

**ICTS SOCIB. R ECOLARV**

**Module 2: ‘Larval Index’**

**R Scripts to retrieve abundance and length frequency data from Access database and estimate larval abundance index considering environmental effects**

**User Manual 2023**

24th November 2023

**M. Pilar Tugores Ferrà**

**R ECOLARV**

**Module 2: Larval Index**

**R Scripts to retrieve abundance and length frequency data from ECOLARV database and estimate larval abundance index considering environmental effects**

**User Manual 2023**

|  |  |
| --- | --- |
| **Document type:** | User Manual |
| **Date:** | 24/11/2023 |

|  |  |
| --- | --- |
| **Description:** | Structure of interlaced R scripts to retrieve abundance and length frequency data from ECOLARV database for one or multiple species. Then estimate larval abundance index for only one species, taking into account environmental effects. |
| **Module:** | LarvInd |
| **Version:** | 1.0.7.4 |
| **Author:** | M. Pilar Tugores |
| **Beta testers:** | Asvin Pérez Torres, Diego Álvarez-Berastegui |
| **Project:** | PANDORA, BLUEFIN, BALEATÚN, Vicenç Mut |

**INDEX**

[1. GENERAL PURPOSE AND STRUCTURE 7](#_Toc96529512)

[2. RETRIVE ACCESS DATA: “00\_Get\_tabs\_fromAccess.r” 10](#_Toc96529513)

[2.1. Purpose, requirements, structure and usage 10](#_Toc96529514)

[2.2. Inputs section of the script 10](#_Toc96529515)

[*2.2.1.* *Setting databases and paths* 11](#_Toc96529516)

[*2.2.2.* *Process specifications* 11](#_Toc96529517)

[2.3. Processing section of the scripts and usage 12](#_Toc96529518)

[*2.3.1.* *Abundance tables* 13](#_Toc96529519)

[*2.3.2.* *Length-frequency data tables* 14](#_Toc96529520)

[2.4. Summary of the script outputs 16](#_Toc96529521)

[3. ABUNDANCE AND LENGTH DATA EXPLORATION 17](#_Toc96529522)

[3.1. Set input files (“01\_0\_set\_input\_files.r”) 17](#_Toc96529523)

[3.2. Abundance data exploration and mapping (“01\_1\_explore\_data\_&\_map\_samples.r”) 18](#_Toc96529524)

[3.3. Explore length distributions (“01\_2\_explore\_length\_distributions.r”) 20](#_Toc96529525)

[4. ABUNDANCE STANDARDISATION BY LENGTH AND GEAR 22](#_Toc96529526)

[4.1. Process exponential retrocalculation (“02\_1\_process\_exponential\_retrocalc.r”) 22](#_Toc96529527)

[4.2. Calculate larvae at 2 mm and perform gear standardisation between B60 and B90 (“02\_2\_abund\_standardisation\_2mm\_&\_gears.r”) 23](#_Toc96529528)

[4.3. Nominal Index plots and summary (“02\_3\_sampling&abund\_summary\_&\_nominal\_Index.r”) 24](#_Toc96529529)

[5. ENVIRONMENTAL DATA 25](#_Toc96529530)

[5.1. Read operational variables (“03\_0\_read\_operational\_vars\_todevelop.r”) 25](#_Toc96529531)

[5.2. Interpolate insitu & derived variables (“03\_1\_interpolate\_insitu\_&\_derived\_vars.r”) 26](#_Toc96529532)

[5.3. Make habitat quality variables (“03\_2\_make\_habitat\_quality\_vars\_20180917.r”) 27](#_Toc96529533)

[5.4. Explore environmental variables (“03\_3 explore\_environ\_vars.r”) 28](#_Toc96529534)

[6. MODELLING 29](#_Toc96529535)

[6.1. Bluefin tuna strict update (07\_v3\_BFT\_strict\_update\_VPA\_lineal\_lsmeans\_20211214.r) 29](#_Toc96529536)

[6.2. Bluefin tuna actualization (“10\_v1.1\_BFT\_strict\_update\_ MSE\_gam\_boostrap\_20211214.r”) 30](#_Toc96529537)

[6.3. Albacore strict update (“11\_v6.1.0\_ALB\_Rev\_2\_GAMd\_emmeans\_20211214.r”) ???? 31](#_Toc96529538)

[6.4. Albacore strict update (“11\_v6.1.1\_ALB\_Rev\_3\_GAMd\_emmeans\_20211214.r”) ???? 32](#_Toc96529539)

[6.5. Albacore actualization (“11\_v6.1\_ALB\_AC\_202005\_GAMd\_emmeans\_20211214.r”) 33](#_Toc96529540)

[ANNEX I. STRUCTURE OF THE GENERATED TABLES 34](#_Toc96529541)

# GENERAL PURPOSE AND STRUCTURE

**Purpose:**

The **Module 2 & 3:** ‘**Larval Index’** is intended to extract **abundance and length frequency** data from ECOLARV database **for one or several species at a time** and then perform larval index estimates for one of these species. It also retrieves CTD data from the CTD database.

**Structure:**

The main folder contains:

1. Access data bases (2),
2. Scripts in R language (16)
3. RData files (2) and
4. Folders (2)

The **Access databases**:

|  |  |
| --- | --- |
| File | Information |
| “ecolarv\_comunidades\_actual.mdb” | Abundance and length data for the different icthyoplantonic species in the Balearic Islands |
| “CTD\_para\_bluefin\_actual.mdb” | CTD data for BLUEFIN stations |

The **R scripts:**

They are **structured** in *five groups of scripts*, according to the main task they perform:

1. Retrieve information from the access data bases
2. Set the files to be used for processing and do some data exploration (for abundance and length frequency)
3. Standardise abundance estimates to larvae of 2 mm length and gears to obtain CPUA
4. Generate, retrieve and explore environmental data.
5. Model the abundance of larvae (previously standardised in step 3) in relation to environmental data and produce an estimate of the larval index.

|  |  |
| --- | --- |
| Script file | Purpose |
| 1. *Retrieving data* | |
| “00\_Get\_tabs\_fromAccess.r” | Retrieve data from Access database for single or multiple species (i.e. abundance, length and hydrographic data), applies some filters and saves outputs .csv and .xlsx files |
| 1. *Species to process and data exploration* | |
| “01\_0\_set\_input\_files.r” | Select the species for which the larval index will be estimated |
| “01\_1\_explore\_data\_&\_map\_samples.r” | Produces abundance plots, summary tables and maps of samples and presence-absence for the target species |
| “01\_2\_explore\_length\_distributions.r” | Produce plots and summary tables for the length-frequency data of the target species |
| 1. *Abundance standardisation by length and gear* | |
| “02\_1\_process\_exponential\_retrocalc.r” | Fit an exponential decay model to length frequency data to compensate abundance for mortality or loads a pre-existing model computed in 2018 (mainly, for strict updates) |
| “02\_2\_abund\_standardise\_2mm\_&\_gears.r” | Applies the fitted exponential decay model to standardise abundances of larvae at the different lengths observed in the samples to a reference length (currently set to 2 mm) |
| “02\_3\_sampling&abund\_summary\_&\_nominal\_Index.r” | Produces some summary tables and plots showing the amount of samples and nominal index, among some other information |
| 1. *Environmental variables processing and exploration* | |
| “03\_0\_read\_operational\_vars\_todevelop\_NEW.r” | Still not developed. Intended to incorporate operational variables and other external environmental data such as geostrophic velocities. |
| “03\_1\_interpolate\_insitu\_&\_derived\_vars.r” | Add interpolated data to CTD environmental data showing some NA using linear models or generalized linear models. It also produces derived environmental variables like annual anomalies of temperature or spatial variability of temperature (temperature after removing annual variability) |
| “03\_2\_make\_habitat\_quality\_vars\_20220221.r” | Produces an annual model of habitat suitability using Walter’s model but it is not finally incorporated to the process table |
| “03\_3\_explore\_environ\_vars.r” | Produces some plots to explore environmental variables |
| 1. *Modelling* |  |
| “04\_model\_ALB\_11\_v6.1.0\_Rev\_2\_GAMd\_emmeans\_20210407.r” | Model 1 for albacore |
| “04\_model\_ALB\_11\_v6.1.1\_Rev\_3\_GAMd\_emmeans\_20210408.r” | Model 2 for albacore |
| “04\_model\_BFT\_07\_v3\_strict\_update\_VPA\_lineal\_lsmeans.r | Model 1 for Bluefin tuna |
| “04\_model\_BFT\_10\_v1.1\_strict\_update\_MSE\_gam\_boostrap.r” | Model 2 for Bluefin tuna |
| “04\_model\_BFT\_11\_v6.1\_model\_BFT\_AC\_202005\_GAMd\_emmeans\_gapyears.r” | Model 3 for Bluefin tuna, accepted by ICCAT 2021 |

