FRAMEWORK CONTRACT – MARE/2016/22 “Strengthening regional cooperation in the area of fisheries data collection” Annex III “Biological data collection for fisheries on highly migratory species”

Project acronym: RECOLAPE

WP1 - Large Pelagic Regional Coordination group structure

Participating partners:

D.1.1 - Proposal on the organizational structure for the future RCG-LP

Deliverable coordinated by Jon Ruiz (AZTI)

Index

Executive summary 3
1. Background 5
2. Objective 6
3. Organization of the future RCG LP’s groups and subgroups 7
4. Synergies with other RCG 8
   4.1 Regional sampling plan Pan-regional subgroup 16
   4.2 Data management 16

Executive summary

This deliverable is part of the Work Package 1 within the project MARE/2016/22: "Strengthening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species" (acronym RECOLAPE). It proposes a series of prerequisites necessary for the correct establishment and operation of the RCG-LP. It is expected that this proposal will foster the implementation of large pelagic RCG, so that RCG can become fully operational, and not get entangled into long discussions that can take several meetings/ years.

This proposal considers the holding of two preparatory meetings or stages prior to the RCG-LP main meeting (figure 1). Far too many meetings (subgroups) can hinder the participation of LP experts, thus preparatory meetings are not new meetings at all, the aim is to integrate existing ones under the umbrella of the future RCG-LP. The first stage would have the purpose to identify data gaps and prioritize LP data needs, including tuna RFMO data requirements and data transmission failures. Second stage will be much more gear/stock specific. This second stage would design RSP (Regional Sampling Plans) by coordinating dockside/onboard sampling for the different stocks. Finally, the main RCG meeting (third stage) will evaluate the outputs from the previous two stages, and it would be where the final decisions of greater importance and approval of regional sampling plan should be made.

With respect to the first stage, identification of data needs, DG MARE is already organizing a meeting with a similar goal, where DG MARE desk officers dealing with RFMOs and EU scientists participating in RFMO activities are involved. Thus, considering this meeting as part of the RCG-LP could be advantageous, instead of organizing a new one. Once data gaps are identified and data needs/priorities are stablished, in a second stage, data collection shall be coordinated among MS. Ideally this coordination should be done in methodological groups dealing with specific fisheries. The proposal includes four parallel subgroups based on stocks/gears.

a) Tropical Tuna Treatment (T3): Focus on PS fleet and YFT/SKJ/BET
b) Focus on Longline Fisheries outside MED
c) MED LP fisheries: Focus on LL.
d) Bluefin tuna sampling

In addition to the two stages mentioned above, it is necessary to hold a main RCG-LP meeting, where decisions are made based on the output from the previous stages (RSP approval, etc.). It is not necessary, in this case, such a technical profile of the participants, but they must have the capacity to make decisions.

On the other hand, synergies among RCGs, would be necessary and beneficial. The main common fields identified for cooperation among the RCGs are the ones related with the design of the regional sampling plans based on statistically sound sampling designs and data management, which includes the archiving of data and the processes of quality assurance and quality control (i.e. development of guidelines to evaluate the quality of the data, development of common software tools in R etc.).

1. Background

As provided for in Article 25 of Regulation (EU) No 1380/2013[1] and in the article 9 of the Recast European Regulation 1004/2017[2], Member States (MS) shall coordinate their data collection activities with other MSs in the same marine region. In order to facilitate this regional coordination, RCGs (Regional Coordination Groups) shall be established, with the aim at developing and implementing procedures and methods for collecting and processing data. RCGs shall draw up and agree on rules of procedures and its own organization. In this context, this document makes a proposal for the future structure of the RCG-LP (Large Pelagic) and propose optimal engage of this group among the various defined RCGs/regions.

Data collection on large pelagic fisheries outside the Mediterranean Sea was initially under the scope of RCM-LDF (Regional Coordination Meeting-Long Distance Fisheries), while Mediterranean LP fisheries were under the scope of the RCM Med&BS. However, as decided by Liaison Meeting in 2013[3], a coordination group for LP covering areas of competence of RCM LDF, North Atlantic (NA), Mediterranean and Black Sea (Med&BS) and dealing with all large pelagic species and fisheries was created. This group has been initially associated with RCM-MED&BS in order to limit the number of meetings and allow Mediterranean experts on LP fisheries and stocks assessment to participate in RCM-LP subgroup while also participating in RCM-MED&BS. Between 2014 and 2017 the RCM MED&BS-LP was therefore a joint RCM with two co-chairs, one for MED&BS and one for LP.

However, as it is stated in the RCM MED&BS-LP 2016 report, from the RCM-LP subgroup perspective, it was particularly unfortunate that the annual meeting of this group took place always in September very close to unavoidable ICCAT (International Commission for the Conservation of Atlantic Tunas) scientific activities, resulting in the absence to the meeting of some LP data end user (ICCAT), many EU scientist participating in tuna RFMOs (Regional Fisheries Management Organizations) and Commission Unit dealing with RFMOs. RCM-LP believed that these absences should be avoided, as far as possible, in order to ensure that data needs from RFMOs are reflected in the National Programmes for data collection. In addition, from the RCM-LP subgroup

[1] REGULATION (EU) No 1380/2013 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 11 December 2013 on the Common Fisheries Policy, amending Council Regulations (EC) No 954/2003 and (EC) No 1224/2009 and repealing Council Regulations (EC) No 2371/2002 and (EC) No 639/2004 and Council Decision 2004/585/EC.

[2] REGULATION (EU) 2017/1004 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 May 2017 on the establishment of a Union framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the common fisheries policy and repealing Council Regulation (EC) No 199/2008 (Recast)

[3] Report of the 10th Liaison Meeting. Meeting between the Chairs of the RCMs, the chair of ICES PGCCDBS, the chair of PGMED, the chair of the Regional Database Steering Committee, the ICES representative, the Chairs of STECF EWG's DC-MAP and PGECON and the European Commission

perspective, common topics between the two subgroups, RCM Med&BS and RCM-LP, were very limited. Thus, considering the specificities of these fisheries it seemed logic to have a specific thematic group (Large Pelagic Group). In 2017, LP and Med&BS experts had separated meetings, even if both subgroups where still members of the same RCG. At that time, the group agreed that the RCG-LP should be an RCG independent from Med&BS, and ideally should hold its meetings in the second quarter every year in order to avoid periods of heavy workload. This idea was presented, as an RCM-LP's recommendation, during the 14th Liaison Meeting[4]. The new group will not be regional, but global, based in highly migratory species more than in a restricted geographic region; as the Regulation Recast 1004/2017 refers to coordination of activities in marine regions only (no reference to species or groups of species), RCG-LP set its region as 'all regions'. Thus, during 2018 the RCG-LP acted as an independent group, where all MSs involved on LP fisheries where invited: CYP, ESP, FRA, GBR, GRC, HRV, IRL, ITA, MLT and PRT.

In this context, this document proposes a series of prerequisites necessary for the correct establishment and operation of the RCG-LP. It is expected that this proposal will foster the implementation of large pelagic RCG, so that RCG can become fully operational, and not get entangled into long discussions that can take several meetings/ years. Additionally, in the last section, synergies with other RCMs/RCGs are explored, trying to identify points for possible Pan-regional (inter-RCG) cooperation.

2. Objective

This deliverable is part of the Work Package 1 within the project MARE/2016/22: "Strengthening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species" (acronym RECOLAPE). The overall objective of the project is to strength the regional cooperation in the area of biological data collection for fisheries on large pelagic fish. This purpose will be valuable to improve the coordination among European MS in the fisheries data collection field in support of stock assessment and fisheries advice. At the same time, this seeks to provide solutions to certain needs in terms of data collection identified by scientists involved in the stock assessment of tuna RFMOs and by expert groups like the RCG-LP.

To reach this purpose, the RECOLAPE project addresses several objectives:

- Facilitate the evolution of the RCM-LP towards the RCG-LP: the goal is to evolve from a single meeting to a continuous process that will have greater responsibilities in support of stock assessment and fisheries advice.

[4] Report of the 14th Liaison Meeting. Meeting between the chairs of RCGs (and sub-group on Large Pelagics), PGECON, STECF meetings on DCF, SCRDB, key end users (ICES, RFMOs), JRC and the Commission.

- Design a RSP (Regional Sampling Plan) for large pelagic stocks: facilitating the transition from individual national work plans towards regional ones.

- Develop data collection strategy and tools regarding additional data (not yet collected on a routine basis) on FADs. Such additional data could be used in combination with traditional CPUE or for building alternative abundance indices.

- Test alternative data collection methods for those cases where traditional methods present data deficiencies, for example for data collected using Electronic Monitoring System (EMS).

- Facilitate cooperation among MS in order to improve and develop common data quality assessment procedures at national and regional levels.

- Identify points of consensus and/or disagreement that may arise during the coordination process among organizations dealing with large pelagic fisheries data collection. The idea is to identify a framework of rules and feedback to improve future coordination or expand it on other fisheries/species.

The present deliverable, D.1.1, deals with the future organization of the RCG-LP and proposes a future structure that allows this group to be fully operational. Additionally, in a second step, synergies of this group among the various defined RCGs/regions are explored, including mechanisms for coordination between them.

A first draft of this deliverable (D.1.1-*Proposal on the organizational structure for the future RCG-LP*) was presented during the annual meeting of the RCG-LP 2018, held in June in Heraklion (Greece). This document already considers all the comments made by the group.

3. Organization of the future RCG-LP's groups and subgroups

There is currently a draft document where the RCG-LP Rules of Procedure (ROP) are established. This document is not yet officially adopted, although it has already been reviewed by a large part of the MS involved in LP fisheries. Therefore, this section will not focus on these ROPs, but will make an operational proposal for the future RCG-LP:

The overall objective of the RCG-LP is to strengthen the regional cooperation in the area of biological data collection for fisheries on highly migratory species. In the current context where regional cooperation will evolve from a single meeting (RCM) to a continuous process that will have greater responsibilities (RCG), it is foreseen that this proposal will be valuable to improve the coordination among EU Member States in the fisheries data collection field in support of data transmission to tuna RFMOs, stock assessment and fisheries advice.

This proposal considers the holding of two preparatory meetings or stages prior to the RCG-LP main meeting (figure 1). Far too many meetings (subgroups) can hinder the participation of LP experts, thus preparatory meetings are not new meetings at all, the aim is to integrate existing ones under the umbrella of the future RCG-LP. The first stage would have the purpose to identify data gaps and prioritize LP data needs, including tuna RFMO data requirements and data transmission. Second stage will be much more gear/stock specific. This second stage would design RSP by coordinating dockside/onboard sampling for the different stocks. Finally, the main RCG meeting (third stage) will evaluate the outputs from the previous two stages, and it would be where the final decisions of greater importance and approval of regional sampling plan should be made.

With respect to the first stage, identification of data needs, DG MARE is already organizing a meeting with a similar goal, where DG MARE desk officers dealing with RFMOs and EU scientists participating in RFMO activities are involved. Thus, considering this meeting as part of the RCG-LP could be advantageous, instead of organizing a new one. The aim of this meeting will be to identify and prioritize LP data needs and to improve communications between EU scientist involved in data collection, scientist involved in stock assessment, DG MARE and the rest of the end user (tuna RFMOs).

Once data gaps are identified and data needs/priorities are stablished, in a second stage, data collection shall be coordinated among MS. Ideally this coordination should be done in methodological groups dealing with specific fisheries. In this regard, France and Spain coordinate, since many years ago, their tropical tuna purse seine fisheries in an annual coordination meeting with participation of scientists from both countries through a data analysis Working Group (Tropical Tuna Treatment, T3). During this meeting, sampling methodological issues are discussed; tools and sampling protocols are shared, discussed and eventually revised. Regional sampling coordination and possible bilateral agreements are also discussed. Scientists from non-EU landing countries like Seychelles, Madagascar, Côte d'Ivoire, Senegal and Ghana who participate in the data collection are also invited. Specific or common scientific contributions, as well as data calls, to tuna RFMOs (ICCAT/IOTC) are jointly elaborated.

The existing trend in coordination for the tropical tuna purse seine fisheries can be understood as an example towards the regional sampling scheme. Regarding the rest of the species and fleets, such regional coordination does not currently exist. Thus, expanding the scope, and organizing similar technical meetings to coordinate data collection on other LP stocks, in which matters such as data acquisition, sampling methodology, and sampling coordination issues are discussed, could be beneficial. In this regard, RCG-LP 2017 has already made, during the 14th Liaison Meeting, a recommendation to hold a workshop for exploring the possibility to launch a permanent group for longline fleets outside Mediterranean waters as a complement of the existing tropical tuna T3 group. In this scenario, where the longline and purse seine fisheries outside Mediterranean waters would have specific regional data collection groups, only the fleets operating in the Mediterranean, and fleets targeting temperate tunas (bluefin tuna and albacore) in the North Atlantic would remain out from the sampling coordination network. Thus, it makes sense the creation of such specific groups with the aim of coordinating sampling of these fisheries; Mediterranean LP fisheries (focused in longline as it is the main gear), and a coordination group focused on bluefin tuna sampling (the unique stock which includes both Mediterranean and North Atlantic). Regarding north albacore, it is true that it is outside this coordination framework. However, main countries contributing to albacore landings in the north Atlantic, Spain and France, do not share same gear/metiers. Thus, there is not much room for dockside sampling coordination. Even if coordination of other biological sampling (e.g. maturity, ages) could be beneficial, and specific data analysis subgroup has not been considered necessary.

Figure 1 shows the diagram of the possible meeting for the future RCG-LP including three stages.

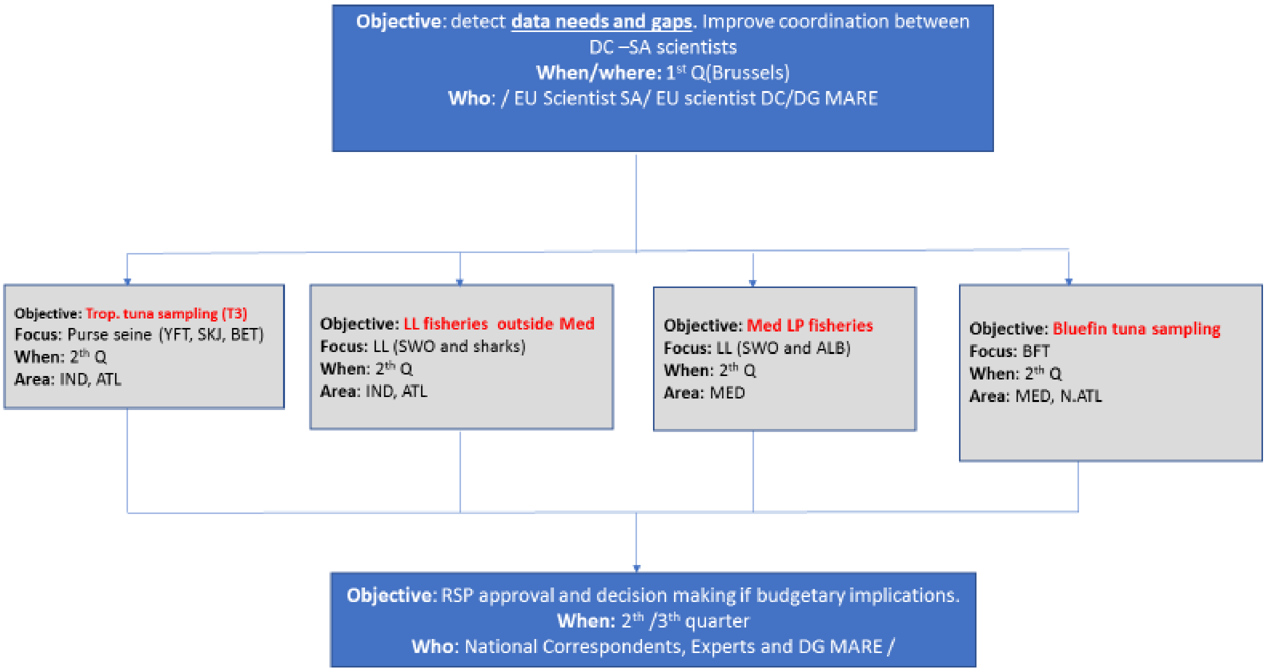


Figure1. Proposal for RCG-LP subgroups.

Bellow, details of each of these subgroups are provided ; identifying the main MS involved and the expertise that is needed.

3.1. Subgroup on data need and data gaps (stage 1):

The specific objective of this subgroup is to address issues of common interest for tuna scientist regarding data collections issues. Identify research priorities in terms of data collection, based on data gaps and data needs identified by the end users (stock assessment groups within the tuna RFMOs), and finally to improve coordination between data collection scientists and stock assessment scientists.

This same group would also be responsible of identifying needs outside the commercial fisheries sampling, such as proposing scientific surveys on highly migratory species, or the need to include new recreational fisheries in addition to those already included in table 3 under Commission Implementing Decision (EU) 2016/1251[5].

Since 2011, DG MARE has been organizing, early in the year, these types of meetings, with the overarching objective of improving the coordination between DG MARE and UE scientists working in different RFMOs. This is also a good opportunity for scientists and DG MARE Officers to exchange views on the achievements experienced at RFMOs level during the previous years and provide response to the priorities of the year in these fora. In case this existing meeting is used as part of the RCG-LP instead of creating a new group, the participation should not be by DG MARE invitation, and should be open, where each MS decides which experts participate. Thus, the participation will be based on MS initiative following a reminder/invitation to the national correspondents by the RCGLP Chair.

Participants and expertise needed: Scientist participating in stock assessment and tuna RFMO species groups + scientists participating in data collection, including RCG meeting, DCF EWG and RFMO subgroups in data collection and statistics+ DG MARE.

Dates: First quarter every year.

Goal: Identify gaps and priorities in terms of data collection. Improve communication among scientists involved in data collection and stock assessment.

3.2. Subgroup on data analysis and Regional Sampling Design (stage 2):

The aim of this group should be the standardization and coordination of the sampling at fleet level, through the sharing of raw samples (e.g. length frequencies) collected by fleets, improving the estimates of specific annual catches and spatialized specific catches as indicated by the RFMOs. Thus, it is expected the participation of the institutes involved in the monitoring of these specific fleets, who should share and agreed on tools for data acquisition (including Data Base), data collection protocols (including codes), and data quality checks. Additionally, these groups could prepare tuna RFMO participation including data analysis and data calls. These goals have been set based on the tropical tuna T3 group experience and successful trajectory in data collection coordination. Number of parallel subgroups in this stage2 and stocks/gears allocation among subgroups has been based on the RCG – LP 2018 participants feedback. However, they should be dynamic and flexible groups, which may vary in the future, and where the group itself must decide

[5] COMMISSION IMPLEMENTING DECISION (EU) 2016/1251 of 12 July 2016 adopting a multiannual Union programme for the collection, management and use of data in the fisheries and aquaculture sectors for the period 2017-2019

on its operational aspects (e.g. inclusion of specific TORs (terms of reference), how often to meet or election of the Chair-person).

The proposal includes four parallel groups based on stocks/gears.

a) Tropical Tuna Treatment (T3): Focus on PS fleet and YFT/SKJ/BET
b) Focus on Longline Fisheries outside Mediterranean waters
c) Mediterranean LP fisheries: Focus on longline.
d) Bluefin tuna sampling

a) Tropical Tuna Treatment (T3)

French and Spanish fisheries research institutes (IRD and IEO) have developed some procedures to establish compatible databases aiming at compiling dockside sampling, observer and fishing dependent information. In particular, these research institutes have developed specific procedures to consolidate catches in tropical tuna purse seine fisheries by cross-checking information extracted from log-books (catch estimations and landing notes), sales notes, observer reports and sampling programmes. Additionally, methodologies, IT tools, as well as consolidated data are even shared with partner fisheries research institutes of coastal states.

This meeting is organized yearly between tropical tuna purse seine fishing countries (France, Spain, Seychelles, and some years ago Ghana and Mauritius) to produce consolidated Task 1 and Task 2 data for tuna targeted species (skipjack, yellowfin tuna and bigeye tuna) as stipulated by ICCAT[6] and IOTC[7]. The meeting allows to share all the raw samples of length frequencies collected by fleets to improve the estimates of specific annual catches and spatialized specific catches at the 1°x 1° month resolution (As indicated by the RFMOs). Additionally, bycatch sampling conducted by onboard observers is also coordinated.

Participants and expertise needed: Scientific institutes (Spain and France) and invited purse seine fishing Coastal States. Within these institutes, experts involved in port sampling, observers' programs and stock assessment.

Dates: Second quarter.

Goal: Design sampling plan and produce consolidated Task 1 and Task 2 data, for tropical tuna purse seiner fisheries. In addition, coordinate bycatch sampling in tropical tuna purse seiner fisheries.

[6] ICCAT Rec [05-09] Recommendation by ICCAT on compliance with statistical reporting obligations.
[7] Resolution 15/02 Mandatory Statistical Reporting Requirements for IOTC Contracting Parties and Cooperating Non-contracting Parties (CPCs)

b) Coordination of the longline Fisheries outside Mediterranean waters

This subgroup would be responsible for coordinating longline fisheries sampling outside the Mediterranean waters in a way similar to the T3 group. Spain and Portugal (and France to a lesser extent) are the MS involved. Swordfish is the man target species, even if data collection on pelagic sharks and some other billfish should be also coordinated.

Participants and expertise needed: Scientific institutes (Spain, Portugal and France). Within these institutes, experts involved in observers' programs and stock assessment. *Dates:* Second quarter *Goal:* Design sampling plan and produce consolidated Task 1 and Task 2 data, for longline fisheries targeting swordfish outside Mediterranean waters. In addition, coordinate bycatch sampling for longline fisheries targeting swordfish outside Mediterranean.

c) Coordination of the Mediterranean LP fisheries data collection

This coordination group would focus in logline fleets operating in the Mediterranean, as it is the main gear used for exploiting large pelagic species in the area. In terms of stocks, coordination would consider swordfish and albacore, that are the main target species further to bluefin tuna. Both stocks are heavily exploited and MS are among the main producers for both species, with a 70-80% contribution to the total Mediterranean landings, in the recent years. According to the latest ICCAT records these stocks are ranked among the most poor ones in terms of data availability (Figure 2). The primary goal of the coordination group will be to promote regional cooperation and identify standardized sampling procedures that would improve data collection from the given fisheries. In practical terms, the subgroup will consider harmonized sampling approaches in line with those developed by the T3 group.

Table #	Fishery	SCORE	Score type
1	ALB-N stock	7.0	score3
2	ALB-S stock	5.5	score3
3	ALB-M stock	1.9	score3
4	BFT-E stock (ATE region)	5.7	score3
5	BFT-E stock (MED region)	3.2	score3
6	BFT-W stock	7.9	score3
7	BET-A stock (AT + MD)	6.7	score3
8	YFT-E region	6.7	score3
9	YFT-W region	4.3	score3
10	SKJ-E stock	7.0	score3
11	SKJ-W stock	4.7	score3
12	SWO-N stock	7.6	score3
13	SWO-S stock	6.5	score3
14	SWO-M stock	3.8	score3
15	BUM-A stock (AT + MD)	4.2	score2
16	WHM-A stock (AT + MD)	4.8	score2
17	SAI-E stock	2.9	score2
18	SAI-W stock	3.6	score2
19	SPF-E stock	2.4	score2
20	SPF-W stock	3.3	score2
21	BSH-N region	3.3	score2
22	BSH-S region	3.7	score2
23	BSH-M region	0.6	score2
24	POR-N region	1.3	score2
25	POR-S region	0.8	score2
26	SMA-N region	2.5	score2
27	SMA-S region	3.3	score2

Figure 2. ICCAT scorecard on fisheries data availability by stock based on information provided in ICCAT/SCRS 2018 plenary meeting (Document SCRS 08/2018)

Participants and expertise needed: Scientific institutes from MS exploiting swordfish and albacore stocks in the Mediterranean (practically all Mediterranean MS). Within these institutes, experts involved in sampling programs and stock assessment.

Dates: Second quarter

Goal: Design sampling plan and produce consolidated Task 1 and Task 2 data, for Mediterranean longline fisheries targetting swordfish and albacore. In addition, coordinate bycatch sampling for Mediterranean longline fisheries.

d) Coordination of bluefin tuna fisheries data collection

The bluefin tuna may be the most complex case; it is captured both in the Mediterranean and Atlantic, and it is exploited by a great number of gears and MS. Figure 3 shows catches in 2016 by gear and flag. Thus, it makes sense to have a specific group.

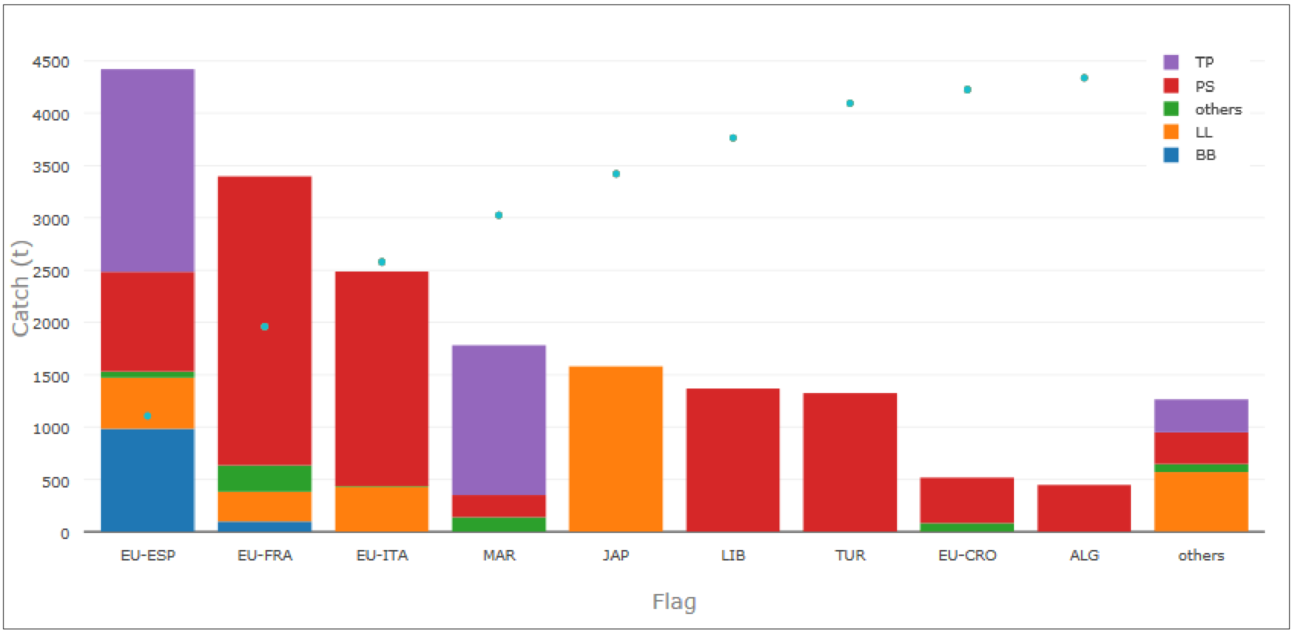


Figure 3. Bluefin tuna Task I reported values in 2016 by flag and gear.

Participants and expertise needed: Scientific institutes (Spain, France, Italy, Croatia, Malta and Portugal). Within these institutes, experts involved in port sampling, observers' programs and stock assessment. *Dates:* Second quarter *Goal:* Design sampling plan and produce consolidated Task 1 and Task 2 data, for bluefin tuna fisheries.

3.3. Main RCG-LP meeting (stage 3):

In addition to the two stages mentioned above, it is necessary to hold a main RCG-LP meeting, where decisions are made based on the output from the previous stages (RSP approval, etc.). It is not necessary, in this case, such a technical profile of the participants, but they must have the capacity to make decisions.

Participants and expertise needed: National Correspondent or designated representative + Scientist involved in DCF+ DG MARE.

Dates: Second quarter

Goal: Decision making (mainly if there are budget implications)

4. Synergies with other RCG

This section explores synergies with other RCGs where contractors have tried to identify which fields are of common interest among 'regions' (RCGs), and where mechanisms are proposed, if needed, for coordination between relevant RCGs. In general terms, these synergies among RCGs, would be necessary and beneficial.

The main common fields identified for cooperation among the RCGs are the ones related with the design of the regional sampling plans based on statistically sound sampling designs and data management, which includes the archiving of data and the processes of quality assurance and quality control (i.e. development of guidelines to evaluate the quality of the data, development of common software tools in R etc.).

It would not be specifically the design of the regional sampling plan for LP stocks (which as explained in the previous section, would be design and coordinated through specific gear/area related subgroups), but to be in contact with other regions to put in common tools to design a statistically sound sampling RSP, as well as tools to evaluate the quality of the data. DCF related groups convened by other end users as ICES (e. g. PGDATA[8]) could be a good example, where issues of common interest, and applicable to other areas/fisheries, are discussed. Ensuring appropriate dissemination and communication of the findings of these DCF related groups, would be beneficial.

This requires (Pan-regional) intersessional work among the different RCGs. This intersessional work will facilitate a cooperation on a supra regional level and develop the work that is needed to fulfil future coordination tasks in a broad sense.

For this purpose, the first stage would be the creation of specific Pan-regional subgroups for the common fields mentioned above (regional sampling plan subgroup and data management subgroup). Once the subgroups are created, it´s the responsibility of the RCG-LP to identify the experts with the needed skills by region that will work together to these issues.

---

[8] Planning Group of Data Needs for Assessments and Advice

4.1 Regional sampling plan Pan-regional subgroup

A regional sampling plan includes: agreed objectives based on end-users needs, an integrated regional sampling design, standardized sampling protocols, a common approach to quality assurance and regional tools for the management and dissemination of data.

The objectives based on 'end-users' needs will be case specific for each RCG and also the involvement and responsibility by MS. However, a core group of experts with the skills required under this Pan-regional subgroup would contribute in all questions regarding the sampling design and quality assurance issues.

To work efficiently, each RCG should be the responsible to identify which are the specific tasks to be covered by this Pan-regional subgroup and a responsible to lead these tasks. The assigned responsible will organize the meeting for this intersessional work (face to face meeting, skype etc.), the ToRs and finally provide the outputs to the RCGs. All this work requires important resources and it should be supported under the RCGs umbrella.

4.2 Data management Pan-regional subgroup

The overall objective of this Pan-regional subgroup should be to have a Quality Assurance Framework (QAF) for a regional sampling plan that covers the whole process from sampling design to data transmission with common and agreed quality checks, data validation and quality indicators.

In previous MARE/2014/19 projects (fishPi[9] and MED&BS[10] projects), under WP4 and intersessionally between experts from both grants, a list of quality checks, together with code details, to be used on national and regional data sets were produced. Some of these quality checks are using the most advanced science in outlier detection. The quality checks were built on a specific exchange format and a data structure was also proposed.

As for the RSP Pan-regional subgroup, it´s the responsibility of each of the RCGs to identify the experts with the needed skills by region that will work together to these issues.

The mechanism adopted for this intersessional work among the RCGs, would be the same that is mentioned for the other Pan-regional subgroups.

---

[9] The fishPi project – Regional co-operation in fisheries data collection

[10] Agreement Number – MARE/2014/19 -SI2.705484 - Strengthening regional cooperation in the area of fisheries data collection in the Mediterranean and Black Sea.

A key issue to work in these common fields is to have a regional database (RDB), with data stored in common formats, ensuring transparency and consistent standards for data processing and dissemination. This is the current situation for the 3 northern RCGs (RCG NA, RCG NS&EA and RCG Baltic). RCG LP is working to adopt the same formats and structures to be able to upload to integrate data in the RDB. With this goal in mind, it´s essential for RCG-LP to appoint a member representing Large Pelagic fisheries in the RDB Steering Committee.

In all this process of coordination and cooperation it´s essential that MS take the compromise to allocate the necessary human resources needed for the work identified under these Pan-regional subgroups. The responsibility to assure the appointment of these experts under these sub-groups must fall under the RCGs.

FRAMEWORK CONTRACT - MARE/2016/22 « Strengthening regional cooperation in the area of fisheries data collection » Annex III « Biological data collection for fisheries on highly migratory species »

Strengthening Regional cooperation in the area of large pelagic fisheries data collection (Acronym: RECOLAPE)

Task 2.1 – Definition of data needs and priorities

D.2.1 – List of data needs for the development of a RSP for the SWO-Med & TROP tunas

Responsible: Mathieu Depetris (IRD) & George Tserpes (HCMR)

Index

Executive summary 3
1. Introduction 4
2. Objectives 4
3. Case studies used for RSP's design 5
3.1. Mediterranean swordfish 5
3.2. Major tropical tunas in the Atlantic Ocean 5
4. Definition of data needs and priorities 7
References 13

Executive summary

This document is in relation to the work package 2 of the project MARE/2016/22: "Strengthening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species" (acronym RECOLAPE). The aim of this work package is to propose a Regional Sampling Plan (RSP) on large pelagic fish. In this context, this document answer to the first task of RSP's design: define data needs and prioritize them within the data collection scheme.

For this RSP's design, the focus is on two case studies: Mediterranean swordfish (*Xiphias gladius*) and major tropical tunas in the Atlantic Ocean (skipjack tuna, *Katsuwonus pelamis*; yellowfin tuna, *Thunnus albacares*; bigeye tuna, *Thunnus obesus*).

Regarding the major tropical tunas in the Atlantic Ocean, 8 groups of data are defined as priority:

- quantities of dead discards,
- quantities of bycatch released alive,
- dataset of catch at size estimations (ICCAT task 2[1]),
- data on support vessel activity,
- number of Fish Aggregating Devices (FADs) deployed by support vessels,
- information about maturity,
- information about age,
- data from the local market (so-called "faux poisson").

Concerning the Mediterranean swordfish, 9 groups of data are selected as priority:

- quantities of dead discards,
- quantities of bycatch released alive,
- catch data and fishing effort data,
- size frequency,
- dataset of catch at size estimations (ICCAT task 2),
- datasets on National Observer programs,
- information about maturity,
- information about reproduction and fecundity,
- information about age.

[1] ICCAT Rec [05-09] Recommendation by ICCAT on compliance with statistical reporting obligations.

1. Introduction

Currently, we have seen increasing concern over whether fisheries can sustainably provide seafood without overfishing fish populations (Pons et al., 2018). Large pelagic fishes, such as tunas and billfish, are important contributors to food security and income in many developed and developing countries. The sustainability of these straddling stocks and their management are under the responsibility of different Regional Fisheries Management Organization (RFMO) like the International Commission for the Conservation of Atlantic Tunas (ICCAT) and the Indian Tuna Commission (IOTC).

Stocks of large pelagic fishes are exploited by several national fleets and most of these stocks have been categorized as overfished (ICCAT, 2018a). In this context, it’s essential to develop regional cooperation and coordination of monitoring activities among Member States (MS). The objectives behind are to develop common procedures (for sampling, data storage/exchange, data quality assessment, etc.) and to improve the existing fisheries data collection schemes.

The project MARE/2016/22: “Strenghtening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species” (acronym RECOLAPE), aims to propose solutions to support RFMO’s fishery management and generally improve large pelagic fisheries data collection.

2. Objectives

The overall objective of the project is to strength the regional cooperation, particularly among European MSs, in the area of biological data collection for fisheries on large pelagic fish. This will not only improve coordination among European MSs regarding the use of fisheries data for stock assessment purposes and formulation of management advice, but it will also trigger the concept of RSPs at the RFMO level. At the same time, this project seeks to provide solutions to certain needs in terms of data collection identified by scientists involved in the stock assessment of tuna RFMOs and by expert groups like the Large Pelagic Regional Coordination Group (RCG-LP).

To reach this purpose, the RECOLAPE project addresses several objectives:

- Facilitate the evolution of the Large Pelagic Regional Coordination Meeting (RCM-LP) towards the RCG-LP: the goal is to evolve from a single meeting to a continuous process that will have greater responsibilities in support of stock assessment and fisheries advice.
- Design a RSP for large pelagic stocks: facilitating the transition from individual national work plans towards regional ones.

- Develop data collection strategy and tools regarding additional data (not yet collected on a routine basis) on FADs. Such additional data could be used in combination with traditional CPUE or for building alternative abundance indices.
- Test alternative data collection methods for those cases where traditional methods present data deficiencies, for example for data collected using Electronic Monitoring System (EMS).
- Facilitate cooperation among MS in order to improve and develop common data quality assessment procedures at national and regional levels.
- Identify points of consensus and/or disagreement that may arise during the coordination process among organizations dealing with large pelagic fisheries data collection. The idea is to identify a framework of rules and feedback to improve future coordination or expand it on other fisheries/species.

The present deliverable is in relation to the Work Package 2: propose a RSP on large pelagic fish. In particular, the goal of this deliverable D.2.1 is to answer the first task of RSP's design: define data needs and prioritize them within the data collection scheme.

3. Case studies used for RSP's design

Regarding the design of a RSP for large pelagic fish, the focus is on two case studies: Mediterranean swordfish and major tropical tuna in the Atlantic Ocean (skipjack tuna, yellowfin tuna and bigeye tuna).

3.1. Mediterranean swordfish

Swordfish in the Mediterranean is considered as a stock that is heavily exploited by several countries which target swordfish mainly using two types of longlines: either surface drifting longlines, or mesopelagic longlines that have been gradually introduced since 2009 and nowadays have replaced the surface gear in several Italian and Spanish swordfish fishing fleets. This is particularly noteworthy, as these fisheries are among the largest within the stock area, and the gear changes have implications on catch rates and the size composition of catches. EU fleets are extending their activities throughout the Mediterranean basin far beyond their national waters and, according to ICCAT records, their landings account for about 75% of the total Mediterranean swordfish landings.

Management of Mediterranean swordfish is within the Convention area of the ICCAT. The stock is considered to be overexploited and ICCAT has recently adopted a multi-annual plan (Recommendation 16-05, https://tinyurl.com/yasv3key) aiming to the recovery of the stock. Apart from management measures the plan establishes a series of rules regarding the collection of fisheries data and the biological monitoring of the stock.

3.2. Major tropical tunas in the Atlantic Ocean

With 250, 000 tons caught in 2016, the European Union (EU) is by far the major contracting party in terms of catch volume of ICCAT (38% of the total catches, nominal catch information from ICCAT statistical databases). Within the EU tropical catches, the major gear used is the purse seine targeting tropical tunas (49% of the EU catches, nominal catch information from ICCAT statistical databases). Therefore, this case study will be focused only on the purse seine fishery. In 2016, the European purse seine fishing fleet operating in the Atlantic Ocean was composed of 21 vessels under French and Spanish flags, with an individual capacity above 600 tonnes. In total, EU purse seiners accumulated more than 6 300 fishing sets, either made on free schools, or sets made on tuna schools associated with drifting floating objects, artificial such as FADs or natural logs.

In 2017, catches were composed by: 58% of skipjack tuna, 31% of yellowfin tuna, 10% of bigeye tuna and 2% of other species.

Skipjack tuna is a gregarious species that is found in schools in the tropical and subtropical waters of the Pacific, Atlantic and Indian Ocean (ICCAT, 2014). Skipjack is the predominant species aggregated to FADs, where it is caught in association with juvenile yellowfin tuna, bigeye tuna and with other species of epipelagic fauna. The spatial distribution, movements and catchability of skipjack tuna are affected by environmental conditions, such as prey availability, temperature and dissolved oxygen concentration (Barkley et al., 1978; Brill, 1994; Brill and Lutcavage, 2001; Dueri and Maury, 2013). Focusing on the Atlantic area (east and west), only the two skipjack tuna stocks are not overfished and not in an overfishing situation. Furthermore, despite the absence of evidence that the eastern stock is overexploited, there is a lack of several information (for example the definition of a fishing effort associated with FADs or the difficulty of taking into account changes in catchability) necessary to improve the stock assessment (ICCAT, 2018a).

Yellowfin tuna is a cosmopolitan species distributed mainly in the tropical and subtropical oceanic waters (ICCAT, 2014). They are opportunistic feeders, eating on a great variety of prey species, like crustaceans, fish, cephalopods and gelatinous organisms (Potier et al., 2004). This allows to this species fast growth and high reproductive outputs (Pecoraro et al., 2017). Furthermore, yellowfin tuna are characterized by several peculiar anatomical and physiological traits, as for example a fusiform and elongate body shape, which improves movement through the water and minimizes hydrodynamic lift (Brill, 1996; Pecoraro et al., 2017). For management purposes, yellowfin tuna is divided into four stocks: Atlantic, Indian Ocean, western and central Pacific and eastern Pacific. Despite distinct spawning areas or substantial heterogeneity in the distribution, a single stock for the entire Atlantic Ocean is currently assumed (ICCAT, 2018a). Moreover, the last stock assessment of 2016 (using catch and effort data until 2014) estimated that the yellowfin tuna Atlantic stock was overfished (the next assessment is planned for 2020).

Bigeye tuna are distributed in the Atlantic Ocean between 50°N and 45°S (ICCAT, 2018a). This species is known to perform diel vertical migrations (deeper than other tropical tuna species) and exhibit a preference for deep, cold water during the daytime (Boggs, 1992; Brill, 1994;

Holland et al., 1999; Ohshimo et al., 2018). The main prey taxa of bigeye tuna are reported to be fish and squid, suggesting that it plays an important role as top predators in pelagic ecosystems (Young et al., 2010). Spawning takes place in tropical waters when the environment is favourable. From nursery areas in tropical waters, juvenile fish tend to diffuse into temperate waters as they grow. Catch information from surface gears indicate that the Gulf of Guinea is a major nursery ground for this species (ICCAT, 2018a). Bigeye tuna exhibits relatively fast growth (about 105 cm fork length at age three) and young fish form aggregations with other tunas such as yellowfin tuna and skipjack. These aggregations are often associated with drifting objects, whale sharks and sea mounts. This association weakens as bigeye tuna grow. The last stock assessment for bigeye tuna was conducted in July 2018 (ICCAT, 2018b). This estimation shows that the Atlantic stock was overfished since 2014.

4. Definition of data needs and priorities

All data identified in this section are essential for performing robust stock assessments and are categorized, when possible, following the ICCAT frame of data requests. The notion of data priority shows which data need particular attention or improvement regarding collection and/or parameter estimation. This means that the data are either not available or the current format includes bias or mistakes and needs to be improved to provide the correct information. It is very important to understand that all the data selected (priority or not) need to be collected to make robust estimates of stock parameters and catch levels. All these data are summarized in table 1 and refer to both fisheries and biological information.

Table 1: Definition of data needs and priority for performing robust estimates of stock parameters and catch levels of Mediterranean swordfish and major tropical tunas (in the Atlantic Ocean) stocks.

	Dataset	Description	Atlantic yellowfin tuna	Atlantic bigeye tuna	Atlantic skipjack tuna	Mediterranean swordfish
Mandatory by ICCAT	**Task 1 fleet characteristics** Fishing vessels with positive fishing effort and fishing for any of the ICCAT species in the convention area	Vessel and fleet ID, other attributes (gear group, length overall, tonnage and tonnage type), total fishing effort and fishing activity				
	Task 1 nominal catches Yearly total catch (targeted, non-target and by-catch, recreational and sport fisheries) estimates in live weight (kg) by species	Caught quantities landed (dead)				
		Caught quantities discards (dead)	Priority	Priority	Priority	Priority
		Caught quantities discards (alive)	Priority	Priority	Priority	Priority
	Task 2 catch & effort Montly catch and effort by species and geographical squares (5x5 longline and 1x1 for other gear)	Catch data and fishing effort data				Priority
	Task 2 size sampling Montly size frequencies (size/weigth classes) with number of fish sampled by sampling area and/or geographic square	Size frequency				Priority
	Task 2 catch at size estimations Montly estimations of size composition of the catch by fleet, gear and area strata (use relation weight-length)	/	Priority	Priority	Priority	Priority
	Support Vessel activity Montly activity of tropical support vessels authorized to operate in the ICCAT Convention area	Activity of support vessel (location and number of days at sea) and fishing vessel association	Priority	Priority	Priority	Not applicable
	FAD Montly number of FADs deployed by statistical rectangles (1°x1°)	FAD characteristics (presence/absence of beacon/buoy, type of FAD and beacon/buoy/echo-sounder, average numbers of beacons/buoyes active/deactivated/active lost)				Not applicable
		Number of FADs deployed by support vessels	Priority	Priority	Priority	Not applicable
	National Observer programs	Data collected under national observer programmes (observed by-catch including interactions between fleets with sea turtles and incidental seabird by-catch), data and information collected from sampling programme, meta-data like percent coverage, sample scheme or geographical regions				Priority
Non-mandatory but recommended	**Information about maturity**	For example region specific size and age at maturity		Priority	Priority	Priority
	Information about reproduction and fecundity	For example estimation of spawner and recruit proporiton in the catches				Priority
	Information about age	Based on fin-rays, otolith, vertebra, ect.	Priority	Priority	Priority	Priority
	Catch data from local market	Part of fish sold on the local market	Priority	Priority	Priority	

Empty cells indicate that the relevant information is already fully available. Furthermore, the main difference between task 1 and task 2 data is that the second includes a fine geographical dimension (like statistical square of 1° by 1°). In opposition task 1 data are reported at scale of ICCAT area.

Regarding the major tropical tunas in the Atlantic Ocean (second study case), 8 groups of data are priority.

An overview of the entire catch (retained, released alive and discarded dead catch) enables a more comprehensive albeit still incomplete estimate of the total fishing mortality (Gilman et al., 2017, 2013). Moreover, the drive to ecosystem-based fisheries management has made it more imperative to understand bycatch in fisheries as a factor contributing to the destabilization of oceanic communities (Essington and Punt, 2011; Hall et al., 2017; Hilborn, 2011). In this context, dead discards and by-product (part of the bycatch sold in local markets) in the task 1 nominal catch are a data priority. The main reason is that, for now, there are not common raising methods for data provided to ICCAT (and by analogy to IOTC). Data submission in 2017 to ICCAT and IOTC were submitted by using the raising method adopted by Amandè et al. (2010). This method is not fully appropriate, and it is necessary to discuss common alternative raising methods for bycatch. This raising method is based in "bycatch/tuna production" ratio, and there is not a clear linear relation between tuna production and bycatch quantities. Otherwise, this necessity was already mentioned during the European Tropical Tuna Observer Meeting, the 15th and 16th of May 2018 at Pasaia (Spain) (Ruiz *et al.*, 2018). Developing alternative raising methods is also the opportunity to reinforce the current observer sampling protocol or to show some weaknesses in the data collection and improve it accordingly.

One other potential bias is about the length-weight relationships. These relationships are essential to make estimations of size composition of the catch, for example in the task 2 catch at size or in the specific composition of the task 1 nominal catches. The strength of the outcome estimations is directly linked with the robustness of this relationship. For now, the length-weight relationships have been established at the beginning of the 80s (Caverivière, 1975; Cayré and laloë, 1986; Parks et al., 1982) and are applied directly in the T3 processes (Tropical Tuna Treatment[2]) to correct the data reported in logbooks, without considering the fishing mode (free school sets vs dFADs sets), the time period, the season, the area or any change in the fish condition along the time. It is necessary to test these relationships through a spatiotemporal variability analysis and by fishing mode in the Atlantic Ocean and in the Indian Ocean. In case of significant differences with the current relationships, the new length-weight parameters should be integrated through the T3 process to improve the task I and II estimates.

Activity of support vessels and associate number of FADs deployed is another group of data priority. The increasing use of FADs in the tropical tuna fisheries has a negative impact on the fishing mortality of juveniles of tunas (especially bigeye and yellowfin tuna). Furthermore, support vessels increase the fishing capacity of purse seine vessels in an uncontrolled manner by setting FADs. Moreover, the number of support vessels has increased over the years. In Atlantic Ocean, the flag of a support vessel is not necessarily the same as the purse seiner. This may cause concerns to evaluate the amount of fishing effort exerted by the associated purse

[2] The T3 process is a series of statistical treatments applied on data in order to correct it and make them exploitable to more academic questions of expertise.

seiners and the related change in fishing efficiency over the years. For example, in the Indian Ocean, through Res [17-01] (on an interim plan for rebuilding the Indian Ocean yellowfin tuna stock), IOTC does not allow the association of a support vessel with purse seiners if they have different flag State. Furthermore, by the recommendation 16-02, ICCAT suggested establishing an ad hoc working group on FADs to notably identify management options and common standards of FADs management. Based on these considerations, a focus is on the regulation of support vessels, in particular the link established in fishing operations between support vessels and individual fishing vessels. Moreover, all these questions and concerns are in relation to some objectives of the European project "Catch, effort and ecosystem impacts of tropical tuna fisheries 2" (CECOFAD2, EASME/EMFF/2016/008), continuity of the previous research project "Catch, Effort and eCOsystem impact of FAD-fishing" (CECOFAD, MARE/2012/24).

Investigation of the reproductive biology of tuna's population, like for example age at maturity, can provide useful information for stock assessment and management purpose. This biological data collection is planned among national working plans of EU Member States through Data Collection Multi Annual Program. For example, estimates of the size and age at sexual maturity are necessary inputs in size and age-structured stock assessment models (Sun et al., 2013). Furthermore, information on maturity gives important knowledge of reproduction for management and enables separation of estimates of abundance into values representing the immature and mature stages of the populations (Kolding and Giordano, 2002; Zhu et al., 2010). Studies on the reproductive biology of bigeye tuna were recently made in the Atlantic Ocean (Zhu et al., 2011), but there are still large uncertainties about the population biology. Moreover, skipjack tuna has been considered by most tuna RFMOs as a notoriously difficult species to assess due to its biology (skipjack recruitment occurs all over the year in many areas due to its large spawning period (ICCAT, 2014). Even if we dispose of knowledge about reproductive biology for these two species, it's necessary to have a review and an evaluation of all information available and update it if necessary (with a focus on the Atlantic Ocean for match with our study case).

Information concerning age is very important and represents an enormous effort invested in fishery science. Ageing fish is one of the major activities of fisheries science largely because individual fish growth is a fundamental biological process in population dynamics. Furthermore, most population dynamics models to assess and manage fish populations are age-based. Until the 60s, there was an increase of "growth study" on tunas and fish generally. However, while the number of "fish growth study" increases more until the 90s, the number of "tuna growth paper" kept similar between 1960 and 2000 (Murua et al., 2017). The decline of these studies could be due to the difficulty to estimate growth in tunas. For example, it's difficult to applying tagging programs on highly migratory species. Moreover, it's not always easy to bring back samples from these species (hard part, like otoliths or vertebrae used in ageing evaluation) because of their high commercial value (this sample collection often requires the purchase of the fish). Furthermore, even if these hard part is available, the perceived difficulty of ageing tuna from this part explain why the age of tunas is not widely and routinely estimated and used in stock assessment in comparison to other important commercial fish (Prince and Pulos, 1983). Focusing on the study case of major Atlantic tropical tunas, for now there is no consensus on the most appropriate method to age yellowfin tuna and bigeye tuna. In addition, there have been no published studies directly comparing the effects of different laboratories and methods of age estimation of these two species (Murua et al., 2017). Regarding the skipjack tuna, a large number of investigations were carried out between the 60s and 80s. In the Atlantic Ocean, these studies represent 25% of all studies on this species. Even if information of skipjack's age is more available than the two other major tunas of the study case, data knowledge should be updated.

The last data priority is in relation to the flow of tuna's catches sold on the local market, the « Faux poisson » (Amande et al., 2017). The landing of purse seiners in the Eastern Atlantic Ocean, particularly in Abidjan (Côte d'Ivoire), is comprised of tunas directed to canneries or cargos for exportation and a part of catches (bycatch or not) destined for the local market (Monin et al., 2017). This local market fish comprises small individuals of major tunas and tuna-like species but also billfishes and other bony fishes (Monin et al., 2017). Moreover, this flow increases since the 80s and has become today a very important socio-economic part of the local fisheries. Furthermore, because of lack of detailed knowledge on this tuna catches sold on the local market (e.g., in terms of species composition or in terms of the representation of the flow), the data of Faux-Poisson reported at ICCAT are likely biased. Improving our knowledge of the flow from landings to the local market and quantifying the task I of Faux-Poisson, by time and area, is a priority (one of the main objectives should be the design of a sampling plan). Moreover, these estimations are a part of mandatory data requested by the French fisheries administration (specify in the last data collection agreement between the French National Research Institute for Sustainable Development and the French Direction of Fishery and Aquaculture).

Regarding the Mediterranean swordfish, identification and prioritizing of data needs was based on aspects discussed in the relevant Assessment and Species Groups of ICCAT, as well as recommendations mentioned in the latest reports of the Scientific Committee for Research and Statistics (SCRS). According to them the data requirements listed in Table 1, apart from providing essential information on the swordfish fisheries activities will assist clarification of important questions related to:

- Size and age at maturity: As there are ecological differences between the east and west Mediterranean, fine scale biological sampling will allow to explore possible differences in swordfish life-history at the spatial scale.
- Discards: Recently adopted management measures and particularly the minimum catch size limitations established through ICCAT Recommendation 16-05[3], may have increased

[3] ICCAT 16-05. Recommendation by ICCAT replacing the Recommendation [13-04] and establishing a multi-annual recovery plan for Mediterranean swordfish.

discard levels, therefore it is important to improve estimates on the discard levels of undersized swordfish.

- Stock mixing and management boundaries: This refers to knowledge about stock boundaries between the Mediterranean and North Atlantic swordfish stocks. Answers to these questions demand collaborative and multidisciplinary research, including population genetics, electronic tagging and detailed analysis of life history parameters. The collection of fine scale spatio-temporal fisheries and biological data mentioned in Annex I will facilitate detection of potential life history differences among swordfish population units.

References

Amandè, M.J., Ariz, J., Chassot, E., de Molina, A.D., Gaertner, D., Murua, H., Pianet, R., Ruiz, J., Chavance, P., 2010. Bycatch of the European purse seine tuna fishery in the Atlantic Ocean for the 2003–2007 period. Aquatic Living Resources 23, 353–362. https://doi.org/10.1051/alr/2011003

Amande, M.J., Dewals, P., Amalatchy, J.N., Pascual, P., Cauquil, P., Irie, D.B.Y., Floch, L., Bach, P., Scott, J., Restrepo, V., 2017. Retaining by-catch to avoid wastage of fishery resources: how important is by-catch landed by purse seiners in abidjan? Collect. Vol. Sci. Pap. ICCAT 73, 6.

Barkley, R.A., Neill, W.H., Gooding, R.M., 1978. Skipjack tuna, Katsuwonus Pelamis, habitat based on temperature and oxygen requirements. Fishery Bulletin 76, 653–662.

Boggs, C.H., 1992. Depth, capture time, and hooked longevity of longlinecaught pelagic fish: timing bites of fish with chips. Fishery Bulletin 90, 642–658.

Brill, R.W., 1996. Selective advantages conferred by the high performance physiology of tunas, billfishes, and dolphin fish. Comparative Biochemistry and Physiology Part A: Physiology 113, 3–15. https://doi.org/10.1016/0300-9629(95)02064-0

Brill, R.W., 1994. A review of temperature and oxygen tolerance studies of tunas pertinent to fisheries oceanography, movement models and stock assessments. Fisheries Oceanography 3, 204–216. https://doi.org/10.1111/j.1365-2419.1994.tb00098.x

Brill, R.W., Lutcavage, M.E., 2001. Understanding Environmental Influences on Movements and Depth Distributions of Tunas and Billfishes Can Significantly Improve Population Assessments 20.

Caverivière, A., 1975. Longueur prédorsale, longueur à la fourche et poids des albacores (Thunnus albacares) de l'Atlantique (Rapport prelimminaire No. Document SCRS/75/73).

Cayré, P., laloë, F., 1986. Relation poids-longueur de listao (KATSUWONUS PELAMIS) de l'Océan Atlantique. Proc. ICCAT Intl. Skipjack Yr. Prog. 1 335–340.

Dueri, S., Maury, O., 2013. Modelling the effect of marine protected areas on the population of skipjack tuna in the Indian Ocean. Aquatic Living Resources 26, 171–178. https://doi.org/10.1051/alr/2012032

Essington, T.E., Punt, A.E., 2011. Implementing Ecosystem-Based Fisheries Management: Advances, Challenges and Emerging Tools: Editorial. Fish and Fisheries 12, 123–124. https://doi.org/10.1111/j.1467-2979.2011.00407.x

Gilman, E., Suuronen, P., Chaloupka, M., 2017. Discards in global tuna fisheries. Marine Ecology Progress Series 582, 231–252. https://doi.org/10.3354/meps12340

Gilman, E., Suuronen, P., Hall, M., Kennelly, S., 2013. Causes and methods to estimate cryptic sources of fishing mortality [a]: cryptic fishing mortality. Journal of Fish Biology n/a-n/a. https://doi.org/10.1111/jfb.12148

Hall, M., Gilman, E., Minami, H., Mituhasi, T., Carruthers, E., 2017. Mitigating bycatch in tuna fisheries. Reviews in Fish Biology and Fisheries 27, 881–908. https://doi.org/10.1007/s11160-017-9478-x

Hilborn, R., 2011. Future directions in ecosystem based fisheries management: A personal perspective. Fisheries Research 108, 235–239. https://doi.org/10.1016/j.fishres.2010.12.030

Holland, K.N., Kleiber, P., Kajimura, S.M., 1999. Different residence times of yellowfin tuna, Thunnus albacares, and bigeye tuna, T. obesus, found in mixed aggregations over a seamount. Fishery Bulletin 97, 392–395.

ICCAT, 2018a. Report for Biennial Period, 2016-17, Part II, Vol. 2 434.

ICCAT, 2018b. Report of the 2018 iccat bigeye tuna stock assessment meeting. Pasaia (Spain).

ICCAT, 2014. Report of the 2014 ICCAT east and west atlantic skipjack stock assessment meeting. Dakar, Senegal.

Kolding, J., Giordano, W.U., 2002. AdriaMed Training Course on Fish Population Dynamics and Stock Assessment (FAO-MiPAF Scientific Cooperation to Support Responsible Fisheries in the Adriatic Sea).

Monin, J.A., Amalatchy, J.N., N'Goran, D.K., Chris, M.N., Kouadio, F.K., Nadège, C.K.A., Dewals, P., Restrepo, V., 2017. Utilization and trade of faux poisson landed in abidjan. Collect. Vol. Sci. Pap. ICCAT 73, 749–754.

Murua, H., Rodriguez-Marin, E., Neilson, J.D., Farley, J.H., Juan-Jordá, M.J., 2017. Fast versus slow growing tuna species: age, growth, and implications for population dynamics and fisheries management. Reviews in Fish Biology and Fisheries 27, 733–773. https://doi.org/10.1007/s11160-017-9474-1

Ohshimo, S., Hiraoka, Y., Sato, T., Nakatsuka, S., 2018. Feeding habits of bigeye tuna (Thunnus obesus) in the North Pacific from 2011 to 2013. Marine and Freshwater Research 69, 585. https://doi.org/10.1071/MF17058

Parks, W., Bard, F.X., Cayré, P., Kume, S., 1982. Length-weight relations for bigeye tuna captured in the eastern Atlantic Ocean. Col.Vol.Sci.Pap. ICCAT 17, 214–225.

Pecoraro, C., Zudaire, I., Bodin, N., Murua, H., Taconet, P., Díaz-Jaimes, P., Cariani, A., Tinti, F., Chassot, E., 2017. Putting all the pieces together: integrating current knowledge of the biology, ecology, fisheries status, stock structure and management of yellowfin tuna (Thunnus albacares). Reviews in Fish Biology and Fisheries 27, 811–841. https://doi.org/10.1007/s11160-016-9460-z

Pons, M., Melnychuk, M.C., Hilborn, R., 2018. Management effectiveness of large pelagic fisheries in the high seas. Fish and Fisheries 19, 260–270. https://doi.org/10.1111/faf.12253

Potier, M., Marsac, F., Lucas, V., Sabatié, R., Hallier, J.-P., Ménard, F., 2004. Feeding Partitioning among Tuna Taken in Surface and Mid-water Layers: The Case of Yellowfin (Thunnus albacares) and Bigeye (T. obesus) in the Western Tropical Indian Ocean 12.

Prince, E.D., Pulos, L.M., 1983. Proceedings of the International Workshop on Age Determination of Oceanic Pelagic Fishes: Tunas, Billfishes, and Sharks (NOAA Technical Report NMFS No. 8).

Ruiz, J., P. Bach, I. Krug, H. Murua, J. Robinson, C. Shearlock, 2018. SIOTI support for improving information on bycatch for management of the Indian Ocean purse seine tuna fishery. IOTC–2018–WPEB14–19.

Sun, C.L., Yeh, S.Z., Chang, Y.J., Chang, H.Y., Chu, S.L., 2013. Reproductive biology of female bigeye tuna *Thunnus obesus* in the western Pacific Ocean: Reproductive biology of female *thunnus obesus*. Journal of Fish Biology 83, 250–271. https://doi.org/10.1111/jfb.12161

Young, J.W., Lansdell, M.J., Campbell, R.A., Cooper, S.P., Juanes, F., Guest, M.A., 2010. Feeding ecology and niche segregation in oceanic top predators off eastern Australia. Marine Biology 157, 2347–2368. https://doi.org/10.1007/s00227-010-1500-y

Zhu, G., Dai, X., Xu, L., Zhou, Y., 2010. Reproductive biology of Bigeye Tuna, Thunnus obesus, (Scombridae) in the eastern and central tropical Pacific Ocean. Environmental Biology of Fishes 88, 253–260. https://doi.org/10.1007/s10641-010-9636-7

Zhu, G.P., Dai, X.J., Song, L.M., Xu, L.X., 2011. Size at Sexual Maturity of Bigeye Tuna Thunnus obesus (Perciformes: scombridae) in the Tropical Waters: a Comparative Analysis 8.

FRAMEWORK CONTRACT – MARE/2016/22 “Strengthening regional cooperation in the area of fisheries data collection” Annex III “Biological data collection for fisheries on highly migratory species”

Project acronym: RECOLAPE

Task 2.2 - Definition of data-sharing mechanism

Participating partners:

D.2.2 - Outline of actions needed to ensure data sharing among Member States (v2)

Deliverable coordinated by Maria Teresa Spedicato (COISPA)

Index

Executive summary ... 3

1. Introduction ... 5
2. Objective ... 5
3. Type of data to be shared and their use ... 7
4. Existing Data Sharing Mechanisms and Agreement ... 9
5. Criteria and actions for data sharing mechanism ... 9
    - 5.1 Subjects involved and data sharing supports ... 9
    - 5.2 Data Storage and Access ... 11
    - 5.3 Confidentiality ... 11
    - 5.4 Data Transfer and Policy for Use of Data ... 12

Executive summary

The stocks of large pelagic are exploited by several national fleets, but despite this high degree of joint interest there is not yet any formalised data sharing mechanism among the different Member States to facilitate data analyses.

The project RECOLAPE, under the EU Call for Proposals MARE/2016/22, seeks to provide solutions to certain needs identified both by scientists involved in the stock assessment of the tuna RFMOs and by the Large Pelagic Regional Coordination Group (RCG-LP) established in 2017.

The present outline of actions needed to ensure data sharing among Member States (DS) has been drawn up by the partners involved in the RECOLAPE project, to facilitate the provision of data necessary for fulfilling the objectives of the project by all the Member States (MSs).

This DS is intended to be communicated to all relevant National Correspondents (NC).

The role assigned to the RCGs by the Regulation (EU) 2017/1004 (recast) makes them more responsible of the implementation/functioning of the DCF at regional and supra-regional levels.

The RCG-LP is thus the subject tasked to develop and implement procedures, methods, quality assurance and quality control for collecting and processing data with a view to enabling the reliability of scientific advice to be further improved. To facilitate this process, the RCG-LP is also the subject tasked to develop and implement electronic supports and tools for data sharing, including regional databases. Currently no regional database is present that allows storing large pelagic stock's data, and according to the RCG-LP, the development of such a support is urgent to allow an efficient use of the data collected under DCF, as recognised also by the DG MARE International Directorate (RCG_LP report, 2017). At present a new Regional Database and Estimation System (RDBES) is under development by ICES. RDBS is already used by northern RCGs and is currently limited to fisheries in the ICES area for demersal and small pelagic stocks. The RDBES is a tool that would allow improving the standardisation of quality among MS, and it also would allow responding to different data calls without extra work. Thus, expanding the scope of this RDBES to LP stocks would facilitate the data sharing in the RCG-LP framework. However, RDBES data model should be tested on the LP stocks first (more details are provided under task 2.5).

For the time being the data sharing can be supported through specific data call by the RCG-LP or conducted by projects such RECOLAPE in agreement with the RCG-LP. Indeed, this DS is intended to represent a reference also for future similar task at RCG-LP level.

Data analyses will be performed within the RECOLAPE project and thereinafter by specific working groups of the RCG-LP.

Therefore, regarding the data sharing of **swordfish (*Xiphias gladius*)**, all the European countries of the **Mediterranean** Sea are the involved subjects. Fisheries involved are: **longlines** (**surface and mesopelagic**).

Regarding the data sharing of **major tropical tunas (skipjack, *Katsuwonus pelamis*; yellowfin tuna, *Thunnus albacares*; and bigeye tuna, *Thunnus obesus*),** Spain France are the main European countries involved, though, from actions taken at RCG-LP level, the cooperation could be extended to other countries. Fisheries involved are: **purse seines**.

The exchange of data needed for the RECOLAPE activities can be supported by the project SharePoint and by the SmartDots online platform, the latter for the exchange of samples of hard structures for fish ageing.

The SharePoint of the RECOLAPE projects will host data and information derived from the national databases. As regards SmartDots the stored and exchanged images of biological hard structures and related information will be the ones made available by the Research Institutes involved in the RECOLAPE project.

National Correspondents will be part of this process in all the steps, and also the NCs of the countries not participating to the RECOLAPE project will be involved.

Access to the data to be shared shall be restricted to persons (project participants, other end-users and stakeholders) who have been granted (by the project coordinator or the RCG chair), a personal user name and a password.

This DS will represent the basis for the future flow of information towards the RDBES under development by ICES that will be managed by the RCG-LP, keeping under review the types of data to be shared. The RCG-LP will prioritize and develop road maps for data uploads as well as identify areas for further development.

The project partners agree and acknowledge that the data shared will be regarded as Confidential Information.

Dedicated sections of the RECOLAPE SharePoint will be made available to each National Correspondents who will receive a specific authentication to access and upload the data.

The data provision will be made according to the access restrictions deriving from the Regulation (EU) 2017/1004 (recast), while for non-EU countries in accordance with the limitations given by the owners of the data.

Data shared within RECOLAPE project will be used only for the purposes of this project.

1. Introduction

The evolving process of regional cooperation towards a greater responsibility for the Regional Coordination Groups entails the progress of coordination among EU Member States in the fisheries data collection field, in support of stock assessment and fisheries advice.

According to the Regulation (EU) 2017/1004 (recast)[1] (article 5; point 4) for the purpose of the multiannual Union programme, the Commission shall take into account, *inter alia*:

i. regional specificities and regional agreements concluded in regional coordination groups;
ii. the international obligations of the Union and its Member States.

The Regulation (EU) 2017/1004 (recast) (article 9; point 3) also establishes that the Regional coordination groups shall aim at developing and implementing procedures, methods, quality assurance and quality control for collecting and processing data, with a view to enabling the reliability of scientific advice to be further improved. For that purpose, regional coordination groups (RCGs) shall aim to develop and implement regional databases.

The stocks of large pelagic are exploited by several national fleets, consequently the fisheries are interacting, and the stocks are managed at the level of the Convention area of the International Commission for the Conservation of Atlantic Tunas (ICCAT) (e.g. swordfish in the Mediterranean) or at the level of the Indian Ocean Tuna Commission (IOTC) (e.g. tropical tuna). Despite this high degree of joint interest there is not yet any formalized data sharing mechanism among the different Member States to facilitate data analyses, but only some processed data, as for example for the swordfish stock, are widely available through ICCAT.

The project: "Strengthening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species – RECOLAPE", under the EU Call for Proposals MARE/2016/22, seeks to provide solutions to certain needs identified both by scientists involved in the stock assessment of the tuna RFMOs and by the Large Pelagic Regional Coordination Group (RCG-LP) established in 2017.

The present outline of actions needed to ensure data sharing among Member States (DS) has been drawn up by the partners involved in the RECOLAPE project, to facilitate the provision of data necessary for fulfilling the objectives of the project by all the Member States (MSs).

This DS is intended to be communicated to all relevant NCs.

2. Objectives

The objectives of the RECOLAPE project address several issues:

1. design a Regional Sampling Plan for large pelagic fisheries;

[1]Regulation (EU) 2017/1004 of the European Parliament and of the Council of 17 May 2017 on the establishment of a Union framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the common fisheries policy and repealing Council Regulation (EC) No 199/2008 (recast).

2. compare the data collected using Electronic Monitoring System (EMS) to the data collected by observers on board longline fleet, clarifying the strengths and weaknesses of the EM to become a complement to human observers;
3. develop tools and protocols for collecting new data needs around the FADs (Fish Aggregating Devices)[2];
4. collect "new" data on a routine basis (e.g. technology, crew…), to be used in combination with traditional data for tropical tuna purse seiner CPUE standardization;
5. facilitate cooperation among Member States in order to improve the procedures to assess the quality of biological data on large pelagic stocks, both at the national and regional levels.

Regarding the objective 1. the focus of RECOLAPE is on two case studies: a) Mediterranean swordfish and 2) major tropical tuna in the Atlantic Ocean.

These case studies are thus specifically tackled in the present deliverable 2.2, which has the objective of setting a frame of criteria and actions to support data sharing.

Regional data collection requires a common sampling protocol, raising method, data format, but also a common evaluation of data quality. Even if there is not a regional data base for large pelagic species, RCM-LP has used a common SDEF format (Standard Data-Exchange format) for data exploratory analysis, giving the opportunity this way to use IT tools, developed under the COST project (AA.VV., 2006[3]; ICES, 2010[4]), for evaluation of aspects of data quality. RDBES is instead a new format of the old RDB (FishFrame), which includes new tables. SmartDots is an online platform for sharing and comparing images of age structures that aims to standardize age determination among readers.

Data sharing plays thus a key role for the RECOLAPE objectives and the aim of the task 2.2 of the project is indeed to outline the necessary framework that would allow data sharing among all Member States exploiting the Mediterranean swordfish stock and major tropical tunas.

The data sharing would include special agreements between countries exploiting these stocks or any other relevant action, as for example the frame of the agreement/procedure used between France and Spain for tropical tuna fisheries.

The outline of actions for data sharing of the present deliverable takes advantage of the results/frames achieved in other projects/programs tackling similar issues (e.g. MARE2014/19_Med&BS; MARE2014/19_fishpi; Data Collection Reference Framework-GFCM; ICES Data Policy of FishFrame).

---

[2]FAD data are not specifically included under EU Multi Annual Program (EU Reg. 2016/1251), but the RCM-LP recommended during the 2016 meeting to include a number of FADs data under effort variables in table 4 (Fishing activity variables) from EU Multi Annual Program annex.

[3]AA.VV. 2006 COST project Final Report. Common tool for raising and estimating properties of statistical estimates derived from the Data Collection Regulation. Studies and Pilot projects for carrying out the common fisheries policy. Call for proposal ref : FISH/2006/15 – lot 2. Project no : SI2.467814: 118 pp.

[4]ICES, 2010. Report of the Workshop on the implementation of the Common Open Source Tool (COST), 13-16 April 2010, Nantes, France. ICES CM 2010/ACOM:42. 20 pp.

3. Type of data to be shared and their use

The Article 3 of the regulation EU 2017/1004 (recast) establishes that data collected under the Data Collection Framework (DCF) can be classified as follows:

a) '*primary data*', i.e. data associated with individual vessels, natural or legal persons or individual samples;
b) '*metadata*', i.e. data giving qualitative and quantitative information on the collected primary data;
c) '*detailed data*', i.e. data based on primary data in a form which does not allow natural persons or legal entities to be identified directly or indirectly;
d) '*aggregated data*', i.e. the output resulting from summarizing the primary or detailed data for specific analytic purposes.

Data classified under point d) are the ones usually available through the RFMOs databases (e.g. ICCAT), which can be associated with metadata (point b). This type of data is commonly used by experts in working groups dedicated to stock assessments and sometime called by research organizations and projects for scientific purposes.

The process of sharing data types as described at the points a) and c) is instead quite new, at least at the level of Mediterranean countries, except some specific case (for example bilateral agreement between Member States). The process started quite recently, in association with the needing of design regional sampling plans, optimizing sampling effort and programs, setting data check procedures, improving data quality, implementing quality indicators, in compliance with the EU regulation on DCF. Quality evaluation of the data is indeed the first step considered essential by end-users for the quality of stock assessment and advice.[5]

The Regulation (EU) 2017/1004 (recast) also indicates the requirements for data management (art. 13), data quality control and validation processes (art. 14). Regarding this last point, Member States are responsible for the quality and completeness of the primary data (checks for errors by appropriate quality control procedures) collected under national work plans, and for the detailed and aggregated data (validation before detailed or aggregated data transmission to end-users of scientific data). The quality procedures should be in accordance with the ones adopted by the international scientific bodies, regional fisheries management organizations, STECF and regional coordination groups.

The stock of swordfish is considered to be overexploited and ICCAT has recently adopted a multi-annual plan (Recommendation 16-05) aiming to the recovery of the stock. Therefore, apart from the regular ICCAT demands, as it is shown in task 2.1, the Regional Sampling Plan for swordfish to be designed in RECOLAPE will consider the additional questions mentioned in Part V (Scientific Information) of ICCAT Reg 16-05 that establishes a recovery plan. Thus, it will help to fill the identified knowledge gaps in:

- region specific size and age at maturity;
- habitat use for comparison of the availability of swordfish to the various fisheries;

[5] Scientific, Technical and Economic Committee for Fisheries (STECF) – Quality Assurance for DCF data (STECF 17-11). Publications Office of the European Union, Luxembourg, 2017, ISBN 978-92-79-67483-9, doi:10.2760/680253, JRC107587

- the impact of the mesopelagic longline fisheries in terms of catch composition, CPUE rates, size distribution of the catches;
- monthly estimation of spawners and recruit proportion in the catches.

Filling the above gaps and a regional sampling plan design will require that data are shared for common analyses.

For the major tropical tuna purse seine fishery, the aim of the sampling is to cover the fishery activities all over spatial and temporal scales, to increase the sampling coverage in all the strata (area/period/set type). To reach this objective all samples must be shared between countries as already done by France and Spain during their annual «Tropical Tuna Treatment, T3» working group. However, some indicators must be developed to analyze the variability (for example standard error or coefficient of variation) of estimates in regards to the sampling strategy (temporal and spatial coverage).

In addition, for the objectives of the task 5.3 on the Regional data improvement of RECOLAPE project and, more in detail, for the task 5.3.1 - Comparison of age-length keys between Member States and exploratory analysis, a comparison of age-length keys (ALKs) of the swordfish, *Xiphias gladius*, is planned among MSs of the Mediterranean, for the identification of the differences and differences' source between ALKs. The data required are including age-length keys and relevant metadata (e.g. sampling scheme, sampling procedures, ageing criteria, experience level of readers, ageing scheme).

For the implementation of the above case studies and tasks the following data (possibly for the period 2015-2017) should be shared:

1. **Landings data,** by fishing activity category European level 6 (métier), quarter, GSA and/or country (for tropical tunas only country is required), aggregated data;

2. **Biological data** by métier, sampling trip/event, at trip or haul level, detailed or detailed data;

3. **Fishing activity data,** related to sampling fishing trips with available biological data, detailed or primary data;

4. **Data** to estimate **CPUE** by stock and sampling unit, detailed data;

5. **Age length keys,** related to the age at length in the samples and proportions of age groups by length classes; detailed and aggregated data.

Guidelines for the establishment of common data exchange formats that can facilitate in-deep analyses and reporting and details on the type of data under the points from 1 to 4 are provided **in the tasks 2.3 and 2.5 of RECOLAPE**, while those related to the point 5 are provided in the data call issued for asking age data of swordfish **(WP5, task 5.3.1)**.

In case of missing data during the required period, alternative periods may be selected on a case-by-case level.

In case data in the proper format are already available, only an authorization to the data access is needed to be granted by National Correspondents.

The actions and criteria outlined in the present data sharing are coordinated with a similar task of the STREAM project[6] under MARE_2016_22 framework.

4. Existing Data Sharing Mechanisms and Agreement

The research Institutes of France (IRD) and Spain (IEO) have signed since 2011 (still in force) a Memorandum to define the scientific collaboration in the field of data collection of purse seine fisheries for tropical tuna. It is followed by a Monitoring Committee between the two institutions for evaluation of results and solution of possible issues. The collaboration of the Memorandum is centred on the following fields:

- fulfil the objectives of the DCF as regards tropical tuna fisheries, except the socio-economic aspects;
- follow-up of basic data of the fishery (effort, catches, catch composition and demographic structure);
- pilot projects on the estimates of by-catch and discards by observers on board;
- studies on biological traits (growth and reproduction of yellowfin tuna and bigeye tuna).

Besides this formal cooperation agreement, France and Spain through the research institutes involved in the data collection of tropical tuna purse seine fisheries, i.e. IEO (Spain), AZTI (Spain) and IRD (France), coordinate their sampling methodological approach, tools and sampling protocols. Since 2005, alternatively in Spain and France, an annual coordination meeting, the T3 group meeting, takes place with participation of scientists from both countries, to discuss regional sampling coordination and possible bilateral agreements. Scientists from non-EU countries like Seychelles, Madagascar, Ivory Coast, Senegal and Ghana who participate in the data collection of tropical tuna are also invited. Specific or common scientific contributions, as well as data calls, to tuna RFMOs (ICCAT/IOTC) are jointly elaborated.

The existing trend in coordination for the data collection in the tropical tuna purse seine fisheries can be understood as an example towards the regional sampling scheme.

5. Criteria and actions for data-sharing mechanism

5.1 Subjects involved and data sharing supports

The Regulation (EU) 2017/1004 (recast) indicates that Member States should cooperate among themselves, with third countries and with relevant stakeholders for the collection of data regarding the same marine region.

The role assigned to the RCGs by the Regulation (EU) 2017/1004 (recast) makes them more responsible of the implementation/functioning of the DCF at regional and supra-regional levels.

The RCG-LP is thus the subject tasked to develop and implement procedures, methods, quality assurance and quality control for collecting and processing data with a view to enabling the reliability of scientific advice to be further improved. To facilitate this process, the RCG-LP is also

[6]STrengthening REgional cooperation in the Area of fisheries biological data collection in the Mediterranean and Black Sea, STREAM (SI2.770115).

the subject tasked to develop and implement electronic supports and tools for data sharing, including regional databases.

Currently no regional database is present and, according to the RCG-LP, the development of such a support tool is urgent to allow an efficient use of the data collected under DCF, as also recognized by the DG MARE International Directorate (RCG_LP report, 2017). In order to facilitate data provision for the future regional coordination, RCM-LP, in 2017 meeting, recommended expanding the scope of the Regional Data Base FishFrame, hosted by ICES, to include EU Large Pelagic fisheries data. This would imply the inclusion of specific fields and codes of interest for the LP fisheries, as well as some LP expertise should be added to the Steering Committee of the RDB. However, two main aspects require further clarification: the adaptability of FishFrame to host the (highly diversified) data regarding LP fisheries and the new fields or tables specific to LP fisheries that are not currently in the current RDB. However, at present a new Regional Database and Estimation System (RDBES) is under development by ICES, which is already used by northern RCGs and which.is currently limited to fisheries in the ICES area for demersal and small pelagic stocks. The RDBES is a tool that would allow improving the standardisation of quality among MS, and it also would allow responding to different calls without extra work. Thus, expanding the scope of this RDBES to LP stocks would facilitate the data sharing in the RCG-LP framework. However, RDBES data model should be tested on the LP stocks first"

Regarding the data sharing of large pelagic fleets, except tropical tuna purse seine fisheries, a regional coordination does not exist, thus the RCM-LP recommended in 2017 and more recently in 2018 (details are reported in WP1) the organization of a workshop for the longline LP regional sampling coordination, treating Mediterranean separately and Long Distance Fishery (LDF).

For the time being the data sharing can be supported through specific data call by the project RECOLAPE. This will represent an example also for the future data sharing and issuing data calls at RCG-LP. Data analyses will be performed within the RECOLAPE project and in the future by specific working groups of the RCG-LP.

Therefore, regarding the data sharing of **swordfish (*Xiphias gladius*)**, all the European countries of the **Mediterranean** Sea are involved subjects. Fisheries involved are: **longlines** (**surface and mesopelagic**).

Regarding the data sharing of **major tropical tunas (skipjack, *Katsuwonus pelamis*; yellowfin tuna, *Thunnus albacares*; and bigeye tuna, *Thunnus obesus*)** Spain and France are the main European countries involved, though from actions taken at RCG-LP level the cooperation could be extended to other countries. Fisheries involved are: **purse seines**.

The exchange of data needed for the RECOLAPE activities can be supported by the project sharepoint and by the SmartDots online platform, the latter for the exchange of images from biological samples of hard structures for fish ageing.

5.2 Data Storage and Access

As stated in the Regulation (EU) 2017/1004 (recast) (art. 13), primary data and any other type of data to be shared shall be safely stored and all necessary measures shall be taken to ensure that these data are treated as confidential.

In addition, all necessary technical measures should be taken to protect such data against any accidental or illicit destruction, accidental loss, deterioration, or unauthorized consultation or distribution.

