

## User Guide for CMSY++

Rainer Froese, Nazli Demirel, Gianpaolo Coro and Henning Winker

Cite as: Froese, R., Demirel, N., Coro, G. and Winker, H. 2021. User Guide for CMSY++. GEOMAR, Germany, 17 pp.

Published online at <http://oceanrep.geomar.de/52147/> in March 2021

This is an accompanying document for Froese, R., Winker, H., Coro, G., Palomares, MLD., Tsikliras, A.C., Dimarchopoulou, D., Touloumis, K., Demirel, N., Scarcella, G., de Souza Vianna, G.M., Liang, C. and Pauly, D. 2021. *Catch time series as the basis for fish stock assessments: CMSY++ and its applications*, submitted in March 2021.

CMSY++ is an advanced state-space Bayesian method for stock assessment that estimates fisheries reference points ( $MSY$ ,  $F_{msy}$ ,  $B_{msy}$ ) as well as status or relative stock size ( $B/B_{msy}$ ) and fishing pressure or exploitation ( $F/F_{msy}$ ) from catch and (optionally) abundance data, a prior for resilience or productivity ( $r$ ), and broad priors for the ratio of biomass to unfished biomass ( $B/k$ ) at the beginning, an intermediate year, and the end of the time series. For the purpose of this User Guide, the whole package is referred to as CMSY++ whereas the part of the method that deals with catch-only data is referred to as CMSY (catch MSY), and the part of the method that requires additional abundance data is referred to as BSM (Bayesian Schaefer Model). Both methods are based on a modified Schaefer surplus production model (see paper cited above for more details). The main advantage of BSM, compared to other implementations of surplus production models, is the focus on informative priors and the acceptance of short and incomplete (i.e., fragmented, with missing years) abundance data. This document provides a simple step-by-step guide for researchers who want to apply CMSY++ to their own data.

The CMSY++ version referred to in this guide is a further development of the CMSY method presented in Froese et al. (2017<sup>1</sup>). The main differences are (i) the use of a full Bayesian approach with MCMC (Markov chain Monte Carlo) modelling also for the catch-only (i.e. CMSY) analysis, (ii) faster execution, (iii) the use of an AI (Artificial Intelligence) neural network to predict default biomass priors from catch, and (iv) more emphasis on graphical outputs including various analytical plots. A major improvement for both CMSY and BSM is the introduction of multivariate normal priors for  $r$  and  $k$  in log space, replacing the previous uniform prior distributions. This allowed also for a simplified determination of the 'best'  $r$ - $k$  pair in CMSY and faster run times. In addition, when CPUE data from commercial fisheries are used, assumed annual

---

<sup>1</sup> Froese R., Demirel N., Coro G., Kleisner K. and H. Winker, 2017. Estimating fisheries reference points from catch and resilience. *Fish and Fisheries*, 18: 506-526

<sup>2</sup>Palomares, M.L.D., and D. Pauly. 2019. On the creeping increase of vessels' fishing power. *Ecology and Society*, 24(3):31

increase in efficiency of the fishery, i.e., effort-creep and resulting increase in catch per unit of effort (CPUE) can be indicated, e.g. as 2% per year (Palomares and Pauly 2019<sup>2</sup>).

The R-code (CMSY++16.R), a file with the trained neural network data (ffnn.bin), and some example input files (Train\_Catch\_9e.csv, Train\_ID\_9e.csv) can be downloaded from <http://oceanrep.geomar.de/52147/>

### Installation instructions

1) Install a recent version of R on your computer. CMSY++ was tested under different R versions up to 4.0.4, 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 from <http://www.rstudio.com/products/rstudio/download/>.

3) Install the Gibbs sampler 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 and you get error messages, try commenting out (#) all the lines in the code containing the windows(..) function, because display of graphs in dedicated windows is operating system specific. We tried to capture this in the code, but it may not work in all cases.*

4) If you want to generate reports in PDF format, you have to install a LaTeX program. See the steps required in the Appendix to this document. Also, you have to set write.pdf to TRUE in the “General settings for the analysis” section of the code.

5) In order to run the code, several R-packages are required. These are downloaded and installed automatically when you run the code for the first time and are connected to the Internet. If R still alerts you of a missing package, install it through the Tools option in RStudio.

6) A file with the trained neural network data (ffnn.bin) and two data files are required by CMSY++, all of which should be placed in the same directory as the R-script. Examples are provided and their structure is explained in the next section below. The names of these files are specified in the code section “Required settings, File names”. You can replace the names of the data files with your own file names (see 8).

7) The code should set its own directory automatically as the working directory. Otherwise, use the tab “Session” in RStudio 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 for the *catch\_file* and *id\_file* variables in the “Required settings, File names” section of the code (Lines 48-49). 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 consistent use of commas; semi-colons are not accepted and will result in cryptic error messages).

9) The R-code can either analyze all stocks in the data files or a single stock as specified in the “Select stock to be analyzed” section of the code. To specify the stock to analyze just enter the unique stock name (e.g. `stocks <- “ple.27.7d”`). To make the code run on all of the stocks included in the files, comment out this line (insert # at the beginning of line 59). You can analyze the stocks in alphabetic order or in the sequence they appear in the ID file or by Region or subregion, according to settings in the “Analyze stocks” section, lines 403 and below. Uncomment (remove #) the row of code you want to use.

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 as well as in the graph windows, which can be saved manually in different formats. If `write.output` is set to TRUE in the “General settings for the analysis” section (Lines 61 and below), another output file in .csv format is produced that contains the results in Table format, which can then be opened in, e.g., Excel for documentation or further analysis.

12) **For Mac users:** Some commands are only available in Windows. For getting the plots right, Mac users should change all the *windows()* commands to *quartz()* which starts a graphics device driver for the macOS system and supports plotting both on screen and to various graphics file formats. The *savePlot()* command works fine. In some cases, Mac users might get an error with the *antialias* argument. In that case, use *antialias="default"* instead of *antialias="cleartype"*.

## 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.27.7e-k”), repeated for each year. Avoid special characters in the stock ID such as / or : or ñ or Á, as these may cause errors.

**yr:** the reporting year of the catch (e.g., 2004). One row for each year. Years have to be consecutive from the first to the last year without any missing years. If years are missing, add them and interpolate the catch. Missing years will cause an error.

**ct:** catch in tonnes (e.g. 12345.6). Do NOT use comma as indicator of thousands (12,345.6 is WRONG). Use one row for each year. Gaps with no entries are not accepted and must be filled by interpolating missing or incorrect values, e.g., do not accept zero or NA as entry if data are missing, instead use mean of adjacent values to fill any gaps.

**bt:** the value of the biomass (in tonnes, e.g. 345.67), 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 bt, i.e., abundance data can be fewer than catch data.

### Structure of ID file (.csv)

In all text variables below, avoid special characters such as / or : or ñ or Å, as these may cause errors.

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

**Continent:** a string specifying the continental shelf, such as “Europe” or “North America” or “Asia”

**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).

**Group:** optional; the functional group that a species belongs to, e.g., “Large predators” or “Pelagic plankton feeders” or “Benthic organisms”.

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

**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. If present, SpecCode can be used to lookup Resilience and r.low, r.hi priors, e.g., from the ‘Estimates based on models’ section of <https://www.fishbase.org/summary/69>, where 69 is the SpecCode of Atlantic cod.

**Source:** optional; the source where the data were taken from, e.g.,

<http://www.ices.dk/sites/pub/Publication%20Reports/Advice/2014/2014/gfb-comb.pdf>

**MinOfYear:** the first year with catch data.

**MaxOfYear:** the last year with catch data.

**StartYear:** the first year of the catch time series to be used for the analysis (from when on the data are thought to be reliable).

