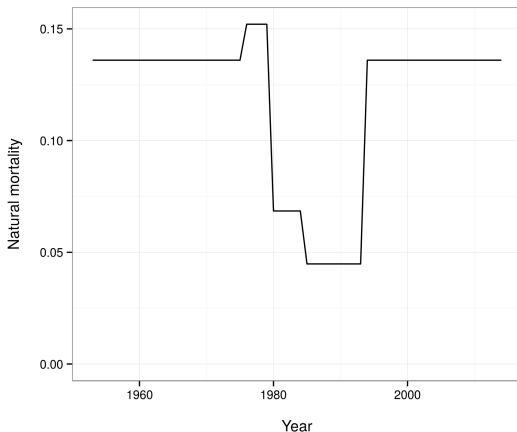
## Natural mortality: option 3

If time-varying natural mortality is specified using the **cubic spline** option, the model constrains  $M_{h,i}$  to be a cubic spline process at specified knots. For example



## Natural mortality: option 4

If time-varying natural mortality is specified using the **blocked changes** option, the model constrains  $M_{h,i}$  by the variance ( $\sigma_M^2$ ). For example, setting  $\sigma_M^2 = 0.04$  and four specific years (1976, 1980, 1985, 1994) we get



# Selectivity, retention and fishing mortality

Symbol	Dimensions	Description
$a_{h,i,k}$	1	Length at 50% selectivity
$\sigma_{h,i,k}^s$	1	Standard deviation in length at selectivity
$s_{h,i,k}$	$\ell \times 1$	Length at 50% selectivity in length interval $\ell$
$r_{h,i,k}$	1	Length at 50%
$\sigma_{h,i,k}^y$	1	Standard deviation in length at retention
$y_{h,i,k}$	$\ell \times 1$	Length at 50% retention in length interval $\ell$
$\xi_{i,k}$	1	Discard mortality rate for gear $k$ in year $i$
$\nu_{h,i,k}$	$\ell \times 1$	Vulnerability due to fishing mortality for sex $h$
$\bar{f}_k$	$i \times 1$	Average fishing mortality rate for gear $k$
$\Psi_{i,k}$	1	Fishing mortality deviate for gear $k$ in year $i$
$F_{i,k}$	1	Fishing mortality rate for gear $k$ in year $i$

# Selectivity and retention

The probability of catching an animal of sex  $h$ , in year  $i$ , in fishery  $k$ , of length  $\ell$  (i.e. selectivity) is

$$s_{h,i,k} = \left(1 + \exp\left(-(\ell - a_{h,i,k})/\sigma_{h,i,k}^s\right)\right)^{-1}.$$

The probability of an animal of sex  $h$ , in year  $i$ , in fishery  $k$ , of length  $\ell$  being retained is

$$y_{h,i,k} = \left(1 + \exp\left(-(r_{h,i,k} - \ell)/\sigma_{h,i,k}^y\right)\right)^{-1}.$$

# Selectivity, retention and fishing mortality

The joint probability of vulnerability due to fishing and discard mortality is

$$\nu_{h,i,k} = s_{h,i,k} [\mathbf{y}_{h,i,k} + (1 - \mathbf{y}_{h,i,k})\xi_{i,k}],$$

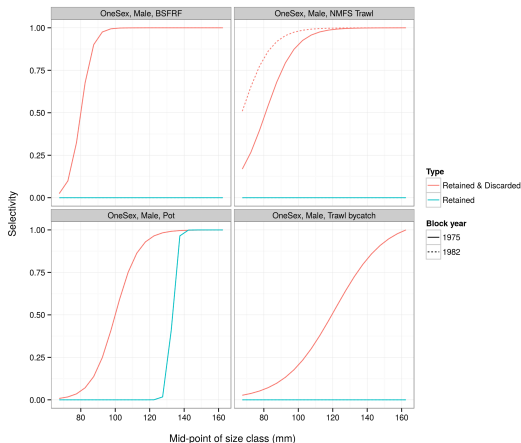
where  $\xi_{i,k}$  is the discard mortality rate for fishery  $k$  in year  $i$ . Finally the fishing mortality is calculated as

$$\mathbf{F}_{h,i} = \sum_k \exp(\bar{\mathbf{f}}_k + \Psi_{i,k}) \nu_{h,i,k},$$

The vector  $\mathbf{F}_{h,i}$  represents all mortality associated with fishing, including discards in directed and non-directed fisheries.

