

WKBIOPTIM 2 REPORT 2018

ICES ECOSYSTEM OBSERVATION STEERING GROUP

ICES CM 2018/EOSG:23

REF WGFAST, ACOM & SCICOM

Report of the Workshop on Optimization of Biological Sampling (WKBIOPTIM 2)

29-31 May 2018

Nantes, France



International Council for
the Exploration of the Sea

Conseil International pour
l'Exploration de la Mer

International Council for the Exploration of the Sea Conseil International pour l'Exploration de la Mer

H. C. Andersens Boulevard 44–46

DK-1553 Copenhagen V

Denmark

Telephone (+45) 33 38 67 00

Telefax (+45) 33 93 42 15

www.ices.dk

info@ices.dk

Recommended format for purposes of citation:

ICES. 2019. Report of the Workshop on Optimization of Biological Sampling (WKBIOP-TIM 2). WKBIOP-TIM 2 Report 2018 29–31 May 2018. Nantes, France. ICES CM 2018/EOSG:23. 172 pp.

The material in this report may be reused using the recommended citation. ICES may only grant usage rights of information, data, images, graphs, etc. of which it has ownership. For other third-party material cited in this report, you must contact the original copyright holder for permission. For citation of datasets or use of data to be included in other databases, please refer to the latest ICES data policy on the ICES website. All extracts must be acknowledged. For other reproduction requests please contact the General Secretary.

This document is the product of an Expert Group under the auspices of the International Council for the Exploration of the Sea and does not necessarily represent the view of the Council.

Contents

Executive summary	1
1 Introduction	2
1.1 Terms of Reference	2
1.2 WKBIOPTIM 2 participants and agenda.....	2
1.3 Background to WKBIOPTIM 2 and report outline	2
2 Data and script preparation	4
2.1 Analyses on the number of individuals collected from biological samples.....	5
2.1.1 Rationale of the analyses	5
2.1.2 Developments during WKBIOPTIM 2	5
2.1.3 Analyses of simulation results.....	7
2.2 Multi-level analysis of sampling effort needed to produce estimates of catch numbers for stock assessment	7
2.2.1 Rationale of the analyses	7
2.3 Sampling Design Tool v.2.....	8
2.3.1 Rationale of the analyses	8
2.3.2 Development of the scripts	9
3 Quality Indicators	10
3.1 Which sample size is adequate to estimate a length structure?.....	11
4 Sample level analysis.....	14
4.1 Baltic Herring SD25	14
4.1.1 Introduction	14
4.1.2 Work developed	14
4.1.3 Conclusions and follow-up	31
4.1.4 References.....	31
4.2 Blue whiting in the ICES Division 27.9.a.....	31
4.2.1 Blue whiting sampling data.....	31
4.2.2 Descriptive analysis	32
4.2.3 Simulation results.....	34
4.2.4 Discussion.....	39
4.2.5 Next steps	39
4.3 Beam trawl survey data 2014–2017 for sole and plaice - Optimizing the Belgian BTS survey with respect to sample sizes for ages.....	40
4.3.1 History of the BTS survey	40
4.4 International Baltic Acoustic Survey (IBAS) (Finland 2015)	41
4.4.1 Descriptive analysis	41
4.4.2 Simulation on the entire dataset.....	43
4.4.3 Smoothing of the data.....	46

4.4.4	Data simulation for optimizing sampling effort	46
4.5	Data Collection Framework (DCF) - Greece 2014–2016	48
4.5.1	Descriptive analysis	48
4.5.2	Simulation on the entire dataset.....	49
4.5.3	Smoothing of the data.....	52
5	Multi-level analysis.....	53
5.1	French data - ICES 27.7.d.....	53
5.1.1	Sampling design	53
5.1.2	Data	53
5.1.3	Stratification.....	56
5.1.4	Length structure	56
5.1.5	Quality of the estimates.....	57
5.1.6	Simulation	58
5.1.7	Results.....	58
5.1.8	Conclusion.....	62
5.1.9	References.....	63
5.2	Swedish data (2016).....	63
5.2.1	Sampling design	63
5.2.2	Data	63
5.2.3	Exploratory data analysis.....	63
5.2.4	Need for optimization (what do you want to test)	69
5.2.5	Scenarios.....	69
5.2.6	Next steps	74
5.3	Italian data in South Adriatic Sea (GSA 18 and GSA 19) for deep-water pink shrimp and blue and red shrimp.....	75
5.3.1	Fisheries description	75
5.3.2	Dataset available.....	75
5.3.3	Data investigation	76
5.3.4	Analyses on sampling optimization in terms of trips	89
5.3.5	Analyses on sampling optimization in terms of number of individuals to be measured.....	94
5.3.6	Possible scenarios varying number of trips and measured individuals.....	101
5.3.7	Summary of results	108
5.4	Greek data for Red Mullet in the Aegean Sea (GSA 22).....	108
5.4.1	Sampling design	108
5.4.2	Exploratory data analysis.....	110
5.4.3	Need for optimization.....	111
5.4.4	Next steps	122
6	ToR c - Discuss progress achieved in implementation at national level since WKBIOPTIM 1.....	122
6.1	Sampling optimization of Blue whiting - ICES 27.9.a (Patrícia Gonçalves – IPMA, Portugal).....	122

6.2	How to determine sample size for biological parameters (Esha Mohamed - SLU Aqua, Sweden)	123
6.3	Estimating uncertainty of North Sea IBTS Indices (Natoya O.A.S. Jourdain - IMR, Norway).....	123
6.3.1	ALK Estimators	123
6.3.2	The North Sea cod data	126
6.4	Optimization of Reference Sampling of <i>Pandalus borealis</i> (Nuno Prista and Annelie Hilvarsson - SLU Aqua, Sweden)	129
7	Conclusions and future work	130
8	References	132
	Annex 1: List of participants	133
	Annex 2: Agenda.....	134
	Annex 3: List of oral presentations	136
	Annex 4: Scripts and code	157
	Annex 5: Proposal: The third Workshop on Optimization of Biological Sampling (WKBIOPTIM 3).....	165

Executive summary

The Workshop on Optimization of Biological Sampling (WKBIOPTIM 2), chaired by Ana Cláudia Fernandes (Portugal) and Maria Teresa Facchini (Italy) was held in Nantes, France, 29–31 May 2018. Fourteen participants from eight countries within the ICES and Mediterranean communities were represented.

This second workshop continued to focus on the practical aspects of optimization of sampling and on the development and improvement of the R-scripts presented in WKBIOPTIM 1. With regards to the sample level analyses, the scripts were reorganized and extended to multiple biological parameters (e.g., age, sex, maturity), for the multi-level analysis, the work was extended to integrate space, time, gear and species in the scenarios and analyses in the context of a concurrent sampling framework. In additionally, for this workshop, two other sets of R-scripts were presented and made available for participants to use and test in their own case studies: one related to simulating for ages (number of otoliths selected by length class, with different sampling stratification options: sex, period (month, quarter,...), port, metier at sample level optimization and the other related to developments and updates of the Sampling Design Tool developed in MARE/2014/19 Med & BS project (Deliverable 2.5) (SD Tool v.2) for the multi-level sampling. Concerning the data format, the exchange format of the Regional Database (RDB format) continued to be used as the input data but the intention is to have also the possibility of using e.g the DATRAS format when using survey data in these analyses. In this workshop a function to convert that data in the RDB CA table format was prepared and made available for participants to test.

In this workshop, the improvements on the scripts and analyses developed last year were presented, tested and discussed. Participants brought their own case studies along with other suggestions to improve the optimization of sampling, and were separated into two subgroups to work and test the sample level and the multilevel procedures. Work developed in this area of improving the biological sampling at national level both for commercial and survey sampling was also presented and discussed during the workshop. Some code adaptations and work on case studies were only finalised after the workshop.

WKBIOPTIM 2 agreed that the compilation of the scripts and procedures being developed, improved and tested during these workshops should be compiled and documented in a Toolbox (e.g. R-Package) so national institutes can analyse their own data and improve their resources allocation and/or distribution. The group think that the main part of the procedures has already been tested in several case-studies and it can be adapted from now on to include more suggested improvements. Along with this, a guide for adequate use of sampling optimization procedures should also be prepared since there are some important rules to take into account.

1 Introduction

1.1 Terms of Reference

The **second Workshop on Optimization of Biological Sampling (WKBIOPTIM 2)** chaired by Ana Cláudia Fernandes (Portugal) and Maria Teresa Facchini (Italy) met in IFREMER Nantes, 29–31 May 2018 to:

- a) Further develop catch-sampling evaluation toolbox (following WKBIOPTIM 1): Improvements will be considered based on additional case studies (i.e. stocks or fisheries), consideration of additional metrics (e.g age and maturity) and considerations for methods to calculate effective sample size for these metrics
- b) Development of quality indicators: evaluate a second set of quality indicators
- c) Discuss progress achieved in implementation at national level since WKBIOPTIM 1

1.2 WKBIOPTIM 2 participants and agenda

The list of participants and the agenda for the workshop can be found in Annex 1 and Annex 2 respectively.

1.3 Background to WKBIOPTIM 2 and report outline

The work developed by WKBIOPTIM 1 aimed to look at how resources can be optimized without compromising the data quality in respect to biological parameters. It consisted in developing tools to overcome some of the issues highlighted by several ICES EGs (e.g. PGCCDBS 2012, PGDATA 2015, WKCOSTBEN 2016). Those issues are related to oversampling in lower stages of national sampling programs (e.g. number of trips, hauls within trips, fish within hauls) and inefficient sampling effort distribution that may not provide additional information on the sample itself nor on the population. Also, with the actual EU-MAUP's, a multipurpose sampling scheme, more information is starting to be requested (e.g. variables for multispecies, unassessed stocks, recording and sampling for PETs) and so there is an increasing need for efficient resource allocation also in terms of costs and time. For example, numbers of length and age measurements and maturity may need to be reduced because the costs and time spent by national laboratories in these tasks is needed for other analyses. Some of the national labs are already developing statistical tools with the aim of analyzing and optimize biological sample sizes by reducing on some clear-cut cases of excessive sampling at sample-level or increase when information collected is not sufficient. Regarding all this, WGCATCH 2016 and PGDATA 2016 proposed the development of R-scripts based on the widely available RDB format as the main scope for WKBIOPTIM 1 so national labs can make a more effective use of the code developed to quantify the effects of different sampling intensities and sampling designs, and support their discussions on the advantages and disadvantages of different sampling strategies in terms of time and costs savings. The R-scripts prepared in WKBIOPTIM 1 were tested, discussed and used by some institutes to analyze the improvements in sampling for some oversampled species. The work performed during WKBIOPTIM 1 was presented in STECF Data Quality, WGBIOP, WGCATCH and in the RCGs and it was seen important concerning the development, discussion and implementation of common tools for optimizing sampling for biological parameters.

In WKBIOPTIM 2, further development of those scripts was performed extending them to more biological parameters at sample level analysis and integrating space, time, gear

and species in the scenarios used for multilevel sampling analysis. The main part of this work was done prior the workshop and then it was presented, discussed and some suggestions of improvements were given by participants. The code was tested and applied in a set of case studies. Finalization of this work was only completed after the workshop. New approaches for the optimization of sampling were also presented during the workshop along with some examples showing the implementation of optimization tools at national level. In what concerns to surveys, a function for incorporating information on length stratification into the CA table when using surveys data was presented and used in some case studies. It is expected that some of the tools developed in WKBIOPTIM 1 and WKBIOPTIM 2 can also be used for survey data. More information on possible quality indicators and optimization procedures for the collection of biological parameters already being used for surveys may be needed along with the input of stock assessors using this data.

2 Data and script preparation

Preparatory work for WKBIOPTIM 2 can be broadly divided into a) improvement of R-scripts developed for sample-level and multi-level analysis by WKBIOPTIM 1, b) development and preparation of new different approaches/procedures for optimization, and c) analyse the possibility of integrating or not this new approaches in the R-scripts that already exist. This work was steered by the chairs of the workshop that coordinated with script developers in several skype meetings prior and after the workshop.

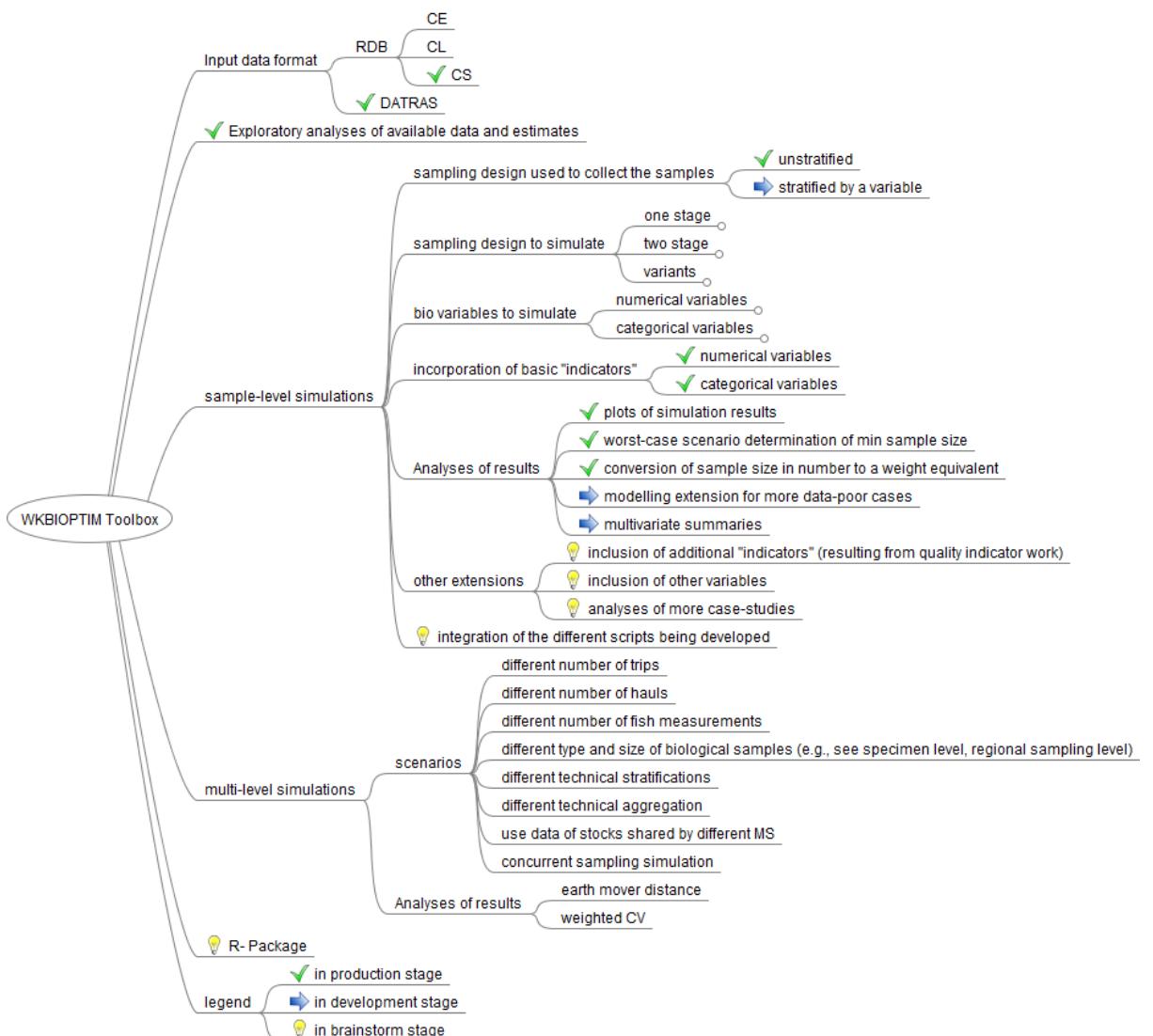


Figure 2.1 Improvement of WKBIOPTIM Toolbox.

2.1 Analyses on the number of individuals collected from biological samples

2.1.1 Rationale of the analyses

See WKBIOPTIM 1 report, Section 2.1.1.

2.1.2 Developments during WKBIOPTIM 2

2.1.2.1 Overview of Sample-level Algorithm

A simplified overview of the scripts developed in WKBIOPTIM 1 and WKBIOPTIM 2 is provided in Figure 2.1.2.1.

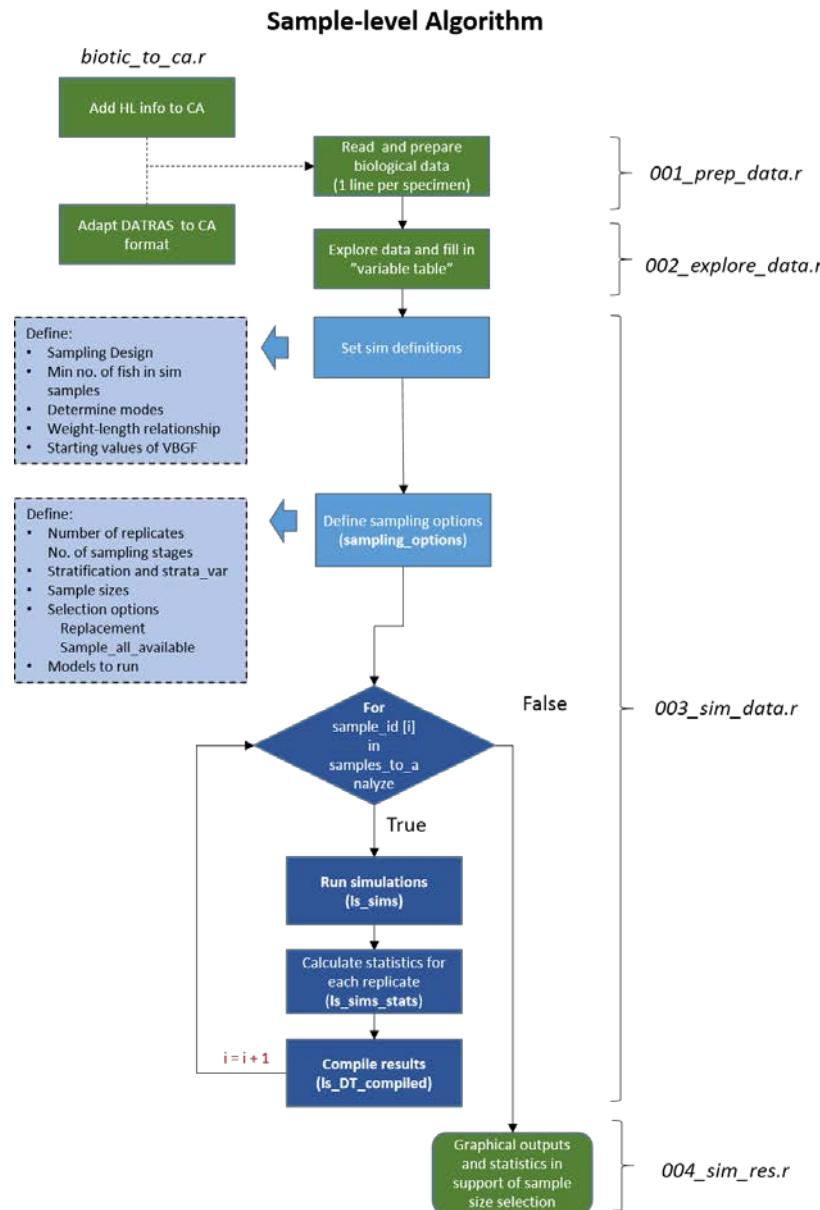


Figure 2.1.2.1 Flowchart of the script used to analyse the number of individuals that should be collected from each biological sample.

2.1.2.2 Input data

Function for incorporating information on length stratification into the CA data

CA-HLdataFunction

Function biotic_to_ca in file takes CA and HL files and produces a CA file with new rows generated according to the numbers at length in the HL files.

The generation of data rows is done by comparing counts of individual rows in CA-file to the values at length in HL files and then generating individual rows. The rows are generated with no biotic data for length classes corresponding to the number of length minus the number of individual row for that specific length class. The function also preserves biotic data for non-generated rows.

In order to operate on different data sets, the function allows specifying the sample key columns, length class columns for CA and HL data and number at length columns for HL data. It is assumed that CA and HL files share the same sample key columns.

Attempt was made to apply the *CA-HLdataFunction* on North Sea International Bottom Trawl Survey data, but the large number at length values made running the script inconvenient. Testing could proceed with a subsample of the NS-IBTS data. There were further plans on applying the script to Beam Trawl Survey data. Data collection framework data sets do not need generated CA records, since those datasets do not have missing individual records.

Script for general data preparation ("001_prep_data.r")

The data preparation part of the script developed during WKBIOPTIM 1 was individualized into a separate, more-autonomous, script ("*001_prep_data.r*"). This script now carries out column renaming (based on a external .csv file) and variable formatting that can be specific for each different project. The script was also streamlined, and its annotations were improved.

2.1.2.3 Exploratory analyses

Script for general exploratory analyses ("002_explore_data.r")

The data exploration part of the script developed during WKBIOPTIM 1 was individualized into a separate, more-autonomous, script ("*002_explore_data.r*"). This allows the formal separation of the exploratory analyses that support the project from the simulations themselves. Further quality checks were added, together with some analyses that help determine the threshold number of individuals per sample required for the samples to enter the simulations (*min_n*). The script was also streamlined, and its annotations were improved.

2.1.2.4 Simulation of sampling under different sampling sizes and strategies

Script for general simulation analyses ("003_sim_data.r")

The data simulation part of the script developed during WKBIOPTIM 1 was individualized into a separate script ("*003_sim_data.r*"). Some significant structural changes and improvements were also made on the code, namely:

- The definition of a specific object ("sampling_design") that specifies the sampling design underlying the data collection, namely the existence (or not) of stratification (e.g., by size)

- The separation of the many functions used in the script into independent r files that can be more adequately documented and version controlled.
- The automatic calculation of weight-length and Von Bertalanffy growth model from available data. These parameters are used to estimate the weight of the simulation replicates and as starting values for non-linear Von Bertalanffy growth modelling of the replicates (where applicable)
- Adaptation of the core simulation code to multivariate biological data
- Inclusion of new statistical tests proposed and discussed in WKBIOPTIM 1 and 2 (e.g. ADV (Acceptable Dissimilarity Value)).
- Inclusion in simulation outputs of Von Bertalanffy growth modelling
- Implementation of a versioning system on the final simulation object.

Additionally, code development continued with regards to the incorporation of stratified sampling and two stage sampling strategies¹. Finally, the script was streamlined, and its annotations improved with a few bugs corrected (e.g., breaking of the code at particularly low sample sizes when NAs were present in the variables).

2.1.3 Analyses of simulation results

The analyses part of the script developed during WKBIOPTIM 1 was individualized into a separate script (“*004_sim_analyses.r*”)². **An extensive set of new analyses was developed during and post-workshop** that included several types of univariate and bivariate plots of individual samples. With regards to sample size determination the worst-case strategy was further developed with outputs on exploratory tables that help to investigate conservative minimum sample sizes that attain pre-specified goals. Finally, a first attempt at sample size determination from data-limited samples was also coded. This involved a whole new set of functions for the fitting, diagnosis and prediction of exponential models of, e.g., MWCV or CV vs. sample size. For more details and examples of these analyses see, e.g., sections 4.1 and 4.2.

2.2 Multi-level analysis of sampling effort needed to produce estimates of catch numbers for stock assessment

2.2.1 Rationale of the analyses

The rationale of the analyses follows the framework presented in the WKBIOPTIM 1 report (ICES, 2018).

Data collected in the EU-MAP framework are aimed first to provide stock information for the assessment working groups. National monitoring program and sampling plan have to be tailored to fit this need. Consequently, optimization of the sampling plan has to follow the same constraint: to provide the estimates requested by the end-users.

In this framework, analyses were carried out to test how sampling plan changes can affect population estimates. The estimator used in this exercise was the length distribution estimate. Compared to age estimates or discards estimates, this estimate is the simplest to provide to assessment working groups. Moreover, measuring length is the archetypal activity in field work related to commercial fisheries and data are widely available in the national databases. Starting from the existing sampling database, simulations were used to modify the sampling plan. The simulation process includes

¹ This part of the code is still being developed and not yet incorporated in the main part of the script

² This script presently also contains explorations into appropriate sample sizes - it is envisioned this part will later be individualized into its own independent script.

changes in the sampling effort by decreasing the amount of samples or measurements and then computing new length distribution estimates using the modified sampling database. The new estimates were compared with the original length distribution to assess the effect of reducing samples on the estimates.

During the first WKBIOPTIM workshop (ICES, 2018), impact on length distribution was assessed at a stock level. During this workshop, impact on length distribution was assessed for a given area for all the main species included in the sampling plan. The main objective was to assess sampling plan optimization in a concurrent sampling framework.

2.3 Sampling Design Tool v.2

2.3.1 Rationale of the analyses

During the workshop the approach developed in the MARE/2014/19 Med & BS project (Deliverable 2.5) implemented in the *Sampling Design tool* for optimization of sampling intensity (based on COST tools) was presented (see Annex 3, Section 3.), as well as some updates made on the routine aiming at the fulfilment of the objectives of the MARE/2016/22 STREAM project. Specifically, the analyses will support the design and proposal of a regional sampling plan (RSP) for 2019 covering commercial fisheries/stocks/métiers (RSP-CF).

The new SD Tool v.2, generalizing the previous SD Tool, allows through bootstrap technique to resample the historical data studying the Coefficient of Variation (CV) in association with the number of primary sampling units (e.g. trips) of a given species.

The statistical principle at the basis of the adopted approach is that the value of CV decreases with the increase of the number of sampling units, defining a curve. In the SD tool the part of the curve where the tangent changes and begins to flatten (i.e. the curvature range) is considered as a suitable trade-off between the precision and the sampling effort, and then the **size sample** (in term of sampling units) corresponding to that part of the curve is proposed as "**optimal" sample size**".

This new version includes options allowing a flexible definition of the sampling scheme and allows to carry out simulations on:

- different technical stratifications introducing options to define the technical strata on the basis of gear (level 4) and/or métier, so grouping strata with similar characteristics;
- different temporal aggregation in order to make flexible the stratification by quarter and/or semester, depending on fisheries and target species specifications;
- data of stocks considered shared among MS in order to get results on the whole area of the stock (not only by GSA)

In addition, the new tool allows to perform **possible scenarios varying number of trips and measured individuals**. Based on the results of the sampling optimization for all the target species (implemented with **SD Tool v.2**) the following **scenarios** can be simulated:

- Simulate different number of trips taken from the “optimal” sampling size ranges (e.g. from 5 to 50 trips) allocating them in the **stratification scheme of the past sampling**.
- Simulate different number of trips taken from the “optimal” sampling size ranges (e.g. from 5 to 50 trips) **collapsing strata by means of the aggregation of métiers**.
- Simulate different number of trips taken from the “optimal” sampling size ranges (e.g. from 5 to 50 trips) **collapsing strata by means of aggregation of quarters**.
- Simulate different number of trips taken from the “optimal” sampling size ranges (e.g. from 5 to 50 trips) allocating the trips according to the past sampling. In addition, for each number of trips the number of **measured individuals is reduced by means of subsampling**. For example: for all the selected scenarios of type 1, the scenarios of type 4, i.e. with reduction of length measures can be performed comparing for example a scenario (A), with 5 trips and all individuals measured, and scenario (B), with 15 trips sampled and reduced number of individuals measured.

The main output of a simulated scenarios for each species included in the case studies consists of: total CV, CV per length class and raised length distribution for the whole population. To express the performance of each scenario the values obtained for each species of two indicators will be compared to the respective values of each species in the baseline: **Coefficient of Variation** (CV) and **Earth Mover Distance** (EMD). A summary of changes to the past is done in order to have a complete overview by species and by scenario.

2.3.2 Development of the scripts

The R scripts of SD Tool have been updated in order to read datasets into 2 input data format:

- SDEF format (ICES Standard Data Exchange Format) used in COST project (Jansen *et al*, 2009), i.e. TR, HH, SL, HL and CL tables (see COST - Standard Data Exchange Format .pdf file in /SamplingDesign tool/COST-man folder)
- RCM Med&BS-LP Data Call format: a simplified format identified as common standard format to foster data exchange and dissemination in MARE/2014/19 Med&BS project.

In addition, generalization has been done in order to perform analyses according to different spatial, temporal and technical aggregation allowing the following rules:

- Space: GSA, Country or ALL
- Time: Y (year), Q (quarter) or S (semester)
- Technical: lev4 (level 4 gear), lev6 (level 6 metier) or NONE (overall)

Also the calculation of the Earth Mover’s Distance (EMD), not present in the previous version, has been implemented in order to use it to assess performance of scenarios respect to the baseline.

3 Quality Indicators

Some relevant questions were raised in WKBIOPIM 1 regarding how are the sampled numbers by species defined and how can the sample optimization procedures be applied while ensuring quality of the requested indices. Input from both WGBIOP and WGCATCH on additional quality indicators for length frequency data and for other biological parameters was not clear as this requires a broader overview of the goals to achieve, according to end user need and data collection programmes.

Coming from last year report, a presentation on the ongoing implementation of GFCM quality indicators presented the last progress on this topic. Seven quality indicators were presented: timeliness, completeness, conformity, stability, coherency, accuracy (bias and precision).

Timeliness is the length of time between the data transmission and the deadline as defined by the related datacall. This indicator is behind the scope of the WKBIOPIM work, but it really represents the degree 0 of the quality of the data: the availability to the end-user. **Completeness** is the extent to which the expected data as requested by the datacall is transmitted to the end-user. This indicator underlines the necessity to provide information for the fishery of a country targeting a given stock. Again, this indicator is behind the scope of the WKBIOPIM work, but its implementation by ICES should be straightforward. This indicator needs only to cross-check the official landings with the data provided by stock and area. **Conformity** is the extent to which the transmitted data adheres to the datacall standards (codifications and formats). This indicator is well implemented in the different facilities where the European countries provide their data (namely ICES via Intercatch and Fishframe RDB, and the JRC upload facility for the FDI and new FDI datacall). For these facilities, if the data are not in line with the datacall standards (for example if a metier code is not present in the reference list of ICES or the JRC), then the upload cannot be done, and the data transmitter has to provide correction. The data transmission inside the GFCM do not follow such a strict framework. Data can be transmitted even if they not follow the standard, and errors are reported later. One can ask the question of the impact of submission constraints on data quality during uploading. Knowing that only the provision of data in time is evaluated in the European framework, it is tempting for the data provider to make a quick correction on the data to force the upload without this correction being relevant for the data. These quick and often dirty correction could have a negative impact on the data quality, and are not assessable. Four data quality indicators are related to the consistency of the data: stability, coherency and precision, which includes bias and precision. The consistency is the extent to which submitted data is within a range of plausible values and consistent to previous data submission, as well as to other data sources. **Stability** indicator checks if the data submitted for a given year is in line with the data submitted the previous years. This indicator is again behind the scope of the WKBIOPIM work, but it should be easily implemented inside the ICES or JRC data submission facilities. ICES in Intercatch provides some diagnostic tables regarding this fact, pointing out the differences of landings for example by stratum from one year to another. Sadly, this analysis seems not be reported to end-user or data submitter. For the **coherency**, cross-checks between data are done on the parameters present several times in different data sources. This indicator is then based on the redundancy of information inside a datacall. In the European framework, this could be implemented by cross-checking the data present in Intercatch and in the Fishframe RDB for example, and reported to the end-user. **Bias** and **precision** were addressed in the previous WKBIOPIM report and were implemented in some scripts developed during the

WKBIOPTIM 2. Also some improvements were performed for sample level scripts in order to include some of the statistical analysis suggested in WKBIOPTIM 1 for age and length distribution comparisons, relating to completeness and comparability of the biological data. A work developed on this subject was presented during the workshop (Annex 3 – Topic 7.) and its description, showing the several approaches analysed, is described below (Section 3.1).

In respect to survey sampling, input from what is already being used to assess quality for biological parameters is needed so that survey optimization procedures can be better accommodated in the work developed in WKBIOPTIM. The stock assessors may have an important role in this part of the work.

3.1 Which sample size is adequate to estimate a length structure?

Determination of required sample size providing reliable information about length structure is the critical component of sampling design. A sample size larger than necessary may be wasteful, but can also lead to misleading results. So, the differences in length structure between full and reduced samples can be real, but small in magnitude and biologically irrelevant. Several approaches to the problem of estimation of the adequate sample size were discussed during the workshop.

Let's consider the following data set df1 (CA table data format).

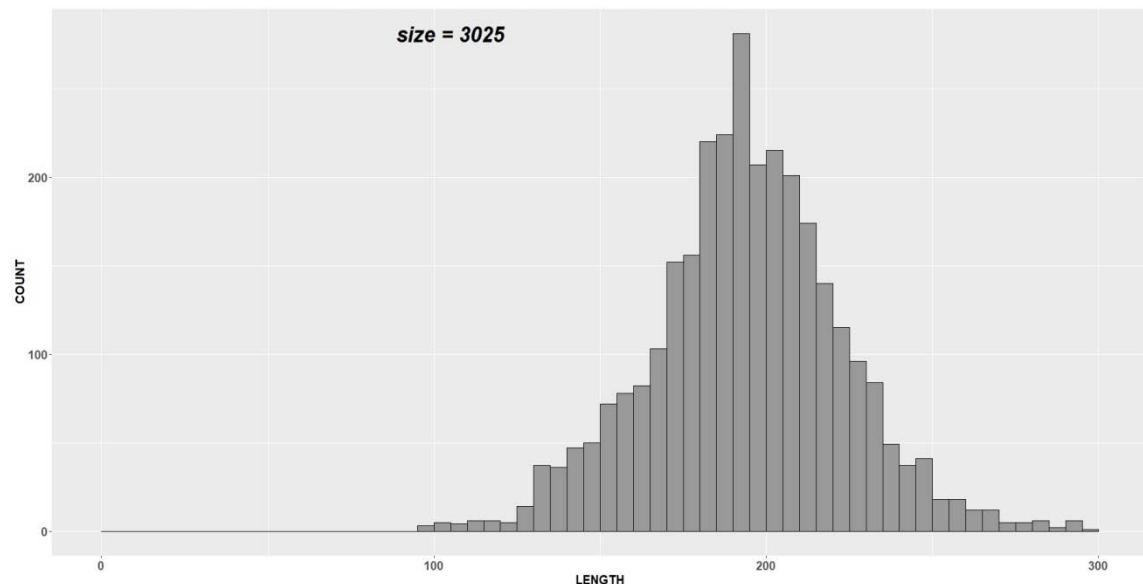


Figure 3.1.1 Original length frequency of *Clupea harengus* in a sample.

Approach 1

This approach is based on application of different tests for distribution (Kolmogorov-Smirnov test, Mooses test etc.) for comparison of original data set and reduced set (subsample). This approach was discussed in framework of WKBIOPTIM 1.

Approach 2

This approach implements bootstrapping and subsampling procedures.

We calculate the bootstrap version of the statistic of interest (e.g. mean length, see R-code approach-2_1.R in Annex 4) and get the 95% bootstrap confidence intervals. Then, applying the subsampling procedure by different subsample sizes, we determine a required sample size via a power analysis approach. So we can vary the sample size until we achieve the desired power. For example, applying the code to our data set:

Power for sample size = 2500 under 10000 replicates: 0.926

Power for sample size = 2000 under 10000 replicates: 0.8953

Power for sample size = 1500 under 10000 replicates: 0.838

Power for sample size = 1000 under 10000 replicates: 0.7419

Power for sample size = 500 under 10000 replicates: 0.5788

So, the power equal to 0.9 will be achieved for the sample size about 2100-2300 individuals.

Comment 1: To preserve a correlation structure in the data, we would strongly recommend the block bootstrap procedure, where each haul/station is supposed to be a block.

Comment 2: Instead of choosing any statistic of interest, we can apply the bootstrapping to the empirical cumulative function (see R-code approach-2_2.R in Annex 4). Bootstrap confidence bands (see Figure 3.1.2) developed from original data set can serve as lower and upper boundaries by subsample selection, i.e. an empirical CDF of the well-selected subsample should be located inside the confidence bands.

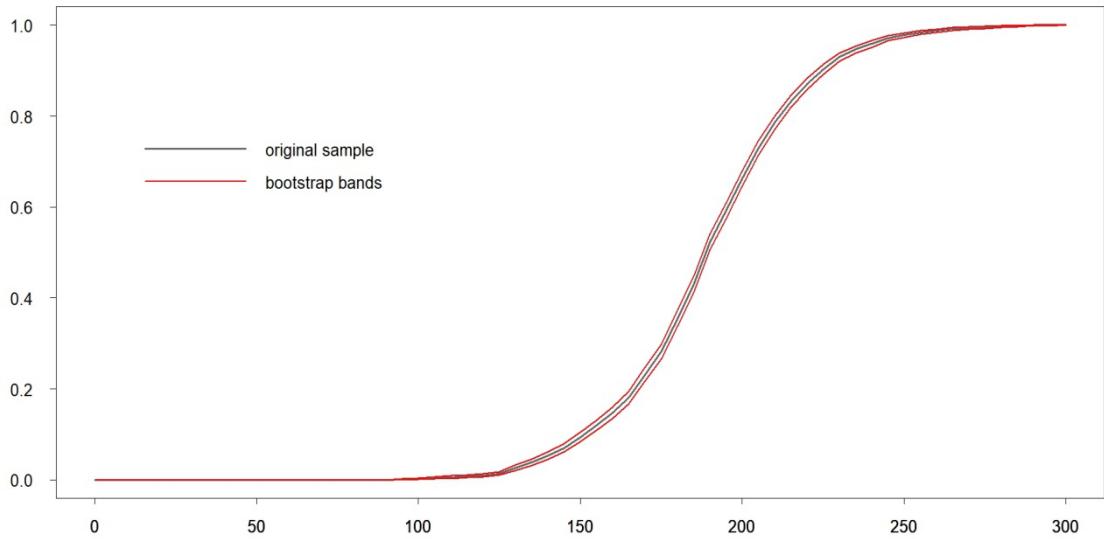


Figure 3.1.2 CFD and confidence bands developed from original dataset of *Clupea harengus*.

Approach 3

To define a length structure of a certain species we have to study a shape of its length distribution (see presentation Julia Wischnewski and Matthias Bernreuther, Thünen institute of sea fisheries, Germany in Annex 3). Typically, distributional shape is a set of modes/bumps and antimodes/dips. The preserving the shape of the length structure seems to be more important than keeping

The algorithm searching the modes and antimodes in a given data set is presented in Annex 4 (modes_and_antimodes.R). Applying this function to our data set:

```
modes_and_antimodes(df1,5)[[1]]: 130 190 200 245 (modes)
modes_and_antimodes(df1,5)[[2]]: 90 135 195 240 295 (antimodes)
```

The subsample will be constructed by condition, that it preserves all extreme points (modes and antimodes) of the original data set as well as keeps almost constant (e.g. 90%) amplitudes between adjacent robust extreme points (see R-code approach_3.R in Annex 4). In other words, the subsampling process continues until the abnormal structures in distributional shape don't arise. The resulting subsample and original data set are presented in the figure below (Figure 3.1.3). The sample size proposed by the algorithm is about 2750 individuals.

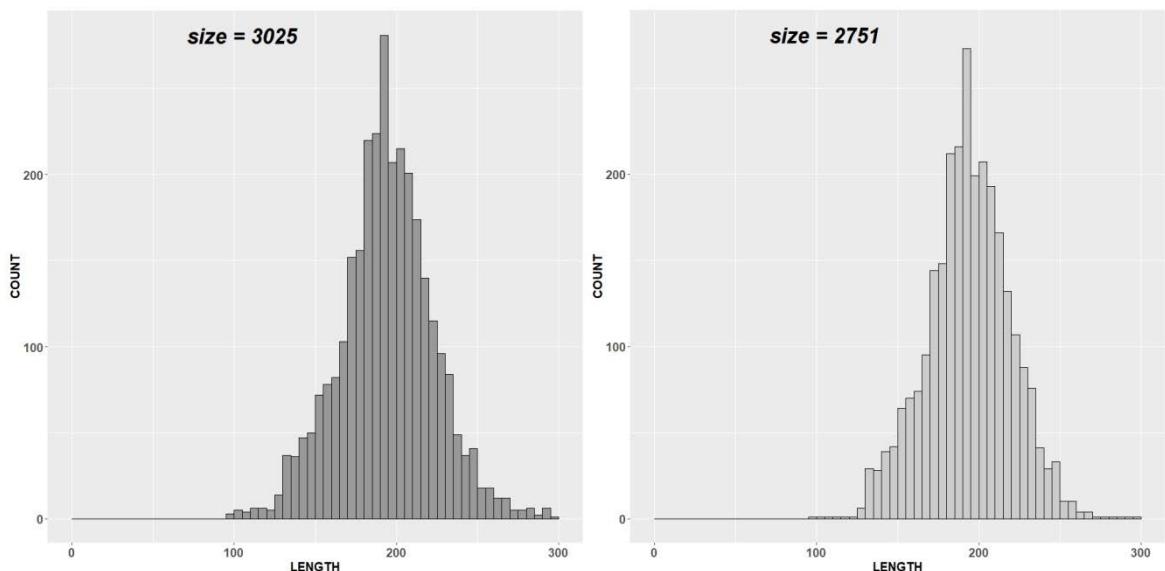


Figure 3.1.3 Length frequency of the original sample and of the subsample with the sample size proposed by the algorithm.

Comment 1: Some modes and antimodes can be just artefacts of insufficient sampling. The approach is expected to be improved by introducing condition relying on “importance” of these values.

Comment 2: The procedure has to be separated for large/small length classes (“shoulders” or “tails” of length frequency histogram) and length classes which seem to be relative oversampled (middle part of length frequency histogram). This can help to save important information about rare length classes which can be poorly represented in subsample, and, at the same time, to reduce a sampling effort.

4 Sample level analysis

The following case studies were developed both during and after the workshop.

- a) Baltic Herring SD25
- b) Blue whiting ICES 27.9.a
- c) Beam trawl survey data 2014–2017 for sole and plaice
- d) Survey analysis using the Finnish stations from IBAS 2015
- e) Greek data DCF 2014 and 2016 (*Merluccius*)

4.1 Baltic Herring SD25

4.1.1 Introduction

Sweden samples its Baltic commercial herring fishery using market sampling. The sampling frame is a hypothetical list of fishing trips landing the species. A stratification by quarter and subdivision is employed with sampling effort being 8 to 10 trips per strata. The fishing trips are selected by first hand buyers that are requested to spread the samples throughout the quarter. The overall goal is to biologically sample 400 individuals per quarter in subdivisions 24–29S and 800 individuals per quarter in subdivisions 29N–31. To meet these goals, a box of fish from each trip is processed at the lab. In general, the entire content of the boxes is measured and biologically sampled, but subsamples of 50–150 fish are also done when samples are particularly large and quarterly*subdivision goals are not at risk.

The staff time costs annually spent processing herring samples for length and age is considerable so there is interest in investigating how effective the present sampling goals are in terms of the quality of length and age frequencies obtained from the samples. The objective of this case study was to illustrate and test the use of the sample-level tools developed under WKBIOPTIM2 in investigating that matter.

4.1.2 Work developed

Data: The data analysed were biological samples of herring from SD 25 collected under the Swedish National Programme (DCF) between 2014 and 2017. A total of 116 samples were included in the dataset, comprising 8454 individuals. Each individual had records for length, age, and a few other biological variables (e.g., sex, maturity stage). For sake of simplicity of this case-study only length and age data were considered.

Data preparation: The data were prepared using script *001_prep_data.r*. The initial format of the data was very similar to the RDB CA format so only minor adaptations to column names were needed for it to enter the WKBIOPTIM scripts.

Data exploration: The data were explored using script *002_explore_data.r*. The number of individuals per sample varied between 20 and 300 fish with a dominance of sample sizes 50, 100 and 25 (Figure 4.1.2.1).

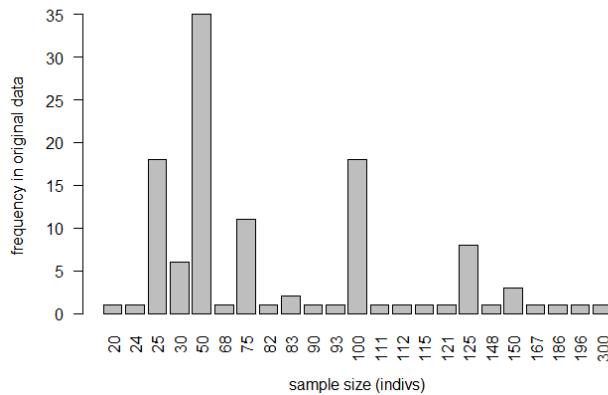


Figure 4.1.2.1 Distribution of sample sizes of the original herring SD25 dataset.

All 8454 fish in the dataset had been measured and nearly all had also been aged. From the $n = 54$ that had no ages assigned, $n = 50$ had been aged but their ages were considered unreliable³ and not registered in the database. General bar plots of biological data (all samples combined) indicate the individuals ranged 95 to 315 mm in length (in 5 mm increments) and 0 to 14 years in age (Figure 4.1.1.2). A few rare classes of particularly small (<150 mm) or young (<2 yr.), and large (>250 mm) or old individuals (>8 yr.) can also be identified (Figure 4.1.2.2).

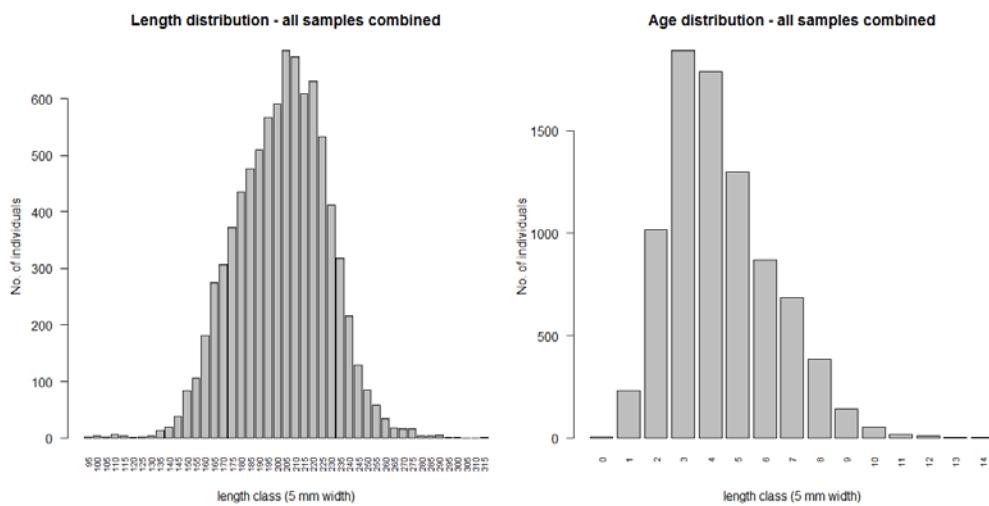


Figure 4.1.2.2 Distribution in length and age of the herring SD25 dataset.

³ A quality indicator ageQuality is also present in the dataset. ageQuality == 4 means the age determination was signalled as unreliable

Preliminary analysis of the MWCV and CV of the mean length and age distributions of the samples indicated a strong correlation between the two statistics (Pearson correlation 0.87 and 0.75, respectively) (Figure 4.1.2.3). At around $n = 100$ individuals the variation of these statistics seem to attenuate with sample size (Figure 4.1.2.4). Consequently, $n = 100$ was chosen as the minimum number of fish required for samples to enter simulations (\min_n). Using $\min_n \geq 100$ criteria, $n = 38$ samples (32%) were selected as “representative-enough” for the simulation study.

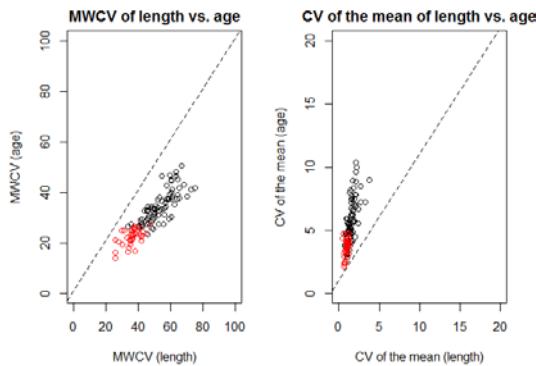


Figure 4.1.2.3 Correlation between MWCV and the CV of the mean age and length in the initial herring SD25. Red dots correspond to samples selected for simulation.

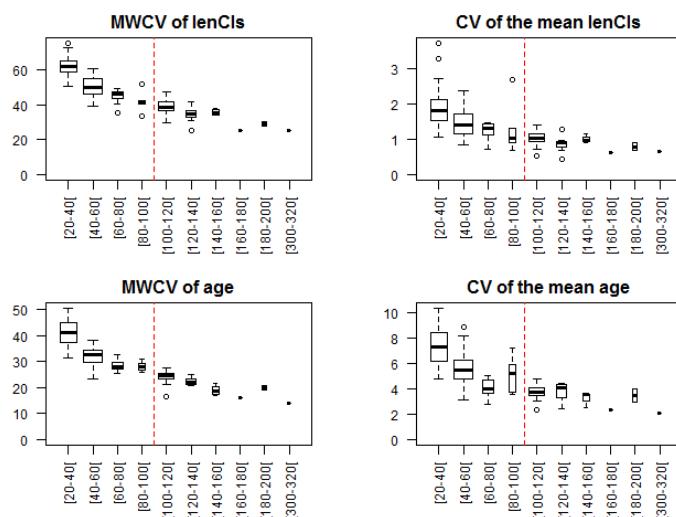


Figure 4.1.2.4 Distribution of MWCV and CV of the mean length and age by sample size. Correlation between MWCV of age and length in the initial herring SD25. Red line indicates the threshold of samples selected for simulation ($n \geq 100$).

These 38 samples included samples from different quarters of the year with size distributions of quite different complexity.

The MWCV of the selected length samples varied between 25.5% (sample “2016_2005” with $n = 300$ individuals in 27 length classes) and 47.2% (sample 2014_2024 with $n = 100$ individuals in 26 length classes) with a mean value of 36.2%. The original MWCV of the age samples varied between 13.9% (sample “2016_2005” with $n = 300$ individuals

in 10 age classes) to 27.4% (sample 2014_2024 with n = 100 individuals in 10 age classes) with a mean value of 22.4%.

The CV of the mean length of the selected samples varied between 0.4% (sample "2014_2033" with n = 125 individuals in 15 size classes) and 1.4% (sample 2014_2024 with n = 100 individuals in 26 size classes) with a mean value of 0.9%. The CV of the mean age of the selected samples varied between 2.1% (sample "2016_2005" with n = 300 individuals in 10 age classes) and 4.8% (sample "2015_2032" with n = 100 individuals in 8 age classes) with a mean value of 3.6%.

Relatively to the original quarterly sample distribution, the selected samples were found to slightly over represent Q3 (+21%) relatively to Q1 (-14%) and Q2(-9%) but this effect is minor and the subset retained quite a lot of the original variability in number of length and age classes. Considering that n = 100 would be the lowest sample size considered in the simulations, the mode detection algorithm was run with 5% threshold for mode detection, i.e., a minimum 5 individuals were required to positively identify a mode (*min_proportion_to_accept_mode* = 0.05). A smooth size span of 2 times the original class width (i.e., $2 \times 5\text{mm} = 10\text{ mm}$) was set up as smoothing parameter for further length analyses. No smoothing was employed age data. Examples of the outputs of the mode detection algorithm are displayed in Figure 4.1.2.5.

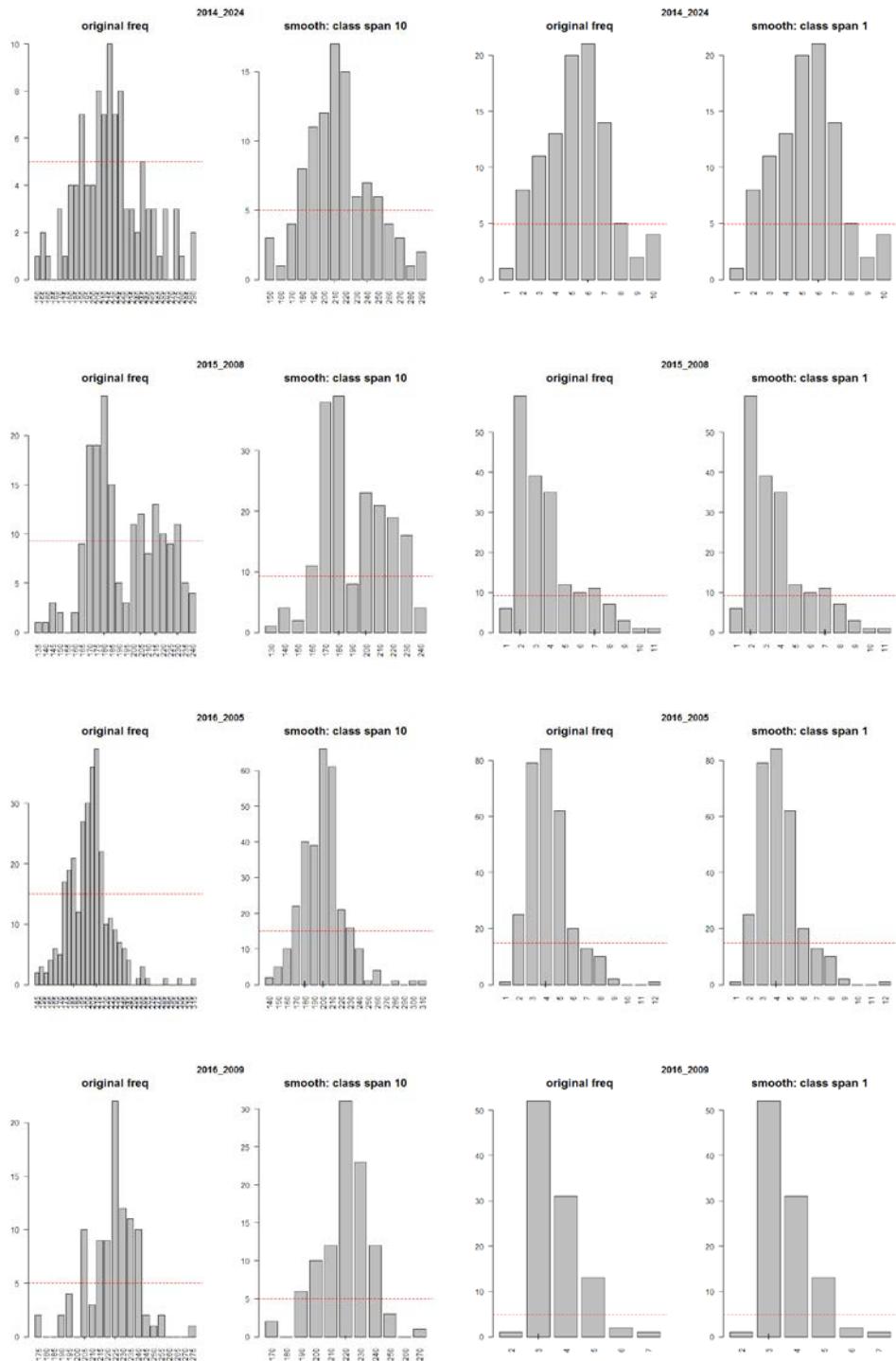


Figure 4.1.2.5 Example of frequency distributions and mode detection in 4 samples selected for simulation analyses. Right graphs: length distribution; Left graphs: age distribution.

Simulations:

The simulations of different sample sizes were done with script *003_sim_data.r*. The following parameterization was employed:

- Minimum number of individuals per sample: 100 fish
- Target biological = c("lenCls", "age")⁴
- Smooth class span (lenCls) = 10; Smooth class span (age) = 1
- Threshold for mode identification = 5% of length distribution
- Weight-length relationship [estimated from original data, via log-log linear model]
 - o a = -12.173888
 - o b = 3.044068
- VBGF starting values [estimated from original data]:
 - o Linf = 215.6136
 - o K = 0.4365
 - o t0 = -2.7327
- sampling_design
 - o stratified = FALSE
 - o strata_var = ""
- Sampling_options:
 - o n_sims = 500,
 - o stages = "one",
 - o stratified = FALSE
 - o strata_var = "none",
 - o replacement=FALSE,
 - o sample_all_available = FALSE,
 - o sample_all_available_warning = FALSE,
 - o stage1_samp_size = NA
 - o samp_sizes = c(seq(10,90, by=10), nrow(df0[df0\$sampId == sampId,]))
 - o replacement = FALSE
 - o sample_all_available = TRUE
 - o sample_all_available_warning = TRUE
 - o models = c("weight-length", "VBGF")

The above mentioned setup resulted in the simulation of 500 replicates containing fish sampled randomly without replacement for each of 9 sample sizes (10 to 90, 10 mm intervals). The simulations took ca. 1 hours to run in an HP ZBook 15 G2 equipped with a Processor Intel(R) Core(TM) i7-4710MQ CPU @ 2.50GHz, 2501 Mhz, 4 Core(s), 8 Logical Processor(s) with 6 clusters allocated in the parallel processing stage. The set of "indicators" included in *func_make_summary_numeric.r* was calculated for each replicate and variable (length and age). These include mean, standard error of the mean, CV of the mean; minimum, maximum, and median; number of size/age classes sampled; number of modes, number of modes identical to the modes of original sample; number of modes (after smoothing), number of modes identical to the modes of origi-

⁴ "sex", "matStage", "mature" were also included but their results are not handled in this report

nal sample (after smoothing); two sample t-test for equal means; two sample Kolmogorov-Smirnov test and mean weighed coefficient of variation (MWCV) of the distribution (according to Gerritsen and McGrath 2007).

Exploration of simulation results:

Exploratory analyses to determine the most appropriate sample size for herring samples of SD25 were carried out using script *004_sim_analysis.r*. To simplify the results, in this case-study the results and discussion are focused on the MWCV and CV of the mean of the length and age distributions. It is possible, but to our knowledge still to be fully demonstrated, that these constitute good summary “indicators” of quality of frequency data.

Graphical outputs obtained from script *004_sim_analysis.r* indicated a fast reduction in both CV of the mean and MWCV with increasing sample size in length and age distributions of all analysed samples. Illustrative examples of summary plots of simulation results are presented in Figure 4.1.2.6. In these examples it is noticeable, e.g., that the CV of the mean length and mean age are already quite low at samples sizes of 20 individuals. Also, when one contrasts the results of sample “2016_2009” and “2014_2024” (both with n=100 individuals but with different MWCV and number of length classes) (Figure 4.1.2.6) with their length and age distribution (Figure 4.1.2.5) one of the characteristics of MWCV as a “indicator” of quality of frequency distributions becomes apparent, namely the its positive relationship to the breath of the size distribution (i.e., wider size distributions require larger sample sizes to attain a similar MWCV).

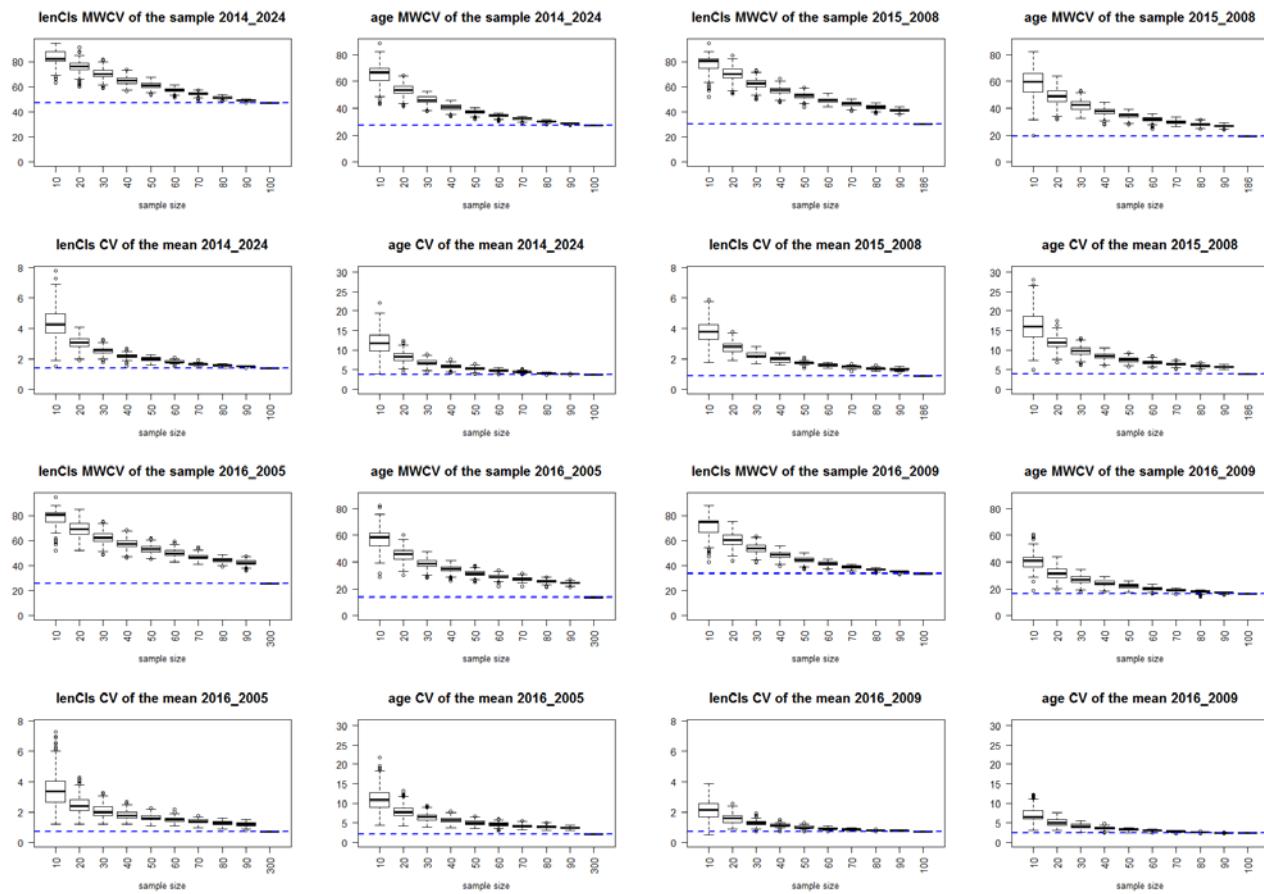


Figure 4.1.2.6 Simulation results for 4 samples (“2014_2024”, “2015_2008”, “2016_2005”, and “2016_2009”). For each sample, the MWCV and CV of the mean length and age distributions of replicates of different sample sizes are displayed. Blue line represents the true value of the sample (at the sample size originally collected). Outputs can be compared to those of Figure 4.1.2.5).

The overall goal of WKBIOPTIM sample-level scripts is to identify conservative target sample sizes that, however, do not imply ineffective cost and time expenditures. Using the provided scripts, the minimum sample sizes required to attain a pre-specified MWCV or CV level can be estimated by various means (see Section 3.1 for other alternatives).

A worst-case scenario (and very conservative) approach is to select as the target sample size the one that always ensured less than $x\%$ MWCV or CV in replicates produced by the simulations⁵, with $x\%$ being defined based on reasonable expectations for the case-study at hand and the overall objectives of the data collection.

To exemplify this procedure, the maximum MWCV and CV obtained for each sample size across all samples can be calculated for each variable analysed (Tables 4.1.2.1–4.1.2.4). From these tables one can infer, e.g., that if one sets the goal ($x\%$) at, e.g., 60% MWCV, then all replicates of sample size 70 and higher met this condition. One can also observe that, under that on samples of that sample size ($n = 70$), one would expect at max 36.5% in the MWCV of the age distribution, 2% of CV in mean length and 7.5% of CV in the mean age. If these values are judged acceptable for the purpose at hand, then $n = 70$ fish hand, then $n = 70$ fish would be an appropriate (albeit conservative) sample size to collect for herring in SD25. That sampling target can also be expressed in terms of fish weight⁶ from which a very conservative value would be 5500 g per sample (the max weight registered in all replicates of sample size 70) and a less conservative value would be 4100 g per sample (the mean weight of all replicates of sample size 70).

One difficulty with the worst-case scenario approach is that it is restricted to the universe of simulated sample sizes, i.e., it does not allow the exploration of the sample size needed to attain a MWCV or CV that is lower than that provided by *min_n*. In the case of herring in SD25, that situation would come up if one wished to estimate, e.g., the sample size that would routinely yield MWCV of 40% in length; or 20% in the age distribution or, 1% in CV of mean length or 5% in the CV of mean age of the samples. Those type of situations occur frequently, e.g., when dealing with more data poor situations where many samples collected end up being left out of the simulations due to their insufficient number of individuals, or, when the objectives of data collection change and higher precision becomes needed.

In such situations, one possible alternative is to model the MWCV and CV using, e.g., an exponential model. Such modelling approach was developed in the follow-up of WKBIOPTIM2 meeting and its results should for now be viewed with caution as they largely result from extrapolations outside the sample sizes used to derive the model and may therefore be prone to significant error and between dataset variability.

Script *004_sim_analysis.r* runs a set of functions that fit and diagnose the models for each sample and output model predictions (e.g., what MWCV or CV is expected for a given sample size) and inverse predictions (e.g., what sample size would correspond to a pre-specified target MWCV or CV). In the end a set of graphs can be produced that allows an evaluation of the expected improvement in the “indicator” chosen from collecting a large sample size, but also a rough estimate of the sample size that is expected to yield a pre-specified targets for that “indicator”. Bearing in mind the previously

⁵ Less conservative approaches can be used when many samples are available (e.g., the sample size that produces $x\%$ MWCV in, e.g., 90% of the simulated samples).

⁶ It is frequently better to state sample size in terms of weight as it avoids the “cherry-picking” that may be involved in selecting a pre-determined number of fish from a large box.

mentioned cautionary note, the latter targets can be outside the range of those observed in the samples.

Modelling results for some herring SD25 selected samples are displayed in Figure 4.1.2.7a-d. The model fits were done on the MWCV of length over the range of sample sizes simulated (10 to 90, in 10 increments of 10 individuals), i.e., excluding the actual sample itself from the modelling. The most appropriate fits were obtained with mod3, i.e., a model fit to boxcox-transformed MWCV. Selected boxcox-lambda for both MWCV of length and age varied between -1.6 and -2.9. Overall, the model fits appeared quite reliable with relatively minor departures from expected being registered even outside the range of sample sizes originally modelled (see e.g., "2015_2008" and "2016_2005") (Figure 4.1.2.8)

Table 4.1.2.1 Maximum MWCV registered in the 500 replicates of length distribution for different sample sizes. Top row: simplified sample ID (n = 38 samples). 1st column on the left: tested sample size.