**EndYear:** the last year of the catch time series to be used.

**Flim, Fpa, Blim, Bpa, Bmsy, MSYBtrigger, Fmsy, last\_F:** optional; fisheries reference points from assessments, for comparison, not used in the analysis.

**Resilience:** prior estimate of resilience or productivity, 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) for fish and from [www.SeaLifeBase.org](http://www.SeaLifeBase.org) for invertebrates.

**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 default ranges of the Resilience categories (see Table 2). Check [www.FishBase.org](http://www.FishBase.org) and [www.SeaLifeBase.org](http://www.SeaLifeBase.org) for prior estimates.

**stb.low / stb.hi:** the prior biomass range relative to the unexploited biomass (B/k) at the beginning of the catch time series (see next section). Set to NA to have the neural network make prior suggestions (note: these may not be appropriate for your stock).

**int.yr:** a year in the time series for an intermediate biomass level. Set it to NA to have it estimated by default rules.

**intb.low / intb.hi:** the prior biomass range relative to the unexploited biomass (B/k) at int.yr. Set to NA to have the neural network make prior suggestions (note: these may not be appropriate for your stock).

**endb.low / endb.hi:** the prior biomass range relative to the unexploited biomass (B/k) at the end of the catch time series (see next section). Set to NA to have the neural network make suggestions (note: these may not be appropriate for your stock).

**e.creep:** An indication of assumed increase of catchability q per year in percent for commercial CPUE data, typically in the range of 1 - 5%, with 2% being a good default assumption (Palomares and Pauly 2019<sup>2</sup>). This will cause a decrease in the CPUE considered by BSM. The difference between the provided and the corrected CPUE can be made visible by setting ‘e.creep.line’ to TRUE in code line 79. This will cause the provided CPUE to be plotted in green in the biomass plot, while the used CPUE is plotted in red. No correction is needed for standardized CPUE data such as resulting from scientific surveys.

---

<sup>2</sup> Palomares, M.L.D., and D. Pauly. 2019. On the creeping increase of vessels’ fishing power. Ecology and Society, 24(3):31

**btype**: the type of information in the bt column of the catch file. Allowed values are: “biomass” (when total biomass is reported, but note that this setting gives misleading results if the biomass does not represent the whole stock that is exploited, such as if reported biomass refers to spawning stock biomass (SSB) but juveniles are already fished; set btype to CPUE instead); “CPUE” (when CPUE or CPUE index or SSB are reported); or “None” (if no abundance data are available). Remember that entries are case-sensitive.

**force.cmsy**: set to TRUE if the management analysis should use the CMSY results rather than available BSM results. Useful when the abundance data are deemed unreliable. Default is FALSE or F.

**Comment**: a comment on the stock or the quality of the analysis or special settings. 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, as this is a very common error.

### Suggestions for parameters settings

Table 1 reports a set of questions that can help to set the CMSY++ input parameters. Please note that priors are preferably derived with other stock assessment tools, such as length-frequency analysis with the LBB package (Froese et al. 2018<sup>3</sup>).

**Table 1.** Example of questions to be addressed by 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 and represent all catches?
Relative start and end biomass <i>B/k</i>	What is the most likely stock status for the beginning and end of the time series: lightly fished, fully exploited, or overfished?
Relative intermediate biomass <i>B/k</i>	Is there an intermediate year where biomass is considered to have been particular low, e.g., exploitation changed from light to full, or recruitment failure affects the stock?
Resilience prior <i>r</i>	What is your best guess for the range of values including natural mortality of adults ( <i>M</i> )? Consider the empirical relationship $r \approx 2 M$ .
Resilience prior <i>r</i>	What is your best guess for the range of values including maximum sustainable fishing mortality ( $F_{msy}$ )? Considering the relationship $r \approx 2 F_{msy}$ <i>Use this question to reinforce or change the answer to previous question</i>
Resilience prior <i>r</i>	Alternatively and better, get the prior range of <i>r</i> from the section “Estimates based on models” in the species summary page of FishBase ( <a href="http://www.fishbase.org">www.fishbase.org</a> ) or SeaLifeBase ( <a href="http://www.sealifebase.org">www.sealifebase.org</a> ).

<sup>3</sup> Froese R., Winker H., Coro G., Demirel N., Tsikliras A.C., Dimarchopoulou D., Scarcella G., Probst W.N., Dureau M., and D. Pauly, 2018. A new approach for estimating stock status from length frequency data. ICES Journal of Marine Science, 75: 2004-2015

Table 2 reports the  $r$  ranges automatically assigned by CMSY++ based on resilience categories, which can be found in [www.FishBase.org](http://www.FishBase.org) and [www.SealifeBase.org](http://www.SealifeBase.org).

**Table 2.** Prior ranges for parameter  $r$ , based on classification of resilience.

Resilience	prior $r$ range
High	0.6 – 1.5
Medium	0.2 – 0.8
Low	0.05 – 0.5
Very low	0.015 – 0.1

Table 3 suggests ranges for relative biomass to be used as input parameters, depending on the depletion status of the stock.

**Table 3.** Prior relative biomass ( $B/k$ ) ranges for CMSY++.

Depletion level	$B/k$ range	Alternative descriptions of stock status or fishery
Very strong	0.01-0.2	strongly overfished; severely depleted; collapsed; closed; abandoned; unprofitable; minimal catch; truncated size/age structure; strongly reduced recruitment; only sporadic recruitment; low abundance in much of previous area
Strong	0.01-0.4	overfished; depleted; outside safe biological limits; reduced recruitment; reduced catch; increased cost of fishing; increased effort; reduced profits; reduction of mean size in the catch and in surveys; disappearance of fish from areas where they used to be
Medium	0.2-0.6	fully exploited; maximally sustainably fished; high catch; high fishing effort; high cost of fishing but still reasonable profits; first signs of decline in average size and reduced abundance in some areas; occasional low recruitment
Low	0.4-0.8	pretty good catch; good catch per effort; high profits; many large fish; healthy size/age structure; high abundance throughout area; regular recruitment; healthy fishery
Very low	0.75-1.0	underdeveloped fishery; underfished; low market demand; only occasional catches; only bycatch; not vulnerable to common gears

When choosing the  $B/k$  prior for the intermediate year, it often improves the CMSY analysis if the intermediate  $B/k$  prior is placed after a period of sustained very high catches that are suspected to have led to low biomass by specifying a respective relative range, e.g. as 0.01-0.3. In general, the range of the  $B/k$  prior 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. If the stock was nearly unexploited, 0.75-1.0 is appropriate for the relative start biomass. Setting a range of 0.01 to 1 is also possible, and would indicate that absolutely no information about stock status is available. However, this is unlikely. If a stock is fished, it must be smaller than the unexploited level of 1.0. If it is delivering decent catches, there must be fish in the water and thus it must be larger than 0.01. See Table 1 and 3 for guidance on how to get priors from interviews with fishers or experts (or yourself).

Note that if abundance information is provided, an additional surplus production analysis is performed with BSM. This analysis is then used as default for management advice, because it is based on more information. If instead you want to use the CMSY results because you do not trust the CPUE data, set “force.cmsy” to TRUE in the respective row in the stock ID input file.

### Results of CMSY analysis

When running CMSY++ it will first do an analysis based only on catch and priors for  $r$  and  $B/k$ . Progress will be indicated on screen and in graphs (Figure 1 B). If CMSY does not find any viable points, review all your priors to check whether they are indeed plausible. Increase the final prior biomass range if it is very narrow (e.g. change 0.01-0.1 to 0.01-0.3). If at least 3 years with abundance data have been provided, the BSM model will run a surplus production analysis using catch and abundance.

