

RoME

R Code to Perform Multiple Checks on MEDITS Survey Data

Second Training on quality checks
18/04/2024

Walter Zupa,
Loredana Casciaro,
Cosmidano Neglia,
Isabella Bitetto



COISPA

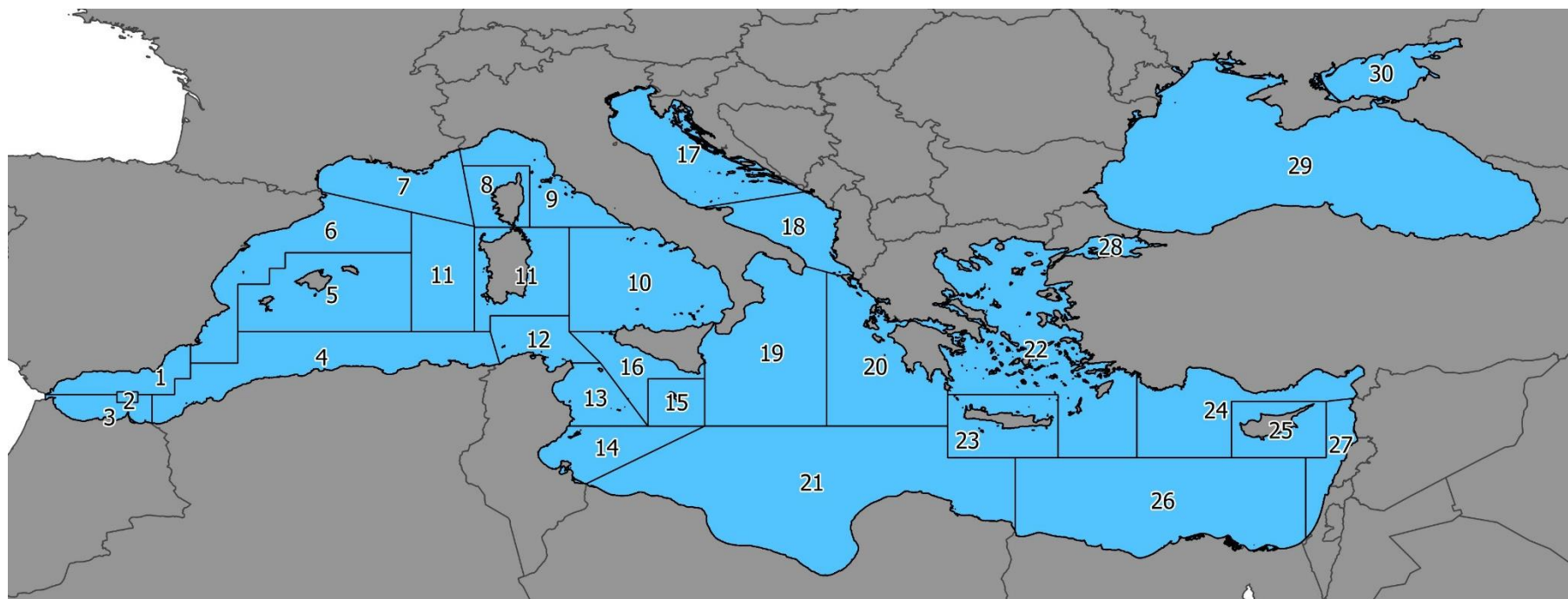


CNR
IRBIM
Istituto Nazionale
per lo Studio e la
Ricerca in
Bioscienze Marine



Data quality check tool

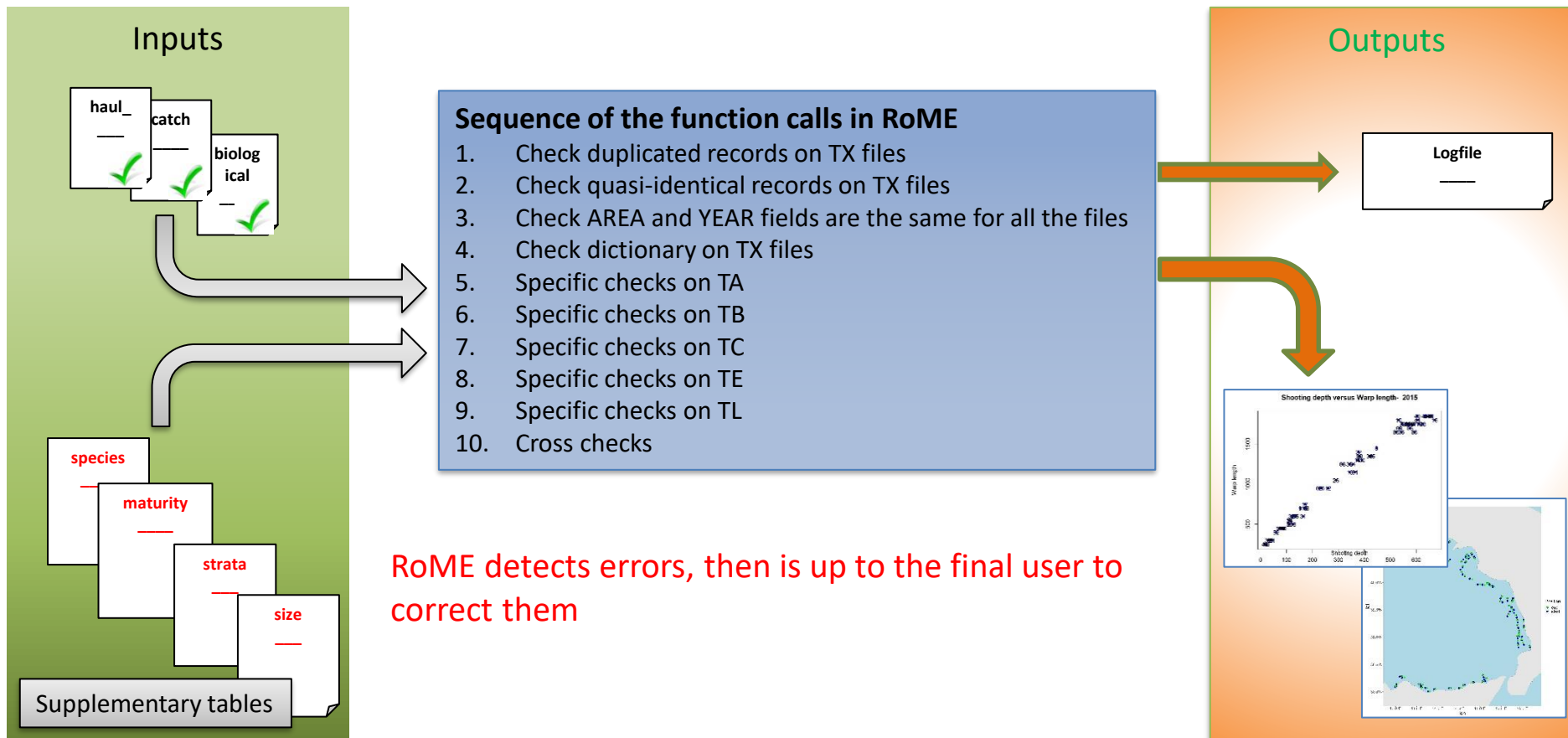
- RoME was presented for the first time in the MEDITS Coordination meeting held in Nantes (March 2011)
- RoME represents a common tool to perform the data checks with a standardized procedure within MEDITS coordination.
- New updated version (v.0.1.35) with specific function to be run in RDBFIS environment



Mean features

Input: Exchange data tables in .csv format (TA, TB and TC, TE, TL) and additional tables from MEDITS manual (TM list, maturity scales table, stratification, etc..)

Output: Report of errors, Text file (Logfile.dat) with the outcome of each check and errors/warnings specifications and plots for qualitative controls.



```
#####  
# RoME v.0.1.22 R Code to Perform Multiple Checks on MEDITS Survey Data  
# W. Zupa, I. Bitetto, M.T. Spedicato  
#  
# In case of use of the software, the Authors should be cited.  
