

BioIndex (v. 2.1.3) - Tutorial

Walter Zupa, Loredana Casciaro, Isabella Bitetto, Maria Teresa Spedicato

Coispa Tecnologia & Ricerca - Stazione sperimentale per lo Studio delle Risorse del Mare

February 18, 2020

BioIndex routine (version 2.1.3) is developed with Rstudio running R version 3.6.1.

Download the precompiled binary distribution of the base system (version 3.6.1) from the **Comprehensive R Archive Network (CRAN)** <https://cran.r-project.org/link> and install it in your preferred folder. When the software is completely installed start the R software or open Rstudio with the right R version.

BioIndex routine needs that some supplementary libraries are installed (this step is necessary only the first time that you run the routine on your computer):

- sp
- hms
- mgcv
- rgdal
- tcltk2
- ggplot2
- gridExtra
- svDialogs

There are three possibilities to install the libraries:

1. run commands from the console (in this case you need an internet connection):

```
install.packages("sp")
install.packages("hms")
install.packages("mgcv")
install.packages("rgdal")
install.packages("tcltk2")
install.packages("ggplot2")
install.packages("gridExtra")
install.packages("svDialogs")
```

2. install from repositories: (also in this case you need an internet connection):
 - open the menu *Packages* and then *Install package(s)*
 - select the closest mirror to you position

- select the zip file(s)
- select the package(s) to be installed

3. install from zip files:

- download libraries from the **CRAN** <https://cran.r-project.org/link>
- open the menu *Packages* and then *Install package(s) from local files...*
- select the zip file(s)

Initialization of the routine

The BioIndex routine is distributed as a zip file: *BioIndex_2.1.3.zip*. After you have extracted all the files from the zip in a preferred folder on your PC, you will find the main script *BioIndex_2.1.3.R* and three folders: *input*, *output*, *scripts*. Other files and folders are part of the *BioStand* routine.

The TA, TB and TC files (.csv format with semicolon as values' separator) should be placed in the respective directories of the *input* folder. In the *input* folder there is also a .csv file (*maturity_sizes.csv*) that should be compiled with the values of the cutoffs to be used for the selection of recruits and spawners (adults).

The *output* folder is used by the routine to store the outputs of the analyses. At the end of each analysis it is strongly recommended to store the outputs in another folder to exclude the possibilities of overwriting the files.

The folder *scripts* contains all the functions and utilities used by the routine to perform the analyses.

Running the routine

First open the main script of the routine with the name *BioIndex_2.1.3.R* in R environment or in Rstudio.

- in R environment: open the menu *File* and then the sub-menu *Open script*. Select from the folder the file *BioIndex_2.1.3.R*.
- in Rstudio environment: open the menu *File* and then the sub-menu *Open file*. Select from the folder the file *BioIndex_2.1.3.R*.

The first part of the routine consists in loading the previously installed libraries.

To run the command in R select with the mouse the row of the script with the command you want to run and press the following buttons combination: <CTRL>+R

if you are using Rstudio you need to press the following buttons combination: <CTRL>+<ENTER>

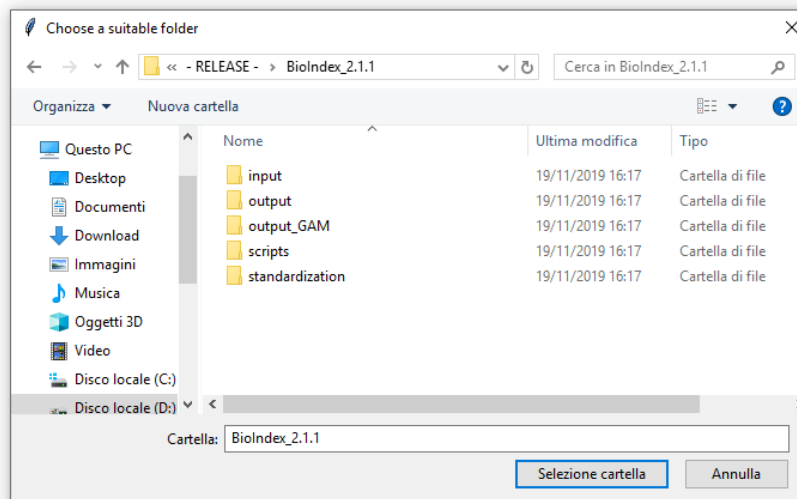
At this step you have 2 alternatives:

- 1) run the whole script at once with default settings clicking the button "source" in Rstudio.
- 2) You can run the code step by step following the instructions below described.

```
#####
#
# BioIndex v.2.1.3
# Developed on R 3.6.1 (x86-64bit)
# February 2020
#
# Authors:
# Walter Zupa, Loredana Casciaro, Isabella Bitetto,
# Maria Teresa Spedicato
# Coispa Tecnologia & Ricerca
# Stazione sperimentale per lo Studio delle Risorse del Mare
#
# For info and suggestions please contact zupa@coispa.eu
#
#####
```

Once all the libraries are correctly loaded without errors you have to select the working directory. Running the following commands the routine opens a window for the assisted selection of the folder. The routine runs correctly only whether the folder containing the *BioIndex_2.1.3.R* file is selected.

```
#-----
# Selection of the working directory
#-----
wd <- tk_choose.dir(getwd(), "Choose a suitable folder")
setwd(wd)
```

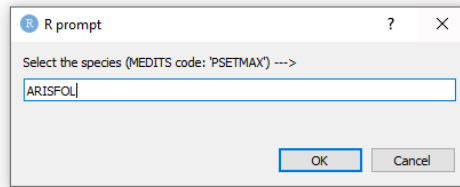


The initialization of the routine consists in the selection of the species for the analysis.