The sharepoint of the RECOLAPE projects will host data and information derived from the national databases. As regards SmartDots the stored and exchanged images and information on hard structures taken for ageing will be the ones made available by the Research Institutes involved in the RECOLAPE project.

Using compatible data storage and exchange systems, as those above described, will also facilitate future dissemination of information and results to other interested parties (e.g. end-users, EU and non-EU scientists, etc..).

National Correspondents will be part of this process in all the steps, and also the NCs of the countries not participating in the RECOLAPE project will be involved.

Access to the data to be shared shall be restricted to persons who have a user name and a password, being a user name for the sole use of that individual. When the user is logged in, the access to the data and functionality is a role based. Each role defines the user's access to functionality, data groups and the minimum aggregation level for those data. A list should be provided of focal persons, with role(s) assigned to the different profiles. A given user has only access to detailed data from her/his country, unless access to detailed data from additional countries is specifically stated in the user profile. All identified users have access to aggregated data from all countries in the resolution that will be defined by RECOLAPE and thereinafter by the RCG-LP.

Following the principles of the Regulation (EU) 2017/1004 (recast), access to viewing and analyzing other countries data does not entail permission to download, copy or publish detailed data. Such permission can only be granted by the National Correspondent of each country. The focal point in each EU MS is the National Correspondent. For non-EU countries a ICCAT, GFCM or other RFMO delegate can be considered the focal point.

This DS will represent the basis for the future flow of information towards the RDBES under development by ICES that will be managed by the RCG-LP, keeping under review the types of data to be shared. The RCG-LP will prioritize and develop road maps for data uploads as well as identify areas for further development.

5.3 Confidentiality

The project partners agree and acknowledge that the data shared will be regarded as Confidential Information.

The Parties must treat with confidentiality any data and data protection will be ensured[7] according to the following confidentiality policy:

- Data shall be held securely and not make publicly available.
- Data will not be freely available for dissemination.
- In case primary data and/or detailed data are requested/used to derive any metrics these data cannot be published at raw state. Anonymisation of data should be put in place in case data include information relating to identify or identifiable natural persons;
- Data sharing will undertake the access and the transfer only to Depersonalized Data[8] and Non-Personal Data[9];
- If a third party is authorized to access to the data, the project coordinator and thereafter the RCG chairs will ensure that any such third party complies with the terms of the data-sharing rules.

The data sharing must follow the rule of correct professional conduct and any further data exploitation (e.g. scientific publication, if allowed following the legal provisions of the Regulation (EU) 2017/1004 (recast)) should be agreed among the concerned Parties.

5.4 Data Transfer and Policy for Use of Data

Dedicated sections of the RECOLAPE sharepoint will be made available to each National Correspondent who will receive a specific authentication to access and upload the data.

The data provision will be made according to the access restrictions deriving from the Regulation (EU) 2017/1004 (recast) and for non-EU countries in accordance with the limitations given by the owners of the data.

Data shared will be used only for the purposes of the RECOLAPE project and in the future for the RCG-LP activities.

Correct and appropriate data interpretation with regard to scientific ethics is solely the responsibility of data users.

Data sources (individual data providers, if relevant) must be duly acknowledged.

Data Users must respect any and all restrictions on the use or reproduction of data such as restrictions on use for commercial purposes.

Data Users should inform RGC of any suspected problems with the data.

RCG should provide the Member States concerned and the Commission with references to the results of the use of the data.

Data Users will not forward the requested data to third parties without consent from the Member State concerned.

---

[7] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 - On the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[8] Information that relates to individuals where it is not possible to identify individuals from that information, whether in isolation or in conjunction with any other information.

[9] Information that does not relate to people including information about organizations, resources, projects or information about people that has been aggregated to a level that is not about individuals.

FRAMEWORK CONTRACT - MARE/2016/22 « Strengthening regional cooperation in the area of fisheries data collection » Annex III « Biological data collection for fisheries on highly migratory species »

Strengthening Regional cooperation in the area of large pelagic fisheries data collection (Acronym: RECOLAPE)

Task 2.3 - Development of an agreed sampling protocol

D.2.3 - Guidelines and sampling protocols, including reference lists with standard codes, establishing a common way for data reporting

Responsible: Mathieu Depetris (IRD), Antoine Duparc (IRD) & Fulvio Garibaldi (UNIGE)

Index

Executive summary 3

1. Introduction...4

1.1. Background and context 4

1.2. Purpose of this document 4

2. Proposal for the Mediterranean swordfish 4

3. Proposal for the major tropical tunas in the Atlantic Ocean 5

References..8

Executive summary

This document is in relation to the Work Package 2 of the project MARE/2016/22: "Strengthening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species" (acronym RECOLAPE). The main objective of this WP will make a proposal for a Regional Sampling Plan for large pelagic fisheries. According to that, the aim of this document is to develop and propose a common sampling protocol for large pelagic data collection, before testing it in the next task.

Regarding the Mediterranean swordfish, there is no common sampling protocols among Member States. Such protocols should be developed through RCG-LP, taking into account ICCAT code listing and the particularities of each national fishery. The most important aspects that should be considered refer to the spatiotemporal resolution of the sampled/reported fisheries data, the effort units and the size frequency composition of the catch. Furthermore, a protocol for sampling basic biological parameters, such as gonad maturity stage, should be developed.

In the case of the Atlantic tropical tunas, an onshore protocol was already shared, between France and Spain. This sampling schema will be the foundation of the current proposal. However, recent studies highlighted the necessity to improve some aspect of the sampling design. Regarding that, two proposals were made:

- modify the sampling stratification, moving from the current (hierarchically structured by large sampling areas, quarter and school types) to a regular 5-degree grid.
- reduce the number of individuals to measure per sample and still assess accurately the size distribution of the major tunas.

These two proposals of design improvement of the Regional Sampling Plan have to be carefully tested and discussed in the next task (associated with the D.2.4, simulation of the regionalised sampling plan) before being adopted.

1. Introduction

1.1. Background and context

The project MARE/2016/22: "Strengthening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species" (acronym RECOLAPE), aims to propose solutions to support RFMO's fishery management and generally improve large pelagic fisheries data collection.

The WP2 of this project aims to make a proposal for a Regional Sampling Plan for large pelagic fisheries with a focus on two studies cases: Mediterranean swordfish targeted by longline fisheries and major tropical tunas in the Atlantic Ocean targeted by purse seiner fishery.

1.2. Purpose of this document

The main aim of this document is to develop a common sampling protocol for large pelagic data collection. According to previous deliverables of the WP2, the proposal will focus on the two studies cases of the RECOLAPE project:

- Mediterranean swordfish targeted by the longline fisheries,
- Major Atlantic tropical tunas targeted by the purse seiner fishery.

Furthermore, data need and gaps identified in the D.2.1 (List of data requirements for the development of an RSP for the SWO-Med & TROP tunas) will be integrated in the current proposal.

2. Proposal for the Mediterranean swordfish

Regarding the Mediterranean swordfish fisheries there are not any shared protocols among countries, neither for port nor for onboard sampling. Each country follows its own protocol that it is based on the coding and requirements mentioned in ICCAT website (https://www.iccat.int/en/stat_codes.html). However, ICCAT coding is quite flexible, as it is designed to cover a variety of large pelagic fisheries with different characteristics. For instance, it allows reporting fisheries information (e.g. catch, effort) at various levels of spatial resolution and analogous flexibility exist when reporting effort units, as well as the size composition of the catch. Additionally, it does not provide specific guidance for sampling biological parameters (e.g. maturity).

Given that all swordfish individuals are considered to compose a common stock, it is necessary the establishment of common sampling protocols than ensure harmonized sampling and facilitate subsequent data processing and stock parameter estimates. It is suggested that the development of such protocols should be discussed within the RCG-LP in order to take into account the particularities of each national fishery.

The ICCAT coding should form the basis for such a discussion but clarifications should be made for a series of data requirements and most importantly for: (a) the spatiotemporal resolution of the sampled/reported fisheries data, (b) the effort units, and (c) the size frequency composition of the catch including definition on the required number of observations. Furthermore, a protocol for sampling basic biological parameters, such as gonad maturity stage should be developed.

3. Proposal for the major tropical tunas in the Atlantic Ocean

So far, a port sampling protocol has been shared between France and Spain, for tropical tunas in the Atlantic and Indian Ocean (Bach et al., 2018). This protocol performs well at assessing the total weight, due to scaling of the catches during the vessel landing at the harbour. However, recent studies (Duparc et al., 2018; Fonteneau et al., 2017; Herrera and Baez, 2018) highlighted the necessity to improve some aspect of the sampling design, especially the species composition and the size distribution of species, both of which are keystones for the stock assessments carried out by the tuna Regional Fisheries Management Organisations and by the Regional Coordination Group on Large Pelagics.

Based on the current protocol (available through the following link http://hal.ird.fr/ird-02132072), optimal modifications of the existing sampling schemes have been suggested:

- modify the sampling stratification, moving from the current (hierarchically structured by large sampling areas, quarter and school types) to a regular 5-degree grid.
- reduce the number of individuals to measure per sample and still assess accurately the size distribution of the major tunas.

Currently, the sampling design is hierarchically structured by statistical area, quarter and school types, in order to define strata as homogeneous as possible in terms of species composition and size distribution (Pianet, Pallarés and Pett, 2000). At each unloading, the purse seiner’s wells (which are the sampling units) are selected to ensure that fish originates from sets of the same (and in some instances neighbouring) stratum. However, these zones, which were established in the 90s based on the small amount of data available, were unequal in terms of surface (some of the areas 12 times larger than others) and were not uniformly used by the fishing fleets (see figure). This design could lead to an under-sampling of the area where catch events are less frequent. Furthermore, the fishing ground has evolved from year to year while sampling zones remained fixed. As a consequence, some fishing events were observed outside the sampling zones which raises the question of representativeness of the species compositions in these new fishing areas. In order to solve these issues, we propose to modify the design by sampling according to a regular grid, of a size to be defined, instead of the current zonation (figure 1 below).

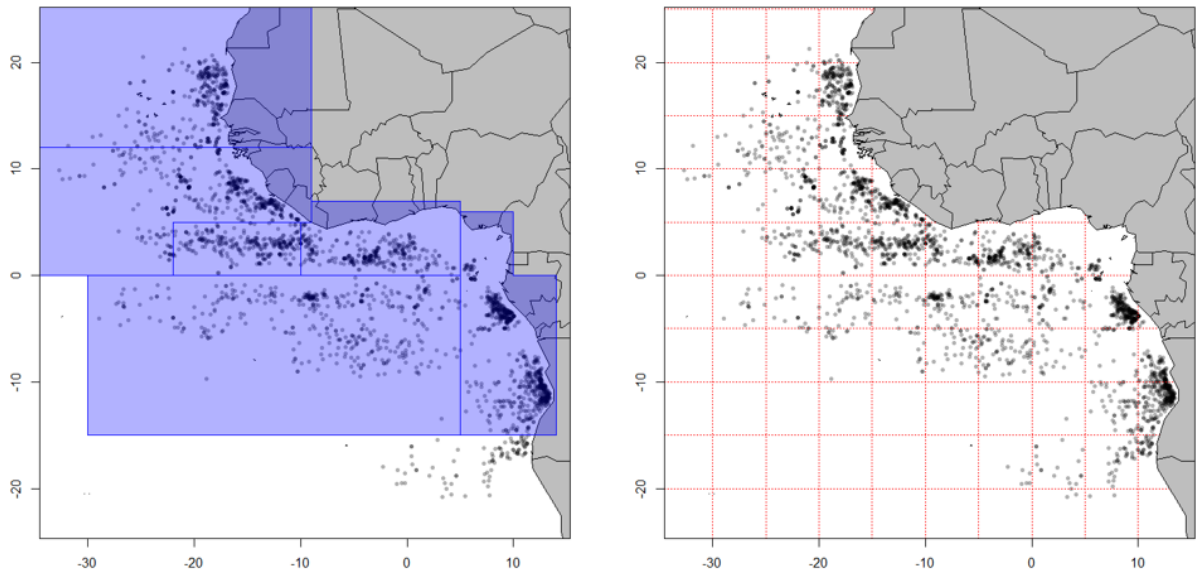


Figure 1: Sampled catches of the Spanish and French purse seiners fleet in 2017. The left panel: blue areas represent the current sampling zones used in the current protocol (for free schools). The right panel: Regular grid (with the example of 5° square) as a proposal of design modification.

Regarding each sample, the size distribution and the species composition is derived from the identification and measurement of a given number of individuals. These measurements are the most time-consuming part of the protocol. A recurrent question is whether the number of individuals to measure can be reduced and still assess accurately the size distribution and species composition of the major tunas. Indeed, by reducing the handling time of the fishes, the technical team should be able to sample more wells. An improvement of the RSP could so be to assess the optimal number of fishes to measure in each sample keeping a precise size distribution.

These two proposals of design improvement of the Regional Sampling Plan must be carefully tested and discussed in the next task (associated with the D.2.4, simulation of the regionalised sampling plan) before being adopted.

Finally, it is important to stress that the current sampling protocol is only valid for specimens frozen in brine. Recently, some fishing companies ultra-frozen a part of the catch, separating some individuals and storing them in specific wells. Thus, a specific protocol for these cases is needed.

References

Bach, P., Cauquil, P., Depetris, M., Duparc, A., Floch, L., Lebranchu, J., Sabarros, P., 2018. Procédures d'échantillonnage au port des thonidés tropicaux dans les océans atlantique et indien.

Duparc, A., Cauquil, P., Depetris, M., Dewals, P., Floch, L., Gaertner, D., Herve, A., Lebranchu, J., Marsac, F., Bach, P., 2018. Assessment of accuracy in processing purse seine tropical tuna catches with the T3 methodology using French fleet data. Presented at the 20th working party on tropical tunas, IOTC, Mahé, Seychelles.

Fonteneau, A., Alayón, P.J.P., Chassot, E., 2017. From large fixed to small mobile spatio-temporal strata: improving estimates of species and size composition of the landings of the European purse seine fishery in the Atlantic Ocean 21.

Herrera, M., Baez, J.C., 2018. On the potential biases of scientific estimates of catches of tropical tunas of purse seiners, the EU and other countries report to the ICCAT and IOTC. IOTC–2018-WPTT20–17.

Pianet, R., Pallarés, P., Petit, C., 2000. New sampling and data processing strategy for estimating the composition of catches by species and sizes in the European purse seine tropical tuna fisheries 37.

FRAMEWORK CONTRACT - MARE/2016/22 « Strengthening regional cooperation in the area of fisheries data collection » Annex III « Biological data collection for fisheries on highly migratory species »

Strengthening Regional cooperation in the area of large pelagic fisheries data collection (Acronym: RECOLAPE)

Task 2.4 - Simulation of the proposed RSP (including cost and implication)

D.2.4 - Comparison of the designed RSP and the "business as usual" approach in terms of costs/benefits – Case study: Tropical Tunas in the Atlantic Ocean.

Duparc Antoine*[1], Depetris Mathieu[1]

*Corresponding author

[1] Institut de Recherche pour le Développement (IRD), MARBEC UMR 248 (IRD, Ifremer, CNRS, Univ. Montpellier), Sète, France

Index

Executive summary 3

1. Simulation of the RSP for the tropical tunas in the Atlantic Ocean 4

1.1. Spatial and temporal effect sizes on the species composition 4

1.1.1. Aims 4

1.1.2. Methods 4

1.1.3. Results and recommendations 5

1.1.4. Take-home message 6

1.2. Spatial simulation of the sampling effort for species composition 6

1.2.1. Aims 6

1.2.2. Methods 6

1.2.3. Results and recommendations 8

1.2.4. Take-home message 8

1.3. Simulation of the sampling intensity for size distribution assessment by species 11

1.3.1. Aim 11

1.3.2. Methods 11

1.3.3. Results and recommendations 12

1.3.4. Take-home message 12

1.4. Other source of bias : Catch representativeness in samples 13

1.5. Conclusion of simulations 15

2. Simulation of the RSP for the Mediterranean swordfish **¡Error! Marcador no definido.**

References 17

Appendices 18

Executive summary

This document is in relation to the work package 2 of the project MARE/2016/22: "Strengthening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species" (acronym RECOLAPE).

The aim of the task 2.4 is, through statistical simulations, to test for potential optimizations and adjustments of the Regional sampling plan (RSP) designed in the task 2.3. The RSP will take advantage of the work done under the project MARE2014/19-Med&BS and the sampling experience acquired this last decade. Considering the RSP of the tropical tuna, the total weight was relatively well known, due to scaling at the port. The main issue of the RSP was so to accurately sample the species composition variability (temporally and spatially) and the size distribution of species which both were keystones for the stock assessment of the tuna RFMO and by the RCM-LP. The task 2.4, simulation of the proposed RSP, so mainly focus on these two points although other sources of bias were explored. The following document gives the description of methods as well as the results and interpretation of the simulation. Finally, given results, recommendations for the RSP optimization will be discussed.

The major change in the RSP should be the modification of the hierarchical spatial design. Indeed the use of a regular grid, instead of large zones of unequal areas, should improve the estimations of the species composition and the size distribution of species. Thus, simulation results demonstrated that in the current RSP, over-sampling occurred in several squares whereas others could be under-sampled. Therefore, theoretical thresholds of sampling were estimated, according to the amount of error on the mean composition, to help make decisions on the RSP optimizations. These optimal numbers of samples appeared to be dependent of the species, BET being the most difficult species to assess. It was also proved that the resolution of the RSP should be finer as possible. At the end, the use of a grid of 5° square should be a good compromise considering a sampling effort as "business as usual". In such condition, it seems to be challenging to assess mean species composition with less than 5% of error because it involves too much effort in terms of samples number by square (at the 5° square scale).

The analyses also suggested that the number of fishes currently measured seems to be suitable to assess the size distribution of major tropical tunas. Thus YFT and BET, for which all individuals were actually measured in the protocol, required more than 150 individuals to maintain an accurate representativeness of their size distribution in the sample whatever the school type. It could be considered, after further analyses, to increase the number of SKJ measured in FOB catches.

Finally, one of the major constraints in the RSP is the mixture of sets in wells of the vessels. Indeed, wells composed of different school type, or numerous sets (more than 5 sets sometimes), or covering a large extent (sets far from each other) are not suitable to assess species composition and size distribution. In these cases, measured fishes could not be attributed to a particular set and so do not enable to reconstruct species composition and size distribution of heterogeneous catch sources. Therefore, increase the number of sampled wells improve the RSP accuracy only until all wells suitable for sampling were sampled.

Associated to this document, R scripts developed for simulations were stored in a public repository located on Github (https://github.com/OB7-IRD/RECOLAPE/tree/master/WP2).

1. Simulation of the RSP for the tropical tunas in the Atlantic Ocean

1.1. Spatial and temporal effect sizes on the species composition

1.1.1. Aims

Fishing is a seasonal activity correlated to the biology of species, such as reproduction or feeding period. As consequence, vessels exploit the area according to the period of the year, and so spatial use of the ocean should be correlated to the time dimension. This assumption has to be confirmed to enhance the RSP design.

Therefore, before simulated sampling effort, assessment of the importance of spatial and temporal dimension at explaining species composition is required to make a decision on the optimal resolution at which sampling should occur.

1.1.2. Methods

1.1.2.1. Relationship between space and temporal scale

We investigated the relationship between space and time in the fishing activity. We so estimated the association between space and time on the number of samples with Cramer's V coefficient, which is based on Chi-Squared statistics (Cramér 1946). Cramer's V coefficient varies from 0, for no association, to 1, complete association between the nominal variables (Table 1).

Table 1: Interpretation of the Cramer's V value

Cramer's V	
0.25 or higher	Very strong relationship
0.15 to 0.25	Strong relationship
0.11 to 0.15	Moderate relationship
0.06 to 0.10	weak relationship
0.01 to 0.05	No or negligible relationship

Samples were harvested at landing on the French and Spanish fleet during the years 2015 to 2017 and were categorized between free school (FSC: 1405 samples) and school under floating objects (FOB: 2291 samples). We assumed that the number of well sample was correlated to the total number of samples in a given square. We so calculated the number of samples by square and period for different scenarios, which varied in square size (1, 5 and 10°) or in period duration (1 and 2 weeks, 1 to 3 months). We then calculated Cramer's V coefficient to estimate the association intensity.

1.1.2.2. Effect size

We tested how much the temporal and the spatial scale of sampling explained the species composition of samples. To do so, we first performed multivariate analyses of variance (MANOVA, Krzanowski 2000) with species composition (percent of SKJ and YFT) as response variable and year, period, spatial square of the sample as explanatory variables. We ran MANOVA variating the square size (from 1 to 10°) and the period duration (week, 2 weeks and 1 to 3 months).

Catches by sets and by species come from sampling at the harbor on the French and Spanish fleet for the years 2015 to 2017. Species composition was the percent of the catch of major tuna species (BET, SKJ, and YFT). Information of school type was recorded (free school – FSC: 1213 samples, school under floating objects – FOB: 2489 samples).

We estimated effect size of the 3 explanatory variables (square, year and period of the year) calculating, Partial eta-squared (η^2), the partial association for linear models (Muller et al. 1992). η^2 describes the proportion of total variation attributable to a given factor, excluding other factors from the total non-error variation. Finally, we computed 95% CI on η^2 using bootstrap on samples (100 replicates with replacement).

This methodology was applied separately for catch on FSC and FOB considering distinct ecological processes and species composition.

1.1.3. Results and recommendations

The number of samples by square was strongly associated with the period of aggregation whatever the square size and the period duration (see details in Appendix 1). The Cramer's V coefficients were always larger than 0.27 and decreased along with the square size (Figure 1). For instance, the number of samples by 5° square was strongly correlated to the month for both school types. Indeed Cramer's V coefficients are 0.38 and 0.37 respectively for FOB and FSC. We also noted that association between spatial and temporal dimension was very similar for the FSC than FOB whatever the spatio-temporal resolution.

We could so conclude from these results that each period of the year represents only a part of the whole fishing ground. Therefore, it is essential to sample catches all along the year to cover the entire fishing ground.

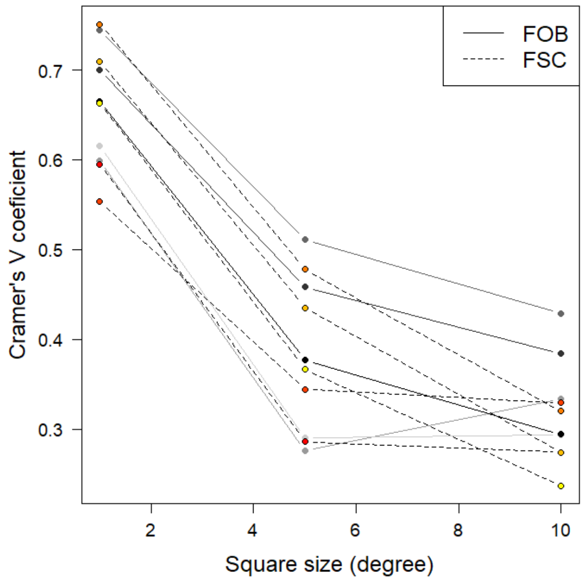


Figure 1: Cramer's V coefficient on contingency table of the number of samples by square and period. Color gradient correspond to different period duration (1 and 2 weeks, 1 to 3 months)

Second, spatial dimension explained larger part of the variability in species composition of catches than the period. Indeed Eta square (η^2) of the square size calculated from MANOVA ranged from 0.65 to 0.20 both for FOB and FSC whereas η^2 of the period ranged from 0.17 to 0.004 (Figure 2). For instance, effect size of spatial dimension at 5° resolution is 11 times more informative than the month resolution for FOB and 5.4 times for FSC. The spatial dimension is slightly higher for the FSC species composition than FOB. The year effect was always negligible with a mean around 0.004 and 0.002 respectively for FOB and FSC (see details in appendix 2).

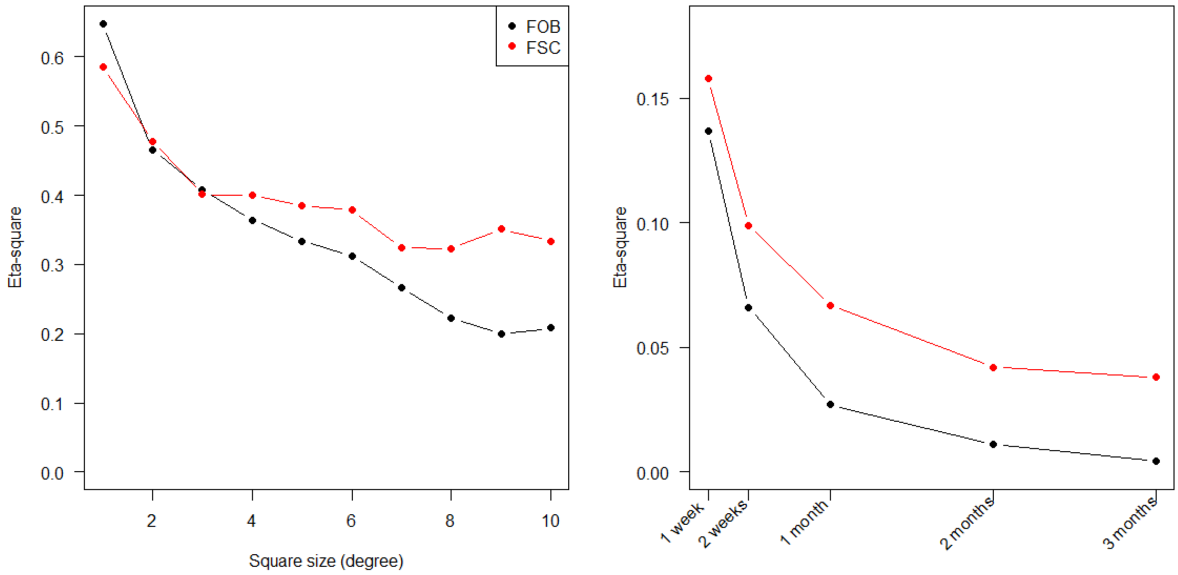


Figure 2: Partial association (Eta-square) for MANOVA on mean species composition in major tunas catches according to square size with period fix for 1 week (left) and the period with square size fix to 1° (right) by school types

In other words, the species composition of a set is more influenced by its location than is catching date. Sampling should so focus first on covering maximum of the spatial extend exploited during a given period. Furthermore, the finer the spatial resolution and temporal resolution will be, the more accurate will be the assessment of species composition.

1.1.4. Take-home message

These analyses brought two major results. First and as expected, spatial and temporal use of fishing ground were highly correlated. Second, finer was the resolution of the sampling more accurate were the representativeness of the fishing ground and the species composition of samples. As a consequence, sampling should occur all over the areas exploited with the finest resolution possible for a given period but in continue all along the year to cover all the fishing ground. Given that, sampling effort has to be balanced between the need for quality data (representativeness and robustness) and the operating cost.

1.2. Spatial simulation of the sampling effort for species composition

1.2.1. Aims

The RSP aim to cover as much as possible the fishing ground. It was so proposed in the task 2.3 to use systematic sampling along a grid instead of large zones of unequal area and unbalanced sampling effort. This change will enable a better spatial covering.

First analyses on the resolution at which sampling have demonstrated the necessity to keep the finest resolution possible to assess the species composition accurately. Then, simulation procedure aims to determine the sampling intensity to be applied by a unit of area to accurately assess the species composition.

1.2.2. Methods

1.2.2.1. Database

Catches by sets and by species come from sampling at the harbor on the French and Spanish fleet for the period 2015-2017. We retain for the analyses the square with 12 samples at least. As sampling occurred on the well of the vessel, we kept for the analyses only wells composed of a maximum 3 sets as distant of less than 3° and all of the same school type to insure for spatial representativeness of the sample (FSC=1204 samples, FOB= 1568 samples). For each species, only catches with the presence of the species were used in the analysis.

1.2.2.2. Statistical analyses

We aimed at estimating percent of errors on the mean catch composition in 5-degree squares according to the sampling effort (in number of samples by squares) and for each species. We performed the analyses only with squares of 5°, because we did not dispose of enough samples to test for the sampling effort at a smaller scale considering the actual cost allocated to sampling.

The main issue is that the variability in species catch composition is correlated with the number of samples available until a certain threshold, which is a function of the heterogeneity inside square. We should so insure to select for square containing enough samples to efficiently assess the species composition.

To this end, we separately perform the following procedure for each species.

We proceeded in 3 steps:

1- Simulations of sampling effort

We first simulated a variation of sampling effort using bootstrap method. For each square, we sampled randomly 3 to N_{max}-2 samples, with N_{max} maximum number of samples available, and calculated mean species composition. We repeat each step 50 times. We finally calculated the coefficient of variation (CV) dividing the SD of each simulation by the mean calculated with all samples available in the square (μ_{ref} considered as the true mean estimator of the population, Figure 3, left panel).

Then for each 1% of CV:

2- Link between the number of samples and percent of error of the mean species composition

In each square, we determined the minimal number of smaples required to reach each percent of CV of the mean (N_{sampCV}). In Figure 3 left panel, we represented the N_{sampCV} for a CV of 20% (N = 20 samples).

3- Estimation of the "minimal" number of samples without dependence to the sampling effort

We modeled the N_{sampCV} against N_{max} in square using linear mixed effect model. Variables Year and Square were random effects to account for repeated measures. We repeat the modeling removing, one by one, squares with the less number of samples available until correlation become nonsignificant, i.e., independence of the N_{sampCV}.

At each repetition, we compared the null model against the model including N_{max} as a fixed effect using Akaike's information criterion (AICc) with second-order adjustment to correct for small sample bias (Burnham and Anderson 2002, Barton 2015,package "MuMIn"). We considered models significantly different when delta AICc >2. Finally, we estimated mean and 95% CI of the "minimal" number of samples (Figure 3, right panel).

This methodology was applied separately for catch on FSC and FOB considering distinct ecological processes and species composition.

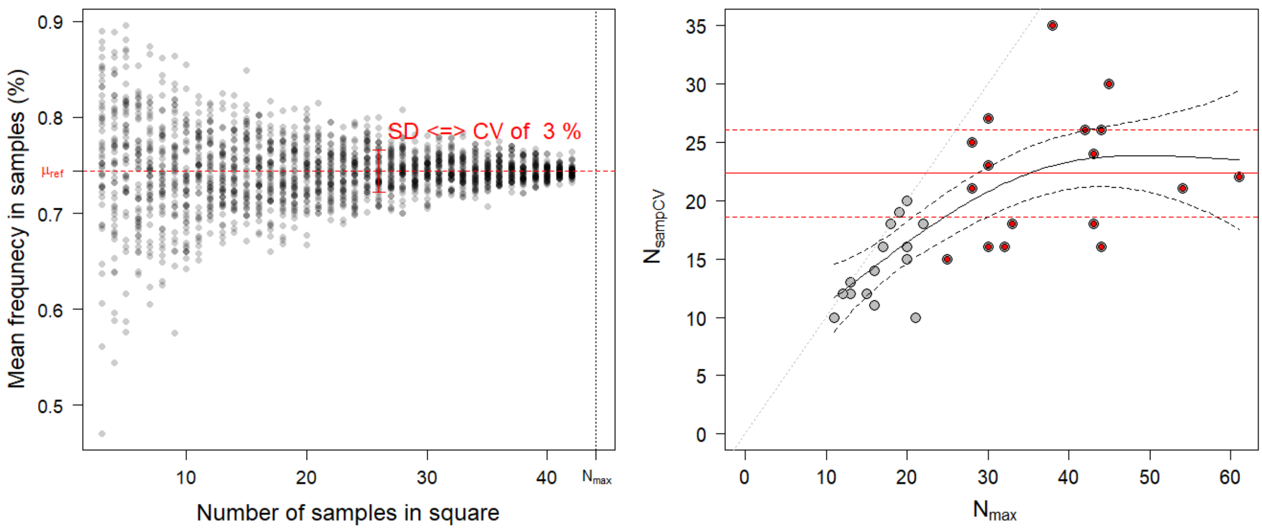


Figure 3: Example of the method to assess percent of error in frequency of SKJ in catches on FOB according to the sampling effort in at 5 degrees square scale. (left) Simulation of a sample increase by bootstrap (rep=50) for one square (centroid: Lat=-2.5°,Lon=7.5°) with 44 sampled samples available (N_{max}); μ_{ref} corresponds to the mean SKJ frequency in samples calculated with all samples. b) Number of samples (N_{sampCV}) by square required to assess the mean with 3% of error (CV). Gray points represent squares with not enough samples to correctly assess the mean composition (with a CV> 3%), i.e. correlated to the Nmax. Solid and dashed red lines represent mean and 95% CI of the number of samples required to have 3% of error on the mean composition of SKJ fitted from mixed model. Solid and dashed black lines represent mean and 95% CI fitted from gam model. Dotted gray line represents the straight line of slope 1 and intercept 0.

1.2.3. Results and recommendations

As expected, species compositions strongly differ between the school types (Figure 4). Catches on FOB were dominated by SKJ (mean of 0.65 ± 0.03%) and catches on FSC were dominated by YFT (mean of 0.65 ± 0.08% of the total catches). BET was always species in the minority.

Heterogeneity in square decreased with the number of samples available (Figure 5), which confirm the hypothesis that a minimum number of samples is required to assess the species composition.

From the simulation, this minimum number of samples increased exponentially with the precision of the estimate of mean proportion in catches (i.e., a decrease of CV, Figure 6). We noted that for the CV < 0.05, number of square containing enough sample for the simulation was only about 10 or less. We also noted that the 95% CI were extremely large for the SKJ and YFT composition in FSC resulting from a strong variation in catches (see details in Appendix 3).

Further, this number is mainly dependent of the species (Figure 7). Thus SKJ is the species that need less sampling effort, whereas mean proportion of BET is always the most difficult to assess as it needs more samples for the same CV. Surprisingly, the school types play a minor role in the sampling effort for the species composition assessment, except for the BET. Assessment of catches on FOB always needed less sampling effort than those on FSC.

1.2.4. Take-home message

From these results, it seems to be challenging to assess mean species composition with less than 5% of error because it involves many more effort in terms of sample number by square (at the 5° square scale). We could so recommend dividing the sampling effort over all the squares. Knowing that not all squares were fished with the same frequency and intensity, it appears important to adjust sampling priority in function of this fishing effort, focusing on rare fishing events. In the same way, BET, which is the most difficult species composition to assess, should be systematically sampled when possible.

We performed the analyses only with squares of 5°, because we did not dispose of enough samples to test for the sampling effort a smaller scale considering the actual cost allocated to sampling. If more resources are provided, we recommend applying the RSG at the smaller square possible.

Finally, it is essential to base the number of samples on the upper confidence interval limit (and not the mean) to assess the species composition in 95% of the squares.

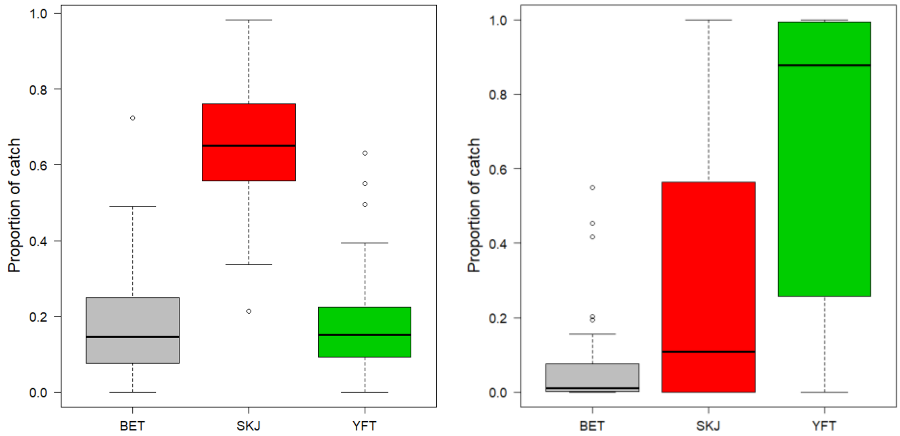


Figure 4: Boxplot of mean the species composition (in weight) of catches for FOB (left panel) and FSC (right panel) for major tunas in 5° squares. BET: Big eye tuna, SKJ: Skipjack tuna, YFT: Yellowfin tuna.

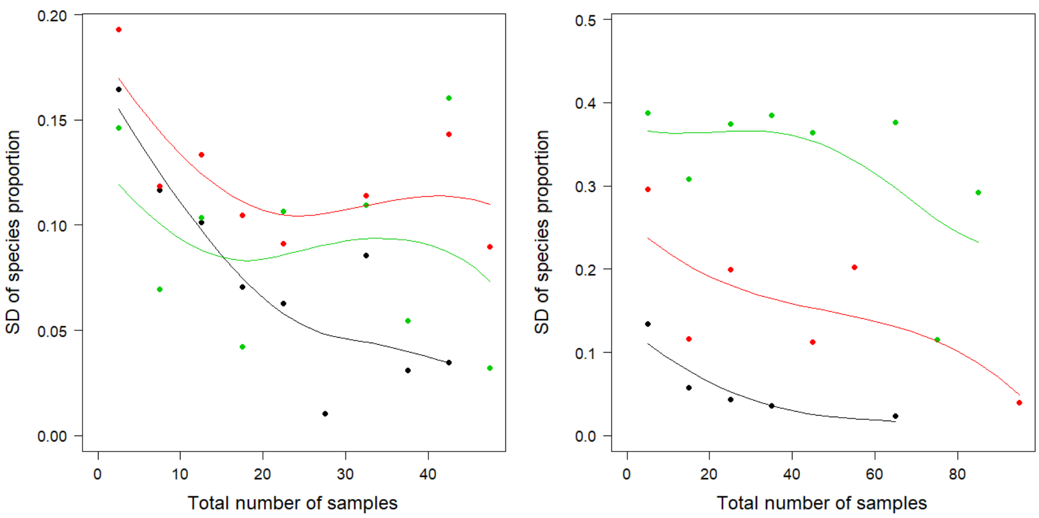


Figure 5: Variation (SD) in species proportion of major tunas according to the number of samples available (split in categories) in each 5° square and by school types (left panel: Floating object right panel: Free school). Solid lines represent a cubic smoothing s smoothing splines. BET: Big eye tuna, SKJ: Skipjack tuna, YFT: Yellowfin tuna

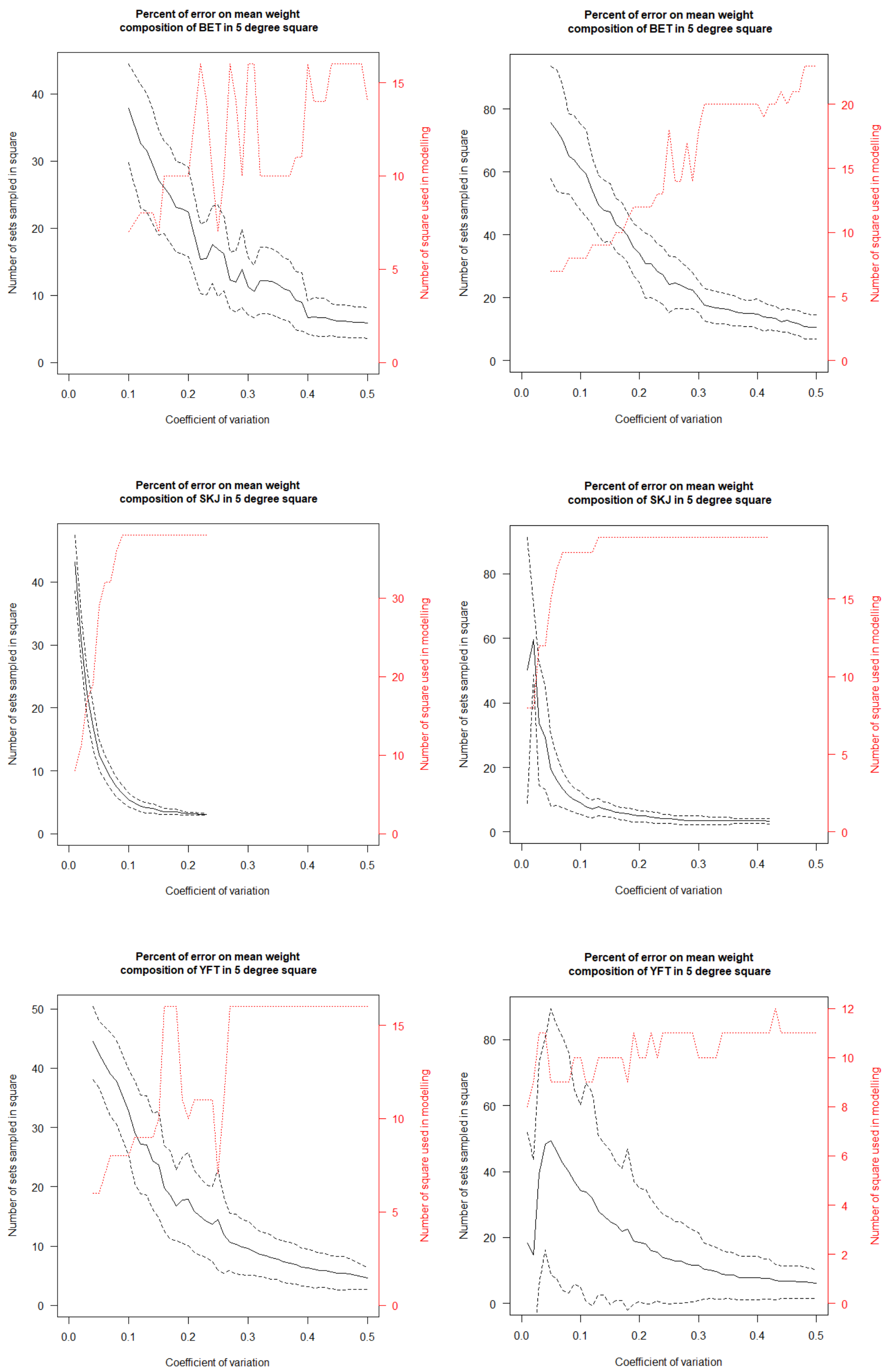


Figure 6: Mean and 95%CI of the Number of samples in a 5-degree square according to CV of the mean in composition per species and per school type: FOB on left panels (School under Floating object) and FSC on right panels (Free school). Red line correspond to the number of square-year used in the model

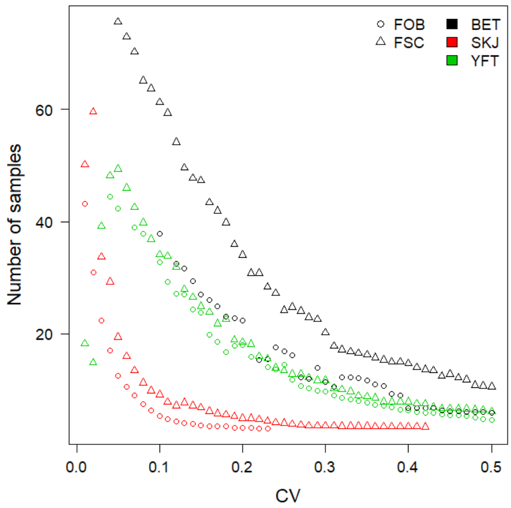


Figure 7: Mean number of samples in a 5-degree square according to CV of the mean per species and school type.

1.3. Simulation of the sampling intensity for size distribution assessment by species

1.3.1. Aim

The assessment of size distribution by species is as crucial point for the stock management of tuna. To fix this issue, the number of fishes measured in each sample is the parameter, which have to be evaluated. In this section, we so aims at simulate the number of fishes measured to ensure a representativeness of the size distribution of the 3 major tunas (BET,SKJ and YFT).

1.3.2. Methods

1.3.2.1. Database

The number of individuals by species comes from sampling at the harbor on the French and Spanish fleet for the period 2015-2017 (N=2291 samples for FOB and N=1100 for FSC). Samples with fewer than 100 fishes measured were removed from the analyses.

1.3.2.2. Statistical analyses

We first simulated a variation of sampling effort using bootstrap method. For each sample, we sampled randomly 50 to N_{max} fishes (with a step of 10), with N_{max} maximum number of fishes measured, and calculated size distribution of each species. We repeat each step 100 times.

Then we compared size distribution of each repetition to its reference, defined as the size distribution using all individuals measured in the same sample (S_{ref}), calculating Pearson correlation coefficient R. We considered size distribution as equal when R is equal to 0.95. We so calculated, for each species, each sample and each repetition, the minimum number of total fishes (N_{min}) to measure for which size distribution is similar to its S_{ref}.

Finally, we modeled N_{min} against N_{max} and test for their dependency using simple linear model. We repeat the modeling removing samples with less number of fishes available (10 by 10) until correlation become nonsignificant. When we met this condition, we estimated mean and 95% CI overall the remaining N_{min}. We also estimated the mean number of fishes of each species when N_{min} fishes were measured in a sample.

This methodology was applied separately for catch on FSC and FOB considering distinct ecological processes and species composition. Noted that for the SKJ, number of fishes measured was limited to 100 individuals in the protocol in order to shorten the sample time. We should so expect N_{min} be quickly independent of the N_{max} due to this sampling bias if the size distribution is homogenous as it was hypothesis under this limit of 100 measured individuals.

1.3.3. Results and recommendations

Both for the samples on FOB and FSC, we could have estimated YFT a minimum number of total fishes on average to measure in order to conserve good representativeness of the YFT size distribution (219 ± 13 fishes and 228 ± 16 fishes for FOB and FSC respectively). Under such condition the number of YFT were on average 107 ± 15 for samples on FOB and 152 ± 12 for samples on FSC.

Considering the BET, we only could estimate the N_{min} for FSC (343 ± 43 fishes), under which the number BET was on average of 124 ± 42. However, the N_{min} under FOB was never independent of the total number of fishes measured, meaning that the size distribution was not perfectly known in such condition (Figure 8).

Considering the SKJ and contrary to what we expected, we found that the N_{min} was very high on samples under FOB (343 ± 37 fishes) and is based on only 12 samples which is weak. Worst, the N_{min} was never independent of the N_{max} in FSC, meaning again that the size distribution was not perfectly known whatever the number of total fishes measured below 100. One reason for these results is that the threshold of 100 fishes measured could not be reach in FSC because SKJ were in two low proportions.

1.3.4. Take-home message

YFT and BET, for which all individuals were actually measured in the protocol, more than 100 individuals were on average measured to maintain an accurate representativeness of their size distribution in the sample whatever the school type. As expected the BET being the less abundant species, the minimal number of total fishes to measure in the sample was higher than for the YFT in samples on FSC and it was never reach in samples on FOB, with a R=0.95 (it would with R=0.92).

Moreover, the limitation of 100 individuals measured for SKJ, actually apply in the protocol, appeared to be too low to accurately estimated the size distribution.

From our results, we should so recommend maintaining at least the number of 500 fishes count in samples and measured a minimum of 150 fishes of each species (when possible) whatever the school type.

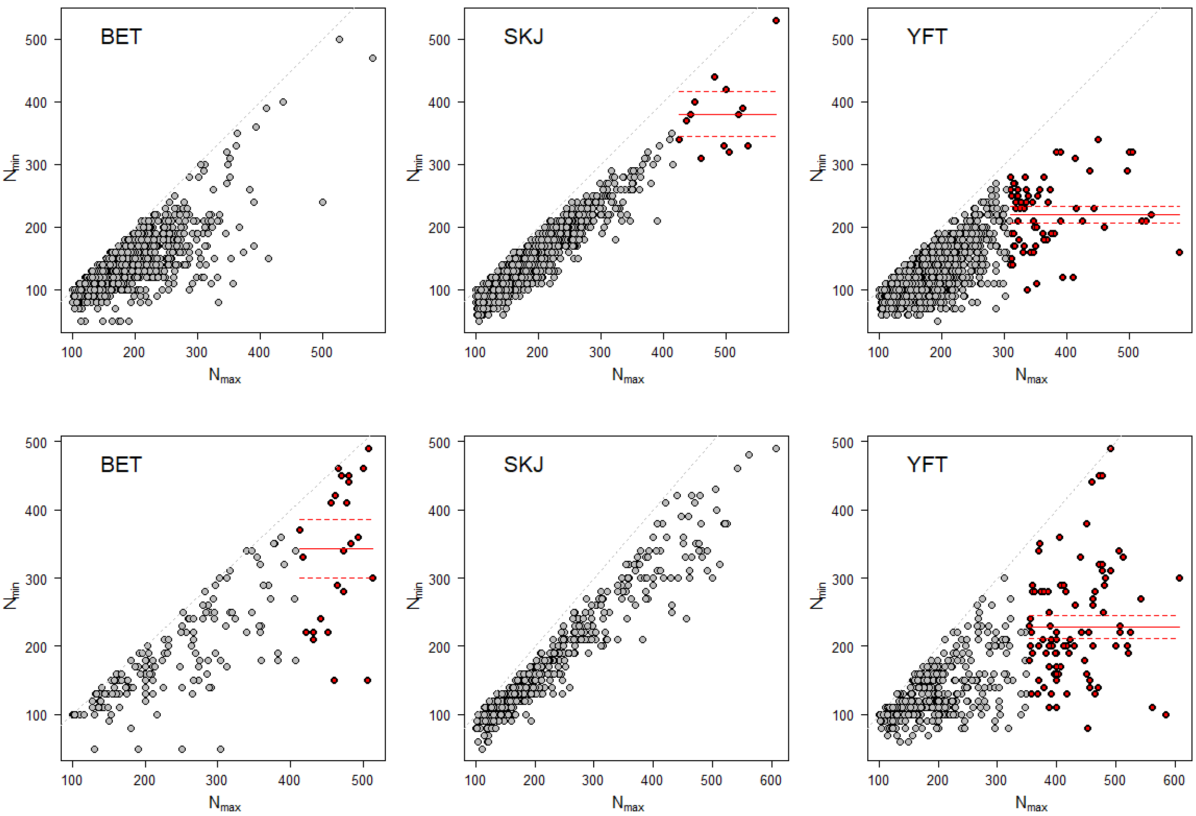


Figure 8: Minimum number of fishes to measure (N_{min}) to keep similar size distribution to the total sample (R=0.95) against the total number of fishes measured available in the sample (N_{max}). The 3 top panels are samples for the catches on FOB and the 3 bottom panels the catches on FSC. Red points are the samples for which N_{min} and N_{max} are not correlated anymore. Straight and dashed red lines represent the mean and 95% CI of the N_{min}

1.4. Other source of bias : Catch representativeness in samples

Well are composed mainly of more than one set, which is the main issue for species composition prediction. Well were so not sampled at random in order to maximize the homogeneity in species composition in wells. Therefore, wells composed of sets with different school types were avoided and wells only composed of one or two sets were preferred. As consequence, largest catches were almost systematically sampled (Figure 9), and wells composed of many small catches were under-sampled (<20t, Figure 10). This pattern was observed whatever the school type (Appendix 4). This issue can lead to bias if species composition varies according to the size of the catches. Unfortunately, there is no easy way to deal with this issue at the vessel landing because sets were mixed on board during the trip. However, the size of the catch could be taken into account in the modeling process.

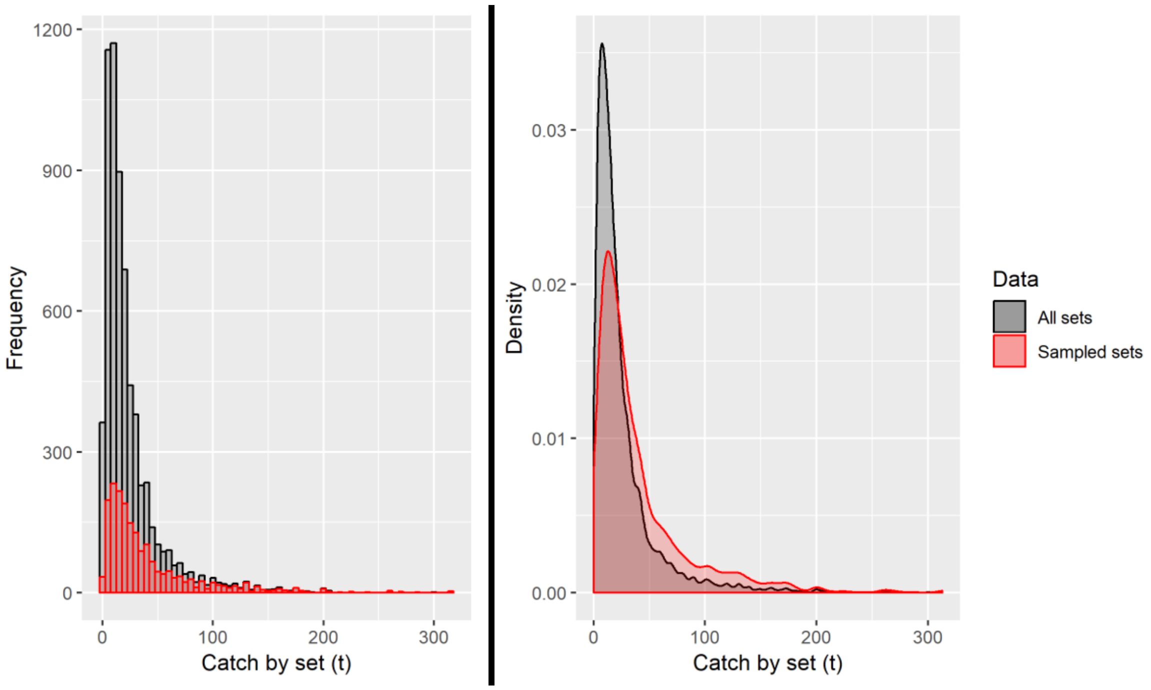


Figure 9: Frequency and density in catch per sets for all sets and sampled sets only from 2015 to 2017

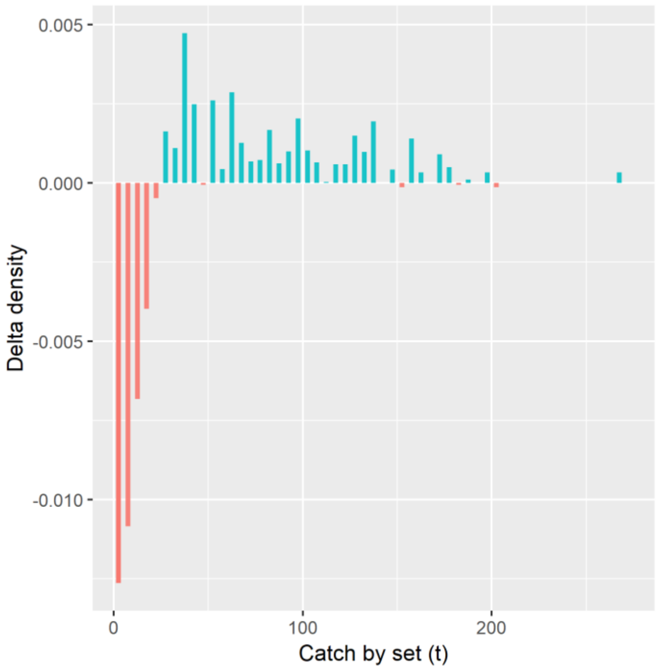


Figure 10: Differences in density (delta) between sampled sets and all sets according to catch by sets from 2015 to 2017

1.5. Conclusion and recommendations

During the study period (2015-2017), the yearly number of samples was 1200 on average (with about 750 on FOB and 450 on FSC) for the French and Spanish fleets. Considering, consistency in cost and human resources allocated to the sampling, and knowing the number of squares fished according to their size (Table 2.4.1), sampling effort should tend toward a sampling of 17 of well by squared of 5 degrees per year. From the simulations, it seems to be challenging to assess mean species composition with less than 5% of error because it involves many more effort in terms of samples number by square (at the 5° square scale). Obviously, this number is an indication, which should be adjusted all over the year function of the fishing activities. Indeed spatio-temporal aspect of the fishing should be taken into account. Not all squares could be sampled at each period because not all squares were fished all along the year (see spatio-temporal correlation of the fishing). Not even the "optimal" number of samples can be reached in square not frequently fished. The sampling effort should so be adjusted dynamically according to the vessel landings and their well plan characteristics all along the year.

Regarding the measure effort in each sample, the current total number of fishes measured seems to be suitable. Even If for the YFT the size distribution is already well estimated with fewer fishes, the size distribution for BET need such an effort. YFT and BET, for which all individuals were actually measured in the protocol, required more than 150 individuals to maintain an accurate representativeness of their size distribution in the sample whatever the school type. Regarding the SKJ, increase the number of measured fishes on free school samples should not change the accuracy of the size distribution because they are not abundant enough to reach 100 individuals (current threshold) or more in the samples. However, it should be considered to increase the number of SKJ measured in FOB catches.

The under-sampling of the smallest sets (<20t) identified as bias cannot be solved in the sampling at the vessel landing because sets were previously mixed on board. However, this sampling bias is an issue only if the species composition and the size distribution change accordingly to the set size. In that case, this size effect should be integrated in the model, which assesses the catch data. Further analyses so have to be performed to conclude with the procedure to follow.

Finally, one of the major constraints in the RSP is the mixture of sets in wells. Indeed, wells composed of different school type, or numerous sets (more than 5 sets sometimes), or covering a large extent (sets far from each other) are not suitable to assess species composition and size distribution. In these cases, measured fishes could not be attributed to a particular set and so do not enable to reconstruct species composition and size distribution of heterogeneous catch sources. Therefore, increase the number of sampled wells improve the RSP accuracy only until all wells suitable for sampling were sampled.

From all the simulation results, we so could make some recommendations by order of priority for the RSP improvement:

- Try to work on a grid of the smallest resolution as possible considering the allocated sampling effort and cost implication.

- It is more informative to sample a new square (new part of the fishing ground) than having numerous samples in only a few squares.

- Always sample the square which has been the least sampled

- It is also important to sample each square several times during the period it is exploited: by quarter at least but could be monthly for the densest catch area.

- According to the well plan stated by vessel crews, BET (and YFT on FOB) catches should be preferred in cases of choice between several suitable wells.

- Maintain at least the number of 500 fishes counted in samples

- Try to reach the measurement of 150 individuals when possible for BET and YFT

Table 2: Theoretical yearly number of samples by square according to square size and per school type considering 1200 samples per year on average (with about 750 on FOB and 450 on FSC)

Square size (degree)	Number of square fished	Samples by square	School type
1	443	2	FOB
2	173	4	FOB
3	94	8	FOB
4	61	12	FOB
5	43	17	FOB
6	35	21	FOB
7	24	31	FOB
8	18	42	FOB
9	18	42	FOB
10	15	50	FOB
1	212	2	FSC
2	91	5	FSC
3	56	8	FSC
4	39	12	FSC
5	27	17	FSC
6	22	20	FSC
7	16	28	FSC
8	14	32	FSC
9	13	35	FSC
10	12	38	FSC

References

Barton, K. 2015. MuMIn: Multi-Model Inference.

Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multimodel Inference. Springer New York, New York, NY.

Cramér, H. 1946. Mathematical methods of statistics (PMS-9). Princeton university press.

Krzanowski, W. 2000. Principles of multivariate analysis. OUP Oxford.

Muller, K. E., L. M. Lavange, S. L. Ramey, and C. T. Ramey. 1992. Power calculations for general linear multivariate models including repeated measures applications. Journal of the American Statistical Association 87:1209–1226.

Appendices

Appendix 1: Details of the CramersV index, comparing fishing activities according to square size (CWP in degree) and period length (1 week , 2 weeks, 1 to 3 months) and per school type

CWP	Period	School type	CramersV
1	week	FOB	0.599
1	week2	FOB	0.615
1	mon	FOB	0.665
1	mon2	FOB	0.7
1	mon3	FOB	0.744
1	week	FSC	0.553
1	week2	FSC	0.595
1	mon	FSC	0.663
1	mon2	FSC	0.709
1	mon3	FSC	0.751
5	week	FOB	0.276
5	week2	FOB	0.29
5	mon	FOB	0.377
5	mon2	FOB	0.458
5	mon3	FOB	0.511
5	week	FSC	0.344
5	week2	FSC	0.286
5	mon	FSC	0.367
5	mon2	FSC	0.435
5	mon3	FSC	0.478
10	week	FOB	0.334
10	week2	FOB	0.293
10	mon	FOB	0.294
10	mon2	FOB	0.384
10	mon3	FOB	0.429
10	week	FSC	0.329
10	week2	FSC	0.274
10	mon	FSC	0.237
10	mon2	FSC	0.274
10	mon3	FSC	0.32

Appendix 2 : Mean and 95% CI of Eta square from MANOVA on species composition in major tunas (BET,SKJ, YFT) according to square size (CWP in degree) and period length (1 week , 2 weeks, 1 to 3 months) and per school type

Formula	Eta cwp	Eta_cwp inf	eta_cwp sup	Eta year	eta_year inf	eta_year sup	Eta period	eta_period inf	eta_period sup	Type
week+year+CWP1	0.648	0.644	0.651	0.0038	0.0032	0.0043	0.137	0.134	0.14	FOB
week2+year+CWP1	0.638	0.635	0.641	0.0037	0.0033	0.0042	0.066	0.064	0.068	FOB
mon+year+CWP1	0.625	0.623	0.629	0.0038	0.0033	0.0043	0.027	0.025	0.028	FOB
mon2+year+CWP1	0.629	0.626	0.632	0.0037	0.0031	0.0042	0.011	0.01	0.012	FOB
mon3+year+CWP1	0.633	0.63	0.636	0.0037	0.0033	0.0042	0.0042	0.0037	0.0046	FOB
week+year+CWP1	0.648	0.644	0.651	0.0038	0.0032	0.0043	0.137	0.134	0.14	FOB
week+year+CWP2	0.465	0.462	0.469	0.0042	0.0038	0.0045	0.101	0.099	0.104	FOB
week+year+CWP3	0.408	0.405	0.412	0.0039	0.0034	0.0044	0.1	0.098	0.0102	FOB
week+year+CWP4	0.364	0.36	0.367	0.0038	0.0033	0.0043	0.101	0.099	0.103	FOB
week+year+CWP5	0.334	0.331	0.338	0.0049	0.0044	0.0053	0.106	0.104	0.108	FOB
week+year+CWP6	0.312	0.308	0.315	0.0051	0.0046	0.0056	0.118	0.116	0.12	FOB
week+year+CWP7	0.266	0.262	0.269	0.0034	0.0031	0.0037	0.121	0.119	0.123	FOB
week+year+CWP8	0.223	0.22	0.226	0.0032	0.0029	0.0037	0.129	0.127	0.13	FOB
week+year+CWP9	0.199	0.196	0.202	0.0029	0.0025	0.0032	0.113	0.111	0.116	FOB
week+year+CWP10	0.208	0.205	0.211	0.0032	0.0029	0.0036	0.109	0.106	0.111	FOB
week+year+CWP1	0.585	0.58	0.589	0.0175	0.0162	0.0189	0.158	0.155	0.162	FSC
week2+year+CWP1	0.581	0.577	0.585	0.0193	0.0181	0.0205	0.099	0.096	0.102	FSC
mon+year+CWP1	0.575	0.57	0.58	0.018	0.0168	0.0192	0.067	0.065	0.069	FSC
mon2+year+CWP1	0.578	0.573	0.583	0.015	0.014	0.016	0.042	0.04	0.044	FSC
mon3+year+CWP1	0.571	0.567	0.575	0.02	0.019	0.021	0.038	0.036	0.04	FSC
week+year+CWP1	0.585	0.58	0.589	0.0175	0.0162	0.0189	0.158	0.155	0.162	FSC
week+year+CWP2	0.477	0.473	0.48	0.0152	0.0144	0.0161	0.132	0.129	0.135	FSC
week+year+CWP3	0.401	0.398	0.404	0.0113	0.0104	0.0121	0.123	0.119	0.128	FSC
week+year+CWP4	0.4	0.397	0.403	0.012	0.0111	0.0129	0.0123	0.12	0.126	FSC
week+year+CWP5	0.385	0.382	0.387	0.0156	0.0146	0.0165	0.138	0.135	0.142	FSC
week+year+CWP6	0.379	0.376	0.382	0.012	0.011	0.013	0.11	0.107	0.113	FSC
week+year+CWP7	0.324	0.321	0.326	0.0116	0.0108	0.0123	0.128	0.125	0.13	FSC
week+year+CWP8	0.323	0.32	0.326	0.0082	0.0075	0.009	0.134	0.131	0.137	FSC
week+year+CWP9	0.351	0.348	0.354	0.0115	0.0108	0.0123	0.151	0.149	0.154	FSC
week+year+CWP10	0.334	0.331	0.337	0.0196	0.0186	0.0206	0.169	0.166	0.171	FSC

Appendix 3 : Mean and 95 CI of the number of well to samples to assess mean proportion in catch for each major tuna (BET, SKJ, YFT) according to the CV on this mean. Min_N sample is the minimal number of samples in squares used in the analysis. N square is the number of 5° squares retained for the analysis.

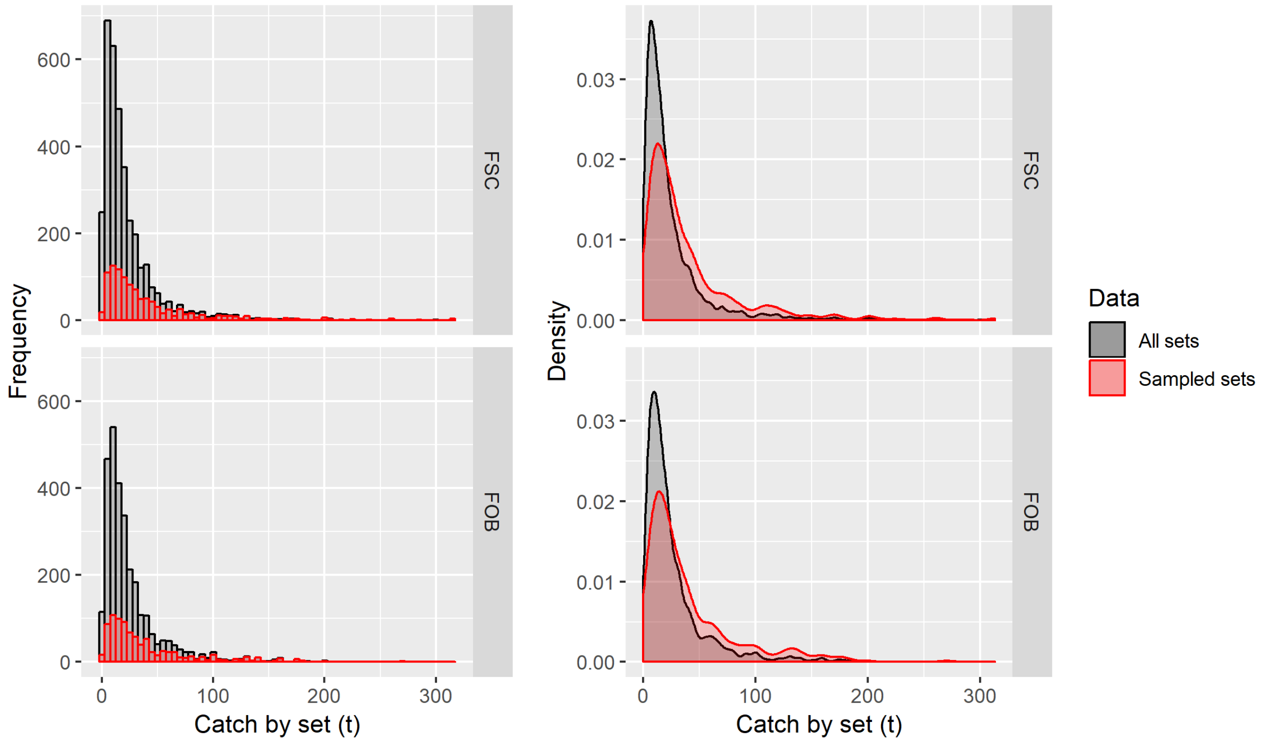
BET	FOB					FSC				
CV	Mean	95CI inf	95CI up	Min_N sample	N square	Mean	95CI inf	95CI up	Min_N sample	N square
0.01										
0.02								No match		
0.03										
0.04										
0.05			No match			75.6	57.9	93.8	16	7
0.06						73.0	54.0	92.3	39	7
0.07						70.3	53.2	87.8	39	7
0.08						65.1	53.1	78.5	39	8
0.09						63.7	50.7	77.7	46	8
0.1	37.9	29.8	44.4	43	7	61.2	47.7	75.2	46	8
0.11			No match			59.4	45.8	73.4	46	8
0.12	32.5	22.9	41.4	39	8	54.2	43.2	64.6	39	9
0.13	31.6	22.5	40.0	39	8	49.5	39.7	59.0	39	9
0.14	29.4	20.7	37.8	39	8	47.7	37.7	57.4	39	9
0.15	27.1	19.0	34.5	43	7	47.4	37.9	56.4	39	9
0.16	26.0	19.2	33.0	33	10	43.4	34.8	51.7	34	10
0.17	24.9	17.8	32.0	33	10	41.9	33.2	50.3	34	10
0.18	23.1	16.4	29.9	33	10	39.8	31.3	47.0	30	11
0.19	22.8	16.2	29.7	33	10	35.9	26.9	43.6	26	12
0.2	22.4	15.8	29.1	33	10	34.0	24.9	42.3	26	12
0.21			No match			30.8	19.9	40.6	26	12
0.22	15.3	10.2	20.6	26	16	30.8	20.1	39.6	26	12
0.23	15.5	10.1	21.0	29	14	28.4	19.1	37.3	24	13
0.24	17.6	11.8	23.3	33	10	27.2	17.9	36.2	24	13
0.25	16.8	9.7	23.5	43	7	24.2	15.2	33.3	16	18
0.26	16.2	10.6	21.8	33	10	24.8	16.6	33.0	23	14
0.27	12.2	8.0	16.4	26	16	24.0	16.6	31.4	23	14
0.28	12.0	7.5	16.5	29	14	22.9	16.3	29.8	19	17
0.29	13.9	8.2	19.7	33	10	22.6	16.6	28.1	23	14
0.3	11.3	7.1	15.7	26	16	20.2	15.3	25.3	16	18
0.31	10.6	6.7	14.5	26	16	17.8	12.6	23.0	14	20
0.32	12.2	7.2	17.1	33	10	17.2	12.1	22.4	14	20
0.33	12.2	7.2	17.1	33	10	16.8	11.8	22.1	14	20
0.34	12.0	7.1	17.0	33	10	16.6	11.7	21.7	14	20
0.35	11.6	6.8	16.5	33	10	16.3	11.6	21.1	14	20
0.36	10.9	6.3	15.6	33	10	15.8	11.0	20.7	14	20
0.37	10.6	6.0	15.3	33	10	15.3	11.1	19.8	14	20
0.38	9.2	4.9	13.6	31	11	15.0	10.9	19.2	14	20
0.39	9.0	4.6	13.3	31	11	15.0	10.9	19.2	14	20
0.4	6.7	4.3	9.1	26	16	14.7	10.2	19.6	14	20
0.41	6.7	3.9	9.7	29	14	14.0	9.6	18.5	15	19
0.42	6.7	3.9	9.6	29	14	13.7	9.8	17.7	14	20
0.43	6.7	3.9	9.6	29	14	13.5	9.6	17.5	14	20
0.44	6.4	4.0	8.8	26	16	12.5	9.0	16.1	13	21
0.45	6.2	3.8	8.5	26	16	12.8	9.1	16.7	14	20
0.46	6.2	3.8	8.5	26	16	12.1	8.5	16.1	13	21
0.47	6.1	3.7	8.5	26	16	11.8	8.0	16.0	13	21
0.48	6.0	3.7	8.3	26	16	10.8	6.8	15.0	12	23
0.49	6.0	3.7	8.3	26	16	10.7	6.9	14.7	12	23
0.5	5.8	3.6	8.1	29	14	10.6	6.8	14.5	12	23

SKJ

SKJ	FOB					FSC				
CV	Mean	95CI inf	95CI up	Min_N sample	N square	Mean	95CI inf	95CI up	Min_N sample	N square
0.01	43.2	38.6	47.4	34	8	50.1	8.8	91.4	3	8
0.02	31.0	27.7	35.2	31	11	59.6	48.8	70.6	46	8
0.03	22.3	18.5	26.0	23	17	33.7	14.5	52.4	24	12
0.04	17.0	13.5	20.5	21	19	29.3	13.1	45.0	24	12
0.05	12.6	10.2	14.9	14	29	19.4	8.1	30.3	14	15
0.06	10.5	8.5	12.4	13	32	16.0	8.2	23.5	12	17
0.07	8.9	7.1	10.8	13	32	13.5	7.6	19.0	3	18
0.08	7.4	5.9	9.0	12	36	11.2	6.6	15.6	3	18
0.09	6.3	5.0	7.7	3	38	9.9	6.0	13.6	3	18
0.1	5.4	4.3	6.4	3	38	9.1	5.5	12.6	3	18
0.11	4.8	3.9	5.8	3	38	7.9	4.7	10.9	3	18
0.12	4.4	3.5	5.2	3	38	7.2	4.3	9.9	3	18
0.13	4.1	3.3	4.9	3	38	7.7	5.1	10.3	3	19
0.14	4.0	3.3	4.7	3	38	7.2	4.7	9.3	3	19
0.15	3.7	3.0	4.3	3	38	6.8	4.5	8.8	3	19
0.16	3.5	3.1	4.0	3	38	6.2	4.1	8.1	3	19
0.17	3.5	3.1	3.9	3	38	5.8	3.7	7.6	3	19
0.18	3.5	3.0	3.9	3	38	5.6	3.5	7.4	3	19
0.19	3.3	3.0	3.6	3	38	5.3	3.2	7.2	3	19
0.2	3.2	3.0	3.4	3	38	4.9	3.1	6.6	3	19
0.21	3.2	3.0	3.3	3	38	4.9	3.1	6.6	3	19
0.22	3.1	2.9	3.3	3	38	4.7	2.8	6.4	3	19
0.23	3.1	2.9	3.3	3	38	4.4	2.7	6.0	3	19
0.24	< 3					4.2	2.7	5.5	3	19
0.25	< 3					4.1	2.8	5.4	3	19
0.26	< 3					3.8	2.5	5.1	3	19
0.27	< 3					3.7	2.3	5.1	3	19
0.28	< 3					3.6	2.2	4.9	3	19
0.29	< 3					3.6	2.2	4.9	3	19
0.3	< 3					3.6	2.2	4.9	3	19
0.31	< 3					3.6	2.3	4.8	3	19
0.32	< 3					3.5	2.3	4.7	3	19
0.33	< 3					3.5	2.3	4.7	3	19
0.34	< 3					3.5	2.3	4.7	3	19
0.35	< 3					3.5	2.3	4.7	3	19
0.36	< 3					3.4	2.7	4.2	3	19
0.37	< 3					3.4	2.7	4.2	3	19
0.38	< 3					3.4	2.7	4.2	3	19
0.39	< 3					3.4	2.7	4.2	3	19
0.4	< 3					3.4	2.7	4.2	3	19
0.41	< 3					3.4	2.7	4.2	3	19
0.42	< 3					3.4	2.5	4.1	3	19

YFT	FOB					FSC				
CV	Mean	95CI inf	95CI up	Min_N sample	N square	Mean	95CI inf	95CI up	Min_N sample	N square
0.01						18.3	0.0	51.8	13	8
0.02			No match			14.8	0.0	43.6	15	9
0.03						39.2	5.7	72.6	19	11
0.04	44.5	38.0	50.4	39	6	48.2	16.2	80.0	26	11
0.05	42.3	36.7	47.9	39	6	49.4	8.7	89.4	39	9
0.06			No match			46.0	7.4	84.0	39	9
0.07	39.0	31.8	45.9	39	8	42.6	4.0	80.7	39	9
0.08	37.8	30.5	44.8	39	8	39.8	3.3	76.0	39	9
0.09			No match			36.9	5.9	64.9	34	10
0.1	32.7	25.4	39.9	39	8	34.1	5.0	60.2	34	10
0.11	29.2	20.6	37.9	34	9	33.9	0.7	66.6	39	9
0.12	27.2	18.8	35.5	34	9	31.9	0.0	64.0	39	9
0.13	27.0	18.5	35.4	34	9	27.9	2.6	50.8	34	10
0.14	24.4	16.2	32.4	34	9	26.5	2.5	48.4	34	10
0.15	23.7	14.9	32.6	33	10	24.9	0.0	46.3	34	10
0.16	19.8	12.5	27.0	26	16	23.9	1.0	43.1	34	10
0.17	18.6	11.2	26.0	26	16	21.8	0.9	40.6	34	10
0.18	16.8	11.0	22.8	26	16	22.6	0.0	46.9	39	9
0.19	17.8	10.6	25.1	31	11	19.0	0.0	37.1	30	11
0.2	17.9	10.1	25.8	33	10	18.5	0.5	34.8	34	10
0.21	15.9	8.9	22.8	31	11	18.2	0.0	34.7	34	10
0.22			No match			16.0	0.0	31.6	30	11
0.23	14.1	8.0	20.2	31	11	15.5	0.7	29.0	34	10
0.24	13.7	7.4	20.0	31	11	13.9	0.0	27.1	30	11
0.25	14.5	6.0	22.9	43	7	13.5	0.0	26.2	30	11
0.26	11.8	5.4	18.2	31	11	12.8	0.1	24.9	30	11
0.27	10.7	5.8	15.5	26	16	12.8	0.1	24.9	30	11
0.28	10.3	5.3	15.4	26	16	12.1	0.3	23.4	30	11
0.29	9.8	5.1	14.5	26	16	11.7	0.5	22.3	30	11
0.3	9.6	5.1	14.2	26	16	11.7	1.0	21.4	34	10
0.31	9.1	5.0	13.3	26	16	10.3	1.4	18.4	34	10
0.32	8.5	4.8	12.4	26	16	10.1	1.6	17.7	34	10
0.33	8.3	4.6	12.2	26	16	9.7	1.3	17.1	34	10
0.34	8.0	4.4	11.8	26	16	8.9	1.4	16.0	30	11
0.35	7.7	4.4	11.1	26	16	8.8	1.6	15.6	30	11
0.36	7.3	3.9	10.8	26	16	8.6	1.4	15.5	30	11
0.37	7.1	3.7	10.6	26	16	7.9	1.2	14.3	30	11
0.38	6.9	3.6	10.3	26	16	7.9	1.2	14.3	30	11
0.39	6.5	3.3	9.6	26	16	7.9	1.2	14.3	30	11
0.4	6.3	3.2	9.5	26	16	7.9	1.2	14.3	30	11
0.41	6.0	2.9	9.2	26	16	7.5	1.3	13.4	30	11
0.42	5.9	3.0	8.8	26	16	7.5	1.3	13.4	30	11
0.43	5.9	3.1	8.6	26	16	6.9	1.2	11.9	26	12
0.44	5.6	2.8	8.4	26	16	6.7	1.5	11.4	30	11
0.45	5.4	2.6	8.2	26	16	6.7	1.5	11.4	30	11
0.46	5.4	2.6	8.2	26	16	6.7	1.5	11.4	30	11
0.47	5.3	2.6	7.9	26	16	6.6	1.5	11.3	30	11
0.48	5.1	2.7	7.4	26	16	6.5	1.5	10.9	30	11
0.49	4.8	2.7	6.9	26	16	6.4	1.6	10.7	30	11
0.5	4.6	2.7	6.5	26	16	6.1	1.6	10.1	30	11

Appendix 4: Frequency and density in catch per sets and per school type for all sets and sampled sets only from 2015 to 2017



FRAMEWORK CONTRACT - MARE/2016/22 « Strengthening regional cooperation in the area of fisheries data collection » Annex III « Biological data collection for fisheries on highly migratory species »

Strengthening Regional cooperation in the area of large pelagic fisheries data collection (Acronym: RECOLAPE)

Task 2.4 - Simulation of the regionalized sampling plan

D2.4 - Development of a regionalized sampling plan - Case study: Mediterranean swordfish

Responsible: George Tserpes (HCMR)

Index

EXECUTIVE SUMMARY 3
1. Conceptual approach 4
2. Data availability 4
3. Bootstrap analysis 5
3.1 Eastern Levantine - Cypriot fisheries 6
3.2 Aegean and Cretan seas - Greek fisheries 8
3.3 GFCM/GSA 15 - Maltese fisheries 10
3.4 Central Mediterranean - Italian fisheries 12
4. Conclusions - proposed vs current sampling 14
5. References 17

EXECUTIVE SUMMARY

The current document deals with the optimization of regionalized sampling schemes for the Mediterranean swordfish based on data from four national longline fisheries (Cyprus, Greece, Malta, Italy), exploiting different Mediterranean regions. Estimates of optimal sampling rates have been obtained by means of a simulation approach and take into account the data needs and priorities that were defined under Deliverable 2.1, based on the specific recommendations of the International Commission for the Conservation of Atlantic Tunas (ICCAT).

The analysis identified the sampling levels needed to ensure sampling quality standards by national fishery, exploited region and year-quarter. Results showed that, depending on the size of the exploited area and season, a sample size of 60-120 individuals per year-quarter would provide sufficient levels of precision. Generally, as "rule of thumb" it could be suggested that quarterly sampling of about 70-100 samples is sufficient, at the GFCM/GSA level. However, this estimate is based on size measurements and ignores other biological parameters, such as sex ratio and maturity stage.

The suggested sampling levels are generally not in agreement with the current sampling schemes that seem to be exclusively based on the landings' volume. As both under and over sampling rates are currently observed, depending on the fishery and season, optimization of the temporal allocation of sampling effort is needed.

