Gmacs: Generalized Modeling for Alaskan Crab Stocks

Athol R. Whitten[1,2]**, André E. Punt**[1]**, James N. Ianelli**[2,1]

[1] School of Aquatic and Fishery Sciences, Box 355020, University of Washington, Seattle, WA 98195
[2] Resource Ecology and Fisheries Management Division, Alaska Fisheries Science Center, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, 7600 Sand Point Way NE, Seattle, WA 98115

A. Summary

This document describes a generalized and flexible size-structured modelling framework called Gmacs (Generalized Modeling for Alaskan Crab Stocks). The first version of Gmacs (Version 1.0) has been designed to develop stock assessment models for the stocks of red king crab in Bristol Bay and Norton Sound, Alaska. The framework makes use of most of the available data sources for both male and female crabs, including survey and fishery indices of abundance and fishery and survey size-compositions. Gmacs has been applied to data for red king crab stocks to test its efficacy and determine priorities for expansion of the model so it may be used to develop assessment models for other Alaskan crab stocks.

B. Introduction

Gmacs is a statistical size-structured population modeling framework. It is designed to be flexible, scalable, useful for both data-limited and data-rich situations, and to be applied to Alaskan crab stocks. Gmacs can incorporate multiple data types from a variety of fisheries or surveys by combing all data in the form of an integrated analysis (Maunder and Punt, 2013; Punt *et al*., 2013). Integrated analysis allows various kinds of data with different (and sometimes incomplete) collection histories to provide complementary information about fished stocks. This is important when modelling Alaskan crab stocks, as most are represented by numerous data sources, some of which are characterised by short or incomplete time-series.

The population sub-model of Gmacs is typical of the population dynamics represented in most statistical size-structured models (e.g., Punt *et al*., 2013). However, the observation sub-model is designed to utilise the wide variety of data types typical of Alaskan crab stocks. Observation data that can be used in Gmacs include: fishery indices of abundance or effort; survey indices of abundance; discard data; size-composition and weight-at-size data; and tag-recapture data. The current framework allows for the specification of time-varying natural mortality, time-varying molt probability, and time-varying fishery selectivity.

Gmacs has initially been designed for application to the red king crab stocks of Bristol Bay and Norton Sound. Models of these stocks will serve as a testing ground for the first version of the modeling framework (Gmacs Version 1.0). Gmacs is coded using Auto-Differentiation Model Builder (ADMB, Fournier *et. al*., 2012), so inherits its capability to efficiently estimate hundreds of parameters using maximum likelihood. The code is written in a manner that should encourage contributions from multiple developers. It is heavily commented and formally structured, allowing for easy expansion. Furthermore, many of the population dynamics functions, and other generic statistical functions, are inherited from user-contributed libraries to ADMB, including the 'Cstar' library of functions, currently under development (Cstar, 2013). Gmacs output processing can be handled through a purpose-built R package, also currently under development. The Gmacs model code, example model files, and associated R

code is publicly available at https://github.com/awhitten/gmacs. Details of the underlying population dynamics used in Gmacs Version 1.0 are provided here.

C. Generalised population dynamics model
C.1 Main population equations

Gmacs generalised population dynamics are governed by a size-structured model that accounts for numbers-at-time, -sex, -maturity, -shell-condition, and -size-class. The model also accounts for growth, natural mortality, fishery removals, and recruitment. The dynamics of the population are modelled using the equations:

$$N_{v,m,t+1,l}^{s} = \begin{cases} \phi_{t,l}^{s} \sum_{l'} m_{\text{imat},t,l}^{s} Q_{v,\text{imat},t,l'}^{s} P_{l,l',t}^{s} + \sum_{l'} m_{\text{mat},t,l'}^{s} Q_{v,\text{mat},t,l'}^{s} P_{l,l',t}^{s} & \text{if } v = \text{new}; m = \text{mat} & (C.1.1a) \\ (1-\phi_{t,l}^{s}) \sum_{l'} m_{\text{imat},t,l}^{s} Q_{v,\text{imat},t,l'}^{s} P_{l,l',t}^{s} + \bar{R}\, e^{\varepsilon_t} \Pr_l^{s} & \text{if } v = \text{new}; m = \text{imat} & (C.1.1b) \\ (1-m_{\text{mat},t,l}^{s}) Q_{v,\text{mat},t,l}^{s} & \text{if } v = \text{old}; m = \text{mat} & (C.1.1c) \\ (1-m_{\text{imat},t,l}^{s}) Q_{v,\text{imat},t,l}^{s} & \text{if } v = \text{old}; m = \text{imat} & (C.1.1d) \end{cases}$$