ATTENTION: Use both uppercase and lowercase letters without spaces to write the name of the species in the form of the MEDITS code, as they are reported in the MEDITS manual (AAVV, MEDITS-Handbook. Version n. 9, 2017. <http://www.sibm.it/MEDITS2011/principaledownload.htm>)

Running the following code a box will pop-up for the selection of the species. “ARISFOL” is set as default value.

```
#-----  
# initialization  
#-----  
source(paste(wd, "/scripts/initialization.r", sep=""))
```



For the analysis the routine uses meta-database files that have origin from the merge of the TA file with TB and TC ones. Running the following code the routine will produce two tables containing the merge results that will be saved in the output folder as .csv files:

- *mergeTATB_ARISFOL.csv*
- *mergeTATC_ARISFOL.csv*

```
#-----  
# Merge TA-TB files  
# Merge TA-TC files  
#-----  
source(paste(wd, "/scripts/Merge_TA-TB-TC.R", sep=""))
```

In case the study area is a GSA in which 2 or more countries are included, the code allows to perform the analysis at GSA level or at country level. Hence, the user is asked to answer to the following question:

```
Countries  
1      ALB  
2      ITA  
3      MON  
There are 3 countries in the TA file.  
Do you want to perform the analysis on the entire GSA area?  
  
1: Yes  
2: No  
  
selection: |
```

In case the choice is 2 (NO), the user is asked to select the reference country:

```
select the country for the analysis (use only numbers) ->  
  
1: ALB  
2: ITA  
3: MON  
  
selection: 2|
```

```
## TA file correctly read
## TB file correctly read
## TC file correctly read
## Merging TA-TB files
## TA-TB files correctly merged
## Merge TA-TB files saved in the following folder: 'F:/COISPA/R_BioIndex_1.4/output/mergeTATB_ARISFOL.csv'
## Merging TA-TC files
## TA-TC files correctly merged
## Merge TA-TC files saved in the following folder: 'F:/COISPA/R_BioIndex_1.4/output/mergeTATC_ARISFOL.csv'
```

In the next step a check of the survey is performed assessing whether the following condition:

- Was the same gear used along the years?
- Was the same vessel used along the years?
- Was the same number of hauls used along the years?
- Were hauls allocated in the same position along the years?

The first three checks were done by the routine:

```
#-----
# checks of the survey
#-----
source(paste(wd, "/scripts/checks.r", sep=""))

During the survey 3 different gears were used. The standardization of the indices
with GLM/GAM models is suggested.

During the survey 3 different vessels were used. The standardization of the indices
with GLM/GAM models is suggested.

During the survey a different number of hauls per year was used. The standardization
of the indices with GLM/GAM models is suggested.
```

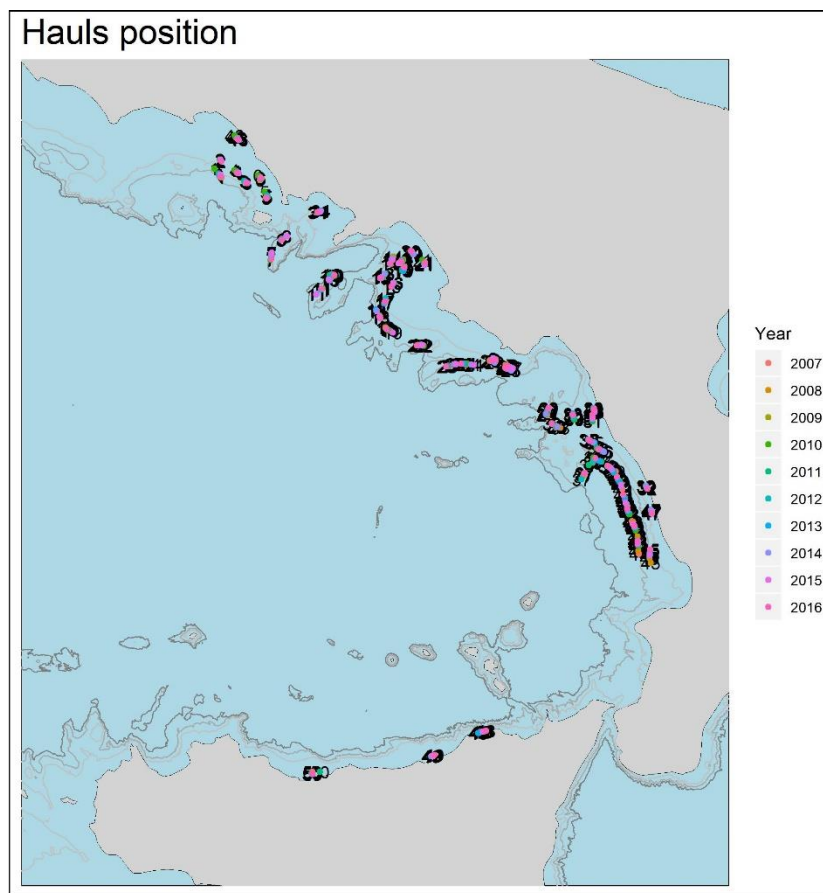
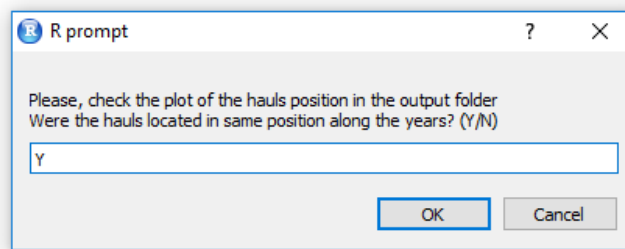
A plot of the hauls position is generated to allow the qualitative check of the hauls allocation during the years. Hence, the user is asked to answer to the following question.

ATTENTION: check that the coordinate range of the GSA is uploaded in the file
"~/script/utilities/GSAs_coordinates.csv"

An example of the check results is reported hereafter:

```
Please, wait a moment. Elaboration of the hauls plot in progress
Regions defined for each Polygons
Regions defined for each Polygons
Bubble plot of Hauls position correctly saved
Please, check the plot of the hauls position in the output folder
```

3 checks over 4 indicate/s that the standardization of the indices with GLM/GAM models is recommended



In case the checks have negative answers, the user is addressed toward the standardization of the indices with GLM/GAM models.

Some analyses use the 30" GFCM geographical grid (approximately 30NM on the latitude). Running the following commands, the catch data and biological data are merged with the grid producing two meta-DB files:

- *ARISFOL - allGSAs_metaDB_catch in GRID.csv*
- *ARISFOL - allGSAs_metaDB_biological in GRID.csv*

```
#-----
# Overlay metaDB with the grid
#-----
source(paste(wd, "/scripts/overlayDBinGRID.r", sep=""))

## Catch metaDB saved in the following folder: 'F:/COISPA/R_BioIndex_2.1.3/output/
ARISFOL - allGSAs_metaDB_catch in GRID.csv

## Biological metaDB saved in the following folder: 'F:/COISPA/R_BioIndex_2.1.3/ou
tput/ARISFOL - allGSAs_metaDB_biological in GRID.csv
```

In case some of the records included respectively in catch data and biological data merged with the GFCM grid fall out of the grid, a warning message is returned with the list of these critical records:

```
Warning messages:
MergeTATB: Some records have the mean coordinate out of the grid. They could be li
kely on the land.
Check and correct the returned list of data.
If you continue, these data will be removed from the spatial analysis.

MergeTATC: Some records have the mean coordinate out of the grid. They could be li
kely on the land.
Check and correct the returned list of data.
If you continue, these data will be removed from the spatial analysis.
```

