



Documentation of the

**AT-SEA 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, I. Salinas,

J. Acosta and J. Cebrian

## Table of Contents

Table of Contents .....	2
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 .....	11
6.1. Evaluation of data accuracy .....	11
6.2. Editing and imputation methods .....	11
6.3. Validation of the final dataset.....	12
ANNEX I .....	13
ANNEX II .....	15
ANNEX III .....	20

## FIGURES

Figure 1. Flow diagram of the vessel/trip selection protocol for at-sea sampling. ....	7
Figure 2. Measurement of length for fish, crustaceans and cephalopods. ....	13
Figure 3. Sex identification of crustaceans: male (left), female (right).....	14
Figure 4. Sex identification of elasmobranchs: male (left), female (right). ....	14
Figure 5. Sampling sheet for fishing trip general data. ....	15
Figure 6. Sampling sheet for catches of bottom otter trawl and pair trawl métiers. ....	16
Figure 7. Sampling sheet for catches of set gillnets métiers.....	17
Figure 8. Sampling sheet for catches of the purse seine fleet. ....	18
Figure 9. Sampling sheet for length frequency distributions.....	19
Figure 10. Cook's distance plots to detect outliers ( <i>e.g.</i> applied for length frequency distributions). ....	20

## TABLES

Table 1. Strata of the IEO at-sea sampling program of the Spanish (non-Basque) fleet of European Atlantic waters.....	5
---	---

## 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 at-sea 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 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)).

---

1 English translation: Monitoring and Analysis of Spanish Fisheries in the Northeast Atlantic.

2 Spanish Institute of Oceanography.

## 1. DESCRIPTION OF THE POPULATION

The **IEO at-sea 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 length samples from catches and by-catch at sea for all species listed in Table 1 of the EU-MAP Delegated Decision annex. At the same time, information on bycatch of protected, endangered, and threatened species (PETS) is also collected, as well as marine litter data. The sampling scheme covers the *métiers* more susceptible to producing discards, due to the use of less selective fishing gears.

The target population consists of **fishing trips** from trawlers and gill netters operating in the Atlantic Spanish fishing grounds, as well as the purse seiners of the Gulf of Cadiz and the Spanish bottom otter trawl targeting megrims in European non-Iberian Atlantic waters (mainly, ICES Subarea 7). By operability, the sampling population has been stratified in **5 strata** according to the official lists of licensed vessels:

Nº	Strata	Fishing ground	Nº of sampled trips	Fishing gear
1	IEO_P1_S_CN_GNS	Cantabrian-Northwest	32	Set gillnets ("volanta" and "rasco")
2	IEO_P1_S_CN_TB	Cantabrian-Northwest	104	Bottom otter and pair trawl
3	IEO_P1_S_GC_OTB	Gulf of Cadiz	108	Bottom otter trawl
4	IEO_P1_S_GC_PS	Gulf of Cadiz	44	Purse seine
5	IEO_P1_S_S7_OTB	ICES Subarea 7	8	Bottom otter trawl (targeting megrims)

**Table 1.** Strata of the IEO at-sea sampling program of the Spanish (non-Basque) fleet of European Atlantic waters.

## 2. SAMPLING PROTOCOL

In relation to the procedure for selecting sampling units, the vessel represents the Primary Sampling Unit (PSU), which is randomly selected from official lists of boats with fishing license, taking for sampling the fishing trip immediately following the phone call. For strata 1 to 4, the PSU selection is done by Simple Random Sampling With Replacement (**SRSWR**).

However, the vessel-trip selection for stratum 5 ("IEO\_P1\_S\_S7\_OTB"), the only one operating in non-Spanish waters, is carried out directly by the fishery association for operational reasons. Therefore, in this case the sampling design must be defined as Non-Probabilistic Quasi Simple Random Sampling Without Replacement (**NPQSRSWOR**).

The entire target population is included in the sampling frame, with the exception of the small-scale gillnetters and trammel nets without habitability for observers. So that the sampling of gillnets is focused on those targeting hake (locally called "*volanta*") and white anglerfish ("*rasco*"). Meanwhile, all vessels of the purse seine and trawl *métiers* are susceptible to be sampled.

Considering the vessel-trip combination as PSU, the secondary sampling unit (SSU) is the fishing operation (haul). All of them are sampled in the *métiers* of the national fishing ground with daily trips (*métiers* 1 to 4). However, *métier* 5, whose trips last 12 days on average, require alternating hauls for sampling.

