

# Gmacs: A size-structured assessment modeling framework for hard-to-age species

Athol Whitten, Andre Punt, Dave Fournier, and James Ianelli

September 5, 2014

## Abstract

Gmacs is a statistical size-structured stock assessment modelling framework for molting crustacean species. The framework makes use of a wide variety of data, including both fishery- and survey-based size-composition data, and fishery-dependent and -independent indices of abundance. Gmacs has initially been designed for application to the king crab stocks of Alaska. Models of these stocks serve as a testing ground for the first versions of the modeling framework. Gmacs is coded using Auto-Differentiation Model Builder, so inherits its capability to efficiently estimate hundreds of parameters using maximum likelihood. In this paper we describe details of the underlying population dynamics and statistical framework. The description is based upon modelling for crustaceans that undergo molting and with each subsequent molt increase in size.

## Introduction

Statistical catch age models have several advantages over simple production type models in that age and size composition data can be used to better inform structural features such as recruitment variability, and total mortality rates. There are a number of generic age-structured models in use today, but there are very few generic size-based, or staged-based models that are used in stock assessment. In this paper we describe a generalized statistical catch-at-size model that is well suited for animals that cannot be aged, and where only precise length measurements are available. The description is based on a crustaceans that undergo molting and with each subsequent molt increase in length.

## Methods

The analytical details of the generalized model is summarized using tables of equations (e.g., Table 2). These tables serve two purposes: (1) to clearly provide the logical order in which calculations proceed, and (2) organization of a relatively large integrated model into a series

of sub-models that represent specific components such as population dynamics, observation models, reference point calculations, fisheries dynamics, and the objective function. We first start with a description of the population dynamics under steady-state (equilibrium) conditions. Model notation and a description of symbols are provided in Table 1.

## Equilibrium considerations

Parameters for the population sub model are represented by the vector  $\Theta$  (T2.1 in Table 2), which consists of the natural mortality rate, average-recruitment, initial recruitment in the first year, parameters that describe the size-distribution of new recruits, and stock-recruitment parameters (see Table 1 for notation). Constraints for these model parameters are defined in (T2.2). Assuming the molt increments are linear, growth following each molt is a parametric function with the parameters defined in  $\Phi$ .

The model is based on a set of user defined size categories. We assume at any time the population consists of a vector where each component of the vector consists of a number of individuals in some size category. The size category intervals and mid points of those intervals are defined by (T2.3). Average molt increments from size category  $l$  to the next is assumed to be sex-specific, and is defined by a linear function (T2.4). The probability of transitioning from size category  $l$  to  $l'$  assumed that variation in molt increments follows a Gamma distribution (T2.5), and the size-transition matrix for each sex  $h$  is denoted as  $\mathbf{G}_h$ .

The size distribution of new recruits is assumed to follow a gamma distribution (T2.7) with the parameters  $\alpha_r$  and  $\beta_r$ . The gamma distribution is scaled such that  $\alpha_r$  is the mode of the distribution and could potentially be obtained from empirical size composition data. The vector of new recruits at each time step (T2.8) assumes a 50:50 sex ratio.

For unfished conditions that are subject only to sex-specific natural mortality  $M_h$  rates, we assume that each year members of the population grow and experience mortality. The basic assumption is that this process is a linear function of the numbers in each size category, where the categories are separated by sex to accommodate differential growth and survival rates. Survival and growth at each time step in unfished conditions is based on (T2.9), where  $(\mathbf{I}_n)_{l,l'}$  is the identity matrix and  $M_h$  is a scalar. It's also possible to accommodate size-specific natural mortality rates in (T2.9) where  $M_h$  represents a vector of length-specific natural mortality rates.

Assuming a non-zero steady-state fishing mortality rate vector  $\mathbf{f}$ , the equilibrium growth and survival process is represented by (T2.10). The vector  $\mathbf{f}_h$  represents all mortality associated with fishing, including mortality associated with discards in directed and non-directed fisheries.

Assuming unit recruitment, then the growth and survivorship in unfished and fished conditions is given by the solutions to the matrix equations (T2.11) and (T2.12), respectively. The vectors  $\mathbf{u}_h$  and  $\mathbf{v}_h$  represent the unique equilibrium solution for the numbers per recruit in each size category. The total unfished numbers in each size category is defined as  $R_0\mathbf{u}_h$ .