# If you have any comments or suggestions please contact the following  
# e-mail address: zupa@coispa.it  
#  
# This program is free software; you can redistribute it and/or  
# modify it under the terms of the GNU General Public License  
# as published by the Free Software Foundation; either version 2  
# of the License, or (at your option) any later version.  
#  
# This program is distributed in the hope that it will be useful,  
# but WITHOUT ANY WARRANTY; without even the implied warranty of  
# MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the  
# GNU General Public License for more details.  
#####
```

Environment

R version > 4.1.2

Rstudio: v. 2022.12.0

License: GPL-3

Version 0.1.35

NEW FEATURES

- Completely updated version, compatible with the more recent versions of R software and dependencies
- Multiplatform (Windows and Linux for RDBFIS)
- 67 standalone check functions
- Documentation embedded in the package
- Runs embedded in RDBFIS and stand-alone
- Improved quality of plots and maps;
- Specific functions for RDBFIS
 - RoMEcc to analyse the MEDITS tables without interruption and report all the errors occurred (error report + logfile)
 - all the outputs of the analysis are also stored in a local zip file

Initialization

RoME 0.1.26 can be downloaded directly from COISPA github:

<https://github.com/COISPA/RoME>

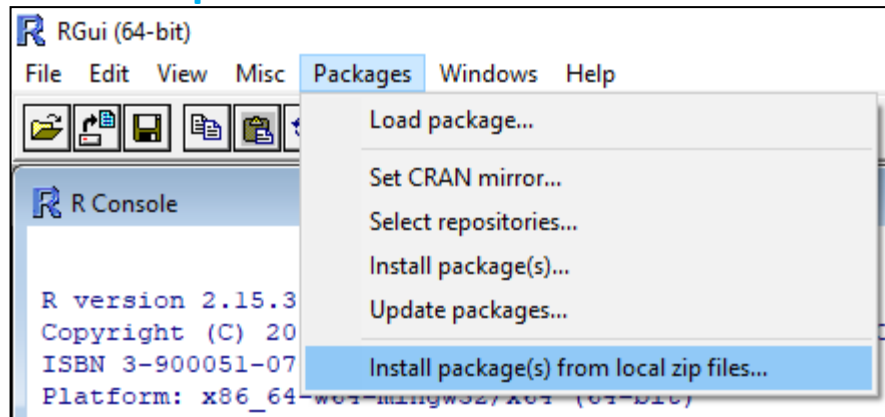


- RoME is distributed as a package to be loaded directly in R console.
- RoME has been produced and runs with the R version > 4.1