Table 4.1.2.2 Maximum MWCV registered in the 500 replicates of age distribution for different sample sizes. Top row: simplified sample ID (n = 38 samples). 1st column on the left: tested sample size.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
10	82,7	82,2	88,5	80,9	88,5	82,2	82,2	75,9	82,2	88,5	82,2	82,2	82,2	75,9	75,9	82,7	82,2	74,6	75,9	76	82,2	88,5	60,6	82,2	75,9	82,2	75,9	75,9	80,9	75,9	76	75,9	82,2	88,5	72,4	82,		
20	65,4	61,8	64,5	60,4	64,7	56,6	57,3	57,9	55,2	64	66,5	63,7	61,5	57,9	63,4	56,8	53,2	64,1	61,8	52,6	56,1	59,5	60,1	62,6	43,8	60,9	55,5	59,5	55,9	60,1	59,5	52,3	59,1	54,6	57,9	62,4	56,8	60,
30	53,1	52,1	53	51	52,4	46,1	47	47,5	45,4	53,1	55	52,3	49,9	47,3	52,4	46	42,6	51,2	50	41,7	47,6	47,6	47,8	50,7	34,4	49,7	44,3	49,2	45,2	48,4	48,1	41,7	48	44,5	47	50,2	42,6	50,
40	45,7	45,5	45,7	43,1	45,8	39,7	40,1	41,1	39,3	44,7	48,8	43,4	42,5	40,9	45,8	40,3	36,7	43,5	42,9	35,5	41,3	41,1	40,8	44,1	29,5	42,2	38,8	42	38,5	41,8	41,9	36	40,7	38,4	40,6	44	36,9	42,
50	41	39,9	40,2	37,9	40,5	35,1	35,6	36,2	34,8	39,3	43,5	39,8	38,1	36,5	40,6	35,2	32,7	38,7	38	32,1	35,8	35,8	37,6	38,7	25,6	37,6	33,8	37,2	34,1	37,1	37	31,8	36,8	34,3	36,1	38,9	32,8	37,
60	37,4	36,9	36,6	34,3	36,5	31,8	32,4	33	31,4	36,1	39,4	36,2	34,7	33,1	36,4	31,8	29,7	34,8	34,4	28,9	32,3	32,9	33,8	34,8	23,1	34,4	30,8	33,4	30,7	33,4	33,6	29	32,7	31,2	32,8	35,3	29,1	33,0
70	34,4	33,5	34	31,3	33,9	29,6	29,9	30,3	28,8	33,5	36,5	32,9	31,9	30,5	33,3	29	27,4	32,2	31,8	26,5	29,9	29,9	31,4	31,9	20,8	31,3	28,3	30,8	27,8	30,6	30,8	26,5	29,9	28,9	30,3	32,6	26,7	31,
80	31,9	31,2	31,5	29	31,3	27,3	27,8	28,4	26,6	31,5	33,9	30,7	29,5	28,3	30,7	26,9	25,6	29,6	29,7	25	27,5	27,8	29	29,5	19,2	29,2	26,1	28,5	26	28,5	28,4	24,8	27,8	26,9	28,3	30,3	24,6	28,2
90	30,1	29,1	29,7	27	29,3	25,3	26,1	26,5	24,8	29,2	31,7	29	27,7	26,6	28,5	25,2	24,2	27,7	27,6	23,2	25,7	26,2	27,1	27,5	17,7	27,3	24,3	26,5	24,1	26,5	26,5	23,2	26,3	25,3	26,4	28,4	22,7	26,

Table 4.1.2.3 Maximum CV registered in the 500 replicates of length distribution for different sample sizes. Top row: simplified sample ID (n = 38 samples). 1st column on the left: tested sample size.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
10	3,8	6,8	5,4	4,6	7,8	5,7	2,7	3,5	4,7	5,9	4,8	6,8	5,2	4	6,2	4,4	5,8	7	5,5	7,5	7,6	5,6	7,3	5,1	3,9	5,2	5,5	5,7	6,4	6,4	6,2	4,5	5,5	6,4	4,4	4,8	4,5	5,8
20	2,4	4	3,4	3,1	4,1	3,5	1,7	1,9	2,9	3,8	2,9	3,9	3,2	2,5	3,5	2,6	3,5	3,7	3,5	4,5	4,4	3,6	4,3	3,2	2,6	3,4	3,1	3,2	3,7	3,8	2,7	3,4	3,9	2,8	2,9	2,7	3,6	
30	1,9	2,9	2,5	2,2	3,3	2,6	1,2	1,3	2,2	2,8	2,3	2,8	2,4	1,8	2,6	2	2,7	2,8	2,6	3,3	3,3	2,7	3,3	2,3	1,9	2,5	2,4	2,3	2,7	2,8	2,9	2	2,7	2,9	2,2	2,3	2,2	2,8
40	1,5	2,5	2,1	1,9	2,7	2,2	1	1,1	1,8	2,4	1,9	2,2	1,9	1,5	2,1	1,7	2,2	2,2	2,2	2,7	2,7	2,1	2,7	2	1,5	2,1	2	1,9	2,2	2,3	2,4	1,6	2,2	2,4	1,8	1,9	1,7	2,4
50	1,3	2,1	1,8	1,6	2,3	1,9	0,9	0,9	1,6	2,1	1,6	1,9	1,7	1,3	1,8	1,5	2	1,9	1,9	2,4	2,3	1,8	2,3	1,7	1,3	1,8	1,7	1,7	1,8	2,1	2,1	1,4	2	2,1	1,6	1,7	1,5	2
60	1,2	1,9	1,6	1,4	2,1	1,7	0,8	0,8	1,4	1,8	1,5	1,6	1,5	1,1	1,6	1,3	1,8	1,6	1,7	2,1	2,1	1,6	2,2	1,5	1,1	1,6	1,5	1,5	1,6	1,8	1,9	1,2	1,7	1,9	1,4	1,4	1,8	
70	1	1,7	1,5	1,2	1,9	1,5	0,7	0,7	1,3	1,7	1,3	1,5	1,4	1,1	1,4	1,2	1,6	1,5	1,6	2	1,9	1,4	1,8	1,3	1	1,5	1,3	1,4	1,6	1,7	1,1	1,6	1,6	1,7	1,2	1,3	1,2	1,6
80	0,9	1,5	1,3	1,1	1,7	1,3	0,6	0,7	1,2	1,6	1,2	1,3	1,3	0,9	1,3	1,1	1,5	1,3	1,4	1,8	1,8	1,3	1,6	1,2	0,9	1,4	1,2	1,2	1,3	1,5	1	1,4	1,5	1,1	1,2	1,1	1,5	
90	0,9	1,4	1,2	1	1,5	1,2	0,6	0,6	1,1	1,5	1,1	1,2	1,2	0,9	1,2	1	1,4	1,2	1,3	1,6	1,6	1,2	1,5	1,1	0,8	1,2	1,1	1,1	1,2	1,3	1,4	0,9	1,3	1,4	1	1,1	1	1,4

Table 4.1.2.4 Maximum CV registered in the 500 replicates of age distribution for different sample sizes. Top row: simplified sample ID (n = 38 samples). 1st column on the left: tested sample size.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
10	19,7	22,8	23,5	20,9	22,1	21,9	24,7	22,8	22,2	28,1	20,9	21,8	18,5	16,1	23,8	18,2	26,9	26,7	25,2	21,6	30,3	26,4	21,8	19,3	12,3	18,9	18,9	20,4	18,5	20,7	25,2	16,7	16,8	13,9	14,4	15,5	19,7	21,7
20	11,3	14	14,5	12,6	12,5	12,3	14,1	14,6	13,7	17,5	13	13,8	11,6	9,6	13,1	12,5	14,1	15,5	14,1	13,9	17,2	16,9	13,4	12,7	7,6	10,3	11,6	11,4	11,2	12,2	13,5	10,2	10	9,1	8,1	10,5	11,3	13,8
30	8,3	11,1	11	9,8	9,1	9,4	10,7	11,1	10,4	13,1	9,9	9,8	9,2	7,9	9,7	8,9	10,5	11,9	11,4	10,3	12,6	12,3	9,6	9,6	5,6	8	8,8	9	8,2	9,8	10,2	7,4	7,8	7,1	6,4	7,5	9,2	10,2
40	6,9	8,5	8,9	8,1	7,7	7,7	9,4	9,3	8,7	10,6	8,4	8,3	7,2	6,3	7,8	7,6	8,8	9,4	9,3	8,8	10,6	10,1	7,9	7,5	4,8	6,4	7,4	7,3	6,9	7,9	8,3	6,3	6,1	5,7	5,4	6,4	7,3	8,3
50	6	7,6	7,8	7,4	6,6	6,7	8,1	7,7	7,5	9,4	7,5	7,3	6,5	5,5	6,5	6,4	7,4	8	8	7,6	8,9	8,7	6,7	6,4	3,9	5,6	6,2	6,3	6,1	6,8	6,9	5,3	5,5	5	4,4	5,5	6,3	7,2
60	5,5	6,8	7	6,2	5,6	5,7	7,1	6,9	6,8	8,5	6,3	6,4	5,5	4,9	5,8	5,7	6,7	6,9	6,9	7,8	7,4	6	5,7	3,5	4,9	5,7	5,6	5,3	6	6,4	4,8	4,8	4,4	4,1	4,9	5,5	6,4	
70	4,8	6,1	6,3	5,6	5,2	5,1	6,5	6,3	6,2	7,5	5,8	5,5	5	4,4	5,3	5,3	6,1	6,3	6,3	7	6,8	5,6	5,1	3,1	4,4	5	5	4,8	5,5	5,7	4,2	4,3	4,1	3,6	4,5	4,9	5,7	
80	4,4	5,6	5,7	5	4,6	4,7	6,1	5,8	5,7	7	5,2	5,2	4,6	4,1	4,7	4,7	5,5	5,8	5,7	5,8	6,3	5,1	4,6	2,9	4,1	4,5	4,4	4,4	4,9	5,1	4	3,9	3,7	3,3	4	4,5	5,2	
90	4	5,2	5,2	4,5	4,2	4,3	5,5	5,2	5,2	6,4	5,1	4,7	4,3	3,7	4,3	4,3	5,2	5,2	5,2	5,8	5,7	4,6	4,2	2,6	3,7	4,2	4,2	4	4,5	4,7	3,7	3,4	3,1	3,7	4,1	4,9		

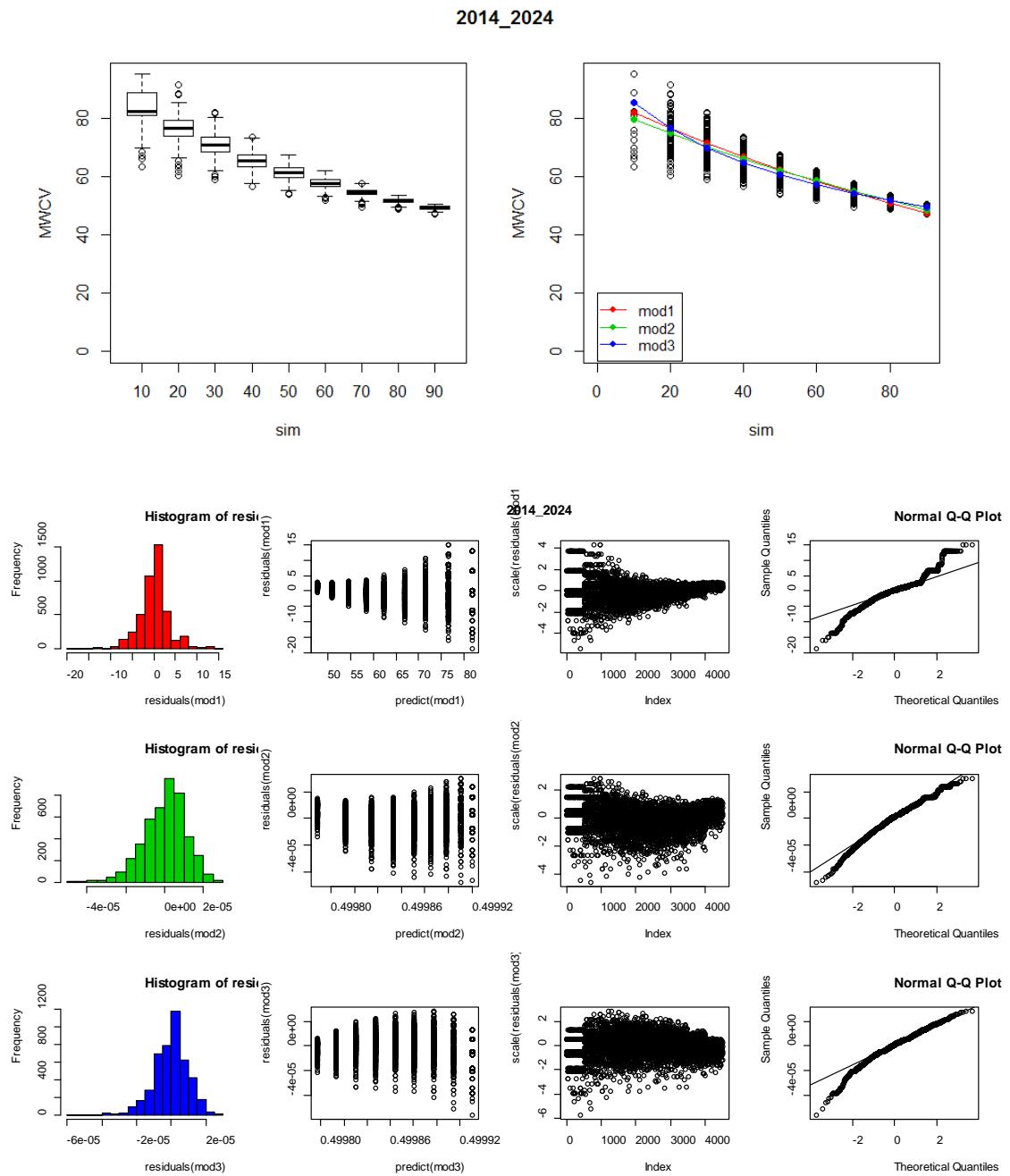


Figure 4.1.2.7a Exponential model fits and diagnostic tests on MWCV of length distribution: sample “2014_2024”.

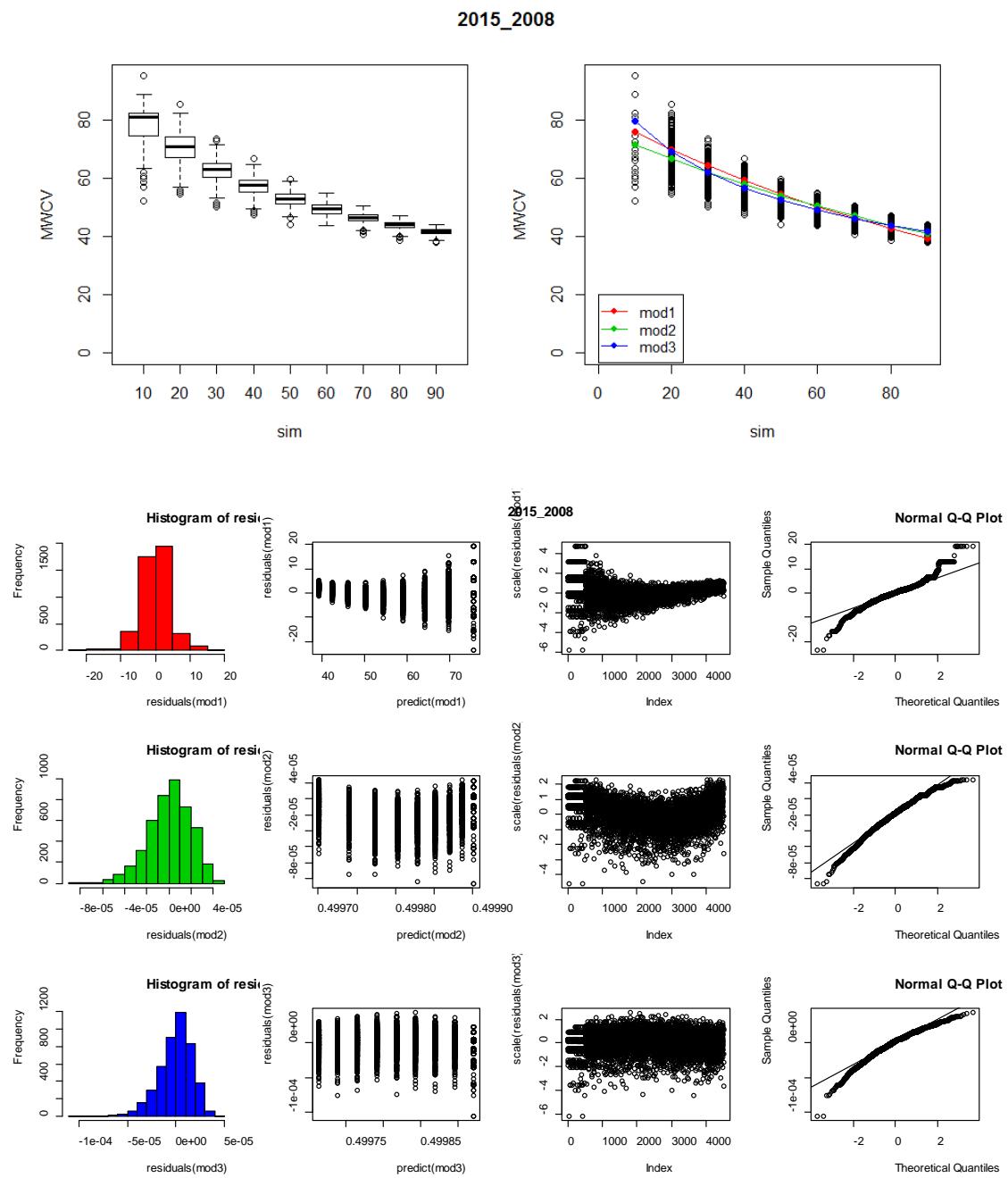


Figure 4.1.2.7b Exponential model fits and diagnostic tests on MWCV of length distribution: sample “2015_2008”.

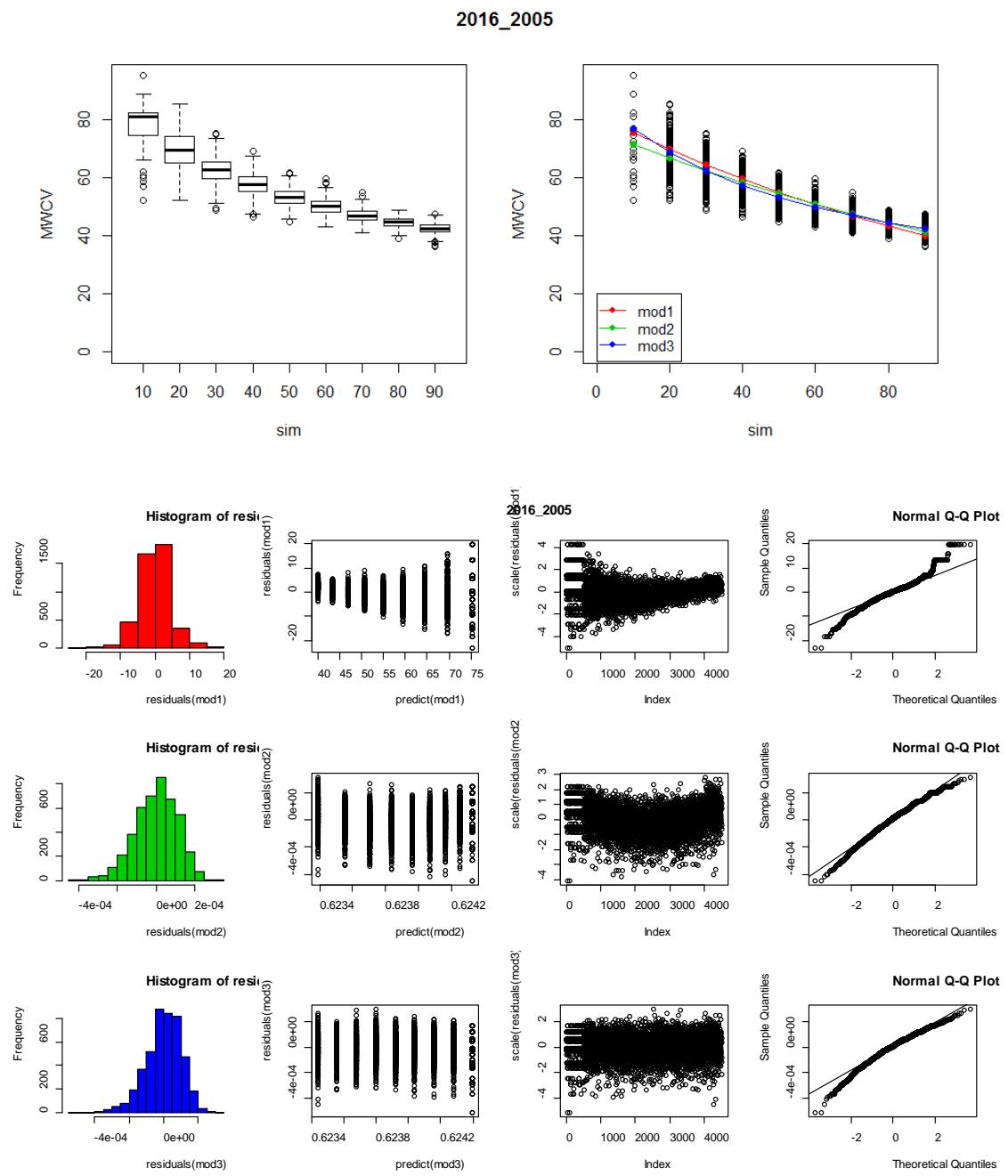


Figure 4.1.2.7c Exponential model fits and diagnostic tests on MWCV of length distribution: sample “2016_2005”.

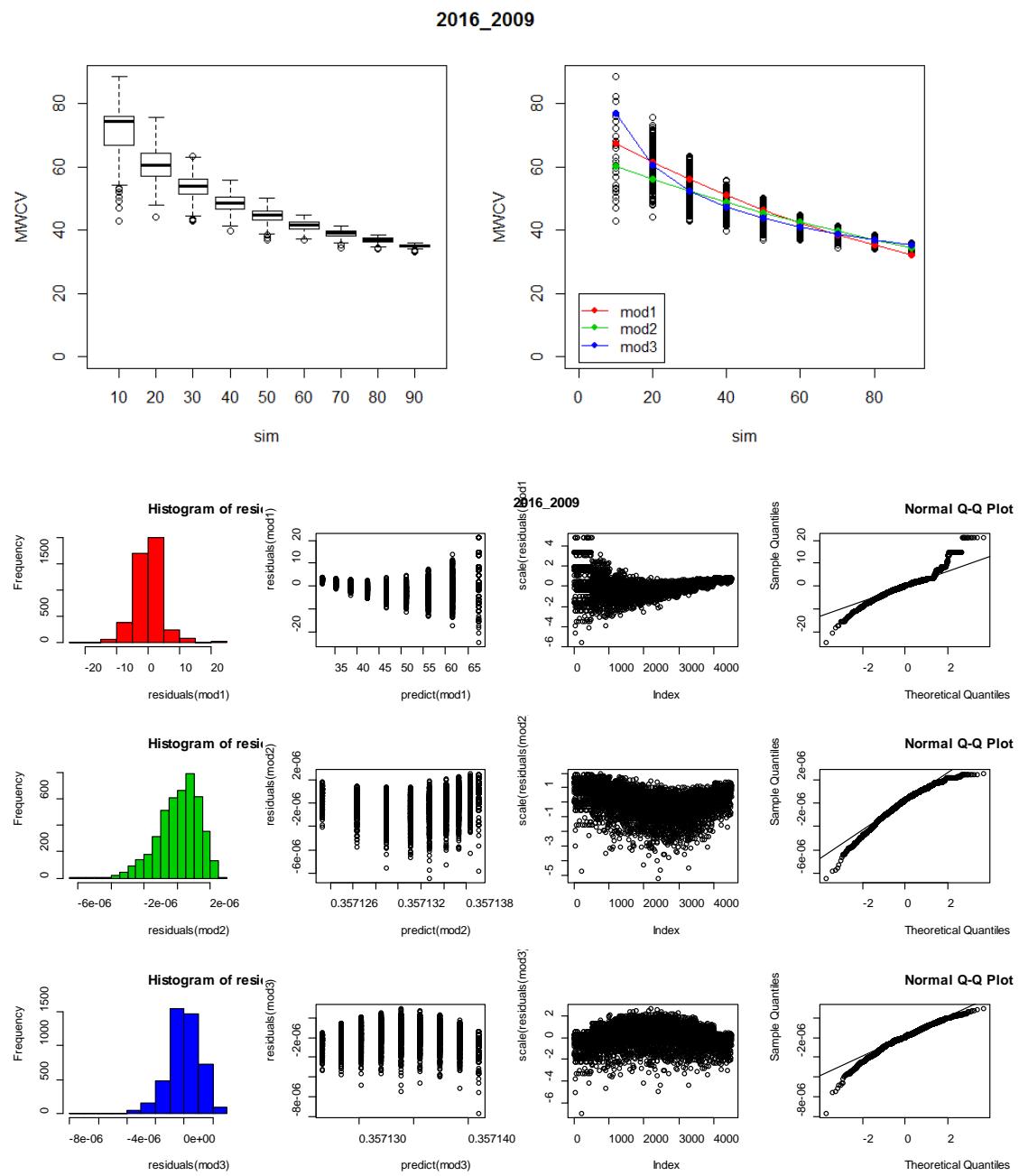


Figure 4.1.2.7d Exponential model fits and diagnostic tests on MWCV of length distribution: sample “2016_2009”.

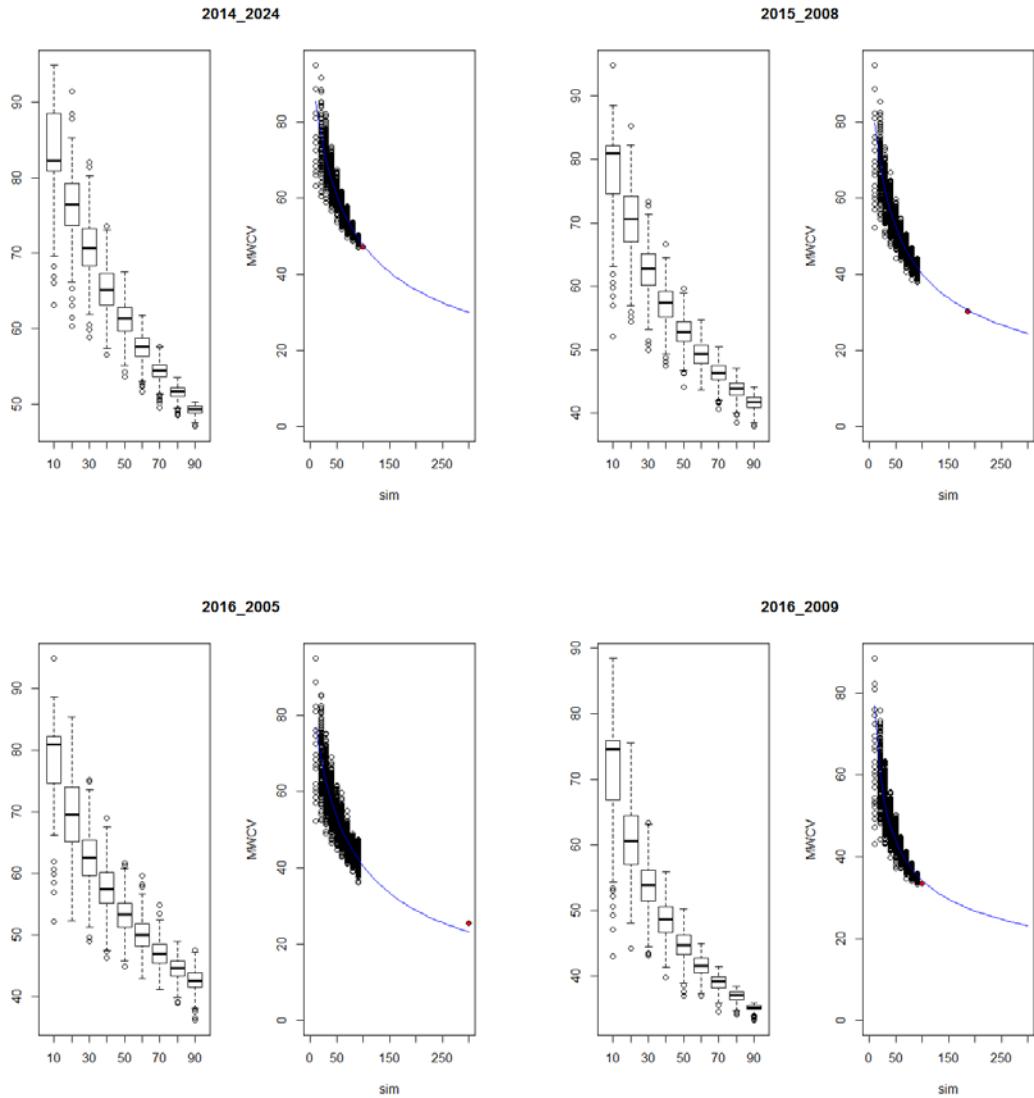


Figure 4.1.2.8 Examples of final model fit (mod3) on MWCV of length distribution including model extrapolation outside the universe of sample sizes used to fit it. Red point indicates the actual sample size and MWCV of length distribution not included in the fit.

The graph of improvements with increased sample size indicate that above $n = 90$ individuals less than 5% improvement in MWCV is obtained by adding an additional 10 individuals (Figure 4.1.2.9); and that above $n = 250$ individuals this improvement is less than 2.5% improvement for a similar increase in sampling effort. Inverse prediction graph indicates that a target MWCV of 40% in length distribution may be obtained with ca. 200 individuals per sample but that $n = 100$ individuals will already yield that MWCV (or less) in 50% of the occasions (Figure 4.1.2.9). The latter may be a suitable compromise when one considers that in most applications it is the quality of the probabilistically combined length frequencies to national or even international level samples that is the ultimate objective of the data collections.

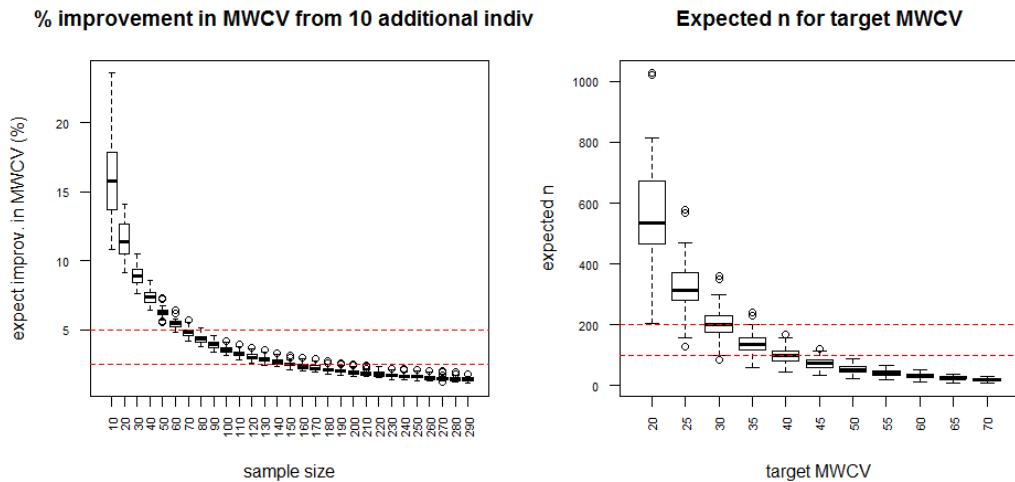


Figure 4.1.2.9 Expected % improvement in the MWCV of the length distribution resulting from +10 individuals being collected (left). Expected n for a range of different target MWCV (right). “red lines”: see text.

4.1.3 Conclusions and follow-up

The results obtained in the present case-study should be viewed with caution as the code is still being developed and MWCV and CV may not represent the full range of characteristics of length or age frequencies. Furthermore, it is also objective of the Swedish sampling programme to collect data on sex and maturity of herring and the behaviour of these variables in relation to sample size has not been analysed here. Outputs for these other variables and other indicators must be pondered and a consultation process with the end-users of the data is needed before a conclusion is reached.

4.1.4 References

Gerritsen H. D. and D. McGrath (2007). Precision estimates and suggested sample sizes for length-frequency data. *Fishery Bulletin* 106:116-120.

4.2 Blue whiting in the ICES Division 27.9.a

4.2.1 Blue whiting sampling data

The blue whiting data used to test the script for the sampling effort optimization procedure was collected under the Portuguese Data Collection Framework Programme (DCF). For estimation biological parameters the blue whiting samples are usually length stratified (collection of 10 individuals by each length class) and from these all the biological parameters were estimated. Thus, most of the blue whiting data available are already length stratified, with few individuals sampled for biology (length, sex, maturity and age). Although, for this workshop the dataset used in the tests, from 2008, were randomly collected by length class.

Blue whiting 2008 dataset has been converted to the ICES RDB format.

Data description:

DCF (Data collection framework);

Country: Portugal;

Species: Blue Whiting;

Year: 2008

4.2.2 Descriptive analysis

4.2.2.1 ALK R-script

The blue sampling data from 2008 ($n = 638$ (first semester), $n = 715$ (second semester)) was applied to test the ALK R-script. This tool, was been develop aiming to determine the number of otoliths that should be annually collected and read by each length class. The generally idea was optimizing the number of otoliths annually read without compromise the annual age-length key (ALK) used for stock assessment (see Section 6.1. for details).

Blue whiting presents sexual dimorphism on growth (female's growth faster than males) and there are significant differences in the weight-length relation between semesters due to spawning occurrence during the first semester.

Taking into account the main aspects of this species biology the algorithm developed for the simulation process presents the following options:

(1) Number of otoliths selected by one of the following options:

- (1.1) Length class for each sample (Period=sample);
- (1.2) Length class for each quarter (Period=quarter);
- (1.3) Length class for each semester (Period=semester);

(2) The selection of individuals in (1) include an option to define the sex-ratio (equally to 1 - same number of males and females by length class; from 1–0.5 more females than males; from 0.5–0 more males than females).

The parameters from the von Bertalanffy growth models (VBGM) were determined for each dataset from the simulation process. Root mean squared error (RMSE) and mean absolute percentage error (MAPE) from the von Bertalanffy growth parameters estimates were determined through (repeated) K-fold cross-validation, considering the different scenarios of this case-study.

The scenarios tested were based on the number of otoliths read by length class ($n = c(1,2,4,5,10,20,40,50,100)$) and by semester, with the sex-ratio set to 1. In the length classes where these numbers of otoliths were not available, the total number of otoliths was used instead. A total of 100 sampling simulations, *without* replacement, were performed in order to test each of those hypotheses.

The simulation results of the different scenarios were presented in Section 4.2.2.1.

4.2.2.2 Sample-level R-script

The blue whiting sample named as: "2008_2008" is used to illustrate the sample-level R-script test results. This sample comprises a total number of 108 individuals. Blue whiting length distribution in the original sample is shown in Figure 4.2.2.1. In Figure 4.2.2.2 is presented the age distribution from the "2008_2008" sample. The sex distribution of individuals in this sample is in Figure 4.2.2.3.

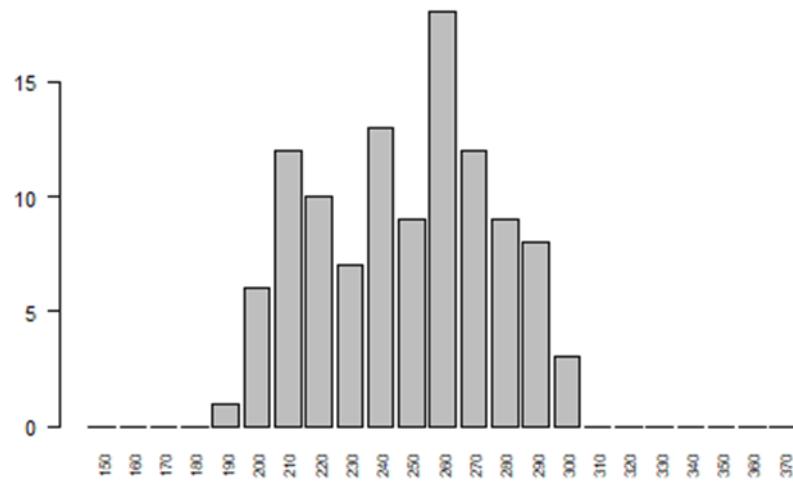


Figure 4.2.2.1 Length distribution (mm) of blue whiting from sample "2008_2008".

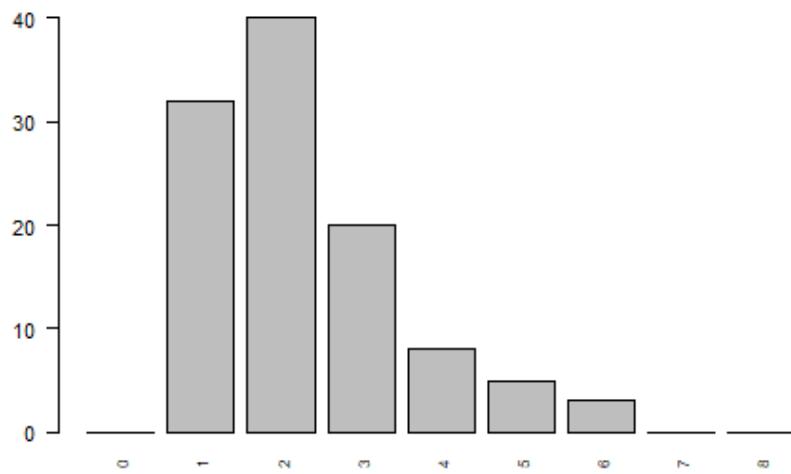


Figure 4.2.2.2 Age distribution of blue whiting from sample "2008_2008"

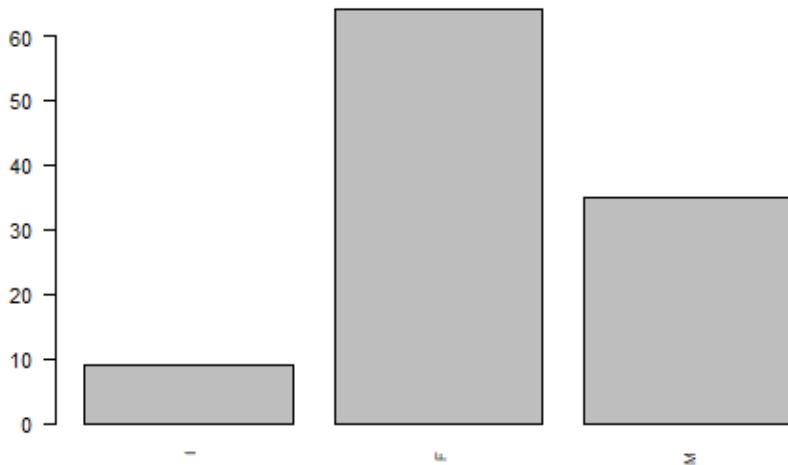


Figure 4.2.2.3 Sex distribution of blue whiting from sample “2008_2008”. (I - undetermined; F - female; M - male).

This dataset was applied to generate bootstrap simulations in order to evaluate the differences in the blue whiting structure (length, age and sex) from samples with a lower number of individuals, from $n = 100$ to $n = 10$ (intervals of 10 of magnitude). For each sample size option, a total of 50 runs were performed. Furthermore, the parameters of VBGM applied to the age and length data, from each set of simulations, were compared.

4.2.3 Simulation results

4.2.3.1 ALK R-script

The 2008 blue whiting annual ALK is shown in the figure above (Figure 4.2.3.1.1.)

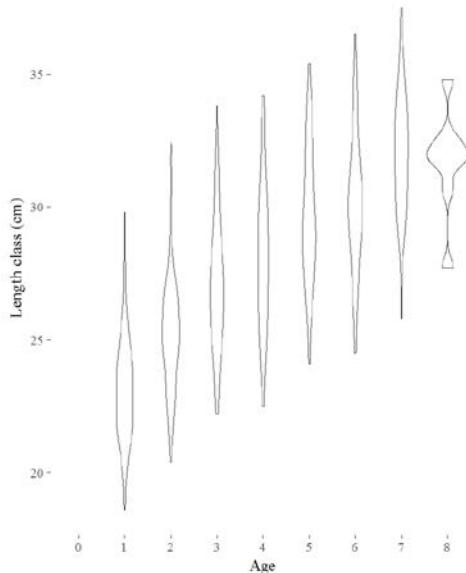


Figure 4.2.3.1.1 2008 blue whiting annual age-length key.

The parameters of the VBGM adjusted to the 2008 data were $L_{inf} = 36.78$, $K = 0.16$ and $t_0 = -5.25$. The parameters obtained from the simulations taking into account the different number of otoliths simulation scenarios are presented in Figure 4.2.3.1.2.

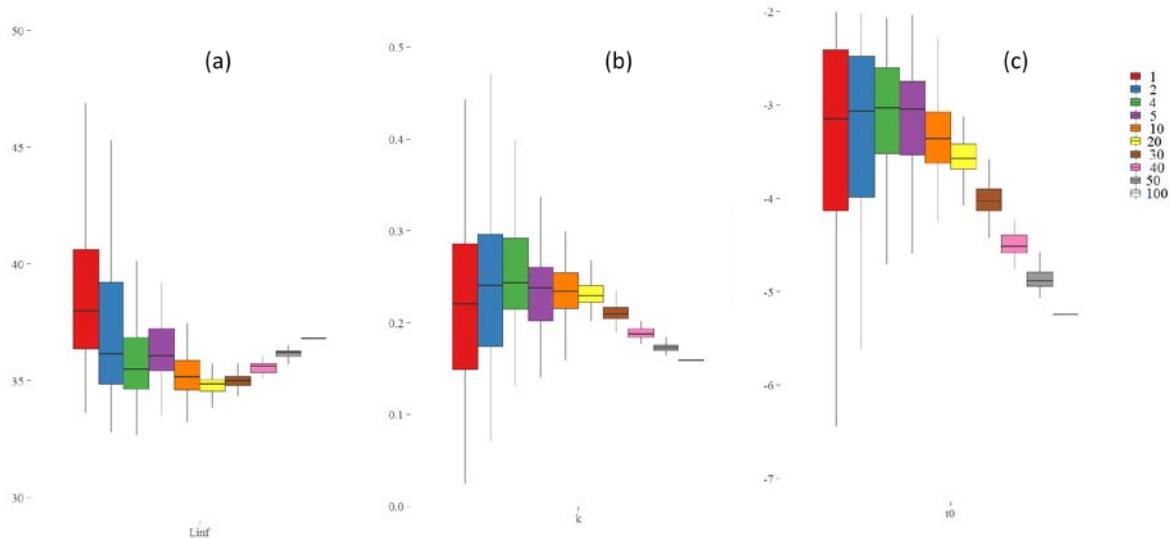


Figure 4.2.3.1.2 - Parameters from the VBGM adjusted to the data simulations by semester and by length class $n = c(1, 2, 4, 5, 10, 20, 30, 40, 50$ and 100) otoliths read. ((a) L_{inf} , (b) K and (c) t_0)

The root mean square error (RMSE) and the mean absolute percentage error (MAPE) from the VBGM estimates were determined via (repeated) K-fold cross-validation, considering the hypotheses described above and presented on the next table (Table 4.2.3.1.). The prediction errors were smaller and very similar from a fixed number of 40 to 100 otoliths by length class and by semester.

Table 4.2.3.1 The prediction errors (RMSE and MAPE) from the VBGM estimates through cross-validation by the fixed number of otoliths by semester by length class (cm) (2008 all data: $L_{inf} = 36.78$, $K = 0.16$ and $t_0 = -5.25$). The upper (UCI) and lower (LCI) 95% confidence intervals for the von Bertalanffy parameters.

Data	RMSE	MAPE	L_{inf}			k			t_0		
			Median	95% LL	95% UL	Median	95% LL	95% UL	Median	95% LL	95% UL
Fixed number of otoliths by length class (cm) and by semester	2008		36.86	33.99	44.24	0.16	0.09	0.23	-5.26	-7.59	-3.99
			24.7	23.99	36.38	35.79	37.07	0.29	0.27	0.32	-1.87
			24.1	22.15	34.05	33.68	34.54	0.34	0.32	0.36	-1.62
			23.8	23.16	33.31	33.05	33.55	0.37	0.35	0.38	-1.57
			23.9	22.97	33.42	33.19	33.65	0.37	0.35	0.38	-1.57
			23.7	23.07	32.54	32.39	32.69	0.39	0.38	0.40	-1.53
			23.6	22.79	32.32	32.21	32.43	0.38	0.37	0.39	-1.66
			23.5	22.76	32.34	32.23	32.45	0.35	0.34	0.36	-1.96
			23.4	22.59	32.58	32.46	32.72	0.32	0.31	0.33	-2.24
			23.4	22.41	33.01	32.87	33.14	0.29	0.28	0.29	-2.54

4.2.3.2 Sample-level R-script

The blue whiting dataset was used to simulate length distribution in samples with different number of individuals ($n = c(100, 90, 80, 70, 60, 50, 40, 30, 20, 10)$) and the results from simulations with replacement are shown in Figure 4.2.3.2.1. For the same simulation procedure, the age and sex distributions of blue whiting in the different samples sizes are presented in Figure 4.2.3.2.2 and Figure 4.2.3.2.3.

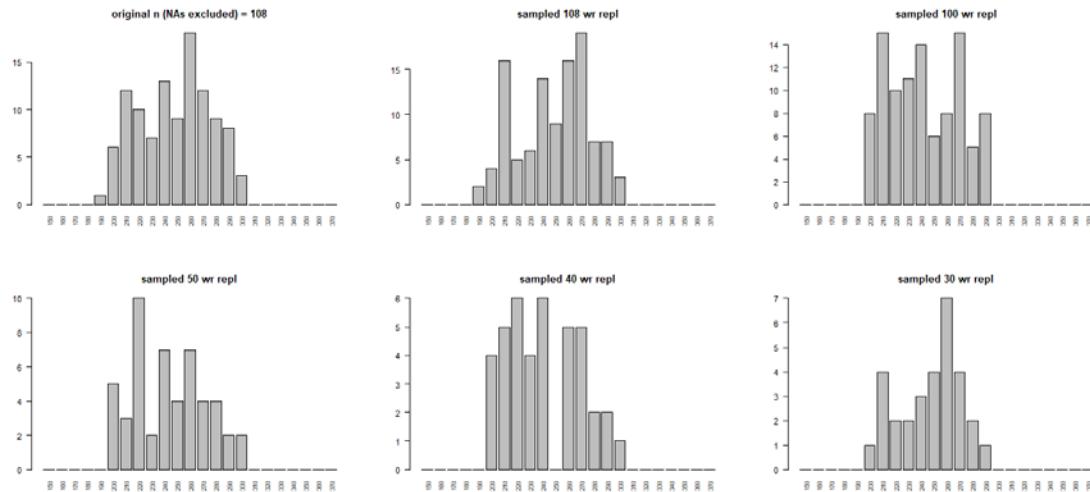


Figure 4.2.3.2.1 Length distribution (mm) from the bootstrap blue whiting data *with replacement* from sample “2008_2008”.

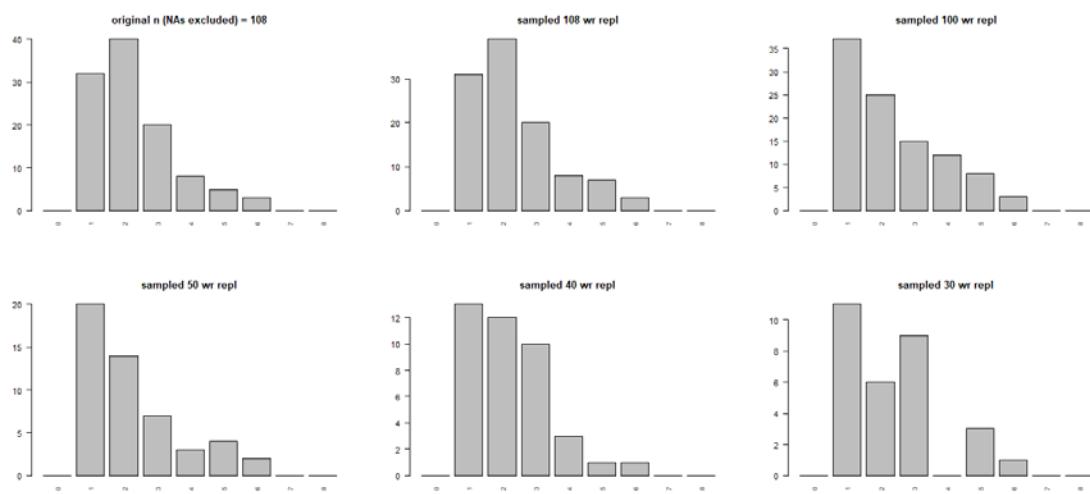


Figure 4.2.3.2.2 Age distribution from the bootstrap blue whiting data *with replacement* from sample “2008_2008”.

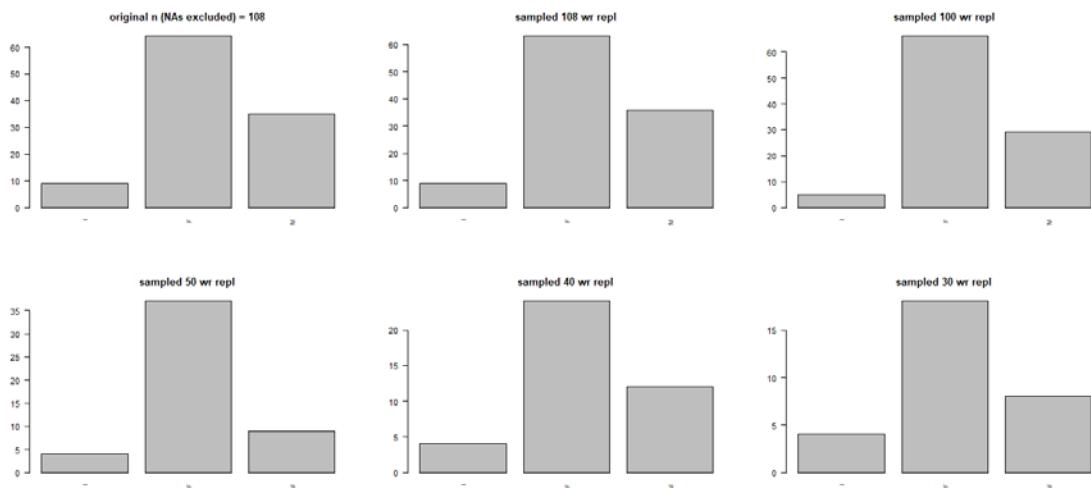


Figure 4.2.3.2.3 Sex distribution from the bootstrap blue whiting data *with* replacement from sample “2008_2008”.

The coefficient of variation (CV) of the mean length and mean the weighted CV (MWCV) of the blue whiting samples are shown in Figure 4.2.3.2.4.

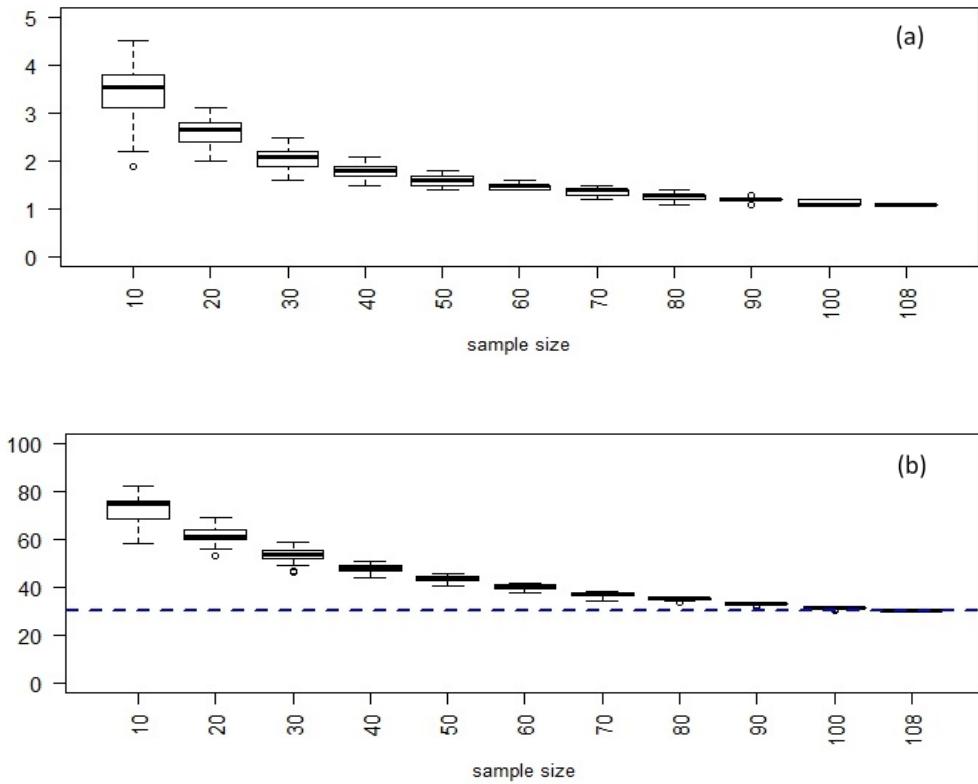


Figure 4.2.3.2.4 Blue whiting length data from simulations on sample “2008_2008”: (a) coefficient of variation (CV) and (b) mean weighted CV (MWCV).

The mean length and the coefficient of variation of the mean from the simulations for each sample size are represented in Figure 4.2.3.2.5.

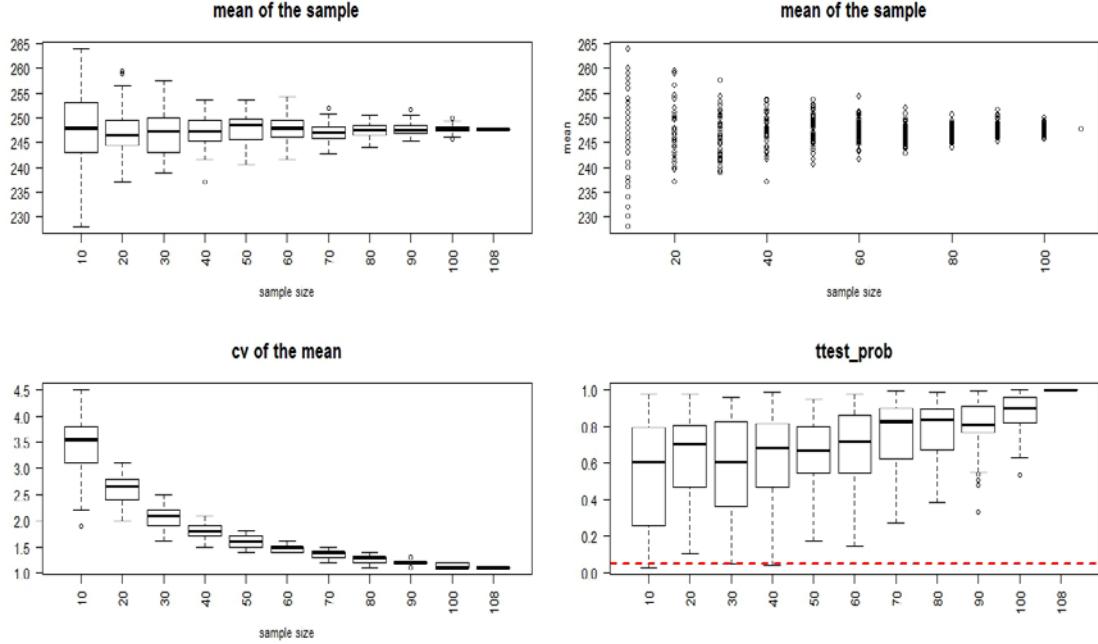


Figure 4.2.3.2.5 Mean length and cv of the mean by sample size from simulations on sample “2008_2008”.

Age and length data resulting from the different simulated scenarios on the blue whiting sample “2008_2008” were used to adjust the VBGM, the parameters L_{inf} , k and t_0 from the simulations are shown in Figure 4.2.3.2.6.

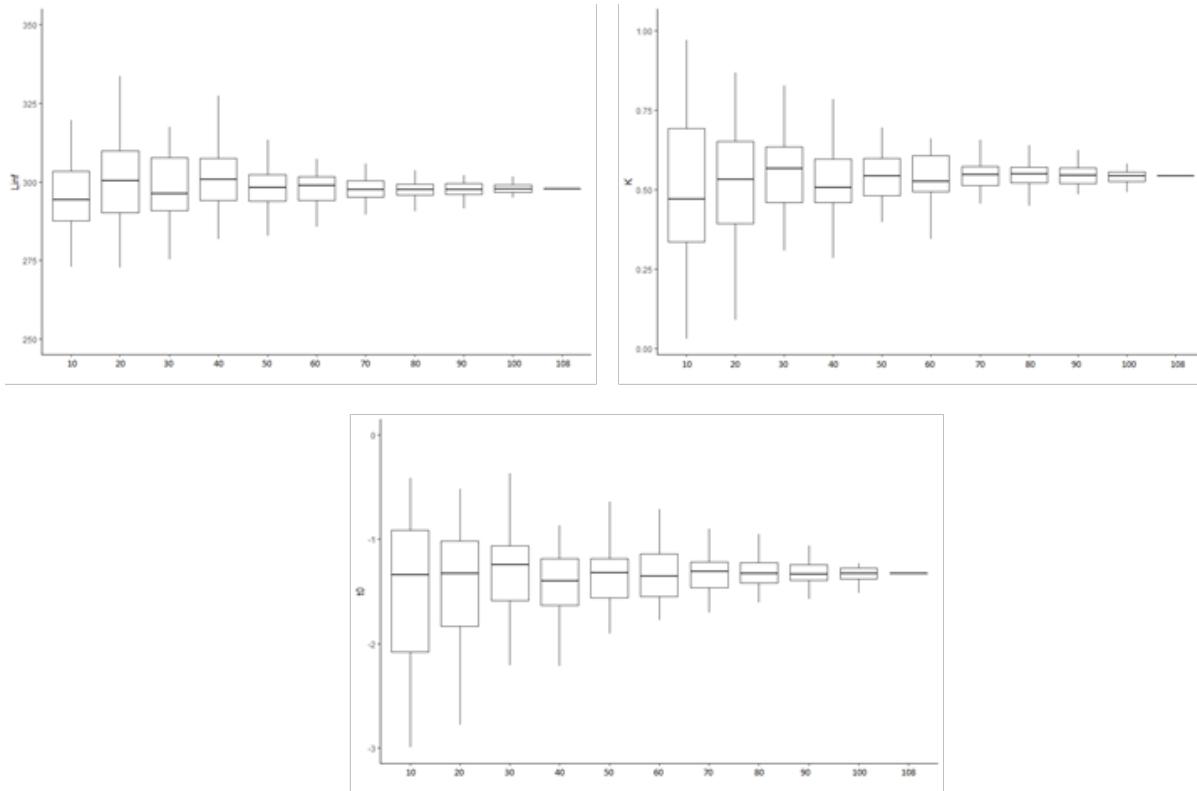


Figure 4.2.3.2.6 - Boxplots of the parameters from the VBGM: L_{inf} , K and t_0 from simulations on sample “2008_2008”.

4.2.4 Discussion

Currently in the blue whiting sampling a subsample covering all length classes, with the number of individuals ranging between 1 and 10, was already adopted as a sampling procedure for biological parameters. Thus, the studies that are now being conducted are focused on testing the number of otoliths to collect (read) by sample and length class to construct the age-length keys used for stock assessment (see Section 4.2.2.1 for details). Although, the total amount of samples by quarter, by semester or in an annually basis is also currently being evaluated. This is mainly due to the fact that the length ranges on landings could significantly change by harbour and fleet, and in those cases the number of samples should be different (higher) from cases were no significant changes occurs.

In the blue whiting sample (“2008_2008” dataset), the number of individuals by length class varied between 3 and 15, with a length range from 190 mm until 300 mm. Thus, the results concerning sampling optimization for the number of individuals sampled by length class indicates just a small reduction. The VBGM parameters in samples with less than 70 individuals reveal differences in the median when compared with the original data. Although, some concerns should be made by using the VBGM parameters as a quality indicator for each sample. The use of VBGM parameters by sample or by all the annual data combined should be further evaluated.

4.2.5 Next steps

The next steps are:

- compare the mean length at age in the different scenarios;

- improve the script in order to be possible to identify the optimal sampling level, which in some cases could imply to increase the number of samples collected or the number of individuals sampled;
- identify the thresholds for quality indicators to be used as decisions makers on sampling optimization.

4.3 Beam trawl survey data 2014–2017 for sole and plaice - Optimizing the Belgian BTS survey with respect to sample sizes for ages

4.3.1 History of the BTS survey

The Belgian offshore beam trawl survey, collecting fisheries-independent data primarily for plaice and sole in the North Sea (area 4.b.c), started in 1992. The continuous time-series using a 4 m-beam trawl as standard gear, started in 1992. The 10-day survey takes place at the end of August/beginning of September. 62 fixed stations are fished for 30 min at 4 knots. There is no fixed order in which the stations are fished but a similar yearly pattern is executed as much as possible. Although the target species are plaice and sole, all fish species are measured since 2010. All epibenthic species are recorded (numbers).

Objectives of the BTS survey:

- ✓ Create a fisheries-independent stock estimate for plaice and sole for the sampled area
- ✓ Collection of data on all fish species for ecosystem purposes
- ✓ Collection of data on epibenthos species for ecosystem purposes

The indices are supplied to the relevant ICES stock assessment working groups.

Catch monitoring and data collection/storage process of the BTS survey:

- ✓ The whole catch of each of the hauls is weighted. The catch is sorted for all fish species, rays, sharks and shellfish and total weight by species is collected.
- ✓ A selection of species is measured a.o. *Solea solea* and *Pleuronectes platessa*. Subsampling by species may occur for the length measurements when the numbers are extremely high.
- ✓ For a selection of species a.o. *Solea solea* and *Pleuronectes platessa*, biological data (age, length, weight,...) is collected. Before 2017, 5 otoliths per cm class per rectangle were collected for sole and plaice. Since 2017, 3 otoliths per cm class per rectangle were collected and also age information was gathered for individuals smaller than 15 cm.
- ✓ All data are available in DATRAS-format (HH, HL, CA).

Optimizing the Belgian BTS survey with respect to sample sizes for ages:

The total number of otoliths collected for sole and plaice in the last years is quite large. For plaice (*Pleuronectes platessa*) 986, 969, 861 and 658 otoliths were collected in 2014, 2015, 2016 and 2017 respectively. For sole (*Solea solea*) 681, 689, 547 and 561 otoliths were collected in 2014, 2015, 2016 and 2017 respectively. At the Institute for Agricultural and Fisheries Research (ILVO), the following questions were raised:

- Is a reduction (<3) of the age readings of sole and plaice possible? → Budget re-allocation. What is the ‘minimum’ number?
- Is age sampling for individuals smaller than e.g. 8 cm useful? (probably all of them have age 0)
- Is stratification by length class useful? → more variation in age for larger individuals (compared to small fish) → more sampling needed for larger individuals?

This case was presented (see Annex 3) and discussed at the WKBIOPTIM 2. ILVO will use the input of the discussion and the scripts developed by WKBIOPTIM 2 to further explore the potential for optimization of the Belgian BTS survey design.

4.4 International Baltic Acoustic Survey (IBAS) (Finland 2015)

The data used to test the proposed sample level algorithm is the International Baltic Acoustic Survey (IBAS) data for the year 2015 for the species herring (*Clupea harengus*). The data consist of herring samples from herring stock SD30 and Finnish parts of herring stock SD25–32. There are 29 hauls in the data set, but we have used haul ID 21 for our analysis. The original data is in the ICES acoustic data format and was transformed to the ICES RDB format containing necessary variables from the CA, HL and HH records using a CA-HL data function developed for this purpose.

The IBAS Surveys currently takes place once a year between Q3 and Q4 (typically last week of Sept and first week of Oct). The aim is to collect acoustic and biological data. Biological data on distribution, relative abundance, and biological information together with acoustic data are used to estimate abundance indices for Baltic herring. There are nine nations participating: Estonia, Finland, Germany, Denmark, Latvia, Lithuania, Poland, Russia and Sweden.

CA-HL data function was applied on the Finnish IBAS data.

Attempt was made to apply the function on North Sea International Bottom Trawl Survey data, but the large number at length values made the running of the script inconvenient. Testing could proceed with a subsample of the NS-IBTS data. There were further plans on applying the script to Beam Trawl Survey data. Data collection framework data sets do not need generated CA records, since those datasets do not have missing individual records.

Data description:

IBAS (Baltic International Acoustic Survey);

Country: Finland;

Species: Herring;

Year: 2015

4.4.1 Descriptive analysis

Descriptive analysis of the Baltic Sea herring species from 29 stations in the Baltic sea in year 2015 are provided below. Figure 4.4.1.1 shows a bar plot of the age distribution of herring. The age of herring ranged from 0–21 years and catch rates of younger fishes, particularly 1 and 2-year old are higher compared with older fish.