The text screen output for European seabass (*Dicentrarchus labrax*) in the northern and central Bay of Biscay (bss.27.8ab) is shown below, first with only catch and priors as input (btype=None in the ID file). The text should be largely self-explanatory. For comparison, the official assessment for this stock can be found here: <https://www.ices.dk/sites/pub/Publication%20Reports/Advice/2020/2020/bss.27.8ab.pdf>

-----  
CMSY++ Analysis, Fri Mar 05 15:41:46 2021  
-----

Files Train\_Catch\_9b.csv , Train\_ID\_9c.csv , ffnn.bin read successfully  
Processing bss.27.8ab , Dicentrarchus labrax  
startbio= 0.256 0.721 default , intbio= 1999 0.19 0.583 default , endbio= 0.194 0.59 default  
Running MCMC analysis with only catch data....

-----  
Species: Dicentrarchus labrax , stock: bss.27.8ab , Seabass (Dicentrarchus labrax) in divisions 8.a-b (northern and central Bay of Biscay)  
Seabass (Dicentrarchus labrax) in divisions 8.a-b (northern and central Bay of Biscay)  
Region: Northern Europe , NA  
Catch data used from years 1985 - 2019 , abundance = None  
Prior initial relative biomass = 0.256 - 0.721 default  
Prior intermediate rel. biomass= 0.19 - 0.583 in year 1999 default  
Prior final relative biomass = 0.194 - 0.59 default  
Prior range for r = 0.35 - 0.84 expert , prior range for k = 9.03 - 37.5 , MSY prior = 2.51

Results of CMSY analysis

-----  
r = 0.431 , 95% CL = 0.314 - 0.596 , k = 25.7 , 95% CL = 18.3 - 35.6  
MSY = 2.76 , 95% CL = 2.44 - 3.19  
Relative biomass in last year = 0.472 k, 2.5th perc = 0.26 , 97.5th perc = 0.663  
Exploitation F/(r/2) in last year = 0.888 , 2.5th perc = 0.516 , 97.5th perc = 1.79

Results for Management (based on CMSY analysis)

-----  
Fmsy = 0.216 , 95% CL = 0.157 - 0.298 (if B > 1/2 Bmsy then Fmsy = 0.5 r)  
Fmsy = 0.216 , 95% CL = 0.157 - 0.298 (r and Fmsy are linearly reduced if B < 1/2 Bmsy)  
MSY = 2.76 , 95% CL = 2.44 - 3.19  
Bmsy = 12.8 , 95% CL = 9.16 - 17.8  
Biomass in last year = 11.9 , 2.5th perc = 6.16 , 97.5 perc = 19.8  
B/Bmsy in last year = 0.945 , 2.5th perc = 0.52 , 97.5 perc = 1.33  
Fishing mortality in last year = 0.192 , 2.5th perc = 0.106 , 97.5 perc = 0.4  
Exploitation F/Fmsy = 0.888 , 2.5th perc = 0.516 , 97.5 perc = 1.79  
Comment: F-age: 4-15 ; type of abundance: SSB  
-----

**Figure 1** shows the CMSY assessment for European seabass in the northern and central Bay of Biscay. The black curve in **A** shows the time series of catches and the blue curve shows the smoothed data with indication of highest and lowest catch in red, as used in the estimation of prior biomass by the default rules. The dashed purple line shows the prior estimate for MSY. Panel **B** shows the explored log  $r$ - $k$  space, in light grey potential  $r$ - $k$  pairs, and in dark grey the  $r$ - $k$  pairs which were found by the model to be compatible with the catches and the prior information. The rectangle indicates the range of the  $r$  and  $k$  priors provided in the ID file. The point in the center of the blue cross is the most likely  $r$ - $k$  pair, while horizontal and vertical error bars approximate 95% confidence limits, which are shown again in a closer view in Panel **C**. The blue curve in **D** shows the median of the biomass trajectories estimated by CMSY. Dotted lines indicate the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles. Vertical purple lines indicate the prior biomass ranges, dotted if provided by the neural network (as is here the case) and solid if set by the user. The dashed horizontal line indicates  $B_{msy}$  and the dotted line indicates half of  $B_{msy}$  as a proxy for the border to reduced recruitment. Panel **E** shows median exploitation ( $F/F_{msy}$ ) as a blue curve, with the dotted curves indicating the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles. Panel **F** shows the Schaefer equilibrium curve of catch/MSY relative to  $B/k$ , indented at  $B/k < 0.25$  to account for reduced recruitment at low stock sizes. The blue curve shows the predictions by CMSY, from first year (square) to last years (triangle).

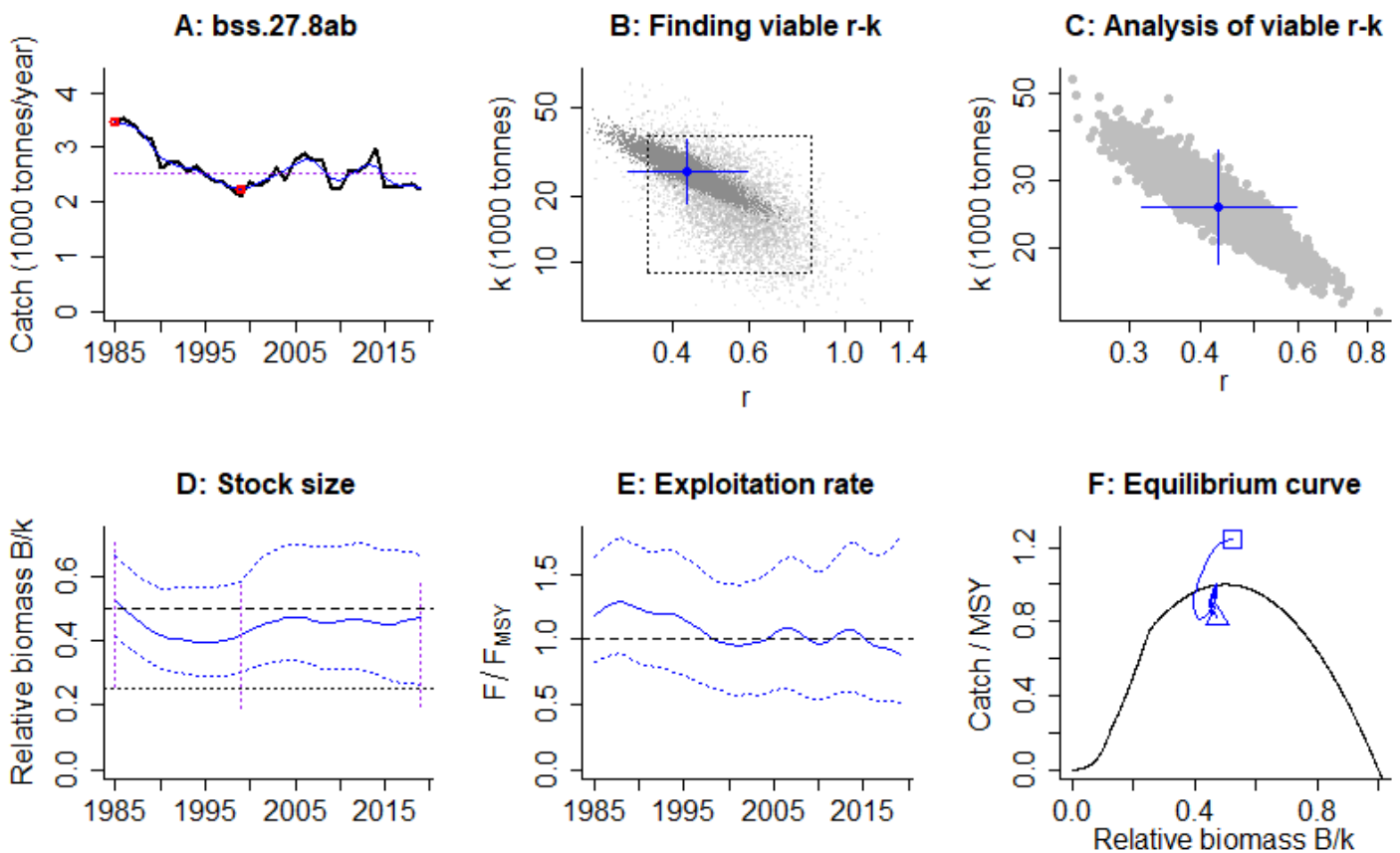


Figure 1. Results of CMSY analyses for European seabass in the northern and central Bay of Biscay.