1. Conceptual approach

The quality of the biological sampling schemes employed by the different member states in the frame of the data collection framework (DCF) is commonly assessed through estimates of the coefficient of variation (CV) values of the sampled fish lengths. Typically, CV is decreasing with increasing sample size, but this decreasing trend is not linear and slows down at large sample sizes. The objective is to estimate the trade-off between sample number and precision, i.e. the number of samples after which the gains in precision are not important. As the sample size-CV curve tends to flatten after a certain point, the optimal sampling range is indicated by the flattening of the curve and it is assumed that this empirical range ensures acceptable sampling quality. In the present case, this approach is followed in order to provide quantitative information on the sample size required to meet the data needs and priorities that were defined under Deliverable 2.1, based on the specific recommendations of the International Commission for the Conservation of Atlantic Tunas (ICCAT) for the Mediterranean swordfish.These recommendations mention that biological data should be collected on a high spatiotemporal resolution (i.e cover local fisheries on a monthly/quarterly basis).

2. Data availability

The data used concerned the size distribution of swordfish landings of different national fisheries exploiting different Mediterranean regions. Sizes were expressed in terms of Lower Jaw Fork Length (LJFL) and were aggregated by fishing trip and date. Although the data were provided in the SDEF format, they did not fully meet the protocol requirements as this format is not used for submission to ICCAT, which is the responsible Regional Fisheries Management Organization (RFMO). The available data covered the longline swordfish fisheries of Cyprus, Greece, Italy ans Malta for the period 2015-2017. Data were lacking harmonization in terms of reported size-interval and spatial resolution. Some of them were reported in 5ºx5º squares (Greece, Cyprus), while others by GFCM Geographic Sub Area - GSA (Malta) or with geographical coordinates indicating major areas but not geographical squares (Italy). However, there was no spatial overlap between observations from the different national fisheries. Another issue was that the number of fishing operations accomplished within each fishing trip was highly variable (ranging from 1-60) and that size data were not available by fishing operation. As such, it was not possible to simulate the impact of the number of sampled trips on the CV of fish lengths and only explicit changes on the sample size, expressed in terms of numbers of individuals, could be simulated.

3. Bootstrap analysis

The relationship between the sample size and the precision was examined by means of a bootstrap approach, in line with that followed in Deliverable 2.5 of the MARE 2014/19 project. Similar to that, 100 sub-samples were created for each given sample size and the CV was computed for each sub-sample. This allows to estimate the variability of the precision expected from a given sample size. Plotting CV against sample size allows to identify in an empirical way, the optimal number of samples needed to ensure sampling quality.

As the approach that was followed relies on re-sampling real data, data-sets of sufficient size are necessary in order to generate oversampling. Although using small-size data-sets is technically possible, this could provide biased results because re-sampling a small data-set results in the re-use of the same sub-samples for several times; thus variance is underestimated. In such cases, the “CV versus number of samples” curves would not help in evaluating the sampling scheme, as the samples are very quickly recycled within the bootstraps.

Preliminary trials suggested that a minimum of 50 observations per case was necessary for performing the analysis, which was realized separately for each national fishery/region. Given that the fisheries exploitation pattern varied between years, and fishing activities were not homogeneously distributed throughout any given year, there were several months with very few samples that could not included in the analysis. Hence, it was considered suitable to aggregate the data on a quarterly basis. In order to account for inter-annual variations and sampling deficiencies, data from all available years were grouped together.

The discrepancies regarding the number of fishing operations by trip, as well as the lack of standardized reporting of length frequencies that were following different size intervals did not allow to apply existing tools (e.g. COST); thus new scripts in R language (R Core Team 2019) were developed. Bootstrapping was accomplished using the R package “boot” (Canty & Ripley, 2017).

3.1 Eastern Levantine - Cypriot fisheries

The available data were refering to fisheries exploiting the 5°x5° square extending from 30-35°E and 30-35°N. Figure 1 shows the size frequency of the samples by quarter. The CV vs sample-size curves show that the optimum sampling size is about 60-70 individuals per quarter and this sampling level produces CV rates smaller than 10% (Figure 2).

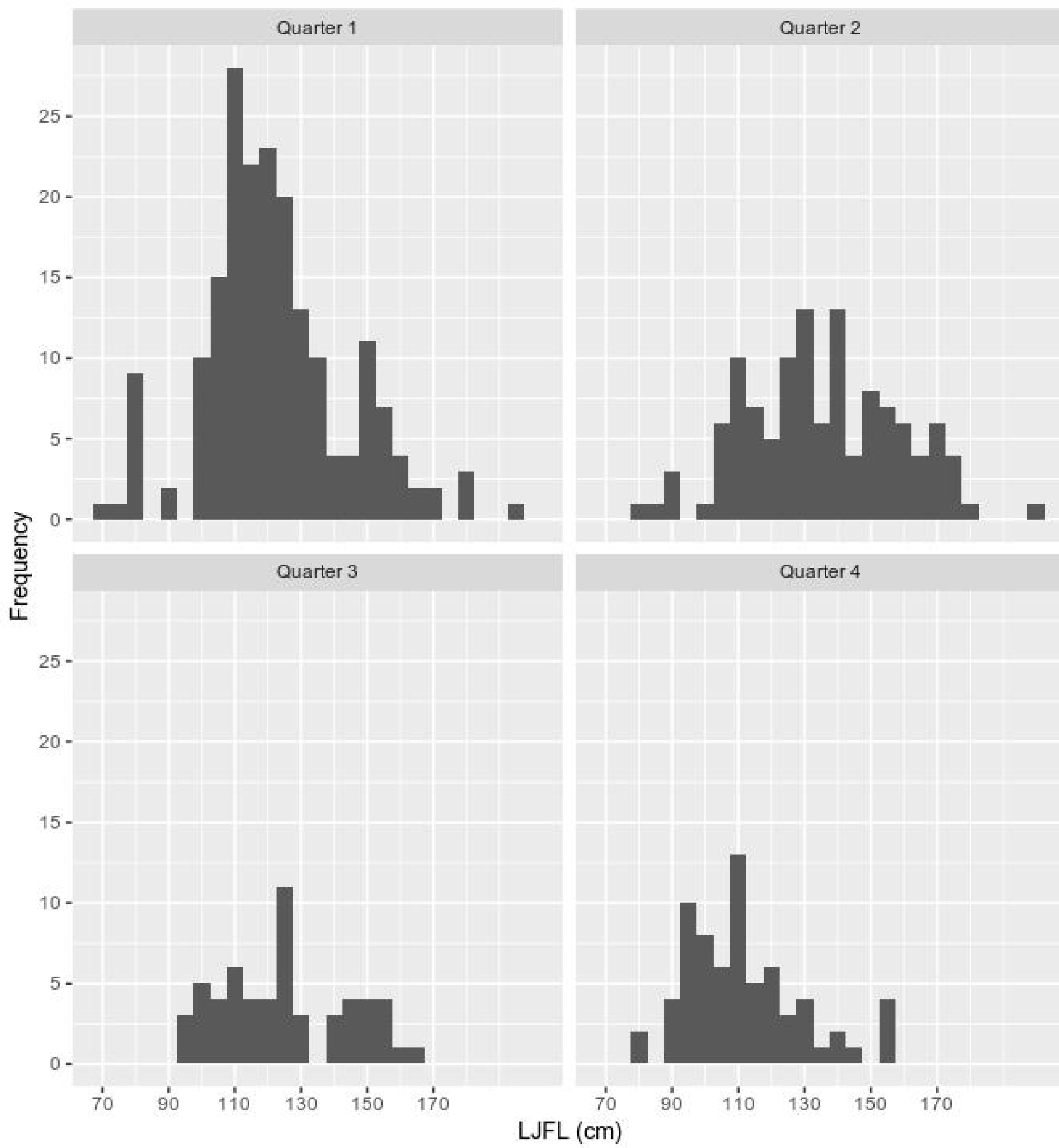


Figure 1: Sample frequency histograms by year quarter.

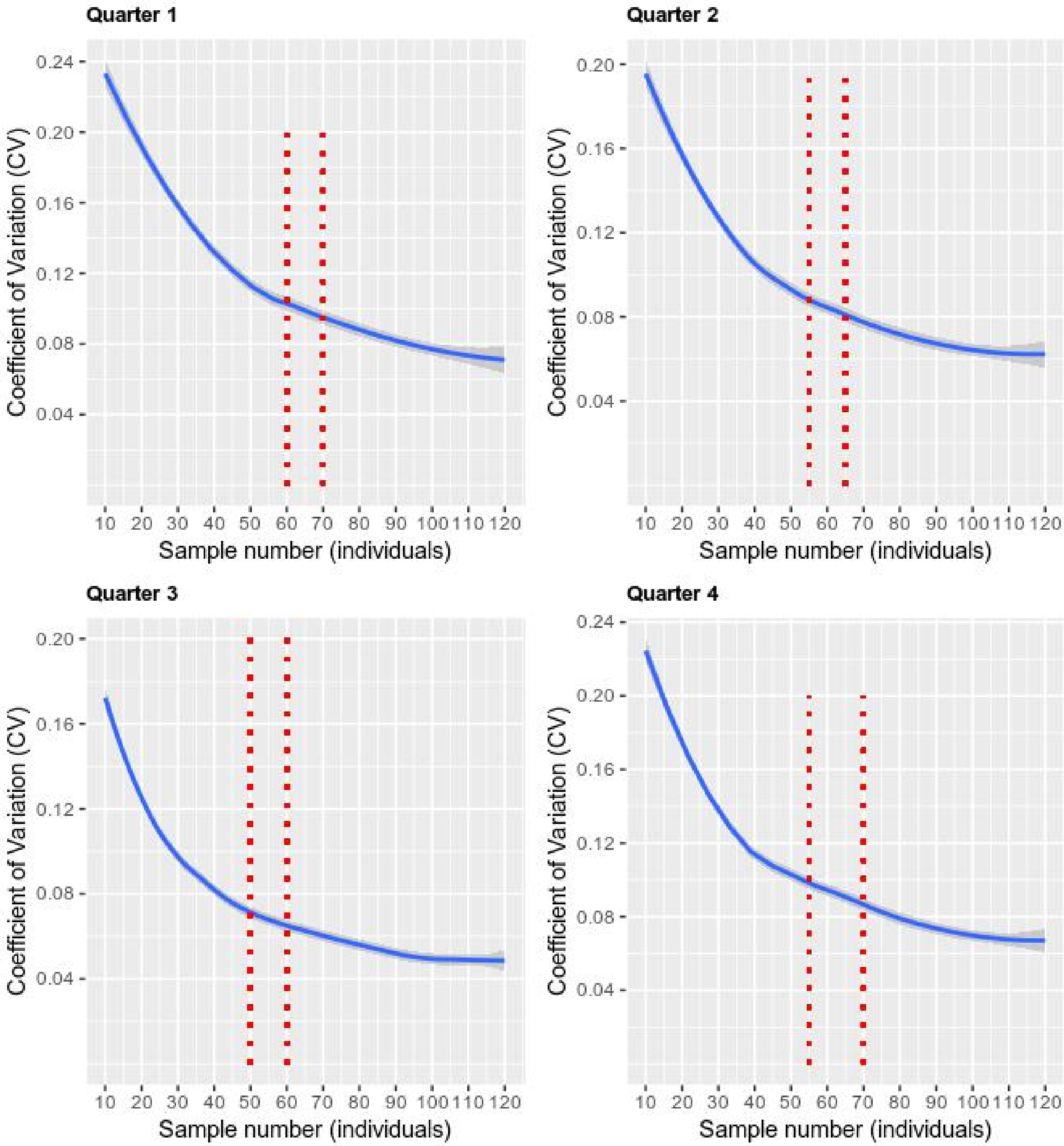


Figure 2: Coefficient of variation against sample size derived from bootstrapping quarterly aggregated data. Vertical lines define the optimal sampling range.

3.2 Aegean and Cretan seas - Greek fisheries

The available data covered the the fisheries exploiting the 5ºx5º square extending from 20-25ºE and 35-40ºN (mainly Aegean and Cretan seas). Figure 3 shows the size frequency of the samples by quarter. The CV vs sample size curves show that the optimum sampling size is about 80-100 individuals per quarter. This sampling level corresponds to CV rates smaller than 10% (Figure 4).

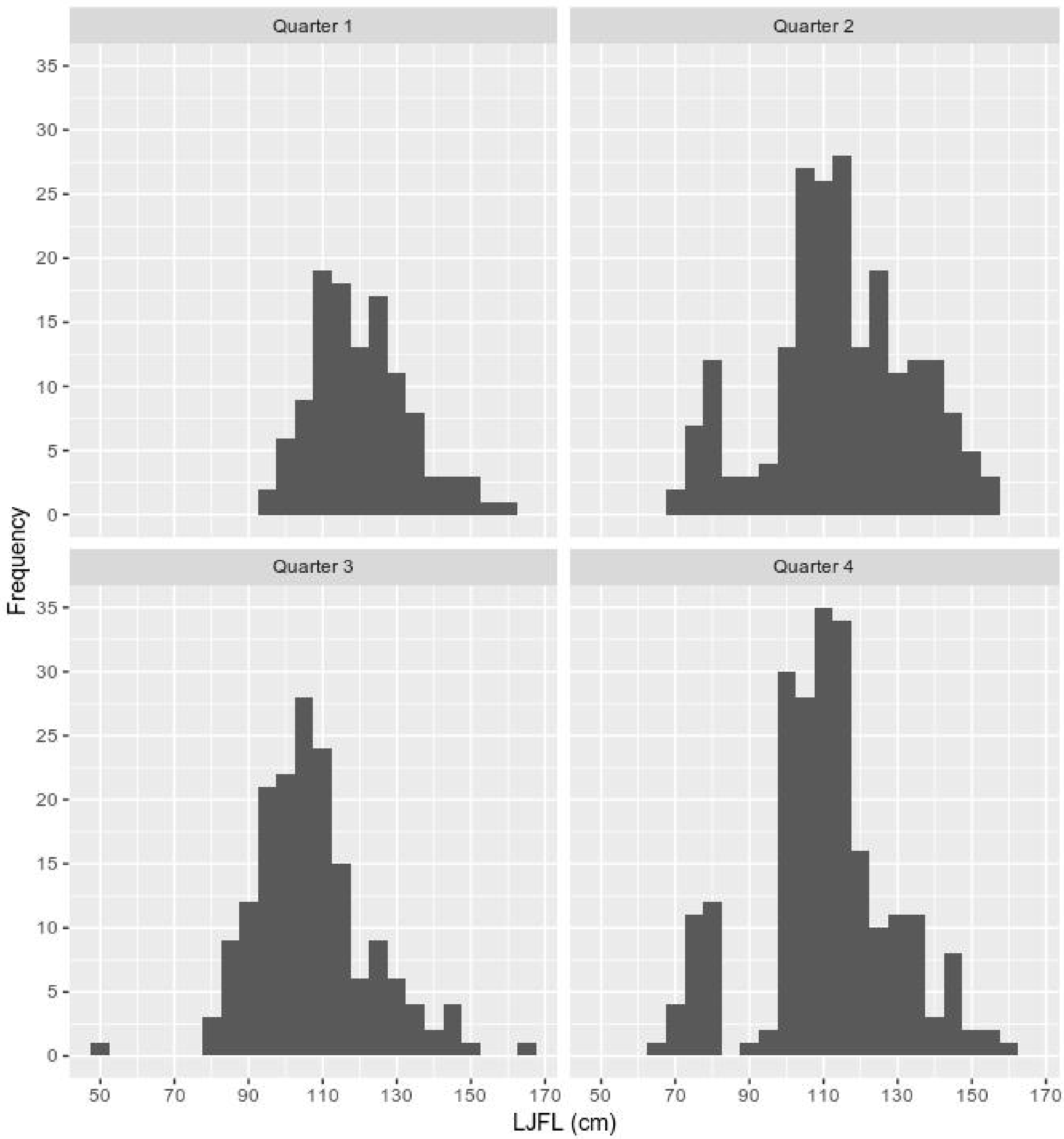


Figure 3: Sample frequency histograms by month.

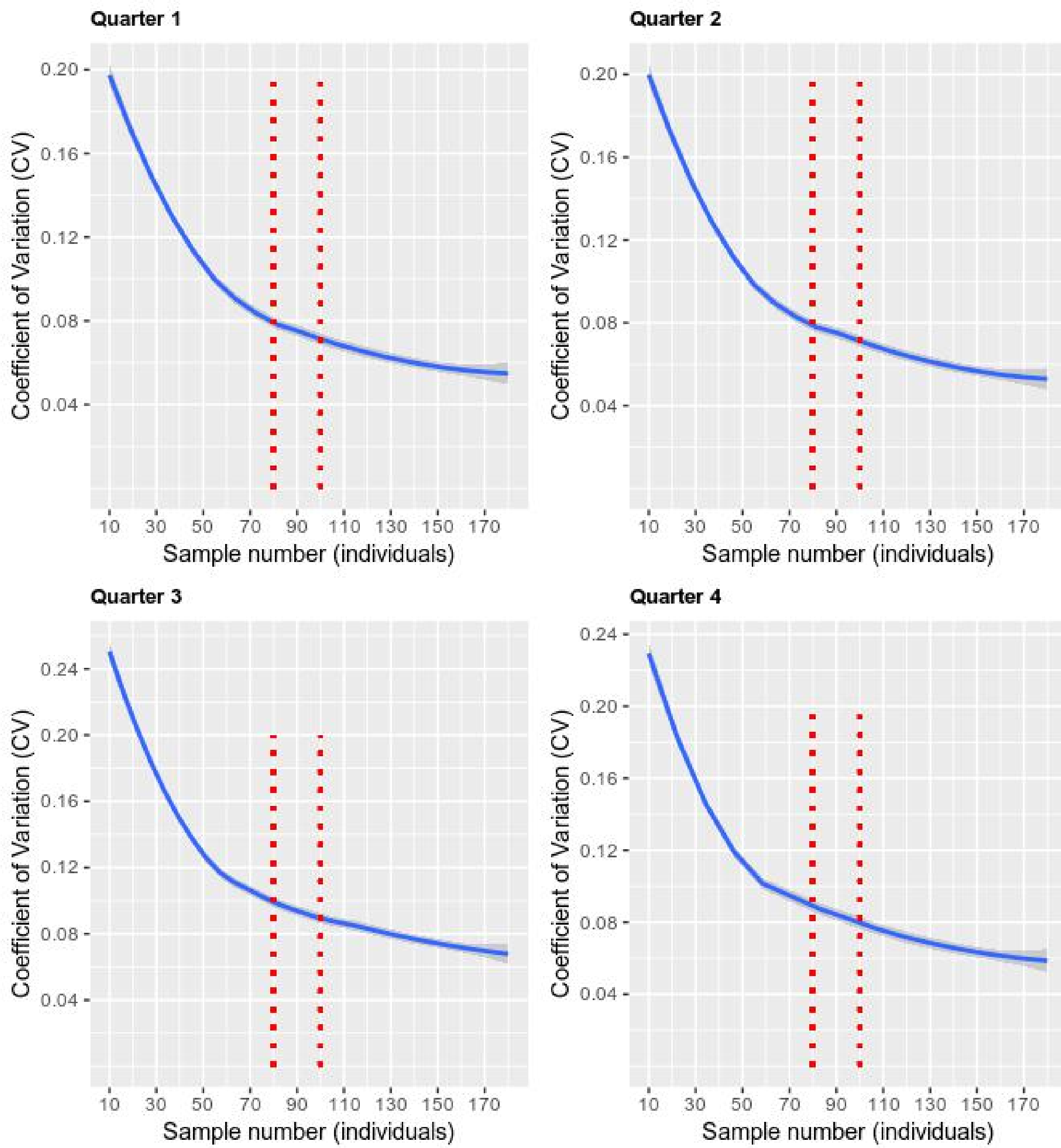


Figure 4: Coefficient of variation against sample size derived from bootstrapping quarterly aggregated data. Vertical lines define the optimal sampling range.

3.3 GFCM/GSA 15 - Maltese fisheries

The available data were refering to fisheries exploiting the GSA 15 and Figure 5 shows the size frequency of the samples by quarter. The second and fourth quarter of the year were excluded from the analysis due to lack of sufficient data. The CV vs sample size curves show that the optimum sampling size is about 60-80 individuals for the first quarter and this sampling level corresponds to a rather low CV rate (~7%). For the third quarter however precision remains low (CV>20%), even at sampling levels up to 120 individuals. Considering a trade-off between sampling intensity and precision a sample number of 85-100 individuals is suggested in this case (Figure 6).

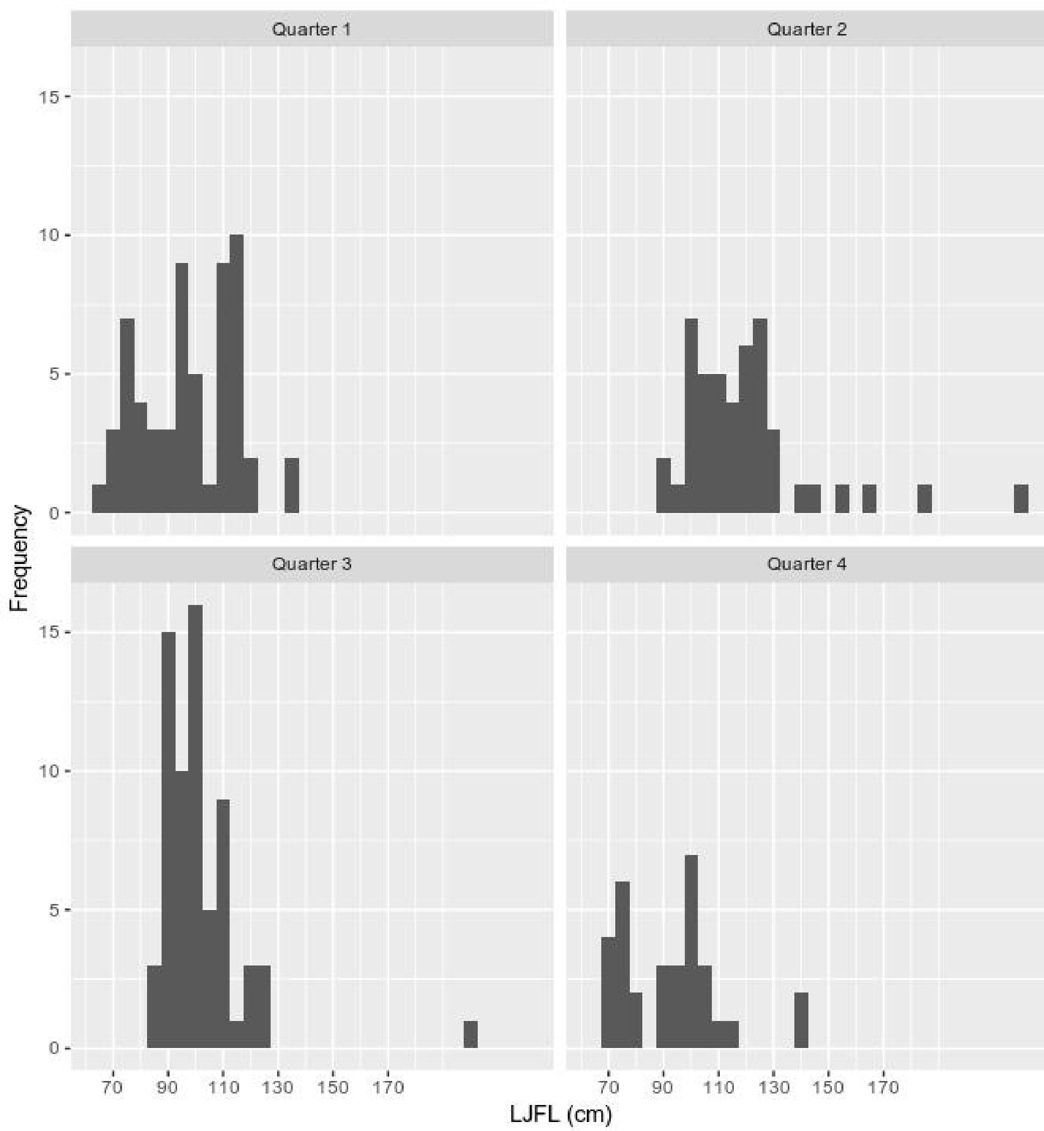


Figure 5: Sample frequency histograms by month.

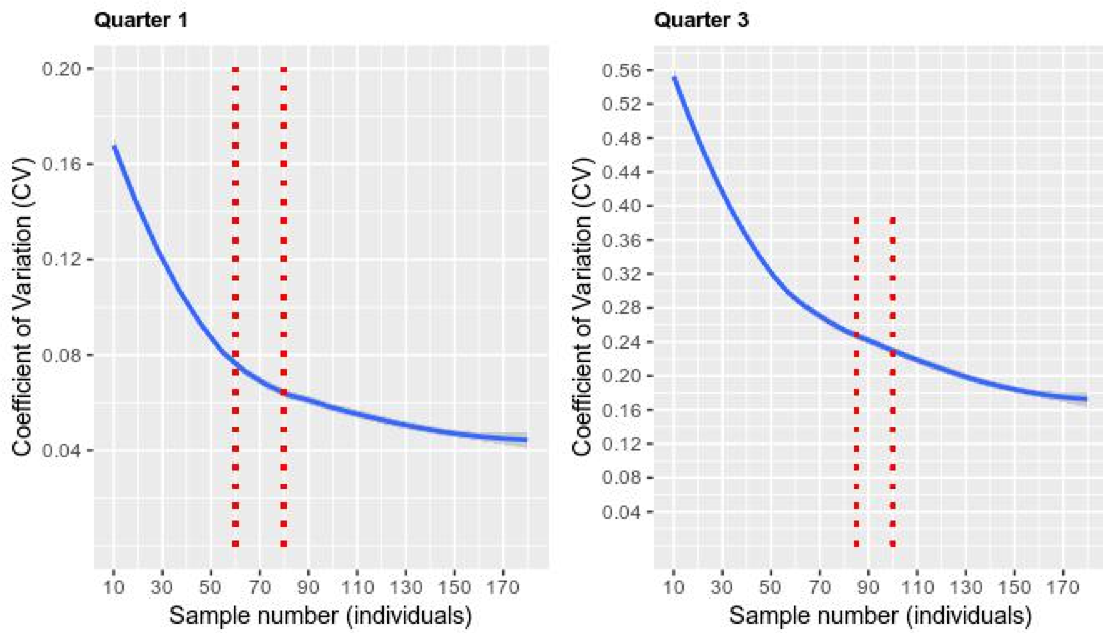


Figure 6: Coefficient of variation against sample size derived from bootstrapping quarterly aggregated data. Vertical lines define the optimal sampling range.

3.4 Central Mediterranean - Italian fisheries

The data were refering to fisheries exploiting the Ligurian, Tyrrhenian, W. Ionian and Adriatic seas but the available information was insufficient for grouping the samples by specific area. Figure 7 shows the size frequency of the samples by quarter. As size measurements were reported in 5cm intervals, previously to the analysis they were classified in 1cm intervals through a randomization function in R language. The CV vs sample size curves show that the optimum sampling size ranges from 90-120 individuals per quarter. This sampling level corresponds to CV rates less than 10% (Figure 8).

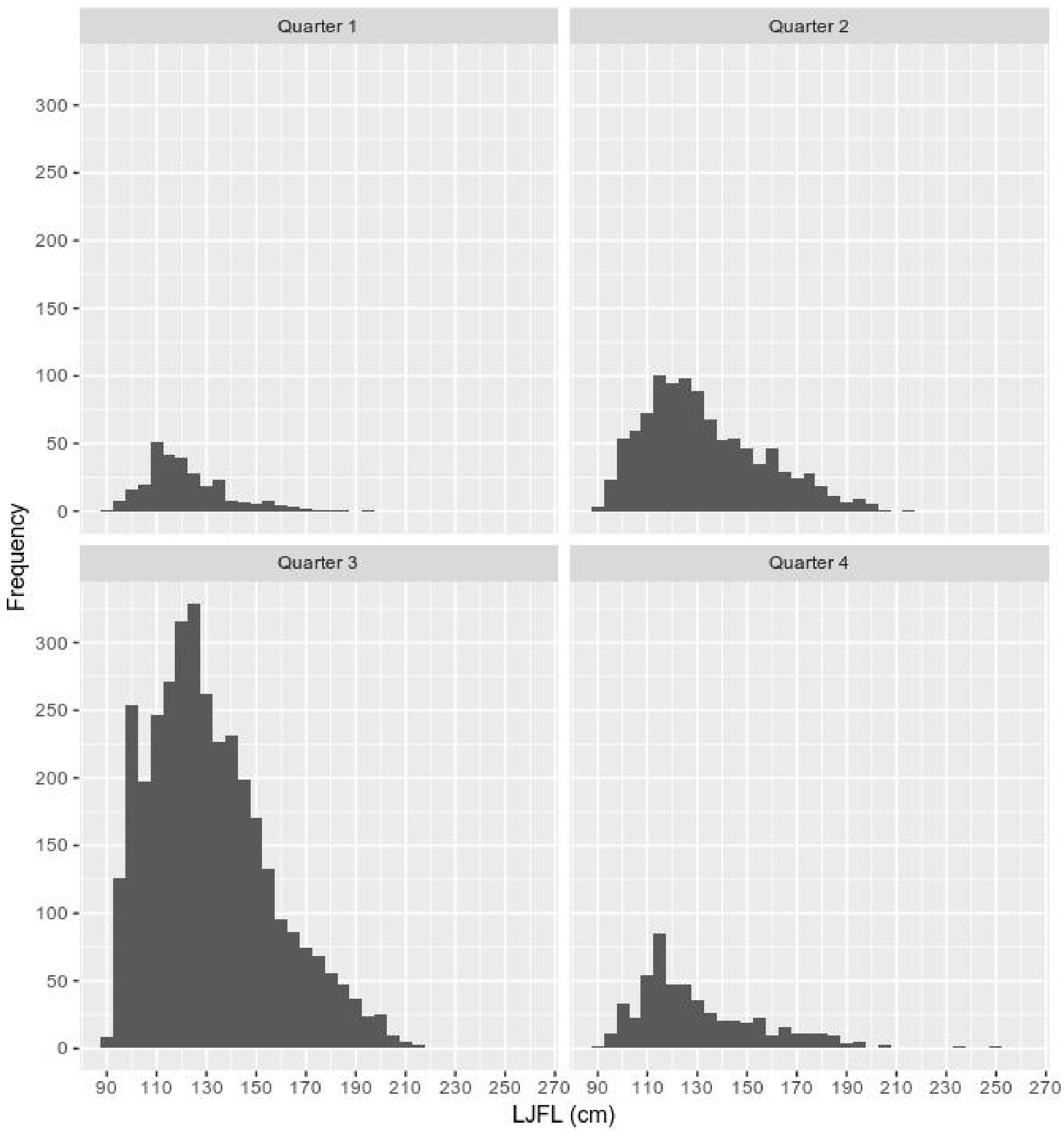


Figure 7: Sample frequency histograms by month.

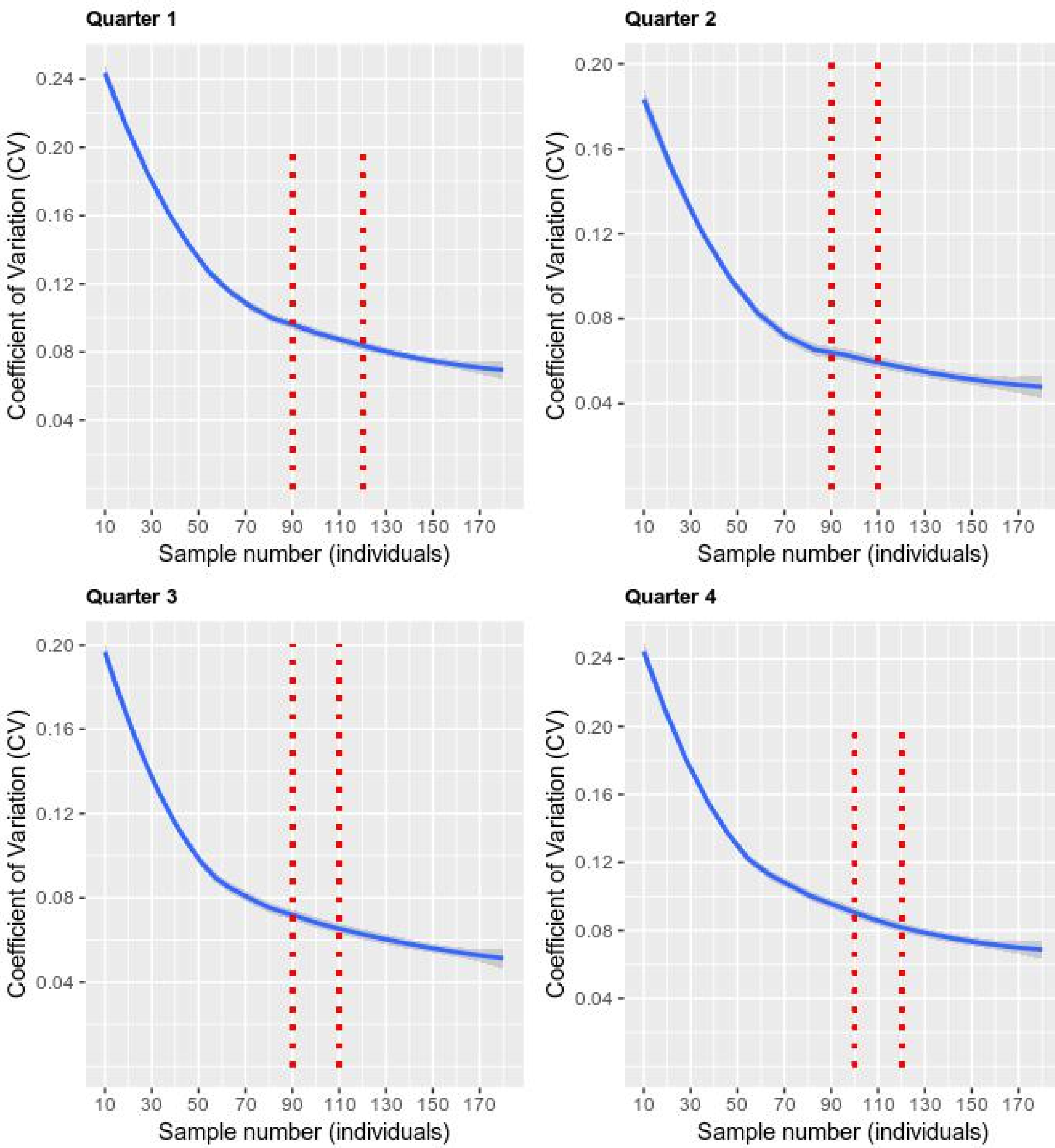


Figure 8: Coefficient of variation against sample size derived from bootstrapping quarterly aggregated data. Vertical lines define the optimal sampling range.

4. Conclusions - proposed vs current sampling

The bootstrapping approach allowed to estimate optimal sampling rates that would allow to meet the requirements identified in Deliverable 2.1. Though the available data did not cover the whole Mediterranean, they included information from several fisheries exploiting different parts of the basin and provided useful information on the sampling frequency requirements by fishery/region. As it would be expected, the larger the exploited area the higher the number of samples required to ensure quality sampling, given the migratory movements of the fish during its life-cycle. Generally, as a "rule of thumb", it could be suggested that quarterly sampling of about 70-100 samples per GSA provides sufficient precision levels (CV<10%). However, this estimate is based on size measurements and ignores other biological parameters, such as sex ratio and maturity stage. It also ignores the size composition of discards, as such information was not available.

Tables 1-4 compare the current sampling rates with the optimal ones for the years available in the current study. In general, the suggested sampling levels are not in agreement with the current schemes that seem to be exclusively based on the volume of landings. As both under and over sampling rates are currently observed, depending on the fishery and season, optimization of the temporal allocation of sampling effort is needed. The highest oversampling rates were found for the Italian fisheries but in this case findings may have differed if the data allowed to consider sub-regions of the examined large area. Taking into account the existing ecological differences among different Mediterranean areas, ICCAT reccomended the colection of biological data at a regional scale, which however is not precisely clarified (ICCAT, 2019). Apart from data reporting discrepancies, the lack of spatial overlapping in the exploitation patterns of the examined national fisheries did not allow to develop regional sampling schemes commonly coordinated by different member states. It should be noted, however, that harmonized data reporting in standardized formats is crucial for the potential future development of coordinated sampling schemes in commonly exploited regions, as well as for stock assessment purposes.

Table 1: Current and optimal sampling rates by year and quarter for the Cypriot fisheries.

Year	Quarter	Sample.size	Optimum	Coverage
2015	1	156	60-70	~240%
2015	2	80	60-70	~123%
2015	3	37	60-70	~56%
2015	4	19	60-70	~29%
2016	1	35	60-70	~53%
2016	2	24	60-70	~36%
2016	4	50	60-70	~76%
2017	1	1	60-70	Temporal closure
2017	2	13	60-70	~20%
2017	3	19	60-70	~29%

Table 2: Current and optimal sampling rates by year and quarter for the Greek fisheries.

Year	Quarter	Sample.size	Optimum	Coverage
2015	1	114	90-100	~120%
2015	2	151	90-100	~158%
2015	3	55	90-100	~57%
2015	4	162	90-100	~170%
2016	2	57	90-100	~60%
2016	3	113	90-100	~118%
2016	4	60	90-100	~63%

Table 3: Current and optimal sampling rates by year and quarter for the Maltese fisheries

Year	Quarter	Sample.size	Optimum	Coverage
2015	1	35	60-70	~53%
2015	2	20	Non defined	Non defined
2015	3	31	85-100	~32%
2015	4	15	Non defined	Non defined
2016	1	24	60-70	~36%
2016	2	22	Non defined	Non defined
2016	3	25	85-100	~26%
2016	4	8	Non defined	Non defined
2017	2	4	Non defined	Non defined
2017	3	10	85-100	~10%
2017	4	9	Non defined	Non defined

Table 4: Current and optimal sampling rates by year and quarter for the Italian fisheries

Year	Quarter	Sample.size	Optimum	Coverage
2015	1	239	90-120	~227%
2015	2	263	90-120	~250%
2015	3	1118	90-120	~1064%
2015	4	325	90-120	~309%
2016	1	46	90-120	~43%
2016	2	533	90-120	~507%
2016	3	1849	90-120	~1760%
2016	4	95	90-120	~90%
2017	1	2	90-120	Temporal closure
2017	2	231	90-120	~220%
2017	3	541	90-120	~515%
2017	4	116	90-120	~110%

5. References

Canty, A. & Ripley R., 2017. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-20.

ICCAT, 2019. Report for biennial period, 2018-19 PART I(2018)-Vol. 2, SCRS. URL: http://www.iccat.int/Documents/BienRep/REP_EN_18-19_I-2.pdf.

R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

FRAMEWORK CONTRACT – MARE/2016/22 "Strengthening regional cooperation in the area of fisheries data collection" Annex III "Biological data collection for fisheries on highly migratory species"

Project acronym: RECOLAPE

Task 2.5 – Propose guidelines for data storage/management/analysis solution at regional level

Participating partners:

D.2.5 – Guidelines for data storage/management/analysis solution at regional level.

Deliverable coordinated by Ioannis Thasitis (DFMR), Jon Ruiz (AZTI)

and Mathieu Depetris (IRD)

Index

1. Introduction 3
2. Objectives 4
3. Fisheries description 5
3.1. Tropical tuna purse seine fishery in the Atlantic Ocean 5
3.2. Mediterranean Swordfish (EU Cyprus) 7
4. Sampling description 9
4.1. Port sampling in Africa (Abidjan and Dakar) 9
4.2. Cyprus Swordfish sampling 12
4.2.1. On-board 12
4.2.2. On-shore 13
4.2.3. Remarks applied to for both cases (on-board & on-shore) 13
5. Populating RDBES data model for tropical tuna sampling 14
6. Final conclusino and remarks 24
7. References 26

1. Introduction

This deliverable is part of the project MARE/2016/22: "Strengthening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species" (acronym RECOLAPE). The overall objective of the project is to strengthen the regional cooperation of large pelagic fisheries biological data collection. This endeavour will be valuable in improving the coordination among European Member States (MS), and subsequently the efficiency in the fisheries data collection field supporting better stock assessment quality and advice formulation. At the same time, it seeks to provide solutions to certain needs already identified by scientists involved in the tuna RFMOs' stock assessment and by the RCG-LP (Regional Coordination Group – Large Pelagic).

Within RECOLAPE project, Work Package (WP) 2 aims to make a proposal for a Regional Sampling Plan (RSP), which will replace the relevant parts of the MS' National sampling Work Plans. It includes two case studies (Mediterranean swordfish and tropical tuna purse seine fishery in the Atlantic Ocean) and deals with all the necessary elements to conduct an RSP; definition of data needs, data sharing protocols, development of an agreed sampling protocol, regional sampling design and solutions to store the data at regional level. This deliverable proposes concrete solutions to this last point, identifying the elements of a system that allows to store the large pelagic stock data at regional level.

In this context, during the 2016 RCM Med&BS-LP meeting the LP subgroup recommended expanding the scope of the Regional Data Base FishFrame (RDB) to Large Pelagic specificities. RDB, which is hosted and maintained by the International Council for the Exploration of the Sea (ICES) Secretariat, is a regionally coordinated database platform for fisheries assessments. Funds for hosting and maintenance of the RDB/RDBES come from the Grant Agreement between the European Commission (EC) and ICES. It addresses fishery management needs related to the European Union Common Fisheries Policy and covers fisheries in the scope of the RCG North Sea and Eastern Arctic, the RCG North Atlantic and the RCG Baltic Sea. Thus, having a single system hosting the data from different regions (including LP fisheries) appeared as the best solution in term of harmonisation of procedures, and from a cost-benefits point of view. The RDB would be also a tool that would allow improving the standardisation of quality among MS, and it also would allow responding to different end user's data calls without extra work. Thus, RCM-LP (2016) recommended to include EU-LP fisheries data in the RDB. This would imply the inclusion of specific fields and codes of interest for the LP fisheries. Furthermore, this option was in line with the expressed interest from the DG MARE International Directorate to support the development of a database holding the data for fisheries in non-EU waters. Thus, both DG MARE and RCM-LP subgroup supported the inclusion of data relevant to the RCM-LP into the RDB. However, despite this recommendation appeared in 2016, discussions held during the RCG-LP meeting in 2017, showed that before taking a clear decision some aspects require to be further discussed.

Thus, during 2017 the RCG-LP did not advance in the concept of any system to store data on highly migratory species.

In 2017 a new development of the RDB started, with the aim of including information about sampling design and quality. Thus, in the new Regional Data Base, RDBES (Regional Data Base and Estimation System), statistically sound sampling information can be quality checked and stored and statistical sound estimations can be executed and documented. It is able to handle so far 32 different ways of sampling, and aims to:

- Support the Regional Coordination Groups with harmonised data for coordination.
- Improve data quality by using common quality checks across all countries' data (at national and regional level)
- Automatically deliver data for different data calls. Including tuna RFMOs (Regional Fisheries Management Organizations)
- Ensure that the estimation methods used are transparent and documented
- Document data submissions to the RDBES
- From a cost benefit point of view, it is beneficial to develop and maintain of one system compared with many similar systems

In this context, and in view of the advantages of being part of this process (listed above), the RCG-LP 2018 participants decided to move forward and to check if the data for specific LP stocks could be inserted into the current RDBES data model. This document shows main findings resulted from this exercise of testing RDBES data model to host the Mediterranean swordfish and tropical tuna purse seine sampling data.

2. Objectives

The present deliverable D.2.5 deals with the solutions for data storage, management and analysis at regional level. It aims specifically to check the feasibility of the RDBES to host sampling data for LP stocks, and concretely to host the data for the two study cases included in the project; Mediterranean swordfish and tropical tuna purse seine. The specific objectives are listed below:

- Find out which RDBES data model's Hierarchy[1] fits better European tropical tuna sampling in west Africa ports (Abidjan/Dakar) and Mediterranean swordfish sampling in Cyprus.
- Fill in the tables in the selected Hierarchy with sample data.
- Identify gaps/extra fields needed/find gaps in master tables (areas, species…)

[1] ICES. 2018. Report of the Workshop on new data model for the Regional Database (WKRDB-MODEL), 15 – 18 January 2018, ICES HQ, Copenhagen, Denmark. ICES CM 2018/ACOM:41, 44 pp.

3. Fisheries description

3.1. Tropical tuna purse seine fishery in the Atlantic Ocean

The fishery: Purse seine boats fishing for major tropical tuna (skipjack tuna, yellowfin tuna and bigeye tuna) now make up a very modern fleet that is constantly developing in terms of both size and in their fishing technology and techniques. It is the surface gear that contributes most to the catch of yellowfin and skipjack globally (Majkowski *et al*., 2011). In the purse seine fishery, two main fishing strategies are used to capture tunas: (1) targeting fish swimming in free schools, (2) targeting fish swimming around drifting floating objects (FOBs). In the first approach, called a free-school set, a school of fish is identified from evidence in the water's surface, as presence of birds, and it is captured by encircling it. In the second approach, a drifting object, where fish are aggregated, is encircled with the net. Within this second strategy, there are a subset of techniques including sets on natural floating objects (log on animal, ANLOG, or vegetal origin, VNLOG) and sets on artificial log resulting from human activity, related to fishing activities (FALOG) or not (HALOG). Table 1 included in annex 1 presents the FOB codification based on ICCAT Recommendation 16-06. Additionally, DFADs (Drifting fish aggregating devices) are manmade floating objects placed in the fishing areas by the fishers to attract fish, and to facilitate their aggregation and capture. Additionally, DFADs are often outfitted with a satellite buoy to help fishers locate them. The strategy of using this kind of floating objects was developed in the 1980s, but greatly increased in use during the 1990s, and is currently responsible of the major component of the purse seine bycatch and discards (Hallier and Parajua, 1992; Fonteneau *et al*, 2000; Davies *et al*, 2014). The composition by species and sizes of this type of floating objects fishing varies a lot with respect to the traditional one on free schools, catching mainly SKJ and increasing the proportion of YFT and BET of smaller size.

21 vessels compound the EU purse seine fleet targeting tropical tunas in the tropical Atlantic Ocean, 11 Spanish and 10 French. There are also several associated flag vessels. Average length is around 80 meters and average carrying capacity around 1.000 tones. Figure 1 shows the fishing area, situated between latitudes 30º N and 15º S, comprising most of the fishing activity in the eastern area. Each colour or box in the map represents a different sampling area (ET areas described in Pallares *et all*, 1997). Fishing trips last between 3 and 7 weeks, where different areas could be visited, as well as different fishing techniques used (free-school sets and sets around floating objects objects). Retained catches are frozen in brine and stored in wells/tanks (number of wells per vessel could vary between 12 and 20 based on the total length). After each fishing operation the catch is stowed, without sorting by species, in one or several well(s) depending on the amount captured and wells capacity. In the same way, a single well may contain capture of different fishing operations if the quantity captured per set is less than the total capacity of the wells. Then, retained catches are landed in African ports; Adibjan (Ivory coast) is the main port (around 80% of landings), while Dakar (Senegal) may be relevant mainly

during a specific time period. There are also some secondary ports (e.x. Mindelo (Cape Verd) and Tema (Ghana)).

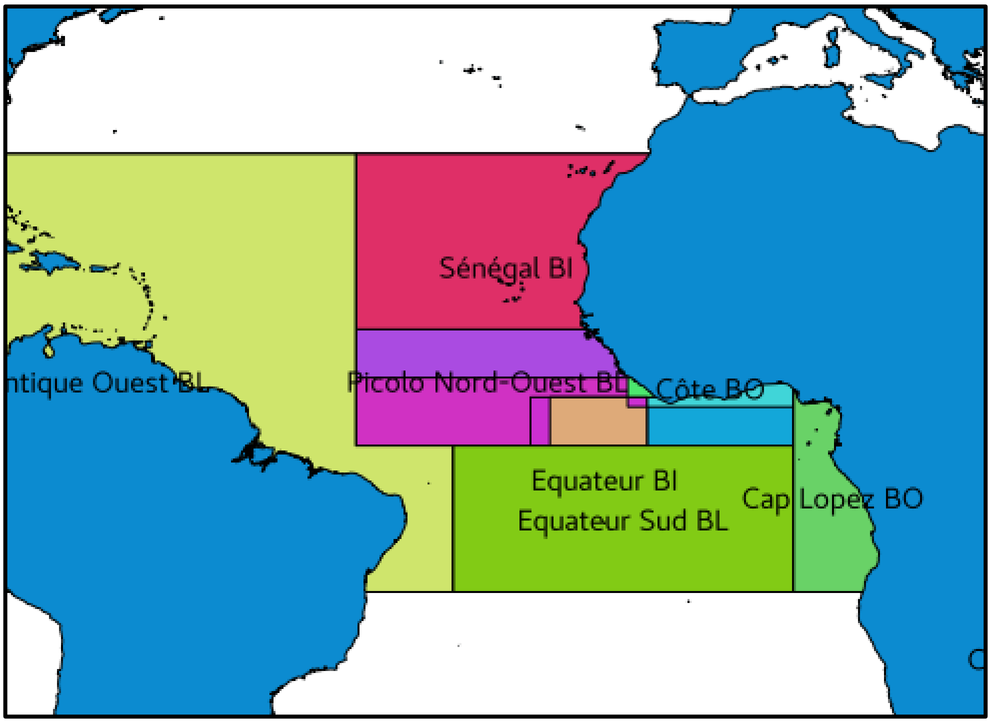


Figure 1. Fishing ground and ET sampling areas (ET areas described in Pallares et all, 1997).

Target species and catch volume per species: Three tropical tuna species are targeted, bigeye (*Thunnus* obesus, BET), skipjack (*Katsuwonus pelamis*, SKJ), and yellowfin (*Thunnus albacares,* YFT). For management purposes, a single stock is considered in the Atlantic Ocean for BET and YFT, while two stocks are considered for the SKJ (the eastern and western stocks). The contribution of each species within the Atlantic Oceans is shown in the figure 2. Based on the landings in 2017, the purse seine EU fleet is responsible for the 52%, 37% and 13% of the total YFT. SKJ and BET catches in the Atlantic Ocean respectively. Since 2003 the relative contribution of SKJ has been greater than YFT and BET comprising the 50 % of catches in comparison to 30 % of catches of YFT and 20% BET in 2010.

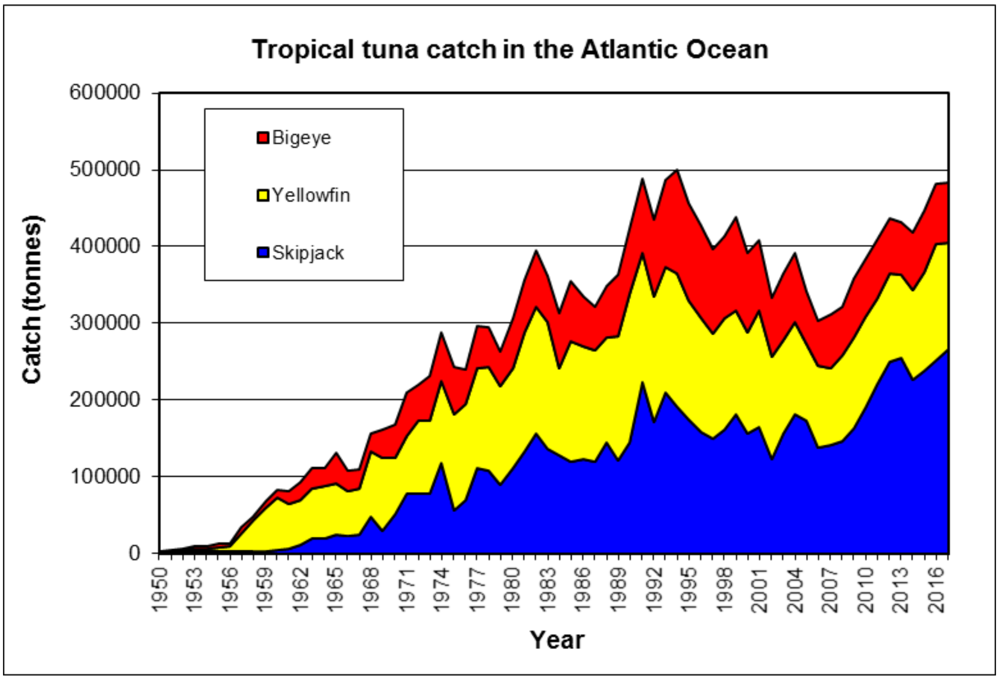


Figure 2. Tropical tuna catches in the Atlantic Ocean, by species for skipjack, yellowfin and bigeye tunas from 1950 to 2017. (Source: AZTI)

3.2. Mediterranean Swordfish (EU Cyprus)

In Cyprus large pelagic species are targeted by the polyvalent fleet. Polyvalent vessels have an overall length between 12-24 m and are engaged in two fisheries. Most part of their effort is allocated in the large pelagic fishery using drifting longlines and operating around Cyprus waters and the Eastern Mediterranean area GSAs 22,24,25,26 (Figure 3). During closed periods and/or seasons with low abundance of large migratory species they operate in the inshore demersal fishery using mostly bottom set nets and bottom longlines. Around 35 vessels are licenced for this segment annually. The fleet conforms with closed seasons, restrictions on the use of gears and landing sizes, in accordance with national and community regulations.

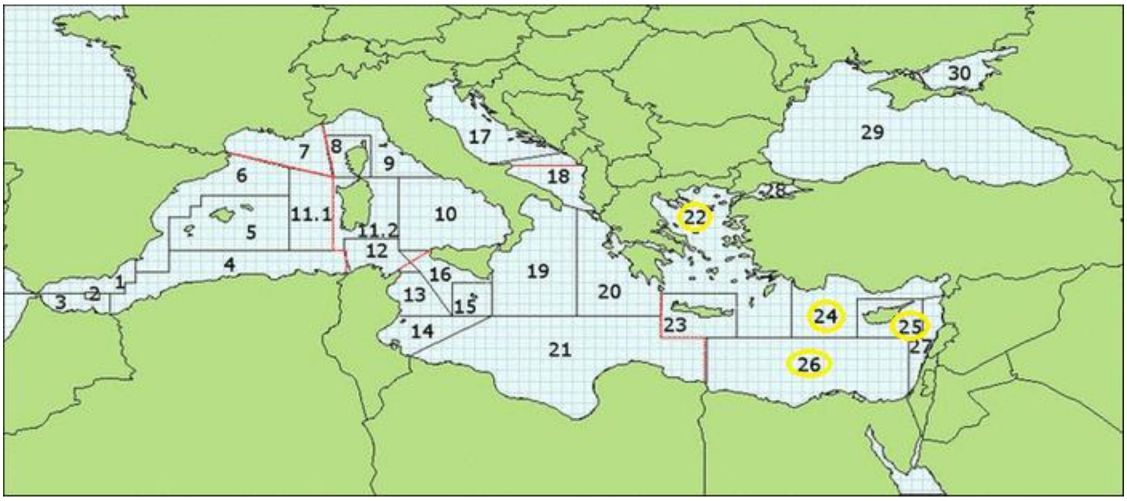


Figure 3. Cyprus polyvalent fleet area of operation (in yellow) targeting large pelagic species.

Target species and volume per species: Large pelagic fishery initiated in 1975 with vessels targeting swordfish. Catches grew steadily until 1997 when bluefin tuna started to be a considerable part of the landings and vessels started actively fishing for it. From 2001 onwards there is a clear preference of the fleet towards albacore fishing (June to September) as this is a less risky and convenient activity with single day trips that result in steady catch rates/incomes. A great reduction of the production occurs in 2008 as a result of a considerable reduction of fleet capacity with the scraping plan of 12 vessels. Cyprus catches of albacore take a considerable proportion of EU Mediterranean landing. Nowadays swordfish is actively targeted from a small number of vessels. However, vessels targeting bluefin tuna and albacore are also catching swordfish as bycatch. Bluefin tuna is targeted by a handful of vessels mainly in GSA 22 area. A graphical presentation of the landings trajectories over time is given in Figure 4.

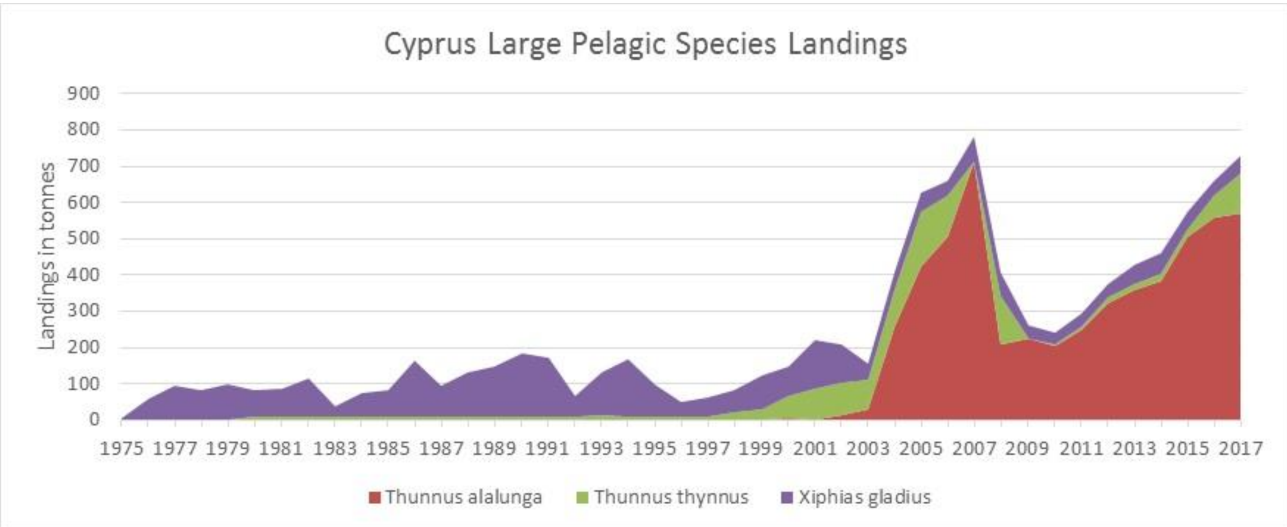


Figure 4. Large pelagic landings by species from 1975 – 2017.

4. Sampling description

4.1. Port sampling in Africa (Abidjan and Dakar)

The multi-specific nature of tropical tuna purse seine fisheries makes it difficult to estimate basic catch by species and sizes. Already in the 80s, ICCAT Tropical Tuna Working Group observed biases in the species composition declared in logbooks, especially affecting mainly juveniles (Cayré, 1984). In 1997, a new sampling system was proposed within the European Program for the correction of this bias based on the multi-specific sampling of the landings, a procedure that is still in place at present for purse seine landings both in the Atlantic and Indian Oceans. In the case of the Atlantic Ocean, sampling is conducted in three ports; Abidjan as the main port and Dakar and Tema as secondary ports. Thus, the current sampling system has been defined with the aim of improving the accuracy of the statistics considering most of the variables that influence this fishery. This sampling is the basis of the procedure which allows to estimate the species and size composition of the total tuna catch. This estimation process, that includes port sampling and subsequent raising, is known as T3 (Tropical Tuna Treatment).

Thus, this sampling program has two objectives; (1) estimate the species composition of the main target tuna species, thus correcting the estimates included in official logbooks, and (2) estimate size distribution for each of the target tuna species.

Port Sampling: The sampling team is in the port whenever a landing event occur, so that all trips are susceptible to being sampled. However, not all the catch within a fishing trip is susceptible to being sampled. Vessel's wells are unloaded simultaneously, and only some of these wells can be selected for conducting the sampling. Sampling is stratified; thus, a fish well could be selected for sampling only when all the catches stored in it come from fishing operations or sets recorded for the same strata (fishing area, time period and fishing mode). Wells that do contain catches from one single fishing set are the priority, while wells that do contain several sets from the same strata are secondary targets for sampling. Wells that include catches from sets conducted in more than one stratum are taken out from the sampling frame

The stratification has been defined based in the analyses within the European project ET (Pallarès P. and Hallier J.P. 1997), and it is done according to three criteria: fishing zone, period and fishing mode. One of the important results of the analysis was that it is not necessary to add a supplementary stratum "fleet", so the ship's flag is not considered for stratification at this moment.

- Regarding fishing mode, two strata are distinguished; (1) free school sets and (2) sets associated to drifting floating objects (either natural or artificial objects).
- Time-period, with each year broken by quarter (January-March, April-June, July-September, October-December).

- Six fishing areas are distinguished, which slightly vary between fishing modes. Areas are divided as follows (Figure 1):

Free school set areas:

ZONA 1	Senegal	30ºN-12ºN	35ºW-línea de costa
ZONA 2	Norte Piccolo	12ºN-5ºN 5ºN-0º	35ºW-línea de costa 35ºW-20ºW
ZONA 3	Piccolo	5ºN-0º	20ºW-10ºW
ZONA 4	Noreste Ecuador	latitud costa-0º	10ºW-5ºE
ZONA 5	Cabo López	latitud costa-15ºS	5ºE-línea de costa
ZONA 6	Sur de Ecuador	0º-15ºS	35ºW-5ºE 35ºW-línea de costa

FOB associated set areas:

ZONA 1	Senegal	30ºN-12ºN	35ºW-línea de costa
ZONA 2	Norte Piccolo	12ºN-5ºN 12ºN-0º	35ºW-11ºW 35ºW-20ºW
ZONA 3	Piccolo	5ºN-0º 4ºN-0º	22ºW-10ºW 20ºW-10ºW
ZONA 4	Costa	7ºN-4ºN	11ºW-5ºE
ZONA 5	Cabo López	latitud costa-15ºS	5ºE-línea de costa
ZONA 6	Sur de Ecuador	4ºN-0º 0º-15ºS	10ºW-5ºE 35ºW-5ºE 35ºW-línea de costa

In a first step, the sampling team obtain the wells' plan from all vessels at port. Each fishing trip has a specific wells' plan, which shows how the different fishing sets included in the logbook are distributed among wells. Following the criteria specified above, wells are divided in sampleable wells (i.e. wells containing catch from a unique stratum) and non-sampleable ones (i.e. catch from several strata mixed). Sample or well selection method is **non-probability quota sampling (NPQS)**. This is a non-probability selection method, which takes place until a preassigned number of units is attained. The number of samples (sampled wells) recommended in each stratum is 25, the minimum number being 15. Depending on the landings, the priority in the choice of wells will be made trying to achieve this objective, in order to achieve the minimum number in all the strata considered. Although the recommended number of samples over a stratum has been exceeded, the team will continue collecting samples as long as there are no other priority strata on the same day. Once wells are selected, two random samples are measured; the first will be done shortly after opening the well, and the second a few hours later, as separated as possible from the first, to avoid biases due to fish position or stratification within the well. Number of individuals to be measured by sample vary based on the commercial category, where 300 fishes are measured for smaller categories and 100 if larger individuals are landed.

Table 1 below summarizes the tropical tuna sampling design in the Atlantic Ocean. This sampling is the basis of the procedure which allows to estimate the species and size composition of the total tuna catch (T3 process).

The T3 processing was built about 30 years ago in order to correct biases of the logbook on species composition and to provide more accurate estimates of catch by species for the European purse seine fleet (Duparc *et al.* 2018).

The T3 processing is divided into three major components.

- The first part aims at standardizing the logbooks catches (level 1),
- The second part aims to standardize and enhance size samples (level 2),
- Based on results of the first two stages, level 3 aims to correct the specific composition of the catch by commercial category reported in the logbooks by applying the standardized samples composition to them.

Table1. Tropical tuna sampling design in the Atlantic Ocean.

	Sampling design	Comments
Country	ESP & FRA	Samples are shared among countries; sampling not stratified by flag. There are other associated flags (Spanish companies, but vessels flagged in third countries)
Frame	Landings of tropical tuna (SKJ, YFT, BET) in the ports of Abdijan, Dakar and tema, targeted by PS fleet in the ATL	Some ports out from the sampling frame. Globally, these ports represent <10% of total landings.
1st SU	Day*Port	Three ports are sampled: Abidjan, Dakar and Tema. These ports represent the >85% of the landings of this fleet. Some minor ports (e.g. Mindelo) are not sampled, thus out from the sampling frame
stratification of 1st SU	Different sampling teams by flag, who follow same protocol. Later samples (raw data) are shared for raising purposes.	There is not real stratification of the 1st sampling unit. Sampling team is always at port if there is a landing event. Sampling is design in a yearly basis but sometimes you can have a trip overlapping two year (for example at the end or the begging of the year).
Selection of 1st SU	Almost all port*days are sampled. There is not a real selection of trips, as the sampling team is in the port whenever a landing event occur	Selection of PSU could be seen as a CENSUS, as you aim to sample whenever a landing event occur. In practice. It would be also similar to consider Simple random sampling with a very high probability of being sample
2nd SU	Well	Each 'port_day' some wells that are unloaded that day are selected to be

		sampled, thus the well is the 2nd SU. However, trip and fishing operation (set) data are needed as auxiliary variables.
Stratification of 2nd SU	Area*Fishing mode *trimester	There are some wells which are not sampleable (sampling probability=0) because they contain catch from different strata mixed in the same well. These wells are never sampled and could be registered in a separate stratum "mixed". It is also desirable to distinguish between priority wells with one single set (mono-set) and wells with more than one set (multi-set), even if they belong to the same strata.
Selection of 2nd SU	Quota sampling from a list of wells which is provided each landing day to the sampler	
3rd SU	Fish (length)	
Stratification of 3rd SU	Scientific species	
Selection of 3rd SU	- Two subsamples are measured; the first is taken shortly after opening the tank, and the second a few hours later, as separated as possible from the first '- Number of individuals to be measured by sample vary based on the commercial category, where 300 fishes are measured for smaller categories and 100 if larger individuals are landed '- Three target species sampled (SJK, YFT and BFT)	

4.2. Cyprus Swordfish sampling

Sampling of Swordfish in Cyprus involves three metiers as apart from the targeted activity (LLD_LPF_0_0_0 (SWO)) of the species, it is present as bycatch in the other two major large pelagic activities (LLD_LPF_0_0_0 (ALB), LLD_LPF_0_0_0 (BFT).

Sampling is performed at sea and on shore (landing sites), based on a probability sample survey.

4.2.1. On-board

The sampling frame consists of the list of the licensed vessels for the reference years, with the vessel being the primary sampling unit (PSU) and as secondary sampling unit (SSU) the trips by each vessel on a randomly selected time frame.

Temporal stratification is employed for the collection of the data, in order to estimate the metier-related variables on a quarterly basis. The overall number of sampled trips is designed to fulfil the minimum requirements which have been selected considering the relatively low average number of trips in the reference years.

Up until 2016 scientific observers coming from consulting firms were used to implement the on-board and on-shore activities through tenders. Due to increased number of denials and difficulties for on-board sampling there was a tender for direct hiring of scientific observers' services which was cancelled as there were no applicants. In order to cover the sampling needs there was an agreement with control observers to cover this aspect 2017 onwards. The data that has been used for the practical testing of RDBES model come from 2016 as this is a configuration coming from a probabilistic scheme. However, RDBES is designed to accommodate data coming from control agencies, fishers self-sampling, ad-hoc and non-probabilistic approaches.

4.2.2. On-shore

The on-shore sampling of the 2016 dataset can fit in Hierarchy 6 where a sampling takes place at landing site from a random selection of known fishing trips. Although Swordfish is the main target species for the fleet, as well as the observers at shore, any other species that come as bi-catch and is landed is also measured. Additionally, data of weight and length comes from control inspectors during random checks and/or in the cases that a BFT is included in the catch. The number of individuals per boat is relatively small allowing for census measurement of the catch. Samples can also come from albacore and BFT landings.

4.2.3. Remarks applied to for both cases (on-board & on-shore)

The small number of Cypriot vessels operating in this segment as well as the relatively smaller catches, allow for census sampling of all landing events from all returning fishing trips during visits. This might not be the case for other countries were a decision must be made further down to select which of the landing events will be processed. Additionally, the fact that area and vessel number are confined compared with other areas (targeting Mediterranean Swordfish) saves on strata and design complexity. Thus, this example cannot be considered suitable a priori for the rest of Member states. Further investigation with additional more complex cases is needed.

5. Populating RDBES data model for tropical tuna sampling and mediterranean swordfish

In the following we explain the decisions made when populating the RDBES data model with data (samples) from tropical tuna and Mediterranean swordfish sampling program.

Within this process to populate the data model someone must have in depth knowledge of the sampling design and its real-life application in terms of assumptions, specificities and goals in order to produce a valid output. Any deviations from the plan must be explicitly reported as they can influence the achieved outcomes which will be derived *post mortem* from the estimation process.

Considering Hierarchies defined within RDBES data model (ICES, 2018), once the suitable one is identified for a specific sampling program, these tables consist a very useful hands on exercise to test the sampling design results and identify gaps, flows or grey areas in the design that need further documentation or improvements. They can also be very helpful in pinpointing every minute detail and info one has to collect while on the field easing the communication of the requested technical specifications between sampling supervisors and observers.

Some issues have been identified during these exercise: In general, data model can facilitate the registration of these LP data and only minor adjustments are needed, most of which are simple data model provisions: such as expanding allowable active areas to include ICCAT 5x5 degree squares and ET tropical tuna sampling areas, inclusion of new métiers that allow distinguishing free school sets and sets on floating objects, or provision for allowing hook size in the gear characteristics based on ICCAT principles (SWO, BFT, ALB).

In large pelagic context it can happen that the primary sampling unit for on-shore sampling is the vessel and then other attributes follow in the lower selections. This case is not provisioned in landing Hierarchies 5 to 8. One can try to work with the on-board dedicated Hierarchies to check if they can accommodate that scenario. Furthermost this case can be explicitly developed for that purpose.

5.1. Tropical tunas in West Africa

5.1.1. Identifying Hierarchy

In this sampling program each sampled fish comes from an unsorted landing, and each landing is found by a visit to a port on a specific day (onshore event). The primary sampling unit is therefore the onshore event, the secondary sampling unit is a landing of a well, where we know the vessels and trip details, and from these wells' landing we obtain our ultimate samples which is 200-600 fish of the 3 target tuna species.

Thus, considering hierarchies defined within RDBES data model (ICES, 2018), the best fit seems to be Hierarchy 5, where it will be completed *Design table*, *Sampling Details table*, each sampled port-day in the *On-Shore table*, each sampled well in the *Landing Event table*, including the auxiliary vessel and trip information in the *Vessel Details table*. Finally, it will be fill in the tables: *Species Selection, Species List Details and Sample*. However, after discussing it with the RDBES Core Group, it is not enterily clear whether the sampling design can be fit in the existing hierarchyes (i.e. Hierarchy 5) or a new hierarchy is needed to acomodate a selection level for weels.

5.1.2. Populating Design Table

- '*DEstratum*': Stratum could be simply "TROP_TUNA_PS_ATL" (purse seine targeting tropical tunas in the Atlantic Ocean). Flag shouldn't be considered different strata, even if there are different sampling teams by flag, as they follow the same protocol. The different sampling teams can be identified in an existing filed for samplers information

- '*DEhierarchyCorrect*': Hierarchy that best fit seems to be Hierarchy 5. However, after discussing it with the RDBES Core Group, it is not entirely clear whether a new hierarchy is needed, with a new selection level for wells. In relation with this, it is important to confirm how the wells are selected once they are in the port. Is there a selection of wells directly, or is there a selection first of vessels and second of wells?

TABLE: Design						
DEid	**DErecordType**	**DEsamplingScheme**	**DEyear**	**DEstratum**	**DEhierarchyCorrect**	**DEhierarchy**
1	DE	T3 sampling	2017	TROP_TUNA_PS_ATL	YES	5

5.1.3. Populating Sampling Details table

- *'SDcountry' & 'SDinstitution':* In many cases, sampling coordinated by EU countries (i.e. ESP and FRA), but field work done by local institutes (CRO in Ivory Coast and SFA in Seychelles).

TABLE: SamplingDetails				
SDid	**DEid**	**SDrecordType**	**SDcountry**	**SDinstitution**
1	1	SD	FRA, ESP	IRD, IEO

5.1.4. Populating Vessel Details' table (auxiliary variable)

Vessel selection is not done in each 'day_port', and sampling team directly select one or some specific well(s) each 'port_day'. However, information referred to the list of vessels and vessel's details is needed for raising, and it is available in 'Turbobat' file (file with vessels characteristics, such as length, capacity, horse power…).

- 'lengthCategory' and 'VdsizeUnit': Using length to distinguish Vdcategories is possible. However, for this fleet carrying capacity (tonnes) is normally used to separate categories.

TABLE: VesselDetails **(Auxiliary)**										
VDid	**VDrecordType**	**VDencryptedCode**	**VDhomePort**	**VDflagCountry**	**VDlength**	**VDlengthCategory**	**VDpower**	**VDsize**	**VDsizeUnit**	**VDtype**
1	VD	1		ESP	78	801 - 1200 tonnes	5000	1200	GRT	4

5.1.5. Populating On Shore table

- *OSstratification:* Each port sampled (main ports) have a different sampling team, thus stratification by port has been considered (three strata: Abidjan, Dakar, Tema). Regarding temporal stratification, year is the strata.
- *OSclustering:* "No", as ports are not grouped. Each port refers to an independent stratum.
- *OStotal, OSsampled, OssampProb and OSselection method: OStotal* refers to the total on-shore events (landing days) in this stratum – (e.g. In Abidjan port during 2017 landing occurred 215 days). As commented above, sampling teams are present every day, thus selection method is '*Census*'. However, due to logistics some days could be without any sampling (200 days where some sampling occurred in Abidjan port during 2017). PSU can be seen as a CENSUS, as you aim to sample whenever a landing event occur. In practice, it would be similar to consider Simple random sampling with a very high probability of being sample.
- *OSsampler:* CRO (Centre Recherche Océanographique)

TABLE: OnshoreEvent									
OSid	SDid		OSrecordType	OSnationalLocationName	OSstratification	OSlocation	OSsamplingDate	OSsamplingTime	OSstratum
1	1		OS	Abidjan	Y	Abidjan	01/03/2017		ABJ-PORT

OSclustering	OSclusterName	OSsampler	OStimeUnit	OStimeValue	OStotal	OSsampled	OSsampProb	OSselectionMethod
No	-	CRO			215	200	0.93	Census

OSlocationType	OSselectionMethodCluster	OStotalClusters	OSsampledClusters	OSclustersProb	OSreasonNotSampled
Port	-	-	-	-	-

5.1.6. Populating Landing Event

We have used the table Landing event (LE) to include information on the selection of the Secondary Sampling Untis, which in this case are the well. However, depending the hierarchy finally selected, this may change.
There are some wells which are not sampleable (sampling probability=0) because they contain catch from different strata mixed in the same well. These wells are never sampled and could be registered in a separate stratum "mixed strata".

- LEstratification: Stratification by area and fishing mode (FSC; Free school and FOB: Floating object). Both areas and fishing mode should be added to the code list.
- *LEtotal, LEsampled, LEsampProb and LEselection method: LEtotal* refers to the total wells landed in this stratum – (e.g. wells unloaded during 1st quarter in Abidjan port containing exclusively FSC catches from Piccolo area (112 wells unloaded). *LEsampled* shows that 23 wells have been sampled. As commented above, sampling teams have some targets, 25 samples per stratum. Thus, selection method is 'quota sampling'.
- *LEreasonNotSampled.* Include specific reasons in the code list.

LEid	OSid	FTid	VSid	VDid	LErecordType	LEstratification	LEsequenceNumber
1	1			1	LE	Y	1

LEhaulNumber	LEstratum	LEclustering	LEclusterName	LEsampler	LEmixedTrip	LEcatchReg	LElocation
2	FSC_Piccolo			CRO	No	Lan	Abidjan

LElocationType	LEcountry	LEdate	LEtime	LEeconomicalZone	LEarea	LErectangle	LEsubpolygon
Port	CIV	01/03/2017			Piccolo		

LEfunctinalUnit	LEnationalCategory	LEmetier5	LEmetier6	LEgear	LEmeshSize	LEselectionDevice	LEselectionDeviceMeshSize
			PS_LPF_0_0_0	PS			

LEtargetSpecies	LEtotal	LEsampled	LEsampProb	LEselectionMethod	LEselectionMethodCluster	LEtotalClusters	LEsampledClusters

LPF	112	23	0.20	NPQS	-	-	-

LEclustersProb	LEreasonNotSampled
-	-

5.1.7. Populating species selection

For this sampling program we always sample the landed fraction of all target species catches (YFT, SKJ and BET). We therefore fill in SSselectionMethod as 'CENSUS'. SSstratification is 'No' and SSclustering is 'No'.

'*SSspeciesList*' should include one specific code for "target tunas", as sample is not sorted by species, and this code will be used in the '*Sample*' table.

'FOid' identifies the fishing operation → trip data and fishing operation data needed as auxiliary variable.

SSid	LEid	FOid	Slid	SSstratification	SSspeciesListID	SScatchFraction	SSstratum
1	1	1	1	N	1	Lan	U

SSclustering	SSclusterName	SSsampler	SSspeciesListName	SStotal	SSsampled	SSselectionMethod	SSselectionMethodCluster
No		CRO	SPSlist (target tunas)	3	3	CENSUS	

SStotalClusters	SSsampledClusters	SSclustersProb	SSreasonNotSampled

5.1.8. Populating sample

When sampling a well, selection of fish (200-600 individuals) is assumed to be simple random without replacement (separated in two subsamples in order to avoid the possible stratification within a well), so we write 'SRSWOR' in the SAselectionMethod field. The tank is sampled as a whole. This is, the samples are not sorted, so we write 'U' both for '*SAstratum*' and '*SAsex*'.

The sampling is also used to estimated species composition; thus the sample is not sorted by species and the three target tuna species are sampled together. Thus, in the field *'SAcomercialSpecies'* (and in the *SSspeciesList* 'target tunas') the three target tunas should be included; YFT, SKJ & BET. Weight categories (>10 Kg, <10Kg) are later used for raising purposes.

Retained catch, which is sold to canneries (HUC), is landed frozen and whole.

‘SAunitType’ is “number”, as the target is a determined number of fishes to be measured by sample.

‘SAliveWt’ refers to the total catch in a well (e.g. 60 tones). ‘*SAsampleWeight*’ refers to the sample weight (e.g. 4 tones).

‘SAtotal’ ‘SAsampled’ and ‘SAsampProb’ can be calculated if needed using W-L relationship. Even if the sample is selected without stratification, weight categories (>10 Kg, <10Kg) are later used for raising purposes.

SAid	**SAparentID**	**SSid**	**SArecordType**	**SAnationalCode**	**SAstratum**	**SAspeciesCode**	**SAcommercialSpecies**
1		1	SA	YFT	U	403215	YFT

SApresentation	**SAcatchFraction**	**SAlandingCategory**	**SAcommSizeCatScale**	**SAcommSizeCat**	**SAsex**	**SAunitType**	**SAliveWt**
Whole, frozen	LAN	HUC			U	number	60,000,000

SAsampleWeight	**SAtotal**	**SAsampled**	**SAsampProb**	**SAselectionMethod**	**SAlowerHierarchy**	**SAsampler**	**SAreasonNotSampledFM**	**SAreasonNotSampledBV**
4,000,000				SRSWOR	c	CRO		

5.1.9. Populating biological variable

Once that two subsamples have been selected as described in previous section (SRSWOR), all individuals within the sample are measured (length), thus the ‘*BVselectionMethod’* is CENSUS and ‘*BVsampProb*’ =FADs,,,he rest of the variables (weight, maturity, age…) are under a different sampling program). All individuals are measured by hand with caliper; larger individuals (>70 cm) to the lower ½ cm and using first dorsal length, while smaller individuals (<70 cm) are measured to the lower cm and fork length.

TABLE: BiologicalVariable									
BVid	**SAid**	**FMid**	**BVrecordType**	**BVfishID**	**BVstratum**	**BVtype**	**BVvalue**	**BVunitValue**	**BVunitScaleList**
1	1		BV	1	U	Length	70	cm	

BVmethod	BVtotal	BVsampled	BVsampProb	BVselectionMethod	BVsampler
by hand with caliper			1	CENSUS	CRO

5.2. Mediterranean swordfish sampling in Cyprus

5.2.1. Identifying Hierarchy

The ideal Hierarchy for facilitating 2016 dataset identified to be number 1 (Figure x). On board sampling starts with the selection of vessels as PSU and a potential day for trip as SSU. While on board, observer collects biological data for every single fish coming on the deck. As the assigned number of trips are on the minimum requirements and catch quantities have wide variation it is necessary to sample all fish and observe the activity 100% in order to reach the allocated quota sampling. Additionally, length sampling of discarded species occurs for which stock-related variables are collected.

