Report on Services to meet ICCAT Circular # 4363/2011

GBYP 10/2011 — Stock Assessment Modelling

(ICCAT/GBYP Phase 2–2011)

Modification and extension of ASPIC software for enhanced integration with the R statistics environment

Submitted by:

Michael H. Prager, Ph.D.

2124 SE Grant Street

Portland, Oregon 97214 USA

mike.prager@mhprager.com

http://www.mhprager.com

December 15, 2011

1 Executive summary

Programming and testing were completed to extend the ASPIC software, a surplus-production modeling program used by ICCAT and other regional, national, and international fishery management bodies. The major enhancements were (1) to allow operation under both Linux and Windows; (2) provide better integration with the R statistics language, so that ASPIC can be used directly from R and corresponding ASPIC results analyzed directly in R. These enhancements should allow better use of ASPIC in a Management Strategy Evaluation (MSE) framework, as being developed by the SCRS for developing robust advice consistent with the Precautionary Approach.

2 Linux compatibility

In execution of this project, both ASPIC and ASPICP were ported to Ubuntu Linux, using a commercial Fortran compiler purchased for this work. Correct operation under Linux was verified on several sample data sets.

The porting effort took more than half the allocated time, as it involved several tasks:

- Installing and configuring the Linux operating system
- Converting makefile to Gnu make; working around unimplemented features of Gnu make
- Replacing proprietary file-handling routines in ASPIC and ASPICP with Posix-compliant routines and a few Fortran routines written for this project
- Acquiring, installing and configuring the Absoft compiler
- Resolving quirks and minor bugs in the compiler
- Testing the resulting programs on known data

Versions of both programs compatible with Linux were supplied to ICCAT by way of a Dropbox folder established by the ICCAT Population Dynamics Expert (Dr. Kell).

3 R-compatible output

3.1 ASPIC fitting mode

Existing, but rudimentary R-compatible output from ASPIC in fitting mode was enlarged and refined substantially. An example is given immediately below.

In this example, the R function `ls.str` is used to illustrate the structure and partial contents of the R object. It does not provide a full listing of the object. The symbol ➔ indicates a line broken for printing but continuous in the source.

```
> bsb <- dget("bsb.rdat")
> ls.str(bsb)
diagnostics : List of 9
 $ error.code  : num 0
 $ error.text  : chr "Normal convergence"
 $ contrast    : num 0.879
 $ nearness    : num 1
 $ rsquare     : num [1:2] 0.789 -0.473
 $ obj.fn.value: num 3.43
 $ cpu.time    : Named num [1:3] 0 0 7.01
 $ index.corr  : num [1:2, 1:2] 1 0.57 NA 1
 $ index.n     : num [1:2, 1:2] 30 14 NA 14
estimates :  Named num [1:14] 2 0.814 1340.546 0.143 9357.234 ...
info : List of 5
 $ date      : chr "Thursday, 15 Dec 2011 at 16:22:58"
 $ program   : chr "ASPIC"
 $ version   : chr "5.49"
 $ run.type  : chr "FIT"
 $ input.file: chr "E:/mike.prager/Documents/Prof/ICCAT/ASPIC 2011/ASPIC ➔
    tests/BSB.inp"
inputs : List of 7
 $ int           : Named num [1:7] 54 0 2 1 10000 ...
 $ bool          : Named num [1:7] 1 1 1 1 1 0 0
 $ real          : Named num [1:10] 1.0e-08 3.0e-08 1.0e-04 4.0 1.4e+03 1.0➔
    e+02 5.0e+04 1.2e+04 1.0e+03 5.0e+05
 $ char          : Named chr [1:4] "Black Sea Bass off SE United States -- ➔
    ICCAT Example" "LOGISTIC" "YLD" "SSE"
 $ series.titles : Named chr [1:2] "HB Index (WPUE), Yield ext @ Rec = 2➔
    xCom" "MARMAP Chevron Trap"
 $ series.types  : Named chr [1:2] "CC" "I1"
 $ series.weights: num [1:2] 1 1
t.series : 'data.frame':        55 obs. of  16 variables:
 $ year     : num  1950 1951 1952 1953 1954 ...
 $ F.total  : num  0.03779 0.02629 0.01988 0.01362 0.00795 ...
 $ B        : num  15228 15441 15772 16128 16504 ...
 $ B.bar    : num  15339 15613 15957 16323 16702 ...
 $ Y.tot.obs: num  580 410 317 222 133 ...
 $ Y.tot.est: num  580 410 317 222 133 ...
 $ Y.01     : num  580 410 317 222 133 ...
 $ sp       : num  793 741 674 597 514 ...
 $ F.Fmsy   : num  0.2638 0.1835 0.1388 0.0951 0.0555 ...
 $ B.Bmsy   : num  1.63 1.65 1.69 1.72 1.76 ...
 $ U.01.ob  : num  NA NA NA NA NA NA NA NA NA NA ...
 $ U.01.pr  : num  9.09 9.25 9.46 9.67 9.9 ...
 $ U.02.ob  : num  NA NA NA NA NA NA NA NA NA NA ...
 $ U.02.pr  : num  10.2 10.4 10.6 10.9 11.1 ...
 $ wt.01    : num  1 1 1 1 1 1 1 1 1 1 ...
 $ wt.02    : num  1 1 1 1 1 1 1 1 1 1 ...
>
```

