Description of data processing scripts created for FLBIEA version of the CelticSea Mixed fishery model

# Introduction

Mixedfisheries models combine multiple sources of data that would normally be fed into individual advisory working groups. As part of the process of creating a mixed fisheries model this data needs to be cleaned and filtered for the data specifically used in the assessment. There is also the need to account for the different data requirements for different working groups, which leads can lead to variations in how the same data can be represented. In principle there are 3 sources of data. Accession’s data which has the landings and value of the catch, Intercatch data that has discard rates and numbers at age and stock objects which come directly from the assessments. The data also need to be crosschecked for consistency as all sources of data should contain the same data and all should be consistent with published advice sheet. A principal difference between FLR and FLBEIA is the inclusion of, and the need to process age data, as such the data processing scripts have been completely revised and streamlined to improve transparency and traceability of data.

# Lookup tables

In an effort to improve script readability a series of lookup tables has been established, these are read in as part of processing and are matched on existing errors or differences in the data. The correct values are then taken from lookup table and applied to the data. The advantage of this is that it removes script-based fixes that can obfuscate the purpose of the code and cluttered the script as additional amendments are added. An additional advantage is that these fixes need to be applied across several scripts and reading them in from a central location reduces the likelihood of coding errors as the scripts are updated yearly. A list of lookups is bullet pointed bellow.

* Area
* Metier level 4
* Vessel lengths
* Quarter
* Country
* Stocks

# data\_01\_clean\_accesions

Data script1 focuses on processing and cleaning Accession’s data. This takes the form of Landings submitted directly to the working group outside of Intercatch. Due to the disparate nature of the data sources corrections need to be applied to the data via the lookup tables. In addition to the lookup tables a small number of specific fixes need to be applied in the scripts. This is because they are very specific in nature and the adding the correction to the lookup table would create a broadbrush approach that may “correct” good data. These corrections need to be applied to both the catch and effort data. The cleaned data (Catch and effort) is written out as an intermediate product, the environment cleaned and the data read back in. The data is then filtered to only the Celtic Sea and  
further adjustments made to the megrim. This is needed as the stock is landed as a combination of two species but can be proportioned out to specific species via a known “split”. Monkfish and angler fisher are also adjusted using a similar ratio. These values are established by the expert working groups responsible for the respective species and stocks. The final step in script is to join the catch data with the effort data. Unmatched data is filtered to separate data frames and written out for checking, Table 1 contains a generalised list of inputs, intermediate outputs and outputs from the script.

Table 1 Inputs and outputs from script 1

|  |  |  |
| --- | --- | --- |
| Input | Intermediate product | Output |
| Catch data | clean\_accessions\_landings.csv | Unmatched\_clean\_accessions\_landings.csv |
| Effort data | clean\_accessions\_effort.csv | Unmatched\_clean\_accessions\_effort.csv |
|  |  | NA\_Catch.csv |
|  |  | NA\_Effort.csv |
|  |  | Matched\_clean\_accessions\_landings.csv |
|  |  | Matched\_clean\_accessions\_effort.csv |

# Data\_02\_a\_intercatch\_discard\_rates

Discard rates for all stocks are generated from intercatch files submitted to Mixfish. Similar to the accession data the data has been prepared for different working groups in ICES. The same lookup tables applied to the Accession data are also applied to the data here. A table of all inputs, intermediate outputs and outputs can be found bellow Table 2. Additional script-based fixes have also had to be applied due to the data quality. The data is submitted as zipped files containing the “caton data”, this refers to landings, discard and effort data by stock and area as well as information on how the data was used. Five caton files are read in as part of creating the discard rates, 3 stock specific file for Cod, haddock and whiting and two general files that contain the rest of the stocks.   
The landings data should be either submitted or converted to kg, however this assumption needs to be checked as inconsistency have been noted for some stocks. Effort is likewise submitted in kilowatt effort or fishing days; however, a column is included to indicate what unit of effort has been provided. Intercatch data forms the basis of the advice sheets and for the stock modelled in the Celtic Sea mixfish model the values from Intercatch should match those present in the single species advice. We can use the IcesSAG package which allows us to directly draw data used in the standard assessment graphs on the single species advice. This allows us to validate the data we have been given against values present in the published advice. The IC data is split into individual stocks and compared against the SAG (standard assessment graph) data, due to mismatched between the submitted IC files for hake and monkfish, script specific alteration have been applied to these stocks. The values used in the model have been corrected to the values present in the single species advice sheets.