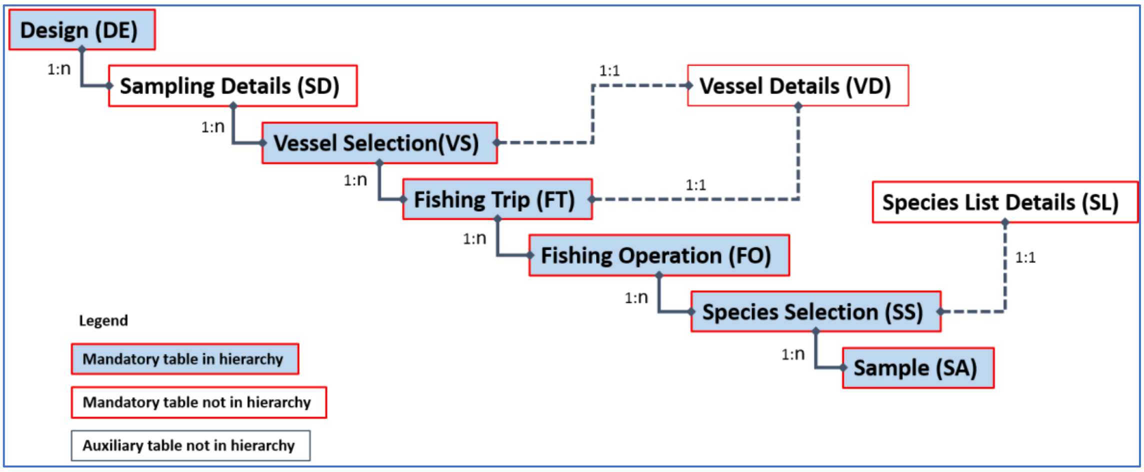


Figure 5 Schematic workflow representation of Hierarchy 1

5.2.2. Populating Design Table

DEstratu': Sampling is stratified per quarter in a way that fulfils the minimum requirements set.

DEhierarchyCorrect: Hierarchy rational seems to capture sampling design accurately thus a Y standing for yes is assigned.

TABLE: Design						
DEid	**DErecordType**	**DEsamplingScheme**	**DEyear**	**DEstratum**	**DEhierarchyCorrect**	**DEhierarchy**
1	DE	LLS	2016	Q1	Y	1
2	DE	LLS	2016	Q2	Y	1

3	DE	LLS	2016	Q3	Y	1
4	DE	LLS	2016	Q4	Y	1

5.2.3. Populating Sampling Details table

- *SDcountry & SDinstitution:* Sampling is coordinated by DFMR in Cyprus and field work was outsourced to observers through tender.

TABLE: Sampling Details				
SDid	**DEid**	**SDrecordType**	**SDcountry**	**SDinstitution**
1	1	SD	CYP	DFMR
2	2	SD	CYP	DFMR
3	3	SD	CYP	DFMR
4	4	SD	CYP	DFMR

5.2.4. Populating Vessel Selection Table

Vessel selection is done through equal probability simple random sampling without replacement, through a list of suitable licenced vessels

- *VStotal*: is the total number of vessels operating in this segment for 1st quarter
- *VSdsampled*: The total number of vessels sampled in the above-mentioned activity
- *VSsampProb*: As this is an equal probability sampling inclusion probability will be calculated automatically in the estimation process
- *VSselectionMethod*: The acronym for the sampling method used

TABLE: Vessel Selection						
VSid	**SDid**	**VDid**	**TEid**	**VSrecordType**	**VSstratification**	**VSstratum**
1	1	1		VS	N	U
2	2	2		VS	N	U
3	3	3		VS	N	U
4	4	4		VS	N	U

VSclustering	**VSclusterName**	**VSsampler**	**VStotal**	**VSsampled**	**VSsampProb**	**VSselectionMethod**
No	None	Observer	32	1		SRSWOR
No	None	Observer				SRSWOR
No	None	Observer				SRSWOR
No	None	Observer				SRSWOR

5.2.5. Populating Fishing Trip table

- *VFTarrivalLocation:* ERS in this case indicates that details will come from that systems but in real situation

TABLE: Fishing Trip								
FTid	**OSid**	**VSid**	**VDid**	**SDid**	**FTrecordType**	**FTnationalCode**	**FTstratification**	**FTstratum**
1		1	1	1	FT	LLS54	Y	Q1

FTclustering	**FTclusterName**	**FTsampler**	**FTnumberOfHauls**	**FTdepartureLocation**	**FTdepatureDate**	**FTdepartureTime**	**FTarrivalLocation**
No	None	Observer	2		23-02-16	14:50:00	ERS

FTarrivalDate	**FTarrivalTime**	**FTtotal**	**FTsampled**	**FTsampProb**	**FTselectionMethod**
25-02-16	11:05:00	ERS	2		SYSS

5.2.6. Populating Species Selection table

Bycatch species will also be included either as landing or discards depending on their fate.

TABLE: Species Selection						
SSid	**LEid**	**FOid**	**SLid**	**SSrecordType**	**SSstratification**	**SScatchFraction**
1		1	1	SS	N	LAN
2		1	2	SS	N	DIS
3		2	1	SS	N	LAN
4		2	2	SS	N	DIS

SSstratum	**SSclustering**	**SSclusterName**	**SSsampler**	**SSspeciesListName**	**SStotal**	**SSsampled**	**SSselectionMethod**
U	No	None	Observer	All species	1	1	CENSUS
U	No	None	Observer	All species	0	0	CENSUS
U	No	None	Observer	All species	1	1	CENSUS
U	No	None	Observer	All species	1	1	CENSUS

5.2.7. Species List table

In here it will be necessary that species codes in databse must allow for the inclusion of Swordfish fishery related species both bicatch and discards.

TABLE: Species List Details					
SLid	**SLrecordType**	**SLlistName**	**SLspeciesCode**	**SLcommercialSpecies**	**SLcatchFraction**

1	SL	All species	NA		LAN
2	SL	All species	NA		DIS

5.2.8. Populating Sample table

TABLE: Sample								
SAid	**SAparentID**	**SSid**	**SArecordType**	**SAnationalCode**	**SAstratum**	**SAspeciesCode**	**SAcommercialSpecies**	**SApresentation**
1		1	SA	LLS541	LAN	2		whole
2		1	SA	LLS542	LAN	2		whole
3		1	SA	LLS543	LAN	2		whole
4		1	SA	LLS544	LAN	2		whole
5		1	SA	LLS545	LAN	2		whole
6		1	SA	LLS546	LAN	2		whole
7		1	SA	LLS547	LAN	2		whole
8		1	SA	LLS548	LAN	2		whole
9		3	SA	LLS549	LAN	2		whole
10		3	SA	LLS5410	LAN	2		whole
11		3	SA	LLS5411	LAN	2		whole
12		3	SA	LLS5412	LAN	2		whole
13		3	SA	LLS5413	LAN	2		whole
14		3	SA	LLS5414	LAN	2		whole
15		4	SA	LLS54	DIS	x		whole

SAcatchFraction	**SAlandingCategory**	**SAcommSizeCatScale**	**SAcommSizeCat**	**SAsex**	**SAunitType**	**SAliveWt**
LAN	HUC	EU	120	M	number	17920
LAN	HUC	EU	106	F	number	12320
LAN	HUC	EU	110	M	number	14560
LAN	HUC	EU	118	M	number	16240
LAN	HUC	EU	126	M	number	16240
LAN	HUC	EU	118	F	number	14000
LAN	HUC	EU	118	F	number	15120
LAN	HUC	EU	124	M	number	18480
LAN	HUC	EU	124	F	number	19040
LAN	HUC	EU	120	F	number	16800
LAN	HUC	EU	127	M	number	22960
LAN	HUC	EU	110	M	number	15120
LAN	HUC	EU	114	F	number	19600
LAN	HUC	EU	116	M	number	19040
DIS						

SAsampleWeight	**SAtotal**	**SAsampled**	**SAsampProb**	**SAselectionMethod**	**SAlowerHierarchy**	**SAsampler**

17920	17920	1		CENCUS	A	Observer
12320	12320	1		CENCUS	A	Observer
14560	14560	1		CENCUS	A	Observer
16240	16240	1		CENCUS	A	Observer
16240	16240	1		CENCUS	A	Observer
14000	14000	1		CENCUS	A	Observer
15120	15120	1		CENCUS	A	Observer
18480	18480	1		CENCUS	A	Observer
19040	19040	1		CENCUS	A	Observer
16800	16800	1		CENCUS	A	Observer
22960	22960	1		CENCUS	A	Observer
15120	15120	1		CENCUS	A	Observer
19600	19600	1		CENCUS	A	Observer
19040	19040	1		CENCUS	A	Observer

6. Final conclusion and remarks

This section shows the main conclusions reached during the feasibility study of the RDBES to host sampling data for LP stocks, and concretely to host the data for the tropical tuna port sampling and Mediterranean swordfish sampling programs. Within the process, the RDBES Core Group has been contacted, and both the RCG-LP and the Core Group will look at the results of this exercise and respond to specific questions or adapt the data model and documentation as required. Some issues have been identified and are documented in this report. However, it is not thought that any of the issues raised are serious impediments to moving forward with the RDBES data model. Nevertheless, this is a process that does not end here. Sampling programs of highly migratory species are diverse, as diverse as the difficulties that some of them could face in the process of uploading data to the RDBES. Probably, sampling Bluefin tuna cages have little to do with tropical tuna or swordfish sampling. In addition, RDBES development is relatively new process and will surely change in the near future; implementation of raising procedures, etc. Moreover, a regional data base is one of the main prerequisites for the development of a Regional Sampling Plan, for data standardization and for quality assurance. Thus, so that this process whereby LP data are part of the RDBE succeed, active involvement of RCG LP is needed.

- The first objective was to find out which RDBES data model's hierarchy fits better both cases studies. Mediterranean swordfish sampling matches without major issues: on board sampling matches using Hierarchy 1, and onshore sampling matches using Hierarchy 6. On the other hand, for tropical tuna onshore sampling the best fit seems to be Hierarchy 5. However, after discussing it with the RDBES Core Group, it is not entirely clear whether the sampling design can be fit in the existing hierarchies (i.e. Hierarchy 5) or a new hierarchy is needed to accommodate a selection level for wells.

- Once the correct hierarchy has been identified, there have been no major problems when populating the different tables. However, both for Mediterranean swordfish and tropical tunas, some new codes should be included in the master tables: such as the ET sampling areas, ICCAT 1 degree and 5 degree statistical squares, new metiers that allows distinguishing free school sets and sets on floating objects, as well as provision for allowing hook size in the gear characteristics based on ICCAT principles.

- Taking into consideration the RDBES road map (Table 2) already announced in the SCRDBES 2018 report (ICES, 2019), it is certain that any new requirements depicted for LP will not come into place in the first releases of the platform. Funds and time (dedicated representatives or projects – as this big goal must not rely in volunteering work) must also be provisioned for dedicated LP developments. This requires involvement of the RCG LP, the RDB Steering Committee and the Development Core Group to succeed.

- A big landmark here is the estimation procedure which will be unveiled and developed in detail after the WKRDBES-EST (estimation process for selected stocks) that will take place in October 2019. The Workshop aim is to produce estimation scripts and in there, inconsistencies, issues or extra requirements on Hierarchies might emerge. Complete datasets of large pelagics sampling (with current scheme and proposed regional sampling plan) are needed to be tested in WKRDBES-EST. The objectives of the sampling should be fulfilled in the estimation process: (1) estimate the species composition of the main target tuna species, thus correcting the estimates included in official logbooks, and (2) estimate size distribution for each of the target tuna species.

- One of the big potentials of the system apart from the obvious improving of data checks, quality assurance, uniformity among users etc. is the save in time. The system will be able to assist member states in reporting fast and accurately to end users (ICCAT, EU etc) as centrally developed (and peer reviewed scripts) will be able to extract and prepare the data in the requested formats.

As mentioned before, all the above dictate the need for active involvement of RCG LP in these steps.

Table 2. RDBES road map.

	RDB System	InterCatch	RDBES	ICES Secretariat	Core Group	WGCATCH/ PGDATA	Countries	WG and Benchmark	RCGs
2019	Production Data in/out	Production Data in/out	Development Test data in/out	System development Test web between Feb-Oct. RDBES operational Dec.	Specifications WKRDB-POP Upload data WKRDB-EST. Start specifying Large Pelagic needs.	WGCATCH: Reviewing the outcome WKRDB-EST and suggest improvement. Guideline and algorithms for general estimations (ratio/statistical/ Design based). PGDATA: Describe how the RDBES fits into the QAF.	WKRDB-POP Upload data after POP. Participate and prepare for RDBES format and estimations WKRDB-EST		RCG chairs to request countries to participate in the WKRDB-POP. RCG support the countries to allocate sufficient time for these WKs.
2020	Production Data in/out	Production Data in/out	Test by selected stocks	Large Pelagic development (funding needed).	WKRDB-POP (Jan-Feb) target selected stocks. WKRDB-EST (incl. imputation) Estimations for selected stocks		Data call. Upload data for selected stocks after WG WKRDB-POP WKRDB-EST	Upload data. WG estimations. Trans Data compilation WK. Laurent/Core Group will help selected WG with selected stock (10-20) (1)	
2021	Production Data in/out	Production Data in/out	Test by all stocks		WKRDB-POP (Jan-Feb) target all stocks. WKRDB-EST Estimations finalised				
2022	Stay alive Data out	Stay alive Data out	Data call for 2021 data						
2023	Stay alive Data out	Stay alive Data out	Data call for 2022 and all year						
2024	Terminated	Terminated							

7. References

Cayré P. 1984. Procédure de suivi pour la révision de la composition spécifique des statistiques thonières FISM. Col.Vol.Sci.Pap. ICCAT, 21 (2): 102-107

Davies, T., Mees, C. and Milner-Gulland, E.J. 2014. The past, present and future use of drifting fish aggregating devices (FADs) in the Indian Ocean. Marine Policyn 45: 163-170.

Duparc A., Cauquil P., Depetris M., Dewals P., Floch L., Gaertner D., Hervé A., Lebranchu J., Marsac F. and Bach P. Assessment of accuracy in processing purse seine tropical tuna catches with the T3 methodology using French fleet data. IOTC-2018-WPTT20-16

Fonteneau, A., Pallares, P., Pianet, R. 2000. A worldwide review of purse seine fisheries on FADs. In J.Y. Le Gall, P. Cayre´, M. Taquet (Eds.), Pêche thonière et dispositifs de concentration de poisons,28, Actes Colloques-IFREMER, pp15-35

Hallier J.P. and Parajua J.I. 1992. Review of tuna fisheries on floating objects in the Indian Ocean. In: Fishing for tunas associated with floating objects. La Jolla : IATTC, 25 p. International Workshop on Fishing for Tunas Associated with Floating Objects, La Jolla (USA), 1992/02/11-14

ICES. 2018. Report of the Workshop on new data model for the Regional Database (WKRDB-MODEL), 15 – 18 January 2018, ICES HQ, Copenhagen, Denmark. ICES CM 2018/ACOM:41, 44 pp.

ICES.2019. Report of the Steering Committee of the Regional Data Base & Estimation System (SCRDBES), 4-6 December 2018, ICES HQ, Copenhagen, Denmark. ICES CM 2018/ACOM:29. 43 pp.

Majkowski, J., H. Arrizabalaga, F. Carocci & H. Murua 2011. Tuna and tuna like species. In FAO 2011. Review of the state of world marine fishery resources. Rome: FAO Fisheries and Aquaculture Technical

Pallarès P. and Hallier J.P. 1997. Analyse du schéma d'échantillonnage multi-spécifiques des thonidés tropicaux. Rapport scientifique. IEO/ORSTOM, Programme N°95/37. Paper No. 569, pp. 227-244

ANNEX1. Floating object's (FOB) codification included in ICCAT Recommendation 16-01.

Table 1. Codes, names and examples of different types of floating object that should be collected in the fishing logbook as a minimum data requirement. Table from 2016 SCRS report (section 18.2 Table 7).

Code	*Name*	*Example*
DFAD	Drifting FAD	Bamboo or metal raft
AFAD	Anchored FAD	Very large buoy
FALOG	Artificial log resulting from related to human activity (and related to fishing activities)	Nets, wreck, ropes
HALOG	Artificial log resulting from human activity (not related to fishing activities)	Washing machine, oil tank
ANLOG	Natural log of animal origin	Carcasses, whale shark
VNLOG	Natural log of plant origin	Branches, trunk, palm leaf

FRAMEWORK CONTRACT - MARE/2016/22 « Strengthening regional cooperation in the area of fisheries data collection » Annex III « Biological data collection for fisheries on highly migratory species »

Strengthening Regional cooperation in the area of large pelagic fisheries data collection (Acronym: RECOLAPE)

Task 2.6 - Lessons learned

D.2.6 - Set of rules and recommendations that would allow transferring of the proposed Regional Sampling Plan (RSP) to other large pelagic fisheries with similar characteristics

Responsible: George Tserpes (HCMR), Mathieu Depetris (IRD) & Antoine Duparc (IRD)

Index

Executive summary 3

1. Introduction...4

2. Large pelagic longline fisheries in the Mediterranean Sea 4

3. Large pelagic fisheries targeting tropical tunas in the Indian Ocean............................5

4. References...7

Executive summary

The current document discusses the potential extension of Regional Sampling Plan for other Mediterranean and Indian Ocean large pelagic fisheries, further to those studied in Task 2.4 regarding the swordfish in the Mediterranean and tropical tunas in the Atlantic Ocean.

Based on the findings of D.2.4, they have been proposed road maps for establishing regionalized sampling plans for the Mediterranean albacore fishery. Furthermore, slight modifications of the already developed R-scripts will facilitate the extension of the sampling schemes in other fisheries with similar characteristics.

For large pelagic fisheries targeting tropical tunas in the Indian Ocean, the current protocol share between France and Spain is already applied in both ocean (Atlantic and Indian). Recommendations and advice made in the D.2.4 are relevant for the Indian Ocean purse seiner fishery. Furthermore, more studies and analyses should be conducted for adopted the current proposal (especially to define the minimum number of samples required). All the data need and priority define in the D.2.1 are valid and all the hierarchies of the Regional Data Base and Estimation System tested under the D.2.5 for data management are relevant for the Indian Ocean.

1. Introduction

The project MARE/2016/22: “Strengthening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species” (acronym RECOLAPE), aims to propose solutions to support RFMO’s fishery management and generally improve large pelagic fisheries data collection.

The WP2 of this project aims to make a proposal for a Regional Sampling Plan for large pelagic fisheries with a focus on two studies cases: Mediterranean swordfish targeted by longline fisheries and major tropical tunas in the Atlantic Ocean targeted by purse seiner fishery. Furthermore, optimal modifications of the existing sampling schemes have been suggested. The proposed schemes were compared to the current situation aiming to identify existing sampling deficiencies. Moreover, they have been addressed questions related to harmonized reporting of data from the different fisheries. The aforementioned aspects have been discussed in Deliverable 2.4.

Based on the findings of D2.4, the current document attempts to compile a basic set of rules and recommendations that could help to establish regionalized sampling plans to other large pelagic fisheries with similar characteristics, further to the examined ones in the Mediterranean Sea and in the Indian Ocean.

2. Large pelagic longline fisheries in the Mediterranean Sea

Drifting longlines of various types are widely used in the Mediterranean for fishing large pelagic species, such as swordfish and tunas, and various national fleets are involved in those fisheries. By far, the most important longline fisheries are those targeting swordfish, as they operate all over the basin throughout the year with the exception of specific seasonal closures established by ICCAT. Other longline fisheries are those targeting bluefin tuna and albacore. The bluefin tuna fishery is strictly controlled and, through the existing ICCAT regulations, full details of all catches are recorded. The situation, however, is different for the albacore fishery, which is poorly monitored. As a result, stock estimates are based on data poor methods and the state of the stock is highly uncertain (ICCAT, 2019). Hence, the development of a harmonized sampling plan is essential for the management of the albacore stock.

The albacore fishery shares similar characteristics with the swordfish fishery concerning the distribution of the fishing activities and the numerous national fleets involved.

Besides, similarly to swordfish, a Mediterranean-wide stock is assumed for management purposes. Hence the approach followed in D2.4 could be used for monitoring catches and collecting data on their size composition. For the development of a biological sampling plan for the Mediterranean albacore fisheries, the following steps are suggested:

- Identification of exploitation patterns for the national fleets involved in the albacore fisheries.
- The coefficient of variation (CV) estimates for different sample sizes by appropriate region and season, based on the approach followed in D2.4.
- Determination of optimal sample sizes from the CV vs sample size curves.
- In case of spatial overlapping among different national fleet segments, the required sample size for the given region should be split among fleets based on their catch volume.

The already developed R-scripts, with slight modifications, would facilitate the realization of the above goals.

3. Large pelagic fisheries targeting tropical tunas in the Indian Ocean

In the Indian Ocean for 2017, 95% of the total catches of tropical tunas made by the European fleet (Bulgaria, France, Germany, Italy, Portugal, Spain and the United Kingdom) were associated with the purse seiner fishery. The remaining part of the total catches was mostly composed of longline targeting swordfish, costal longline or hand lines represented the last part of total catches (nominal catch information from ICCAT statistical databases). Therefore, testing the possibility of transferring the proposed RSP to the purse seiner fishery in the Indian Ocean seems to be relevant.

The current protocol shared between France and Spain (core work of the proposal RSP for tropical tunas) is already applying in the Atlantic Ocean but also in the Indian Ocean. Major proposals for improvement made in the D.2.3 and 2.4 are relevant for the case of the Indian Ocean. For example, sampling design was hierarchically structured by tuna sampling zone, quarter and school types to define strata as homogeneous as possible in terms of species composition and size distribution (Pianet et al., 2000). This kind of stratification did not get used anymore to the reality of the fisheries and ask the question of representativeness of the species compositions. For example, the fishing ground evolved from year to year whereas sampling zones remained fixed. Further analyses were necessary to adjust the method and recommendations to the

Indian purse seine fishery (as it was done in the D.2.4), but apply the new hierarchical design by sampling according to a regular grid, in size to define, instead of the current zonation should improve the RSP accuracy (figure 1).

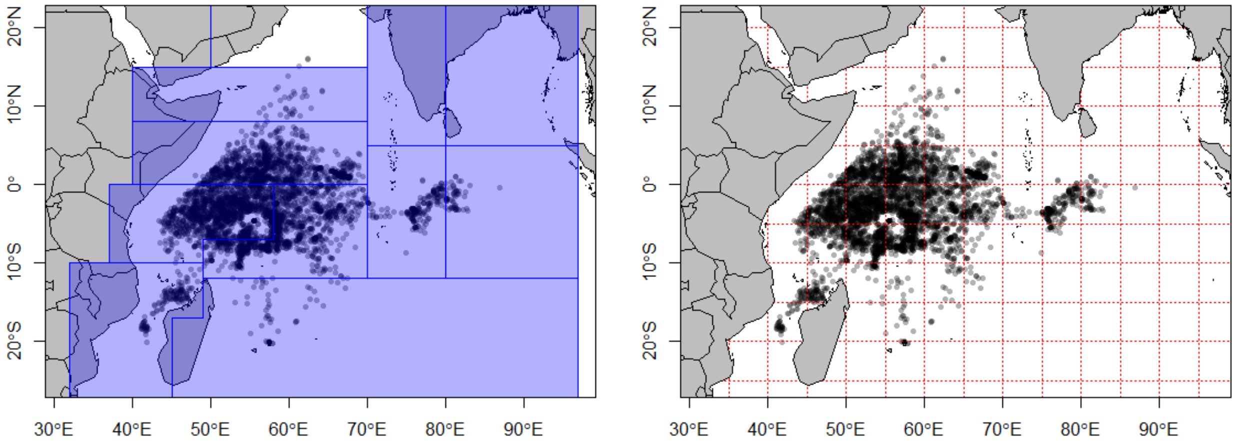


Figure 1: Current sampling zones used for free schools in the current protocol for the Indian Ocean (left panel) with an example of a regular grid of 5° square. Black points represent catch locations of the French fleet in 2017.

Globally and to give a framework a next potential improvement, the current RSP proposal for purse seiner fleet in the Indian Ocean should:

- Try to work on a grid of the smallest resolution as possible considering the allocated sampling effort and cost implication.
- After design and selected the best grid, the square which has been the least sampled should be sampled in priority.
- Each sampling square should be sampled several times during the period it is exploited: by quarter at least but could be monthly for the densest catch area.
- Analysis is necessary to define the minimum number of samples required. The same methodology applied in the D.2.4 could be used in Indian Ocean data.

Otherwise, all the data need and priorities identified in the D.2.1 are according to statistical data mandatory by the IOTC (IOTC, 2015) and so, no additional data are necessary, the recommended one except. With a focus on data priority, 8 group elements are defined:

- quantities of dead discards,
- quantities of bycatch released alive,
- dataset of catch at size estimations,
- data on support vessel activity,
- number of Fish Aggregating Devices (FADs) deployed by support vessels,

- information about maturity,
- information about age,
- data from the local market (so-called "faux poisson").

Furthermore, the existence and the achievement of the RSP proposal in the Indian Ocean for tropical tunas need a regionally coordinated database, as the proposal of the RBDES (Regional Data Base and Estimation System) tested in the D.2.5. Data collected on tropical tunas have exactly the same structure in the Atlantic Ocean and in the Indian Ocean, and fit with the same hierarchies describe in the document D.2.5.

4. References

ICCAT, 2019. Report for sbiennial period 2018-19 Part I - Vol. 2. ICCAT, Madrid, Spain.

IOTC, 2015. Resolution 10/02 Mandatory statistical requirements for IOTC Members and Cooperating Non-Contracting Parties (CPC's).

Pianet, R., Pallarés, P., Petit, C., 2000. New sampling and data processing strategy for estimating the composition of catches by species and sizes in the European purse seine tropical tuna fisheries 37.

FRAMEWORK CONTRACT – MARE/2016/22 "Strengthening regional cooperation in the area of fisheries data collection" Annex III "Biological data collection for fisheries on highly migratory species"

Project acronym: RECOLAPE

Participating partners:

Deliverable D 3.1.1- Minimum data field that can serve for all RFMOs

Deliverable coordinated by Jose Carlos Báez (IEO) and Maitane Grande (AZTI)

Index

Executive summary 3

1. Introduction 5

2. Tuna RFMOs requirements 6

   2.1. International Commission for the Conservation of Atlantic Tunas 6

   2.2. Indian Ocean Tuna Commission 7

   2.3. Inter-American Tropical Tuna Commission 8

   2.4. Western and Central Pacific Commission 9

3. Best standards for Data Collection 9

   3.1. Template format: 9

   3.2 Data to be recorded: 10

4. Best standards for Reporting Requirement 10

   4.1 Format of the templates: 11

   4.2 Definition of terms: 11

   4.3. Data to be requested: 11

References 11

Executive summary

This deliverable is part of the work package 4 Task 4.1 within the project MARE/2016/22: “Strengthening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species” (acronym RECOLAPE). This project seeks to provide solutions to certain needs, in terms of data collection, identified both by the scientists involved in the stock assessment of the tuna RFMOs (Regional Fisheries Management Organizations) and by the RCG-LP (Regional Coordination Group – Large Pelagic), including the development of protocols for collecting new data needs identified by end users around the FADs (fish aggregating devices).

The use of FOBs has continuously increased in tropical tuna purse seine fishery, with FOB-associated catches now exceeding those on free schools in the case of the European Fleet. Despite the importance of this fishery, little information is available on FOB use worldwide which is crucial for the understanding, monitoring and management of FOBs use and the impacts on pelagic ecosystems. As a result, t-RFMOs have called for FAD management plans, including data collection and reporting on deployment and use of FOBs by purse seiners and supply vessels. For example:

- IOTC: Res. 15/01; Res. 15/02; Res. 18/08, and provides specific form for reporting of FOB statistics, 3FA form;
- ICCAT: Rec. 16-01, para 21, Annex 2 FAD logbook form, Annex 3 on the nomenclature of FADs and activities; and Rec. 16-01, para 22, Annex 4 form [list of deployed FADs and buoys], and has developed and updated the ST08-FadsDep form for CPCs for data submission to the RFMO on activities with buoys and FOBs (ST08A) and buoy density (ST08B);
- IATTC: Resolution C-18-05 (Article 2 and Annex I) and C-17-02 established data collection and reporting requirements for purse seiner vessels operating with FADs, has developed and update the FAD Form 09/2018 for skippers, and request information on operational buoys through the INF1 and INF2;
- WCPFC: Collects information on FOB activities through the fishing logbooks and the Regional Observer Programs).

Although efforts are being made to record and report information on FOBs due to the complexity of this fishing strategy and the lack of unified data collection and reporting requirements (e.g. an absence of harmonized definitions for relevant terms or ambiguity among t-RFMOs), there are significant data gaps (Ramos et al., 2017; Lopez et al., 2018) and the information collected and reported has been of limited utility. Recently, several works have addressed collection and submission related problems to propose potential solutions, such as interpretations of requirements and new templates for the data collection/reporting (Báez et al., 2017a; Báez et al., 2017b; Ramos et al., 2017; Gartner et al., 2016; Grande et al., 2018; Lopez et al., 2018). Some of these proposals have been implemented by some users, regionally or at t-RFMOs level, but not at a global level. However, standardization among CPCs and t-RFMOs would be highly desirable. The aim of the present work is to present the proposals for a standardize FOB-related data collection and submission to tRFMOs. The proposals included in this document are the result of a collaborative work between scientists and the fishing industry,

<u>Best Standards for Data Collection on FOBs by skippers</u>

Skippers should collect information on FOBs by the use of FOB logbook on board. All interaction with FOBs (FADs or other floating objects) and buoys if present, should be recorded in the logbook. The record of each activity should provide information on the vessel, trip ID, date and time (GMT), position, buoy attached if present (including the ID of the manufacturer and ownership), type of activity, specifications on the FOB type (the information provided should allow the scientists classifying the activities and FOBs types in CECOFAD categories), and structure of the FOB allowing the assessment of the dimension, entangling character (given by the mesh size if present and configuration) and nature of the material in the floating and submerged structure, as well as the catch of fishing sets (i.e. target species and bycatch) when applicable. Some purse-seine vessels work in collaboration with other purse seiners and/or with supply vessels. In these cases, every vessel should register its own activities, even when they are supporting other vessels (Ramos et al., 2017). If vessels working in collaboration are of different flag states, the details on activities should be shared with the corresponding CPC or t-RFMOs.

If excel files are used for on-board data collection, we recommend using a unique form to record all activities on FOBs as proposed by Ramos et al. (2017). This will require to eliminate the second form or FOB inventory form which was previously used in the Spanish FAD Management Plan and it is now used in the IATTC area, which has been shown to be of limited utility. FOB inventory form was not a practical tool on board as it requires a daily update of the list, and hardly provided good quality data (Ramos et al., 2017). The information of the FOB dynamics (including activities and materials used in the construction) could be deduced from the FOB activity form (if detailed information is given in each record). On the other hand, in case of purse seiners with Electronic Reporting System (ERS) the FOB logbook and fishing logbook should be linked to minimize the errors due to double recording.

Best Standards for data reporting requirements to t-RFMOs

Based on previous experiences (Báez et al., 2017a; Báez et al., 2017b; Ramos et al., 2017; Lopez et al., 2018) and data sources, the group recommends using two specific templates adjusted to the data collections sources (FOB logbook vs. buoy tracks~~)~~:

- One dedicated form to report activities on FOBs and buoys. The information should be derived from FOB logbooks. The activities with buoys and FOBs, as well as FOB types should be in line with CECOFAD categories. Aware of the difficulties of logbook analysis we recommend reducing the request to certain activities: deployment, tagging and loss (CECOFAD categories), until the development and implementation of a standardized data collection tool is available.
- A second template dedicated to report information of density of FADs, which should be derived from buoys transmission information. Information on buoy density should be requested stratified by month and 1ºx1º (i.e. average number of operational buoys belonging to the vessels over the month and 1ºx1º, by summing up the total number of operational buoys recorded per day over the entire month in each grid and dividing by the total number of days in the month). This information should be extracted from buoy transmissions provided by buoy manufactures and not from FOB logbooks.

1. Introduction

Tropical tuna purse seiners operate globally fishing on free schools and on Floating Objects (FOBs), including man-made Fish Aggregating Devices (FADs) and other floating objects. Since the late 90s with the development of satellite-linked echo-sounder buoys for tracking FOBs (Lopez et al. 2014), the use of FOBs has continuously increased (Fonteneau et al., 2013), with FAD-associated catches now exceeding those on free schools in the case of the European Fleet. For example, the European tropical tuna purse seine fishery operating in the Indian Ocean has increased the percentage of FOB sets from 40% in 1990-1994 to 73% in 2010-2014 (Chassot et al., 2015, Ramos et al., 2017), following similar trend in the Atlantic Ocean. In this document the term Floating Objects (FOBs), includes the man-made Fish Aggregating Devices (FADs) and other floating objects (Gaertner et al., 2016).

The increasing use of FOBs has introduced worldwide major changes in the tropical tuna purse seiners fishing patterns which could have affected the marine environment. In this sense, potential effects associated with the increased number of FOB deployments at sea has been described: alteration of normal movements of tuna (Marsac et al., 2001; Hallier and Gaertner, 2008), increased skipjack catches (the principal target species), reduction in yield per recruit of yellowfin and bigeye (from which small specimens co-occur in the catches with skipjack), increase in bycatch, potential impacts on coastal habitats and source of pollution (Dagorn et al., 2012, Maufroy et al., 2015, Davies et al., 2017). Despite these concerns, little information is available on FOB use worldwide while it is crucial for the understanding, monitoring and management of the impacts of FOBs on pelagic ecosystems. As a result, Tuna Regional Fisheries Management Organizations (t-RFMOs) have called for FAD management plans, including data collection on deployment and use of FOBs by purse seiners and supply vessels and data reporting requirements on FOBs to CPCs/t-RFMOs (ICCAT, 2016a, 2016b).

Although efforts are being made to record and report information on FOBs, including man-made FADs and other natural floating objects, due to the complexity of this fishing strategy and the lack of unified data collection and reporting requirements (an absence of harmonized definitions for relevant terms or ambiguity among t-RFMOs), there are significant data gaps (Ramos et al., 2017; Lopez et al., 2018) and the information collected so far by the skippers and available for analysis has been of limited utility. Several works have been conducted recently to analyze data collection and submission related problems and have proposed potential solutions, such as interpretations on the data collection and submission requirements or new FAD logbook templates to improve the quality of the data recorded (Báez et al., 2017a; Báez et al., 2017b; Ramos et al., 2017; Lopez et al., 2018). Some of these proposals have been implemented regionally or by some users. However, standardization among CPCs and t-RFMOs would be highly desirable. Therefore, efforts from all stakeholders are required to improve data collection and submission on FOBs. In this sense, the RECOLAPE project (MARE/2016/22, “Strengthening Regional cooperation in large pelagic fisheries data collection”), which seeks to improve the coordination among EU Member States in the fisheries data collection field in support of stock assessment and fisheries advice, aims to develop protocols for FOB data collection and data storage tools to meet the requirements of the tuna t-RFMOs. The aim of the present document is to summarize the results of the workshop which took place in the frame of RECOLAPE project during 24th and 25th of May in AZTI (Sukarrieta) in which t-RFMO requirements and other procedures in place were reviewed and standards for the collection and submission of FOB-related data were proposed. The proposals included in this document are the result of a collaborative work between scientists and the fishing industry.

2. Tuna RFMOs requirements

t-RFMOs have called for FAD management plans, including data collection on deployment and use of FOBs by purse seiner and supply vessels, and data reporting requirements on FADs to CPCs/t-RFMOs (**Table 1**). Recent works reviewed these t-RFMOs requirements including a detailed analysis of the data gaps, data requested on FAD-logbooks and other data submission forms (Báez et al., 2017a; Báez et al., 2017b; Ramos et al., 2017, Lopez et al., 2018), which are not repeated here. We briefly summarize and discuss the issues detected in each t-RFMO.

2.1. International Commission for the Conservation of Atlantic Tunas

The International Commission for the Conservation of Atlantic Tunas (ICCAT) through Recommendation 16-01 (Rec. 16-01, 21, Annex 2 form [FAD logbook], Annex 3 on the nomenclature of FADs and activities; and Rec. 16-01, 22, Annex 4 form [list of deployed FADs and buoys]), proposed specific forms for data collection on FOBs including CECOFAD codes for type of floating objects and activities. In these forms an identification code is proposed for marking the FOBs in addition to the buoy ID. This marking scheme was previously applied with not promising results, and therefore the 2nd FAD Working Group of ICCAT concluded that the FADs should be marked/tracked by the buoy unique ID attached to the FAD (given by the buoy manufacturer), recording in the logbook details of all changes (ICCAT 2016a, Ramos et al., 2017). On the other hand, in ICCAT two templates are provided for recording activities with FOBs, instead of one, as proposed by Ramos et al. (2017). In this sense the forms included in the Annex 2 and 4 (Rec. 16-01) are not in line with the recommendations made from previous experience and reviews on data collection (ICCAT 2016a, Ramos et al., 2017). ICCAT recommendations also establishes the obligation by CPCs to provide data on FOBs. According to the management recommendations: Rec. 16-01, Rec. 13-01(paragraph 2), ICCAT developed ST08-FadsDep form for data submission to the t-RFMO. Paragraph 23 of Rec. 16-01 requested that the CPCs should provide to the t-RFMO information on (i) the number of deployed FADs with and without beacon, (ii) the average number of active beacons, (iii) the average number of deactivated beacons followed per vessel, (iv) the average number of active lost and (v) the number of FADs deployed by support vessel by month, 1 x 1 square (this spatial stratification only specified for some data), FAD and beacon type.

During the 2nd FAD Working Group of ICCAT, the ICCAT Secretariat provided the data received so far from Form ST08 regarding FAD deployments. The Secretariat highlighted that very few CPCs provided data using the recently modified ST08 forms. In addition, several problems with the received submissions were noted. In one case information was provided by 5º x 5º rather than 1º x 1º degree squares, which may be due to a misinterpretation, as the spatial stratification is not specified for all data requested (i.e. number of buoys activated and deactivated) (Báez et al., 2017a). This provides an idea of the problems in FAD data submission and underlines the need for standardization and homogenization of the criteria for filling the forms.

In relation to this, Báez et al. (2017a) summarizes the interpretation of EU-Spain with regards to the ICCAT's data reporting requirements for activities on FADs aiming to describe the difficulties, posing questions and providing interpretations on the FAD data collection requirements under ST08-Rec 16-01 to allow standardizing the data collection and reporting of FAD information.

The main observations and recommendations from Báez et al. (2017a) were:

- Harmonization of the request made in the Rec. 16-01 under paragraph 23 and the file ST08 FAD Form provided to CPCs to report the data, taking into account the data collection mechanism available.
- Definition of terms and detailed description of each field (i.e. deployed FAD, active beacon, deactivated beacon, lost beacon)
- Harmonization between required information and codes between different Regional fisheries management organizations (t-RFMOs) (e.g. FAD and beacon types)

2.2. Indian Ocean Tuna Commission

The Indian Ocean Tuna Commission (IOTC) through IOTC´s Resolution 13/08[1] includes standards for the collection and reporting of data on fishing activities around FOBs, both drifting and anchored, undertaken by purse seine and pole-and-line fisheries. This resolution has been reviewed and updated by Resolution 15/08 superseded by 17/08, and then by Resolution 18/08. Resolution 18/08 stablish guidelines for FOBs management plans including more strict limitations on the numbers of FOBs, more detailed specifications of data collection from visits to FOBs (Annex I of Res. 18/08) including date, position, identifier, FOB type, design, type of visit and catch if the visit is followed by a set. In addition, Resolution 15/01 (which superseded Res. 13/03) on the recording of catch and effort data for fishing vessels aims to harmonize data collection and to further monitor FOBs use. It also defines minimum requirements on data collection on FOBs deployments and sets on FOBs (Annex I and II of Res. 15/01). Although minimum requirements on data collection are provided, none of the resolutions presents specific forms for data collection on FOBs to be used onboard.

Currently, as specified in Resolutions 15/02 and 18/08, and according to the guidelines for the reporting of fishery statistics to the IOTC (Form 3FA, IOTC Secretariat, 2014), CPCs must provide catch-and-effort data in relation to: (i) total number (by type) of FADs deployed by purse seiners and support vessels by month/quarter and fleet, (ii) effort data expressed as the total number of FOB visits per type of FOB, type of visit, 1° grid area and month; and (iii) total catches of target IOTC species and bycatch species taken on FOBs, at the same level of resolution. However, some of the information requested is unclear and the requirements are not harmonized in Resolution 18/08 and Form 3FA (e.g., spatial stratification, or interpretation of the types of visits) (Báez et al., 2017b). The ambiguity in the interpretation of FOB data requirements may result in the development of FAD logbooks not adjusted to the requirements.

[1] "Procedures on a fish aggregating devices (FADs) management plan, including more detailed specifications of catch reporting from FAD sets, and the development of improved FAD designs to reduce the incidence of entanglement of non-target species". For the purposes of this Resolution, the term "Fish-Aggregating Device" (FAD) means anchored, drifting, floating or submerged objects deployed and/or tracked by vessels, including through the use of radio and/or satellite buoys, for the purpose of aggregating target tuna species for purse-seine fishing operations.

Báez et al. (2017b) described the difficulties, raised questions and provided interpretations on the FOB collection requirements under Form 3FA to allow standardization among the data submission. Finally, this paper proposes a reorganization of Form 3FA, using CECOFAD conclusions for FOB types and activities.

2.3. Inter-American Tropical Tuna Commission

The Inter-American Tropical Tuna Commission (IATTC) through resolutions C-18-05 (Article 2 and Annex I) and C-17-02 established data collection and reporting requirements for purse seiner vessels operating with FADs on the IATTC Convention area. From 1st of January 2017 the skippers shall collect, and report information contained in the Annex I of the C-18-05 which referred to activities with FADs, including position, date, hour, FAD identification, FAD type, FAD design characteristics, type of the activity, the result of the catch when resulting in a set, and buoy characteristics if any attached to the FAD. To record this information, the working group on FADs designed and proposed a FAD form to be used on board (i.e. IATTC Form: FAD Form 9/2016 which have been recently updated with the FAD Form 09/2018). This new form is composed by two files, one dedicated to record activities on FADs (following the requirements stablished in C-18-05, Annex I) and a second one which should be used as an inventory of active FADs including specifications of the raft and hanging structure. In these IATTC forms, a unique identification is given to FADs, being allowed to use the buoy ID attached or to follow the FAD identification scheme proposed by the IATTC which assigns an independent ID for each FAD. This form structure (activity and inventory in separate forms) and using and independent ID for FADs is not in line with the recommendations made from previous experience and reviews which aim to simplify and adapt the form to be use on board (ICCAT, 2016c; Ramos et al., 2017).

During 2017, with the establishment of new measures for FADs including limits on the number of active FADs (as refer in the resolution), new reporting requirements were designated (C-17-02). From 1st of January of 2018 CPCs shall report monthly to the Secretariat, with a delay between 60 to 90 days, daily information of all active FADs following the guidelines established by the *Ad Hoc* Permanent Working Group on FADs. In this sense, two files should be reported, which are still under discussion (Lopez et al., 2018), including information about the number of active buoys per vessel and day, and a monthly summary of the activated, deactivated and average number of active FADs followed by vessel and 1º square grid (INF1 and INF2, respectively). The information used to monitor the number of active FADs should be provided by the FAD tracking services directly to the designated verification body of each CPC (and/or to the IATTC staff if so requested by the CPC).

Lopez et al., (2018) recently reviewed the data collection and reporting requirements identifying data gaps regarding FAD logbooks and active FAD information. The IATTC proposed modifications in the CIAT Form 09/2016, which has been conducted in the FAD Form 09/2018, aiming to collect detailed data on FOB (as information about buoys-swapping, re-deployment, including activities with natural objects). However, the form maintains two files (activity and inventory form) and an independent marking scheme for FADs and buoys. To standardize and improve the data collection on FOBs as described in the C-18-05 (Article 2 and Annex 1) and reporting to IATTC, this t-RFMO proposes a web application as data collection tool (Lopez et al., 2018). Finally, aiming to assess the compliance with the C-17-02, the provision of fine scale buoy transmission data from buoy manufactures and VMS data are recommended.

2.4. Western and Central Pacific Commission

In the case of the Western and Central Pacific Commission (WCPFC), new FAD/buoy control measures are in force limiting the number of activated instrumented buoys attached to FADs at any given moment to 350 (CMM 2017-01). There are not specified FOB logbooks for skippers and for data submission to the t-RFMO. The master of each vessel shall ensure that information on relevant activities with FADs are recorded in the logbook as requested in CMM-2013-05. The fishing logbook (SPC / FFA Regional Purse-Seine Logsheet) give the possibility to collect some activities with FOBs (e.g. Investigate floating object; Deploy - raft, FAD or payao; Retrieve - raft, FAD or payao) and have the option to characterize the FOB (drifting log, debris or dead animal"; "drifting raft, FAD or payao"; "anchored raft, FAD or payao"; "live whale"; and "live whale shark"). Since 2010, purse seine vessels operating in the Convention Area of this t-RFMO have a 100% observer coverage (as established by CMM2008-01 and following Conservation and Management Measures). The Regional Observer Program includes data collection on FOB activities (WCPFC 2017).

3. Best standards for Data Collection

The lack of unified criteria among t-RFMOs on FOBs data collection, specific guidelines and a standard and simple templates for the fleet has resulted in a non-harmonized data collection; which hampers its use for scientific purposes (Ramos et al., 2017). During 2016 and 2017 various works were conducted and presented in t-RFMOs´ working groups to address the problem (Gaertner et al., 2016; Báez et al., 2017a; Báez et al., 2017b; Ramos et al., 2017). Specific details requested by the t-RFMOs are reviewed and discussed, and best standards for data collection are proposed for each requirement.

3.1. Template format:

The forms propose among t-RFMOs (i.e. ICCAT 16/01 – Annex 2 and Annex 3; and IATTC FAD Form 09/2018) are not harmonized and not in line with the recommendations made from previous experience and reviews (ICCAT 2016a, Ramos et al., 2017), which proposed to simplify the marking scheme and structure of the form. When excel files are proposed for data collection, we recommend using a unique form to record all activities on FOB, including detailed information as date, hour, position, buoy identifier, ownership, FAD Type, FAD design and catch if the visit is followed by a set as proposed by Ramos et al., 2017; and eliminating the second form or inventory which was previously used in the Spanish FAD Management Plan with limited used. This inventory was designed to record the relation and design, or type of the FOBs used. However, it is not a suitable tool to be used on board as it requires a daily update of the list, and hardly provided good quality data (Ramos et al., 2017). Moreover, the information of the dynamics of FOB use can be deduced from the FOB activity form (if detailed information is given in each record) and information on buoy transmissions if they are made available for scientific purposes to the research institutions or bodies responsible for the verification of compliance with buoy limitations in force. In this situation, the inventory does not provide additional relevant information and, thus, it could be removed to facilitate data collection on board.

On the other hand, in case of purse seiners with Electronic Reporting System (ERS) the FOB logbook and fishing logbook should be linked somehow to minimize the errors due to double recording.

3.2 Data to be recorded:

All interaction with FOBs (FADs or other floating objects) and buoys if present, should be recorded in the FOB logbook while only sets should be recorded on the fishing logbook.

The record of each activity should provide information on buoy attached if present (including the ID of the manufacturer and ownership), specifications on the FOB type and structure allowing the assessment of the entangling and nature of the material, as well as the occurrence and catch of fishing sets, when applicable. Overall, the information provided should also allow the scientists classifying the activities and FOBs in CECOFAD categories (Gaertner et al., 2016).

Some purse-seine vessels work in collaboration with other purse seiners and/or with supply vessels. In these cases, every vessel should register its own activities, even when they are supporting other vessels (e.g., deployment of buoys for another vessel) (Ramos et al., 2017). If vessels working in collaboration are of different flag states, the details on activities or on the collaboration should be shared with the corresponding CPC or RFMOs for effort assessment.

Details of each specific information to be collected are included in the tables of this document document. The tables include details of the information required by the t-RFMOs (IATTC, ICCAT, IOTC, and WCPFC) regarding the marking scheme (Table 2), spatial and seasonal dynamics (Table 3), FOB type (Table 4), FOB structure (Table 5), activity with FOB and buoys (Table 6), and information on the fishing set/catch (Table 7) and other requirements (Table 8). In each case, best standards for data collection and minimum details to be recorded are proposed for a standardize data collection in each case.

4. Best standards for Reporting Requirement

The t-RFMOs aiming to assess the effort on FOBs and potential impacts have strength the data reporting requirements and, in some cases, specific templates has been provided to CPCs for data submission on FOBs. However, some data gaps have been identified for the different t-RFMOs (Báez et al., 2017a; Báez et al., 2017b; Ramos et al., 2017; Lopez et al., 2018), indicating a generalized problem in data collection and reporting schemes stablished. Some of the potential sources of un-reporting are identified as un-harmonized spatial and temporal stratification of the data required, misinterpretation of the request due to un-specific guidelines, lack of definitions of the terms and variables to be recorded, or complex templates where information extracted from different sources cannot be integrated in a single template (i.e., information from FOB or FAD logbooks vs. information from buoy transmissions).

In order to provide the t-RFMOs with good quality information on FOBs and facilitate CPCs the collection and submission of data, we reviewed the t-RFMO data reporting requirements and identified best standards for the spatial and temporal resolution requested (Table 9), floating object types (Table 10), activities with floating objects (Table 11), activities with buoys (Table 12), FOB number or density (Table 13); information on the catch on FOBs (Table 14).

4.1 Format of the templates:

Regarding to the previous experiences the group recommends using two specific templates adjusted to the data collections sources (FOB logbook vs. buoy tracks~~)~~: one dedicated form to report activities on FOB (based in CECOFAD categories) which are extracted from the FOB or FAD logbooks; and another template dedicated to report information on number or density of followed and/or owned buoys or FADs, which is extracted from buoys transmission information (examples are included in the Annex 1 and 2 of this document, following those proposed by a small working group that met during the ICCAT SCRS 2018 meeting).

4.2 Definition of terms:

The activities with buoys and FOBs, as well as FOB types should be in line with CECOFAD categories.

4.3. Data to be requested:

The information on buoy density should be requested stratified by month and 1ºx1º. This information should be extracted from buoy transmissions provided by buoy manufactures and not from FAD or FOB logbooks. It should be requested by all t-RFMOs.

The data on FOB and buoy activities should be extracted from FOB logbooks. This information should be requested in an independent template. The group aware of the difficulties of logbook analysis and recommends reducing the request to certain activities: deployment, tagging and loss (CECOFAD categories), until the development and implementation of a standardized data collection tool is available and implemented.

References

Báez, J.C., Ramos, M.L., López, J., Santiago, J., Grande, M., Herrera, M.A., Rojo, V., Moniz, I., Muniategi, A., Pascual, P.J., Murua, H., Abascal, F.J., 2017a. Interpreting ICCAT's data reporting requirements for activities on FADs: An overview from EU-Spain. Documento de trabajo presentado en el

Standing Committee on Research and Statistics (SCRS) de ICCAT, número SCRS/2017/217. Madrid 25-29 de septiembre 2017.

Báez, J.C., Bach, P., Capello, M., Floch, L, Gaertner, D., Goujon, M., Grande, M., Herrera, M.A., López, J., Marsac, F., Maufroy, A., Moniz, I., Muniategi, A., Murua, H., Pascual, P.J., Ramos, Mª.L., Rojo, V., Sabarros, P.S., Santiago, J., Abascal, F.J., 2017b. Interpreting IOTC`s data reporting requirements for activities on floating objects: an outlook from EU scientist and fishing operators. Submitted to 13th Working Party on Data Collection and Statistics. IOTC-2017-WPDCS13-27.

Chassot, E., Assan, C., Soto, M., Damiano, A., Delgado de Molina, A., 2015. Statistics of the European Union and associated flags purse seine fishing fleet targeting tropical tunas in the Indian Ocean 1981-2014. IOTC-2015-WPTT17-12. 17th session of the IOTC Working Party on Tropical Tunas, Montpellier, France.

Dagorn, L., Holland, K. N., Restrepo, V. and Moreno, G., 2012. Is it good or bad to fish with FADs? What are the real impacts of the use of drifting FADs on pelagic marine ecosystems? Fish and Fisheries, 14: 391–415.

Davies, T. K., Curnick, D., Barde, J. and Chassot, E., 2017. Potential environmental impacts caused by beaching of drifting Fish Aggregating Devices and identification of management solutions and uncertainties. 1st joint t-RFMO FAD WG.

Fonteneau, A., Chassot, E. and Bodin, N., 2013. Global spatio-temporal patterns in tropical tuna purse seine fisheries on drifting fish aggregating devices (DFADs): Taking a historical perspective to inform current challenges. Aquatic Living Resources, 26: 37–48.

Gaertner, D., Ariz, J., Bez, N., Clermidy, S., Moreno, G., Murua, H., Soto, M., Marsac, F., 2016. Results achieved within the framework of the EU research project: Catch, Effort, and eCOsystem impacts of FAD-fishing (CECOFAD). IOTC-2016-WPTT18-35. 18th Session of the IOTC Working Party on Tropical Tunas, Mahé, Seychelles.

Hallier, J.P. and Gaertner, D., 2008. Drifting fish aggregation devices could act as an ecological trap for tropical tuna species. Marine Ecology Progress Series 353: 255-264.

ICCAT, 2016a. Second meeting of the ad hoc Working Group on FADs (Bilbao, Spain, 14-16 march 2016).

ICCAT, 2016b. Recommendation by ICCAT on a Multi-Annual Conservation and Management Program for Tropical Tunas. Rec. 16-01.

Lopez, J., Moreno, G., Sancristobal, I., and Murua, J., 2014. Evolution and current state of the technology of echo-sounder buoys used by Spanish tropical tuna purse seiners in the Atlantic, Indian and Pacific Oceans. Fisheries Research, 155: 127–137.

Lopez, J., Altamirano, E., Lennert-Cody, C., Maunder, M., Hall, M., 2018. Review of IATTC resolutions C-16-01 and C-17-02: available information, data gaps, and potential improvements for monitoring the fad fishery. 3rd Meeting of the Ad Hoc Working Group on FADs La Jolla, California USA, 11-12 May 2018

Marsac F, Fonteneau A, Ménard F., 2000. Drifting FADs used in tuna fisheries: an ecological trap? In: Le Gall JY, Cayré P, Taquet M (eds) Pêche thonière et dispositifs de concentration de poisons. Actes Colloques-IFREMER 28:537–552

Maufroy A, Chassot E, Joo R, Kaplan D.M., 2015. Large-Scale Examination of Spatio-Temporal Patterns of Drifting Fish Aggregating Devices (dFADs) from Tropical Tuna Fisheries of the Indian and Atlantic Oceans. PLoS ONE 10(5): e0128023. doi: 10.1371/journal.pone.0128023

Ramos, Mª.L., Báez, J.C., Grande, M., Herrera, M.A., López, J., Justel, A., Pascual, P.J., Soto, M., Murua, H., Muniategi, A., Abascal, F.J., 2017. Spanish FADs logbook: solving past issues, responding to new global requirements. Joint t-RFMO FAD Working Group meeting Doc. No. j-FAD_11/2017. April 19-21, 2017 Madrid, Spain.

Table 1. t-RFMO data collection and reporting requirements on FOBs

t-RFMOs	Data Collection Requirements	Data Reporting Requirements
IOTC	Resolution 17/08 (para. 10) - [Annex I and Annex 2]. Res. 15/01 - [Annex I and Annex 2] No form provided	Resolution 18/08 (para. 9); Resolution 15/02 (para. 6); Guidelines for the reporting of fisheries statistics to the IOTC - Form 3FA
ICCAT	Rec. 16-01 (para. 21) – [Annex 2 form, FAD logbook; Annex 3, minimum standards]; Rec. 16-01 (para. 22) – [Annex 4 form, list of deployed FADs and buoys]	Rec. 16-01 (para. 23); Rec. 13-01 Form: ST08-FadsDep form
IATTC	C-18-05 (para. 2) Annex I (Amendment of Resolution C-16-01) FAD Form 9/2016 C-17-02	C-18-05 (para. 3) (Amendment of Resolution C-16-01); C-17-02 (para. 11, 12); Guidance in reporting on FADs in accordance with IATTC Resolution C-17-02: INF1; INF2
WCPFC	CMM-2013-05 Report - tenth meeting of the Tuna fishery Data Collection Committee	Not specified in the Resolutions

Table 2. Summary of the identification criteria on activities with FOBs in FOB logbooks as defined by t-RFMOs and the best standards proposed by the group. The identification of each activity with FOBs should be linked with the name of the vessel and IMO number, and starting and end date of the trip. As activities with FOBs could be given between fishing trips (e.g. lost), records between the trips will belong to the next starting trip. Each FOB should be identified by the buoy ID if present. The identification of the buoy in the FOB should be noted (model and identification number) and the ownership of the buoy if known (name of the vessel owing the buoy). The date, time and position of each specific activity (included in the next table) are also crucial for the identification of each record.

General Data	t-RFMOs Data Collection Requirements	IATTC	ICCAT	IOTC	WCPFC	Standards for data collection	Minimum Details
Identification	Vessel	Required	Not required	Required	-	Required	Name of the vessel fulfilling the form and conducting the activity
	nº of trip/ Identification of the trip	Calendar year of the start of the trip and the consecutive number of the trip for that calendar year in the spaces provided. For example:' 2015-001', denotes the first trip in 2015.	Not required	Not required	-	Required	(*) Start of the trip and its end [= when arriving at port], same as in the logbook
	Register number	Required	Not required	Required	-	Required	IMO number
	Identification (of the locating buoy):	Unique identification number of the locating buoy. If this is a satellite buoy, it must be the unique serial number. If it is another type of locating buoy, use a unique identification code self-provided to the FAD or the locating buoy and that could be used as reference for future encounters.	Required	Required	-	ID Buoy required	Model and identification number
	FAD ID	CPCs shall obtain unique alphanumeric codes from the IATTC staff, or in the alternative, if there is already a unique FAD identifier associated with the FAD (e.g., the manufacturer identification code for the attached buoy), the vessel owner or operator may instead use that identifier as the unique code for each FAD that may be deployed or modified.	FAD Marking and buoy ID or any information allowing to identify the owner. If ID are absent or unreadable, the FAD shall not be deployed	DFAD Marking or beacon ID or any information allowing to identify the owner	-	Not required	Given by the buoy identifier
	Other information not requested				-	Ownership required	Name of the vessel owning the buoy if present

(*) *As indicated for the DEA, the fishing activity is considered to be finished with the arrival at port, the unloading document or the end of the trip (http://www.mapama.gob.es/es/pesca/temas/control-e-inspeccion-pesquera/informacion-sobre-actividad-pesquera/preguntas_diario_electronico_pesca.aspx). For scientific issues, the arrival date should coincide with the unloading date and the date registered in the DEA/ERS.*

Table 3. Summary of the seasonal and spatial details requested by the t-RFMOs on activities and the best standards proposed by the group.

General Data	t-RFMOs Data Collection Requirements	IATTC	ICCAT	IOTC	WCPFC	Standards for data collection	Minimum Details
Seasonal and Spatial dynamics	Time	The local time of the event in a 24-hour format (13:00 = 1 pm).	hh:mm	24-hour format, GMT or local time	-	Required	Time* of the activity in UTC (HHMM) If a loss of the buoy, information of the last transmission should be provided
	Position	Write the geographic location of the event (Latitude and Longitude) in degrees and minutes. Note the corresponding hemisphere (N=North, S=South, E=East, W=West).	N/S/mm/dd or E/W/mm/dd In case of loss, last registered position	Not specified format	-	Required	Position* of the activity.
	Date	The date of the event in the format DD/MM/YY (day/month/year)	dd/mm/yy	YYYY/MM/DD	-	Required	Date* of the activity.

***** *If a loss of the buoy, information of the last transmission should be provided*

Table 4. Summary of the Floating Object (FOB) type requested by the t-RFMOs and the best standards proposed by the group. The FOB type should include all types of floating objects and not only FADs. The group recommends recording enough information on the FOB logbook to allow researchers to classify on CECOFAD categories or giving as choice to the fleet the CECOFAD categories (Gaertner et al., 2016): DFAD (Drifting FAD); AFAD (Anchored FAD); FALOG (Artisanal log resulting from human activity, related to fishing activities); HALOG (Artificial log resulting from human activity, not related to fishing activities); ANLOG (Natural log of animal origin); VMLOG (Natural log of plan origin).

General Data	t-RFMOs Data collection Requirements	IATTC	ICCAT	IOTC	WCPFC	Standards for data collection	Minimum Details
FOB TYPE	FAD Type	1. Natural (log, ropes, pallets/racks, fronds, dead animal); 2. FAD owned by your vessel; 3. FAD owned by another vessel; 4. Anchored object	anchored FAD, drifting natural FAD, drifting artificial FAD: DFAD; AFAD; FALOG; HALOG; ANLOG; VNLOG	drifting natural FAD, drifting artificial FAD, anchored FAD	Not specific fad logbook provided. Given in the fishing logbook drifting log, debris or dead animal"; "drifting raft, FAD or payao"; "anchored raft, FAD or payao"; "live whale"; and "live whale shark".	The information collected should allow to classify in CECOFAD codes	CECOFAD codes could be provided by skippers or enough information to allow a posterior analysis on FOBs and classification on CECOFAD codes

Table 5. Summary of details on floating object (FOB) structure requested by the t-RFMOs and the best standards proposed by the group. The information given should allow evaluating the potential of entanglement of the FOB and the nature of the materials (synthetic or natural and/or biodegradable).

General Data	t-RFMOs Data collection Requirements	IATTC	ICCAT	IOTC	WCPFC	Standards for data collection	Minimum Details
FOB Structure	FOB Dimension	Dimensions and material of the floating part (in meters); W –Width -, L –Length–, D –Depth Dimensions of the underwater hanging structure (Not specified format)	Required	Required. Not specified format	-	Dimensions for the floating and hanging structure	Floating structure [aaxbb] (width and length) Hanging structure: depth in m
	Components of the surface structure	Raft: 1. Bamboo Rack; 2. Bamboo in a sausage form; 3. Metallic; 4. PVC or plastic; 5. No raft; 6. Other Wrapping/covering: 1. Entangling net; 2. Non-entangling net; 3. Cloth; 4. Palm fronds; 5. No wrapping; 6. Other Floating devices: 1. Net corks; 2. Plastic buoys; 3. Plastic containers; 4. No floats; 5. Other	Material of the floating part and the entangling or non-entangling feature of the underwater hanging structure	Material of the floating part and of the underwater hanging structure	-	non-entangling character based in ISSF classification scheme and biodegradable character	- Type of material: Natural and biodegradable; or other synthetic materials in the FOB. - Entangling potential of the external mesh size (if present)
	FOB hanging structure (tail)	Components 1 and 2: 1. Nylon; 2. Palm fronds; 3. Bamboo; 4. No tail; 5. Other Config. (Configuration): 1. Sausage; 2. Ropes; 3. Cloth; 4. Other Mesh size: If the tail is made of net, indicate the mesh size. Otherwise, leave blank.	Material of the underwater hanging structure and the entangling or non-entangling feature of the underwater hanging structure	Material of the floating part and of the underwater hanging structure	-	non-entangling character based in ISSF classification scheme and biodegradable character	-Type of material: Natural and biodegradable or synthetic - Entangling potential of the hanging structure (reference to the mesh size and configuration, i.e. open or coiled)

Table 6. Summary of type of activity on floating object (FOB) requested by the t-RFMOs and the best standards proposed by the group. The group recommends recording enough information on the logbook to allow researchers to classify on CECOFAD categories or giving as choice to the fleet the CECOFAD categories (Gaertner et al., 2016). When any part of the FOB is modified, or the buoy or ownership are changed, the specification prior and after the change should be recorded.

General Data	t-RFMOs Data collection Requirements	IATTC	ICCAT	IOTC	WCPFC	Standards for data collection	Minimum Details
FOB Activity	Type of the activity on FOB	Set, deployment, hauling, retrieving, loss, other	Recommends using CECOFAD terms. FOB: Encounter, visit, deployment, strengthening, remove FAD, fishing.	deployment, hauling, retrieving, loss	Not specific FOB/FAD logbook. Given in the fishing logbook as: Set; Searching; Transit; No fishing - Breakdown; No fishing - Bad weather; In port; Net cleaning set; Investigate free school; Investigate floating object; Deploy - raft, FAD or payao; Retrieve - raft, FAD or payao"	CECOFAD activities with FOBs	Recommend using the CECOFAD activities on FOB: Encounter, visit, deployment, strengthening, remove FAD, fishing.
BUOY Activity	Type of the activity on BUOY	intervention on electronic equipment	Recommends using CECOFAD terms. Buoy: Tagging, remove buoy, loss	intervention on electronic equipment	-	CECOFAD activities with buoys	Recommend using the CECOFAD activities on Buoy: Tagging, remove buoy, loss.

Table 7. Summary of details of the catch on Floating Object (FOB) requested by the t-RFMOs and the best standards proposed by the group. The FAD logbook should be preferably linked with the fishing logbook when using ERS or dedicated software for standardize data collection and catch obtained from fishing logbook. The destiny of the catch should be included (i.e. retained, discarded or released in case of sensitive species). If the FAD logbook is not linked with the fishing logbook specific fields for the catch should be included in the FAD form.

General Data	t-RFMOs Data collection Requirements	IATTC	ICCAT	IOTC	WCPFC	Standards for data collection	Minimum Details
Catch	Target species	If the event is a set, the catch in metric tons of each of the tuna species denoted. When the catch includes other tunas (OTH), record the quantities and species under Comments.	If the visit is followed by a set, the results of the set in terms of catch. If the visit is not followed by a set, note the reason (e.g. not enough fish, fish too small, etc.). Estimated catches expressed in metric tons.	If the visit is followed by a set, the results of the set in terms of catch	Not specific FOB/FAD logbook- Reported. in the fishing logbook	Required. Preferably linked to fishing logbook in ERS and obtained from fishing logbook	Target species (tn). Destiny should be included [retained, discarded]. When the catch includes other tunas (OTH), record the quantities and species as bycatch
	Bycatch	For the groups noted (Sharks – SHRK –, Turtles – TURT –, Billfishes – BILL –, Manta rays – MANT – and Other vertebrates – OTR –), present in the set, indicate either the number of individuals (N) or metric tonnage (t) caught. Use the line below to record the quantity of these, released alive.	If the visit is followed by a set, the results of the set in terms of by-catch whether retained or discarded dead or alive (in case of release expressed as number of specimen.). Estimated catches expressed in weight or in number.	If the visit is followed by a set, the results of the set in terms of bycatch.	Not specific FOB/FAD logbook.. Reported in the fishing logbook	Required. Preferably linked to fishing logbook in ERS and obtained from fishing logbook	little tuna; other bony fishes; billfishes; sensible species; (n or tones). Destiny should be included [retained, discarded or released in case of sensitive species].

Table 8. Summary of other requirements referred to the specification of the buoy attached to the FOB requested by the t-RFMOs and the best standards proposed by the group. Buoy technical specifications are given by the buoy model and therefore it is not necessary to include another field different from the one provided to the buoy identification.

General Data	t-RFMOs Data collection Requirements	IATTC	ICCAT	IOTC	WCPFC	Standards for data collection	Minimum Details
Others	Characteristics of any attached buoy or positioning equipment	1. GPS, SHERPE type; 2. Satellite with eco-sounder; 3. Satellite with no eco-sunder; 4. Other	E.g. GPS, sounder, etc. If no electronic device is associated to the FAD, note this absence of equipment	Serial number required	-	Given by the buoy model	

Table 9. Summary of spatial and seasonal resolution requested to CPCs by t-RFMOs on FOB data and the best standards proposed by the group. The guidelines to CPCs for data reporting in terms of spatial and temporal resolution are not specified for all data requested and not harmonized among t-RFMOs, as it refers to 1º or 5º grid square size and to monthly or quarterly basis. This has resulted in a misinterpretation of the request and inadequate submissions of data (Báez et al., 2017a, 2017b). The group recommends the harmonization to 1º grid square and monthly basis.

General Data	t-RFMOs Data Reporting Requirements	IATTC	ICCAT	IOTC	Information extracted from FAD Logbook	Information extracted from Buoys transmissions	Standards for data reporting
Seasonal and spatial distribution	Grid size	1x1	1x1 (but not specified for all data required)	1x1	X	X	Harmonize grid size:1ºx1º
	Time scale	Monthly	Monthly	Is not harmonized. [Monthly and Quarterly]	X	X	Harmonize time scale to a monthly basis

Table 10. Summary of the Floating Object (FOB) type requested to CPCs by t-RFMOs and the best standards proposed by the group. The information on FOB types described in each t-RFMO are various, and the group recommend using a single classification based in CECOFAD categories: DFAD (Drifting FAD), AFAD (Anchored FAD); FALOG (Artisanal log resulting from human activity, related to fishing activities); HALOG (Artificial log resulting from human activity, not related to fishing activities); ANLOG (Natural log of animal origin); VMLOG (Natural log of plan origin). The information on FOB type comes from the FAD logbooks and, thus, it should be request in independent template different from the one provided for buoy density (information coming from buoy transmission).

General Data	t-RFMOs Data Reporting Requirements	IATTC	ICCAT	IOTC	Information extracted from FAD Logbook	Information extracted from Buoys transmissions	Standards for data reporting
FOB TYPE	FAD type	Not required	FAA Anchored FAD FADN Drifting Natural FAD FADA Drifting artificial FAD	IOTC FADs codes: LOG, LGT, NFD, NFT, FAD, FDT, ANF, DFR, DRT	X		CECOFAD categories, information coming from FAD logbooks

Table 11. Summary of the activities on Floating Object (FOB) requested to CPCs by t-RFMOs and the best standards proposed by the group. The activities should refer to activities described in CECOFAD. The activities are extracted from FOB logbooks and should be requested by t-RFMOs in a separated template, different from the one designated to record information from buoy transmissions.

General Data	t-RFMOs Data Reporting Requirements	ICCAT	IATTC	IOTC	Information extracted from FAD Logbook	Information extracted from Buoys transmissions	Standards for data reporting
Activities with FOBs	Number of FAD visits per type of FAD	Not required	Not required	Total number of FAD visits (deployment, retrieval/encounter, hauling, revisiting or loss) by purse seiners, support vessels	X		Given by CECOFAD activities with FOB
	Number of FADs deployed	The number of FADs deployed on a monthly basis per 1°x1° statistical rectangles, by FAD type (Type: FAA - Anchored FAD; FADN - Drifting Natural FAD; FADA Drifting artifical FAD) indicating the presence or absence of a beacon/buoy or of an echo-sounder associated to the FAD and specifying the number of FADs deployed by associated support vessels, irrespective of their flag;	INF2: No. Deployed belonging to the vessel over the month in 1º degree square	Required (1°x1° statistical and month)	X		Given by CECOFAD activities with FOB
	Numbers of lost FADs	Average numbers of lost FADs with active buoys on a monthly basis	Not required	Required (1°x1° statistical and month)	X		- Given by CECOFAD activities with buoys -The term '*lost*' should refer to the end of the transmission of the buoy, in line with CECOFAD
	Number of sets			Required (1°x1° statistical and month)			Should not be included in FOB related templates as it is provided by other means.

Table 12. Summary of the activities on buoys requested to CPCs by t-RFMOs and the best standards proposed by the group. The activities should refer to activities described in CECOFAD: Tagging (deployment of a buoy on FOB which includes three aspects: deploying a buoy on a foreign FOB, transferring a buoy which changes the FOB owner and changing the buoy on the same FOB which does not change the FOB owner); Remove a buoy (Retrieval of the buoy equipping the FOB); Loss (Loss of the buoy/End of transmission of the buoy). Specific terms used in t-RFMOs as “activated” or “deactivated” which are poorly defined should be harmonized, by adopting common terms of “deploying” or “Tagging” or “Loss” in CECOFAD. The activities should be extracted from FOB logbooks and should be requested by t-RFMOs in a separated template different from the one designated to record information on buoy density which is derived from buoy transmissions.

General Data	t-RFMOs Data Reporting Requirements	IATTC	ICCAT	IOTC	Information extracted from FAD Logbook	Information extracted from buoy transmissions	Standards for data reporting
Activities with buoys	Number and type of beacons/buoys deployed	Not required	Examples for the type of beacon: e.g. radio, sonar only, sonar with echo-sounder; deployed on a monthly basis per 1°x1° statistical rectangles;	The number of deployments refer to FADs	X		- Given by CECOFAD activities with buoys
	Numbers of beacons/buoys activated and deactivated	No of deactivated belonging to the vessel over the month in 1º degree square	The average numbers of beacons/buoys activated and deactivated on a monthly basis that have been followed by each vessel; the spatial resolution is not specified.	The number of instrumented buoys activated, deactivated on each quarter during 2016 its purse seine vessel under the confidentiality rules set by Resolution 12/02. Required by quarter			- Given by CECOFAD activities with buoys -When referring to the submission of activities with buoys the activated buoy should refer to tagging. - The deactivated buoy should reflect the loss

Table 13. Summary of information on number of FOBs requested to CPCs by t-RFMOs and the best standards proposed by the group. The FOB density should be requested. It is estimated by the analysis of daily buoy transmissions which are provided by the buoy manufacturers to the organism responsible of the verification of the compliance with buoy limitation. This information should be provided to t-RFMOs in a separate template different from the one designated to report data on FOB and buoy activities. The information provided by the CPCs to t-RFMOs should include at least the average number of buoys owned/followed by vessel in each 1ºx1º square and month.

General Data	t-RFMOs Data Reporting Requirements	IATTC	ICCAT	IOTC	Information extracted from FAD Logbook	Information extracted from Buoys transmissions	Standards for data reporting
FOB number	Active FADs / buoys	Daily information on all active FADs to the Secretariat, in accordance with guidance developed under Paragraph 12, with reports at monthly intervals submitted with a time delay of at least 60 days, but no longer than 90 days: INF1: Number of active FADs/date INF2: Average number of active FADs belonging to the vessel over the month (by summing up the total number of active beacons recorded per day over the entire month and dividing by the total number of days) in 1 degree square	Average No. Active beacons followed per vessel.	Res 17-08 (9) - the number of instrumented buoys active on each quarter during 2016 its purse seine vessel under the confidentiality rules set by Resolution 12/02		X	Average number of active buoys that is transmitting a signal and is drifting in the sea in 1ºx1º and month Should be reported in a separated form

Table 14. Summary of information on the catch requested to CPCs by t-RFMOs and the best standards proposed by the group. The catch data are generally obtained by other sources and in order to avoid data duplication and facilitate the data reporting to CPCs this information shouldn´t be provided in templates designated to report activities on FOBs or data on buoy densities.

General Data	t-RFMOs Data Reporting Requirements	ICCAT	IATTC	IOTC	Information extracted from FAD Logbook	Information extracted from Buoys transmissions	Standards for data reporting
Catch	Catches and effort	when the activities of purse seine are carried out in association with bait boat, report catches and effort in line Task I and Task II requirements as "purse seine associated to bait boats" (PS+BB).	Not required	Total catches of target IOTC species and bycatch species taken on FOBs, at the same level of resolution (1ºx1º and month) Retained catches: catches for each species retained on board in live weight and/or number. Discard levels: discard levels for each species in live weight or number.			Shouldn´t be requested related to information on FOB activities or buoy densities as it is provided in other Tasks

Annex 1 – FOB logbook

Flag (current) cod.	Month	Lat	Lon	Number of vessels	Vessel Type	FOB type	Buoy Type	No. buoy[2] Deployed	No. FOB[3] Lost

[2] Total number of buoys deployed in the 1-degree square refers only to the first deployment of a FAD with its buoy, the deployment of a buoy on a log [see CECOFAD categories] that was not previously tracked by any vessel, i.e. buoy transfer events are not reported here (i.e. the change of buoy).
[3] FOB that can no longer be tracked by a vessel because the information of the buoy attached is no longer received. It is estimated by summing up the total number of FOB lost per entire month and 1-degree square.

Annex 2 – Buoy transmission Data

Flag (current) cod.	Month	Number of vessels	Lat	Lon	Buoy type	Average No.[4] Of Operational buoy[5]

[4] Average number of operational buoys belonging to the vessels over the month (by summing up the total number of operational buoys recorded per day over the entire month and dividing by the total number of days). It should be provided in 1ºx1º scale

[5] Active buoy that is transmitting a signal and is drifting in the sea

FRAMEWORK CONTRACT – MARE/2016/22 “Strengthening regional cooperation in the area of fisheries data collection” Annex III “Biological data collection for fisheries on highly migratory species”

Project acronym: RECOLAPE

WP3 – Specific Pilot Studies

WP 3.2.1 - EMS capabilities and functionalities to monitor longline fisheries targeting swordfish

Deliverable coordinated by Pascal Bach (IRD) with the contribution of Jon Ruiz and Iñigo Krug (AZTI)

Index

Executive summary 3
1. Background 6
2. Objective 8
3. Methodology 9
   Selection of longliners 10
   Description of the EMS deployed 10
      Information of the EM system deployed on “Le Grand Morne” 11
      Information of the EM system deployed on “Le Bigouden” 14
      Sensor to trigger the EMS while fishing 15
   Material and methods 18
      Data collection from electronic monitoring 18
      Image analysis 18
4. Results 20
   Time necessary for image analysis 20
   Comparative analysis for data of the longline characteristics 21
   Comparative analysis for the data of catches 22
      Electronic Monitoring versus human observer data 22
      Electronic Monitoring versus extended self-reporting data 25
5. Conclusions and recommendations 26
6. References 27

Executive summary

This deliverable is a part of the Work Package 3 (WP 3) within the project MARE/2016/22: "Strengthening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species" (acronym RECOLAPE). This WP 3 dedicated to specific pilot studies was split into two sub-work packages:

- WP 3.1 aiming to identify first minimum data fields set that can server for all the tuna RFMOs and second to develop a data collection and data storage tools,
- WP 3.2 aiming to propose a detailed listing of EMS capabilities for pelagic longline fisheries. Four 4 different tasks in this sub-work package were undertaken:
  1- Literature review on EMS experiences carried out for pelagic longline fisheries worldwide,
  2- Experiment design for the data collection,
  3- At sea trial,
  4- Comparison of data collection methods (i.e. human observer or fishermen versus electronic monitoring).

- Based on all tuna RFMOs feedback (IOTC, ICCAT, IATTC, WCPFC), an increase in observer coverage is needed for some (many in the IO) longline fleets because the minimum requirement of 5% coverage although already low is not always achieved.
- Logistical difficulties are often pointed out for the embarkation of observer on board fishing units less than 16 m length overall (LOA).
- For fleets of large longliners up to 40 m LOA, the duration of the trip can be a constraint and the at-sea sampling from the observer program is often biased because the spatio-temporal representativeness of observed fishing operations is not respected.
- There is a scant of studies in regard to the implementation of the electronic monitoring to collect the equivalent of observer data on pelagic longliners. Probably one of first report published as grey literature was published in 2008.
- Two pelagic longliners belonging to the fleet based in La Réunion (Indian Ocean) were involved in a pilot study aiming to assess the feasibility in terms of strengths and weaknesses of the electronic monitoring deployment on board small vessels.
- The two longliners equipped belonged to the ENEZ DU fishing company. The two vessels called "Le Grand Morne" and "Le Bigouden" have a length overall of 15.8 m and 20.9 m, respectively. They were equipped with the system the Electronic Eye (eEYE™) v6.2 provided by the fishing company Marine Instruments. The eEye v6.2 is a system with three cameras recording images at a frequency of 0.5 frame per second, a V6 antenna with GPS connection and a network attached storage (NAS) installed into the wheelhouse to speed up image retrieval.
- A rotation sensor was installed on a side of the drum where the mainline is stored in order to identify the fishing activities (setting and hauling) and trigger the collection of images. This design permits to avoid intervention on the hydraulic circuit of the boat.

- The data collection from images was similar to those carried out in the frame of the longline observer program ongoing in La Réunion in which observer data come from human observer embarked and self-reporting information from volunteer captain.
- After each fishing trips, the images were analysed through the Beluga software developed by Marine Instruments.
- Images from EM were analysed by two desk-based observers for 36 fishing operations targeting swordfish. For 15 of them on one part and 26 of them on the other part fishing activities and capture data were collected by human observers on-board and by a captain volunteer filling a dedicated logbook, respectively.
- The two fishing operations, namely setting and hauling, considered in this pilot study corresponded to an average setting period of 5.2 hours and an average hauling period of 8.2 hours, i.e. un total of 13.4.
- The time average necessary to analyse images of the setting and the hauling took 16% and 45% of the real time of each operation, respectively. For the two fishing operations, a total time of about 4.5 hours on average (33% of the total real time) was necessary for the image analysis.
- The time for image analysis for a trip of 10 days with 7 fishing operations will last on average 31.5 hours for well-trained desk-based observers.
- The congruence for both the description of the horizontal shape of the longline and the fishing effort estimation (number of hooks) was high.
- The frequency of image records (0.5 frame per second) was too low to distinguish accurately i) the deployment of sensor (for example depth recorders to monitor the maximum fishing depth of the mainline), ii) the deployment of light attractors like lightsticks, iii) the type and size of hook.
- For both sources of human data collection, the congruence with the EM estimates for the species kept on board (swordfish, tunas, marlins, other finfish) was high.
- For both sources of human data collection, the congruence with the EM estimates for the species discarded was low, particularly for sharks which are not hauled on board for safety reasons.
- The congruence between EM estimates and human data collection on board for the capture of sensitive species belonging seabirds, sea turtles and marine mammals species groups, was high. Interactions with marine mammals, sea turtles and Mobulids were observed during trials at sea.
- Due to calibration issues of the eEye v6.2 the opportunity to collect data could not be tested, however we proved it was possible to implement at least on "The Grand Morne" for which the camera aiming to monitor the hauling operation was installed properly to get electronic size data.
- As for other fleets/metier already tested, this pilot study aimed to demonstrate again that the electronic monitoring system (EMS) is a relevant alternative to collect main observer data to increase the level of observer coverage and collect observer data on both small longliners and large longliners having long trips at sea.
- Conclusions obtained from comparative analysis matches perfectly with those already obtained in electronic monitoring programs applied to pelagic longline fisheries.

