



Documentation of the

**ON-SHORE SAMPLING PROGRAM**

**OF THE SPANISH (NON-BASQUE) FLEET OF EUROPEAN ATLANTIC WATERS**

developed by

**IEO**

*(Instituto Español de Oceanografía)*

**Date:**

28/09/2021

**Authors:**

J. Castro, D. Espino, M. Amez, J. Cebrian,

J. Acosta and B. Castro

## Table of Contents

FIGURES .....	3
INTRODUCTION .....	4
1. DESCRIPTION OF THE POPULATION .....	5
2. SAMPLING PROTOCOL.....	6
3. SAMPLING IMPLEMENTATION .....	7
4. DATA CAPTURE.....	8
5. DATA STORAGE .....	9
5.1. National database .....	9
5.2. International database .....	10
5.3. Quality check and data validation .....	10
6. DATA PROCESSING .....	12
6.1. Evaluation of data accuracy .....	12
6.2. Editing and imputation methods .....	12
6.3. Validation of the final dataset.....	13
ANNEX I .....	14
ANNEX II .....	16
ANNEX III .....	18

## FIGURES

Figure 1. Structure and geographical allocation of the IEO's data base "SIRENO" .....	9
<b>Figure 2.</b> Monthly dump sampling revision in the national database SIRENO. ....	11
Figure 3. Measurement of length for fish, crustaceans and cephalopods. ....	14
Figure 4. Sex identification of crustaceans: male (left), female (right).....	15
Figure 5. Sex identification of elasmobranchs: male (left), female (right). ....	15
<b>Figure 6.</b> Sampling sheet for landings.....	16
<b>Figure 7.</b> Sampling sheet for length frequency distributions. ....	17
Figure 8. Cook's distance plot to detect outliers ( <i>e.g.</i> length frequency distributions). ....	18

## TABLES

Table 1. Strata of the IEO on-shore sampling program of the Spanish (non-Basque) fleet of European Atlantic waters, with their respective ports. ....	5
Table 2. Total DCF métiers of the Spanish (non-Basque) fleet of European Atlantic waters, indicating their inclusion (Y) or not (N) in the IEO on-shore sampling program.....	6

## INTRODUCTION

The present document has been developed by the SAP project ("*Seguimiento y Análisis de las Pesquerías Españolas en el Atlántico Nordeste*"<sup>1</sup>) of IEO ("*Instituto Español de Oceanografía*"<sup>2</sup>) and constitutes the sampling quality documentation of the EU MAP work plan for the on-shore sampling program of the Spanish (non-Basque) fleet of European Atlantic waters. This document describes the elements, actions and measures to be taken by the team members to ensure the quality of the data under the responsibility of the SAP project. It has been written as a summary in English of the detailed internal protocol in Spanish in order to facilitate the sampling design and corresponding quality control process to the EU MAP reviewers.

This document has been structured according to the EU MAP guidelines, adapting our internal quality control process, which can be summarized in the following four steps:

1. **Supervision:** monthly monitoring of the sampling coverage to check that the number of sampling units determined in the sampling plan has been fulfilled.
2. **Verification:** checking of the integrity of computerized data to verify its correct recording in the database, taking as reference the original sampling forms. This procedure is automated by applying an R script to facilitate the revision of a number of variables.
3. **Matching:** crossing the sampled trips with the official logbooks to assign the same trip ID, as well as the DCF *métier*. Also automated by an R script.
4. **Validation:** final acceptance of the sampling data for scientific use, through statistical analysis of a number of variables of the set of sampled trips by weighting domain (*métier*-quarter). It allows the detection of outliers. Also automated by an R script.

For further information or specific details please contact Jose Castro ([jose.castro@ieo.es](mailto:jose.castro@ieo.es)).

---

<sup>1</sup> English translation: Monitoring and Analysis of Spanish Fisheries in the Northeast Atlantic.

<sup>2</sup> Spanish Institute of Oceanography.