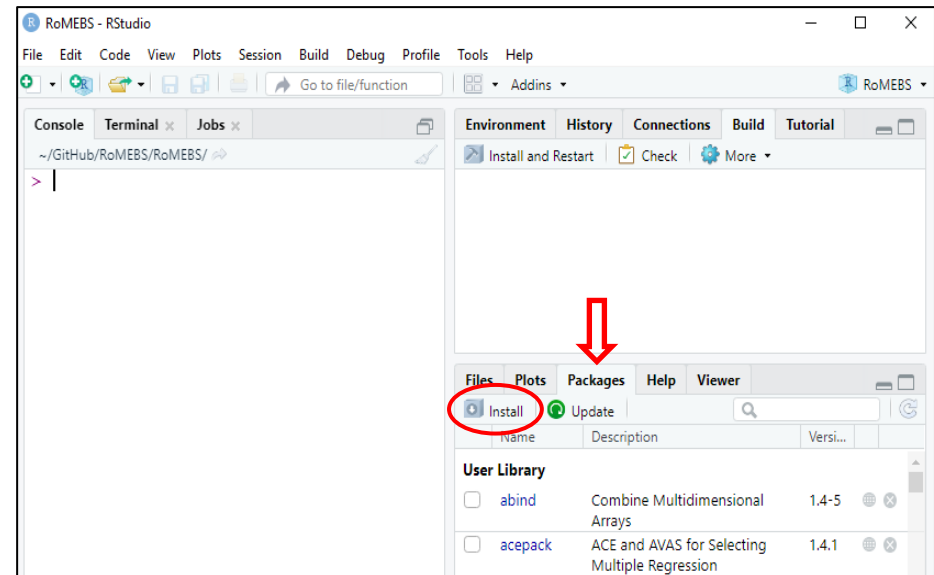
Dependencies

- timeDate
- Stringr
- ggplot2
- rnatrualearth
- rnatrualearthdata

Install dependencies



Install also the **RoME_0.1.35.zip** package



Library installation

```
### Install RoMEBS packages and dependencies ###  
library(devtools)  
install_github("COISPA/RoME")
```

Load RoME library

```
library(RoME)
```

Info about RoME function

?RoME

?RoMEcc

RoME {RoME}

R Documentation

R code to perform multiple checks on MEDITS Survey data (TA, TB, TC, TE and TL files)

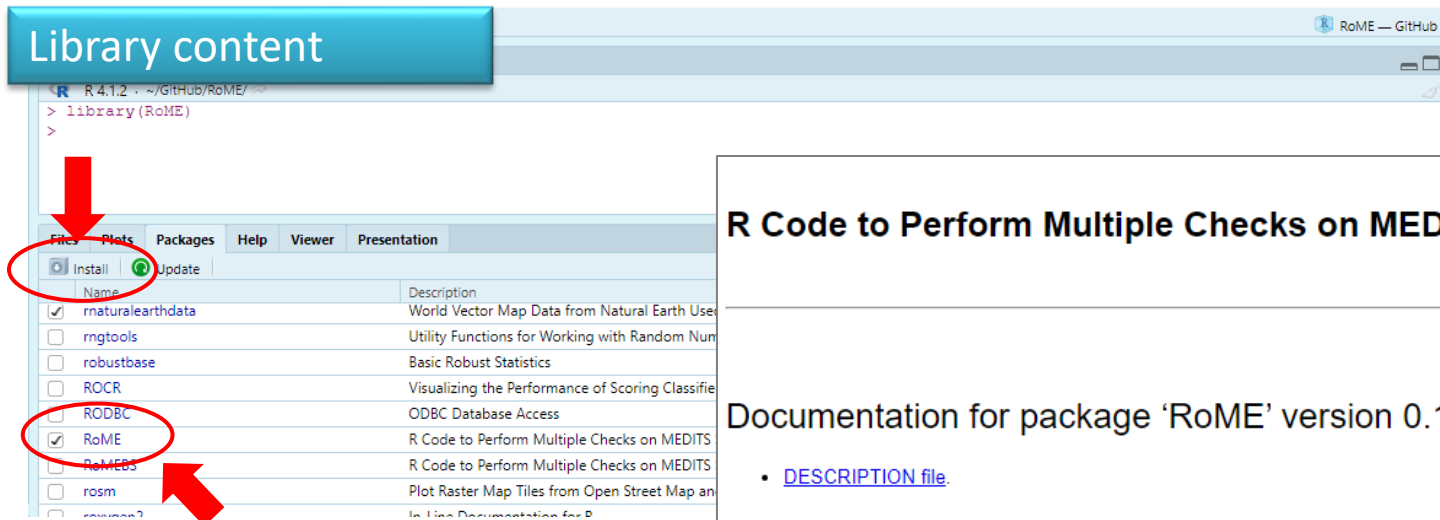
Description

The function calls all the functions built in the package in an ordered way to perform a complete quality check of TX data available. The check is performed simultaneously on the files that can contain also data of more than one year.

