

**RoME (version 1.4):
R code to perform multiple checks on MEDITS Survey
data
(TA, TB, TC, TD, TT and TE files)**

Authors: Isabella Bitetto, Maria Teresa Facchini, Maria Teresa Spedicato

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

If you have any comments or suggestions please contact the following e-mail address:

bitetto@coispa.it



Index

1. Objectives.....	3
2. Functions contained in RoME	3
2.1 Check on TA file.....	4
2.2 Check on TB file.....	7
2.3 Check on TC file.....	8
2.4 Check on TE file.....	11
2.5 Check on TD-TT files.....	12
2.6 Cross-checks	13
2.5 Creation of R-Sufi files.....	14
3. Order of the checks.....	14
4. Function RoME ()	15
5. Instructions for the user	15

1. Objectives

In order to unify the checks that are made independently over the MEDITS data by the 18 GSAs (Geographical Sub-Area) participating to MEDITS Survey, an R code performing all the main checks on exchange MEDITS data tables was developed.

The first version of RoME has been presented for the first time in the MEDITS Coordination meeting held in Nantes (March 2011). The use of RoME by experts belonging to different GSAs participating to MEDITS Programme brought to the implementation of several adjustments of the R code.

In the version 1.3, RoME has been transformed in a package structured in 55 different functions: the run is performed by means of the function `ROME()`. Each function is related to a specific check and is recalled in a specific order to avoid cascade errors. This is also mantaned in RoME 1.4.

The software does not correct the data, but it detects the errors, warning the user that there is the possibility of one or more errors, specifying the type of the error and easing the data correction.

In this version the user can check files only in .csv format (separator “;”).

The check is performed simultaneously on the files that can contain also data of more than one year. For this reason the files should be stored together in the same folder.

RoME 1.4 works with R versions $\geq 2.15.2$.

2. Functions contained in RoME

In `ROME` package are contained the functions related to the single checks for a total number of 55 functions (+ 9 “facility functions”) associated with a specific check in TA, TB, TC, TD, TT, TE and TL tables. All the checks have been designed according to INSTRUCTION MANUAL VERSION 5 MEDITS 2007 specifications for the old MEDITS formats and according to INSTRUCTION MANUAL VERSION 9 MEDITS 2017 for the new MEDITS formats.

The list of the functions is reported below:

1. `check_identical_records`
2. `check_quasiidentical_records`
3. `check_consistencyTA_duration`
4. `check_consistencyTA_distance`
5. `check_hauls_TATB`
6. `check_hauls_TBTA`
7. `check_rubincode`
8. `check_species_TBTC`
9. `check_haul_species_TCTB`
10. `check_raising`
11. `check_length`
12. `check_weight`
13. `check_mat_stages`
14. `check_nbtotTB`
15. `check_length_class_codeTC`
16. `check_nb_per_sexTC`
17. `check_distance`
18. `check_position`
19. `check_dictionary`
20. `check_bridles_length`
21. `graphs_TA`

22. check_position_in_Med
23. check_area_year
24. check_depth
25. check_stratum
26. check_quadrant
27. check_step_length_distr
28. check_unique_valid_haul
29. check_weight_tot_nb
30. check_spawning_period
31. check_sex_inversion
32. check_smallest_mature
33. check_no_empty_fields
34. check_dm
35. check_0_fieldsTA
36. check_temperature
37. check_stratum_code
38. check_date_haul
39. check_TE_TC
40. check_individual_weightTE
41. check_individual_weightTC
42. check_nm_TB
43. check_nb_TE
44. check_G1_G2
45. scheme_individual_data
46. check_subsampling
47. check_hauls_TATL
48. check_hauls_TLTA
49. check_associations_category_TL
50. check_0_nbTL

Functions for R-sufi file creation:

51. create_strata
52. create_length
53. create_haul
54. create_catch
55. RSufi_files

2.1 Check on TA file

The checks specific for TA already present in RoME 1.2 are summarized in the table below:

CHECK	Function
The WING_OPENING and VERTICAL_OPENING fields have to be expressed in dm	check_dm
WING_OPENING, WARP_DIAMETER and	check_0_fieldsTA