## 1. DESCRIPTION OF THE POPULATION

The **IEO on-shore sampling program of the Spanish (non-Basque) fleet of European Atlantic waters** (*i.e.* EEZ of UK, Ireland, France, Spain and Portugal in Subareas ICES 6, 7, 8 and 9) follows a sampling scheme type based on commercial fishing trips. This sampling scheme aiming at collecting length samples from commercial landings for all species listed in Table 1 of the EU MAP Delegated Decision annex.

The target sample population is made up of **on-shore events (port\*day)**, from a selection of 22 major national ports (from 165 total ports, both international and national), which host around 44% of the target trips and 66% of the landings. Thus, a combination of major ports and weekdays is obtained, excluding foreign ports, the national minor ports and weekends, when fishing is banned in Spain. This target sample population is **stratified in 3 geographical fishing grounds with different fleets landing at their respective ports**:

Nº	Strata identifier	Description	Nº of ports	Port names
1	IEO_P1_M_CN	Spanish (non-Basque) fleet operating in the national fishing ground of Cantabrian-Northwest.	17	A Coruña, Avilés, Burela, Cedeira, Celeiro, Fisterra, Gijón, Llanes, Luarca, Marín, Muros, Ribeira, San Vicente de la Barquera, Santander, Santoña, Suances and Vigo
2	IEO_P1_M_GC	Spanish fleet operating in the national fishing ground of the Gulf of Cadiz.	5	Barbate, Isla Cristina, Rota, Sanlúcar de Barrameda and Tarifa
3	IEO_P1_M_NEAFC	Spanish (non-Basque) fleet operating in the NEAFC non-Spanish waters.	6	A Coruña, Avilés, Burela, Celeiro, Gijón and Vigo

**Table 1.** Strata of the IEO on-shore sampling programme of the Spanish (non-Basque) fleet of European Atlantic waters, with their respective sampled ports.

## 2. SAMPLING PROTOCOL

In relation to the procedure for selecting primary sampling units (on-shore event=port\*day), the visits to the fixed number of ports established in the sampling plan are carried out trying to maintain a temporary periodicity throughout the year (generally weekly). The selection of the sampling day is not random, but depends on the availability of trips landed in port according to the seasonality of the fisheries and the fishing strategy of the métiers. Besides, only weekdays are included in the sampling schedule, since Spanish legislation prohibits fishing activity on weekends. Therefore, the method of selecting on-shore events for sampling is defined as Non-Probabilistic Judgement Sampling (“**NJPS**”).

After the selection of the PSU, the secondary sampling unit (**SSU**) is determined by the **landing event**. This allows avoiding the assumption of full access to the total landing of the fishing trip, which may not be true and must be confirmed later after checking the official declarations. The selection of landing events is concentrated in 17 of the 24 total *métiers*:

Nº	DCF métier	Sampling
1	FPO_FIF_>0_0_0	N
2	FPO_MOL_>0_0_0	Y
3	GND_SPF_31-39_0_0	N
4	GNS_DEF_>=220_0_0	Y
5	GNS_DEF_100-119_0_0	Y
6	GNS_DEF_120-219_0_0	Y
7	GNS_DEF_50-59_0_0	Y
8	GNS_DEF_60-69_0_0	Y
9	GNS_DEF_90-99_0_0	Y
10	GTR_DEF_50-59_0_0	Y
11	GTR_DEF_60-69_0_0	Y
12	LHM_CEP_0_0_0	N
13	LHM_DEF_0_0_0	N
14	LHM_DWS_0_0_0	Y
15	LHM_SPF_0_0_0	Y
16	LLS_DEF_0_0_0	Y
17	OTB_DEF_100-119_0_0	Y
18	OTB_DEF_70-99_0_0	Y
19	OTB_MCD_55-64_0_0	N
20	OTB_MPD_55-64_0_0	Y
21	PS_SPF_0_0_0	Y
22	PTB_MPD_55-64_0_0	Y
23	SDN_MCF_55-64_0_0	N
24	TBB_MOL_55-64_0_0	N

**Table 2.** Total DCF métiers of the Spanish (non-Basque) fleet of European Atlantic waters, indicating their inclusion (Y) or not (N) in the IEO on-shore sampling programme.