The **.RData files**:

|  |  |
| --- | --- |
| File | Information |
| “autosetting\_work\_dir.RData” | Shortcut to avoid having to set manually the working directory at the beginning of each R script.  Open the file and start running the scripts.  Warning: It can produce errors depending on the R version; not checked in RStudio. |
| “20180917\_cuadgrid\_balsea\_1DD.RData” | Contains ‘cuadgrid’, an object of class “SpatialPolygons” a grid of 10 x 6 polygons spanning around the Balearic Islands (longitude range: -1.95, 8; latitude range: 36, 42.1) with a grid size of 0.5 degrees |

The **Folders**:

|  |  |
| --- | --- |
| Folder name | Information |
| “/\_drivers\_odbc/”: | Contains a file (“Connect R to Access database via ODBC.docx”) with instructions to run the library “RODBC” for the first time on a computer. |
| “/Rsource/”: | Contains: i) 20 R scripts including functions and other source code to allow retrieving data from the Access databases and producing some data summaries and plots ii) 2 subfolders, one containing FAO codes for species and another with shapefiles of the coastline and NOAA bathymetry. |

The **content** of the **Rsource** folder:

|  |  |  |
| --- | --- | --- |
| Rsource  content | File/folder  name | Content |
| Scripts | “02\_1\_get\_Abundance\_table\_withCTD.r” | Code to retrieve abundance data from ECOLARV |
|  | “02\_1a\_check01\_duplicated\_collectors\_stations.r” | Code to check the existence of duplicated collectors and stations in the retrieved data |
|  | “02\_1b\_apply\_filters&create\_species\_vars.r” | Apply standard filters for larval index production and generates variables such as Julian date (jd), normalised hour (hournorm), etc. |
|  | “02\_2\_get\_LFD.r” | Code to retrieve length-frequency data from ECOLARV database |
|  | “02\_2a\_check02\_comprueba\_nlarvas\_y\_tllnumero.r” | TO BE COMPLETED |
|  | “02\_2b\_apply\_filters\_4\_t\_length.r” | TO BE COMPLETED |
|  | “02\_2b\_apply\_filters\_4\_t\_length\_1.r” | Equal than the former but a threshold length is set for BTF and ALB  The thresholds are:   * BFT: 10.92 mm (historical max length for the period 2001-2019, excluding 2018) * ALB: 8.5 mm (historical max length for the period 2001-2019, excluding 2018) |
|  | “02\_2b\_apply\_filters\_4\_t\_length\_2.r” | Equal than the former but a threshold length is set for BTF and ALB  The thresholds are:   * BFT: 8.5 mm (historical max length for the period 2001-2005 and B60 hauls) * ALB: 8.5 mm (historical max length for the period 2001-2019, excluding 2018) |
|  | “base.functions.r” | TO BE COMPLETED |
|  | “db.functions.r” | TO BE COMPLETED |
|  | “derived.envirvars.r” | TO BE COMPLETED |
|  | “gettabs\_v.r” | TO BE COMPLETED |
|  | “relative\_folderpaths.r” | TO BE COMPLETED |
|  | “relative\_folderpaths\_m3.r” | TO BE COMPLETED |
|  | “saveModel\_outputs.r” | TO BE COMPLETED |
|  | “settings\_m2.r” | TO BE COMPLETED |
|  | “settings\_m3.r” | TO BE COMPLETED |
|  | “summary.functions.r” | TO BE COMPLETED |
| Subfolders | **/fao\_codes/** | “ASFIS\_sp\_2019.txt”: table with FAO codes for marine species |
|  | **/sig/** | “country.shp”: polygon shapefile with coastline |
|  |  | “mallorca\_cabrera.shp”: polygon shapefile with Balearic Islands coastline |
|  |  | “ETOPO1\_BalearicSea.tiff”: geotiff file with NOAA bathymetric data |

# RETRIVE ACCESS DATA: “00\_Get\_tabs\_fromAccess.r”

# Purpose, requirements, structure and usage

**Functionality:**

The R script **“00\_main\_Get\_tabs\_fromAccess.r”** retrieves abundance with CTD data and length-frequency data from the ECOLARV and CTD databases.

**Requirements:**

* **Access databases:**
  + Copy paste in the main directory of the LarvalIndex scripts:
    - ecolarv database (“ecolarv\_comunidades\_actual.mdb”)
    - CTD database (“CTD\_para\_bluefin\_actual.mdb”).
* **Other requirements:** libraries and functions are automatically loaded within the script by calling the source code “*settings\_m2.r*”

**Structure:**

The script is structured in **two parts**, **one** devoted to incorporate **input information** (section 2.1, divided in two subsections “2.1.1 Setting databases and paths” and “2.1.2. Process specifications”, respectively) and a **second part** devoted to **processing & checking** (section 2.2) where all the processing is performed through calling several settings, scripts and functions to generate the output files and some checks on them.

**Usage:**

1. Copy a version of the database in the main folder ‘Gettabs\_LarvInd’
2. Double click on ‘autosetting\_work\_dir.RData” to set the R working directory to the main folder
3. Run the lines of the main script containing the parameters that want to be used for processing
4. Run the lines devoted to processing, one by one and being sure every line has run properly before running the following

# Inputs section of the script

The inputs section is divided in two subsections, one where the user sets the names and paths of the databases and a second subsection where the user selected some parameters to retrieve data from the databases like the species to be processed or the type of hauling structure.

### *Setting databases and paths*

In this first part, the user sets the names and paths of the databases and an additional parameter regarding CTD data. The data bases should be copy-pasted in the main directory.

* ***dbpath:*** ECOLARV database path (typically, the directory of the main scripts)
* ***databaseversion:*** “ecolarv\_comunidades\_actual.mdb” ECOLARV database name
* ***dbpathCTD:*** CTD database path (typically, the directory of the main scripts)
* ***databaseCTD:*** “CTD\_para\_bluefin\_actual.mdb” CTD database name
* ***independentCTD:***  by default it is set to FALSE; if set to TRUE, an independent file containing CTD data is generated # Temporary parameter intended for future developments of the scripts. Do not change unless necessary.

