**A Simple User Guide for CMSY and BSM (CMSY\_2019\_5.R)**

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

published online at <http://oceanrep.geomar.de/33076/> in February 2019

This is an advanced version of the accompanying document for Froese R., Demirel N., Coro G., Kleisner K.M., Winker, H. (2017) Estimating Fisheries Reference Points from Catch and Resilience. *Fish and Fisheries*, 18: 506-526 DOI: 10.1111/faf.12190. CMSY is a Monte-Carlo method that estimates fisheries reference points (*MSY*, *Fmsy*, *Bmsy*) as well as relative stock size (*B/Bmsy*) and exploitation (*F/Fmsy*) 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 (BSM). 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 (= fragmented) abundance data. This document provides a simple step-by-step guide for researchers who want to apply CMSY or BSM to their own data.

The CMSY version referred to in this guide (CMSY\_2019\_5.R) is newer than the one used in Froese et al. (2017). The main differences are faster execution because of parallel processing and more emphasis on management, e.g. by the addition of an optional Kobe plot. Also, estimation of draft priors has been improved and some labels in the input files have changed, as indicated below. When CPUE data from commercial fisheries are used, assumed annual increase in efficiency of the fishery, i.e., effort-creep and resulting increase in catch per unit of effort can be indicated, e.g. as 2% per year.

The R-code and some example input files (O\_Stocks\_Catch\_15\_Med.csv and O\_Stocks\_ID\_19\_Med.csv) can be downloaded from <http://oeanrep.geomar.de/33076/>

# Installation instructions

1) Install a recent version of R on your computer. CMSY was tested under different R versions up to 3.5.2, 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 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, please comment (#) all the lines in the code containing the windows(..) function.*

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 line 64.

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

6) Two different data files are required by CMSY, which should be placed in the same directory as the script. Examples are provided and their structure is explained in next section. The names of these files are specified in the code lines 37 and 38. You can replace them with your own file names (see 9).

7) Make sure that the data files and the R script are in the same directory.

8) Open CMSY in RStudio. Use the tab “Session” and select “Set Working Directory” -> “To Source File Location”, so that the code will find the data files.

9) 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. 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).

10) 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, line 48. To specify the stock to analyze just enter the unique stock name (e.g. stocks <- “cod-scow”). To make the code run on all the stocks in the files, comment out line 50 (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 248 – 250. Uncomment (remove #) the code you want to use.

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

12) 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, another output file in .csv format contains the results in Table format, which can then be opened in e.g. Excel for documentation or further analysis.

# 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 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 bt, i.e., abundance data can be fewer than catch data.

**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. “Large predators” or “Pelagic plankton feeders” or “Benthic organisms”.

**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 start year of the catch report

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

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

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

**Flim, Fpa, Blim, Bpa, Bmsy, FMSY, MSYBtrigger, B40, M, Fofl, last\_F**: optional; fisheries reference points from assessments, for comparison, 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.ca](http://www.FishBase.ca) 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).

**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 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 prior relative biomass (B/k) range at the end of the catch time series (see next section). Set to NA if you want to use the defaults.

**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).

**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 (Pauly and Palomares, in review). 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 62. 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.

**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. 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.

# 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  *B/B0* | 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  *B/B0* | 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  *r* | What is your best guess for the range of values including natural mortality of adults (*M*)? Consider the empirical relationship *r* ≈ *2 M.* |
| Resilience prior  *r* | What is your best guess for the range of values including maximum sustainable fishing mortality (*Fmsy*)? Considering the relationship *r* ≈ 2 *Fmsy*  *Use this question to reinforce or change the answer to previous question* |

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

*r* ≈ 2 *M* ≈ 2 *Fmsy* ≈ 3 *K* ≈ 3.3/*tgen* ≈ 9/*tmax*

where *r* is the intrinsic rate of population increase, *M* is the rate of natural mortality, *Fmsy* is the maximum sustainable fishing mortality, *K* is the somatic growth rate (from the von Bertalanffy growth equation), *tgen* is generation time, and *tmax* 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. Consider that low annual fecundity (<100) will further reduce *r*. This is already considered in prior *r* estimates available from FishBase (www.fishbase.org).