Usage

```
RoME(TA,TB,TC,TE=NA,TL=NA,wd,suffix=NA,  
create_RSufi_files = FALSE, create_global_RSufi_files=FALSE,  
Year_start=NA,Year_end=NA,verbose =TRUE,Stratification=RoME::stratification_scheme,  
TM_list=TM_list,DataTargetSpecies=DataTargetSpecies,  
Maturity_parameters=Maturity_parameters,  
ab_parameters=RoME::LW,  
stages_list=RoME::mat_stages,assTL=assTL)
```

Library content



R 4.1.2 - ~/GitHub/RoME/

```
> library(RoME)
```

Files Plots Packages Help Viewer Presentation

Install Update

Name	Description
☒ rnatuarearthdata	World Vector Map Data from Natural Earth Use
☐ rngtools	Utility Functions for Working with Random Num
☐ robustbase	Basic Robust Statistics
☐ ROCR	Visualizing the Performance of Scoring Classifie
☐ RODBC	ODBC Database Access
☒ RoME	R Code to Perform Multiple Checks on MEDITS
☐ RoMEBS	R Code to Perform Multiple Checks on MEDITS
☐ rosm	Plot Raster Map Tiles from Open Street Map an
☐ rosm2	In-Line Documentation for R

R Code to Perform Multiple Checks on MEDITS Survey Data

Documentation for package 'RoME' version 0.1.22

- [DESCRIPTION file.](#)

Help Pages

Function	Description
assTL	TL association between categories and sub-categories
checkHeader	Function to check the correctness of the headers.
check_0_fieldsTA	Checks the presence of 0 fields in TA
check_area	Check if TX files have the same area
check_associations_category_TL	Check correctness of TL categories
check_bridles_length	check of bridles length correctness
check_class	Check of field's class
check_consistencyTA_distance	Consistency check of distance in TA
check_consistencyTA_duration	Consistency check of hauls duration in TA
check_date_haul	Check of date consistency
check_depth	Check between start depth and end depth
check_dictionary	Check of the dictionary of specific fields
check_distance	Check of distance consistency
check_dm	Check of "WING_OPENING" and "VERTICAL_OPENING" fields
check_G1_G2	Check of length measurements for G1 and G2 species
check_hauls_TATB	Check of TA hauls in TB
check_hauls_TATL	Check presence of TA hauls in TL
check_hauls_TBTA	Check of TB hauls in TA
check_hauls_TLTA	Check presence of TL hauls in TA
check_haul_species_TCTB	Check species of TC in TB
check_identical_records	Check of identical records in TX tables
check_individual_weightTC	Check of observed and estimated total weight in the haul
check_individual_weightTE	Consistency of individual weights (according to length-weight relationship)

Complete documentation for each element included in the library



Check total number of individuals in TB

Documentation format

Description

Briefly describe what the function does

The function checks that the total number of individuals is consistent with the sum of the individuals per sex

Usage

Synopsis of the function(s) and variables documented

```
check_nbtotTB(DataTB, wd, suffix)
```

Arguments

Details on the characteristics of the parameters to pass to the function

`DataTB` Catch data table according to MEDITS protocol (TB)
`wd` working directory path defined by the user
`suffix` Suffix string of the Logfile

Details

Provides any other important details of how the function operates

The function checks that the total number of individuals is consistent with the sum of the individuals per sex

Value

Description of the function's return value (outputs)

The function returns TRUE if no error occurs, FALSE if one or more inconsistencies in the individuals number is detected.

Author(s)

I. Bitetto, W. Zupa

References

Anonymus. 2017. MEDITS-Handbook. Version n. 9. MEDITS Working Group. 106 pp. <https://www.sibm.it/MEDITS/2011/principaledownload.htm>

Examples

provides executable R code showing how to use the function in practice

```
wd <- tempdir()
suffix="2020-03-05_time_h17m44s55"
DataTB = MEDITS::TB
check_nbtotTB(DataTB, wd, suffix)
```

Launch RoMEcc function (RoMEBSc)

1. Load RoME library in R memory

```
library(RoME)
```

2. Load TA, TB, and TC file (mandatory)
3. Optionally, load TE and/or TL files

RoME processes data from a single GSA per run

4. run RoME function to check survey data

```
RoMEcc(TA, TB, TC, TE, TL,  
       wd, suffix, verbose=TRUE,  
       Stratification=Stratification,  
       Ref_list=TM_list,  
       DataTargetSpecies=DataTargetSpecies,  
       Maturity=Maturity_parameters,  
       ab_parameters=ab_parameters,  
       stages_list=stages,  
       assTL=assTL)
```

Input TX tables

Output parameters

Other useful tables

Demo tables already
included in the R package.