### *Process specifications*

In this section the user selects the parameters required to retrieve information from the ECOLARV database.

In the script, the user manually creates an object (**input\_process\_specifs**) of class **list** containing **five elements** which can consist of one value or a vector of values. The parameters required are:

* ***grupo\_para\_procesar:*** some of the valid names included in the ECOLARV database in the variable “ttn\_id\_target\_num”, which are:
  + Tunidos
  + Ictio sin tunidos
  + Paralarvas
  + Decapodos
  + Phyllosomas
  + Ephyras
* ***tes\_nombre\_cientifico:***some of the valid names included in the ECOLARV database, in the variable “tes\_nombre\_cientifico”, i.e. “*Thunnus thynnus*”, “*Thunnus alalunga*”, “*Katsuwonus pelamis*”, etc.
* ***tc\_estructura:*** type of gear that is going to be included. Typically set to “B60” and “B90” for larval index.
* ***tc\_cont\_estructura:*** typically set to 1 for larval index # No sé bien qué es
* ***tpt\_larval\_index:*** for standard species like *Thunnus thynnus* and *Thunnus alalunga* it is typically set to 1. For non-standard species like *Katsuwonus pelamis*, it is better that it is set to c (1, 2, 3, 4, NA).

# Processing section of the scripts and usage

**Observations:**

The outputs have the following colour code:

- orange: objects within the R Console

- lila: files saved in a specific folder

- **bold**: final output files

- non-bold: intermediate output files, generally processing files

This section is where all the processing is performed.

Usage: **Warning!!** The lines should be run one by one and in the order included in the script. If one does not properly work it is probable that the following step is not going to work properly.

The script calls the following settings, functions and scripts in the order specified below:

* ***"settings\_m2.R":*** it mainly performs the following actions:
  + i) loads required libraries (if they are not installed, the script previously install them)
  + ii) calls other scripts that contain required functions (e.g. “base.functions.R”)
  + iii) sets relative paths and generate folders that will be latter used (“relative\_folderpaths.R”). Created folders:
    - **/data/:** where all the data retrieved from the databases are saved
    - **/data/processfiles/:** where all the intermediate processing files are saved
    - **/outtables/:** where tables with some additional processing are saved
* ***formatPS (input\_process\_specifs, databasefile)***:

Output: function that, taking the object ‘input\_process\_specifs’ created by the user with the processing specifications, generates another object of class list in the working space (*'process\_specifs'*) in which some extra required information is added and it is also saved as an RData file (‘*00\_0\_process\_specifs\_BFT\_ALB\_SKJ\_2021-10-19.RData’*) to have a recording of the processing applied and in case it is required in posterior steps of the processing or in posterior processing attempts. The name of the file changes according to the System.Date and according to the species processed. The code for the species is that used within FAO (ex. SKJ: Katsuwonus pelamis; BFT: Thunnus thynnus; ALB: Thunnus alalunga).

Output file folder: **/data/processfiles/**, where all the checks on retrieved data from ECOLARV database will be saved.

### *Abundance tables*

* ***“02\_1\_get\_Abundance\_table\_withCTD.r"***: the main objective of the script is to generate a data.frame containing the abundance information for the larvae for one or more species, together with the CTD information for the same stations.

It also generates an intermediate output containing the tables from the ECOLARV database that are required for further processing and filtered according to ‘process\_specifs’

1. Intermediate/processing output

Output: this scripts calls a function in the script ‘db.functions.r’ to connect to the ECOLARV database through ODBC and creates an object of class list in the working space (*‘ecolarv\_filt’*) which is also saved as a .RData (*‘00\_1\_ecolarv\_DB\_filt\_BFT\_ALB\_SKJ\_2021-10-19.RData’*). This object contains the required tables of the ECOLARV database with the filters imposed by the ‘process\_specifs’ object.

Output folder: **/data/processfiles/**, where all the checks on retrieved data from ECOLARV database will be saved.

1. Script output: a data.frame containing the abundance information for the larvae of one or more species, together with the CTD information for the same stations.

Output: an R object (*‘****Abundance\_df****’*) that is also saved as .csv files, containing raw larval abundance data for one species (**‘*abundance\_larv.index\_BFT \_2021-10-19.csv***’) or for multiple species (‘***abundance\_larv.index\_multispecies\_BFT\_ALB\_SKJ\_2021-10-19.csv****’*).

The object and the .csv file contain one column or several columns with the information regarding the species amount of collected larvae. The column/s have the name of the species scientific name, with a lower dash instead of a blank space (ex. “Thunnus\_thynnus”). See Table 1 in **Annex I. STRUCTURE OF THE GENERATED TABLES** to have an overview of the structure of the data.frame.

If *‘independentCTD’*is set to TRUE it also produces an independent.csv file containing only the CTD data (‘***CTD\_larvind\_BFT\_2021-10-19.csv****’* or ‘***CTD\_larvind\_multispecies\_BFT\_ALB\_SKJ\_2021-10-19.csv****’*).

Output file folder: **/data/,** where all the data extracted from ECOLARV database or CTD database will be located.

* ***“02\_1a\_check01\_duplicated\_collectors\_stations.r":*** check no duplicated collectors and stations for a given ‘grupo\_para\_procesar’.

Output: two .txt files, one for the duplicated collectors *(‘check1\_Duplicated\_collectors\_Tunidos\_2021-10-19.txt’*) and the second with the duplicated stations (*‘check1\_Duplicated\_stations\_Tunidos\_2021-10-19.txt’*).

Output file folder: **/data/processfiles/**, where all the checks on retrieved data from ECOLARV database will be saved.

Usage: open the .txt files and check there are not duplicated collectors and stations.

* ***"02\_1b\_apply\_filters&create\_species\_vars.r":*** this scripts applies further filters, creates some variables and save larval index file.

It removes subsurface hauls, MEDIAS0710 survey data, duplicated stations for gap years and removes years 2002 and 2003 for albacore.

Output: object of class ‘list’ (‘*abund\_list’*), each element being a data.frame for each of the analysed species. The data.frame consists of the abundance file retrieved from the Access database in which some additional filters have been applied and some additional variables have been created saved. It is also as a .csv files, one for each of the elements of the list and, hence, one for each of the processed species (**‘*20211026\_t\_analisis\_larvalindex\_SWO\_abs.csv’***). See Table 2 in **Annex I. STRUCTURE OF THE GENERATED TABLES** to have an overview of the structure of the data.frame.

Output file folder: **/outtables/**, where processed tables are saved.

### *Length-frequency data tables*

* **“*02\_2\_get\_LFD.r*":**

Output: according to the specifications detailed in the object ‘process\_specifs’ the script generates an R object of class list (***‘outQ23’***) containing a data.frame that incorporates length frequency data for each species (each species is an element of the list). Each element of the list is saved as a .csv file, hence we will have one file for each species being processed (**‘*length.frec\_abundance\_340\_BFT\_201-10-19.csv*’, ‘*length.frec\_abundance\_342\_ALB\_201-10-19.csv*’**, etc.).

If the species for which the processing is performed do not have length data incorporated within the ECOLARV database an error warning is produced within the R console, as follows: “Error in eval(ei, envir) : No length data in 'ecolarv' DB for the species processed!!”