where $N_{v,m,t,l}^{s}$ is the number of crab of sex s in size-class l which are in maturity state m (mature/immature) and shell state v (new/old shell) at the start of time-step t (generally years), $P_{l',l,t}^{s}$ is the proportion during time t of crab of sex s in size-class l' which grow into size-class l given that they molted, $m_{m,t,l}^{s}$ is the probability that a crab in maturity state m and size-class l will molt during time-step t (the possibility of a terminal molt is implemented by specifying that $m_{\text{mat},t,l}^{s} = 0$ [1]), $\phi_{t,l}^{s}$ is the probability of crab of sex s in size-class l maturing during time-step t, $Q_{v,m,t,l}^{s}$ is the number of crab of sex s in maturity state m, shell-condition v and size-class l that survive fishing and natural mortality during time-step t:

$$Q_{v,m,t,l}^{s} = N_{v,m,t,l}^{s} e^{-M_{t,l}^{s}} - (C_{v,m,t,l}^{s} + D_{v,m,t,l}^{s}) e^{(y_t-1)M_{t,l}^{s}} \qquad (C.1.2)$$

$M_{t,l}^{s}$ is the rate of natural mortality on crab of sex s in size-class l during time-step t, $R_{t+1} = \bar{R}e^{\varepsilon_t}$ is the recruitment of crab of sex s during time-step t, $\bar{R}$ is mean recruitment, ε_t is the recruitment deviation for time-step t, $\Pr_l^{s}$ is the proportion of recruitment that occurs to sex s and size-class l, $C_{v,m,t,l}^{s}$ is the total landings (in numbers) of crab of sex s in maturity state m, shell state v and size-class l during time-step t, $D_{v,m,t,l}^{s}$ is the total (over all fisheries and sources of discard) number of discarded catch of crab of sex s in maturity state m, shell state v and size-class l during time-step t, and y_t is the time in years (if any) between a survey and the middle of the catch period during time-step t.

C.2 Growth

Growth increments are assumed to be gamma distributed with a mean value linearly dependent on pre-moult size, i.e.:

---

[1] If $m_{\text{mat},t,l'}^{s} = m_{\text{imat},t,l'}^{s}$ the dynamics of immature and mature crab is the same

$$P^s_{l,l',t} = \int_{L_l-\Delta L/2}^{L_l+\Delta L/2} \frac{x^{\alpha^s_{L_{l'},t}} e^{x/\beta^s}}{(\beta^s)^{\alpha^s_{L_{l'},t}} \Gamma(\alpha^s_{L_{l'},t})} dx \qquad \alpha^s_{L_l,t}\beta^s = a^s_t + b^s_t L_l \tag{C.2.1}$$

where L_l is the mid-point of size-class l, ΔL is the specified width of each size-class, a^s_t, b^s_t are the parameters of the linear size–growth increment relationship for sex s during time-step t, and β^s is the parameter determining the variance of the growth increment for sex s. The probability of molting as a function of size is specified by a separate parameter for each size-class or an inverse logistic function, i.e. for a single year and maturity state:

$$m_l = \frac{1}{1+e^{\tilde{\beta}(L_l-\tilde{L}_{50})}} \tag{C.2.2}$$

where $\tilde{\beta}$ and $\tilde{L}_{50}$ are the slope and inflection point of the relationship between size and the probability of molting.

C.3 Landings and discards
Landings are only made by directed fisheries while discards can occur in both directed fisheries and other fisheries:

$$C^{mal}_{v,m,t,l} = \frac{V^{dir}_l S^{mal,dir}_{t,l} F^{dir}_t}{F^{mal}_{t,l}} N^{mal}_{v,m,t,l} e^{-y_t M^{mal}_{t,l}} (1-e^{-F^{mal}_{t,l}}) \tag{C.3.1a}$$

$$D^{mal}_{v,m,t,l} = D^{dir}_{v,m,t,l} + \sum_f D^f_{v,m,t,l} \tag{C.3.1b}$$

$$D^{dir}_{v,m,t,l} = \frac{\phi^{dir}(V^{dir}_l+1-V^{dir}_l) S^{mal,dir}_{t,l} F^{dir}_t}{F^{mal}_{t,l}} N^{mal}_{v,m,t,l} e^{-y_t M^{mal}_{t,l}} (1-e^{-F^{mal}_{t,l}}) \tag{C.3.1c}$$