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 forCMSY.

|  |  |  |  |
| --- | --- | --- | --- |
| **Very strong depletion** | **Strong depletion** | **Medium depletion** | **Low depletion** |
| 0.01 – 0.2 | 0.01 – 0.4 | 0.2 – 0.6 | 0.4 – 0.8 |

Table 3 reports the *r* ranges automatically assigned by CMSY based on resilience categories.

**Table 3**. 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 |

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 – 0.3. 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 height 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. If CMSY does not find any viable points, double-check all your priors. Increase the final prior biomass range if it is very narrow (e.g. change 0.01-0.1 to 0.01 – 0.3). Afterwards, if at least 3 years with abundance data have been provided, the BSM model will run a full surplus production model on catch and abundance.

The screen output for Anchovy (*Engraulis encrasicolus*) in Lions Gulf (ENGRENC\_LI) is shown on the next two pages.

-------------------------------------------

CMSY Analysis, Tue Jan 22 15:18:45 2019

-------------------------------------------

Parallel processing will use 8 cores

Files O\_Stocks\_Catch\_15\_Med.csv , O\_Stocks\_ID\_18\_Med.csv read successfully

Processing ENGRENC\_LI , Engraulis encrasicolus

startbio= 0.5 0.9 expert , intbio= 2005 0.01 0.4 expert , endbio= 0.01 0.3 expert

First Monte Carlo filtering of r-k space with 10000 points...

Found 640 viable trajectories for 381 r-k pairs

Shrinking k space: repeating Monte Carlo in the interval [ 17.2 , 135 ]

Attempt 1 of 3 with 10000 additional points...

Found altogether 2218 viable trajectories for 1107 r-k pairs

Final sampling in the tip area above r = 0.356 with 20000 additional points...

Found altogether 3483 viable trajectories for 2093 r-k pairs

Running MCMC analysis....

---------------------------------------

Species: Engraulis encrasicolus , stock: ENGRENC\_LI

Anchovy in Lions Gulf

Region: Mediterranean , Lions Gulf

Catch data used from years 1970 - 2014 , abundance = CPUE

Prior initial relative biomass = 0.5 - 0.9 expert

Prior intermediate rel. biomass= 0.01 - 0.4 in year 2005 expert

Prior final relative biomass = 0.01 - 0.3 expert

Prior range for r = 0.26 - 1.2 expert, , prior range for k = 9.72 - 173

Prior range of q = 1.63 - 6.89

Results of CMSY analysis

-------------------------

Altogether 3483 viable trajectories for 2093 r-k pairs were found

r = 0.799 , 95% CL = 0.559 - 1.14 , k = 38.7 , 95% CL = 24.1 - 62.2

MSY = 7.72 , 95% CL = 6.12 - 9.74

Relative biomass in last year = 0.137 k, 2.5th perc = 0.0197 , 97.5th perc = 0.29

Exploitation F/(r/2) in last year = 0.896

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

------------------------------------------------------------

q = 2.66 , lcl = 1.98 , ucl = 3.59

r = 0.581 , 95% CL = 0.407 - 0.827 , k = 47 , 95% CL = 35.9 - 61.7

MSY = 6.83 , 95% CL = 6.07 - 7.68

Relative biomass in last year = 0.261 k, 2.5th perc = 0.149 , 97.5th perc = 0.355

Exploitation F/(r/2) in last year = 0.53

Results for Management (based on BSM analysis)

-------------------------------------------------------------

Fmsy = 0.29 , 95% CL = 0.204 - 0.414 (if B > 1/2 Bmsy then Fmsy = 0.5 r)

Fmsy = 0.29 , 95% CL = 0.204 - 0.414 (r and Fmsy are linearly reduced if B < 1/2 Bmsy)