CHECK	Function
VERTICAL_OPENING ¹ fields have to be not equal to 0	
The DURATION, SHOOTING_TIME and HAULING_TIME fields have to be consistent	check_consistencyTA_duration
The DURATION and DISTANCE fields have to be consistent	check_consistencyTA_distance*
The distance has to be consistent with the coordinates at the start and at the end of the haul	check_distance*
The hauls have to be on reasonable positions	check_position**
Check of the dictionary of specific fields (e.g. validity can be only V or I)	check_dictionary
All the fields, except to HYDROLOGICAL_STATION and OBSERVATIONS, must be not empty for valid hauls	check_no_empty_fields
The field BRIDLES_LENGTH can assume value 100 between 10-200 m of depth or 150 between 200-800 m	check_bridles_length*
The difference between start depth and end depth should be not greater than 20%	check_depth*
Start depth and end depth of each haul should be in the same stratum	check_stratum*
The start and end coordinates of each haul must be in the Mediterranean Sea	check_position_in_Med
Among hauls with the same code, only one must be valid	check_unique_valid_haul
The shooting quadrant and the hauling quadrant should be the same	check_quadrant*
Check consistency between shooting depth and warp opening and between warp length and wing opening	graphs_TA**
There must not be duplicated	check_identical_records

¹ If vertical opening and warp diameter fields equal 0, only a warning message is given, because these values are not necessary for the calculation of swept area.

** This check gives a graphical output that has to be checked by the user. Graphs are automatically saved in *RoME/Graphs* directory.

* This check gives only a warning message. The check procedure is not stopped.

CHECK	Function
records	
There must not be quasi-identical records	check_quasiidentical_records

Some details about check functions are listed below:

- `check_dm` : the interval used for wing and vertical opening expressed in dm are respectively (50, 250) and (10, 99);
- `check_consistencyTA_distance` : the consistency between distance and duration of the haul is evaluated considering the fact that in a 30 minutes haul the vessel covers a distance of about 1852 m×1.5 nm (nautical miles) and in a 60 minutes haul about 1852 m×3 nm; then, in general, a distance of 1852·m×haul duration/20 is covered; moreover, a tolerance of 15% is applied. A warning message instead of an error message is given in the checks involving a certain tolerance;
- `check_distance` : the comparison between distance in TA and computed distance is performed with a tolerance of 30%. The formula used to compute the distance from the start and end coordinates is:

$$d = \left| 60 \cdot \frac{lat_{end} - lat_{start}}{\cos(N_3)} \right| \cdot 1852$$

where:

$$N_3 = \arctan\left(\frac{\pi \cdot (long_{end} - long_{start})}{180 \cdot (\ln(\tan(N_2)) - \ln(\tan(N_1)))}\right)$$

and:

$$N_1 = \frac{\left(\frac{lat_{start}}{2} + 45\right) \cdot \pi}{180}$$

$$N_2 = \frac{\left(\frac{lat_{end}}{2} + 45\right) \cdot \pi}{180}$$

All the coordinates involved in the formulas above are in decimal degrees; the routine converts automatically the coordinates listed in TA. For the hauls with a computed distance quite different from the distance recorded in TA a plot is produced and stored in Graphs directory to allow an easier correction.

- `check_position`: the check on the correctness of hauls position is a qualitative check, made by means of 3 graphs: the first one contains the start and end position of the hauls, the second plot contains only the start points with the hauls code labels and the third one, in a similar way, the end points with the respective labels;
- `check_quasiidentical_records`: two or more “quasi-identical records” occurred when all the fields are respectively equal, except:
 - TYPE_OF_FILE, AREA, GEAR, VESSEL, YEAR, RIGGING, DOORS for TA table;
 - TYPE_OF_FILE, AREA, VESSEL, YEAR for TB and TC tables.
- `graphs_TA`: the consistency between shooting depth and warp length and between warp length and wing opening is a qualitative control (scatter plots stored automatically in Graphs directory) allowing to detect the occurrence of outliers;

- `check_bridles_length`: if bridles length equals 150 m for hauls with mean depth >500m RoME gives only a warning message and does not stop, according to instruction in the MEDITS manual.
- `check_stratum`: the function gives also a warning message, because sometimes it could be possible that a haul starts in a stratum but ends in another;
- `check_quadrant`: the function gives a warning.