- The implementation of an electronic monitoring program (EMP) is not only the deployment of cameras on a fishing vessel. Before its implementation the coordinator of the program must to present properly the requirements of the program to the fishing industry and the crew of vessels involved. Afterwards, the implementation is going to be effective through a Memorandum of Understanding (MoU).
- To be effective the EMP will need a collaboration of the crew to enhance the quality of the data collection, particularly to control the dirt of the lens of cameras.
- The dynamic of the fishing with a longline either at setting or hauling implies a frequency of image records higher than 4 frames per second to analyse images properly.
- The deployment of the EMS must be vessel-based taking into account the installation of all the material on the deck particularly at both setting and hauling sites.
- The counting of some discarded individuals of sensitive species groups like sharks proves to be a serious issue for EMS in pelagic longlining. An underwater wide lens installed next to the hauling door of the freeboard deck might be the appropriated design to satisfy this data collection requirement.

1. Background

In the ecosystem-based approach to fishery, stock management and decision-making process are based on the collection of fishery data either independent or dependent (Garcia et al., 2003). Fishery dependent data (FDD) collection meet requirements for scientific purposes such as scientific advices for management based on fish stock assessment (FSA) aiming to propose a management strategy evaluation measuring the relative effectiveness of several potential management decisions. This FDD is generally collected at landings (port sampling of length frequency and biological data) or at-sea (paper logbook or electronic logbook filled by fishermen and embarked observer data). Fisher's knowledge allows collect self-reported information in both manual and electronic logbook. However, in general fishermen do not have time enough to be involved in training to improve both the quality and the quantity of data they might collect particularly for discarded individuals (fish, sea turtle, marine mammals, seabird). Moreover, economic or regulatory issues may hamper their submission of accurate data even on logbook.

Human observers embarked on fishing vessel are one of the most valuable sources of scientific information to describe fishing activities, to identify and inventory species either caught or interacting with the fishing gear, to quantify catches of both target species and bycatch including ETP (endangered, threatened and protected) species while recording at-vessel mortality for discarded species and to collect length frequency data and biological samples. Therefore, human observers provide the main source of information in the frame of the ecosystem-based fishery management (EBFM), (Gilman et al., 2017). However, despite the benefits of embarked observer data, regular weaknesses of observer program are pointed out by regional fishery organization such as difficulties to reach a prerequisite level of coverage of the fishing effort, the lack of the representativeness of the fishing activity (fishing effort by metier for example) and the maladjustment of the executed program to meet management objective (Williams et al., 2016; Gilman et al., 2017). Moreover, the implementation of human observer program is considered as costly and sometimes difficult to set up due to the size of vessels, the trip duration and the crew acceptance. To coping these constraints, the electronic monitoring was proposed as an alternative measure to collect observer data. Electronic monitoring systems (EMS) on fishing vessels have been developing rapidly during the last decade thanks to the fast improvement of the EMS technology (Anonymous, 2016). Now, experiences have shown that EMS is not only the deployment of cameras and sensors onboard and the development of minimum standards is necessary before implementing the EMS program. These minimum standards must be considered at the three steps of the program (Ruiz, 2018), (table 1):

1 - Before the trip (acknowledged requirements with stakeholders, installation, certification, audits),

2 – during the trip (configuration required for the data collection),

3 – after the trip (data traceability and analysis).

Table 1 – Summary of the operations to consider at the three steps (before, during, after) of the EMS program implementation (modified from Ruiz, 2018).

When	Object or task	Description
Before	Setting the scene	Prior to both the installation and trials requirements having to be satisfied must be adopted
	EMS installation	Adapt an EMS configuration fitting at a vessel level in regard to requirements
	Agreement with stakeholders	Collaboration with the fishing industry (at least boat owner and boat captain) is a fundamental requirement (ideally through MoUs)
	Pilot study	EMS trials are a key for the proof of concept. Ideally trials must be undertaken through an iterative process in order to shape the EMS in regard to requirements
During	Cameras	Digital cameras installed to observe the areas of interest with a frame rate and image quality selected in regard to requirements
	EMS design	The whole system must resist to rough conditions at-sea and wet and salty environment. Tamper proof system with encrypted data, near-real-time remote online "health statements" and GPS linked imagery. Autonomous with a minimal maintenance by the crew
	Storage	Autonomy of the system adapted to the trip duration of the vessel and the image analysis frequency as defined as requirement by end-users
After	Desk-observer	Desk-based observer (ideally well-trained observer at-sea) must be qualified for the image analysis
	Image analysis	An integrated software to analyse images collected should be designed as a part of the EMS. The output format of data must be compatible with standard data flow and relational database management system
	Traceability	A hard drive chain of custody ensures the traceability of all hard drives and information saved on
	Data analysis	Analysis and submission of data should be achieved by independent bodies involved in human observer program in regard to requirements

In the case of large pelagic fisheries, EMS has already been tested for both purse seine and pelagic longline fisheries.

For purse seine fisheries in the worldwide ocean (Pacific, Atlantic and Indian) different EMS pilot studies were carried out, and EMS capabilities have been proven (Monteagudo et al., 2015; Ruiz et al., 2015; Briand et al., 2017) and minimum standards were established (Ruiz et al., 2017).

For pelagic longline fisheries, for the moment pilot studies were carried out mostly in the Pacific Ocean (Solomon Islands, New Caledonia, Taiwan, Hawaii and Australia), (Mc Elderry et al., 2020; Anonymous, 2016; Emery et al., 2018). In the Western and Central Pacific Ocean, the EMS was considered as a tool to satisfy international data collection and exchange obligations while reaching the minimum level of the observer coverage of 5%. The results of the study using the Western and Central Pacific Fisheries Commission as a case-study shown that 78% of the mandatory longline fields can informed by the current EM technology and 84% of information collected in those fields used in scientific analyses. Regarding comparisons of data sets of species (list of species and number of individuals caught), no differences between EMS and observer data were observed for retained catch (target and bycatch) while a significant underestimation was detected for discarded individuals, particularly for sharks. In Hawaii, EMS is implemented to comply with the regulation requirement of the observer coverage of 100% of shallow-set and 20% of deep-set longline fishing trips. Results obtained from 165 longline hauls are similar for those of the WCPFC case-study analysis. EM was proved successful as a human observer alternative for estimating i) catches for retained species, ii) discards for commercial species and iii) discards for protected species. However, for the shark group, EM underestimated the number of discarded individuals by more than 50%. This underestimation was a consequence of a safety measure for the crew due the potential danger of mandatory weight installed on branchline to mitigate seabird interactions.

As for EMS implemented for purse seiners, EMS deployed on pelagic longliner is nowadays mature enough for a wide implementation in the global ocean. Available systems are adaptable and scalable, and providers must adapt their system to clients' requirements. The quality of the EMS is linked to the clarity of the objectives of the program. In the case of indecisive objective, the process necessary to frame the adapted EMS will be time consuming. Different technologies exist and the final solution corresponds in general in a trade-off between several constraints like storage capacity, image resolution, image recording frequency, image analysis software. Finally, at least for the moment all EM systems need the implication of crew members to control the operability of the system. For successful implementation, the coordinator of the EM program must consider mechanisms that will incentivize the crew for its collaboration.

2. Objective

Based on tuna RFMOs feedback, an increase in observer coverage is needed for some longline fleets, where the minimum requirement of 5% coverage although already low is not always achieved. The logistical difficulties to embark observer on board is usually mentioned as a main issue to launch a human observer program on longline fishery composed of fishing units less than 16 m length overall (LOA). Many longline fleets in coastal fishing countries worldwide are characterized by fishing units with small or medium sizes compared to long distance longliners (Japan, Taiwan, China, Spain, Portugal) on which the embarkation of observer is logistically easier than on a small vessel. However, the duration of the trip can also be a constraint. Hence, EMS successfully deployed in several fisheries worldwide could be an alternative and/or a complement to human observers onboard. However because EMS is not simply putting camera on board, EMS before to be implemented should be tested in order i) to identify strengths and weaknesses of the system used and ii) to propose improvements if necessary to at least reach the minimum standards required similar to the approach implemented for European purse seine tuna fisheries in the Indian and Atlantic oceans (Ruiz et al., 2017).

Therefore, the main objective of this pilot study involving two pelagic longliners from the pelagic longline fleet based in La Réunion island (France, Indian Ocean) is to compare the data set collected by each sampling method i.e. EMS versus both human observer and expanded self-reported data. This dataset concerns both the set of variables informing the fishing activity (characteristic of the mainline, branchline, floatline, buoy, hook, bait and deployment of devices to mitigate negative impacts) and values for variables quantifying catches kept on board and discarded by species or group of species. This comparative approach will permit to determine if the EMS deployed can be used to reliably collect unbiased data on board, first on fishing boats equipped and second, on pelagic longline vessel in general. Finally, this pilot study will allow to clarify the strengths and weaknesses of the EMS implemented with the goal of defining an electronic autonomous system as an alternative or a complement to human observer embarked on pelagic longliners.

3. Methodology

Before presenting the details of the methodology carried out in the frame of this EMS pilot study, we summarize in the table 2 the time schedule achieved to produce the present deliverable T.3.2.1.

Table 2 – Summary of the time schedule of main operations realized in the frame of the EMS pilot study**.**

Period	Description of tasks achieved
March 2018	Presentation of the project to the fishing industry in La Réunion and selection of four potential longliners involved in the pilot study
April 2018	Sending of fishing vessel construction plans to the EMS provider
May 2018	Installation of EM system on two vessels
July 2018	Complement of the EMS installation on one longliner
August to October 2018	Process of the recruitment of two observers for at-sea data collection and desk-based image analysis
December 2018	First at-sea observations with EMS problems to record information on one vessel. Statement of maritime security service prohibiting boarding of a human observer on a vessel equipped with EMS

January to March 2019	Data collection from observer at-sea, from self-reporting and EMS. Image analysis with the Marine Instruments Beluga software
April 2019	Data analysis of all datasets (human observer, self-reporting data and EMS data)

Selection of longliners

The pelagic longline fleet in La Réunion is composed of vessels with an overall length ranged from 14 m to 30 m. The selection of longliners for this pilot study was based on the resolution 11/04 of the Indian Ocean Tuna Commission in regards to the paragraph 2 of it "In order to improve the collection of scientific data, at least 5 % of the number of operations/sets for each gear type by the fleet of each CPC while fishing in the IOTC area of competence of 24 meters overall length and over, and under 24 meters if they fish outside their Exclusive Economic Zone (EEZ) shall be covered by this observer scheme. For vessels under 24 meters if they fish outside their EEZ, the above mentioned coverage should be achieved progressively by January 2013".

Therefore, vessel construction plans of longliners belong to the fishing company "ENEZ DU" were sent to the company Marine Instruments (Marine Instruments link) selected as the vendor's small-scale vessel equipment option or Electronic Eye v6.2 (eEYE™). Two longliners, "Le Bigouden" of 20.9 m LOA and "Le Grand Morne" of 15.8 m LOA, were proposed by the fishing company for the deployment of the EMS. Both exploit fishing grounds principally outside La Réunion EEZ.

Description of the EMS deployed

The EM system deployed on the two longliners corresponds to the Electronic Eye (eEYE™) v6.2 (Figure 1). This system is a full featured remote Electronic Monitoring System with automatic image capture for fishing monitoring and bycatch control onboard vessels. For this project, 3 camera Electronic Eye V6 systems were installed on each vessel. The selected vessels being small sized, some areas are double covered to avoid blind spots blocked by fishing gear and marginally benefit in case of camera technical failure. Besides the internal GPS, a rotation sensor was installed on the drum to identify fishing activity and trigger the image records. A NAS was installed into the wheelhouse to speed up image retrieval. Cameras and EM system were configured to work both during day and night fishing operations. Data were recorded onto an external hard drive of 2TB estimated to last 4 months corresponding to about 80 fishing operations with a deployment of 1500 hooks per set.

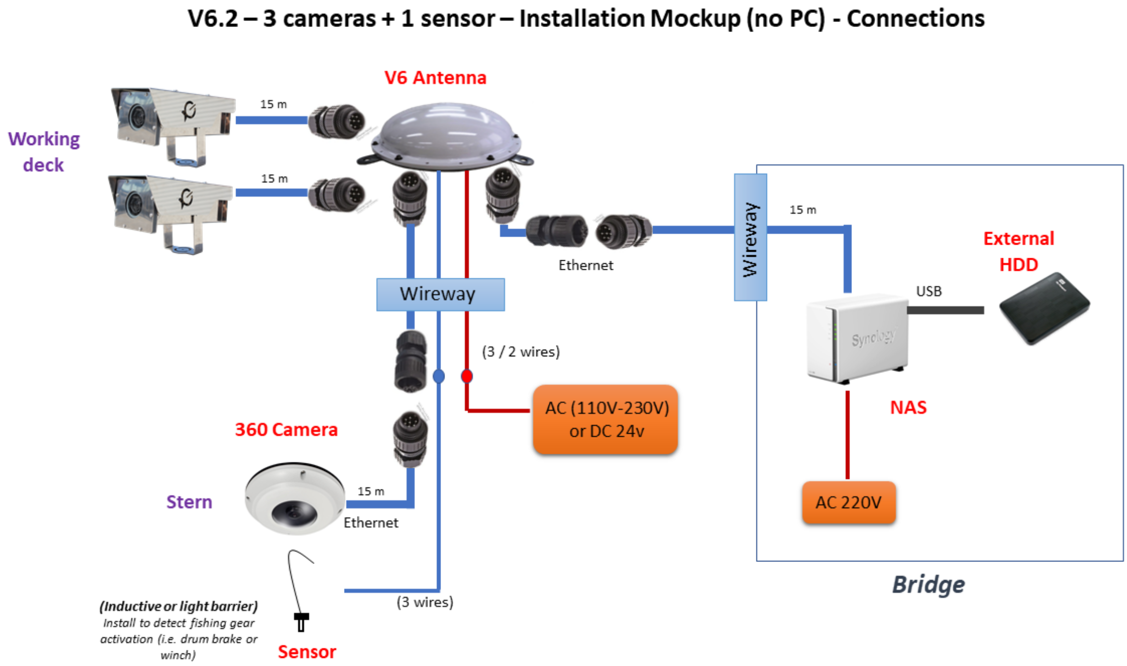


Figure 1 – Diagram of the Electronic Eye (eEYE™) v6.2 installed on longliners

Information of the EM system deployed on "Le Grand Morne"

The longliner "Le Grand Morne" is presented on the figure 2. This is a longliner with a fiberglass hull, a capacity of 5 fishermen operating fishing trips of a maximum of 10 days.

Three cameras (Table 3, Figure 3) were installed, 1) one to have a wide view of hauling operations and to observe potential interactions with marine mammals (depredation), 2) a second to observe the setting and eventually the deployment of tori lines the gear and seabirds from the rear deck, 3) a third to observe the hauling of the catch on board and status of the individual at hauling and eventually at release in the case particularly for sea turtles which are hauled on board to remove the hook before releasing them.

Length measurements of fish had to be carried out from images collected by the second camera. Unfortunately, due to a calibration problem, this operation could not be undertaken.

Figure 2 – The longliner “Le Grand Morne”

Table 3 – Details of cameras of the EMS installed on "Le Grand Morne"

	Camera 1	Camera 2	Camera 3
Location	Starboard side (roof)	Up to the drum (roof)	Port side (roof)
Aiming	Stern / starboard side	Stern	Starboard side
Type of view	General	Setting	General
Action monitored	Hauling. Catch handling. Discard operations. Interactions	Gear deployment. Tori line deployment	Hauling. Catch handling
Frame speed	0.5 fps	0,5 fps	0,5 fps

Figure 3 – Cameras (1 top, 2 below left, 3 below middle – number refers to table 3) and main unit at the mast installed on "Le Grand Morne".

Information of the EM system deployed on "Le Bigouden"

The second longliner "Le Bigouden" equipped with the EMS is presented on the figure 4. This is a longliner with a steel hull, a capacity of 7 fishermen operating fishing trips of a maximum of about two weeks.

Three cameras (Table 4, Figure 5) were installed, 1) one to have a wide view of hauling operations and to observe potential interactions with marine mammals (depredation), 2) a second to observe the setting and eventually the deployment of tori lines the gear and seabirds from the rear deck, 3) a third to observe the hauling of the catch on board and status of the individual at hauling and eventually at release in the case particularly for sea turtles which are hauled on board to remove the hook before releasing them. The main unit with the VMS was installed at the mast (Figure 6).

Figure 4 – The longliner "Le Bigouden"

Table 4 – Details of cameras of the EMS installed on "Le Bigouden"

	Camera 1	Camera 2	Camera 3
Location	Starboard side (roof)	Starboard side (above hauling area)	Stern (roof)
Aiming	Bow / starboard side	Stern / starboard side	Stern / starboard side
Type of view	General	Detail view	General view (360° surrounding view)
Action monitored	Hauling. Catch handling. Discard operations. Interactions	Hauling. Catch handling	Gear deployment. Tori line deployment
Frame speed	0.5 fps	0,5 fps	0,5 fps

Sensor to trigger the EMS while fishing

In order to collect images during the setting and hauling operations only, a sensor was used for detection of fishing activities. In some trials the sensor was installed on the hydraulic system. In our case, the pilot study being limited in time it was decided in agreement with the fishing company that the sensor will be installed without any intervention on the hydraulic circuit. Therefore, an inductive sensor detecting the rotation movement of the drum storing the mainline was installed to trigger image records during both setting and hauling operations (Figure 7).

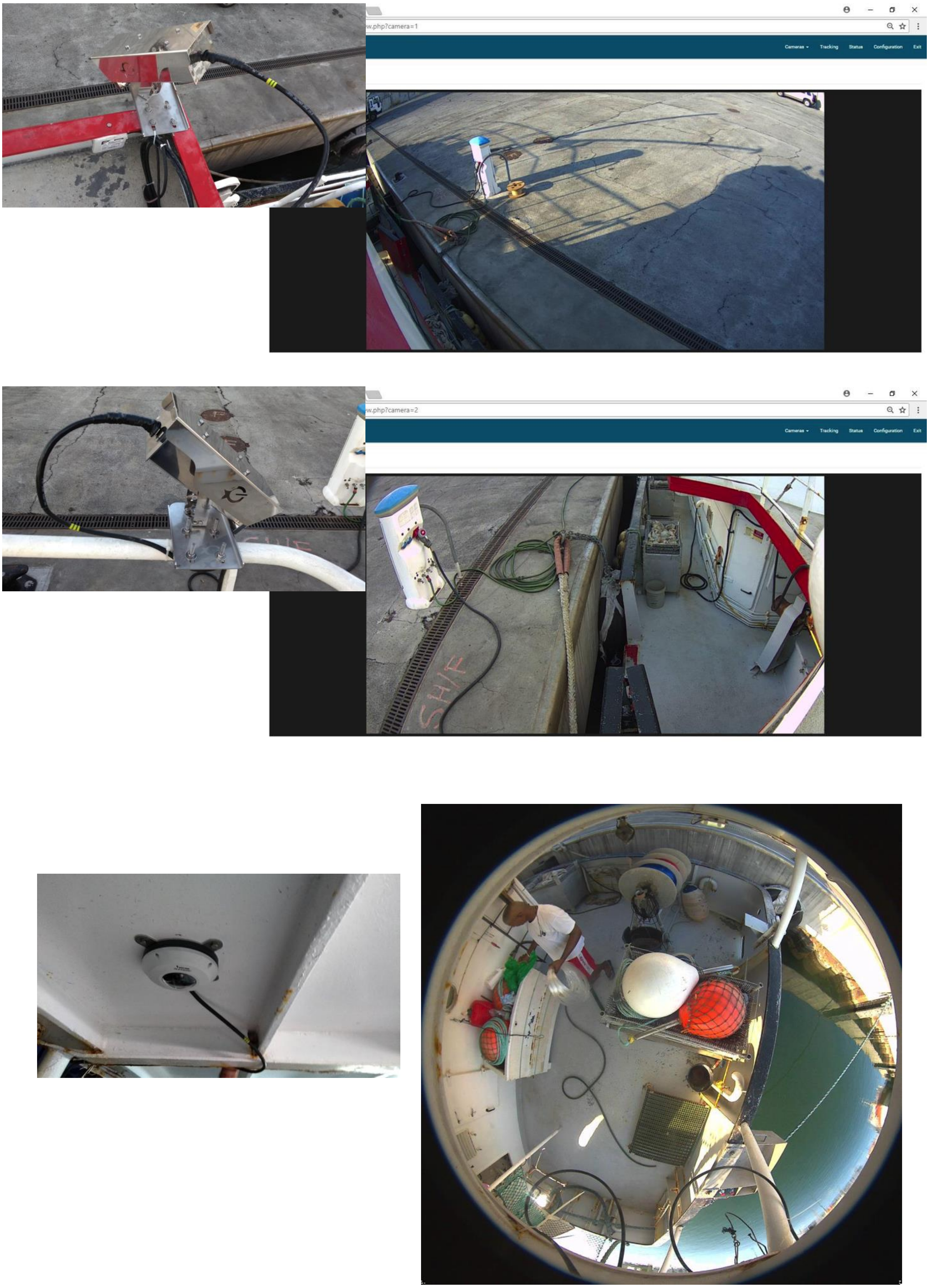


Figure 5 – Cameras 1, 2, 3 (number refers to table 4) installed on “Le Bigouden” on the left and corresponding view for each on the on right.

Figure 6 – Main unit at the mast installed on "Le Bigouden".

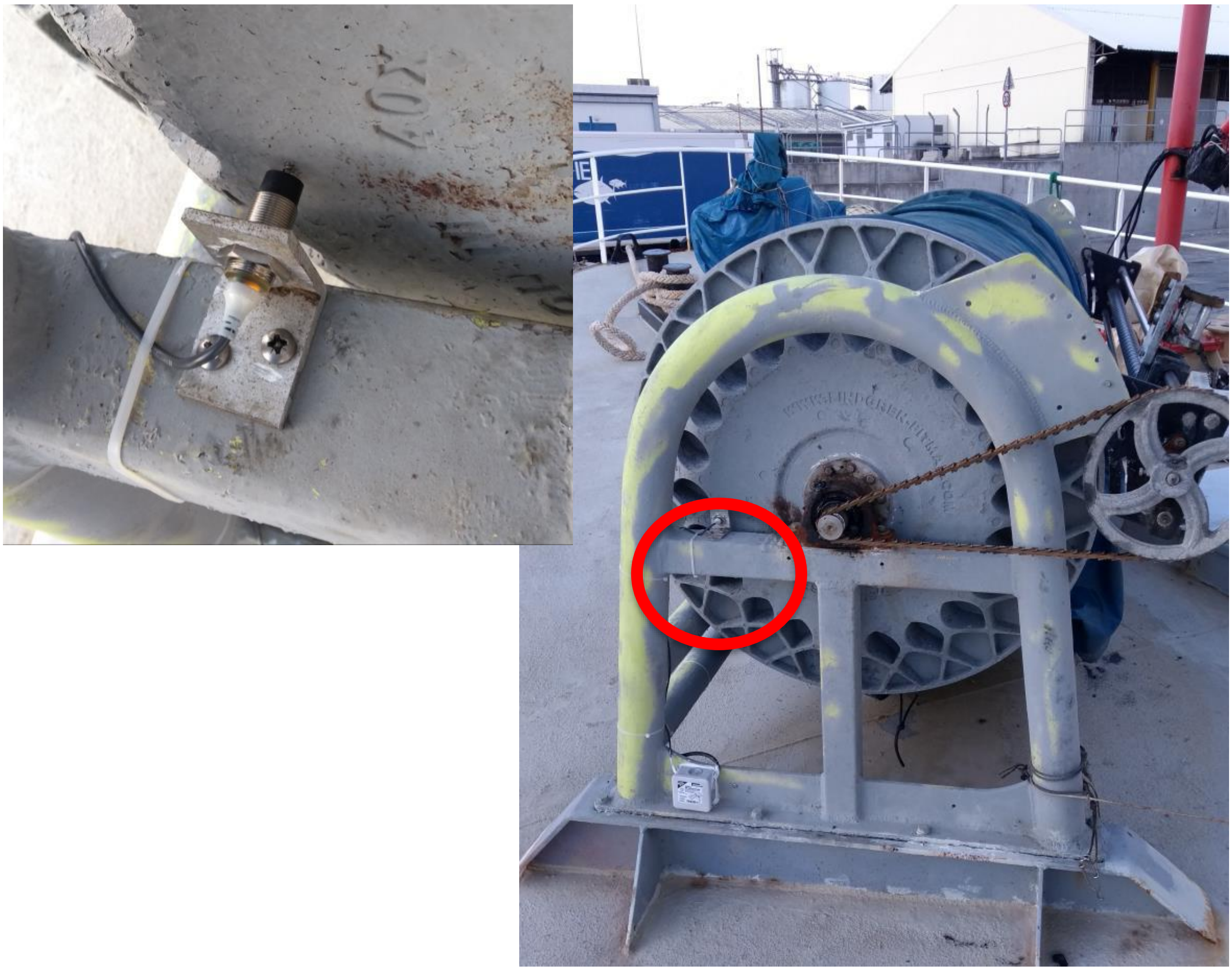


Figure 7 – Sensor installed on the drum to trigger image records**.**

Material and methods

Data collection by human observers

One data source came from on board data collected by human observers (15 fishing operations) and a captain volunteer (26 fishing operations) filling a dedicated logbook. The collection of data followed the scientific observer protocol set up in the frame of observer program of the pelagic longliner fleet based in La Réunion in respect to the IOTC resolution 11/04 (Bach and Sabarros, 2018),

Data collection from electronic monitoring

Based on the protocol used by scientific observers embarked to collect data on longliners based in La Réunion, the minimum data monitoring needs that EMS should cover on longline vessels targeting large pelagic species would optimally be the following:

- Date, time and position (latitude and longitude) of the gear deployment for both setting and hauling operations,
- Horizontal shape (linear, U, L, Z) of the longline,
- Mainline material, branchline material, hook and bait types. Use and number of light attractants deployed (chemical light-sticks or electrolume),
- Number of sections, number of baskets, hooks between floats and total number of hooks deployed,
- Use of line shooter, towed buoy and tori lines or water spray,
- Catch species, hook position, status at hauling and/or release,
- Number of interactions with protected species.

The disposition of the three cameras on board each vessel as presented above was defined in order that the electronic monitoring be able to achieve these requirements.

Image analysis

Images from EM were analysed by two trained desk-based observers for 36 fishing operations targeting swordfish.

In order to assess the feasibility of the EMS in terms of time necessary for the data collection from images by desk-based observers, for each fishing operations of each set was measured for each phase of the fishing operation, *i.e* setting and hauling. These estimates were performed after a habituation of the observers to the Beluga software, this period lasted about 10 days. Moreover, each desk-based observer analysed images for a cruise observed by the second observer, the two data sources electronic monitoring and observer data on a given fishing operation being collected by the two different means. Examples of such images from the EM analysed by the desk-based observers are presented on the figure 8 A, B.

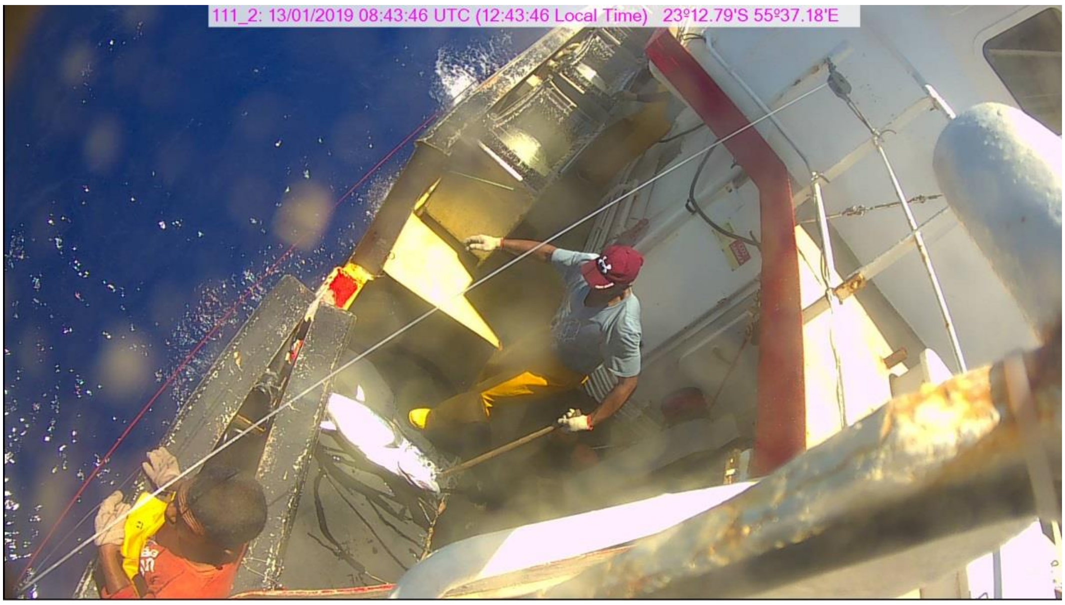


Figure 8 A, B – Picture from the camera on the Port side on “Le Grand Morne” showing a yellowfin capture (A – above) and picture from the camera on the starboard side on “Le Bigouden” showing a capture of al albacore (B - below).

4. Results

The 36 longline fishing operations were achieved around La Réunion and in the Mauritius EEZ (Figure 9).

Time necessary for image analysis

The two phases of a longline fishing operation, namely setting and hauling, lasted on average 5.2 hours and 8.2 hours, respectively, corresponding to a total time of 13.4 hours. The time average necessary to analyse images of the setting and the hauling took, respectively, 16% and 45% of the real time of each operation (Figure 10). For the two fishing operations, a total time of about 4.5 hours on average (33% of the total real time) was necessary for the image analysis (Figure 10).

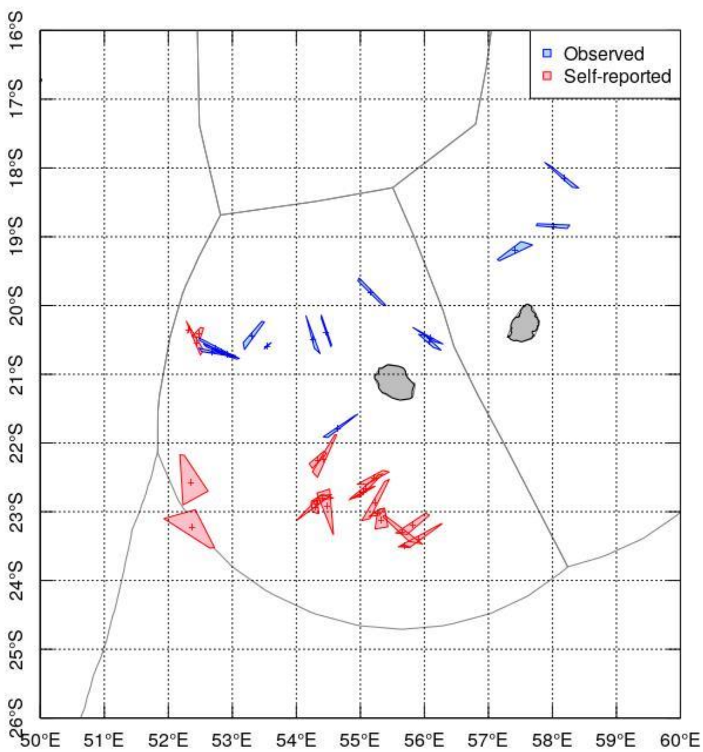


Figure 9. Position of longline fishing operations with an electronic monitoring data collection implementation carried out from December to February 2019.

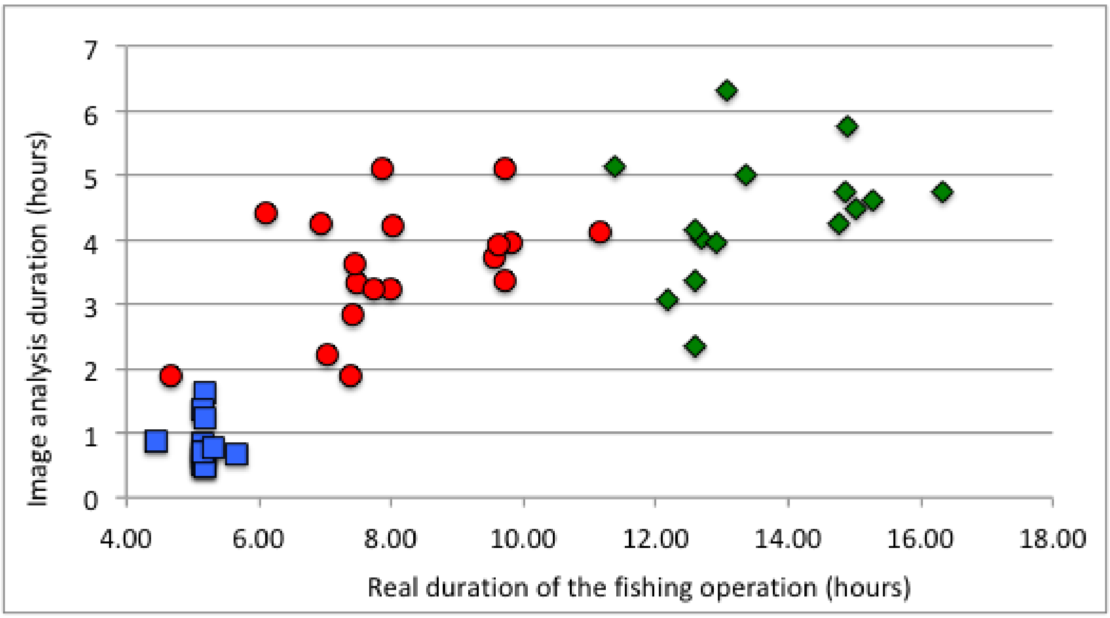


Figure 10. Comparison of the real time of the fishing operation (setting in blue, hauling in red, sum of both in green) with the respective time necessary to analyse the images recorded for each of them**.**

This value of 33% certainly corresponds to the maximum percentage to allocate for image analysis because some improvements might be done in the next future regarding the strategy of image analysis, the frequency of image capture (the 1 image per 2 seconds used in the pilot study is not well adapted for the fishing practice on interest) and the ergonomics of the software interface.

Comparative analysis for data of the longline characteristics

The EMS data obtained from image analysis were compared with data collected by observer on-board and self-reporting data collected by the captain (Table 5). For the two data series, the coordinates of the longline in time and space showed a high similarity. EMS allows to complete all the fields while some of them may be forgotten by observer (the captain in this study). Moreover, the similarity between EMS data and observer data sources was very high for the horizontal shape of the mainline.

The data fields corresponding to the description of the gear (material and size of mainline, floatline, branchline) cannot be collected through EMS. However, such information is in general vessel/captain dependent and it can be found in the logbook or in an observer report if the vessel is able to embark it. For other fields describing the fishing strategy (electrolume, weight on branchine) the frequency of image records (0.5 frame per second) was to fill these fields. However, the deployment of sensor, the hook type and the bait type could be identified.

Correlation coefficient values between the fishing effort estimates from the two data sources were significant with a high level of accuracy for EMS estimates whatever the method used to calculate the number of hooks (the method 1 is based on the counting of the number of sections, the number of baskets per section and the number of hooks between floats, the method 2 is based on the estimate of the number of baskets by using the total time of the setting divided by the average time to deploy a basket and the number of hooks between floats).

Finally, some fields like the using of a shooter to set the line or the deployment of tori lines to mitigate seabird interactions were not presented in the table 5 because such operations or events were not observed during the fishing operations of this pilot study.

Comparative analysis for the data of catches

Electronic Monitoring versus human observer data

For the 15 fishing operations with the two data sources collected, the number of total catches recorded by the electronic monitoring and the human observer was 419 and 425 respectively. Number of catches per fishing set recorded by the two methods are highly correlated and can be considered as similar, the slope of the regression line having a value of 0.987 not different of 1 corresponding to the slope of the identity line (Figure 11).

However, the comparison of catches by group of species pointed out two major differences between the two methods. EMS overestimates at a level of 100% the records of undetermined individuals which do not exist in observer records (Table 6). In the meantime, EMS underestimates at a level of -154.5% the number of catches of sharks, certainly some undetermined individuals likely being shark individuals. For the swordfish as the main target species of the fishery the congruence between the two methods is rather satisfying with an underestimation by the EMS close to 10% (Table 3).

Regarding the fate of individuals (Table 7), the records of individuals kept on board are similar, however compared to human observed records EMS leads to underestimate the level of discards, a result likely linked to the underestimation of sharks by this technic.

Table 5. Comparison of information collected by EMS and human observers for the longline deployment characteristics (The information in the cell correspond to the similarity of values between two data sources (ranged from 0 to 1) or Y versus N if the collection is possible or the correlation significant).

Variables	EMS versus Observer	Comments	EMS versus Self Reporting	Comments
Date, time of the setting	1		1	
Position of the setting	1		1	
Date, time of the hauling	1		1	
Date, time of the hauling	1		1	
Horizontal shape of the longline	1		0.86	Due to incomplete image records
Mainline material	0	Impossible to be collected with EMS. Can be in the logbook. Dependent of the vessel/captain	0	Impossible to be collected with EMS. Can be in the logbook. Dependent of the vessel/captain
Mainline diameter	0		0	
Branchline material	0		0	
Branchine length	0		0	
Floatline material	0		0	
Floatline length	0		0	
Weight on the branchline	0		0	
N. sections	0.8	Image frequency to low	0.95	Fishing effort data not declared by the captain for one set. Difference less than 1% for method 1. Overestimate of ~3% by EMS for method 2.
N. baskets/section	1		0.95	
N. hooks (1)	Y	Difference less than 1%.	Y	
N. hooks (2)	Y	Overestimate of ~3% by EMS	Y	
Hook type	1		1	
Hook size	N		N	
Bait type	1		1	
Bait size, status	N		N	
Electrolume deployment	Y		Y	
Type of electrolume	Y		Y	
Frequency of electrolume deployment	0	Image frequency to low to estimate the number of electrolume	0	Image frequency to low to estimate the number of electrolume
Deployment of sensors on the mainline	Y		Y	

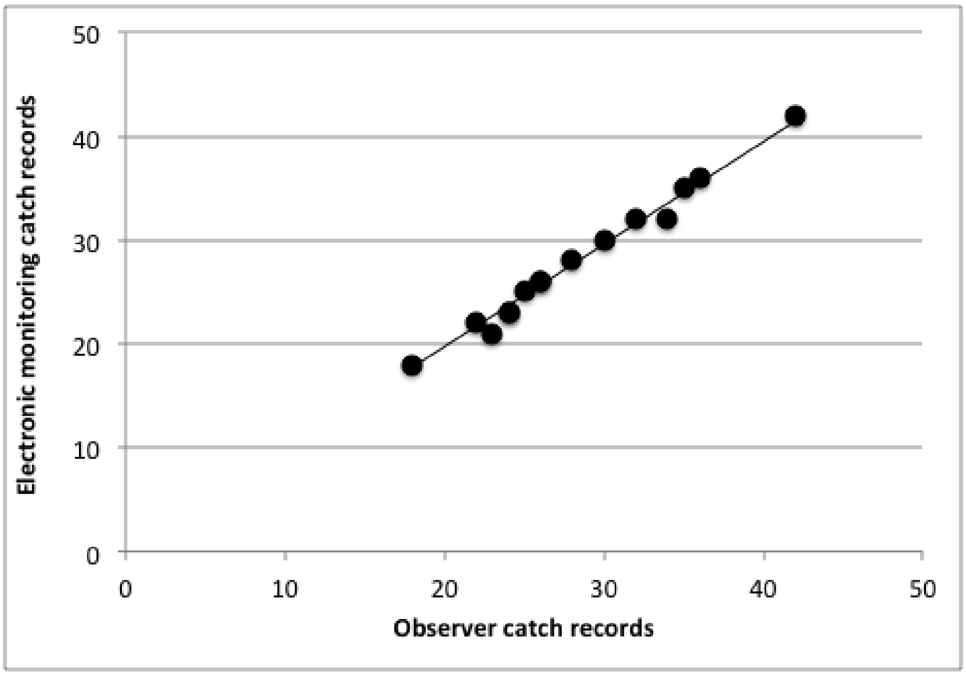


Figure 11. Relationship between the records of catches per set obtained from human observer (horizontal axis) and EMS (vertical axis). The line corresponds to the identity line (intercept = 0, slope = 1).

Table 6. Comparison of records of catches by species group obtained by human observer (OBS) and EMS (UND = undetermined, BILL = billfish, FINF = finfish, RAYS = rays; SHARK = sharks, SWO = swordfish, TUNA = tunas).

	OBS	EMS	Diff (%)
UND	0	52	100
BILL	8	8	0
FINF	86	83	-3.6
RAYS	11	9	-22.2
SHARK	56	22	-154.5
SWO	177	160	-10.6
TUNA	87	85	-2.4
TOTAL	425	419	-1.4

Table 7. Comparison of the fate of catches all species aggregated obtained by human observer (OBS) and EMS.

FATE	OBS	EMS
DISCARDED	214	189
ESCAPED	7	1
KEPT	204	205
UNKNOWN		24
TOTAL	425	419

Electronic Monitoring versus extended self-reporting data

For the comparison of catch records between the electronic monitoring and extended self-reporting data collected by a volunteer captain catch per set and per group of species at the total level of the pilot study as well as the fate of individuals at the total level.

For the 21 fishing operations with the two data sources collected, the number of total catches recorded by the electronic monitoring and the self-reporting was 580 and 600 respectively. This similarity must be noted as it highlights somehow the quality of data self-reported by the captain. Except one outlier, the number of catches per fishing set recorded by the two methods are correlated and can be considered as similar, the slope of the regression line having a value of 0.956 not different of 1 corresponding to the slope of the identity line (Figure 12).

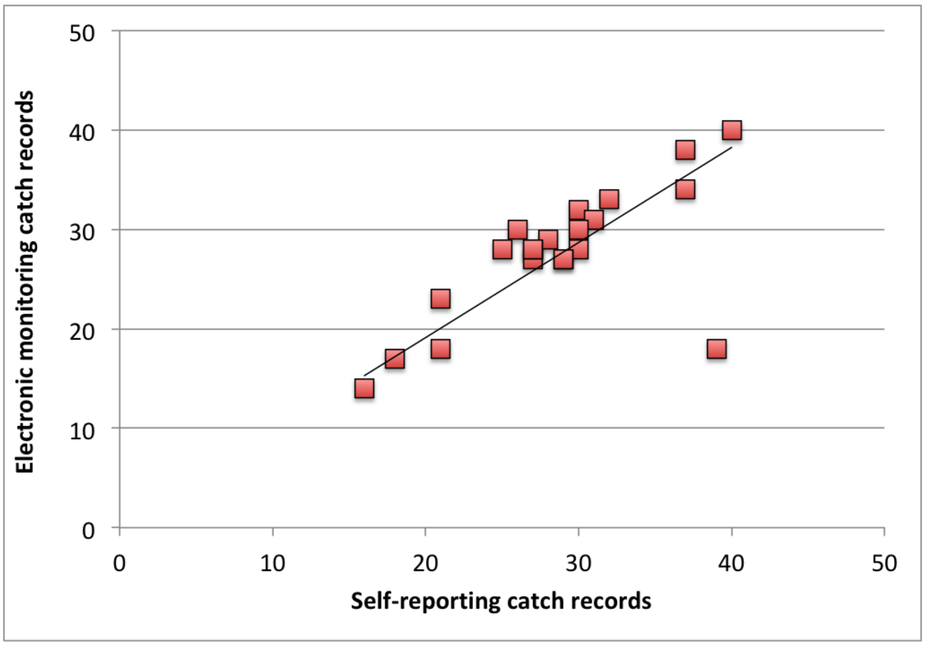


Figure 12. Relationship between the records of catches per set obtained from self-reporting data (horizontal axis) and EMS (vertical axis). The line corresponds to the identity line (intercept = 0, slope = 1).

However, as mentioned previously the comparison of catches by group of species pointed out two important differences between the two methods for discarded groups, finfish and sharks. The underestimation of catches for EMS reached about 60% and 40% respectively (Table 8). For the target species group, swordfish and tuna, the estimates of catches are rather similar, opposite to the group of undetermined catches overestimate by the EMS but corresponding in some extent to the level of underestimates described for the finfish and shark groups. It must be noted the record of an individual of a marine mammal and a see turtle reported by the data sources. However, an individual of Mobulids counted in the RAYS group in the table 6 was self-reported but not detected through EM.

Table 8. Comparison of records of catches by species group obtained by the self-reporting (SR) and EMS (UND = undetermined, BILL = billfish, FINF = finfish, RAYS = rays; SHARK = sharks, SWO = swordfish, TUNA = tunas, SEA TURTLE = sea turtle, MAM = marine mammal).

	SR	EMS	Diff (%)
BILL	25	25	0
FINF	112	71	-57.7
RAYS	49	47	-4.3
SHARK	111	79	-40.5
SWO	221	210	-5.2
SEA TURTLE	1	1	0
TUNA	81	77	-5.2
UND	0	69	100
MAM	1	1	0
TOTAL	601	580	-3.6

5. Conclusions and recommendations

As for other fleets/metier for which EMS has already been tested, this pilot study aimed to demonstrate again that the electronic monitoring system (EMS) is a relevant alternative to collect main observer data to increase the level of observer coverage and collect observer data on both small longliners and large longliners having long trips at sea.

Based certainly of a high level of 33% of the real time need to analyse the EMS data of longline fishing operations, we can estimate that the time necessary for image analysis for a trip of 10 days with 7 fishing operations will last on average 31.5 hours for well-trained desk-based observers.

For both sources of human data collection, the congruence with the EM estimates for the main species group kept on board (swordfish, tunas, marlins) was high. On the other hand, the congruence with the EM estimates for the species discarded was low, particularly for sharks which are not hauled on board for safety reasons. The counting of some discarded individuals of sensitive species groups like sharks proves to be a serious issue for EMS in pelagic longlining. An underwater wide lens installed next to the hauling door of the freeboard deck might be the appropriated design to satisfy this data collection requirement.

Our current dataset of this pilot study was rather limited due to the short time dedicated for the data collection and constraints to embark an observer on one boat occurring after the installation of the EM. However, one of positive findings concerns the congruence between our results with those already published in the literature from EM studies implemented to collect fishery dependant data at-sea for scientific and control purposes.

This pilot study showed us that the implementation of an electronic monitoring program (EMP) is not only the deployment of cameras on a fishing vessel. Before its implementation the coordinator of the program must to present properly the requirements of the program to the fishing industry and the crew of vessels involved. Afterwards, the implementation is going to be effective through a Memorandum of Understanding (MoU). To be effective the EMP will need a collaboration of the crew to enhance the quality of the data collection, particularly to control the dirt of the lens of cameras.

The dynamic of the fishing with a longline either at setting or hauling implies a frequency of image records higher than 4 frames per second to analyse images properly. Moreover, the deployment of the EMS must be vessel-based taking into account the installation of all the material on the deck particularly at both setting and hauling sites. In the meantime, the rotation sensor on the drum used as the trigger system to switch on the system we implemented in this pilot study has proved its efficiency without interfering with crucial devices on board such as the hydraulic circuit.

Due to calibration issues of the eEye v6.2 the opportunity to collect electronically length data could not be tested, however we proved it was possible to implement it at least on "The Grand Morne" for which the camera aiming to monitor the hauling operation was installed properly.

6. References

Anonymous, 2016. Summary report of the international workshop on application of electronic monitoring systems in tu!na lmongline fisheries (No. 2016– 07), ISSF Technical Report. International Seafood Sustainability Foundation, Washigton D.C., USA.

Bach, P., Sabarros, P., 2018. Manuel de l'observateur embarqué à bord des palangriers pélagiques réunionnais (Document technique). Observatoire des Ecosystèmes Pélagiques Tropicaux exploités, Ob7, IRD, UMR 248 « MARBEC ».

Briand, K., Bonnieux, A., Le Dantec, W., Le Couls, S., Bach, P., Maufroy, A., Relot-Stirnemann, A., Sabarros, P.S., Vernet, A.-L., Jehenne, F., Goujon, M., 2018. Comparing Electronic Monitoring System with observer data for estimating non-target species and discards on French tropical pirse seine vessels, Collect. Vol. Sci. Pap. ICCAT. Madrid.

Emery, T.J., Noriega, R., Williams, A.J., Larcombe, J., Nicol, S., Williams, P., Smith, N., Pilling, G., Hosken, M., Brouwer, S., Tremblay-Boyer, L., Peatman, T., 2018. The use of electronic monitoring within tuna longline fisheries: implications for international data collection, analysis and reporting. Rev. Fish Biol. Fish. 28, 887–907. https://doi.org/10.1007/s11160-018-9533-2

Garcia, S.M., Zerbi, A., Aliaume, C., Do Chi, T., Lasserre, G., 2003. The ecosystem approach to fisheries. Issues, terminology, principles, institutional foundations, implementation and outlook (No. 443), FAO Fisheries Tehnical Paper. FAO, Rome.

Gilman, E., Weijerman, M., Suuronen, P., 2017. Ecological data from observer programmes underpin ecosystem-based fisheries management. ICES J. Mar. Sci. 74, 1481–1495. https://doi.org/10.1093/icesjms/fsx032

McElderry, H., Pria, M.J., Dyas, M., McVeigh, R., 2010. A Pilot Study Using EM In The Hawaiian Longline Fishery (Unpublished report prepared for the Western Pacific Fishery Management Council). Archipelago Marine Research Ltd., Victoria British Columbia, Canada.

Monteagudo, J., Legorburu, G., Justel-Rubio, A., & Restrepo, V. (2015). Preliminary study about the suitability of an electronic monitoring system to record scientific and other information from the tropical 1270 tuna purse seine fishery. SCRS/2014/132. Collective Volume of Scientific Papers ICCAT, 71, 440-459.

Ruiz Jon, 2018 – Five years of electronic monitoring onboard tropical tuna purse seiners. In Kennelly, S.J., Borges, L. (Eds.), 2018. Proceedings of the 9th International Fisheries Observer and Monitoring Conference. Vigo, Spain, p. 395.

Ruiz, J., Bach, P., Krug, I., Briand, K., Murua, H., Bonnieux, A., 2017. Strength and weakness of the data elements currently collected through Electronic Monitoring Systems in the Indian Ocean (IOTC-2017-WPDCS13 No. 26). Indian Ocean Tuna Commission, Victoria, Seychelles.

Ruiz, J., Batty, A., Chavance, P., McElderry, H., Restrepo, V., Sharples, P., Santos, J., Urtizberea, A., 2015. Electronic monitoring trials on in the tropical tuna purse-seine fishery. ICES J. Mar. Sci. J. Cons. 72, 1201–1213. https://doi.org/10.1093/icesjms/fsu224

FRAMEWORK CONTRACT – MARE/2016/22 “Strengthening regional cooperation in the area of fisheries data collection” Annex III “Biological data collection for fisheries on highly migratory species”

Project acronym: RECOLAPE

Participating partners:

Deliverable D.4.1- List of explanatory factors for standardizing CPUE series

Deliverable coordinated by Daniele Gaertner (IRD) and Maitane Grande (AZTI)

Index

Executive summary 3

1. List of explanatory variables 5

Executive summary

This deliverable is part of the work package 4 Task 4.1 within the project MARE/2016/22: "Strengthening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species" (acronym RECOLAPE). This project seeks to provide solutions to certain needs, in terms of data collection, identified both by the scientists involved in the stock assessment of the tuna RFMOs (Regional Fisheries Management Organizations) and by the RCG-LP (Regional Coordination Group – Large Pelagic), including the development of protocols for collecting new data needs identified by end users around the FADs (fish agregating devices)

Stock assessments that rely on time series of abundance indices, derived from commercial catch per unit effort (CPUE) data, generally assume that catchability remains constant over time. However, when fishing efficiency is improved, as is the case of the tuna purse seiners, the catchability increases, and the adopted assumptions are not met which could derive in a poor scientific advice. The introduction of FADs in the tropical tuna fishery and FAD associated technology is the most significant innovation introduced historically in this fishing sector which has broken the link between searching time and effective fishing effort for DFAD sets. The remote detection of satellite-tracked buoys attached to FADs and aggregated biomass monitoring by means of echo sounder integrated in the buoy has reduced the searching time and increase the proportion of successful sets. In addition, the use of supply vessels, which can visit DFADs and inform purse seiners on the fish aggregations around them, also contributes to the efficiency of some purse seiners. Because of abundance indices for tuna are derived from commercial CPUE, distinguishing between the impacts of technological innovation and natural variations in fish abundance is crucial. In this scenario, and as recommended by CECOFAD and from the 2016 EU WG on CPUE standardization, the access to non- official data for standardizing the CPUE on FADs is fundamental. Information on the non- official data that reflect the technology evolution and increase on efficiency as is the use of FADs, relationship with supply vessels and other information not provided in the Data Collection Framework (DCF) should be recovered and evaluated.

On this basis, the main goal of this Task (4.1.) is to gather the useful information needed to correct raw CPUE series. The group has mainly work on the census of the candidate variables and quality protocols from different sources (DCF and other), to check the quality of the data, the level of resolution (i.e., set by set, 1° square, etc.) and eventually to propose proxies when the most accurate variable is lacking, for the useful information needed to correct tropical tuna purse seiner CPUE series, specifically for DFAD-fishing activities. This is, we are focused mainly on the recollection of the potential explanatory variables and quality protocols, rather than on the CPUE based-GLM standardization models. This way, deliverables from WP4 will be direct inputs for future EU PS CPUE standardization workshops, such as task 1.3 of CECOFAD2.

On this regard, the group has done a census of data, identification of the data source in each case, and has proposed templates to collect and integrate EU data in a standardize format. Some of the information was already available in the research centers (i.e. IEO, IRD, AZTI), but certain information has been provided by the fishing companies and satellite buoy providers. In addition, before any data exchange and integration, fishing companies should give consent for the data use in the required resolution. Thus, an informative document has been developed for informing the industry on the progress and needs of this working package (Annex 1).

1. List of explanatory variables

Variables*	Time scale	Spatial resolution	Historic	Description of Raw data	Source of information				Responsible for data preparation	Deadline
					Observer data	FAD Logbooks	Buoys providers	Fishing companies		
FAD (buoy) density	Monthly	1x1	2010-2017	Output from RECOLAPE W.P.4.3.			X		AZTI/IRD	15/08/2018
Information on buoys: Model	Ideally daily from individual positions/activities (for buoy transmission data high resolution would be helpful)	Ideally from individual positions/activities (for buoy transmission data high resolution would be helpful)	2010-2017	The model of the buoy (unique ID) give us information about the buoy specific characteristics (echo-sounder, frequency…etc.). High resolution data will be ideal to proceed with variable selection for CPUE standardization. Different data source will be explored: (i) The one provided by the observer and FAD-logbooks which give information on the buoy model in each activity (activities with followed buoys and other not followed but found buoys); (ii) the one extracted from buoy transmissions which give information of the followed buoys model.	X	X	X		AZTI/IRD/IEO	15/08/2018
Followed Buoys	Ideally daily (high resolution data)	Ideally of individual positions (high resolution data)	2010-2017	Spanish - Information of the owned buoy by vessels is available. This information should be completed with information on the collaborations between purse seiners (working groups at sea) which should be provided by the fishing companies. French - Information of the followed buoys or ID of the vessels that follow each buoy is available			X	X	AZTI/IRD/fishing companies	15/08/2018
Relationship between support and purse seiner	Yearly		2010-2017	% of dedication to each purse seiner: #year #purse seiner (name codified) # support vessel (name codified) # dedication to the purse seiner (0-1)				X	AZTI/IRD/fishing companies	15/08/2018
Activities with FOB	Daily	High resolution	2010-2017	Information about the number of visits, deployments and deactivation/lost (for each vessel, including other flags if possible)	x (historical 2010-2017)	x (2015-2017)			IRD/IEO	15/08/2018

Annex 1

Strengthening Regional cooperation in the area of large pelagic fisheries data collection (RECOLAPE)

Data collection strategy for use in standardization of CPUE or in alternative abundance indices in tropical tuna fisheries

The objective of this document is to inform tropical tuna purse seiner fishing sector about the objectives of the RECOLAPE project, its progress, and the use of non-official data that would be carried out in the frame of the project for the standardization of CPUE and research on alternative abundance indices.

General Objective of the project:

The overall objective of the project is to strengthen the regional cooperation in the area of biological data collection for fisheries targeting highly migratory species. In the current context where regional cooperation will evolve from a single meeting (RCM – Regional coordination meeting) to a continuous process that will have greater responsibilities (RCG – Regional coordination group), the project will seek to enable Member States to build up experience in new areas of regional cooperation, that allow them to properly advance in the establishment of the Large Pelagic Regional Coordination Group (RCG -LP). This project will be valuable to improve the coordination among EU Member States in the fisheries data collection field in support of stock assessment and fisheries advice. At the same time, this project seeks to provide solutions to certain needs, in terms of data collection, identified both the by scientists involved in the stock assessment of the tuna RFMOs and by the RCM-LP.

During the project, different actions will be developed, such as the design of regional sampling plans for large pelagic stocks, the development of tools and protocols for collecting new data needs around the FADs (fish aggregating devices), test alternative onboard data collection methods, or design a proper regional framework to assess the data quality. Finally, project outputs will be disseminated through a regional consultation process where all Member States participating in large pelagic data collection (whether they are part or not of the project) identify points of consensus and/or disagreement.

For this purpose, this project will address the following specific objectives:

- Facilitate the evolution of the large pelagic RCM towards the large pelagic RCG
- Design of a Regional Sampling Plan for large pelagic fisheries, thus facilitating the transition from individual national Work Plans towards regional Work Plans
- Develop tools and protocols for collecting new data needs identified by end users around the FADs

- Test alternative data collection methods for those cases where traditional methods present data deficiencies
- Facilitate cooperation among Member States in order to improve the procedures to assess the quality of biological data on large pelagic fisheries, both at the national and regional levels.
- Identify points of consensus and/or disagreement that may arise during the coordination process among Member States dealing with large pelagic fisheries.

To this end the project is structured along the following work packages:

WP1- Large Pelagic Regional Coordination group structure

WP2-Design of Regional Sampling Plan for 2019

WP3-Specific pilot studies:

WP3.1- Development of tools for FAD data collection/transmission:

WP3.2- Electronic Monitoring System (EMS) feasibility study for longlines

WP4- Data collection strategy for use in standardization of CPUE or in alternative abundance indices in tropical tuna fisheries.

WP5-Procedures to assess the quality of biological data collected at regional level

WP6- Regional consultation and training of Member States

To reach the objectives defined in each of the working packages the project relies on the collaboration of the scientific and fishing sector. WP4 is directed to the improvement of the management advice for the tropical tuna purse-seine fisheries, and, to make progress, the input of the fishing sector is essential.

WP4- Data collection strategy for use in standardization of CPUE or in alternative abundance indices in tropical tuna fisheries.

Overarching objectives under this WP will be to develop a data collection strategy of non-official data ("new" data not collected on a routine basis; technology, crew…), to be used in combination with traditional data for tropical tuna purse seiner CPUE standardization, or in alternate abundance indices in tuna fisheries.

The relationship between catch per unit effort (CPUE) and abundance is central to stock assessment models and thus, changes in this relationship will ultimately result in changes in scientific diagnostic and associated management advice. In tuna fisheries, commercial data are traditionally used to compute CPUE and to derive indices of abundance for stock assessments, due to the lack of fishery-independent information. In the lack of direct estimate of abundance, at least until recently, an important number of tuna assessments are conducted based on CPUE from a few fleets/countries, as access to relevant data differs between Member States. This results in partial coverage and associated uncertainty in the interpretation of outcomes (e.g. whether observed trends in CPUE reflect actual changes in abundance or changes in catchability due to improved fishing efficiency and fishing strategy of the fleets). In the particular case of tropical tunas, there is no established approach to discriminate the fishing effort of purse seiners targeting free schools and FADs. For example, in the FAD fishing mode, there is no clear criteria on how to integrate the assistance of the support vessels in the calculation of the fishing effort, and how to estimate their contribution to the improvement in efficiency of the tropical purse seiners.

Alternatively, non-conventional information, such as the buoy acoustic signals may be used to directly estimate the local and regional relative abundance of tunas under drifting FADs, i.e., the Buoy-derived Abundance Index (BAI). Such type of information must be compared between different types and models of buoys in a particular time and area strata and then combined on a larger scale to implement direct indices of abundance which could be compared to the conventional CPUE-based indices in recent years.

Task 4.1 – Selection of variables to be collected

In order not to overlap with other actions/projects funded by DG MARE, this WP will focus on the census of the candidate variables and quality protocols from different sources (conventional and non-official data), to check the quality of the data, the level of resolution (i.e., set by set, 1° square, etc) and eventually to propose proxies when the most accurate variable is lacking, for the useful information needed to correct tropical tuna purse seiner CPUE series, specifically for DFAD-fishing activities. This is, we will focus mainly on the recollection of the potential explanatory variables and quality protocols, rather than on the CPUE based-GLM standardization models. This way, deliverables from WP4 will be direct inputs for future EU PS CPUE standardization workshops, such as task 1.3 of CECOFAD2

Regarding non-official data, the variables included in the Table 1 has been identified as candidates. The information of the buoy transmission and FAD use is available on the research centers (i.e. IRD, IEO and AZTI). During the CPUE standardization the data will be introduced as explanatory variables, always safeguarding the confidentiality of the enterprises. Other information as the collaboration between purse seine-supply vessels or purse seine-purse seine should be completed or reported, to have reliable indicators of the fishing effort. The research centers should work with the fishing enterprises to try to complete the information required.

Table 1. Non-official data to be explored as potential candidate explanatory variables into the GLM "Lasso" method.

Variables*	Time scale	Spatial resolution	Historic	Description of Raw data	Source of information				Responsible for data preparation	Deadline
					Observer data	FAD Logbooks	Buoys providers	Fishing companie s		
FAD (buoy) density	Monthly	1x1	2010-2017	Output from RECOLAPE W.P.4.3.			X		AZTI/IRD	15/08/2018
Information on buoys: Model	Ideally daily from individual positions/activities (for buoy transmission data high resolution would be helpful)*	Ideally from individual positions/activities (for buoy transmission data high resolution would be helpful)*	2010-2017	The model of the buoy (unique ID) give us information about the buoy specific characteristics (echo-sounder, frequency...etc.). High resolution data will be ideal to proceed with variable selection for CPUE standardization. Different data source will be explored: (i) The one provided by the observer and FAD-logbooks which give information on the buoy model in each activity (activities with followed buoys and other not followed but found buoys); (ii) the one extracted from buoy transmissions which give information of the followed buoys model.	X	X	X		AZTI/IRD/IEO	15/08/2018
Followed Buoys	Ideally daily (high resolution data)**	Ideally of individual positions (high resolution data)**	2010-2017	Spanish - Information of the owned buoy by vessels is available. This information should be completed with information on the collaborations between purse seiners (working groups at sea) which should be provided by the fishing companies. French - Information of the followed buoys or ID of the vessels that follow each buoy is available			X	X	AZTI/IRD/fishing companies	15/08/2018
Relationship between support and purse seiner	Yearly		2010-2017	% of dedication to each purse seiner: #year #purse seiner (name codified) # supply vessel (name codified) # dedication to the purse seiner (0-1)				X	AZTI/IRD/fishing companies	15/08/2018
Activities with FOB	Daily	High resolution	2010-2017	Information about the number of visits, deployments and deactivation/lost (for each vessel, including other flags if possible)	x (historical 2010-2017)	x (2015-2017)			IRD/IEO	15/08/2018

**Information of the EU and associated flag states if possible (anonymous ID for vessels)*

*** If high resolution cannot be obtained the best resolution available should be provided*

Task 4.2 – Alternative indices of abundance

In the case of alternative indices of abundance, through the use of echo sounder buoys attached to DFADs, the specific data collection/treatment needed are the following:

- Comparing and improving the current algorithms used to filter out erroneous and non-valid data from the echo sounder buoy databases (wrong positions, wrong biomass estimation, on-board positions)

- Comparing and tentatively improving the current algorithms used to provide biomass estimates for tropical tuna species for different buoy models.

To progress in the identified tasks IRD and AZTI will work in close collaboration. While IRD is working with Marine Instruments buoys, AZTI has great experience in Satlink and have made progress in Marine Instruments. Therefore, both organizations will work on:

- Comparison and standardization of the filtering criteria for each buoy model (bathymetry, voltage…)
- Assess common indices of uncertainty for biomass estimates from echo sounder buoys (for presence/absence and eventually catch categories). The assessment will be carried out comparing the catch data (logbooks and/or observed data) and echo sounder estimates. IRD scientists will estimate the indices of uncertainty for Marine Instruments buoys, using the echo sounder and observers' database from the French Fleet. AZTI scientists will be in charge of assessing such indices for Satlink buoys, using the echo sounder data from the Spanish Fleet and observers' and/or logbook data provided by the IEO the.

Task 4.3- Developing dedicated algorithms to provide the total number of active beaconed FADs at a spatial and temporal stratum

Disposing of the total number of active buoys at sea is key to provide alternative abundance indices as well as to improve the standardization procedure. Consequently, one task of this WP will be devoted in developing dedicated algorithms to provide the total number of active beaconed FADs at a spatial and temporal stratum from the confidential databases provided to national scientists by each EU fleet.

AZTI and IRD would work in close collaboration in this task on the following specific duties:

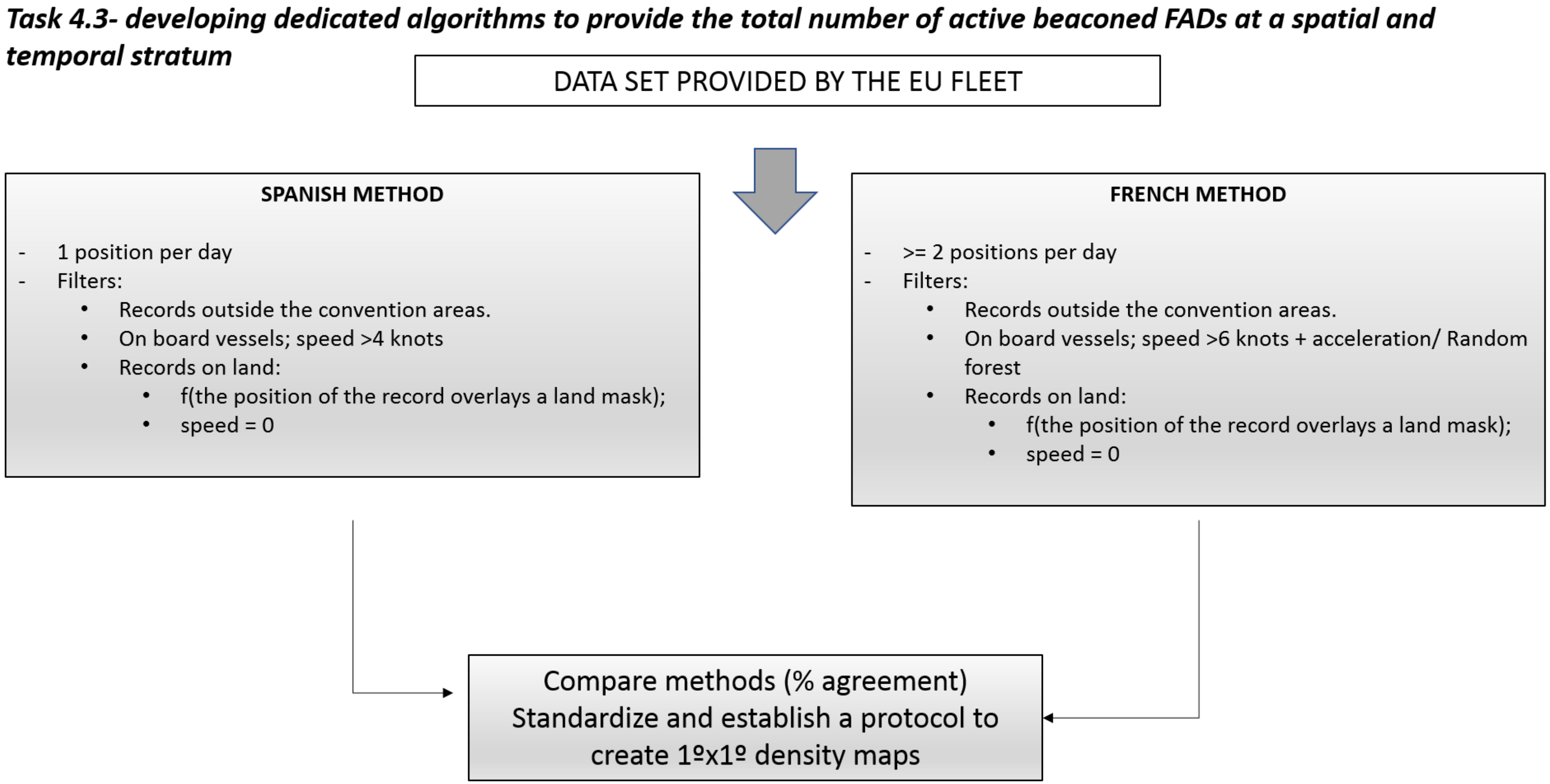
- Specification of the buoy position data (e.g. number of positions per day for each buoy brand)
- Description of the database of buoys positions (For each ocean and year: Number of buoys, brand/type of buoy) for French/Spanish
- Description of data filtering algorithms used in Spanish/French method
- Run the two algorithms on a common database (French + Spanish) and comparison of outputs (number of on board/at-sea positions, number of wrong positions filtered)
- Adopt a common protocol for FAD density estimates that will be used to provide data for CECOFAD2 and for RFMOs

Working on a common database is essential to compare the outputs of the French/Spanish data filtering algorithms. Using a subset of the buoy positions' data over a reduced time window in the past (e.g. a subset of the buoys used during one month during 2016 in the Indian and Atlantic Ocean) will be sufficient to test the algorithms. To this purpose, this WP involves the creation of a common working file merging the positions of the buoys deployed by both the French and Spanish during such time window.

The data exchange will be subjected to the data exchange agreement signed by AZTI, IRD and IEO.

The working plan diagram is included in Figure 1.

Figure 1. Working plan on the frame of RECOLAPE Task 4.3- developing dedicated algorithms to provide the total number of active beaconed FADs at a spatial and temporal stratum



FRAMEWORK CONTRACT – MARE/2016/22 "Strengthening regional cooperation in the area of fisheries data collection" Annex III "Biological data collection for fisheries on highly migratory species"

Project acronym: RECOLAPE

WP4 – Data Collection Strategy for use in Standardization of CPUE or in Alternative Abundance indices in tropical tuna

D.4.2 - Documented algorithms for cleaning the acoustic signal by type of buoy and for providing comparable Buoy-derived Abundance Index (BAI) between fleets

Deliverable coordinated by Manuela Capello (IRD) and Maitane Grande (AZTI)

Index

Executive summary 3
1. Background 5
2. Objective 6
3. Acoustic Data filtering criteria 7
4. Algorithms for converting acoustic Data into Biomass 8
5. Definition of common indicators of uncertainty 13
6. Preliminary estimates of the algorithms´ performance 14
7. Conclusions and Recommendations 18

Executive summary

The fishing efficiency and dynamics of the tropical tuna purse seine fleets are evolving very rapidly due to the fast technological development (Torres-Irineo et al, 2014) and the sharp increase of the use of Fish Aggregating Devices (FADs) (Scott & Lopez, 2014). This fact makes it difficult to obtain reliable CPUE indices for tropical tunas from purse-seine fisheries with FADs. Therefore, initiatives such as the EU funded RECOLAPE project, is focusing on understanding the use of FADs in tropical purse seine tuna fisheries and trying to provide reliable estimates of abundance indices. The instrumented echosounder buoys attached to FADs have been identified as being a valuables platform for the observation of presence of tuna and biomass underneath the FADs, which can be used to develop alternative indices of abundances of tropical tuna, independent from catch (Baidai et al. 2018; Orue et al., 2019). The information provided for these tools is being used for fisheries purposes, but with a suitable evaluation of the information provided, the estimates could be used for evaluation of the tropical tuna stocks. As such, the WP 4 and specifically the Task 4.2 is devoted to developing and test methods for the estimation of reliable estimates of tuna presence and abundance underneath the FADs.

EU Research institutions (i.e. AZTI and IRD) have worked in close collaboration with EU tropical tuna associations (ORTHONGEL, ANABAC and OPAGAC) in the recovery of acoustic information provided by buoys (see D.4.3) and the development of methods for the use of acoustic information for scientific uses. First, to obtain valid acoustic tuna echoes, a set of filters has been defined and applied to the raw database. Then, IRD has focused on the analysis of information provided by Marine Instruments (MI) buoys and AZTI has worked on Satlink buoys.

IRD developed a procedure for estimating the presence/absence of tuna and the size class of the tuna aggregation based on the acoustic data obtained from M3I buoys. A supervised learning algorithm (random forest classification algorithm) is applied, for each ocean, to translate the raw outputs provided by the buoys into metrics of tuna presence and abundance. The training datasets for each ocean were constructed by cross-matching the observer's data and the daily acoustic matrices corresponding to the same buoy ID, selecting the acoustic sounding of the day before the set. The random forest approach has shown a very good efficiency for pattern recognition of presence and absence of tuna aggregation under FADs regardless of the ocean (accuracy 0.75 and 0.85 in the Atlantic and Indian Oceans, respectively). The procedure is less accurate for estimating the precise range of aggregation sizes (accuracy 0.5 and 0.45 in the Atlantic and Indian Oceans, respectively).

AZTI developed a procedure to be applied for Satlink buoys, based on the model developed by Lopez et al., 2016. The model is based on the best available knowledge of the vertical behaviour of species and sizes at FADs, and their corresponding target strength (TS) and weight values by group of species for corrected biomass estimations. An echo-integration procedure was conducted repeatedly by applying all possible combinations of depth limits between small and large tuna in the entire depth range. The selected depth limit was the one that had the best coefficients of correlation (r) and determination (r^2) between predicted biomass and the real catch. Finally, to correct the predicted biomass, the error (in tonns) of the uncorrected predicted biomass was modelled using different regression models (polynomials of order 2 (POL2) and 3 (POL3), generalized linear models (GLM), and generalized additive models (GAM) (Hastie and Tibshirani, 1990; Venables and Dichmont, 2004; Wood, 2006)) as a function of the uncorrected predicted biomass. Functions obtained by regression models were used to adjust biomass estimates and obtain the final corrected biomass values. The polynomial of order 3 was selected as the main model. The results showed that the model used in this study (based on existing knowledge of the vertical distribution of non-tuna and tuna species at FADs and mixed TS and weights)

slightly improves the biomass estimates provided by the manufacturer. The improvement is not as large as expected, which could indicate that the large variability in these data is not easily explained by a single model.

1. Background

Fishery stock assessment models are demographic analyses designed to determine the effects of fishing on fish populations and to evaluate the potential consequences of alternative harvest policies (Methot & Wetzel, 2012). Assessing reliable abundance indices is key for any fish stock assessment model and it is commonly one of the most difficult tasks. This is even more complicated in the case of highly migratory fish stocks, such as tuna, were conventional fishery-independent surveys are in general not applicable. In the absence of fishery-independent information, most of the abundance indices used in tuna stock assessments are derived from estimates of Catch per Unit Effort (CPUE), corresponding to the biomass of fish caught as a function of effort (Quinn & Deriso, 1999).

Relative abundance indices based on CPUE data are problematic (Maunder et al., 2006), as catch data is usually biased by fishing effort, spatial and temporal coverage, and other limiting factors of fisheries-dependent data. The primary assumption behind a CPUE-based abundance index is that changes in the index are assumed to be proportional to changes in the actual stock abundance (Maunder & Punt, 2004), being catchability (q) -the portion of the stock captured by one unit of effort – assumed as a constant coefficient of proportionality. One of the associated difficulties is that q is rarely constant and depends on several different components, such as those related to changes in the fishing efficiency and dynamics of the fleet.

Tropical tuna purse seining is one of such fisheries where both factors, fishing efficiency and dynamics of the fleet, are evolving very rapidly due to the fast technological development (Torres-Irineo et al, 2014) and the sharp increase of the use of Fish Aggregating Devices (FADs) (Scott & Lopez, 2014). This fact makes it difficult to obtain reliable CPUE indices for tropical tunas from purse fisheries fishing with FADs. Therefore, initiatives such as the EU funded RECOLAPE project, is focusing on understanding of the use of FADs in tropical purse seine tuna fisheries and trying to provide reliable estimates of abundance indices. The collaboration between science and industry, in the context of this and other projects (i.e. CECOFAD), is clearly improving the understanding of the FAD use but also the availability of data with great potential for improving CPUE indices and for developing new novel abundance indicators.

The collaboration with the French and Spanish vessel-owners associations (ORTHONGEL, ANABAC and OPAGAC) and the buoy-providers companies (Marine Instruments, Satlink and Zunibal), has made it possible the recovery of the information recorded by the satellite tracking buoys used by the French and Spanish tropical tuna purse seiners and associated fleets in the Atlantic and Indian Oceans for the period 2006-2018 (see D.4.3). These instrumental buoys inform fishers remotely in near real-time about the accurate geolocation of the FAD and, for echosounder buoys, the presence and abundance of tuna aggregations underneath them.

Apart from its unquestionable impact in the conception of a reliable CPUE index from the tropical purse-seine tuna fisheries operating on FADs, echosounder buoys have also the potential of being a privileged observation platform to evaluate the presence of tuna and abundances of tunas and associated species using catch-independent data (Dagorn et al., 2006; Lopez et al., 2014; Capello et al., 2016; Santiago et al., 2016; Baidai et al., 2017, 2018). This alternatively, non-conventional information, such as the buoy acoustic signals may be used to directly estimate the presence of tuna and the local and regional relative abundance of tunas under drifting FADs, i.e., the Buoy-derived Abundance Index (BAI). Such type of information must be compared between different types and models/brands of buoys, over different time and area strata, and then combined at a larger scale to implement direct indices of abundance which could be compared to the conventional CPUE-based indices in recent years. In this sense the objective of this task is to recover historical acoustic information on echosounder buoys and the development of protocols for the analysis of the acoustic information for each buoy model used by the EU fleet in the Atlantic and Indian Oceans.

2. Objective

This deliverable is part of the Work Package 4 within the project MARE/2016/22: "Strengthening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species" (acronym RECOLAPE). The overall objective of the project is to strength the regional cooperation in the area of biological data collection for fisheries on large pelagic fish. This purpose will be valuable to improve the coordination among European MS in the fisheries data collection field in support of stock assessment and fisheries advice. At the same time, this seeks to provide solutions to certain needs in terms of data collection identified by scientists involved in the stock assessment of tuna RFMOs and by expert groups like the RCG-LP.

To reach this purpose, the RECOLAPE project addresses several objectives:

- Facilitate the evolution of the RCM-LP towards the RCG-LP: the goal is to evolve from a single meeting to a continuous process that will have greater responsibilities in support of stock assessment and fisheries advice.
- Design a RSP (Regional Sampling Plan) for large pelagic stocks: facilitating the transition from individual national work plans towards regional ones.
- Develop data collection strategy and tools regarding additional data (not yet collected on a routine basis) on FADs. Such additional data could be used in combination with traditional CPUE or for building alternative abundance indices.
- Test alternative data collection methods for those cases where traditional methods present data deficiencies, for example for data collected using Electronic Monitoring System (EMS).
- Facilitate cooperation among MS in order to improve and develop common data quality assessment procedures at national and regional levels.
- Identify points of consensus and/or disagreement that may arise during the coordination process among organizations dealing with large pelagic fisheries data collection. The idea is to identify a framework of rules and feedback to improve future coordination or expand it on other fisheries/species.

In the case of alternate indices of abundance, through the use of echosounder buoys attached to FADs, the specific data collection/treatment needed are the following:

- Comparing and improving the current algorithms used to filter-out erroneous and non-valid data from the echosounder buoys databases (wrong positions, wrong biomass estimation, on-board positions)

- Comparing and tentatively improving the current algorithms used to provide biomass estimates for tropical tuna species for different buoy models.

To progress in the identified tasks IRD and AZTI worked in a close collaboration. While IRD is working with Marine Instruments buoys, AZTI has more experience in Satlink. In order to meet with the objective of this working package both organisms have worked on the following specific tasks:

- Definition of the acoustic data-filtering criteria
- Development and description of algorithms for converting acoustic data into biomass data
- Definition of common indicators of uncertainty in biomass estimates and estimation uncertainty.