Table 2. Inputs and outputs from script 2a

|  |  |  |
| --- | --- | --- |
| Input | Intermediate product | Output |
| Caton data (distribution) |  | caton\_summary.csv |
| Caton data (no distribution) |  |  |
| Caton data (Cod) |  |  |
| Caton data (Haddock) |  |  |
| Caton data (Whiting) |  |  |
| Standard assessment graph data (ICES singles species advice) |  |  |

# Data\_02\_b\_intercatch\_age

Age data is also extracted from intercatch files submitted to the Mixfish working group. These files are separate from the data used to generate the discard rates but have come from the same source and are formatted the same. They consist of four “canum” files, one general and 3 stock specific, for Cod, Haddock and Whiting. An additional alk.r file has been provided for Hake in the Bay of Biscay.

Monkfish numbers and weights are also provided as separate inputs and lengths are converted to ages via life history parameters and age length key. A full list of inputs, intermediate outputs and outputs can be found in Table 3. The correction and adjustments from the lookup tables applied to the discard rate IC data are also applied here. Some ages also need to be changed to plus groups for some stocks as there may be ages present in the data that exceed the model range. The data is also cross validated against the ICES SAG data and values corrected to the data used in the single species advice as needed. The canum data is also checked against the product of script 2b “Caton summary” to ensure consistency between scripts and as an additional form of safety check for both scripts.

Table 3. Inputs and outputs from script 2b

|  |  |  |
| --- | --- | --- |
| Input | Intermediate product | Output |
| Canum data |  | Canum summary |
| Canum data (Cod) |  |  |
| Canum data (Haddock) |  |  |
| Canum data (Whiting) |  |  |
| Caton summary data (product of 2a) |  |  |
| Monkfish number (landings and discards) |  |  |
| Monkfish weight (landings and discards) |  |  |
| Hake alk data |  |  |
|  |  |  |

# Data\_02\_c\_stock\_objects

Stock objects used in the respective SAM assessments for each stock are loaded into R and converted to an FLBEIA stock objects. The stock objects are taken directly from the assessments and should be the same as those used in the single species advice. A table of all inputs, intermediate outputs and outputs can be found bellow Table 4.

Table 4 Inputs and outputs from script 2c

|  |  |  |
| --- | --- | --- |
| Input | Intermediate product | Output |
| Model.Rdata (model data for all SAM stock, individual files) | Other.discard.rates.csv (Neprhops) | “stock”.RData object (one for each stock) |

# Data\_03\_combine\_accessions\_intercatch

This script combines the outputs of the preceding files, specifically it add the discard rates generated from the intercatch files to the accession data. A unique discard ID is generated from the year, country, species area and lvl4 metier. Successive discard ID’s are also created removing components to create broader aggregations of DR. The data is then joined together via the discard ID’s, mismatched data is filter off and joined to the next level of discard ID. This is repeated until no more matches can be made. A weighted mean is used to when summarising the data so that a mean discard based on the value of the landings is applied to the accessions data. Landings that cannot be matched are assign a 0 value for discards. A fleet is also assigned based on the country, a general grouping and the vessel lengths. A combined Catch and effort data frame is created, this is then filtered by a percentage value of landings in fleet, 0.01. This will remove from a given fleet any metier that contributes less than 1% of the total landings for that fleet and assign it to an “other” fleet. The step is repeated at the metier level to remove landings that contribute less than 1% to a metier to an “other metier. The purpose of this is to remove incidental catches of a stock that may change the fleet or metier composition and give a distorted perception a fleet acting on a stock. A table of all inputs, intermediate outputs and outputs can be found bellow

Table 5. Inputs and outputs from script 3

|  |  |  |
| --- | --- | --- |
| Input | Intermediate product | Output |
| Caton Data | Catch per country file | Catch file for script 4 |
| Canum Data |  | Effort file for script 4 |
| Accession landings |  |  |
| Accession effort |  |  |

# Data\_4\_make\_FLBiols.R

The 4th script takes information from the submitted stock objects and extracts the relevant biological information for each stock, a list of inputs, intermediate products and outputs can be found in Table 6. The stock objects should contain the following information compiled by the stock assessors for each respective stock.