# Selectivity and retention

Assuming that selectivity for the NMFS trawl fishery is split into two blocks (1975-1981 and 1982-2014) and that retention is constant with time  $y_{h,i,k} = y_{h,k}$

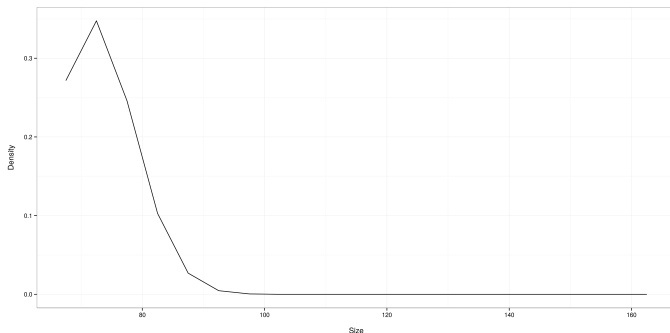


# Recruitment

## Recruitment size-distribution

$$\alpha = \frac{\alpha_r}{\beta_r},$$

$$p[x_\ell - 0.5\Delta x \leq x \leq x_\ell + 0.5\Delta x] = p[x] = \int_{x_\ell - 0.5\Delta x}^{x_\ell + 0.5\Delta x} \frac{x^{\alpha-1} \exp\left(\frac{x}{\beta_r}\right)}{\Gamma(\alpha)x^\alpha} dx.$$





# Recruitment

Initial recruitment

$$r_{h,i} = 0.5p[x]\bar{R} \quad \text{for } i = 1.$$

Recruitment

$$r_{h,i} = 0.5p[x]\bar{R}e^{\delta_i} \quad \text{for } i > 1,$$

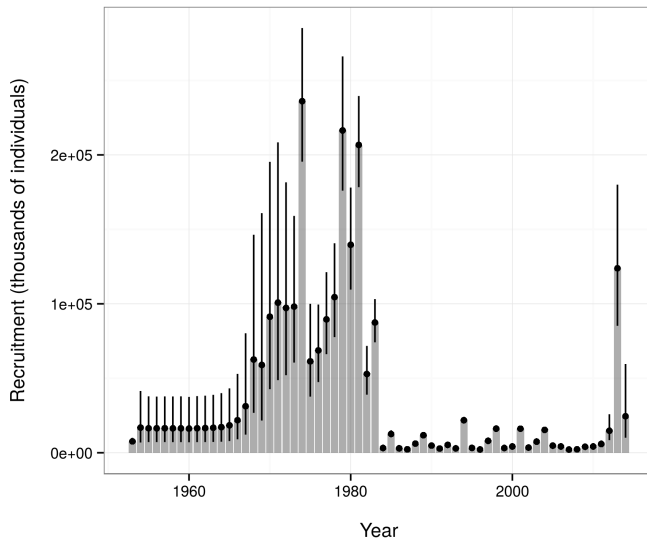
where

$$\delta_i = \log(r_i) - (1 - \rho) \log(\bar{R}) - \rho \log(r_{i-1}) + 0.5\sigma_R^2$$

and

$$r_i = \sum_h r_{h,i}.$$

# Recruitment



# Growth and survival

Growth and survival process are combined, represented by

$$\mathbf{A}_{h,i} = \begin{cases} \mathbf{G}_h [\exp(-\mathbf{M}_{h,i})\mathbf{I}] & \text{for } i = 1 \\ \mathbf{G}_h [\exp(-\mathbf{Z}_{h,i})\mathbf{I}] & \text{for } i > 1 \end{cases}.$$

where

$$\mathbf{Z}_{h,i} = \mathbf{M}_{h,i} + \mathbf{F}_{h,i}.$$

Growth and survivorship in unfished and fished conditions is given by the solution to the equation

$$\mathbf{u}_{h,i} = -(\mathbf{A}_{h,i} - \mathbf{I})^{-1}(p[x]) \quad \forall i.$$

The vector  $\mathbf{u}_{h,i}$  represent the unique equilibrium solution for the numbers per recruit in each size category.