### 3. SAMPLING IMPLEMENTATION

The selection process of sampling units includes the annotation of the number of trips rejected for sampling. This information is used to improve the design and sampling protocols, for example to detect changes in port logistics that advise looking for alternative ports. However, work is being done to identify the rejected trips so that later they can be located in the logbooks, since this will allow to carry out bias analysis.

The general coordination of the on-shore sampling program requires the monitoring of the sampling progress throughout the year. For this, the monthly monitoring of the sampling coverage is carried out, adapting the sampling intensity when variations in fishing activity occur, so as to guarantee the quarterly robustness of the scientific data. This process corresponds to the first of the 4 steps indicated in the Introduction section: **“SUPERVISION”**.

## 4. DATA CAPTURE

For each trip, the sampler proceeds to the biological sampling of the available **landings categories**, which are usually classified in **commercial categories**. This determines a new level of sampling stratification, where all the commercial categories of each species must be sampled. The sampling is of a **concurrent** type, that is, it covers all the species landed. The data to be collected are the **taxonomic identification** of the species and the **length** of the individuals.

The latter consists of the measurement and annotation of the total length (cm) for fish, the cephalothorax length (mm) for crustaceans, and the mantle length (cm) for cephalopods. Fish and cephalopods measurements are taken with a ruler, while the cephalothorax of crustaceans is measured by calliper. All species are measured to the lower centimeter (cm), except anchovy (*Engraulis encrasicolus*) and sardine (*Sardina pilchardus*), which are measured to the lower half centimeter (1/2 cm), and Norway lobster (*Nephrops norvegicus*), at the lower millimeter (1 mm). In the case of crustaceans and elasmobranchs sex is also recorded ([ANNEX I](#)).

In addition to the information derived from the biological sampling, the sampler must also take note of the **technical data** of the fishing trip: name of the vessel, number of vessels participating in the trip (in the case of pairs of trawlers made up of three rotating vessels), *métier*, etc.

On-shore sampling data is collected by **voice recording** or written directly on the sampling sheets designed specifically for this. In the first case, the sampler themselves transfer their voice recordings to paper later. The objective is to have all the sampling data in **paper format** for later filing and consultation ([ANNEX II](#)).

Subsequently, this information is then **computerized**, either by the observer himself or by specialized staff with direct contact with the sampler to remedy any doubt interpretation. The digitized data are collected in flat files for later uploading to the IEO's biological-fisheries database.



## 5. DATA STORAGE

All data recorded must be legible (readable) and permanent, in order to guarantee their accessibility throughout the data life cycle. This includes the storage of human-readable metadata that may be recorded to support an electronic record.

### 5.1. National database

The data collected by the IEO on-shore sampling program of the Spanish non-Basque fleet of European Atlantic waters are stored in the IEO database on Ocean Natural Resources, called “SIRENO<sup>3</sup>”.

SIRENO is a standardized storage system of multidisciplinary oceanographic data. This computer application integrates different modules in which to register both biometric samplings of the Spanish commercial catches and oceanographic data from the IEO’s scientific surveys. In this way, the application allows the collection, storage and management of the basic data necessary for oceanographic and fisheries research, in order to evaluate the fluctuations of the stocks and the biotic and abiotic factors that condition them.

SIRENO has been developed on a relational database environment in ORACLE language. The database is located on a computer that acts as a data server, with INTEL technology, under Windows NT operating system. The IEO users can access from any of its 8 coastal laboratories through a computer connected to the IEO’s intranet and duly identified by its IP.

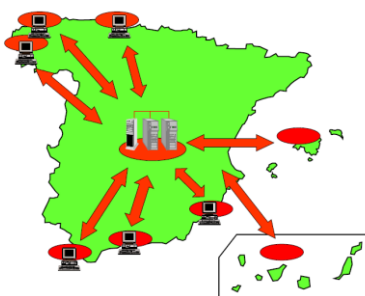


Figure 1. Structure and geographical allocation of the IEO’s data base “SIRENO”.