In RoME >=1.3 two additional checks on TA table (new format, see INSTRUCTION MANUAL VERSION 7 MEDITS 2013) have been implemented, in order to perform a control also on temperature and stratum code:

CHECK	Function
Correctness of stratum code	<code>check_stratum_code.r</code> *
Consistency depth versus temperature	<code>check_temperature.r</code> **

- `check_stratum_code.r`: this function checks if the stratum code associated to each haul is consistent with the code reported in MEDITS manual and in the table Stratification scheme. **ATTENTION: this check is applied to TA if the format “after_2012” is selected, while is applied to TT (if this file is provided and stored in RoME working directory) if the format “before 2012” is selected.**
- `check_temperature.r`: this function checks if the temperature by haul is in the range]10,30[°C; moreover, a plot depth versus temperature is produced and stored in the Graph folder. **ATTENTION: this check is applied to TA if the format “after_2012” is selected, while is applied to TD if the format “before 2012” is selected.**

In RoME 1.4 the format of TA table was updated, including also information of salinity and including additional checks on allowed values in the fields SHOOTING_DEPTH (now even 0 is allowed), HAULING_DEPTH(now even 0 is allowed), DISTANCE, WING_OPENING, WARP_LENGTH, WARP_DIAMETER, BOTTOM_TEMPERATURE_BEGINNING (now even NA is allowed), BOTTOM_TEMPERATURE_END(now even NA is allowed),, BOTTOM_SALINITY_BEGINNING (even NA is allowed), and BOTTOM_SALINITY_END(even NA is allowed).

The strata codes was updated according to the INSTRUCTION MANUAL VERSION 9 MEDITS 2017).

2.2 Check on TB file

The checks specific for TB, already present in RoME 1.2 are summarized in the table below:

CHECK	Function
Correctness of species MEDITS code and faunistic category according to reference list in Tables directory	<code>check_rubincode</code> *
$NB_TOT=Nb_F+Nb_M+Nb_U$	<code>check_nbtotTB</code> *

** This check gives a graphical output that has to be checked by the user. Graphs are automatically saved in *RoME/Graphs* directory.

* This check gives only a warning message. The check procedure is not stopped.

CHECK	Function
The total weight and total number in the haul have to be consistent	check_weight*
If total weight is different from 0, total number must be different from 0 (only if the category of the species is different from “E”) and vice versa (for all faunistic categories)	check_weight_tot_nb*
Check of the dictionary of specific fields	check_dictionary
All the fields must be not empty	check_no_empty_fields
There must not be duplicated records	check_identical_records
There must not be quasi-identical records	check_quasiidentical_records

Some details about check functions are listed below:

- `check_rubincode` : in case the format “*before_2012*” has been selected, for the check on species MEDITS code was used the FM list corresponding to Annexe XV INSTRUCTION MANUAL VERSION 5 MEDITS 2007 (contained in *Tables* directory and names *FM list_old*); in case the selected format was “*after_2012*”, the TM list contained in the INSTRUCTION MANUAL VERSION 7 MEDITS 2013 is taken as reference to check the correctness of species code and category.
- `check_weight` : the check of the consistency of total weight and total number in TB is performed after computing the mean weight for each species in each haul; a list of the weight intervals is provided in *Tables* directory (*Species_LEN_WEIGHT.csv*); this list was filled in on the basis of the data collected in GSA 10 and 18 from 1994 to 2014. Of course, the user should update and complete as much as possible the table in order to allow RoME to detect errors on the basis of the updated weight intervals. This check is divided in a quantitative control (using the ranges mentioned above) and a qualitative control (using plots of the mean weight). The function gives a warning message. A maximum of 20 graphs will have been displayed in the R console and a maximum of 20 .tif files (6 plots per file) will have been stored in *Graphs* directory. The species with at least an occurrence of 10 hauls are plotted and saved.
- `check_weight_tot_nb`: in this check RoME verifies for the format “*after_2012*”, if for the records with total weight not null, there is a total number not null, except for categories V, G, H, D and E, as reported in MEDITS manual.

An additional check has been introduced, only for the format “*after_2012*”:

CHECK	Function
Check of consistency in number per sex set “not mandatory” in TB	check_nm_TB

- `check_nm_TB`: for the species G1 are not allowed that the fields related to total number, number of females, number of males and number of undetermined are simultaneously null, according to MEDITS manual version 7 of 2013.

