

Quality Assurance Framework

Institute full name: Spanish Institute of Oceanography
Institute acronym: IEO
Institute webpage: <http://www.ieo.es>
Project full title: Seguimiento y análisis de pesquerías
Project acronym: SAP6
Project webpage: <http://www.proyectosap.es>
Team full title: Equipo de Muestreos SAP
Team acronym: EM-SAP

Deliverable:
Bias, precision and related data validation



| | |
|---|--|
| Deliverable Name | Bias, precision and related data validation |
| Deliverable No. | 3 |
| Date of preparation of this version: | 22/05/2019 |
| Authors: | Jose Rodríguez, Jose Luis Cebrián, José Miguel García, Eva Velasco, Jose Castro |
| Status (F: final; D: draft; RD: revised draft): | Final |
| Version: | 1.0.0 |
| Copyright |  This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-sa/4.0/.. |

Revision History

| Version No. | Date | Details |
|-------------|------------|-------------------------------------|
| 1.0.0 | 22/05/2019 | <i>1st final version</i> |
| | | |
| | | |

List of abbreviations

| Abbreviation | Definition |
|--------------|--|
| EM-SAP | Sampling team of SAP project |
| EU | European Union |
| QAF | Quality Assurance Framework |
| NVDP | Spanish official fisheries data (logbooks and sales notes) |
| OAB | On-board sampling network (ICES area) |
| RIM | On-shore sampling network (ICES area) |
| SAP | IEO project in charge of fisheries sampling and data provision (ICES area) |
| SGPIDS | ICES Study Group on Practical Implementation on Discard Sampling Plans |
| SIRENO | IEO database (Spanish acronym for “Integrated Monitoring of Ocean Natural Resources”) |
| WGCATCH | ICES Working Group on Commercial Catches |
| WKPICS | ICES Workshop on Practical Implementation of Statistical Sound Catch Sampling Programs |



Table of Contents

| | |
|--|-----------|
| 1. Introduction | 4 |
| 1.1. About the Quality Assurance Framework | 4 |
| 1.2. QAF version in English | 4 |
| 2. Bias general introduction | 5 |
| 3. Comparative analysis sampling vs official information | 6 |
| 3.1. Pairing/crosschecking process trips | 6 |
| Pairing methodology | 6 |
| 3.2. Comparative analysis of LPUEs | 6 |
| 4. Bias due to selection of PSUs: on-board sampling | 8 |
| 4.1. Methodology | 8 |
| List creation | 8 |
| 4.2. Contact protocol and responses registration | 8 |
| 4.3. Bias Monitoring analysis | 9 |
| 5. Length distribution analysis: COST | 10 |
| 5.1. Methodology | 10 |
| 5.2. Goals | 10 |
| 5.3. Validation | 10 |
| 5.4. Responsible and time | 11 |
| 6. Main bibliography of interest | 12 |



1. Introduction

The present document has been developed by EM-SAP, in the frame of the SAP project and constitutes a deliverable as part of the overall Quality Assurance Framework.

1.1. About the Quality Assurance Framework

QAF describes the framework for the quality assurance of the team, and particularly the elements, actions and measures to be taken by the team members to ensure the quality of the project and its outcomes.

The main objectives of the QAF are:

- To ensure that all inputs of the EM-SAP are consistent; and
- To ensure that all outputs of the project are of the highest quality.

In order to ensure the achievement of the above objectives, the QAF aims to define procedures for ensuring best practices for every process undertaken by the team thus affecting to all process included in the team protocol.

The aim for the QAF is to be applicable to all EM-SAP activities in the long term.

The QAF will be a dynamic document throughout the lifecycle of EM-SAP, and will be updated as required.

1.2. QAF version in English

QAF addresses specific actions, measures and individual responsibilities and it's therefore produced and restricted to the members of the team. This includes the fact that EM-SAP works and produces most part of its work in Spanish.