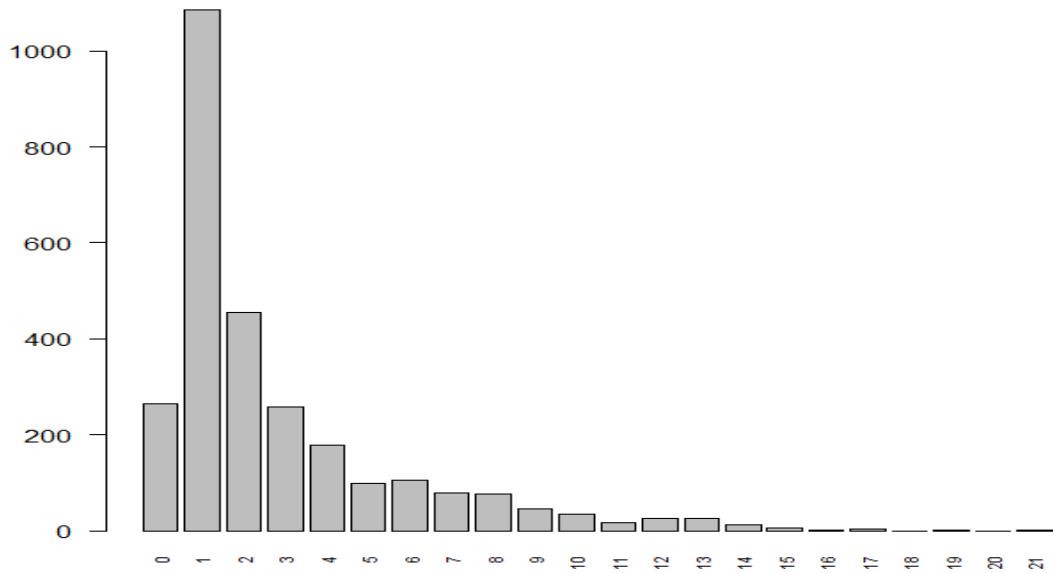


Figure 4.4.1.1 Barplot showing age information for herring species from IBAS 2015

From visual inspection of Figure 4.4.1.2, it is observed five modes occurring at length classes: 70 mm (7.0 cm), 110 mm, 135 mm, 155 mm, and 245 mm, with the highest mode occurring at 135 mm. However, it is possible to threshold the length classes so that at minimum two individuals are included to form a mode. Also catch rates of female herring are higher compared with males as shown in Figure 4.4.1.3. There are at least 200 more females caught than males.

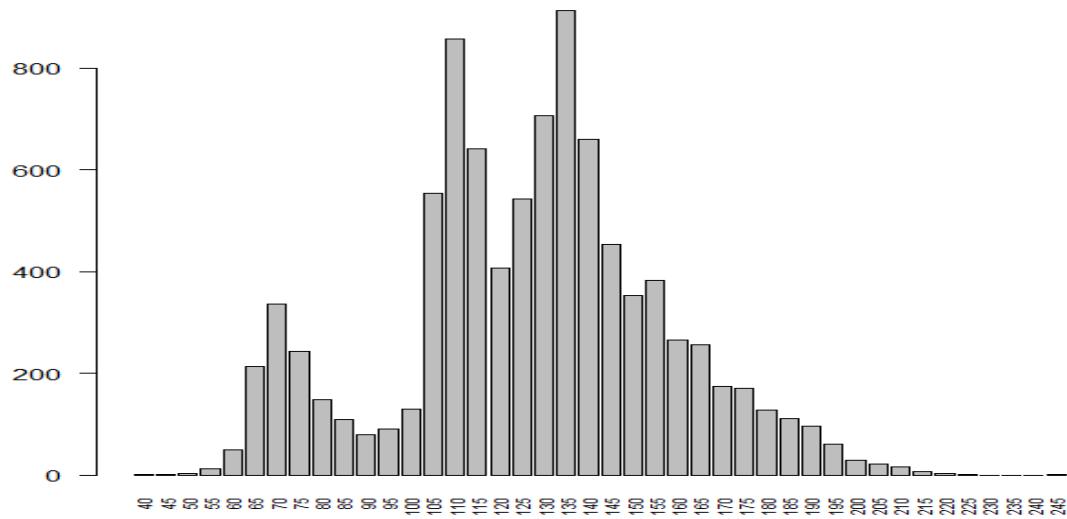


Figure 4.4.1.2 Histogram of length distribution of herring in the Baltic Sea.

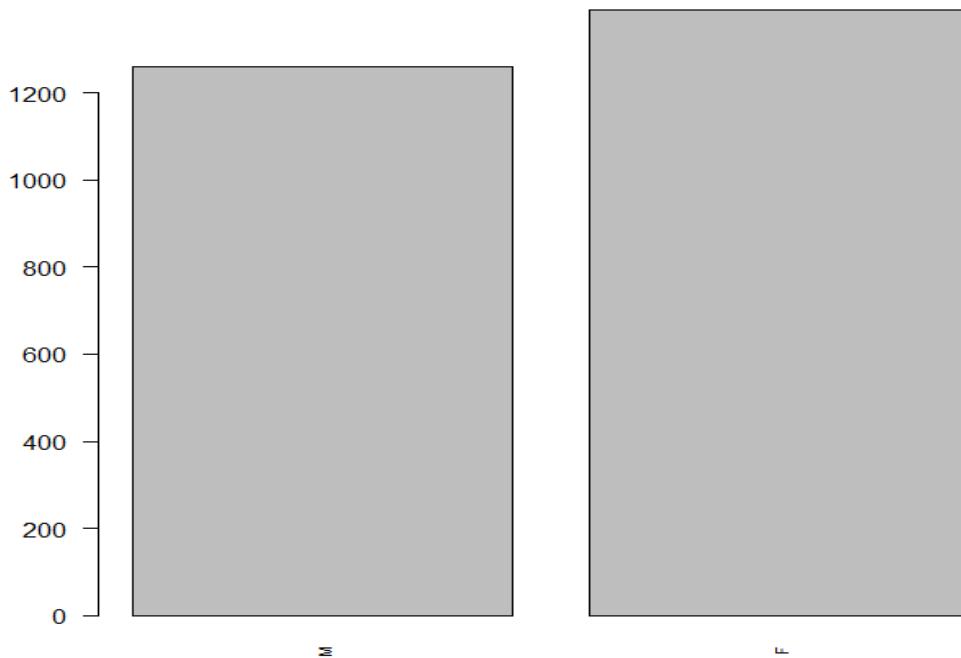


Figure 4.4.1.3 Barplot showing the distribution of sex of the herring species

4.4.2 Simulation on the entire dataset

As an example, it is used haul number 21 to simulate lengths of herring in samples of 319, 200, 150, 100, 50 from an original sample of 319 herring. Data was simulated with replacement (Figure 4.4.2.1) and without replacement (Figure 4.4.2.2). The first simulation of sample size 319, there are 6 modes in the data set occurring at length classes 75 mm, 115 mm, 150 mm, 165 mm, 195 mm, and 200 mm. The distributions of lengths for all sample sizes, except sample size 50, are similar with modes occurring at similar length classes. However, further analysis is required to determine whether a sample size of 100, which is the smallest of the sample sizes, that is similar to the original data set is sufficient for use of making inference about the population parameters.

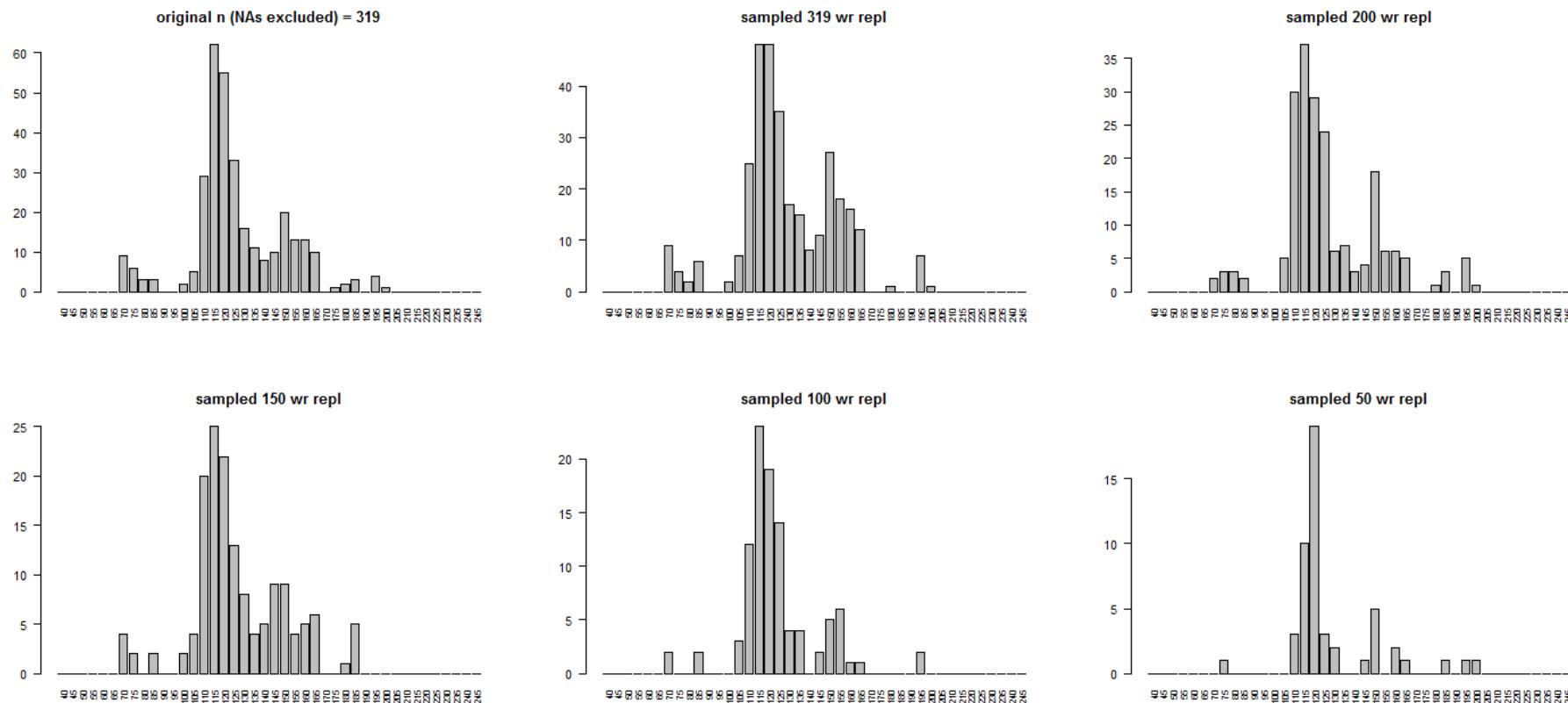


Figure 4.4.2.1 Bootstrap herring data *with* replacement from haul number 21 at IBAS 2015.

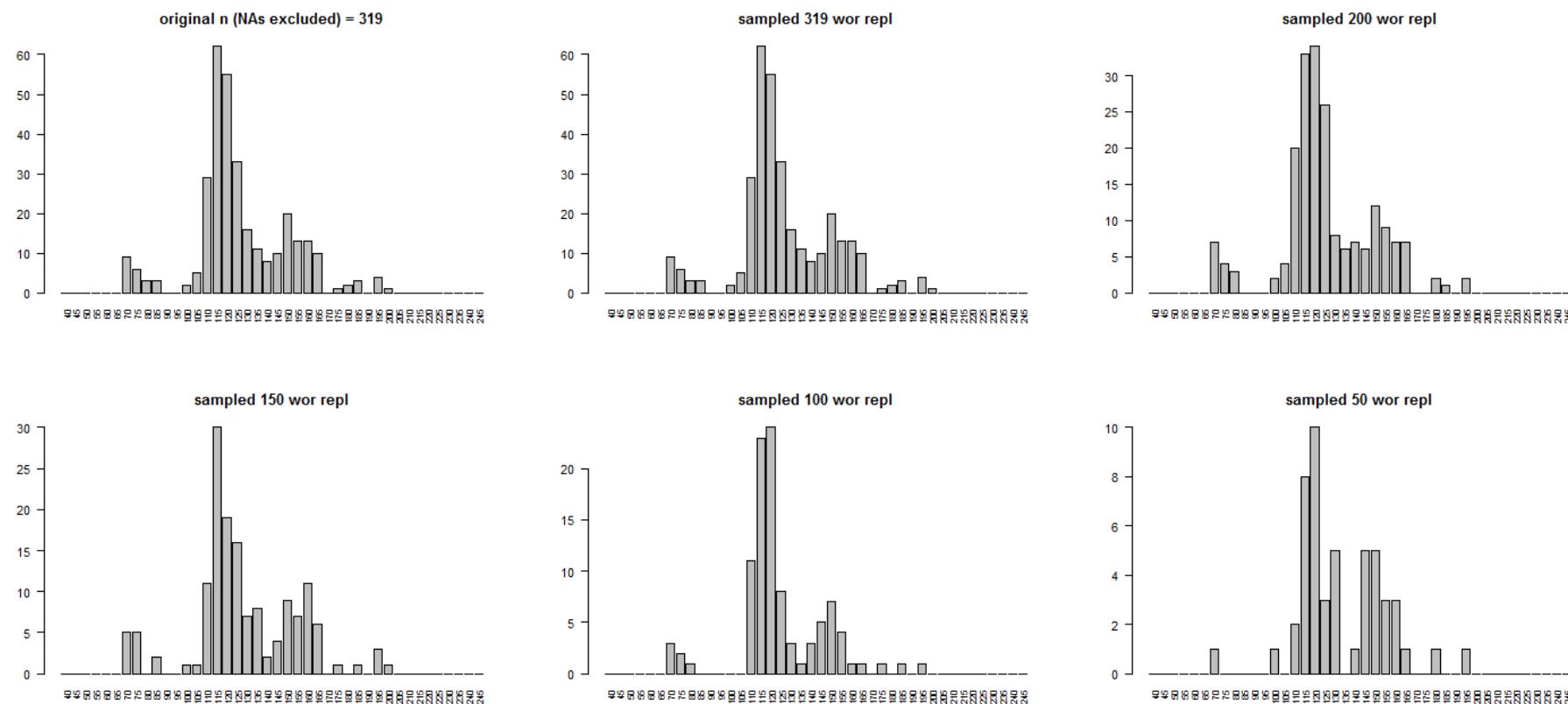


Figure 4.4.2.2 Bootstrap herring data *without* replacement from haul number 21 at IBAS 2015.

4.4.3 Smoothing of the data

Lengths of herring in haul number 21 were used to simulate smoothing the data. Data was smoothed from half cm length classes to one cm length classes. The threshold for modes was set to 1%, that is more than 1% of fish measured must be included to form modes. The original data has four modes occurring at lengths 70 mm, 115 mm, 150 mm and 195 mm. The smoothed data has four modes occurring at lengths 70 mm, 110 mm, 150 mm and 180 mm. (Figure 4.4.3.1)

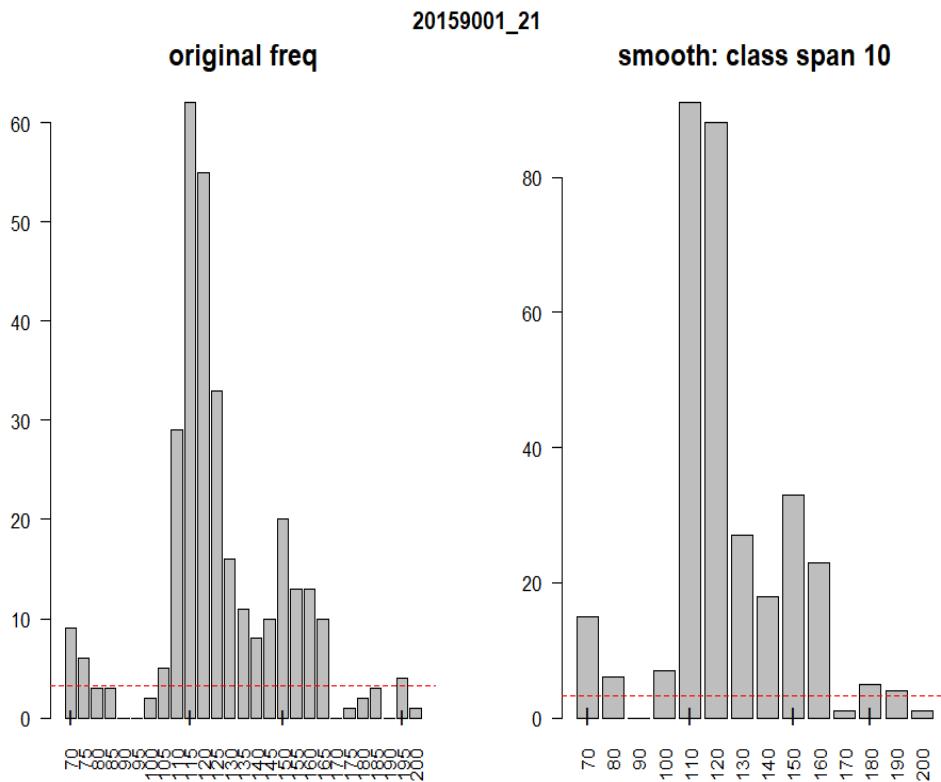


Figure 4.4.3.1 Smoothing of herring data from haul number 21 at IBAS 2015

4.4.4 Data simulation for optimizing sampling effort

Lengths of herring in hauls number 1, 11 and 21 was used to simulate sampling effort optimization. Simulation was done on sample sizes ranging from 30 to 300 with increments of 30. Simulation results are well behaved and suitable for visual inspection from the selected hauls. Visualization of the simulation results of the CV of the mean and MWCV of the sample is shown in Figure 4.4.4.1.

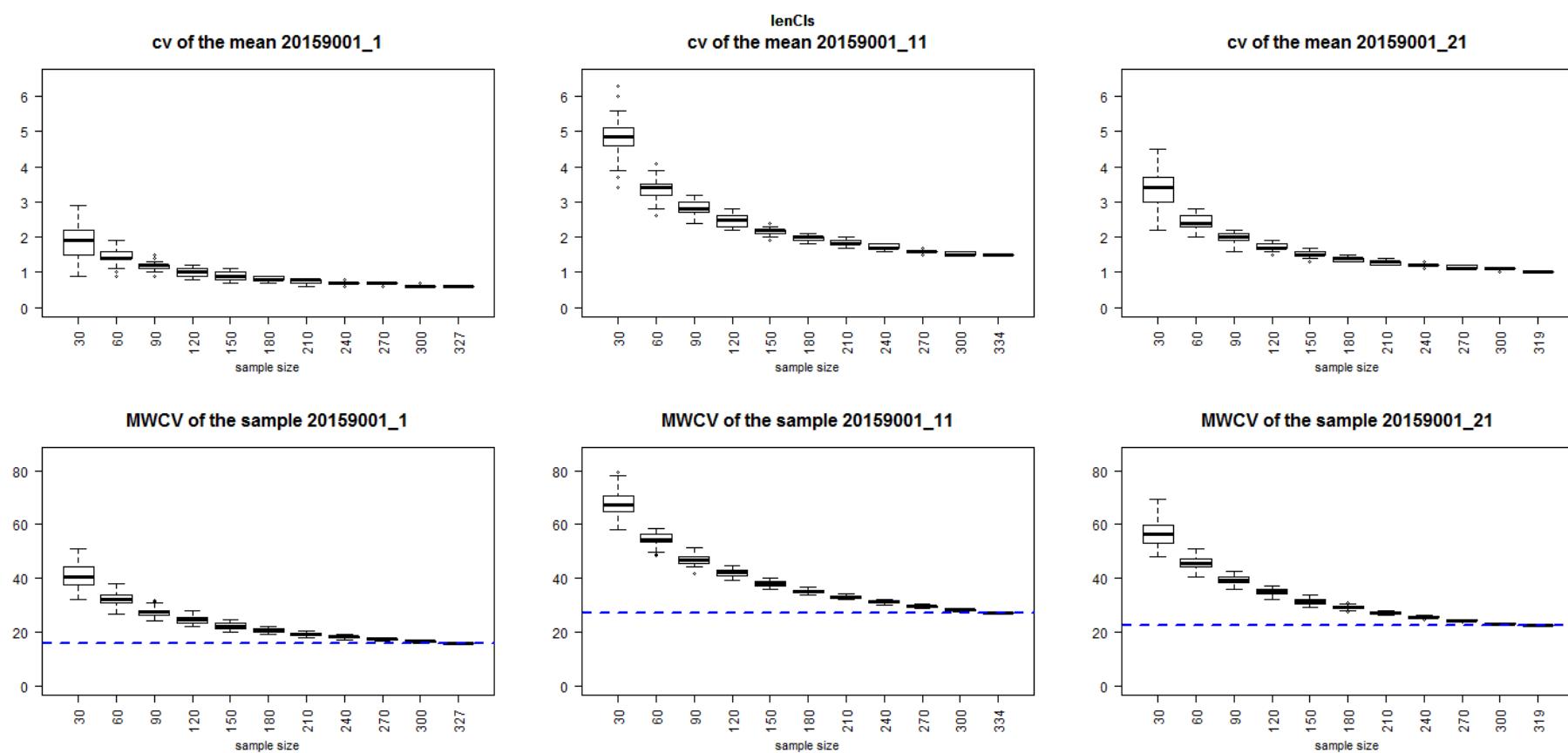


Figure 4.4.4.1 Data simulation of herring length data from hauls number 1, 11 and 21 at IBAS 2015.

4.5 Data Collection Framework (DCF) - Greece 2014–2016

The data used is from the Greek part of the European Data Collection Framework used to collect fisheries data for Greece. Biological data is collected during the survey, and comprises sex, maturity, length and age for various species. The survey takes place on board various fishing vessels which use all main fishing gears (trawlers, purse-seiners, long lines, traps). Data is from all quarters of years 2014 and 2016, since no surveys took place during any other years. Hake was used while testing a version of the sample-level script being developed at WKBIOPTIM. There are 320 hauls in total with various numbers of length measurements in each of them. Age is measured for a portion of each haul individuals. Data is stored in the HCMR's fisheries database and it had to be converted to the ICES RDB format.

Data description:

DCF (Data collection framework);

Country: Greece;

Species: Hake;

Year: 2014–2016

4.5.1 Descriptive analysis

A descriptive analysis of the Greek hake data is below. Figure 4.5.1.1 shows a bar plot of the age distribution of hake. Hake age ranges from 0–10 years and catch rates of younger fishes, are higher compared with older fish.

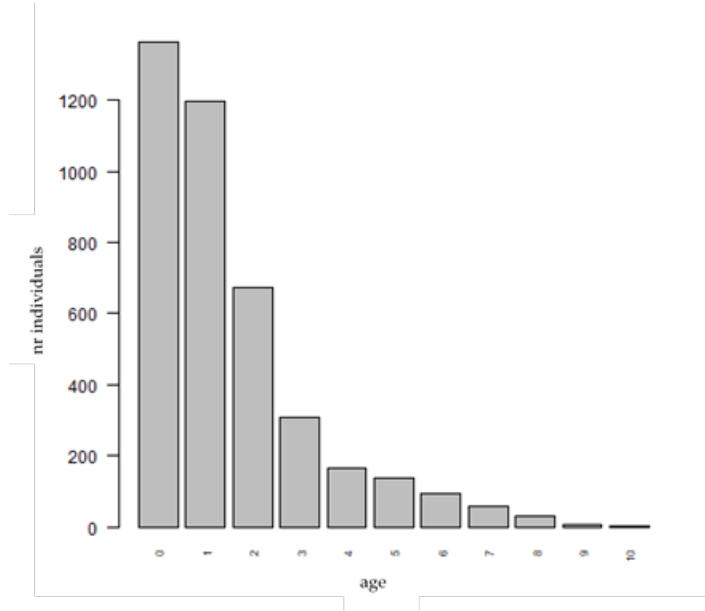


Figure 4.5.1.1 Barplot showing age information for hake species from Greek DCF 2014–2016.

We also examine the length distribution of hake which is given Figure 4.5.1.2. The length distribution of hake ranges from 50 mm–900 mm with at least 10 modes.

For the distribution of sex of hake (Figure 4.5.1.3), and for those individuals for which sex was identified, the data shows that the catch rate for female hake is about twice as high as the male hake, however, at least 1000 fish were unidentified for sex.

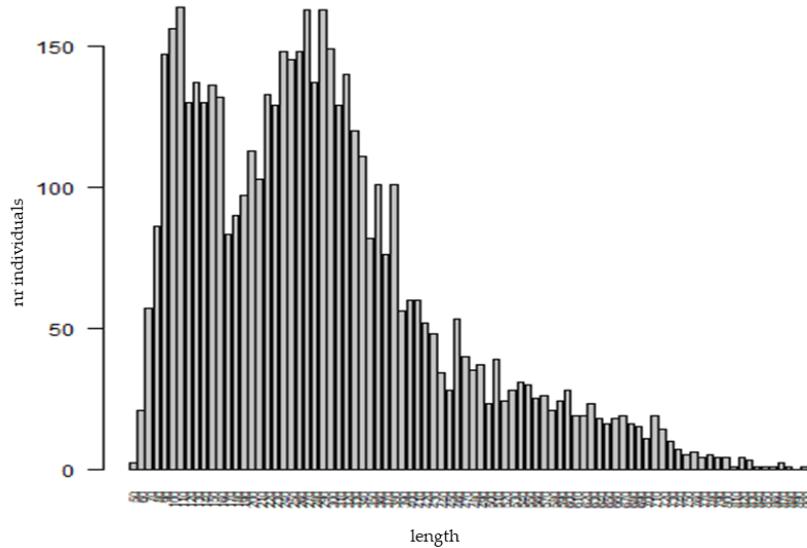


Figure 4.5.1.2 Length distribution of hake species from Greek DCF 2014–2016.

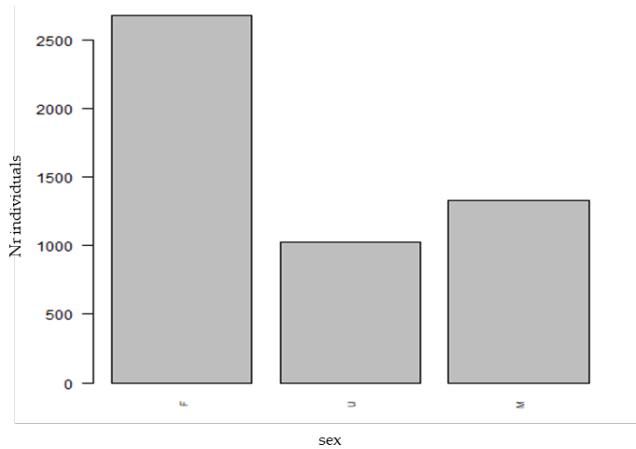


Figure 4.5.1.3 Sex distribution of the hake species from Greek DCF 2014–2016.

4.5.2 Simulation on the entire dataset

As an example, we have used haul number DCF20167008odb6 to simulate lengths of hake in samples of 100, 50 and 25 from an original sample of 137 individuals of hake. Data was simulated with replacement (Figure 4.5.2.1) and without replacement (Figure 4.5.2.2). The first simulation of sample size 137 there are 9 modes in the data set. The distribution of lengths for sample size 100 is the only one that is similar to the original data of size 137. However, further analysis is required to determine whether a sample size of 100, which is the smallest of the sample sizes, that is similar to the original data set is sufficient for use of making inference about the population parameters.

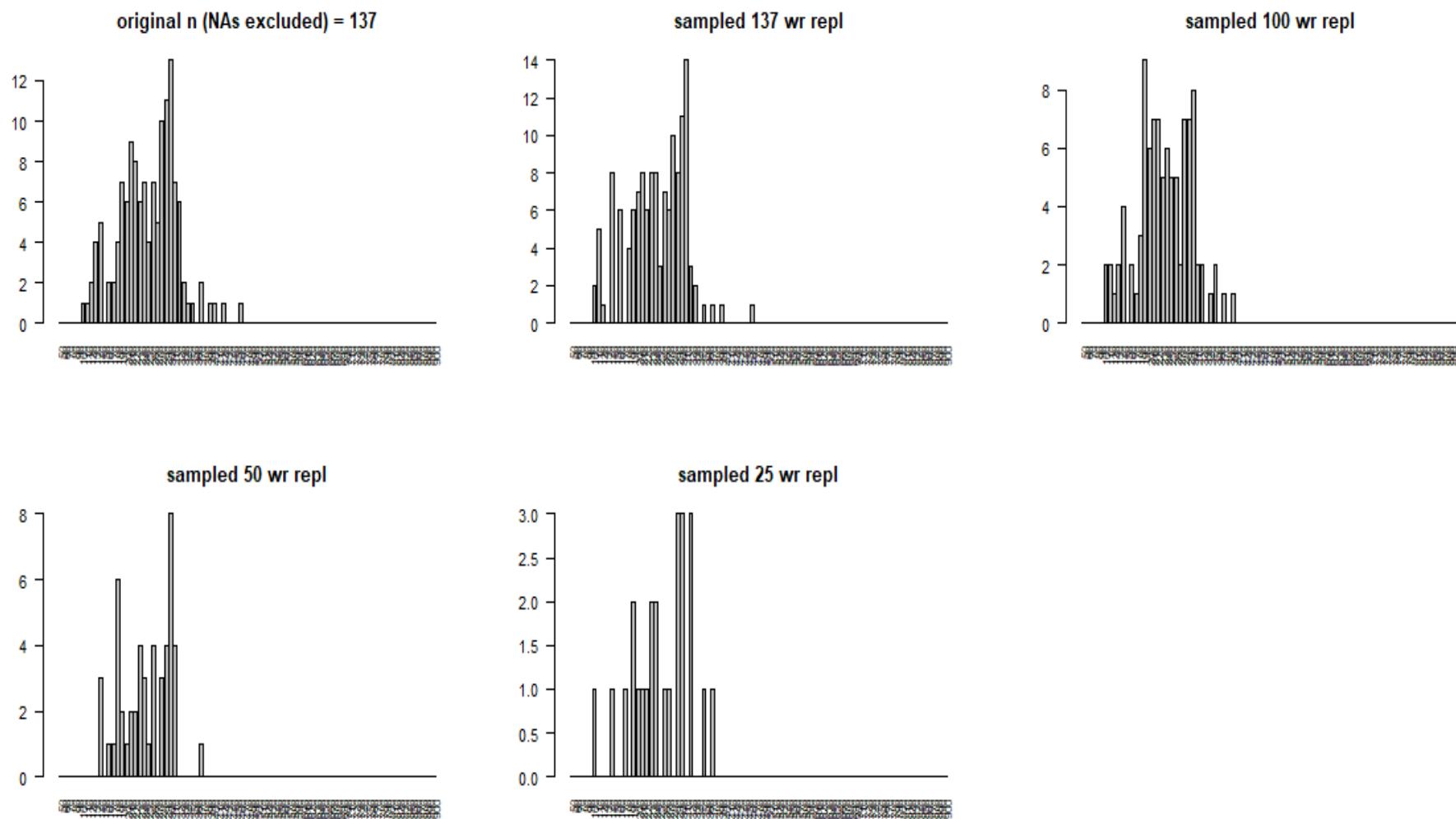


Figure 4.5.2.1 Bootstrap hake data *with* replacement from haul number DCF20167008odb6 from DCF-Greece 2016.

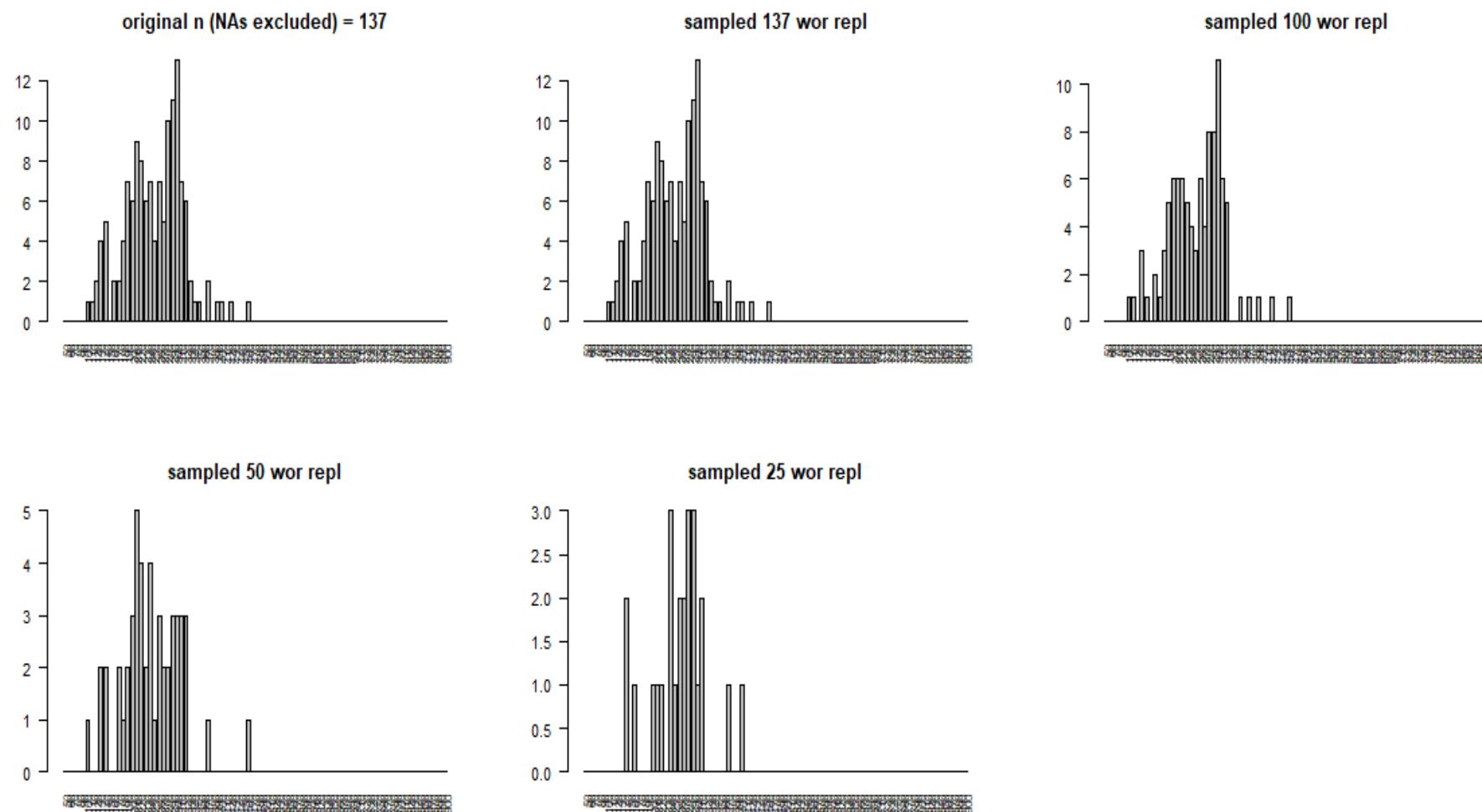


Figure 4.5.2.2 Bootstrap hake data *without* replacement from haul number DCF20167008odb6 from DCF-Greece 2016.

4.5.3 Smoothing of the data

In the original length frequency distribution there were some gaps and some spikes in the plot and particularly at the series edges. Moreover, there are many modes showing in the original like in lengths 140 mm, 190 mm, 270 mm, 300 mm, 370 mm and more. After smoothing these spikes and gaps disappear and the series has only 3 modes at lengths 120 mm, 200 mm and 280 mm. In this way the distribution becomes smoother and continuous (Figure 4.5.3.1).

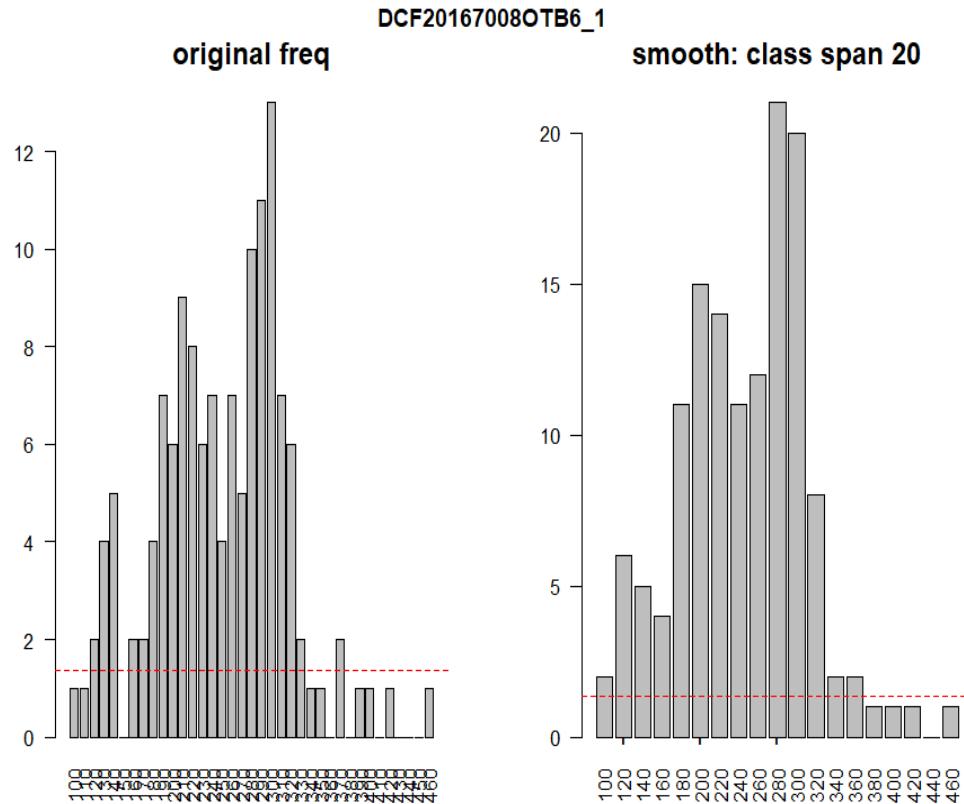


Figure 4.5.3.1 Smoothing of hake data - example.

5 Multi-level analysis

The analyses were performed on data from different countries:

- French data 27.7.d
- Swedish data (2016)
- Italian data in South Adriatic Sea (GSA 18 and GSA 19) for deep-water pink shrimp and blue and red shrimp
- Greek data for Red Mullet in the Aegean Sea (GSA 22)

5.1 French data - ICES 27.7.d

A multi-level analysis was performed on the French data available on the area 27.7.d for the year 2015. This area is of main interest for the WGNSSK because regarding the French fishery 8 stocks estimates have to be provided to this WG each year. Some of these stocks are only located in the area 27.7.d (sol.27.7d, ple.27.7d) and for some others, 90% of the catches are located in this area (mur.27.3a47d...).

5.1.1 Sampling design

Landings are collected using the national fishery declaration system (combining and cross-checking the logbook information and the sells notes), for all the vessels trips. At-sea and in auction samplings cover both large and small vessels related to this fishery. The sampling coverage is given each year in relation to the importance of each fleet segment. The primary sampling unit (PSU) is vessel x trip (as a proxy for a trip). For at-sea observation, a haul (within a trip) is defined as the secondary sampling unit (SSU). Selecting a vessel x trip (PSU) is done by a random draw from a vessel list, linked to a geographical area and a gear use. Haul selection follows some hierarchical rules linked to working conditions (at least 1/3 of the hauls of a metier have to be sampled). Species sampling selection is sampling plan dependent. The sampling of a species is linked to the sampling plan.

5.1.2 Data

Landings and sampling efforts are presented in the Figure 5.1.2.1. Sampling effort cover the whole area spatially and temporally.

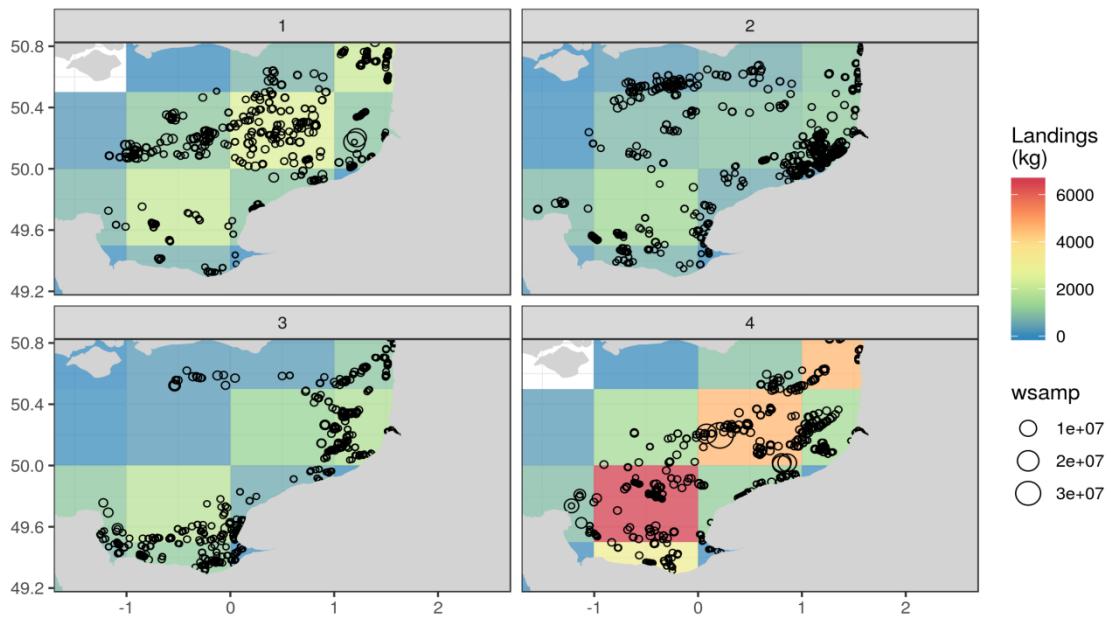


Figure 5.1.2.1 Total landings and sampling position and landing sampled weights for the French data in area 27.7d by quarter and statistical rectangle for the year 2015.

The sampled species by gear in landings during the year 2015 in this area and the related landings are presented in the Figure 5.1.2.2. 67 species are available in the dataset. To simplify the analyses and keep the simulation tractable, (1) gears sampled at least one time in each quarter were selected and (2) 13 species were selected for the simulation if they were sampled in at least 100 trips and if their total landings were exceeding 100 tonnes. The Figure 5.1.2.3 shows the selected species in relation to their total landings and trips sampled.

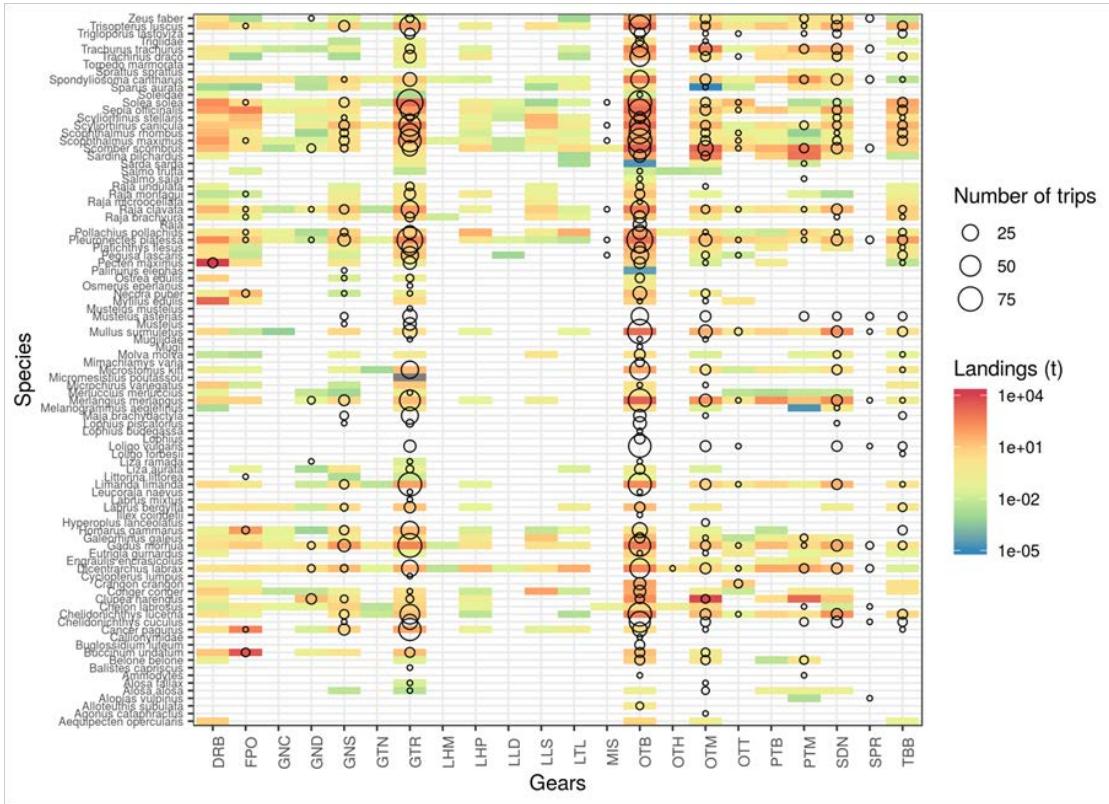


Figure 5.1.2.2 Total landings and number of trips sampled by gear and species for the French data in area 27.7d for the year 2015.

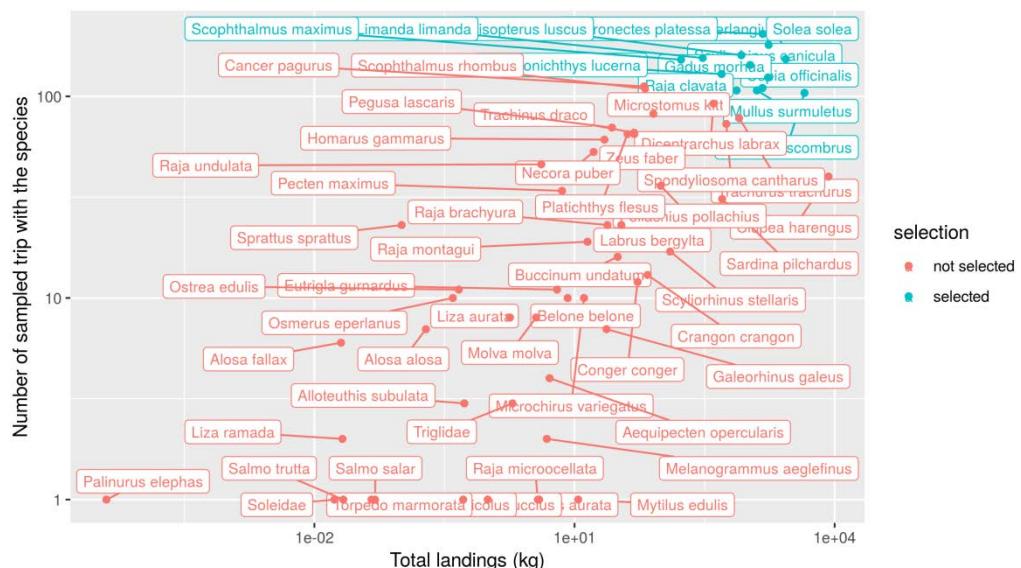


Figure 5.1.2.3 Species represented by the total landings and the number of trips sampled including this species. The blue species were retained in the simulation.

5.1.3 Stratification

The new dataset includes only the trips where at least one of the 13 selected species were sampled. The stratification to build up the length structure was defined following partly the ICES data call standard: the quarter for the temporal stratification, ICES division for the spatial stratification, and gear for the technical stratification. The metier level 6 stratification used to answer the ICES data call can be rather artificial due to the arbitrary labelling of the metier at this level. Moreover, using gear ensures an objective definition (and consequently automatic) of the technical level. Figure 5.1.3.1 shows the selected species represented by the total landings and the number of trips sampled in the stratification. Interestingly, if the sampling coverage seems to be balanced, the selection steps leads to a realistic dataset: some species are not well sampled for some quarter or/and some gear, following the fishing seasonality and the associated sampling plan.

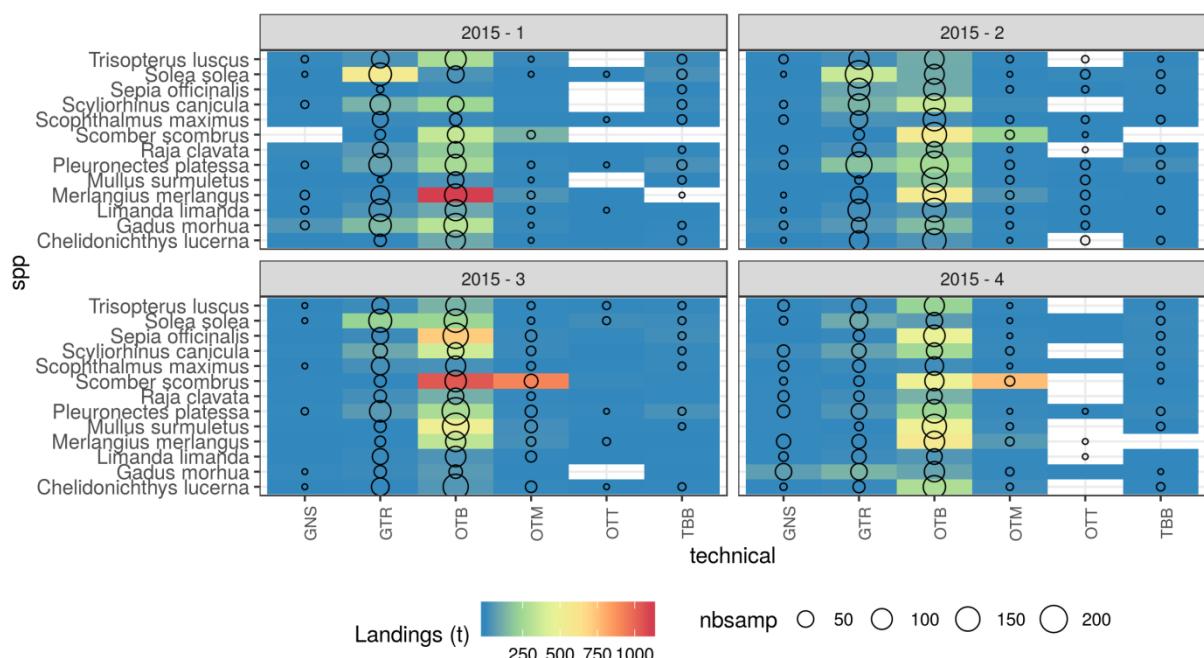


Figure 5.1.3.1 Selected species represented by the total landings and the number of trips sampled (nbsamp) by temporal, spatial (here 27.7.d, not shown) and gear.

5.1.4 Length structure

For the strata described above, the length structures of the 13 species raised to the population using ratio estimator based on the population landings were calculated. The Figure 5.1.4.1 shows the estimates for sole. For each stratum, results were filtered using the same criterion used to answer ICES data call: at least 3 samples and 50 measured fishes have to be present in a stratum to be officially transmitted to ICES. Length structure estimates poorly sampled are then removed from the analyses. This objective threshold removes the need of expert assessment of the quality of the estimates. For instance, the length structure of the OTR gear in the Figure 5.1.4.1 could pass a visual inspection by an expert, but this stratum includes only 2 samples and, consequently, no confidence intervals are available for the length estimates.

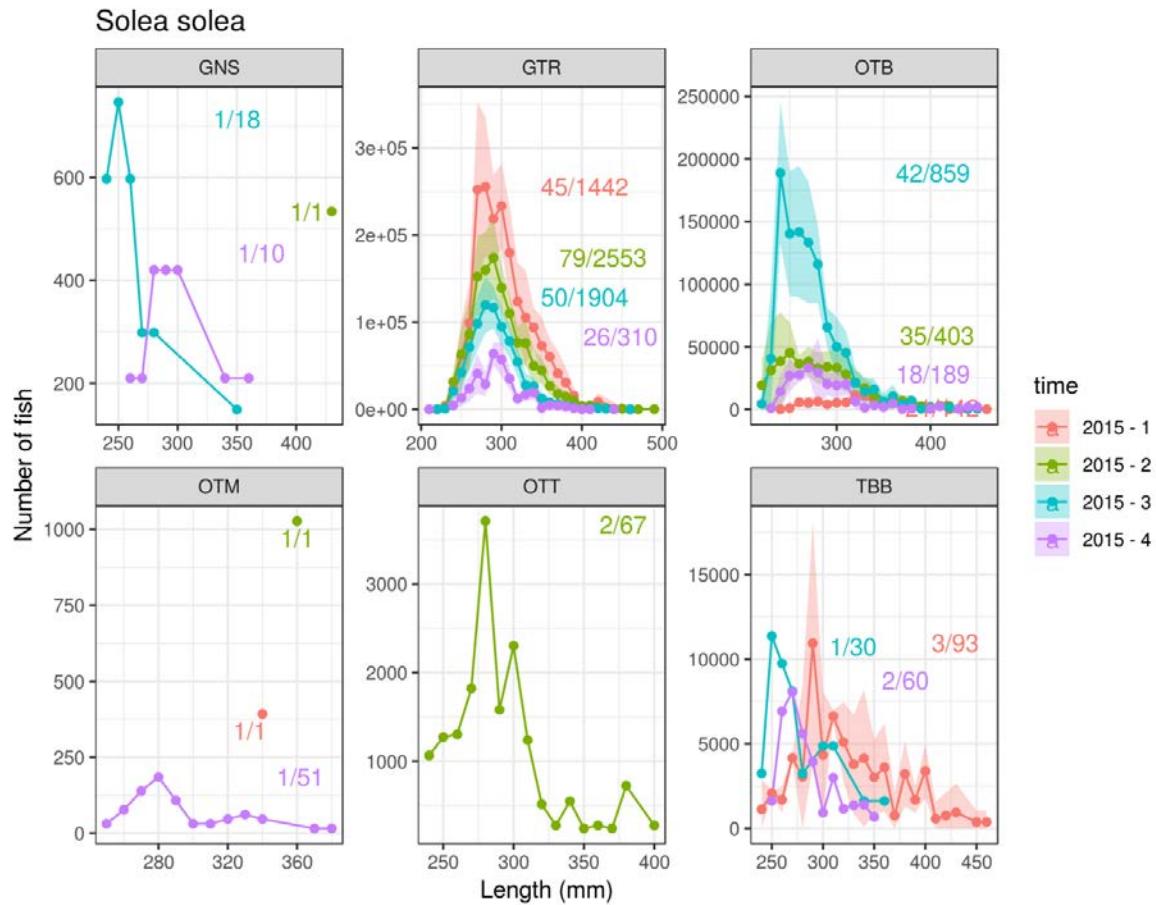


Figure 5.1.4.1 Length structure estimates for Sole in area 27.7.d raised to the population. Annotation give the number of samples/number of measurements used to compute the estimates. Length structure based on low number of samples will be removed from the analysis (for example the OTT length structure in the second quarter, where only 2 samples were used to compute the length distribution). The coloured ribbons show the confidence interval of the estimates.

5.1.5 Quality of the estimates

In the simulation framework, quality of the estimates in each scenario and replicates was assessed using a distribution metric called the Earth Mover's Distance (EMD). The length distribution is estimated based on the new samplings base (this base being defined by the scenario). The results are compared to the actual estimates using the original number of samplings. To compare the 2 distribution, the EMD is used. In statistics, the EMD is a measure of the distance between two probability distributions over a region D (named Wasserstein metric in mathematics, see Rubner *et al.* (1998)). Informally, if the distributions are interpreted as two different ways of piling up a certain amount of fishes over the region D, the EMD is the minimum cost of turning one pile into the other; where the cost is assumed to be the number of fish moved times the distance -the length class interval in this study- by which it is moved. To complete this metrics, the occurrence of the length distribution for a given stratum was computed: it is the number of times, the length distribution can be effectively computed. For some sampling rates, if the stratum is poorly sampled, in some replicates of the scenario the samples are not available, and no information is available for this stratum. Then the coefficient of variation of the length distributions was used to assess estimates stability through the simulation.

5.1.6 Simulation

In order to test the effect of down-sampling the number of trips, an exploratory analysis was performed to investigate the number of trips where each species was sampled for each stratum. The Figure 5.1.6.1 presents these numbers. They were requested to set up efficient simulation: the number of trips resampled should be in line with the number of trips available in each stratum.

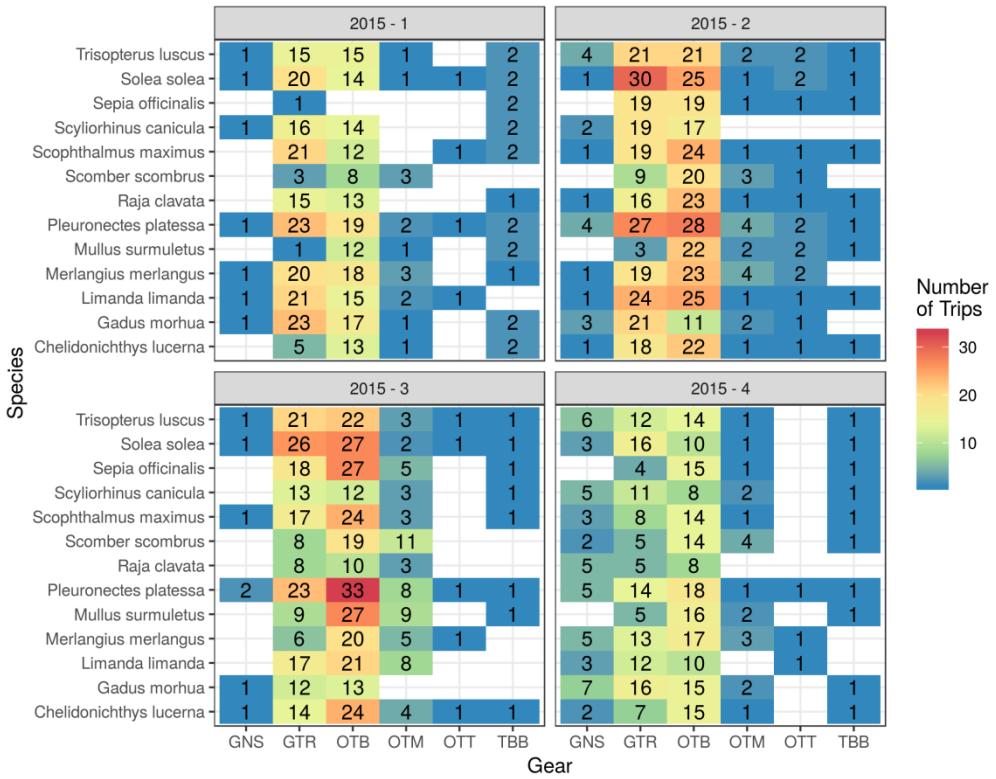


Figure 5.1.6.1 Number of trips where species were sampled by strata.

The simulation followed two scenarios. In the first one, trips were sampled randomly without replacement. In the second scenario, sampling was stratified by stratum (gear, ICES area, quarter): trips belonging to one stratum were selected randomly without replacement, and this selection was repeated for each available stratum, if the number of trips available was enough to select the number of trips requested by the simulation. Each simulation was repeated 100 times.

5.1.7 Results

5.1.7.1 Scenario 1

In this scenario, trips were down-sampled randomly with replacement. The number of trips available in the data being 215, the number of trips varied from 10 to 10 until 210. The estimates of the length distribution raised to the population were then computed using the new pool of trips. Each simulation was repeated 100 times. Figure 5.1.7.1.a shows the evolution of the Earth Mover Distance (EMD) in relation to the number of trips used to compute the estimates, by stratum. As expected, the EMD is increasing with the increase of the number of trips: the length distribution estimates are getting closer to the actual distributions computed with the whole dataset. The occurrence of the species in each stratum is presented on the same figure. Interestingly, if for some species or stratum the number of trips is high enough to provide information even if the original information is degraded (see the plot for OTB for example), for other stratum

and species, the occurrence is very low when the number of trips is low, and for other, the EMD is almost constant over all the simulations. Some stratum and species are poorly sampled (in term of number of trips), and if only 2 or 3 trips are related to a stratum or a species, bootstrapping them do not change the EMD values, but only occurrence values. The figure highlights the fact that if in some strata the number of sampled trips can be down sampled without high consequences in the length estimates for some species (it is the case for plaice caught by OTB in the 4th quarter), for some other species in the same stratum, the estimates can disappear (for *Raja clavata* in the same stratum, the occurrence is under 50% when the number of sampled trips is less than 100).

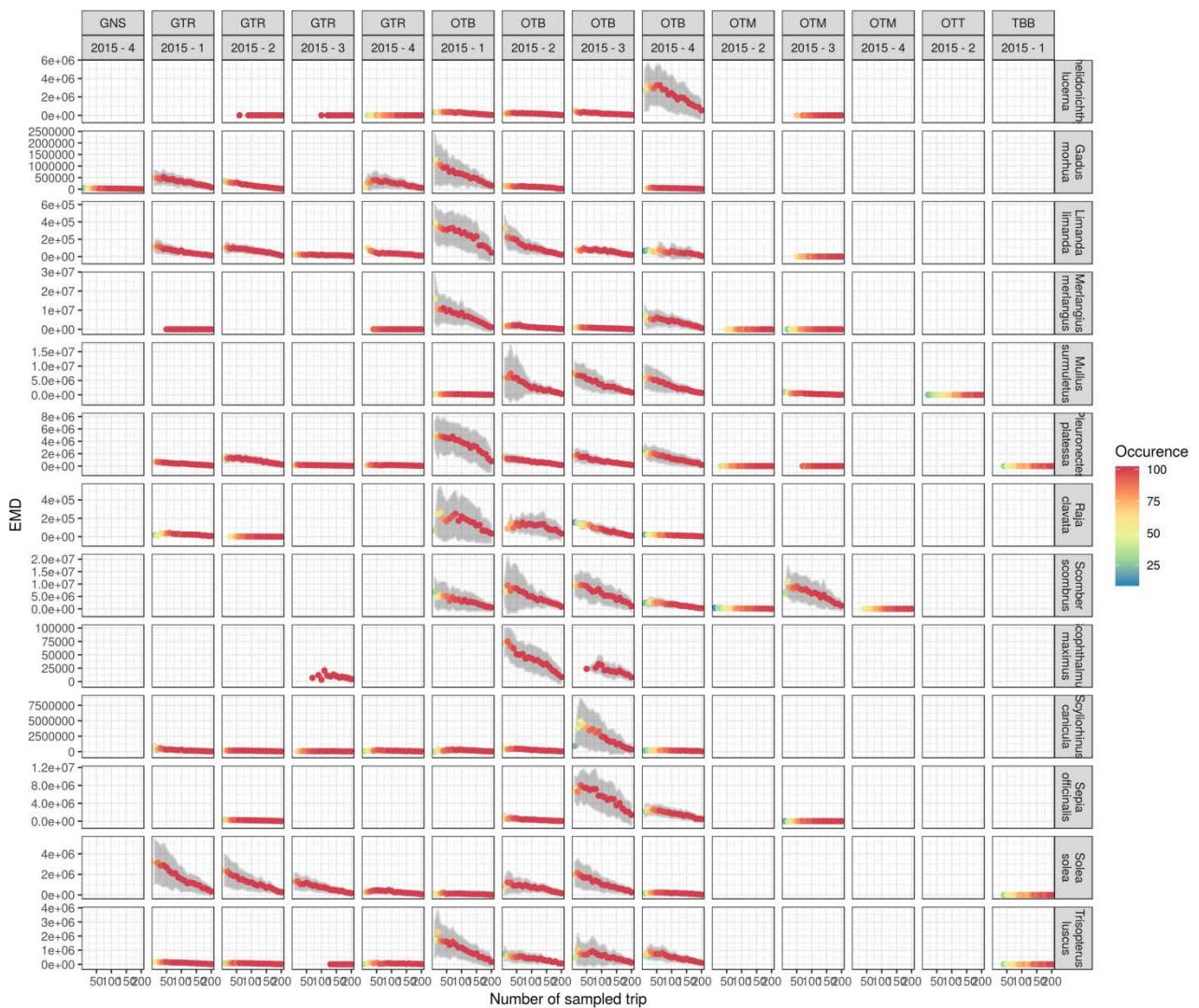


Figure 5.1.7.1.1 Earth Mover Distance (EMD) between the original length distribution and the simulated one, in relation to the number of sampled trips for each stratum (gear, quarter and species). The colour scale is associated with the occurrence of the length distribution in the simulation replicates.

Figure 5.1.7.1.2 summarizes the evolution of the EMD with the occurrence in relation to the number of trips used in the simulation. The averages were computed by gear and quarter for each species. The EMD is decreasing with the number of trips used in the simulation, but occurrence (the number of time by replications a given length distribution for a species and a stratum can be computed) is depending on the species, the quarter and the gear. Here, "rare" species,

like *Raja clavata* or *Mullus surmuletus*, show low occurrences if the number of trips is low. These species are not the main species targeted by the sampling plan.