2.3 Check on TC file

The checks specific for TC, already present in RoME 1.2, are summarized in the table below:

CHECK	Function
Correctness of	check_length_class_codeTC

CHECK	Function
LENGTH_CLASSES_CODE	
Consistency of LENGTH_CLASS	check_length*
Consistency between sum of NB_LON and NB_SEX	check_nb_per_sexTC
Consistency of maturity stages, according to the faunistic category, sex and species	check_mat_stages*
Correctness only of <u>species MEDITS code</u>	check_rubincode*
Check of the dictionary of specific fields	check_dictionary
All the fields must be not empty (different from NA)	check_no_empty_fields
Fishes and cephalopods length classes must have full or half step (in case of LENGTH_CLASSES_CODE=1 only full). All the measures , must be integer numbers.	check_step_length_distr
Check consistency of size of mature individuals compared with the size of smallest mature individual reported in bibliography	check_smallest_mature
Check consistency of maturity stages using information about spawning period, L ₅₀ and size of smallest mature individual collected from literature.	check_spawning_period*
The user will be informed if information about sex-inversion size for hermaphrodite species (at the moment only for <i>Pagellus spp.</i> and <i>Spicara spp.</i>) is stored in Maturity parameters.	check_sex_inversion*
There must not be duplicated records	check_identical_records
There must not be quasi-identical records	check_quasiidentical_records

- `check_length`: For the check of the lengths, intervals based on the data collected in GSA 10 and 18 from 1994 to 2014 were used. The ranges mentioned above can be changed directly changing the values in `Species_LEN_WEIGHT.csv` table.
- `check_mat_stages`: for the format “before_2012”, the check on maturity stages is done on the species listed in the table `Species_LEN_WEIGHT.csv` (Tables directory) with faunistic category field not empty. As the check is done not for all species, if the user wants to check the maturity stages for one or more species with empty faunistic category field, he needs to fill in the field and the check will be done for that species as well. The faunistic category codes are the same of those contained in the old FM list except for the selachians that are classified with the S code and not with A (fish) code. While, for the format “after_2012”, the check on maturity stage is performed for all the species of the new TM list, where selachians and bony fish are distinguished.

This function also gives a warning message, because it is difficult to define for all GSAs the year of passing from the “old” MEDITS maturity scale to the current MEDITS scale. However,

* This check gives only a warning message. The check procedure is not stopped.

2007 was chosen as a reference year in order to give the warning message to the user. Up to 2006 no warning message is reported if the “old” maturity scale is used in TC files. Moreover, if the year is 2012 or after 2012 the new maturity scales is used as reference (see INSTRUCTION MANUAL VERSION 7 MEDITS 2013). The user can verify the maturity scales used by RoME in the file MATURITY_STAGES_from_2012.csv and MATURITY_STAGES.csv, contained in Tables folder.

- `check_nb_per_sexTC`: If the field number per sex is found completely empty, the routine will stop and will produce automatically a .csv file (`TC_file_with_computed_nb_per_sex.csv`) with the nb per sex column filled in. The user will have to copy and paste the column in the original file and run again the code.
- `check_smallest_mature`, `check_spawning_period` and `check_sex_inversion`: in these checks are involved the information stored in *Maturity_parameters.csv* table, L_{50} , spawning period and sex-inversion size (for hermaphrodite species). For the first one, mature individuals with length smaller than size of the smallest mature individuals reported in literature are detected. In the second check the mature individuals caught outside the spawning are detected; moreover, the mature individuals caught outside the spawning with size smaller than the size of smallest mature individual in bibliography and the immature individuals caught during the reproductive period, but with size greater than maximum $L_{50} + 15\% * L_{50}$ are detected. The last check is performed only for proterogonic and proterandrous target species; the user will be only informed if information about sex-inversion size for hermaphrodite species (at the moment only for *Pagellus spp.* and *Spicara spp.*) is present in *Maturity_parameters.csv* (folder Tables) in order to be used by the user to check the sex data of these species. *Maturity_parameters.csv* should be as much as possible shared and complete.

As for TB, in RoME ≥ 1.3 the new TM list is used in `check_rubincode` function for format “*after_2012*”. For the format “*before_2012*” the old FM list is used (MEDITS handbook 2007).

A new check on the field `WEIGHT_OF_THE_SAMPLE_MEASURED` has been added in order to control the difference between observed and estimated weight of the sample. For this check a new table containing the length-weight relationship coefficients by species is used. This table is stored in Tables folder (*L-W.csv*) and can be modified or updated by the user.

Moreover, another function that checks if for G1 and G2 species the length measurements have been collected has been created.

Finally, the check of faunistic category also in TC has been introduced for the format “*after_2012*” as well a check in case the sub-sample (the `WEIGHT_OF_THE_SAMPLE_MEASURED`) is smaller than the 10% of the total weight in the haul (`WEIGHT_OF_THE_FRACTION`), giving only a warning message.