Nevertheless, there is a need to transmit properly the processes and actions taken around the production, handling and management of the sampling information in a wider EU Data Collection context, where this team is incorporated.

Therefore, a reduced version of the QAF is produced in English thus allowing end-users, partners, reviewers and the European Commission to understand what and how is being done.

For further information or specific details please contact Jose Rodriguez (jose.rodriguez@ieo.es).



2. Bias general introduction

Several issues related to the sampling affect the accuracy and therefore, the quality of the data and estimates obtained. WGCATH pointed out (WGCATCH, 2017) three basic groups: aspects of bias related to design; aspects of bias related to implementation, and precision.

This document explain some of the work developed by EM-SAP mostly affecting the three groups, nevertheless specific sampling design elements will be explained in separate QAF document.

- Unbiased and precise estimates of commercial fisheries data are essential inputs to stock assessment and fisheries management. Concerning these parameters, the objective of the sampling is to obtain data representative of the population that can be used to estimate the population characteristics of interest accurately (unbiased) and precisely (low variability).
- Precision is a measure of the size of the closeness of agreement among individual measurements. Accuracy is a measure of the closeness of measurements to the true value. Precision is high when measurements uncertainties are relatively small.



3. Comparative analysis sampling vs official information

3.1. Pairing/crosschecking process trips

IEO receives official fisheries data (logbooks, sales notes and the operational fishing fleet census) during the first months of the following year from the Ministry. IEO is in charge of the hierarchical classification into metiers (Decision of the Commission of November 6, 2008), which constitute the fundamental units of aggregation of fishery information.

SAP team builds the NVDP database of official data that our team uses to review the design of the sampling plan, monitor fisheries dynamics, process sampling information, etc.

The pairing/crosschecking process between the sampled trips and the official data consists in crossing both sources through an R script in order to assign to each sampled trip the corresponding fishing trip of the NVDP.

Pairing is done for both, the on-shore sampled trips (RIM) and on-board sampled trips (OAB); including contacted OAB trips which were rejected.

This procedure is set after the annual sampling review (see “QAF EMSAP Data Integrity, Data quality” deliverable) with the aim of:

- Assign the ID logbook of the NVDP to the sampled trips.
- Contrast and consolidate the information of the sampled trips.
- For RIM trips:
 - Record catches profile and catches by species.
 - Confirm the fleet activity.
 - Georeference the sampled trip: fishing Division and ICES rectangle.
 - Obtain specific trip variables not collected by the samplers (e.g. fishing days, deep, etc).
 - Obtain the sale location (sales notes cross checking), the location where landings are accessible for sampling (e.g. relevant to assess the coverage done to Spanish fleet landing abroad)
- For OAB trips, same information is collected. In this case, since most part of variables are collected on-board by the observers, information from logbooks are specially relevant for those trips where observer couldn't get on-board (refusals, etc).

Pairing methodology

This process is done through R after preparation of both datasets. Match is mainly done based in:

- NVDP: Landing Date/Fishing date and vessel ID code.
- Sampling: Sampling Date and vessel ID code.

Matches are reviewed based on supervisors knowledge of specific port or fleets dynamics.

3.2. Comparative analysis of LPUEs

Comparative analysis of DPUEs between Logbooks (DP), sale notes (NV) and sampling data allows



the identification of problems in:

- Selection of trips.
- Metiers coverage.
- Concurrence of sampling (e.g. not access to certain species)
- Accessibility problems in ports to certain species, categories or part of the catches.

The landings comparison report is a routine task within the team to perform after the annual closure of samplings.

The results will be examined at the sampling team meeting so that the team can investigate the problems of **representativeness** and **concurrence** in sampling.



4. Bias due to selection of PSUs: on-board sampling

4.1. Methodology

Vessel selection is one of the major sources of bias for on-board sampling programs worldwide. Based on different ICES groups recommendations such as WGCATCH, SGPIDS and WKPICS, EM-SAP moved towards a probability-based approach in 2016 to reduce potential bias¹. Thus changing from a previous ad-hoc selection -more likely to produce bias- to a random selection with replacement.