Before continuing with the estimation of the indices, the user is asked to answer to other questions in order to categorize the sampling protocol adopted, in order to continue the estimation of the indices using the opportune formulas.

```
#-----
# Questionnaire
#-----
source(paste(wd, """/scripts/Questionnaire.r ", sep=""))

## Catch metaDB saved in the following folder: 'F:/COISPA/R_BioIndex_2.1.3/output/
ARISFOL - allGSAs_metaDB_catch in GRID.csv

## Biological metaDB saved in the following folder: 'F:/COISPA/R_BioIndex_2.1.3/ou
tput/ARISFOL - allGSAs_metaDB_biological in GRID.csv
```

```
Was any stratification criteria used in the survey?
```

```
1: Yes  
2: No
```

```
selection: 1
```

```
Was the allocation of the hauls in the strata proportional to the strata surface?
```

```
1: Yes  
2: No
```

```
selection: 2
```

```
the sampling schema is classified as  
'Random stratified sampling with post-stratification'
```

```
Please, continue with the opportune analysis  
> |
```

At this point the user should continue running the part of the script corresponding to the sampling protocol

Case 1

```
#####  
##### Simple Random sampling #####  
#####
```

Case 2

```
#####  
##### Random stratified sampling #####  
#####
```

Case 3

```
#####  
##### Random stratified sampling with post-stratification #####  
#####
```

The routine continues estimating the time series of abundance (n/km^2) and biomass (kg/km^2) indices.

Running the following code a box pops-up for the selection of the depth range to be applied to the analysis. The box will not appear in case of the "Simple Random Sampling"

```
#-----  
# Abundance and biomass indices per GSA in the timeseries  
#-----  
source(paste(wd, "/scripts/Indices_timeseries_(Random_Penalized_Sampling).R ", sep  
=""))  
## [1] "Select the depth range for the analysis"
```

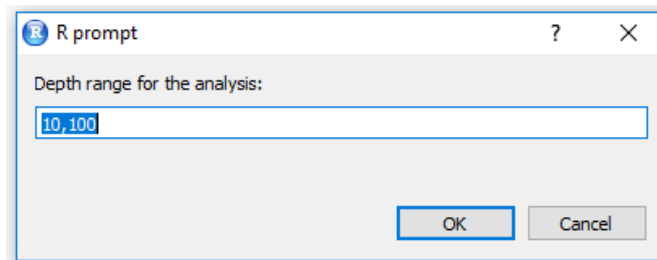

To estimate the indices time series in the case of Random Stratified (RSS) and Random Penalized Sampling (RPS, Random stratified sampling with post-stratification) the user have to define the stratification scheme adopted during the survey. BioIndex is able to use up to 6 strata defined by the user.

ATTENTION: check that the strata information were updated in the “~\scripts\utilities\strata.csv” file.

GSA	COUNTRY	CODE	MIN_DEPTH	MAX_DEPTH
10	ITA	1	10	50
10	ITA	2	50	100
10	ITA	3	100	200
10	ITA	4	200	500
10	ITA	5	500	800
...

The estimation of the indices could be performed using all the different assemblage of the contiguous strata, defining the minimum and maximum values of the depth range to be used:

ATTENTION: separate the values only with comma.



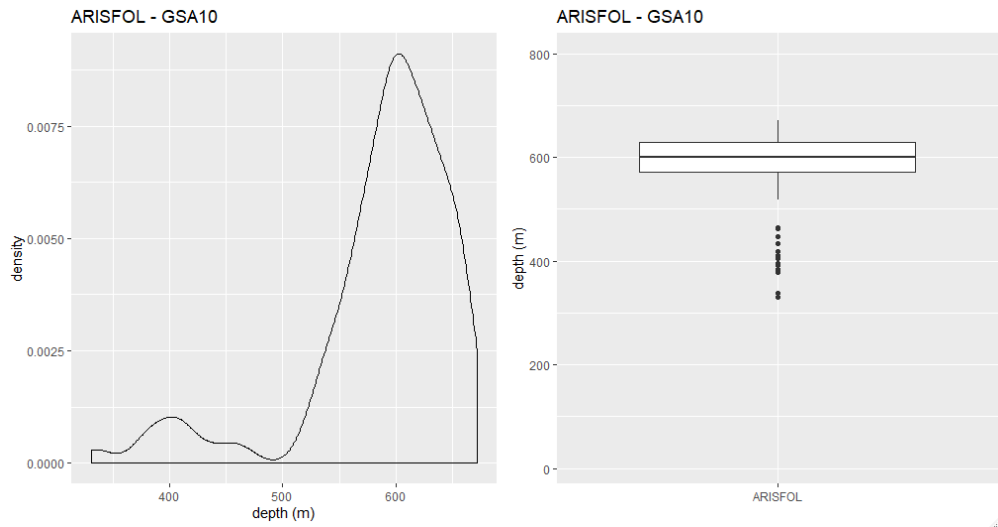
R prompt

Depth range for the analysis:

10,100

OK Cancel

The extension of the depth range for the studied species could be assessed using the plots saved in the output folder showing the occurrence frequency and a box plot of positive hauls in the 0-800m depth range.

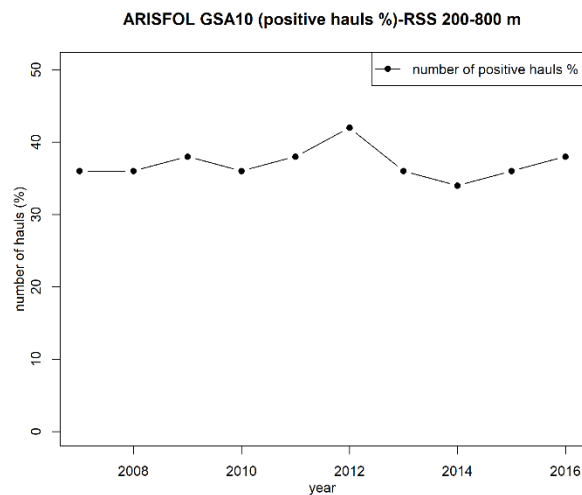


The following files are saved in the output folder:

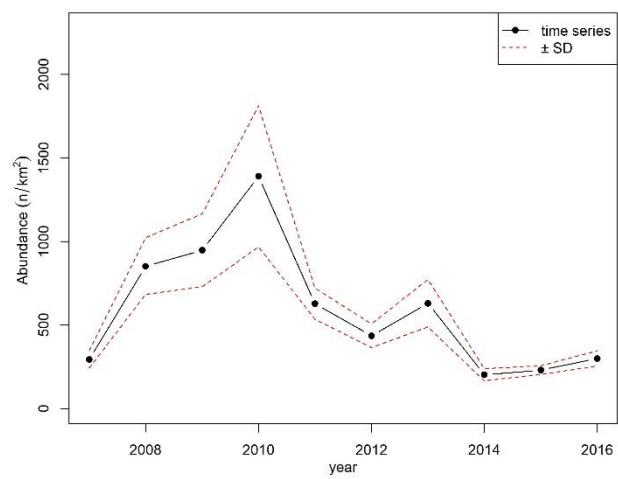
- *depth.distribution_(boxplot)ARISFOL_GSA10.jpg*
- *depth.distribution_ARISFOL_GSA10.jpg*