Personalized by user

Already included in the R
package.
Modified tables could be
used instead

RoMEcc analysis

```
[1] Check Headers TA in progress...
[1] Check Headers successfully completed!
[1] Check Headers TB in progress...
[1] Check Headers successfully completed!
[1] Check Headers TC in progress...
[1] Check Headers successfully completed!
[1] Check CLASS TA in progress...
[1] Check CLASS successfully completed!
[1] Check CLASS TB in progress...
[1] Check CLASS successfully completed!
[1] Check CLASS TC in progress...
[1] Check CLASS successfully completed!
[1] Check YEAR in progress...
[1] Check YEAR successfully completed!
[1] Check TYPE_OF_FILE in progress...
[1] Check TYPE_OF_FILE successfully completed!
[1] Checking year 2007
[1] Check identical record TA in progress...
[1] Check identical record TA successfully completed!
[1] Check identical record TB in progress...
[1] Check identical record TB successfully completed!
[1] Check identical record TC in progress...
[1] Check identical record TC successfully completed!
[1] Check quasi-identical record in TA in progress...
[1] Check quasi-identical record in TA successfully completed!
[1] Check quasi-identical record in TB in progress...
[1] Check quasi-identical record in TB successfully completed!
[1] Check quasi-identical record in TC in progress...
[1] Check quasi-identical record in TC successfully completed!
[1] Check consistency of area TA, TB, TC, TE, TL in progress...
[1] Check consistency of area TA, TB, TC, TE, TL successfully completed!
```

Order of the checks

- CHECK HEADERS
- CHECK CLASS
- CHECK YEAR
- CHECK TYPE_OF_FILE
- CHECKS by YEAR
 - Check identical record
 - Check quasi-identical record
 - Check consistency of area
 - Dictionary checks
 - Check numeric ranges
 - Check no empty fields
 - Specific checks on TA
 - Specific checks on TB
 - Specific checks on TC
 - Specific checks on TE
 - Specific checks on TL
 - Cross checks

CRITICAL
CHECKS



TA table (haul data)

1. Check the class of the fields included in TA (e.g. numeric)
2. Check the correctness of the headers
3. Check the presence of duplicated rows
4. Check quasi-identical record
5. Check the presence of empty fields
6. Check the consistence of the values contained in specific fields with the relative allowed values (dictionaries)
7. Check numeric range according to MEDITS protocol
8. Check uniqueness of valid hauls

Syntax checks

9. Visual check of the haul positions
10. Check start quadrant and end quadrant by haul
11. Check if the coordinates are in the Mediterranean or Black Sea and/or inside the GSA border
12. Check consistency between distance and duration of the haul
13. Check consistency between duration and time by haul
14. Check consistency of the hauls coordinates with the distance (difference not greater than 30%)

Checks on hauls position

15. Check if start depth and end depth are in the same stratum by haul
16. Check consistency of stratum code according to the mean depth of the haul
17. Check difference between start depth and end depth (not greater than 20%) by haul

Checks on depth

18. Check consistency of bridles length according to MEDITS protocol
19. Check the presence of 0 fields, according to MEDITS protocol, where not allowed (e.g. WING_OPENING, WARP_DIAMETER and VERTICAL_OPENING)
20. Check if the values in "WING_OPENING" and "VERTICAL_OPENING" fields are in the allowed ranges according to MEDITS protocol
21. Check on the relation between shooting depth and warp length, and between warp length and wing opening
22. Check consistency of temperature by haul

