

# A Simple User Guide for CMSY

R. Froese, N. Demirel, G. Coro, K. Kleisner and H. Winker

published online at <http://oceanrep.geomar.de/33076/> on 21 June 2016

This is an accompanying document for Froese, R., Demirel, N., Coro, G., Kleisner, K., Winker, H. (2016), Estimating Fisheries Reference Points from Catch and Resilience (accepted for publication in Fish and Fisheries in June 2016). CMSY is a Monte-Carlo method that estimates fisheries reference points ( $MSY$ ,  $F_{msy}$ ,  $B_{msy}$ ) as well as relative stock size ( $B/B_{msy}$ ) and exploitation ( $F/F_{msy}$ ) from catch data and broad priors for resilience or productivity ( $r$ ) and for stock status ( $B/k$ ) at the beginning and the end of the time series. Part of the CMSY package is an advanced Bayesian state-space implementation of the Schaefer surplus production model. This document provides a simple step-by-step guide for researchers who want to apply the CMSY methods to their own data.

The CMSY version referred to in this guide (CMSY\_O\_7f.R) is newer than the one used in Froese et al. (2016). The main differences are faster execution because of parallel processing and more emphasis on management than on evaluating CMSY. Also, some labels in the input files have changed, as indicated in the header of the code.

The required R-code and some example input files (O\_ICES&BS\_Catch\_7.csv and O\_ICES&BS\_ID\_8.csv) can be downloaded from <http://oceanrep.geomar.de/33076/>

## Installation instructions

1) Install a recent version of R on your computer. CMSY was tested under R version 3.2.3, available from <http://www.r-project.org/>, but newer versions should also work.

2) We suggest using RStudio as an R development environment. RStudio is a free software that is available for several Operating Systems (Windows, OS, Linux, ...) and can be downloaded at <http://www.rstudio.com/products/rstudio/download/>

3) Install the Gibbs samples JAGS for your Operating System from the following web site: <http://sourceforge.net/projects/mcmc-jags/files/JAGS/3.x/>

*If you are using an Operating System different from MS Windows, please comment (#) all the lines in the code containing the windows(..) operation.*

4) In order to run the code, several R-packages are required. In the R Console execute the following commands (i.e. cut & paste into the Console window, then hit Enter and wait):

```
install.packages("R2jags")
```

```
install.packages("coda")
```

```
install.packages("lattice")
```

```
install.packages ("parallel")
```

```
install.packages ("foreach")
```

```
install.packages ("doParallel")
```

```
install.packages ("gplots")
```

(pay attention to possible required sub-packages, as alerted in warning/error messages from the R installer, and install missing packages if needed)

5) Two different data files are required by CMSY, which should be placed in the same directory as the script. The names of these files are specified in the code (lines 32 and 33). Examples are provided with the code and their structure is specified in next section.

6) Make sure that the source file and the downloaded R script are in the same directory.

7) Open the downloaded CMSY R script (CMSY\_O\_7e.R) in RStudio. Use the tab "Session" and select "Set Working Directory" -> "To Source File Location", so that the code will find the data files.

8) If you want to use your own input files, just change the file names in the *catch\_file* and *id\_file* variables in the "Required settings, File names" section of the code. If you create your own input files, make sure you use the same headers (case sensitive) as in the provided example files. Make sure you are using comma-delimited (.csv) files (look at the data in a simple text editor such as Notepad to check for commas).

9) The R-code can either analyze all stocks in the data files (default) or a single stock can be specified in the "Select stock to be analyzed" section of the code, according to the stock identification specified at line 44. To specify the stock

to analyze just enter the unique fish stock name or identifier of the stock there (e.g. `stocks <- "cod-scow"`). To make the code run on all the stocks in the files, just comment out line 44 (put # in front). You can analyze the stocks in alphabetic order or in the sequence in the ID file or by Region or subregion, according to settings in lines 214 – 216. Uncomment (remove #) one of these lines.

10) In RStudio, click on “Source” (or press Ctrl+A followed by Ctrl+R or Ctrl + Shift + S) to execute the code. [

11) When the analysis is complete, results can be found in the console window as well as in the CMSY graphs window (which can be saved manually) and in an output .txt file (having the “Out\_” prefix), whose name depends on the date of the execution. An output file in .csv format containing the results is also created in the same directory, which can be opened in Excel for example. The titles of the columns are in the same order as the results in the console.