Output file folder: **/data/,** where all the data extracted from ECOLARV database or CTD database will be located. See Table 3 in **Annex I. STRUCTURE OF THE GENERATED TABLES** to have an overview of the structure of the data.frame.

* **"02\_2a\_*check02\_comprueba\_nlarvas\_y\_tllnumero.r*":** It compares the differences between the variable ‘tg\_nlarvas’ and ‘tll\_numero’ for each of the analysed species. It prints, in the R Console, information regarding the range of values of the differences between the two variables and if there is some haul in which the differences are greater than 2. It generates a .csv file for each species being analysed (‘*check2\_on\_length.frec\_339\_SKJ\_2021-10-19.csv*’), saved in the processfiles folder (“/data/processfiles/”), in which the two variables to be compared are included, together with their difference and some other information to be able to identify the station.

Output file folder: **/data/processfiles/**, where all the checks on retrieved data from ECOLARV database will be saved.

Usage: open the .txt file and check if the difference is greater than 2, in case the RConsole has adviced you to do so.

* **"02\_2b\_*apply\_filters\_4\_t\_length.r*":** this scripts applies further filters and save length-frequency table file for larval index analysis.

It removes: 1) subsurface hauls, 2) MEDIAS0710 survey data, 3) duplicated stations for gap years and 4) removes years 2002 and 2003 for albacore.

It also substitutes lengths of larvae bigger than the historical values by these historical maxima to avoid huge impacts on the index. The **maximum historical lengths** are 10.92 mm and 8.50 mm, for Bluefin tuna and Albacore, respectively.

Output: object of class ‘list’ (‘*length\_list’*), each element being a data.frame for each of the analysed species. The data.frame consists of the length-frequency data retrieved from the Access database with the additional filters. It is also saved as a .csv and .xlsx files, one for each of the elements of the list and, hence, one for each of the processed species (‘***20211026\_t\_length\_larvlind\_BFT.csv’, ‘20211026\_t\_length\_larvlind\_ALB.csv’, 20211026\_t\_length\_larvlind\_SKJ.csv’,*** etc.).

Output file folder: **/outtables/**, where processed tables are saved.

# Summary of the script outputs

**Outputs:**

The amount of outputs varies according to the amount of species being processed and if there are available length information (tuna species) or not (e.g. other accompanying species).

* + - **/data/:** where all the data retrieved from the databases are saved
* *‘Abundance\_df’*: data.frame containing abundance data for single or multiple species. Also saved as one .csv file, with the name slightly different if data is for single or for multiple species:
  + single species ‘*abundance\_larv.index\_BFT \_2021-10-19.csv*’
  + multiple species ‘*abundance\_larv.index\_multispecies\_BFT\_ALB\_SKJ\_2021-10-19.csv’*
* If *‘independentCTD’*=TRUE: only CTD data is saved as .csv file: e.g.
  + ‘*CTD\_larvind\_BFT\_2021-10-19.csv’*
  + *‘CTD\_larvind\_multispecies\_BFT\_ALB\_SKJ\_2021-10-19.csv’).*
* *‘outQ23’*: list of data.frames each element containing length information for one of the species for which data is retrieved. Also saved as separate .csv files, one for each species:
  + *‘length.frec\_abundance\_340\_BFT\_201-10-19.csv’*
  + *‘length.frec\_abundance\_342\_ALB\_201-10-19.csv’*
    - **/data/processfiles/:** where all the intermediate processing files are saved
* *'process\_specifs'* ‘*00\_0\_process\_specifs\_BFT\_ALB\_SKJ\_2021-10-19.RData’*
* *‘ecolarv\_filt’**‘00\_1\_ecolarv\_DB\_filt\_BFT\_ALB\_SKJ\_2021-10-19.RData’*
* *‘check1\_Duplicated\_collectors\_Tunidos\_2021-10-19.txt’*
* *‘check1\_Duplicated\_stations\_Tunidos\_2021-10-19.txt’*
* ‘*check2\_on\_length.frec\_339\_SKJ\_2021-10-19.csv*’
  + - **/outtables/:** where tables with some additional processing are saved
* *‘abund\_list’ ‘20211026\_t\_abundance\_larvalindex\_BFT\_abs.csv’, ‘20211026\_t\_abundance\_larvalindex\_ALB\_abs.csv’*
* ‘*length\_list’*, *‘20211026\_t\_length\_larvlind\_BFT.csv’* and *‘20211026\_t\_length\_larvlind\_ALB.csv’,*

# ABUNDANCE AND LENGTH DATA EXPLORATION

**Observations:**

All scripts have a final part where information regarding the script processing are saved to a .txt log file, saved in the main directory, with the following name:

“LarvInd\_settings\_log\_”, **faocode**, “\_”, **processingdate**, “\_”, **currentdate**, “.txt”

* faocode: BFT, ALB, SKJ, SWO
* processingdate: date when abundance and length data were processed
* currentdate: date as set in the sytem Sys.date()

“LarvInd\_settings\_log\_BFT\_20211221\_2021-12-21.txt”

## Set input files (“01\_0\_set\_input\_files.r”)

The script **“01\_0\_set\_input\_files.r”** contains the names of the species to be processed and the names of the files where the required information is located.

The user MUST modify this first script according to the species and files one wants to use for processing. Within the script, the user will set the following parameters:

**species** <- "Thunnus thynnus"

**flen** <- "20211115\_t\_length\_larvind\_BFT.csv"

**fabund** <- "20211115\_t\_abundance\_larvind\_BFT\_abs.csv"

**fprocess** <- "00\_0\_process\_specifs\_BFT\_ALB\_SKJ\_2021-11-15.RData"

Details

* **species:** species scientific name
* **flen:** name of the file containing length data
* **fabund:** name of the file containing abundance data
* **fprocess:** name of the file containing the process specifications used for extracting the abundance and length data from the ecolarv database

These objects will be used in posterior scripts to recall files generated in the previous script “00\_Get\_tabs\_Access.r” that NEED to be correctly produced and cannot be moved to another folder, as specified in the **Requirements** section bellow.

**Requirements:**

Files generated in the previous script:

* **/outtables/:** 
  + “20211115\_t\_abundance\_larvind\_BFT\_abs.csv”: abundance file
  + “20211115\_t\_length\_larvind\_BFT.csv”: length file
* **/data/processfiles/:** 
  + “00\_0\_process\_specifs\_BFT\_ALB\_SKJ\_2021-11-15.RData” containing the object ‘process\_specifs’ that has been used to create the abundance and length files.

## Abundance data exploration and mapping (“01\_1\_explore\_data\_&\_map\_samples.r”)

The script **“01\_1\_ explore\_data\_&\_map\_samples.r”** performs several data exploration on the abundance file data.

**Inputs:**

Scripts:

* "/Rsource/settings\_m3.r" and all dependencies (i.e. scripts in Rsouce folder): script installing and loading all required libraries and dependencies
* "01\_0\_set\_input\_files.R”: script containing the input parameters as specified at **section 3.1**

Files produced in “00\_Get\_tabs\_Access.r”:

* "outtables/20211115\_t\_abundance\_larvind\_BFT\_abs.csv" (fabund): abundance data frame
* "data/process\_files/00\_0\_process\_specifications\_340\_342\_2021-04-19.RData" (fprocess)

**Analysis**

The script performs the two sets of data exploration and three sets of sample mapping, as follows:

1. **Summary tables of samplings:** the script shows within the R Console the following summarie of the amount of samples and towing depths:
   1. The amount of samples by year (Nsamp\_y)
   2. The amount of samples by year and gear (Nsamp\_yg)
   3. A data.frame combining the two previous (Nsamp)
   4. The amount of samples by fishing type (Nsamp\_ft)
   5. A summary of the towing depths of the hauls by fishing type, gear and year (TowDep\_summary)
2. **Boxplots of volume filtered by fishing gear:** the script shows in the R Console and saved as a .tiff file boxplot of the filtered volumes for B60 and B90
   1. “outfigures/boxplot\_volumes\_B60&B90\_BFT.tiff

As long as we have performed some species specific filters in the Module 2, the output of this plots CAN be species dependent.

