# rsimGmacs

William Stockhausen, AFSC

“rsimGmacs” is an R package I developed using RStudio (http://www.rstudio.com/) for simulating datasets with which to debug and test Gmacs models. The package, as well as its source code and sample configuration files with which to run a simulation, is hosted in a public repository on GitHub at <https://github.com/seacode/rsimGmacs.git>. The current version (as of 5/04/2015) is 0.6.0. rsimGmacs (“rsim”, for short) is intended to be a port of the Gmacs “operating model” (e.g., the Gmacs population dynamics and observation models) to a code base completely independent of the ADMB-based model to facilitate code debugging and model testing. R was selected for the code base because it is a rich and flexible programming, statistical, and visualization environment, as well as because R forms the principal basis for analyzing and visualizing Gmacs model output. Creating an rsim simulation consists of: 1) writing a “model configuration file” that defines the model configuration (range of years, number of sexes, number of fisheries, etc.), parameter values for the population dynamics and observation models, and data-related Gmacs model options and 2) reading in the model configuration file and running the simulation by invoking the R function “runSim.BBRKC()”. Model output consists of 1) an output data file suitable for use in a Gmacs model run, 2) a series of plots reflecting the simulated model dynamics (available as a pdf, if so specified), and 3) an R list object encapsulating the configuration and arrays reflecting the model processes, dynamics, and observations for later comparison with Gmacs model runs against the simulated data.

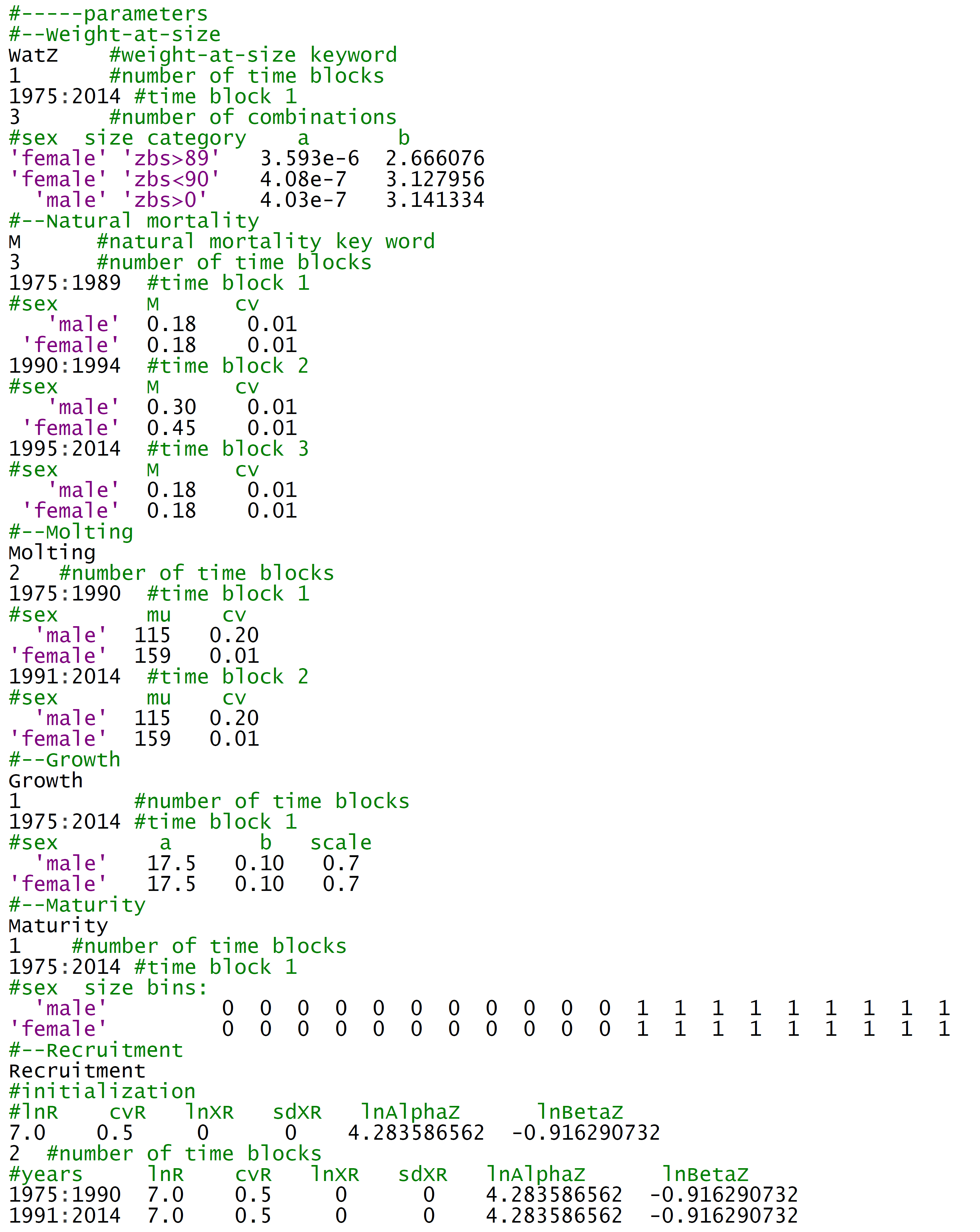
The following text boxes outline a basic “configuration file” for running an rsim simulation. The first section of the configuration file (see Text Box 1) begins with the keyword “ModelConfiguration”, followed on the next line by “KC”, indicating the simulation is for king crab. Model dimensions are specified next. These consist of the beginning and ending years (min y, max y) for the simulation, the size bins included in the model, the number and names of all sexes in the model, the number and names of all shell conditions included in the model, the number and names of all fisheries in the model, and the number and names of all surveys in the model. Assigning names to the model dimensions reduces the scope for accidentally assigning, say, male parameter values to female crab. In contrast to Gmacs, which does not distinguish between fisheries and surveys per se, rsimGmacs distinguishes between the two to simplify the simulation code and the configuration file. In this respect, a “survey” produces only indices of abundance/biomass and/or size compositions but it is not a source of mortality. A “fishery”, on the other hand, is a source of mortality. Fishery-dependent indices, such as a CPUE time series, can be produced in a simulation by defining a survey with selectivity characteristics similar to the fishery in question.