* Numbers at age
* fishing mortality
* weights at age
* landings numbers
* discard numbers
* range(stock) units for f and plus group range(stock) ("minfbar","maxfbar")
* Fish stock objects contain data up to the assessment year -1, however Nephrops stock objects contain data up the assessment year itself.

The conversion from stock object to FLBiol is handled by ICES TAF function sam\_to\_FLStock and QC\_FLStock, which respectively convert the SAM stock object to an FLStock object and run a quality check on the reformatted data. These were created by ICES and represent a standardised way to convert stocks and check for basic problems in the conversion processes. The stocks are processed individually and save as an Rdata object. These objects are also combined into an R list of FLStocks,  
this allows later scripts to access a single object for stock data.

Table 6 Inputs and outputs from script 4

|  |  |  |
| --- | --- | --- |
| Input | Intermediate product | Output |
| SAM Stock objects | FLStock objects | CS\_FLBiols\_FLStocks (R list of all stock object) |

Description of Model scripts created for FLBIEA version of the CelticSea Mixed fishery model

In order run an FLBEA model the clean and formatted data is then processed via a series of scripts to constitute individual fleets and metier that are nested within the overall model. The next set of steps can be broken down into roughly three arcs, although the number of scripts is more than this to allow for compartmentalisation of code and ease of editing. initially the fleets are constructed and then fleet, stock objects and advice (ICES) are conditioned, and covariates added as required. Secondly, we check model validation, run the intermediate year forecast of the stocks and run the chosen Mixfish Scenarios. The final stage is to reproduce the advice sheets from our model and format a report on the scenarios run by the framework.

# Model\_01\_make\_FLeetsExt

This script takes the products of data script 03 to produce an FLfleets object, a table of inputs and outputs can be found in Table 7 . The data can be aggregated at this point to reduce the overall number of fleets. This is done via a percentage threshold of what metiers and what fleets contribute to the total landings of a stock. This can be applied at the fleet or metier level independently (1% and 2% respectively). Landings and effort are not removed they are simply assigned to an “other” fleet that is not explicitly represented in the model. The data is also filtered to the last three years of the time series as this is the relevant data for the projection.  
Landings and discard data also needs to be disaggregated by age to create the appropriate population structure for age assessed stocks. The data is first matched on complete information (Country, metier, area, stock and year), If a match cannot be made successive elements are removed to fill in data gaps. Ideally the majority of data can be matched at the highest resolution, although this is not always the case. The script also checks that the finished fleet object contains the correct information and are internal consistent. For example, ensure that all age slots are fill out for a given age range or that the data is in the correct units. The data is disaggregated by age, for age structured stocks, this is done by matching landings data to intercath age data at the lowest level of aggregation. If a complete match is not found a higher aggregation of data is used to disaggregate the age data. The levels of aggregation are complete match, no area and finally stock area. Non category 1 stocks are filled directly with landings and discard data from intercatch. Once disaggregation and assignment of discards has taken place the sum of product is checked against the known landings and intercatch data. The “other fleet” (that catch that does not meet the 1% and 2% thresholds) is taken from the difference between the stock objects and fleets object created in the script. This should account for any residual catch. Finally, the data is aggregated where possible to condense the overall size of the fleet object created.

Table Inputs and outputs from model script 1

|  |  |  |
| --- | --- | --- |
| Input | Intermediate product | Output |
| Catch data | fleet\_data\_workup.RData | Fleet object |
| Effort data |  | Data\_workup\_fleets\_landings.png |
| Canum data |  | Data\_workup\_fleets\_discards.png |
|  |  | Data\_workup\_fleets\_landings\_at\_age.png |
|  |  | Data\_workup\_fleets\_discards\_at\_age.png |
|  |  | FLFleets.RData |

# Model\_02\_Stock\_Conditioning

The stock conditioning script takes reference year data and the FLBiols generated in the data script 4, a table of inputs and outputs can be found in Table 8. The Biol data is either extended or restricted to the reference year period. Setting for the FLBEIA run are also specified including the growth structure, for the Celtic sea this is either ASPG (age structured population growth) or fixed growth. The observation model and assessment model are also set, perfectObs and NoAssessment. The Biols data also allows for the creation of the Stock recruitment relationships. All the control objects are created as separate arguments to the FLBEIA function. This allows for individual options to be tweaked quickly.