The unfished spawning stock biomass is defined as the equilibrium unfished recruitment multiplied by the sum of products of survivorship per recruit, weight-at-length, and proportion mature-at-length (T2.13). The definition of spawning biomass may include only the

Table 1: Mathematical notation, symbols and descriptions.

Symbol	Description
<u>Index</u>	
$g$	group
$h$	sex
$i$	year
$j$	time step (years)
$k$	gear or fleet
$l$	index for length class
$m$	index for maturity state
$o$	index for shell condition.
<u>Leading Model Parameters</u>	
$M$	Instantaneous natural mortality rate
$\bar{R}$	Average recruitment
$\dot{R}$	Initial recruitment
$\alpha_r$	Mode of size-at-recruitment
$\beta_r$	Shape parameter for size-at-recruitment
$R_0$	Unfished average recruitment
$\kappa$	Recruitment compensation ratio
<u>Size schedule information</u>	
$w_{h,l}$	Mean weight-at-length $l$
$m_{h,l}$	Average proportion mature-at-length $l$
<u>Per recruit incidence functions</u>	
$\phi_B$	Spawning biomass per recruit
$\phi_{Q_k}$	Yield per recruit for fishery $k$
$\phi_{Y_k}$	Retained catch per recruit for fishery $k$
$\phi_{D_k}$	Discarded catch per recruit for fishery $k$
<u>Selectivity parameters</u>	
$a_{h,k,l}$	Length at 50% selectivity in length interval $l$
$\sigma_{s_{h,k}}$	Standard deviation in length-at-selectivity
$r_{h,k,l}$	Length at 50% retention
$\sigma_{y_{h,k}}$	Standard deviation in length-at-retention
$\xi_{h,k}$	Discard mortality rate for gear $k$ and sex $h$

Table 2: Mathematical equations and notation for a steady-state length based model.

model parameters

$$\Theta = (M_h, \bar{R}, \ddot{R}, \alpha_r, \beta_r, R_0, \kappa) \quad (\text{T2.1})$$

$$M_h > 0, \bar{R} > 0, \ddot{R} > 0, \alpha_r > 0, \beta_r > 0, R_0 > 0, \kappa > 1.0 \quad (\text{T2.2})$$

$$\Phi = (\alpha_h, \beta_h, \varphi_h) \quad (\text{T2.3})$$

length-schedule information

$\vec{l}, \vec{x}$  vector of length intervals and midpoints, respectively

$$a_{h,l} = (\alpha_h + \beta_h l) / \varphi_h \quad (\text{T2.4})$$

$$p(l, l')_h = \mathbf{G}_h = \int_l^{l+\Delta l} \frac{l^{(a_{h,l}-1)} \exp(l/\varphi_h)}{\Gamma(a_{h,l}) l^{(a_{h,l})}} dl \quad (\text{T2.5})$$

recruitment size-distribution

$$\alpha = \alpha_r / \beta_r \quad (\text{T2.6})$$

$$p[\mathbf{r}] = \int_{x_l-0.5\Delta x}^{x_l+0.5\Delta x} \frac{x^{(\alpha-1)} \exp(x/\beta_r)}{\Gamma(\alpha) x^\alpha} dx \quad (\text{T2.7})$$

$$\mathbf{r}_h = 0.5p[\mathbf{r}] \ddot{R} \quad (\text{T2.8})$$

growth and survival

$$\mathbf{A}_h = \mathbf{G}_h[\exp(-M_h)(\mathbf{I}_n)_{l,l'}] \quad (\text{T2.9})$$

$$\mathbf{B}_h = \mathbf{G}_h[\exp(-M_h - \mathbf{f}_{h,l})(\mathbf{I}_n)_{l,l'}] \quad (\text{T2.10})$$

survivorship to length

$$\mathbf{u}_h = -(\mathbf{A}_h - (\mathbf{I}_n)_{l,l'})^{-1}(p[\mathbf{r}]) \quad (\text{T2.11})$$

$$\mathbf{v}_h = -(\mathbf{B}_h - (\mathbf{I}_n)_{l,l'})^{-1}(p[\mathbf{r}]) \quad (\text{T2.12})$$