Checks on gear data

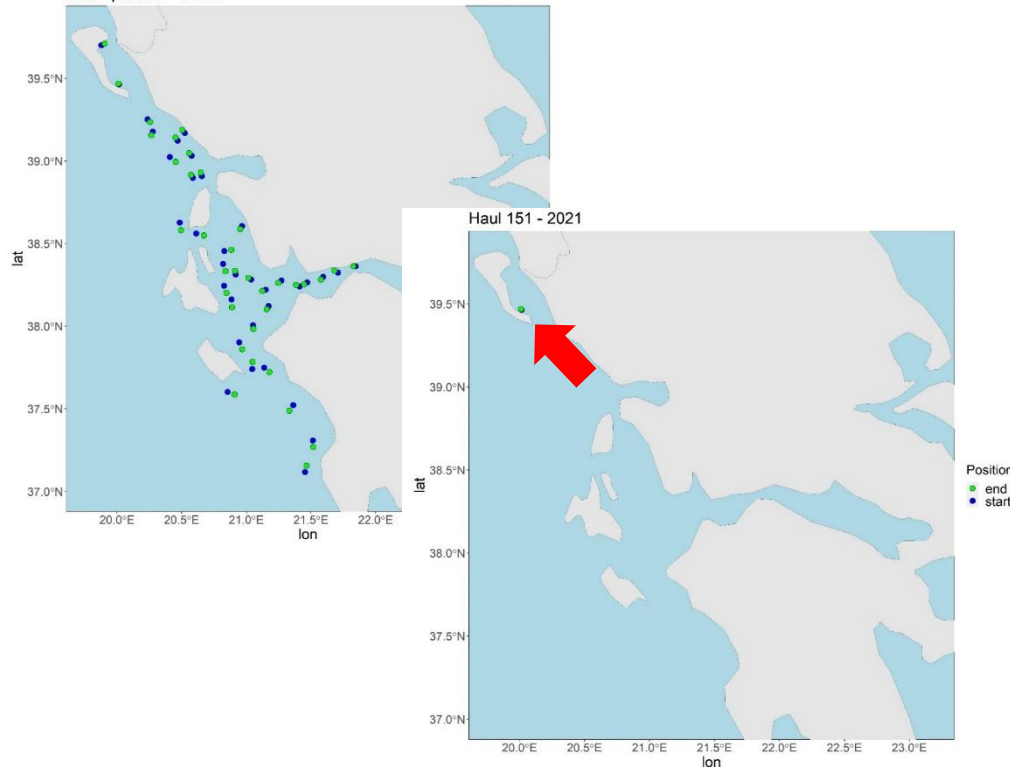


FRAMEWORK CONTRACT – EASME/EMFF/2020/OP/021

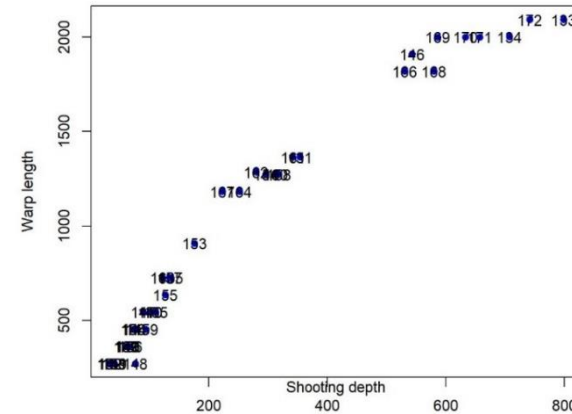


Check distance

Hauls position - 2021

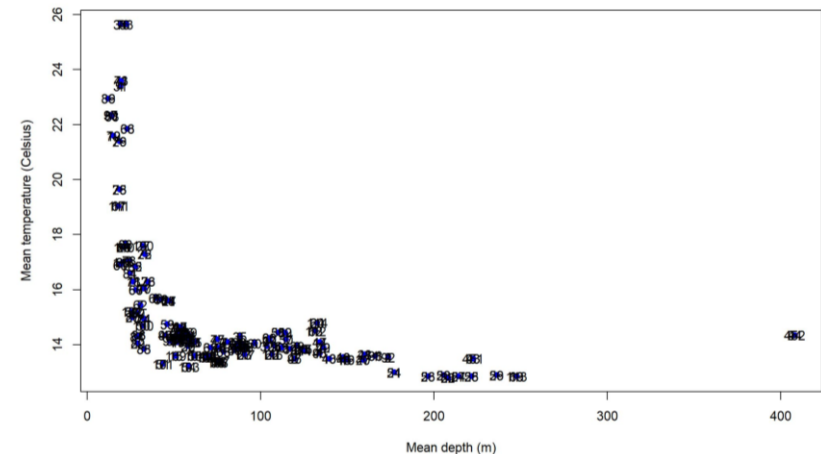


Shooting depth versus Warp length- 2021



Check temperature

Temperature data - 2021



TB table (catch data)

1. Check the class of the fields included in TB (e.g. numeric)
2. Check the correctness of the headers
3. Check the presence of duplicated rows
4. Check quasi-identical record in TB
5. Check the presence of empty fields
6. Check dictionary according to MEDITS protocol
7. Check numeric range according to MEDITS protocol
8. Check correctness of species codes according to MEDITS protocol

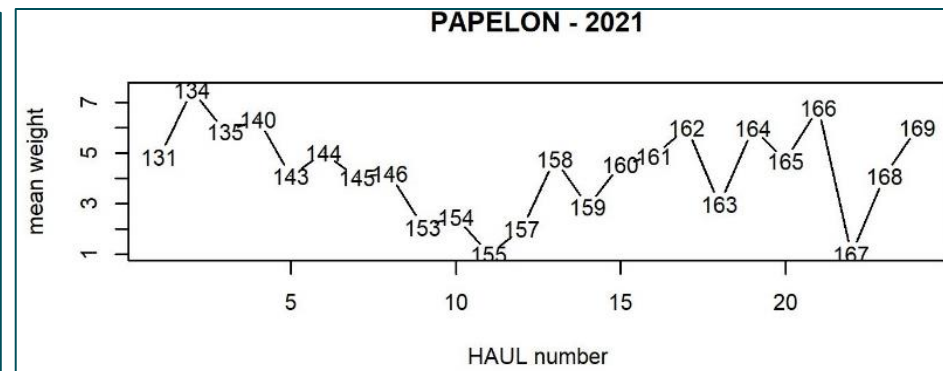
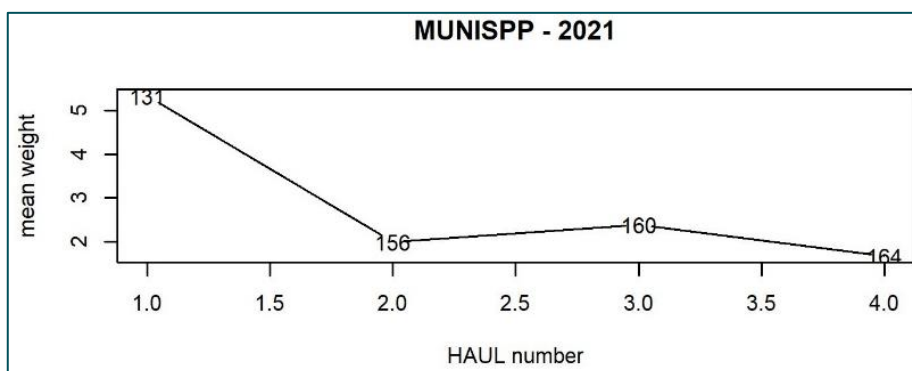
Syntax checks

1. Check consistency between not null weight and not null total number
2. Check if the total number of individuals is consistent with the sum of the individuals per sex
3. Check if number of individuals and total weight are consistent
4. Check presence of total number and number per sex TB for species G1

Checks on numbers and weights

Check mean weight

Plots of mean weight by haul and species are produced for the qualitative control to identify possible outliers.