**Figure 2** shows the graphs meant to inform management about the stock status. The upper left panel shows catches relative to  $MSY$  (dashed line) as estimated by CMSY, with indication of 95% confidence limits in light grey. The upper right panel shows the time series of predicted total biomass relative to  $B_{msy}$  (dashed line) and to the border of reduced recruitment (dotted line), with the light grey area indicating uncertainty. The lower left graph shows relative exploitation ( $F/F_{msy}$ ). The lower-right panel shows the trajectory of relative stock size ( $B/B_{msy}$ ) as a function of fishing pressure ( $F/F_{msy}$ ). The “banana” shape around the assessment of the final year triangle indicates uncertainty with yellow for 50%, grey for 80% and dark grey for 95% confidence levels.

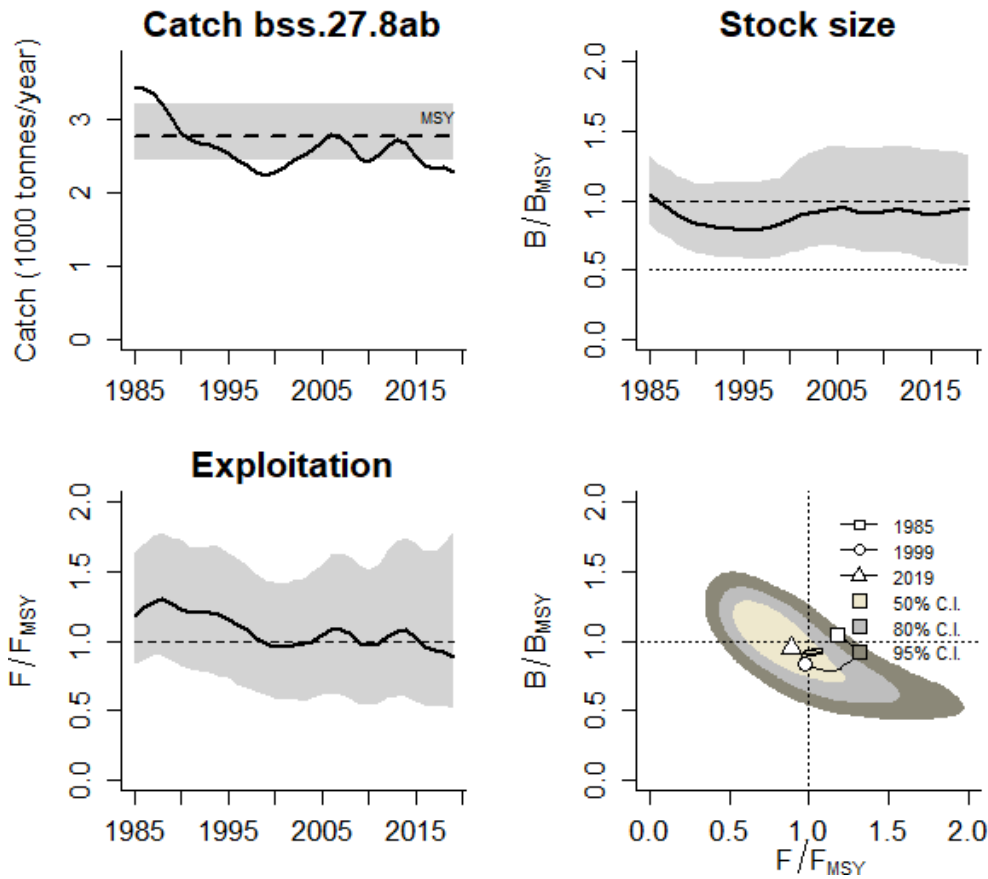


Figure 2. Graphical output of CMSY for management purposes, for European seabass in the northern and central Bay of Biscay.

**Figure 3** is a Kobe phase plot, representing the time series of pressure ( $F/F_{MSY}$ ) on the Y-axis and of state ( $B/B_{MSY}$ ) on the X-axis. The plot is divided into four quadrants, defined for the stock biomass and fishing mortality relative to  $B_{MSY}$  and  $F_{MSY}$ , respectively. The orange area indicates healthy stock sizes that are about to be depleted by overfishing. The red area indicates that the stock is overfished and is undergoing overfishing, with biomass levels being too low to produce maximum sustainable yields. The yellow area indicates reduced fishing pressure on stocks recovering from still too low biomass levels. The green area is the target area for management, indicating sustainable fishing pressure and healthy stock size capable of producing high yields close to MSY. The “banana” shape around the assessment of the final year triangle indicates uncertainty with yellow for 50%, grey for 80% and dark grey for 95% confidence levels. The legend in the upper right graph also indicates the probability of the last year falling into one of the colored areas, i.e., in this example there is a 40.1% probability that the stock is in the green area and a 34.6% probability that it is in the red area. Target would be, e.g., a higher than 75% probability that the stock is in the green area.