steady-state conditions

$$B_0 = R_0 \sum_h \lambda_h \sum_l \mathbf{u}_{h,l} w_{h,l} m_{h,l} \quad (\text{T2.13})$$

$$\tilde{B} = \tilde{R} \sum_h \lambda_h \sum_l \mathbf{v}_{h,l} w_{h,l} m_{h,l} \quad (\text{T2.14})$$

stock-recruitment parameters

$$s_o = \kappa R_0 / B_0 \quad (\text{T2.15})$$

$$\beta = (\kappa - 1) / B_0 \quad (\text{T2.16})$$

$$\tilde{R} = \frac{s_o \phi - 1}{\beta \phi} \quad (\text{T2.17})$$

females, or only males, or some combination thereof. To accommodate various definitions of spawning biomass the parameter  $\lambda_h$ , with the additional constraint  $\sum_h \lambda_h = 1$ , assigns the relative contribution of each sex to the spawning biomass. For example, if  $\lambda = 1$  then the definition of spawning biomass is determined by a single sex. If  $\lambda = 0.5$ , the spawning biomass consists of an equal sex ratio.

Under steady-state conditions where the fishing mortality rate is non zero, (T2.14) defines the equilibrium spawning biomass based on the survivorship of a fished population. In this case the equilibrium recruitment  $\tilde{R}$  must be defined based on a few additional assumptions; the first of which being the form of the stock-recruitment relationship. Assuming recruitment follows the familiar asymptotic function, or Beverton-Holt relationship:

$$\tilde{R} = \frac{s_o \tilde{B}}{1 + \beta \tilde{B}}, \quad (1)$$

where  $\tilde{B}$  is the equilibrium spawning biomass,  $s_o$  is the slope at the origin, and  $s_o/\beta$  is the asymptote of the function. The parameters of this model can be derived from the unfished recruitment  $R_0$  and the recruitment compensation ratio  $\kappa$ . The slope at the origin, or  $s_o$ , is defined as (T2.16) with the additional constraint that  $\kappa > 1$  for an extant population. Substituting (T2.16) into the Beverton-Holt model (1), and solving for  $\beta$  yields (T2.17).

Given (T2.12) is defined as the vector of individuals per recruit in a fished population, the relative reproductive potential per individual recruit is defined as the sum of products of weight-at-length, maturity-at-length and survivorship-to-length:

$$\phi = \sum_h \lambda_h \sum_l \mathbf{v}_{h,l} w_{h,l} m_{h,l}$$

The total equilibrium spawning biomass is defined as  $\tilde{B} = \tilde{R}\phi$ . Substituting this expression into (1) and solving for  $\tilde{R}$  results in (T2.17).

The equilibrium model defined in Table 2 is a very concise system of equations from which fisheries reference points are easily derived. The minimum amount of information that is necessary to derive SPR-based reference points is an estimate of the natural mortality rate, fisheries selectivity, the size-transition matrix (or growth based on molt increment information). These data alone are sufficient enough to calculate  $F_{\text{SPR}}$ , and the only additional requirement for  $B_{\text{SPR}}$  is to have an estimate of unfished recruitment or a specified average recruitment.

## SPR-based reference points

Reference points based on the Spawning Potential Ratio (SPR) are frequently used as proxies to determine Maximum Sustainable Yield (MSY). The spawning potential ratio is defined as the ratio of reproductive output in fished conditions relative to unfished conditions. In the notation used in Table 2 the SPR is defined as

$$SPR = \tilde{B}/B_0. \quad (2)$$

Fishing mortality rates that results in SPR values of 0.3-0.4 are often desirable for approximating maximum sustainable yield (Clark, 2002). To calculate  $F_{\text{SPR}}$  an iterative solution is required to determine the instantaneous fishing mortality rate that is required to achieve the desired SPR level. For a single fishery this is relatively simple; the fishing mortality rate-at-length vector is the product of selectivity-at-length and fishing effort. A simple bisection algorithm can then be used to iteratively solve for the value of  $F_{\text{SPR}}$  that results in the target SPR. In cases where there is one or more fisheries involved, and the selectivities differ among fisheries, the target  $F_{\text{SPR}}$  will depend on the relative differences in selectivities among gears. Moreover, the allocation of total fishing mortality to each gear has to be specified *a priori*. The solution to this problem must be negotiated as part of the management process as there are an infinite number of allocation combinations that would result in achieving  $F_{\text{SPR}}$ . Given a vector of predefined fishing mortality rate ratios,  $F_{\text{SPR}}$  can then be numerically determined using the same algorithm as the single fishery example.