## Structure of the input files

### Structure of the catch file (.csv)

For each stock, the following information must be specified (in the corresponding columns):

**Stock:** a unique fish stock name or identifier (e.g. “cod-2532”), repeated for each year

**yr:** the reporting year of the catch (e.g. 2004). One row for each year.

**ct:** catch value, in tonnes (e.g. 12345). One row for each year. Gaps with no entries are not accepted at the moment and must be filled by interpolating missing or incorrect values.

**bt:** the value of the biomass (in tonnes, e.g. 34567), or the value of the CPUE or stock size index (e.g. 0.123), or NA if there is no information. Gaps filled with NA are acceptable for abundance.

## Structure of ID file (.csv)

For each stock, the following information must be specified (in the corresponding columns):

**Region:** a string indicating the catch area, e.g. “Northeast Atlantic”

**Subregion:** a string indicating the subarea, e.g. “Baltic Sea”

**Stock:** a unique fish stock name or identifier (corresponding to the one in the ‘Stock’ column in the catch file)

**Name:** optional; a common name of the species, e.g. “Makala”

**EnglishName:** optional; a common English name of the species, e.g. “Greater forkbeard”

**ScientificName:** optional; the scientific name of the species, e.g. “Phycis blennoides”

**SpecCode:** optional; the code number used in FishBase for fish or SealifeBase for non-fish

**Group:** optional; the functional group that a species belongs to, e.g. “Deep water” or “Small pelagics”

**Source:** optional; the source where the catch statistics were taken from, e.g. <http://www.ices.dk/sites/pub/Publication%20Reports/Advice/2014/2014/gfb-comb.pdf>

**MinOfYear:** the start year of the catch report

**MaxOfYear:** the last year of the catch report

**StartYear:** the start year of the Catch time series. This value can also be used to specify a subset of the year data to use for the CMSY.

**EndYear:** the end year of the Catch time series. This value can also be used to specify a subset of the year data to use for the CMSY.

**Flim, Fpa, Blim, Bpa, Bmsy, FMSY, MSYBtrigger, B40, M, Fofl:** optional; fisheries reference points from assessments, not used in the analysis

**Resilience:** prior estimate of resilience, corresponding to intrinsic growth rate ranges (see next section). Allowed values are “High”, “Medium”, “Low”, “Very Low”. Get default values from [www.FishBase.org](http://www.FishBase.org)

**r.low / r.hi:** an optional pair of parameters to specify the range of intrinsic growth rate for the species. Set these to NA to use the range specified in Resilience. If values are given, the Resilience column will be ignored. Check [www.FishBase.org](http://www.FishBase.org) for prior estimates.

**stb.low / stb.hi:** the estimated biomass range relative to the unexploited biomass (B/k) at the beginning of the catch report period (see next section).

**int.yr:** the year in the catch time series range, to be taken as reference for intermediate biomass level. Set it to NA to have it estimated from the year with maximum or minimum catch.

**intb.low / intb.hi:** the estimated intermediate relative biomass range (see next section). Set it to NA to have it estimated from maximum or minimum catch, according to some simple rules (note: these may not be appropriate for your stock).

**endb.low/ endb.hi:** the estimated relative biomass (B/k) range at the end of the catch time series (see next section).

**q.start / q.end:** the start and end year for determining the catchability coefficient. Set to a recent period of at least 5 years where catch and abundance were relatively stable or had similar trends. If set to NA the default is last 5 years (or last 10 years in slow growing species).

**btype:** the type of information in the bt column of the catch file. Allowed values are “biomass” (when total biomass is reported), “CPUE” (when cpue or cpue index or SSB are reported) or “None” (if no abundance data are available).

**force.cmsy:** set to TRUE if the management analysis should use the CMSY results rather than available BSM results. Default is FALSE or F.

**Comment:** a comment on the stock or the quality of the analysis. This comment is shown in the output.

Remember that the files must be saved in “csv” (comma delimited) format. Double-check that indeed a comma (and not a semi-colon) is used as delimiter.

## Suggestions for parameters settings

Table 1 reports a set of questions that can help to set the CMSY input parameters. Please note that priors can also be derived with other stock assessment tools, such as length frequency analysis or catch per unit of effort.