Moreover, the following .csv and .tiff files of the estimated indices are saved:

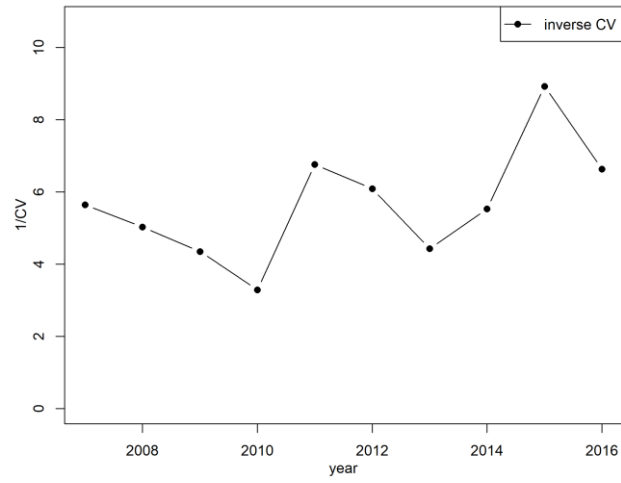
- *ARISFOL_GSA10_(positive hauls)-Random_Stratified_Sampling_200-800 m_Timeseries.tiff*
- *ARISFOL_GSA10_(abundance)-Random_Stratified_Sampling_200-800 m_Timeseries.csv*
- *ARISFOL_GSA10_(abundance)-Random_Stratified_Sampling_200-800 m_Timeseries.tiff*
- *ARISFOL_GSA10_(abundance)-FEMALES-Random_Stratified_Sampling_200-800 m_Timeseries.tiff*
- *ARISFOL_GSA10_(abundance)-FEMALES-Random_Stratified_Sampling_200-800 m_Timeseries.csv*
- *ARISFOL_GSA10_(abundance)-MALES-Random_Stratified_Sampling_200-800 m_Timeseries.tiff*
- *ARISFOL_GSA10_(abundance)-MALES-Random_Stratified_Sampling_200-800 m_Timeseries.csv*
- *ARISFOL_GSA10_(inverseCV of abundance)-Random_Stratified_Sampling_200-800 m_Timeseries.tiff*
- *ARISFOL_GSA10_(biomass)-Random_Stratified_Sampling_200-800 m_Timeseries.csv*
- *ARISFOL_GSA10_(biomass)-Random_Stratified_Sampling_200-800 m_Timeseries.tiff*
- *ARISFOL_GSA10_(MIW)-Random_Stratified_Sampling_200-800_Timeseries.csv*
- *ARISFOL_GSA10_(MIW)-Random_Stratified_Sampling_200-800 m_Timeseries.tiff*



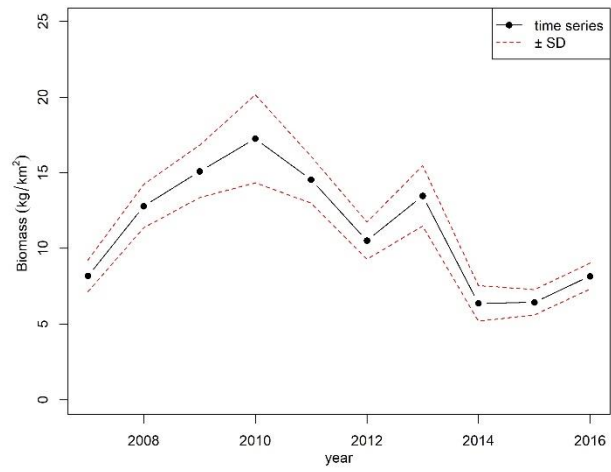
ARISFOL GSA10 (abundance)-RSS 200-800 m

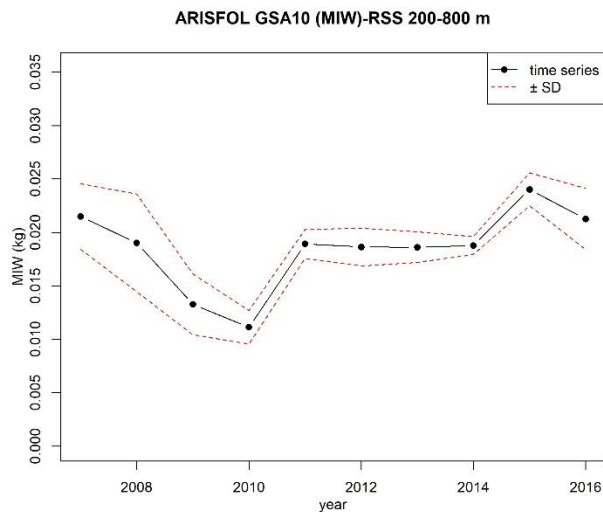


ARISFOL GSA10 (1/CV of mean abundance)-RSS 200-800 m



ARISFOL GSA10 (biomass)-RSS 200-800 m





```
## Estimation of abundance indices completed
##
## Estimation of biomass indices completed
##
## Estimation of MIW completed
```

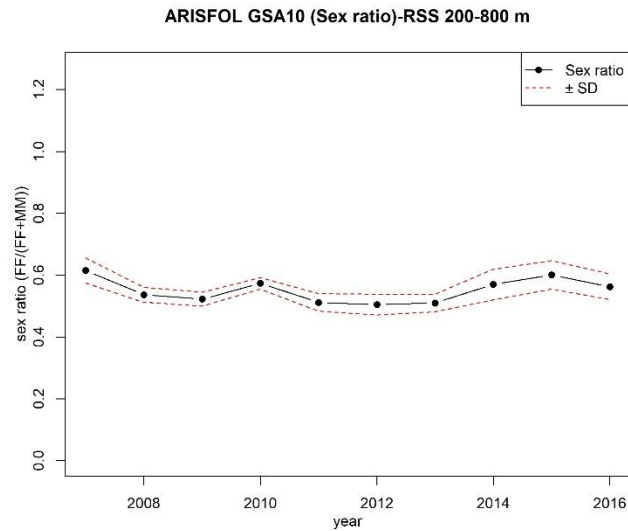
To estimate the time series of the species sex ratio in the selected GSA run the following code.