The expected spawning biomass level when fishing at  $F_{\text{SPR}}$  is referred to as  $B_{\text{SPR}\%}$ . To calculate  $B_{\text{SPR}\%}$  we first evaluate (T2.10) using  $F_{\text{SPR}}$  and then calculate the steady state biomass using (T2.14). An alternative way to calculate the SPR-based reference points is to use per recruit quantities, or incidence functions (Table 3). For example, the SPR is the ratio of spawning biomass per recruit in fished and unfished populations, or (T3.5). For each incidence function the primary difference between the fished and unfished states is the equilibrium numbers at length. In Table 3 we used  $\phi$  and  $\check{\phi}$  to represent fished and unfished states.  $B_{\text{SPR}\%}$  is given by (T3.6) and the total catch at  $F_{\text{SPR}}$  is given by (T3.7). Other quantities such as the equilibrium retained catch, or discards, when fishing at  $F_{\text{SPR}}$  or any other rate can easily be computed using survivorship-at-length schedules and the equilibrium recruitment (Table 3).

## Fishing mortality

Many invertebrate fisheries impose a minimum size limits and prohibit the retention of females as a conservation tool. The probability dying due to fishing comes in two forms, an individual is captured and retained, or the same individual is discarded and dies prematurely due to handling effects. The retained portion of the catch is generally a measurable quantity; however, the discard mortality component cannot be directly measured in situ. To account for mortality associated with both retention and discarding we use the following joint probability to model vulnerability due to fishing mortality by gear  $k$  at length category  $l$  (ignoring sex for clarity)

$$\nu_{k,l} = s_{k,l}[y_{k,l} + (1 - y_{k,l})\xi_k], \quad (3)$$

where  $s_{k,l}$  is the probability of catching an animal of length  $l$  (or selectivity),  $y_{k,l}$  is the probability of retaining an individual in fishery  $k$ , and  $\xi_k$  is the discard mortality rate for fishery  $k$ . Both parametric functions and non-parametric can be used to parameterize selectivity and retention functions. In this application we use a simple logistic function to represent selectivity (T4.3), and retention (T4.4) The discard mortality rate  $\xi_k$  in (3) and (T4.5) assumes the rate is size-independent. However, alternative size-based functions or

Table 3: Incidence functions for equilibrium calculations based on unfished and fished conditions, and SPR-based reference points.

<u>Incidence functions</u>	
<u>unfished</u>	<u>fished</u>
$\phi_B = \sum_h \lambda_h \sum_l u_{h,l} w_{h,l} m_{h,l},$	$\tilde{\phi}_B = \sum_h \lambda_h \sum_l v_{h,l} w_{h,l} m_{h,l} \quad (\text{T3.1})$
$\phi_{Q_k} = 0,$	$\tilde{\phi}_{Q_k} = \sum_h \sum_l \frac{v_{h,l} w_{h,l} s_{h,k,l} (1 - \exp(-Z_{h,l}))}{Z_{h,l}} \quad (\text{T3.2})$
$\phi_{Y_k} = 0,$	$\tilde{\phi}_{Y_k} = \sum_h \sum_l \frac{v_{h,l} w_{h,l} y_{h,k,l} (1 - \exp(-Z_{h,l}))}{Z_{h,l}} \quad (\text{T3.3})$
$\phi_{D_k} = 0,$	$\tilde{\phi}_{D_k} = \sum_h \sum_l \frac{v_{h,l} w_{h,l} d_{h,k,l} (1 - \exp(-Z_{h,l}))}{Z_{h,l}} \quad (\text{T3.4})$
<u>SPR-based reference points</u>	
$SPR = \tilde{\phi}_B / \phi_B$	$(\text{T3.5})$
$B_{\text{SPR}\%} = \tilde{R} \tilde{\phi}_B \big _{F=F_{\text{SPR}}}$	$(\text{T3.6})$
$C_{\text{SPR}\%} = \tilde{R} \tilde{\phi}_Q \big _{F=F_{\text{SPR}}}$	$(\text{T3.7})$

Table 4: Size-based selectivity, retention and fishing mortality.