MSY = 6.83 , 95% CL = 6.07 - 7.68

Bmsy = 23.5 , 95% CL = 17.9 - 30.9

Biomass in last year = 12.3 , 2.5th perc = 6.99 , 97.5 perc = 16.7

B/Bmsy in last year = 0.523 , 2.5th perc = 0.297 , 97.5 perc = 0.709

Fishing mortality in last year = 0.154 , 2.5th perc = 0.113 , 97.5 perc = 0.271

Exploitation F/Fmsy = 0.53 , 2.5th perc = 0.391 , 97.5 perc = 0.933

Comment: Catch=landings from FishStat (Spain, France), Biomass from MEDIAS for GSA7. RF

2005 0.01-0.4, final 0.3. GS OK

----------------------------------------------------------

Figure 1 shows assessments for anchovy (*Engraulis encrasicolus*) in the Lions Gulf. 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 log *r-k* space and in dark grey the *r-k* pairs which were found by the 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 *Bmsy* = 0.5 *k*, and blue the biomass trajectory estimated by CMSY. Dotted lines indicate the 2.5th and 97.5th 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.

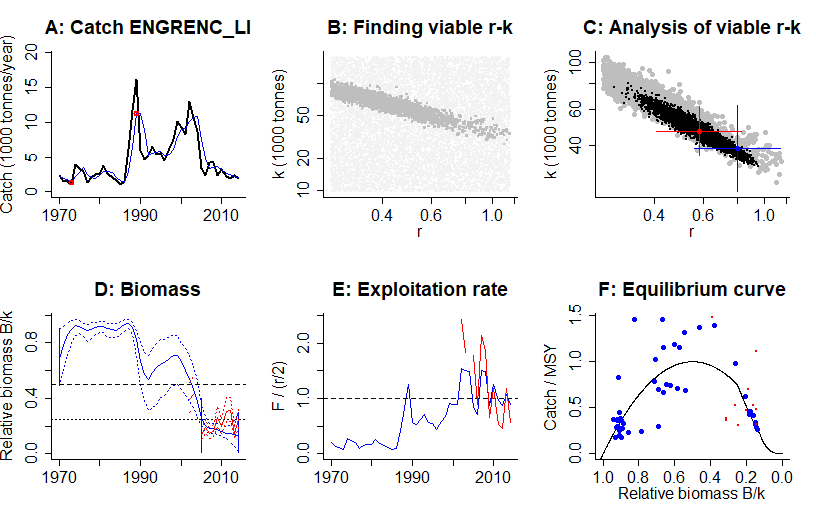


Figure 1. Results of CMSY and BSM analyses for anchovy in the Lions Gulf.

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/Bmsy*), with the grey area indicating uncertainty. The lower left graph shows relative exploitation (*F/Fmsy*), with *Fmsy* corrected for reduced recruitment below 0.5 *Bmsy*. The lower-right panel shows the trajectory of relative stock size (*B/Bmsy*) as a function of fishing pressure (*F/Fmsy*). 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.

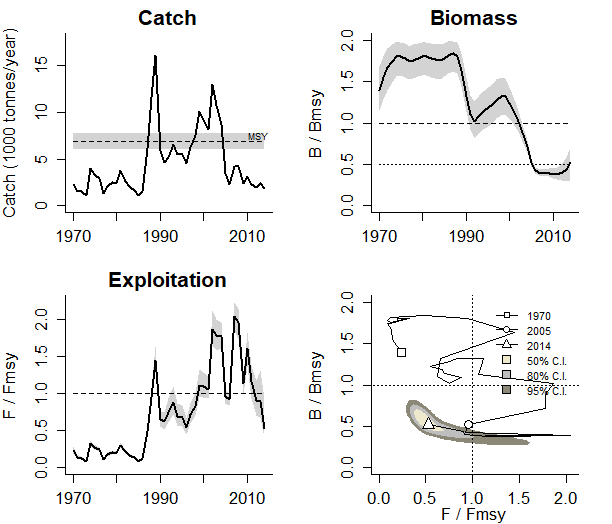


Figure 2. Graphical output of the CMSY code for management purposes, here based on the BSM assessment.