The routine uses the depth range previously selected for the abundance and density indices.

```
#-----
# Sex ratio per GSA in the timeseries
#-----
source(paste(wd, "/scripts/sex_ratio_timeseries_(Random_Stratified_Sampling).R ",
sep=""))
```

The output are the following .csv and .tiff files:

- *ARISFOL_GSA10_(Sex ratio)-Random_Stratified_Sampling_200-800 m_Timeseries.csv*
- *ARISFOL_GSA10_(Sex ratio)-Random_Stratified_Sampling_200-800 m_Timeseries.tiff*



The spawners are selected using the threshold value inserted in the *maturity_sizes.csv* file ("*~\scripts\utilities\maturity_sizes.csv*") during the initialization phase of the routine. Before running the code **check** you have updated the *maturity_sizes.csv* file.

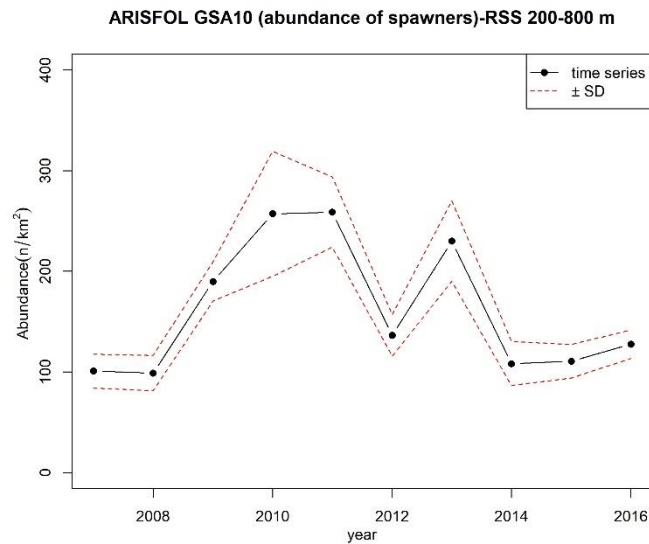
ATTENTION: Only female specimens are selected for the estimation of the spawners indices.

```
#-----
# Abundance indices of spawners in the timeseries
#-----

#-----> check the threshold in the file "~/input/maturity_sizes.csv"
source(paste(wd, "/scripts/Indices_spawners_timeseries_(Random_Stratified_Sampling)
).R", sep=""))
```

The output are the following .csv and .tiff files:

- *ARISFOL_GSA10_(abundance of spawners)-Random_Stratified_Sampling_200-800_m_Timeseries.csv*
- *ARISFOL_GSA10_(abundance of spawners)-Random_Stratified_Sampling_200-800_m_Timeseries.tiff*



The recruits are selected using the threshold value inserted in the *maturity_sizes.csv* file ("*~\scripts\utilities\maturity_sizes.csv*") during the initialization phase of the routine. Before running the code **check** you have updated the *maturity_sizes.csv* file.

```
#-----
# Abundance indices of recruits in the timeseries
#-----

#-----> check the threshold in the file "~/input/maturity_sizes.csv"
source(paste(wd, "/scripts/Indices_recruits_timeseries_(Random_Stratified_Sampling)
).R", sep=""))
```

The outputs are the following .csv and .tiff files:

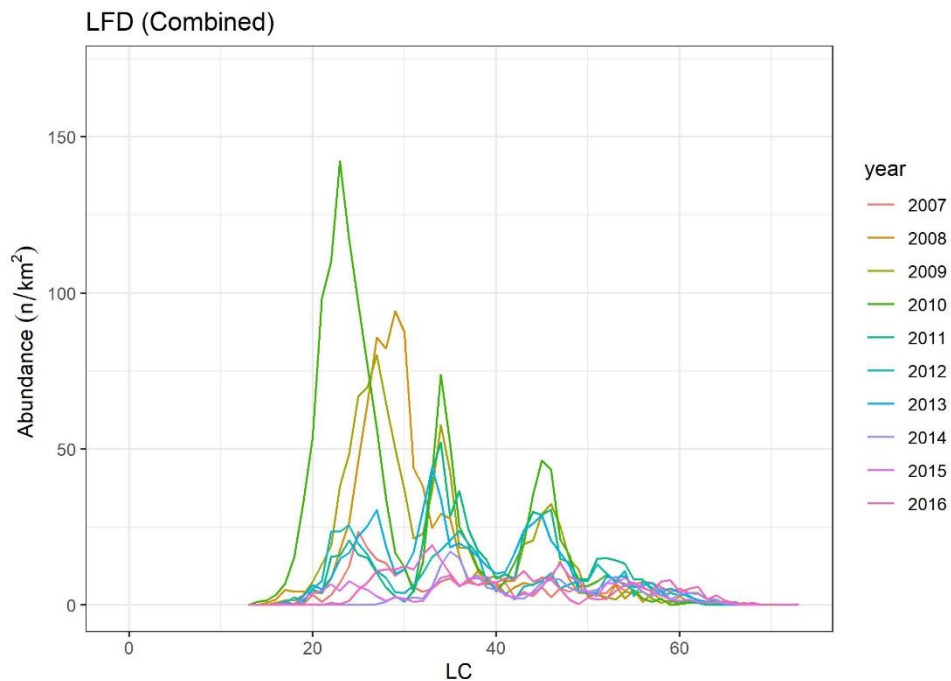
- *ARISFOL_GSA10_(abundance of recruits)-Random_Stratified_Sampling_200-800 m_Timeseries.csv*
- *ARISFOL_GSA10_(abundance of recruits)-Random_Stratified_Sampling_200-800 m_Timeseries.tiff*

To estimate the median (50th) and the length at 95th percentile is first needed to estimate the length class frequency distribution (LFD) running the following code:

```
#-----
# LFD - L0.50 - L0.95
#-----
source(paste(wd, "/scripts/LFD_by_year_(Random_Stratified_Sampling).R", sep=""))
source(paste(wd, "/scripts/L050_L095_RSS.r", sep=""))
```

The LFDs are estimated per sex and the following csv and jpg files are saved:

- *ARISFOL_GSA10_LFD_(Combined)_RSS.csv*
- *ARISFOL_GSA10_LFD_(Females)_RSS.csv*
- *ARISFOL_GSA10_LFD_(Males)_RSS.csv*
- *LFD_(Combined)_ARISFOL_GSA10_RSS.jpg*
- *LFD_(Females)_ARISFOL_GSA10_RSS.jpg*
- *LFD_(Males)_ARISFOL_GSA10_RSS.jpg*



When all the LFD are computed, the LFD of sexes combined is used to estimate the $L_{0.50}$ and $L_{0.95}$ running the following code:

```
source(paste(wd, "/scripts/L050_L095_RSS.r", sep=""))
```

The results are reported in a table saved in a csv file: *ARISFOL_GSA10_L50_L95_RSS.csv*.

