GREECE

Sampling scheme

and

Data Quality Assurance Framework

National Data Collection Programme

2019

Contents

1. Sampling Plan for Biological Data

Biological data on stocks exploited by the Hellenic commercial fishery are collected through the implementation of a temporally and spatially stratified sampling scheme, within the frame of the National Data Collection Programme. The targeted species, the type of collected data, the fraction of the commercial fleet monitored, as well as the temporal and spatial allocation of the sampling effort are defined by the relevant EU Commission Implementing Decisions, the GFCM-DCRF and the Regional Co-ordination Group for the Mediterranean and Black Sea. In order to be able to adjust to the interannual variability of fishing effort and stock biomass, various aspects of the implemented sampling scheme are properly modified based on reference points. These reference points are usually calculated based on the last three years' data.

1.1 Targeted Species-Type of collected data

The list of species for which biological data are collected are defined by the Tables 1A, 1C of the multi-annual Union Programme as well as by the GFCM-DCRF Annexes A.1, A.2, A.3. For all the species included in the aforementioned Tables and Annexes, the volume (Total weight and Total number per haul/trip) and the length frequency of all catch fractions (landings, discards, unwanted catches) are recorded.

For a subset of species, additional biological data, such as individual weight, age, sex-ratio and sexual maturity are collected. These species are selected based on specific thresholds (see the Fig. 1). More specifically, according to Chapter II of the Annex of Implementing Decision (EU) 2019/909 additional biological data are collected for species with Total annual Landings >200 ton. or share >10% of the average total EU landings in the previous 3 years. GFCM-DCRF dictates that additional biological data should be collected for species with mean Total Landings greater than 2% of total landings of the country for the reference years.

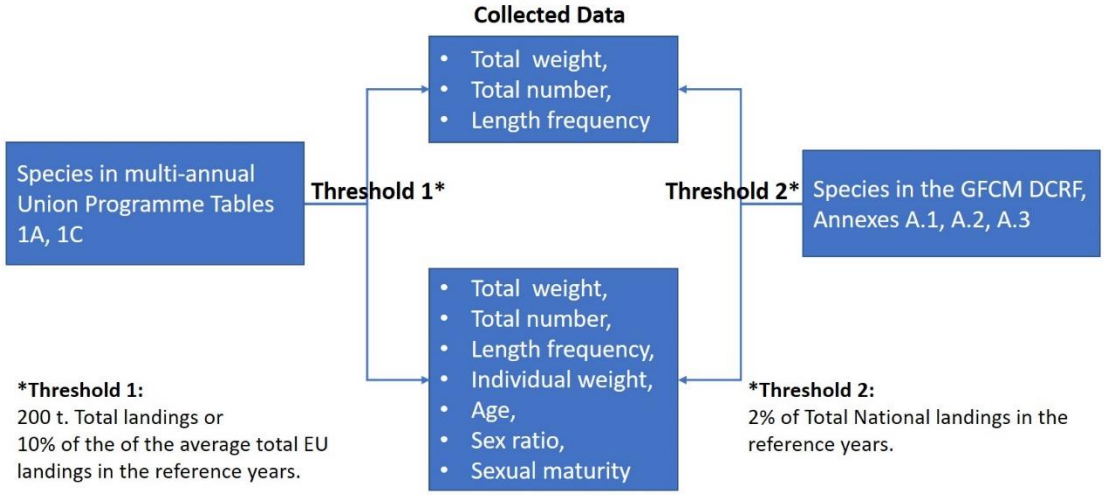


Figure 1. Species for which data are collected. According to Implementing Decision (EU) 2019/909, additional biological data (Individual weight, Age, Sex ratio and Sexual maturity) are collected for

species with mean Total Landings >200 ton. or share >10% of the average total EU landings in the three previous years. GFCM-DCRF dictates that additional biological data should be collected for species with mean Total Landings greater than 2% of total landings of the country for the reference years.