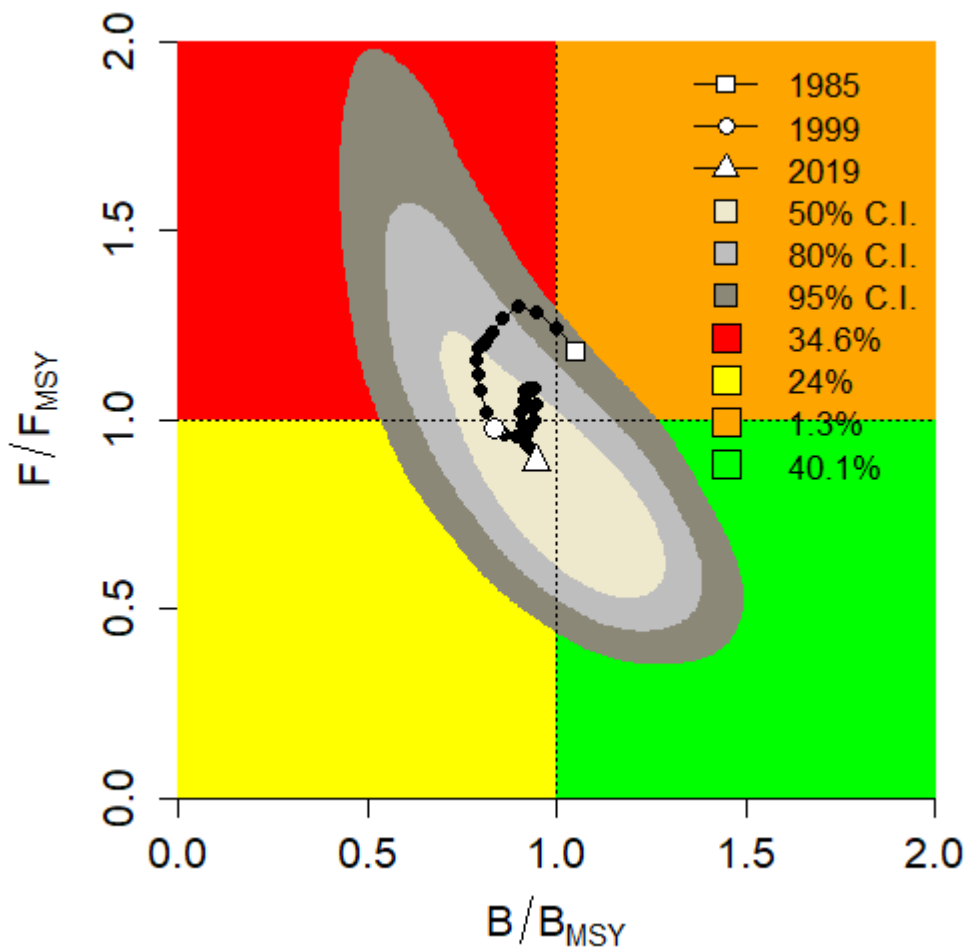


Figure 3. CMSY results for European seabass presented in a Kobe plot.

A common misconception of Bayesian analyses is that the priors determine the results. It is true that if grossly wrong priors are provided as input to CMSY, the results will be wrong. But that is true for any model provided with wrong data. If instead reasonable priors are provided, as **Figure 4** nicely shows, the priors (light grey) inform the results, with posterior understanding (dark grey) of the stock clearly improved compared to prior perceptions. The lower the prior-posterior variance ratio (PPVR), the more the posterior knowledge is improved relative to prior knowledge. If CPUE data are available, a similar graph is produced for BSM results. Display of these graphs is triggered by setting `pp.plot <- TRUE` in the “General settings for the analysis” section in the R-code.

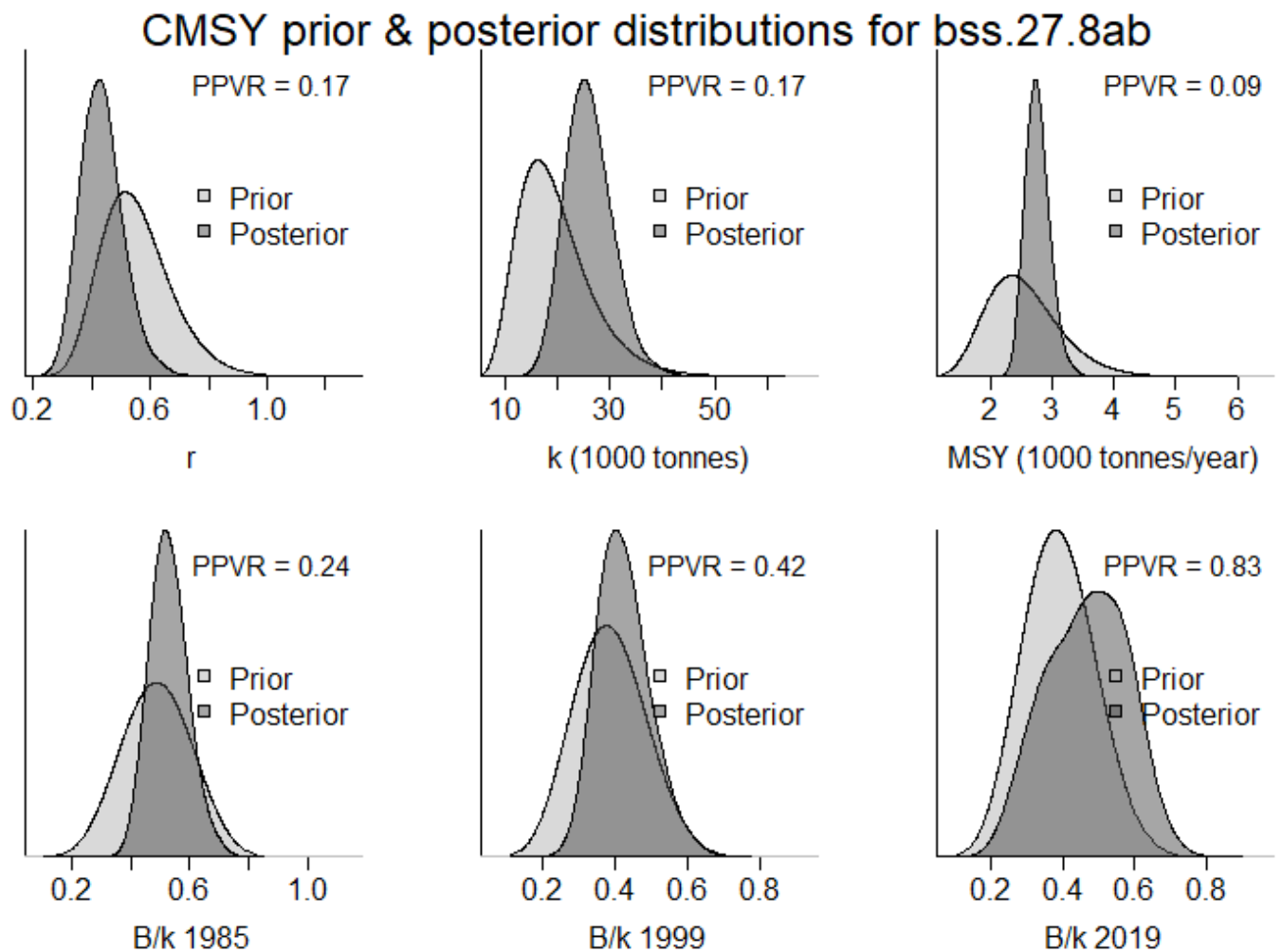


Figure 4. Comparison of prior and posterior densities (same area under curves) for resilience or productivity ( $r$ ), unexploited stock size ( $k$ ), maximum sustainable yield (MSY), and relative stock size ( $B/k$ ) at the beginning, the end, and an intermediate year of the available time series of catch data, for European seabass in the northern and central Bay of Biscay.

CMSY++ also provides an option for a retrospective analysis, i.e., a comparison of results if the last one, two, or three years of data are omitted from the analysis. The retrospective results are displayed on screen for each set of years and a new graph (Figure 5) is produced for comparing predicted time series of exploitation ( $F/F_{msy}$ ) and relative stock size ( $B/B_{msy}$ ). In the example for European seabass in the northern and central Bay of Biscay (**Figure 5**), the results are not changed much by omitting years. If, however, the predictions for all years differ substantially from those without the last year, i.e. in the presence of a strong retrospective discrepancy, then it might be prudent to, e.g. not increase allowed catch until the data for the last year are confirmed. To activate the retrospective analysis, set `retros <- TRUE` in line 82 of the code, in section General Settings for Analysis. The default setting is FALSE.