<u>Parameters</u>	
$\bar{\mathbf{f}}_k, \Psi_{k,i}, a_{h,k}, \sigma_{s_{h,k}}, r_{h,k}, \sigma_{y_{h,k}}$	(T4.1)
$\sum_i \Psi_{k,i} = 0$	(T4.2)
<u>Selectivity &amp; retention</u>	
$s_{h,k,l} = (1 + \exp(-(l - a_{h,k})/\sigma_{s_{h,k}}))^{-1}$	(T4.3)
$y_{h,k,l} = (1 + \exp(-(r_{h,k} - l)/\sigma_{y_{h,k}}))^{-1}$	(T4.4)
$\nu_{h,k,l} = s_{h,k,l}[y_{h,k,l} + (1 - y_{h,k,l})\xi_{h,k}]$	(T4.5)
<u>Fishing mortality</u>	
$\mathbf{F}_{k,i} = \exp(\bar{\mathbf{f}}_k + \Psi_{k,i})$	(T4.6)
$\mathbf{f}_{h,i,l} = \sum_k \mathbf{F}_{k,i} \nu_{h,k,l}$	(T4.7)
<u>Penalized negative loglikelihoods</u>	
$\ell_{F_k} = 0.5 \ln(2\pi) + \ln(\sigma_{F_k}) + 0.5(\bar{F}_k - \hat{F})^2/\sigma_{F_k}^2$	(T4.8)

other covariates such as temperature could be accommodated, but the estimates would have to be obtained independently as its not possible to quantify the discard mortality rate in situ.

In cases where only one sex is allowed to be retained (i.e., male only fisheries) the retention probability  $y_{h,k,l} = 0$  for all size-classes. In this case, it is not necessary to estimate parameters for a sex-specific retention curve, as all females for example, would be discarded.

To improve numerical stability during early phases of the parameter estimation scheme, a penalized likelihood component (T4.8) is used to constrain the average fishing mortality  $\bar{F}_k$  to some user defined value of  $\hat{F}_k$ , where the weight associated with this penalty is defined by  $\sigma_{F_k}$ . During the last phase of estimation, the value of  $\sigma_{F_k}$  is arbitrarily set to a large value (e.g., 2.0) to ensure its is not influencing the overall estimate of the mean fishing mortality rate.

## Natural mortality

Natural mortality is assumed to be sex-specific, size-independent, and may or may not be constant over time. In the case of time-varying natural mortality, the model assumes that changes in  $M_{h,i}$  is a random walk process

$$M_{h,i+1} = M_{h,i} e^{\delta_i}$$



Table 5: Model and constraints for natural mortality

Parameters	
$M_h, \boldsymbol{\delta}_i, \sigma_{M_h}$	(T5.1)
Natural mortality	
$M_{h,i} = \begin{cases} M_h, & i = 1 \\ M_{h,i-1} \exp(\delta_i) & i > 1 \end{cases}$	(T5.2)
Penalized negative log-likelihood	
$\ell_{M_h} = \begin{cases} 0.5 \ln(2\pi) + \ln(\sigma_{M_h}) + 0.5\sigma_{M_h}^{-2} \sum_{i=1}^N \delta_i^2, \\ 0.5 \ln(2\pi) + \ln(\sigma_{M_h}) + 0.5\sigma_{M_h}^{-2} \sum_{i=1}^{N-1} (\delta_{i+1} - \delta_i)^2 \end{cases}$	or (T5.3)

where  $\delta_i$  is assumed to be normal with a known variance  $\sigma_M^2$ . There is an additional constraint where  $\sum_i \delta_i = 0$  to ensure that the scale of  $M_h$  remains estimable. The model further assumes that deviations in natural mortality rate are independent of sex. Key parameters for parameterizing natural mortality are defined in (T5.1). Note that  $\sigma_{M_h}$  is not estimable and must either be fixed or use an informative prior.

The parameter vector  $\boldsymbol{\delta}_i$  with all elements set equal to 0 is equivalent to time-invariant  $M$ . To implement time-varying changes in natural mortality rates,  $\boldsymbol{\delta}$  is treated as an estimated vector of parameters; one parameter for each year in the model, or a fixed number of nodes and the series is interpolated using a bicubic spline. The number of nodes and the placement of the nodes is specified *a priori*.