$$D^f_{v,m,t,l} = \frac{\phi^f S^{s,f}_{t,l} F^f_t}{F^s_{t,l}} N^s_{v,m,t,l} e^{-y_t M^s_{t,l}} (1-e^{-F^s_{t,l}}) \tag{C.3.1d}$$

where $F^s_{t,l}$ is the total fishing mortality rate during time-step t on crab of sex s in size-class l (assumed to be independent of maturity stage and shell condition):

$$F^s_{l,t} = \begin{cases} S^{mal,dir}_{t,l}(V^{dir}_l + \phi^{dir}[V^{dir}_l+1-V^{dir}_l])F^{dir}_t + \sum_f \phi^f S^{mal,f}_{t,l} F^f_t & \text{if } s = \text{male} \quad (C.3.2a) \\ S^{fem,dir}_{t,l}\phi^{dir} F^{dir}_t + \sum_f \phi^f S^{fem,f}_{t,l} F^f_t & \text{if } s = \text{female} \quad (C.3.2b) \end{cases}$$

where $S^{s,f}_{t,l}$ is the selectivity for crab of sex s and in size-class l by fishery f (where fishery f can be the directed fishery or another fishery) during time-step t, V^{dir}_l is the probability of males being landed (retained) in the directed fishery by size, ϕ^f is the handling mortality for fishery f (the proportion of crab which die due to being returned to the water following capture), and F^f_t is the fully-selected fishing mortality by fishery f during time-step t:

$$F^f_t = \bar{F}^f e^{\eta^f_t} \tag{C.3.3}$$

$\bar{F}^f$ is the reference level of fishing mortality for fishery f (directed and other), and η^f_t is the deviation from the reference fishing mortality, $\bar{F}^f$, for fishery f and time-step t.

C.4 Selectivity

Sex-specific selectivity functions can be defined for any fishery or survey. Selectivity patterns for bycatch in a directed or non-directed fishery can be specified separately. Other selectivity patterns that can be specified include:

- Logistic function of size
- A separate parameter for each size-class, subject to a smoothness penalty
- Double-normal dome-shaped function of size

C.5 Retention

Retention is specified as either a separate parameter for each size-class or a logistic function of size:

$$V_l = \frac{1}{1 + e^{-\beta^V (L_l - L_{50}^V)}}$$ (C.5.1)

where β^V and L_{50}^V are the slope and inflection point of the relationship between size and the probability of a crab being retained by a directed fishery.

C.6 Recruitment

The proportion of recruits entering each size-class is assumed to be governed by a gamma function, i.e.:

$$\Pr_l^s = \Omega_R \int_{L_l - \Delta L/2}^{L_l + \Delta L/2} \frac{x^{\alpha_R^s} e^{x/\beta_R^s}}{(\beta_R^s)^{\alpha_R^s} \Gamma(\alpha_R^s)} dx$$ (C.6.1)

where α_R^s and β_R^s are the sex-specific parameters which define the mean and variance of the gamma function and Ω_R is a constant to ensure the integral over all size-classes is equal to 1.

C.7 Maturation

The probability of a crab of sex s in size-class l maturating, ϕ_l^s, is given by a separate parameter for each size-class (subject to a smoothness penalty) or a logistic function of size:

$$\phi_{t,l}^s = \frac{1}{1 + e^{-\beta_{\phi,t*}^s (L_l - L_{\phi,50,t*}^s)}}$$ (C.7.1)

where $\beta_{\phi,t}^s$ and $L_{\phi,50,t}^s$ are the slope and inflection point of the relationship between size and the probability of a crab of sex s maturity for time-step t. t^* in Equation C.7.1. denotes that a variable such as maturity may be assumed to change over time in blocks of time-steps.

C.8 Mature male biomass

The size of the reproductive component of a population can be quantified by the biomass of mature males at the time of mating (set at some specified date), i.e.:

$$MMB_t = \sum_l w_l^{mat} (N_{new,mat,t,l}^{mal} + N_{old,mat,t,l}^{mal}) e^{\tau M_{t,l}^{mal}} e^{-F_{t,l}^{mal}}$$ (C.8.1)

where w_l^{mal} is the weight of a male in size-class l, and τ is the fraction of time-step t related to the time of mating.