## Initial population

The mean weight at length ( $\ell$ ) by sex ( $h$ ) is represented by the  $\ell \times 1$  vector  $\mathbf{w}_h$  and can take any form the user wishes

$$\mathbf{w}_h = f_w(\ell, \theta)$$

Similarly, the average proportion mature at length ( $\ell$ ) by sex ( $h$ ) is represented by the  $\ell \times 1$  vector  $\mathbf{m}_h$  and can take any form the user wishes

$$\mathbf{m}_h = f_m(\ell, \theta)$$

Steady-state conditions

$$B_0 = \textcolor{red}{R}_0 \sum_h \lambda_h \sum_{\ell} \mathbf{u}_{h,i} \mathbf{w}_h \mathbf{m}_h \quad \text{for } i = 1,$$

$$\tilde{B} = \tilde{R} \sum_h \lambda_h \sum_{\ell} \mathbf{u}_{h,i} \mathbf{w}_h \mathbf{m}_h \quad \text{for } i > 1.$$

# Population evolution

The total unfished numbers in each size category is defined as  $R_0 \mathbf{u}_{h,i=1}$ . Initial numbers at length

$$\mathbf{n}_{h,i} = \left[ -(\mathbf{A}_h - \mathbf{I})^{-1} \mathbf{r}_{h,i} \right] e^{\epsilon} \quad \text{for } i = 1,$$

where  $\epsilon$  is an  $\ell \times 1$  vector of initial recruitment deviates. The numbers in each size-class in the following time-step ( $\mathbf{n}_{h,i+1}$ ) is the product of the numbers in each size-class in the previous time-step ( $\mathbf{n}_{h,i}$ ), size-specific growth and survival ( $\mathbf{A}_{h,i}$ ), plus new recruits ( $\mathbf{r}_{h,i}$ )

$$\mathbf{n}_{h,i+1} = \mathbf{n}_{h,i} \mathbf{A}_{h,i} + \mathbf{r}_{h,i} \quad \text{where } i \geq 1.$$

# Likelihoods and penalties

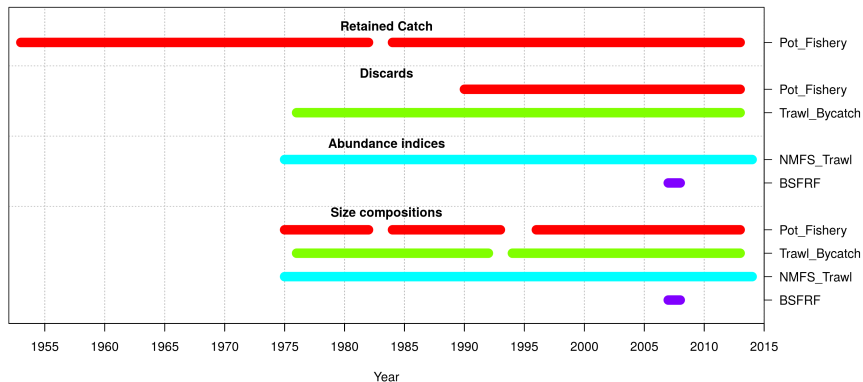
## Likelihoods

- likelihood of catch (log-normal)
- likelihood of relative abundance (weighted log-normal)
- likelihood of size compositions (multinomial, robust multinomial, Dirichlet)
- likelihood of recruitment deviations (log-normal)
- likelihood of growth increment data (log-normal)

## Penalties

- constrain  $\log(\Psi_{i,k})$  to ensure they sum to zero
- constrain mean  $\bar{\mathbf{f}}_k$  to regularize the solution
- constrain  $M_{h,i}$  in random walk (log-normal)

## Data by type and year



## Log-likelihood: catch

The standard deviation of the catch ( $\sigma_{i,k}$ ) is calculated from the CV as

$$\sigma_{i,k} = \sqrt{\log(1 + c_{i,k}^2)}.$$

The expected catch is calculated using the Baranov catch equation

$$\hat{C}_{i,k} = \sum_{\ell} \left[ n_{h,i} w_h \frac{F_{h,i}}{Z_{h,i}} (1 - e^{-Z_{h,i}}) \right] \quad \text{where} \quad Z_{h,i} = M_{h,i} + F_{h,i}.$$

The log-likelihood is

$$\ell(C_{i,k}) = 0.5 \log(2\pi) + \log(\sigma_{i,k}) - \frac{1}{2\sigma_{i,k}^2} (\log(C_{i,k}) - \log(\hat{C}_{i,k}))^2.$$