Text Box 1.

Parameter values for model processes are specified next (Text Box 2). Most model processes can be defined over several time blocks, using R vector syntax to define each time block (e.g. “1975:2014”). Parameter values can be repeated in several contiguous time blocks to create an effective non-contiguous time block with the same parameterization.

Parameter values for weight-at-size follow the “WatZ” keyword (Text Box 2). Within each time block, parameter values *a* and *b* describing weight-at-size, modeled as , where *W* is weight in grams and *Z* is carapace length in mm, are specified by sex and size category for each time block. Size categories are specified using the format “*zbs* ? *Z*”, where ‘?’ is either ‘>’ or ‘<’ and ‘*Z*’ is carapace length in mm.



Text Box 2.

Natural mortality (Text Box 5.2), following the “M” keyword, is specified in time blocks by sex and maturity state. A non-zero cv can also be specified to introduce process error in the rate of natural mortality.

Annual probabilities of molting by sex are specified following the “Molting” keyword (Text Box 2). These are defined in time blocks using sex- and size-specific ascending logistic functions characterized by size at 50% probability of molting (*z50*) and its slope (*sdv*).

The growth transition matrix for molting crab is specified following the “Growth” keyword (Text Box 2) by time block using sex-specific parameters describing the mean growth increment as a 2-parameter power-law function of size (*a* and *b*) and a gamma distribution scale factor (*scale*).

Following the “Maturity” keyword, the probability that a crab of a given size is mature is specified by sex within time blocks for every model size bin (Text Box 2). Values should range from 0 (all immature) to 1 (all mature).

Recruitment (Text Box 2) is treated somewhat differently from other processes, in that parameters describing the equilibrium recruitment pattern used to define the initial size composition of the stock are specified following the “Recruitment” keyword, as well as parameters describing sex and size-specific recruitment in time blocks. Parameters used to specify recruitment processes include the mean ln-scale recruitment (*lnR*), the cv for recruitment (*cvR*), the logit-scale sex ratio (*lnXR* = ln(pr(male)/pr(female)), annual process error in the sex ratio (*sdXR*), and parameters describing the size composition of recruiting crab using a gamma distribution (*lnAlphaZ* and *lnBetaZ*).

Fishery characteristics are defined following the “Fisheries” keyword. For each fishery listed in Text Box 1, characteristics are specified by time block following the fishery name (Text Box 3). For each time block, parameters are specified related to discard mortality, fully-selected male capture rates (“mean F”, “sd F”), female capture rates (“female offset”), and observation error (“cv” and “add observation error”). Additionally, sex-specific selectivity/retention functions are defined for each for each time block.



Text Box 3.

Survey characteristics are specified in a similar fashion to those for the fisheries, following the “Surveys” keyword (Text Box 4).

Growth increment data can be included in the simulation output by specifying a non-zero number of years for this type of data following the “GrowthIncrementData” keyword (Text Box 4).

The input configuration file is terminated with the “Done” keyword.



Text Box 4.