C.9 Initial conditions

The initial conditions can be estimated in terms of initial population numbers, or initial recruitment deviations that build-up the initial population numbers. If estimating initial numbers, then numbers-at-size by sex, maturity state, and shell condition are estimated parameters, and are subject to a penalty (see Section F), i.e.:

$$N^s_{v,m,t_1,l} = e^{\lambda^s_{v,m,l}}$$

(C.9.1)

where $\lambda^s_{v,m,l}$ is the parameter which determines the initial (time-step t_1) numbers of crab of sex s, maturity state m, and shell condition v in size-class l. If estimating initial recruitment deviations, the model estimates recruitment deviations beginning in year $t_{1-\delta}$ where δ is the user-specified number of years before the model start year. The model estimates deviations from mean recruitment so as to fit the initial observed abundance indices and length composition data. This results in δ number of extra parameters, but can be considerably less parameters than might be required by the initial population numbers formulation when the number of length classes in the model is large.

D. Likelihood function

D.1 Catch data

The contribution of the catches to the negative of the logarithm of the likelihood function is given by:

$$L_1 = \sum_{f^*} \frac{1}{2(\lambda_{1a})^2} \sum_t \left(\ell n C^{f^*}_t - \ell n C^{f^*,\text{obs}}_t \right)^2$$

$$+ \sum_{f^*} \frac{1}{2(\lambda_{1b})^2} \sum_t \left(\ell n D^{f^*}_t - \ell n D^{f^*,\text{obs}}_t \right)^2$$

(D.1.1)

$$+ \sum_f \frac{1}{2(\lambda^f_{1c\ldots})^2} \sum_t \left(\ell n D^{f}_t - \ell n D^{f,\text{obs}}_t \right)^2$$

where λ_j is a weighting factor, $C^{f^*,\text{obs}}_t$ is the observed retained catch of male crab by a directed fishery f^* during time-step t, $D^{f^*,\text{obs}}_t$ is the observed discarded catch by the corresponding directed fishery f^* during time-step t, and $D^{f,\text{obs}}_t$ is the observed discard by fishery f during time-step t. The observed catches can be in mass or in numbers. The model-predicted catches depend on whether the fishery concerned is a directed fishery or another fishery which discards crab (see Equations C.3.1a-C.3.1d).

D.2 Catch size-frequency data

The model is fitted to the size-frequency of the retained catch of males, the size-frequency of the total catch of males, the size-frequency of females in the directed fishery, the sex-specific size-frequency of the catch by the other fisheries, and the size-frequency of males and females during surveys, under the assumption that data are multinomially distributed. There are, in principle, separate likelihood components for each of the catches by sex, maturity state and shell condition, although the data for some of these population components can be combined. The model-predictions by sex, maturity state and shell condition, and size-class, are given in Equations C.3.1a – C.3.1d.

D.3 Abundance indices

The model is fit to abundance indices separately by sex, i.e.:

$$L_3 = \frac{1}{2\lambda_3^2}\sum_s\sum_t\left(\ell\mathrm{n}I_t^s - \ell\mathrm{n}\hat{I}_t^s\right)^2 / (\sigma_t^s)^2 \tag{D.3.1}$$

where I_t^s is the index of abundance for sex s and time-step t, $\hat{I}_t^s$ is the model-estimate corresponding to I_t^s:

$$\hat{I}_t^s = \sum_l\sum_m\sum_v S_{t*,l}^{I,s}\, w_l^s\, N_{v,\mathrm{mat},t,l}^s \tag{D.3.2}$$

σ_t^s is the (pre-specified) standard error of I_t^s and $S_{t*,l}^{I,s}$ is the selectivity of the gear for crab of sex s in size-class l during year $t*$. Survey selectivity is modelled in the same way as fishery selectivity (see Section C.4). The w_l^s term is dropped from Equation D.3 if the index is specified in terms of numbers rather than mass.

D.4 Effort data

The fully-selected fishing mortality rates by fishery can be assumed to be related to a time-series of effort values, i.e.:

$$L_4 = \sum_f\frac{1}{2(\lambda_4^f)^2}\sum_t(\ell\mathrm{n}E_t^f - \ell\mathrm{n}(q^f F_t^{s,f}))^2 \tag{D.4.1}$$

where E_t^f is the effort for fishery f and time-step t, and q is a constant of proportionality.

D.5 Penalty components

There are several potential penalty components applied to the objective function:

1) A penalty is placed on the deviations in recruitment from average recruitment:

$$P_1 = \frac{1}{2(\lambda_6)^2}\sum_t\varepsilon_y^2 \tag{D.5.1}$$

2) A first and/or second derivative penalty can be place on between-size-class variation in the initial size-structure:

$$P_2 = \frac{1}{2(\lambda_9)^2}\sum_s\sum_v\sum_m\sum_{l=1}^{n_L^s-1}(-\ell\mathrm{n}\lambda_{v,m,l}^s + \ell\mathrm{n}\lambda_{v,m,l+1}^s)^2 \qquad \text{First derivative} \tag{D.5.2a}$$

$$P_2 = \frac{1}{2(\lambda_9)^2}\sum_s\sum_v\sum_m\sum_{l=1}^{n_L^s-2}(\ell\mathrm{n}\lambda_{v,m,l}^s - 2\ell\mathrm{n}\lambda_{v,m,l+1}^s + \ell\mathrm{n}\lambda_{v,m,l+2}^s)^2 \qquad \text{Second derivative} \tag{D.5.2b}$$

where n_L^s is the number of size-classes in the first year for sex s.