Two alternative methods are used to constrain the natural mortality rate deviations: (1) the first is using a penalized likelihood, and (2) using cubic spline interpolation with a fixed number of nodes. There are two options for penalizing the likelihood (T5.3) for deviations in natural mortality: (a) allow for deviations around a mean, and (b) allow for drift in natural mortality rate using a likelihood penalizes the first differences in  $\delta_i$ .

## Initial conditions

Initialization of the numbers-at-length, annual recruitment, can either be done assuming steady-state conditions using the equations laid out in Table 2, or treated as unknown parameters to be estimated by fitting the model to data. The ladder method requires size-composition data in the initial years in order to estimate the initial numbers-at-length. Table 6 summarizes the proceeding steps required to initialize the number-at-length for the initial conditions.

Estimated parameters at this stage include the natural mortality rate, the initial recruits, parameters for the size distribution of new recruits, and a vector of deviates  $\mathbf{v}$ , which are

Table 6: Initialize population model  $\{i = 1\}$ 

Parameters	
$\Theta = (M_h, \ddot{R}, \alpha_r, \beta_r, \mathbf{v})$ , where $\sum_l \mathbf{v}_l = 0$	(T6.1)
$\Phi = (\alpha_h, \beta_h, \varphi_h)$	(T6.2)
Growth transition	
$a_{h,l} = (\alpha_h + \beta_h l) / \varphi_h$	(T6.3)
$\mathbf{G}_h = \int_l^{l+\Delta l} \frac{l^{(a_{h,l}-1)} \exp(l/\varphi_h)}{\Gamma(a_{h,l}) l^{(a_{h,l})}} dl$	(T6.4)
Recruitment vector	
$\alpha = \alpha_r / \beta_r$	(T6.5)
$p(\mathbf{r}) = \int_{x_l-0.5\Delta x}^{x_l+0.5\Delta x} \frac{x^{(\alpha-1)} \exp(x/\beta_r)}{\Gamma(\alpha) x^\alpha} dx$	(T6.6)
$\mathbf{r}_h = 0.5p(\mathbf{r})\ddot{R}$	(T6.7)
Growth and survival	
$\mathbf{A}_h = \mathbf{G}_h[\exp(-M_h)(\mathbf{I}_n)_{l,l'}]$	(T6.8)
Initial numbers-at-length $i = 1$	
$\mathbf{n}_{h,i} = [-(\mathbf{A}_h - (\mathbf{I}_n)_{l,l'})^{-1}(\mathbf{r}_h)]e^{\mathbf{v}}$	(T6.9)

constrained to sum to zero, that represent historical variation in recruitment and or mortality (T6.1). The initial numbers-at-length can be found by solving the same matrix equation outlined in Table 2; however, in this case the total number of recruits is used, rather than assuming unit recruitment. This first proceeds by calculating the growth transition matrix (T6.4), the distribution of new recruits (T6.6), and assume a 50:50 sex ratio (T6.7) for the new recruits in each size class. The transition and survival from length class  $l$  to  $l'$  is assumed to involve only natural mortality  $M_h$  in (T6.8). Note that  $(\mathbf{I}_n)_{l,l'}$  is the identity matrix. The initial numbers at length, for each sex, is given by (T6.9), which is just the equilibrium vector of numbers-at-length multiplied by a vector of deviates  $\mathbf{v}$ .

It is also possible to assume unfished equilibrium conditions for this model, where  $\ddot{R}$  is replace with  $R_0$  in (T6.7), and the vector of deviates  $\mathbf{v} = 0$ . In this case, the length-based model, with any number of size-classes, is initialized with just seven parameters assuming sex-independent growth and survival, and just 11 parameters with sex-dependent growth and survival.

## Population dynamics

Updating the numbers ( $\boldsymbol{n}_{h,i}$ ) in each size-class at each time step is the product of size-specific growth and survival ( $\boldsymbol{A}_h$ ) times the numbers-at-length plus new recruitment ( $\boldsymbol{r}_{h,i}$ ):

$$\boldsymbol{n}_{h,i+1} = \boldsymbol{n}_{h,i}\boldsymbol{A}_h + \boldsymbol{r}_{h,i} \quad (4)$$

## References

Clark, W. (2002). F 35% revisited ten years later. *North American Journal of Fisheries Management*, 22(1):251–257.