TC table (length and aggregated biological data)

1. Check the class of the fields included in TC (e.g. numeric)
2. Check the correctness of the headers
3. Check the presence of duplicated rows
4. Check quasi-identical record in TC
5. Check the presence of empty fields
6. Check dictionary according to MEDITS protocol
7. Check numeric range according to MEDITS protocol
8. Check correctness of species codes according to MEDITS protocol

Syntax checks

Cross checks

1. Check the correctness of the number and weight in TB and in TC
2. Check on date by haul
3. Check presence in TA of TB hauls
4. Check presence in TB of TA hauls
5. Check presence in TB of TC species
6. Check presence in TC of TB target species

9. Check correctness of LENGTH_CLASSES_CODE according to MEDITS protocol
10. Check if the length measures are reported with the correct precision
11. Check presence of lengths for G1 and G2 MEDITS species
12. Check consistency of length distribution, whether the length classes by species are included in the range reported in the DataSpecies dataset

Checks on length data

13. Check about the presence of subsamples < 10% of the total catch
14. Check correctness of number per sex in TC
15. Check total weight in the haul in TC

Checks on numbers and weights

16. Check consistency of maturity stages according to MEDITS protocol
17. Check consistency of maturity stages TC by the comparison with the length of smallest mature individuals reported in bibliography

Checks on maturity stages



Output of the analysis

Outputs in the console

```
...  
[1] Check presence in TB of TC species in progress...  
[1] Check presence in TB of TC species successfully completed!  
[1] Check correctness of the number per sex in TB in case of sub-sampling in TC in progress...  
[1] Check correctness of the number per sex in TB in case of sub-sampling in TC successfully  
completed!  
[1] Check on date by haul TB in progress...  
[1] Check on date by haul TB successfully completed!  
[1] Check on date by haul TC in progress...  
[1] Check on date by haul TC successfully completed!  
[1] All the checks have been performed!
```

Outputs in summary of the errors

errors	check type	table	variable	year
	Check Headers	TA	NA	NA
	Check Headers	TB	NA	NA
	Check Headers	TC	NA	NA
X	Check CLASS	TA	NA	NA
	Check CLASS	TB	NA	NA
	Check CLASS	TC	NA	NA
X	Check YEAR	ALL	NA	NA

ERRORS

Outputs in logfile

```
----- check consistency of TOTAL_NUMBER_IN_THE_HAUL and number per sex in TB - 2021  
Error: Haul 171 ARIT ANT : NB_TOTAL doesn't equal NB_F+NB_M+NB_I TB  
  
----- check consistency between duration and distance TA - 2015  
Warning: in haul 40 distance measure 3430 inconsistent with the duration of the haul ( 60 min )  
No error occurred  
  
----- check consistency of the hauls coordinates with the distance - 2015  
Warning: Haul 40 : the distance in TA 3430 is quite different from the computed distance  
5315.8/13 (haul duration: 60 min)  
Some of the hauls coordinates may be inconsistent with the computed distance. For a visual  
check, look at the jpeg files in Graphs directory  
No error occurred
```

WARNINGS

logfile

Wed Jan 18 17:36:57 2023

CHECK RESULTS

----- check of HEADERS
No error occurred

----- check of HEADERS
No error occurred

----- check of HEADERS
No error occurred
No error occurred
No error occurred
No error occurred

----- check on YEAR field
No error occurred

----- check on TYPE_OF_FILE field
No error occurred

----- check identical records - 2015
TA:
No error occurred
TB:
No error occurred
TC:
No error occurred

----- check quasi identical records - 2015
TA:
No error occurred
TB:
No error occurred
TC:
No error occurred

Warning

----- check consistency between duration and distance TA - 2015
Warning: in haul 40 distance measure 3430 inconsistent with the duration of the haul (60 min)
No error occurred

----- check consistency of the hauls coordinates with the distance - 2015
Warning: Haul 40 : the distance in TA 3430 is quite different from the computed distance 5315.8713 (haul duration: 60 min)
Some of the hauls coordinates may be inconsistent with the computed distance. For a visual check, look at the jpeg files in Graphs directory
No error occurred

Error

logfile

----- check consistency of TOTAL_NUMBER_IN_THE_HAUL and number per sex in TB - 2021
Error: Haul 171 ARIT ANT : NB_TOTAL doesn't equal NB_F+NB_M+NB_I TB