Figure 3 is a Kobe plot representing the time series of pressure (*F*/*FMSY*) on the Y-axis and of state (*B*/*BMSY*) on the X-axis. The brown area indicates healthy stock sizes that are about to be depleted by overfishing. The red area indicates ongoing overfishing while the stock is too small to produce maximum sustainable yields. The yellow area indicates reduced fishing pressure on stocks recovering from still too small biomass. 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 percentage of the 95% confidence area falling into the respectively colored areas.

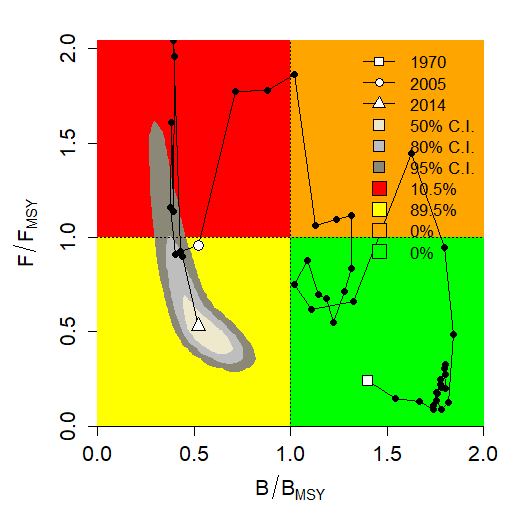


Figure 3. CMSY results presented in a Kobe plot.

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)

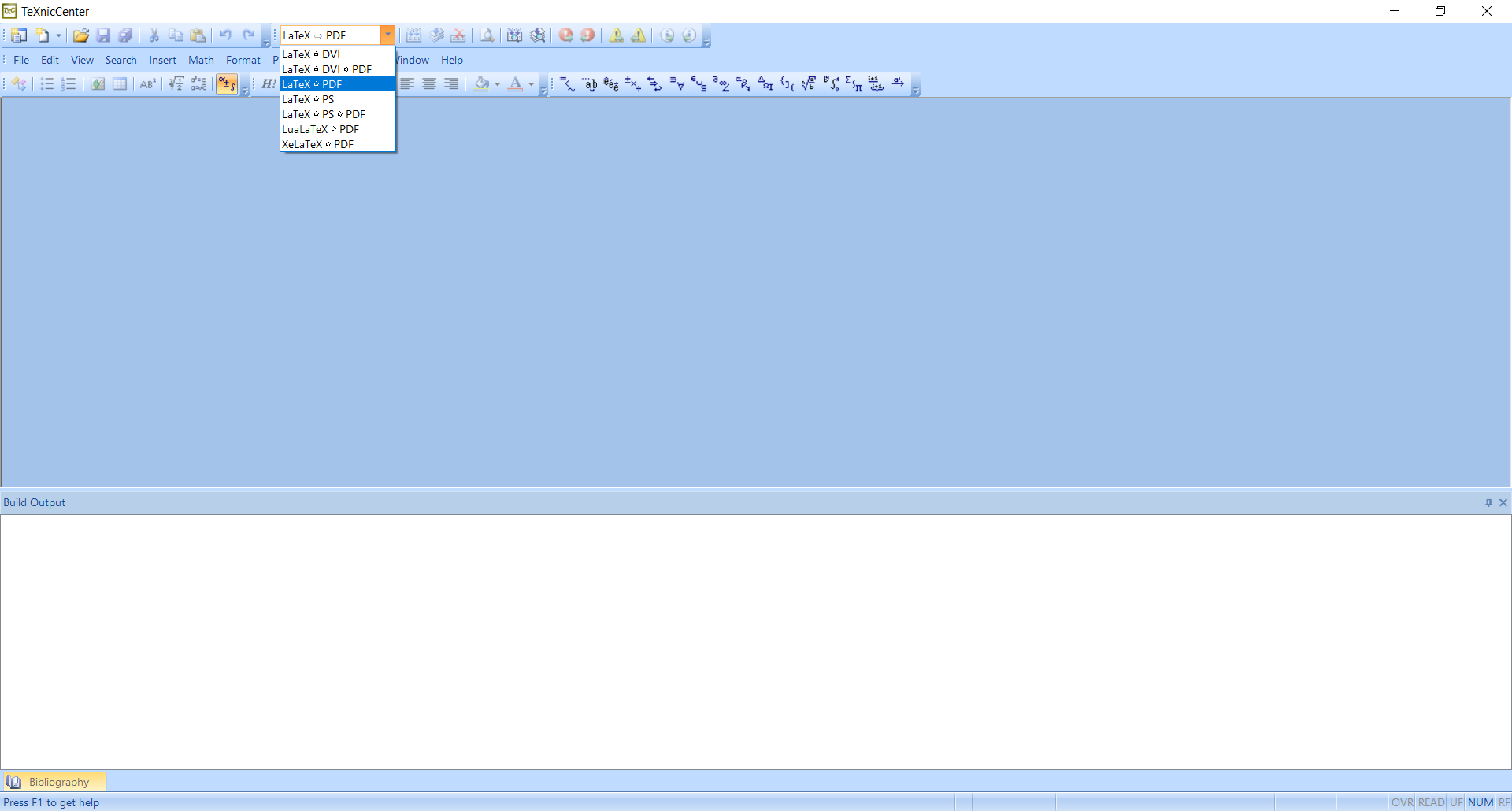
Nazlı DEMİREL [nazdemirel@gmail.com](mailto:nazdemirel@gmail.com)