The SIRENO application is structured in a series of screens, first accessing its home screen where the main menu appears:

- Masters: main tables from which the entire application feeds.
- Data storage: screens for entering fisheries management data.
- Processes: standard calculation procedures.
- Surveys: entry of trawl surveys data.
- Acoustic: entry of acoustic surveys data.
- Reports: exploitation of the data entered by generating lists, reports or files.
- Utilities: general information screens and application services.

---

<sup>3</sup> “Seguimiento Informático de los Recursos Naturales Oceánicos”

## 5.2. International database

On a supranational scale, the data collected by the IEO on-shore sampling program of the Spanish non-Basque fleet of European Atlantic waters are stored in the “Regional DataBase”. This **RDB** is a regionally coordinated database platform for fisheries assessments, which covers fisheries in the North Atlantic Ocean, the North Sea and the Baltic Sea. It addresses fishery management needs related to the European Union Common Fisheries Policy.

The platform was developed by the National Institute of Aquatic Resources at the Technical University of Denmark (DTU-Aqua). Since 2012, the RDB is hosted and maintained by ICES for the preparation and analysis of commercial catch and landing data received from the cooperating countries: <https://www.ices.dk/data/data-portals/Pages/RDB-FishFrame.aspx>

## 5.3. Quality check and data validation

During the process of computerization of the sampling data, the second step of quality control is carried out: “**VERIFICATION**”. To do this, the SAP team has established a **data integrity** assurance system, in which both physical and logical integrity are considered. In the first one, elements associated with the management, physical and electronic storing are included, as well as the detection of erroneous data, mostly related to human-induced errors in the computerization process. In the second one, different algorithms are applied to ensure the correctness or rationality of the data, thus including referential integrity, check constraints, etc.

This process is carried out in several phases that involve different quality checks and treatments of the information received which can be summarized in 3 steps:

1. Pre-dump revision: once the digitized data are received, an R script (<https://github.com/Eucrow/IPDtoSIRENO>) is used to homogenize the information with respect to the SIRENO structure. In addition, the files are exported to the appropriate format required for direct dumping into SIRENO.
2. Dump: it consists of pouring information from the previous phase into SIRENO (*i.e.* SOP weights, etc.). It is carried out by the SIRENO's computer service.
3. Post-dump revision: from the SIRENO output reports, an R script ([https://github.com/Eucrow/revision\\_volcado\\_R](https://github.com/Eucrow/revision_volcado_R)) is used to detect errors (objective mistake which must be fixed) and warnings (possible error). Errors and warnings are reviewed by the on-shore sampling supervisor. It usually entangles revision of originally paper sampling sheets and/or communication with the observer.

The periodicity of this process is monthly, however after closing the sampling year, at the latest before mid-February of the following year, the last step is repeated on the whole set of the annual data to broaden perspective and confirm the monthly corrections. Then the “checking box” is marked in the trip header in the SIRENO database, indicating that the sampled trip can be moved to the validation step.