### Retrospective analysis for bss.27.8ab

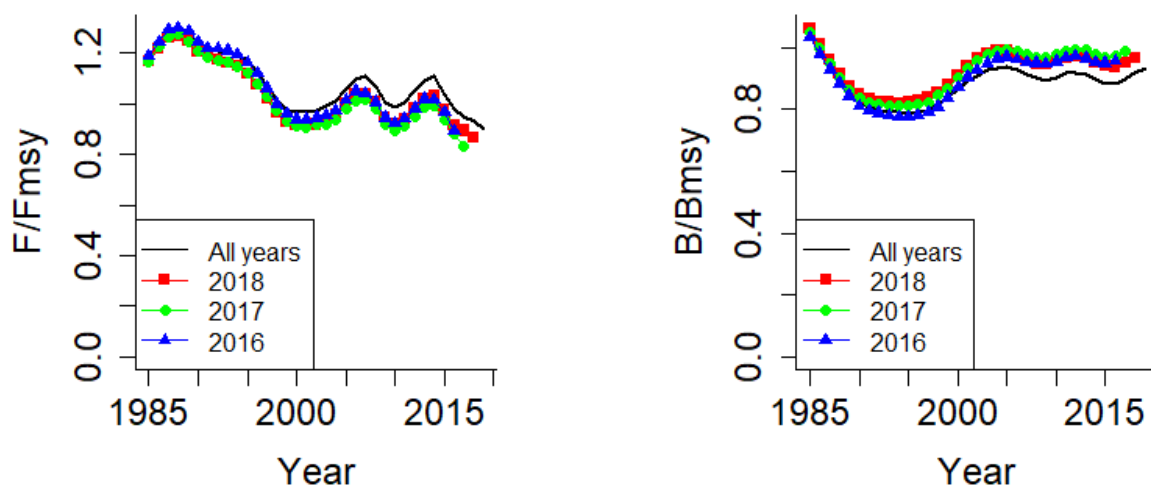


Figure 5. Comparison of predictions for exploitation ( $F/F_{msy}$ ) and relative stock size ( $B/B_{msy}$ ) when the last 1-3 years are omitted from the analysis, here for European seabass in the northern and central Bay of Biscay.

If CPUE data are available in addition to the catch data, a Bayesian state-space implementation of a full Schaefer model (BSM) is performed automatically by the CMSY++ package. The screen output then changes as indicated on the next page. If a BSM analysis is available, it is used automatically for management advice unless `force.cmsy` is set to TRUE in the ID file. Note that in this example, for demonstration purposes an increase of efficiency of fishers to find and catch the species of 1% per year was assumed by setting `e.creep` to 1 in the ID file (Palomares and Pauly 2019<sup>2</sup>).

-----  
CMSY++ Analysis, Fri Mar 05 17:17:02 2021  
-----

Files Train\_Catch\_9b.csv , Train\_ID\_9c.csv , ffnn.bin read successfully  
Processing bss.27.8ab , Dicentrarchus labrax  
startbio= 0.25 1 default , intbio= 2016 0.132 0.351 default , endbio= 0.127 0.339 default  
Running MCMC analysis with only catch data....  
Running MCMC analysis with catch and CPUE....  
-----

Species: Dicentrarchus labrax , stock: bss.27.8ab , Seabass (Dicentrarchus labrax) in divisions 8.  
a-b (northern and central Bay of Biscay)  
Seabass (Dicentrarchus labrax) in divisions 8.a-b (northern and central Bay of Biscay)  
Region: Northern Europe , NA  
Catch data used from years 1985 - 2019 , abundance = CPUE  
Prior initial relative biomass = 0.25 - 1 default  
Prior intermediate rel. biomass= 0.132 - 0.351 in year 2016 default  
Prior final relative biomass = 0.127 - 0.339 default  
Prior range for r = 0.35 - 0.84 expert , prior range for k = 8.79 - 38.2 , MSY prior = 2.51  
B/k prior used for first year in BSM and intermediate year and last year  
Prior range of q = 0.0633 - 1.1 , assumed effort creep 1 %

#### Results of CMSY analysis

-----  
r = 0.422 , 95% CL = 0.303 - 0.583 , k = 23.7 , 95% CL = 17.5 - 33.1  
MSY = 2.5 , 95% CL = 2.22 - 2.85  
Relative biomass in last year = 0.229 k, 2.5th perc = 0.138 , 97.5th perc = 0.334  
Exploitation F/(r/2) in last year = 2.21 , 2.5th perc = 1.18 , 97.5th perc = 6.19

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

-----  
q = 0.192 , lcl = 0.135 , ucl = 0.274 (derived from catch and CPUE)  
r = 0.44 , 95% CL = 0.309 - 0.63 , k = 23.6 , 95% CL = 16.9 - 34.1 , r-k log correlation = -0.95  
MSY = 2.59 , 95% CL = 2.33 - 2.92  
Relative biomass in last year = 0.266 k, 2.5th perc = 0.194 , 97.5th perc = 0.342  
Exploitation F/(r/2) in last year = 1.67 , 2.5th perc = 1.11 , 97.5th perc = 3

#### Results for Management (based on BSM analysis)

-----  
Fmsy = 0.22 , 95% CL = 0.155 - 0.315 (if B > 1/2 Bmsy then Fmsy = 0.5 r)  
Fmsy = 0.22 , 95% CL = 0.155 - 0.315 (r and Fmsy are linearly reduced if B < 1/2 Bmsy)  
MSY = 2.59 , 95% CL = 2.33 - 2.92  
Bmsy = 11.8 , 95% CL = 8.44 - 17.1  
Biomass in last year = 6.25 , 2.5th perc = 4.08 , 97.5 perc = 9.37  
B/Bmsy in last year = 0.533 , 2.5th perc = 0.388 , 97.5 perc = 0.684  
Fishing mortality in last year = 0.362 , 2.5th perc = 0.217 , 97.5 perc = 0.61  
Exploitation F/Fmsy = 1.67 , 2.5th perc = 1.11 , 97.5 perc = 3  
Comment: F-age: 4-15 ; type of abundance: SSB  
-----

If CPUE data are present and a BSM analysis is performed, the graphical output changes as shown below in **Figure 6**. The red cross in panel **B** now indicates the best  $r$ - $k$  estimate of BSM. In panel **C**, the black dots are the viable  $r$ - $k$  pairs found by BSM, with indication of a red cross for best estimate with 95% confidence limits. The red curves in panel **D** show the BSM predictions for relative biomass, the dots indicate the CPUE data scaled by BSM and corrected for effort creep, and the green line indicates the uncorrected CPUE. The red curves in panel **E** show the BSM predictions for exploitation, with the dots showing catch per CPUE as scaled by BSM. The red curve in panel **F** shows the BSM predictions for exploitation and relative stock size, with the dots showing predicted catch per predicted biomass as scaled by BSM. The management graph, the Kobe plot and the prior-posterior graphs will look similar as shown above for CMSY, but will be based on BSM results (not shown here).