CHECK	Function
Correctness also of faunistic category according to reference list in Tables directory	<code>check_rubincode*</code>
Check of observed estimated total weight in the haul	<code>check_individual_weightTC.r</code>

* This check gives only a warning message. The check procedure is not stopped.

Check if for G1 and G2 species the length measurements are present in TC	check_G1_G2.r
Check if the sub-sample is smaller than the 10% of the total weight in the haul.	check_subsampling.r

- `check_individual_weightTC.r`: The 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. The file *Comparison_estimated_observed_weight_in_TC.csv* is automatically saved in the working directory in order to easily detect the samples with this differences in total weight. For all the records the percentage difference between observed and estimated weight is reported. This check is based on the table `L-W.csv` contained in Tables folder, where the length-weight relationship coefficients are reported by species, area and sex.

ATTENTION: In order to allow RoME to perform this check, the user has to enter length-weight relationship coefficients for the area.

- `check_G1_G2.r`: this check uses a new support table (in Tables folder) containing the list of MEDITS G1 and G2 species and verify if the length has been collected for the selected species for each haul. If the length is lacking for any species in any haul, a warning message is given in the logfile.

2.4 Check on TE file

From 2012 TE table has been introduced in the exchange formats of MEDITS data, containing the individual data. For this table specific checks have been foreseen:

CHECK	Function
Correctness of species MEDITS code and faunistic category according to reference list in Tables directory	check_rubincode*
Consistency of individual weights (according to length-weight relationship)	check_individual_weightTE.r
All the fields except the last three must be not empty (different from NA)	check_no_empty_fields.r
Consistency of number of individuals sampled for weight and ageing in TE	check_nb_TE
Consistency of maturity stages, according to the faunistic category, sex and species	check_mat_stages

- `check_individual_weightTE.r`: for each individual is calculated the estimated weight according to length-weight relationship coefficient stored in `L-W.csv` (Tables folder) and the difference between observed and estimated individual weight in percentage. If for at least one record this difference is greater than 20% a warning is given in `Logfile.dat` and a table named *TE_with_estimated_weights.csv* is automatically produced in order to allow the user to easily

eventually detect the errors. For all the records the percentage difference between observed and estimated weight is reported.

ATTENTION: In order to allow RoME to perform the check, the user has to enter length-weight relationship coefficients for the area.

Moreover, this function checks if for G1 species has been collected the weight or if has been entered the value ND, that is not allowed.

- `check_nb_TE`: this function verify the consistency of the check-fields:
 - `NO_PER_SEX_MEASURED_IN_SUB_SAMPLE_FOR_OTOLITH`;
 - `NO_PER_SEX_MEASURED_IN_SUB_SAMPLE_FOR_WEIGHT`;
 - `NO_PER_SEX_MEASURED_IN_SUBSAMPLE_FOR_AGEING`.

These fields are compared to the number of records present in TE by sex, length class and haul.

2.5 Check on TD-TT files

For the format “before 2012” the same checks introduced in TA (new format) on temperature and stratum code has been implemented, referring to TT (MEDITS table containing stratum code for each haul) and TD (table containing temperature data for each haul) files.

CHECK	Function
Correctness of stratum code (TT)	<code>check_stratum_code.r*</code>
Consistency depth versus temperature (TD)	<code>check_temperature.r*</code>

- `check_stratum_code.r`: this check is applied to TA if the format “*after_2012*”, while is applied to TT for the format “*before_2012*”.
- `check_temperature.r`: this check is applied to TA if the format “*after_2012*”, while is applied to TD for the format “*before_2012*”.

2.6 Check on TL file

For the data on marine litter a set of ad hoc checks were developed:

CHECK	Function
Check of the dictionary of specific fields (category and sub-category)	<code>check_dictionary</code>
Check correctness of associations between category and sub-category on Litter data	<code>check_associations_category_TL</code>
Check if the number is always filled in on Litter data	<code>check_no_empty_fields</code>
Check identical records	<code>check_identical_records</code>
Check if the number is always not null on Litter data	<code>check_0_nbTL</code>

- `check_stratum_code.r`: this check is applied to TA if the format “*after_2012*”, while is applied to TT for the format “*before_2012*”.
- `check_temperature.r`: this check is applied to TA if the format “*after_2012*”, while is applied to TD for the format “*before_2012*”.