The analysis of trend could be carried out with two different tests. The first is the Spearman's test. Run the following code:

```
#-----
# Spearman test of trends on timeseries
#-----
source(paste(wd, "/scripts/spearman_RSS.R", sep=""))
```

A pop-up box appears for choosing the extension of the time series to be used in the analysis of trends. Separate values with the comma. The default value is "2006,2012"

The results of the test are saved in a csv file (*ARISFOL - Spearman summary_RPS.csv*) and visualized in the console as follows:

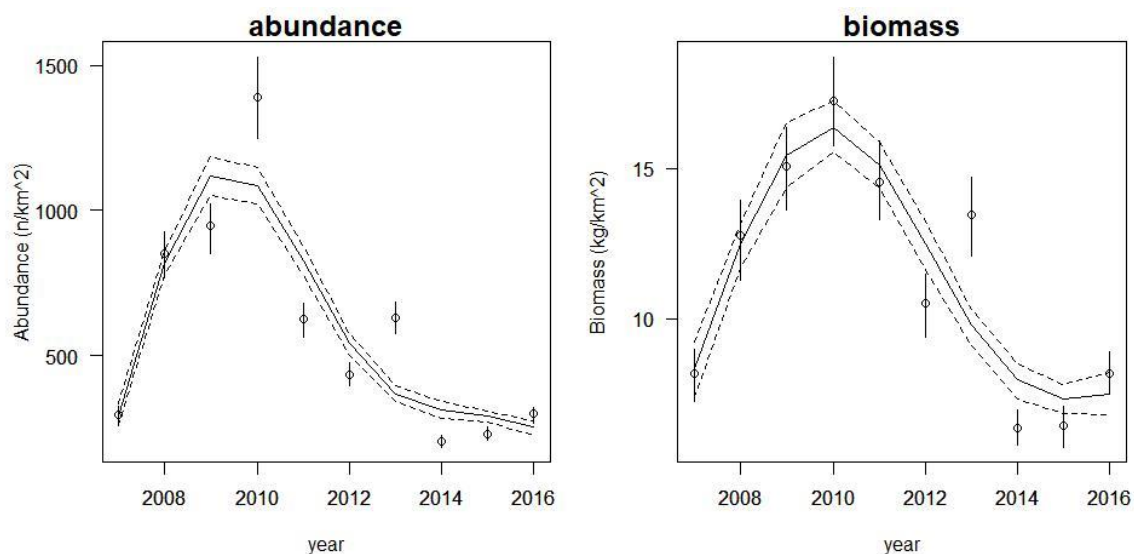
```
[1] "Select the year range for the analysis"
      index      r      t      p
1 abundance 0.08571429 0.1720618 0.8717434
2 biomass 0.25714286 0.5321812 0.6227872
```

A second test that could be applied to the time series is the Intersection Union Test. First the *lastn* parameter should be set (the default value is 5). This parameter is used by the routine to set the number of last years to be considered during the analysis. Run the following code:

```
#-----
# Intersection Union Test
#-----
lastn=5          #number of years to be considered in test
source(paste(wd, "/scripts/IUT_RSS.r", sep=""))
```

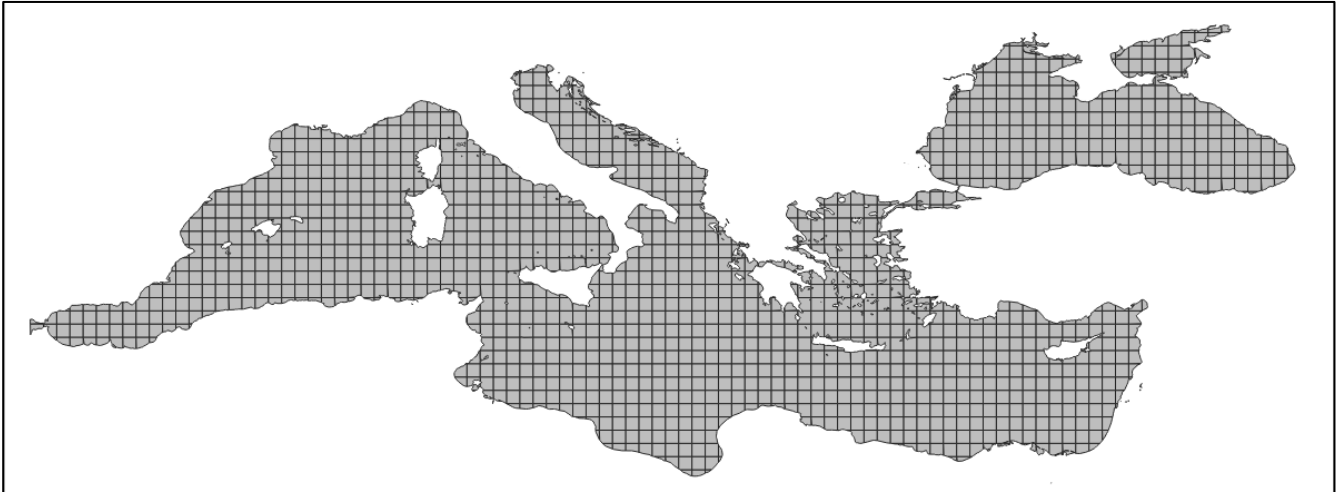
An intermediate file is generated by the routine (*GSA10IUTest5_RSS.csv*) while the results are stored in the following files:

- ARISFOL_GSA10_SmoothedIndicators_RSS.jpg
- ARISFOL_GSA10_IUT_results_5years_RSS.csv



The spatial indicators are estimated with the resolution of the 30" GFCM grid (approximately 30NM on the latitude) considering only the last 5 years (default value) of the time series. The user can modify the value of the variable "n_last_years" to select the number of the last years to be used in the analysis. In case there are less than "n_last_years" years of survey data a box will pop-up to insert the number of the last years to be considered in the analysis.