**Table 1.** Example of questions to be put to experts to establish priors for CMSY analysis.

| Prior                                     | Question to experts  |
|---|--|
| Start year for catch time series          | From what year onward are catch data deemed reliable?  |
| Relative start and end biomass<br>$B/B_0$ | What is the most likely stock status for the beginning and end of that time series: lightly fished, fully exploited, or overfished?  |
| Relative intermediate biomass<br>$B/B_0$  | Is there an intermediate year where biomass is considered to have been particular low or high, e.g., exploitation changed from light to full, or where an extraordinary large year class entered the fishery?  |
| Resilience prior<br>$r$                   | What is your best guess for the range of values including natural mortality of adults ( $M$ )? Consider the empirical relationship $r \approx 2 M$   |
| Resilience prior<br>$r$                   | What is your best guess for the range of values including maximum sustainable fishing mortality ( $F_{msy}$ )? Considering the relationship $r \approx 2 F_{msy}$<br><i>Use this question to reinforce or change the answer to previous question</i> |

Alternatively, you can get preliminary estimates of  $r$  from the following empirical relations:

$$r \approx 2 M \approx 2 F_{msy} \approx 3 K \approx 3/t_{gen} \approx 9/t_{max}$$

where  $r$  is the intrinsic rate of population increase,  $M$  is the rate of natural mortality,  $F_{msy}$  is the maximum sustainable fishing mortality,  $K$  is the somatic growth rate (from the von Bertalanffy growth equation),  $t_{gen}$  is generation time, and  $t_{max}$  is maximum age. If point estimates are very close to each other, assume a range of uncertainty of +/- 50%. Give more weight to traits giving low estimates of  $r$ , as these will act as bottle neck for population growth.

Table 2 suggests ranges for relative biomass to be used as input parameters, depending on the depletion status of the stock.

**Table 2.** Prior relative biomass ( $B/k$ ) ranges for CMSY.

| <b>Strong depletion</b> | <b>Medium depletion</b> | <b>Low depletion</b> |
|-------------------------|-------------------------|----------------------|
| 0.01 – 0.4              | 0.2 – 0.6               | 0.5 – 0.9            |

Table 3 reports the  $r$  ranges automatically associated by CMSY to the resilience parameter values.

**Table 3.** Prior ranges for parameter  $r$ , based on classification of resilience.

| <b>Resilience</b> | <b>prior <math>r</math> range</b> |
|-------------------|-----------------------------------|
| High              | 0.6 – 1.5                         |
| Medium            | 0.2 – 0.8                         |
| Low               | 0.05 – 0.5                        |
| Very low          | 0.015 – 0.1                       |

When setting an intermediate biomass, it often improves the CMSY analysis if the end of a period with low biomass is indicated by setting the intermediate year to the last year with low biomass, and indicating a respective relative range, e.g. as 0.01 – 0.4. Similarly, indicate a period of large biomass by setting the intermediate year to the last year with high biomass and indicate a respective range, e.g. as 0.4 – 0.8. In general, the width of relative biomass windows should not be less than 0.4, unless the stock is known to be very strongly depleted, in which case ranges of 0.01-0.3 or 0.01 – 0.2 are appropriate. Setting a range of 0 to 1 is also possible, and would indicate no information at all about stock status, which is, however, unlikely. If a stock is fished it must be smaller than 1. If it is delivering decent catches, it must be larger than 0.01. See Table 1 for guidance on how to get priors from interviews with fishers or experts (or yourself).

Note that if abundance information are provided, an additional surplus production analysis is performed with an advanced Bayesian state-space implementation of a Schaefer model (BSM). This analysis is then used for management advice, because it contains more information. If instead you want to use the CMSY results, set “force.cmsy” to TRUE in the ID file.

## Results of CMSY analysis

When running CMSY it will first do a Monte-Carlo analysis of catch and priors for  $r$  and  $B/k$ . Progress will be indicated on screen and in graphs. Afterwards, if at least 3 years with abundance data have been provided, it will run a full surplus production model on catch and abundance.

The screen output for Sole (*Solea solea*) in the Celtic Seas (sol-iris) is shown on the next two pages. This text is also available in one of the output files.