1.2 Targeted métiers-Technical stratification

Data are collected and reported by métier, at the aggregation level 6 (see Table 2, Commission Delegated Decision (EU) 2019/910). The list of metiers that was agreed at Regional level during the RCM Med&BS 2009, and have been updated annually since then, has been set as a reference for the selection of the metier to be sampled. On this reference list, a ranking system described in the Commission Decision 2010/93/EU is applied in order to designate the metiers to be sampled on a national level (technical stratification of the sampling scheme). More specifically, by using the average values of the reference years, the metiers are first ranked according to their share in the total commercial landings. The shares are then cumulated, starting with the largest, until a cut-off level of 90 % is reached. All metiers belonging to the top 90 % are selected for sampling. This process is then repeated according to the total value of the commercial landings and, a third time, according to the total effort in days-at-sea. All the metiers in the top 90 % in the above formulated lists are added to the selection. Additionally, a metier could also be picked up if it has special importance in terms of management. Based on this process, Greek sampling scheme usually covers the metiers in Table 1.

Table 1. The métiers usually included in the Greek sampling scheme.

Metier (level 6)	Spatial coverage
OTB_DEF_>=40_0_0	all GSAs
PS_SPF_>=14_0_0	all GSAs
FPO_DEF_0_0_0	GSA 22
GNS_DEF_>=16_0_0	all GSAs
GTR_DEF_>=16_0_0	all GSAs
LLD_LPF_0_0_0	all GSAs
LLS_DEF_0_0_0	all GSAs
SB_SV_DEF_0_0_0*	GSAs 20, 22

**The metier has been selected because it is in a Management plan*

1.3 Primary Sampling Unit, target and frame population

The Primary Sampling Unit (PSU) of the Greek sampling scheme is the fishing trip. The Target Population is all the fishing trips conducted by the fraction of the Greek fishing fleet using the targeted métiers, selected by the ranking system. The Sampling Frame is populated by data from the Official National Fleet Registry and by the DCF data collection system in the reference years.

1.4 Temporal and Spatial stratification

Sampling is temporally stratified by quarter. Greek fishing activities cover three GSAs: (a) Aegean Sea (GSA 22), (b) Ionian Sea (GSA 20) and (c) Cretan Sea (GSA 23) (see Fig. 2). The Hellenic coastline and marine area of the 3 aforementioned GSAs are divided in 12 sub-areas which constitute the next level of stratification within each metier (Fig. 3).

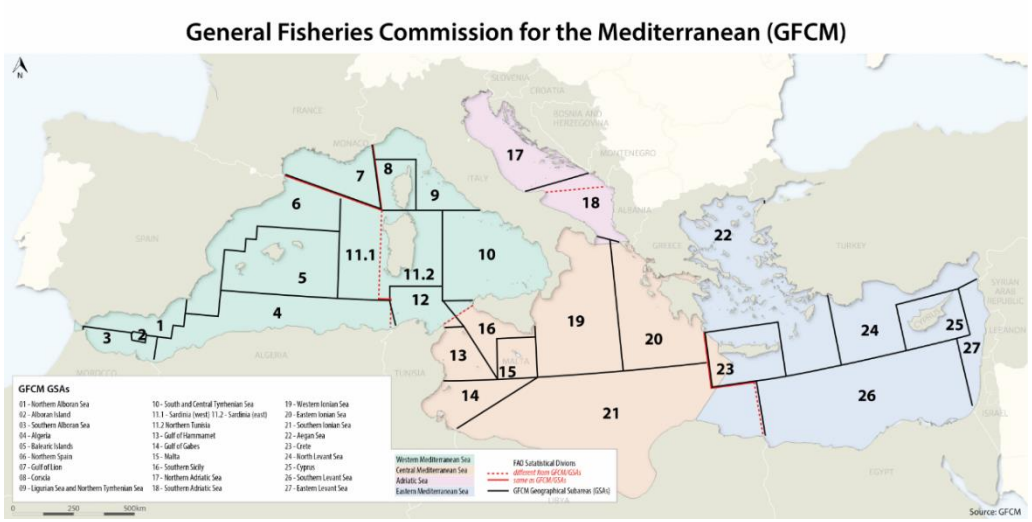


Figure 2. The division of the Mediterranean Sea in 27 Geographical Sub-Areas (GSAs). Greek sampling scheme covers the GSAs 20, 22 and 23.