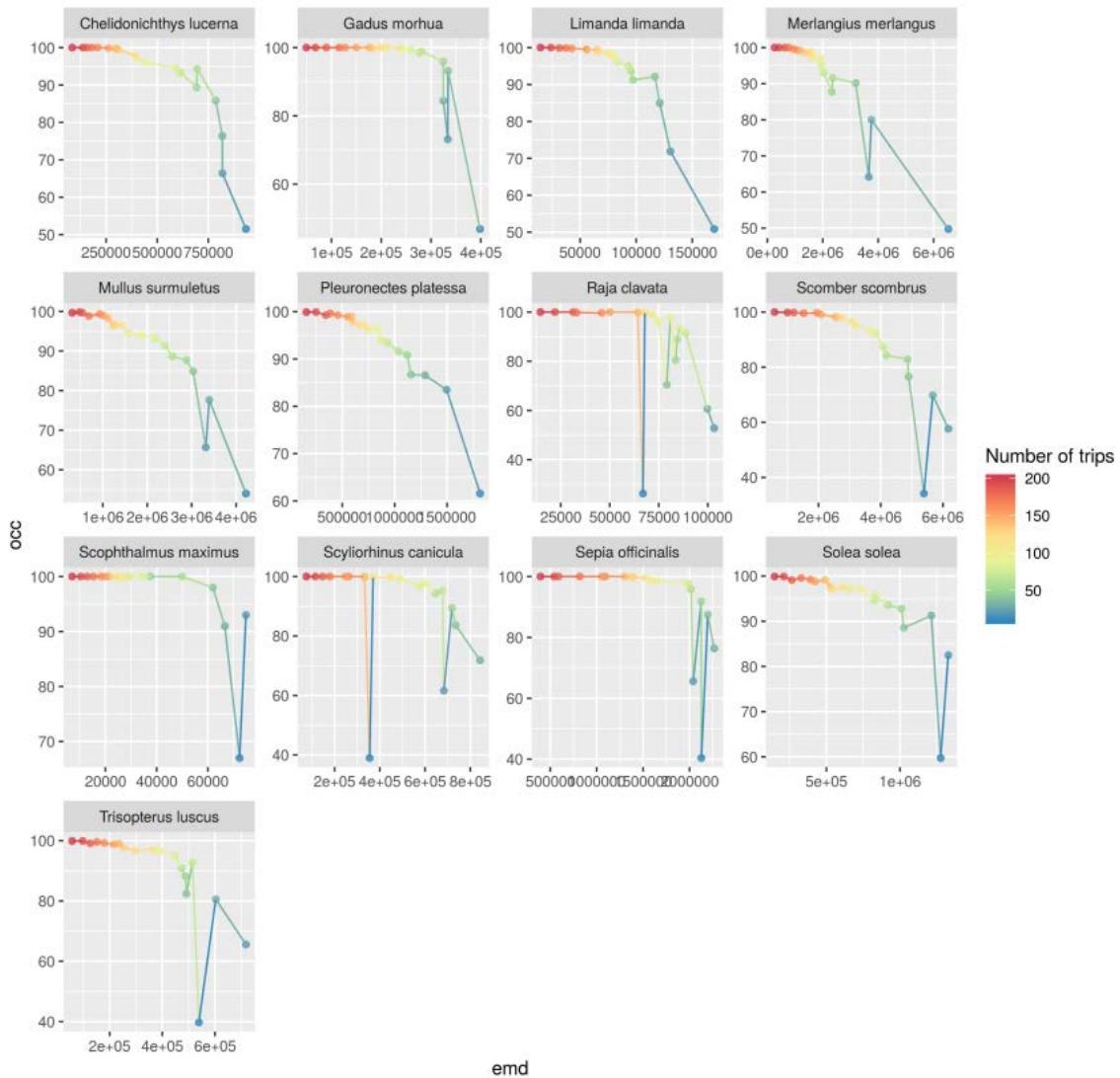


Figure 5.1.7.1.2 Earth Mover Distance (EMD) evolutions according the occurrence of the length distribution in the simulated estimates (occ) and the number of trips used in the simulation (colour scale).

5.1.7.2 Scenario 2

In this scenario, trips were down sampled randomly with replacement by stratum. For each stratum, trips belonging to a given stratum (a combination of gear, ICES area and quarter), were selected randomly and estimates were computed using the new trips subset. The maximum number of trips available in the stratum data being 20, the number of trips varied from 1 to 1 until 20. The estimates of the length distribution raised to the population were then computed using the new pool of trips. Each simulation was repeated 100 times. Figure 5.1.7.2.1 shows the evolution of the Earth Mover Distance (EMD) in relation to the number of trips used to compute the estimates, by stratum. As expected, the EMD is increasing with the increase of the number of trips: the length distribution estimates are getting closer to the actual distributions computed with the whole dataset. The occurrence of the species in each stratum is presented on the same figure. As in the scenario 1, occurrence is low for some stratum when the number of trips is low.

For some other strata, simulations were not available for high number of trips (estimates for OTT, OTB or OTM gears for example).

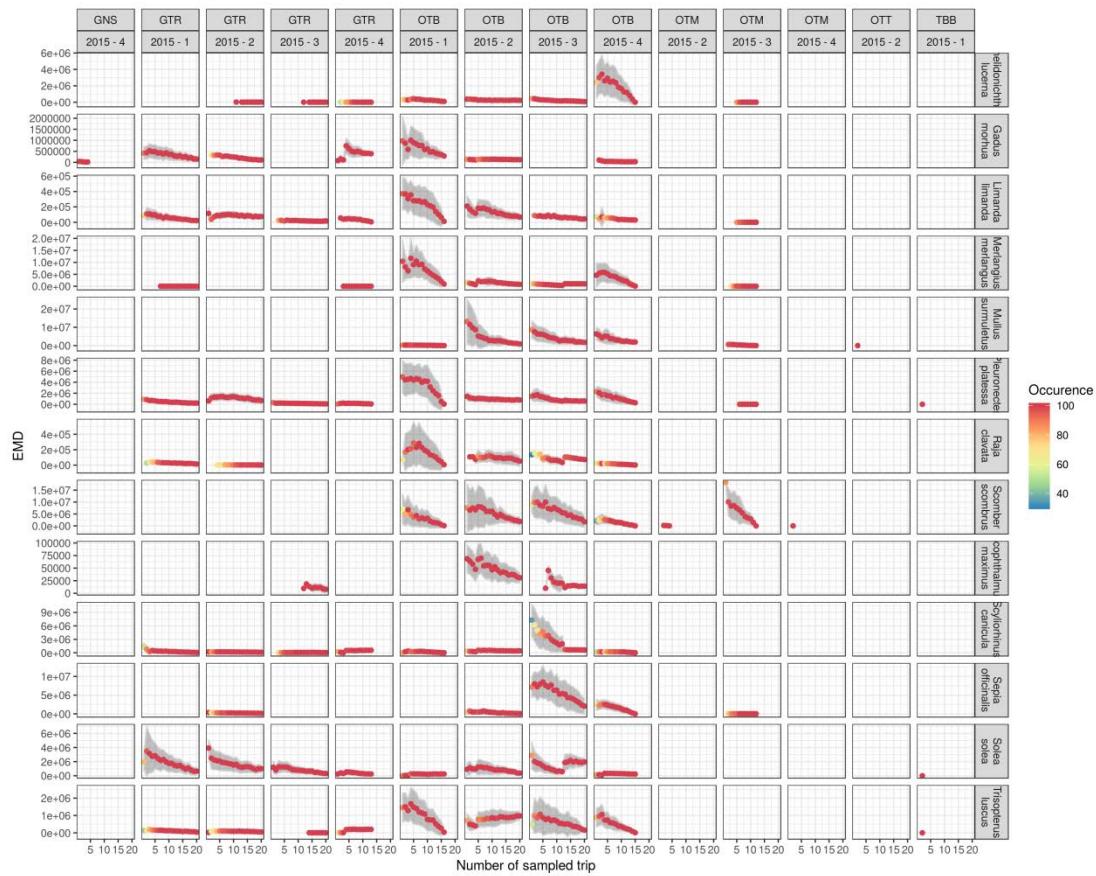


Figure 5.1.7.2.1 Earth Mover Distance (EMD) between the original length distribution and the simulated one, in relation to the number of sampled trips for each stratum (gear, quarter and species). The colour scale is associated with the occurrence of the length distribution in the simulation replicates.

To complete this simulation work, a focus is made on the OTB gear, and the evolution of the coefficient of variation of the length distribution in relation to the number of trips used to compute the estimates by stratum. Figure 5.1.7.2.2 shows the results. In general, the CVs are less variable when the number of trips by stratum increases. This result is very dependent of the species under consideration. For example, in the 4th quarter, if the number of trips needed to reach stability in the CVs is around 10 for plaice. Decisions could be taken to decrease the cost of the sampling scheme related to this species, by selecting this number of trips for this stratum. But regarding *Raja clavata*, such threshold will lead to decrease by 50% the probability to get information on this species for this stratum (occurrence is around 50).

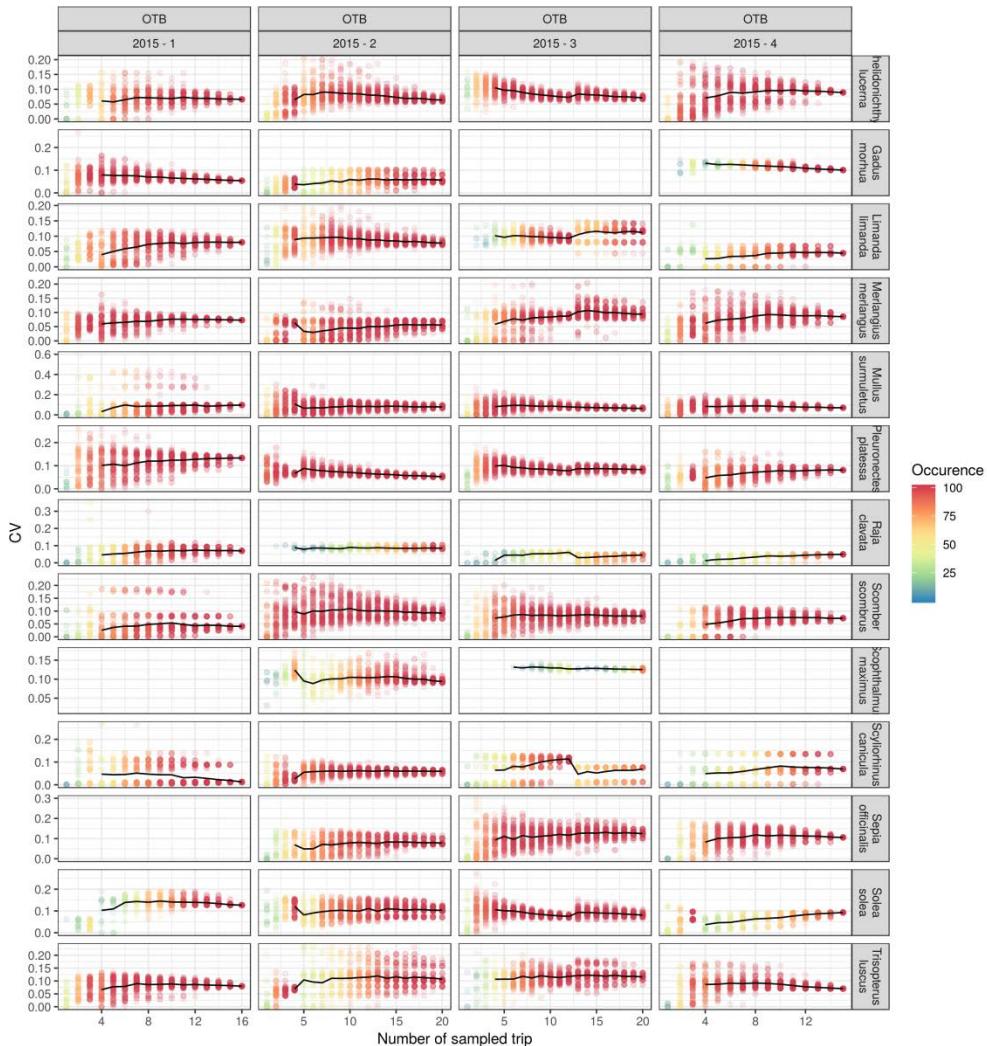


Figure 5.1.7.2.2 Coefficient of variation (CV) of the length distribution in relation to the number of trips used to compute them when down sampling if stratified for the OTB gear.

5.1.8 Conclusion

The work done on this case study highlights the complexity to use bootstrapping simulation to improve the sampling plan (in term of number of trips sampled) in the case of concurrent sampling. If for the main targeted species (landed in high quantity and sampled accordingly), the improvement is still possible, for "rare" species (low landings, low sampling numbers) a slight modification on the number of sampled trips can decrease dramatically the quality of the length distribution estimates, but more importantly can lead to the absence of these species in the samples and in the stratum. This exercise was performed on landings data only. For discards, the results should be even more dramatic (as all the species are recorded). Three metrics were used in this framework: the Earth Mover Distance to assess how the length distribution changes according the actual length distribution, the occurrence to investigate the number of time the species/stratum was seen in the simulation replicates, and the coefficient of variation of the length distribution to assess variability in the estimates. In the framework of the concurrent sampling, occurrence metrics seems to be essential to quantify the loss of information, not seen by the other metrics.

5.1.9 References

Rubner, Y., C. Tomasi, and L. J. Guibas. 1998. "A Metric for Distributions with Applications to Image Databases." IEEE International Conference on Computer Vision, 59–66.

5.2 Swedish data (2016)

Sweden has a wish to get better insight into possible ways to explore ongoing sampling designs for possible areas to optimize and also to get better insight about analytical methods to check if sampling can be optimized.

5.2.1 Sampling design

In Kattegat, area 27.3.a.21 both sea and market sampling is performed. Sampled métiers at focatEu6 and total number of sampled trips in 2016 are shown in Table 5.2.1.1. In this study initially all Swedish data for 2016 was explored but then the focus narrowed down to area 27.3.a.21.

Table 5.2.1.1 Total number of sampled trips in 27.3.a.21 in 2016

focatEu6	Market sampling	Sea sampling
GNS_DEF_120-219_0_0	1	NA
GTR_DEF_120-219_0_0	2	NA
OTB_CRU_90-119_1_120	NA	3
OTB_DEF_90-119_1_120	4	NA
OTB_MCD_90-119_1_120	NA	2
OTB_SPF_16-31_0_0	1	NA
OTM_SPF_32-69_0_0	4	NA
OTT_CRU_70-89_2_35	NA	13
OTT_DEF_90-119_1_120	30	3
OTT_MCD_90-119_1_120	18	4
PTM_SPF_32-69_0_0	11	NA

5.2.2 Data

The data that has been explored is Swedish data from 2016. As the explorative analyses have been conducted using the COST-package, RDB-data was downloaded and then adjusted in order to be transformed into COST-objects. These adjustments have taken a huge amount of effort and time due to unfamiliarity with the COST-package.

5.2.3 Exploratory data analysis

The analysis started with exploratory analyses of the whole Swedish dataset with the R Markdown (Baumer *et al.*, 2014) script presented at the meeting (01_data.Rmd). The script contained information on how to develop files in the COST-format but to understand the COST library

additional information was gathered from several documents such as: COST User Manual V1_1, COST Common tool for raising and estimating properties of statistical estimates and the package descriptions COSTdbe, COSTeda, and COSTcore. The script largely facilitated the production of informative maps.

Landings

Total Swedish landings are mapped per ICES rectangle and quarter in Figure 5.2.3.1.

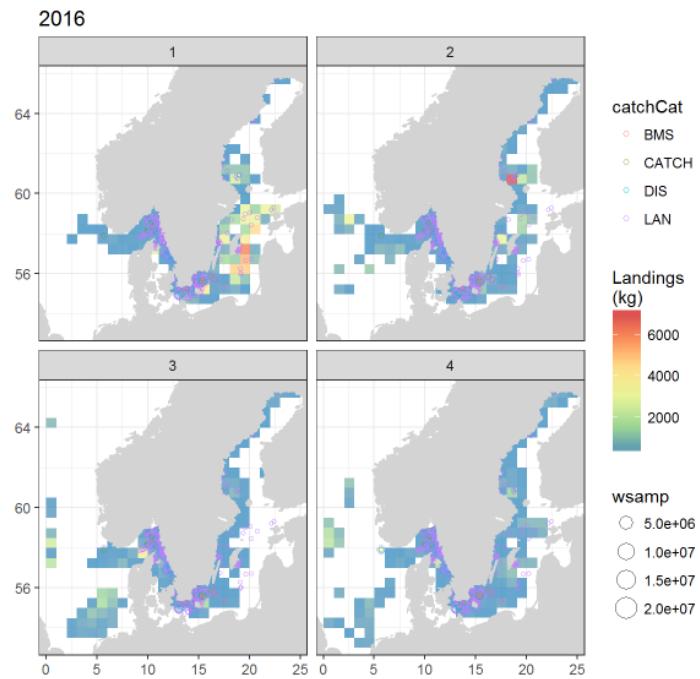


Figure 5.2.3.1 Total Swedish landings per quarter in 2016.

Total landings are also mapped specifically for area 27.3.a.21 in Figure 5.2.3.2.

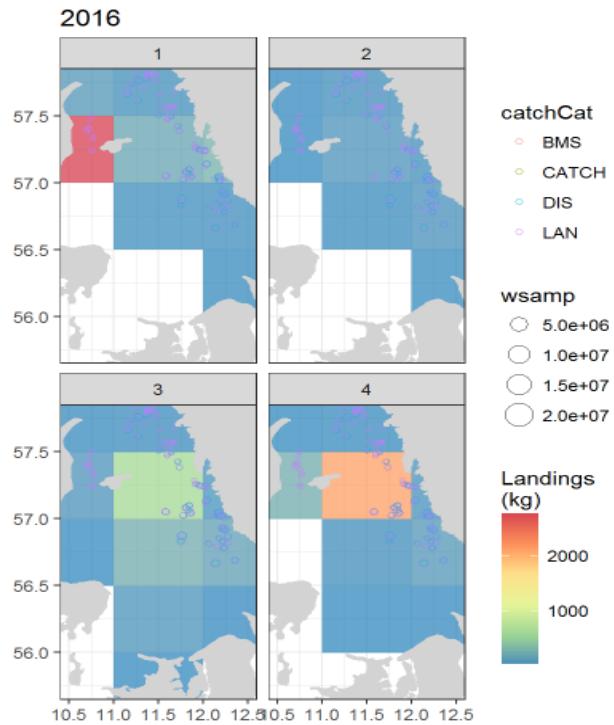


Figure 5.2.3.2 Total landings and sampling weight in 27.3.a.21 in 2016.

Further sub setting was done to map landings and number of samples per species and gear in 27.3.a.21. Only species with more than 10 samples and landings above 10 tonnes are shown in Figure 5.2.3.3.

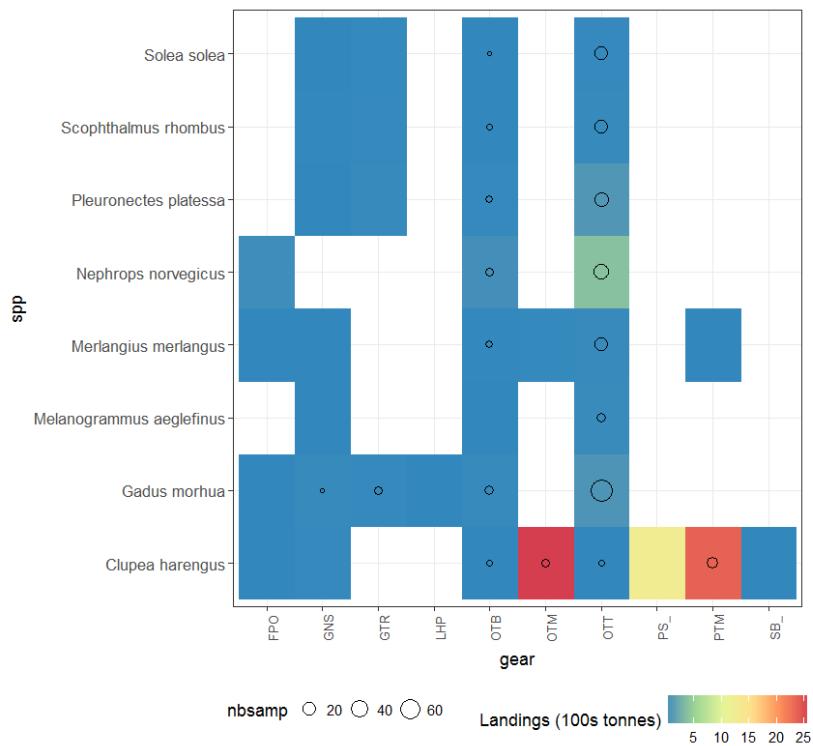


Figure 5.2.3.3 Landings and number of samples (hauls) per species and gear in 27.3.a.21 in 2016. Species selected if nr sampled >10 and landings > 10 tonnes.

Samples and landings as relative rate

Gadus morhua (cod) is a species of particular interest in 27.3.a.21 therefore relative rates of landing weight of cod and number of samples of cod was explored using the tool relativeValue from the COST-library. Number of samples per métier (foCatEu4) follows amount of landings of cod very well as shown in Figure 5.2.3.4.

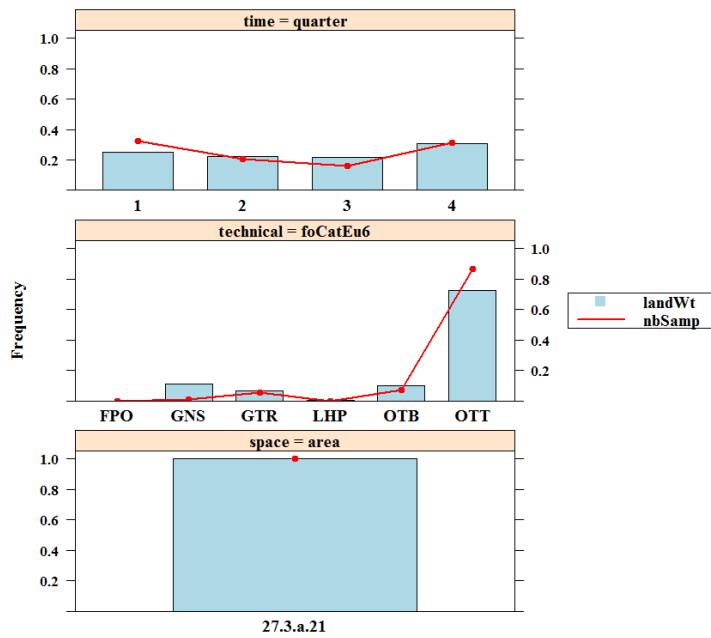


Figure 5.2.3.4 Relative rates of landings and nr samples per quarter, foCatEu4 level and area 27.3.a.21.

Lengths raised to total landings for cod in 27.3.a.21

With the COST-package it is also possible to calculate lengths raised to total landings, which is done for cod in 27.3.a.21 and shown in Figure 5.2.3.5. Variability exists both between metiers and quarters.

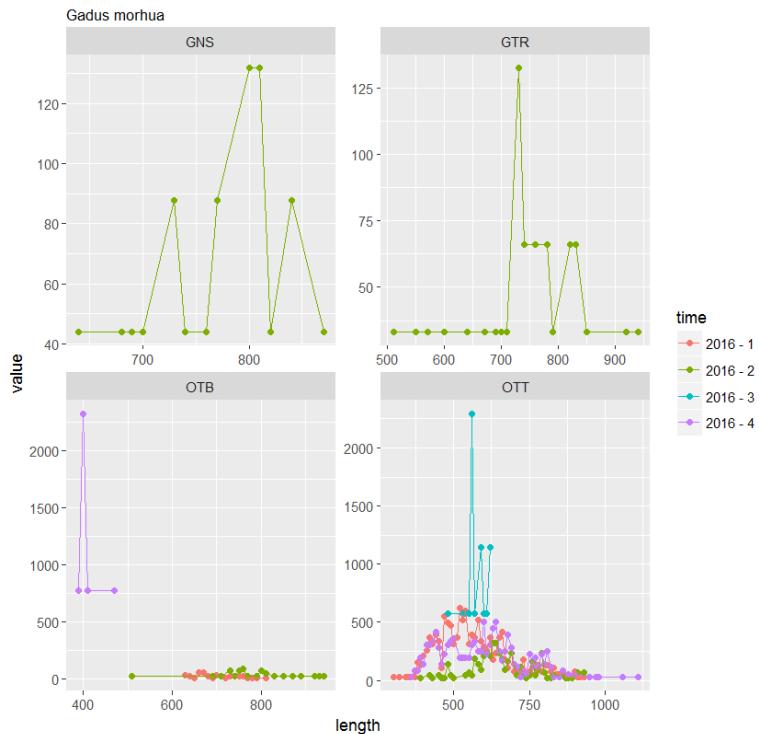


Figure 5.2.3.5 Lengths for cod raised to total landings in 27.3.a.21.

The tool also provides coefficient of variation for the raised lengths. The results show that CV are lowest for the more common lengths as shown for OTT quarter 1 and 4 in Figure 5.2.3.6. When looking at the coefficients of variation it is necessary to remind that GNS only was sampled once and GTR twice.

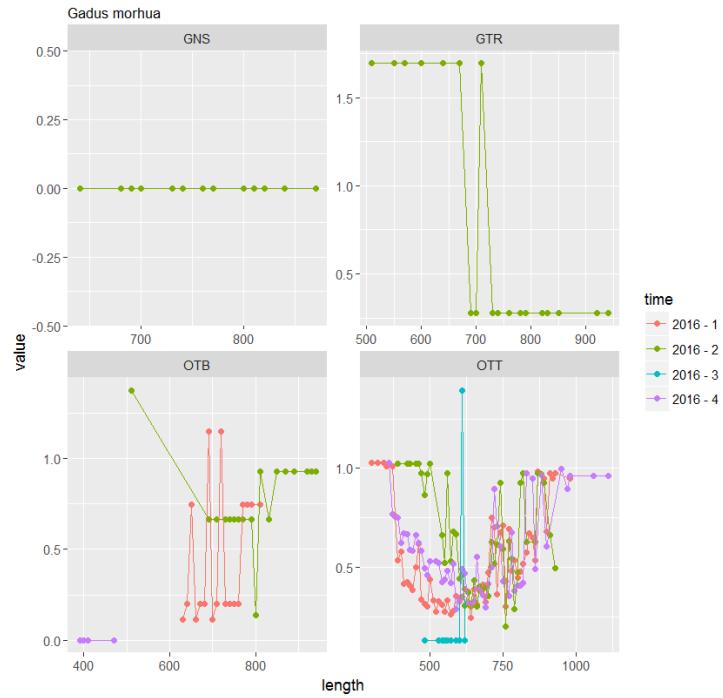


Figure 5.2.3.6 CV for lengths raised to total landings in metiers and quarters in 27.3.a.21.

Number of trips

The number of trips per métier including cod in 27.3.a.21 in 2016 is shown in Figure 5.2.3.7.

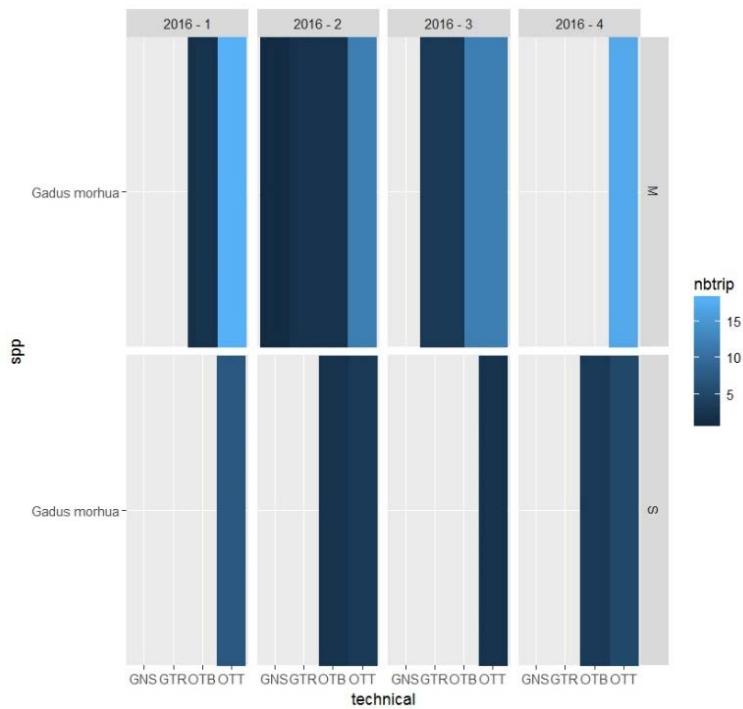


Figure 5.2.3.7 Number of trips with cod, per métier, sampling type and quarter in 27.3.a.21.

5.2.4 Need for optimization (what do you want to test)

Figure 7 does not indicate that optimization (in the form of reducing number of samples) is necessary to do for cod in 27.3.a.21. Most of the métiers and quarters have quite few samples which makes exploratory simulations testing lower number of samples a bit restricted. However, it is still interesting to explore the tools for learning purposes and we focus on the métier with the largest number of samples.

5.2.5 Scenarios

Within the workshop r markdown scripts were provided to test simulations of the data with number of samples and repetitions as optional (02_simulation.Rmd). As we want to pick samples from the total number of samples in 27.3.a.21 a repeated look at the number of actual samples (trips) in 27.3a.21 is shown in Table 5.2.5.1 updated with quarter.

Table 5.2.5.1 Number of samples per metier (foCatEu6), quarter and sampling type.

foCatu6	Market Q1	Sea Q1	Market Q2	Sea Q2	Market Q3	Sea Q3	Market Q4	Sea Q4
GNS_DEF_120-219_0_0	NA	NA	1	NA	NA	NA	NA	NA
GTR_DEF_120-219_0_0	NA	NA	2	NA	NA	NA	NA	NA
OTB_CRU_90-119_1_120	NA	NA	NA	1	NA	NA	NA	2
OTB_DEF_90-119_1_120	2	NA	2	NA	NA	NA	NA	NA
OTB_MCD_90-119_1_120	NA	NA	NA	1	NA	NA	NA	1
OTB_SPF_16-31_0_0	NA	NA	NA	NA	NA	NA	1	NA
OTM_SPF_32-69_0_0	4	NA	NA	NA	NA	NA	NA	NA
OTT_CRU_70-89_2_35	NA	3	NA	3	NA	4	NA	3
OTT_DEF_90-119_1_120	11	1	10	NA	NA	NA	9	2
OTT_MCD_90-119_1_120	7	3	2	NA	1	NA	8	1
PTM_SPF_32-69_0_0	8	NA	NA	NA	NA	NA	3	NA

Sampling is in reality stratified by different metiers (foCatEu6) as they get different catches but for this exercise focus has been métiers at foCatEu4-level. Scenario 1 is that samples are randomly taken from 27.3.a.21 from all quarters and all métiers. Numbers at length are shown for different sample sizes in Figure 5.2.5.1.

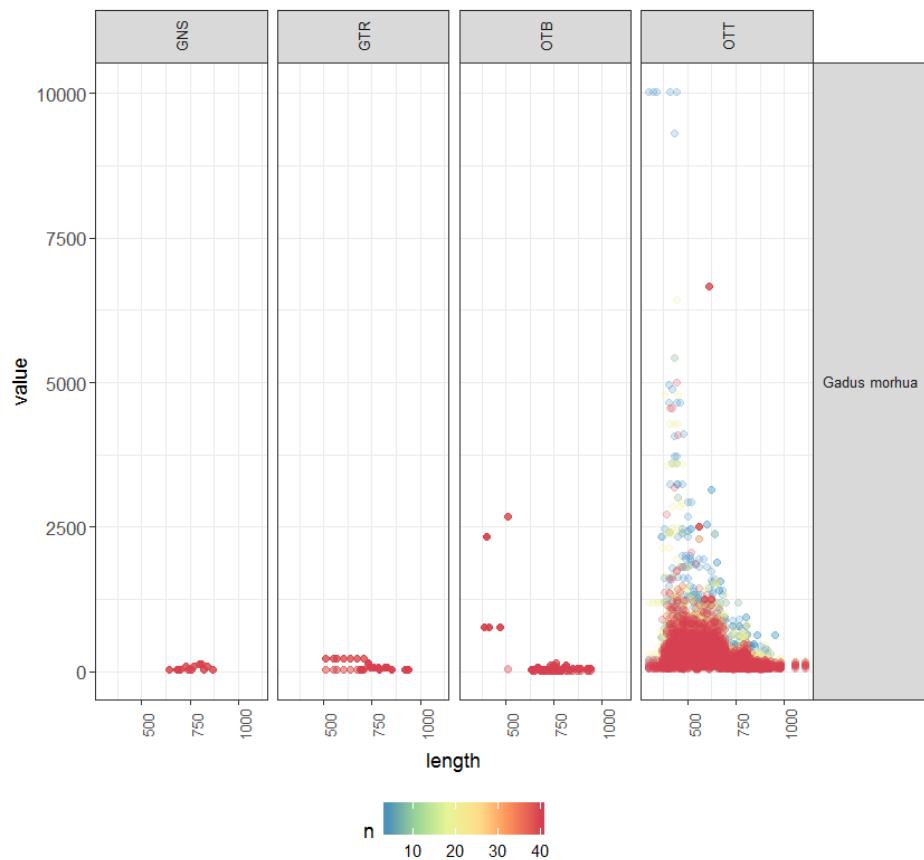


Figure 5.2.5.1 Numbers at length for cod with different sample sizes. Data is raised to total landings.

The script also produces coefficients of variation which are shown in Figure 5.2.5.2.

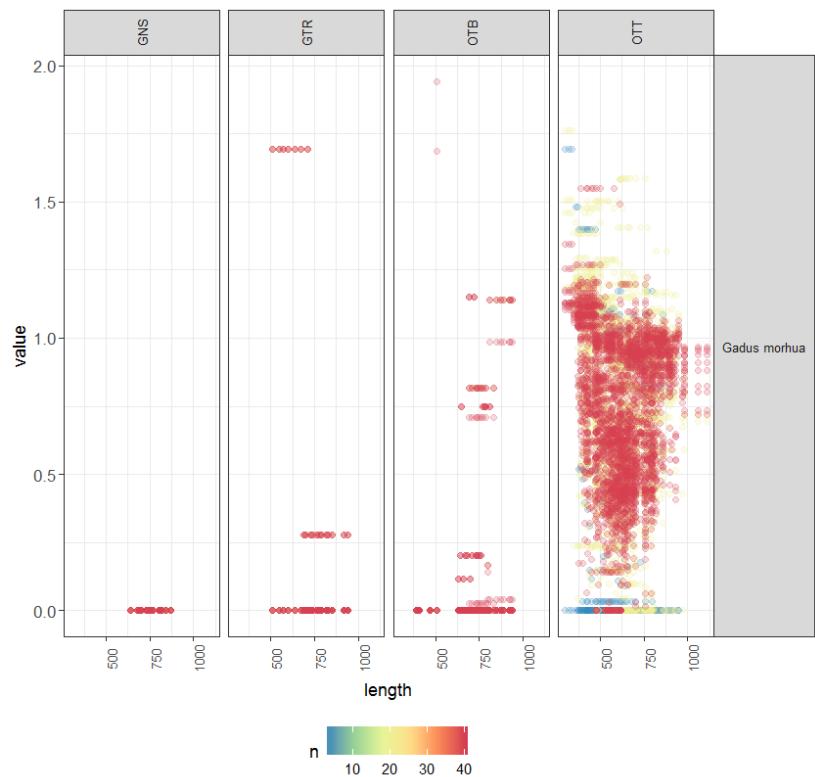


Figure 5.2.5.2 Coefficients of variation for raised lengths with different number of samples.

In scenario two simulations are done only with OTT. A closer look at only cod and specifically OTT is shown in Figure 5.2.5.3.

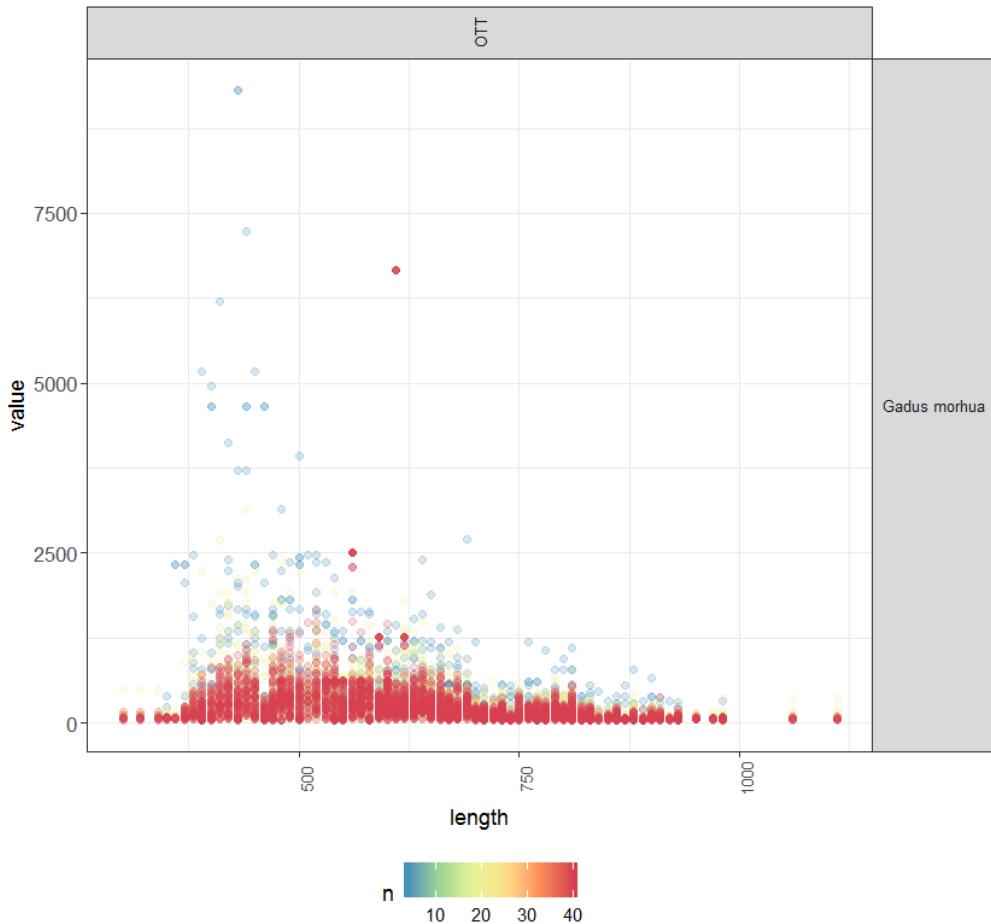


Figure 5.2.5.3 Numbers at length for cod for samples taken from OTT and all quarters in 27.3.a.21. Data is raised to total landings.

Coefficients of variation from raised lengths for cod from samples taken from OTT and all quarters in 27.3.a.21 is shown in Figure 5.2.5.4.

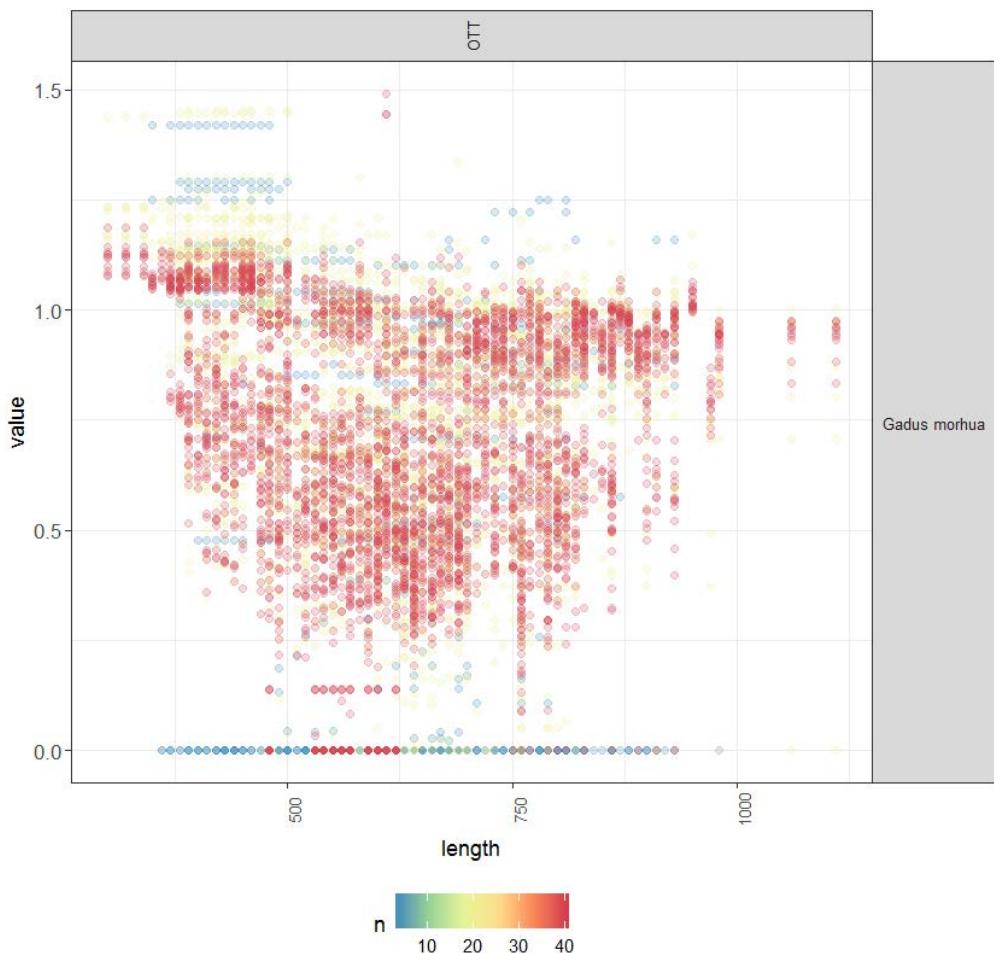


Figure 5.2.5.4 Coefficients of variation for raised lengths with different number of samples taken from OTT and all quarters in 27.3.a.21.

5.2.6 Next steps

The next step for the Swedish part will be to continue stratifying the data even more before running the simulations. The future focus will be to explore other species/area combinations that might have a higher number of samples to be able to see differences between number of samples even clearer.

5.3 Italian data in South Adriatic Sea (GSA 18 and GSA 19) for deep-water pink shrimp and blue and red shrimp

5.3.1 Fisheries description

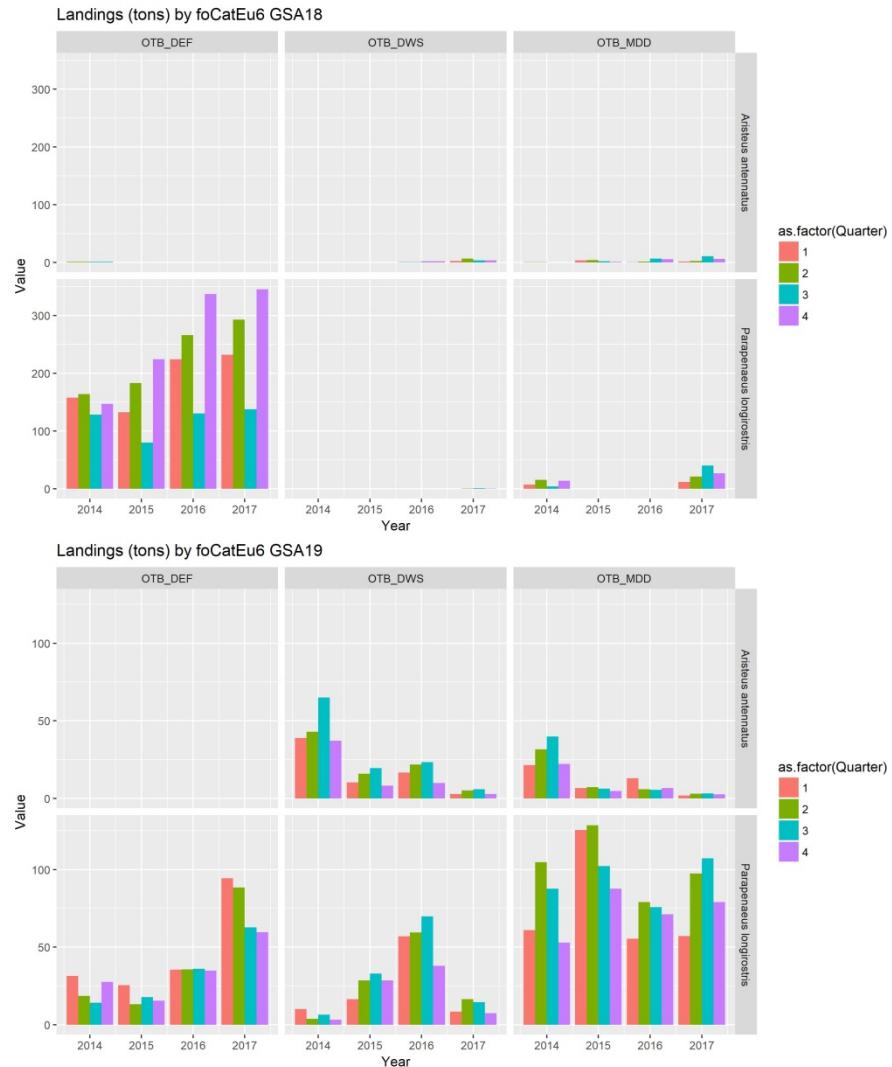


Figure 5.3.1.1 *P. longirostris* and *A. antennatus* in GSA 18- 19 (Italy). Landings by year, quarter and Fishing activity category European lvl 6.

5.3.2 Dataset available

Data have been converted from RCGformat to COST format to be used as input in the SD Tool 2.

Data have also been prepared to be used by the scripts for the analyses of lengths at sample level through **001_prep_data.r** script including the settings for the conversion from CA table to the input format needed for those scripts. The output is saved in the **input_data.rdata** workspace to be used for the analyses.

5.3.3 Data investigation

The analyses on the past data have been performed running the *dataInvestigation.r* script with the following settings.

```

results_path           <- "DPS_ARA_18_19"
cl_all    <- readRDS(paste(results_path, "/input files/costCL.rds", sep=""))
cs_all    <- readRDS(paste(results_path, "/input files/costCS.rds", sep=""))

table_strat_res <- read.csv(paste(results_path, "/input files/05_stratification_results_by_gsa_q_16.csv",
                                     sep=""), sep=";")

target_SPECIES      <- list("DPS"      = "Parapenaeus longirostris",
                           "ARA"     = "Aristeus antennatus")

```

The coverage by GSA, year, fishing activity category European lvl 6 (métier code) and quarter is reported below indicating the number of trips positive to each target species, the number of samples (for the different commercial categories) and the number of measured individual for each species.

Table 5.3.3.1 - *P. longirostris* and *A. antennatus* in GSA 18-19 (Italy). Number of trips positive to the species, samples and measured individuals by GSA, year, fishing activity category European lvl 6, species and quarter.

Area	Year	Metier	Species	TQ1	TQ2	TQ3	TQ4	SQ1	SQ2	SQ3	SQ4	NQ1	NQ2	NQ3	NQ4	annualT	annualS	annualN
GSA18	2015	OTB_MDD	Aristeus antennatus	0	3	2	3	0	0	4	3	0	0	473	241	8	7	714
GSA18	2016	OTB_MDD	Aristeus antennatus	3	3	3	2	0	1	4	222	0	92	193	12	7	507	
GSA19	2014	OTB_DWS	Aristeus antennatus	0	0	4	3	0	0	21	26	0	0	2363	3540	7	47	5903
GSA19	2015	OTB_DWS	Aristeus antennatus	0	4	5	4	0	17	33	19	0	1557	4075	1771	13	69	7403
GSA19	2016	OTB_DWS	Aristeus antennatus	3	3	3	4	20	13	19	25	2506	1463	1517	1858	13	77	7344
GSA19	2017	OTB_DWS	Aristeus antennatus	3	3	1	4	14	18	5	13	1825	1900	1271	2628	11	50	7624
GSA19	2014	OTB_MDD	Aristeus antennatus	0	0	5	4	0	0	4	8	0	0	143	1063	9	12	1206
GSA19	2015	OTB_MDD	Aristeus antennatus	0	5	5	6	0	10	7	17	0	1489	1005	1370	16	34	3864
GSA19	2016	OTB_MDD	Aristeus antennatus	3	5	4	3	9	9	9	10	1098	1036	669	1086	15	37	3889
GSA19	2017	OTB_MDD	Aristeus antennatus	3	3	0	5	7	6	0	17	605	453	0	1196	11	30	2254
GSA18	2014	OTB_DEF	Parapenaeus longirostris	0	0	15	12	0	0	81	66	0	0	17841	13538	27	147	31379
GSA18	2015	OTB_DEF	Parapenaeus longirostris	0	13	9	10	0	83	59	58	0	18864	12988	13140	32	200	44992
GSA18	2016	OTB_DEF	Parapenaeus longirostris	10	13	11	8	56	76	74	44	16198	18490	15269	8154	42	250	58111
GSA18	2017	OTB_DEF	Parapenaeus longirostris	10	9	5	12	57	55	30	86	11687	9922	6755	15360	36	228	43724
GSA18	2014	OTB_MDD	Parapenaeus longirostris	0	0	3	3	0	0	24	26	0	0	4931	6933	6	50	11864
GSA18	2015	OTB_MDD	Parapenaeus longirostris	0	3	2	3	0	28	20	21	0	6491	2795	4514	8	69	13800
GSA18	2016	OTB_MDD	Parapenaeus longirostris	3	3	3	3	18	23	20	21	5786	6841	3848	4603	12	82	21078
GSA18	2017	OTB_MDD	Parapenaeus longirostris	2	2	1	2	10	17	5	12	2740	3547	1141	2586	7	44	10014
GSA19	2014	OTB_DEF	Parapenaeus longirostris	0	0	3	5	0	0	16	43	0	0	5226	13064	8	59	18290
GSA19	2015	OTB_DEF	Parapenaeus longirostris	0	3	4	5	0	18	29	33	0	6756	7161	10184	12	80	24101
GSA19	2016	OTB_DEF	Parapenaeus longirostris	5	3	4	3	28	18	24	24	8488	4516	4768	4958	15	94	22730
GSA19	2017	OTB_DEF	Parapenaeus longirostris	3	2	0	4	21	8	0	29	6344	1508	0	5762	9	58	13614
GSA19	2014	OTB_DWS	Parapenaeus longirostris	0	0	4	3	0	0	8	15	0	0	792	1760	7	23	2552
GSA19	2015	OTB_DWS	Parapenaeus longirostris	0	4	5	4	0	11	20	15	0	1052	2633	2086	13	46	5771
GSA19	2016	OTB_DWS	Parapenaeus longirostris	3	3	3	4	9	9	14	14	2057	1422	1327	1365	13	46	6171
GSA19	2017	OTB_DWS	Parapenaeus longirostris	3	3	1	4	11	12	0	8	1173	2181	0	1058	11	31	4412
GSA19	2014	OTB_MDD	Parapenaeus longirostris	0	0	5	4	0	0	26	22	0	0	6551	4492	9	48	11043
GSA19	2015	OTB_MDD	Parapenaeus longirostris	0	5	5	6	0	24	26	33	0	6261	6133	9877	16	83	22271
GSA19	2016	OTB_MDD	Parapenaeus longirostris	3	5	4	3	14	27	24	10	2589	5571	4948	1866	15	75	14974
GSA19	2017	OTB_MDD	Parapenaeus longirostris	3	3	0	5	11	17	0	24	2053	2817	0	3998	11	52	8868

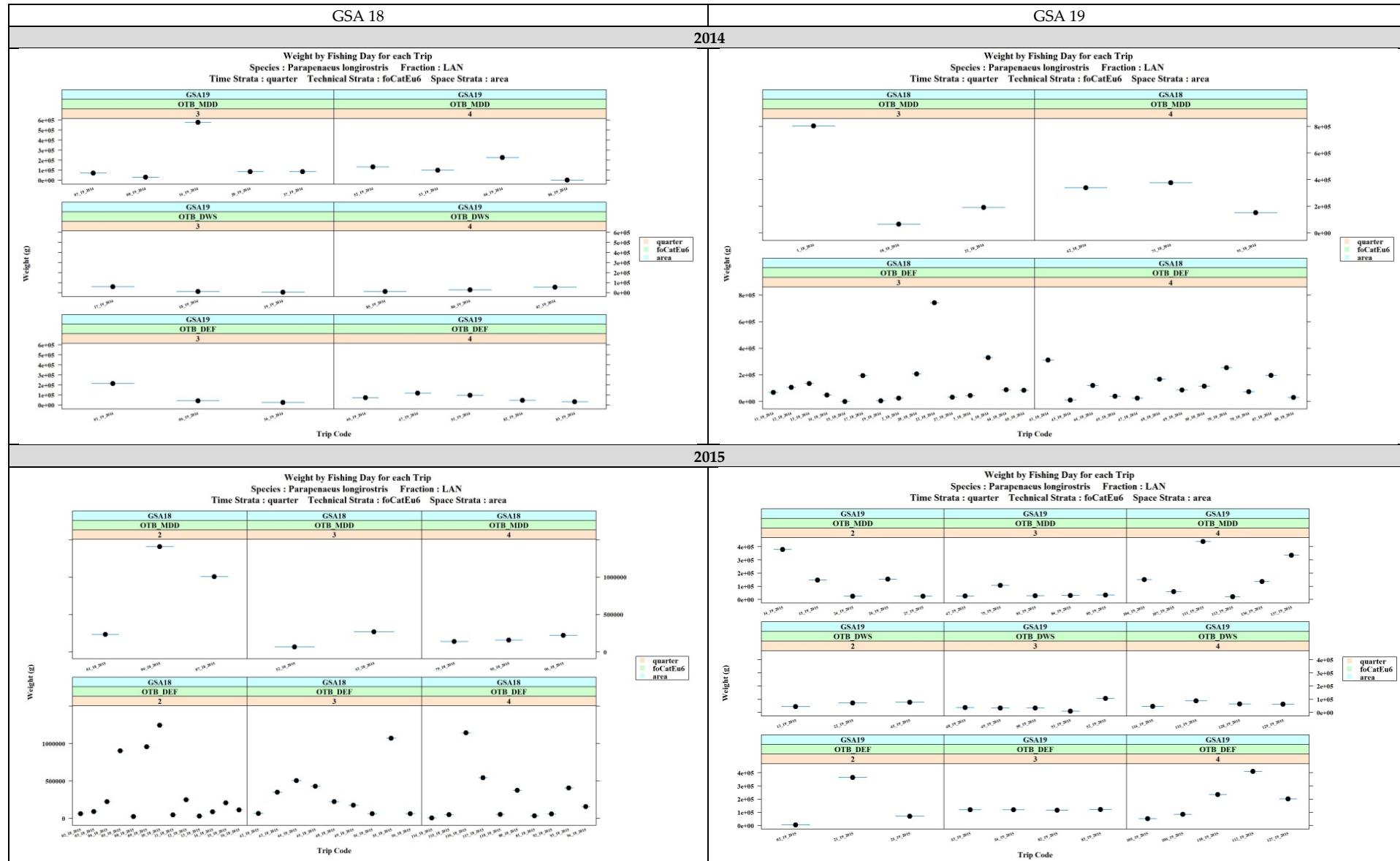


Figure 5.3.3.1 *P. longirostris* in GSA 18-19 (Italy). Sample weight by trip, fishing activity category European lvl 6 and quarter.

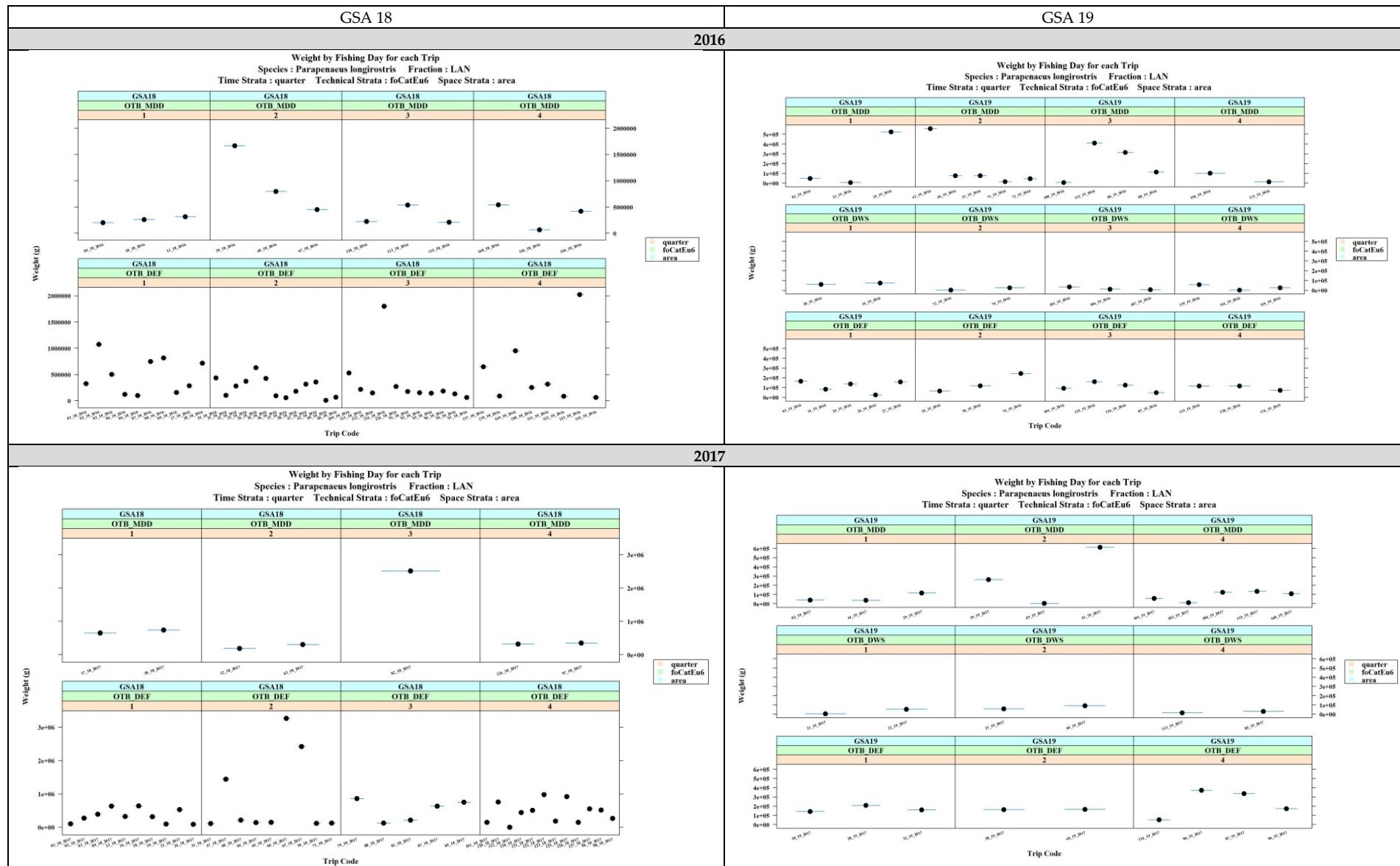


Figure 5.3.3.1 (cont.) *P. longirostris* in GSA 18-19 (Italy). Sample weight by trip, fishing activity category European lvl 6 and quarter.

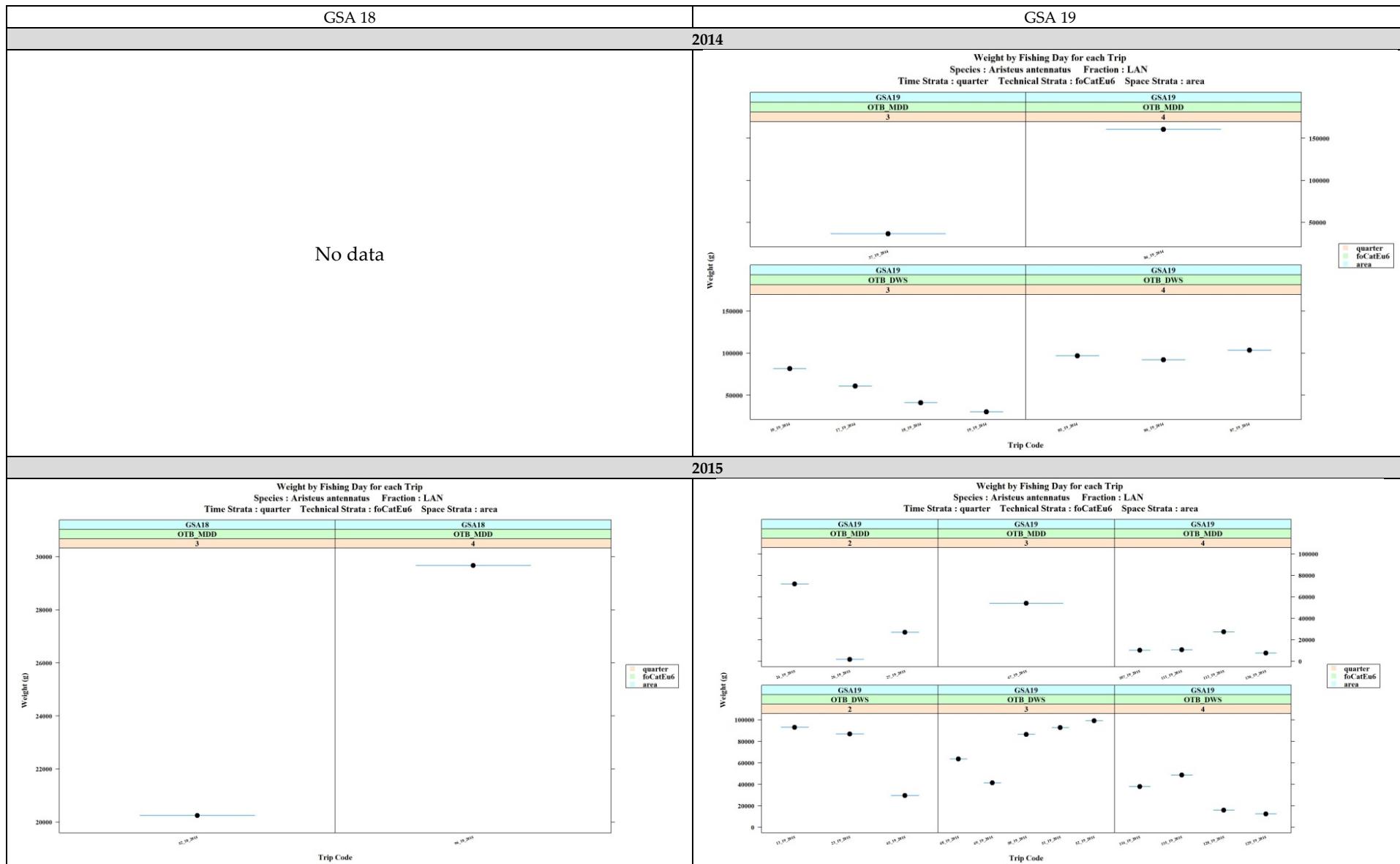


Figure 5.3.3.2 (cont.) *A. antennatus* in GSA 18-19 (Italy). Sample weight by trip, fishing activity category European lvl 6 and quarter.

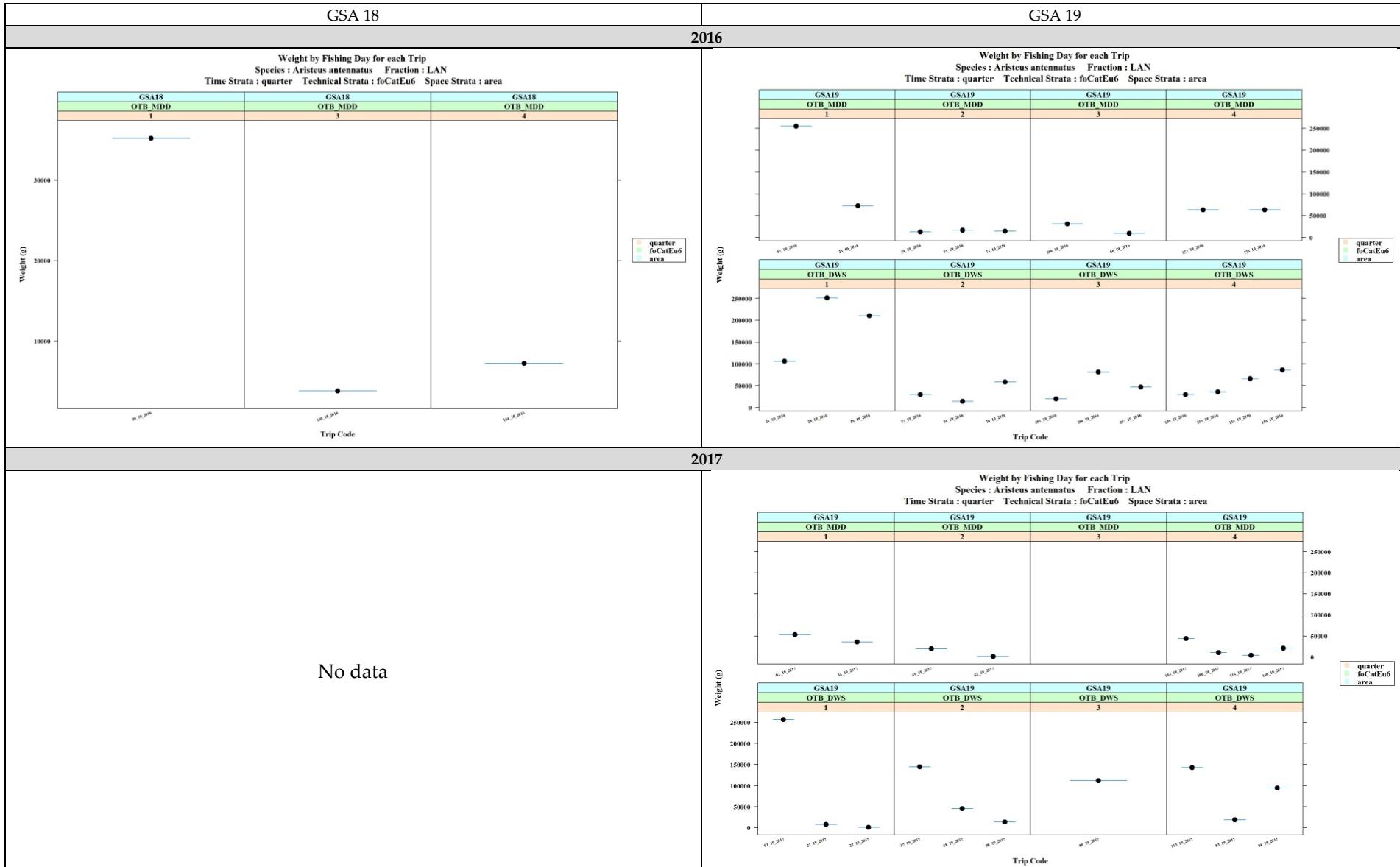


Figure 5.3.3.2 (cont.) *A. antennatus* in GSA 18-19 (Italy). Sample weight by trip, fishing activity category European lvl 6 and quarter.