1. **Yearly maps of samples:**

In this section, the script calls the function ‘*mapSamples*’ contained within the script (**“/Rsource/mapping.functions.r”**), specially created for this application, maps for data regarding larval index 1, 2 and 1&2 are generated. They are plot in the **R Console** and as long as the parameter save2tiff is set to TRUE (**save2tiff=TRUE**) the function also export the maps as **.tiff files**.

* 1. “Maps\_Samples\_in\_larvind\_1\_2001.tiff”
  2. “Maps\_Samples\_in\_larvind\_2\_2001.tiff”
  3. “Maps\_Samples\_in\_larvind\_1&2\_2001.tiff”

1. **Yearly maps of samples with CTD:**

Calling the function ‘*mapSamples\_withCTD*’ which is contained within the script (**“/Rsource/mapping.functions.r”**) and was specially created for this application, the script maps for data regarding samples that have information regarding CTD data. They are plot in the **R Console** and as long as the parameter save2tiff is set to TRUE (**save2tiff=TRUE**) the function also export the maps as **.tiff files**.

* 1. “Maps\_Samples\_withCTD\_2001.tiff”

1. **Yearly maps of presence-absence data for the selected species:**

Calling the function ‘*mapSpecies\_pa*’ which is contained within the script (**“/Rsource/mapping.functions.r”**) and was specially created for this application, the script maps for presence-absence maps for the selected species. They are plot in the **R Console** and as long as the parameter save2tiff is set to TRUE (**save2tiff=TRUE**) the function also export the maps as **.tiff files**.

* 1. “Maps\_presence-absnece\_BFT\_2001.tiff”

## Explore length distributions (“01\_2\_explore\_length\_distributions.r”)

The script **“01\_2\_explore\_length\_distritubions.r”** is aimed at check the length-data from a selected species.

**Inputs**

Scripts

* "/Rsource/settings\_m3.r" and all dependencies (i.e. scripts in Rsouce folder): script installing and loading all required libraries and dependencies
* "01\_0\_set\_input\_files.R”: script containing the input parameters as specified at **section 3.1**

From script “00\_Get\_tabs\_Access.r”

* "outtables/20211115\_t\_length\_larvind\_BFT.csv" (flen)
* "data/process\_files/00\_0\_process\_specifications\_340\_342\_2021-04-19.RData" (fprocess)

**Analysis**

The script performs the following analysis on the length data for the selected species:

1. **Summary tables of length data:** the script shows within the R Console and saves as a **.csv** and **.xlsx file** a summary of the larval lengths (min, max and mean) by year.
   1. "20211115\_summary\_lengths\_BFT.csv"
   2. "20211115\_summary\_lengths\_BFT.xlsx"
2. **Analysis of the rounding errors:** the script shows in the R Console a series of checks to visualise how the rounding of values affects the final result.
3. **Yearly boxplots of length data:**

The script produces a boxplot of length data by year, saved as a **.tiff file**:

* 1. “boxplot\_lengths\_BFT\_allgears.tiff”

1. **Length-frequency histograms:**

The script produces a length-frequency histogram for all the data included in the length data file, saved as a **.tiff file**:

* 1. “histt\_lengths\_BFT\_allgears inc1.tiff”

1. **Length-frequency histograms by fishing gear (B60-B90):**

The script produes a length-frequency histogram for each fishing gear, considering all the data included in the length data file, saved as **.tiff files**:

* 1. “histt\_lengths\_BFT\_B60 inc1.tiff”
  2. “histt\_lengths\_BFT\_B90 inc1.tiff”
  3. “histt\_lengths\_BFT\_B60-B90.tiff”

1. **KS-test to check statistical differences between length-frequencies obtained by each fishing gear (B60 and B90):**

the script shows in the R Console the result of a KS-test to compare if the two length-frequency distributions (B60-B90) are driven from the same distribution or if they differ significantly. If the shown p-value is < 2.2e-16 the two length distributions are significantly different.

1. **Density plots of length data**

the script shows in the R Console density plots of length data.

1. **Annual histograms of length-frequency data**

The script produces anual histograms of length-frequency data, which is saved as a **.tiff file**.

* 1. "20211115\_histt\_annual\_lengths\_BFT.tiff"

# ABUNDANCE STANDARDISATION BY LENGTH AND GEAR

## Process exponential retrocalculation (“02\_1\_process\_exponential\_retrocalc.r”)

The script **“02\_1\_process\_exponential\_retrocal.r”** is aimed at fitting an exponential decay model to larval length data to estimate the amount of larvae of 2 mm that would have been collected in the sampling site according to larval mortality.

**Inputs**

Scripts

* "/Rsource/settings\_m3.r" and all dependencies (i.e. scripts in Rsouce folder): script installing and loading all required libraries and dependencies
* "01\_0\_set\_input\_files.R”: script containing the input parameters as specified at **section 3.1**

Exponential retrocalculation model

* “20180917\_exponential\_retrocalc\_model\_allgears\_BFT.RData”

Files produced in the script “00\_Get\_tabs\_Access.r”

* "outtables/20211115\_t\_length\_larvind\_BFT.csv" (flen)
* "data/process\_files/00\_0\_process\_specifications\_340\_342\_2021-04-19.RData" (fprocess)

**Analysis**

The script performs the following analysis:

1. Check the amount of larvae < 1.5 mm and set to 1.5
2. Plot an histogram ¿??
3. Check if all length data are used ¿???
4. Save an .xlsx file with aggregated abundance by length-frequency
5. Load an old retrocalculated model or run it with all the new length data
6. Plot (I don’t know what)
7. Compute mean length per year

## Calculate larvae at 2 mm and perform gear standardisation between B60 and B90 (“02\_2\_abund\_standardisation\_2mm\_&\_gears.r”)

The script **“02\_2\_abund\_standardisation\_2mm\_&\_gears.r”** is aimed at estimating the abundance of larvae of 2 mm that have been collected as well as to standardized the catches between B60 and B90 gears.