ATTENTION: check whether the coordinate range of the GSA is uploaded in the file
"*~/script/utilities/GSAs_coordinates.csv*"



Run the following code:

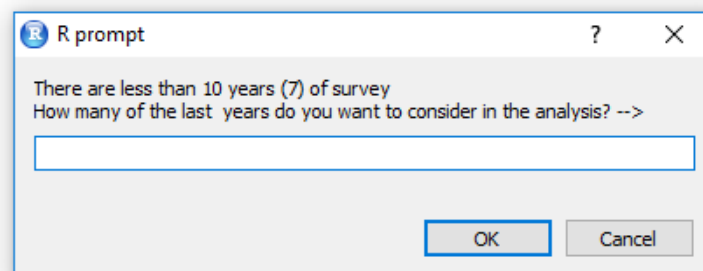
```
#-----
# Abundance indices for statistical squares
# inverse of CV of abundance indices for statistical squares
# Biomass indices for statistical squares
# Mean individual weight for statistical squares
#-----
n_last_years <- 5
source(paste(wd, "/scripts/index_by_grid.r", sep=""))
```

The user is further asked to select the depth strata to be used to cut the grid in the map. Three different depth strata are available: 10-200 m, 10-800 m and 200-800 m.

only the hauls between 200m and 800m are selected for the analysis.
choose the depth strata to be used for the map output ->

```
1: 10-200
2: 10-800
3: 200-800
```

selection: |



```
## Abundance indices for statistical squares correctly estimated
```

```
## inverse of CV of abundance indices for statistical squares correctly estimated
```

```
## file of abundance indices for statistical squares saved in the following folder
: '/output/ARISFOL - GFCM GRID ABUNDANCE.csv

## Biomass indices for statistical squares correctly estimated

## file of biomass indices for statistical squares saved in the following folder:
'/output/ARISFOL - GFCM GRID BIOMASS.csv

## MIW for statistical squares correctly estimated

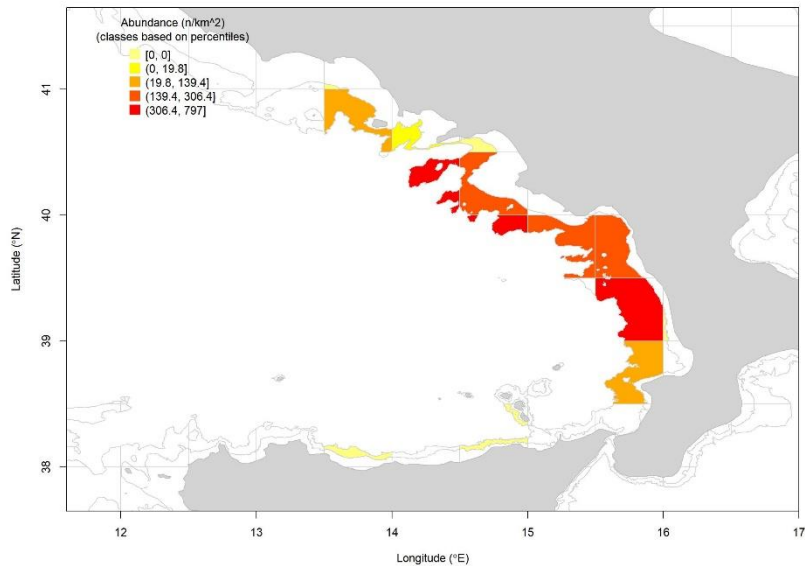
## inverse of CV of MIW for statistical squares correctly estimated

## file of MIW for statistical squares saved in the following folder: '/output/ARI
SFOL - GFCM GRID MIW.csv
```

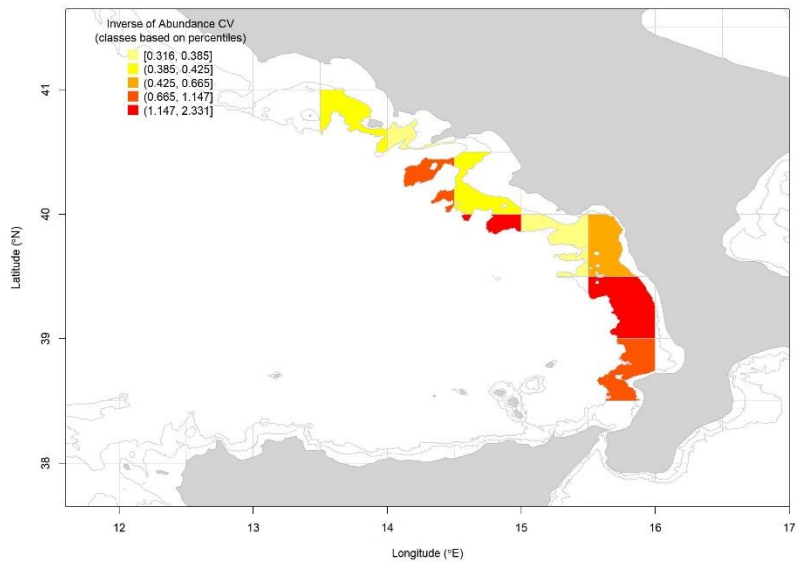
The results are reported in the following .csv and .jpg files:

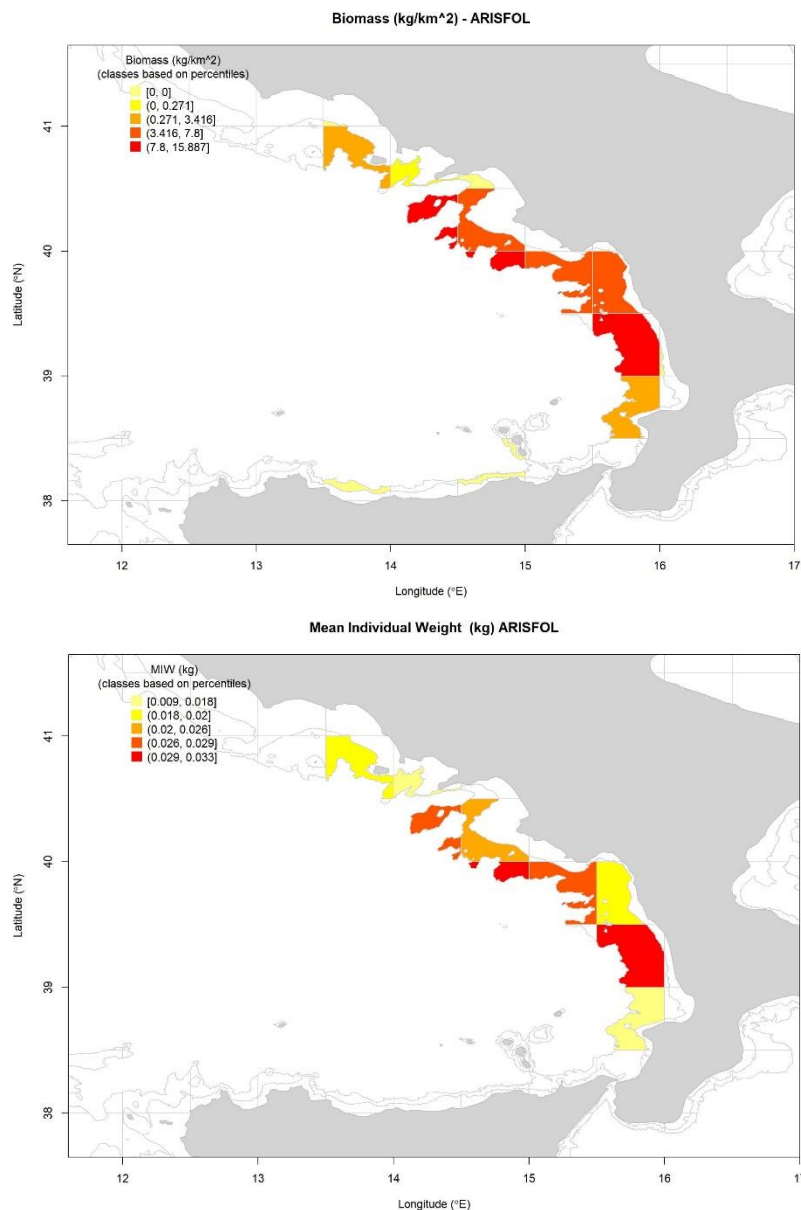
- ARISFOL - GFCM GRID ABUNDANCE.csv
- ARISFOL - GFCM GRID BIOMASS.csv
- ARISFOL - GFCM GRID MIW.csv
- ARISFOL - GFCM GRID ABUNDANCE.jpg
- ARISFOL - GFCM GRID ABUNDANCE Inverse CV.jpg
- ARISFOL - GFCM GRID BIOMASS.jpg
- ARISFOL - GFCM GRID MIW.jpg