In the example, the following components of the output list can be seen:

- A `diagnostics` item (R list), which contains
  - The error code from ASPIC
  - A text explanation of the error code
  - Two ad hoc indices (coverage and nearness) that indicate the probability of a good fit. These indices were reformulated to be compatible with both the Schaefer and Pella–Tomlinson forms of production model. Their calculation is described in Appendix A.
  - The R^2 of the fit to each data series
  - The objective-function value
  - The computer time taken by the run
- The `estimates` item (R vector) contains estimates of model parameters and other quantities of interest
- The `info` item (R list) contains date and time of the run, version number of ASPIC used, input file name, and similar items
- The `inputs` item (R list) contains the user's specifications for the run, such as convergence criteria, starting values, and run title, as well as information about each data series (title, type of data, and statistical weight)
- An R dataframe named `t.series` contains observed and fitted time-series, such as index values, biomass, fishing effort, and others.

3.2 ASPIC bootstrap mode

R-compatible output was added to the bootstrap mode of ASPIC, which previously generated R output only in fitting mode. An example of the bootstrap is given below:

```
> bsb.bot <- dget("bsb-bot.rdatb")
> ls.str(bsb.bot)
bootstrap : 'data.frame':       7 obs. of  11 variables:
 $ B1.K         : num  0.716 0.542 0.957 0.657 0.755 ...
 $ K            : num  18370 15071 42164 15769 28150 ...
 $ MSY          : num  1357 680 1495 1023 1458 ...
 $ q.01         : num  0.000609 0.000212 0.000811 0.000329 0.000737 ...
 $ q.02         : num  0.000609 0.000212 0.000811 0.000329 0.000737 ...
 $ Bmsy         : num  9185 7535 21082 7884 14075 ...
 $ Fmsy         : num  0.1477 0.0369 0.1985 0.0723 0.1861 ...
 $ Bnext.Bmsy   : num  0.147 0.101 0.221 0.118 0.18 ...
 $ Flast.Fmsy   : num  2.68 1.87 3.35 2.16 2.89 ...
 $ Ynext.at.Fmsy: num  212 162 320 193 272 ...
 $ Ynext.equil  : num  369 290 542 340 471 ...
diagnostics : List of 3
 $ error.code     : num 0
```

```
 $ cpu.time       : Named num [1:3] 0 2 38.2
 $ rejected.trials: Named num [1:5] 0 29 0 0 0
info : List of 5
 $ date      : chr "Thursday, 15 Dec 2011 at 16:26:45"
 $ program   : chr "ASPIC"
 $ version   : chr "5.49"
 $ run.type  : chr "BOT"
 $ input.file: chr "E:/mike.prager/Documents/Prof/ICCAT/ASPIC 2011/ASPIC ➔
     tests/BSB-bot.inp"
inputs : List of 7
 $ int           : Named num [1:7] 54 300 2 1 100000 ...
 $ bool          : Named num [1:7] 1 1 1 1 1 0 0
 $ real          : Named num [1:10] 1.0e-08 3.0e-08 1.0e-04 4.0 1.4e+03 1.3➔
     e+02 1.5e+04 1.2e+04 1.0e+03 5.0e+05
 $ char          : Named chr [1:4] "Black Sea Bass off SE United States -- ➔
     ICCAT Example" "LOGISTIC" "YLD" "SSE"
 $ series.titles : Named chr [1:2] "HB Index (WPUE), Yield ext @ Rec = 2➔
     xCom" "MARMAP Chevron Trap"
 $ series.types  : Named chr [1:2] "CC" "I1"
 $ series.weights: num [1:2] 1 1
>
```