**Inputs**

Scripts

* "/Rsource/settings\_m3.r" and all dependencies (i.e. scripts in Rsouce folder): script installing and loading all required libraries and dependencies
* "01\_0\_set\_input\_files.R”: script containing the input parameters as specified at **section 3.1**

Files in the script “00\_Get\_tabs\_Access.r”

* "outtables/20211115\_t\_length\_larvind\_BFT.csv" (flen)
* "data/processfiles/00\_0\_process\_specifs\_BFT\_ALB\_SKJ\_2021-11-15.RData" (fprocess)

Other

* "20180917\_exponential\_retrocalc\_model\_allgears\_BFT.RData" (fretro)
* "20211115\_tmeanlengthsyear\_BFT.RData" (fmeanlength)

**Analysis**

The script performs the following analysis:

1. open length distributin tabla from data species specific data base
2. filter the species of interés
3. remove non-vector variables
4. open exponential model for backcalcuation
5. compute the abundance at reference length specified at sample (only one combined decay curve)
6. open table t anlaisis larval index all vars
7. paste tanalisis and tabundances at 2mm
8. complete n2mm in samples with larvae but no lengths measured
9. compute variables density and abundance
10. compute varaible abundance (gear standardized) with the model published in iccat
11. export table (20211115\_t\_abs\_2mm\_m2\_all\_vars\_mixedgears.csv)

## Nominal Index plots and summary (“02\_3\_sampling&abund\_summary\_&\_nominal\_Index.r”)

The script **“02\_3\_sampling&abund\_summary\_&\_nominal\_Index.r”** generates a table containting sampling summary, plots of the nominal index and plots of the positive abundance.

**Inputs**

Scripts

* "/Rsource/settings\_m3.r" and all dependencies (i.e. scripts in Rsouce folder): script installing and loading all required libraries and dependencies
* "01\_0\_set\_input\_files.R”: script containing the input parameters as specified at **section 3.1**

Files in the script “00\_Get\_tabs\_Access.r”

* "data/processfiles/00\_0\_process\_specifs\_BFT\_ALB\_SKJ\_2021-11-15.RData" (fprocess)

Other

* “outtables/BFT/20211115\_t\_abu\_larvind\_BFT\_inter\_insitu.csv” (fanalisis)
* "20180917\_exponential\_retrocalc\_model\_allgears\_BFT.RData" (fretro)
* "20211115\_tmeanlengthsyear\_BFT.RData" (fmeanlength)

**Analysis**

The script performs the following analysis:

1. Sample summary
2. Plot larval indicators
3. Compute & save nominal CPUA
4. Compute & Plot CPUA with SE
5. Boxplot Mean positive abundance

**Outputs**

FALTA POSAR-LOS, NO???

# ENVIRONMENTAL DATA

## Read operational variables (“03\_0\_read\_operational\_vars\_todevelop.r”)

The script **“03\_0\_read\_operational\_vars\_todevelop.r”** is aimed at incorporating operational variables to the main abundance data.frame **STILL PENDING DEVELOPMENT**

**Inputs**

Scripts

* "/Rsource/settings\_m3.r" and all dependencies (i.e. scripts in Rsouce folder): script installing and loading all required libraries and dependencies
* "01\_0\_set\_input\_files.R”: script containing the input parameters as specified at **section 3.1**

Files from module 2

* "outtables/20211115\_t\_length\_larvind\_BFT.csv" (flen)
* "data/processfiles/00\_0\_process\_specifs\_BFT\_ALB\_SKJ\_2021-11-15.RData" (fprocess)

Other

* "20180917\_exponential\_retrocalc\_model\_allgears\_BFT.RData" (fretro)
* "20211115\_tmeanlengthsyear\_BFT.RData" (fmeanlength)

**Analysis**

The script performs the following analysis:

1. Sample summary
2. Plot larval indicators
3. Compute & save nominal CPUA
4. Compute & Plot CPUA with SE
5. Boxplot Mean positive abundance

## Interpolate insitu & derived variables (“03\_1\_interpolate\_insitu\_&\_derived\_vars.r”)

The script **“03\_1\_interpolate\_insitu\_&\_derived\_vars.r”** is aimed at estimating environmental data at stations were they were not measured by means of some modelling imputation approach. It also estimate variables derived from the in situ variables after calculating anual anomalies and/or removing temporal trends (i.e. SALanom, restemp).

**Inputs**

Scripts

* "/Rsource/settings\_m3.r" and all dependencies (i.e. scripts in Rsouce folder): script installing and loading all required libraries and dependencies
* "01\_0\_set\_input\_files.R”: script containing the input parameters as specified at **section 3.1**

Files in the script “00\_Get\_tabs\_Access.r”

* "data/processfiles/00\_0\_process\_specifs\_BFT\_ALB\_SKJ\_2021-11-15.RData" (fprocess)

From previous scripts in the module

* “outtables/BFT/20211115\_t\_abu\_larvind\_BFT\_2mm\_m2\_mixgears.csv” (fanalisis)

**Analysis**

The script performs the following analysis:

1. Imputation of NA values (fill CTD data where missing)
2. Compute residualtemp from CTD data: removes the effect of the day of the year (jd)
3. Compute environmental variables annual anomalies (Year\_anom)
4. Save .csv & .xlsx with derived variables (“20211115\_t\_abu\_larvind\_BFT\_interp\_insitu.csv”)

## Make habitat quality variables (“03\_2\_make\_habitat\_quality\_vars\_20180917.r”)

The script **“03\_3\_make\_habitat\_quality\_vars\_20180917.r”** generating and habitat quality variable using a model created by Walter Ingram.

**Inputs**

Scripts

* "/Rsource/settings\_m3.r" and all dependencies (i.e. scripts in Rsouce folder): script installing and loading all required libraries and dependencies
* "01\_0\_set\_input\_files.R”: script containing the input parameters as specified at **section 3.1**

Files in the script “00\_Get\_tabs\_Access.r”

* "data/processfiles/00\_0\_process\_specifs\_BFT\_ALB\_SKJ\_2021-11-15.RData" (frpocess)

From previous scripts in the module

* “outtables/BFT/20211115\_t\_abu\_larvind\_BFT\_interp\_insitu.csv” (fanalisis)

**Analysis**

The script performs the following analysis:

1. Generates a presence-absence GAM model including the interaction of longitude and latitude, the day of the year, the residual temperatura and the salinity in the mixed layer using a quasibinomial family
2. Performs the same model excluding one year at a time
3. Add the predicted values to the main abundance data.frame
4. Saves the generated table as .csv file (“20211115\_t\_abu\_larvind\_BFT\_allvars.csv”)

## Explore environmental variables (“03\_3 explore\_environ\_vars.r”)

The script **“03\_2\_explore\_environ\_vars.r”** is aimed at generating several plots to check the environmental variables.

**Inputs**

Scripts

* "/Rsource/settings\_m3.r" and all dependencies (i.e. scripts in Rsouce folder): script installing and loading all required libraries and dependencies
* "01\_0\_set\_input\_files.R”: script containing the input parameters as specified at **section 3.1**

Files in the script “00\_Get\_tabs\_Access.r”

* "data/processfiles/00\_0\_process\_specifs\_BFT\_ALB\_SKJ\_2021-11-15.RData" (fprocess)

From previous scripts in the module

* “outtables/BFT/20211115\_t\_abu\_larvind\_BFT\_allvars.csv” (fanalisis)

**Analysis**

The script performs the following analysis:

1. Boxplots of environmental variables (classical style)
2. Boxplots of environmental variables (Ottman style)
3. Histogram of environmental variables weighted by Abundance

# MODELLING

## Bluefin tuna strict update (07\_v3\_BFT\_strict\_update\_VPA\_lineal\_lsmeans\_20211214.r)

The script **“07\_v3\_BFT\_strict\_update\_VPA\_lineal\_lsmeans\_20211214.r”** is aimed at performing a standardisation of the larval abundance data to in situ and derived environmental data and perform an estimation of Bluefin tuna anual index with confidence intervals.