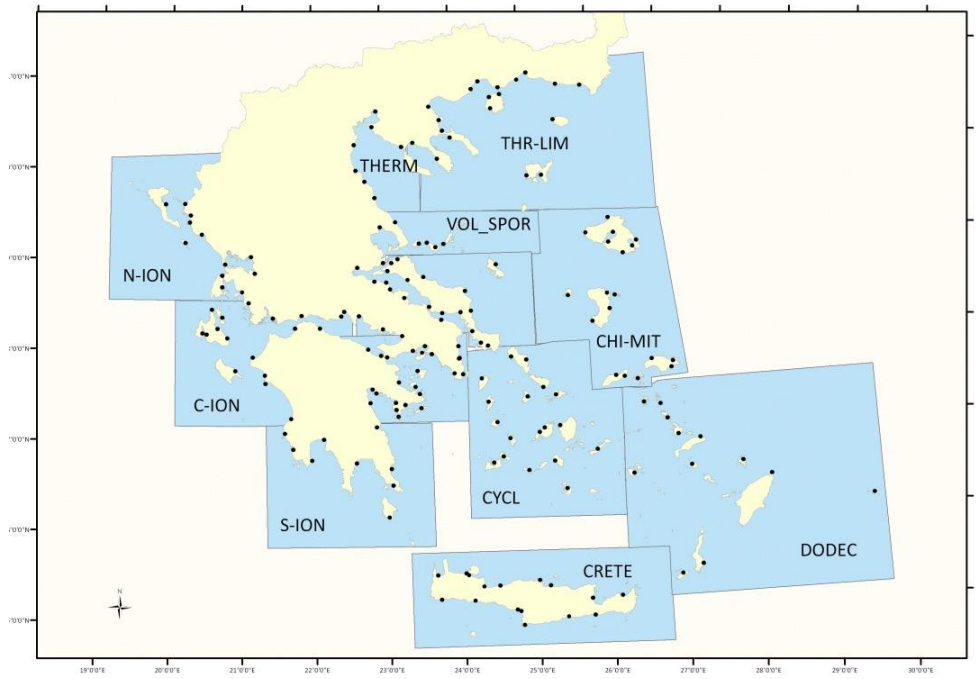


Figure 3. The subdivision of the Greek GSAs in sub-areas. GSA 20 is divided in 3 sub-areas (N-ION, C-ION and S-ION), GSA 22 is divided in 8 sub-areas (THERM, THR-LIM, VOL-SPOR, EVIA, ARGSAR, CHI-MIT, CYCL and DODEC) while GSA 23 corresponds to a single sub-area (CRETE).

1.5 Sampling scheme

Based on the aforementioned, the Greek sampling scheme is divided to two distinct programmes:

1. Sampling for total landings, discards and length distribution.
2. Sampling for additional biological data (individual weight, sex ratio, age and sexual maturity).

The total number of sampling trips and the allocation of sampling effort between the different strata is based on the corresponding distribution of effort, number of vessels and total landings and discards between the different strata in the reference years (see Data Quality Assurance Framework section).

1.5.a Sampling for total landings, discards and length distribution.

For all the species in the Tables 1A, 1C of multi-annual Union Programme and in the GFCM-DCRF Annexes A.1, A.2, A.3, sampling is conducted concurrently by observers either on board (at sea) or on shore (at landing sites) for all the metiers, except OTB_DEF_>=40_0_0 and SB_SV_DEF_0_0_0 where all sampling is conducted on board. For the metiers associated with small scale fisheries, 1/3 or of the sampling trips is performed at-sea and the 2/3 on-shore. Finally, for Purse Seines half of the sampling trips are conducted at-sea and the remaining samples are taken on-shore. The PSU selection is performed through random draw of a trip within each stratum, with replacement in case that the vessel owner refuses to cooperate. Consequently,

the sampling hierarchy is the following: Vessel trips are randomly selected within each stratum (i.e., *technical*-8 metiers x *temporal*-4 quarters x *spatial*-12 sub-areas stratification). At sea, all hauls are selected (no further stratification), and within each haul, the total weight and the number of individuals per species in landings are recorded. Regarding the discards, a 10% subsample of the total discard volume in each haul is used for calculations. On shore, the samples are taken from the whole volume of the landings. Finally, to record the length composition, a random sample of 50-100 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.