The present deliverable, D.4.2 *(Documented algorithms for cleaning the acoustic signal by type of buoy and for providing comparable Buoy-derived Abundance Index (BAI) between fleets)* presents the progress done in establishment of procedures for the analysis of acoustic information obtained from echosounder buoys used by the tropical tuna purse seiner fleet. In order to work on these objectives a working group was organized in AZTI (Pasaia, Spain) from the 15th to 17th of January. The agenda can be found in the Annex 1. Information on the workshop can be found in the following link:

https://azti.sharepoint.com/sites/Proyectos/RECOLAPE/default.aspx?RootFolder=%2Fsites%2FProyectos%2FRECOLAPE%2FDocumentos%20compartidos%2FWP4%2E%20DATA%20COLLECTION%20STRATEGY%20FOR%20USE%20IN%20STANDARDIZATION%20OF%20CPUE%2FAcoustics%20Workshop&FolderCTID=0x012000DF0BAB7BAD07A24C8125BBC93D958985&View=%7B9920D54A%2D4A41%2D4735%2DAE48%2DCF893A173399%7D

3. Acoustic Data filtering criteria

In order to process the acoustic information obtained from buoy echo-sounders at sea the first step is to apply the filtering criteria defined in the Deliverable 4.3 or Task 4.3 (filtering out erroneous GPS positions, buoy on land and on-board that can give false positives). Additional filters applied to the acoustic data (besides those applied for filtering position data described in the Deliverable 4.3) associated with the bathymetry of the buoy and the battery level:

- Bathymetry: Using high-resolution bathymetry data (British Oceanographic Data Centre, UK, www.gebco.net, resolution of 15 arc-second intervals), acoustic records from buoys in areas with a depth smaller than 150 m (in case of IRD, MI buoys) or 200 m (in case of AZTI,MI and Satlink buoys) are excluded, which prevents including in the analysis false-positive echoes coming from the sea floor and allows the exclusion of acoustic records of FADs that have drifted to coastal areas where tunas are less likely to be found.
- Battery Voltage: According to the buoys manufacturer (MI), data obtained with a voltage of 11.5 V have poor reliability (in terms of location and acoustic measurements). Therefore, this parameter is also being used as a filtering criterion on MI buoys.
- Vertical boundaries: According to the buoy technical specifications, buoys operate with a blind area of 3 (Satlink) to 6 meters (Marine Instruments) which is excluded from the analysis.

Depending on the algorithm used to estimate the presence of tuna or tuna biomass, additional filters can be used to eliminate noise and obtain a representative signal of tuna biomass. On one hand, the algorithm developed by IRD considers the acoustic information contained in the whole sampled water column (3-150m) during a full day (24h), and selects the most important depth layers and time periods, which can vary between oceans, through machine learning (Baidai et al. 2018). On the other hand, the following filters apply to the approach developed by AZTI:

- Vertical boundaries: acoustic information from the shallower layers, <25m, is used as the vertical boundary between non-tuna species and tunas, at about 25 m (Lopez, 2016; Robert et al., 2013). Therefore, depending on the algorithm used, this vertical boundary could be applied to eliminate the noise from the non-tuna species associated with the FAD.
- Time of the day: Samples obtained around sunrise, between 4 a.m. and 8 a.m., are supposed to capture the echosounder biomass signals that better represents the presence and abundance of fish under the FADs. This is the time when tuna is observed closely aggregated around the FADs (Brill et al., 1999; Josse et al., 2000; Moreno et al., 2007a; Harley et al., 2009). For the specific case of comparing the acoustic data with abundance it is important that the echosounder measurements are received when the signal is more representative of the biomass around the FAD (Orue et al., 2019).

4. Algorithms for converting acoustic Data into Biomass

Random forest algorithm

Baidai et al. (2018) developed a procedure for estimating the presence/absence of tuna and the size class of the tuna aggregation based on the acoustic data obtained from M3I buoys. The input data for this algorithm, for each buoy and day, is a 50 × 12 matrix, reporting the acoustic scores recorded at different depth layers (50 layers) and different times of the day (every 2 hours in the default operating mode of the buoy, corresponding to 12 columns). To reduce the dimensionality of the data this matrix is pre-processed, and the temporal and spatial information aggregated. First, in order to aggregate the acoustic samples over time, the data is aggregated over 6 slices of 4 hours each, each slice containing the acoustic sample whose sum of scores correspond to the maximum recorded acoustic energy over the period considered. This results in a matrix of 6 columns (one for each time slot), and 50 rows for the different depth layers. Secondly, clustering methods are used for identifying groups of homogeneous layers. This pre-processing allows obtaining a daily matrix of 6 rows (groups of layers) and 6 columns (time slots), referred to as "daily acoustic matrix" (Figure 1).

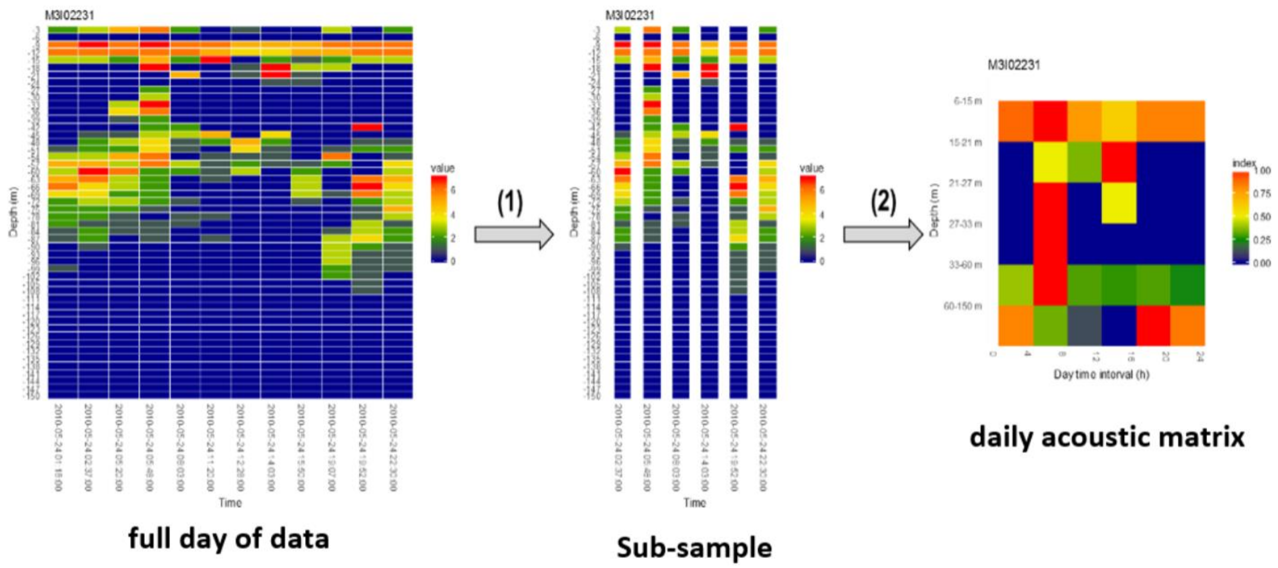


Figure 1. Pre-processing step for standardization and reduction of data dimensionality before classification

A supervised-learning algorithm (random forest classification algorithm) is then applied, for each ocean, to translate the raw outputs provided by the buoys into metrics of tuna presence and abundance. Figure 2 provides a schematic view of the algorithm. The training datasets for each ocean are constructed by cross-matching the observer data and the daily acoustic matrices corresponding to the same buoy ID. Two types of classification algorithms are considered:

- a binary classification algorithm describing the absence or presence of tuna
- a multiclass classification considering different sizes of aggregations under FAD (no tuna, less than 10 tons, between 10 and 25 tons, more than 25 tons).

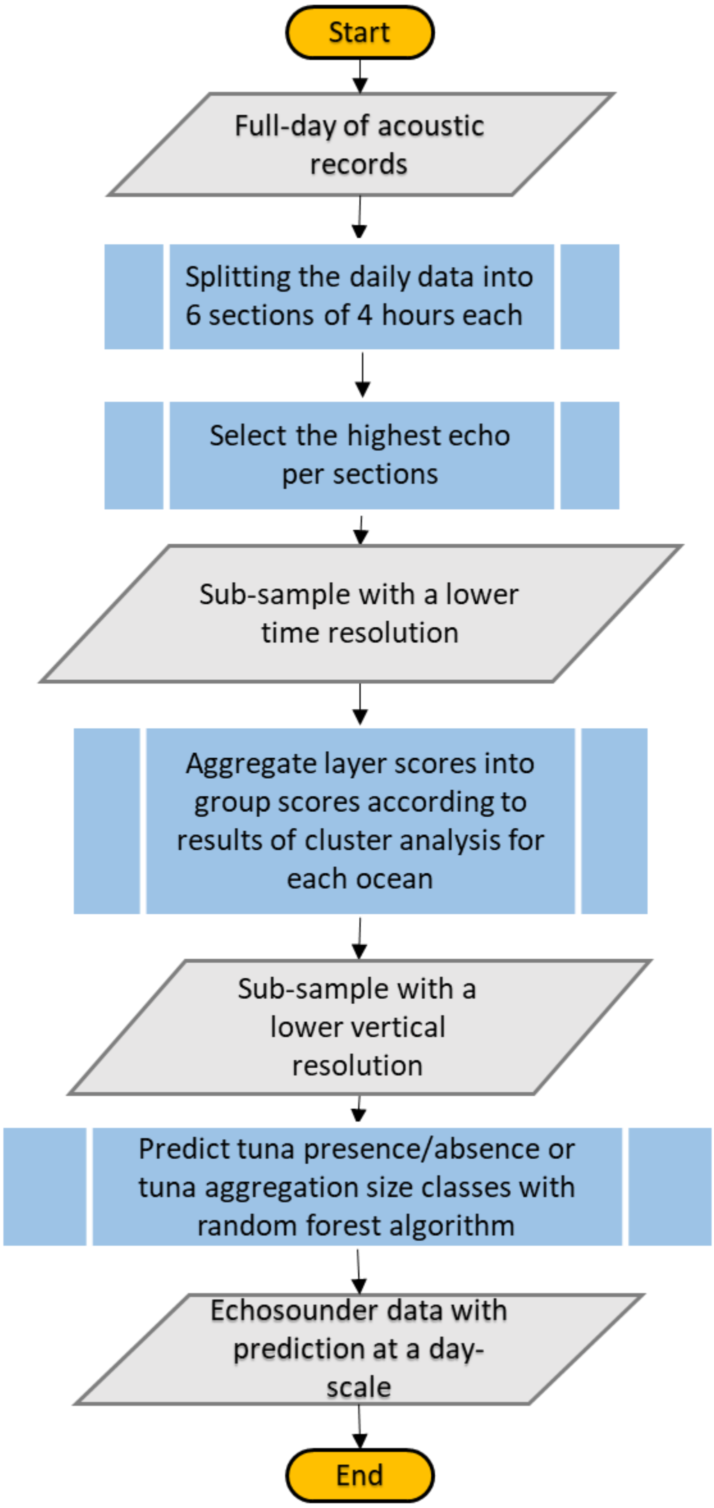


Figure 2: Schematic view of the random forest algorithm used for estimating tuna presence/absence and size classes developed in Baidai et al. (2018)

Algorithm for improving biomass estimates

AZTI has developed a procedure to be applied for Satlink buoys based on the model developed by Lopez et al 2016. This model was based on the best available knowledge of the vertical behaviour of species and sizes at FADs, their corresponding target strength (TS), weight values by group of species (Fig. 3). Following the steps indicated in the Figure 3 the corrected biomass estimations are obtained. First, a depth boundary limiting non-tuna from tuna species at 25 m is established, based on experimental evidences from tagging and acoustic surveys around FADs (Dagorn et al., 2007b; Moreno et al., 2007a; Moreno et al., 2007b; Taquet et al., 2007; Leroy et al., 2009; Govinden et al., 2010; Filmalter et al., 2011; Mitsunaga et al., 2012; Govinden et al., 2013; Schaefer and Fuller, 2013; Matsumoto et al., 2014; Forget et al., 2015). Second, a preliminary limit between small and large tuna at 80 m according to previous studies showing potential segregation of size with depth is established (Moreno et al., 2007a) (Figure 3, step 1). The next step is the election of the most appropriate TS and weight values for non-tuna species, small and large tuna (Figure 3, step 2). For non-tuna species biomass, a TS value of –42 dB was used based on previous field studies (Josse et al., 2000; Doray et al., 2006; 2007; Lopez et al., 2010). The mean weight used for the biomass characterization of this community was 1 kg ind^{-1}, which was estimated from the mean length of the most representative non-tuna species at FADs, and their corresponding weights (Lopez et al., 2016). Because no consistent TS-length relationships exist for yellowfin and bigeye tuna, although it is known for skipjack (Boyra et al., 2018), and the 3 tuna species are usually mixed in similar depth ranges, difficulties exist to accurately know the acoustic backscatter contribution by each species (Josse and Bertrand, 2000). Thus a TS corresponding to mixed species aggregations was chosen (Moreno et al., 2007a) to apply to the supposedly mixed tuna layers. These TS values were measured in situ at FADs for thousands of acoustic shoals at different depth ranges using scientific echo-sounders in the Indian Ocean (Moreno et al., 2007a). These mixed species acoustic shoals showed the following TS values: (i) -35.1 dB for acoustic shoals found at shallower-medium depths (25-80 m), likely corresponding to small tuna and (ii) -29.9 dB for acoustic shoals occupying deepest layers (greater than 80 m), likely corresponding to large tunas. According to the most common tuna sizes caught at DFADs (Chassot et al., 2013; Fonteneau et al., 2013), the depth range for tuna shoals shallower in the water column was considered to be populated by skipjack, yellowfin and bigeye tuna of a mean mass of 2 kg ind^{-1}, whereas the depth for acoustic shoals found at greater depths was assumed to be occupied by larger yellowfin and bigeye tuna individuals with a mean weight of 21 kg ind^{-1}. Then, the predicted biomass is calculated using a depth layer eco-integration procedure (Maclennan et al., 2002)(Figure 2, step 3). A specific acoustic backscattering cross-section value (σ_{bs}, m^2, TS in linear scale; MacLennan et al., 2002) was used to obtain number of individuals for each of the echo-sounder buoy´s layer (n=1, 2, …, 10) according to the presence of each group (non-tuna, tuna at shallow depth layers and tuna at deep layers) in each depth layer. The number of fish per group and layer ($N[n, gr]$) were estimated as follows:

$$\mathrm{N(n, gr)} = \frac{\mathrm{s_a(n)}}{\sigma_{\mathrm{bs(gr)}}} \cdot \mathrm{A(n)} \qquad (1)$$

Where:

$s_a(n)$ = the TVG-corrected (time-varied-gain, a correction function to compensate the signal for spreading and absorption losses; Simmonds and MacLennan, 2005) area backscattering coefficient (Maclennan et al., 2002) in each layer (n);

$\sigma_{(bs(gr))}$ = the mean TS of a group in linear scale and

$A(n)$ = the mean cross sectional area sampled by the beam of the cone for each layer (n).

Then, the total number of fish per group N(gr) were obtained by summing for all layers (2):

$$\mathrm{N(gr)} = \sum_{\mathrm{n}} \mathrm{N(n, gr)} \qquad (2)$$

The estimated number of fish per group (N[gr]) was converted into biomass per group (B[gr], in t) by multiplying the total amount of individuals by their corresponding mean weight (w, in kg) and dividing by 1000.

$$\mathrm{B(gr)} = \frac{\mathrm{N(gr).w(gr)}}{1000} \qquad (3)$$

Where:

$B(gr)$ = the biomass estimated per fish group (in t);

$N(gr)$ = the number of individuals per group; and

$w(gr)$ = the average weight of an individual of a particular group (in Kg) used to convert number of individuals in weight.

The echo-integration procedure was conducted repeatedly by applying all possible combinations of depth limits between small and large tuna in the entire depth range (i.e., having the virtual limit in 25 m, 36m, 47m, 59m, 70m, 92m, 104m and 115m) (Figure 3, step 4). The selected depth limit was the one that had the best coefficients of correlation (r) and determination (r^2) between predicted biomass and the real catch (Figure 3, step 5). Finally, to correct the predicted biomass, the error (in tonns) of the uncorrected predicted biomass was modelled using different regression models (polynomials of order 2 (POL2) and 3 (POL3), generalized linear models (GLM), and generalized additive models (GAM) (Hastie and Tibshirani, 1990; Venables and Dichmont, 2004; Wood, 2006)) as a function of the uncorrected predicted biomass. Functions obtained by regression models were used to adjust biomass estimates and obtain the final corrected biomass values (Figure 3, step 6) (see Lopez et al. (2016) for details).

.

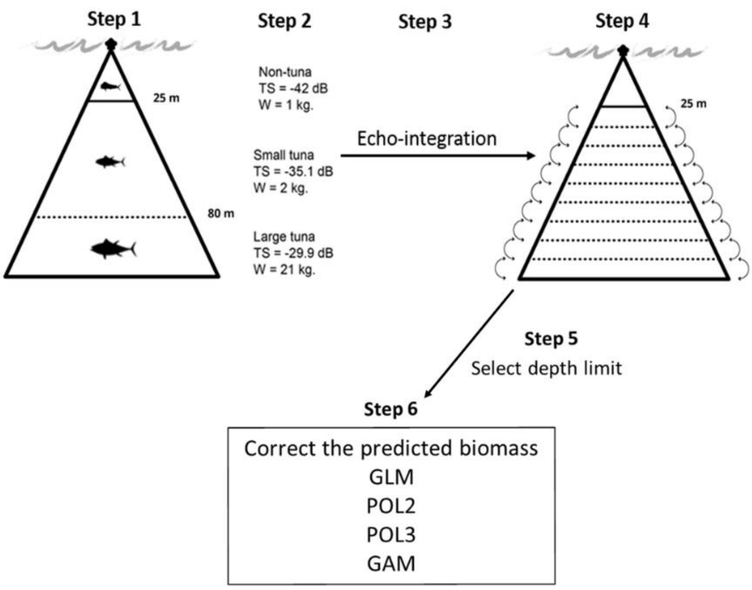


Figure 3. Steps of the AZTI model proposed

5. Definition of common indicators of uncertainty

Several indicators are proposed to evaluate the accuracy and precision of the biomass estimates obtained from different algorithms. Those indicators are based on the comparison between the biomass predictions and catch data reported by observers/logbooks matching the same ID of the buoy.

For the presence/absence of tuna and size class of the catch, a confusion matrix by buoy model and ocean is defined as follows:

		Observed	
		True	False
Predcited	True	True Posiitive (**TP**)	False positive (**FP**)
	False	False negative (**FN**)	True negative (**TN**)

In this case the common indices of uncertainty are defined as follows:

-*Accuracy*: proportion of correctly predicted
Accuracy = (TP + TN) / (TP + TN + FP + FN)

-*Kappa*: statistic measurement of inter-rater concordance which considers the possibility of the agreement occurring by chance

$$\textbf{Kappa} = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

where *Pr(a)* is the total proportion of agreement between the two classifications and *Pr(e)* is the theoretical proportion of agreement expected by chance.

- *Sensitivity*: (recall or true positive rate) measures the efficiency of the algorithm in correctly classifying positive cases.
  Sensitivity = TP / (TP + FN)

- *Specificity*: (or true negative rate) measures the efficiency of the algorithm in correctly classifying negative cases
  Specificity= TN / (FP + TN)

- *Precision*: (also called positive predictive value) is the fraction of correctly predicted presence among tuna presence prediction

  Precision = TP / (TP + FP)

- *F1 Score:* is the harmonic average of the precision and sensitivity

$$\mathbf{F1} = 2 \times \frac{\text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}}$$

For the total biomass for each buoy model a correlation coefficient (r^2) is estimated and residual analysis included.

6. Preliminary estimates of the algorithms´ performance

a. Classification performance of tuna presence

The algorithm developed by Baidai et al. (2018) provides the following performance indicators for assessing tuna presence (Table 1):

Table 1. Performance indicators for assessing tuna presence

Evaluation Metrics	Atlantic	Indian
Accuracy	0.75 (0.02)	0.85 (0.01)
Kappa	0.50 (0.04)	0.69 (0.01)
Sensitivity	0.82 (0.03)	0.78 (0.01)
Specificity	0.67 (0.03)	0.91 (0.01)
Precision	0.72 (0.03)	0.89 (0.01)
F1 score	0.77 (0.03)	0.84 (0.01)

These results imply a very good efficiency for pattern recognition of presence and absence of tuna aggregations under FADs regardless of the ocean. Assessment of importance of predictive factor in the classification, performed through analysis of mean decrease accuracy (Breiman 2001), revealed that detection and characterization of tuna aggregations from echosounder buoys were typically more effective during daytime periods and at ocean-specific depths (figure **4**).

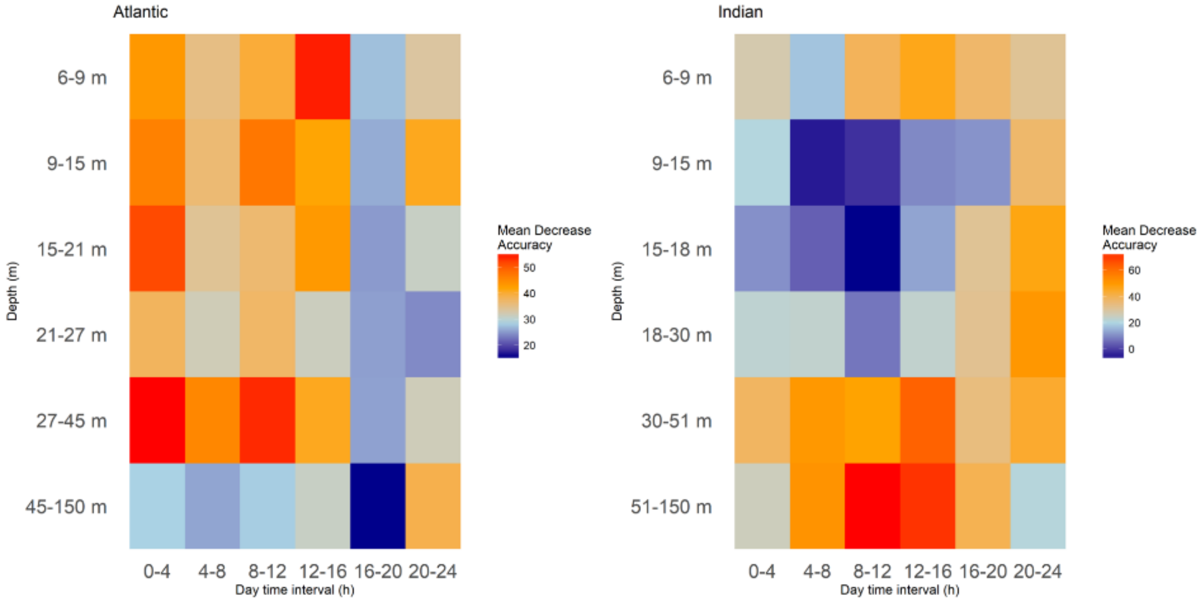


Figure 4: Importance of depth layers and day period in presence/absence classification for the Atlantic and Indian oceans. Each cell represents a combination between a depth and a time period. Color indicates the relevance of the predictor in the classification

b. Classification performance of tuna size aggregation

The algorithm developed by Baidai et al. (2018) provides the following performance indicators for assessing the size class of tuna aggregations (Table 2 and Table 3):

Table 2. performance indicators for assessing the size class of tuna aggregations in the Atlantic Ocean

Atlantic				
	No tuna	**<10 tons**	**[10 , 25 tons]**	**> 25 tons**
Sensitivity	0.66 (0.04)	0.40 (0.05)	0,28 (0.07)	0.37 (0.06)
Specificity	0.84 (0.02)	0.79 (0.02)	0.83 (0.02)	0.87 (0.03)
Precision	0.80 (0.03)	0.33 (0.03)	0.23 (0.05)	0.33 (0.05)
F1 score	0.72 (0.03)	0.36 (0.04)	0.25 (0.05)	0.35 (0.05)
Accuracy	0.50 (0.02)			
Kappa	0.28 (0.03)			

Table 3. performance indicators for assessing the size class of tuna aggregations in the Indian Ocean

Indian				
	No tuna	**<10 tons**	**[10 , 25 tons]**	**> 25 tons**
Sensitivity	0.88 (0.03)	0.19 (0.02)	0.23 (0.03)	0.51 (0.03)
Specificity	0.78 (0.01)	0.89 (0.01)	0.83 (0.01)	0.77 (0.01)
Precision	0.57 (0.01)	0.37 (0.04)	0.31 (0.03)	0.43 (0.02)
F1 score	0.69 (0.02)	0.26 (0.03)	0.26 (0.03)	0.46 (0.02)
Accuracy	0.45 (0.01)			
Kappa	0.27 (0.01)			

Globally, these results imply a lower accuracy for the classification of the size of the aggregation than the assessment of presence/absence only. Assessment of the importance of predictive factor in the classification confirm ocean-specific depth patterns and remains more effective during daytime periods, see Figure 5.

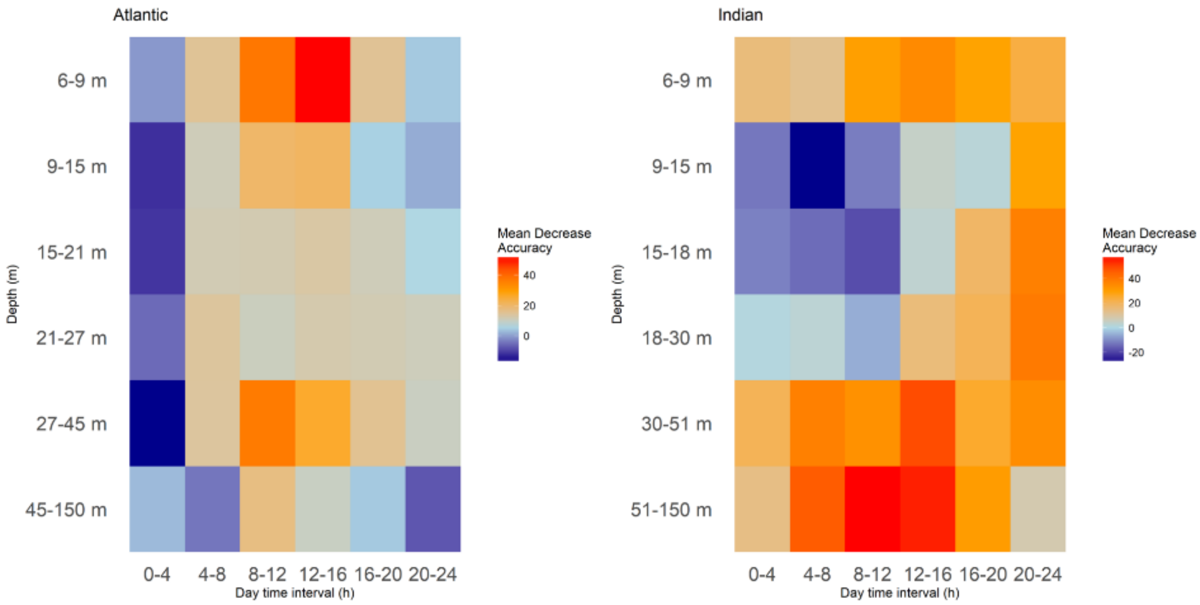


Figure 5: Importance of depth layers and day period in multiclass classification for the Atlantic and Indian oceans. Each cell represents a combination between a depth and a time period. Color indicates the relevance of the predictor in the classification

c. Estimates of biomass underneath the FADs

After applying all possible combinations of depth limits for tunas occupying shallow layers (likely being smaller) and tuna occupying deeper layers (likely being larger), we select the one with the best correlation and determination coefficients between the uncorrected predicted biomass and catch. For Indian Ocean in 287 sets the best correlation value corresponded to limit at 25m or 115m, which suggests that there is not a clear limit between small and large tunas. However, the application of the method by areas showed different potential depth limits between small and large tunas for each zone (Somalia 59 m, Seychelles NW 104 m, Seychelles SE 70 m and Mozambique Channel 104 m). Then, using these depth limits for each region and non-limit in all sets together, we corrected the predicted tuna biomass using four regression models. The corrected tuna biomass estimates using the different regression models and manufacturer biomass estimates were compared with catch of the same fishing set (Table 4).

Table 4. Coefficients of determination (r2) between catch and biomass estimated (manufacturer biomass, Manuf.; predicted biomass, Before correction; and corrected biomass obtained after different model corrections (GLM=generalized linear model; POL2=polynomial of order 2; POL3=polynomial of order 3; GAM=generalized additive model)) for all sets and each region.

Zone	Manuf.	Before correction	GLM	POL2	POL3	GAM
All sets	0.022	0.021	0.021	0.028	0.03	0.027
Somalia	0.025	0.025	0.025	0.026	0.029	0.025
Seychelles NW	0.047	0.050	0.050	0.158	0.158	0.159
Seychelles SE	0.065	0.073	0.073	0.093	0.093	0.073
Mozambique Channel	0.011	0.012	0.012	0.012	0.084	0.012

An improvement is observed when the biomass is corrected by polynomial regressions and GAMs. The corrected biomass obtained with the correction of the GLM model hardly improves, the biomass provided by the manufacturer. We selected polynomial of order 3 as the main model for all sets and regions. The results showed that the model used in this study (based on existing knowledge of the vertical distribution of non-tuna and tuna species at FADs and mixed TS and weights) improves the biomass estimates provided by the manufacturer. This improvement varies by area, being highest in NW Seychelles and in Mozambique Channel, while the improvement is very slight in the Somalia area (Figure 6). However, the improvement is not as large as expected, which could indicate that the large variability in these data is not easily explained by a single model.

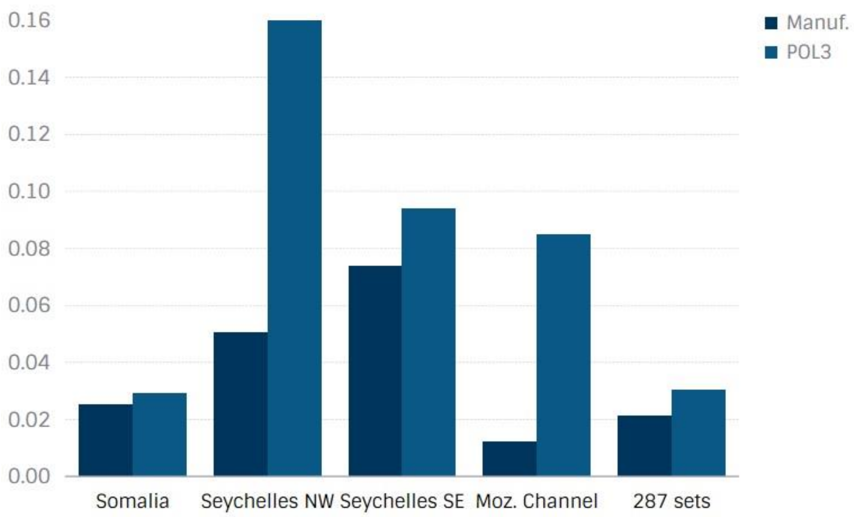


Figure 6. Coefficients of determination (r^2) between catch and biomass estimated by manufacturer (Manuf.) and between catch and final biomass estimations corrected by polynomial model of order 3 (POL3)

7. Conclusions and Recommendations

The collaboration with the purse seiner EU fishing associations has allowed to advance in the recovery of acoustic information provided by the buoys and the development of methods for the use of acoustic information for scientific uses. Common data filtering protocols have been defined and applied in order to select valid acoustic records for the analysis. The algorithm developed for MI buoys has shown a very good efficiency in pattern recognition of presence and absence of tuna aggregation under FADs, regardless of the ocean. This procedure is less accurate for estimating the precise range of aggregation sizes. The method applied on Satlink buoys (based on existing knowledge of the vertical distribution of non-tuna and tuna species at FADs and mixed TS and weights) improves slightly the biomass estimates provided by the manufacturer. However, the improvement of the biomass estimates was not as large as expected, so it should be further improved.

To further advances in the detection of tuna and the estimation of biomass aggregated underneath the FADs, the following recommendations are produced:

- The effect of additional factors on the acoustic signal, such as spatio-temporal and environmental factors could be explored.
- New TS values should be considered in future assessments and new experiments should be conducted for the estimation of new TS (e.g. yellowfin TS).
- Accounting for the spatio-temporal variation on the species and size distribution could help in the estimation of the biomass.
- Further electronic tagging studies should be conducted to assess the associative behavior of fish for different FAD densities and environmental characteristics, since this behavior affects the amount of associated biomass and thus the abundance index.

- In order to assess differences among buoy models further studies should be conducted attaching different buoy models in the same FAD.
- The following technical features of the buoy could be improved to overcome the current limits in the biomass assessment of tuna at FADs through echosounder buoys:
    a. Using two frequencies to improve the discrimination among species, with the same angle for all frequencies
    b. Using split beam echosounders instead of single beams
    c. Using wide bands to have the full spectrum of the response
    d. Transmitting the skippers settings (biomass limits set to transmit the information)

-

References

Baidai, Y., M. Capello, N. Billet, L. Floch, M. Simier, P. Sabarros, and L. Dagorn. 2017. Towards the derivation of fisheries-independent abundance indices for tropical tunas: progress in the echosounders buoys data analysis Y. BAIDAI. 2017–WPTT19–22_Rev1, Mahé.

Baidai Y, Amandé JM, Gaertner D, Dagorn L, Capello M. (2018) Recent advances on the use of supervised learning algorithms for detecting tuna aggregations under fads from echosounder buoys data IOTC-2018-WPTT20-25_Rev1 20th Working Party on Tropical Tunas (WPTT). Victoria, Seychelles.

Boyra, G., Moreno, G., Sobradillo, B., Pe, I., Sancristobal, I., & Demer, D. A. (2018). Target strength of skipjack tuna (Katsuwanus pelamis) associated with fish aggregating devices (FADs). *ICES Journal of Marine Science*, *75*, 1790–1802. https://doi.org/10.1093/icesjms/fsy041

Breiman, L. 2001. Random forests. Machine Learning 45(1):5–32.

Brill, R., Block, B., Boggs, C., Bigelow, K., Freund, E., Marcinek, D., 1999. Horizontal movements and depth distribution of large adult yellowfin tuna (Thunnus albacares) near the Hawaiian Islands, recorded using ultrasonic telemetry: implications for the physiological ecology of pelagic fishes. Mar. Biol. 133, 395-408.

Capello, M., Deneubourg, J. L. , Robert, M. , Holland, K. N. , Schaefer, K. M., & Dagorn, L. "Population assessment of tropical tuna based on their associative behavior around floating objects." *Sci Rep*. 6 (2016): 36415.

Chassot, E., de Molina, A.D., Assan, C., Dewals, P., Cauquil, P., Areso, J., Rahombanjanaharyk, D., Floch, L., 2013. Statistics of the European Union and associated flags purse seine fishing fleet targeting tropical tunas in the Indian Ocean 1981–2012, IOTC-WPTT-13, 44pp.

Dagorn, L., Holland, K., Puente, E., Taquet, M., Ramos, A., Brault, P., ... Aumeeruddy, R. (2006). *FADIO (Fish Aggregating Devices as Instrumented Observatories of pelagic ecosystems): a European Union funded project on development of new observational instruments and the behavior of fish around drifting FADs* (No. IOTC-2006-WPTT-16).

Doray, M., Josse, E., Gervain, P., Reynal, L., Chantrel, J., 2006. Acoustic characterisation of pelagic fish aggregations around moored fish aggregating devices in Martinique (Lesser Antilles). Fish. Res. 82, 162-175.

Doray, M., Josse, E., Gervain, P., Reynal, L., Chantrel, J., 2007. Joint use of echosounding, fishing and video techniques to assess the structure of fish aggregations around moored Fish Aggregating Devices in Martinique (Lesser Antilles). Aquat. Living Resour. 20, 357-366.

Fonteneau, A., Chassot, E., Bodin, N., 2013. Global spatio-temporal patterns in tropical tuna purse seine fisheries on drifting fish aggregating devices (DFADs): Taking a historical perspective to inform current challenges. Aquat. Living Resour. 26, 37-48.

Forget, F.G., Capello, M., Filmalter, J.D., Govinden, R., Soria, M., Cowley, P.D., Dagorn, L., 2015. Behaviour and vulnerability of target and non-target species at drifting fish aggregating devices (FADs) in the tropical tuna purse seine fishery determined by acoustic telemetry. Can. J. Fish. Aquat. Sci. 72, 1398-1405.

Gaertner, D, Ariz, J., Bez, N., Clermidy, S., Moreno, G., Murua, H., & Soto, M. (2016). Objectives and first results of the CECOFAD project. *Collect. Vol. Sci. Pap. ICCAT*, *72*(2), 391–405.

Gaertner, Daniel, Ariz, J., Bez, N., Clermidy, S., Moreno, G., Murua, H., ... Marsac, F. (2016). *Results achieved within the framework of the EU research project: Catch, Effort, and eCOsystem impacts of FAD-fishing (CECOFAD)* (No. IOTC-2016-WPTT18-35).

Govinden, R., Dagorn, L., Filmalter, J., Soria, M., 2010. Behaviour of Tuna a ssociated with Drifting Fish Aggregating Devices (FADs) in the Mozambique Channel, IOTC-2010-WPTT-25.

Govinden, R., Jauhary, R., Filmalter, J., Forget, F., Soria, M., Adam, S., Dagorn, L., 2013. Movement behaviour of skipjack (Katsuwonus pelamis) and yellowfin (Thunnus albacares) tuna at anchored fish aggregating devices (FADs) in the Maldives, investigated by acoustic telemetry. Aquat. Living Resour. 26, 69-77.

Harley, S., Williams, P., Hampton, J., 2009. Analysis of purse seine set times for different school associations: a further tool to assist in compliance with FAD closures? Western and Central Pacific Fisheries Commission WCPFC-TCC5-2009/IP-02.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized additive models. CRC press.

Josse, E., Dagorn, L., Bertrand, A., 2000. Typology and behaviour of tuna aggregations around fish aggregating devices from acoustic surveys in French Polynesia. Aquat. Living Resour 13, 183-192. Devices, Springer, pp. 161-179.
Leroy, B., Phillips, J.S., Nicol, S., Pilling, G.M., Harley, S., Bromhead, D., Hoyle, S., Caillot, S., Allain, V., Hampton, J., 2012. A critique of the ecosystem impacts of drifting and anchored FADs use by purse-seine tuna fisheries in the Western and Central Pacific Ocean. Aquat. Living Resour. 26, 49-61.
Lopez, J., Moreno, G., Sancristobal, I., & Murua, J. (2014). Evolution and current state of the technology of echo-sounder buoys used by Spanish tropical tuna purse seiners in the Atlantic , Indian and Pacific Oceans. *Fisheries Research*, *155*, 127–137. https://doi.org/10.1016/j.fishres.2014.02.033
Lopez, J., Moreno, G., Boyra, G., Dagorn, L., 2016. A model based on data from echosounder buoys to estimate biomass of fish species associated with fish aggregating devices. Fish. Bull. 114.
Maclennan, D.N., Fernandes, P.G., Dalen, J., 2002. A consistent approach to definitions and symbols in fisheries acoustics. ICES J Mar Sci. 59, 365-369.
Maunder, M. N., & Punt, A. E. (2004). Standardizing catch and effort data: a review of recent approaches. *Fisheries Research*, *70*(2–3), 141–159. https://doi.org/10.1016/j.fishres.2004.08.002
Maunder, M. N., Sibert, J., Fonteneau, A., Hampton, J., Kleiber, P., & Harley, S. J. (2006). Interpreting catch per unit effort data to assess the status of individual stocks and communities. *ICES Journal of Marine Science*, *63*(8), 1373–1385. https://doi.org/10.1016/j.icesjms.2006.05.008
Matsumoto, T., Satoh, K., Toyonaga, M., 2014. Behavior of skipjack tuna (Katsuwonus pelamis) associated with a drifting FAD monitored with ultrasonic transmitters in the equatorial central Pacific Ocean. Fish. Res. 157, 78-85.
Methot, R. D., & Wetzel, C. R. (2012). Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research*. https://doi.org/10.1016/j.fishres.2012.10.012
Moreno, G., Dagorn, L., Sancho, G., Itano, D., 2007a. Fish behaviour from fishers´ knowledge: the case study of tropical tuna around drifting fish aggregating devices (DFADs). Can. J. Fish. Aquat. Sci. 64.
Moreno, G., Josse, E., Brehmer, P., Nøttestad, L., 2007b. Echotrace classification and spatial distribution of pelagic fish aggregations around drifting fish aggregating devices (DFAD). Aquat. Living Resour. 20, 343-356.
Orúe B., Jon Lopez, Gala Moreno, Josu Santiago, Guillermo Boyra, Maria Soto, Hilario Murua 2019. Using fishers' echo-sounder buoys to estimate biomass of fish species associated with drifting fish aggregating devices in the Indian Ocean. Revista de Investigación Marina, AZTI, 26(1): 1-13
Quinn, T. J., & Deriso, R. B. (Richard B. . (1999). *Quantitative fish dynamics*. Retrieved from https://books.google.es/books/about/Quantitative_Fish_Dynamics.html?id=5FVBj8jnh6sC&redir_esc=y
Santiago, J., Lopez, J., Moreno, G., Murua, H., Quincoces, I., & Soto, M. (2016). Towards a tropical tuna buoy-derived abundance index (TT-BAI). *Collect. Vol. Sci. Pap. ICCAT*, *72*(3), 714–724.
Schaefer, K.M., Fuller, D.W., 2013. Simultaneous behavior of skipjack (Katsuwonus pelamis), bigeye (Thunnus obsesus), and yellowfin (T. albacares) tunas, within large multi-species aggregations associated with drifting fish aggregating devices (FADs) in the equatorial eastern Pacific Ocean. Mar. Biol. 160, 3005-3014.
Scott, G. P., & Lopez, J. (2014). *The use of FADs in Tuna Fisheries*.
Torres-Irineo, E., Gaertner, D., Chassot, E., & Dreyfus-León, M. (2014). Changes in fishing power and fishing strategies driven by new technologies: The case of tropical tuna purse seiners in the eastern Atlantic Ocean. *Fisheries Research*, *155*, 10–19. https://doi.org/10.1016/j.fishres.2014.02.017
Venables, W.N., Dichmont, C.M., 2004. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. Fish. Res. 70, 319-337.
Wood, S., 2006. Generalized additive models: an introduction with R. CRC press.

Annex 1

RECOLAPE WORKSHOP

DAY 1 : Data description and Position Filtering (WP4.3)

- Metadata and Data description tables
- Review positions filtering criteria
- Outputs on filtering for positions on common Data set and further test (if required)
- Protocol for buoy density estimates

Deliverables: Metadata & Data description tables, List of position filtering criteria, Outputs on common dataset, protocol for density estimates.

DAY 2 : Acoustic Data analysis I (WP4.2)

- Acoustic data filtering criteria
- Algorithms for converting acoustic data into biomass
- Definition of common indicators of uncertainty on biomass estimates.

Deliverables: List of acoustic data filtering criteria, Table Synthetising the Algorithms characteristics, Table on indicators of uncertainty

DAY 3 : Acoustic data analysis II (WP4.2 & BIOFAD)

- Application of the algorithms for converting acoustic data into biomass on the BIOFAD echosounder data
- Derivation of indicators of uncertainty of biomass estimates on BIOFAD echosounder data

FRAMEWORK CONTRACT – MARE/2016/22 "Strengthening regional cooperation in the area of fisheries data collection" Annex III "Biological data collection for fisheries on highly migratory species"

Project acronym: RECOLAPE

WP4 – Data Collection Strategy for use in Standardization of CPUE or in Alternative Abundance indices in tropical tuna

D.4.3 - Developing dedicated algorithms to provide the total number of operational beaconed FADs

Deliverable coordinated by Manuela Capello (IRD) and Maitane Grande (AZTI)

Index

Executive summary 3

1. Background.. 5

2. Objective 6

3. Instrumented buoys derived data description.. 8

    3.1 Technical characteristics of the buoys used by the EU fleet 8

    3.2 Metadata Description tables 11

    3.3 Position and acoustic Databases Description 17

4. Data Processing Protocols 25

    4.1 Data filtering criteria 26

    4.2 Description of algorithms for density estimates 31

Executive summary

In the tropical tuna purse seine fishery, the fishing efficiency and dynamics of the fleet are evolving rapidly due to the fast technological development (Torres-Irineo et al, 2014) and the increase of the use of Fish Aggregating Devices (FADs) (Scott and Lopez, 2014). This evolution makes it difficult to obtain reliable CPUE indices for tropical tunas from purse fisheries fishing with FADs. Therefore, initiatives such as the EU funded RECOLAPE project is focusing on understanding of the use of FADs in tropical purse seine tuna fisheries and trying to provide reliable estimates of abundance indices. As such, one of the objectives of WP 4 of the RECOLAPE EU project is to develop a data collection strategy of non-official data on FADs and standardized protocols to provide indicators of the total number of operational buoys at sea and to improve the CPUE standardization procedure, which requires the efforts from all the stakeholders. This deliverable presents the progress done in buoy data collection for filling data gaps on FADs and presents the work done for the establishment of procedures for buoy data pre-processing (i.e. data filtering protocols) for its use in support of stock assessment and tuna fisheries management.

Under specific data-exchange agreements signed between research organisms (i.e. AZTI and IRD) and EU tuna purse seiner associations (i.e. ORTHONGEL[1], ANABAC[2] and OPAGAC[3]), historical data on buoy positions and acoustics has been gathered. The data from three buoy brands (i.e. Zunibal, Satlink and Marine Instrument) has been gathered in the Atlantic and the Indian Ocean, covering periods 2006 - 2018 and 2010 – for the French and Spanish fleets, respectively. In addition, the technical specifications of each buoy model and brand have been recovered, in order to better understand the functioning of each model.

To develop common indicators of the number of buoys at sea, the raw data need to be pre-processed in order to filter out erroneous locations, data related to failures in satellite communication and location data acquisition, land positions and on-board positions. In order to assess and compare the performances of the filtering methods used by each research center and develop a standardized filtering algorithm, a common EU database was created and shared (i.e. subset of the total number of buoys recovered), integrating the tracks of 2000 buoys (i.e.,1000 buoys from the Spanish and 1000 buoys from the French fleet for each ocean) during 1 month in the Atlantic and Indian Ocean, respectively.

The comparison of the outputs of the filtering algorithm run by IRD and AZTI on the common database demonstrated a high rate of agreement between the two algorithms which were developed based on specific common criteria, validating both methods for data pre-processing. The main differences occurred in the land classification, where the shapefile resolution could impact the filtering of land positions, as well as the size of the land along which the buoys are considered on land (see the Appendix). In addition, minor differences among the two methods occurred in the number of buoys classified as on-board. These differences were higher for the Spanish dataset in the Indian Ocean, since the performances of the algorithms are affected by the characteristics of the databases (i.e. lower performance on shorter tracks

[1] Organisation des producteurs de thon tropical congelé et surgelé
[2] Asociación Nacional de Armadores de Buques Atuneros Congeladores
[3] Organización de Productores Asociados de Grandes Atuneros Congeladores

and smaller temporal resolution). In this sense, in order to minimize the misclassification, the use of high-resolution data is recommended if available. Finally, the methodology for the estimation of the buoy density is defined and applied in the common database.

1. Background

Tropical tuna purse seiners operate globally and have continuously increased their use of Floating Objects (FOBs), including man-made Fish Aggregating Devices (FADs) and logs, since the late 1980s (Fonteneau et al., 2013). The man-made FADs deployed by tuna purse seine fisheries equipped with GPS and echosounder buoys have undoubtedly improved the fishing efficiency of the purse seiners during the last three decades.

Stock assessments that rely on time series of abundance indices derived from commercial catch per unit effort (CPUE) data, generally assume that changes in catchability are accounted for in the standardization process. However, when fishing efficiency is improved, as is the case of the tuna purse seiners FAD fishery, the catchability increases, and the adopted assumptions are not met which could derive in a poor scientific advice (Torres-Irineo et al., 2014). In this sense, the introduction of FADs in the tropical tuna fishery and FAD-associated instrumented buoys (i.e. buoys with echosounder and with GPS location) has broken the link between searching time and effective fishing effort (Torres-Irineo et al., 2014), allowing the detection of satellite-tracked buoys attached to FADs and the monitoring of the biomass aggregated to these FADs by means of echo sounders integrated in the buoys. This has reduced the searching time and increase the proportion of successful sets. Given that abundance indices for tuna are derived from commercial CPUE, distinguishing between the impacts of technological innovation and natural variations in perceived fish abundance is crucial (Torres-Irineo et al., 2014). Thus, while traditionally catch rates for free-swimming schools take into consideration the quantity of fish caught per set and the searching time, catch rate calculations for fishing on FAD-associated schools must take into consideration additional factors which can further complicate the task of achieving good estimates. In this scenario information on the use of FADs is essential.

On the other hand, potential adverse effects on target and non-target species and the marine ecosystem have been identified (Dagorn et al*.* 2012; Gilman et al., 2018). In this sense, potential impacts associated with the increasing FAD deployments, FAD density, drift and distribution of FADs at sea have been described: alteration of normal movements of tuna, increases in skipjack catches (the principal target species), reduction in yield per recruit (juvenile yellowfin and bigeye tuna are regularly caught in FAD sets together with skipjack), increase of bycatch and damage of sensitive coastal habitats and littering of the coastline in case of FADs that are lost and otherwise discarded (Fonteneau *et al.* 2000, Dagorn *et al.* 2012, Maufroy et al., 2015, Gilman et al.,2018).

Despite these concerns, little information is available on FAD use worldwide for a science-based FAD fisheries management, while it is crucial for effort assessment and monitoring the impacts of FADs on pelagic ecosystems and coastal habitats. In this context, filling the data gaps on FADs has become a priority for t-RFMOs and other skate-holders which work on defining standards and procedures for data collection. Based on the t-RFMOs requirements and guidelines, local efforts have been made to develop FADs management plans that should rather be regionally coordinated to improve data quantity and quality. Although efforts are being made from t-RFMOs to record and report information on FADs, due to the complexity of this fishing strategy and the lack of unified data collection and reporting requirements, and an absence of clear guidelines and harmonized definitions for relevant terms or ambiguity among t-RFMOs, still there are significant data gaps on FAD use (Ramos et al., 2017; Lopez et al., 2018). Therefore, efforts from all the stakeholders are required to improve the collection of FAD-related data in a comprehensive way.

In this context, this document presents the data that have been gathered at IRD and AZTI with the collaboration of the fishing industry and buoy providers and presents the current algorithms used for processing these data.

2. Objective

This deliverable is part of the Work Package 4 within the project MARE/2016/22: "Strengthening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species" (acronym RECOLAPE). The overall objective of the project is to strengthen the regional cooperation in the area of biological data collection for fisheries on large pelagic fish. This objective will be valuable to improve the coordination among European MS in the fisheries data collection in support of stock assessment and fisheries management advice. At the same time, the project aims to provide solutions to certain needs in terms of data collection identified by scientists involved in the stock assessment of tuna RFMOs and by expert groups like the RCG-LP.

To reach this objective, the RECOLAPE project addresses several objectives:

- Facilitate the evolution of the RCM-LP towards the RCG-LP: the goal is to evolve from a single meeting to a continuous process that will have greater responsibilities in support of stock assessment and fisheries advice.
- Design a RSP (Regional Sampling Plan) for large pelagic stocks: facilitating the transition from individual national work plans towards regional ones.
- Develop data collection strategy and tools regarding additional data (not yet collected on a routine basis) on FADs. Such additional data could be used in combination with traditional CPUE or for building alternative abundance indices.
- Test alternative data collection methods for those cases where traditional methods present data deficiencies, for example for data collected using Electronic Monitoring System (EMS).
- Facilitate cooperation among MS in order to improve and develop common data quality assessment procedures at national and regional levels.
- Identify points of consensus and/or disagreement that may arise during the coordination process among organizations dealing with large pelagic fisheries data collection. The idea is to identify a framework of rules and feedback to improve future coordination or expand it on other fisheries/species.

Disposing of the total number of operational buoys at sea is key to provide alternative abundance indices as well as to improve the standardization procedure. Consequently, one task of WP4 of RECOLAPE is devoted in developing dedicated algorithms to provide the total number of operational beaconed FADs at a spatial and temporal stratum from the confidential databases provided to national scientists by each EU fleet.

AZTI and IRD have worked in close collaboration in following the steps shown below:

- Specification of the buoy position and acoustic metadata (i.e., format and description of each column in the database)
- Description of the database of buoy positions and acoustics (i.e., for each ocean and year: number or proportion of buoys, brand/type of buoy)
- Description of the data-filtering protocol.
- Running algorithms using a common database (French + Spanish) and comparison of outputs (e.g., number of at-sea positions, number of positions filtered)
- Adoption of a common protocol for FAD density estimates that will be used to provide data for CECOFAD2 and for tuna RFMOs

Working on a common database is essential to compare the outputs of the French/Spanish data filtering algorithms and density estimates. To this purpose, this WP involved the creation of a common working file merging the positions of 2000 buoys deployed by the French and Spanish fleets in the Atlantic and Indian oceans during one month in 2016.

The present deliverable, D.4.3 *(developing dedicated algorithms to provide the total number of operational beaconed FADs at a spatial and temporal stratum)* presents the progress done in buoy data collection for filling data gaps on FADs and the establishment of procedures for the analysis of this data in support of stock assessment and tuna fisheries management. In order to work in these objectives defined a working group was organized in AZTI (Pasaia, Spain) from the 15th to 17th of January. The agenda can be found in the Annex 1. Information on the workshop can be found in the following link:

https://azti.sharepoint.com/sites/Proyectos/RECOLAPE/default.aspx?RootFolder=%2Fsites%2FProyectos%2FRECOLAPE%2FDocumentos%20compartidos%2FWP4%2E%20DATA%20COLLECTION%20STRATEGY%20FOR%20USE%20IN%20STANDARDIZATION%20OF%20CPUE%2FAcoustics%20Workshop&FolderCTID=0x012000DF0BAB7BAD07A24C8125BBC93D958985&View=%7B9920D54A%2D4A41%2D4735%2DAE48%2DCF893A173399%7D

3. Instrumented buoys derived data description

In this section the buoy technical specifications are described, the metadata tables on buoy data available and the composition of the databases available on each research center (i.e. IRD and AZTI).

3.1 Technical characteristics of the buoys used by the EU fleet

EU fleet is mainly working with two buoy brands: Marine Instrument (MI) and Satlink buoys. In a minor extent the Spanish fleet also uses Zunibal buoys and the French fleet uses Thalos buoys. However, the acoustic information of these buoys (i.e. Zunibal and Thalos) still is not being used for scientific purpose for the estimation of tuna biomass. In Tables 1 and 2 the technical characteristics of the main buoy brands and models used by the EU fleet are described (i.e. Satlink and Marine Instrument buoys). This information has been obtained with the collaboration of buoy providers.

Table 1. Technical characteristics of Marine Instruments buoys

MODEL	Buoy model 1	Buoy model 2	Buoy model 3	Buoy model 4
	MSI	**M3I**	**M4I**	**M3I+**
GPS	YES	YES	YES	YES
Number of GPS positions per day (default mode)	**2**	**2**	**2**	**2**
Battery	Solar Panel rechargeable battery and pack of alkaline batteries as a backup	Solar Panel rechargeable battery and pack of alkaline batteries as a backup	Solar Panel rechargeable battery and pack of alkaline batteries as a backup	Solar Panel rechargeable battery and pack of alkaline batteries as a backup
Echosounder	**NO**	YES	YES	YES
Frequency		50 kHz	50, 120, 200 kHz	50 kHz/200kHz
Beam angle (per frequency)		36º(50KHz)	42º (50 kHz), 17° (120kHz), 10° (200kHz)	36º (50kHz)/8º (200kHz)
Power (W)		500	500	500
Resolution per layer (m)		3 m	3 m	3 m
Max depth (m)		150 m	150 m	150 m
Blind area (m)		6 m	3 m	3 m
Frequency of transmission of echosounder data		12 hours	12 hours	3 hours
Frequency of acoustic sampling (ping rate)		5 min	5 min	1 min
Number of echosounder data per day (default mode)		12	12	12 (plus 4 at dawn)

Biomass index specification (range, nature)		Integer 0-7 values for each layer	Integer 0-15 values for each layer	Integer 0-15 values for each layer
Biomass index calculation		Maximum of sum of layers	Maximum of sum of layers	Maximum of sum of layers
maximum detection threshold		230tn	230tn	804tn
minimum detection threshold		0 tn	0 tn	0 tn

Table 2. Technical characteristics of the Satlink buoys.

MODEL	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	D+	**DL+**	**DS+**	**DSL+**	**ISL+**	**ISD+**
GPS	YES	YES	YES	YES	YES	YES
Number of GPS positions per day (default mode)	1	1	1	1	1	1
Batery	non solar; expresed in voltios	solar expresed in %	non solar; expresed in voltios	solar expresed in %	solar expresed in %	solar expresed in %
Echosounder	NO	NO	YES	YES	YES	YES
Frequency			190,5 KHz	190,5 KHz	190,5 KHz	Double frequency (200 KHz and 38 KHz), but the data exported are in 200 KHz
Beam angle (per frequency)			32º	32º	32º	32º
Power (w)			100	100	100	200
Resolution per layer			11,2	11,2	11,2	11,2
Max depth			115	115	115	115
Blind area			3 m	3 m	3 m	3 m
Frequency of transmission of echosounder data			24 hours	24 hours	transmit the data if the value recorded is 50% above the previpous record	transmit the data if the value recorded is 50% above the previpous record
Frequency of acoustic sampling (ping rate)			3 records/daily (1 h to dusk, at dusk, after dusk) if > than 0	3 records/daily (1 h to dusk, at dusk, after dusk) if > than 0	15 min	15 min
Number of echosounder data per day (default mode)			3	3	variable (reset at dusk)	variable (reset at dusk)
Biomass index specification (range, nature)			Mean of 32 individual pings	Mean of 32 individual pings	Mean of 32 individual pings (double sampling when detection)	Mean of 32 individual pings
Biomass index calculation			tn derived from mean SKJ density	tn derived from mean SKJ density	tn derived from mean SKJ density	tn derived from mean SKJ density
maximum detection threshold			63 by layer	63 by layer	63 by layer	63 by layer
minimum detection threshold			1	1	1	1

3.2 Metadata Description tables

The raw data received by each research center is different (even for the same buoy model). The following metadata description tables report, for each research center, the meaning of each field for the raw positions and acoustic data.

3.2.1. Buoy position raw data

- *Marine Instruments Buoys: Raw data received in AZTI*

In the recuperation of the historic data (2013-2018) a csv. file by ocean, year and company has been received in AZTI and then this information has been integrated in a common database. This data has been sent to AZTI by the buoy providers under specific agreements with data owners. Information about the ownership of the buoy can be obtained from other sources (i.e. since 2015 in the Indian ocean and 2016 in the Atlantic Ocean AZTI is receiving also a monthly csv. file with information on buoy positions for the assessment of the number of operational buoys for the verification of the limitation in the number of buoys in place in each RFMOs which includes in the csv. file the IMO of the vessel in each case).

Fecha	Hora	Numero de Boya	Latitud	Longitud	Velocidad
01/01/2017	23.52	M3+502549	4,95	56,42	0,5
01/01/2017	23.2	M3+502666	-1,33	48,06	1,3
01/01/2017	23.05	M3+502702	5,96	64,61	0,8
01/01/2017	21.42	M3+502739	3,92	49,51	0,3
01/01/2017	23.56	M3+502748	-14,76	44,77	0,6

- *Fecha*: Date of the last position of the day
- *Hora*: Hour (GMT)
- *Numero de Boya*: Identification number of the buoy, given by the model code and number by 5-6 digits.
- *Latitud*: Latitud in decimals of the last position of the day
- *Longitud*: Longitud in decimals of the last position of the day
- *Velocity*: v calculated from the distance/time between the last position of the day and the last position of the previous day.
- *Notas*: Empty column

- *Marine Instruments Buoys: Raw data received in IRD*

Since 2006 and until 2015, the fishing companies delivered to IRD the .csv files containing the buoys' positions data for the French fleet.

The names of the .csv datafiles are specific for each buoy model (e.g. M3I). For the first files delivered, the buoy model was not specified, and the filename was labeled IES. For those buoys, the name of the file was used as a complementary timestamp for the data (see below). Below is the specific format of the data for each buoy model.

- IES (unknown model)

Example of file name: 9125_20100131082146.ies.csv

(format: boatID_YYYYMMDDhhmmss.ies.csv; boatID is unknown; timestamp YYYYMMDDhhmmss refers to the reception datetime of the data).

Column 1	Column 2	Column 3	Column 4	Column5	Column 6	Column 7	Column 8	Column 9	Column 10
$XXBSC	23477	0010.03	S	00431.33	E	28.3	19	31.01.07	00.00.00
$XXBSC	23476	0041.71	S	00421.66	E	28.3	19	31.01.07	00.00.00

- *Column1*: Unknown field
- *Column2* Identification number of the buoy (numeric, 5 digits)
- Column3: Latitude (ddmm.mm dd=degrees, mm.mm=decimal minutes)
- Column4: Latitude (N=Nord, S=South)
- Column5: Longitude (dddmm.mm ddd=degrees, mm.mm=decimal minutes)
- Column6: Longitude (E=East, W=West)
- Column7: Sea Water Temperature (°C)
- Column8: Speed (Sea Speed knot)
- Column9: Date-time stamp (DD.MM.hh). Note that the year and minutes are not specified. The year is recovered from the filename.
- Column10: Unknown field

- M3I and M4I buoy models

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8
7383	0264.85	N	00353.37	W	26.8	03.09.11	07.58.00
7383	0264.85	N	00353.37	W	26.8	03.09.11	08.00.00

- *Column1* Identification number of the buoy (numeric, 5 digits)
- Column2: Latitude (decimal degrees*100)
- Column3: Latitude (N=Nord, S=South)
- Column4: Longitude (decimal degrees*100)
- Column5: Longitude (E=East, W=West)
- Column6: Sea Water Temperature (°C)
- Column9: Date (DD.MM.YY)
- Column10: GMT Time (HH.mm.ss)

- *Satlink Buoys: Raw data received in AZTI*

In the recuperation of the historic data (2010-2018) a csv. file by vessel and month has been received in AZTI and then this information has been integrated in a common database. This data has been sent by the buoy providers under specific agreements with data owners. Since 2015 in the Indian ocean and 2016 in the Atlantic Ocean AZTI is also receiving a monthly csv. file with information on buoy positions for the assessment of the number of operational buoys for the verification of the limitation in the number of buoys in place in each RFMOs which includes in the csv. file the IMO of the vessel in each case.

Fecha	Hora	Número de boya	Latitud	Longitud	Velocidad	Notas
01/12/2017	12.49	DSL+134021	-7,03	-11,43	0,35	
01/12/2017	16.36	ISL+139017	4,27	-38,85	0,26	
01/12/2017	19.38	DSL+145062	5,9	-20,84	0,5	
02/12/2017	12.49	DSL+134021	-7,01	-11,54	0,18	
02/12/2017	16.36	ISL+139017	4,43	-38,78	0,44	
02/12/2017	19.38	DSL+145062	5,68	-20,79	0,53	
03/12/2017	12.49	DSL+134021	-6,95	-11,65	0,1	
03/12/2017	16.36	ISL+139017	4,53	-38,65	0,4	

  - *Fecha*: Date of the last position of the day
  - *Hora*: Hour (GMT)
  - *Número de Boya*: Identification number of the buoy, given by the model code and number by 5-6 digits.
  - *Latitud:* Latitud of the last position of the day (in decimals)
  - *Longitud*: Longitud of the last position of the day (in decimals)
  - *Velocity*: v calculated from the distance/time between the last position of the day and the last position of the previous day.
  - *Notas:* Empty column

- *Zunibal Buoys: Raw data received in AZTI*

In the recuperation of the historic data (2013-2018) a csv. file by company and year has been received in AZTI and then this information has been integrated in a common database. This data has been sent by the buoy providers under specific agreements with data owners. Since 2015 in the Indian Ocean and 2016 in the Atlantic Ocean AZTI is also receiving a monthly csv. file with information on buoy positions for the assessment of the number of operational buoys for the verification of the limitation in the number of buoys in place in each RFMOs which includes in the csv. file the IMO of the vessel in each case.

Fecha	Hora	Numero de boya	Latitud	Longitud	Velocidad
01/01/2017	0.01	T8E008175958	0,7231	69,3392	0
01/01/2017	0.07	T7+007165729	-0,0533	78,2753	0
01/01/2017	0.09	T8E005174138	0,805	69,2903	0
01/01/2017	0.24	T8E005174138	0,8069	69,2942	1,03
01/01/2017	0.39	T8E005174138	0,8089	69,2986	1,19
01/01/2017	0.48	T7+001160308	10,3997	59,9175	0
01/01/2017	0.54	T8E005174138	0,8106	69,3028	1,06

- *Fecha*: Date of the last position of the day
- *Hora:* Hour (GMT)
- *Número de Boya*: Identification number of the buoy, given by the model code and number (9 digits)
- *Latitud:* Latitud of the last position of the day (in decimals)
- *Longitud*: Longitud of the last position of the day (in decimals)
- ***Velocidad:*** v calculated from the distance/time between the last position of the day and the last position of the previous day.

3.2.2. Buoy acoustics raw data

- *Marine Instruments Buoys: Raw data on acoustics received in AZTI.*

In the recuperation of the historic data (2013-2018) a csv. file by ocean, year and company has been received in AZTI and then this information has been integrated in a common database. This data has been sent by the buoy providers under specific agreements with data owners.

TransmissionDate	TransmissionHour	FactoryCode	Latitude	Flash	Longitude	Status	Temperature	Vcc	SounderDate	Gain	Resolution	Levels	DeepLevel	DataText					
01/12/2017	0:17:41	M3+502674	-7,2755	OFF	"43.0028"	SLEEP	30,2	13,6	30/11/2017 21:00	1	4	50	150	7	0	0	0	3	3
01/12/2017	0:17:41	M3+502674	-7,2755	OFF	"43.0028"	SLEEP	30,2	13,6	30/11/2017 21:24	1	4	50	150	7	0	0	1	0	0

- *TransmissionDate* Date of buoy transmission by Iridium (dd-MM-yyyy)
- *TransmissionHour* Time of buoy transmission by Iridium (HH.mm)
- *FactoryCode* Identification number of the buoy, given by the model code and number by 5-6 digits
- *Latitude* Latitude (GPS position)
- *Flash* Flash On/Off
- *Longitude* Longitude (GPS position)
- *Status* Buoys status (Search, recovery, sleep, none)
- *Temperature* Water temperatura (°C). In situ. Sea water temperature at the time of sending the message
- *Vcc* Battery level in volts (from 0 to 17.7)
- *SounderDate* Date and time of echo-sounder measurement (dd/MM/yyyy HH:mm:ss)
- *Gain* Gain associated with the buoy. Fixed Value (0,1,2,3)
- *Resolution* Number of bits used in each layer. Fixed value (3 or 4)
- *Levels* Number of layers. Fixed value (50)
- *DeepLevel* Maximum depth (150m)
- *DataText* Nominal data of the echo-sounder (0-7 or 0-15)

- *Folowing 50 columns*: Intensity of the acoustic signal detected (0-7 or 0-15), the best record sampled (the highest biomass)

- *Marine Instruments Buoys: Raw data on acoustics received in IRD*

Since 2016, IRD received the buoys' acoustic data (which also contain position data, see below). The data was delivered directly by the buoy provider (Marine Instruments), who made a data rescue of the buoy acoustic data transmitted since 2010.

NOMBRE	BARCOS	FECHA	LATITUD	LONGITUD	FLASH	VOLTAJE	TEMPERATURA	VELOCIDAD	RUMBO	ACTIVA	DESACTIVA
M3+514875	MALOYA,MORN SESELWA,MORNE BLANC	02/01/2018 11:19	-1.8217	70.1843	OFF	14.4	30.9	1.4	77	14/12/2017	23/03/2018
M3+514875	MALOYA,MORN SESELWA,MORNE BLANC	02/01/2018 11:19	-1.8217	70.1843	OFF	14.4	30.9	1.4	77	14/12/2017	23/03/2018
M3+514875	MALOYA,MORN SESELWA,MORNE BLANC	02/01/2018 14:18	-1.8017	70.2593	OFF	14.2	30.4	1.6	75	14/12/2017	23/03/2018

VALOR.SONDA	FECHA.SONDA	GANANCIA	PROFUNDIDAD.MÁXIMA	Nº.CAPAS	BITS.POR.CAPA	FRECUENCIA	DISPARO
7.000000	02/01/2018 08:06	1	150	50	4	50	7 0 0 0 0 0 0 0 0 1 4 3 3 3 2 1 0 0 2 0 0 0 0 0 0 0 0 0…
7.000000	02/01/2018 09:42	1	150	50	4	200	7 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0…
7.000000	02/01/2018 10:24	1	150	50	4	50	7 0 0 0 0 0 0 0 0 0 3 4 4 2 4 2 1 0 0 0 0 0 0 0 0 0 0 0…

- *Nombre*: Buoy ID (First three alphanumeric characters: buoy model; remaining 5-6 digits: buoy ID).
- *Barcos*: Names of the vessels associated to the buoy at the time of transmission. These vessels can see the buoy's position/acoustic data on the MSB application onboard.
- *Fecha:* Transmission date and time of the buoy. (Format DD/MM/YYYY HH:mm).
- *Latitud:* Latitude, GPS position during the Transmission date and time in degrees (Format: DD.DDDD)
- *Longitud*: Longitude, GPS position during the Transmission date and time in degrees (Format: DD.DDDD)
- Flash: Flash activation status. Values: ON/OFF
- *Voltaje:* Voltage of the Battery in Volts. Values between 0 and 17.7.
- *Temperatura* : Sea water temperature. In degrees Celsius. From 10 °C up to 61.1 °C.
- *Velocidad*: Speed of the buoy in knots. Ranges between 0 and 99. If not possible to calculate it, field is "--"".
- *Rumbo:* Direction of movement of the buoy (heading). In degrees. Ranges between 10° and 365°.
- *Fecha Activa* Date of Activation of the buoy. Format DD/MM/YYYY.
- *Fecha Desactiva* Date of Deactivation of the buoy. Format DD/MM/YYYY. If null, field is "--".
- *Valor Sonda*: Sensor value Estimated tonnage obtained from the echosounder. Numeric value.
- *Fecha Sonda*: Date and time of the echosounder sample. (Format DD/MM/YYYY HH:mm).
- *Ganancia* : Gain. Fixed values [0|1|2|3]
- *Profundidad Maxima* : Maximum depth in meters. Fixed value: 150.
- *N.° Capas* : Number of layers. Fixed value: 50
- *Bits por Capa* : Bits per layer. Numeric value [3|4]
- Frecuencia: Echosounder Frequency in Hertz. Fixed values [50|120|200]

- *Disparo Tir* : Values of the echosounder measure for each layer. Discrete values between [0-7] or [0-15] depending on the buoy model. Values are separated by a space.

- *Satlink Buoys: Raw data on acoustics received in AZTI*

In the recuperation of the historic data (2010-2018) a csv. file by vessel and month has been received in AZTI and then this information has been integrated in a common database. This data has been sent by the buoy providers under specific agreements with data owners.

Name	OwnerName	MD	StoredTime	Latitude	Longitude	Bat	Temp	Speed	Drift	Layer1	Layer2	Layer3	Layer4	Layer5	Layer6	Layer7	Layer8	Layer9	Layer10	Sum	Max	Mag1	Mag2 (1)	Mag3	Mag4 (2)	Mag5	Mag6 (3)	Mag7	Mag8 (4)
DSL+144577	Albacan	168	01/12/2017 1:26							0	0	0	0	0	0	0	0	0	0			0	4	0	0	0	0	0	0
DSL+144577	Albacan	168	01/12/2017 2:26							0	0	0	0	0	2	1	0	0	0			0	1	0	3	0	1	0	0
DSL+144577	Albacan	168	01/12/2017 3:26							0	0	0	0	0	0	0	0	0	1			0	0	0	1	0	0	0	0
DSL+144577	Albacan	168	02/12/2017 1:30							0	0	0	0	0	0	0	0	0	0			0	4	0	0	0	0	0	0

- *Name*: Name of the buoy given by a unique identifier by 5-6 digits.
- *OwnerName*: Name of the buoy owner assigned to a unique purse seine vessel
- *MD*: Message descriptor (160, 161 and 162 for position data, without sounder data, and 163, 168, 169 and 174 for sounder data)
- *StoredTime*: Date (dd/mm/yyyy) and hour (H:MM) of the echo-sounder record
- *Latitude, Longitude*: Not provided (this information is provided in the position file csv.)
- *Bat*: Not provided. (Charge level (in percentage). Except for the D+ and DS+ in voltaje.)
- *Temp*: Temperature (Not provided)
- *Speed*: Speed in knots (Not provided)
- *Drift*: bearing in degrees (Not provided)
- *Layer1-Layer10*: Depth observation range extends from 3 to 115 m, which is split in ten homogeneous layers, each with a resolution of 11.2 m. The buoy has also a blanking zone (a data exclusion zone to eliminate the near-field effect of the transducer between 0 and 3 m. 32 pings are sent from the transducer and an average of the backscattered acoustic response is computed and stored in the memory of the buoy. Manufacturer´s method converts raw acoustic backscatter into biomass in tons, using a depth layer echo-integration procedure based exclusively on an algorithm based on the TS and weight of skipjack tuna.
- *Sum:* Sum of the biomass estimated at each layer
- *Max*: Maximum biomass estimated at any layer
- *Mag1, Mag3, Mag5 and Mag7 are the magnitudes*: It is the count of detected targets according to the TS of the detection peak.

	Limit (dB)	Limit (Linear, ES12 raw data)	Equivalent fish size
Upper limit (Limit 5)	-20	2270	*[-25,-20dB] : Bigeye or very large Yellowfin or other large species*
Limit 4	-25	719	*[-30,-25dB> : Large Yellowfin*
Limit 3	-30	227	*[-35,-30dB> : Yellowfin < 100cm*
Limit 2	-35	72	*[-40,-35dB> : Yellowfin <60cm and skipjack> 60cm*

- Mag1 is Limit2, Mag3 is Limit3, Mag5 is Limit4 and Mag7 is Limit5. Mag2,4,6 and 8 are not used.

3.3 Position and acoustic Databases Description

The amount of data available is described by ocean providing information about the % of the number of buoys by model and the total number of records of the raw data set (depending on the data use agreements number of buoys by ocean and year are also provided).

3.3.1 Position Databases

3.3.1.1 *Raw position database in AZTI*

Regarding to the number of buoys in the Indian and Atlantic Ocean, information on 326,644 buoys has been recovered from 2010 to 2018 period of the companies in Spanish Tuna purse seiners associations ANABAC and OPAGAC. The database has 43,200,556 lines corresponding to individual positions of the buoys in the Atlantic and Indian Ocean. The percentage of buoys used by model for the Atlantic and Indian Ocean is described in the Table 3, Table 4 and Figure 1. In the 2010-2012, period the information of individual positions for Satlink could be obtained, but information on individual buoy positions of Marine Instruments buoys could not be exported and integrated in the data base, due to a technical limitation in data exportation process. From 2013 to 2018 information on all buoys of Satlink (models: D+, DS+, DSL+, ISL+, ISD+), Marine Instrument (models: MSR, MSI, M3I, M4I and M3+) and Zunibal (models: T07, T7+, T8E, Te7, Z07) could be integrated in the data base. In Satlink case, from 2010 to 2014, D+, DL+ and DS+ were replaced by DSL+, which was the most used buoy to 2015. Since 2015 when the ISL+ buoy was developed DSL+ has been progressively replaced by the ISL+. In the case of the MI buoys, M3I is the most used and since 2016 the use of M3+ is increasing. At the beginning of the series about 50% of the buoys were working with echosounder. This percentage has increased gradually and nowadays all buoys deployed by the fleet have echosounder. Nowadays, buoys used by the fleet work mainly with a unique frequency, but since 2016 the two frequency buoys are also in used (i.e. mainly M3+)

Table 3. Percentage of buoys by model and year in the Atlantic Ocean constituting the raw AZTI´s position database.

model	%.2010	%.2011	%.2012	%.2013	%.2014	%.2015	%.2016	%.2017	%.2018
D+	51,83	29,50	8,66	0,97	0,12	0,08	0,00		
DL+	3,03	22,96	26,18	11,46	3,91	1,01	0,16	0,03	0,00
DS+	45,14	41,87	14,82	2,24	0,34	0,07	0,00		
DSL+		5,67	50,34	51,11	61,09	56,17	36,65	17,85	3,86
ISD+							0,17	0,75	0,79
ISL+					0,55	12,23	18,63	28,56	29,65
M3+							1,77	24,34	42,19
M3I				24,43	27,54	27,60	40,80	27,80	17,48
M4I				0,90	0,52	0,01	0,01		
MSI				8,85	5,92	2,82	1,79	0,39	0,10
MSR				0,01					
SLX									0,65
T07				0,02	0,01	0,01			
T7+							0,00		0,02
T8E							0,01	0,27	5,20
T8X									0,05
Te7				0,01	0,01		0,00		0,00
Z07				0,01					

Table 4. Percentage of buoys by model and year in the Indian Ocean constituting the raw AZTI ´s position database.

model	%.2010	%.2011	%.2012	%.2013	%.2014	%.2015	%.2016	%.2017	%.2018
D+	55,07	31,11	3,52	1,11	0,44	0,04	0,00		
DL+	4,96	25,41	22,17	4,44	1,04	0,20	0,03	0,01	0,00
DS+	39,97	34,52	4,72	1,27	0,68	0,07	0,01		
DSL+		8,95	69,59	43,78	50,69	47,00	19,42	4,48	0,97
ISD+							0,12	0,26	1,12
ISL+					0,29	15,05	41,52	48,75	53,18
M3+							2,86	10,52	15,57
M3I				44,56	45,15	36,01	32,72	33,14	29,00
M4I				0,27	0,20	0,05	0,02		
MSI				0,32	0,06	0,02	0,01	0,01	
MSR				0,02					
T07				2,66	0,86	0,20	0,02	0,00	
T17				0,01	0,01				
T7+						0,11	1,67	0,57	0,04
T8E							0,92	2,16	0,09
Te7					0,44	0,86	0,52	0,09	0,01
Te8						0,36	0,13	0,01	
Z07				1,55	0,16	0,05	0,02	0,00	

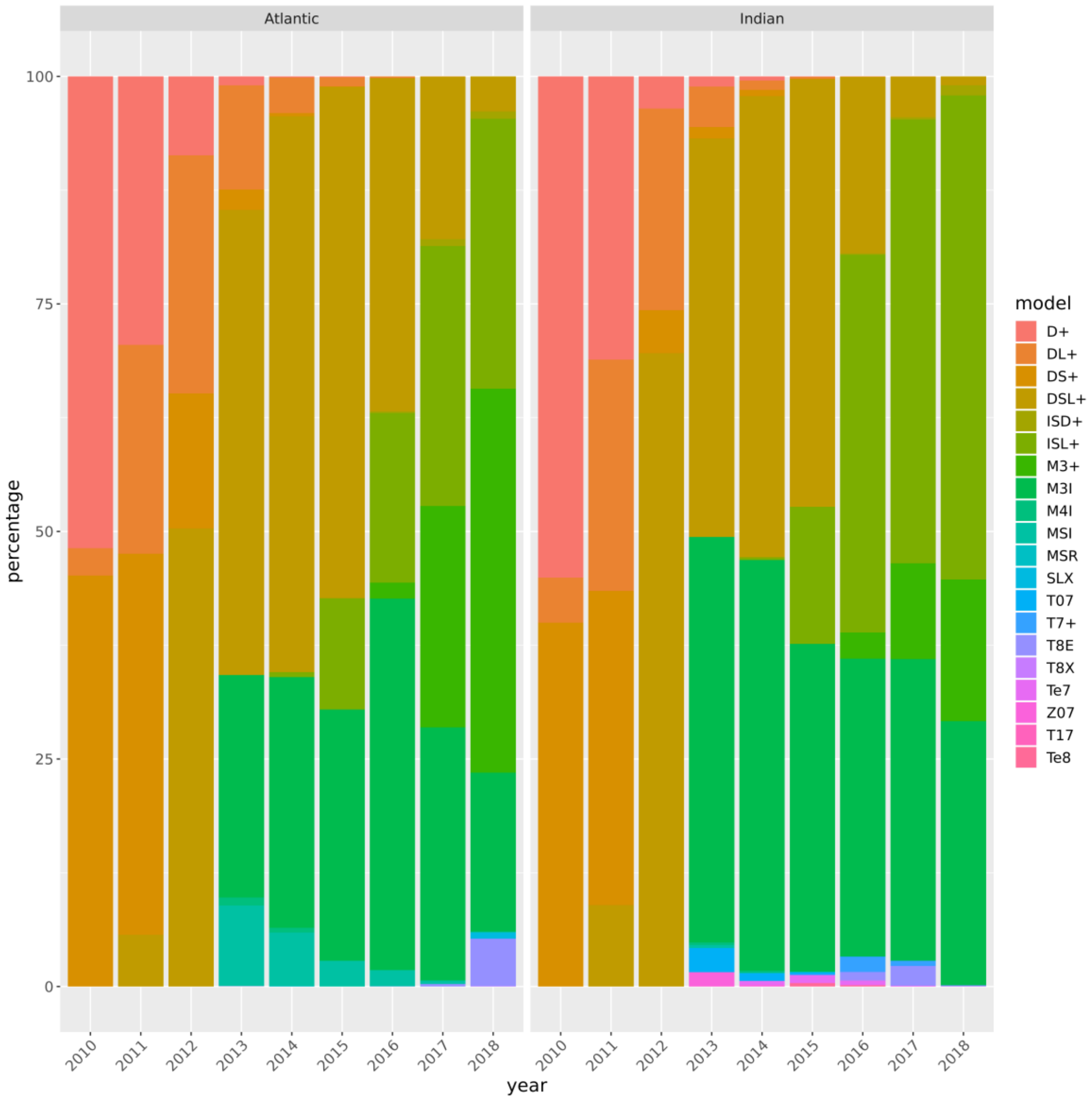


Figure 1. Percentage of buoy by model and year constituting the raw AZTI position database for the Atlantic and Indian Ocean from 2010-2018. Note that for the period 2010 to 2013 Marine Instrument individual buoy positions could not be obtained.