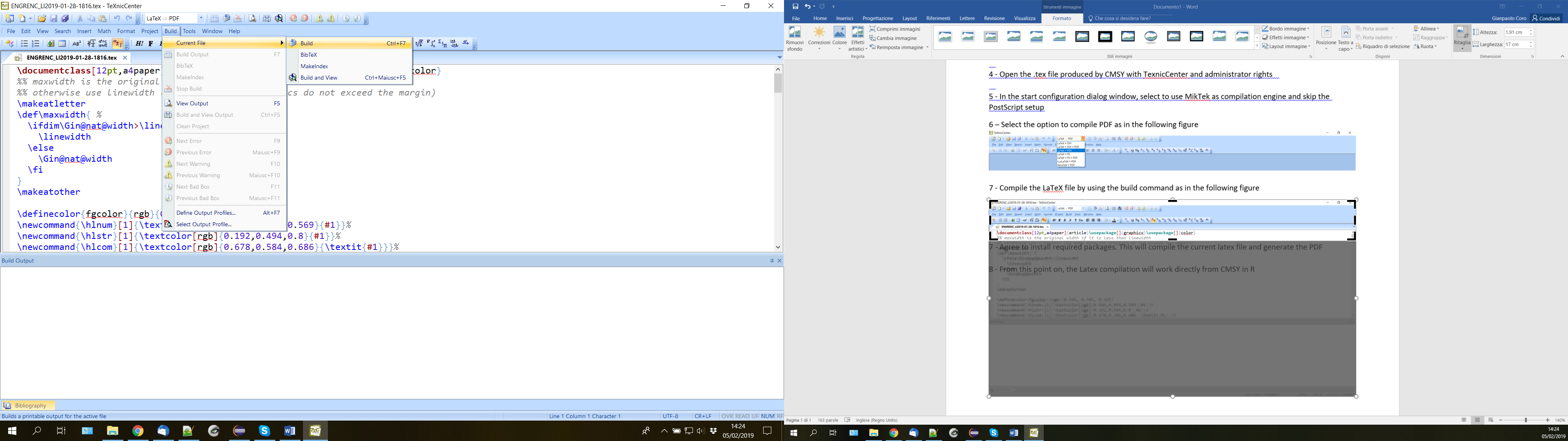
Enjoy using CMSY and BSM.

Appendix: Generating PDFs with CMSY

CMSY requires LaTeX installed and configured on the PC in order to generate PDF documents.

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 focussed 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
6. Select the option to compile PDF, as reported in the following figure
7. Compile the LaTeX file by using the build command as in the following figure



1. A pop-up window will open to ask for automatic installation of required packages
2. The current latex file will be compiled and the first PDF will be generated
3. From this point on, LaTeX compilation and PDF production will work directly and transparently from CMSY in RStudio