List creation

The supervisors (EM-SAP) of each one of these fleets are in charge of the creation of the list at the beginning of the year, so that it is available for the accomplishment of the first sampled survey.

List is generated based on the register of active vessels in the fishery to minimize sampling frame errors.

To create it, an R script is used to generate a random list from the available version of the current year's census fleet.

Within each sampling stratum, vessels are randomly selected without considering any additional factors (port of landing, port of registration, target species of the trip, etc.).

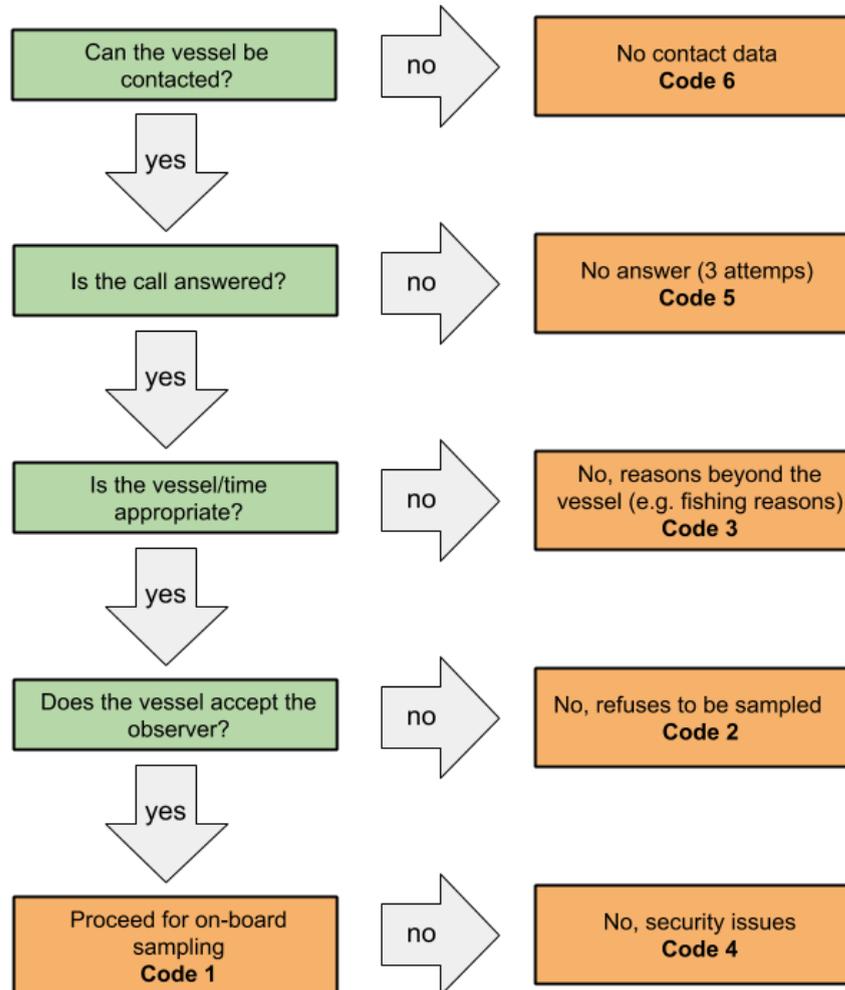
4.2. Contact protocol and responses registration

The protocol includes the full registration of responses into 6 categories including 5 different types of non-responses.

| CODE | Description |
|------|--|
| 1 | Affirmative: shipment completed. |
| 2 | Negative of the skipper (declines collaboration, proposal of another later tide...). |
| 3 | Lack of availability of the vessel or target tide: temporary (repair, lack of punctual space, seasonal activity in another metier) or definitive (no structural space, vessel sold, scrapped, sunk). |
| 4 | Refusal of the observer for safety reasons, etc... |
| 5 | No answer: impossibility of contact. |
| 6 | No contact details (after exhausting all sources of information). |

¹ As an exception, the selection process of vessels for the stratum IEO_S_RP_AC depend on the collaboration with the fishing association ANASOL, to which all vessels belong. Since the beginning of the sampling of this fleet, the sampling has been carried out with the collaboration of this association. The exploration of the sampling data obtained until now (number of unique vessels sampled, catches, discards, etc) indicates that although it is probably not strictly random, it is determined by logistical requirements.