-----  
CMSY Analysis, Tue Jun 21 16:39:03 2016  
-----

Parallel processing will use 8 cores  
Files O\_ICES&BS\_Catch\_7.csv , O\_ICES&BS\_ID\_8.csv read successfully  
Processing sol-iris , Solea solea  
startbio= 0.2 0.6 default , intbio= 2010 0.01 0.4 default , endbio= 0.01 0.4 default

First Monte Carlo filtering of r-k space with 10000 points...  
Found 58 viable trajectories for 55 r-k pairs  
Shrinking k space: repeating Monte Carlo in the interval [ 5.68 , 44.6 ]

Attempt 1 of 3 with 10000 additional points...  
Found altogether 204 viable trajectories for 180 r-k pairs  
Shrinking k space: repeating Monte Carlo in the interval [ 4.96 , 44.2 ]

Attempt 2 of 3 with 20000 additional points...  
Found altogether 470 viable trajectories for 414 r-k pairs  
Shrinking k space: repeating Monte Carlo in the interval [ 4.96 , 44 ]

Attempt 3 of 3 with 30000 additional points...  
Found altogether 900 viable trajectories for 800 r-k pairs

Final sampling in the tip area above  $r = 0.216$  with 30000 additional points...  
Found altogether 1046 viable trajectories for 939 r-k pairs  
Running MCMC analysis...

-----  
Species: Solea solea , stock: sol-iris  
Sole in Division VIIa (Irish Sea)  
Region: Northeast Atlantic , Celtic Seas  
Catch data used from years 1970 - 2014 , abundance = CPUE  
Prior initial relative biomass = 0.2 - 0.6 default  
Prior intermediate rel. biomass= 0.01 - 0.4 in year 2010 default  
Prior final relative biomass = 0.01 - 0.4 default  
Prior range for  $r = 0.18 - 1$  expert, , prior range for  $k = 2.22 - 50.4$   
Prior range of  $q = 0.959 - 4.57$

Results of CMSY analysis  
-----

Altogether 1046 viable trajectories for 939 r-k pairs were found  
 $r = 0.377$  , 95% CL = 0.273 - 0.522 ,  $k = 15.9$  , 95% CL = 11.3 - 22.5  
MSY = 1.5 , 95% CL = 1.21 - 1.85  
Relative biomass in last year = 0.129  $k$  , 2.5th perc = 0.0163 , 97.5th perc = 0.393  
Exploitation  $F/(r/2)$  in last year = 0.468

Results from Bayesian Schaefer model (BSM) using catch & CPUE

-----  
q = 1.22 , lcl = 0.961 , ucl = 1.56  
r = 0.517 , 95% CL = 0.373 - 0.715 , k = 11.1 , 95% CL = 7.9 - 15.5  
MSY = 1.43 , 95% CL = 1.21 - 1.68  
Relative biomass in last year = 0.0776 k, 2.5th perc = 0.0605 , 97.5th perc = 0.0954  
Exploitation F/(r/2) in last year = 0.446

Results for Management (based on BSM analysis)

-----  
Fmsy = 0.258 , 95% CL = 0.187 - 0.358 (if B > 1/2 Bmsy then Fmsy = 0.5 r)  
Fmsy = 0.0802 , 95% CL = 0.0579 - 0.111 (r and Fmsy are linearly reduced if B < 1/2 Bmsy)  
MSY = 1.43 , 95% CL = 1.21 - 1.68  
Bmsy = 5.53 , 95% CL = 3.95 - 7.74  
Biomass in last year = 0.859 , 2.5th perc = 0.669 , 97.5 perc = 1.06  
B/Bmsy in last year = 0.155 , 2.5th perc = 0.121 , 97.5 perc = 0.191  
Fishing mortality in last year = 0.115 , 2.5th perc = 0.0938 , 97.5 perc = 0.148  
Exploitation F/Fmsy = 1.44 , 2.5th perc = 1.17 , 97.5 perc = 1.84

Stock status and exploitation in 2014

Biomass = 0.859 , B/Bmsy = 0.155 , F = 0.212 , F/Fmsy = 2.64  
Comment: OK (RF 15.04.16)

-----  
CMSY also produces two graphical outputs, one for comparing the results of the Monte Carlo model with the full surplus production model, and one for use in management reports. These graphs are shown and explained on the next two pages.