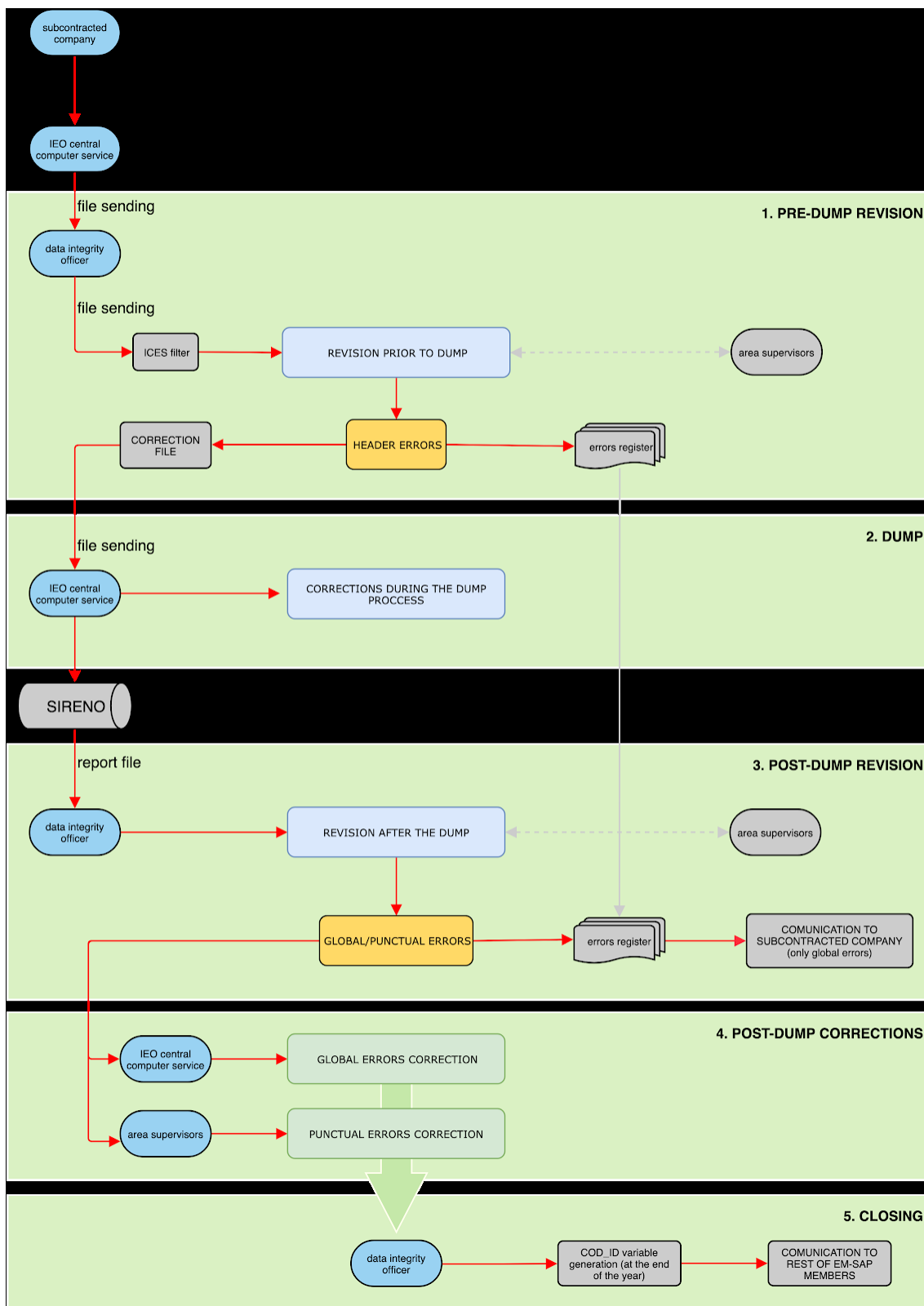


Figure 2. Monthly dump sampling revision in the national database SIRENO.

## 6. DATA PROCESSING

### 6.1. Evaluation of data accuracy

As it stated above, the selection process of on-shore sampling units includes the record of the number of fishing trips rejected for sampling. So far, this information is being used to detect changes in port logistics and access to sample target trips, but without developing bias analysis. Regarding precision analyses, the methodology used is developed in the following sections.