Abundance (n/km²) ARISFOL



Inverse of Abundance CV - ARISFOL



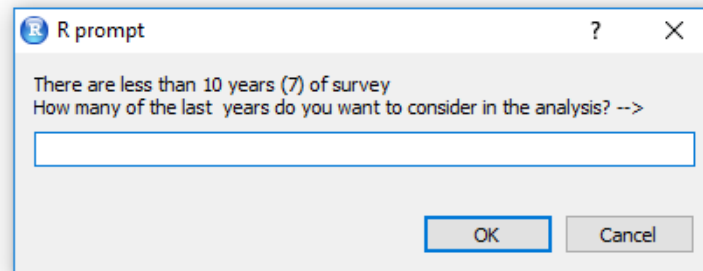


The sex ratio is also estimated over the GFCM grid running the following commands:

ATTENTION: set the threshold of the minimum number of individuals per haul to be considered in the analysis.

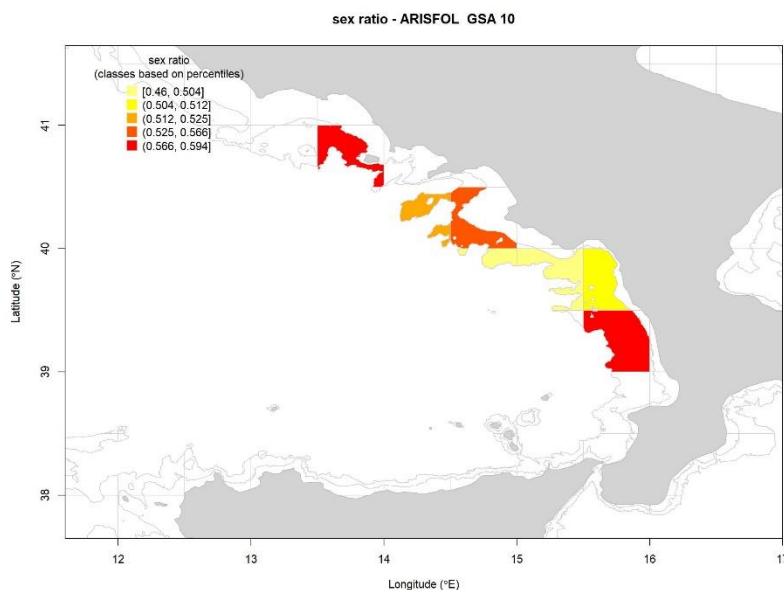
```
#-----
# Sex ratio for statistical squares
#-----
#-----> select the minimum number of individuals per haul to be considered in the
analysis
threshold <- 30
n_last_years <- 5
source(paste(wd, "/scripts/sex_ratio_by_grid.r", sep=""))
```

The user can modify the value of the variable “n_last_years” to select the number of the last years to be used in the analysis (default value is 5). In case there are less than “n_last_years” years of survey data a box will pop-up to insert the number of the last years to be considered in the analysis.



The outputs are stored in the following .csv and .jpg files:

- ARISFOL - GFCM SEX RATIO.csv
- ARISFOL - GFCM GRID Sex Ratio.jpg



Finally, the abundance indices by hauls of the whole time series can be plotted over a map for both recruits and spawners (only female specimens). The user can modify the value of the variable “n_last_years” to select the number of the last years to be used in the analysis (default value is 5).

The recruits and spawners are selected using the threshold values inserted in the *maturity_sizes.csv* file (input folder) during the initialization phase of the routine. Before running the code **check** that you have updated the *maturity_sizes.csv* file.

ATTENTION: check that the coordinate range of the GSA is uploaded in the file
 "~/script/utilities/GSAs_coordinates.csv"

The "GSAs_coordinates.csv" file contains both the coordinates range of the GSA and the preference about the bathymetrical lines to be plotted in the graphs as grey lines. Three reference bathymetrical lines could be selected respectively in the fields "depth1", "depth2" and "depth3" inserting selecting the values in the following list: 50, 100, 200, 500 and 800.

A	B	C	D	E	F	G	H	I	J
GSA	COUNTRY	xmin	xmax	ymin	ymax	depth1	depth2	depth3	notes
10	ITA	12.2	16.7	37.7	41.2	200	500	800	GSA10
18	ITA	15.5	20	39.8	42.5	200	500	800	GSA18
19	ITA	14.8	19.5	35.7	41	200	500	800	GSA19

```
#-----
# Bubble plots - indices of recruits (abundance)
# Bubble plots - indices of spawners (abundance)
#-----

#-----> check the threshold in the file "~/input/maturity_sizes.csv"
source(paste(wd, "/scripts/Bubble_plot_by_haul.r", sep=""))

## Regions defined for each Polygons
## Regions defined for each Polygons
## Bubble plot of recruits correctly saved
## Regions defined for each Polygons
## Regions defined for each Polygons
## Bubble plot of spawners correctly saved
```

The following jpg files are saved:

- ARISFOL_GSA10 -indices of RECRUITS.jpg
- ARISFOL_GSA10 -indices of SPAWNERS.jpg

Abundance of recruits (n/km²)