Diagram protocol of contact for vessels:



4.3. Bias Monitoring analysis

Statistical tests and analysis are based on landings with the aim of looking for potential bias regarding the on-board vessel selection.

Specific objectives:

- Analysis of representativeness and coverage.
- Evaluate if the sampled trips have a fishing behaviour that substantially deviate from the non-observed trips (test statistical significant differences).
- Analysis of the differences for 6 different types of responses received from the phone calls (test statistical significant differences).

5. Length distribution analysis: COST

It consists of the joint exploration of the size structure sampled by species for outliers detection and calculation of precision estimators. It does validate the data and makes them available for use in the rest of processes.

5.1. Methodology

Length analysis is done using the R-developed open-source software COST. This software allows the detection of outliers in length frequencies in the most disaggregated level through the Delta measure (Vigneau and Mahevas, 2007).

This method summarises the length structure of one sample by one value, indicating the discrepancy between the sample and the overall mean.

The goals of this tool are double: (i) to understand and / or to quantify the contribution of individual samples to the overall variance, and (ii) to estimate the similarities between samples.

Two scripts have been developed: one for the mixed species and another for all the others. Input files are generated directly from SIRENO database.

The analysis is carried out taking into account the technical (DCF métier Level 6) and temporal (quarter) stratification that fits the design of our sampling.

Type of graphs used:

- deltaplot graphs to identify outliers,
- graphs for specific review of trips identified as outliers (number of specimens measured by size of the sampled trip versus number of specimens measured by size for all trips),
- joint comparative graphs with all species/stratum size distributions and graphs to review the proportionality between sampling intensity and landings.

5.2. Goals

- Outliers detection based on the joint exploration of the size structure sampled by species. Validation of length data to be used in the subsequent processes.
- Analysis of proportionality between "sampling intensity" and official landing statistics, in order to correct deficits in our sampling strategy.
- Precision statistical estimators: calculation of size coefficients of variation, aggregated quarterly or annually, which were previously required by the DCF (COSTdbe library).

5.3. Validation

Validation consolidates data and enables it to be used in size processes.

To this end an specific tick has been implemented in the national database (SIRENO). All validated



data must be marked in SIRENO. These ticks will be done through the central database service.

5.4. Responsible and time

This task is allocated to a supervisor, in charge of the analysis for all the sampling data. This ensures homogeneity along the areas, fleets and species for all the sampling network.

Validation consolidates data and enables it to be used in size processes. It is not possible to perform validation before the partial closure of samples at the beginning of the following year.

Once closed, species validation is performed successively so that they are available before IEO begins to respond to each WG.



6. Main bibliography of interest

COST project (Common tool for raising and estimating properties of statistical estimates derived from the Data Collection Regulation), 2009. FISH/2006/15 – lot 2. <http://wwz.ifremer.fr/cost>

Gilbert, Richard O. Statistical methods for environmental pollution monitoring. TD193.G55 1987 628.5'028 86-26758. ISBN 0-442-23050-8

ICES. 2017. Report of the Working Group on Commercial Catches (WGCATCH), 7-11 November 2016, Oostende, Belgium. ICES CM 2016/SSGIEOM:03. 142 pp.

Vigneau, J., and Mahévas, S. 2007. Detecting sampling outliers and sampling heterogeneity when catch-at-length is estimated using the ratio estimator. – ICES Journal of Marine Science, 64: 1028–1032.