**Inputs**

Scripts

* "/Rsource/settings\_m3.r" and all dependencies (i.e. scripts in Rsouce folder): script installing and loading all required libraries and dependencies
* "01\_0\_set\_input\_files.R”: script containing the input parameters as specified at **section 3.1**

Files in the script “00\_Get\_tabs\_Access.r”

* "data/processfiles/00\_0\_process\_specifs\_BFT\_ALB\_SKJ\_2021-11-15.RData" (fprocess)

Other

* “outtables/BFT/20211115\_t\_abu\_larvind\_BFT\_allvars.csv” (fanalisis)
* modelID: character vector to identify the model being run

e.g.: “SU\_VPA" # Strict update of the VPA model; or any other character string that help identify the model and its outputs

e.g.:"SU\_MSEboot" # Strict update MSE bootstrap

e.g.:"AC\_deltagam": # Actualization of the delta gam model

**Analysis**

The script performs the following analysis:

1. Apply additional filters (if required)
2. Create definitive analysis table
3. Annual Mean CPUA (=nominalCPUA) summary table and plot
4. Histogram plots of the final datasets
5. Create salinity anomaly variable
6. Modelling: delta GLM
7. Estimate larval index estimate (least-squares means) combining the two parts of the delta model
8. Save outputs

## Bluefin tuna actualization (“10\_v1.1\_BFT\_strict\_update\_ MSE\_gam\_boostrap\_20211214.r”)

The script **“10\_v1.1\_BFT\_strict\_update\_MSE\_gam\_boostrap\_20211214.r”** is aimed at performing a standardisation of the larval abundance data to in situ and derived environmental data and perform an estimation of Bluefin tuna anual index with confidence intervals.

**Inputs**

Scripts

* "/Rsource/settings\_m3.r" and all dependencies (i.e. scripts in Rsouce folder): script installing and loading all required libraries and dependencies
* "01\_0\_set\_input\_files.R”: script containing the input parameters as specified at **section 3.1**

Files in the script “00\_Get\_tabs\_Access.r”

* "data/processfiles/00\_0\_process\_specifs\_BFT\_ALB\_SKJ\_2021-11-15.RData" (fprocess)

Other

* “outtables/BFT/20211115\_t\_abu\_larvind\_BFT\_allvars.csv” (fanalisis)
* modelID: character vector to identify the model being run

e.g.: “SU\_VPA" # Strict update of the VPA model; or any other character string that help identify the model and its outputs

e.g.:"SU\_MSEboot" # Strict update MSE bootstrap

e.g.:"AC\_deltagam": # Actualization of the delta gam model

**Analysis**

The script performs the following analysis:

1. Apply additional filters (if required)
2. Create definitive analysis table
3. Annual Mean CPUA (=nominalCPUA) summary table and plot
4. Histogram plots of the final datasets
5. Create salinity anomaly variable
6. Modelling: delta GAM
7. Estimate annual larval index and confidence intervals (bootstrap)
8. Save outputs

## Bluefin tuna actualization (“10\_v1.1\_BFT\_Rev\_202005 GAMd\_emmeans.r”)

## Albacore strict update (“11\_v6.1.0\_ALB\_Rev\_2\_GAMd\_emmeans\_20211214.r”) ????

The script **“11\_v6.1.0\_ALB\_Rev\_2\_GAMd\_emmeans\_20211214.r”** is aimed at performing a standardisation of the larval abundance data to in situ and derived environmental data and perform an estimation of Bluefin tuna anual index with confidence intervals.

**Inputs**

Scripts

* "/Rsource/settings\_m3.r" and all dependencies (i.e. scripts in Rsouce folder): script installing and loading all required libraries and dependencies
* "01\_0\_set\_input\_files.R”: script containing the input parameters as specified at **section 3.1**

Files in the script “00\_Get\_tabs\_Access.r”

* "data/processfiles/00\_0\_process\_specifs\_BFT\_ALB\_SKJ\_2021-11-15.RData" (fprocess)

Other

* “outtables/BFT/20211115\_t\_abu\_larvind\_BFT\_allvars.csv” (fanalisis)
* modelID: character vector to identify the model being run

e.g.: “SU\_VPA" # Strict update of the VPA model; or any other character string that help identify the model and its outputs

e.g.:"SU\_MSEboot" # Strict update MSE bootstrap

e.g.:"AC\_deltagam": # Actualization of the delta gam model

**Analysis**

The script performs the following analysis:

1. Apply additional filters (if required)
2. Create definitive analysis table
3. Annual Mean CPUA (=nominalCPUA) summary table and plot
4. Histogram plots of the final datasets
5. Create salinity anomaly variable
6. Modelling: delta GAM
7. Estimate annual larval index and confidence intervals (least-squares means) combining the two parts of the delta model
8. Save outputs

## Albacore strict update (“11\_v6.1.1\_ALB\_Rev\_3\_GAMd\_emmeans\_20211214.r”) ????

The script **“11\_v6.1.1\_ALB\_Rev\_3\_GAMd\_emmeans\_20211214.r”** is aimed at performing a standardisation of the larval abundance data to in situ and derived environmental data and perform an estimation of Bluefin tuna anual index with confidence intervals.

**Inputs**

Scripts

* "/Rsource/settings\_m3.r" and all dependencies (i.e. scripts in Rsouce folder): script installing and loading all required libraries and dependencies
* "01\_0\_set\_input\_files.R”: script containing the input parameters as specified at **section 3.1**

Files in the script “00\_Get\_tabs\_Access.r”

* "data/processfiles/00\_0\_process\_specifs\_BFT\_ALB\_SKJ\_2021-11-15.RData" (fprocess)

Other

* “outtables/BFT/20211115\_t\_abu\_larvind\_BFT\_allvars.csv” (fanalisis)
* modelID: character vector to identify the model being run

e.g.: “SU\_VPA" # Strict update of the VPA model; or any other character string that help identify the model and its outputs

e.g.:"SU\_MSEboot" # Strict update MSE bootstrap

e.g.:"AC\_deltagam": # Actualization of the delta gam model

**Analysis**

The script performs the following analysis:

1. Apply additional filters (if required)
2. Create definitive analysis table
3. Annual Mean CPUA (=nominalCPUA) summary table and plot
4. Histogram plots of the final datasets
5. Create salinity anomaly variable
6. Modelling: delta GAM
7. Estimate annual larval index and confidence intervals (least-squares means) combining the two parts of the delta model
8. Save outputs

## Albacore actualization (“11\_v6.1\_ALB\_AC\_202005\_GAMd\_emmeans\_20211214.r”)

The script **“11\_v6.1\_ALB\_AC\_202005\_GAMd\_emmeans\_20211214.r”** is aimed at performing a standardisation of the larval abundance data to in situ and derived environmental data and perform an estimation of Bluefin tuna anual index with confidence intervals.