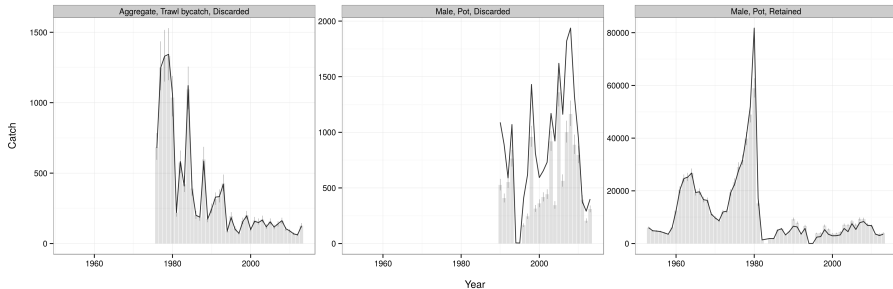
## Log-likelihood: catch

```
for(int kk = 1; kk <= nCatchDF; kk++ )
    for( j = 1; j <= nCatchRows(kk); j++ )
        pre_catch(kk,j) = nal * elem_div(elem_prod(tmp_ft *
            sel, 1.0 - exp(-Z(h)(i))), Z(h)(i));

for(int kk = 1; kk <= nCatchDF; kk++ )
    res_catch(kk) = log(obs_catch(kk)) - log(pre_catch(kk));

// 1) Likelihood of the catch data.
if(verbose == 1) COUT(res_catch(1));
for(int k = 1; k <= nCatchDF; k++ )
{
    dvector catch_sd = sqrt(log(1.0 + square(catch_cv(k))));
    nloglike(1,k) += dnorm(res_catch(k), catch_sd);
}
```

# Log-likelihood: catch



## Log-likelihoods: relative abundance

The catchability coefficient  $q$  is treated as a nuisance parameter and integrated out of the model (Walters & Ludwig 1994).

$$q = \exp \left( \frac{1}{n} \sum_i \log \left( \frac{I_i}{V_i} \right) \right).$$

The standard deviation ( $\sigma_{i,k}$ ) is calculated from the CV as

$$\sigma_{i,k} = \sqrt{\log(1 + c_{i,k}^2)}.$$

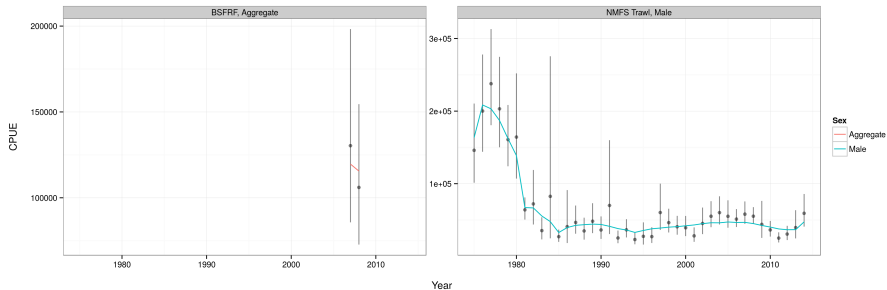
The log-likelihood is

$$\ell(I_{i,k}) = \lambda \left( 0.5 \log(2\pi) + \log(\sigma_{i,k}) + \frac{1}{2\sigma_{i,k}^2} (\log(I_{i,k}) - \log(qV_{i,k}))^2 \right).$$

## Log-likelihood: relative abundance

```
dvar_vector zt = log(obs_cpue(k)) - log(V);  
dvariable zbar = mean(zt);  
res_cpue(k)    = zt - zbar;  
survey_q(k)    = mfexp(zbar);  
pre_cpue(k)    = survey_q(k) * V;  
  
// 2) Likelihood of the relative abundance data.  
if (verbose == 1) COUT(res_cpue(1));  
for (int k = 1; k <= nSurveys; k++ )  
{  
    dvector cpue_sd = sqrt(log(1.0 + square(cpue_cv(k))));  
    nloglike(2,k) += cpue_lambda(k) *  
        dnorm(res_cpue(k), cpue_sd(k));  
}
```

# Log-likelihood: relative abundance



# Log-likelihoods: size composition

Size composition data is assumed to be multinomial distributed

$$\mathbf{P}_{h,i} = (P_\ell)_{h,i} = \text{Multinomial}(n_{h,i}, \mathbf{Q}_{h,i})$$

Alternatively we could use

$$\mathbf{P}_{h,i} = (P_\ell)_{h,i} = \text{Dirichlet}(\lambda_0 n_{h,i} \mathbf{Q}_{h,i}).$$

In this context,  $\lambda_0$  can be thought of as the data weight (which may be estimated in the model) and  $n_{h,i}$  is the relative sample size between years.