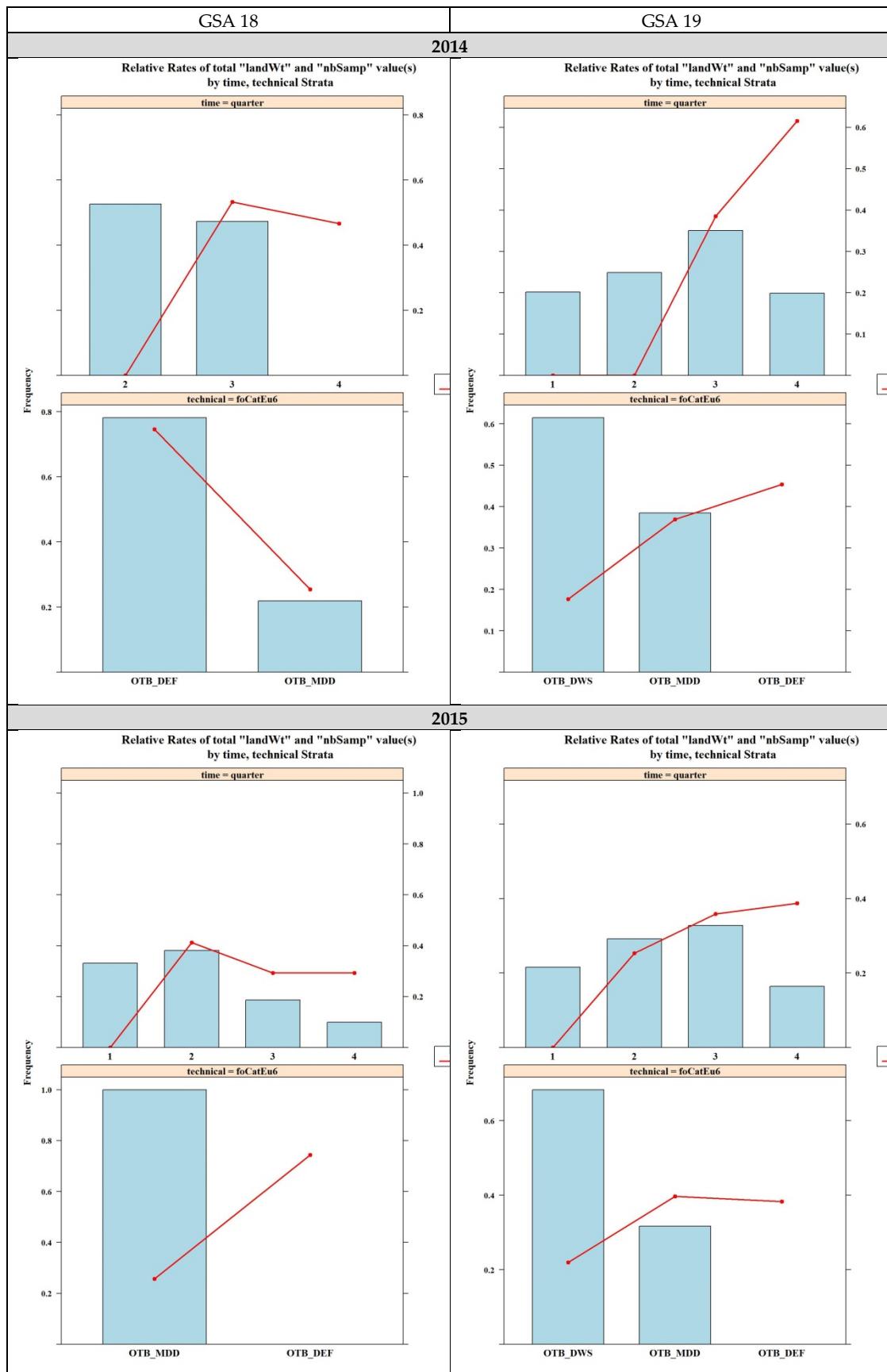


Figure 5.3.3.3 *P. longirostris* in GSA 18-19 (Italy). Relative importance of sampling compared to the relative importance of landing by fishing activity category European lvl 6 and quarter.

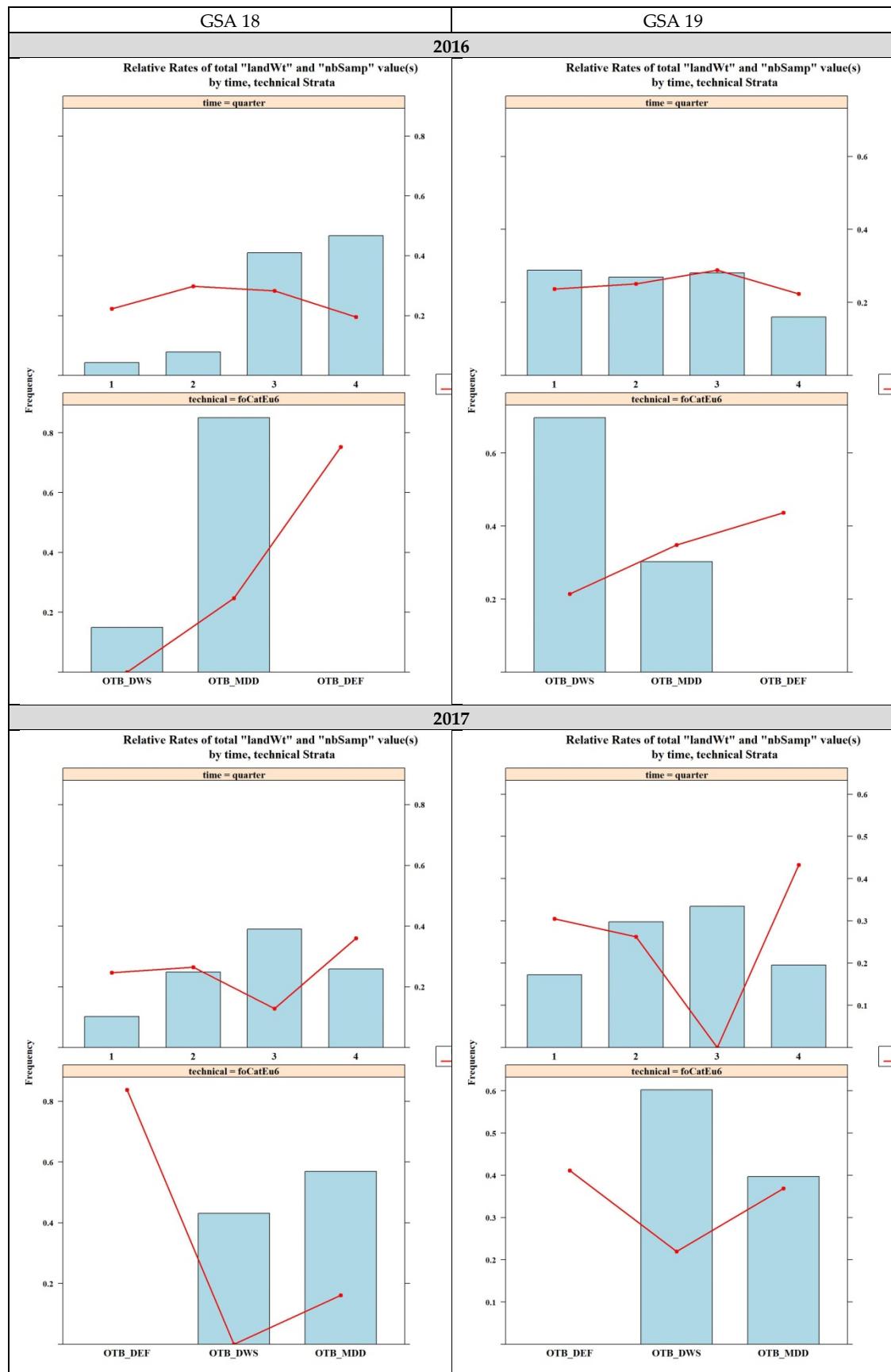


Figure 5.3.3.3 (cont.) *P. longirostris* in GSA 18-19 (Italy). Relative importance of sampling compared to the relative importance of landing by fishing activity category European lvl 6 and quarter.

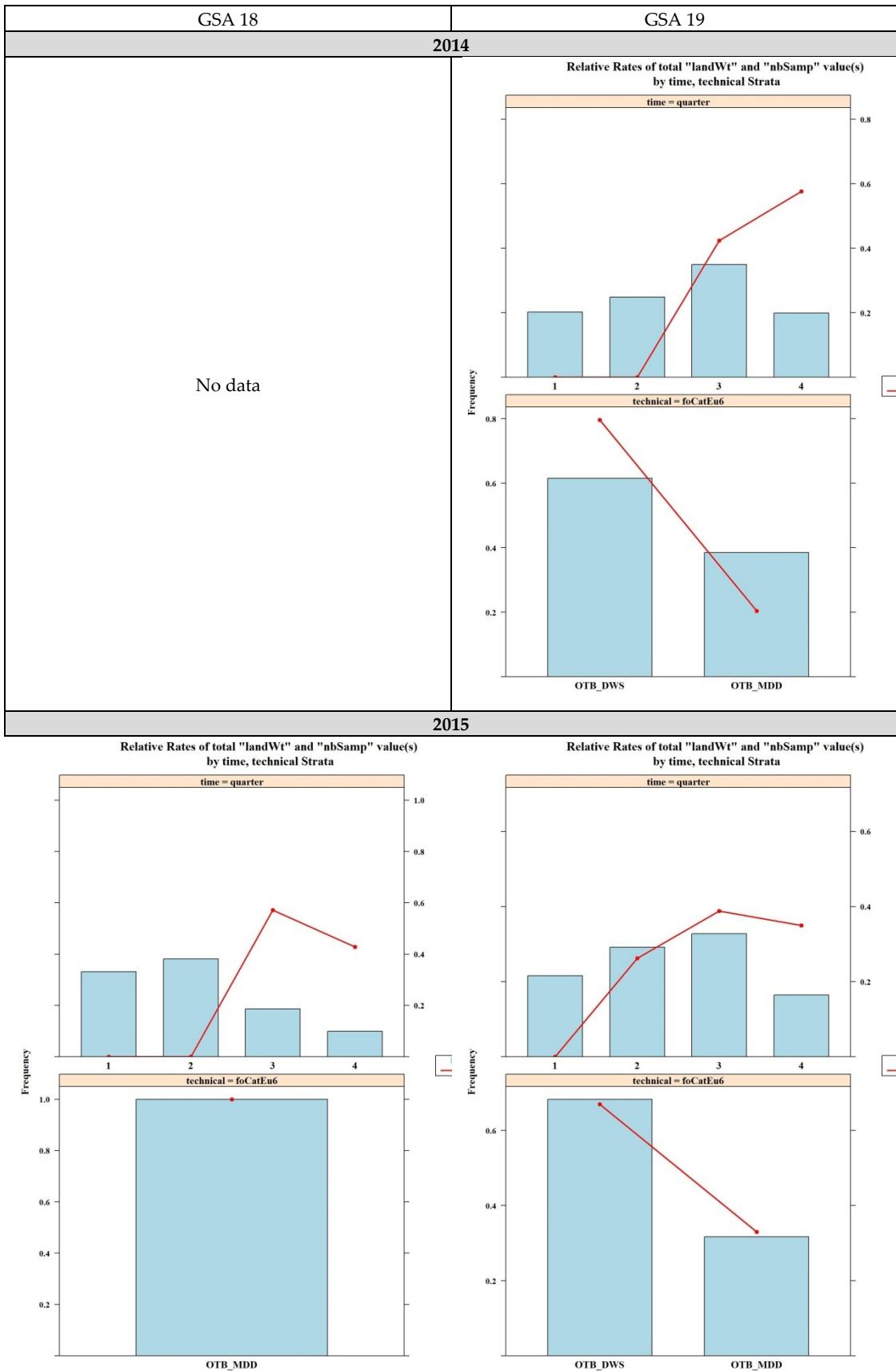


Figure 5.3.3.4 A. *antennatus* in GSA 18-19 (Italy). Relative importance of sampling compared to the relative importance of landing by fishing activity category European lvl 6 and quarter.

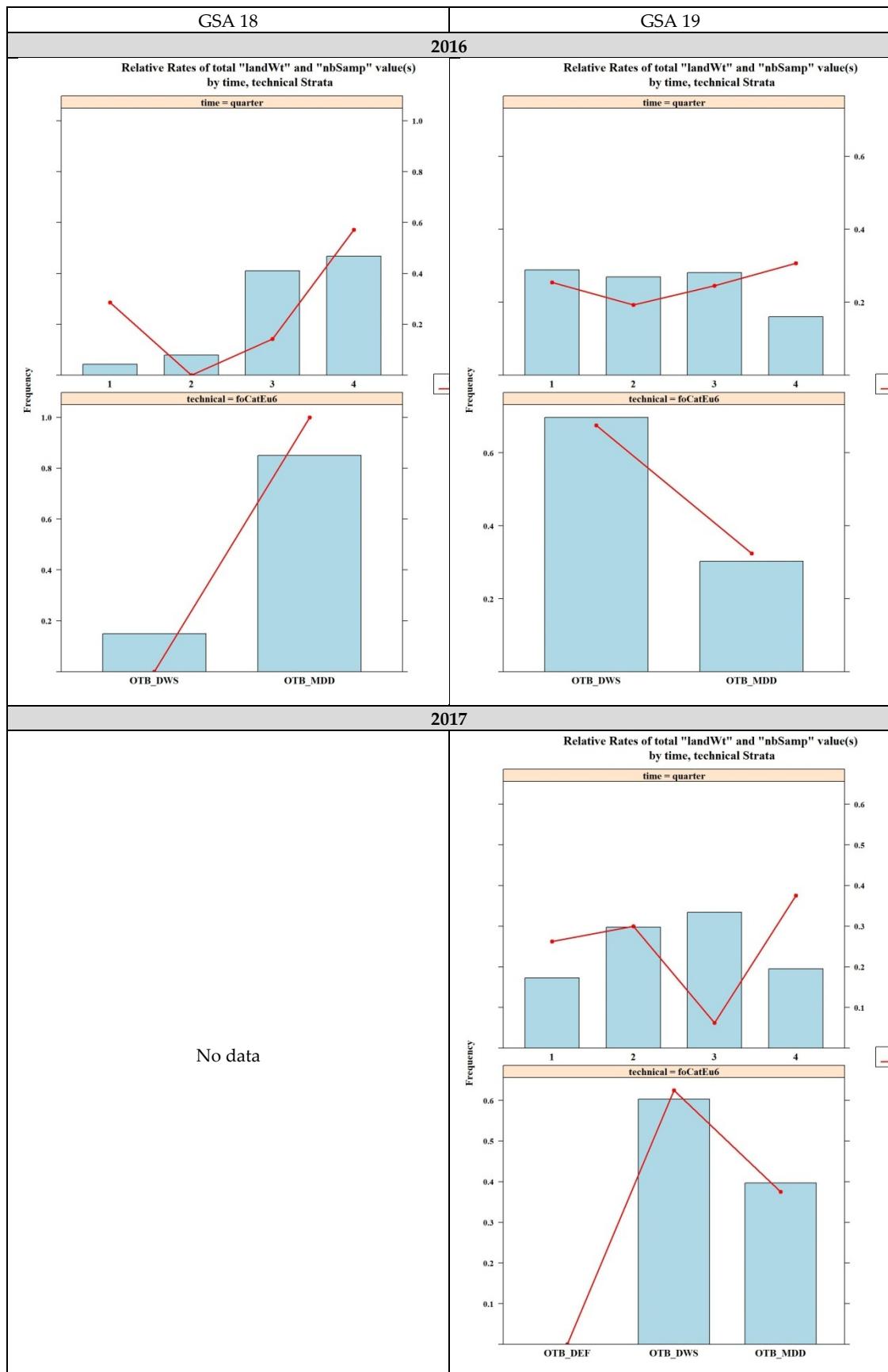


Figure 5.3.3.4 (cont.) *A. antennatus* in GSA 18-19 (Italy). Relative importance of sampling compared to the relative importance of landing by fishing activity category European lvl 6 and quarter

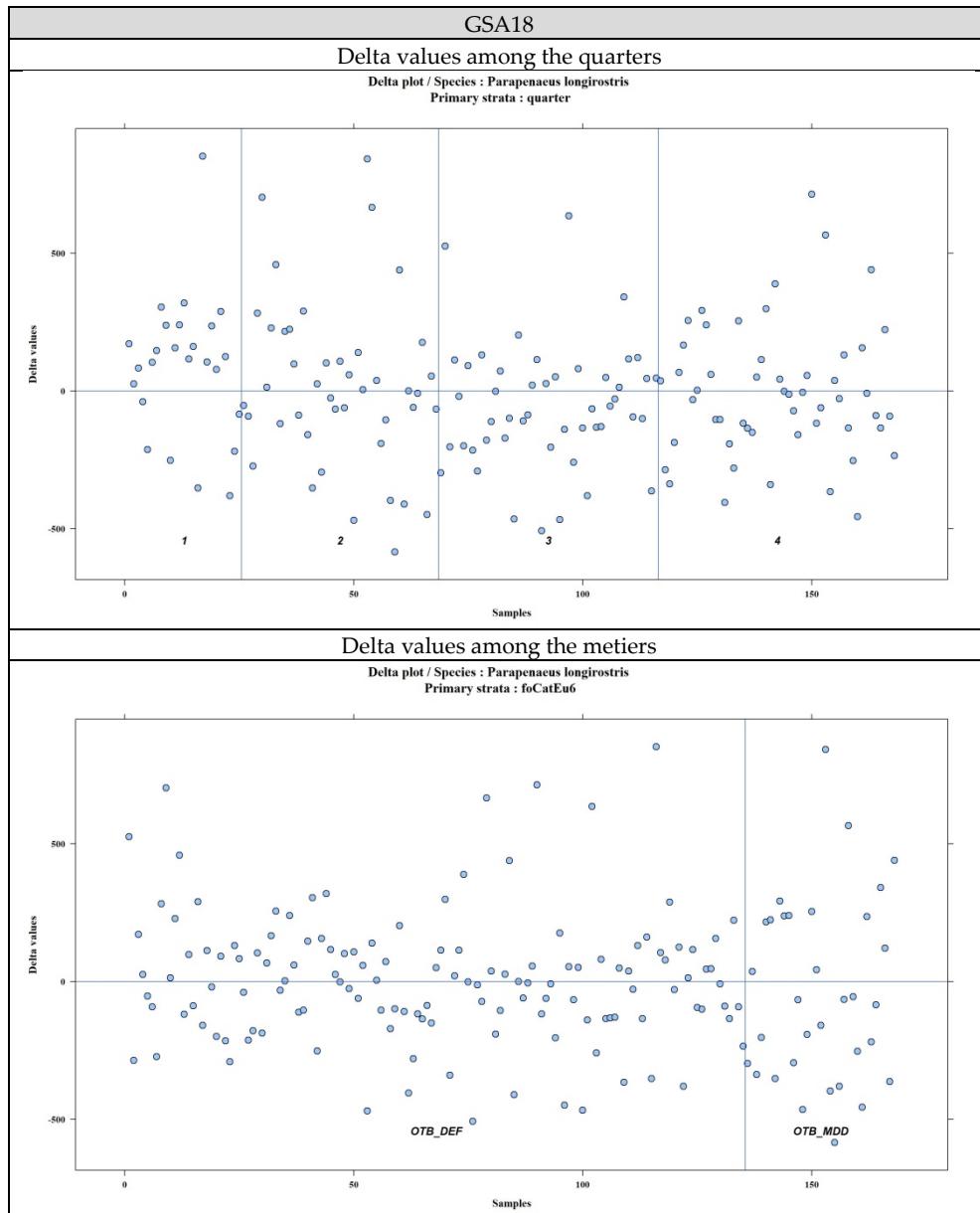


Figure 5.3.3.5 *P. longirostris* in GSA 18-19 (Italy). Variability in the sampling by fishing activity category European lvl 6 and by quarter.

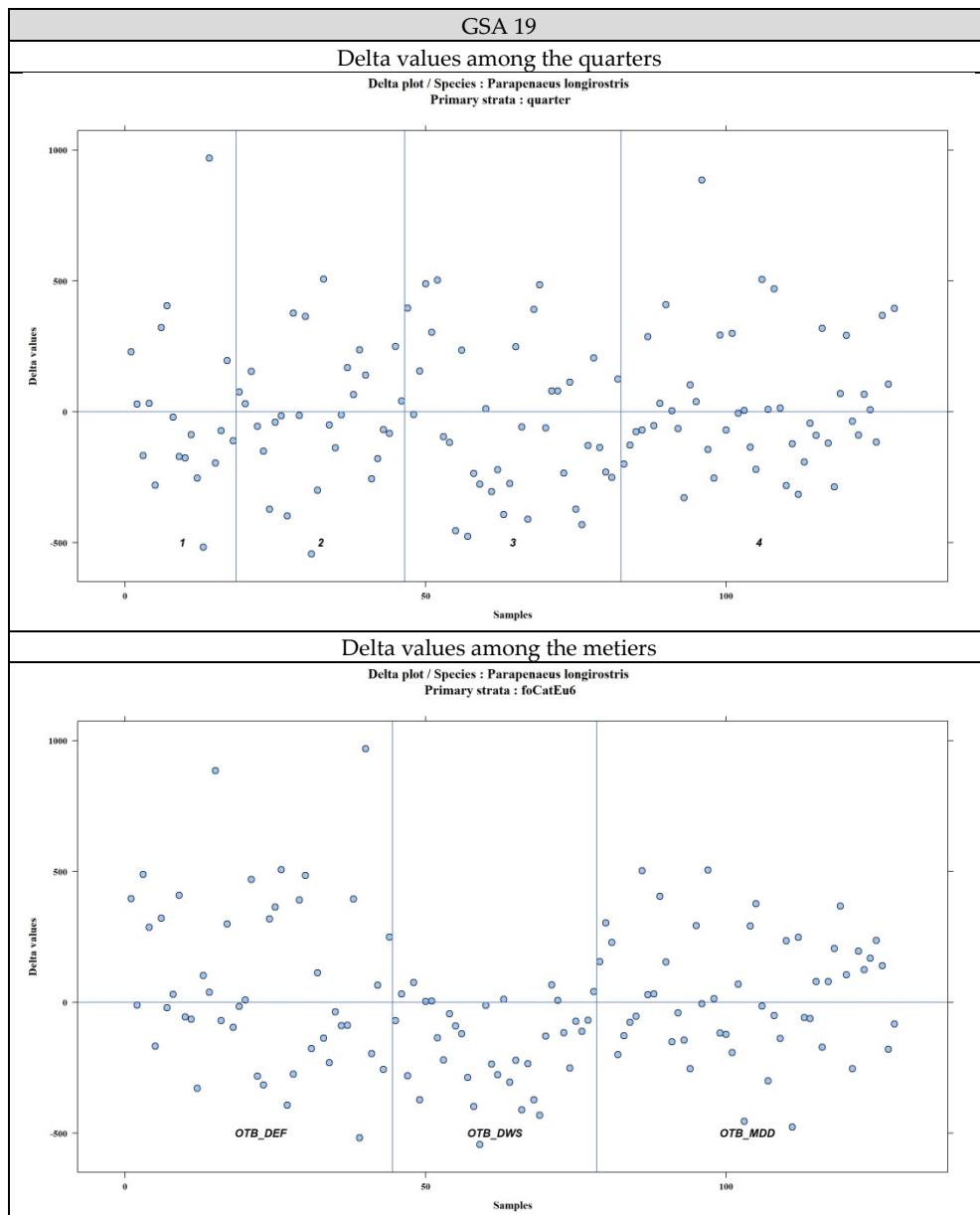


Figure 5.3.3.5 (cont.) *P. longirostris* in GSA 18-19 (Italy). Variability in the sampling by fishing activity category European lvl 6 and by quarter.

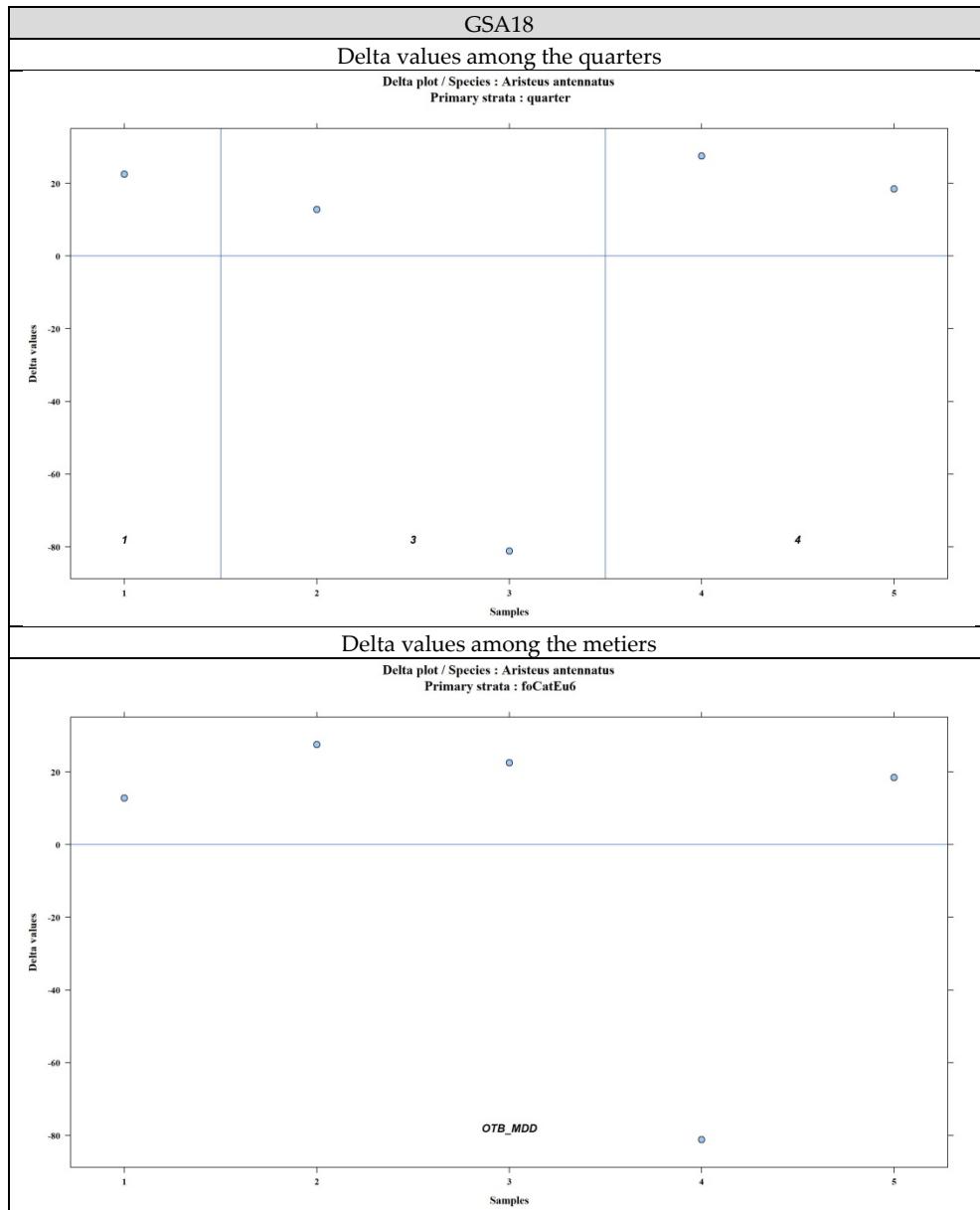


Figure 5.3.3.6 - *A. antennatus* in GSA 18-19 (Italy). Variability in the sampling by fishing activity category European lvl 6 and by quarter.

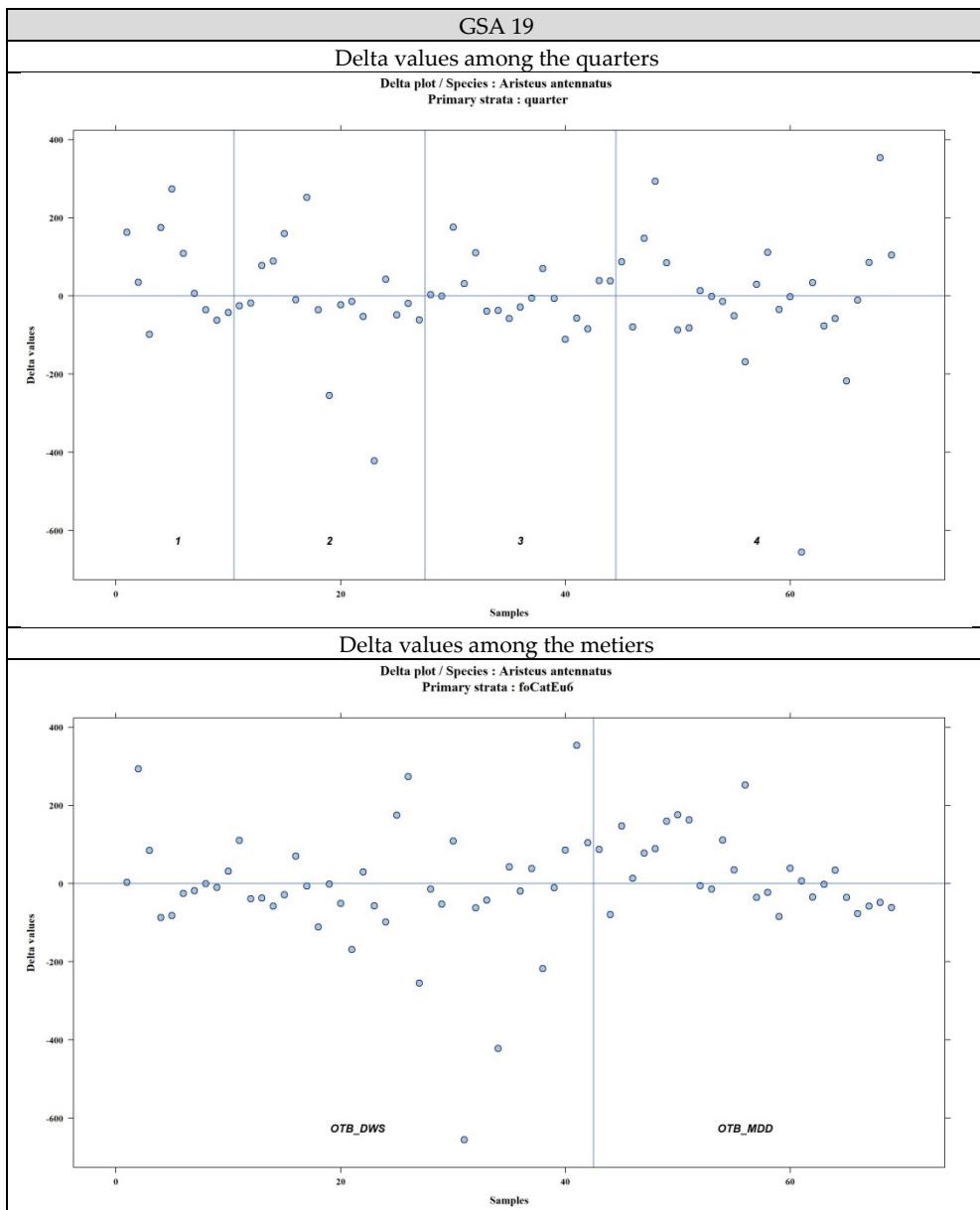


Figure 5.3.3.6 (cont) *A. antennatus* in GSA 18- 19 (Italy). Variability in the sampling by fishing activity category European lvl 6 and by quarter.

Possible outliers have been investigated through the analyses of variability and among the delta values no outliers are founded. Also the level of variability in the length distribution of each trip by métier and quarters has been investigated and it seems that the delta values are comparable among the metiers and the quarters for both the target species.

In this case study not being the variability very different by quarters and by metiers for both species, the analyses can be performed at gear level (level4) and by year (aggregating the quarters). Doing this choice also a wider dataset for the resampling can be used for each level4-year level.

5.3.4 Analyses on sampling optimization in terms of trips

The analyses for sampling optimization to obtain an “optimal” sampling size based on the analyses of the CV of a given species have been performed through the script from the SD Tool 2 package, `toRUNoptimization_from_SDEF_format.r`, with the following settings.

```
REFERENCE_SPECIES <- "Parapenaeus longirostris"
RS_shortcode <- "DPS"

REFERENCE_SPECIES <- "Aristeus antennatus"
RS_shortcode <- "ARA"

cl_all <- readRDS(paste(CASE_STUDY_PATH, "/input files/costCL.rds", sep=""))
cs_all <- readRDS(paste(CASE_STUDY_PATH, "/input files/costCS.rds", sep=""))
pastSituations <- read.csv(paste(CASE_STUDY_PATH, "/04_dataInvestigation_gsa_y_14/Past
situation.csv", sep=""), sep=";")
table_strat_res <- read.csv(paste(CASE_STUDY_PATH, "/input
files/05_stratification_results_by_gsa_y_14.csv", sep=""), sep=";")

nIter <- 100
min_accepted_sample_size <- 12
threshold_for_RecyclingRate <- 0.5
```

The `min_accepted_sample_size` has been set to 12 considering that the analyses are performed by year and by gear then we could accept solution for the minimum number of trips to be sampled that consider at least 3 trips per quarter and by gear.

Table 5.3.4.1 Input table to indicate the level to be used for the optimization process: by GSA, by gear (collapsing the métier) and year (collapsing all the quarters).

space	time	technical
GSA	Y	lev4

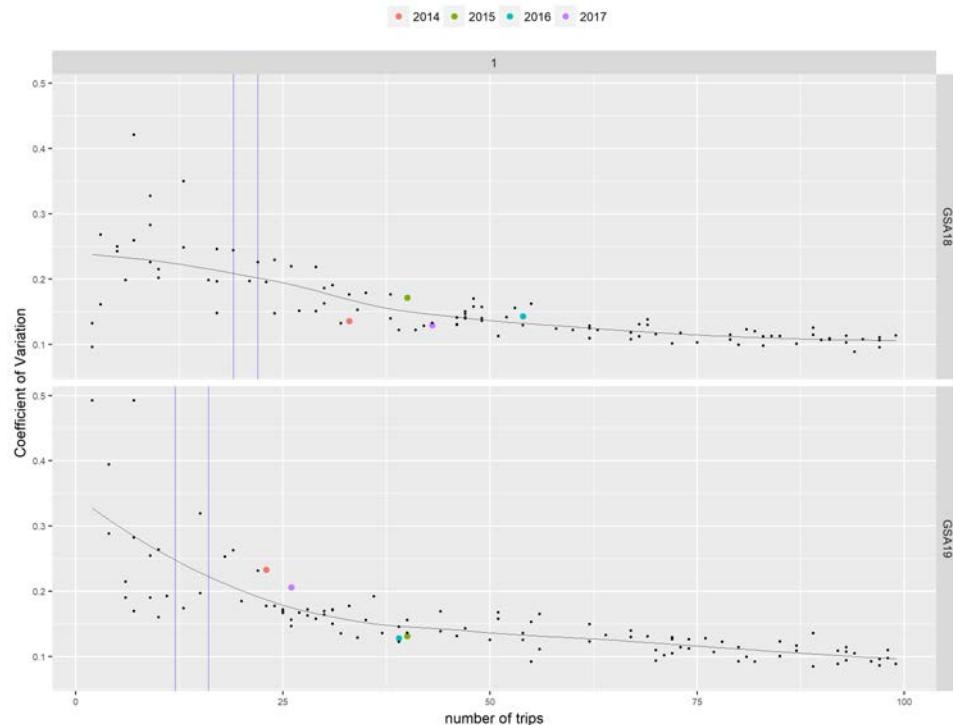


Figure 5.3.4.1 - *P. longirostris* in GSA 18- 19 (Italy). CV versus number of trips. Trip codes by métier (level 4) and year.

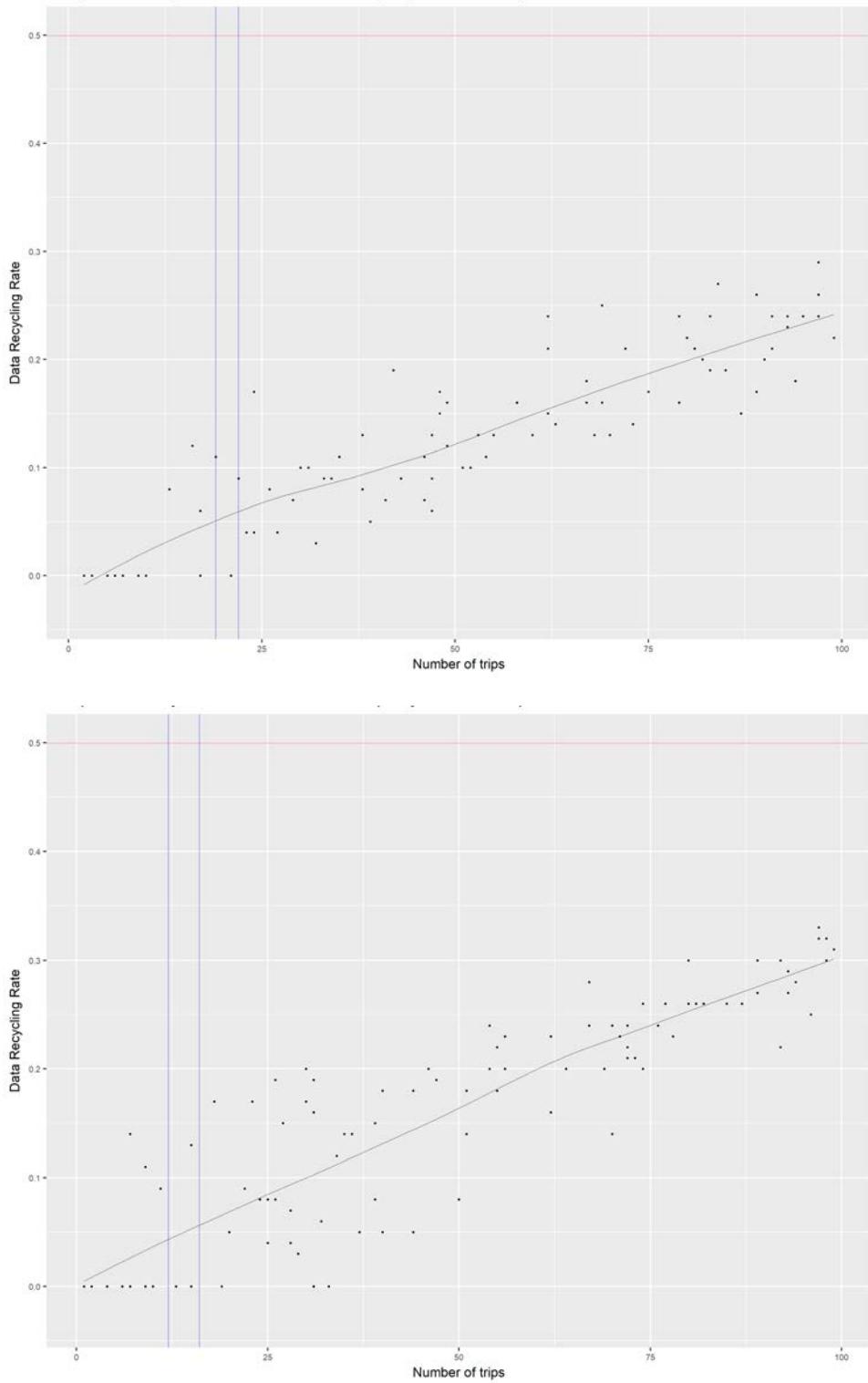


Figure 5.3.4.2 *P. longirostris* in GSA 18-19 (Italy). Recycling rate of the samples against the number of samples by métier (level 4) and year. Vertical blue lines are referred to the optimal sampling size range inferred via the method. Red line corresponded to an arbitrary critical threshold of 50%.

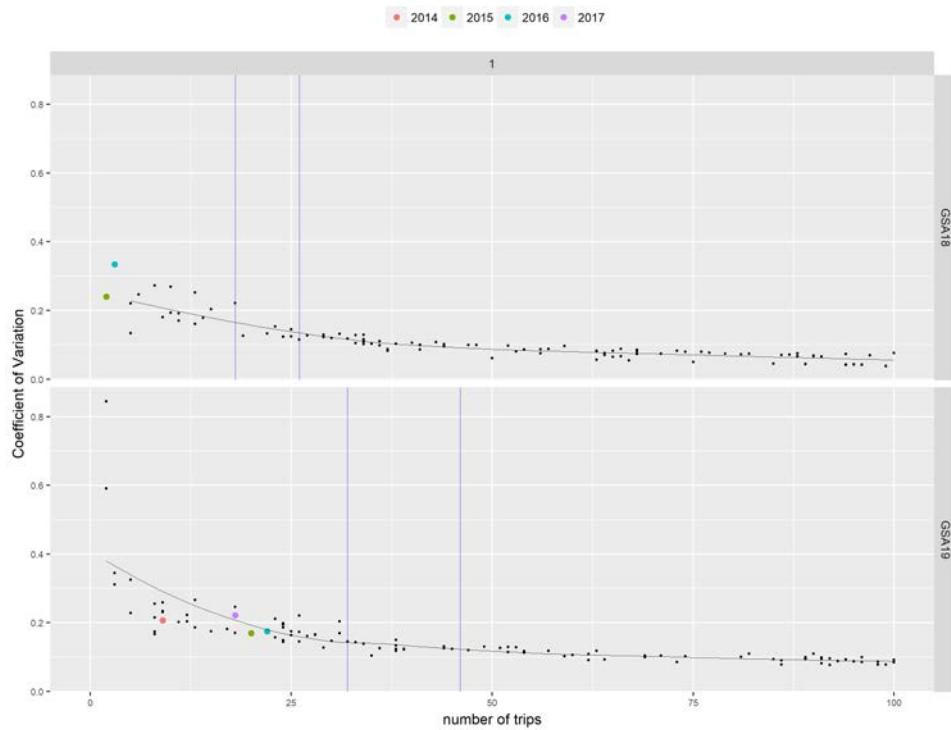


Figure 5.3.4.3 *A. antennatus* in GSA 18-19 (Italy). CV versus number of trip codes. Trips by métier (level 4) and year.

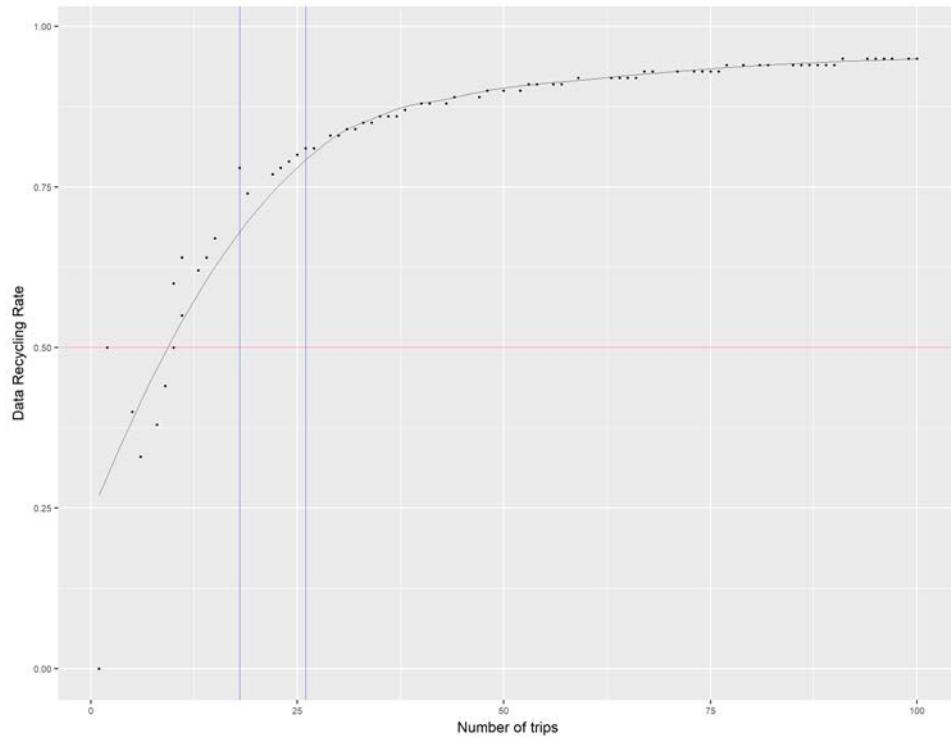


Figure 5.3.4.4 *A. antennatus* in GSA 18-19 (Italy). Recycling rate of the samples against the number of samples by métier (level 4) and year. Vertical blue lines are referred to the optimal sampling size range inferred via the method. Red line corresponded to an arbitrary critical threshold of 50%.

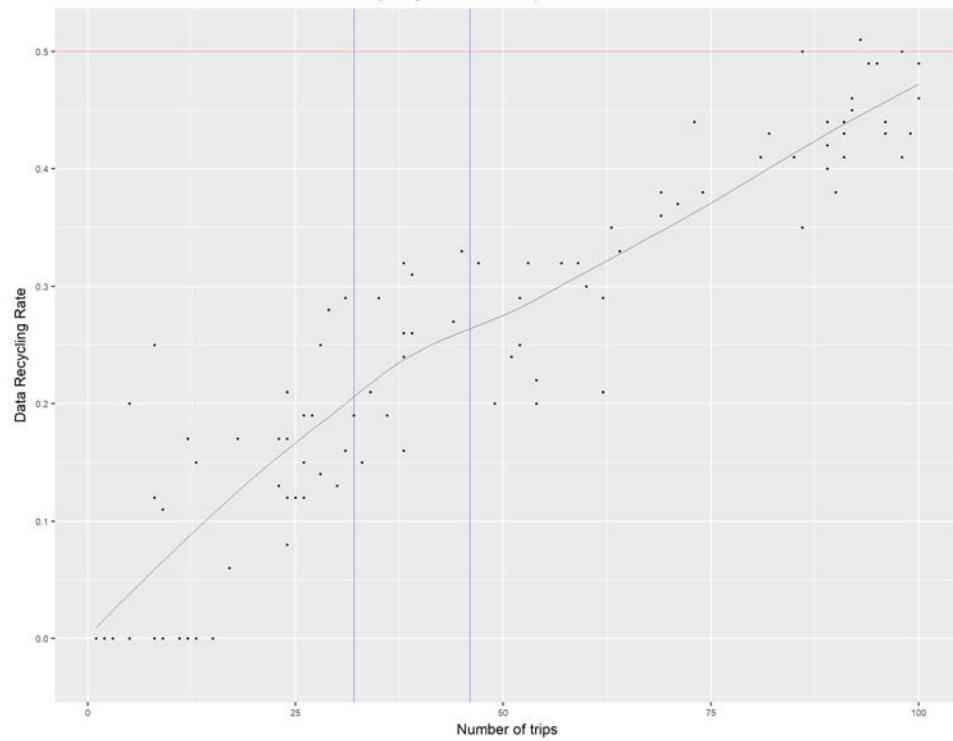


Figure 5.3.4.4 (cont.) *A. antennatus* in GSA 18-19 (Italy). Recycling rate of the samples against the number of samples by métier (level 4) and year. Vertical blue lines are referred to the optimal sampling size range inferred via the method. Red line corresponded to an arbitrary critical threshold of 50%.

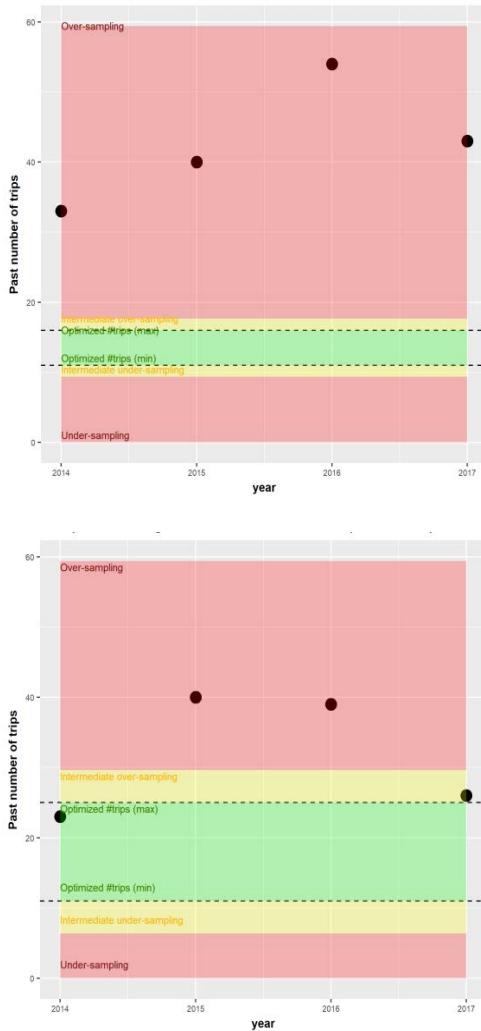


Figure 5.3.4.5 *P. longirostris* in GSA 18-19 (Italy). Optimal sampling size in terms of number of trips by metier (level 4) and year compared to the historical sampling data. Green zone represents the optimal sample size, the yellow is calculated +/- 33% of the range to the limits and the red is +/- 66% of the range to the limits.

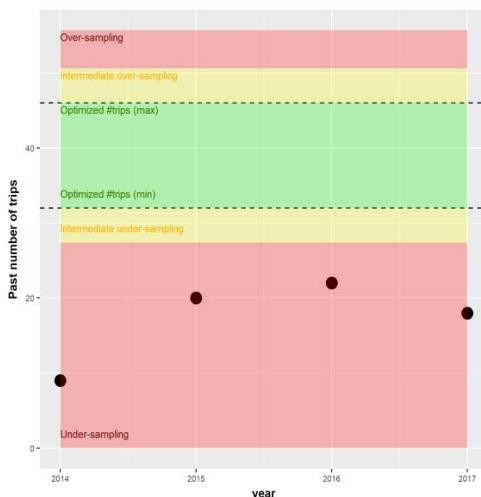


Figure 5.3.4.6 *A. antennatus* in GSA 18-19 (Italy). Optimal sampling size in terms of number of trips by metier (level 4) and by year compared to the historical sampling data.

Table 5.3.4.2 *P. longirostris* and *A. antennatus* in GSA 18-19 (Italy). Optimal sampling size in terms of number of trips by métier (level 4) and by year. In the columns are reported the limits of the optimal sampling range, the mean recycling rate and the mean CV in the range calculated on all the iterations.

	Spatial	Temporal	Technical	min_trips	max_trips	mean_RR	mean_CV
<i>P. longirostris</i>	GSA18	year	OTB	19	22	0.07	0.22
<i>P. longirostris</i>	GSA19	year	OTB	12	16	0.04	0.23
<i>A. antennatus</i>	GSA18	year	OTB	18	26	0.78	0.14
<i>A. antennatus</i>	GSA19	year	OTB	32	46	0.25	0.13

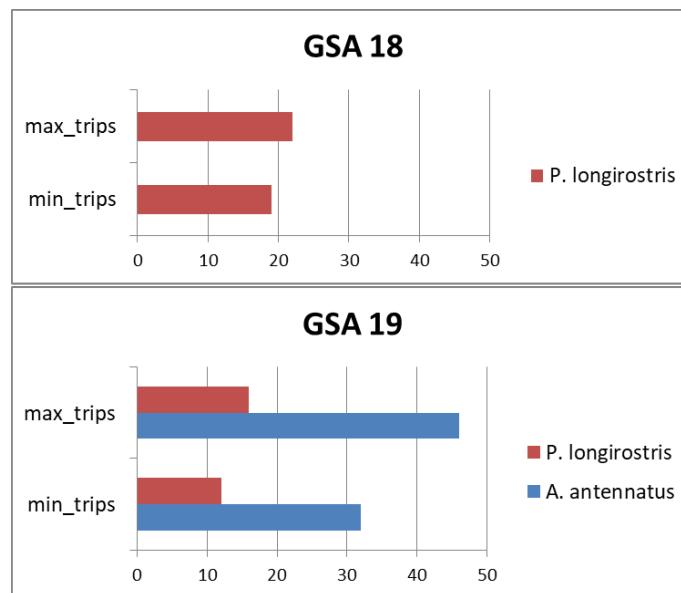


Figure 5.3.4.7 *P. longirostris* and *A. antennatus* in GSA 18-19 (Italy). Optimal sampling size in terms of number of trips by métier (level 4) and by year.

For the GSA 18 only the results of *P. longirostris* have been taken into consideration given that the trips and samples positive to *A. antennatus* in GSA 18 are very few and, consequently, the recycling rate when running bootstrap is too high.

According to the results obtained with the sampling optimization process for all the target species the number of trips to be sampled ranges in [19, 22] for GSA 18 and in [32, 46] for GSA 19. For GSA 19 were optimization process has been based on the analysis of CV values of all the two target species, the minimum number considered is the minimum of the species that requires a higher number of trips (*A. antennatus*).

5.3.5 Analyses on sampling optimization in terms of number of individuals to be measured

Analyses on sampling optimization in terms of number of individuals to be measured have been carried out through the **003_sim_data.r** script with the code at sample level to perform simulations resampling individual length measurements by sample (trip) from the original sampling for a given number of iteration and for different sample size (scenarios) saving indicators (MWCV, n_class_sampled, n_modes_correct) that can be compared with the original sample.

The analyses have been carried out for the *P. longirostris* (in GSA 18) and *A. antennatus* (in GSA 19) with the same settings reported below.

```

species_name <- "Parapenaeus longirostris"
short_name <- "DPS"

GSA <- "GSA18" # "ALL"

# by category or not
by_cat="N" # "Y" or "N"

# expliciting the number of iterations and the sample sizes
n_sims<-50
samp_sizes<-c(seq(30,200, by=10))

# set sampling design of sample data
sampling_design <- list (stratified = FALSE, strata_var = "")

sampling_options <- list (n_sims = n_sims,
    stages="one", # no of stages
    stratified = FALSE, strata_var = "", # stratification details
    stage1_samp_size=NA, samp_sizes = samp_sizes, # samp sizes
    replacement=FALSE, sample_all_available = TRUE,
    sample_all_available_warning = TRUE, # replacement options
    vars_to_keep = c(""))

# Load input data
Load("000_Inputs_DPS\\input_data.rdata")

# setting of the minimum number of individuals considered representative
min_n <- 200

```

Table 5.3.5.1 - *P. longirostris* in GSA 18 (Italy). Results obtained from the simulations on optimization of number of measures to be taken.

commCat	Metier_Gear	quarter	samp_size	EMD	MWCV %
-1	-1	-1	30	0.435	6.5
-1	-1	-1	40	0.432	5.6
-1	-1	-1	50	0.435	5
-1	-1	-1	60	0.430	4.6
-1	-1	-1	70	0.436	4.3
-1	-1	-1	80	0.432	4
-1	-1	-1	90	0.433	3.8
-1	-1	-1	100	0.436	3.6
-1	-1	-1	110	0.428	3.4
-1	-1	-1	120	0.438	3.3
-1	-1	-1	130	0.431	3.1
-1	-1	-1	140	0.434	3
-1	-1	-1	150	0.433	2.9
-1	-1	-1	160	0.435	2.8
-1	-1	-1	170	0.431	2.7
-1	-1	-1	180	0.434	2.7
-1	-1	-1	190	0.432	2.6
-1	-1	-1	200	0.438	2.5
-1	-1	-1	Baseline [1382]	NA	1.0

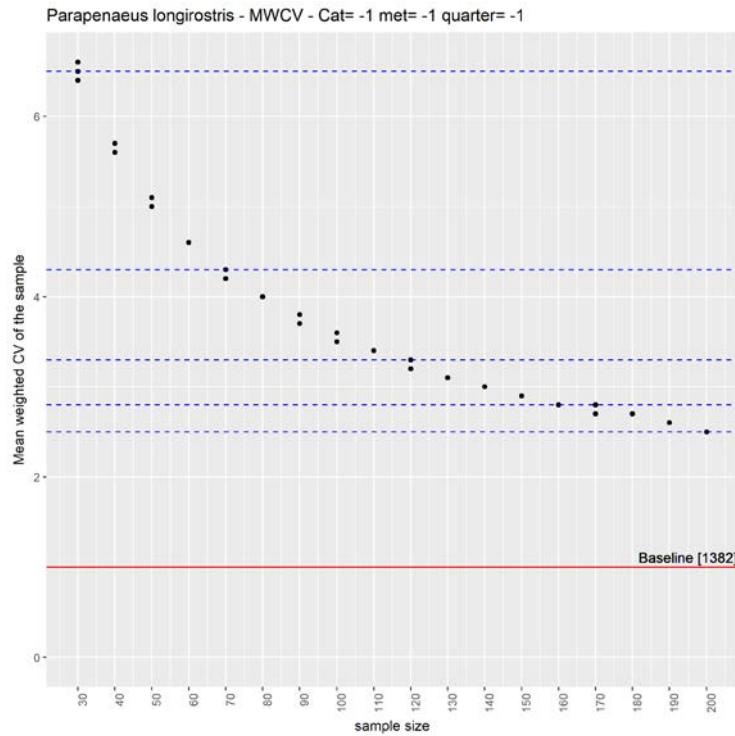


Figure 5.3.5.1 *P. longirostris* in GSA 18 (Italy). Trend of the MWCV calculated on the overall length distribution (obtained summing the length distribution of all the trip selected for the simulations) for each of 50 iterations and each sample size (from 30 to 200, by step of 10).

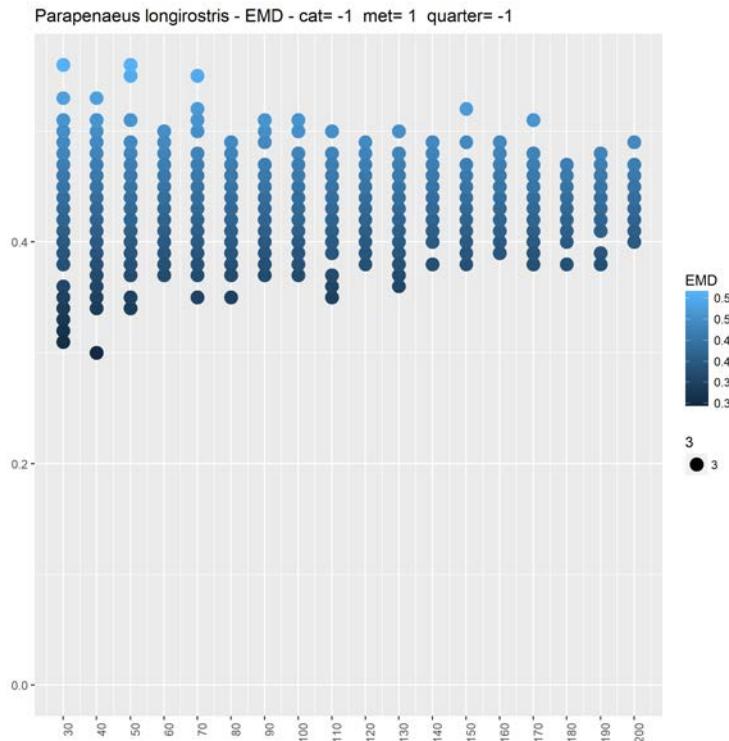


Figure 5.3.5.2 *P. longirostris* in GSA 18 (Italy). Values of the EMD calculated on the overall length distribution (obtained summing the length distribution of all the trip selected for the simulations) for each of 50 iterations and each sample size (from 30 to 200, by step of 10).

Table 5.3.5.2 *P. longirostris* in GSA 19 (Italy). Results obtained from the simulations on optimization of number of measures to be taken.

commCat	Metier_Gear	quarter	samp_size	EMD	MWCV %
-1	-1	-1	30	1.268	7.6
-1	-1	-1	40	1.28	6.6
-1	-1	-1	50	1.261	5.9
-1	-1	-1	60	1.274	5.4
-1	-1	-1	70	1.269	5
-1	-1	-1	80	1.263	4.7
-1	-1	-1	90	1.261	4.4
-1	-1	-1	100	1.265	4.2
-1	-1	-1	110	1.26	4
-1	-1	-1	120	1.268	3.8
-1	-1	-1	130	1.265	3.7
-1	-1	-1	140	1.275	3.5
-1	-1	-1	150	1.267	3.4
-1	-1	-1	160	1.264	3.3
-1	-1	-1	170	1.271	3.2
-1	-1	-1	180	1.267	3.1
-1	-1	-1	190	1.27	3
-1	-1	-1	200	1.27	3
-1	-1	-1	Baseline [1209]	NA	1.1

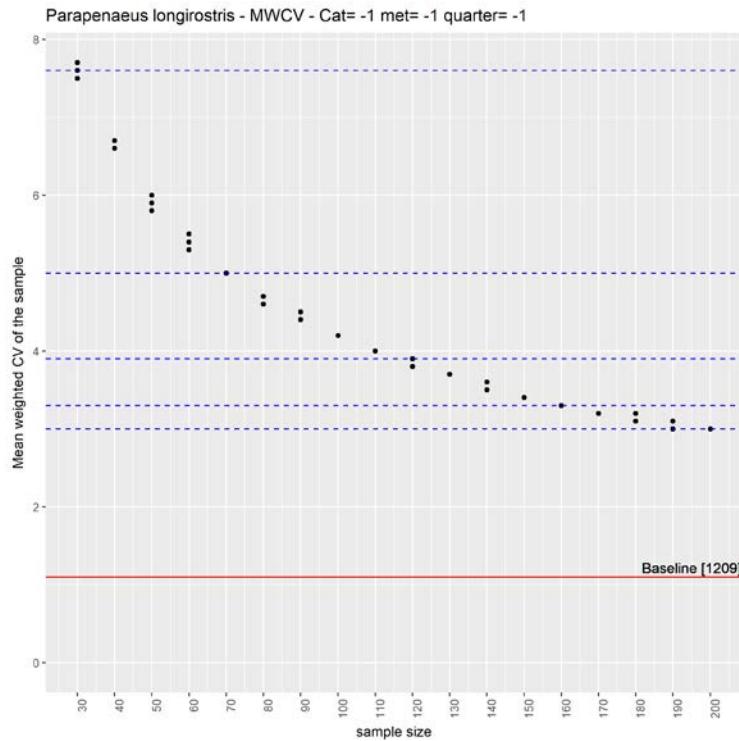


Figure 5.3.5.3 *P. longirostris* in GSA 19 (Italy). Trend of the MWCV calculated on the overall length distribution (obtained summing the length distribution of all the trip selected for the simulations) for each of 50 iterations and each sample size (from 30 to 200, by step of 10).

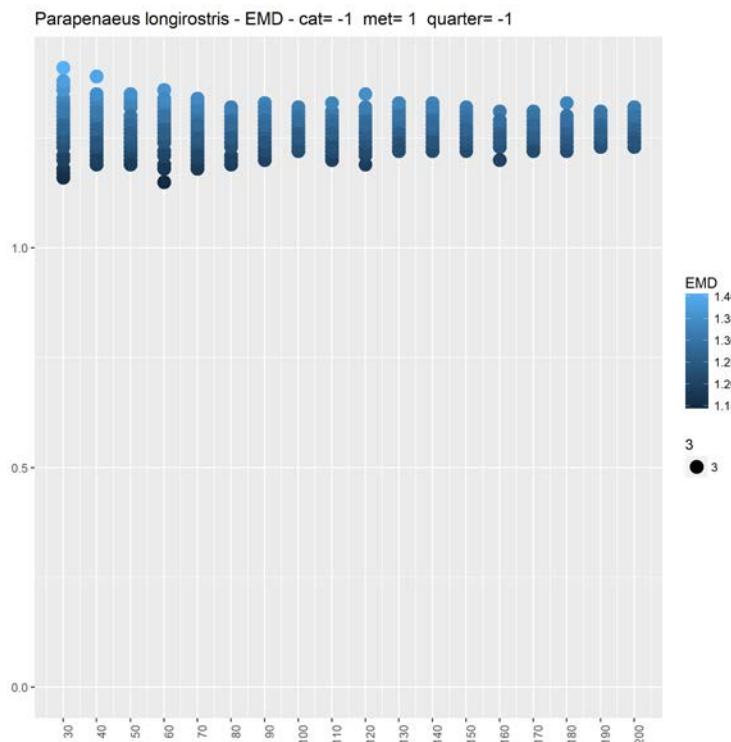


Figure 5.3.5.4 *P. longirostris* in GSA 19 (Italy). Values of the EMD calculated on the overall length distribution (obtained summing the length distribution of all the trip selected for the simulations) for each of 50 iterations and each sample size (from 30 to 200, by step of 10).

Table 5.3.5.3 A. *antennatus* in GSA 19 (Italy). Results obtained from the simulations on optimization of number of measures to be taken.

commCat	Metier_Gear	quarter	samp_size	EMD	MWCV %
-1	-1	-1	30	0.382	13.5
-1	-1	-1	40	0.361	11.7
-1	-1	-1	50	0.372	10.5
-1	-1	-1	60	0.331	9.6
-1	-1	-1	70	0.343	8.9
-1	-1	-1	80	0.324	8.3
-1	-1	-1	90	0.324	7.9
-1	-1	-1	100	0.324	7.5
-1	-1	-1	110	0.334	7.1
-1	-1	-1	120	0.331	6.8
-1	-1	-1	130	0.32	6.6
-1	-1	-1	140	0.324	6.3
-1	-1	-1	150	0.313	6.1
-1	-1	-1	160	0.31	5.9
-1	-1	-1	170	0.295	5.8
-1	-1	-1	180	0.292	5.6
-1	-1	-1	190	0.278	5.5
-1	-1	-1	200	0.277	5.4
-1	-1	-1	Baseline [620]	0	3

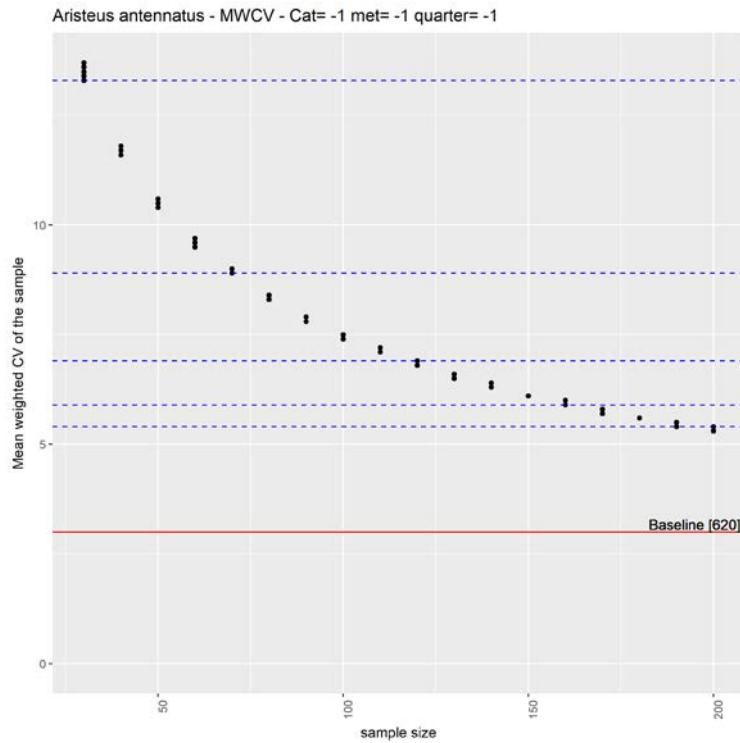


Figure 5.3.5.5 *A. antennatus* in GSA 19 (Italy). Trend of the MWCV calculated on the overall length distribution (obtained summing the length distribution of all the trip selected for the simulations) for each of 50 iterations and each sample size (from 30 to 200, by step of 10).