Figure 1 shows assessments for sole (*Solea solea*) in the Celtic Seas. Panel A shows in black the time series of catches and in blue the three-years moving average with indication of highest and lowest catch, as used in the estimation of prior biomass by the default rules. Panel B shows the explored  $r$ - $k$  log space and in dark grey the  $r$ - $k$  pairs which were found by the CMSY model to be compatible with the catches and the prior information. Panel C shows the most probable  $r$ - $k$  pair and its approximate 95% confidence limits in blue. The black dots are possible  $r$ - $k$  pairs found by the BSM model, with a red cross indicating the most probable  $r$ - $k$  pair and its 95% confidence limits. Panel D shows the available abundance data in red, scaled to the BSM estimate of  $B_{msy} = 0.5 k$ , and in blue the biomass trajectory estimated by CMSY. Dotted lines indicate the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles. Vertical blue lines indicate the prior biomass ranges. Panel E shows in red the harvest rate (catch/abundance) scaled to the  $r/2$  estimate of BSM, and in blue the corresponding harvest rate from CMSY. Panel F shows the Schaefer equilibrium curve of catch/MSY relative to  $B/k$ , here indented at  $B/k < 0.25$  to account for reduced recruitment at low stock sizes. The red dots are scaled by BSM estimates and the blue dots are scaled by CMSY estimates.

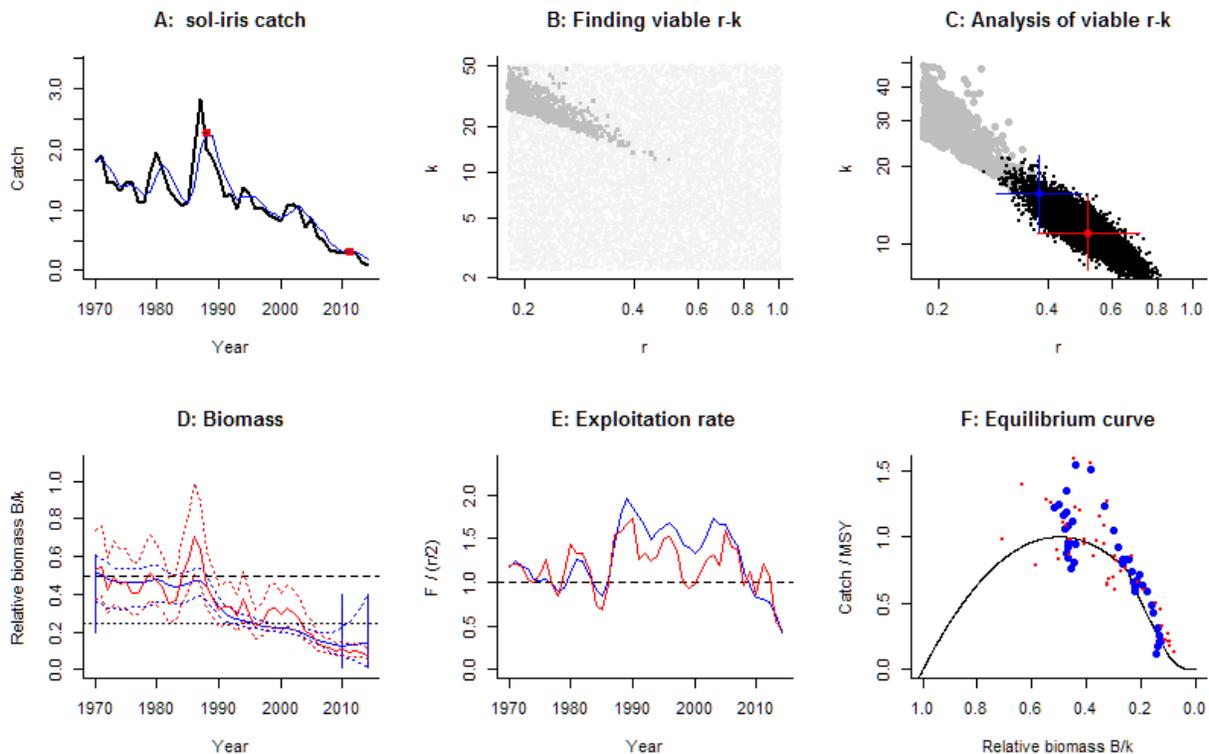


Figure 1. Results of CMSY and BSM analyses for Sole in the Irish Sea.

Figure 2 shows the graphs meant to inform management. The upper left panel shows catches relative to the BSM estimate of  $MSY$ , with indication of 95% confidence limits in grey. The upper right panel shows the development of relative total biomass ( $B/B_{msy}$ ), with the grey area indicating uncertainty. The lower left graph shows relative exploitation ( $F/F_{msy}$ ), with  $F_{msy}$  corrected for reduced recruitment below  $0.5 B_{msy}$ . The lower-right panel shows the trajectory of relative stock size ( $B/B_{msy}$ ) over relative exploitation ( $F/F_{msy}$ ).