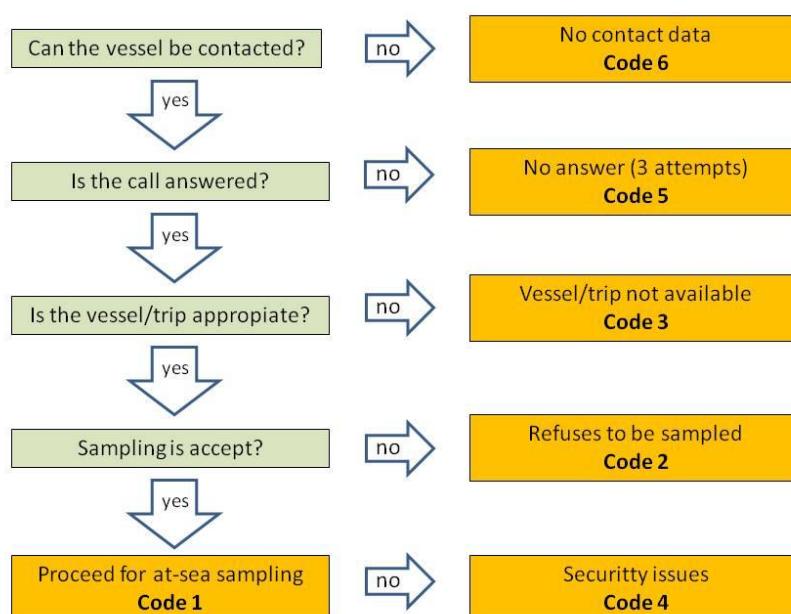
All catch categories available on board, as well as the incidental catches of sensitive species are considered in the sampling scheme.

### 3. SAMPLING IMPLEMENTATION

The call protocol for the vessel-trip selection of the SRSWR at-sea sampling scheme includes the **recording of responses**, which have been classified into the following **6 categories**:

1. Affirmative: sampled trip.
2. Hard refusal: skipper declines collaboration.
3. Soft refusal: temporary unavailability of the vessel/trip (repair, temporary lack of space, seasonally dedicated to other fishing activity...).
4. Observer refusal (security reasons, etc.).
5. No answer: unable to contact.
6. No contact details.

The recording of the responses is made according to the flow diagram shown in Figure 1.



**Figure 1.** Flow diagram of the vessel/trip selection protocol for at-sea sampling.

As explained above, the vessel selection for stratum 5 is carried out directly by the fishery association, so the recording of responses does not apply.

The general coordination of the at-sea sampling program requires the monitoring of the sampling progress throughout the year: “**SUPERVISION**”. For this, the monthly monitoring of the sampling coverage is carried out, adapting the sampling intensity when there are variations in fishing activity, so as to guarantee the quarterly robustness of the data.

## 4. DATA CAPTURE

In each sampled haul, all catch categories are considered for biological sampling: retained catch, discards, and BMS (Below Minimum Size) fraction. The data to be collected are the **taxonomic identification** of the species and the **length** the individuals, as well as **sex** in the case of elasmobranches and Norway lobster.

The length of the individuals 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) ([ANNEX I](#)).

As soon as the gear is emptied on board, the crew proceed to the sort process (fish selection and catch categorization). Within each catch category, the observer collects a random sample in a container of known weight, so that the weight of the subsample can be inferred.

Secondarily, information on **incidental catches of sensitive species** is also collected, as well as marine litter data. When an incidental catch occurs, whether the individuals are brought on board and then discarded or immediately released, the record is noted, as well as the taxonomic identification and the length of the specimens if possible. If the captured animal is alive at the time of being introduced on board, first of all it is tried to free them from the tackle or net cloths with the utmost care, trying not to damage them. They will be kept on board for the minimum precise time to be able to identify the species, take the total length data and if possible the sex of the animal. When returning them to the sea, with the collaboration of the crew, precautions will be taken not to cause further damage. The complete biometry will be done, as long as the survival of the specimen is not endangered.

In addition to the information derived from the biological sampling, the observer must also take note of the **technical data** of the 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. In relation to the incidental catches of marine mammals, data on the use of acoustic deterrent devices (“pingers”) on board are also collected: presence/absence, number and model.

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

Subsequently, this information is **computerized**, either by the observer himself or by specialized staff with direct contact with the observers 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 at-sea 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.



**Map:** 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.

---

3 “Seguimiento Informático de los Recursos Naturales Oceánicos”

## 5.2. International database

On a supranational scale, the data collected by the IEO at-sea 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 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 is used to detect errors (objective mistake which must be fixed) and warnings (possible error). Errors and warnings are reviewed by the at-sea 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.