2.7 Cross-checks

The cross-check among TA, TB, TC already present in RoME 1.2 are listed below:

CHECK	Function
All the hauls in TA must be in TB	check_hauls_TATB
All the hauls in TB must be in TA	check_hauls_TBTA
All the target species in TB must be in TC	check_species_TBTC*
All the species in TC must be listed in TB	check_haul_species_TCTB
All the hauls in TC are in TB	check_haul_species_TCTB
In case of sub-sampling in TC, the Total number and the number per sex in TB must be raised correctly	check_raising
TA, TB and TC must have the same year and area	check_area_year

- check_species_TBTC function gives a warning message; in fact, in the case the user does not have the biological data for target species, the routine has not to stop the process of check, because the user cannot correct this kind of missing data.
- check_raising function takes into account also the possibility of a “differentiated” sampling, according to INSTRUCTION MANUAL VERSION 5 MEDITS 2007 (p. 38).
Extracted from INSTRUCTION MANUAL:
“The word "Fraction" means any sub-group of individual from the total catch of a species (males, females, large sized individuals, small individuals, juveniles, etc.) on which it could be proceed to a sub-sample. For example: total weight = 1000 g which is divided into 100g of big individuals and 900 g of small. The big individuals will be entirely measured (PFRAC = 100; PECHAN = 100). The small ones will be sub –sampled with a ratio of 1/10 (PFRAC = 900; PECHAN = 90)”.
In the check check_raising the comparison between the number in TB and the raised number in TC has been not taking into account decimals, but comparing the integer numbers.

In RoME versions ≥ 1.3 new cross checks related to the introduction of the new formats have been introduced:

CHECK	Function
Check if the individuals by species, length, sex and maturity stage reported in TE are less than the number reported in TC	check_TE_TC

* This check gives only a warning message. The check procedure is not stopped.

Check if the date of the haul in TB,TC and TE is consistent with TA	check_date_haul
Summary of the individual data collected by species	scheme_individual_data
Check if the date in TL is consistent with TA	check_date_haul
Check if the hauls in TL are present in TA	check_hauls_TLTA
Check if the hauls in TA are present in TL	check_hauls_TATL

- `check_date_haul.r`: a check on the correctness of the association, on the basis of TA, between haul and date has been added for the format “*after_2012*”, because in the new format of TB DAY and MONTH fields have been added;
- `check_area_year.r`: this check has been extended also to TT, TD (for the format “*before_2012*”) and TE (for the the format “*after_2012*”).
Moreover, the `check_area_year` function has been applied to TE, TT, TD and TL tables.
- `scheme_individual_data.r`: this check has as output a table (automatically saved in RoME directory) named *sampling_individual_measures.csv* where for each species are stored the number of length measurements, individual weights and number of otoliths taken by length class. This table is useful to the user to evaluate the coverage of the individual measurements collections in order to verify if the sampling is in line with the protocol and to understand how eventually improve the sampling procedure.

2.8 Creation of R-Sufi files

Four of the 51 functions of RoME create the files to run R-Sufi : *traits.csv* (based on TA), *captures.csv* (based on TB), *taille.csv* (based in TC for the format “*before 2012*” and based on TE for the format “*after 2012*”) and *strates.csv* (created automatically from Stratification scheme table). Another function creates the global 4 files from an year to another year.

For the file *taille* the change in maturity scale in 2006 has been taken into account: from 1994 to 2005 the males of crustaceans have stage NA, because they were not staged until 2005. From 2006 they are considered mature for the stages strictly greater than 2A as well as for females of crustaceans. Before 2006 the females of crustaceans are considered mature for stages strictly greater than 1.

Bony fish and cephalopods are considered mature from stage 3 until 2005 and then they are considered mature from stage 2B.

For selachians, the immature are always stage 1 and 2.

When the check procedure is completed for a number of years, it is possible to obtain the R-Sufi files from `Year_start` to `Year_end`.

3. Order of the checks

All the checks listed above have to be performed following a specific sequence in order to avoid “cascade errors”:

Sequence of the function calls
1. Check duplicated records on TA, TB, TC, TL
2. Check quasi-identical records on TA, TB and TC

3. Check AREA and YEAR fields are the same for all the files
4. Check dictionary on TA, TB, TC, TD, TT, TE and TL
5. Specific checks on TA
6. Specific checks on TB
7. Specific checks on TC
8. Specific checks on TD
9. Specific checks on TT
10. Specific checks on TE
11. Specific checks on TL
12. Cross checks

This software manages the function calls according to the outcome of the check: if the check is successfully completed the next function is called, otherwise the procedure stops in order to allow the user to correct the files. After correction, the user has to run RoME from the beginning. In this way, the procedure avoids the introduction by the user of errors not found in the previous launch (i.e. after correcting an error in the step 2 of the above sequence, the user could have introduced an identical record).