Report of the errors

	Check on temperature by haul	TA	NA	2021
	Check correctness of species codes TB	TB	NA	2021
X	Check consistency of NB_TOTAL and number per sex TB	TB	NA	2021
	Check presence of NB_TOTAL and number per sex TB for species G1	TB	NA	2021
	Check consistency between not null weight and not null total number	TB	NA	2021

GENERAL CHECK

Check quasi-identical record

Two or more "**quasi-identical records**" occurred when all the fields are respectively equal, except: (these specific fields are allowed to be identical)

Error

- TYPE_OF_FILE
- AREA
- GEAR
- VESSEL
- YEAR
- RIGGING
- DOORS

TA table

- TYPE_OF_FILE
- AREA
- VESSEL
- YEAR

TB and TC table

Check of identical records

The function checks where there is one or more identical records in the selected type of table

Error

Check consistency of area

checks if TX files have the same area code

Error

GENERAL CHECK

Check of the dictionary of specific fields

The function checks whether the values contained in specific fields are consistent with the **allowed values** of the dictionaries.

Error

Check no empty fields

The function checks whether the values contained in specific fields are **empty**

Error

Check numeric range

The function checks whether values in a specific fields are consistent within the allowed range of values

Error

CHECK for TA

Check consistency between duration and time

The function checks whether the durations reported in the haul data (TA) are consistent with the differences between HAULING_TIME and SHOOTING_TIME.

Error

Check consistency between distance and duration

The function checks whether the distances reported in the haul data (TA) are consistent with the hauls duration.

Error

CHECK for TA

Check difference between start depth and end depth (not greater than 20%)

Check if that difference between start depth and end depth is not greater than 20%

Warning

Check consistency of stratum code

Checks whether start depth and end depth of each haul are in the same stratum.

Warning

verifies the correctness of the stratum code, following the stratification scheme table in the MEDITS protocol.

Warning

Check position

Visual check of the haul positions

Warning

Check if the coordinates are on the Sea

Warning

Check consistency of the hauls coordinates with the distance

The function checks whether there are inconsistencies between the DISTANCE field values and the computed distance.

Warning

CHECK for TB

Check correctness of species codes

Function checking the correctness of species code and faunistic category according to TM reference list

Warning

Check consistency of TOTAL number and number per sex in TB

The function checks that the total number of individuals is consistent with the sum of the individuals per sex

Error

Check consistency of weight and number TB

Function that checks if number of individuals and total weight collected in the haul are consistent. The check is quantitative if there is information related to average individual weight. Alternatively the check is qualitative (through a plot).

Warning

Check of consistency in number per sex in G1 species

The function checks if total number, number of females, males and undetermined for G1 species are simultaneously null

Error

CHECK for TC

Check correctness of species codes

Function checking the correctness of species code and faunistic category according to TM reference list

Error

Check correctness of LENGTH_CLASSES_CODE

The function checks whether the LENGTH_CLASS_CODE by species are consistent with those reported in the TM list.

Error

Check consistency of length classes TC

The function checks whether the length classes by species are included in the range reported in the DataTargetSpecies table.

Warning

Check correctness of number per sex in TC

The function checks the consistency of the number of individuals by sex measured (NO_OF_INDIVIDUAL_OF_THE_ABOVE_SEX_MEASURED field in TC) with the sum of the individuals by sex, length class and maturity stage (NUMBER_OF_INDIVIDUALS_IN_THE_LENGTH_CLASS_AND_MATURITY_STAGE)

Error

CHECK for TC

Check sub-sampling

Function to warn the user about the presence of subsamples' ratio < 0.1 of the total catch

Warning

Check of observed and estimated total weight in the haul

warning is given when difference between the sum of **estimated individual weights** (by haul, species and sub-samples) and the **WEIGHT_OF_THE_SAMPLE_MEASURED** is greater than 50% for at least one record. This check is based on the LW table contained in the package.

Warning

Check consistency of length distribution TC

The function verifies that in TC the length measures are reported with the correct precision

Error

Check correctness of species codes

Function checking the correctness of species code and taxonomic category according to TM reference list

Error

cross checks

Check presence in TB of TA hauls - Check presence in TA of TB hauls

The function checks:

- presence of the TA hauls in the TB
- presence of the TB hauls in the TA

Error

Check presence in TC of TB target species - Check presence in TB of TC species

The function checks whether

- all the species present in TC are listed in TB
- all the species in the TB are in the TC

Error

Warning

Check correctness of the number per sex in TB in case of sub-sampling in TC

Function checking if, in case of sub-sampling in TC, the Total number and the number per sex in TB is raised correctly

Error

Check on date by haul

Check if in TB, TC and TE the date by haul is the same of the one reported in TA

Error

Conclusions

The RoME package:

- ❑ **Open source** code, publicly available on **GithHub**
- ❑ Can be used with 3 different functions: **RoME, RoMEcc, RoMEBScc**
- ❑ Supports the **fast identification of errors** and inconsistencies in survey data
- ❑ Allow to **harmonize** the MEDITS survey data quality checks **among the Countries**
- ❑ Designed to facilitate a more **systematic** and **well-defined schedule** for quality checks prior the relevant data calls
- ❑ Globally **increases the data quality**