3.3.1.2 *Raw position database in IRD*

The raw position database of IRD contains information on the buoys deployed by the French fleet between 2006 and 2015. This data was provided to IRD directly by the fishing companies (see section 3.2), under a data-exchange agreement signed between IRD and ORTHONGEL. For the following years, the information on the buoy position data was part of the acoustic database (see next section). The total number of lines constituting the raw position database is 6,300,358 (each line corresponds to a buoy position). The number and percentage of buoys by model and year for the Atlantic and Indian Ocean that constitute the database is reported in Tables 5 and 6 and Figures 2 and 3.

Table 5. Number of buoys by model and year constituting the raw IRD position database for the Atlantic ocean.

Atlantic ocean	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
IES	113	276	297	696	1117	749	228	31	0	0
M3I	0	0	0	0	115	653	1044	1046	121	175
M4I	0	0	0	0	0	0	15	245	912	1045
MSI	0	0	0	0	2	258	38	0	0	0
MSR	0	0	0	0	0	0	0	0	0	0
Total	**113**	**276**	**297**	**696**	**1234**	**1660**	**1325**	**1322**	**1033**	**1220**

Table 6. Number of buoys by model and year constituting the raw IRD position database for the Indian ocean.

Indian ocean	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
IES	17	566	1668	1906	2089	1425	472	66	0	0
M3i	0	0	0	0	61	807	1896	1914	2360	4199
M4i	0	0	0	0	0	0	20	366	1616	1377
MSi	0	0	0	0	24	117	420	422	355	57
MSR	0	0	0	0	0	0	0	7	0	0
Total	**17**	**566**	**1668**	**1906**	**2174**	**2349**	**2808**	**2775**	**4331**	**5633**

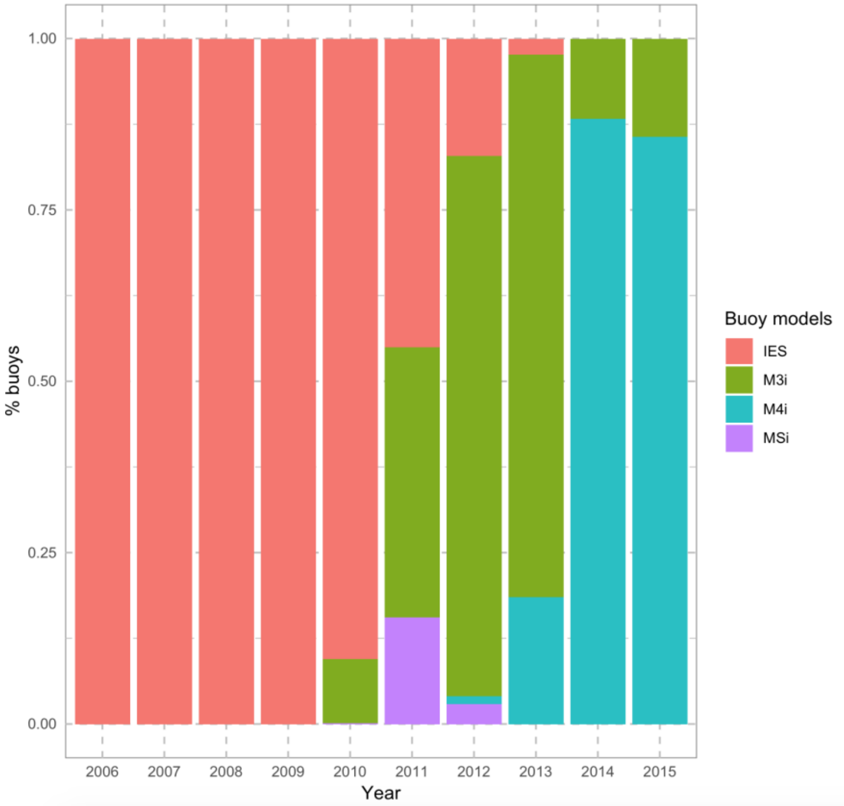


Figure 2. Percentage of buoys by model and year constituting the raw IRD position database for the Atlantic Ocean.

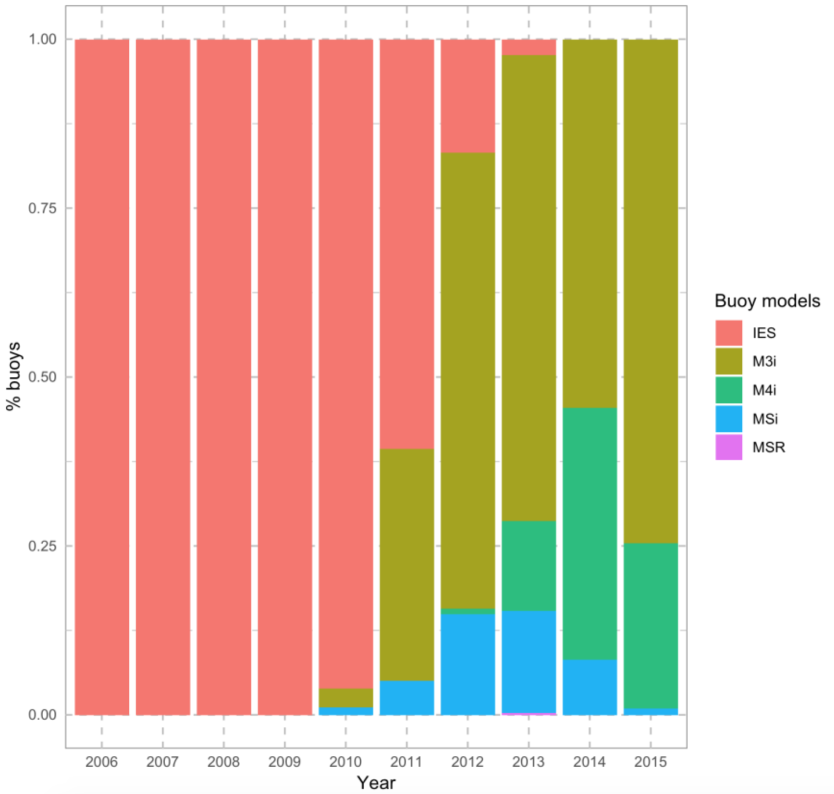


Figure 3. Percentage of buoys by model and year constituting the raw IRD position database for the Indian Ocean.

3.3.2 Acoustic Databases

3.3.2.1 *Raw acoustic database in AZTI*

The raw acoustic database of AZTI contains information on the echosounder 274,208 buoys deployed by vessels in ANABAC and OPAGAC fleet between 2010 and 2018. In the 2010-2012, period the information of individual acoustic records for Satlink could be obtained, but information on individual buoy records of Marine Instruments buoys could not be exported and integrated in the data base, due to a technical limitation in data exportation process. In addition, the recovery of acoustic information of Satlink buoys in the Indian Ocean is on-going. The database contains 185,095,740 lines, each line corresponding to an acoustic record in the Indian and Atlantic Ocean (Table 7 and 8, respectively). This data is provided to AZTI directly by the buoy providers (Marine Instrument and Satlink), under a data-exchange agreement signed between AZTI and data owners. In this data base information of the main buoy brands has been integrated (i.e MI and Satlink).

Table 7. Percentage of buoys by model and year in the Atlantic Ocean constituting the raw AZTI´s acoustic database.

model	%.2010	%.2011	%.2012	%.2013	%.2014	%.2015	%.2016	%.2017	%.2018
DS+	100,00	91,55	21,55	2,66	0,28	0,06			
DSL+		8,45	78,45	74,97	68,49	56,25	38,48	18,13	
ISD+									
ISL+					0,63	12,26	19,48	29,18	
M3+							1,20	24,12	72,65
M3I				21,73	30,08	31,26	40,82	28,57	27,35
M4I				0,65	0,52	0,16	0,01		

Table 8. Percentage of buoys by model and year in the Indian Ocean constituting the raw AZTI´s acoustic database. Note that the recovery of acoustic information of Satlink buoys of 2018 in the Indian Ocean is on-going.

model	%.2010	%.2011	%.2012	%.2013	%.2014	%.2015	%.2016	%.2017	%.2018
DS+	100,00	87,24	5,32	1,43	0,65	0,06	0,01		
DSL+		12,76	94,68	54,44	51,32	45,96	20,06	4,38	0,22
ISD+								0,00	0,11
ISL+					0,29	13,48	43,30	48,56	19,84
M3+							2,80	15,40	37,28
M3I				44,12	47,48	40,40	33,81	31,65	42,54
M4I					0,27	0,10	0,02		

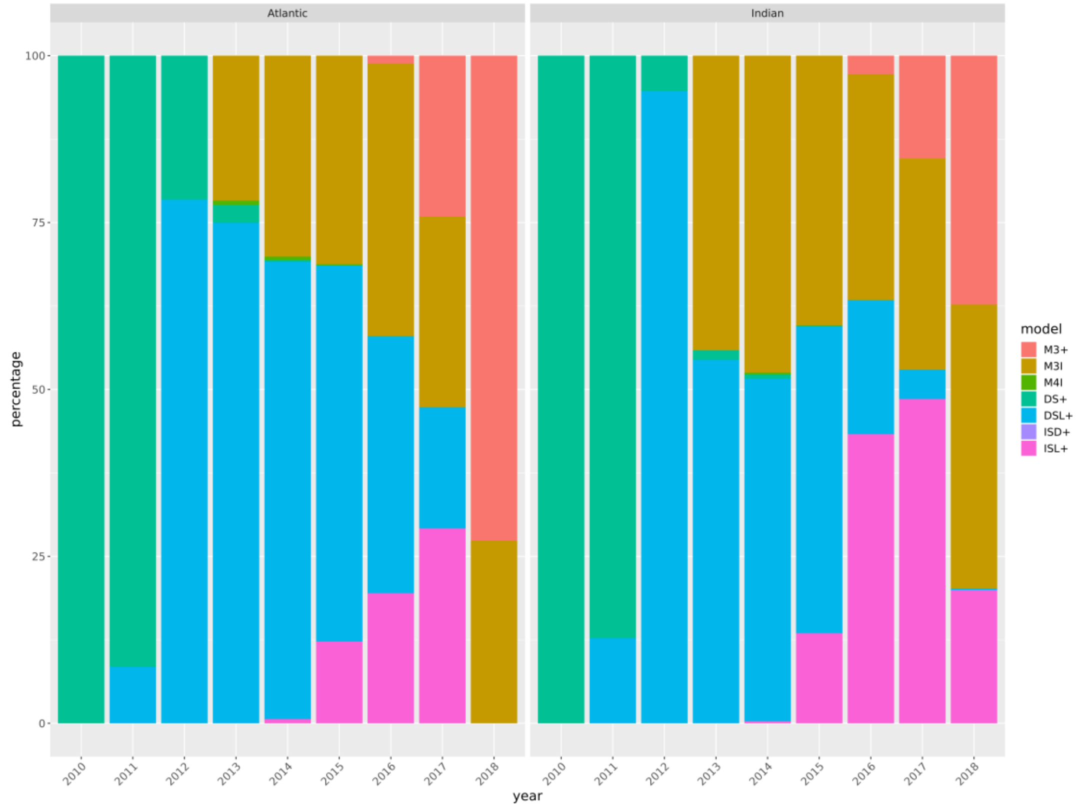


Figure 4. Percentage of buoy by model and year constituting the raw AZTI´s acoustic database for the Atlantic and Indian Ocean from 2010-2018. Note that for the period 2010 to 2012 acoustic information on Marine Instrument could not be obtained. In addition, information on Satlink buoys in the Indian Ocean is in the recovery process.

3.3.2.2 ***Raw acoustic database in IRD***

The raw acoustic database of IRD contains information on the echosounder buoys deployed by the French fleet between 2010 and 2018. This data is provided to IRD directly by the buoy providers (Marine Instruments, see section 3.2), under a data-exchange agreement signed between IRD and ORTHONGEL in 2016. The data correspond to the buoy positions and echosounder data received by the French fleet and are delivered to IRD on a quarterly basis. During the first data delivery in 2016, the buoy provider (i.e. Marine Instruments) made a data rescue of the whole the buoy acoustic data stored on its servers since 2010. Therefore, this database also contains acoustic data during the period 2010-2016 and partly overlaps with the position database for the years before 2016. The total number of lines constituting the raw acoustic database is 152,893,867 (each line corresponds to one acoustic record). The number and percentage of buoys by model and year for the Atlantic and Indian Ocean constituting the database is reported in Tables 9 and 10 and Figures 5 and 6.

Table 9. Number of buoys by model and year constituting the raw IRD acoustic database for the Atlantic ocean.

Atlantic Ocean	2010	2011	2012	2013	2014	2015	2016	2017	2018
M3+	0	0	0	0	0	0	227	2653	4424

M3I	208	569	1202	1220	1019	1215	1613	640	118
M4I	0	0	17	283	1197	1212	915	592	109
MSI	11	0	0	0	0	2	10	9	10
Total	**219**	**569**	**1219**	**1503**	**2216**	**2429**	**2765**	**3894**	**4661**

Table 10. Number of buoys by model and year constituting the raw IRD acoustic database for the Indian ocean.

Indian Ocean	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
M3+	0	0	0	0	0	0	70	2709	3570
M3I	61	1165	2275	2335	3081	6878	10827	11489	11187
M4I	0	0	29	458	1971	1772	818	274	68
MSI	23	117	469	785	424	80	14	7	0
Total	**84**	**1282**	**2773**	**3578**	**5476**	**8730**	**11729**	**14479**	**14825**

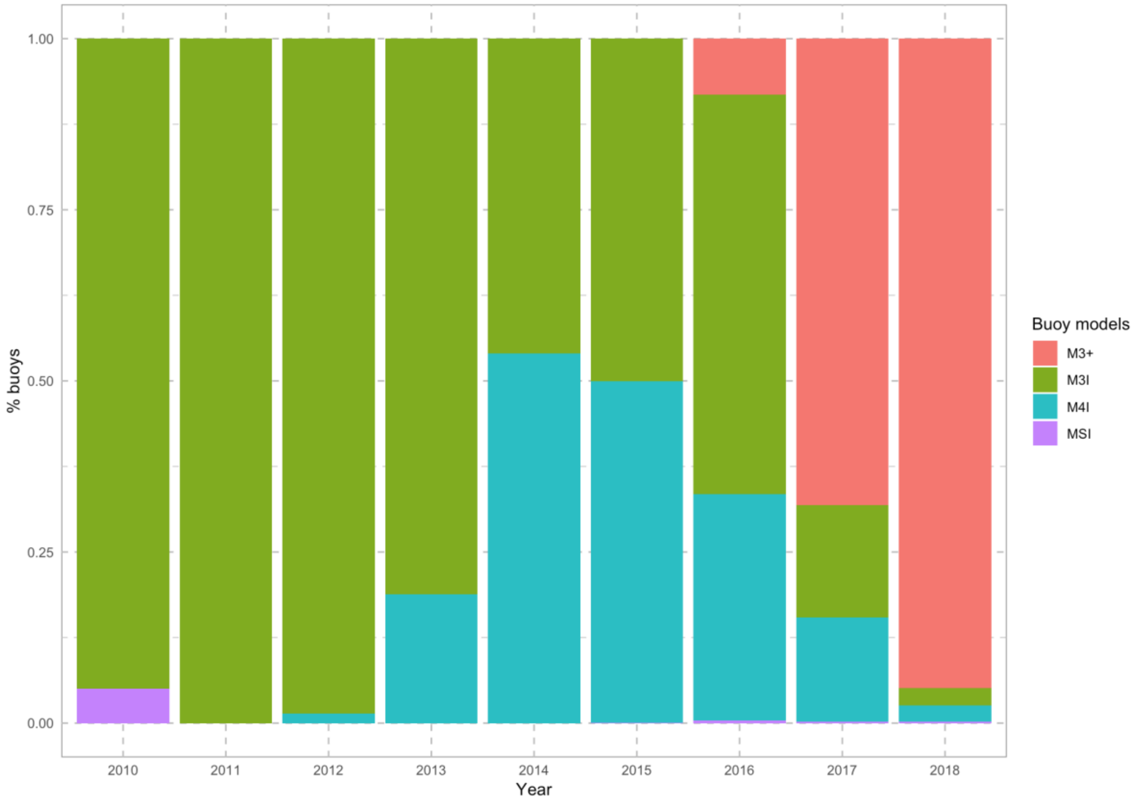


Figure 5. Percentage of buoys by model and year constituting the raw IRD acoustic database for the Atlantic Ocean.

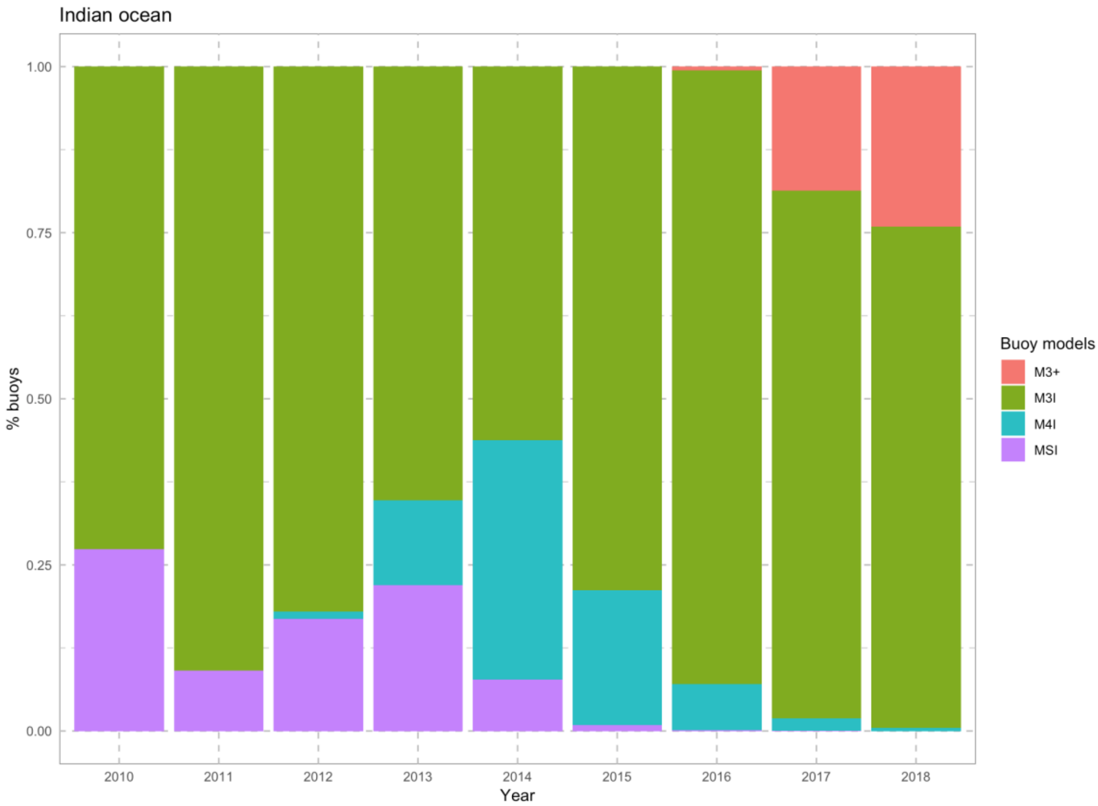


Figure 6. Percentage of buoys by model and year constituting the raw IRD acoustic database for the Indian Ocean.

4. Data processing protocols

In order to compare different methods and agree a common approach for data filtering and estimation of buoy density at sea, a common EU database was created, integrating the position data recorded by 2000 buoys (i.e.,1000 buoys from the Spanish and 1000 buoys from the French fleet for each ocean) during 1 month in the Atlantic and Indian Ocean, respectively. The working scheme adopted to compare the filtering methods is presented in Figure 7.

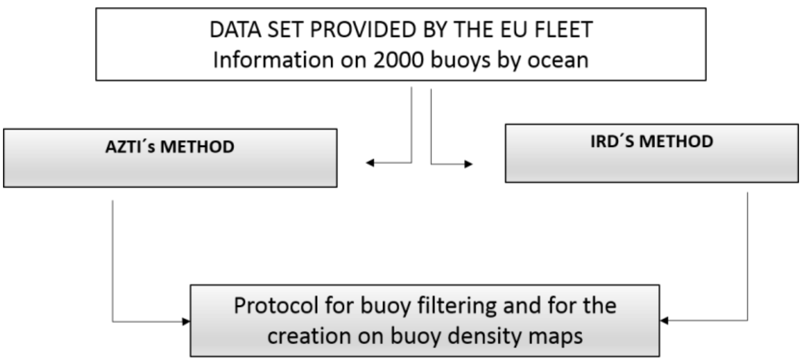


Figure 7. Working scheme to test different algorithms on buoy position data.

4.1 Data filtering criteria

In order to estimate the number of buoys at sea, the raw data need to be pre-processed by filtering erroneous location, data related to failures in satellite communication and location data acquisition; identifying buoys on land positions; and identifying buoys data recording on-board positions. In the table 11 the filters defined for preprocessing the raw data are described (see also Appendix 1 and 2). These filters have been applied in a sequential way.

Table 11. Filters defined for pre-processing raw position data.

FILTER	Description
F1. Isolated	Isolated Position (>48 hours from another position or estimated speed above 35 knots relative to next/previous position)
F2. Duplicated	Duplicated data (all fields are the same)
F3. Land and stationary	Data on land with speed <0.01 knots
F4. Land	Data on land with speed >0.01 knots
F5. Ubiquity	Data entry from the same date/time, different positions
F6. Not classified	Position not on the land and not classified by at sea/on board algorithm
F7. Onboard	Buoys on board
F8. Water	Buoys at sea. Operational buoys: *Active buoy that is transmitting a signal and is drifting in the sea (definition from RECOLAPE)*

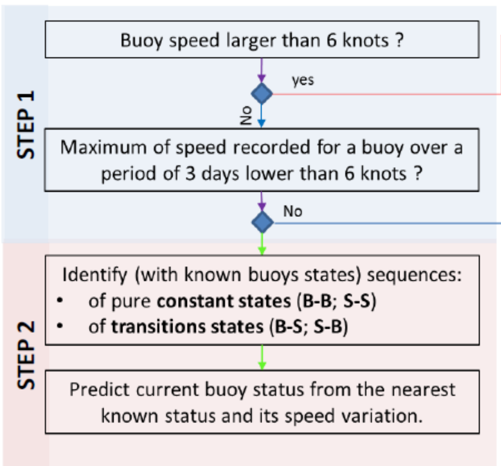
Different algorithms have been developed for the classification of buoys on board. On the one hand IRD has developed the Random Forest algorithm (Maufroy et al. 2015) and the kinetic algorithms (Baidai et al 2017). The random forest algorithm is a machine-learning type of classification method trained on a subset of the buoy positions whose on-board/at sea positions have been previously classified, see Maufroy et al. 2015 for details. The kinetic algorithm is based on the analysis of variations in buoy speed and acceleration along the buoy trajectory. First, the mean speed value (from the position of two subsequent points in the trajectory) and mean acceleration (from the difference in speed between two consecutive points of the trajectory) are estimated. Then, the algorithm applies the two following classification rules (Figure 8):

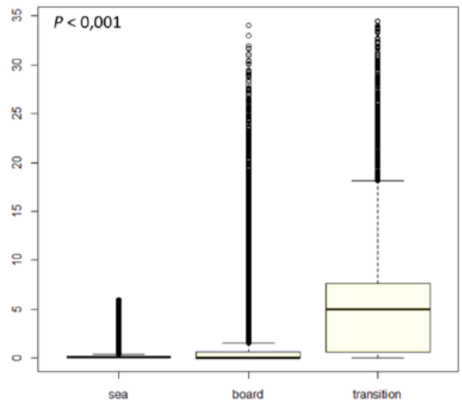
(i) Positions with associated mean speed greater than 6 knots are considered “on-board”.
(ii) Positions with associated mean speeds lower than 6 knots during a continuous period of 3 days (before the position to be assigned) are considered as emitting “at-sea”.

The sequences of the buoy trajectory with no state variation (i.e. portions of the trajectory presenting “on-board – on-board” or “at-sea – at-sea” states only, herein called *constant sequences*) and the sections of buoy trajectory where the buoy state changed from one to another state (i.e. “on-board” – “at-sea” or viceversa, herein called *transition sequences*) are then identified. Finally, the positions not yet classified are assigned to the “on-board” or “at sea” category, by comparing their speed and acceleration relative to those found in the constant and transition sequences (see Figure 8, Appendix 1 and Baidai et al 2017 for further details).

The validation of this classification algorithm was performed by comparing its classification outputs with observer data. Indeed, onboard observers perform buoy data collection (model and serial number of the buoy associated with the FAD) at each operation on a FAD (visit, retrieving or deployment of FADs by the vessel). Based on these data, we could classify "on-board" and "at-sea" states for a subset of the buoys trajectories that were present within our database (around 27000 positions). Comparisons of the two classifications revealed a good similarity between them (more than 90% of similarities between observer-based and kinetic-based approaches).

The kinetic algorithms outputs were also compared to classification results deriving from random forest models developed by Maufroy et al (2015). Running the two approaches on the common EU position database, exhibited a strong percentage of agreement with respectively 94.1 and 89 % of agreement for French and Spanish dataset (in Atlantic and Indian oceans).





Boxplots of the mean acceleration (absolute values) for constant (sea and board) and transition states with p-value at ANOVA tests

Figure 8. Description of the kinetic classification algorithm (left panel) and differences in mean acceleration of constant sequences and transition sequences (right panel).

On the other hand, AZTI has developed an algorithm based also in a random forest classification approach to classify the buoys at sea/onboard using information from another brand of buoys, the Zunibal buoys, which can identify true positions at sea employing a conductivity sensor. The sensor measures the ionic content between two electrodes and determines, through a simple algorithm, whether the buoy is in the water. The predictors variables used in the RF analysis were: distance between two points (km), velocity (km/h), change in velocity (km/h), acceleration (km/h2), azimuth (degree), change in azimuth (degree) and time since the first and last observation of the corresponding buoy trajectory (days). We use these data to train the model, randomly dividing this data set into 80% train and 20% test and executing 5 runs of the classification algorithm with 5-fold cross validation in order to avoid overfitting and to achieve stable results (Kohavi, 1995). The most important variables in the RF model are velocity and variation (Fig. 9).

The average validation indices for sensitivity, specificity, True Skill Statistic (TSS) (Allouche et al., 2006), Kappa (Wood, 2007) and Area Under the Curve (AUC) (Fawcett, 2006) were estimated to evaluate the performance effectiveness and efficiency of the RF classification (Table 12 and Figure 9). These validation indices are calculated using a confusion matrix which evaluates the predictive accuracy of presence-absence models on a set of test data for which the true values are known. The confusion matrix is defined by the true positive rate (TP, presence was correctly predicted by the model), the true negative rate (TN, absence was correctly predicted by the model), the false negative rate (FN, the model incorrectly predicted absence) and the false positive rate (FP, the model incorrectly predicted presence).

	Values
Specificity	0.89
Sensitivity	0.99
Kappa	0.87
AUC	0.94

Table 12. Specificity, sensitivity, AUC and Kappa values for the RF model

The most important variables in the RF model are velocity and variation (Table 12 and Fig. 9).

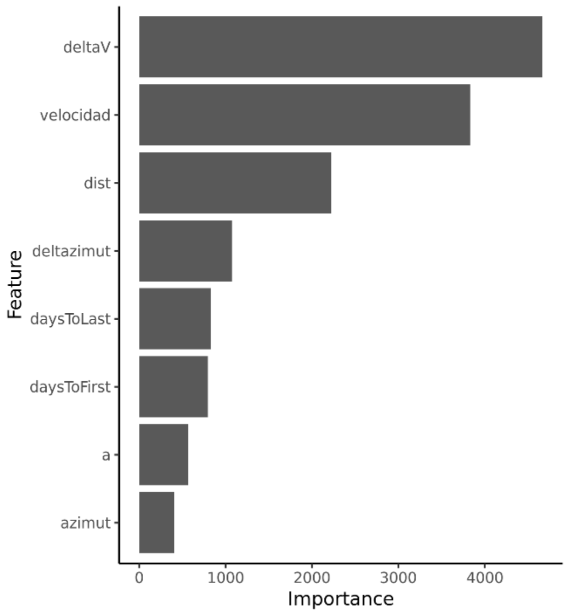


Figure 9. Variable Importance of the Random Forest Model. Name “deltaV” is the change in velocity, “velocidad“ is the velocity, “dist“ is the spatial distance between two points, “deltaazimut“is the change in azimuth, “daysToLast“ is the time since the last observation, “daysToFirst“ is the time since the first observation, “a“ is the acceleration and “azimuth” is the azimuth.

Results on the common dataset

The above-described data filtering procedures were applied on the common dataset composed by 2000 buoy tracks from the EU fleet by ocean. Results for the Atlantic and Indian ocean are presented in Table 13 and Table 14, respectively.

Table 13. Results from the application of the filtering protocol by AZTI and IRD in the atlantic ocean common Data Base

Filters	FR_ATL (AZTI)	% FR_ATL (AZTI)	FR_ATL (IRD)	% FR_ATL (IRD)	ESP_ATL (AZTI)	% ESP_ATL (AZTI)	ESP_ATL (IRD)	% ESP_ATL (IRD)
f.1.Duplicated	47	0.075	47	0.075	0	0	0	0
f.2. Isolated	38	0.06	38	0.06	80	0.316	91	0.36
LAND (f3 and f4)	4915	7.813	4595	7.305	325	1.285	232	0.916
f.5. Ubiquity	11	0.017	11	0.017	0	0	0	0
f.6. Not-Classified	2010	3.195	103	0.164	1924	7.604	1562	6.173
f.7. Onboard (RF_Ocean)	2746	4.366	3473	5.521	122	0.482	4	0.016
Water	53135	84.473	54635	86.857	22853	90.314	23415	92.535
TOTAL	62902	100	62902	100	25304	100	25304	100

Table 14. Results from the application of the filtering protocol by AZTI and IRD in the indian ocean common Data Base

Filters	FR_IND (AZTI)	% FR_IND (AZTI)	FR_IND (IRD)	% FR_IND (IRD)	ESP_IND (AZTI)	% ESP_IND (AZTI)	ESP_IND (IRD)	% ESP_IND (IRD)
F1. Duplicated	94	0.154	94	0.154	0	0	0	0
F2. Isolated	46	0.075	48	0.078	154	0.69	174	0.775
LAND (f3 and f4)	2352	3.844	1574	2.572	333	1.48	138	0.614
F5. Ubiquity	11	0.018	11	0.018	86	0.38	149	0.663
F6. Not-Classified	2076	3.392	165	0.27	1941	8.64	1473	6.558
F7. Onboard (RF_Ocean)	595	0.972	496	0.811	971	4.32	14	0.062
Water	56020	91.545	58806	96.098	18976	84.48	20513	91.327
TOTAL	61194	100	61194	100	22461	100	22461	100

The inspection of the outputs of the filtering algorithm run by IRD and AZTI on the common database demonstrated a high rate of agreement between the two algorithms. The main differences occurred in the land classification. The shapefile resolution could impact the filtering of land positions (low resolution shoreline from GSHHG[4] buffered with 0.05° for IRD) and the land buffer chosen to assign land positions. In addition, minor differences among the two methods occurred in the number of buoy classified as on-board. These differences were higher for the Spanish dataset, since the performances of the algorithms are affected by the resolution of the databases (i.e. one position per day in the Spanish database relative to high-resolution position data in the French database).

The on-board classification algorithms leave a subset of positions unclassified. In the case of the algorithms developed by AZTI it refers to the first position of the track of the buoy. In the case of the kinetic algorithm of IRD unclassified positions are both due to the presence of short trajectories and, for the Spanish dataset, to the low resolution of the data. For these classification algorithms that leave a subset of positions unclassified, it was agreed that the unclassified position should not be eliminated from the dataset and included in the buoy density estimates (i.e., buoys with unclassified positions will be still considered as buoys “at water”).

The final comparisons of the performance of the algorithms for classifying the buoys at water were carried out through the calculation of simple matching coefficient (Sokal and Michener, 1958), estimated from confusion matrices derived from the outputs of the two classification methods. Results are included in Table 15, Table 16, Table 17, and Table 18. Overall, the two methods show high matching coefficients (>94%) in all oceans and datasets. In the Atlantic Ocean, the performances of the classification protocol by IRD and AZTI to classify the buoys at water are >96%. The weaker agreement (94%) is observed in the Indian Ocean in the Spanish data set, possibly due to the characteristics of this data set with shorter tracks and smaller temporal resolution (i.e. a position per day).

	IRD	
AZTI	water	not water
water	24764	13
not water	213	314

Table 15. Confusion matrix on AZTI´s filtering and IRD filtering on the Spanish buoys in Atlantic Ocean. Simple matching coefficient = 0.991;

	IRD	
AZTI	water	not water
water	53735	1457
not water	1061	6649

Table16. Confusion matrix on AZTI´s filtering and IRD filtering on the French buoys in Atlantic Ocean. Simple matching coefficient= 0.96

[4] Wessel, P., and W. H. F. Smith (1996), A global, self-consistent, hierarchical, high-resolution shoreline database, J. Geophys. Res., 101(B4), 8741–8743, doi:10.1029/96JB00104.

	IRD	
AZTI	water	not water
water	20892	25
not water	1245	299

Table 17. Confusion matrix on AZTI´s filtering and IRD filtering on the Spanish buoys in Indian Ocean. Simple matching coefficient= 0.9435

	IRD	
AZTI	water	not water
water	57843	347
not water	1233	1771

Table 18. Confusion matrix on AZTI´s filtering and IRD filtering on the French buoys in Indian Ocean. Simple matching coefficient= 0.9742

4.2 Description of algorithms for density estimates

The criteria adopted for the estimation of the density of buoys on a 1°/month basis are online with the definitions proposed from RFMOs:

ICCAT: *Average number of operational buoys belonging to the vessels over the month (by summing up the total number of operational buoys recorded per day over the entire month and dividing by the total number of days).*

Average of the Number of different unique bouy IDs counted daily within any 1° cell during 1 month, excluding buoys on board, ubiquitous and land_and_stationary.

Given b(i,r) representing the number of different bouys counted during day *i* for 1° grid cell r.
The 1°/month density will be estimated as:

$$D(r) = \frac{\sum_{i=1}^{N} b(i,r)}{N}$$

Where N is the total number of days for a given month.

Based on this definition, the density maps obtained from the common EU database are shown in Fig. 10 and Fig. 11.

A preliminary analysis was also conducted on additional buoy densities indicators, considering the fact that averaging the density of buoys on a monthly basis may hide some short-term patterns in their spatial distribution. For this reason, the maximum and the standard deviation of the number of buoys in 1° cells on a monthly basis have been explored. This preliminary analysis indicates that the standard deviation is not uniform across the 1° cells and can be affected by the current speed, namely higher standard deviations of the buoy densities correspond to regions where the current speed is expected to be higher, see Figures 10.

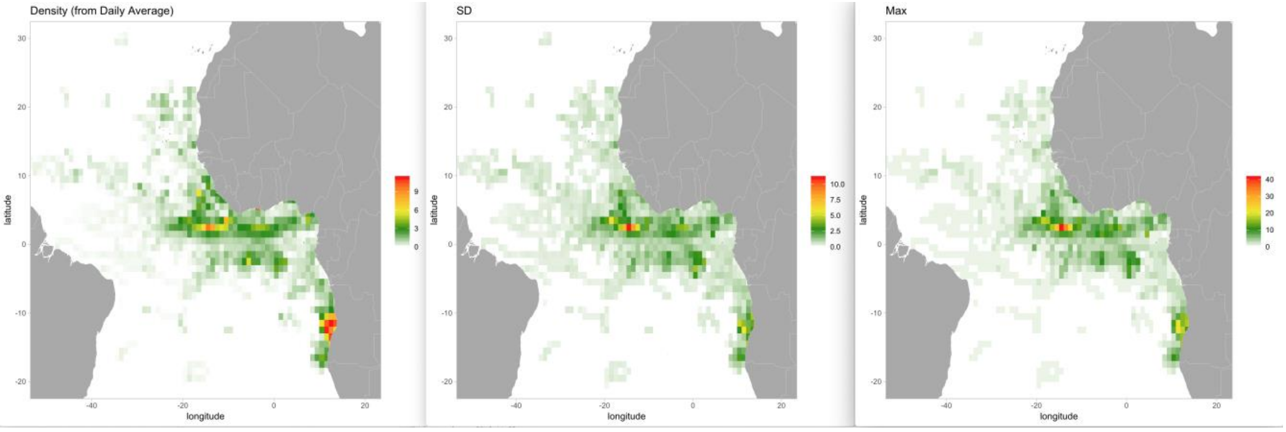


Figure 10: Density maps of the number of buoys on 1° cells/month (left), standard deviation of the number of daily buoys counted on each 1° cells (center) and maximum value the number of daily buoys counted on each 1° cells (right), obtained from the filtered French dataset in the Atlantic Ocean.

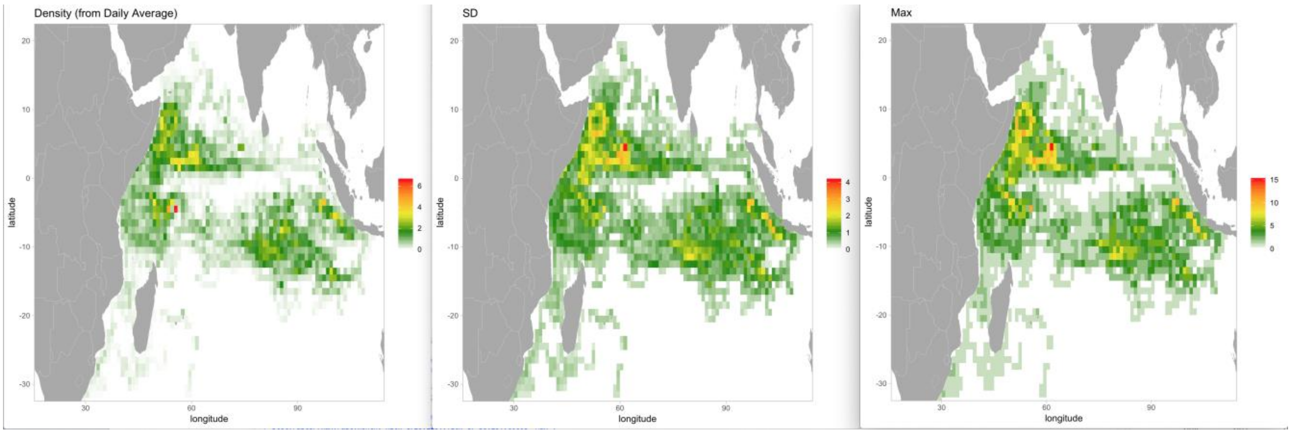


Figure 11: Density maps of the number of buoys on 1° cells/month (left), standard deviation of the number of daily buoys counted on each 1° cells (center) and maximum value the number of daily buoys counted on each 1° cells (right), obtained from the filtered French dataset in the Indian Ocean.

5. Conclusions and Recommendations

The collaborative work conducted by the fishing industry, buoy providers and research institutions has allowed recovering historical information on buoy positions and acoustic data to be used by scientists for developing novel indicators for evaluating and managing tropical tuna stocks. The access to the data has been obtainned thanks to specific agreements with the data owners. IRD has recovered and integrated in the database the data related to all the Marine Instruments buoys used by the ORTHONGEL fleet since 2006 to 2018. In the case of AZTI information of ANABAC and OPAGAC fleeet covering 2010-2018 period has been obtained and integrated in the database. Some buoy providers faced difficulties when exporting historical data, therefore, in the future in order to progress with the recovery of information on buoys, periodical deliveries would be a potential solution.

In this specific exercise, for the analysis of data filtering protocols and the agreement of a common protocol for buoy data pre-processing, a wide set of filters has been defined and test using a common database. Filters run in each research institute were identical except the shapefile for land and onboard filtering, for which a specific algorithm was developed by each institute. The outputs of the filtering algorithms run by each research center show high rates of agreement (>94% agreement on buoys labeled as "in water"), validating both methods for data pre-processing. Minor differences occur on land and on-board positions, for which a specific algorithm has been developed by each research center. These differences were higher for the Spanish dataset in the Indian Ocean, since the performances of the algorithms are affected by the characteristics of the databases (i.e. lower performance on shorter tracks and smaller temporal resolution). In this sense, in order to minimize the misclassification, the use of high-resolution data is recommended if available. In addition, some factors were identified to be valuable to further imporve the filtering and to evaluate the number of buoys followed by each vessel:

a. Water temperature
b. IMO of the vessels receiving the buoy information,
c. Activation and deactivation date
d. Mode of the buoy
e. High-resolution data (all the positions in a day)

Moreover, the addition of an on-board/at sea sensor to the buoys would be a technical improvement that could improve the quality of the data.

References

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 43, 1223–1232.

Baidai Y, Capello M, Amandé J, Billet N, Floch L, Simier M, Sabarros, P and Dagorn L. Towards the derivation of fisheries-independent abundance indices for tropical tunas : progress in the echosounders buoys data analysis. 19th IOTC WPTT, Victoria, Seychelles, 17-22 Oct 2017. IOTC-2017-WPTT19-22

Dagorn. L., Holland, K.N., Restrepo, V., Moreno, G., 2012. Is it good or bad to fish with FADs? What are the real impacts of the use of drifting FADs on pelagic marine ecosystems? Fish and Fisheries 14:391-415.

Maufroy A, Chassot E, Joo R, Kaplan DM (2015) Large-Scale Examination of Spatio-Temporal Patterns of Drifting Fish Aggregating Devices (dFADs) from Tropical Tuna Fisheries of the Indian and Atlantic Oceans. PLoS ONE 10(5): e0128023. https://doi.org/10.1371/journal.pone.0128023

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognit. Lett. 27, 861–874.

Fonteneau, A., Ariz, J., Gaertner, D., Nordstrom, V., Pallares, P., 2000. Observed changes in the species composition of tuna schools in the Gulf of Guinea between 1981 and 1999, in relation with the Fish Aggregating Device fishery. Aquatic Living Resources 13:253-257.

Fonteneau, A., Chassot, E., Bodin, N., 2013. Global spatio-temporal patterns in tropical tuna purse seine fisheries on drifting fish aggregating devices (DFADs): Taking a historical perspective to inform current challenges Aquatic Living Resources 26:37-48.

Gilman, E., B. Bigler, B. Muller, G. Moreno, E. Largacha, M. Hall, F. Poisson, J. Toole, P. He and W.-C. Chiang (2018). "Stakeholder Views on Methods to Identify Ownership and Track the Position of Drifting Fish Aggregating Devices Used by Tuna Purse Seine Fisheries with Reference to FAO's Draft Guidelines on the Marking of Fishing Gear." FAO Fisheries Circular 1163.

Kohavi, R., 1995. A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection. Ijcai, Montreal, Canada, pp. 1137–1145.

Lopez, J., Altamirano, E., Lennert-Cody, C., Maunder, M., Hall, M., 2018. Review of IATTC resolutions C-16-01 and C-17-02: available information, data gaps, and potential improvements for monitoring the fad fishery. 3rd Meeting of the Ad Hoc Working Group on FADs La Jolla, California USA, 11-12 May 2018

Orue, B., Lopez, J., Moreno, G., Santiago, J., Boyra, G., Uranga, J., Murua, H., 2019. From fisheries to scientific data: A protocol to process information from fishers' echo-sounder buoys. Fisheries Research 215 (2019) 38–43

Ramos, Mª.L., Báez, J.C., Grande, M., Herrera, M.A., López, J., Justel, A., Pascual, P.J., Soto, M., Murua, H., Muniategi, A., Abascal, F.J., 2017. Spanish FADs logbook: solving past issues, responding to new global requirements. Joint t-RFMO FAD Working Group meeting Doc. No. j-FAD_11

Scott, G.P., Lopez, J., 2014. The use of FADs in tuna fisheries. European Parliament Policy Department B: Structural and Cohesion Policies: Fisheries IP/B/PECH/IC/2013–123: p 70,

Sokal, R.R. and Michener, C.D. I958. A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, 38: 2.

Torres-Irineo, E., Gaertner, D., Chassot, E., Dreyfus-León, M. (2014) Changes in fishing power and fishing strategies driven by new technologies: The case of tropical tuna purse seiners in the eastern Atlantic Ocean. Fisheries Research, 155: 10-19.

Wood, J.M., 2007. Understanding and Computing Cohen's Kappa: A Tutorial. WebPsychEmpiricist. Web Journal Available from:. http://wpe.info/vault/wood07/Wood07.pdf.

Annex 1

RECOLAPE WORKSHOP

DAY 1 : Data description and Position Filtering (WP4.3)

- Metadata and Data description tables
- Review positions filtering criteria
- Outputs on filtering for positions on common Data set and further test (if required)
- Protocol for buoy density estimates

Deliverables: Metadata & Data description tables, List of position filtering criteria, Outputs on common dataset, protocol for density estimates.

DAY 2 : Acoustic Data analysis I (WP4.2)

- Acoustic data filtering criteria
- Algorithms for converting acoustic data into biomass
- Definition of common indicators of uncertainty on biomass estimates.

Deliverables: List of acoustic data filtering criteria, Table Synthetising the Algorithms characteristics, Table on indicators of uncertainty

DAY 3 : Acoustic data analysis II (WP4.2 & BIOFAD)

- Application of the algorithms for converting acoustic data into biomass on the BIOFAD echosounder data
- Derivation of indicators of uncertainty of biomass estimates on BIOFAD echosounder data

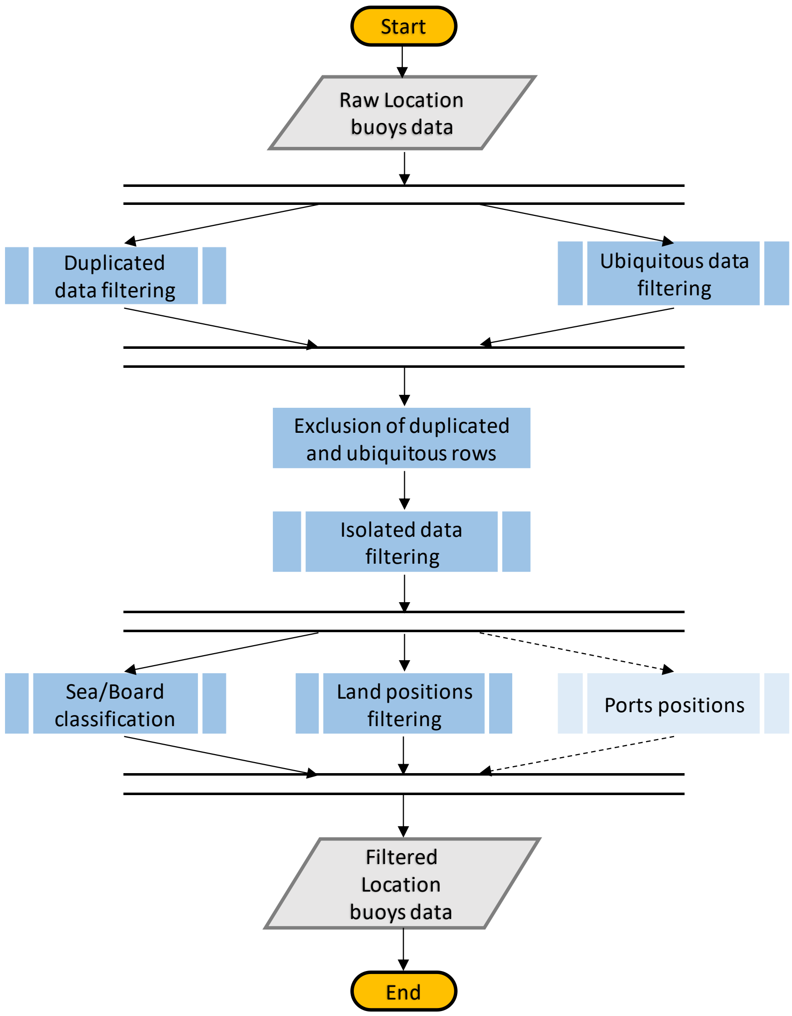
Appendix 1 IRD's detailed database filtering process

1. Flowchart of filtering processing

The filtering process of raw location buoys data consists of 6 main filtering steps and one optional step. Each step adds a new logical column to the data indicating whether or not a row meets the filter criteria. The process ends with a summary column that labels each row with its corresponding status from the filtering operations. Thus, the integrity of raw data is maintained at the end of the filtering process, but all lines are tagged with a label indicating its status at the end of all filtering steps.

The first filtering operations are applied independently of each other. They identify errors related to GPS failures or that may can occur during data compilation (ubiquitous and duplicated data). The concerned data are then labelled and excluded from the rest of the processing operation. The next step detects isolated points in a buoy trajectory, and prepare buoy trajectory for sea/board classification. This step distinguishes buoys at sea from buoys on board a ship, the former being referred to as "water" and the latter as "board» in the final summary column fields. Sea/board classification is applied independently of land positions detection. The figure 1 shows the general flow chart of filtering process. Each filtering steps is further described in detail in the next section of this document.

Figure 1 : General flowchart of the filtering process



1. Duplicated data filtering

Duplicated positions refers to data rows with similar buoy code, time stamp, and positions. Indeed, sometimes due to GPS failures, some lines may have exactly the same value for these parameters but with different speed values (see table 1 for an example on French data in Atlantic Ocean where duplicate lines have *NA* values as speed, and figure 1 for the flowchart).

Tableau 1: Example of duplicated lines for the french dataset in Atlantic ocean

buoy_id	position_date	latitude	longitude	speed
M3I215318	07/11/2016 06:25	-10,4442	10,5058	0,5
M3I215318	07/11/2016 06:25	-10,4442	10,5058	NA
M3I215404	07/11/2016 01:12	-14,182	11,8798	0,3
M3I215404	07/11/2016 01:12	-14,182	11,8798	NA
M3I215410	07/11/2016 02:18	-10,7605	11,446	0,4
M3I215410	07/11/2016 02:18	-10,7605	11,446	NA

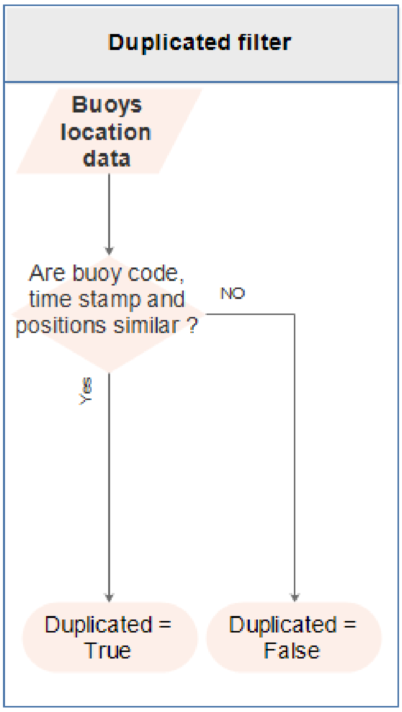


Figure 2: Flowchart of duplicated data filtering

2. Ubiquitous data filtering

Ubiquitous data are referred to as buoys with multiple positions at the same time. Thus rows with similar buoy code and date, but differing in positions (figure 2). A binary choice is made on ubiquitous data. If the two positions are more than 1 km away, the first one is kept and the second is labelled as ubiquitous in the database. If not, the two data are labelled as ubiquitous (see table 2 for ubiquitous data example).

Table 2: Example of duplicated lines for the french dataset in Atlantic ocean

buoy_id	position_date	latitude	longitude	speed
M3+501536	08/11/2016 10:41	1,4133	-18,624	10,5
M3+501536	08/11/2016 10:41	1,4132	-18,6242	1,3
M3I209383	08/11/2016 18:25	4,614	-14,1385	0,2
M3I209383	08/11/2016 18:25	4,614	-14,1383	1,4
M3+501900	12/11/2016 02:04	-15,2997	11,1357	1
M3+501900	12/11/2016 02:04	-15,2995	11,1357	1,5

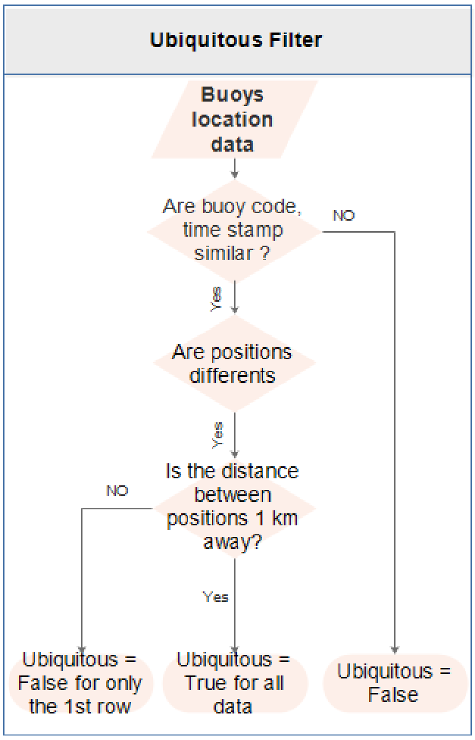


Figure 3 : Flowchart of ubiquitous data filtering

3. Isolated positions filtering

Points isolated for more than 48 hours from the other points on a buoy trajectory are labelled as isolated points. Since the filtering of isolated positions begins to address the trajectory of buoys, this filtering step is applied by excluding ubiquitous and duplicated data. We also consider as isolated points a point separated from the others by an abnormally high speed (35 knots), probably due to GPS failures (Figure 4, 5 and 6).

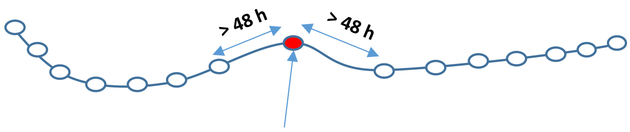


Figure 4: Concept of isolated points (temporal criteria)

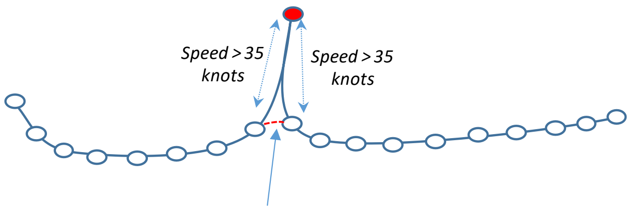


Figure 5: Concept of isolated points (speed criteria)

Table 3 : Examples of isolated points (highligthed in orange and yellow respectively for speed and time criterias)

buoy_id	position_date	latitude	longitude	speed
M3+501536	12/11/2016 10:41	0,8512	-19,5393	0,9
M3+501536	12/11/2016 13:40	0,8245	-19,5782	1
M3+501536	12/11/2016 16:40	0,8013	-19,6217	1
M3+501536	12/11/2016 19:40	1,9843	-23,6	83,2
M3+501536	12/11/2016 22:40	0,7483	-19,6955	81,9
M3+501536	13/11/2016 01:39	0,7208	-19,733	0,9
M3+501536	13/11/2016 04:40	0,6997	-19,7787	1
M3I197959	13/10/2016 06:10	-76,0073	100,6282	0
M3I197959	13/10/2016 18:09	-76,003	100,627	0
M3I197959	14/10/2016 06:11	-76,003	100,6272	0
M3I197959	14/10/2016 18:09	-75,291	100,6027	3,6
M3I197959	17/10/2016 06:09	-87,1322	100,3573	59,3
M3I197959	20/10/2016 13:42	-41,2862	101,0605	NA
M3I197959	20/10/2016 16:46	-42,4798	100,8678	23,6
M3I197959	20/10/2016 19:49	-42,4577	100,8453	0,5

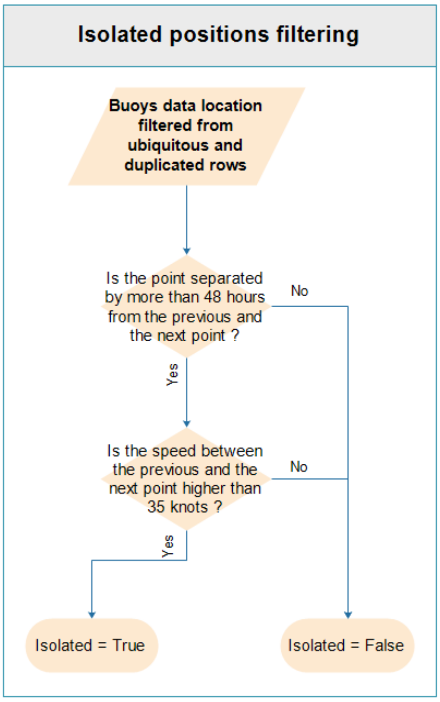


Figure 6 : Flowchart of isolated positions filtering

4. Land and stationary buoys

Land and stationary buoys are defined as buoys on land with a speed lower than a certain thresholds (0.01 knots) during a time window of 24 hours. A buffer zone whose size is defined by a sensitivity analysis (see appendix) is applied around the shoreline.

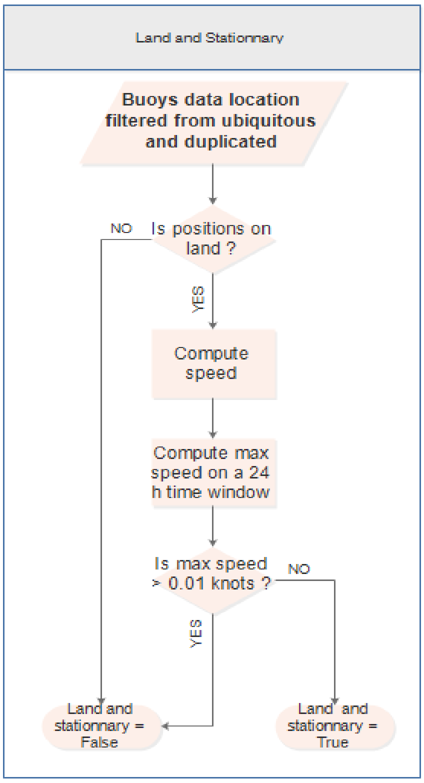


Figure 7 : Flowchart of land and stationary positions filtering

5. Sea/Board classification

The sea/Board classification relies on a simple rule based algorithm that uses the buoy speed and its variations as main classifiers. It consists of two distinct steps. The first one carries out a preliminary classification based on two assumptions:

- a high speed buoy (more than 6 knots) is necessarily on a boat.
- active vessels never drift for very long. Thus a buoy which keep a speed lower than 6 knots for a 3-day time window, is classified as buoys actually at water

This first step provided a set of points for which speed variations can be described in different sequences of states:

- Constant (Sea-Sea or Board – Board)
- Transitions (Sea-Board or Board-Sea)

Then, the analysis of variations in speed between two consecutive points on the same trajectory, makes it possible to classify the remaining points, starting from any point already classified on the trajectory; by determining for each point the type of sequences that corresponds to its speed variation with the following point.

Appendix 2 AZTI's detailed database filtering process

Flowchart of filtering processing

The filtering process of raw location buoys data consists of 6 main filtering steps. Each step reduces the initial database one after another. At each step a new logical column is added to the table indicating whether or not a row meets the filter criteria. The process ends with a summary column that labels each row with its corresponding status from the filtering operations. Thus, the integrity of raw data is maintained at the end of the filtering process, but all lines are tagged with a label indicating its status at the end of all filtering steps. Initial database is reduced. The processing order is shown in the figure 1.

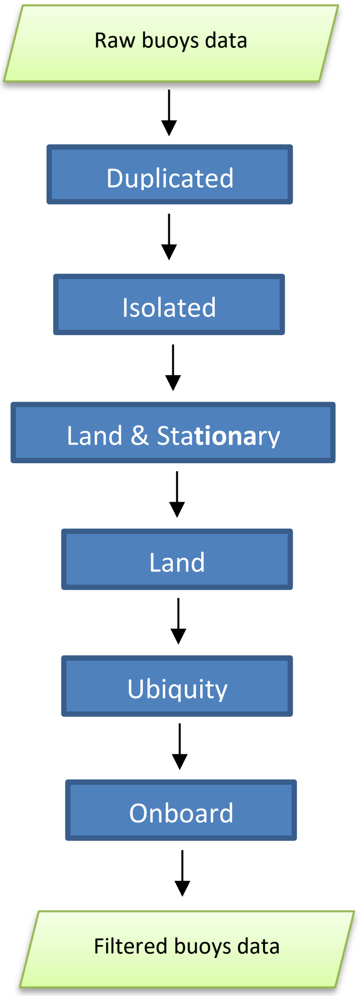


Figure 8 : General flowchart of the filtering process

Each filtering step is further described below:

1. Duplicated data filtering

Duplicated positions refer to data rows with similar buoy code, time stamp, and positions. Indeed, sometimes due to GPS failures or double transmission, some lines may have exactly the same value for these parameters so one row is removed.

2. Ubiquitous data filtering

Ubiquitous data are referred to as buoys with multiple positions at the same time. Thus, rows with similar buoy code and date, but differing in positions are removed.

3. Isolated positions filtering

Points isolated for more than 48 hours from the other points on a buoy trajectory are labelled as isolated points. We also consider as isolated points a point separated from the others by an abnormally high speed (35 knots).

4. Land and stationary buoys

Land and stationary buoys are defined as buoys on land with a speed lower than a certain threshold during a time window of 24 hours. A buffer zone whose size is defined by a sensitivity analysis is applied around the shoreline.

5. Land buoys

Land and stationary buoys are defined as buoys on land during a time window of 24 hours.

6. Sea/Board classification

By the classification model built with the methodology explained at the section 4.1 of this report, values classified as “onboard” are removed from the database.

APPENDIX 3

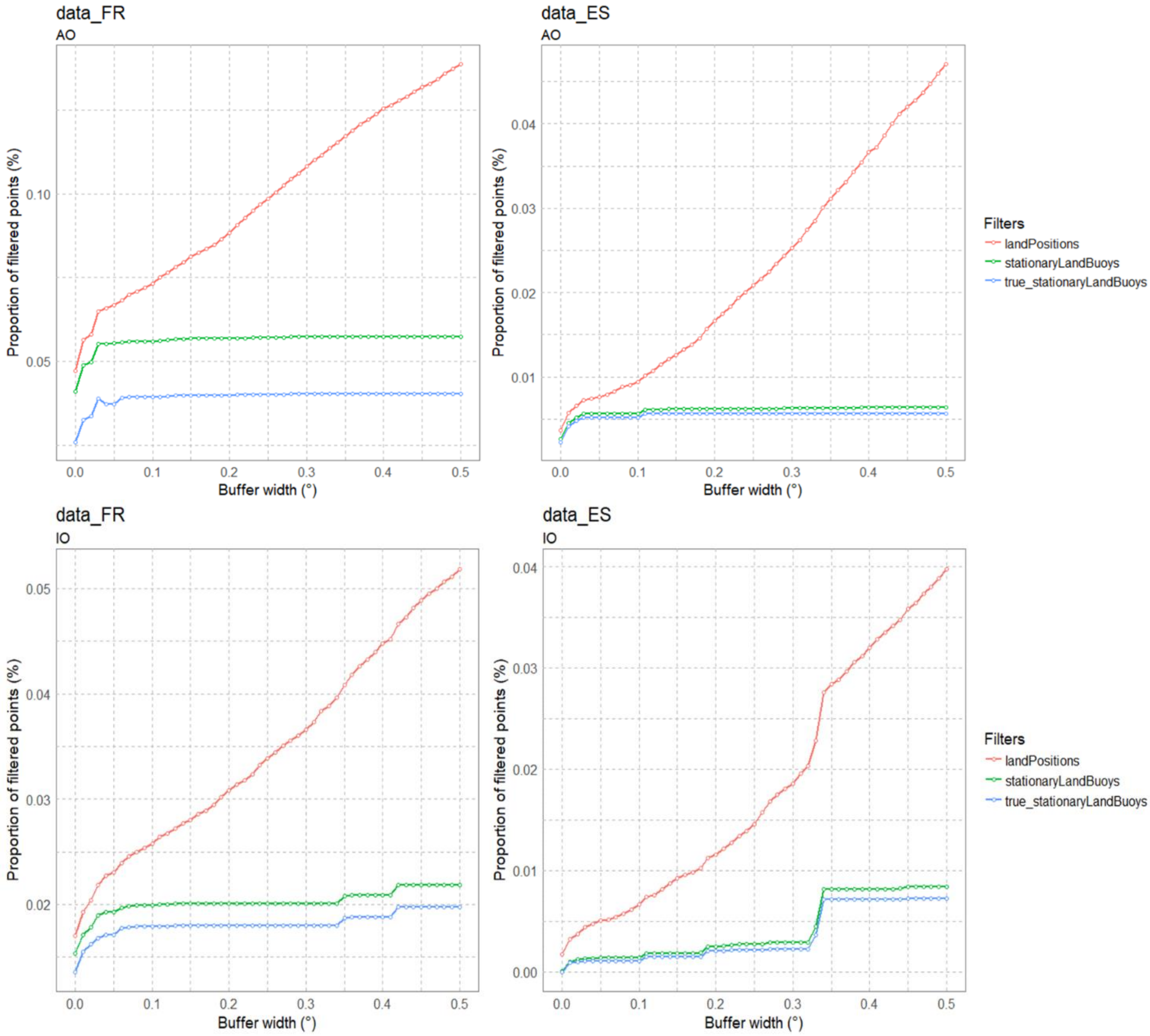


Figure 1: Sensitivity analysis of proportion of filtered points as function of shoreline buffer width for French (data_FR) and spanish data (data_ES), in Atlantic (AO) and Indian (IO) ocean. "landPositions": buoys on land; "stationaryLandBuoys" : buoys within the buffer with a null speed; "true_stationaryLandBuoys" : buoys within the buffer zone that keep a null speed for at least 24 hours.

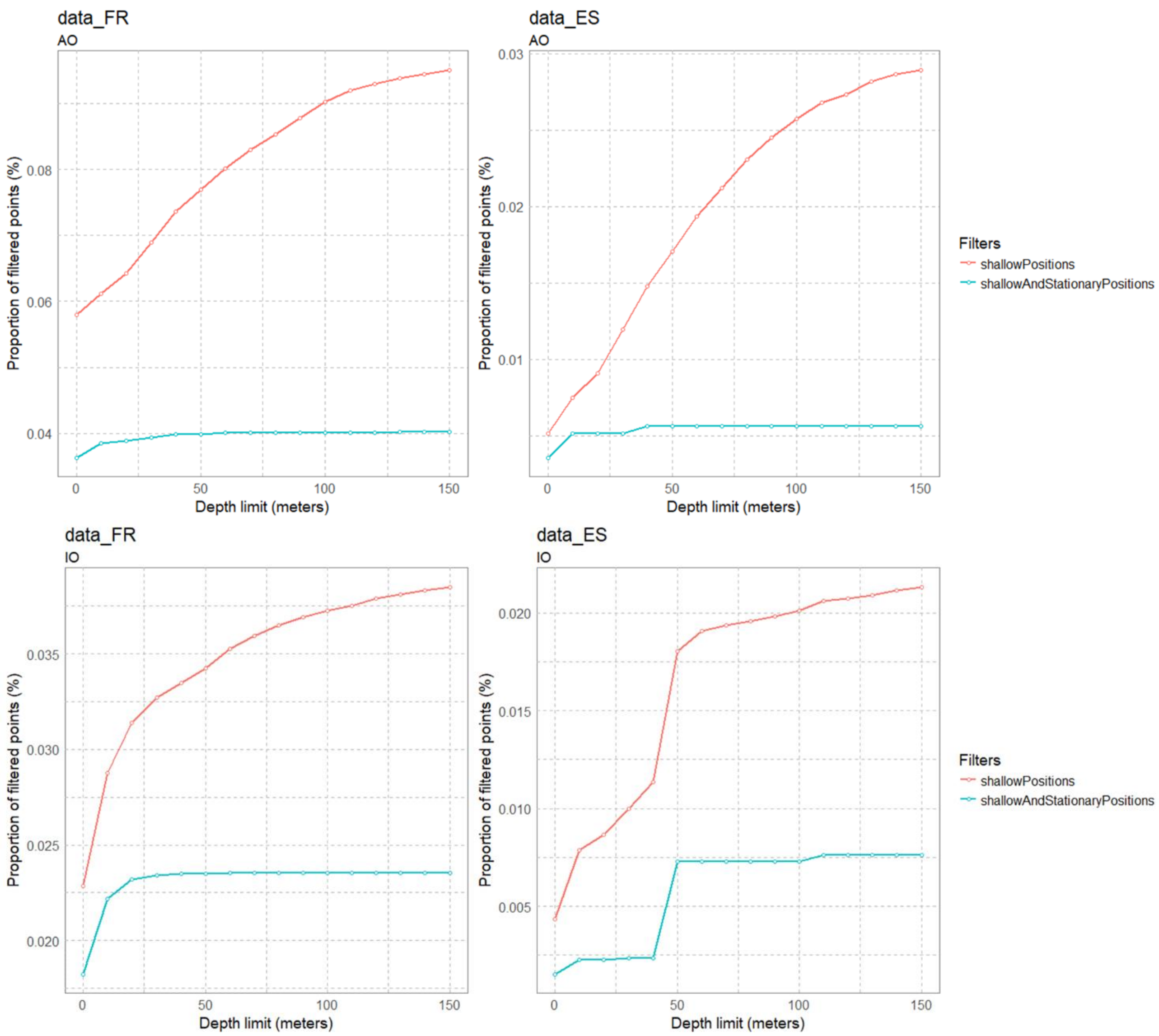


Figure 2: Sensitivity analysis of proportion of filtered points as function of bathymetric threshold for French (data_FR) and spanish data (data_ES), in Atlantic (AO) and Indian (IO) ocean. “landPositions”: buoys on land; “stationaryLandBuoys”: buoys within the buffer with a null speed; “true_stationaryLandBuoys”: buoys within the buffer zone that keep a null speed for at least 24 hours.

FRAMEWORK CONTRACT - MARE/2016/22 « Strengthening regional cooperation in the area of fisheries data collection » Annex III « Biological data collection for fisheries on highly migratory species »

Strengthening Regional cooperation in the area of large pelagic fisheries data collection (Acronym: RECOLAPE)

Task 5.1, National data quality assessment & task 5.2, Regional data quality

D.5.1.1 & 5.2.1 – National and regional data quality improvement

Responsible: Mathieu Depetris (IRD)

Index

Executive summary 3
1. Package dqassess: procedures to assess the quality of large pelagic data 4
  1.1. Background and context 4
  1.2. Aim of this package 4
  1.3. Package framework 4
2. Definition data format, description and utilization 6
  2.1. Using a definition data format 6
  2.2. Definition data format architecture 6
    2.2.1. Sheet "format_infos" 6
    2.2.2. Sheet "slot" 7
    2.2.3. Sheet "slots_hierarchy" 8
    2.2.4. Sheet slot definition 9
    2.2.5. Categories definitions 11
    2.2.6. Codelist category 11
    2.2.7. Numeric category 12
    2.2.8. Logical category 12
    2.2.9. Text category 13
    2.2.10. Date category 13
3. Package methodology 15
  3.1. Global methodology 15
  3.2. Library installation 15
  3.3. Fictive dataset 16
  3.4. Checking data 16
    3.4.1. First example: data on excel file 17
    3.4.2. Second example: data on csv file 19

Executive summary

This document is in relation to the work package 5 of the project MARE/2016/22: "Strengthening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species" (acronym RECOLAPE). The main objective of this package concerns the development of data quality assessment procedures.

To provide answers, a package was designed in R language. Source files of it were stored on a public repository located on GitHub. This package proposes several functions for process checks and controls under data and produce reports associated. Furthermore, a system for tracking bugs or feature requests is available through the following link.

The following document is a compilation of vignettes documentation of the package. To be sure, to read the last version of it, it is advisable to read it directly from the package documentation. To do that, run the following lines in R console (with an internet connection):

```
# Devtools is a necessary package
# If it is not installed, run the following line
install.packages("devtools")
# Load the package from the Git
devtools::install_github("https://github.com/OB7-IRD/dqassess.git", build_opts = c("--no-re
save-data", "--no-manual"))
# Load the library
library(dqassess)
# You can access the package documentation with the following line
?dqassess
```

1. Package dqassess: procedures to assess the quality of large pelagic data

1.1. Background and context

This package was developed under the cover of the European project MARE/2016/22 « Strengthening regional cooperation in the area of fisheries data collection » Annex III « Biological data collection for fisheries on highly migratory species » (acronym: RECOLAPE). He answered the aims of the two first task of the Work Package 5: propose procedures for data quality assessment, at national and regional levels, in the area of large pelagic fisheries data collection.

1.2. Aim of this package

The main objective of this package concerns the development of data quality assessment procedures. Within the framework of the regional cooperation, it's crucial to ensure that the data transmitted to the end users have undergone common quality assessment procedures.

As a minimum, we should check the data quality at two important steps of the data flow:

- At the national scale, before the transmission to the regional coordination group,
- At the regional scale, after building the regional dataset and before the transmission to end users.

1.3. Package framework

This package was designed with a simple thought: quality and associated processes have to be in constant improvement!

Indeed, the structure package was developed to maximize compatibility with future updated and addition of specific controls and tests. Furthermore, it must be seen as a bridge to systems more transversal and interconnected, like the new RDBES (Regional Data Base and Estimation System), under development by the International Council for the Exploration of the Sea (ICES). Otherwise, several other projects, like the WP6 of the FishPi2 project or the future evolution of the COST R package, have common guidelines and goals. These projects were not available yet, but it could be very interesting to working with and develop common processes.

For now, the package source was stored on a public repository located on GitHub. This is web-based hosting service for version control system or tracking changes. In addition to storage and tracking changes, this platform allowed a system to track "to do", bugs or feature requests. For that, you need to use “Issues” menu through the following link. All contributions and feedback experiences increase the relevance and the robustness of this package and serve the interest the large pelagic community people.

2. Definition data format, description and utilization

2.1. Using a definition data format

All controls, and more generally verification process, are focusing around a definition data format. This element contains metadata and all the information necessary to define your data. For example, this definition format was able to answers to the kinds of questions:

- What is the structure of my data? Do I have several tables?
- I there are, what are the relations between my data?
- What is the type of my data? numeric, codelist, free text?
- ...

To explain how we can make and read this definition data format, we will use as an example, the data from the RECOLAPE data call. The characteristics of the data collected were described in the following document.

First, we will use a template for creating an empty definition data format. This template, provide the squeleton or the initial architecture of the format definition. To have it, launch the code below:

```
format_db_empty <- build_template_format_db(format_name = "name_format")
```

This function create an R list. If you want to export it (in xlsx or xls format), you could use the function below:

```
# Don't forget to check function documentation with ?write_format_db_excel or help(write_form
at_db_excel)
write_format_db_excel(format_db = format_db_empty)
```

You can see an example of the empty template for the definition data format here and the example of the definition data format for the RECOLAPE data call here.

The last basic command to know is if you want to read an existing definition data format:

```
format_db <- read_format_db(input_file_path = "path_definition_data_format")
```

2.2. Definition data format architecture

In this section, we will explain the different sheets of the definition data format, though the empty template and the format definition for the RECOLAPE data call.

2.2.1. Sheet “format_infos”

The first sheet, “format_infos”, contains the name of the definition data format (column “format_name”) and the indication of the format’s version (column “format_version”).

Even if the column "format_name" does not need a specific explanation, this parameter is the opportunity to introduce the notion of code style or notation style. Good writing style is like using correct punctuation: you can manage without it, but it sure makes things easier to read and especially avoid conflicts. Several rules and framework exist for maximization compatibility of your text. Here, we are only speaking about the most common advice:

- Use significant words: use only short words, not sentence, which is representative of your global idea. For example, a good format name should inform the reader of the field of the definition data format without having to read it.
- Use short sentences: shorter it will be easier to understand you text will be!
- Avoid specific characters (, ; . : ! ? % & ( ) # / * etc.), diacritics (accent, cedilla, etc.) and empty spaces (replace it by _). These kinds of elements should bring incompatibility and crash functions or code launch.

Dealing with the column "format_version", there's no imposed rule, just like the writing style, but here we decide to use specific semantic versioning: <major>.<minor>.<patch>. The first number (<major>) is related to a major updated. It's when you have added many new features or conceptual changes impacted whose directly to the user interface (typically the new user interface is not compatible with the previous). The second number (<minor>) is when you add functionality in a backwards-compatible manner. The third number (<patch>) is related to bug resolutions and more preciously when you make backwards-compatible bug fixes. If you want to know more about this kind of thing, you should go on Semantic Versioning or X.Org.

You can modify the version through the parameter "format_version" (default on 0.1.0) in the function "build_format_db".

2.2.2. Sheet "slot"

The second sheet, "slot", contain all the tables of your dataset. Each table is called a **slot**.

In the RECOLAPE data call example, the structure of the slot is that:

slot_name	mandatory	definition_table
effort	TRUE	effort_table
landing	TRUE	landing_table
sampling	TRUE	sampling_table

We have 3 slots (=tables), each slot is mandatory (if not we should have FALSE as an argument) and the name of the slot in the definition data format is indicated in the definition_table column. We will see that in the next section, but all these slots need a sheet named according to the definition_table modalities.

2.2.3. Sheet "slots_hierarchy"

This sheet indicates is there being any relations between your slots. This information is related to the cardinality of your data. In database design and more precisely in the relation model, tables can be related as "one to many", "many to many", "one to one", "zero to many", ect... This kind of information leads very powerful constraints and ensure the consistency of your data.

For understanding this specification, which seems complicated for several people foreign in the field of databases. Looks at our data as an example. Here we have 3 slots and potentially 0 relation between them or a maximum of 6 relations (if each slot is related to another, on both sides). Like we said before, we can have a lot kind of relationship between data, but we should consider (for simplification) two levels:

- "One to many" relationship, this is a generalization of the major case (for example, a relation "one to one" is a specific case where many is equal to zero).
- "Zero to many" relationship (this is the same argument as the relation before).

For better understand a relation between two, a solution is to try to make a sentence which explain the relation. Let take an example between the effort slot and the landing slot. In our case, an effort data could be related to no landing (if non-catch during a set for example) or several landings (if we have a partial landing). Relation between the effort slot and the landing slot is a "zero to many" type. Similarly, a landing should be associated with one and only one effort (if landing append that mean an effort related to). The relation between the landing and the effort is a "one to many" type (if we generalize, with many equal to one). In the definition data format, this relation can be specified like that:

link	level_1	level_2	level_3	level_4	level_5	level_6	level_7	level_8
landing_effort	flag_country	vessel_id	year	month	area	fishing_level7		

To define a relation in the definition data format, you do not need to specify what kind of relation is but only direction of the link separated by "_" (here landing to effort). The different levels are all the variables involved in the relation.

For now, the following verification functions can use to check "one to many" relations. However, the code was thought to be upgradable and it's possible to add more verification type (also specific verification like "one to one" associated with her constraints). Furthermore, the default template of the sheet make available 8 levels of hierarchies, but the function could manage more than 8. Just add columns and keep the current nomenclature ("level_x"). Finally, if your data does not present relationship (lucky you are!), just leave this sheet with the empty template.

With our data, we can underscore two more relations, "one to many":

link	level_1	level_2	level_3	level_4	level_5	level_6	level_7	level_8
sampling_landing	flag_country	vessel_id	year	area	species	fishing_level7		
sampling_effort	flag_country	year	vessel_id	area	fishing_level7	species	length_code	length_class

Here we have:

- A sampling could have one and only one landing associated ("one to many" relationship).
- A sampling could have one or several effort data associated ("one to many" relationship in case of partial landings).

2.2.4. Sheet slot definition

The definition data format contains at least one slot definition in relation to slot define in the sheet "slot".

For our example, we would describe the slot definition of the slot "sampling". You could also find the slot definition of the other slot in the definition data format of RECOLAPE data call.

For the slot "sampling"" we have the slot definition "sampling_table":

column_name	nullable	mandatory	pk	type_name	category
sampling_type	FALSE	TRUE	TRUE	sampling_type	codelist
flag_country	FALSE	TRUE	TRUE	country_type	codelist
year	FALSE	TRUE	TRUE	year_type	numeric
trip_code	FALSE	TRUE	FALSE	trip_code_type	text
vessel_id	FALSE	FALSE	TRUE	vessel_type	codelist
nb_set	FALSE	FALSE	FALSE	nb_set_type	numeric
day_at_sea	FALSE	TRUE	FALSE	day_sea_type	numeric

sampling_method	FALSE	TRUE	TRUE	sampling_method_type	codelist
aggregation_level	FALSE	TRUE	FALSE	aggregation_level_type	codelist
station_number	FALSE	TRUE	TRUE	station_number_type	text
catch_registration	FALSE	TRUE	TRUE	catch_registration_type	codelist
date	FALSE	TRUE	FALSE	date_type	date
area	FALSE	TRUE	TRUE	area_type	text
fishing_level6	FALSE	TRUE	FALSE	fishing_l6_type	codelist
fishing_level7	FALSE	TRUE	TRUE	fishing_l7_type	codelist
species	FALSE	TRUE	TRUE	species_type	codelist
catch_category	FALSE	TRUE	TRUE	catch_category_type	codelist
weight	TRUE	TRUE	FALSE	weight_type	numeric
weight_allspecies	FALSE	TRUE	FALSE	weight_allspecies_type	numeric
length_code	FALSE	TRUE	TRUE	length_code_type	codelist
length_class	FALSE	TRUE	TRUE	length_class_type	numeric
number_at_length	FALSE	TRUE	FALSE	number_at_length_type	numeric

Her we have 6 column:

- "column_name", names of the column in our data.
- "nullable", data in the column could be null (TRUE) or not (FALSE).
- "mandatory", column are mandatory (TRUE) or not (FALSE).
- "pk", the column is a primary key (TRUE) or not (FALSE). In relational databases, a primary key is one or a concatenation of variables that uniquely specify an element (a row) in a slot. A primary key could not be repeated and have to be unique. This is a very important parameter in the database model.
- "type_name", this is the name of the type of data present in the associated column.
- "category", category of data present in the associated column.