3) Penalties are placed on the extent of inter-annual variation in deviations in fishing mortality:

$$P_{3,} = \sum_f\frac{1}{2(\lambda_{8,f})^2}\sum_t(\eta_t^f)^2 \tag{D.5.3}$$

4) A first and/or second derivative penalty can be applied to non-parametric selectivity patterns such as those requiring a parameter for each size-class:

$$P_4 = \sum_f \frac{1}{2(\lambda_9)^2} \sum_s \sum_{l=1}^{n_L^s-1} (-\ell n S_l^s + \ell n S_{l+1}^s)^2 \qquad \text{First derivative} \quad \text{(D.5.4a)}$$

$$P_4 = \sum_f \frac{1}{2(\lambda_9)^2} \sum_s \sum_{l=1}^{n_L^s-2} (\ell n S_l^s - 2\ell n S_{l+1}^s + \ell n S_{l+2}^s)^2 \qquad \text{Second derivative} \quad \text{(D.5.4b)}$$

where n_L^s is the number of size-classes in the model for sex s.

E. Application to Bristol Bay red king crab and Norton Sound red king crab

Gmacs has been applied to data for red king crab stocks to test its efficacy and to compare outputs of 'generalised' models with the most recent assessment models for the same stocks. Details of model outputs and differences between current assessment models and Gmacs implemented models for the same species will follow in a later version of this document.

F. Computer code implementation

Gmacs has been coded using ADMB, and takes much of its design from previous crab assessment models and other generalised statistical modeling frameworks. The outline of the code follows four main steps:

1) Read in data at sufficient resolution to be consistent with the equations and model described above (including optional data file for pre-specified size-transition matrices);
2) Read in a "control file" which ensures consistency with the model specification and available data, and allows control of the parameter estimation procedure;
3) Conduct the estimation including alternative ways to evaluate parameter uncertainty;
4) Provide diagnostic output for evaluating model fits and important management parameters.

Step 3) involves evaluating the following generally self-descriptive functions in order:
   o Get_Effort
   o Get_Mortalities
   o Get_Ns (Numbers)
   o Get_Model_Predictions
   o Evaluate_NegLog_Posterior
   o Get_Params_of_Interest

The last function may only be required during the Hessian calculation (or during the MCMC evaluation phase).

Gmacs is open source software and is intended to draw from expertise within the assessment modeling community. As such, a version control system (Git) has been utilised to manage code development for the project. A remote repository for the code is being maintained on Github and is publicly available at https://github.com/awhitten/gmacs. Crab assessment authors are encouraged to access the source code and contribute to development.

G. Future work

Gmacs Version 1.0 is limited in scope and designed for application to less complicated stocks and fishery situations. It is presented here for review and further development before being used for possible assessment and management purposes. The following extensions and improvements have been identified as priorities:

- Extend the modeling framework so it can be used to develop stock assessment models for Snow crab and Tanner crab.

- Increase the allowable dimensions of the modeling framework to account for seasons, multiple areas, and multiple growth morphs:
    o Include tag-recapture data which can be used to aid estimation of movement rates among areas and to estimate growth within the model.

- Improve the ability to describe how parameters change over time:
    o Allow for multiple time-varying options and for parameters to vary in response to environmental and ecosystem factors.

- Extend the ways in which biological and fishery parameters can be specified:
    o Add extra selectivity functions, including functionality for splines;
    o Allow for time-varying and dome-shaped retention functions.

References

Cstar: Common Stock Assessment Routines. 2013. A contributed function library for ADMB. *Available at* https://github.com/awhitten/cstar

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A. and J. Sibert. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software* 27: 233–249.

Maunder, M.N. and A.E. Punt. 2013. A review of integrated analysis in fisheries stock assessment. *Fisheries Research* 142: 61–74.

Punt, A.E., Haung, T-C. and M.N. Maunder. 2013. Size-structured models for stock assessment of hard-to-age crustacean and mollusc species. *ICES Journal of Marine Science* 70: 16–33.