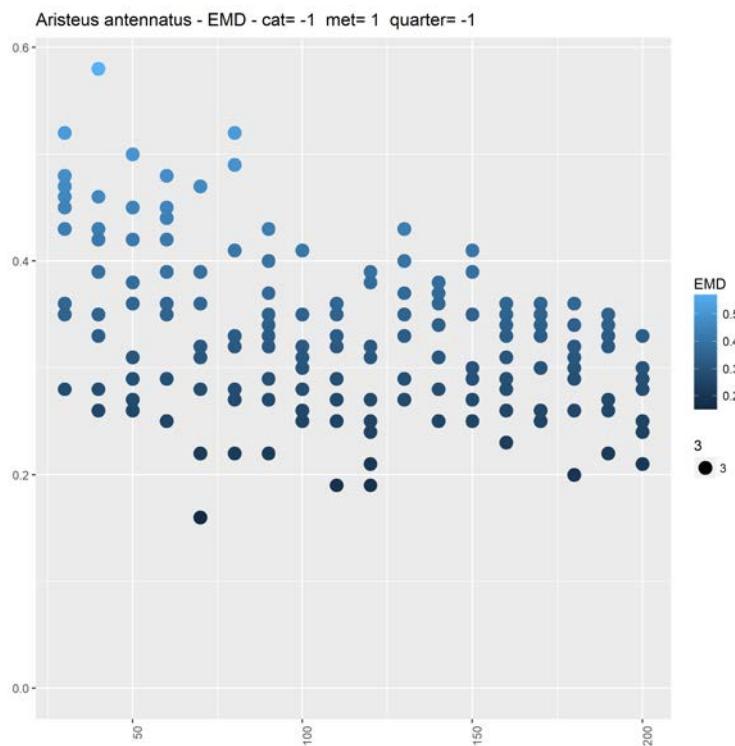


Figure 5.3.5.6 *A. antennatus* in GSA 19 (Italy). Values of the EMD calculated on the overall length distribution (obtained summing the length distribution of all the trip selected for the simulations) for each of 50 iterations and each sample size (from 30 to 200, by step of 10).

From the simulation the minimum number of individuals to be measured is about 170 individuals per trip for both the target species, considering the sample size when the MWCV start to be below the 25th percentile as criterion for selection. Also the values of EMD are coherent with the chosen minimum number.

In the following table (Table 5.3.5.4) the factors calculated by the routine for the subsample needed to reach the minimum number per trip are reported. Subsample factors were calculated averaging the ratios defined for each trip as numbers of lengths in the original dataset to the “optimal” number of lengths suggested from the simulations (only the ratio greater than 1 have been considered in the average). In the columns *Nb_over_Thr* and *Tot* are reporting respectively the number of the samples with a number of lengths greater than the “optimal” number (threshold) and the total number of samples in the original dataset.

Table 5.3.5.4 *P. longirostris* in GSA 18 and *A. antennatus* in GSA 19 (Italy).

Species	Threshold	Sub_sample	Nb_over_Thr	Tot
<i>P. longirostris</i> GSA 18	170	8	168	170
<i>P. longirostris</i> GSA 19	170	8	125	128
<i>A. antennatus</i> GSA 19	170	4	58	69

In the definition of possible scenario simulating a reduction of measures to be taken this minimum number by trip and the factor for subsampling will be considered.

5.3.6 Possible scenarios varying number of trips and measured individuals

The scenario has been defined following the results obtained from the analyses on sampling optimization in terms of number of trips (suggested optimal ranges are [19, 22] for GSA 18 and [32, 46] for GSA 19) and the results from the sampling optimization in terms of number of individuals to be measured at sample level for all the target species.

The scenarios have been defined in the following table and the simulations have been performed through the *toRUNscenario_from_SDEF_format.r* script.

Table 5.3.6.1 *P. longirostris* and *A. antennatus* in GSA 18-19 (Italy). Definition of the scenario to be performed by the SD Tool.

Scenario	Definition	GSA 18	GSA 19
Baseline	The number of trips is equal to the number of trips sampled in the past (average on the years) with the same number of individuals measured in the past.	43 trips allocated to the metiers according the following percentage: OTB_DEF 81% OTB_MDD 19%	35 trips allocated to the metiers according the following percentage: OTB_DEF 32% OTB_DWS 32% OTB_MDD 37%
Scenario 2	The number of trips is equal to the number of trips sampled in the past (average on the years) and the numbers of individuals measured are reduced through subsampling (based on the analyses of lengths to be measured from BiolSim Tool).	43 trips allocated to the metiers according the following percentage: OTB_DEF 81% OTB_MDD 19% Subsampling on: DPS: 1/8 with threshold = 170 ARA: 1/4 with threshold = 170	35 trips allocated to the metiers according the following percentage: OTB_DEF 32% OTB_DWS 32% OTB_MDD 37% Subsampling on: DPS: 1/8 with threshold = 170 ARA: 1/4 with threshold = 170
Scenario 3	The number of trips is equal to the minimum number suggested by the optimization process and with the same number of individuals measured in the past (without subsampling).	19 trips allocated to the metiers according the following percentage: OTB_DEF 81% OTB_MDD 19%	32 trips allocated to the metiers according the following percentage: OTB_DEF 32% OTB_DWS 32% OTB_MDD 37%
Scenario 4	The number of trips is incremented of a percentage of 50% respect to the number of trips sampled in the past (average on the years) and the numbers of individuals measured are reduced through subsampling (based on the analyses of lengths to be measured from BiolSim Tool).	65 trips allocated to the metiers according the following percentage: OTB_DEF 81% OTB_MDD 19% Subsampling on: DPS: 1/8 with threshold = 170 ARA: 1/4 with threshold = 170	53 trips allocated to the metiers according the following percentage: OTB_DEF 32% OTB_DWS 32% OTB_MDD 37% Subsampling on: DPS: 1/8 with threshold = 170 ARA: 1/4 with threshold = 170

The results from the simulations are reported in the table below.

Table 5.3.6.2 *P. longirostris* and *A. antennatus* in GSA 18-19 (Italy). Performance of the scenarios in terms of CV, number of trips (sample size), number of individuals and EMD.

species	Var1	Var2	Var3	scenario	CV %	samp_size	no_indiv	EMD
GSA 18								
<i>Parapenaeus longirostris</i>	GSA18	1	OTB	Baseline	15	43	59333	-
<i>Parapenaeus longirostris</i>	GSA18	1	OTB	Scenario 2	15	42	23197	0.14
<i>Parapenaeus longirostris</i>	GSA18	1	OTB	Scenario 3	20	19	26270	0.04
<i>Parapenaeus longirostris</i>	GSA18	1	OTB	Scenario 4	12	65	32620	0.05
GSA 19								
<i>Aristeus antennatus</i>	GSA19	1	OTB	Baseline	19	35	10079	-
<i>Aristeus antennatus</i>	GSA19	1	OTB	Scenario 2	19	35	5484	0.17
<i>Aristeus antennatus</i>	GSA19	1	OTB	Scenario 3	19	32	9136	0.18
<i>Aristeus antennatus</i>	GSA19	1	OTB	Scenario 4	16	54	8417	0.22
<i>Parapenaeus longirostris</i>	GSA19	1	OTB	Baseline	16	35	38397	-
<i>Parapenaeus longirostris</i>	GSA19	1	OTB	Scenario 2	17	35	15356	0.04
<i>Parapenaeus longirostris</i>	GSA19	1	OTB	Scenario 3	17	32	35291	0.08
<i>Parapenaeus longirostris</i>	GSA19	1	OTB	Scenario 4	14	54	20940	0.05

In GSA 18 the simulations suggest that in *Scenario 2* the value of CV is similar to the CV obtained in the Baseline (15%), though the total number of individuals measured in the year is about 50% of those measured in the past. In *Scenario 3* the number of measured individuals is close to the number of individuals in *Scenario 2*, but the number of trips in *Scenario 3* is about the half of the number of trips sampled in the past. Then, although the number of animals is the same in the two scenarios, the *Scenario 2* with a higher number of trips, showing a lower value of CV, seems to perform better than *Scenario 3*. In *Scenario 4* the number of measures is half respect to the length measurements in the past then the effort in measuring individuals in the laboratories is reduced of 50%, while the effort in sampling trips is increased of 50%. *Scenario 4* shows a higher precision in sampling (12%) so seems to be the best performing from a biological point of view, but it should be assessed also from an economic perspective as the costs for measuring individuals and sampling trips could not be well balanced. It should be also considered that a scenario foreseeing a higher number of trips would be preferable as it implies higher chances to cover the sampled area and allocate the sampling event in a more homogeneous way the trips over the years.

In GSA 19 for all the scenarios the values of CV are similar for both species (about 16% for *P. longirostris* and 19% for *A. antennatus*), as the minimum number of trips suggested by the optimization process is similar to the baseline. While as regards the number of measured individuals, *Scenario 3* is comparable to the baseline and *Scenario 2* shows a lower number (due to the subsampling). Those results show that also in GSA 19 the *Scenario 2* performs better than *Scenario 3*, as the value of CV has not changed although the numbers of measured individuals for both species is about the half of the numbers of individuals measured in the past. *Scenario 4* shows a number of individuals slightly higher than *Scenario 3*, but still below the past (for *P. longirostris* the number of measures is the half respect to the past), and also a slightly higher precision (16% for *A. antennatus* and 14% for *P. longirostris*). Then considering that sampling a

higher number of trips would imply a higher coverage in terms of space and time, also in GSA 19 the *Scenario 4* would be preferred to the others, apart from the economic point of view.

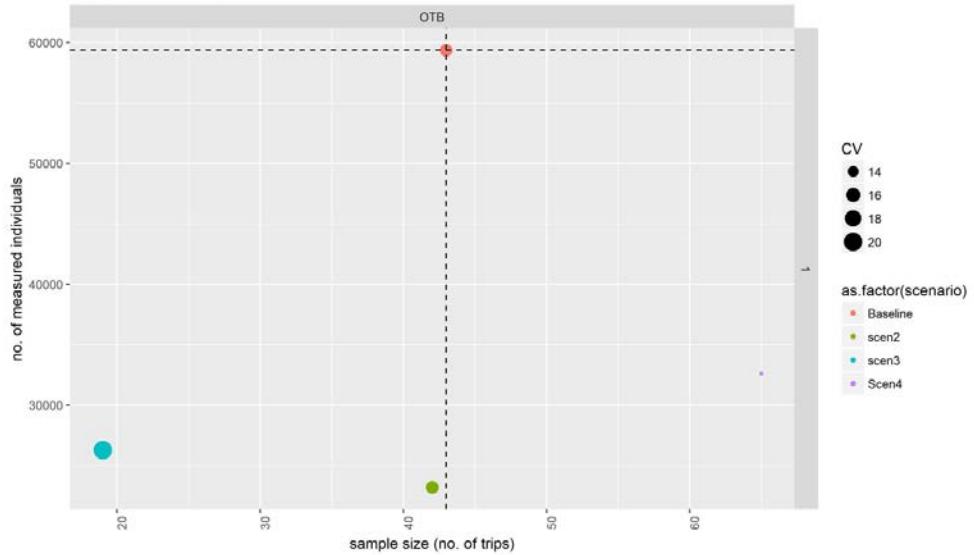


Figure 5.3.6.1 *P. longirostris* in GSA 18 (Italy). Summary of CV values obtained in the performed scenarios with number of trips (x-axis) and number of measured individuals (y-axis).

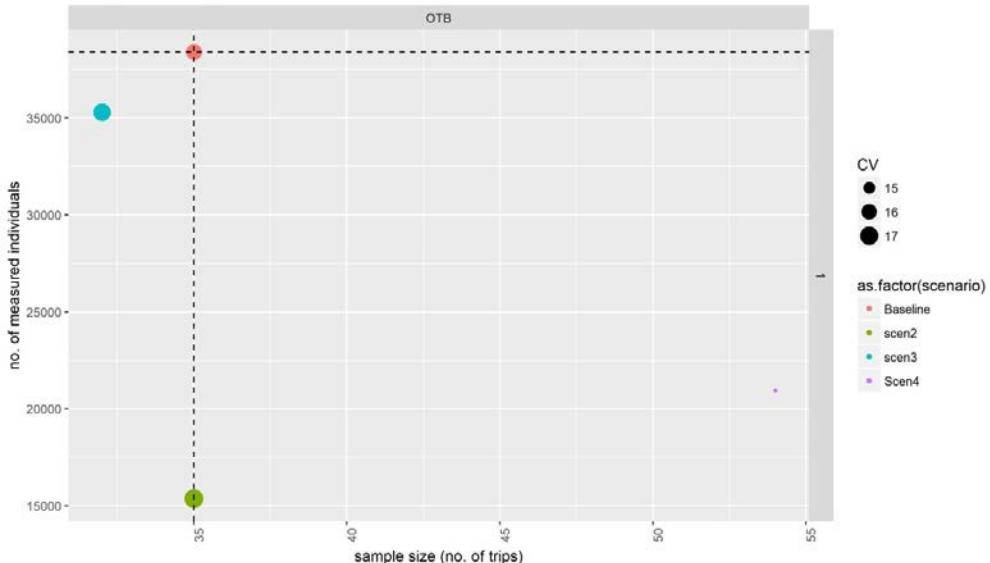


Figure 5.3.6.2 *P. longirostris* in GSA 19 (Italy). Summary of CV values obtained in the performed scenarios with number of trips (x-axis) and number of measured individuals (y-axis).

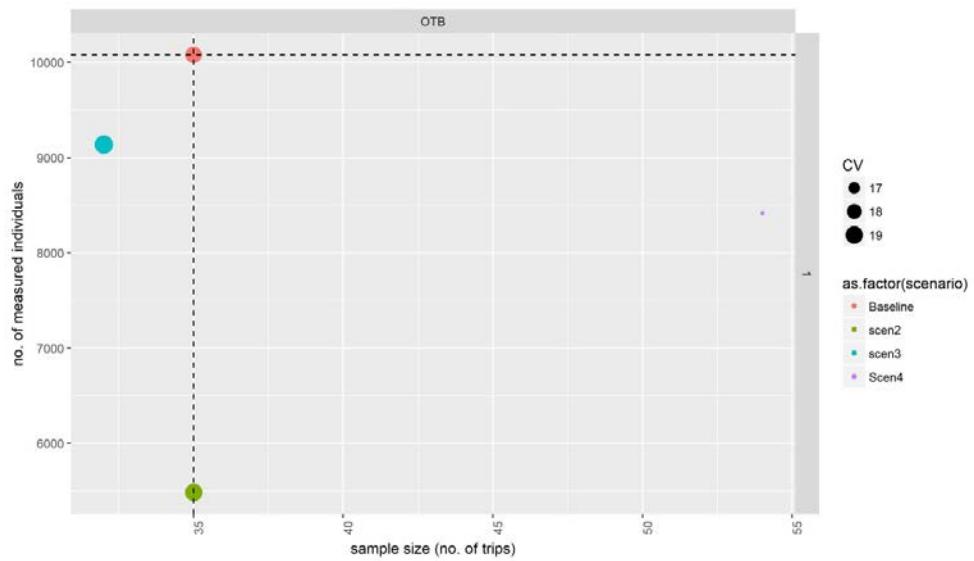


Figure 5.3.6.3 A. *antennatus* in GSA 19 (Italy). Summary of CV values obtained in the performed scenarios with number of trips (x-axis) and number of measured individuals (y-axis).

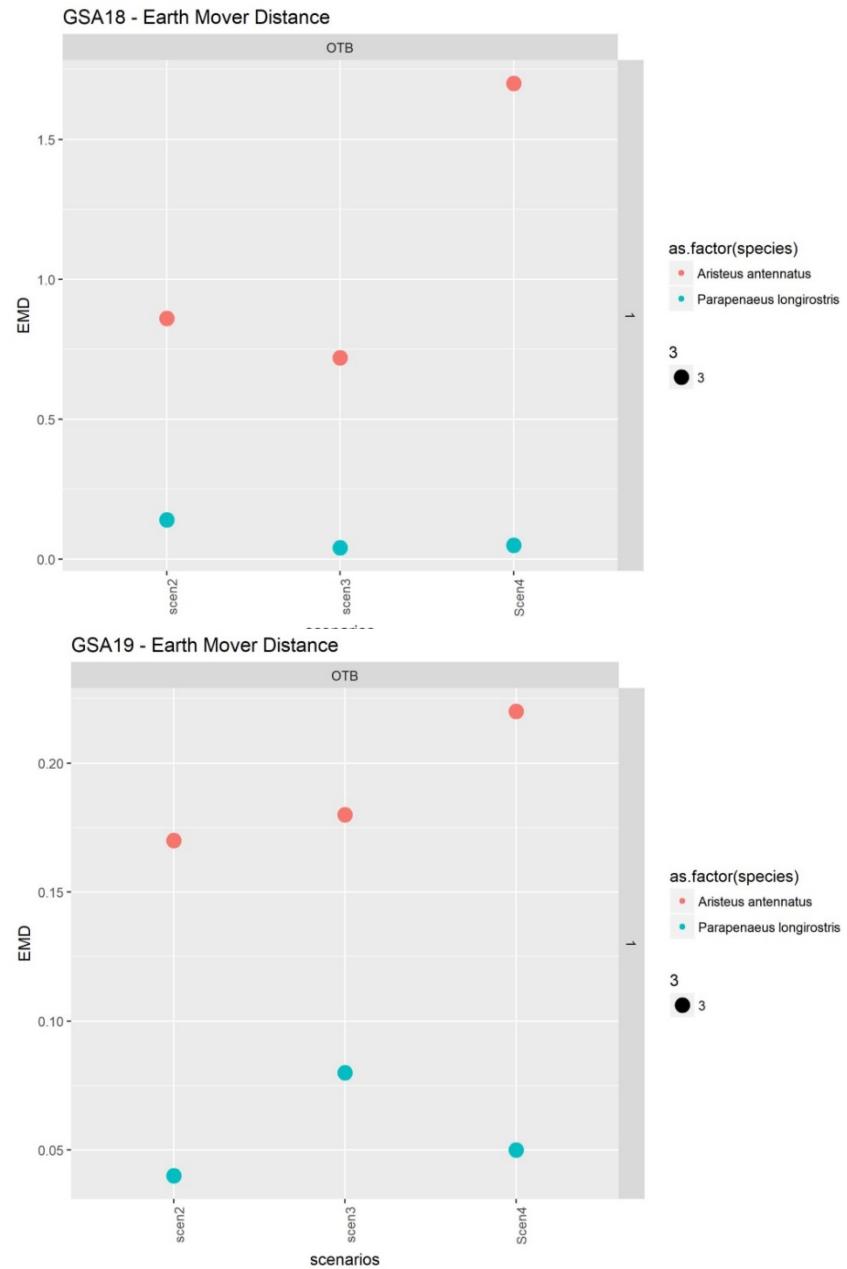


Figure 5.3.6.4 *P. longirostris* and *A. antennatus* in GSA 18-19 (Italy). EMD values for the performed scenarios (in GSA 18 the *A. antennatus* should not be considered, due to the low reliability of the results).

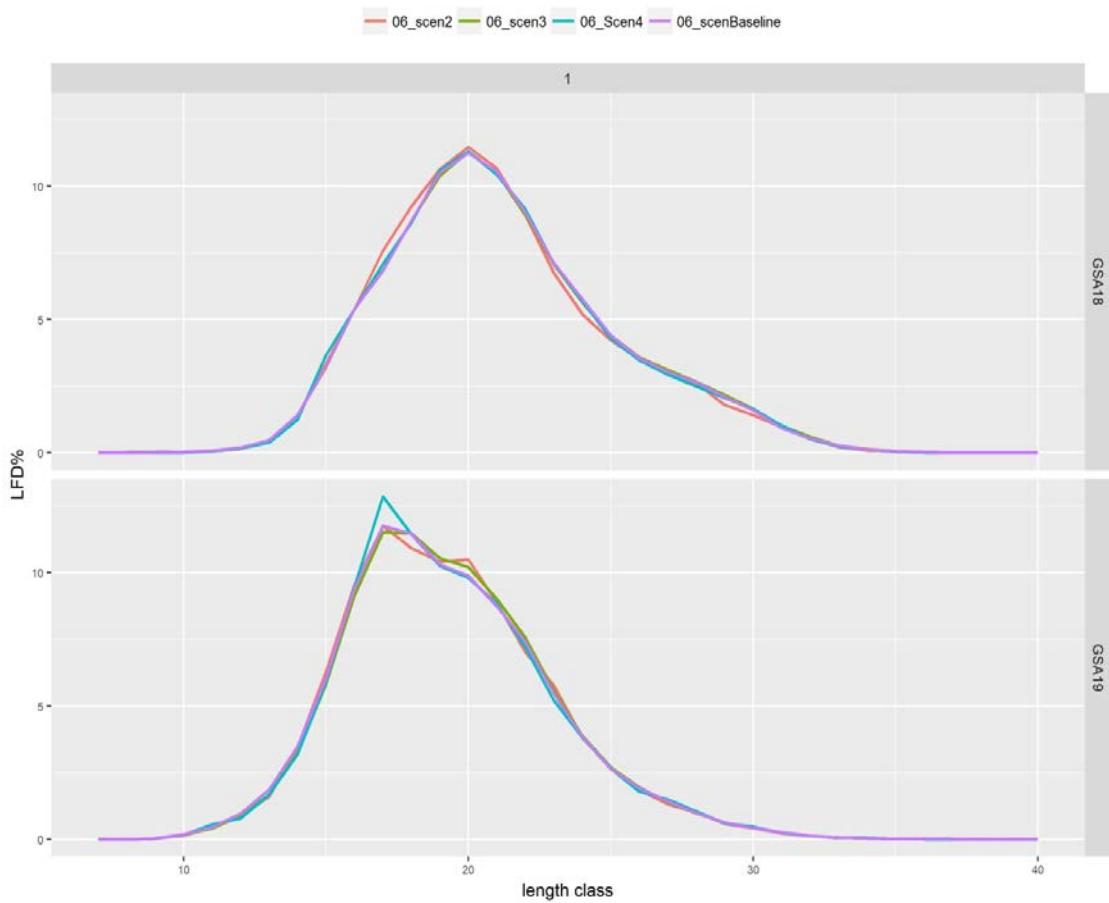


Figure 5.3.6.5 *P. longirostris* in GSA 18-19 (Italy). Raised length frequency distribution (LFD) obtained in the performed scenarios.

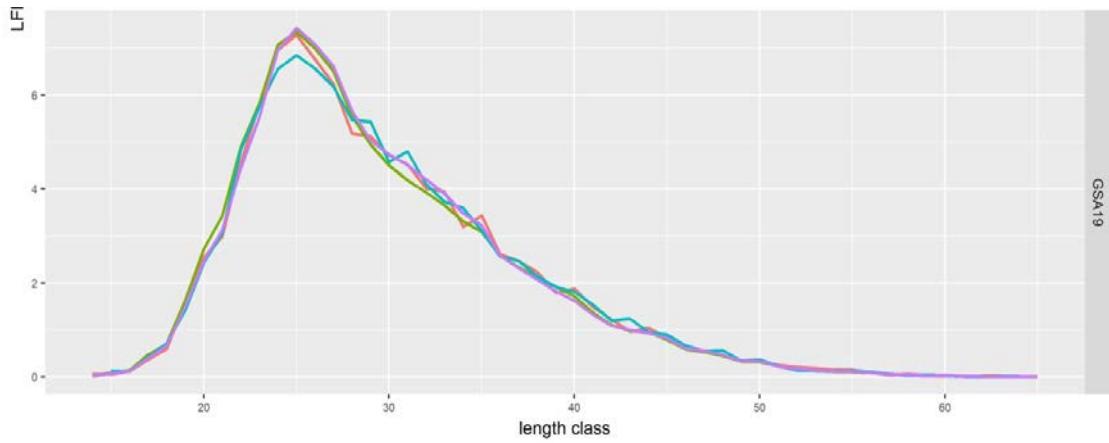


Figure 5.3.6.6 *A. antennatus* in GSA 19 (Italy). Raised length frequency distribution (LFD) obtained in the performed scenarios.

5.3.7 Summary of results

In the following table the number of trips and the number of length measures to be taken in output from the simulations are reported and the respective annual numbers of measures have been estimated considering the number of trips.

Table 5.3.7.1 *P. longirostris* and *A. antennatus* in GSA 18-19 (Italy). Summary table with numbers of trips and numbers of length by trips and annual numbers.

Species	Optimal number of length measurements by trip	Possible subsample	Optimal number of length measurements *nr trips	Optimal nr Trips (Scenario 4)	Length measurements in past years (avg 2014-2017)	Nr Trips in past years
GSA 18 <i>P. Longirostris</i>	170	1/8	11050	65	58740	43
GSA 19 <i>P. Longirostris</i>	170	1/8	9180	54	38699	35
<i>A. antennatus</i>	170	1/4	9180	54	9872	35

5.4 Greek data for Red Mullet in the Aegean Sea (GSA 22)

5.4.1 Sampling design

Red mullet (*Mullus barbatus*) is an important demersal resource for the fisheries in the Aegean Sea (GSA 22). In this area, red mullet is a shared stock exploited by both Greek and Turkish trawlers, gillnetters and trammel netters and, to a lesser extent, by other commercial fisheries.

The sampling scheme for the volume and length of the catch fractions (landings, discards and PETs) is based on the principles of stratified random sampling, employing the métier (level 6) as the basic stratum. The reference list of métiers that was agreed at Regional level during the RCM Med & BS 2009 has been used for the selection of the métiers that have to be sampled. The Hellenic coastline and marine area of GSAs are divided in 12 major sub-areas which constitute the next level of stratification within each métier. The Primary Sampling Unit (PSU) is the fishing trip. The total number of trips to be sampled is defined proportionally to the effort (number of days at sea) for each métier during the reference year. The source of data is the official national fleet registry used to classify vessels by fleet segment and area, and the DCF data collection system of the reference year used for the effort data that were attained based on the sampling scheme. The PSU selection is performed through random-draw of a trip by métier and per GSA, with the option to replace the trip in case that the vessel owner refuses the cooperation. Thus, the sampling scheme is based on the principles of stratified random sampling (8 métiers x 12 sub-areas), implemented through sampling trips performed by observers at sea and on shore (landing sites). The sampling trips are performed quarterly, taking into account the temporal distribution of the effort within each métier and area.

Vessel trips are randomly selected within each stratum (i.e., for every métier within each of the 12 sub-areas, where it is relevant, thus 8 métiers x 12 areas) and then they are equally divided across the quarters. Regarding length composition, a random sample of up to 50 individuals

(depending on availability) per species is selected from the landings and from the discards (separately) per haul (at sea), while on shore the samples are taken from the total amount of landings.

Dataset available

The dataset contains Greek landings, effort and sampling data of *M. barbatus* in GSA 22 for the years 2014–2016. In the year 2015, the data collection was only performed in the last quarter, due to administrative constraints, so this year was not taken into account in the analyses.

In Figure 5.4.1.1 the landings by year, quarter and Fishing activity category European lvl 6 are presented. In Table 5.4.1.1 the number of trips, samples and measurements by fishing activity category European lvl 6 and quarter are presented.

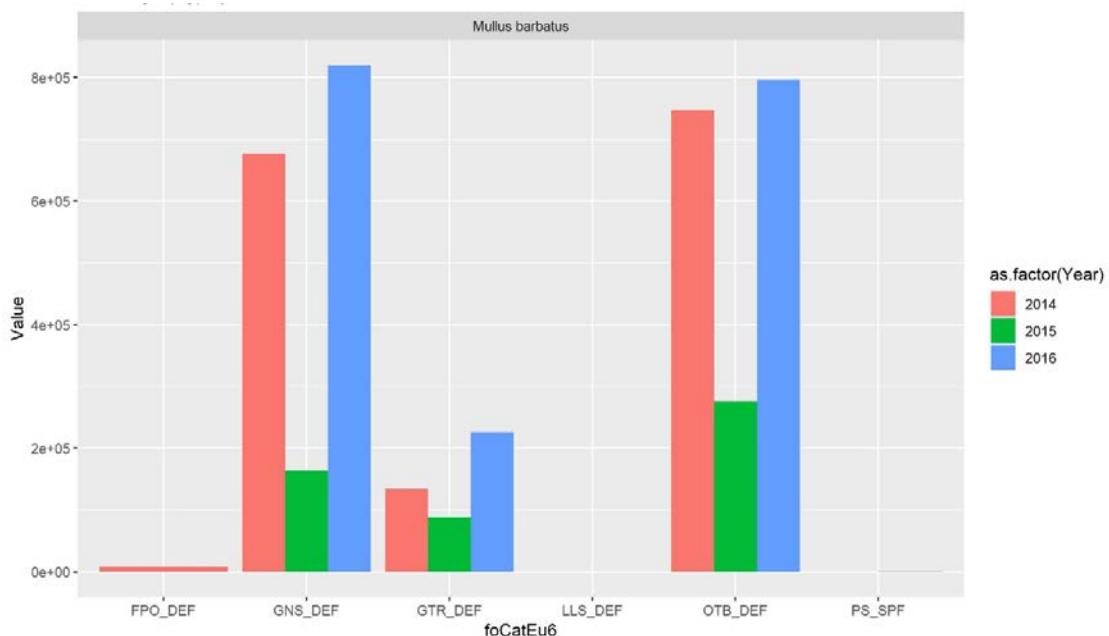


Figure 5.4.1.1 *M. barbatus* in GSA 22 (Greece). Landings by year, quarter and Fishing activity category European lvl 6.

Table 5.4.1.1 *M. barbatus* in GSA 22 (Greece). Number of trips, samples and measurements by fishing activity category European lvl 6 and quarter.

Year	Metier	N. of Trips					N. of samples					N. of individuals				
		Q1	Q2	Q3	Q4	Annualy	Q1	Q2	Q3	Q4	Annualy	Q1	Q2	Q3	Q4	Annualy
2014	GNS_DEF	11	39	52	36	138	4	33	50	34	121	32	724	1058	616	2430
	GTR_DEF	9	35	37	43	124	6	31	37	39	113	23	335	298	414	1070
	OTB_DEF	7	46	0	92	145	9	113	0	355	477	541	4940	0	15097	20578
2015	GNS_DEF	1	0	0	44	45	2	0	0	40	42	2	0	0	663	665
	GTR_DEF	2	0	0	54	56	3	0	0	55	58	35	0	0	535	570
	OTB_DEF	1	1	0	59	61	2	5	0	180	187	6	81	0	6930	7017
2016	GNS_DEF	28	85	69	56	238	26	95	53	47	221	421	1206	978	659	3264
	GTR_DEF	25	61	68	67	221	21	57	70	65	213	55	649	1305	723	2732
	OTB_DEF	35	69	0	124	228	75	158	0	389	622	3027	6307	0	13863	23197
avg	GNS_DEF	13	41	40	45	140	11	43	34	40	128	152	643	679	646	2120
	GTR_DEF	12	32	35	55	134	10	29	36	53	128	38	328	534	557	1457
	OTB_DEF	14	39	0	92	145	29	92	0	308	429	1191	3776	0	11963	16931

5.4.2 Exploratory data analysis

The relative importance of the sampling was generally found proportional to the one of the landings, by year, quarter and métier (figures 5.4.2.1, 5.4.2.2. and 5.4.2.3), with the exception of the 1st quarter in 2016, where the high relative importance of landings data in GNS_DEF is not consistent with the corresponding low samplings data frequency of this métier.

Finally, in Figure 5.4.2.4 the dispersion the Δ index in respect to métier (a), year (b) and quarter (c) is depicted. This index, accounts for the variability of the data and is used to quantify the heterogeneity within the strata to detect possible outlier values. The figure reveals a high variability in the sampling trips of trawlers in some cases in 2014.

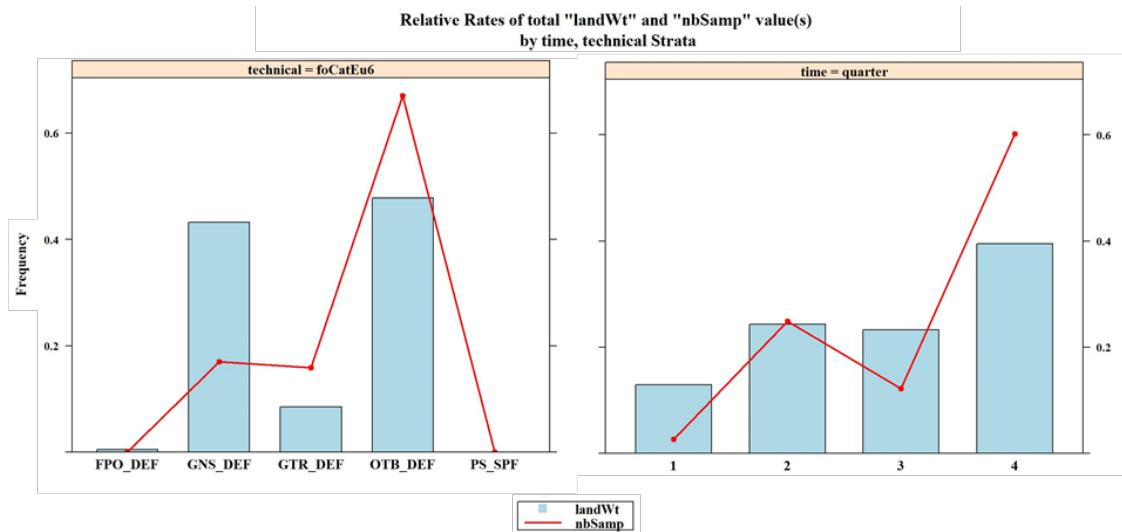


Figure 5.4.2.1 *M. barbatus* in GSA 22 (Greece). Relative importance of sampling compared to the relative importance of landing by fishing activity category European lvl 6 and quarter for the sampling year 2014.

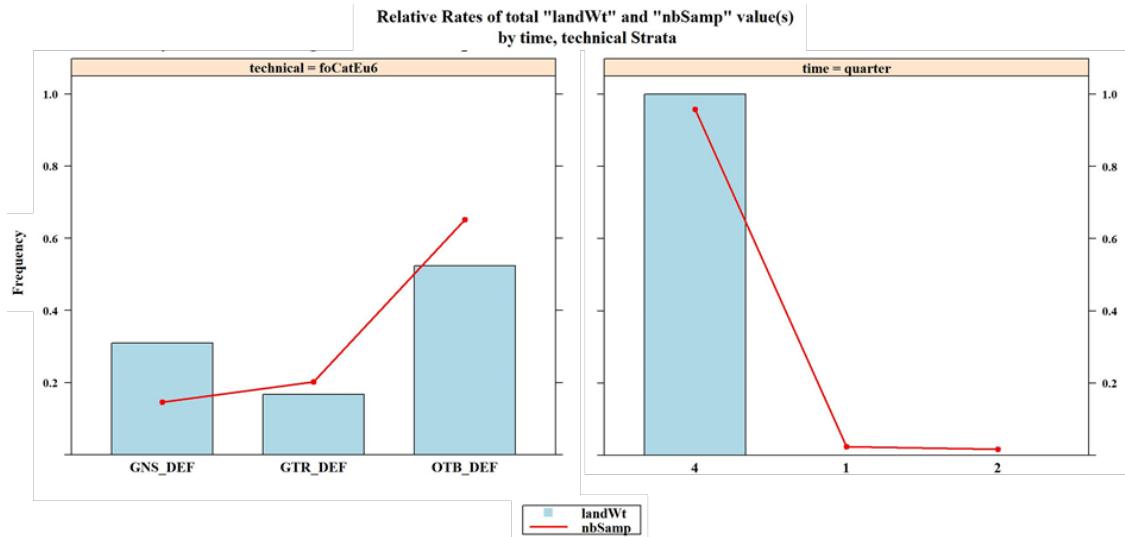


Figure 5.4.2.2 *M. barbatus* in GSA 22 (Greece). Relative importance of sampling compared to the relative importance of landing by fishing activity category European lvl 6 and quarter for the sampling year 2015.

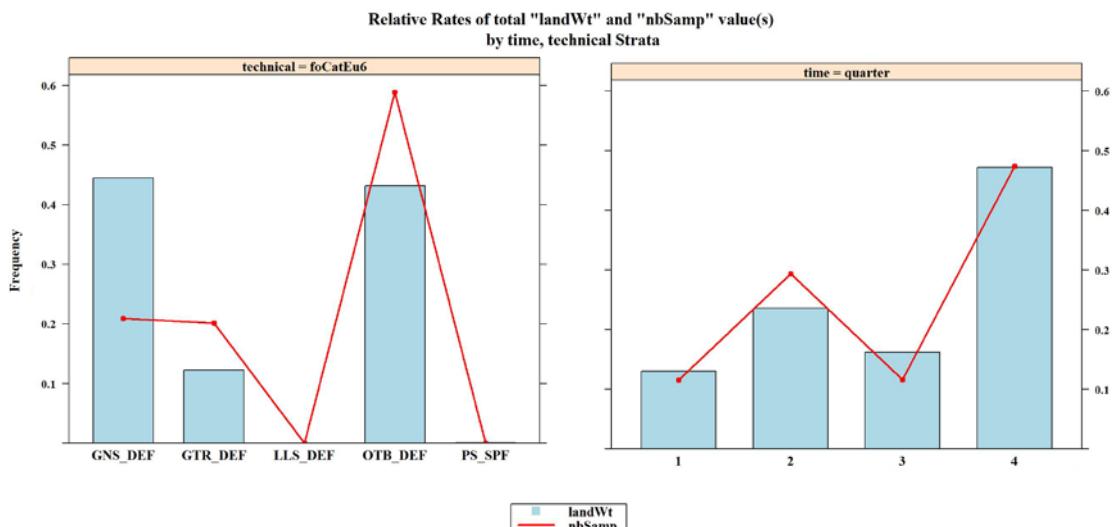


Figure 5.4.2.3 *M. barbatus* in GSA 22 (Greece). Relative importance of sampling compared to the relative importance of landing by fishing activity category European lvl 6 and quarter for the sampling year 2016.

5.4.3 Need for optimization

In the following section, the implementation of the methodology on GNS_DEF, GTR_DEF and OTB_DEF is presented. The number of iterations for the bootstrap resampling method was set to 100.

For GNS_DEF metier, the recycling rate of the samples against the number of trips is depicted in the Figure 5.4.3.1 and the outcomes of the bootstrap resampling method are presented in the Figure 5.4.3.2 (for individuals (a) and trips (b)). Apparently, recycling rate was quite high for the first quarter, due to the low number of implemented trips during this period. As a result, the outcomes of the methodology for this quarter should not be taken into account. For the three remaining quarters the recycling rate was low to moderate and so the corresponding results of

the bootstrap method should be considered as safe. In 2014, under sampling has been noted in the first quarter, the sampling was within the optimal range (trips) in the second quarter and oversampling has been noted in third and fourth quarters. In 2016, oversampling has been recorded in every quarter.

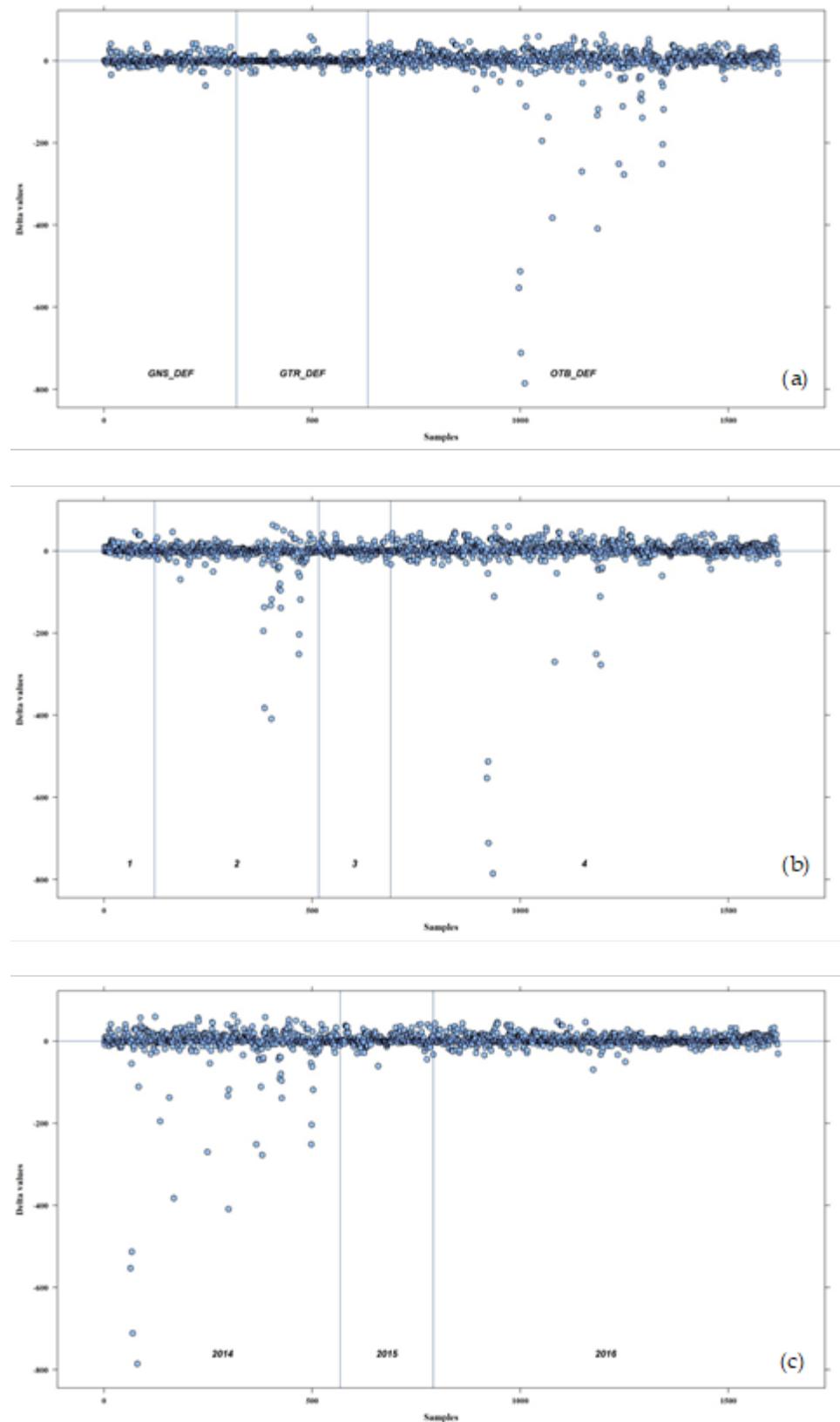


Figure 5.4.2.4 *M. barbatus* in GSA 22 (Greece). Variability in the sampling by fishing activity category European lvl 6 (a), by season (b) and by year (c).

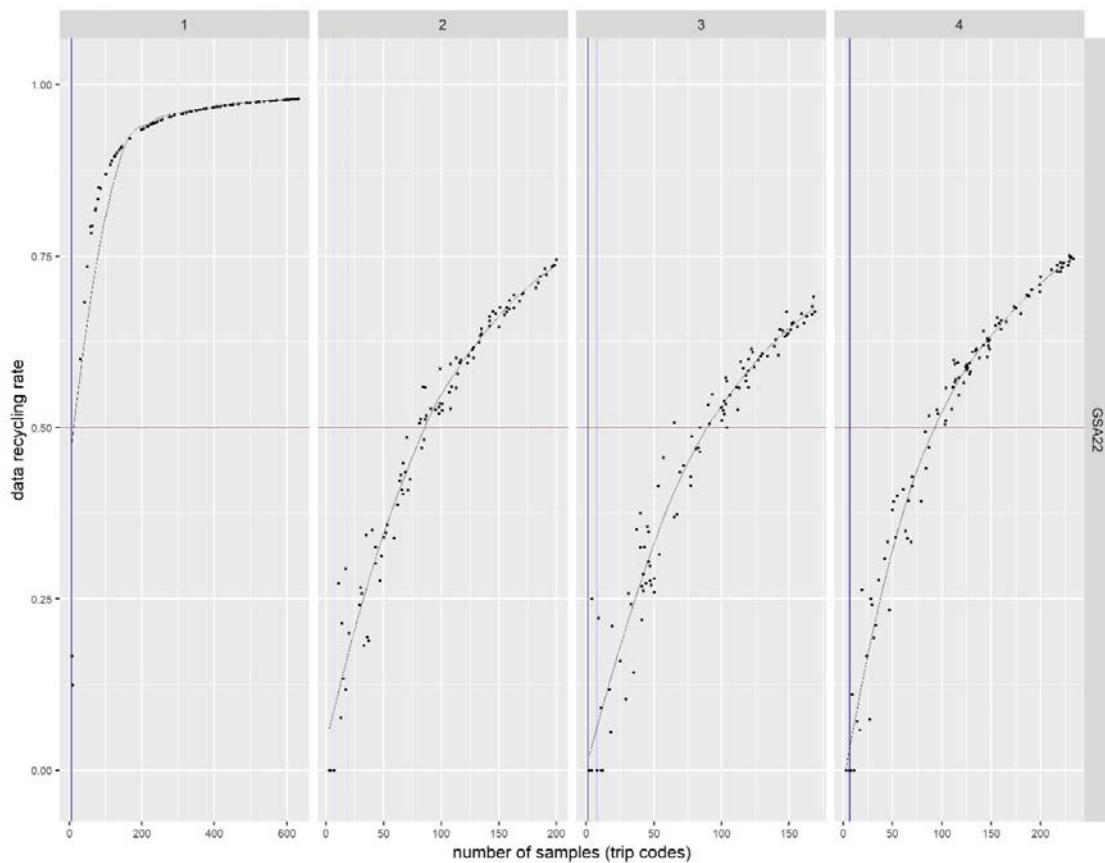


Figure 5.4.3.1 *M. barbatus* in GSA 22 (Greece). Recycling rate of the samples against the number of samples by quarter for GNS_DEF. Vertical blue lines referred to the optimal sampling size range inferred via the method. Red line corresponded to an arbitrary critical threshold of 50%.

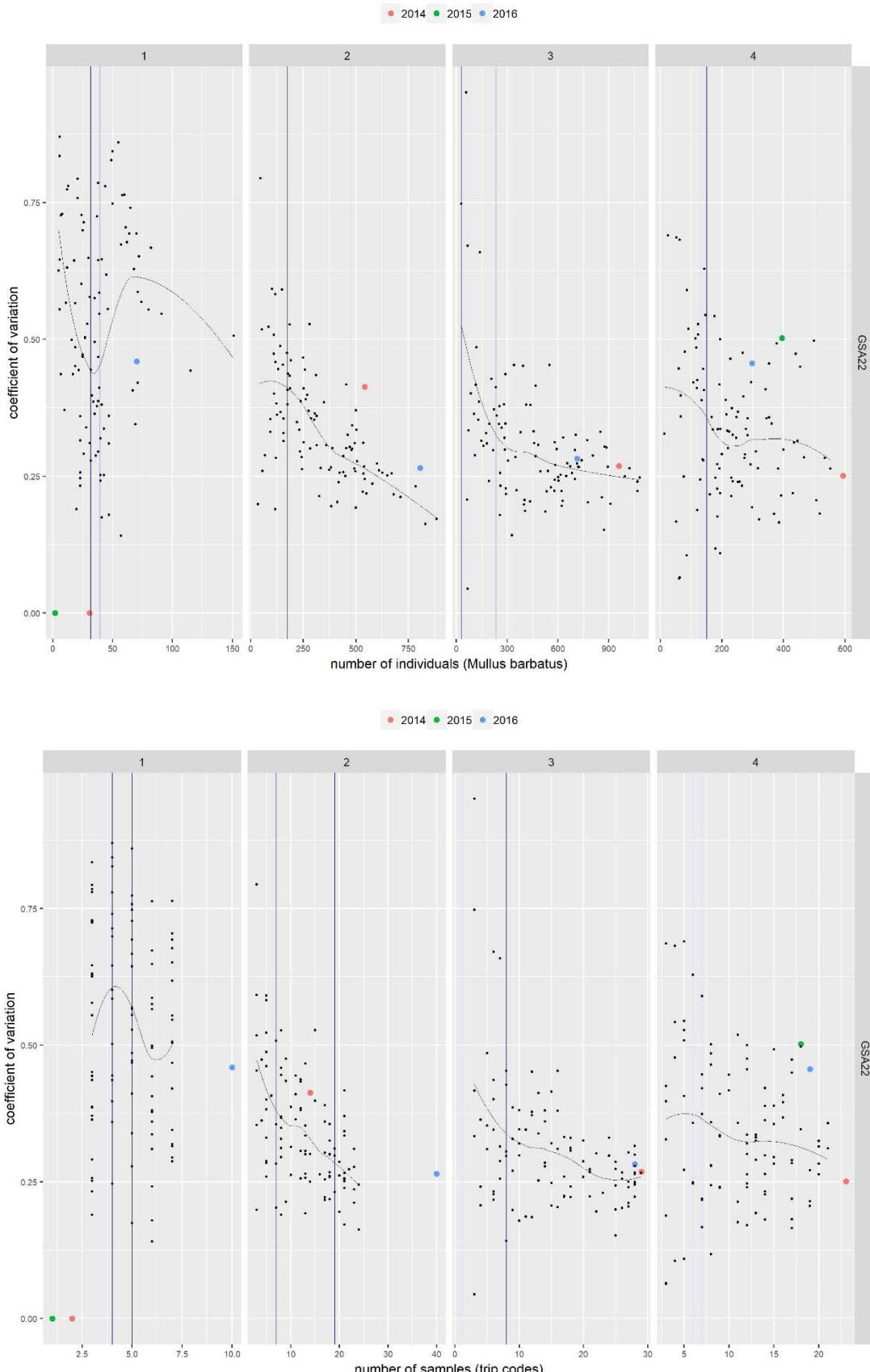


Figure 5.4.3.2 *M. barbatus* in GSA 22 (Greece). CV versus number of samples of (a) individuals and (b) trips by quarter for métier GNS_DEF. The blue lines define the optimal sampling range.

For GTR_DEF the recycling rates for this metier are higher than in the GNS_DEF (Figure 5.4.3.3). In this métier, the realised samplings in the second and third quarter were within or very close to the calculated optimal range while oversampling was recorded again in the fourth quarter in every year (Figure 5.4.3.4).

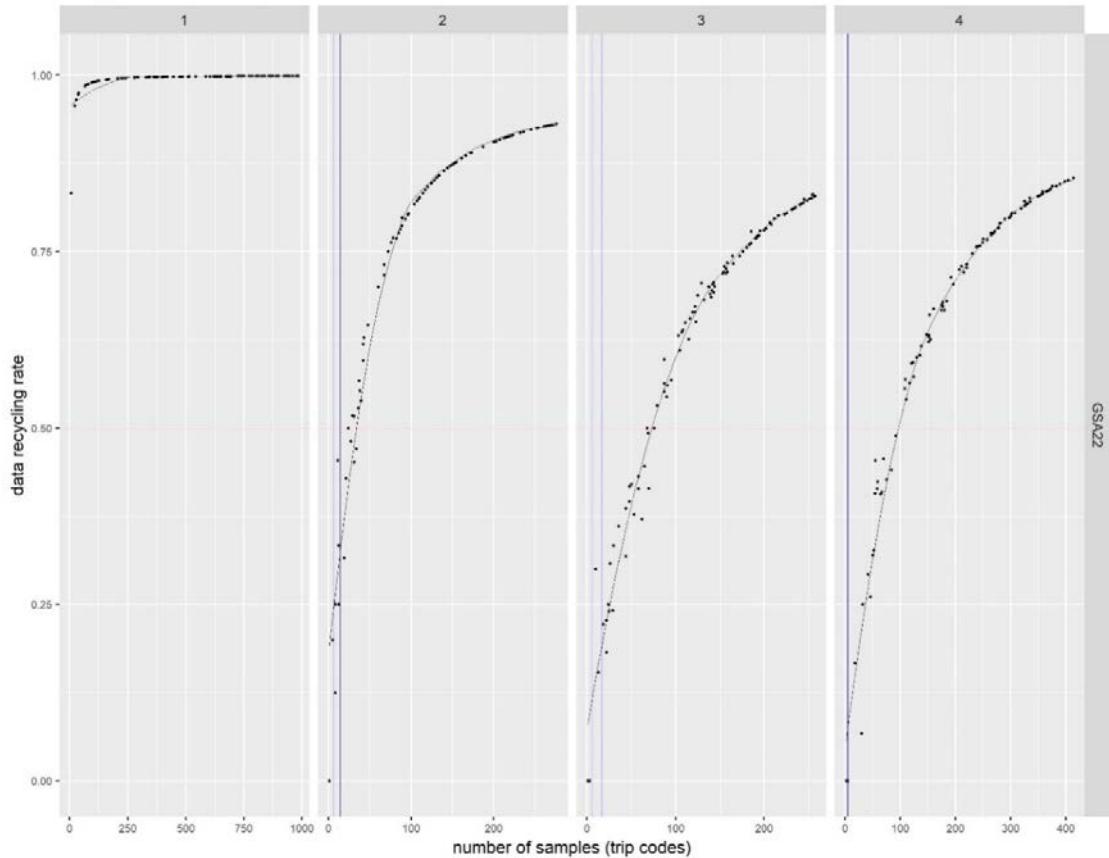


Figure 5.4.3.3 *M. barbatus* in GSA 22 (Greece). Recycling rate of the samples against the number of samples by quarter for GTR_DEF. Vertical blue lines referred to the optimal sampling size range inferred via the method. Red line corresponded to an arbitrary critical threshold of 50%.

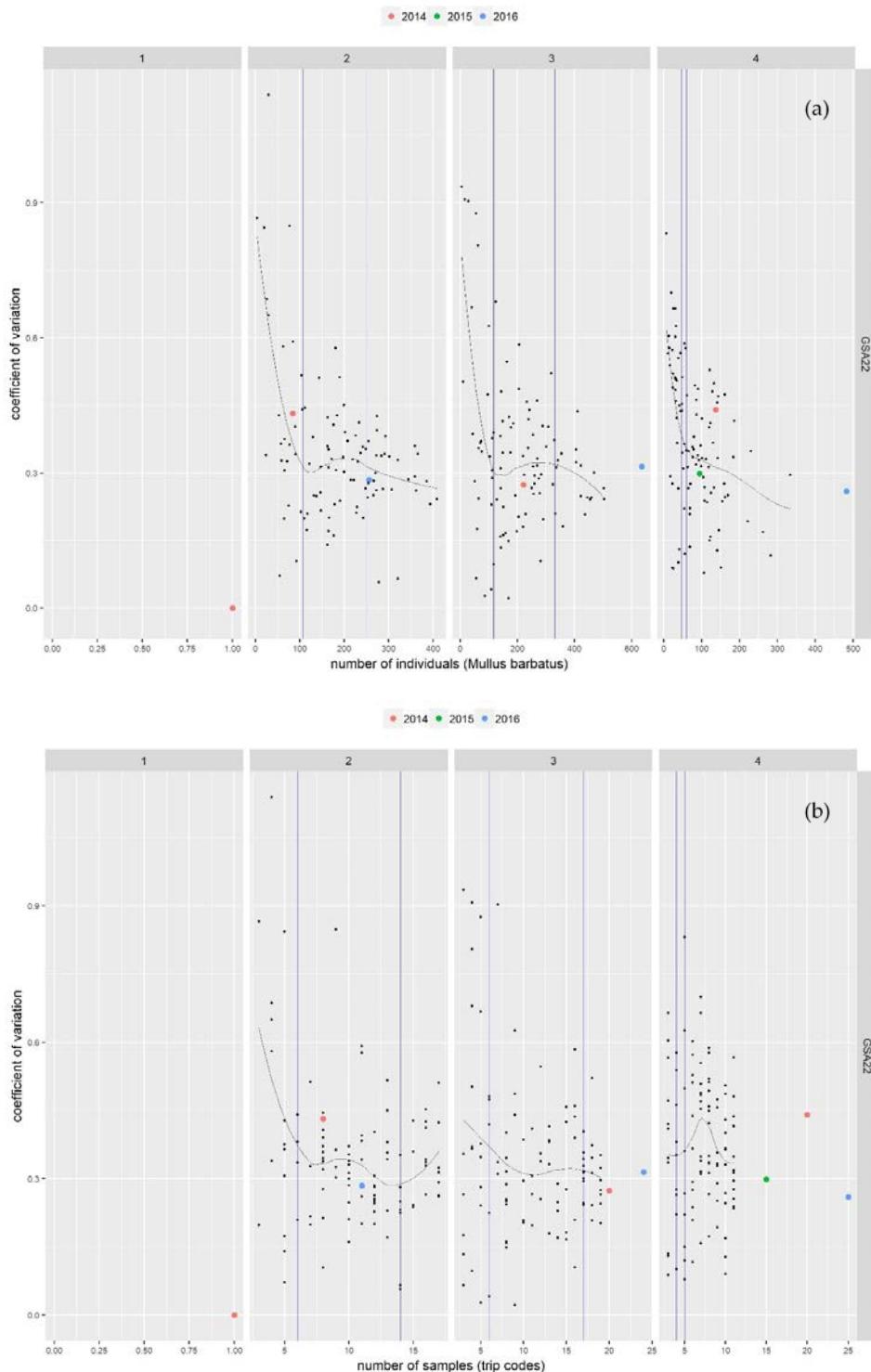


Figure 5.4.3.4 *M. barbatus* in GSA 22 (Greece). CV versus number of samples of (a) individuals measured and (b) trips sampled by quarter for métier GTR_DEF. The blue lines define the optimal sampling range.

Finally, in OTB_DEF, high recycling rate was found for the first quarter, moderate for the second and low for the fourth quarter (Figure 5.3.3.5). It has to be noted that until 2016, Greek trawlers were obliged by the law to stay inactive between May and October and as a result no sampling in OTB_DEF has been done in the third quarter during 2014–2016. In any case, generally, under sampling has been recorded in the first quarter, and oversampling has been recorded for in the

fourth. In the second quarter, sampling was, more or less, within the optimal range (Figure 5.4.3.6).

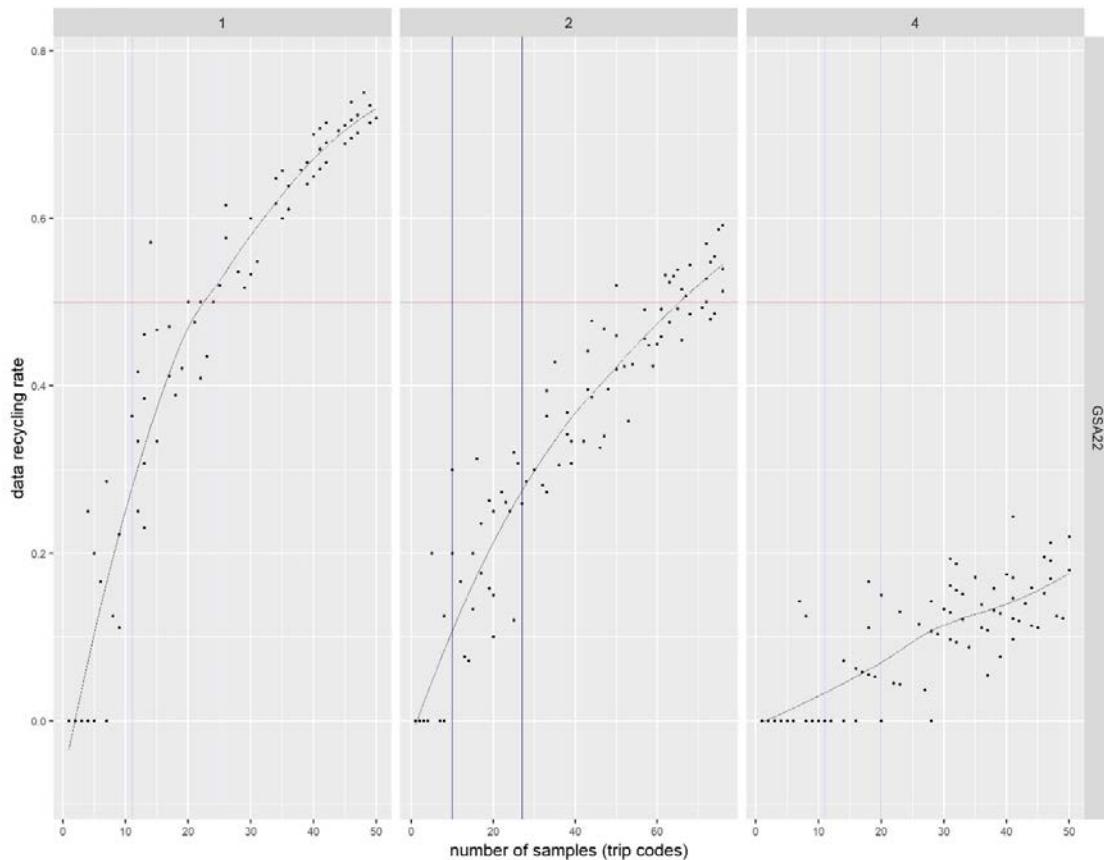


Figure 5.4.3.5 *M. barbatus* in GSA 22 (Greece). Recycling rate of the samples against the number of samples by quarter for OTB_DEF. Vertical blue lines referred to the optimal sampling size range inferred via the method. Red line corresponded to an arbitrary critical threshold of 50%.

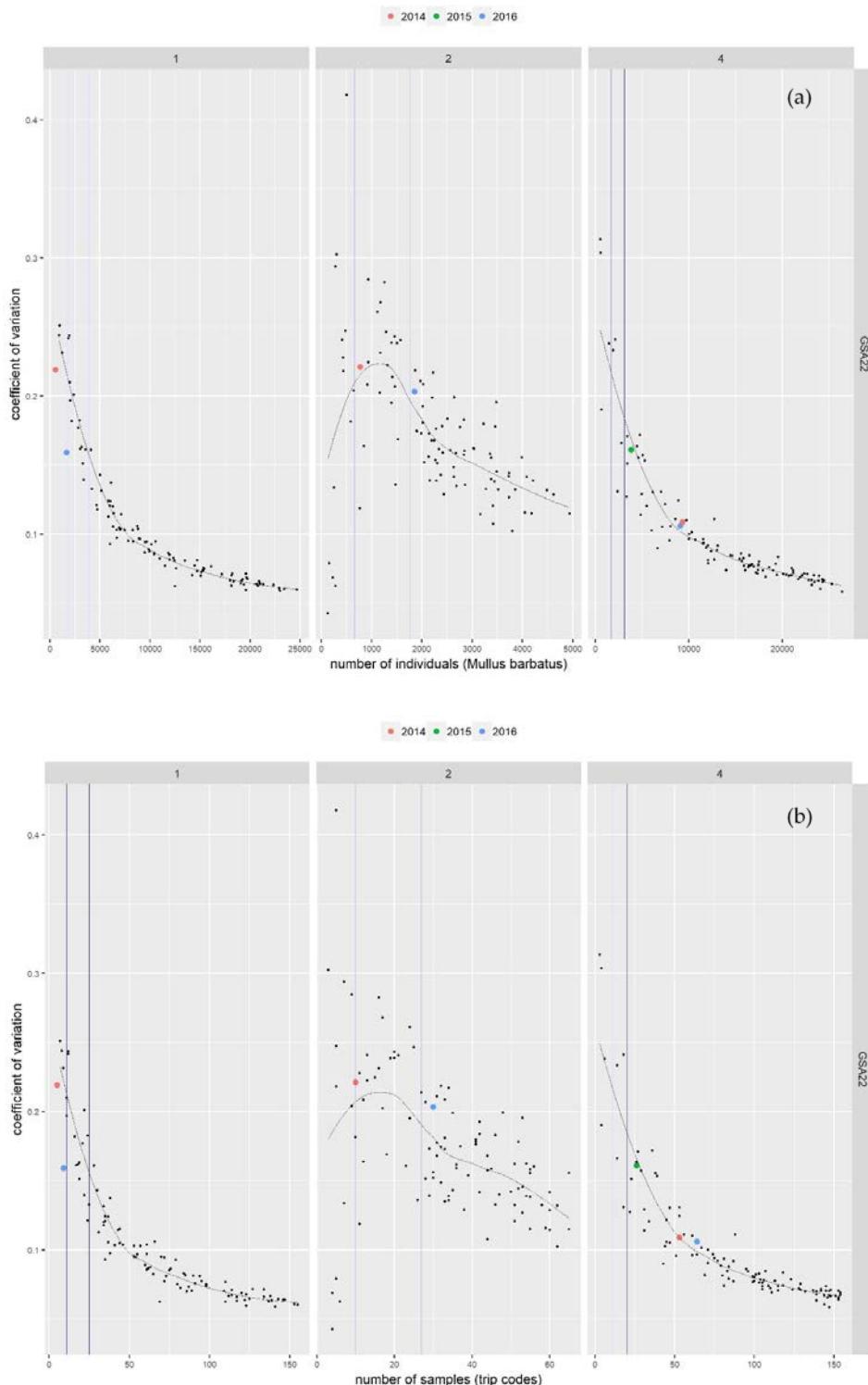


Figure 5.4.3.6 *M. barbatus* in GSA 22 (Greece). CV versus number of samples of (a) individuals and (b) trips by quarter for métier OTB_DEF. The blue lines define the optimal sampling range.

The comparison between the realized and optimal samplings calculated by the applied method (Table 5.4.3.1, figures 5.4.3.7, 5.4.3.8 and 5.4.3.9) indicates that in general terms, the sampling is not well distributed among the four quarters of the year, leading to under sampling in the first and oversampling in the fourth quarter. In the second and third quarter, the GNS_DEF métier is also significantly oversampled.

Table 5.4.3.1 *M. barbatus* in GSA 22 (Greece). Sampling design in terms of number of measured individuals and trips by fishing activity and quarter.