For each check RoME displays the outcome on the R console and prints it in a log file (*Logfile.dat* in RoME directory) adding the details of the check.

4. Function RoME ()

RoME () is the function that the user has to call from R console to start the checks.

After the run, RoME will ask to set (one by one):

- the name of the files to be checked; all the file names (TA, TB, TC, TT, TD and TE) without extension must be set. In the .csv files the separator must be “;”
- if the data are in the old or new Medits format;
- create_RSufi_files and AREA variables (if not set, the R-Sufi files will not be created) for a specified year;
- create_global_RSufi_files, Year_start and Year_end variables for a set of years.

5. Instructions for the user

The steps for running RoME 1.4 are:

- Open R console(version 2.15.2) and set the working directory (File -> Change dir...)
- Load the packages distributed with RoME;
- Install RoME zip package (Packages -> Install packages from local zip files...)
- Type: `require (RoME)`
- Type in the console `RoME()`; automatically the folder Tables containing the tables used in the checks will be created. Tables directory contains:
 - *TM list.csv (updated to 2017)*,
 - *FM list_old.csv*,
 - *MATURITY_STAGES.csv*,
 - *MATURITY_STAGES_from_2012.csv*,
 - *Species_LEN_WEIGHT.csv*,
 - *Stratification Scheme.csv*,

- *Maturity_parameters.csv*,
 - *L-W.csv*,
 - *MEDITS_G1_G2.csv*,
 - *Associations_cat_TL.csv*.
- Also *Graphs* directory, where the graphical outputs are stored will be created
 - *Logfiles* directory where logfiles will be saved will be also created.
1. RoME ≥ 1.3 has been tested on R version 2.15.2 for 64X in Windows.
 2. Eliminate from TA, TB, TC, TD, TT, TE and TL files the empty columns and rows; If the operation above will not be done, you could have some problems in the reading of the files:
 3. The names of the fields in TA, TB, TC, TD, TT, TE and TL files have to be in English. Specifically, the fields requested to run RoME are in the template formats contained in RoME folder².

ATTENTION:

If you are creating .csv format with Microsoft Excel ® or similar software, pay attention to the precision of numbers. In fact, i.e. if I save the number 1.234567, but in Excel it is displayed as 1.2345, when I create the .csv file, the number saved actually will be 1.2345. Then, be careful that all the digits are displayed before creating the .csv file.

ATTENTION: TA, TB and TC tables names are mandatory in both formats. The user can decide to check also TD and/or TT, selecting “before_2012”, as well as decide to check TA, TB and TC files in format “after_2012” without checking TE and/or TL.

If the user decide to create R-sufi files, he/she has to answer Y in the console, indicate the GSA and the year when asked.

If the user decide to create global R-sufi files from one year to another one, he/she has to answer Y in the console, indicate the start and the end year when asked. In this second case, before answering Y, check that all the RSufi files for all the years checked are stored in files *R-Sufi* directory.

ATTENTION: TA, TB,TC, TD, TT, TE and TL files have to be closed during the running.

Even TA, TB, TC, TD, TT, TE and TL files regarding more than one year could be checked by RoME version $\geq 1.3.1$.

4. During the check procedure a message with the outcome of each check will be displayed; if the procedure will stop, you have to open the *Logfile* saved in the folder Logfiles and named with date and time, correct the errors indicated and run again the code.
5. Even if all the checks will be completed, the *Logfile* will contain the outcome of the checks (positive in this case) and the warnings, that should be read as well.
6. At the end of the check procedure you will find in the Graphs directory all the graphical outputs produced.
7. After the check of a single year, if the user answered Y when asked, R-Sufi files will be created and stored in files R-Sufi directory for that year of data.

² see *template formats before and after 2012* files in the RoME templates directory. For the maturity stages the numeric part has to be put in MATURITY and the alphabetic part in MATSUB field.

After the check and a number of years, if the user answered Y in the console indicating `Year_start` and `Year_end`, the global files for R-Sufi routine, will be created and stored in *files R-Sufi* directory.