**RAM Surplus Production Model Fitting Guide**

**Model Fitting Protocol**

To supplement biological reference points (*TBMSY* and/or *ERMSY*) not present in the RAM Legacy Stock Assessment Database, we fit surplus production models to time series data to estimate these missing BRPs. These post-hoc surplus production model fits are only conducted in the Extended RAMLDB version (or ‘model-fits’ version), not in the ‘Assessment-only’ version.

A Pella-Tomlinson model with a flexible shape parameter (*p*) is used (***Appendix 1***), fitting the model to time series of catch and total biomass drawn from the same assessment. This model is fit under an extensive list of data scenarios and shape parameter assumptions (***Table 1***). These cases include different combinations of freely-varying and fixed model parameters (*BMSY, UMSY, p*) as well as different input data sets. The biomass source is total biomass (which, when necessary, is converted from *SSB* as described in RAM B-Conversion Guide....docx’), and annual values of observed surplus production are calculated from total biomass and total catch or landings (***Appendix 1***). The sources of the assessment data are specified in ***Tables 2 and 3***.

The protocol for the model fitting procedure is as follows:

1. Assessments are checked to see if they have the necessary time series (*TB*, which may be converted from *SSB* if necessary, and either *TC* or *TL*), all in metric tons. The number of values of annual surplus production calculated from these time series must also be greater than five years.
2. Model fits are performed across all cases satisfying the data requirements for fixed values (***Table 4***). For example, an assessment that contains an estimate of *BMSY* but not an estimate of *UMSY* will have some models fit to estimate only *UMSY* (holding *BMSY* fixed at the assessment-provided value), and other models fit to estimate both *BMSY*and *UMSY* (while ignoring the assessment-provided value of *BMSY*; these “redundant” fits are done for purposes of cross-validation with assessment-estimated values).
3. If the set of initial parameter values does not result in a converged model fit, a new set is used (***Table 5***). This is repeated if those subsequent initial values do not yield a converged fit until convergence is achieved.
4. Results from converged model fits are filtered through a set of pass/fail criteria (***Table 6***). If the model fits pass all criteria and if the model fit is appropriate for the stock (e.g. a stock with an assessment-provided BMSY would use a model that holds this *BMSY* estimate as a fixed value) then it is considered fully passed and used to supplement missing BRPs. Models that pass all criteria but are not considered appropriate (e.g. if both *BMSY* and *UMSY* are estimated as freely-varying parameters even though an assessment-estimated value of *BMSY* is available) are used in cross-validations, but are not reported in the RAM Legacy Database.
5. A hierarchical list of models is generated from the model fit results, and the estimated parameters from the highest-ranking model that passed and fits the data scenario of the stock are selected.

**Model Fit Cases (Table 1)**

This table presents the possible cases of data scenarios and shape parameter values considered for fitting surplus production models. Reference points *ERMSY* and *TBMSY* are either held fixed at the values drawn from assessments (“Assessment”, if available) or are freely-varying (“Estimated”). Fixed values of shape parameters (“*p* State”) are listed in **Table 4**.

Cases are compared based on cross-validations of estimated reference points between surplus production model fits and assessments to identify the most suitable shape parameter to assume for each case of data availability (only *ERMSY* required, only *TBMSY* required, or both *ERMSY* and *TBMSY* required). Case 19 is considered only for estimating ‘Empirical-taxonomic’ mean values of *p* for each taxgroup. Cases with freely-varying shape parameters (“*p* State” = “Estimated”) were not considered in comparisons for estimating missing values of *ERMSY* and/or *TBMSY* because they were frequently estimated near fuzzy boundaries of the penalty function despite a wide range of allowed values.

|  |  |  |  |
| --- | --- | --- | --- |
| **Case #** | **Umsy State** | **Bmsy State** | ***p* State** |
| **1** | Estimated | Estimated | Estimated (under penalty function) |
| **2** | Estimated | Estimated | Schaefer |
| **3** | Estimated | Estimated | Fox |
| **4** | Estimated | Estimated | Thorson pooled |
| **5** | Estimated | Estimated | Thorson taxonomic |
| **6** | Estimated | Estimated | Empirical-taxonomic |
| **7** | Estimated | Assessment | Estimated (under penalty function) |
| **8** | Estimated | Assessment | Schaefer |
| **9** | Estimated | Assessment | Fox |
| **10** | Estimated | Assessment | Thorson pooled |
| **11** | Estimated | Assessment | Thorson taxonomic |
| **12** | Estimated | Assessment | Empirical-taxonomic |
| **13** | Assessment | Estimated | Estimated (under penalty function) |
| **14** | Assessment | Estimated | Schaefer |
| **15** | Assessment | Estimated | Fox |
| **16** | Assessment | Estimated | Thorson pooled |
| **17** | Assessment | Estimated | Thorson taxonomic |
| **18** | Assessment | Estimated | Empirical-taxonomic |
| **19** | Assessment | Assessment | Estimated (under penalty function) |