## Log-likelihood: size composition

```
// 3) Likelihood for size composition data.  
for (int ii = 1; ii <= nSizeComps; ii++)  
{  
    dmatrix      0 = d3_obs_size_comps(ii);  
    dvar_matrix P = d3_pre_size_comps(ii);  
    dvar_vector log_effn = log(exp(log_vn(ii)) *  
        size_comp_sample_size(ii));  
}
```

## Log-likelihood: size composition

```
switch(nAgeCompType(ii))
{
    case 0: // ignore composition data in model fitting.
        ploglike = NULL;
        break;
    case 1: // multinomial with fixed or estimated n
        ploglike = new acl::multinomial(0,bCmp);
        break;
    case 2: // robust approximation to the multinomial
        if( current_phase() <= 3 || !last_phase() )
            ploglike = new acl::multinomial(0,bCmp);
        else
            ploglike = new acl::robust_multi(0,bCmp);
        break;
}
```



## Log-likelihood: size composition

```
// Compute residuals in the last phase.
if( last_phase() && ploglike != NULL )
{
    d3_res_size_comps(ii) = ploglike->residual(log_effn,P)
}

// now compute the likelihood.
if(ploglike != NULL)
{
    nloglike(3,ii) += ploglike->nloglike(log_effn,P);
}
}
```

## Log-likelihood: recruitment

The standard deviation of the catch ( $\sigma_{i,k}$ ) is calculated from the CV as

$$\sigma_{i,k} = \sqrt{\log(1 + c_{i,k}^2)}.$$

The log-likelihood is

$$\ell(C_{i,k}) = 0.5 \log(2\pi) + \log(\sigma_{i,k}) - \frac{1}{2\sigma_{i,k}^2} (\log(C_{i,k}) - \log(\hat{C}_{i,k}))^2.$$

## Log-likelihood: recruitment

```
switch(nSRR_flag)
{
  case 0: // NO SRR
    res_recruit(byr,nyr) = log(recruits(byr,nyr))
      - (1.0-rho) * logRbar
      - rho * log(++recruits(byr-1,nyr-1))
      + sig2R;
    break;
  case 1: // SRR model
    res_recruit(byr,nyr) = log(recruits(byr,nyr))
      - (1.0-rho) * log(rhat(byr,nyr))
      - rho * log(++recruits(byr-1,nyr-1))
      + sig2R;
    break;
}
```

## Log-likelihood: recruitment

```
// 4) Likelihood for recruitment deviations.
if( active(rec_dev) )
{
    dvariable sigR = mfexp(logSigmaR);
    switch(nSRR_flag)
    {
        case 0:
            nloglike(4,1) = dnorm(res_recruit, sigR);
            nloglike(4,1) += dnorm(rec_ini, sigR);
            break;
        case 1:
            nloglike(4,1) = dnorm(res_recruit, sigR);
            break;
    }
}
```

## Log-likelihood: growth increment

The standard deviation of the catch ( $\sigma_{i,k}$ ) is calculated from the CV as

$$\sigma_{i,k} = \sqrt{\log(1 + c_{i,k}^2)}.$$

The log-likelihood is

$$\ell(C_{i,k}) = 0.5 \log(2\pi) + \log(\sigma_{i,k}) - \frac{1}{2\sigma_{i,k}^2} (\log(C_{i,k}) - \log(\hat{C}_{i,k}))^2.$$

## Log-likelihood: growth increment

```
// 5) Likelihood for growth increment data
if( !bUseEmpiricalGrowth && ( active(Grwth(1)) || active(Grwth
{
    dvar_vector MoltIncPred =
        calc_growth_increments(dPreMoltSize, iMoltIncSex);
    nloglike(5,1) = dnorm(log(dMoltInc) -
        log(MoltIncPred), dMoltIncCV);
}
```

# Penalties

Natural mortality

$$\ell(\textcolor{red}{M}_{h,i}) = 0.5 \log(2\pi) + \log(\textcolor{blue}{\sigma}_M) + 0.5 \frac{1}{\textcolor{blue}{\sigma}_M^2} \sum_i \delta_i^2.$$

# Random thoughts

- Include environmental indices via  $a_f(E_f - \bar{E}_f)$  where  $a_f$  is the shift factor and  $E_f$  the exogenous variable
- the alternative is a more statistical approach whereby selectivity is estimated as a latent state



# References

Walters, C. & Ludwig, D. (1994), 'Calculation of Bayes posterior probability distributions for key parameters', *Canadian Journal of Fisheries and Aquatic Science* **51**, 713–722.