## 6. DATA PROCESSING

### 6.1. Evaluation of data accuracy

The evaluation of data accuracy covers both analyses: bias and precision. In relation to the first, the registration of responses in the selection process of vessels provides a matrix of sampled and rejected trips from their official logbooks. The availability of these matrices allows the application of statistical tests in order to estimate possible sources of bias, evaluating whether the sampled trips have a fishing behaviour that substantially deviate from the non-observed trips (Rodríguez *et al.*, 2018<sup>4</sup>). The methodology used for precision analysis is developed in the following sections.

### 6.2. Editing and imputation methods

The objective of the supervision stage of the quality control process is to avoid the existence of quarters without length frequency distributions (LFD) for landings. Nevertheless, when this happens, an imputation scheme could be applied to fill empty quarter-métier 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 focuses in the following main points:

- Estimation of discards: derived from the estimated volume of discards (tons) of the same quarter-métier domain in the most recent year. However, this procedure has been deprecated since the implementation of the Landing Obligation regulation. Currently, each vessel individually adapts its fishing strategy throughout the year according to its quota consumption; therefore, the imputations derived from previous months/years can cause large deviations from reality.
- Length frequency distributions: derived from the LFD of the same quarter-métier domain in the most recent year. In the same way, it is no longer necessary since the imputation of the volume of discards no longer applies.
- Pairing sampled-declared trips: crosschecking between the sampled trips and the official logbooks to assign a common identifier and confirm the assigned métier.
- Pairing rejected-declared trips: the objective is to match the rejected sampling units with the total population of trips. When the match is not direct, the most likely trip must be assigned (closest landing day of the selected vessel).

In relation to the trip **MATCHING** methodology, the process is carried out by applying R scripts to both datasets, *i.e.* the matrix of sampled/unsampled 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/unsampled trips: Sampling Date and vessel ID code.

---

<sup>4</sup> Rodríguez, J.; Castro, J; Salinas, I.; Araujo, H and Marin, M. 2018. Improving protocol of selection of vessels to reduce bias. 9th International Fisheries Observer and Monitoring Conference (IFOMC). Vigo (Spain), 11-15 June, 2018.

### 6.3. Validation of the final dataset

Finally, a quality check is carried out prior to the final data **VALIDATION**. The quality control of the discard estimates and their respective length 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* to detect outliers. 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 = "IEO_P1_S_S7_OTB")
```

```
>> mod <- lm (mean_length_trip~métier, 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:

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

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.

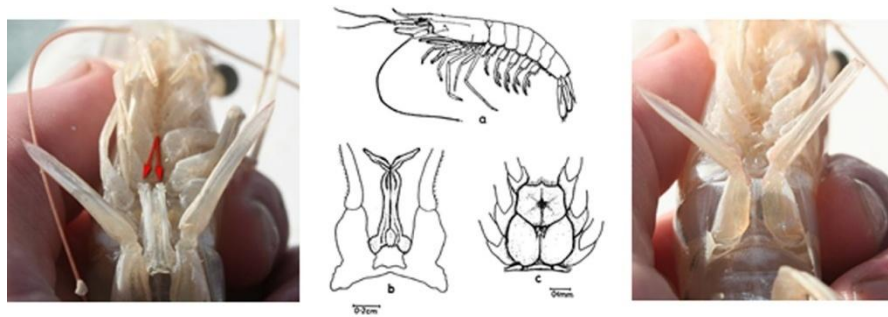
In general use, observations that have a Cook distance greater than 4 times the mean can be classified as influential. This is not a strict limit, since in fact a different value can be used for "lengths" and "discards", based on the experience obtained with the analysis of the sampling of previous years. 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.
- Volume of discard by trip/stock: this allows detecting outliers (very high or very low) that could cause a very large distortion in the DPUE calculation. As for the previous parameters, once detected, the cause must be identified to decide if it should be removed or not.