1.5.b Sampling for additional biological data (individual weight, sex, age and sexual maturity)

For all the species in the Tables 1A, 1C of multi-annual Union Programme with total landings >200 t. or share >10% of the average total EU landings in reference years (the three previous years) in reference years and for those in the GFCM-DCRF Annexes A.1, A.2, A.3 with mean Total Landings greater than 2% of total landings of the country for the reference years, additional biological data are collected. Samples are taken from the landings of at-sea and on-shore sampling trips. Samples obtained from scientific surveys (MEDITS, MEDIAS) can also be used supplementary, mainly for the non-marketable fraction of the stocks, and for the closed season of the trawl fishery. In addition, samples from the market or from discards can also be used, if the quota for each size group has not been achieved through the sampling trips, especially for the largest and the smallest specimens.

The sampling scheme is stratified random sampling, and PSU is the fishing trip. The spatial stratification of the sampling effort is limited to the GSA level and the technical stratification (different metiers) is not taken into account, since the aim is to derive the biological data on the stocks level, irrespectively of the fishing gears. Thus, the sampling hierarchy consist of the random selection of trip within each stratum (*temporal*-4 quarters X *spatial*-3 GSAs stratification) for each species. The biological variables (age, weight, sex ratio, maturity) are collected quarterly to detect seasonal differences in the structure and composition of the species examined. Regarding age distribution, quota sampling is employed, with the aim to collect 5-10 specimens (depending on the species) for each size class.

2. Data Quality Assurance Framework

The quality of the data collected within the Data Collection Framework (DCF) are assessed based on their precision and accuracy. Securing the precision of the provided data could be achieved through the use of indicators designed to assess the adequacy of the applied sampling scheme to provide precise data. The accuracy of the provided

data is related to the application of methods to mitigate data bias. Accuracy failures and bias sources in the collection, processing and reporting of fisheries data are many and varied (for a review see ICES, 2008 and ICES, 2010). In order to assure the quality of the data collected within the National Data Collection Programme, Greece has set up a Data Quality Assurance Framework, that could be analyzed in two major components: the deployment of mechanisms to detect and correct errors during the recording and the registration of the data to a database, as well as methods to assess and adjust the applied sampling scheme to avoid precision failures and mitigate biases.

2.1 Detection and correction of errors

Inaccuracies in the collection of data for DCF could stem from possible errors at the collection and registration of data to local databases. These errors increase the bias and reduce the precision of the collected data. A list of relevant sources of bias is reported in the Workshop on Methods to Evaluate and Estimate the Accuracy of Fisheries Data used for Assessment (WKACCU workshop, ICES, 2008). The practical methods applied within the Greek Data Quality Assurance Framework to avoid, detect and correct these errors are summarized in the following section.

2.1.a Protocols

-The sampling protocols that observers use are designed in a manner that reduces the possibility of misreporting.

2.1.b Species identification

-The stuff is frequently trained on species identification and taxonomy.

-Every haul as well as every rare species caught is photographed in order to facilitate species identification.

-If necessary, molecular techniques (DNA barcoding) are applied to quickly and accurately identify species.

-Contemporary species identification keys are available to the stuff.

2.1.c Age and Maturity

-The stuff is frequently trained on otolith-based age determination as well as on sexual maturity definition. The methods used are based on relevant international expert workshops.

2.1.d Database

The local databases are appropriately designed to prevent and detect possible errors during the data registration process. Specific treatments are integrated in order to

prevent misreporting of data, to detect duplicate entries and to detect unrealistic values on specific fields.

-All the fields in each database are masked, preventing the entry of erroneous data types or preventing blanks.

-Databases have integrated validators with automatic warning alert when a duplicate entry is registered.

-Databases have integrated validators with automatic warning alert when a registered species length or individual weight is not within the allowed species-specific ranges.

-Databases have integrated modules allowing the dynamic reporting of specific types of data. More specifically, these modules allow reporting in real time of: The total number of sampling trips (taking into account spatial and temporal distribution of sampling effort), the number of processed specimens for aging, sex and sexual maturity (total and per quarter), species specific length-weight relationship (with detection of outliers), species specific age-length relationship, age-length key and length frequency.

2.1.e *A priori* data quality checks

Additional data quality checks and verification procedures for the different stages of the sampling scheme, including the data quality checks for data capture and data validation, are applied using appropriately designed scripts within R programming environment (scripts are available at request). These scripts integrate the guidelines and the tools developed in the MARE/2016/22 – SI2.770115 "Strengthening Regional cooperation in the Area of fisheries biological data collection in the Mediterranean and Black Sea (STREAM)" as well as the tools developed within the COST Project (2009). Based on the relative classification introduced by the STREAM project, these checks fall into the category of *a priori* quality checks, since they are applied directly on sampling data.