Table Inputs and outputs from model script 2

|  |  |  |
| --- | --- | --- |
| Input | Intermediate product | Output |
| year\_references.RData |  | FLBiols.RData |
| FLBiols.RData |  | biols\_ctrl.RData |
|  |  | obs\_ctrl.RData |
|  |  | assess\_ctrl.RData |
|  |  | SRs.RData |

# Model\_03\_fleet\_conditioning

Similar to the stock objects the fleet object need to be conditioned and a set of control objects created, a table of inputs and outputs can be found in Table 9. The fleet conditioning sets effort, capital and price models to fixedEffort, fixedCapital and fixedPrice respectively. The important controls applied here are catch restrictions and the catch model. The current FLBEIA model uses a CobbDouglasAge structured catch model. This has implications on the model outcome that are currently being explored by the mixfish working group. All of these control inputs are combined to calculate fleet catchability, formally this was done using R code; however, this took a large amount of time due to limitation in the R Language. A C++ alternative has been coded specifically for the Celtic Sea model which allows the fleet to be condition more time efficient manner. A saving somewhere in the order of days to minutes depending on the size of the model and number of fleets. The script this calls function via the Rcpp library, which allows R to pass argument to a compiled C++ function.

Table Inputs and outputs from model script 3

|  |  |  |
| --- | --- | --- |
| Input | Intermediate product | Output |
| FLFleetsExt.RData |  | FLFleetsExt.RData |
|  |  | fleets\_ctrl.RData |

# Model\_04\_advice\_conditioning

The advice conditioning set the advice tac and quota share for the model, a table of inputs and outputs can be found in Table 10. The advice control is set to fixedAdvice, while the TAC share is input manually. For Nephrops a simple mean of the reference years is used, while exact numbers are input for the finfish stocks. Quota share are set based on the observer landings split from the reference years. The advice control in this case is simply to use the advice, while the advice object contains the actual tac and quota share.

Table Inputs and outputs from model script 4

|  |  |  |
| --- | --- | --- |
| Input | Intermediate product | Output |
| year\_references.RData |  | advice.RData |
| FLFleets.RData |  | advice\_ctrl.RData |
| FLBiols.RData |  |  |

# Model\_05\_covars

Script 5 is place holder script for the future inclusion of covariates, it does produce a covers object, however this is blank.

# Model\_06\_model\_valiadation

Script 6 is place holder script for the future inclusion of model validation.

# Model\_07\_ intermediate \_year

Script 7 is place holder script for the future inclusion of intermediate years as a separate script. The intermediate year is currently included in the scenario script.

# Model\_08\_MIXFISH\_scenario

This script runs the actual FLBEIA function and the scenarios the Celtic Sea subgroup has chosen to run. It uses the output from the preceding model scripts, a table of inputs and outputs can be found in Table 11. The FLBEIA function is run twice, once to produce the fishing, advice and stock history and then to run the actual scenarios using this information. FLBEIA is run in parrel when running scenarios in order to speed up the process.

Table Inputs and outputs from model script 8

|  |  |  |
| --- | --- | --- |
| Input | Intermediate product | Output |
| advice |  | Intermediate\_year\_diag.png |
| advice\_ctrl |  | ScenarioResults |
| assess\_ctrl |  |  |
| biols\_ctrl |  |  |
| covars |  |  |
| FLBiols |  |  |
| fleets\_ctrl |  |  |
| FLFleetsExt |  |  |
| obs\_ctrl |  |  |
| SRs |  |  |

# Model\_09\_reproduce\_the\_advice

A final step is to take the same inputs to the FLBEIA model and reproduce a facsimile of the ICES advice. This allows for comparison and identification of