## ANNEX I



**Figure 2.** Measurement of length for fish, crustaceans and cephalopods.




**Figure 3.** Sex identification of crustaceans: male (left), female (right).



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

## ANNEX II

 <b>ESTADILLO DE MAREA</b>	
METIER:	
<b>AÑO:</b>  <b>MAREA:</b>  <b>OBSERVADOR:</b>  <b>PUERTO BASE BARCO:</b>  <b>PUERTO EMBARQUE OBSERVADOR:</b>  <b>PUERTO DESEMBARQUE:</b>  <b>PUERTO VENTA PESCADO:</b>	<b>FECHA EMBARQUE:</b>  <b>FECHA DE DESEMBARQUE:</b>  <b>DIAS DE EMBARQUE:</b>  <b>DIAS DE PESCA:</b>  <b>Nº DE LANCES:</b>  <b>LANCES MUESTREADOS:</b>  <b>LANCES NULOS:</b>
<b>CARACTERISTICAS BARCO</b>	
<b>NOMBRE BARCO:</b>  <b>DISTINTIVO DE LLAMADA:</b>  <b>CABALLOS DE VAPOR:</b>  <b>TONELADAS DE REGISTRO BRUTO:</b>  <b>ESLORA</b>	<b>Nº TRIPULANTES TOTALES:</b>    <b>Nº TRIPULANTES EN CUBIERTA:</b>    <b>CAPACIDAD BODEGA</b>
<b>ARMADOR</b>	
<b>NOMBRE ARMADOR:</b>  <b>DIRECCIÓN:</b>  <b>NOMBRE PATRON:</b>  <b>DIRECCION:</b>	  <b>TELEFONO:</b>    <b>TELEFONO:</b>
<b>OBSERVACIONES:</b>	

**Figure 5.** Sampling sheet for fishing trip general data.








[illegible]

## DISTRIBUCIÓN DE TALLAS



*Tipo Muestreo*

D= Descartes  
R= RETENIDA  
BMS= BELOW MINIMUM SIZE

MAREA

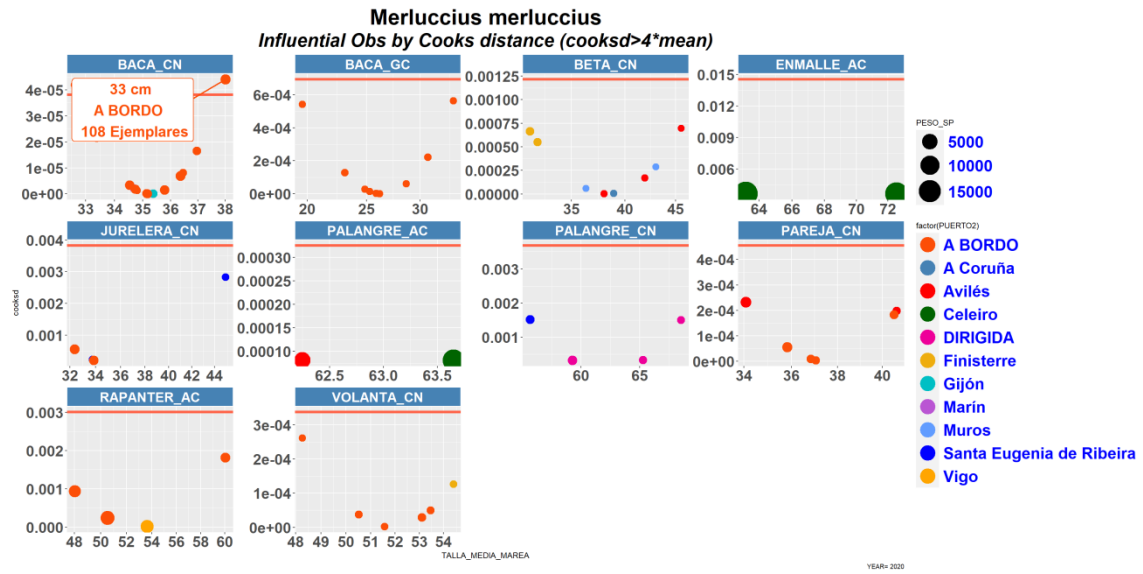
LANCE

FECHA

Especie									
Código									
Categoría									
Tipo Descartes									
Sexo									
Peso muestreado									
Talla inicial									
Talla final									
0			0			0			0
1			1			1			1
2			2			2			2
3			3			3			3
4			4			4			4
5			5			5			5
6			6			6			6
7			7			7			7
8			8			8			8
9			9			9			9
0			0			0			0
1			1			1			1
2			2			2			2
3			3			3			3
4			4			4			4
5			5			5			5
6			6			6			6
7			7			7			7
8			8			8			8
9			9			9			9
0			0			0			0
1			1			1			1
2			2			2			2
3			3			3			3
4			4			4			4
5			5			5			5
6			6			6			6
7			7			7			7
8			8			8			8
9			9			9			9
0			0			0			0
1			1			1			1
2			2			2			2
3			3			3			3
4			4			4			4
5			5			5			5
6			6			6			6
7			7			7			7
8			8			8			8
9			9			9			9
0			0			0			0
1			1			1			1
2			2			2			2
3			3			3			3
4			4			4			4
5			5			5			5
6			6			6			6
7			7			7			7
8			8			8			8
9			9			9			9

**Figure 9.** Sampling sheet for length frequency distributions.

## ANNEX III



**Figure 10.** Cook's distance plots to detect outliers (e.g. applied for length frequency distributions).