2.1.f Methodologies and quality assurance procedures used for estimations

The estimation of the length structure of the catches is made using the methodology described in the Workshop on Sampling Calculation and Methodology for Fisheries Data (WKSCMFD, ICES 2004), which allows estimating the precision, in terms of coefficient of variation (CV) for each length class and for the whole Length Frequency Distribution (LFD) at métier level. Age distribution of landings is calculated using the estimated quarterly LFD of landings and the quarterly proportion of landings of age ***i*** in the length class ***j***, estimated from the age-length keys (ALKs). ALKs by quarter derive from age sub-samples of the length samples, taken as fixed number per length class. For the estimation of the age distribution both analytical and bootstrap methods are used. For discards the length frequency distributions is calculated from the

proportions of sampled discard weights, numbers and total landings. Calculations are made by size class and quarterly before being summed to give total annual discards by size class. Discards are added to landings data to give the total catches, which then is used for stock assessment.

As far as biological variables are concerned, estimation procedures are the following: The age compositions are estimated from two-stage sampling where random length samples are taken, and length-stratified age samples are used to construct an ALKs. The length distributions are obtained from random samples. A length-weight relationship is created and fitted to estimate weight-at-length, and weight-at-age is estimated from this using an age-length-key. Sex-at-age is estimated using a sex-age-length-key. Maturity-at-age is estimated using a maturity-age-length-key or, if appropriate, a sex-maturity-age-length-key. CVs for maturity-at-age and size-at-age, for all sampled species, is calculated annually at GSA level. Estimates are made using the "open source" tool developed in the frames of the COST project (Vigneau, 2008), which allows to assess the accuracy of the biological parameter estimates collected for stock assessment purposes.

2.1.g *A posteriori* data quality checks

Regarding *a posteriori* quality checks, scripts developed under the STREAM project are used, contacting quality checks on the EU MED&BS Data Call formats. The checks focus on the time and spatial coverage among the strata (i.e. quarter, métier) and on the assessment of the completeness of biological information.

2.2 Sampling scheme assessment methods

According to WKACCU Workshop (ICES, 2008), some major causes of bias regarding the design of the sampling schemes are: (1) incomplete sampling frame, and (2) biased sampling procedures for selecting vessels from the sampling frame or by factors preventing the deployment of observers on all selected vessels. Another important aspect of the sampling scheme is the sampling effort and the way it is distributed temporally and spatially, taking into account the optimal use of available resources. Greece applies a series of assessment methods to evaluate the applied sampling scheme, to detect and prevent possible sources of bias, and to be able to adjust and improve the applied sampling scheme annually. In order to qualitatively assess the sampling scheme, a scorecard developed under the WKACCU Workshop (ICES, 2008) is used, which is "*a practical tool to evaluate the quality of data sources used for stock assessments, and can help reduce bias in future data collections by identifying steps in the data collection process that could be improved*".

2.2.a Sampling frame

The Sampling Frame used for the Greek sampling scheme is populated by data from the official national fleet registry. However, the Greek fishing fleet consists of numerous vessels (it is the largest fishing fleet in the EU) of low tonnage and power. According to the National Fleet Register (31/12/2018), the fleet consists of 14,123 fishing vessels with the great majority (~95%) of the fleet being small vessels (average length 7.5 m) exploiting the extensive coastline of the mainland and of the numerous Greek islands (15,000 km, covering more than 6,000 islands and islets), targeting coastal fishing stocks. A subset of this part of the fleet, although registered in the official national fleet registry, could be virtually inactive due to a variety of reasons. Provided that the part of the inactive fleet is substantially large and unevenly spatially distributed, it could reduce the accuracy of the Sampling Frame leading raising methods to overestimations.

In order to avoid this kind of bias, each year within the DCF, the level of activity of the small scales fishing fleet is assessed through in situ tracing of the active fleet at selected harbors.

2.2.b Sampling procedures for selecting vessels

For the selection of the sampling trips within each stratum a stratified random sampling selection with replacement is applied, which means that any vessel within a stratum has the same chance of being selected, even if it has been previously selected. In this process, bias could be introduced if specific vessels are systematically unavailable for sampling, due to reluctance to cooperate or due to inability to accommodate on-board observers (i.e. in smaller vessels).