**Time Series Sources (Table 2)**

This table displays the sources of time series used in the model fit. The ordering of the sources represents the data preference for that field.

|  |  |  |
| --- | --- | --- |
| **Time Series** | **Source 1** | **Source 2** |
| **Total Biomass** | *TB* | *TB* converted from *SSB* (see “RAM B-Conversion Guide\_\_\_.docx”) |
| **Catch** | *TC* | *TL* |

**Reference Point Sources (Table 3)**

This table displays the sources of reference points used in the model fit. The ordering of the sources represents the data preference for that field. See “Supplemental Data Info\_\_\_.xlsx” for further descriptions of these sources and methods.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reference Point** | **Source 1** | **Source 2** | **Source 3** | **Source 4** |
| ***TBMSY*** | *TBMSY* | *TBMSY* proxy | *TBMSY* calculated from *TB*/*TBMSY* | *TBMSY* calculated from *MSY* and *ERMSY* |
| ***ERMSY*** | *ERMSY* | *ERMSY* proxy | Calculated *ERMSY* | *ERMGT* |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reference Point** | **Source 5** | **Source 6** | **Source 7** | **Source 8** |
| ***TBMSY*** | *TBMGT* | *TBMSY* converted from *SSBMSY* | *TBMSY* converted from *SSBMSY* proxy | *TBMSY* converted from *SSBMSY* that was calculated from *SSB*/*SSBMSY* |
| ***ERMSY*** | *ERMSY* only usable in model-fits version due to request by assessors | Calculated *ERMSY* with value of *TB* converted from *SSB* |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reference Point** | **Source 9** | **Source 10** | **Source 11** | **Source 12** |
| ***TBMSY*** | *TBMGT* converted from *SSBMGT* | *TBMSY* converted from *TB0* | *TBMSY* converted from *SSBMSY* that came from *SSB0* | *TBMSY* converted from *SSBMSY* that came from derived Method B |
| ***ERMSY*** |  |  |  |  |

**Fixed Parameters (Table 4)**

This table shows the possible fixed values of parameters used in the set of model cases listed in **Table 1**.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***ERMSY*** | Assessment Value | |  |  |  |
| ***TBMSY*** | Assessment Value | |  |  |  |
| ***p*** | Schaefer Value | Thorson Taxonomic Value | Thorson Pooled Value | Empirical Taxonomic Value | Fox Value |

Expanded fixed p values used in the above categories are listed below.

|  |  |  |
| --- | --- | --- |
| ***p* Parameter Category** | ***p* Parameter Subcategory** | ***p* Parameter Value** |
| Schaefer Value | N/A | 2.000 |
| Thorson Taxonomic Value | Pleuronectiformes | 1.406 |
| Thorson Taxonomic Value | Gadiformes | 2.027 |
| Thorson Taxonomic Value | Perciformes | 0.799 |
| Thorson Taxonomic Value | Clupeiformes | 1.427 |
| Thorson Taxonomic Value | Scorpaeniformes | 3.377 |
| Thorson Taxonomic Value | Other | 1.026 |
| Thorson Pooled Value | N/A | 1.736 |
| Empirical Taxonomic Value | Bivalves-gastropods | 1.153 |
| Empirical Taxonomic Value | Carangids-mackerels | 2.224 |
| Empirical Taxonomic Value | Cephalopods | 0.959 |
| Empirical Taxonomic Value | Crabs-lobsters | 0.799 |
| Empirical Taxonomic Value | Echinoderms | 1.153 |
| Empirical Taxonomic Value | Eels | 1.00 |
| Empirical Taxonomic Value | Elasmobranchs | 0.750 |
| Empirical Taxonomic Value | Forage fish | 1.369 |
| Empirical Taxonomic Value | Gadids | 0.975 |
| Empirical Taxonomic Value | Other marine fish | 1.00 |
| Empirical Taxonomic Value | Other marine invertebrates | 0.799 |
| Empirical Taxonomic Value | Other marine percoidids | 1.019 |
| Empirical Taxonomic Value | Other scorpaenids | 1.161 |
| Empirical Taxonomic Value | Pleuronectids | 0.978 |
| Empirical Taxonomic Value | Sebastids | 0.990 |
| Empirical Taxonomic Value | Shrimps | 0.799 |
| Empirical Taxonomic Value | Tuna-billfish | 1.199 |
| Fox Value | N/A | 1.001 **\*** |