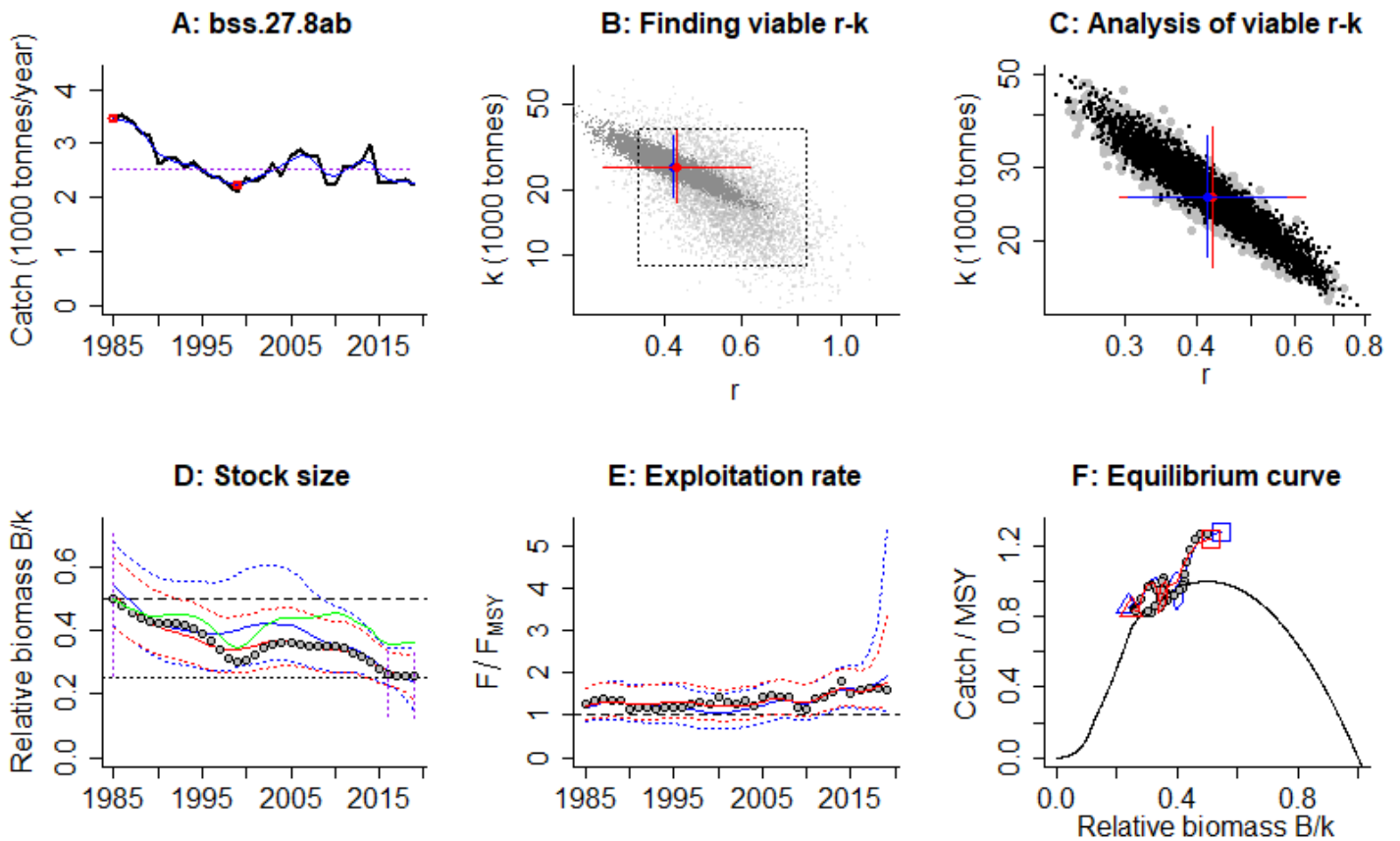


Figure 6. Analysis of European seabass in the northern and central Bay of Biscay with CMSY and BSM.

If CPUE data are available and `BSMfits.plot <- TRUE` in the “General settings for the analysis” section in the R-code, the analytical graph shown below is produced (**Figure 7**). The upper left panel shows the fit represented by the median of predicted catch posterior, with 95% confidence limits (grey shaded area), compared to the observed catch (points). The upper right panel shows a similar graph for predicted versus observed CPUE. The lower left panel shows the deviation between deterministic expectation (surplus production minus catch) and the stochastic realization (after adding process error), where a strong deviation of the bold curve from the dashed line would indicate that changes in biomass diverge from the Schaefer model expectations due to, e.g., (1) strong environmental variation, (2) CPUE not properly describing the abundance or (3) the priors being mis-specified. The lower right graph shows an analysis of the log-CPUE residuals, which should preferably be randomly distributed.

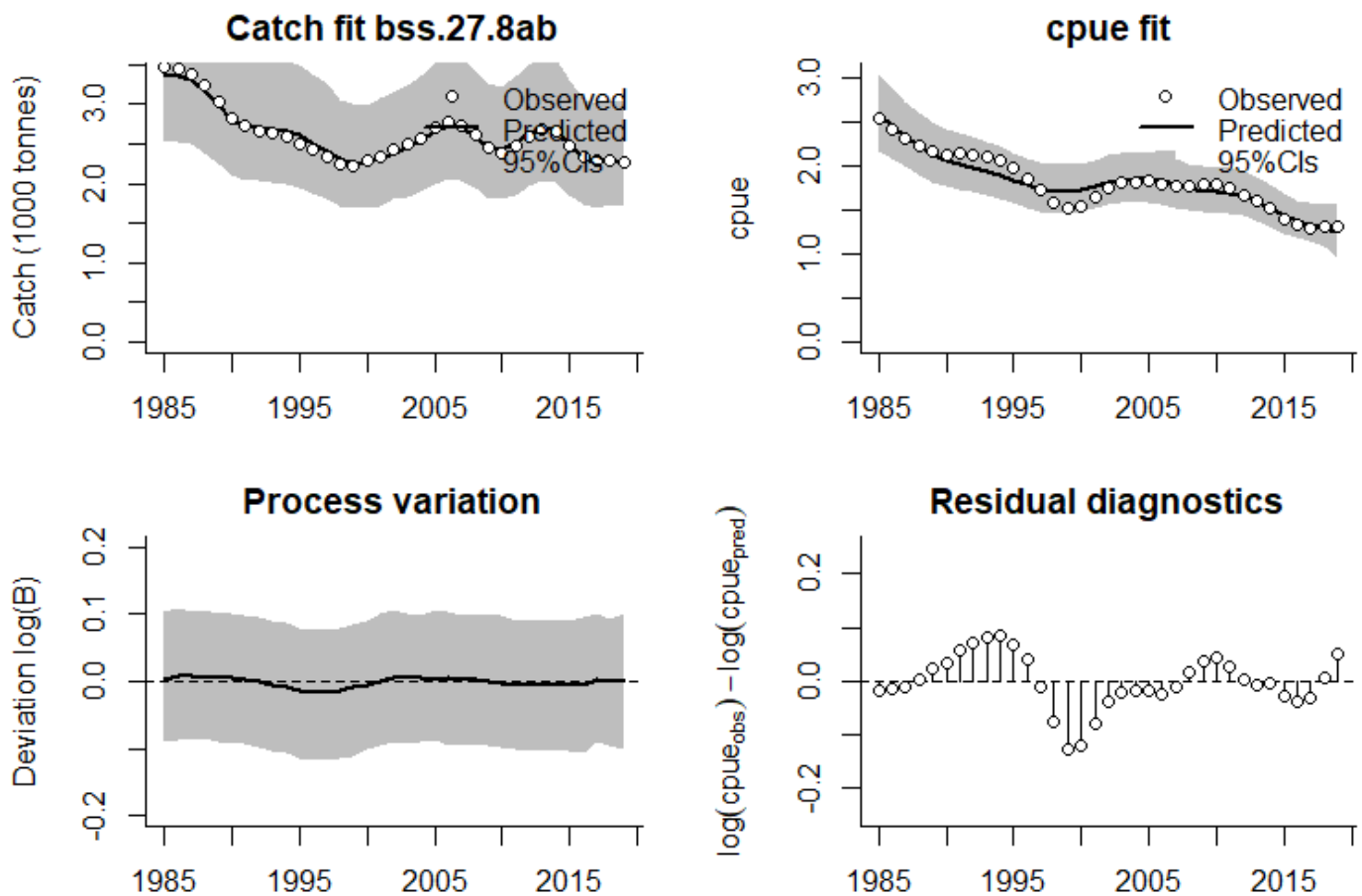


Figure 7. Analytical graph for BSM analysis of European seabass in the northern and central Bay of Biscay, showing the fit of the predicted to the observed catch, the fit of predicted to observed CPUE, the deviation from observed to predicted biomass, and an analysis of the log-CPUE residuals.