**Inputs**

Scripts

* "/Rsource/settings\_m3.r" and all dependencies (i.e. scripts in Rsouce folder): script installing and loading all required libraries and dependencies
* "01\_0\_set\_input\_files.R”: script containing the input parameters as specified at **section 3.1**

Files in the script “00\_Get\_tabs\_Access.r”

* "data/processfiles/00\_0\_process\_specifs\_BFT\_ALB\_SKJ\_2021-11-15.RData" (fprocess)

Other

* “outtables/BFT/20211115\_t\_abu\_larvind\_BFT\_allvars.csv” (fanalisis)
* modelID: character vector to identify the model being run

e.g.: “SU\_VPA" # Strict update of the VPA model; or any other character string that help identify the model and its outputs

e.g.:"SU\_MSEboot" # Strict update MSE bootstrap

e.g.:"AC\_deltagam": # Actualization of the delta gam model

**Analysis**

The script performs the following analysis:

1. Apply additional filters (if required)
2. Create definitive analysis table
3. Annual Mean CPUA (=nominalCPUA) summary table and plot
4. Histogram plots of the final datasets
5. Create salinity anomaly variable
6. Modelling: delta GAM
7. Estimate annual larval index and confidence intervals (least-squares means) combining the two parts of the delta model
8. Save outputs

# ANNEX I. STRUCTURE OF THE GENERATED TABLES

Table 1. Structure of the abundance files for multiple species.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Variable | Type | Units | Range | Observations |
| 1 | idIBAMAR | int |  |  |  |
| 2 | id\_netcolector | int |  |  |  |
| 3 | survey | chr |  |  |  |
| 4 | station\_id | chr |  |  |  |
| 5 | station\_order | int |  |  |  |
| 6 | id\_operation | int |  |  |  |
| 7 | revisita | int |  |  |  |
| 8 | fecha\_sistematic | int |  |  |  |
| 9 | tipo\_estacion | int |  |  |  |
| 10 | fecha\_estacion\_llegada | chr |  |  |  |
| 11 | hour | chr |  |  |  |
| 12 | hournorm | num |  |  |  |
| 13 | Lat | num | decimal degrees |  |  |
| 14 | lon | num | decimal degrees |  |  |
| 15 | fishing\_type | chr |  |  |  |
| 16 | gear | chr |  |  |  |
| 17 | cont\_estructura | int |  |  |  |
| 18 | malla | int | µm |  |  |
| 19 | towdepth | int | m |  | Maximum depth of the collector |
| 20 | volume | num | m3 |  |  |
| 21 | selec\_analisis | int |  |  |  |
| 22 | larval\_index | int |  | 1,2,3,4 |  |
| 23 | grupo\_para\_procesar | chr |  |  |  |
| 24 | Katsuwonus\_pelamis | int | n larvae |  | Number of larvae |
| 25 | Thunnus\_alalunga | int | n larvae |  | Number of larvae |
| 26 | Thunnus\_thynnus | int | n larvae |  | Number of larvae |
| 27 | date | chr |  |  |  |
| 28 | year | int |  |  |  |
| 29 | month | int |  |  |  |
| 30 | day | int |  |  |  |
| 31 | jd | num |  | 1-365 |  |
| 32 | cathour | factor |  |  |  |
| 33 | daynight | chr |  |  |  |
| 34 | lunar\_illum | num |  | 0-1 |  |
| 35 | lunar\_phase | factor |  | 8 phases | New, Waxing crescent, First quarter, Waxing gibbous,  Full, Waning gibbous, Third quarter, Waning crescent |
| 36 | bottom\_dep | num | m |  | From a local geotiff file of NOAA global bathymetry (1 degree resolution) |
| 37 | slope | num | degrees |  | Bottom sea slope |
| 38 | distance2coast | num | Km |  | Distance to the nearest shore |
| 39 | towdepthtemp | num | m | 5-70 |  |
| 40 | prof\_mld | num | m |  |  |
| 41 | TMEZCLA | num | ºC |  |  |
| 42 | SMEZCLA | num |  |  |  |
| 43 | OMEZCLA | num | ml l-1 |  |  |
| 44 | suma\_flu\_zmezcla | num | mg m-3 |  |  |

Table 2. Structure of the table for larval index analysis (t\_analysis) for a single species. Variables from 34 to 45 (besides 41, which contains the amount of larvae of the selected species) have been added in the script (“*00\_1b\_apply\_filters.r”)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Variable | Type | Units | Range | Observations |
| 1 | idIBAMAR | int |  |  |  |
| 2 | id\_netcolector | int |  |  |  |
| 3 | survey | chr |  |  |  |
| 4 | station\_id | chr |  |  |  |
| 5 | station\_order | int |  |  |  |
| 6 | id\_operation | int |  |  |  |
| 7 | revisita | int |  |  |  |
| 8 | fecha\_sistematic | int |  |  |  |
| 9 | tipo\_estacion | int |  |  |  |
| 10 | fecha\_estacion\_llegada | chr |  |  |  |
| 11 | hour | chr |  |  |  |
| 13 | hournorm | num |  |  |  |
| 14 | lat | num |  |  |  |
| 15 | lon | num |  |  |  |
| 16 | fishing\_type | chr |  |  |  |
| 17 | gear | chr |  |  |  |
| 18 | cont\_estructura | int |  |  |  |
| 19 | malla | int |  |  |  |
| 20 | towdepth | int |  |  |  |
| 21 | volume | num |  |  |  |
| 22 | selec\_analisis | int |  |  |  |
| 23 | larval\_index | int |  |  |  |
| 24 | grupo\_para\_procesar | chr |  |  |  |
| 25 | Thunnus\_thynnus | int |  |  |  |
| 26 | date | chr |  |  |  |
| 27 | year | int |  |  |  |
| 28 | month | int |  |  |  |
| 29 | day | int |  |  |  |
| 30 | jd | num |  |  |  |
| 31 | cathour | factor |  |  |  |
| 32 | daynight | chr |  |  |  |
| 33 | lunar\_illum | num |  | 0-1 |  |
| 34 | lunar\_phase | factor |  | 8 phases | New, Waxing crescent, First quarter, Waxing gibbous,  Full, Waning gibbous, Third quarter, Waning crescent |
| 35 | bottom\_dep | num | m | >0 | From a local geotiff file of NOAA global bathymetry (1 degree resolution) |
| 36 | slope | num | degrees |  | Bottom sea slope |
| 37 | distance2coast | num | Km |  | Distance to the nearest shore |
| 38 | towdeptemp | num | m | 5-70 |  |
| 39 | prof\_mld | num | m |  |  |
| 40 | TMEZCLA | num | ºC |  |  |
| 41 | SMEZCLA | num |  |  |  |
| 42 | OMEZCLA | num | ml l-1 |  |  |
| 43 | suma\_flu\_zmezcla | num | mg m-3 |  |  |
| 44 | lpres | num |  | 0-1 |  |
| 45 | nlarvae\_m3 | num | n larvae m-3 |  | Number of larvae by volum of water filtered |
| 46 | CPUAnom | num | n larvae m-2 |  | Number of larvae by volum of water filtered and multiplied by towing depth |
| 47 | CPUAnom\_temp | num | n larvae m-2 |  | Number of larvae by volum of water filtered and multiplied by towing depth temp (towdeptemp) |

Table 3. Structure of the length-frequency data files for a single species.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Variable | Type | Range | Observations |
| 1 | id\_grupo\_autonum | int |  |  |
| 2 | larval\_index | int |  |  |
| 3 | net\_colector | int |  |  |
| 4 | codestacion | chr |  |  |
| 5 | survey | chr |  |  |
| 6 | gear | chr |  |  |
| 7 | idoperacion\_num | int |  |  |
| 8 | volume | num |  |  |
| 9 | hournorm | num |  |  |
| 10 | horallegada | chr |  |  |
| 11 | fecha\_estacion\_llegada | chr |  |  |
| 12 | towdepth | int |  |  |
| 13 | ssp\_name | chr |  |  |
| 14 | length | num |  |  |
| 15 | nsampled | int |  |  |
| 16 | sppid | int |  |  |
| 17 | frec | num |  |  |
| 18 | year | int |  |  |