The bootstrap ASPIC output contains the following components—

- An R dataframe named `bootstrap`, with one column for each estimated parameter or other quantity of management interest. Rows of the dataframe are named and hold the following statistics:
  - Point estimate
  - Upper and lower 80% confidence bounds
  - Upper and lower 50% confidence bounds
  - Interquartile range
  - Standard error (added for this project)
- The `diagnostics` item (R list), which contains
  - The error code from ASPIC
  - The computer time taken by the run
  - the number of bootstrap trials rejected for each potential cause (parameters at bounds; lack of convergence)
- The `info` item (R list) holds the date and time of the run, ASPIC version, input file name, and similar items
- The `inputs` item (R list)holds the user's specifications for the run, and information about the data series

4 Comments and acknowledgements

Due to this work, the ability of ASPIC to run in an R environment has increased markedly. This is a positive development that should facilitate management strategy evaluation and other simulation studies.

Because of the short time frame of this contract, extensive testing was not possible. Testing on several known data sets indicated proper operation. As the software is further used by ICCAT, the author will endeavor to participate in further testing and development to ensure that the interface between R and ASPIC meets the needs of ICCAT scientists.

The author acknowledges with appreciation the financial support of ICCAT in making ASPIC and related programs more useful to the stock assessment community. The assistance of Dr. Antonio Di Natale (the ICCAT GBYP Coordinator), and Dr. Laurie Kell (the ICCAT Population Dynamics Specialist) is particularly appreciated.

Appendix A Nearness and coverage indices

ASPIC output includes two ad hoc statistics for guidance on likely usefulness of a production model fit. The nearness index N^* and coverage index C^* are each calculated from the estimated biomass trajectory and B_{MSY} of the fit.

A.1 Nearness index

The nearness index N^* indicates how close to B_{MSY} a biomass trajectory has come. Values of N^* range from zero to unity, with higher values indicating a trajectory that has come close to (or attained) B_{MSY}.

For two related reasons, high N^* suggests that a fit will be informative. First, B_{MSY} is the most nonlinear part of the production curve, the part with the most information on which effort levels are sustainable and which not. Data including this stock level are (in the author's experience) more likely to produce good estimates. Second, failed fits of production models frequently estimate nearly constant biomass trajectories that are either far below or far above B_{MSY}. Either case results in a low value of N^*.

The index is calculated as follows—

1. Definitions

- Let the estimated biomass trajectory be $B_t, t = 1 \ldots T$
- Let B_t^* be the value of B_t closest to B_{MSY}

2. Computation of N^* depending on range of B_t

   - If any $B_t \leq B_{\mathrm{MSY}}$ and any other $B_t \geq B_{\mathrm{MSY}}$, the trajectory has crossed B_{MSY}, and

     $$N^* = 1.0$$

   - If $B_t^* \leq B_{\mathrm{MSY}}$,

     $$N^* = \frac{B_t^*}{B_{\mathrm{MSY}}}$$

   - If $B_t^* > B_{\mathrm{MSY}}$,

     $$N^* = \frac{K - B_t^*}{K - B_{\mathrm{MSY}}}$$

A.2 Coverage index

The coverage index C^* indicates how much of the space between zero and K the biomass trajectory has covered. Values of N^* range from zero to unity, with higher values indicating a trajectories that have covered more of the space. Coverage of each of two segments of the space—above and below B_{MSY}—is weighted equally.

1. Definitions

   (a) Let the estimated biomass trajectory be $B_t, t = 1 \ldots T$
   (b) Let the mininum biomass estimated be B_{min}
   (c) Let the maximum estimated biomass be B_{max}

2. Compute C^* by one of three formulas

   (a) If all $B_t \leq B_{\mathrm{MSY}}$,

   $$C^* = \frac{B_{\mathrm{max}} - B_{\mathrm{min}}}{2\, B_{\mathrm{MSY}}}$$

(b) If all $B_t > B_{\text{MSY}}$,

$$C^* = \frac{B_{\max} - B_{\min}}{2(K - B_{\text{MSY}})}$$

(c) If some $B_t \leq B_{\text{MSY}}$ and some $B_t > B_{\text{MSY}}$,

$$C^* = 0.5\left[\frac{B_{\text{MSY}} - B_{\min}}{B_{\text{MSY}}} + \frac{B_{\max} - B_{\text{MSY}}}{K - B_{\text{MSY}}}\right]$$