Quarter	metier	Individuals		Trips	
		Optimized mean	Optimized st.dev.	Optimized mean	Optimized st.dev.
1	GNS_DEF	35	19	4	5
	GTR_DEF	NA	NA	NA	NA
	OTB_DEF	2821	789	11	25
2	GNS_DEF	338	147	7	19
	GTR_DEF	179	87	6	14
	OTB_DEF	1233	463	10	27
3	GNS_DEF	140	103	1	8
	GTR_DEF	221	111	6	17
4	GNS_DEF	136	45	6	7
	GTR_DEF	60	38	4	5
	OTB_DEF	2485	604	11	20

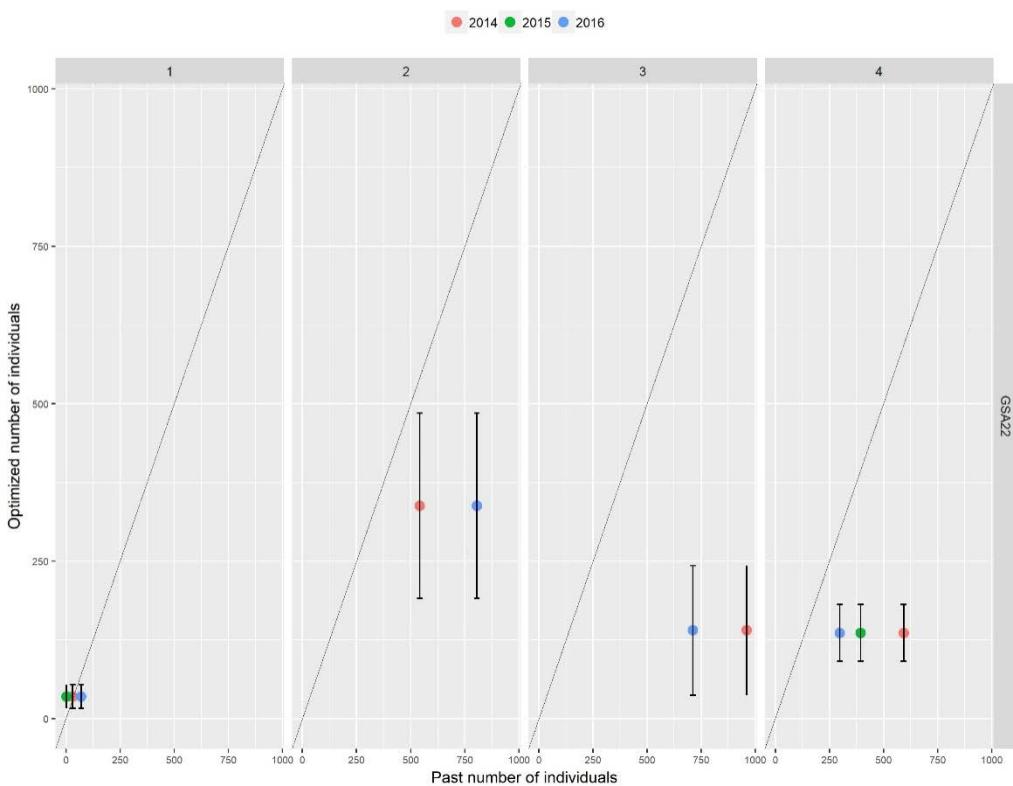


Figure 5.4.3.7 *M. barbatus* in GSA 22 (Greece). Optimized versus past number of individuals for GNS_DEF.

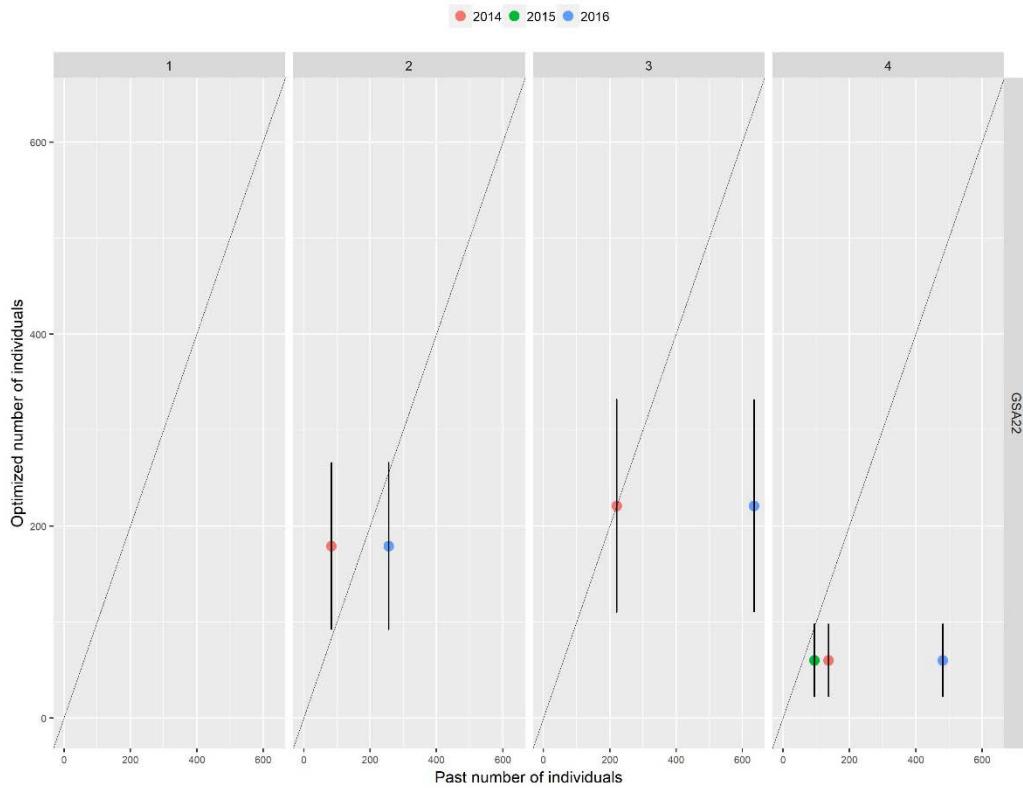


Figure 5.4.3.8 *M. barbatus* in GSA 22 (Greece). Optimized versus past number of individuals for GTR_DEF.

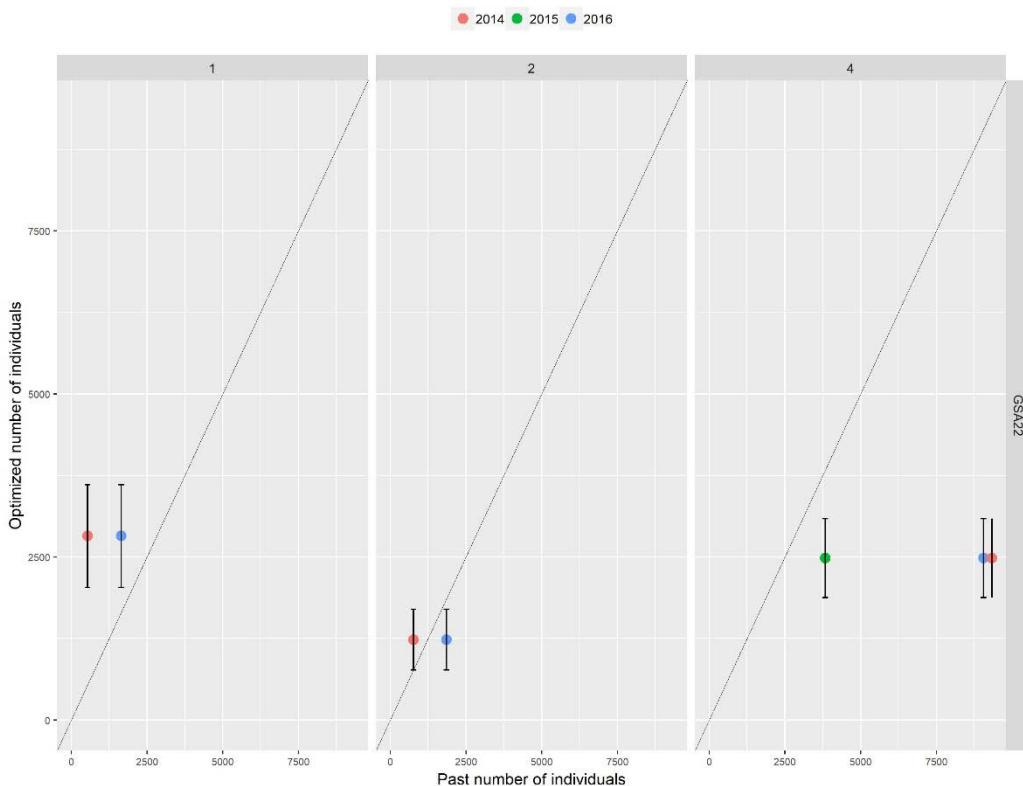


Figure 5.3.3.9 *M. barbatus* in GSA 22 (Greece). Optimized versus past number of individuals for OTB_DEF.

5.4.4 Next steps

Given the multispecies sampling scheme, a potential reduction of trips has to ensure data quality for all species targeted by each métier. Regarding the number of individuals, we consider that also the other biological variables measured (sex, weight, maturity and especially, age) should be taken into consideration.

6 ToR c - Discuss progress achieved in implementation at national level since WKBIOPTIM 1

During the workshop presentations on the optimization procedures tested, used and/or adopted by national institutes were showed. They are listed and presented in the Annex 3 and a summary of their work is here below.

6.1 Sampling optimization of Blue whiting - ICES 27.9.a (Patrícia Gonçalves – IPMA, Portugal)

Blue whiting is under the Portuguese sampling programme. This species length and age data is used as input into the stock assessment model. In order to evaluate if the current blue whiting sampling level in ICES Division 27.9.a have space to improvements, a series of bootstrap simulation were conducted. The main aiming of the simulation process was to address the following questions:

- a) Concerning the blue whiting length sampling for commercial vessels at the harbour: could the length distribution on samples be the same if the number of measured fish are reduced?
- b) When sampling for biological parameters what should be the length distribution of our subsample?
- c) How many otoliths should be read by length class to construct an age-length key?

In (c), the bootstrapping simulations were based on blue whiting sampling data from 2004 (total n = 907) and 2008 (total n = 1353).

The simulation results revealed that the effective number of individuals measured for catch composition could be reduced, without compromise the catch length distribution (a).

For allowing the length distribution proportionality of the subsample to the original sample, an application for Android was been developed, which will help to determine the number of individuals at each length class that should be processed to obtain data on biological parameters (b). This Android application, called BIOWHB is under tests in IPMA. Soon, will be made freely available at *github*.

A number of simulations have been made to test the different options considering the number of otoliths by length class used on the construction of ALKs (c). The results revealed that a lower number of otoliths by length class and sample could have no significant changes on the annual ALKs. Furthermore, the ALKs from the different simulated scenarios were used to obtain the annual catch number at age and a comparison between them was been made (c). Although, the obtained results seem promising this simulation procedure should be repeated using data from more years, before modify the actual number of age reading otoliths by length class. Included in the next steps, are also testing the ALKs from the different scenarios directly on the stock assessment model (SAM).

6.2 How to determine sample size for biological parameters (Esha Mohamed - SLU Aqua, Sweden)

A presentation was made that demonstrated the use of the formula for calculating sample size (Thompson 2012, Cochran 1977, Lohr 2010, Chaudhuri 2014) to herring data that was obtained from onboard commercial sampling (see slides for the formula and their description). The aim was to determine parameter values that will give a distribution of biological parameters (say length and age) with a given level of precision.

A range of values for each parameter α and d which gave different values of the required minimum sample size were used in the simulation. Results showed that the margin of error d has a larger influence on the sample size. This behavior has already been reported in literature. From the simulation results, larger values of d resulted in the distribution of length classes with lots of gaps hence not portraying the “true” distribution of length classes for herring. In conclusion, the values of d and α can be determined by the amount of resources available (such as of time, funds, personnel), level of precision desired etc.

References

- Chaudhuri, Arijit. 2014. *Modern survey sampling*. CRC Press
 Lohr, Sharon. 2010. *Sampling: Design and analysis*. Second edition. Brooks/Cole cengage learning
 Thompson, Steven. 2012. *Sampling*. Third edition. Wiley
 Cochran, William. 1977. *Sampling techniques*. Third edition

6.3 Estimating uncertainty of North Sea IBTS Indices (Natoya O.A.S. Jourdain - IMR, Norway)

The North Sea International Bottom Trawl Survey (IBTS) was started by the International Centre for the Exploration of the Sea (ICES) in 1990. Seven research vessels using standardized fishing methods participates in the survey. The survey with these vessels, which allows fishing also on rough ground provides information on seasonal distribution of stocks, abundance, hydrography and the environment which is then used for stock assessments. Estimates of abundance indices based on age-length keys (ALK) are provided without any assessment of their accuracy. We present a model-based ALK estimator, and a stratified design-based ALK estimator for estimating abundance at age. Both estimators take into account the spatial differences in age-length structures. These estimators are compared with the designed-based ALK estimator proposed by ICES for IBTS, which does not account for spatial differences in the age-length structure. As the proposed ALK estimator by ICES is a combination of age data over a large area, this can result in strongly biased estimates of numbers-at-age. An example of cod (*Gadus morhua*) in ICES subareas 4.a and 4.b is used to illustrate spatial differences in the proportions of age-at-length, and estimates of uncertainty are presented using nonparametric bootstrapping. In general, the model-based ALK estimator provides a more accurate coverage probability compared with the other estimators.

6.3.1 ALK Estimators

6.3.1.1 DATRAS ALK Estimator

- a) Datras assumes that the age-length compositions are homogeneous over relatively large areas

- b) The ALK is an aggregation of individual samples from a haul combined over a RFA. Violations of assumption will give bias results.
- c) A single ALK is produced for ICES round fish areas (RFA) in the North Sea

There are no estimates of variance for the abundance indices of catch-at -age, but the bootstrap procedure is as follows

Pool all hauls in a RFA and sample with replacement, placing hauls in the relevant statistical rectangle. For the new sample, a haul from a different statistical rectangle can be placed in a statistical from which it did not originate, hence, the location of the trawl hauls is not preserved. Hence, we propose a ***stratified bootstrap procedure***:

Sample all hauls in each statistical rectangle in a RFA with replacement and the new sample is placed in the relevant statistical rectangle. This preserve both the location of the trawl and the age observations within each length class.

1. Haul-based ALK

- a) Assumes variation in the age-length structures within a larger area, for example ICES RFA. The plot below shows clear variation in age-length structures for a cod of age 2 given that it is length 40 cm.
- b) An ALK is produced for each trawl, and the spatial variation in the data is accounted for.

The variance is estimated using nonparametric *stratified* bootstrapping approach described above.

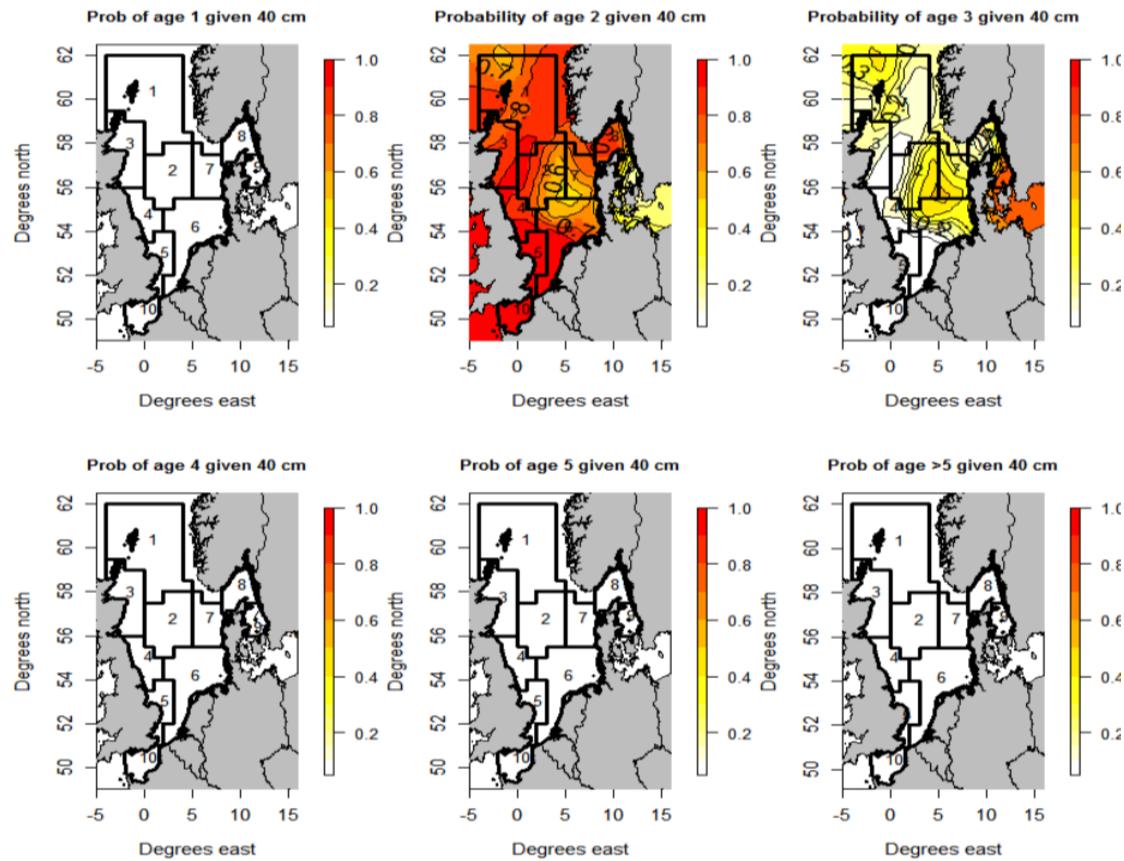


Figure 6.3.1 Estimated probability of age a 40-cm cod in the first quarter of year 2015. The probability that the fish is of age one, three or older is approximately zero. The polygons marked 1 to 10 are the round fish areas (RFAs) where the ALK is assumed to be constant in the currently used estimators of the official CPUEs.

2. Model-based ALK Estimator

- Spatial model-based ALKs are widely used in fisheries (Berg and Kristensen, 2012; Gerritsen *et al.*, 2006). Statistical models
 - create a distribution of age given length and possibly other covariates such haul location
 - fill in of missing values in a more objective and robust manner,
 - accounts for uncertainty arising due to sampling variability
- We consider Logits: a type of model for categorical response data (e.g., age groups)
- Model
- $\Pr(\text{age} \mid \text{length}, \text{location}, \text{haul}) = \text{function}(\text{length}_{\text{age}} + \text{location}_{\text{age}} + \text{haul}_{\text{age}})$
 - location_a : Will capture spatial variation in the ALK
 - haul_{age} : Will capture trawl haul variation e.g., a haul made may “hit” a school of fish of a certain age
- Provides an ALK for each trawl haul

The variance is computed using the *stratified* bootstrap procedure described above. The uncertainty in the ALK is considered by the model (Logit) and the variance-covariance matrix is extracted from the estimated model in TMB. This preserves the positions of trawl hauls in statistical rectangles.

6.3.2 The North Sea cod data

Brief description of cod data in Q1 of 2015

Table 6.3.2.1. Summary of North Sea IBTS cod data for the first quarter of 2015

Data	Description
Trawl hauls	Total of 387 trawl hauls (303 with age information of cod)
Age	The age of cod varied between 1 to 8 years.
Length	Length information in cm of each cod varied between 8 to 112 cm
Date	Date of catch varied between 13.01.2015 to 19.02.2015
Statistical rectangle	The stratum in which at least two trawl hauls are made
Coordinates	Geographic coordinates of each trawl haul in a statistical rectangle
Duration of haul	Mean duration is 25.9 minutes, with 15 to 30 minutes as 90% coverage interval.
Total count per age (Total _{age})	3017 ₁ , 2629 ₂ , 1773 ₃ , 1051 ₄ , 460 ₅ , 194 ₆ , 58 ₇
Total count for all ages	7605 cod in the first quarter of 2015.

Table 6.3.2.2 shows the fraction of trawl hauls with length recordings that also had age recordings from 2010–2017. In 2015, 89% of the trawl hauls with length observations also had an age observation. Conditioning on $l > 50 \text{ cm}$ the percentage is higher. The probability generally increased over the years, almost 10% between 2010–2017.

Table 6.3.2.2. Fraction of trawl hauls with length recordings that also had age recordings.

Year	cod	
	Pr($a l$)	Pr($a l > 50 \text{ cm}$)
2010	86.4 %	83.2 %
2011	90.7 %	92.4 %
2012	87.6 %	92.0 %
2013	88.1 %	93.0 %
2014	89.8 %	93.7 %
2015	89.3 %	93.8 %
2016	93.0 %	93.7 %
2017	95.5 %	98.8 %

Table 6.3.2.3 shows lower catch rates for larger age-length groups and higher catch rates for smaller, younger fish. Table 6.3.2.4 gives the number at ALK for cod in the first quarter for all years.

Table 6.3.2.3 Age-length composition of cod and number at ALK in Q1 of 2015.

Age (a)	NoAtALK	Length (<i>l</i> in cm)
1	460	9.0 - 38.0
2	1191	16.0 - 63.0
3	676	24.0 - 84.1
4	284	30.0 - 93.0
5	101	52.0 - 94.2
6	63	62.0 - 104.0
7	12	75 - 98
8	1	113

Table 6.3.2.4 Age-length composition of cod and number at ALK in Q1 for all years.

Age (a)	NoAtALK	Length (<i>l</i> in cm)
1	1423	8.0 - 38.0
2	1561	16.0 - 63.0
3	1099	24.0 - 84.1
4	665	30.0 - 93.0
5	293	51.0 - 94.2
6	128	57 - 105.0
7	34	66 - 112
8	9	60 - 113

Results

- Model-based ALK generally performs better in terms of uncertainty estimation. Accounts for Spatial differences in age-length structures
- DATRAS procedure generally gave smaller estimates of uncertainty as it lacks the potential to account for spatial variation in the data
- Estimated CPUE at age is captured within a 95% CI for all methods

Table 6.3.2.5 Estimates of abundance indices ($mCPUE_{1,a}$) for cod in RFA 1 in the Q1 of 2015. Estimated standard error estimates of the approximate 95% confidence intervals (CI).

Age (a)	DATRAS ALK DATRAS Bootstrap			Haul-based ALK Stratified Bootstrap			Model-based ALK Stratified Bootstrap		
	Abundance estimate	Standard error	Relative standard error	Abundance estimate	Standard error	Relative standard error	Abundance Estimate	Standard error	Relative standard error
0	0	0	—	0	0	—	0	0	—
1	0.764	0.26	34%	0.60	0.24	40%	0.70	0.36	51%
2	21.989	6.76	31%	22.21	4.15	19%	22.11	4.28	19%
3	11.285	2.19	19%	10.58	1.20	11%	10.99	1.77	16%
4	3.265	0.71	22%	3.67	1.28	35%	3.50	0.87	25%
5	1.147	0.34	30%	1.27	0.42	33%	1.20	0.48	40%
6+	1.276	0.38	30%	1.40	0.70	50%	1.21	0.42	35%
Approximate 95% CI from bootstrap procedures									
0	0	(0, 0)		0	(0, 0)		0	(0, 0)	
1	0.764	(0.31, 1.33)		0.60	(0.31, 0.91)		0.70	(0.35, 1.48)	
2	21.898	(12.73, 37.15)		22.21	(15.64, 30.72)		22.11	(14.76, 30.36)	
3	11.285	(6.31, 15.02)		10.58	(8.74, 13.65)		10.99	(8.61, 15.42)	
4	3.265	(1.49, 4.21)		3.67	(2.81, 4.74)		3.50	(1.96, 5.60)	
5	1.147	(0.40, 1.75)		1.27	(0.67, 2.31)		1.20	(0.56, 2.78)	
6+	1.276	(0.44, 1.82)		1.40	(0.78, 2.69)		1.21	(0.70, 2.43)	

Planned work

- Include trawl haul in the model-based ALK estimator
- Derive an abundance-at-age estimator for the whole North Sea, and its variance estimator
- **Compare ALK estimators**
 - Use ALK estimators in an assessment model, e.g. SAM or XSAM
- **Optimize sampling effort:** Removal of trawl hauls and otoliths to determine if there is any effect on the variance
 - As a means of justifying sampling effort e.g., number of days at sea and number of stations sampled and number of samples taken
 - Consider Hierarchical bootstrapping approach (**completed**)
- **Fully model-based approach for estimating abundance at age**

References

- Berg, C. W. and Kristensen, K.. 2012. Spatial age-length key modelling using continuation ratio logits. *Fisheries Research*, 129:119-126
- Gerritsen, H., McGrath D. and Lordan C.. 2006. A simple method for comparing age-length keys reveals significant regional differences within a single stock of haddock (*Melanogrammus aeglefinus*). *ICES Journal of Marine Science*, 63(6):1096-1100.
- ICES. 2013. NS-IBTS indices calculation procedure. Datras procedure document. International Council for the Exploration of the Sea, 1.1 NS-IBTS indices-2013.

6.4 Optimization of Reference Sampling of *Pandalus borealis* (Nuno Prista and Annelie Hilvarsson - SLU Aqua, Sweden)

During WKBIOPTIM 1 (May–June, 2017), the sample-level scripts were applied to a case-study involving Northern shrimp *Pandalus borealis*. The original data consisted in individual measurements of shrimps from commercial catch samples taken in 2016 by onboard observers of the Swedish shrimp fishery. A total of 64 samples were available involving 229 to 555 length measurements each (average 396 measurements). The objective of the case-study was to explore the consequences of a reduction in the number of measurements as these appeared to be quite large and involve a workload in the lab that could likely be reduced.

In the follow-up of WKBIOPTIM meeting, the sample-level script was further developed and the case-study revised. Then an internal process was started at H-lab (SLU - Aqua) to discuss results. This process involved a consultation between data collectors, data analysts and end-users and culminated in a proposal for a new sample size. An overview of the process is given in Figure 6.4.1. A main meeting to discuss results was held in the end of August 2017. The result from the joint evaluation of the sample level results was a suggestion for a reduction of 25% in the number of measurements taken per sample. Such reduction was expected to result in 200–300 shrimps' measurements per sample. These numbers were judged sufficiently conservative and able to deliver a reduction in workload while securing "good enough" length frequencies for most end-uses of data (both present and future). A joint evaluation of these changes is expected to take place in late 2018.

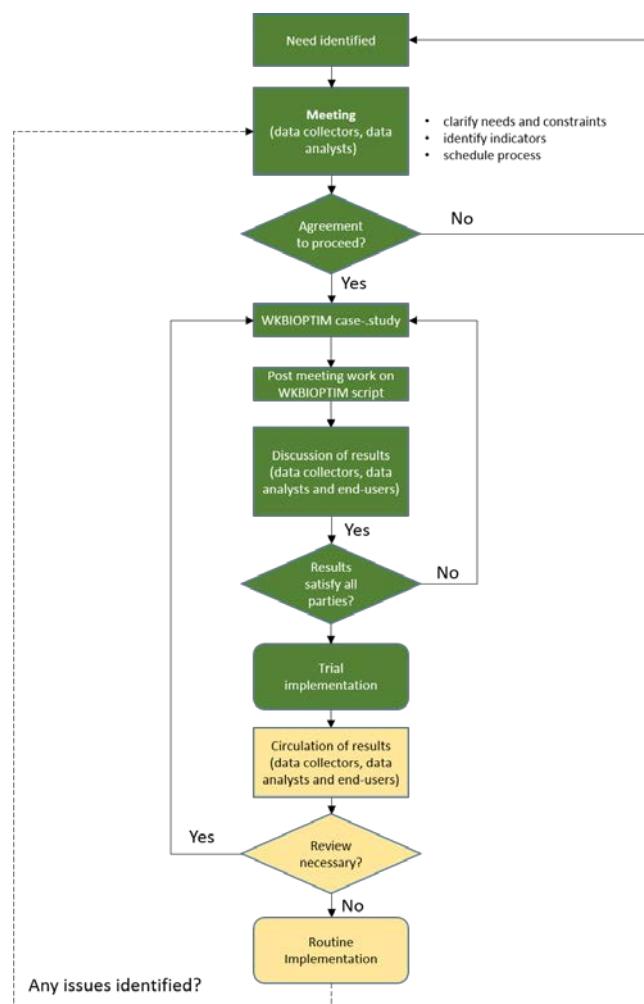


Figure 6.4.1 Suggested roadmap for discussion of optimization processes.

7 Conclusions and future work

The new EU requirements in terms of data collection and advice (e.g., statistically sound sampling schemes, determination of MSY reference points for previously unassessed stocks, by-catch recordings in at-sea sampling, collection of multispecies variables to use in Integrated Ecosystem Assessments) and temporary or permanent budget limitations in many fisheries labs are main drivers for the optimization of the sampling protocols for biological data. From a cost-benefit perspective it is also not defensible that human resources and funds are used to collect more data than what is judged reasonably necessary for the purpose at hand. It is WKBIOPTIM's objective to provide tools for institutes to evaluate strategies for the optimization of their biological sampling programmes, namely those carried out under their EU-MAP work plans.

During WKBIOPTIM 2 participants presented results from the application of tools developed in WKBIOPTIM 1 and discussed other possible approaches for optimizing both for commercial and survey data. Regarding surveys, the work presented and discussed during the workshop resulted in the conclusion that maybe some more input on possible quality indicators for surveys is needed so that survey optimization can be better accommodated in the scripts being developed at WKBIOPTIM, and that stock assessors have a main role in this part.

The practical work of WKBIOPTIM has been divided in two different levels of optimization with different sets of r-code: the sample level and the multilevel sampling. The preparation of the main part of the scripts is performed and discussed prior the workshop between the code developers and the chairs and, during the workshop, further adaptations are performed to accommodate suggestions and improvements proposed by the participants. As an example, during WKBIOPTIM2 the sample-level scripts originally developed by WKBIOPTIM 1 were extended to datasets involving multivariate data collection. So far the results have been explored for the case of numerical variables ("length" and "age") and two "indicators" of quality (MWCV and CV of the mean). An extensive amount of other statistical results is outputted by the scripts that can be analysed. The present setup rends the WKBIOPTIM 2 scripts readily applicable in cases where biological data was randomly sampled and the scripts have already been applied in at least one national lab with sampling goals being significantly optimized (e.g., reference samples of *Pandalus borealis* in Sweden). However, for more widespread application the scripts will need to be extended to other stratified designs, other categorical variables and a possible wider range of quality "indicators". In the case of categorical variables, such as sex and maturity, the present code is equipped to provide a set of outputs (e.g., sex-ratios, length at 50% maturity) but these still need to be explored and summarized. Ultimately an integration of the results of the several indicators and types of variables is needed since the objectives of data collection are frequently multipurpose. Alternatives to the worst-case scenario approach to determine minimum sample size should also be considered as the approach is restricted to the universe of simulated sample sizes and only really applicable to situations where the sample sizes of simulated samples are particularly high. One possible alternative is to model the MWCV and RSE using, e.g., an exponential model. Such modelling approach was developed in the follow-up of WKBIOPTIM 2 meeting and requires further discussion so its results should for now be viewed with caution. The next steps to include in the sample level optimization will involve: scripts in a way that allows for simulations and results from datasets (e.g. some surveys) where some biological variables (e.g., age) were originally stratified by another variable (e.g., length class); finalisation and incorporation of remainder sampling strategies (2-stage random sampling, stratified random sampling by length) in the script; inclusion of additional indicators and analyses proposed in this workshop (see Section 3.1); and further discuss approaches for multivariate summaries and conclusions on sample size; further discuss application of the script to smaller (data-poor) samples.

Concerning multilevel sampling optimization, part of the work developed this year was to assess sampling plan optimization in a concurrent sampling framework. Results obtained in presented case studies highlight the complexity to use bootstrapping simulation to improve the sampling plan in the case of concurrent sampling mainly due to “rare” species events. The procedure shows that a decrease in number of trips sampled can decrease dramatically the quality of length distribution estimates and can lead to the absence of the “rare” species in a stratum. Regarding this, one of the main metrics to be considered when performing these type of analysis and in order to quantify the loss of information, should be the occurrence of species/stratum, by investigating the number of times it is seen in the simulation replicates. However, this optimization procedure probably may be applied/adapted if the number of species to sample in the sampling framework is conditioned to fewer target species, with higher levels of sampling. Regarding the SD Tool approach, improvements were presented regarding simulations with different technical stratifications, different temporal aggregations and using data of stocks considered shared among MS to get results for the whole area. Different scenarios playing with the number of trips and numbers of individuals measured were simulated. The metrics used to evaluate the quality of the results were the CV and the EMD that combined, indicate an optimal number of trips to sample and individuals to measure according to the best scenario results. The next steps for this tool would be to consider that for multispecies sampling scheme, a potential reduction of trips has to ensure data quality for all species targeted by each métier. And regarding the number of individuals obtained, also other biological variables measured (sex, weight, maturity and especially, age) have to be taken into consideration.

WKBIOPTIM 2 agreed that the scripts and procedures being developed, improved and tested during these workshops should be compiled and documented in a Toolbox so national laboratories and institutes can analyse their own data and improve their resources allocation and/or distribution. The group thinks that the main part of the procedures is ready to be used even if final result will benefit from increments to the code. The work proposed for WKBIOPTIM 3 is to finalize this work, documenting the scripts and case-studies and incorporating the results in an R-Package. To achieve this goal, the work will start prior the meeting in an intersessional way in order to make possible the final preparation, compilation and finalisation of all the tasks proposed.

8 References

- Baumer, Ben, Mine Cetinkaya-Rundel, Andrew Bray, Linda Loi, and Nicholas J. Horton. 2014. "R Markdown: Integrating a Reproducible Analysis Tool into Introductory Statistics." *Technology Innovations in Statistics Education* 8 (1).
- Berg, C. W. and Kristensen, K. (2012). Spatial age-length key modelling using continuation ratio logits. *Fisheries Research*, 129:119-126.
- COMM. IMPL. DEC. (EU) 2016/1701 of 19 August 2016 laying down rules on format for the submission of work plans for data in the fisheries and aquaculture sectors.
- Gerritsen H. D. and D. McGrath. 2007. Precision estimates and suggested sample sizes for length-frequency data. *Fishery Bulletin* 106:116-120.
- GFCM, 2017. GFCM Data Collection Reference Framework (DCRF). Version: 2017.1
- ICES. 2012. Report of the Planning Group on Commercial Catches, Discards and Biological Sampling (PGCCDBS). 18–22 February 2013, Belfast, Northern Ireland. ICES CM 2013/ACOM: 49. 128pp.
- ICES. 2013. NS-IBTS indices calculation procedure. Datras procedure document. International Council for the Exploration of the Sea, 1.1 NS-IBTS indices-2013.
- ICES. 2016a. Report of the Working Group on Data Needs for Assessments and Advice (PGDATA), 30 June-3 July 2015, Lysekil, Sweden. 103 pp.
- ICES. 2017a. Report of the Workshop on cost benefit analysis of data collection in support of stock assessment and fishery management (WKCOSTBEN), 28 June -1 July 2016, ICES Headquarters, Copenhagen, Denmark. ICES CM 2016/SSGIEOM: 12. 41pp.
- ICES, 2017b. Report of the Working Group on Commercial Catches (WGCATCH), 7-11 November 2016, Oostende, Belgium. ICES CM 2016/SSGIEM:03. 142pp.
- ICES, 2018. WKBIOPTIM
- Jansen, Teunis, Henrik Degel, Joel Vigneau, and Ernesto Jardim. 2009. "Definition of Standard Data-Exchange Format for Sampling, Landings, and Effort Data Commercial Fisheries." ICES C OOPERATIVE R ESEARCH R EPORT 296. ICES.
- MARE/2014/19 Med&BS project Final report, 2016 - Deliverable 2.5 - Scripts with a detailed document ensuring the reproducibility of the methods and the analyses realized with a special focus on the case studies tests - F. Gontrand, T. Rouyer, 2016.
- MARE/2016/22 STREAM - Strengthening regional cooperation in the area of fisheries data collection
- R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>. <https://www.r-project.org>
- Report of the 2017 Regional Co-ordination Meeting Mediterranean and Black Sea Large Pelagic Fisheries (in prep)
- Rubner, Y., C. Tomasi, and L. J. Guibas. 1998. "A Metric for Distributions with Applications to Image Databases." *IEEE International Conference on Computer Vision*, 59-66.

Annex 1: List of participants

NAME	ADDRESS	E-MAIL
Ana Cláudia Fernandes	Instituto Português do Mar e da Atmosfera, I.P., Rua Alfredo Magalhães Ramalho, 6 1495-006 Lisboa, Portugal	acffernandes@ipma.pt
Eirini Mantzouni	Fisheries Research Institute, Nea Peramos, Kavala, Greece	emantzo@inale.gr
Esha Mohamed (skype participation)	Swedish University of Agricultural Sciences, Department of Aquatic Resources, Institute of Marine Research, Turistgatan 5, 453 30 Lysekil, Sweden	esha.mohamed@slu.se
Ioannis Ntokos	HCMR - Hellenic Centre for Marine Research, Greece	gdokos@hcmr.gr
Julia Wischnewski	Institute of Sea Fisheries, Palmaille 9, 22767 Hamburg-Altona, Germany	julia.wischnewski@thuenen.de
Katja Norén	Institute of Marine Research, Department of Aquatic Resources (SLU Aqua), Swedish University of Agricultural Sciences	katja.noren@slu.se
Laurent Dubroca	Ifremer, Port-en-Bessin Station, PO Box 32, Avenue du Général De Gaulle, 14520 Port-en-Bessin, France	laurent.dubroca@ifremer.fr
Maria Teresa Facchini	COISPA TECNOLOGIA & RICERCA, Stazione Sperimentale per lo Studio delle Risorse del Mare, Via dei Trulli, 18/20 70126, Bari, Torre a Mare, Italy	facchini@coispa.it
Natoya Jourdain	Institute of Marine Research, P.O.Box 1870 Nordnes, 5817 Bergen, Norway	natoya.jourdain@imr.no
Nuno Prista	Swedish University of Agricultural Sciences, Department of Aquatic Resources, Institute of Marine Research, Turistgatan 5, 453 30 Lysekil, Sweden	nuno.prista@slu.se
Patrícia Gonçalves	Instituto Português do Mar e da Atmosfera, I.P., Rua Alfredo Magalhães Ramalho, 6 1495-006 Lisboa, Portugal	patricia@ipma.pt
Perttu Rantanen	Natural Resources Institute Finland (Luke) Itäinen Pitkäkatu 3, 20520 Turku, Finland	Perttu.Rantanen@luke.fi
Petri Sarvamaa	Natural Resources Institute Finland (Luke) Itäinen Pitkäkatu 3, 20520 Turku, Finland	Petri.Sarvamaa@luke.fi
Sofie Vandemaele	ILVO, Ankerstraat 1, 8400 Oostende, Belgium	Sofie.Vandemaele@ilvo.vlaanderen.be



Annex 2: Agenda

Tuesday 29 th May		
9.00 - 9.30	Welcome and logistics	Chairs
9.30 - 10.00	Presentation of ToR's and adoption of agenda	Chairs
10.00 - 10.30	Presentation on the work on determining minimum sampling size for biological parameters	Esha Mohamed
10.30 - 11.00	Sampling optimization on blue whiting (ICES 27.9.a)	Patrícia Gonçalves
<i>Coffee Break</i>		
11.00 – 13:00	Presentation of developments on Sampling Design tool	Maria Teresa Faccini
	Presentation of new development on code used for optimisation at sample level	Nuno Prista
	Presentation of new development on code used for optimisation at sampling design level	Laurent Dubroca
	Presentation on survey IBTS in Belgium	Sofie Vandemaele
	Presentation on Uncertainty estimation of the North Sea IBTS abundance indices	Natoya Jourdain
<i>Lunch break</i>		
14.00 - 14.30	ToR a. Present /Summarize work done to date on QI's followed by plenary discussion on QI's	Julia Wischnewski
14.30 - 15.00	ToR b & c. Present and decide on case studies and sub-groups	Plenary
15.00 - 16.00	Subgroup work	
<i>Coffee break</i>		
16.30 - 17.30	Subgroup work	
17:30 - 18.00	Wrap up of the day	Plenary
Wednesday 30 st May		
9.00 - 9.30	Discussion of issues related to the scripts usage (e.g. code adaptations)	Plenary
9.30 - 11.00	Subgroup work	
<i>Coffee break</i>		

11.30 - 13.00	Subgroup work	
<i>Lunch break</i>		
14.00 - 16.00	Subgroup work	
<i>Coffee break</i>		
16.30 - 17.30	Subgroup work	
17.30 - 18.00	Wrap up of the day	Plenary
<i>Evening Social Dinner</i>		
Thursday 31st May		
9.00 - 11.00	Presentation and discussion of output/work done by subgroups	Plenary
<i>Coffee break</i>		
11.30 - 13.00	ToR c. Discussion on progress achieved in implementation at national level since WKBIOPTIM 1 (including new developments on the catch-sampling evaluation toolbox)	Plenary
<i>Lunch break</i>		
14.00 - 15.00	Report preparation and discussion of future work	Plenary
15.00 - 16.00	Report writing	Plenary
<i>Coffee break</i>		
16.30 - 17.30	Report writing	Plenary
17.30 - 18.00	Wrap up of the meeting	Chairs

Annex 3: List of oral presentations

1. Patrícia Gonçalves (IPMA, Portugal): Sampling optimization of Blue whiting - ICES 27.9.a



Sampling optimization Blue whiting ICES division 27.9.a

Patrícia Gonçalves (IPMA)

Tasks:

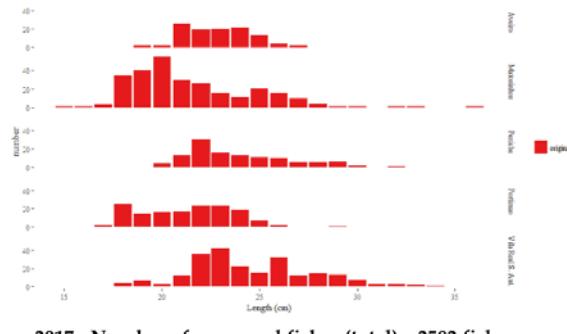
- (a) Commercial sampling – measured individuals (number?);
- (b) Biological parameters – individuals by length class (number?);
- (c) Otoliths/length class (number?) - ALKs.

Sampling BW (auction)

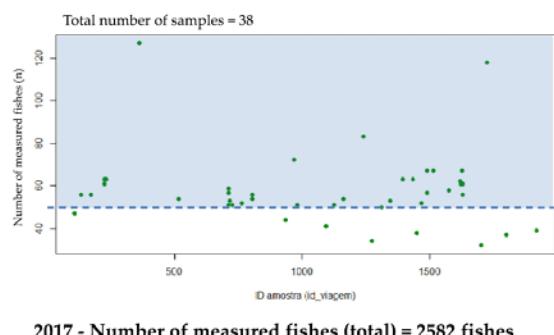
- (a) Sampling for landings length distribution:
number measured by sample?



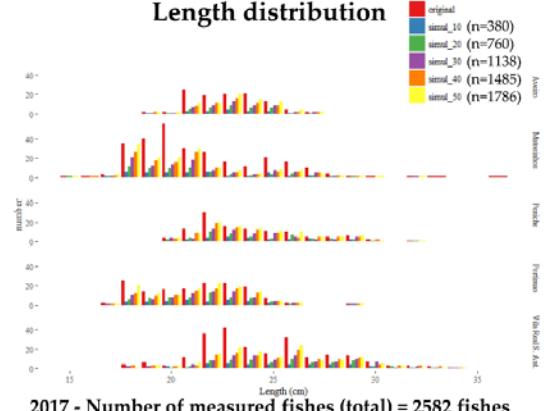
Length distribution



Length distribution



Length distribution



Sampling blue whiting (biology)

(b) Number of individuals by length class?



Biological sampling (lab)

Current procedure (in 5 steps):

1. Individuals are measured
2. Displayed by length class

Biological sampling (lab)

Current procedure (in 5 steps):

1. Individuals are measured

Biological sampling (lab)

Current procedure (in 5 steps):

1. Individuals are measured
2. Displayed by length class
3. Number of individuals in each length - counted
4. Determined the total weight of all individuals
5. 10 individuals by length were selected for biology

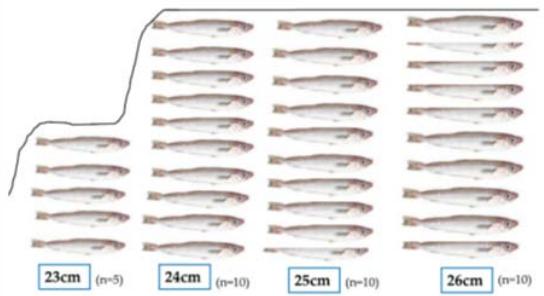
Biological sampling (lab)

Current procedure (in 5 steps):

1. Individuals are measured
2. Displayed by length class
3. Number of individuals in each length - counted
4. Determined the total weight of all individuals

Blue whiting sampling

Length distribution on the sample (?!)



Blue whiting – number of individuals to sample for biology by length class



Blue whiting – number of individuals to sample for biology by length class



app (Android)

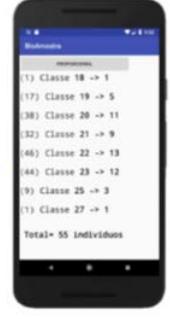
Blue whiting – number of individuals to sample for biology by length class



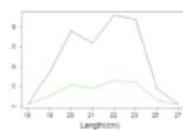
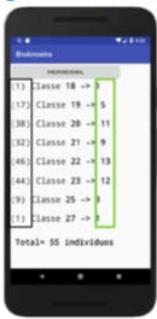
Blue whiting – number of individuals to sample for biology by length class



Blue whiting – number of individuals to sample for biology by length class



Blue whiting – number of individuals to sample for biology by length class



Blue whiting – number of individuals to sample for biology by length class

2017

Total number of individuals sampled (10 ind/length class)

1557

Total number of individuals sampled Bio-WHB app in Android

1460



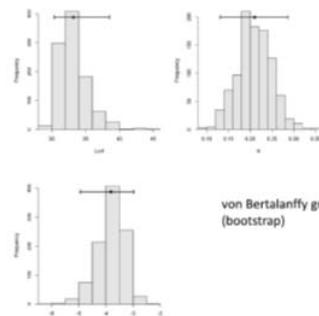
- 97 indiv.

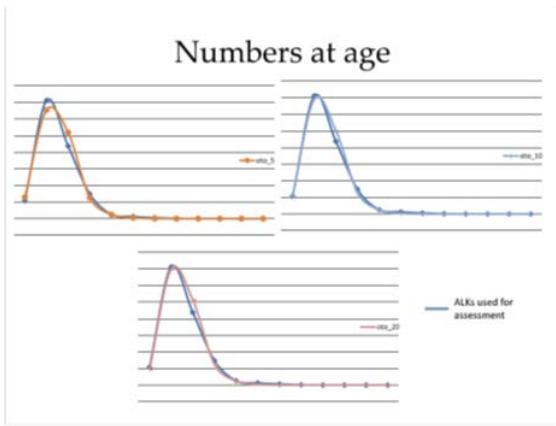
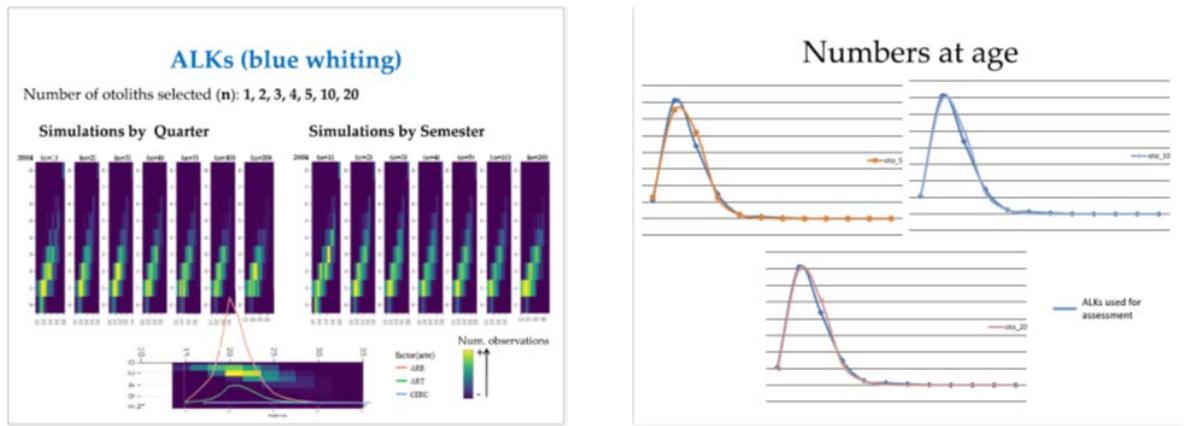
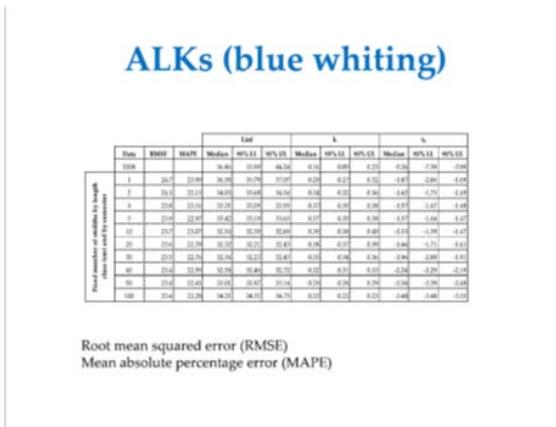
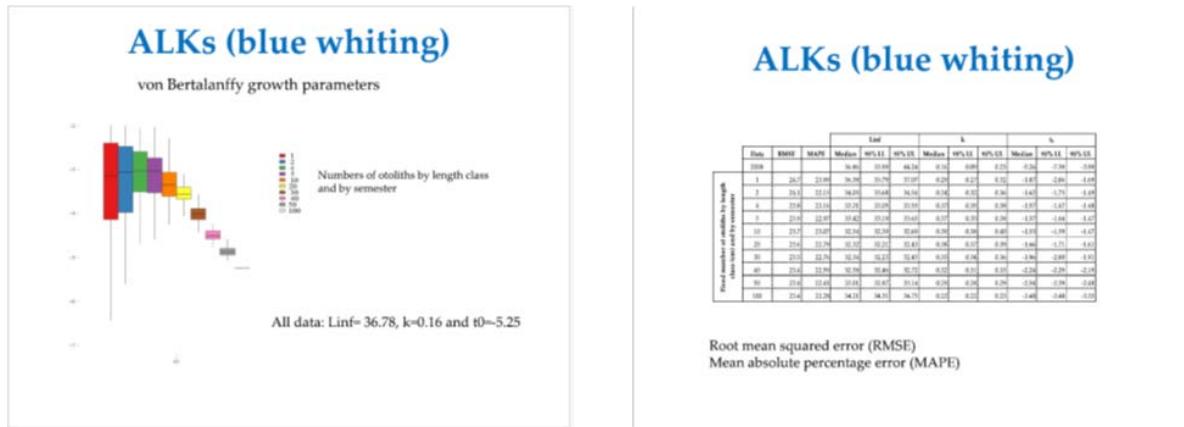
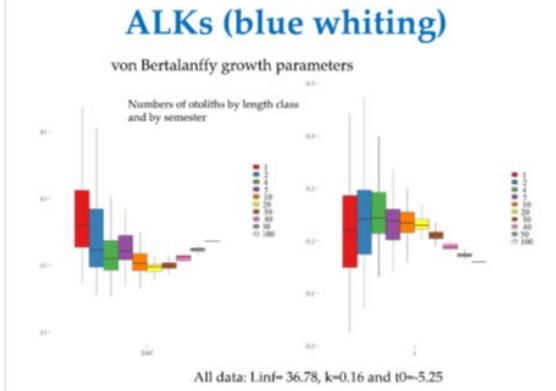
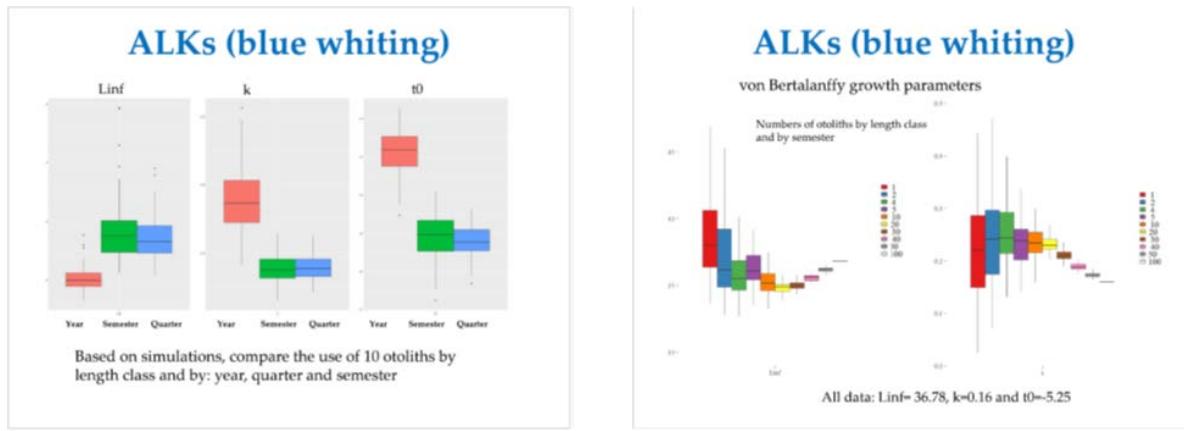
Sampling blue whiting

(c) ALKs (number of otoliths by length class?)

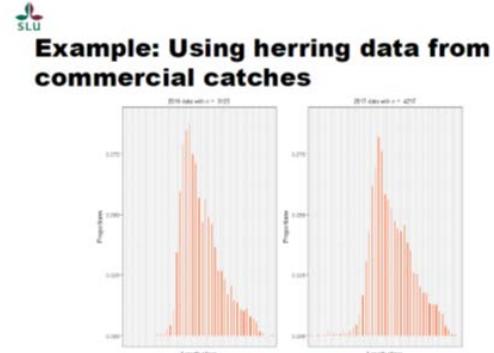
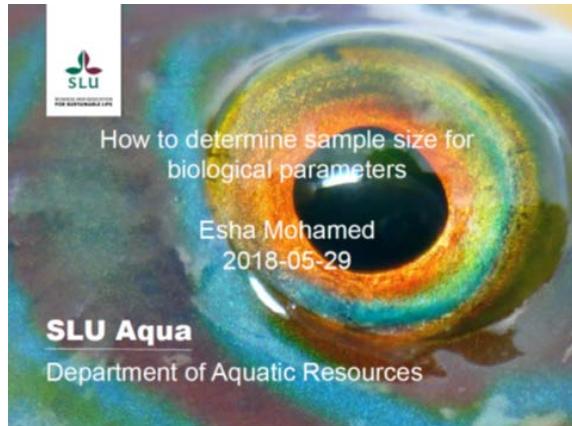


ALKs (blue whiting)





2. Esha Mohamed (SLU Aqua, Sweden): How to determine sample size for biological parameters



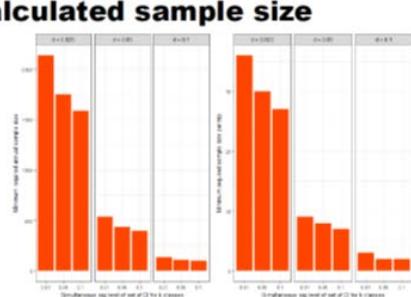
Sample size formula

- $n = \frac{1}{d^2/z_{\alpha/2}^2 s^2 + 1/N}$ (Thompson 2012, Cochran 1977, Lohr 2010, Chaudhuri 2014)
- d : is the margin of error
- α : probability that there will be an error that will exceed d . This is used in calculating confidence interval but presented as $100(1 - \alpha)\%$
- $z_{\alpha/2}^2$: values of the normal distribution
- s^2 : estimated population variance
- N : the population size

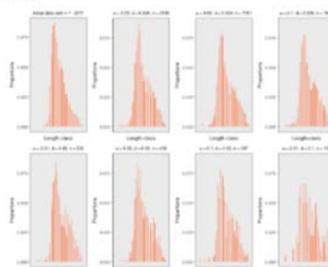
Parameters used

- Length classes: 80:320
- Total number of trips in a year: 60
- $\alpha \in \{0.01, 0.05, 0.10\}$
- $d \in \{0.025, 0.05, 0.10\}$

Calculated sample size



Results



References

- Chaudhuri Arjit (2014) Modern survey sampling. CRC Press
- Lohr Sharon (2010). Sampling: Design and analysis. Second edition. Brooks/Cole cengage learning
- Thompson Steven (2012). Sampling. Third edition. Wiley
- Cochran William (1977). Sampling techniques. Third edition

3. Maria Teresa Facchini (COISPA Italy): Sampling Design tool (SDtool)

WKBIOPIM 2 Workshop on Optimization of Biological Sampling ICES CIEM

29-31 May 2018
Nantes – France

Sampling Design tool (SDtool)

Maria Teresa FACCHINI
(COISPA, Italy)



OVERVIEW

SOME HINTS...

- Sampling Design tool (SDtool) is a methodological framework proposed in MARE/2014/19 Med&BS project for sampling optimization, accounting for cost (sampling effort) and precision (coefficient of variation) trade-offs.
- The tool has been applied to several case studies in Mediterranean Sea during operational sessions of RCM Med&BS-LP in 2016 and 2017.
- March 2017, Cyprus: Training Workshop on Sampling Intensity organized following the Recommendation of 2016 RCM Med&BS-LP for training scientists of different countries involved in the DCF.
- June 2017, Lisbon: results from the training workshop were presented at ICES Workshop on Optimization of Biological Sampling at Catch-Sample Level (WKBIOPIM)

OVERVIEW

METHODOLOGICAL FRAMEWORK FOR SAMPLING OPTIMIZATION, ACCOUNTING FOR COST AND PRECISION TRADE-OFFS

Sampling Design Tool - SDtool
built on COST tools (COST Project, 2009)

OBJECTIVE: Define the amount of fish to be sampled to reach a certain level of precision reflecting an acceptable quality for the sampling, based on CV.

PROCEDURE: (i) Analyze the evolution of the sampling variability with the number of samples, in terms of CV, (ii) assess whether the current plans is over or under-sampling the fisheries and (iii) propose an "optimal" sampling plan (number of trips to sample and number of individuals to measure) disaggregated over strata (e.g. quarters and metiers).

DATA REQUIREMENT: ICES Standard Data Exchange Format (SDEF)

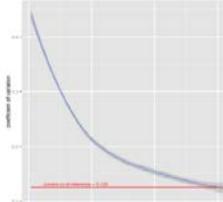
BASIC OPERATIONAL REQUIREMENTS: The procedure relies upon the availability of substantially large datasets. The procedure would not perform well using datasets with a reduced amount of fishing trips and in such cases could give misleading results.

OVERVIEW

METHODOLOGICAL FRAMEWORK FOR SAMPLING OPTIMIZATION, ACCOUNTING FOR COST AND PRECISION TRADE-OFFS

APPROACH

- The CVs decrease as the number of samples increases: as this trend is not linear, the curve tends to flatten from a certain point.
- This flattening underlines a trade-off between the precision and the required number of samples (thus the economic cost).

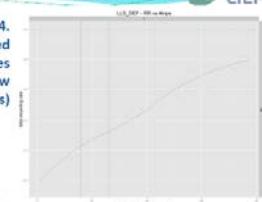


4. The recycling rate of the data is estimated as a function of the number of samples (e.g. it could be used to determine how large are datasets and reliable the results)



Var1	Var2	Var3	Optimized Nbrp (min)	Optimized Nbrp (max)
18	1	G13_DEF	28	35
18	1	GTR_DEF	28	35
18	1	GTR_M00	33	52

5. From the CV curves of each stratum, "optimal" sampling size range is proposed, identifying the part of the curve where its tangent changed and began to flatten. This part is considered as a suitable trade-off between the precision and the cost (in terms of sampling effort).



RESULTS

- Using the "optimal" sampling size range for each stratum, bootstrap procedure is run (e.g. 100 times) in order to reconstruct pools of trips having sizes within the optimal ranges. Then the numbers of individuals related to each pool is derived.
- For each stratum the mean number of measured individuals is proposed as the "optimal" number of individuals (with confidence intervals)

Deliverable 2.5
Project MARE/2014/19-SI2.705484

OVERVIEW

METHODOLOGICAL FRAMEWORK FOR SAMPLING OPTIMIZATION, ACCOUNTING FOR COST AND PRECISION TRADE-OFFS

RESULTS

Var1	Var2	Var3	Optimized Nbrp Ind. (mean)	Optimized Nbrp Ind. (st.dev.)
18	1	G13_DEF	385	385
18	1	GTR_DEF	12672	127
18	1	GTR_M00	11884	2315

L13_DEF - Optimized vs Past Measured ind.



ONGOING DEVELOPMENTS **WKBIOPTIM 2** **ICES CIEM**

NEW ONGOING DEVELOPMENTS FOLLOWING THE OBJECTIVES OF MARE/2016/22, ANNEX I "BIOLOGICAL DATA COLLECTION IN EU WATERS" STREAM PROJECT

OBJECTIVES (WP3)

- Generalizing the existing SDtool in order to include options allowing a flexible definition of the sampling scheme. Update the SDtool to carry out simulations on
 - different technical stratifications introducing options to define the technical strata on the basis of gear (level 4) and/or metier, so grouping strata with similar characteristics;
 - different temporal aggregation in order to make flexible the stratification by quarter and/or semester, depending on fisheries and target species specifications;
 - data of stocks considered shared among MS in order to get results on the whole area of the stock (not only by GSA)

e.g.

Length structure raised to the production

Coefficient of variation values varying with no. of trips and no. of measured individuals

1 - 10 - 1 - Optimized vs Past #trips

ONGOING DEVELOPMENTS **WKBIOPTIM 2** **ICES CIEM**

NEW ONGOING DEVELOPMENTS FOLLOWING THE OBJECTIVES OF MARE/2016/22, ANNEX I "BIOLOGICAL DATA COLLECTION IN EU WATERS" STREAM PROJECT

OBJECTIVES (WP3)

- Introduce procedures for the evaluation of "scenarios" with different sampling stratifications and different trips (PSUs) allocation

e.g.

Scenario A - proportional to LFD variability (CV)

Scenario B - proportional to landings

ONGOING DEVELOPMENTS **WKBIOPTIM 2** **ICES CIEM**

NEW ONGOING DEVELOPMENTS FOLLOWING THE OBJECTIVES OF MARE/2016/22, ANNEX I "BIOLOGICAL DATA COLLECTION IN EU WATERS" STREAM PROJECT

In order to estimate the effective improvement in precision in different scenarios (searching the best option for the Regional Sampling Plan), the Design Effect* could indicate to which extent the actual sampling error differs from the one expected under "Simple Random Sample" (SRS). It is calculated as follows:

Deff = Var design / Var SRS

If the square root of the design effect, called Design Factor (deft)

Deft = 1 there is no effect of sample design on standard error

Deft > 1 the sample design inflates the standard error of the estimate

Deft < 1 the sample design increases efficiency (reduces standard errors) of the estimate.

e.g.

casestudy_folder	RMSE	mean_length	variance	stDev_mean_length	REF_variance	Deff	Deft
sc_A3_GSA_Y_NONE	0.43	24.97	179.36	13.4	99.0	1.81	1.35
sc_B3_GSA_Y_NONE	0.38	20.69	66.33	8.1	99.0	0.67	0.82

* Kish, L. (1995), "Methods for Design Effects," Journal of Official Statistics, Vol. 11, No. 1, pp. 55-65

DATASET PREPARATION **WKBIOPTIM 2** **ICES CIEM**

DATA SET PREPARATION

- Create a new folder in order to collect all the needed input files and output of the case study, e.g. C:\02_SD_TOOL\European hake GSA 18
- Prepare dataset of the selected species-years-area and save them in .csv files. The options available for the data input format are 2:
 - SDEF format** (ICES Standard Data Exchange Format) used in COST project (Jansen et al, 2009), i.e. TR, HH, SL, HL and CL tables (see COST - Standard Data Exchange Format .pdf file in /SamplingDesign tool/COST-man folder)
 - RCM Med&BS-LP Data Call format**: a simplified format identified as common standard format to foster data exchange and dissemination in MARE/2014/19 Med&BS project.

ENVIRONMENT SETTINGS **WKBIOPTIM 2** **ICES CIEM**

SETTING THE ENVIRONMENT

- Download the APP-SDTools-ver3_1904.zip file from /SamplingDesign tool folder on the WKBIOPTIM 2 sharepoint and unzip it in your PC
- Open the R console and install the packages from /SamplingDesign tool/libs folder from the menu **Packages** → **Install packages from local .zip files...**
 - COSTcore**
 - COSTdbe**
 - COSTeda**
 - SDtool**
- Then from the menu **Packages** → **Install packages...** choose a R-CRAN mirror and select one by one the following libraries (if not installed on your R):
 - plyr**
 - knitr**
 - ggplot2**
 - data.table**
 - lattice**
 - reshape**
 - reshape2**

SETTING/RUNNING A CASE STUDY **WKBIOPTIM 2** **ICES CIEM**

CASE STUDY DEFINITION AND SETTING OF THE SCRIPT FOR THE OPTIMIZATION

- Create a new folder in order to collect all the needed input files and output of the case study, e.g. C:\European hake GSA 18
- Use the **stratification_results_TEMPLATE.csv** file to define the level of aggregation of the results for space, time, technical or commercial category (e.g. aggregate results by GSA, Y (year), lev6, N).