In the logistic model of population growth,  $r$  and  $k$  are inversely correlated with  $k = 4 \text{ MSY} / r$  and a slope of  $-1$  in log-space. A random distribution of  $r$ - $k$  pairs generated from that consideration is shown with blue dots in **Figure 8**. Analyzing observed  $r$ - $k$  correlations of 240 stocks gave an empirical slope of  $-0.76$ , with randomly generated dots shown in purple. Orange dots show a distribution of  $r$ - $k$  points derived from JAGS modelling based on the priors for  $r$  and  $k$ . Green dots show the posterior distribution of  $r$ - $k$  points as a result of the Bayesian modelling for European seabass in the northern and central Bay of Biscay.

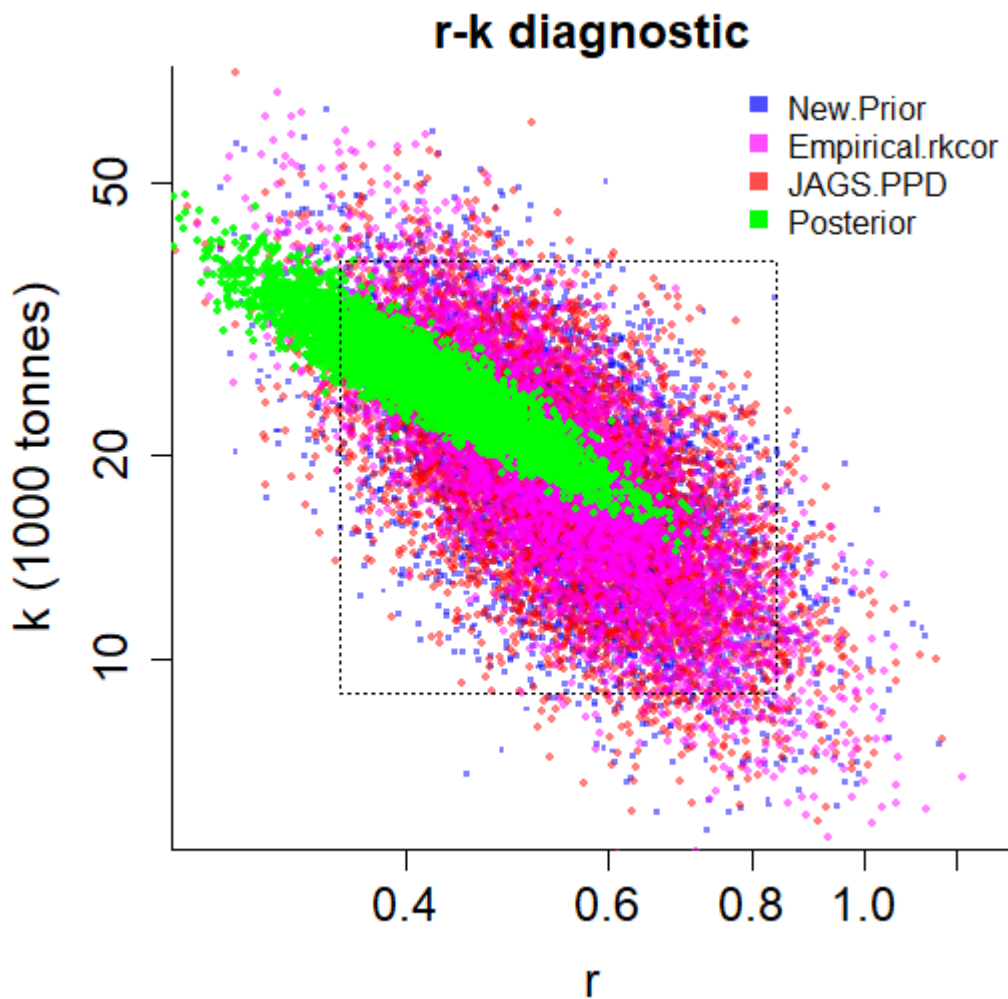


Figure 8. Diagnostic plot of different methods to generate prior distributions for  $r$  and  $k$  that take into account their negative correlation.

If you have any questions, contact us:

Rainer Froese [rfroese@geomar.de](mailto:rfroese@geomar.de)  
 Gianpaolo Coro [gianpaolo.coro@isti.cnr.it](mailto:gianpaolo.coro@isti.cnr.it)  
 Henning Winker [henning.winker@gmail.com](mailto:henning.winker@gmail.com)  
 Nazlı DEMİREL [nazdemirel@gmail.com](mailto:nazdemirel@gmail.com)

Enjoy using CMSY++

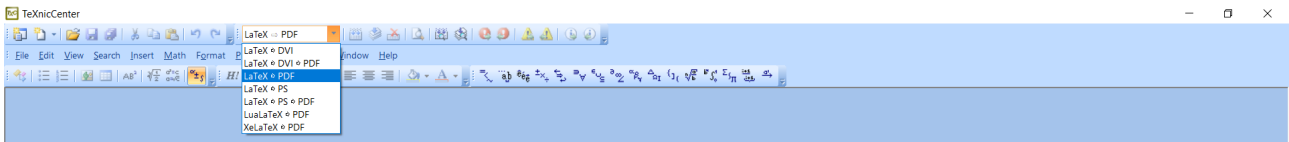


## Appendix: Generating PDFs with CMSY++

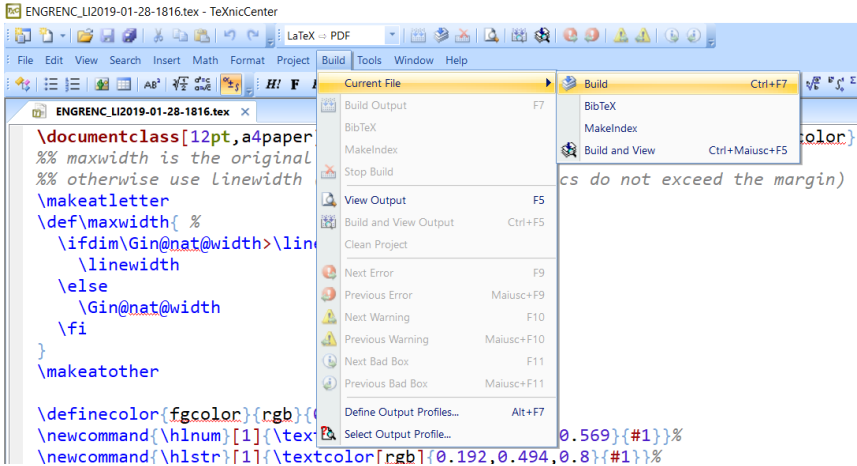
CMSY++ requires LaTeX installed and configured on the PC in order to generate PDF documents. Note that PDF generation from R is somewhat tricky and is not required for proper functioning of CMSY++.

For non-LaTeX users, one easy way to configure LaTeX for the first time and make it work with CMSY++ is to follow these steps:

1. Install MikTeX for “all PC users” and with administrator rights (<https://miktex.org/download>)
2. Install TexnicCenter for “all PC users” and with administrator rights (<http://www.texniccenter.org/download/>)
3. Run CMSY++ focused on one stock
4. Open the .tex file produced by CMSY++ with TexnicCenter and administrator rights
5. In the start configuration dialog window, select to use MikTek as compilation engine and skip the PostScript setup (Note that the installation may succeed even if there is an error message about missing LaTeX files)



7. Compile the LaTeX file by using the build command as in the following figure



8. A pop-up window will open to ask for automatic installation of required packages
9. The current latex file will be compiled and the first PDF will be generated
10. From this point on, LaTeX compilation and PDF production will work directly and transparently from CMSY++ in RStudio.
11. Remember to avoid special characters such as / or : or ñ or Á in texts as these may cause errors.