### 6.2. Editing and imputation methods

The objective of the first step of supervision is to avoid the existence of quarters and *métiers* without length frequency distributions (LFD) for landings. Nevertheless, when this happens, an imputation scheme must be applied to fill empty *métier*-quarter domains. However, this is done at the stage of processing of scientific data, once the sampling phase (from collection to validation of the fisheries sampling data) has concluded. The imputation procedure in the generation of InterCatch files for the assessment of fish stocks can be summarized in the next main points:

- Group of species: the landings declaration for species with a combined species total allowable catch (TAC), such as monkfish or megrims, must be disaggregated for scientific use. This is done by applying the species ratios from biological sampling. In the case of a *métier*-quarter domain without sampling, the ratio of species is imputed from the same closest *métier*-quarter domain sampled.
- Length frequency distributions: derived from the LFD of the same *métier*-quarter domain in the most recent year. Since the complete implementation of the InterCatch Data calls, this procedure is usually avoided, leaving it in the hands of the stock coordinator, who has all the international information available for the stock and can make a more informed decision.
- Pairing sampled-declared trips (**MATCHING**): crosschecking between the sampled trips and the official logbooks to:
  - To assign the ID logbook to the corresponding sampled trip.
  - Contrast and consolidate the information of the sampled trips: landing profile, fleet activity (*métier*), georeference (ICES Division/Rectangle), correspondence between landing event and fishing trip, obtaining specific trip variables not collected by the samplers (*e.g.* fishing days, fishing depth, etc), to obtain the sale location and the location where landings are accessible for sampling (*e.g.* relevant to assess the coverage of the Spanish fleet in foreign waters).
  - Comparative analysis of DPUEs: cross check between logbooks, sale notes and sampling data allows the detection of possible sampling defects in relation to the selection of trips, the *métiers* coverage, the concurrence of sampling (*e.g.* not access to certain species), or port accessibility (not access to certain species, categories or landing categories).

In relation to the trip matching methodology, the process is carried out by applying R scripts to the matrix of sampled trips and the official logbooks, and match is mainly done based on the following fields:

- Logbooks: Landing Date/Fishing date and vessel ID code.
- Sampled trips: Sampling Date and vessel ID code.

### 6.3.Validation of the final dataset

Finally, a quality check is carried out prior to its final **VALIDATION**. The quality control of the length frequency distributions is carried out using the **Cook's distance method**. The procedure consists of the simultaneous exploration of all the length distributions by species and/or *métier*. The Cook's distance is a measure calculated with respect to a given regression model and is therefore only affected by the X variables included in the model. Example:

```
>> mod <- lm (mean_length_trip ~ stock, data = "OTB_DEF_70-99_0_0")
```

```
>> mod <- lm (mean_length_trip~metier, data="HKE")
```

With this command a regression model with the mean length of each trip/stock for a given *métier* is calculated. Mathematically, Cook's distance for observation  $i$  is calculated as:

$$D_i = \frac{\sum_{j=1}^n \left( \hat{Y}_j - \hat{Y}_{j(i)} \right)^2}{p \times MSE}$$

where

- $\hat{Y}_j$  is the value of the  $j$  fitted response when all observations are included.
- $\hat{Y}_{j(i)}$  is the value of the  $j$  fitted response, when the fit does not include observation  $i$ .
- MSE is the mean square error.
- $p$  is the number of coefficients in the regression model.

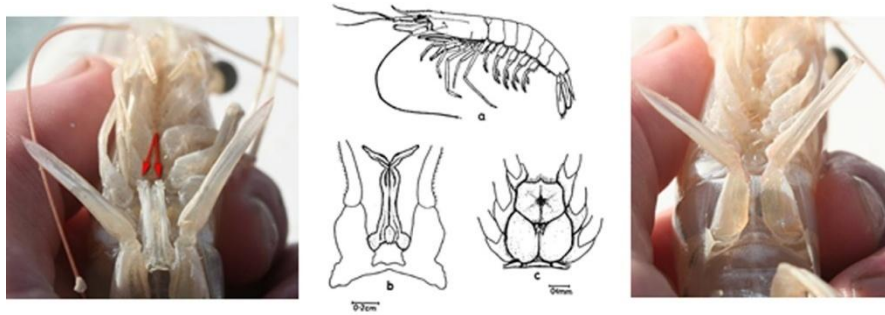
In general use, observations that have a Cook distance greater than 4 times the mean can be classified as influential. This is not a strict limit, since in fact a different value can be used for LFD based on the experience obtained with the analysis of the sampling of previous years. The variables used to detect outliers are ([ANNEX III](#)):

- Weighted mean length by trip/stock: the appearance of an atypical mean length involves looking for the cause of the atypical value and once found, making a decision.
- Raw length data: it allows us to identify any individual length value that is out of the ordinary and that has been masked in the calculation of weighted mean lengths. Once identified, it must be detected to see if it was a typing error or simply an extreme value, and assess whether it should be eliminated or maintained.