To reduce the degree of refusal of cooperation, various incentives are used, such as of scientific support, assistance to fishermen in matters of their concern etc. To include the smaller vessels within the sample, 2/3 or of the sampling trips related to small scale fisheries métiers is performed on-shore, at landing harbors.

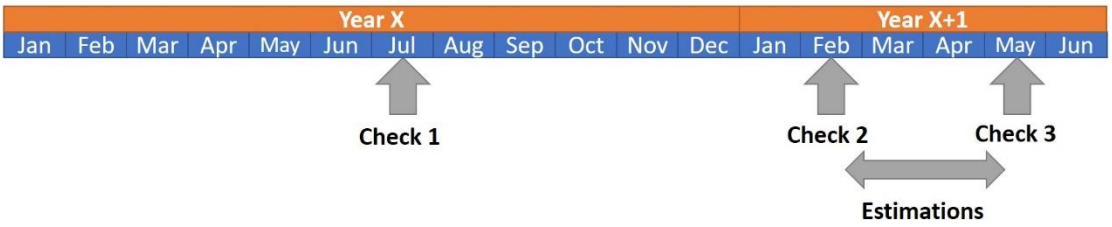
2.2.c Sampling effort and coverage

The estimation of the total sampling effort for biological variables as well as its temporal and spatial distribution are calculated based on a ranking system using the corresponding distribution of fishing effort, number of fishing vessels and total landings and discards between the different strata in the reference years. This process is integrated with the outcomes of the application of the methods delivered under MARE/2014/19Med&BS project (Deliverable 2.5 and 4.2) implementing the Sampling Design tool for optimization of sampling intensity and data quality checks. The methodological framework developed in WP2 of MARE/2014/19 MED&BS is implemented in R scripts built on COSTS tools (Vigneau, 2008) and it allows the analysis of the evolution of the sampling variability with the number of samples, the

assessment of the current plans in terms of over/under-sampling of the fisheries and it facilitates the proposal of an "optimal" sampling plan (in terms of number of trips and individuals to sample) disaggregated over strata (e.g. quarters and métiers). Additional methods for assessing the sampling effort proposed by the ICES WKBIOPTIM Workshop (ICES, 2019) are used when and if necessary.

2.3 Data quality assessment calendar

The temporal distribution of Quality Checks within a year is adjusted to the data calls for submitting catch, effort and biological data (DG MARE MED&BS Data Call, FDI Data Call). Usually the deadlines for these Data Call are starting in early June. Based on this timeframe, three major Quality Checks on annual data are conducted as summarized on the following diagram. Apart from the 3 major Data Quality Checks, periodic comparisons between the hardcopies (data protocols) and the digital information (database entries) are conducted by trained technicians, in order to identify potential omissions and/or erroneous entries in the database.



Check 1

The Quality Check 1 is conducted in July of the year X. *A priori* quality checks are applied on the first semester data. Additionally, the sampling scheme assessment methods are applied by using as reference the three years before Year X. Possible adjustment on the sampling plan for Year X+1 are decided.

Check 2

The Quality Check 2 is conducted in February (or March) of the year X+1. *A priori* quality checks are applied on the second semester data.

Estimations

The estimations on catch, effort and biological data for the Data Calls are conducted between February and May of the year X+1.

Check 3.

A posteriori quality checks are applied on the annual dataset.

References

ICES. 2004. Report of the Workshop on Sampling and Calculation Methodology for Fisheries Data (WKSCMFD), 26-30 January 2004, Nantes, France. ICES CM 2004/ACFM:12. 242 pp.

ICES. 2008. Report of the Workshop on Methods to Evaluate and Estimate the Accuracy of Fisheries Data used for Assessment (WKACCU), 27-30 October 2008, Bergen, Norway. ICES CM 2008/ACOM:32. 41 pp.

ICES. 2010. Report of the Workshop on methods to evaluate and estimate the precision of fisheries data used for assessment (WKPRECISE), 8-11 September 2009, Copenhagen, Denmark. ICES CM 2009/ACOM:40. 43 pp.

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

Vigneau J., 2008. Common tool for raising and estimating properties of statistical estimates derived from the Data Collection Regulation (COST). European Commission, FISH/2006/15– lot 2, Project no: SI2.467814. Title: Studies and Pilot projects for carrying out the common fisheries policy. 118 pp.