NOTE: In the present version technical and commercial category are exclusive options.

space	time	technical	commercial_category
SIAS	Y	Yes	N
Country	Y	No	Y
	Z		
	P		

- Open one of the following scripts and change the settings for your case study:
 - toRUNoptimization_from_SDEF_format_1605.r** for data in **SDEF format**
 - toRUNoptimization_from_RCMDatacall_format_1605.r** for data in **RCM Med&BS-LP Data Call format**

SETTING/RUNNING A CASE STUDY **WKBIOPTIM 2** **ICES CIEM**

CASE STUDY SETTINGS (1)

SDEF Format

```
# Set the absolute path of the COST datasets (sampling tables: TR, HH, SI, HL and landings table: CL)
filepathTR <- "C:\\02_SD_TOOL\\European hake GSA 18\\SDEF CS-TR data.csv"
filepathHH <- "C:\\02_SD_TOOL\\European hake GSA 18\\SDEF CS-HH data.csv"
filepathSI <- "C:\\02_SD_TOOL\\European hake GSA 18\\SDEF CS-SI data.csv"
filepathHL <- "C:\\02_SD_TOOL\\European hake GSA 18\\SDEF CS-HL data.csv"
filepathCL <- "C:\\02_SD_TOOL\\European hake GSA 18\\SDEF CL data.csv"
```

Reading alternative simplified format....

RCM Med&BS Datacall Format

```
# Set the absolute path of the CS and CL datasets
filepathCS <- "C:\\02_SD_TOOL\\CASE_European_hake\\HKE_CS_2014-2016_OSA_10_RCM_Datacall.csv"
filepathCL <- "C:\\02_SD_TOOL\\CASE_European_hake\\HKE_CL_2014-2016_OSA_18_RCM_Datacall.csv"
```

Set the format used for CS data; if COMMERCIAL CATEGORY columns have not been filled set format "OLD"
FORMAT_OLD_NEW <- "NEW"

CASE STUDY SETTINGS (2)

```
# Set the date format in sampling data
DATE_FORMAT <- "dd/mm/aaaa"

# Set the absolute path of the directory where the tool will save results
resultsDirectory <- "C:\\\\02_SD_TOOL\\\\European_hake_GSA_18"

# Set the absolute path of the directory of the SDTool scripts
scriptsDirectory <- "C:\\\\02_SD_TOOL\\\\APP-SDTools-ver3_1904"

# Set the absolute path of the file where the level of aggregation of the results is set (i.e. stratification_results_TEMPLATE.csv)
table_of_stratification <- "C:\\\\02_SD_TOOL\\\\European_hake_GSA_18\\\\stratification_results_HXE_by_mettier.csv"

# Set a short name for the case study (as it will be the name of a folder, NOT USE special characters)
SHORT_CASE_STUDY_NAME <- "European_hake_GSA18"

# Set the codification of the species reported in the CS table (sometimes it could be different from that reported in the CL table)
sppName <- "Merluccius merluccius"
```

CASE STUDY SETTINGS (3)

```
# Set the country of the case study for the selection of data
country <- "ITA" # set the COUNTRY

# Set the years of the time series you want to include in the analyses
years <- c(2014:2016)

# Set the number of iterations for the bootstrap procedures (it is suggested to start with a low number to do the preliminary analysis)
niter <- 10

# Set TRUE if you have already run the results and then you can load the saved workspaces; otherwise FALSE (if you are running the case study for the first time)
READ_PREVIOUS_RESULT <- FALSE

# Set TRUE if you want to calculate calculate the Recycling Rate; otherwise FALSE
CALCULATE_RR <- TRUE

# Set TRUE if you want to save graphs; otherwise FALSE
SAVE_GRAPHS <- TRUE
```

RUN THE OPTIMIZATION SCRIPT

From the menu **File → Source R code...** choose between the files according to your datasets format:

1. [toRUNoptimization_from_SDEF_format_1605.r](#) for data in SDEF format
2. [toRUNoptimization_from_RCMDatcall_format_1605.r](#) for data in RCM Med&BS-LP Data Call format

CASE STUDY DEFINITION AND SETTING THE SCRIPT FOR THE SCENARIO SIMULATION

1. Use the [stratification_results_TEMPLATE.csv](#) file to define the level of aggregation of the results for space, time, technical or commercial category (e.g. aggregate results by GSA, Y (year), lev6, N).

space	time	technical	commercial category
GSA	Y	level	N
Country	O	level	V
S			

e.g. Space=GSA, time=Y, tech:level, comm. cat.=N

2. Use the [sampling_design_definition.csv](#) file to define the criteria for the sampling stratification.

space	time	technical	commercial category
GSA	Y	level	N
Country	O	level	V
S			

e.g. Space=GSA, time=Y, tech:level, comm. cat.=N

3. Use the [sampling_design_trips_sharing.csv](#) file to define the sharing of the trips among the defined strata.

e.g.	space	original_space	time	original_time	technical	original	technical	commercial	original	commercial	no. of trips
ALL	ALL	ALL	ALL	LLS_DEF	LLS_DEF	ALL	ALL	ALL	ALL	ALL	25
ALL	ALL	ALL	OTB_DEF	OTB_DEF	OTB_DEF	ALL	ALL	ALL	ALL	ALL	10
ALL	ALL	ALL	OTB_MDO	OTB_MDO	OTB_MDO	ALL	ALL	ALL	ALL	ALL	15

SCENARIO SETTINGS (1)

1. Open one of the following scripts and change the settings for your scenario:

1. [toRUNscenario_from_RCMDatcall_format_1605.r](#) for data in RCM Med&BS-LP Data Call format
2. [toRUNscenario_from_RCMDatcall_format_1605.r](#) for data in RCM Med&BS-LP Data Call format

```
DATE_FORMAT
resultsDirectory
scriptsDirectory
SHORT_CASE_STUDY_NAME
sppName
country
Years
niter
READ_PREVIOUS_RESULT
```

SCENARIO SETTINGS (2)

```
# Set the absolute path of the file where the level of aggregation of the results is set
table_of_stratification <- "C:\\\\02_SD_TOOL\\\\European_hake_GSA_18\\\\stratification_results_HXE.csv"

# Set the absolute path of the file where definition of the stratification of the sampling you want to reconstruct is set, e.g. random sampling in each stratum defined by GSA, year, level
table_of_sampling_definition <- "C:\\\\02_SD_TOOL\\\\European_hake_GSA_18\\\\sampling_design_definition.csv"

# Set the absolute path of the file where allocation of trips in the given sampling scheme is set
table_of_sharing_of_trips <- "C:\\\\02_SD_TOOL\\\\European_hake_GSA_18\\\\sampling_design_trips_sharing.csv"
```

RUN THE SCENARIO SCRIPT

From the menu **File → Source R code...** choose between the files according to your datasets format:

1. [toRUNscenario_from_SDEF_format_1605.r](#) for data in SDEF format
2. [toRUNscenario_from_RCMDatcall_format_1605.r](#) for data in RCM Med&BS-LP Data Call format

EVALUATION SETTINGS (1)

1. Open the scripts [evaluationSCENARIO_1605.r](#), change the settings as follows and run the script from the menu **File → Source R code...**

```
# Set the codification of the species
sppName <- "Merluccius merluccius"

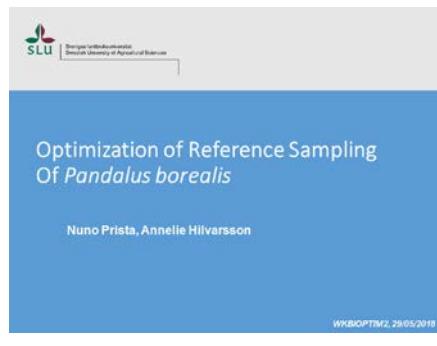
# Set a particular string that is contained in all the scenarios you want to compare
identifies_scenario_folder <- "so"

# Set TRUE if you want to consider a particular scenario as reference for the comparison (e.g. "Simple Random Sampling" SRS); FALSE otherwise
consider_BASELINE <- TRUE

# set the directory name of the scenario considered as reference (e.g. "Simple Random Sampling" SRS)
BASELINE_scenario <- "so_Z3_GSA_Y_NONE"

# Set the absolute path where all the scenarios results to be compared are saved
scenariosDirectory <- "C:\\\\02_SD_TOOL\\\\European_hake_SDtool\\\\results_scenarios_A_and_B"
```

4. Nuno Prista and Annelie Hilvarsson (SLU Aqua, Sweden): Optimization of Reference Sampling for *Pandalus borealis*.



- Main "quality" indicator: Mean Weighted Coefficient Variation (MWCV)

Precision estimates and suggested sample sizes for length-frequency data

Hans D. Grevarek contact author?
David McGrath?

Under the assumption of a multinomial distribution, the standard deviation (σ_x) of the number of fish in a sample that are length category i can be estimated by

$$\sigma_x = \sqrt{np_i(1-p_i)}, \quad (1)$$

where n = the total number of fish in the sample, and p_i = the proportion of the sample that is length i .

The coefficient of variation (CV) of the number of fish at length i is given by

$$CV_i = \frac{\sigma_x}{np_i}, \quad (2)$$

and the mean weighted coefficient of variation (MWCV) is given by

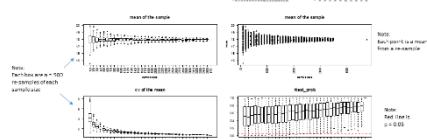
$$MWCV = \sum p_i CV_i = \frac{\sum \sigma_x}{n}. \quad (3)$$

PLA, 2013, 109–120–120–200.

- Some results:

- Sample 1401_1

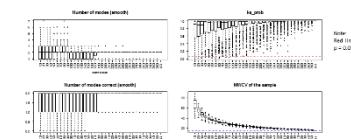
- January
- 477 individuals
- Median: 18 (9-28) mm



- Some results:

- Sample 1401_1

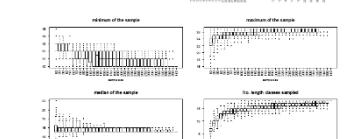
- January
- 477 individuals
- Median: 18 (9-28) mm



- Some results:

- Sample 1401_3

- May
- 448 individuals
- Median: 15 (10-27) mm



- What was done:

In the field (2016):

- All hauls sampled, 1 reference sample per haul; Total 64 samples

- Each sample: range 1.0-2.5 kg (mean: 1.9 kg) of shrimps taken from the unsorted catch

In the lab (2016):

- All *Pandalus borealis* measured

- 225-375 individuals per sample [mean: 356 individuals]

In the computer (WKS00, WKBIOPM, 2016-17)

- Development of a script that simulates length distributions for different sample sizes

Species of *Pandalus*:

- Selection of samples > 350 individuals for analysis (49 samples; 7% of total)

- 500 simulations of each sample size from 10 to 350 in increments of 10 individuals

- Mean Weighted CV used to evaluate "quality" of LFs

- Determination of an optimum (lower) sample size

- Confirmation with an array of other "quality" indicators

- Visual inspection of the results

- Other "quality" indicators:

- Coefficient of Variation of the Mean

- Minimum, Maximum, Median

- No. of classes sampled

- No. of modes correct (smoothed)

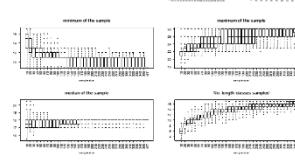
- T-test on means

- Kolmogorov-Smirnov distribution test

- Some results:

- Sample 1401_1

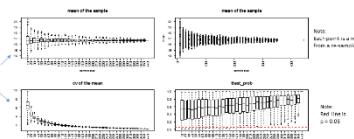
- January
- 477 individuals
- Median: 18 (9-28) mm



- Some results:

- Sample 1401_3

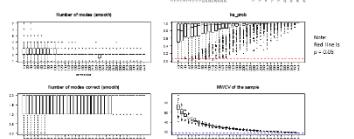
- May
- 448 individuals
- Median: 15 (10-27) mm

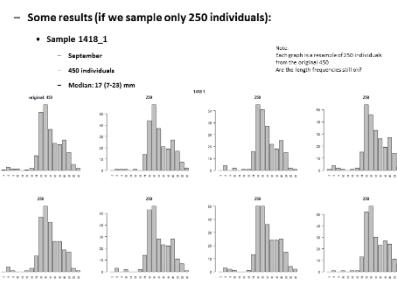
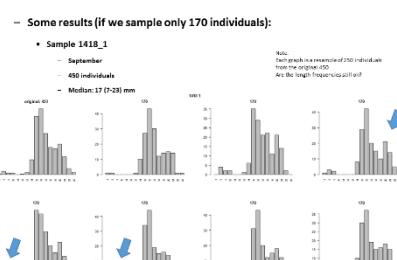
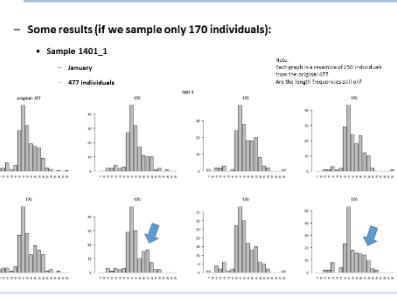
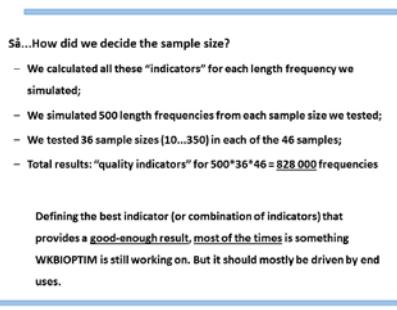
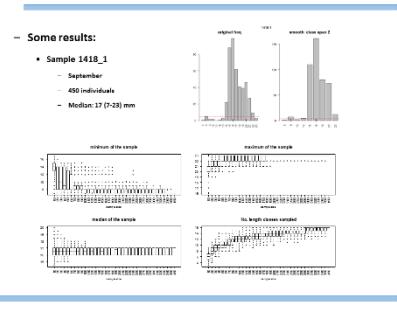
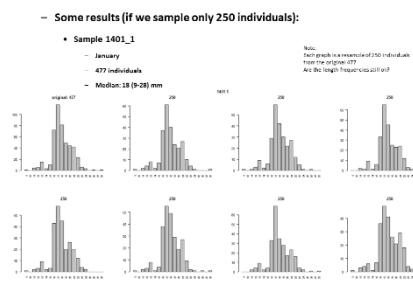
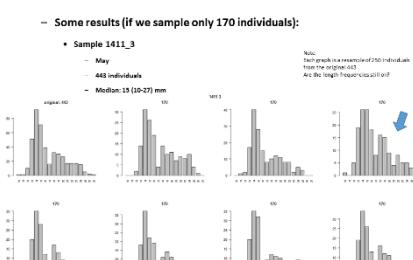
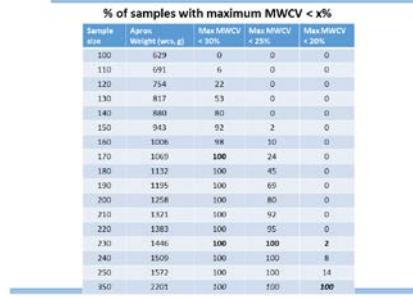
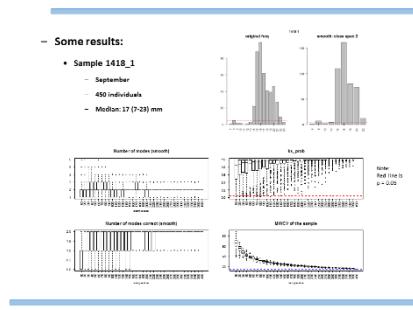
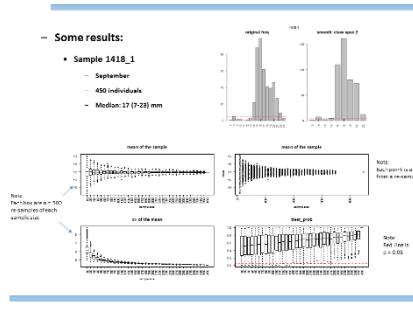


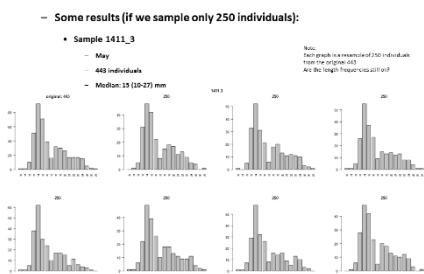
- Some results:

- Sample 1411_3

- May
- 448 individuals
- Median: 15 (10-27) mm





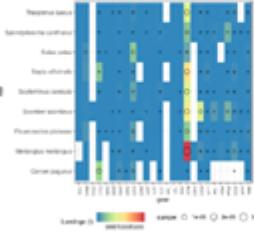
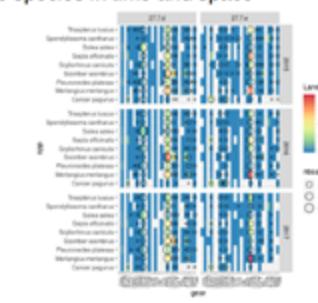
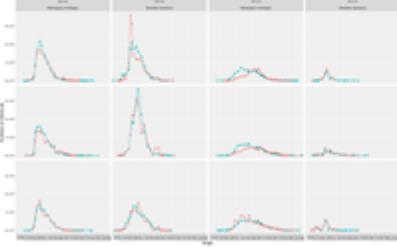
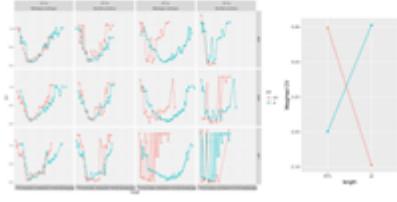
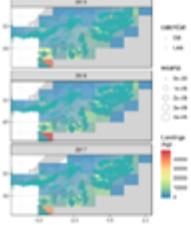


- Recommendations (for now, no particular end-uses considered)
 - Keep the usual sample from the catch (onboard all stay the same)
 - Subsample ca 1.5 kg in the lab (not the hand, use a scoop!)
 - That sample will yield between 200 and 300 shrimps, depending on their size;
 - The change represents a 25% reduction in number of measurements – it is very conservative and should produce good length frequencies in nearly all situations useful for most end-uses (present and future).

Thank you!

5. Laurent Dubroca (Ifremer France): WKBIOPTIM: Optimization procedure at the multi-level sampling.

<h2 style="text-align: center;">WKBIOPTIM2</h2> <p>Optimization procedure at the sampling level</p>	<h3 style="text-align: center;">Optimization procedure at the sampling level (WKBIOPTIM 1)</h3> <ul style="list-style-type: none"> From data to Intercatch estimates <ul style="list-style-type: none"> COST objects (Fishframe) and COST package Ratio estimators Intercatch estimates 												
<h3 style="text-align: center;">Optimization procedure at the sampling level (WKBIOPTIM 1)</h3> <ul style="list-style-type: none"> Scenario : what if the number of trips/hauls/boats... sampled change ? <p>Example of scenario :</p> <ul style="list-style-type: none"> - less trip - less haul - less haul by trip - removing sampling from one country 	<h3 style="text-align: center;">Optimization procedure at the sampling level (WKBIOPTIM 1)</h3> <ul style="list-style-type: none"> Distance between distribution : the earth mover's distance <p>Earth Mover's Distance (EMD) : a measure of the distance between two probability distributions over a region D (namesake Wasserstein metric in mathematics) (Rubner, Tomasi, and Guibas 1998).</p> <p>Informally, if the distributions are interpreted as two different ways of piling up a certain amount of dirt over the same area, the EMD is the minimum cost of turning one pile into the other, where the cost is assumed to be amount of the moved times the distance by which it is moved.</p>												
<h3 style="text-align: center;">WKBIOPTIM 2 : Optimization procedure at the sampling level</h3> <ul style="list-style-type: none"> A simple situation : <p>French landings and samplings 2015-2017 27.7.e & 27.7.d</p>	<h3 style="text-align: center;">WKBIOPTIM 2 : Optimization procedure at the sampling level</h3> <p>Simple ?</p> <table border="1"> <thead> <tr> <th>Category</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>French landings and samplings 2015-2017</td> <td>27.7.e & 27.7.d</td> </tr> <tr> <td>263 taxon</td> <td>263</td> </tr> <tr> <td>205 metier level 6 (26 gears)</td> <td>205</td> </tr> <tr> <td>431597 trips</td> <td>431597</td> </tr> <tr> <td>1543 trips sampled (221 species)</td> <td>1543</td> </tr> </tbody> </table>	Category	Value	French landings and samplings 2015-2017	27.7.e & 27.7.d	263 taxon	263	205 metier level 6 (26 gears)	205	431597 trips	431597	1543 trips sampled (221 species)	1543
Category	Value												
French landings and samplings 2015-2017	27.7.e & 27.7.d												
263 taxon	263												
205 metier level 6 (26 gears)	205												
431597 trips	431597												
1543 trips sampled (221 species)	1543												

<p>WKBIOPTIM 2 : Optimization procedure at the sampling level</p> <ul style="list-style-type: none"> • 10 most landed species with > 2000 samples  <p>French landings and samplings 2015-2017 27.7.e & 27.7.d</p>	<p>WKBIOPTIM 2 : Optimization procedure at the sampling level</p> <ul style="list-style-type: none"> • 10 species in time and space  <p>French landings and samplings 2015-2017 27.7.e & 27.7.d</p>
<p>WKBIOPTIM 2 : Optimization procedure at the sampling level</p> <ul style="list-style-type: none"> • Scenario : assessing sampling optimization effect for all the species  <p>French landings and samplings 2015-2017 27.7.e & 27.7.d</p> <p>Scenario : sample only 25 % of the trips with whiting (length distribution for OTB only here)</p>	<p>WKBIOPTIM 2 : Optimization procedure at the sampling level</p> <ul style="list-style-type: none"> • Scenario : assessing sampling optimization effect for all the species  <p>French landings and samplings 2015-2017 27.7.e & 27.7.d</p> <p>Scenario : sample only 25 % of the trips with whiting (length distribution for OTB only here)</p>
<p>WKBIOPTIM 2 : Optimization procedure at the sampling level</p> <ul style="list-style-type: none"> • Concurrent sampling !  <p>French landings and samplings 2015-2017</p>	<p>WKBIOPTIM 2 : Optimization procedure at the sampling level</p> <ul style="list-style-type: none"> • Extending the work of WKBIOPTIM 1 <ul style="list-style-type: none"> - Assessing objectively the impact of the scenario : Earth mover's distance + CV + weighted CV - How to do that integrating space, time, gear and species in the scenario and the analysis ? <ul style="list-style-type: none"> • Basic scenario ? • Simplify things ? <ul style="list-style-type: none"> - Main species or meter selection ? - Following the data call needs ? • How to deal with multivariate distribution and to assess the overall changes in them ? <ul style="list-style-type: none"> - Hierarchical classification in the CV space ? - How to deal with replicates in this case ? - Computing time ? • Arg...
<p>Tools</p> <ul style="list-style-type: none"> • R + markdown : file .Rmd analyze and document (reproducible research, and easy reporting) • RStudio • Directory position : work with relative path (working with absolute path is dangerous) 	

6. Sophie Vandemaele and Loes Vandecasteele (ILVO, Belgium): Belgian Beam Trawl Survey - optimizing our BTS survey with respect to sample sizes for ages.

Belgian Beam Trawl Survey

optimizing our BTS survey with respect to sample sizes for ages

Sofie Vandemaele (ILVO Belgium)
Loes Vandecasteele (ILVO Belgium)

- Objectives: indices for stock assessment WGs etc.
- Started in 1992
- 4 m-beam trawl as standard gear
- North Sea: area IVbc, several rectangles
- 62 fixed stations (fished for 30 min at 4 knots)
- Target species: plaice and sole
- Data in HH, HL (weight & LFD), CA (biopar) format
- Currently, ages are being determined for 5 species
- Sole & Plaice:
 - SS before 2017: 5 otoliths per cm class per rectangle
 - SS since 2017: 3 otoliths per cm class per rectangle, also age for individuals < 15 cm

Sample sizes for otoliths:

	2010	2011	2012	2013	2014	2015	2016	2017	
BLL	18	12	11	0	18	17	13	13	
COD	9	6	11	8	41	7	2	11	
	PLE	1109	1410	784	624	986	969	861	658
SOL	1197	743	539	442	681	689	547	561	
TUR	6	7	3	0	9	18	20	15	

QUESTIONS TO BE ANSWERED:

- Is a reduction (<3) of the age readings of sole and plaice possible? → budget re-allocation. What is the minimum number?
- Is age sampling for individuals smaller than e.g. 8 cm useful? (probably all of them have age 0)
- Is stratification by length class useful? → more variation in age for larger individuals (compared to small fish) → more sampling needed?

7. Julia Wischnewski and Matthias Bernreuther (Thünen Institute of Sea Fisheries, Germany):
Which sample size is adequate to estimate a length structure?

Which sample size is adequate to estimate a length structure?

Julia Wischnewski, Matthias Bernreuther

Problem statement (1)

- The shape of length frequency distribution is the most important statistical attribute for understanding the basic patterns displayed by length data (rather than summary statistics like mean, variance etc.)
- The ability to describe the distribution shape of a fish population requires sampling an adequate number of individuals, but collecting more fish than needed is inefficient
- How can one determine, whether enough samples have been collected to describe a shape precisely?
- In other words, can one adequately reduce a sample size and still obtain a reasonable size structure?

Problem statement (2)

- The most popular way for comparison of entire and reduced samples (i.e. sample and subsample) are statistical tests (Kolmogorov-Smirnov, Mann-Whitney, χ^2 -test etc.)
- This offers a pure statistical approach to be a "major player", but "adequateness" of subsample has to be considered in both statistical and biological senses, e.g.
 - some moderate difference in central tendency or spread of subsample will be revealed by statistical test, but biologically it still reflects the similar patterns;
 - only large lengths may be affected due to changes in a sample size – so, biologically significant difference between sample and subsample can be present, but many statistical tests will not distinguish between them

Problem statement (3)

- Statistical tests results can be misleading, e.g.:
 - Kolmogorov-Smirnov test is not very much reliable on binned data, if the bins are not small enough. But according to specifics of the Regional Data Base (RDB) Exchange Format, the intervals of 1 cm are recommended for fish that is measured in length, so the original unbinned (continuous) data is not available
 - the "rule of thumb" for the χ^2 -test is that the number of measurements for each length class in original sample should be greater than 5, and this is not always the case with length frequency data

Framework of the proposed approach (1)

- The length frequency data always display a range of modes, bumps and dips. Typically, there is a one major mode and few minor or secondary modes (rather bumps), although some of them can be just artefacts of insufficient samples for certain length classes
- The most frequent length classes might be relatively oversampled
- A well-chosen subsample should preserve a sample data structure, that is, reveal the same modes and capture the information about rare length classes
- At the same time, it has to accord with official sampling plan
- The obtained theoretical reference value of distinction between sample and subsample we will call an acceptable dissimilarity value (ADV)

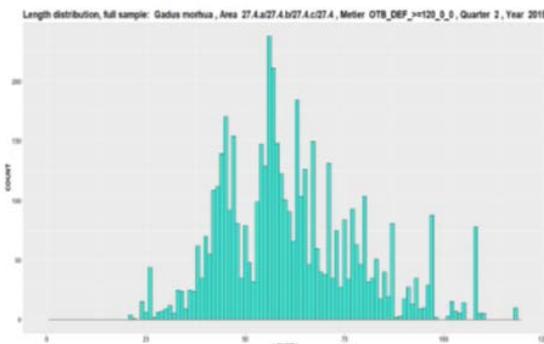
Framework of the proposed approach (2)

- ADV defined as a distance between cumulative distribution functions (CDFs) of sample (F) and subsample (G):

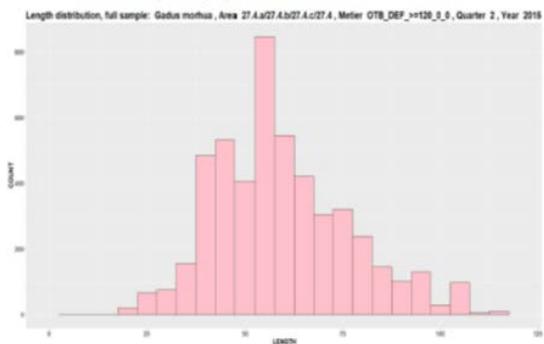
$$L_1(F, G) = \sum_{k=1}^n |F_k - G_k|$$

$$L_2(F, G) = \sqrt{\sum_{k=1}^n (F_k - G_k)^2}$$

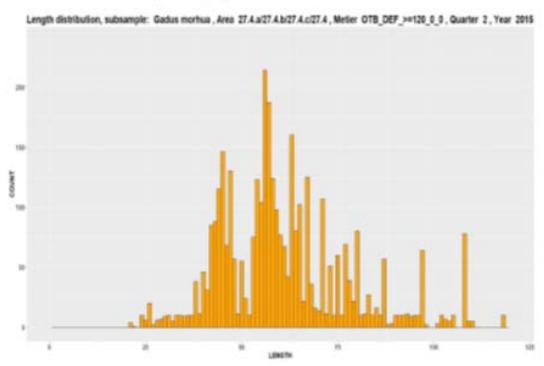
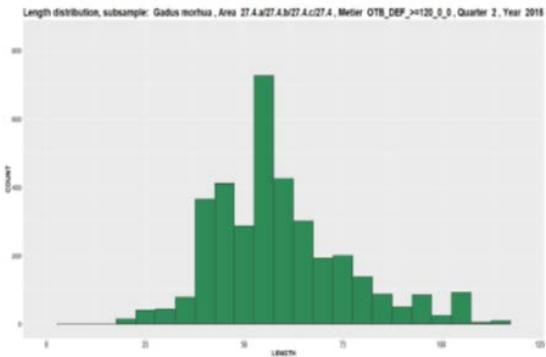
$$L_\infty(F, G) = \max_{k=1+n} |F_k - G_k|$$

Example (1): bin width = 1 cm**Example (2)**

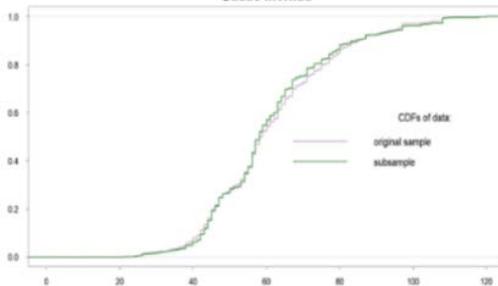
- We have to decrease a number of length measurements in such way that at a certain step the distinction between sample and subsample is considered to be "bordering" ("worst-case-acceptable")?
- However, that the shape of length frequency sample data is completely determined by bin width and bin origin
- A reasonable choice of bin width \Rightarrow process not too sensitive to changes in subsample shape corresponding to small peaks and other specific patterns in the sample, but the dominant modes are still there, not "smoothed away"
- Choice of bin width depends on species (1 cm interval was proposed for 30 cm species, 2 cm interval for 60-cm species and so on)

Publikum fragen!**Example (3): bin width = 5 cm****Procedure description**

1. Use as basic input data the standard RDB data with length rounded to 1 cm (10 mm);
2. Identify all modes in original full sample. A length class is determined as a mode if:
 - a. neighbouring length classes exhibit lower frequency values;
 - b. it contains at least 10 length measurements (sampling plan for Germany?)
3. Stability of revealed modes is ascertained
4. Set so-called reducing permitted length classes (the rare length classes are excluded from the set of reducing permitted length classes, to retain information covered by them: we will preserve all fishes > 105 cm);
5. Remove one length measurement from each reducing permitted length class and see whether all originally existing modes appear in resulted subsample. If yes, repeat the step. If no, go back to the previous subsample and stop

Publikum fragen!**Example (4): bin width = 1 cm****Example (5): bin width = 5 cm****Example (6)**

- Full sample size: **4938**
- Subsample size: **3589**
- $L_1 = 0.967$
- $L_2 = 0.145$
- $L_\infty = 0.038$

Gadus morhua**Example (7)**

- KS-test (5%-level) rejects the null hypothesis that the subsample distribution is identical to the original sample distribution:
One-sample Kolmogorov-Smirnov test
data: species.s\$LengthClass_cm
D = 0.050846, p-value = 1.744e-08
alternative hypothesis: two-sided

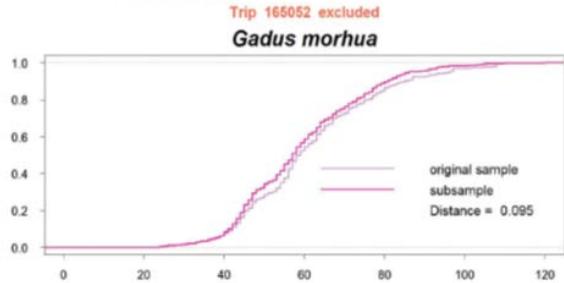
- $ADV = (L_2)^2$ (or L_1 , or L_∞), the theoretical reference value defining the largest acceptable decrease in sampling effort

Example (8)

- How can we reduce sampling effort practically? For instance, trips exclusion
- Trips: 3 trips in Q.2/2015, with cod length sampling:
 - ❑ Trip 165052
 - ❑ Trip 165349
 - ❑ Trip 174955

Example (9)

- 1724 cod length measurements in the trip 165052, distance = $0.095 > \text{ADV} = 0.021$ – trip can not be excluded!

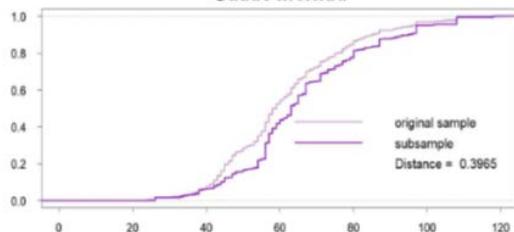


Example (10)

- 2868 cod length measurements in the trip 165349, distance = $0.396 > \text{ADV} = 0.021$ – trip can not be excluded!

Trip 165349 excluded

Gadus morhua

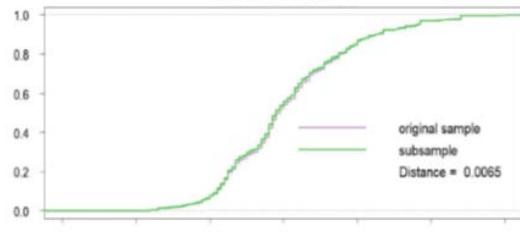


Example (11)

- 346 cod length measurements in the trip 174955, distance = $0.0065 < \text{ADV} = 0.021$ – trip can be excluded!

Trip 174955 excluded

Gadus morhua



Practical application

How can the proposed approach allow some planning in sampling design before the sampling is undertaken:

1. Compute ADV for each stratum (year/quarter/area/metier). In our case study, ADV for Q.2, from 2009 till 2016, in area 27.4, metier OTB_DEF_>=120_0_0
2. Based on annual ADVs, develop intervals for the future values

Discussion

- As example demonstrates, the algorithm has to be improved, e.g.:
 - minimum 10 length measurements per length class not enough;
 - additional condition is required, like: the arithmetic mean (as a measure of central tendency identifying the balance-point in a set of length values) has to stay in CI, or something like this...
-?
- Procedure can be presented in the R-functions terms (find_modes(), find_knots(), and so on)

Conclusion

- More precise procedure is required to determine what level of sampling is the “worst-case-acceptable”, to help to reduce time and cost spent collecting the length data
- The approach proposed in this study provides guidelines that give a rough idea of what to expect when the sample size is changed

8. Natoya Jourdain (IMR, Norway): Uncertainty Estimation of the North Sea IBTS Abundance Indices



Uncertainty Estimation of the North Sea IBTS Abundance Indices

Natoya Jourdain

WGNSSK April 27, 2018

IMR, Norway and Norwegian Computing Centre (NR) Collaboration

Olav Breivik, Edvin Fuglebakke, Jon Helge Vølstad, Sondre Aanes

Overview

• North sea IBTS:

- Important data source for many stock assessments in the North sea
- Sampling intensity of otoliths is quite high and the sampling procedure is *ad hoc* in some round fish areas

• Our Aim:

- Provide uncertainty estimates of abundance indices (input data level), not provided on DATRAS website
- Develop ALK estimators that account for variation in age-length compositions
- Optimizing survey designs with respect to stratification and subsampling for length and age

Uncertainty estimation forms the basis of evaluation of otolith sampling, i.e.,

- Optimize sampling effort – otoliths and number of stations

Overview of Otolith sampling by species

Period	Species	Minimum sampling levels of otoliths per length class
1991–2017		
	herring	Number of otoliths per length class in a RFA
	sprat	8 otoliths per $\frac{1}{2}$ cm group
		16 otoliths per 1 cm length class 0.0 – 11.0 cm
	cod	8 otoliths per 1 cm length class
	red	8 otoliths per 1 cm length class
	haddock	8 otoliths per 1 cm length class
	whiting	8 otoliths per 1 cm length class
	Norway pout	8 otoliths per 1 cm length class
	oarfish	8 otoliths per 1 cm length class
All target species		
	From 2012 Norway and Scotland, and Netherlands from 2016 have been sampling 1 otolith per length class from each trawl haul (0.5 km below for haddock, to 0.5 km below for herring and sprat, and to 1 km below for all other species).	
2018		
	herring	Number of otoliths per length class per trawl haul
	sprat	1 otolith per $\frac{1}{2}$ cm group
		1 otolith per $\frac{1}{2}$ cm length class 0.0 – 11.0 cm
	cod	1 otolith per 1 cm length class
	red	1 otolith per 1 cm length class
	haddock	1 otolith per 1 cm length class
	whiting	1 otolith per 1 cm length class
	Norway pout	1 otolith per 1 cm length class
	oarfish	1 otolith per 1 cm length class
	plaice	1 otolith per 1 cm length class

Actual survey design

Period	0°	0°
1991	0.5	0.5
2000	0.5	0.5
2005	0.5	0.5
2010	0.5	0.5
2015	0.5	0.5
2018	0.5	0.5

1 Depth stratified, semi-randomly
2 Proportion semi-randomized (from database of national safe tows or DATRAS or commercial fishing data)
3 Proportion opportunistic station selection (from database of national safe tows or DATRAS and from commercial data)
4 Fixed Stations
5 Proportion semi-randomized (new positions)

Expected survey design

- Random sampling of trawl locations in statistical rectangle
- Tows must be at least 10 nautical miles apart

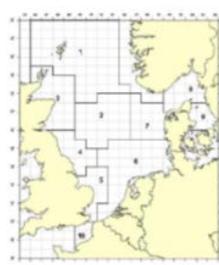
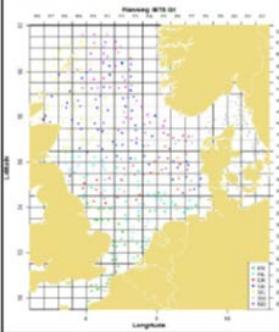


Figure 2: Standard otolith sample need for trawls since 1990. For all standard species after 2010. Additional 10% in order to 2000. For example, the entire 1 inclusion VTF Index Area 1, and an VTF Index coverage

Proposed Trawl locations Q1 2018

- Resampling of valid tows executed between 2000–2017 to increase "randomization"



Issues with non random design

- It is impossible to know how well you are representing the population
- Cannot calculate the probability of any member being selected for a sample
- Clustered stations at edge of rectangles (catches are most likely homogeneous)
- Variation in data within rectangle would not be captured
- Difficult to calculate confidence intervals and identify bias but can be done based on assumptions
 - Design-based: randomization is assumed
 - Model-based: assumption of independent and identically distributed random variables is made
- Parameter estimates must always be accompanied by their uncertainty
- Stock assessment models cannot fix uncertainty issues in assessment - issues must be addressed at the design level (IBTS)
- Uncertainty comes from the design

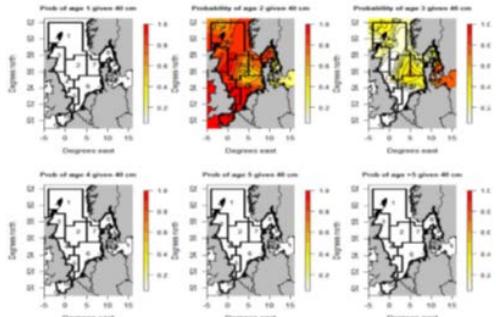
ALK Estimators

• Current ALK Estimators:

1. DATRAS ALK

- An aggregation of individual samples from a haul over a RFA
- Assumes age-length (ALK) compositions are the same over relatively large areas. If violated we would have
 - Biased estimates (Kimura 1977)
 - Underestimation of uncertainty
- A single ALK is produced for a relatively large area (round fish area)
- Imputation of missing samples is done by borrowing and interpolation/extrapolation
- The assumption regarding age-length structures is not valid

- Distribution of cod of length 40cm clearly varies within Central North Sea and Northern North Sea (see 2 year of length 40 cm)



2. Berg et al., (2012) – spatial age-length key modelling (using GAMs)

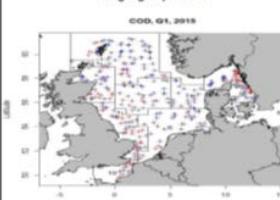
- Accounts for spatial variation in age-length structures between regions

• Our ALK Estimators:

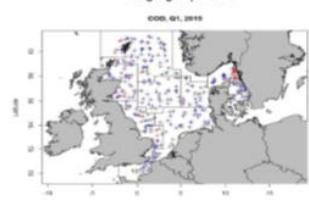
1. Haul-based ALK

- Each trawl haul has an ALK
- Pooled length groups of 5 cm to reduce missing data. Red: hauls with missing age; Blue: hauls with both age and length data

length group = 1 cm



length group = 5 cm



Our ALK Estimators:

- Haul-based ALK (COD)
 - Each trawl haul has an ALK
 - We pooled length groups of 5 cm (first 5 length groups and so on) to reduce missing data
 - Red: hauls with missing age; Blue: hauls with both age and length data

length group = 1 cm

length group = 5 cm

Red: hauls with missing age; Blue: hauls with both age and length data

length group = 1 cm

length group = 5 cm

length group = 1 cm

length group = 5 cm

2) Our Model-based ALK

- Uses logits for the probability of age
- Uses Random Field Theory (RFT)
- Values in a random field are usually spatially correlated in one way or another, i.e., adjacent values do not differ as much as values that are further apart
- Model estimates amount of smoothing
- Linear predictor assumes a zero mean Gaussian spatial random field (models spatial effect)
- Extended to include covariates such as haul (station) effect
- capture haul variation – a haul may hit a school of fish of a certain age
- Issues/ Solutions
 - Small number of observations implies error fields will not be smooth, hence violation of model assumption
 - One solution to this is sub-sampling

ALK Estimators

Berg et al. [2012] Model-based ALK

- Used continuation ratio logits for the probability of age
- Used GAM to model spatial effect
- GAM models spatial effects as a smooth surface, predicting numbers-at-age at the haul level
- Packages in *gamm* allow automatic selection of the amount of smoothing
- Can be extended to include covariates
- Issues/solutions:
 - Overfitting - but can be reduced using cross-validation
 - Statistically challenging to estimate large numbers of smoothing parameters
 - Prediction error criteria (GCV, AIC) can undersmooth substantially – *REML*, which is less problematic, can be used

Our Model-based ALK

Berg et al. [2012] Model-based ALK

- Figure shows Spatial pattern in the probability of being 2 years given 20 cm
- Similar patterns at Berg et al. (2012)
- A 20 cm haddock caught in this area is more likely to be 2 years
- Spatial pattern in the probability of being older than one year given 20 cm
- Spatial contrast in the data with a peak east of the Scottish coast
- A 20 cm haddock caught in this area is more likely to be 2 years or older than 1 year old
- Smoothen probably because of penalized likelihood

Probability of being 2 years given 20 cm

Probability of being older than one year given 20 cm

Fig. 8. Contour plots of the estimated probability (model ALK) of being older than one year given a length of 20 cm in year 2011 Q3.

RFT versus GAM

- Which is the best approach?
 - Further analysis and model comparison are required
 - (Better) Fit to age data?
 - Improved precision?
 - RFT is simply an alternative approach to GAM
 - Species have different characteristics and one model might be preferred over another

DATRAS – Stratified Mean

Berg et al. (2014) Delta-GAM

- Survey indices are computed by age and RFA (currently)
- Proposed a separate ratio estimator for area (north sea) – to be implemented
- Alternatively a combined ratio estimator can be used if deemed more appropriate
 - A comparison of COD estimates from Delta-GAM can then be made
- Survey indices are computed by age and area (North sea) using the Delta-GAM: a more refined alternative to DATRAS
- Models zeros and non-zeros separately
- Corrects for effects: spatial position, possibly depth and time of day
- Eliminates spatial stratification (through GAM)
- Applies bootstrapping to estimate the distribution of the indices

Uncertainty Estimation

- Berg et al. (2012, 2014)
 - Used bootstrapping to estimate uncertainty of abundance indices (not shown in paper)
- Recommended that uncertainties due to ALK estimation be included in stock assessments either by
 - Integrating the ALK estimation within the stock assessment model
 - Estimate uncertainty on the derived indices outside of the model - can be done using bootstrapping
- DATRAS bootstrap procedure (Naïve or simple nonparametric sampling)
 - Pool all hauls in a RFA, sample with replacement and put sampled hauls into relevant rectangles
 - Sample with replacement the pooled age with a given length in a RFA
- Issues:
 - Ignores fine scale stratification at the first stage, hence overestimation of the uncertainty
 - Ignores age-length data collected at the haul level, hence underestimates the uncertainty
 - Biases in both directions
- Stratified bootstrap procedure
 - Take a simple random sample with replacement from the original data in the given stratum (statistical rectangle); repeat independently across strata
 - An ALK is estimated for each haul

Some Results

Table 1: Indices Estimates and standard errors for Cod in Q3 of year 2015 in RFA 1, with approximate 95% confidence intervals (CI) of the different bootstrap methods

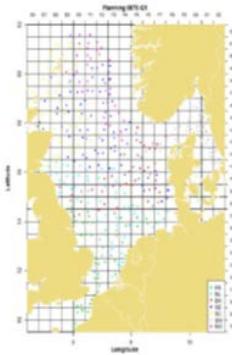
Age (yr)	DATRAS ALK DATRAS Bootstrap			Haul-based ALK Stratified Bootstrap			Model-based ALK Stratified Bootstrap		
	Abundance estimate	Standard error	Relative standard error	Abundance estimate	Standard error	Relative standard error	Abundance estimate	Standard error	Relative standard error
0	0	0	0	0	0	0	0	0	0
1	0.764	0.006	0.80%	0.760	0.006	0.79%	0.770	0.006	0.74%
2	21.889	0.776	3.6%	22.21	0.715	3.2%	22.11	0.706	3.2%
3	11.265	0.719	6.5%	10.58	1.204	11.5%	10.59	1.277	10.5%
4	3.250	0.181	5.6%	2.25	0.174	12.0%	2.26	0.174	12.0%
5	1.147	0.114	10.0%	1.27	0.112	10.0%	1.29	0.112	10.0%
6+1	1.279	0.088	9.6%	1.40	0.079	9.6%	1.38	0.082	9.6%

Approximate 95% CI from bootstrap procedure

Age (yr)	0	1	2	3	4	5	6+1
0	0.764	0.760	21.889	11.265	3.250	1.147	1.279
1	0.764	0.760	21.889	11.265	3.250	1.147	1.279
2	0.764	0.760	21.889	11.265	3.250	1.147	1.279
3	0.764	0.760	21.889	11.265	3.250	1.147	1.279
4	0.764	0.760	21.889	11.265	3.250	1.147	1.279
5	0.764	0.760	21.889	11.265	3.250	1.147	1.279
6+1	0.764	0.760	21.889	11.265	3.250	1.147	1.279

Our Intentions

1. Optimize sampling effort
 - Reduce the number of otoliths sampled
 - Possibly reduce number of stations
2. Possibly look at alternative estimators for abundance indices as Berg et al. (2014)
 - Full model-based approach
 - Estimates abundance indices and uncertainty
3. Alternative survey design (To improve the quality of survey results)
 - Fully randomized survey design
 - Possibly conduct pilot study to determine random "safe tow" areas
 - Sample a proportion of known trawlable locations and a proportion of new tow locations
 - New trawl to replace standard GOV (Robert Kynoch- Scotland) - reduce net damage, suitable for rough bottom
 - Reduce number of strata (fine-grained) – minimize problems of too few or missing values
 - Current design does not address uncertainty issues in the data
 - Uncertainty issues must be addressed at input data level (IBTS)

**Questions and comments**

• Berg

- How does pooling affects younger age groups (0, 1) ? Probability?
- Look at covariance structures when estimating uncertainty (correlation) –model based. What are some issues?
 - France
- How does our method (optimizing) address issues with different species (reducing number of stations e.g., cod abundance in northern north sea while saithe is limited and the increasing or decreasing uncertainty? France has been sampling per trawl haul for cod but for some species by FRA but in a more representative way
 - Knut
- Optimize by value set –put a value to each catch – threatened species (not necessarily monetary)
 - Jose
- We could have a risk target – trade off between uncertainty and catch
- What's the probability of being below some threshold?

Questions and comments

• Jon Helge

- How many stations per RFA? List of possible stations
- How available stations are distributed
- Number of stations actually bootstrapped from possible stations, we assume infinite observations

Annex 4: Scripts and code

The scripts used during the workshop for the case studies will be available in the ICES Sharepoint - WKBIOPTIM 2/Software or can be requested to:

A - Sample level analysis (Section 2.1 and Section 4)

Sample level generic function: nuno.prista@slu.se

Function for sampling number of otoliths by length class for age length key and VBGM parameters estimation can also be requested to patricia@ipma.pt

Function for generating CA data: Petri.Sarvamaa@luke.fi

B - Multilevel analysis (Section 2.2 and Section 5)

Multilevel generic function: laurent.dubroca@ifremer.fr

SD Tool v.2: facchini@coispa.it

C - Quality Indicators (Section 3.1)

Presentation of scripts developed by Julia Wischnewski and Matthias Bernreuther, referred in Section 3.1.

C.1. "approach_2_1.R"

```

library (boot)
B = 10000
n = nrow(df1)

### bootstrap-function

our.bootstrap <- function(data, num) { boot.samples = matrix(sample(data$lenCls, size = B * num, replace
= TRUE), B, num); return(boot.samples) }

### Matrix with bootstrap results

boot.samples <- our.bootstrap(df1, n)

#### Chosen statistics

our.statistics <- apply(boot.samples, 1, mean);

cat("",fill=TRUE)

```

```

q1 <- 0.025
q2 <- 0.975

### Take a subsample of the size = s

for (s in c(2500, 2000, 1500, 1000, 500))
{
  boot.subsamples <- our.bootstrap(df1, s)
  our.statistics.subsample <- apply(boot.subsamples, 1, mean);
  power <- length(which(our.statistics.subsample >= c.int[1] & our.statistics.subsample <= c.int[2]))/B
  cat("",fill=TRUE)
  cat("Power for sample size = ", s, " under ", B, "replicates: ", power, fill=TRUE)
  cat("",fill=TRUE)
}

```

C.2. “approach_2_2.R”

```

library (boot)

B = 10000
n = nrow(df1)

### bootstrap-function

our.bootstrap <- function(data, num) { boot.samples = matrix(sample(data$lenCls, size = B * num, replace
= TRUE), B, num); return(boot.samples) }

### Matrix with bootstrap results

boot.samples <- our.bootstrap(df1, n)

#### CDF

PF.boot.samples <- apply(boot.samples, 1, ecdf);

range_len <- seq(0, max(df1$lenCls), by=5);

M.boot <- matrix(NA, nrow=B*length(range_len), ncol=3)

```

```
P.boot <- c();
s <- 1;

for (i in 1:B)
{
  PF.boot <- PF.boot.samples[[i]];
  T.boot <- data.frame(cbind(knots(PF.boot), PF.boot(knots(PF.boot))));
  names(T.boot) <- c("x","P")

  d <- setdiff(range_len,knots(PF.boot));
  if (length(d)>=1)
  {
    T.diff <- data.frame(cbind(d,rep(NA,length(d))));
    names(T.diff) <- c("x","P");
    T.boot <- data.frame(rbind(T.boot, T.diff));
    T.boot <- T.boot[order(T.boot$x),];
  }

  m <- which(is.na(T.boot$P))
  for (i in 1:length(m))
  {
    if (m[i] > 1)
    {
      T.boot[m[i],]$P <- T.boot[m[i]-1,]$P
    } else T.boot[m[i],]$P <- 0;
  }

  M.boot[((s-1)*length(range_len) + 1):(s*length(range_len)),1] <- rep(s,nrow(T.boot))
  M.boot[((s-1)*length(range_len) + 1):(s*length(range_len)),2] <- T.boot$x;
  M.boot[((s-1)*length(range_len) + 1):(s*length(range_len)),3] <- T.boot$P;

  P.boot <- cbind(P.boot, T.boot$P);

  s <- s+1;
}

names(M.boot) <- c("resample","x","P");
```

```

P.mean <- apply(P.boot,1,mean);

Qu.boot <- c();
for (i in 1:nrow(P.boot))
{
  Qu.boot <- data.frame(rbind(Qu.boot, quantile(P.boot[i],prob=c(q1,q2))))
}
names(Qu.boot) <- c("q1","q2");

df1.boot.mean <- data.frame(cbind(range_len,P.mean));

names(df1.boot.mean) <- c("x","P")

op <- par(mfrow= c(1,1),oma = c(2,4,2,0) + 0.1, las=1, mar=c(2, 2, 3, 2) + 0.1, mgp = c(3, 2, 0))

plot(df1.boot.mean$x, df1.boot.mean$P, xlim=c(0,max(range_len)), col="grey40", cex.axis=1.7, cex.main=2,
      font.main=4, lwd=3, type="l", lty=1)
par(new=TRUE);
plot(df1.boot.mean$x, Qu.boot$q1, , xlim=c(0,max(range_len)), col="red", cex.axis=1.7, main=NA,lwd=2,
      lty=1 ,type="l",yaxt = "n")
par(new=TRUE);
plot(df1.boot.mean$x, Qu.boot$q2, , xlim=c(0,max(range_len)), col="red", cex.axis=1.7, main=NA,lwd=2,
      lty=1, type="l",yaxt = "n")
legend(0, 0.8, c("original sample", "bootstrap bands"), lwd=c(3,2), col=c("grey40","red"), lty=1, cex=1.7,
      bty="n", x.intersp=0.4)

title(paste("CDF AND CONFIDENCE BANDS: ",unique(df1$spp), " , ", paste(unique(df1$vslFlgCtry), col-
      lapse="/"), " , Area ", paste(unique(df1$area), collapse="/"), " , Quarters ", paste(unique(df1$quarter),
      collapse="/"), " , Year ", paste(unique(df1$year),collapse="/")), outer=TRUE, cex.main=1.7)

par(op)

```

C.3. “modes_and_antimodes.R”

```

### min.sample.size.per_lencl : minimal number of individuals in each length class
### data : data in RDB format
### k : bin width, can be 1, 2 etc. (in cm)

```

```

library(FSA)

modes_and_antimodes <- function(data, k) {

  library(FSA)

  u <- lencat(data$lenCls, w=k, startcat=0);
  u <- data.frame(u);
  names(u) <- paste("lenCls_",k);
  z <- cbind(data,u);
  x <- data.frame(table(factor(z[,names(u)], levels=seq(0, max(z[,names(u)]), by=k)))); 
  names(x)<- c("lenCls", "freq");

  modes <- NULL
  antimodes <- NULL

## modes
  for ( i in 2:(length(x$freq)-1) )
  {
    if ( (x$freq[i] > x$freq[i-1]) & x$freq[i] > x$freq[i+1] & (x$freq[i]/sum(x$freq)>0.01) ) {
      modes <- c(modes,i)}
    }
## antimodes
  for ( i in 2:(length(x$freq)-1) )
  {
    if (x$freq[i]==0 & x$freq[i+1] > 0 ) {
      antimodes <- c(antimodes ,i)
    }
    else
      if ( x$freq[i]>0 & x$freq[i] < x$freq[i-1] & (x$freq[i] < x$freq[i+1]) & (x$freq[i]/sum(x$freq)>0.01) ) {
        antimodes <- c(antimodes,i)}
    }
    if ( length(modes) == 0 ) {
      modes = 'Sorry, you have no modes'
    }
    if ( length(antimodes) == 0 ) {
      modes = 'Sorry, you have no antimodes'
    }
  }
}

```

```

l <- x[c(modes),]$lenCls;
d <- x[c(antimodes),]$lenCls;
return(list(as.numeric(as.character(l)),as.numeric(as.character(d)))); }

```

C.4. "approach_3.R"

```

delta1 <- 0.9;
delta2 <- 1.1;

modes <- modes_and_antimodes(df1,5)[[1]]
antimodes <- modes_and_antimodes(df1,5)[[2]]

### modes and antimodes together
all.extreme.points <- sort(c(modes, antimodes));

### only those modes and antimodes where length classes seem to be oversampled
selected.modes <- modes[2:3];
selected.antimodes <- antimodes[3];
selected.extreme.points <- sort(c(selected.modes, selected.antimodes))

M <- max(as.numeric(df1$lenCls));
U <- data.frame(table(factor(df1$lenCls, levels=seq(0, M, by=5))));
names(U) <- c("lenCls", "freq");

diff <- abs(diff(subset(U, lenCls %in% selected.extreme.points)$freq));
subsample <- df1;
diff.s <- diff;
modes.s <- modes;
antimodes.s <- antimodes;
U.s <- U;

t <- 1;
repeat
{
  if(identical(intersect(modes.s, modes), modes) & identical(intersect(antimodes.s, antimodes), antimodes)
    & all(y.diff.s/y.diff >= delta1 & y.diff.s/y.diff <= delta2))

```

```
{  
subsample.previous <- subsample;  
U.s.previous <- U.s;  
n <- which(U.s$freq > 1);  
  
s <- U.s[n,]$lenCls;  
  
for (j in 1:length(s))  
{  
z <- sample(which(subsample$lenCls == s[j]), 1, rep=FALSE);  
subsample <- subsample[-z,];  
}  
  
J <- data.frame(table(factor(subsample$lenCls, levels=seq(0, M, by=5))));  
names(J) <- c("lenCls", "freq")  
  
U.s$freq <- J$freq;  
  
modes.s <- modes_and_antimodes(subsample,5)[[1]];  
antimodes.s <- modes_and_antimodes(subsample,5)[[2]];  
  
all.extreme.points.s <- sort(c(modes.s, antimodes.s));  
  
selected.extreme.points.s <- all.extreme.points.s[4:6]  
  
diff.s <- abs(diff(subset(U.s, lenCls %in% selected.extreme.points.s)$freq));  
  
t <- t + 1;  
  
} else  
break;  
}  
  
subsample <- subsample.previous;  
  
dev.new();  
  
p1 <- ggplot(df1, aes(x=lenCls)) +
```

```

geom_histogram(data = df1, aes(x = lenCls), binwidth = 5, boundary=0, closed="left", colour="black",fill="grey60") +
  xlim(0,M) +
  ylim(0,r) +
  annotate("text", label = paste("size =", nrow(df1)), x = 105, y = r, size = 9, colour = "black", fontface = 4) +
  labs(title="ORIGINAL SAMPLE", x="LENGTH", y = "COUNT") +
  theme(title = element_text(face="bold",size=16), axis.title = element_text(face="bold",size = 15), axis.text =
    element_text(face="bold", size = 14), strip.text = element_text(size=18, face="bold.italic"))

#+ggtitle(paste("Length distribution: ", selected.species, " ", paste(unique(species$FlagCountry), collapse="/"), ", Area ", paste(selected.area, collapse="/"), ", Year ", paste(selected.year,collapse="/"), ", Quarter ",selected.quarter))

p2<- ggplot(subsample, aes(x=lenCls)) +
  geom_histogram(data = subsample, aes(x = lenCls), binwidth = 5, boundary=0, closed="left", colour="black",fill="grey80") +
  xlim(0,M) +
  ylim(0,r) +
  annotate("text", label = paste("size =", nrow(subsample)), x = 105, y = r, size = 9, colour = "black", fontface =
    4) +
  labs(title="SUBSAMPLE", x="LENGTH", y = "COUNT") +
  theme(title = element_text(face="bold",size=15), axis.title = element_text(face="bold",size = 15), axis.text =
    element_text(face="bold", size = 14), strip.text = element_text(size=16, face="bold.italic"))

#+ggtitle(paste("Length distribution: ", selected.species, " ", paste(unique(species$FlagCountry), collapse="/"), ", Area ", paste(selected.area, collapse="/"), ", Year ", paste(selected.year,collapse="/"), ", Quarter ",selected.quarter))

library(gridGraphics)
library(ggplot2)
library(lattice)
library(latticeExtra)
library(gridExtra)
library(grid)
library(plyr)
library(FSA)

grid.arrange(p1, p2, ncol=2, top=textGrob(paste("LENGTH FREQUENCY:", unique(df1$spp), ", ", paste(unique(df1$vslFlgCtry), collapse="/"), ", Area ", paste(unique(df1$area), collapse="/"), ", Quarters ", paste(unique(df1$quarter), collapse="/"), ", Year ", paste(unique(df1$year),collapse="/"))), gp=gpar(fontsize=19, font=2)))

```

Annex 5: Proposal: The third Workshop on Optimization of Biological Sampling (WKBIOPTIM 3)

The **Third Workshop on Optimization of Biological Sampling (WKBIOPTIM 3)** chaired by Ana Cláudia Fernandes (Portugal) and Eirini Mantzouni (Greece) will meet in Lysekil, Sweden, 27-31 May 2019 to:

ToR a) R-Toolbox: Finalization and integrating the different developed scripts, including documentation

ToR b) Quality Indicators: Discuss and conclude on a combination of indicators to evaluate the quality of data under different sample sizes, according to end users' needs

ToR c) Produce a guide for adequate use of sampling optimization procedures at national level, taking into account the results obtained in the analysis of the presented case studies (WKBIOPTIM 1 and WKBIOPTIM 2) and on the ongoing national experiences.

WKBIOPTIM 3 will report by September to the attention of the SSGIEOM Committee.

Supporting Information

Priority	This workshop is considered to have a high priority for already established and new commercial fishery and survey sampling programmes developed under the EU-MAUP. The expectation is that sampling resources (time and costs) will be saved by the development and implementation of the R-toolbox and it will be fundamental to increase data provision on data-limited stocks and environmental variables. The basic toolbox was developed by WKBIOPTIM 1 and further improved in WKBIOPTIM 2 by including different biological parameters and sampling procedures in scripts and testing them in a wide range of different scenarios. There is now the need to compile and document all the work developed to make it available in a more friendly format to the national institutes and end users. WKBIOPTIM 3 proposes to fulfil this goal.
----------	---

Scientific justification	<p>Statistical sound sampling is a requirement of the new EU-MAUP that now specifies that “where data are to be collected by sampling, Member States shall use statistically sound designs” (COM IMPL DEC 2016/1701). One important component of a “statistically sound design” is that sampling effort is optimized and fit for purpose, i.e. that time and costs spent in sampling can be effectively justified in terms of quality of the information finally provided to end-users. There is an increasing demand to determine MSY reference points for an increasing number of stocks, including many data-limited stocks, and, at the same time, to collect additional environmental and biological information. This makes optimisation of the number of length measurements, age and maturity estimation a priority since these tasks involve costs and time that could alternatively be spent in data collection of other stocks and/or variables. It is important that the national laboratories of MS have common tools to quantify the effects, advantages and disadvantages of different sampling intensities and sampling designs so they can optimise sampling in terms of time and costs savings. Several ICES EG’s, including e.g. WKPRECISE 2009, PGCCDBS 2012, PGDATA 2015 and WKCOSTBEN 2016 have pointed out that clustering effects in multistage catch sampling programmes may lead to effective sample sizes much lower than the number of units sampled, e.g. fish caught during one trip or haul often have more similar characteristics than the general population of fish they came from. This effect highlights the likely existence of oversampling in the lower stages of many national catch sampling programmes (e.g. trips, hauls within trips, samples within hauls), where an excessive number of individuals may be being sampled and not accruing significant additional information to estimates provided to end-users.</p> <p>The Workshop on Optimization of Biological Sampling (WKBIOPTIM 1 and 2) developed, improved and tested a set of R-scripts (based on the RBD exchange format) producing a range of statistical and graphical outputs to be used for discussion of appropriate levels of biological sampling of different stocks. Data quality indicators of the biological variables under the optimization procedures carried out at the workshops were discussed and a roadmap for future discussions with end-users outlined. Given the positive feedback both from national labs, RCM’s and other WGs it is recommended that a third workshop takes place to produce an R-Package including its documentation and a guide for adequate use of sampling optimization procedures. WKBIOPTIM is a joint workshop bringing together experts from WGCATCH and WGBIOP and the main results have been brought to further discussion by these two groups. WKBIOPTIM 3 pretends to: finalize and integrate the different sets of developed scripts, including documentation in an R-Toolbox (ToR a); Discuss and conclude on a combination of indicators to evaluate the quality of data under different sample sizes, according to end users’ needs (ToR b) and provide a guide for implementation of the optimization at national level, taking into account the results obtained in the analysis of the presented case studies (WKBIOPTIM and WKBIOPTIM 2) and on the ongoing national experiences (ToR c).</p>
Resource requirements	<p>The data collection programmes which provide the main input to this group are already underway, and resources are already committed. All EU countries already have the datasets required for analysis available in the RDB format and in the case of using survey data, it was developed a function to incorporate information on length stratification for CA table during WKBIOPTIM 2. It is envisioned the inclusion of the input from stock assessors concerning the data from surveys to adapt these methodologies to the surveys sampling design. Preparation work on the development and documentation of the R-package will be required prior to the meeting and it is expected that people involved can give the input from cases studies presented during previous workshops, or produced after, for the compilation of a guide with a set of rules for an adequate use of these optimization tools by national institutes. It is expected that work proposed will only be finalised after the workshop end and more time will be needed before reporting.</p>

Participants	The Workshop is expected to attract wide interest from those involved in WGCATCH and WGBIOP and should include a subset of participants familiar with R-code to the level of “loop coding” and “function building” and a subset of participants experienced in age and reproduction analysis. In view of its relevance to data collection within ICES, the EU-MAUP and regional sampling designs, it should include those involved in the annual planning of sampling and laboratory analysis, including e.g. number of trips to be sampled and fish to be measured and aged/sexed. Members of survey groups located under SSGIEOM should also be among the participants.
Secretariat facilities	Some secretarial support will be needed. The WK should take place in 2019. Therefore it will need to be approved by ACOM and SCICOM in early 2019.
Financial	Member States may fund this through their EMFF programme
Linkages to advisory committees	ACOM and SCICOM
Linkages to other committees or groups	WGCATCH, WGBIOP, PGDATA, SSGIEOM, Survey WGs (IBTS, IBAS, etc.)
Linkages to other organizations	RCGs