**\*** A slight offset from 1 is used to prevent dividing by 0 (see **Appendix 1**).

**Initial Values (Table 5)**

Initial values considered for estimated parameters. Boldfaced values are considered first. All combinations of values (up to 3 x 3 x 3 = 27 combinations if all three parameters are freely varying) are potentially available if previous initial values do not result in a converged model fit.

|  |  |  |  |
| --- | --- | --- | --- |
| ***ERMSY*** | 0.1 | **0.3** | 0.6 |
| ***TBMSY*** | 0.3\**TBMAX* | **0.6\**TBMAX*** | 1.5\**TBMAX* |
| ***p*** | 0.5 | **1.2** | 3 |

**Pass/Fail Criteria (Table 6)**

The following criteria are used to decide whether a converged model is accepted. All criteria (up to 8 filters, as applicable) must be met in order for the fit to be accepted (with the exception of filter 2, for which only one of the two sub-filters a and b passing is necessary for an overall pass of filter 2). In addition to passing all filters, to be considered fully accepted, the model fit must be appropriate for the data available. For example, if a stock has an assessment-estimated *TBMSY*, then a model with a freely varying *TBMSY* parameter would not be considered appropriate, thus is not fully accepted even though it may still pass all filters. Such a model would still be used in cross-validations.

|  |  |
| --- | --- |
| **Filter** | **Criterion for passing** |
| 1 | Time series length (number of observed annual *SP* values) greater than 5 |
| 2a | Less than 50% of observed *SP* values in the two middle quadrants are negative, and the number of *SP* values in these two middle biomass quadrants must also be greater than 5. These biomass quadrants are defined as 4 equal intervals between 0 and estimated *K* (carrying capacity), i.e. the two points where predicated annual surplus production is 0. Both this and 2b must fail together for overall filter 2 to be considered failed. |
| 2b | Sum of observed *SP* values in the two middle biomass quadrants is non-negative, and the number of *SP* values in these two middle biomass quadrants must also be greater than 5). Both this and 2a must fail together for overall filter 2 to be considered failed. |
| 3 | 0.005 < *ERMSY* estimate (ignored if *ERMSY* is not estimated) |
| 4 | *ERMSY* estimate < 0.85 (ignored if *ERMSY* is not estimated) |
| 5 | 0.07\**TBMAX* < *TBMSY* estimate, where *TBMAX* is the maximum observed *TB* in the input time series (ignored if *TBMSY* is not estimated) |
| 6 | *TBMSY* estimate < 2.085\**TBMAX* (ignored if *TBMSY* is not estimated) |
| 7 | *f*(assessment *TBMSY*) > 0, i.e. predicted *SP* evaluated at the assessment-estimated value of *TBMSY* must be positive (ignored if there is no assessment *TBMSY*) |
| 8 | Calculated AIC value of the surplus production model fit to observed annual SP data points is lower than the calculated AIC values of all three linear models fit to the same observed annual *SP* data points. Both *SP* model fits and linear model fits involve surplus production on the vertical axis and total biomass on the horizontal axis. Linear models include: freely varying *SP* intercept with slope = 0 (1 parameter), freely-varying slope with fixed intercept of *SP* = 0 (1 parameter), and freely-varying slope and intercept (2 parameters). |

**Model Functions (Appendix 1)**

This is the form of the Pella-Tomlinson used in the model-fitting:

These predicted values of annual surplus production estimates (pred\_SP) are fit to observed values of annual surplus production estimates (SP), which are calculated from assessment time series data as:

The regression function in ADMB is used with these time series inputs, which calculates the log-likelihood function of the nonlinear least-squares regression. Additionally some command-line options are implemented. The first option (**-rs**) rescales and tries again if no progress is made in the function minimizer. The second option (**-crit 1.e-8**) sets gradient magnitude convergence.

In some cases, the shape parameter *p* is freely-varying. For these cases we incorporate a multiplicative penalty function (based on the Subbotin distribution due to its plateau-like shape) on *p* parameter estimates to prevent outlier values outside the typical range of *p*, while avoiding placing hard boundaries on parameter values.

These are the current values used in the penalty function.

|  |  |
| --- | --- |
| **μ** | *ln*(1) |
| **β** | 10 |
| **σ** | 3 |