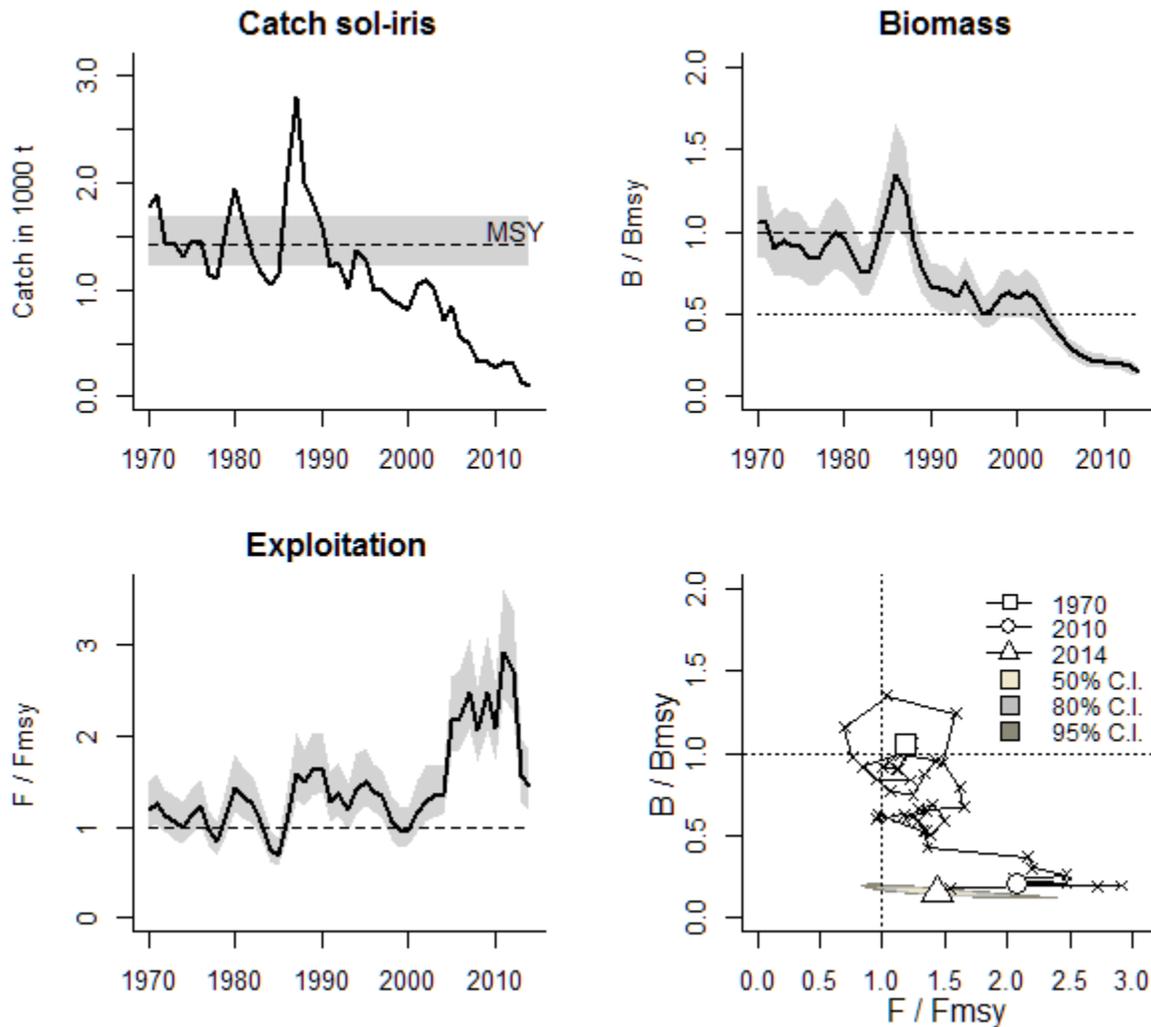


Figure 2. Graphical output of the CMSY code for management purposes, here based on the BSM assessment.

If you have any questions, contact any of the authors of this document.

Enjoy using CMSY.