## ANNEX I



**Figure 3.** Measurement of length for fish, crustaceans and cephalopods.



**Figure 4.** Sex identification of crustaceans: male (left), female (right).




**Figure 5.** Sex identification of elasmobranchs: male (left), female (right).

## ANNEX II

[illegible]

**Figure 6.** Sampling sheet for landings.



		<b>MUESTREO CONCURRENTE</b> <b>TALLAS</b>		mm <input type="checkbox"/> cm <input type="checkbox"/>	Fecha muestreo:	Hoja ____ de ____
Muestreador/a:				Barco:		
esp.		esp.		esp.		esp.
cat.		cat.		cat.		cat.
rango		rango		rango		rango
	ev: <input type="checkbox"/>		ev: <input type="checkbox"/>		ev: <input type="checkbox"/>	
tallas	0	tallas	0	tallas	0	tallas
ejemplares	1	ejemplares	1	ejemplares	1	ejemplares
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8		8		8	
	9		9		9	
	0		0		0	
	1		1		1	
	2		2		2	
	3		3		3	
	4		4		4	
	5		5		5	
	6		6		6	
	7		7		7	
	8					

## ANNEX III

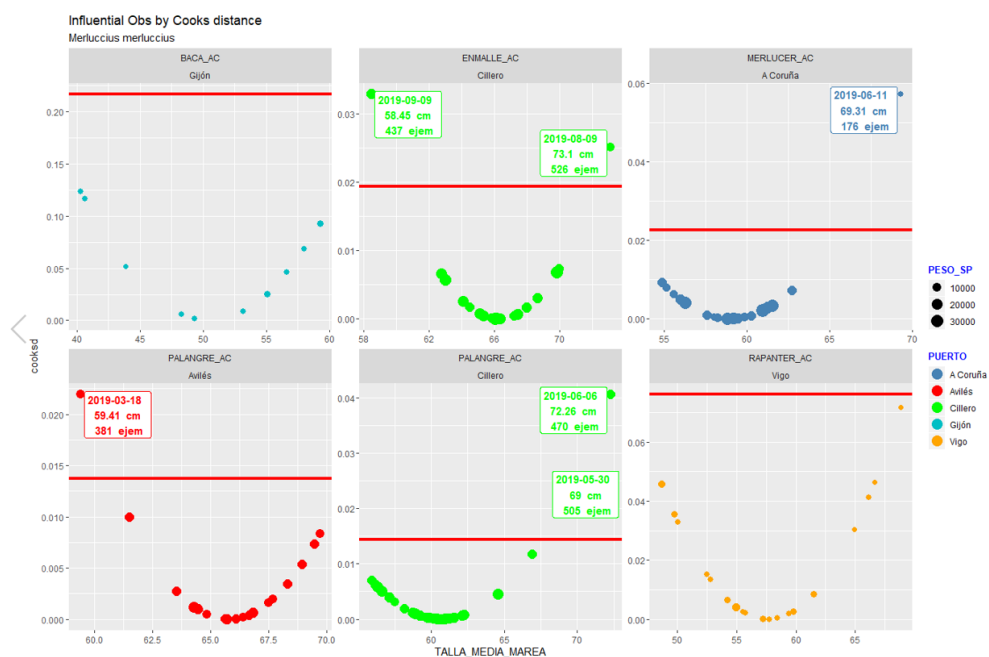


Figure 8. Cook's distance plot to detect outliers (e.g. length frequency distributions).