All of this information will be used for different checks and verification (see section below).

2.2.5. Categories definitions

In the previous sheet, we have defined several categories types. These categories apply specifications on data associated.

For now, there are 5 categories possible: codelist, numeric, logical, text and date.

2.2.6. Codelist category

In our example, we have several codelist category. A codelist is a list of codes or meanings that represent the only allowed values for a particular data item. All the codelist type are referenced in the sheet "codelist_types" and link column "type_name" of the different sheet slot definition with a column "enumeration_table" and a sheet "enumeration_table". This last sheet contains all the codes/meanings of the codelist associated and a description of it.

On the "sampling_table" use before as an example (the first table is the "codelist_types" and the second table are a focus on "codelist_specie"):

type_name	enumeration_table
sampling_type	codelist_sampling_type
country_type	codelist_country
vessel_type	codelist_vessel
sampling_method_type	codelist_sampling_method
aggregation_level_type	codelist_agregation_level
catch_registration_type	codelist_catch_registration
fishing_l6_type	codelist_fishing_level6
fishing_l7_type	codelist_fishing_level7
species_type	codelist_specie
catch_category_type	codelist_catch_category
length_code_type	codelist_lenght_code

code	description
SWO	Xiphias gladius
YFT	Thunnus albacares
SKJ	Katsuwonus pelamis
BET	Thunnus obesus

Warning! To increase the generalization of this package and is appropriation by everyone, a choice was made to let the user complete and update the template of the definition data format. In return, it's very important to keep all the structure of the sheet (column name for example). If you add another codelist sheet (in our example we should have 11 codelist sheets, one for each “type_name”) be careful to use the same template as the sheet “codelist_example” in the empty definition data format.

2.2.7. Numeric category

In our example, we have 7 numeric categories:

type_name	is_integer	min	max
year_type	TRUE	1950	2018
nb_set_type	TRUE	1	50
day_sea_type	FALSE	1	90
weight_type	FALSE	10	5000
weight_allspecies_type	FALSE	10	5000
length_class_type	TRUE	10	300
number_at_length_type	FALSE	0,1	500

Like in the codelist category before, all information about this category is referenced in the sheet “numeric_types” (with a link between with the sheet slot definition through “type_name”). Furthermore, we have 3 new column:

- “is_integer”, does data stock in the column are integers (decimals not allowed, TRUE) or not (FALSE)?
- “min”, minmum value (included) of our data. You have to fill NA in the cell if not applicable.
- “max”, maximum value (included) of our data. You have to fill NA in the cell if not applicable.

2.2.8. Logical category

A logical argument in R only contains TRUE or FALSE values. In R:

- TRUE values could be T, TRUE, True or true.
- FALSE values could be F, FALSE, False or false.

Like before, all information about the logical categories are referenced in the sheet “logical_types”.

In our example, we do not have the kind of category but the template of the sheet “logical_types” could be like that:

type_name
yes_no_type

In the sheet, we see that we have only the information of the “type_name” (link to the sheet slot definition associated). We could think that this sheet should not be relevant because we do not perform any check, except the verification of logical format and we could do that directly from the information of the sheet slot definition associated. This chooses was made from a perspective of the evolution of the package: a new function/check could be easier incremented and not need, a priori, a modification of the definition data format.

2.2.9. Text category

All the information about the text categories are referenced in the sheet “text_types”.

For our example, we have these data referenced as text:

type_name
area_type
trip_code_type
station_number_type

Like in the logical category and for the same reason, we have only the information of the “type_name”.

2.2.10. Date category

The last category is the date category. In our example, we have one data category:

type_name	time_zone_utc	format_1	format_2	format_3	format_4
date_type	TRUE	ymd	ym	yQq	ymd_HMS

We can find:

- “data_type”, link with the sheet slot definition.
- “time_zone_utc”, data are stock in UTC (Coordinated Universal Time, TRUE) or not (FALSE). For now all the data in date format have to be in UTC (by default in the verification function associated). However, it could be possible, if necessary, to update the function and add a dynamic parameter for a specific time zone.
- “format_x”, specification of the date format. The date verification use the function “parse_date_time” of the package lubridate. This function have several specific formats and inherited formats in relation to the function “strptime” of the base package. For more details and format specifications, you could see the help of two functions. Furthermore, the list below contains the most common format used with large pelagic data:
    - y or Y, the first one is year without century (00-99 or 0-99) and the second is year with century (only years 0:9999 are accepted). In the function used here, year matches with century so you can use either indifferently.

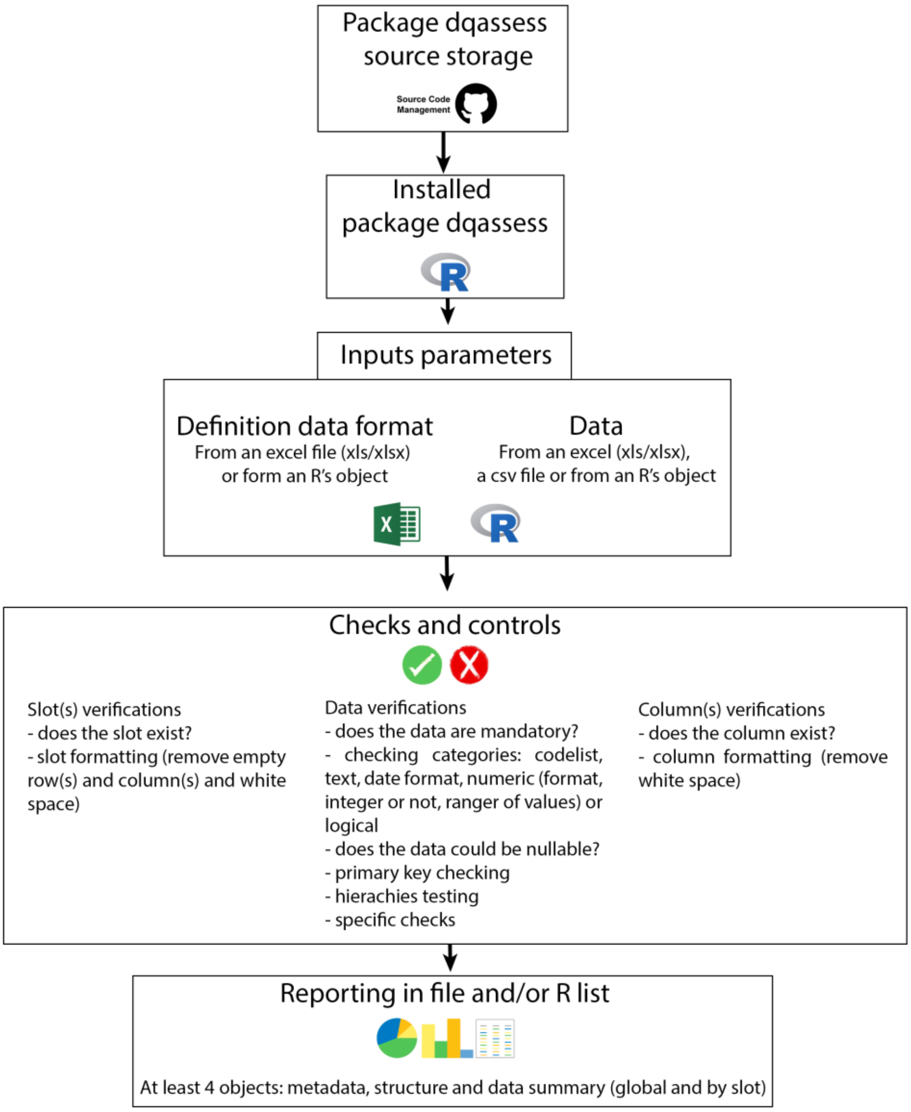
- m, month as decimal number (01-12 or 1-12).
- b, abbreviated or full-month names in the current locale. Be careful because these parameters could be affected by R options for localization. Here only English month names are understood
- d, day of the month as decimal number (01-31 or 0-31).
- H, hours as decimal number (00-24 or 0-24).
- M, minute as decimal number (00–59 or 0–59).
- S, second as decimal number (00–61 or 0–61), allowing for up to two leap seconds (but POSIX-compliant in R implementations will ignore leap seconds).
- q, quarter of the month (1-4). A specification was made for format composed format with quarter. For now, the most common was implemented and should follow template like "yQq" or "yqq" (with always the last q the quarter of the month, for example 2018Q1 or 2018q1).

Like sheet "slots_hierarchy", you can add any number of date format (incrementing by one the formats). Be sure to leave column empty if you do not want to use it (or delete it).

3. Package methodology

This section provides explanations about methodology and process of checking and controls under the package dqassess.

3.1. Global methodology



3.2. Library installation

For now, all the source of the package is stored on GitHub. This is web-based hosting service for version control system or tracking changes. First, you need to install and load the library in R (you need an internet connection):

```
# Devtools is a necessary package
# If it is not installed, run the following line
install.packages("devtools")
# Load the package from the Git
devtools::install_github("https://github.com/OB7-IRD/dqassess.git",  build_opts = c("--no-re
```

```r
save-data", "--no-manual"))
# Load the library
library(dqassess)
# You can access the package documentation with the following line
?dqassess
# If you want the documentation of a specific package function use the same syntax, for exampl
e for the function build_template_format_db
?build_template_format_db
```

3.3. Fictive dataset

As an example of controls, we use a fictive dataset build to be closer to the RECOLAPE data call. This dataset is stored in the data directory of the package source and composed of 3 files:

- Two excel files, test_fictive_data1.xlsx (containing two sheets named "effort" and "landing") and test_fictive_data2.xlsx (containing one sheet named "sampling").
- One csv file, test_fictive_data2.csv which is a copy of the excel file test_fictive_data2.xlsx.

Several errors were introduced in the dataset to providing a panel of different output report and explanation for it. Errors are focusing in red color in the two excel files.

The definition data format used was built according to the data call of the RECOLAPE project (you can find it here) For confidential reason we can have full access to the data of the project, but all of the package was tested under them.

3.4. Checking data

To launch the checking of data, you have to run the following lines:

```r
result_checking <- checking_data(obj,
                                 format_db,
                                 ignore_case_in_codelist,
                                 report,
                                 report_dir,
                                 text_file_sep,
                                 text_file_dec,
                                 file_name_slot)
```

Like explain in the function documentation (run ?checking_data in R console), you have to fill 8 parameters:

- "obj", this is the path of the file or R's object that contain data.
- "format_db", this is the path of the file or R's object that contain the definition format.

- “ignore_case_in_codelist”, by default yes in the function (TRUE). You specify her if you ignored, or not, cases (upper or lower) in the codelist.
- “report”, selected format for the report. You could choose between files, list or both. By default files selected.
- “report_dir”, location of export directory for the report. By default, the function uses the temporary directory.
- “text_file_sep”, if the argument obj is a csv file, specify here the field separator of it. By default the separator is “;”.
- “text_file_dec”, if the argument obj is a csv file, specify here the string use for decimal points. By default the decimal is “.”.
- “file_name_slot”, if the argument obj is a csv file (and by analogy contain only one slot), you have to specify here the name of the slot for a match with the definition data format. By default not provided.

In the following sections, we will run different scenarios and check in detail the output report.

3.4.1. First example: data on excel file

For the first example, we used a dataset composed of 2 slots (effort and landing) from a xlsx file (test_fictive_data1.xlsx). Associated with this data we used the definition data format built during the RECOLAPE project (recolape_definition_data_format.xlsx).

Now, run the following lines (you need to adapt parameters, especially paths, to your configuration):

```
result_checking1 <- checking_data(obj = "path_test_fictive_data1.xlsx",
                                  format_db = "path_recolape_definition_data_format.xlsx",
                                  report_dir = "path_output_directory")
```

Checking of the R console

It’s very important to take a look at the R console, to see what appends and if the function return information, warnings or errors. For our example, you should have this in the R console:

```
## (INFO) Empty enumeration table (code list) "codelist_vessel" in the format definition file.
## (INFO) Empty enumeration table (code list) "codelist_vessel" in the format definition file.
## (INFO) Empty enumeration table (code list) "codelist_vessel" in the format definition file.
## Correct import of the format file definition
## Correct import of data
## Slot effort found
## Checking in progress, be patient or take a coffee
## Slot landing found
## Checking in progress, be patient or take a coffee
## Slot sampling not found
```

The first line gives to us some information. When you put a definition data format in the function, verification was done through a sub function “read_format_db”. This sub function check if the definition data format was relevant. Here, it seems that the sheet “codelist_vessel” was empty but cited in a slot definition. Furthermore, the information was repeated 3 times. If we check in our file, we can see that the “codelist_vessel” was specified in 3 slots definition: effort_table, landing_table and sampling_table.

Lines number 4 and 5 says that the function imported successfully a definition data format and data.

The next following lines explain to users what the function doing. Here we can see that 2 slots were founding (effort and landing) and 1 was missing (sampling). Indeed, in the data file provided we have only these 2 slots. This case could be appended for example where you a very large dataset. If you launch all your data through the function, you could saturate R software and collapse it. It could a better option to split your data in multiple datasets (we will see that in the seconde example).

Checking outputs

Now let check our outputs. There are in the reporting directory (specify in the parameter “report_dir” of the function below) and her names are built through a concatenation between data name, time when the function begins to run and information on global content of the file. In any case, we 4 kinds of csv files:

- A meta.csv file (metadata) which bring global information on definition data format used (name and version) and other information like output format and location.
- A str.csv file with information on the structure of data imported and more precisely if data are according to the structure define in the definition data format (all slots defined are provided ? same question for the columns associated).
- A data.csv file with a summarize, by slots, of tests applied to data.
- One or more (in relation to the number of slots founded in the data/definition data format) slot.csv file with results of each test/verification (VALID or INVALID) made on every data.

Structure and data reports have the same file structure:

- “test” column identify which test/control was applied on data.
- “result” and “message” columns are associated with tests/control output. The first could have 3 modalities with importance increasing: OK, INFO or ERROR. The second provide complementary information, useful to more understand the test/control output.

In this example, all mistakes included in our data are highlighting. Look at outputs reports str.csv and data.csv for identified where problems are, and if necessary, use the slot report for focusing on data associated.

3.4.2. Second example: data on csv file

For this second example, we used the same definition data format as before (recolape_definition_data_format.xlsx) but associated with data in a csv file (by analogy composed of one slot, “sampling”).

For this example, run these lines in R console:

```
result_checking2 <- checking_data(obj = "path_test_fictive_data2.csv",
                                  format_db = "path_recolape_definition_data_format.xlsx",
                                  report_dir = "path_output_directory",,
                                  file_name_slot = "sampling")
```

Checking of the R console

Like before, the R console provides useful information of what append:

```
## (INFO) Empty enumeration table (code list) "codelist_vessel" in the format definition file.
## (INFO) Empty enumeration table (code list) "codelist_vessel" in the format definition file.
## (INFO) Empty enumeration table (code list) "codelist_vessel" in the format definition file.
## Correct import of the format file definition
## Correct import of data
## Slot effort not found
## Process for the next slot available
## Slot landing not found
## Process for the next slot available
## Slot sampling found
## Checking in progress, be patient or take a coffee
```

As we expected, only the table “sampling” was found. For a description of the other outputs, you can look to the previous section (Checking of the R console example 1).

Checking outputs

In this case, we can see 4 outputs: meta.csv file, str.csv file, data.csv file and one file called slot_sampling.csv.

Like the previous example, mistakes included in the “sampling” data are identified. For example, there are troubles with the codelist of the column “flag_country” (“TOF” code are not a valid code according to the definition data format associated). Like before, look at outputs reports str.csv and data.csv for identified where problems are, and if necessary, use the slot report for focusing on data associated.

MARE/2016/22 "Strengthening regional cooperation in the area of fisheries data collection"

Annex III "Biological data collection for fisheries on highly migratory species"

Strengthening Regional cooperation in the area of large pelagic fisheries data collection (Acronym: RECOLAPE)

Deliverable 5.3.1
Analysis, report and guidelines on age issues including full documentation of used methods

COISPA TECNOLOGIA & RICERCA
STAZIONE SPERIMENTALE PER LO STUDIO DELLE RISORSE DEL MARE

Institut de Recherche pour le Développement FRANCE

hcmr ΕΛΚΕΘΕ Hellenic Center of Marine Research

azti tecnalia

Department of Fisheries and Marine Research Ministry of Agriculture, Rural Development and Environment

Coordinators: Pierluigi Carbonara[1]and Fulvio Garibaldi[2]

Collaborators: Sergio Bizzarri[3], Rui Coelho[4], Mathieu Depetris[5], Luca Lanteri[2], Pedro Lino[4], Pablo Quelle[6], Aurelie Guillou[5], Daniela Rosa[4], Ioannis Thasitis[7], George Tserpes[8].

[1]COISPA Tecnologia & Ricerca, Stazione Sperimentale per lo Studio delle Risorse del Mare, Bari, Italy

[2]Università di Genova – Dipartimento di Scienze della Terra dell'Ambiente e della Vita (DISTAV), Genoa, Italy

[3]UNIMAR Soc. Coop., Roma, Italy

[4]Instituto Português do Mar e da Atmosfera (IPMA)Olhão, Portugal

[5]Institut de Recherche pour le Développement (IRD), France

[6]Instituto Español de Oceanografía (IEO)Centro Oceanográfico de Santander, Santander-Cantabria, Spain

[7]Hellenic Centre for Marine Research (HCMR) Athens, Greece

[8]Department of Fisheries and Marine Research, Ministry of Agriculture, Rural Development and Environment, Nicosia, Cyprus

Index

Executive Summary 4

Introduction 8

2. Identification of the main drivers influencing the variability in the age data 9

  2.1 Sampling Strategy of the Hard Structure for the Age Analysis 9

  2.2 Explorative analysis on the Swordfish age data 15

    2.2.1 Age Data and Analysis 15

    2.2.2 Results 17

    2.2.3 Results Interpretation 22

3. Exchange exercise 25

  3.1 Sampling Collection and Participation 25

  3.2 Reading procedures and data analysis 27

  3.3 Results 28

  3.4 Remarks 33

4. Workshop on age reading of *Xiphias gladius* 34

  4.1 Preparation method 35

    4.1.1 Sampling the anal fin 36

    4.1.2 Section location 37

    4.1.3 Measurements 38

  4.2 Age scheme and Age criteria 40

  4.3 Analysis of the Exchange exercise and Smartdot notes 44

  4.4 Reference collection of spines 44

References 51

Executive Summary

The project "Strengthening Regional cooperation in the area of large pelagic fisheries data collection (Acronym: RECOLAPE)" in the context of the framework of MARE/2016/22 (Strengthening regional cooperation in the area of fisheries data collection. Annex III Biological data collection for fisheries on highly migratory species) aims is to strength the regional cooperation in the area of biological data collection for fisheries on highly migratory species.

It is considered strategic in this perspective: facilitate the cooperation among Member States in order to improve the procedures to assess the quality of biological data on large pelagic stocks, both at the national and regional levels. This work has been accomplished through a case study of swordfish *Xiphias gladius*. In particular this work is focused to the age data and two main actions were implemented:

- Identification of the main drivers influencing the variability in the age data, predict missing values and estimate uncertainty linked to the sampling strategy;
- Identify areas of improvement and harmonise ageing approach and protocols.

Information about age composition is useful because it can be used to draw inferences about mortality and growth rates, fishery selectivity, relative cohort strength, and other demographic processes useful to management. However, age information is often costly to obtain. These high costs force many management programs to limit the number of fish aged directly, and to rely on age–length keys (ALK) or on age slicing from growth parameters, to estimate the age composition of the stock and/or catch.

The use of ALKs or of age slicing procedures to provide an unbiased age composition estimate of the sample requires that aged fish are representative of the unaged fish.

The data obtained by the project patterns and/or by a specific data-call addressed to the DCF National Corresponds about the age sampling strategy coming from: Italy, Greece and Cyprus.

The evaluation of the sampling strategy was carried out in term of precision. The precision of the age length keys expressed in terms of coefficient of variation (CV) was estimated for each age group according to the method proposed by Baird (1983).

The sampling strategy are:

- in two cases the sample is stratified by length class: two hard structure (HS) by length class (5 cm)
- in one case the sampling is opportunistic: the HS sampled are chosen among those available without any kind of stratification

The sampling strategy stratified in general given a better result in term of precision and coverage of the Length Frequency Distribution. Anyway, the actual level of age sampling cannot ensure an adequate coverage of the Length Frequency Distribution mostly for the higher LJFL where it is more difficult the sampling activity (e.g. cost of the sample). So, a stratification by length and sex could be more adequate as well as the increasing the number of samples by strata to a minimum of 3 spines for each length class (5 cm). Poor quality ageing data have also contributed in certain cases to misleading evaluation of the population status, sometimes resulting in the stock collapse (Beamish and McFarlane 1995; Liao et al. 2013). For these reasons, an increasing effort has been devoted during the last decades to improve the age data quality (ICES, 2011; 2013), especially in the context of the European Union Data Collection Framework (DCF), which is implementing ageing exchange exercise, workshops and meetings concerning the ageing of the most important species in the European fisheries (ICES, 2018).

In this context, a common ageing protocol could be an important tool to decrease the relative/absolute bias and improve the precision (reduce CV and increase the percentage of agreement) in age determination, and increase the reproducibility among the age readers of the different laboratories (PGCCDBS, 2011). In order to reach this goal, it is useful to assess the effect of the specific factors (i.e. theoretical birthdate, ageing criteria, age scheme, reader’s experience) influencing the age reading variability on Swordfish ageing using a multiparameter approach (Principal Component Analysis). This analysis can represent a first step to standardize the reading protocols aiming at obtaining unbiased Age-Length Keys (ALKs) for *X. gladius*.

The results of the present work confirm previous studies on the high variability, occurring in the age and growth of Swordfish (Arocha et al., 2003; Quelle et al., 2014; Abid et al., 2014). This variability can be affected by several sources, such as: sampling methodologies (Coggins et al 2013), geographical differences (Abid et al., 2014), age estimation criteria, age estimation scheme, skeletal structures used (otolith or spines) (Farley et al., 2016), methodology (direct age estimation or Length Frequency Distribution Analysis) and level of experience of the readers (ICES, 2011; 2013)

The geographical location was found to be the most important factor, influencing significantly the age variability, with Longitude (West-East) being the factor most highly correlated with variability than Latitude (North-South).

The reader’s experience has been identified as an important factor affecting the precision of the age data for many species in both marine and freshwater environments.

In the present analysis, this factor was also found to be important in ageing variability; especially when we compared the results of High versus Low experience readers. Reader experience emerged as a key issue in estimating the age mostly in the first year groups as well as the oldest age group (4 years).

The results of the present analysis have demonstrated the importance of a handbook clarifying and standardizing ageing schemes (e.g. birth date), ageing criteria (e.g. number of false rings before the first winter growth increment) and preparation methods could be a useful action to overcome bias in ageing. Being the reader experience the most important factor in explaining the huge variability in the age data in the Mediterranean basin, workshops, age exercises and exchanges are considered as fundamental tool for improving the precision in the red mullet age analysis (PGCCDBS, 2011). All these actions can be an important contribution to overcoming the ageing uncertainties, thus providing accurate and robust input data for stock assessments.

The exchange exercise was held in the context of the project based on a total of 79 fish sampled from 2003 to 2017 in Mediterranean from 2 sites: Ligurian Sea and Alboran Sea. The pictures of tin section spines (anal fin) were prepared in the same way (Quelle et al., 2014; Lanteri and Garibaldi, 2019). The overall precision reached are in term of Percentage Agreement (PA), Coefficient Variation (CV) and Average Percentage of Error (APE) respectively of 64.4%, 30.8% and 23%. These values are respectively lower and higher than those considered acceptable: 80% PA and 20% CV (PGCCDBS 2011). Moreover, they were no significantly different if they were stratified by readers' experience, so this factor not explained fully the low PA and high CV reach in this exchange exercise. The analysis of the precision indices by age groups showed a negative trend from the first age group to the older one. In addition, the bias analysis on the all data seems highlight an under-estimation for the older age group, while an overestimation for the first age group (0 and 1 year). These results could be explained by the difficult to recognize the first growth increment and mostly growth increments (overlapping of the rings) in the older fish (age > 5 years). The comparison of the age readings among the readers and each reader with modal age highlighted that a group of readers follow a same age criteria. These results are confirmed also by the mean length at age as estimated by each age reader. Indeed, in the first 6 age groups (from age 0 to age 5 years) the mean length at age are comparable for the mostly of readers. All these results were discussed during the next workshop.

The Workshop on age reading of *Xiphias gladius* expected in the context of the RECOLAPE was held in Olhão (Portugal) from the 2th to the 4th April 2019. The meeting was host by IPMA Institute. Eight age readers from 4 countries and 5 laboratories (IPMA, IEO Santander, Genoa University, Unimar, IRD) participated in this workshop

The Term of Reference (ToR) of the meeting are following listed:

ToR a: Preparation method; ToR b: Age scheme and Age Criteria; ToR c: Analysis of the Exchange exercise; ToR d: Develop of a reference collection of spines.

All the ToRs were discussed during the meeting. Regarding the preparation method were evaluated the pros and cons of all protocol from each lab and from the literature. A agreed protocol was fixed in term of preparation of the structure (second ray of anal fin), conservation and thin section procedures. Mostly this last aspect seems to be fundamental to provide unbiased age data (Quelle et al., 2014). For the age scheme were discussed and agreed two age schemes: one based on the 1st January as birthday and one on 1st July. The appropriateness of using one adjustment scheme rather than other was also discussed. Indeed, if the objective is to construct an ALK the adjustment to the 1st of January was considered more useful, while for growth curves the adjustment to the 1st of July was considered more suitable as it takes into account the biology of the species. Moreover, it was fixed a reference collections of Swordfish spines. Accordingly, a reference set of 11spines was developed among the spines that they reached a PA ≥ 80% during the exchange exercise.

The work developed within the Task 5.3 is an example of cooperation in the context of the DCF among the laboratories from several Member state. It is important underline that in this work, some laboratories (IPMA, IEO-Santander) not included in the RECOLAPE consortium were also involved. So, starting from the analysis of the problem about the precision of Swordfish ageing (sampling strategy and explorative analysis) were carried out an exchange exercise and workshop to solve the uncertainties, using the same procedures fixed in the ICES context (PGCCDBS, 2011; ICES, 2011; 2013). At the end of these process were fixed common/agreed procedures (age scheme, age criteria) and methods (preparation of the spines). Therefore, it is recommended that the working group on the Swordfish ageing continue with a new exchange and workshop after three years (ICES, 2011; 2013; 2015) to see if any improvements based on the established procedures and common ageing protocol they will be reached.

Introduction

Since 2002, according to the EC Regulations n° 1543/2000, n° 1639/2001, n° 1581/2004 and n° 199/2008 and to the Commission Decision n° 93/2010, which established the DCR (Data Collection Regulation) and subsequently the Data Collection Framework (DCF) in all the EU countries, biological, environmental, technical, and socio-economic data on the catching, aquaculture and processing sector are regularly collected.

The need of coordination was present since the beginning of the Data Collection and there was a general agreement that regional coordination would have greatly increased the efficiency, effectiveness and integration of the National Programs.

The coordination on biological sampling issues is important, to ensure that the schemes respect statistically sampling procedures to cover the spatial and temporal distribution of the species and fisheries. The project "Strengthening Regional cooperation in the area of large pelagic fisheries data collection (Acronym: RECOLAPE)" in the context of the framework of MARE/2016/22 (Strengthening regional cooperation in the area of fisheries data collection. Annex III Biological data collection for fisheries on highly migratory species) aims is to strength the regional cooperation in the area of biological data collection for fisheries on highly migratory species. These assessments are the basis for the definition of guidelines with the objective of making available more robust key information to end-users. Indeed, it is considered strategic in this perspective: facilitate the cooperation among Member States in order to improve the procedures to assess the quality of biological data on large pelagic stocks, both at the national and regional levels. This work has been accomplished through a case study of swordfish *Xiphias gladius*. In particular two main action were implemented:

- Identification of the main drivers influencing the variability in the age data, predict missing values and estimate uncertainty linked to the sampling strategy;
- Identify areas of improvement and harmonise ageing approach and protocols.

In particular the work related the main drivers influencing the variability in the age data of the selected stocks (*Xiphias gladius*, swordfish) will be accomplished through the following steps:

i. analyse ALKs obtained by MS/partners with meta data regarding year of sampling, sex, geographical area, gear;

ii. compare ALK, considering the stratification in the sampling scheme (area, sex, length etc.) in term of precision reached;

iii. interpret the results to recognize gap in the sampling design, to predict missing values and to estimate uncertainty.

iv. perform an explorative analysis by a multiparametric approach (i.e Principal Component Analysis) to recognize the main drivers influencing the variability in the age data of swordfish in the Mediterranean basin.

Moreover, the work related to the tools to harmonize/coordinate age-reading between Member States (MS) was accomplished by the bottom-up approach (PGCCDBS, 2011). In this approach for the species, where is requested by MS an age harmonization action, this is realized:

i. exchange exercise
ii. workshop

Downstream to this process they are clarified the aspects related to:

- Preparation methods of the hard structure used in the age analysis
- Age scheme
- Age criteria.

2. Identification of the main drivers influencing the variability in the age data

2.1 Sampling Strategy of the Hard Structure for the Age Analysis

Hard structure (HS) available for fish ageing are different: otoliths (sagittae, lapilli, asterischi), scales, vertebrae, spines and opercular bones (Panfili et al., 2002). In the context of the Data Collection Framework (DCF EU Reg. 199/2008) more than 30 species and/or group of species are subject to age sampling and include: small pelagic species (e.g. *Engraulis encrasicolus* (Linnaeus, 1758), *Sardina pilchardus* (Walbaum, 1792), demersal species (e.g. *Mullus barbatus* (Linnaeus, 1758), *Merluccius merluccius* (Linnaeus, 1758)) and large pelagic species (e.g. *Thunnus alalunga* (Bonnaterre, 1788), *Thunnus thynnus* (Linnaeus, 1758), *Xiphias gladius* (Linnaeus, 1758)).

For age determination of the main teleosteans demersal species the sagittae are usually used, while for Lophiusbudegassa and Lophiuspiscatorius the thin transversal section of the illicium (first transformed spine of dorsal fin) is preferred, in Chondrichthyes species the section of vertebra and/or spines is usually used, and eventually in the large pelagic species are used several Calcified Strutctures (CS): otolith, vertebra and spines (Carbonara and Follesa, 2019).

In DCF target species are divided in two main groups: G1 species that drive the international management process, including species under EU management or recovery plans, which assessment is regularly carried out, and G2 species that are important in terms of landing and/or economic values, and for which assessment is regularly carried out with a different calendar.

The G1 species of Large Pelagic include for example *Thynnus thynnus* and for this species the hard structure (HS) for the age analysis are sampled yearly and a fixed number of HS is collected to achieve a total number requested by each MS (RCM MED&BS-LP 2016).

For the G2 species that include for example *Xiphias gladius* and *Thunnus alalunga*, the sampling of HS and age data has a three-year basis.

Information about age composition is useful because it can be used to draw inferences about mortality and growth rates, fishery selectivity, relative cohort strength, and other demographic processes useful to management. However, age information is often costly to obtain. These high costs force many management programs to limit the number of fish aged directly, and to rely on age–length keys (ALK) or on age slicing from growth parameters, to estimate the age composition of the stock and/or catch.

Proportional-age subsampling of the catch is desirable because based on multiple statistical properties, but fixed-age subsampling is frequently used because of improved efficiency in field operations. Instructing field personnel to take a fixed number of fish per length class is easier to execute than having personnel taking fish with lengths in proportion to the abundance of each length-group. The use of ALKs or of age slicing procedures to provide an unbiased age composition estimate of the sample requires that aged fish are representative of the unaged fish. This implies that aged fish are taken with the same gear, season, and spatial location as the unaged fish (Ricker 1975; Kimura, 1977). The performance of ALKs to accurately represents the actual age structure of the entire sample is depending from many factors as the sampling strategy (fixed-age subsampling vs proportional-age subsampling), life span (long-lived fish vs short-lived fish), exploitation status and recruitment strength (Coggins et al., 2013).

The optimum number of otoliths per length class cannot be given in a universal form. A description of the optimum sample size of age readings and length measurements dependent on a universal cost function is given in Oeberst (2000). Moreover according to Mandado and Vasquez (2011) a sample of 20 otoliths in a stratified sampling by length class was considered the optimum for a species with 30-40 length classes. Coggins et al., (2013) show as 10 fish aged by length class, 500 – 1000 fish in total, provided unbiased ALK, respectively for short-lived and long-lived fish. Negligible benefits were achieved collecting more than 10 fish by length class or than 500-1000 fish (Coggins et al., 2013).

The data obtained by the project partners and/or by a specific data-call addressed to the DCF National Corresponds include data about the age sampling strategy from: Italy, Greece and Cyprus.

The evaluation of the sampling strategy was carried out in term of precision. The precision of the age length keys expressed in terms of coefficient of variation (CV) was estimated for each age group according to the method proposed by Baird (1983).

The total number of individuals for a given age group is calculated as:

$$N = \sum_i N_i * p_i$$

where

N_i= number of individuals for length class$_i$;

p_i= proportion of individuals of a given age group for length class $_i$;

$N_i p_i$=number of individuals for length class i belonging to a given age group; $n_i p_i$=number of individuals whose otolith were read for ageing for length class $_i$ belonging to a given age group. Variance for each length class $_i$ is calculate according to Gulland (1966) as:

$$Var(N_i \cdot p_i) = N_i^2 Var(p_i) + p_i^2 Var(N_i)$$

The second term of the above equation is related to the variability associated with the length measurement and can be considered negligible, thus assuming that age groups are distributed by length according to a binomial function we have:

$$\mathrm{var}(p_i) = \frac{p_i \cdot (1 - p_i)}{n_i}$$

Withn$_i$= number of individuals "read" for length class$_i$, i.e. all the fish whose age was estimated in length class$_i$.

The variance of total individuals of a given age group is calculated as sum of variance for each length class in which there are individuals of the age group as:

$$\text{var}(N) = \sum_{i=1}^{L} N_i^2 \text{var}(p_i)$$

where *L* are the length class in which individuals of a given age group are found. Finally CV for a given age group is calculated as:

$$CV = \frac{\sqrt{\text{var}(N)}}{\sum_{i=1}^{L} N_i p_i}$$

In the swordfish the HS used for the age analysis are the first three spines of the anal fin (Lanteri and Garibaldi, 2019). The sampling strategy evaluated are:

- in two cases the sample is stratified by length class: two HS by length class (5 cm)
- in one case the sampling is opportunistic: the HS sampled are chosen among those available without any kind of stratification

In the case of Cyprus, the data are referred to the years 2008-2010. In total were sampled 1064 specimens of Swordfish of these 219 are aged (20.6% of the total). The overall CV is 4.82 %. The Lower Jaw-Fork Length (LJFL) of the specimens sampled are included between 75 to 195 cm and the age group between 1 to 9 years (Fig. 2.1.1). The length class covered/not covered by the age sampling are respectively 27 and 3.

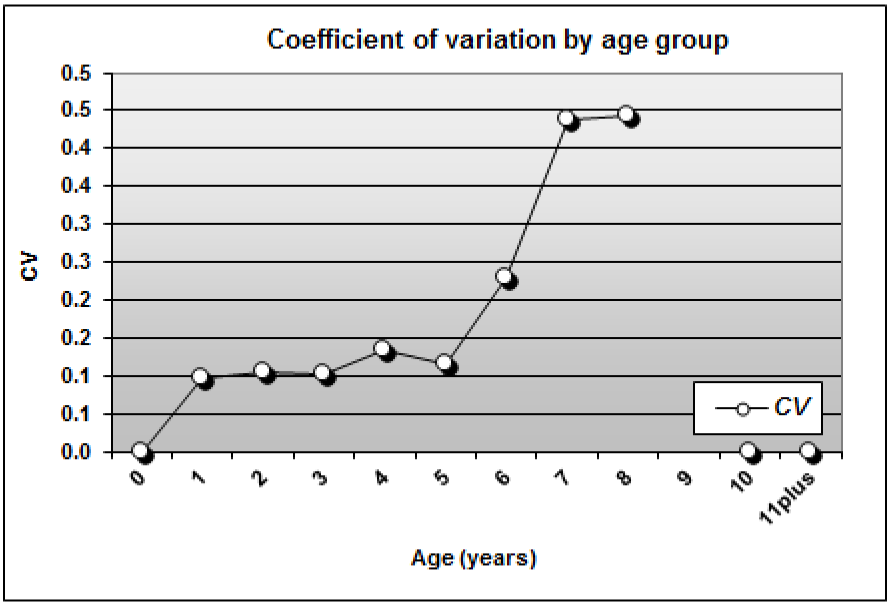


Figure 2.1.1 - The Coefficient of variation by age group in Cyprus samples (2008-2010)

The Figure 2.1.1 reported the Coefficient of variation by age group.

In the case of the Italy the data are referred to the years 2014-2016. In total were sampled 1947 specimens of Swordfish of these 348 are aged (18.9% of the total). The overall CV is 4.54%. The Lower Jaw-Fork Length (LJFL) of the specimens sampled are included between 90 to 260 cm and the age group between 0 to 8 years (Fig. 2). The length class covered/not covered by the age sampling are respectively 23 and 5.

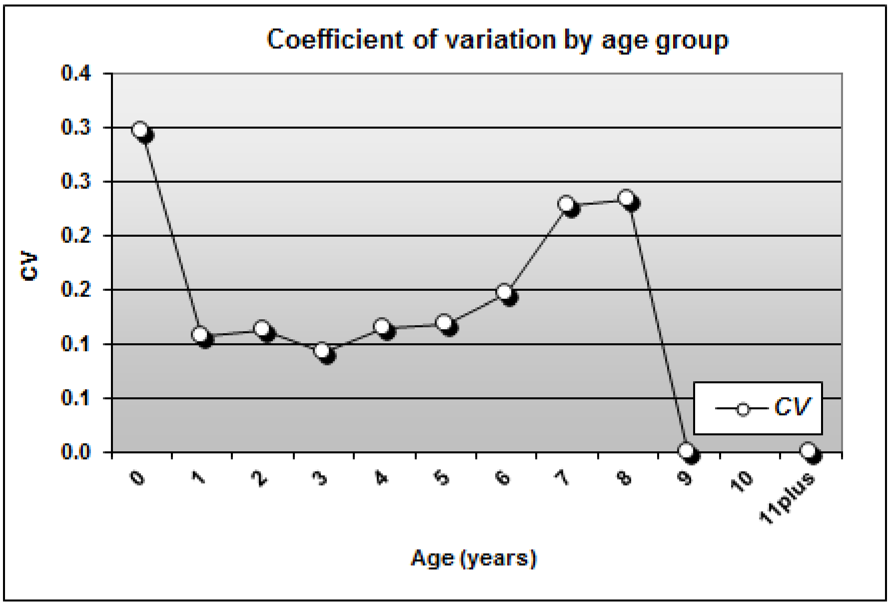


Figure 2.1.2 - The Coefficient of variation by age group in Italian samples (2014-2016)

The Figure 2.1.2 reported the Coefficient of variation by age group.

In the case of the Greece the data are referred to the years 2014-2016. In total were sampled 870 specimens of Swordfish of these 40 are aged (5.63% of the total). The overall CV is 6%. The Lower Jaw-Fork Length (LJFL) of the specimens sampled are included between 60 to 220 cm and the age group between 0 to 7 years (Fig. 2). The length class covered/not covered by the age sampling are respectively 14 and 15.

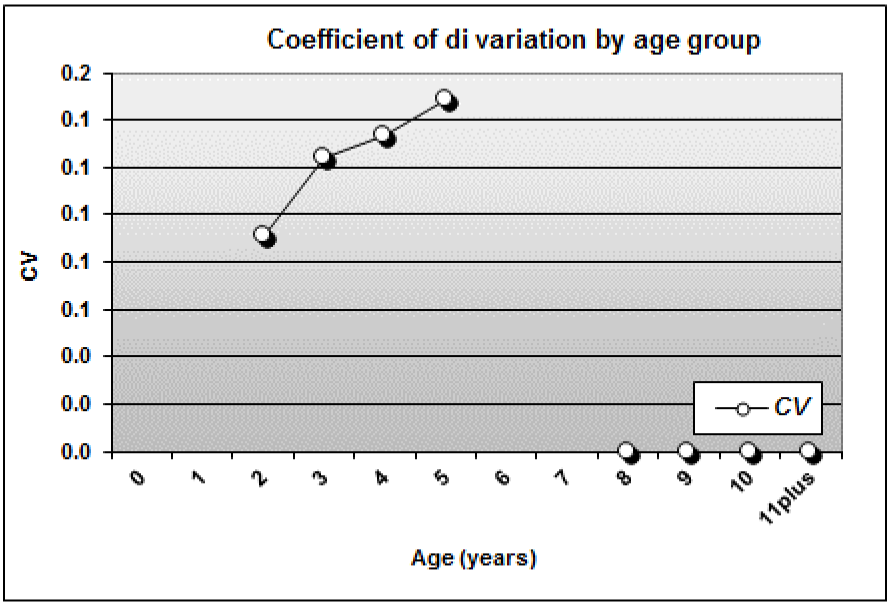


Figure 2.1.3 - The Coefficient of variation by age group in Greece sample (2014-2016)

The Figure 2.1.3 reported the Coefficient of variation by age group.

The sampling strategy stratified in general given a better result in term of precision and coverage of the Length Frequency Distribution. Anyway, the actual level of age sampling cannot ensure an adequate coverage of the Length Frequency Distribution mostly for the higher LJFL where it is more difficult the sampling activity (e.g. cost of the sample). So, a stratification by length and sex could be more adequate as well as the increasing the number sample by strata to a minimum of 3spines for each length class (5 cm). Experiences gathered in the DCF for samplings of commercial catches in Italian GSAs evidenced an acceptable coefficient of variations (around 5%) when 5 otoliths by sex, length class, metièr and quarter were sampled. The following criteria were taken into account to set the sample size for each length class:

- For the smallest size groups, that presumably contain only one age group, the number of HS per length class may be reduced;
- Conversely more otoliths per length are required for the larger length classes.

Analyses carried out in the Baltic sea (AA.VV., 2011) have shown that the necessary number of age readings in a length class depends on:

- the portions of the length classes within the length frequency;
- the maximum variance of the portions of the age-groups within the length class.

Combining data several source (embark and landing sampling) can contribute gain a better coverage for growth estimation.

Stock management needs information on annual basis for several reason such as the inter-annual variation in recruitment which ultimately influenced population abundances and age structure. For the selected stock hake, red mullet, common sole, turbot and sprat (G1 species) it is mandatory collect and analyze otolith yearly (AAVV 2014; ICES 2015).

2.2 Explorative analysis on the Swordfish age data

Age and growth data are among the most important input data in the stock assessment analytical models (Reeves, 2003). However, bias in these data can lead to stock diagnosis failures (Eero et al. 2015). Poor quality ageing data have also contributed in certain cases to misleading evaluation of the population status, sometimes resulting in the stock collapse (Beamish and McFarlane 1995; Liao et al. 2013). For these reasons, an increasing effort has been devoted during the last decades to improve the age data quality (ICES, 2011; 2013), especially in the context of the European Union Data Collection Framework (DCF), which is implementing ageing exchange exercise, workshops and meetings concerning the ageing of the most important species in the European fisheries (ICES, 2018).

In this context, a common ageing protocol could be an important tool to decrease the relative/absolute bias and improve the precision (reduce CV and increase the percentage of agreement) in age determination, and increase the reproducibility among the age readers of the different laboratories (PGCCDBS, 2011). In order to reach this goal, it is useful to assess the effect of the specific factors (i.e. theoretical birth date, ageing criteria, age scheme, reader’s experience) influencing the age reading variability. The objective of this work is to investigate the potential influence of the differences in protocols, reading experience and geographical parameters on Swordfish ageing using a multiparameter approach. This analysis can represent a first step to standardize the reading protocols aiming at obtaining unbiased Age-Length Keys (ALKs) for *X. gladius*.

2.2.1 Age Data and Analysis

The Swordfish were regularly collected within the DCF. In our analysis, we used the Length/Age data collected between 2007-2016 in three countries: Italy, Cyprus and Greece. The data coming from 5 sampling location in Mediterranean (Fig. 2.2.1.1)

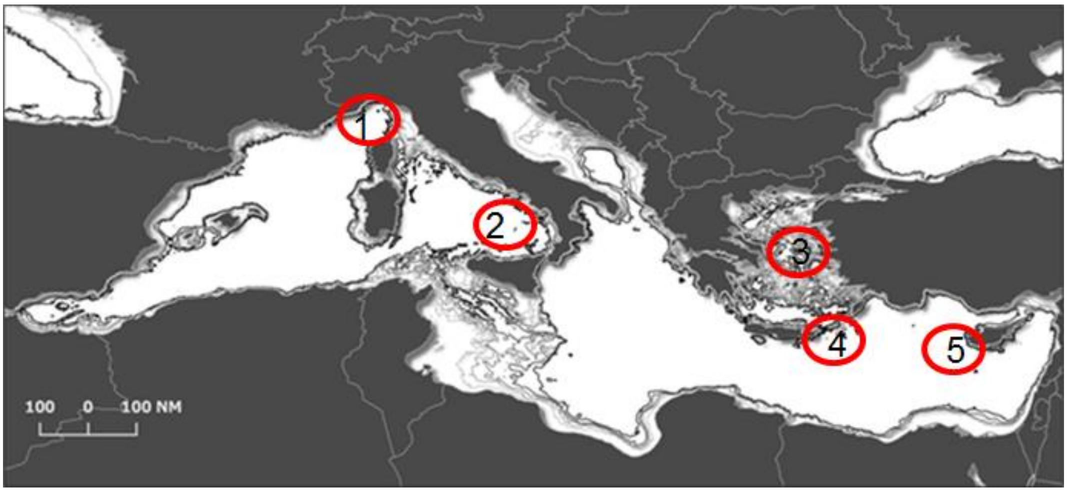


Figure 2.2.1.1 –The sampling location 1 Ligurian sea; 2 South Tyrrhenian sea; 3: Aegean Sea; 4 Crete; 5: Cyprus

Moreover, for each age data set, we considered the following meta-data: sex, used theoretical birth date (e.g. 1st January or 1st July), reader experience (Low: < 200spines, Medium: 200-500spines; High: >500spines of Swordfish read), geographic location as an average between the borders of latitude (Y) and longitude (X) of each GSA and sampling type (at sea, at landing).

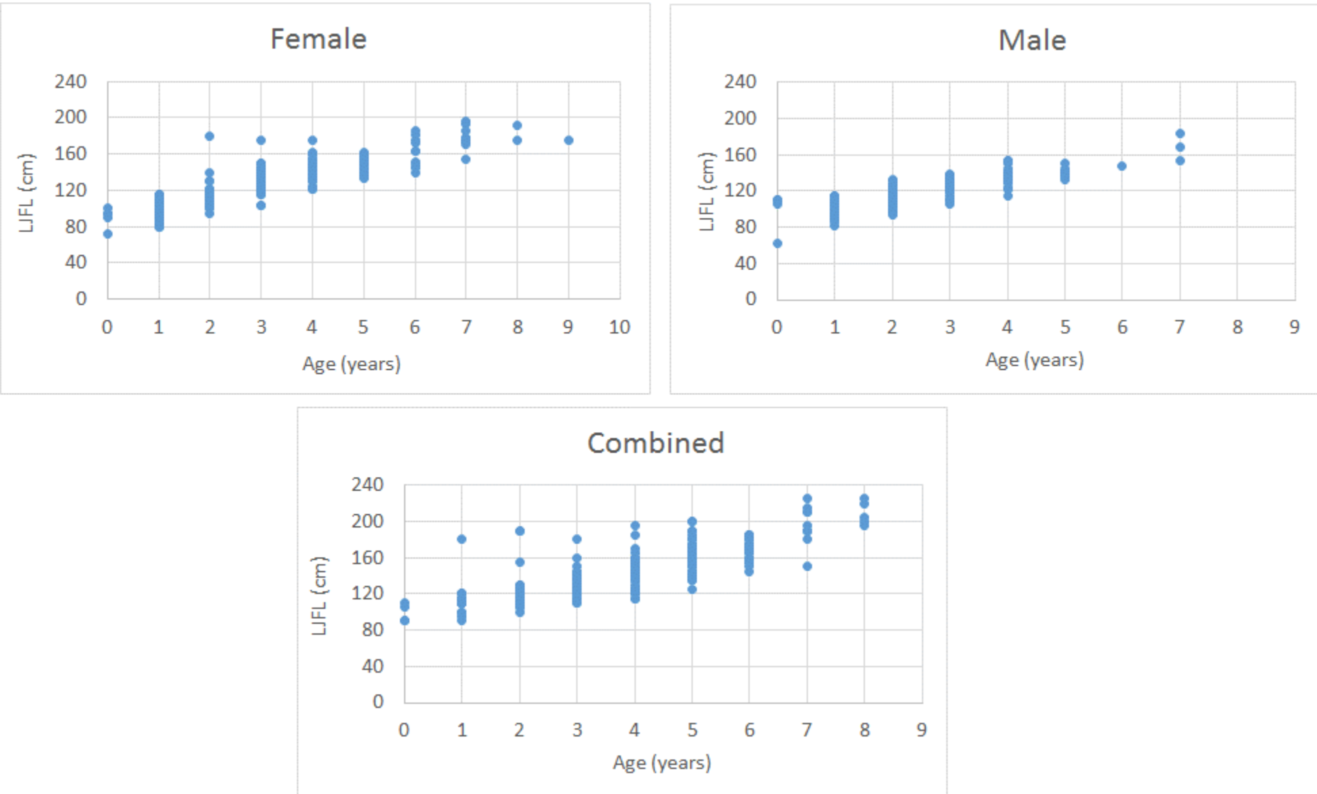


Figure 2.1.1.2 - Length at age data by sex

In total, the spines of 825 specimens (396 Combined, 270 females, 159 male) (Fig. 2.1.1.2) collected in 4 GSAs (Fig. 2.2.1.1) and aged by scientific staff from 4 laboratories.

Principal Component Analysis (PCA) was applied to identify the most informative variables influencing the differences in the ageing data of Swordfish in the Mediterranean basin. The PCA is a multivariate statistical technique (Jolliffe, 2002; Abdi and Williams, 2010) used to reduce a set of variables to a smaller set of orthogonal variables called principal components (PCs). It was performed using the FactoMineR library (Lê, Josse and Husson, 2008) available in R statistical software (R Development Core Team, 2018). The main feature of the FactoMineR library is the ability to perform the analysis using different types of variables (quantitative or categorical).

A first analysis was carried out taking into account all the age groups together using as quantitative variables the LJFL and the age of the specimens and five qualitative variables: sex, theoretical birth date, reader's experience, sampling type. Further 4 PCAs were then performed for each age group from 1 to 4 years using fish size (LJFL) and GSA's coordinates (longitude and latitude) as quantitative variables and theoretical birth date, reader experience, sex and sampling type as qualitative variables.

The number of PCs to be considered for each PCA was determined using the Kaiser's rule (Kaiser, 1960), retaining only those PCs whose variances exceed 1, with any PC with variance lower than 1 being less informative and thus not worth to be retained (Jolliffe, 2002).

2.2.2 Results

In the PCA performed on the whole dataset, the first two principal components were retained, accounting for 95.34% of the total variability (Tab. 2.2.1.1). The remaining 4.66% variability was explained by the other 2 principal components. The first principal component (PC1) was strongly correlated with all four original variables (TL, age, latitude and longitude) while the best correlation was shown by Longitude (Tab. 2.2.1.2). The LJFL, age and latitude variables vary together, being positively correlated with PC1; in contrast, longitude had an opposite effect. Although none of the qualitative variables showed a strong correlation with PC1, except the reader's experience. Longitude had the higher correlation with PC2, showing an opposite behaviour than latitude. Among the qualitative variables, PC2 showed weak significant correlations in descending order with reader's experience, birth date, sex and sampling type.

The PCAs performed on each age class showed a strong geographical effect mostly driving PC1. Indeed, longitude and latitude were the best correlated variables in almost all the age groups, at least in the PC1, but with opposite directions. Moreover, LJFL was mostly correlated with PC2 (Table 2.2.1.2).

Among the qualitative variables, the highest correlation with PC1 was shown for the reader's experience and birth date; the former especially in the lower age classes, while the latter mostly in the highest age groups. The contribution of birthday was important for all age group 1 (PC1) and 4 (PC2) (Tab. 2.2.1.2) (Fig. 2.2.1.3). While the sex and sampling type are less important in term of variability correlation.

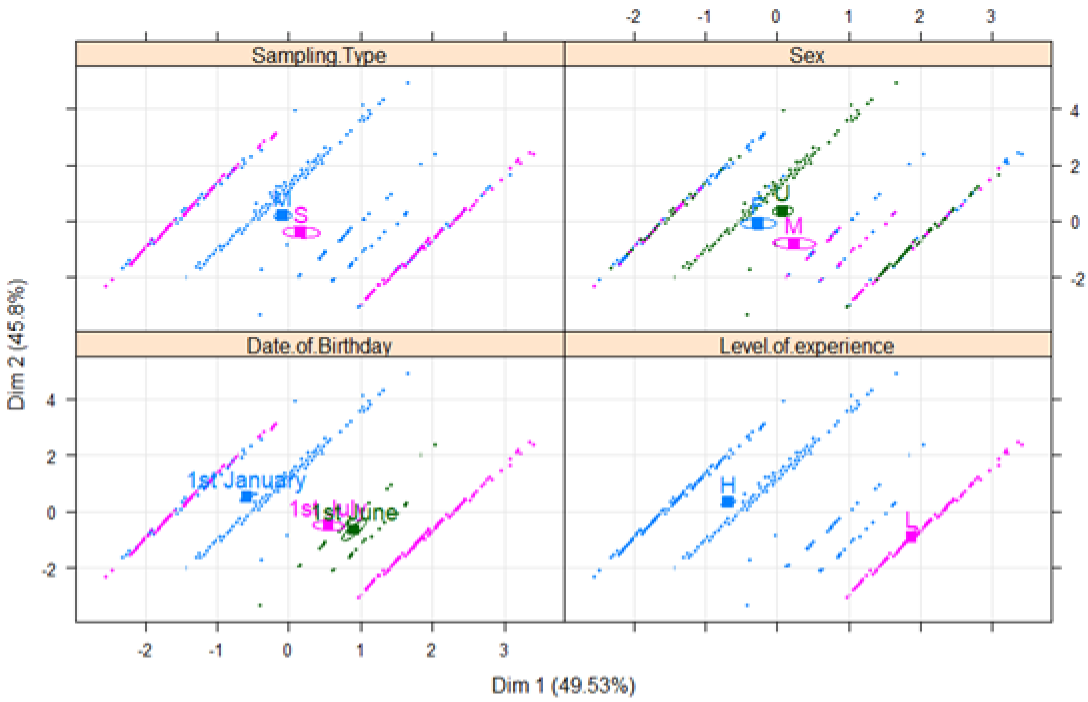
Table 2.2.1.1 - Values of variance (Variance), percentage of variance (% of var.) and cumulative percentage of variance (Cumulative % of var.) accounted for each dimension for the different PCAs performed.

SEX / Age group	Variables	Dim.1	Dim.2	Dim.3	Dim.4
WHOLE DATASET	Variance	1.981	1.832	0.14	0.046
	% of var.	49.535	45.803	3.504	1.158
	Cum. % of var.	49.535	95.338	98.842	100.00
AGE 1	Variance	2.004	0.959	0.037	-
	% of var.	66.809	31.955	1.236	-
	Cum. % of var.	66.809	98.764	100.00	-
AGE 2	Variance	1.974	0.980	0.046	
	% of var.	65.812	32.653	1.535	
	Cum. % of var.	65.812	98.465	100.00	
AGE 3	Variance	2.086	0.866	0.047	-
	% of var.	69.549	28.873	1.578	-
	Cum. % of var.	69.549	98.422	100.00	-
AGE 4	Variance	2.175	0.768	0.057	-
	% of var.	72.487	25.609	1.905	-
	Cum. % of var.	72.487	98.095	100.00	-

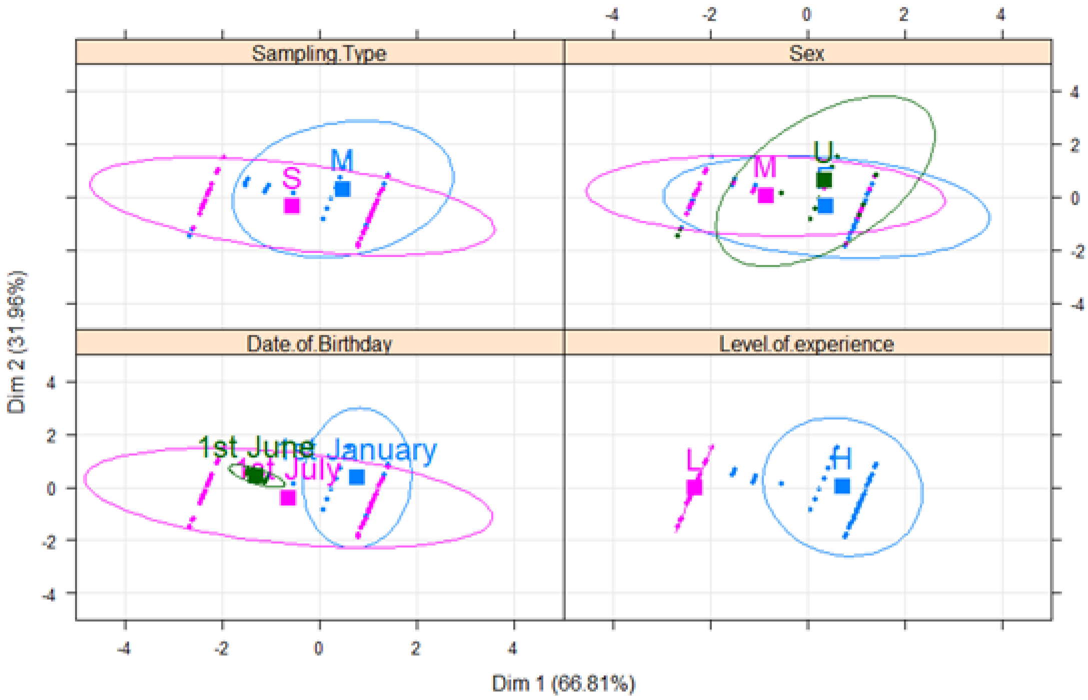
Table 2.2.1.2: Summary of the correlation coefficients of both continuous and qualitative variables with a given principal component of each PCAs. No significant correlation indicated in red (p> 0.05).

SEX / Age group	Variables	Dim.1	Dim.2
WHOLE Data set	Latitude	0.89	-0.42
	Age	0.47	0.84
	LJFL	0.32	0.91
	Longitude	-0.92	0.34
	Exp	0.64	0.17
	Sex	0.02	0.10
	Sampling type	0.01	0.04
	Birth date	0.18	0.14
AGE 1	Longitude	0.96	-0.23
	LJFL	0.31	0.95
	Latitude	-0.98	-0.01
	Experiance	0.83	0.07
	Sex	0.16	0.16
	Birth date	0.28	0.14
	Sampling type	0.13	0.11
AGE 2	LJFL	0.22	0.97
	Latitude	-0.98	-0.01
	Longitude	0.97	-0.16
	Experiance	0.76	0.03
	Sex	0.04	0.01
	Sampling type	0.01	0.06
	Birth date	0.21	0.11
AGE 3	LJFL	0.48	0.87
	Latitude	-0.96	0.19
	Longitude	0.95	-0.23
	Experiance	0.79	0.14
	Sex	0.01	0.01
	Sampling type	0.01	0.01
	Birth date	0.37	0.07
AGE 4	LJFL	0.60	0.79
	Latitude	-0.96	0.21
	Longitude	0.94	-0.28
	Experiance	0.79	0.01
	Sex	0.05	0.01
	Sampling type	0.01	0.01
	Birth date	0.24	0.06

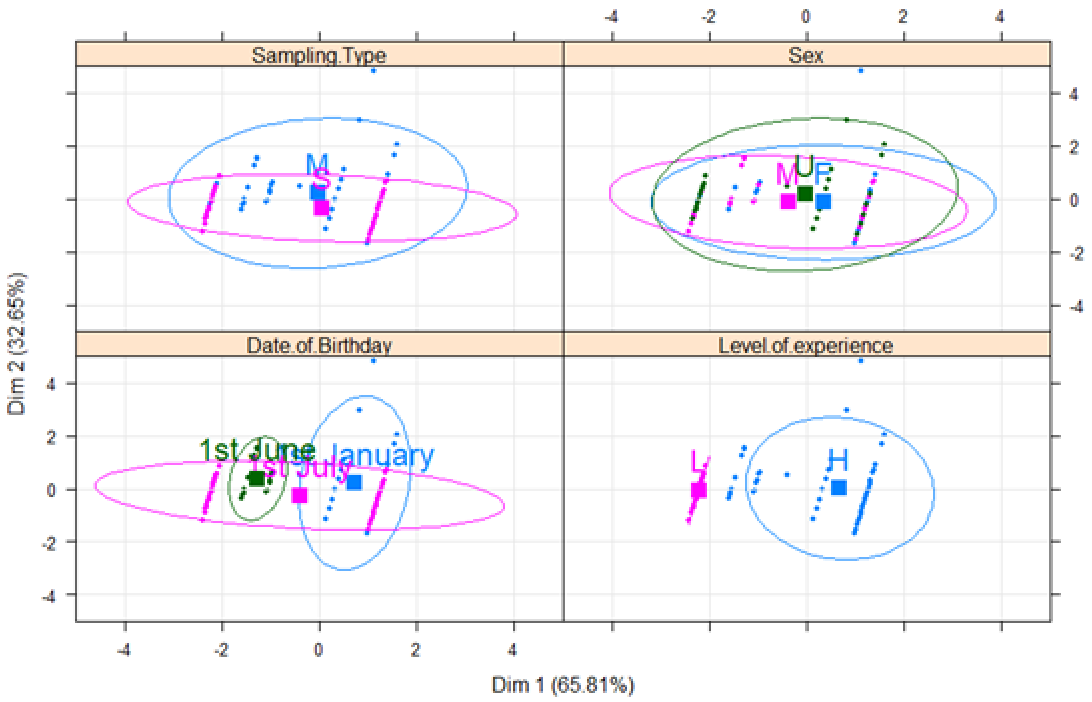
Whole data set



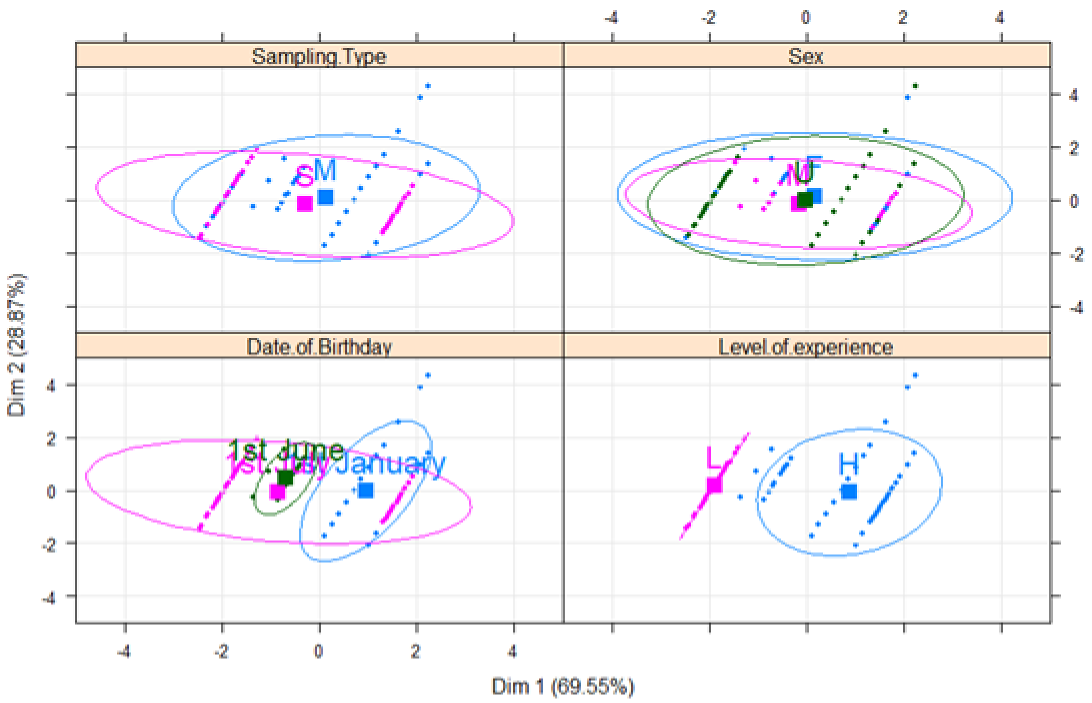
Age 1



Age 2


Sampling.Type
Sex
Date.of.Birthday
Level.of.experience
M
S
U
F
1st June
1st July
January
L
H
Dim 2 (32.65%)
Dim 1 (65.81%)

Age 3


Sampling.Type
Sex
Date.of.Birthday
Level.of.experience
S
M
F
1st June
1st July
January
L
H
Dim 2 (28.87%)
Dim 1 (69.55%)

Age 4

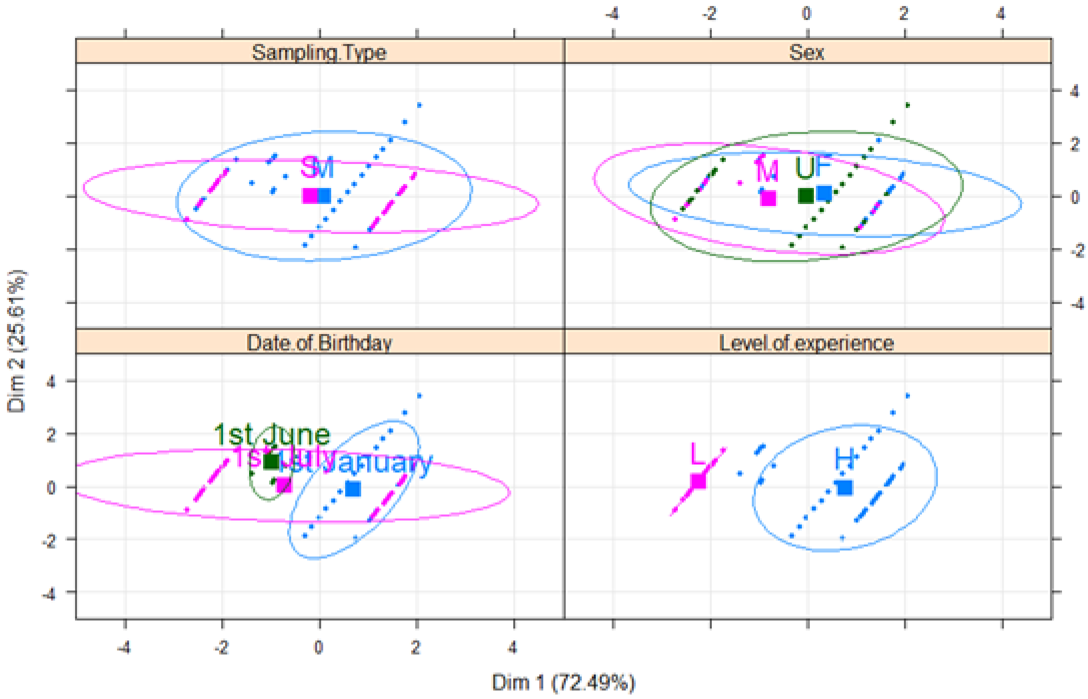


Figure 2.2.1.3 - Confidence ellipses drawn around the levels of the categorical variables considered in each PCA (confidence level = 0.95). Sex F: female; M: male; U: combined. Date of Birthday 1st January, 1st July, 1st June. Level of experiences L: low experience (< 200spines read), H: high experience (> 500spines read). Sampling type S: at sea; M at landing.

2.2.3 Results Interpretation

The results of the present work confirm previous studies on the high variability, occurring in the age and growth of Swordfish (Arocha et al., 2003; Quelle et al., 2014; Abid et al., 2014). This variability can be affected by several sources, such as: sampling methodologies (commercial or survey) (Coggins et al 2013), geographical differences (Abid et al., 2014), age estimation criteria, age estimation scheme, skeletal structures used (otolith or scale) (Farley et al., 2016), methodology (direct age estimation or LFDA-Length Frequency Distribution Analysis) and level of experience of the readers (ICES, 2011; 2013). These factors affect both the accuracy and the precision of the age and growth data producing negative impact on the stock status evaluation and the sustainability of the applied management measures for the Swordfish stocks in the Mediterranean. Most of the stock assessment models used, especially the analytical ones, such as Virtual Population Analysis (VPA) and Statistical-Catch-At-Age (SCAA), require knowledge on the demographic structure of the stocks. One of the first steps to run stock status evaluation is the conversion of the length structure of a stock to the age structure, performed by means of age slicing procedures using growth parameters (e.g. the von Bertalanffy growth formula), or age length keys (ALKs). Inappropriate growth parameters to convert length distribution into age structure or ALKs can lead to unreliable stock status assessment (STECF, 2017). If an age overestimation occurs, the stock assessment will provide an erroneous scenario with a population composed by older individuals and, consequently, affected by lower fishing mortality (F), whereas in the opposite case, fish would be younger with an overestimation of F (Campana, 2001). Moreover, age and growth affect also the estimation of the natural mortality (M) and maturity-at-age data. As a result, they can affect the estimation of recruitment strength and spawning stock biomass. Ultimately, the most important effect is linked to short-term predictions of the stock status and the related management measures (Reeves 2003; Punt et al., 2008; Hüssy et al., 2016; Eero et al. 2015).

The variability of the age results is attributed to several factors, (ICES, 2011, Smith et al, 2016; Hüssy et al., 2016; Anticamara et al., 2011; Kimura and Lyons 1991). In our analysis, we have examined the importance in ageing variability of the age scheme (birth date), the different experience of readers, the geographical differences (latitude and longitude), sampling type (at sea and at landing) and sex.

Our findings revealed high variability in length-at-age for both sexes (Fig. 2.1.1.2). The geographical location was found to be the most important factor, influencing significantly the age variability, with Longitude (west-east) being the factor most highly correlated with variability than Latitude (north-south).

The reader's experience has been identified as an important factor affecting the precision of the age data for many species in both marine and freshwater environments (Smith et al, 2016; Oele et al., 2015; Rude et al., 2013; Kimura and Anderl, 2005; Appelberg et al, 2005; ICES, 2017b). In the present analysis, this factor was also found to be important in ageing variability; especially when we compared the results of High versus Low experience readers. Reader experience emerged as a key issue in estimating the age mostly in the first-year groups as well as the oldest age group (4 years). The identification of the true first growth increment and the overlapping of the growth rings have been mentioned as reasons of disagreement in the Swordfish ageing analysis (ICES, 2009; 2012; 2017).

The theoretical birth date has also been reported as another important element in the process of the age estimation (Morales-Nin and Panfili 2002). In our analysis birth date had a lower influence in the first age group in comparison with the readers experience and sex. In the rest of the age groups, the birth date had a greater influence on the ageing. The specific date of birth at individual and/or population level, as established by studies of reproduction and/or analysis of daily increments, is not always known. Therefore, for convenience during the stock assessment process the conventional birth date for the entire population was established at 1st January (Morales-Nin and Panfili, 2002). The reproduction of the Swordifish in the Mediterranean basin takes place between June to August (Palco et al., 1981). Thus, an age scheme based on the 1st of July as the birthdate of the species, has been suggested as more appropriate avoiding the over estimation of the age in the first year. As a result, considering the 1st of January as birth date, specimens born during the spawning season (June – August) would be aged as 1 year old, even if they are caught after 6 months.

The results of the present analysis have demonstrated the importance of a handbook clarifying and standardizing ageing schemes (e.g. birth date), ageing criteria (e.g. number of false rings before the first winter growth increment) and preparation methods could be a useful action to overcome bias in ageing. The use of a common and standardized protocol by all Institutes is fundamental in order to decrease the relative/absolute bias associated with the activities of age determination and to improve the precision (reproducibility and reduction of the coefficient of variation) of the age readers from the various laboratories, which are involved in the ageing analysis. More importantly, putting all laboratories under a same standardized protocol can ensure the possibility to apply changes horizontally to datasets in case of future breakthroughs and/or ground-breaking discoveries.

Being the reader experience the most important factor in explaining the huge variability in the age data in the Mediterranean basin, workshops, age exercises and exchanges are considered as fundamental tool for improving the precision in the red mullet age analysis (PGCCDBS, 2011). All these actions can be an important contribution to overcoming the ageing uncertainties, thus providing accurate and robust input data for stock assessments.

3. Exchange exercise

The ageing analysis, the examination of the protocols and literature (Rodríguez-Marín et al., 2007; Williams et al., 2013; ICCAT 2006-2016; Lanteri and Garibaldi, 2019; Quelle et al., 2014) on the large pelagic stocks showed some gaps on:

- Ageing scheme;
- Ageing criteria;
- Ageing validation study;
- Preparation method.

These aspects affect both the precision and the accuracy (Panfili et al., 2002) of the age estimation for the selected stocks. To overcome these gaps and improve the precision, workshop and reading exchange (ICES, 2011; ICES, 2013; ICES, 2015) are useful tools, while validation studies are the means to improve the accuracy (Campana, 2001).
In addition, in the case of swordfish, problems in age estimation using spines can be summarized in the following main sources of errors:

• Presence of multiple bands and false bands;
• Progressive disappearance of the inner bands in larger specimens.

The Exchange approach based on supporting tools (SmartDots, Eltink sheet, full scale exchange) (PGCCDBS 2011; ICES 2016, ICES 2017) was utilized to highlight the main source of bias and understand the level of precision of Swordfish

3.1 Sampling Collection and Participation

A preliminary step to the exchange was the collection and calibration on a suitable number of HS images (first three ray of the anal fin). The images of prepared spines have been provided by Genoa University and IEO. In total 79 specimens were sampled from 2003 to 2017 in the Mediterranean area (Tab. 3.1.1; Fig. 1.3.1.1).

Table 3.1.1 - Samples distribution of *Xiphias gladius* by the sampling year and area.

Species	Areas	2003	2004	2005	2007	2008	2009	2010	2011	2016	2017	Tot.
X. gladius	Ligurian Sea (1)				2	1	1	1	3	2	54	64
	Alboran Sea (2)	1	5	9								15
Total		1	5	9	2	1	1	1	3	2	54	79

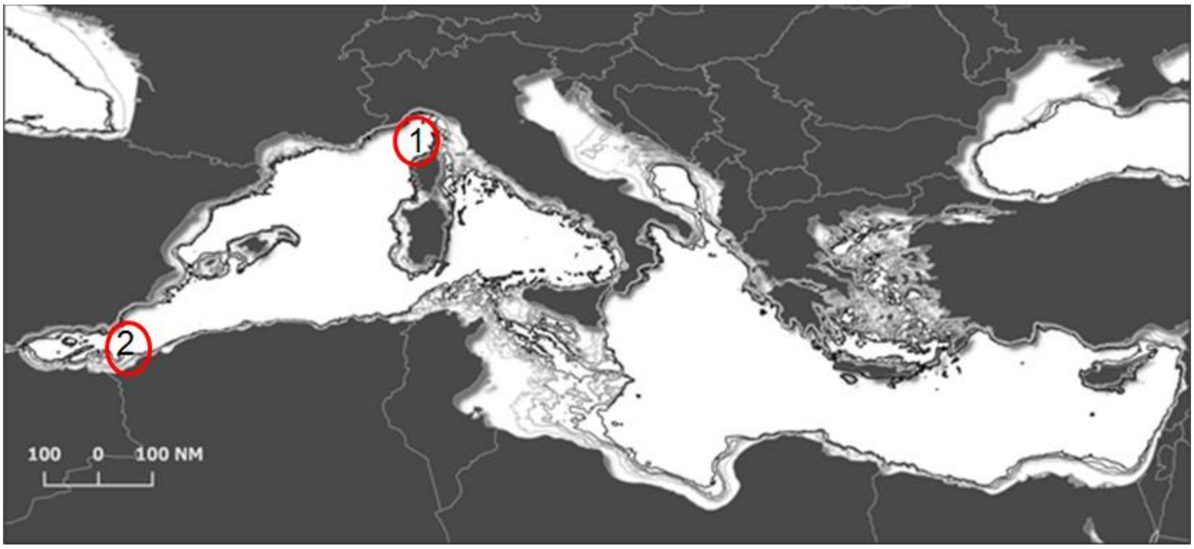


Figure 3.1.1Map of specimens collected: 1Ligurian Sea; 2 Alboran Sea

The length distribution of *X. gladius* (Fig. 3.1.2) there were from 2 different geographical areas. The specimens of Ligurian Sea included the smallest fish below LJFL range included between 69 and 177 cm. Conversely, the fish from Alboran Sea presented the LJFL range from 102to 213 cm (Fig. 3.1.2). In total there were covered a huge range of LJFL that they included juveniles and adult specimens

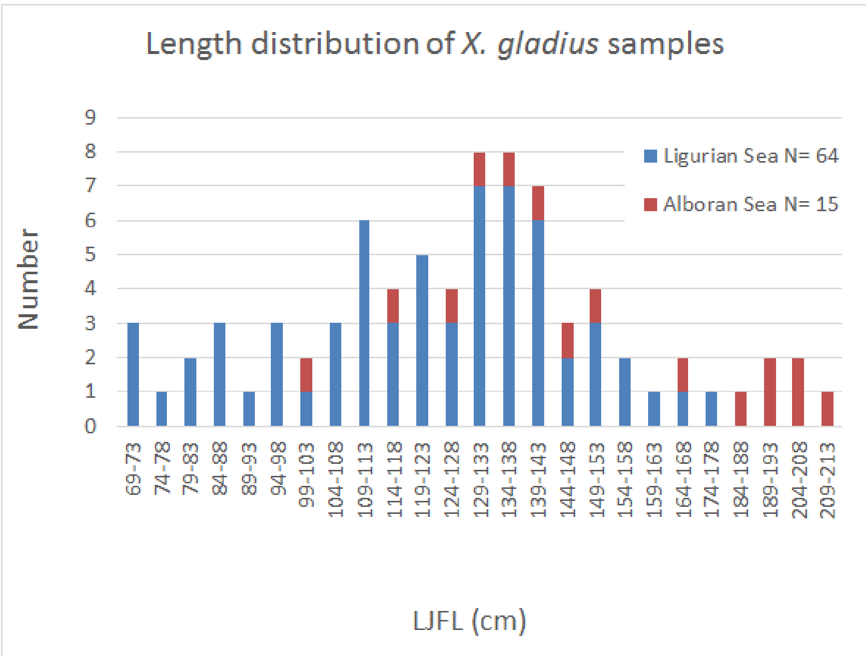


Figure 3.1.2 - Length distribution of *X. gladius* used during the exchange by geographical areas

In total 9 readers participated to the reading exchange exercise from 6 country and 7 laboratories (Tab. 3.1.2). The readers included not only readers from the Institutes involved in the RECOLAPE project but also from others Institution, involved in the DCF.

Table 3.1.2 List of the readers by country and laboratory

Reader	Name	Country	Institution
1	Aurelie Guillou	France	IRD
2	Fulvio Garibaldi	Italy	University of Genoa
3	Sergio Bizzari	Italy	Unimar
4	Daniela Rosa	Portugal	IPMA
5	Luca Lanteri	Italy	University of Genoa
6	Rui Coelho	Portugal	IPMA
7	George Tserpes	Greece	HCMR
8	Ioannis Thasitis	Cyprus	DFMR
9	Pablo Quelle	Spain	IEO

3.2 Reading procedures and data analysis

To all readers were asked to read each digitised image with their own interpretation (positions of the annual rings on a given transect) using the program SmartDOT platform (http://www.ices.dk/marine-data/tools/Pages/smartdots.aspx). SmartDOT is a new set of software tools supports the user in managing all data of ICES age reading workshops and exchanges. The workshop or exchange manager can manage the meta data related to workshops and exchanges, and the age reader can carry out age readings by annotating HS images. All registered data are available in the connected reporting environment.

The instructions, how to use this software in the context of this exchange, are reported in the Annex 1.

The age was assigned taking into account the number of the transparent rings the date of birthday and the edge type. Moreover, the date of capture and the sex were visible by the readers. Then the age for each specimen was assigned following the scheme reported in the Table 3.1.3

Table 3.1.3 - Age scheme used during the exchange

Date of Caprure	Age	Edge
1th semester	N of ring + Internal ring	Trasparent
	N of ring +1 + Internal ring	Opague
2th semester	N of ring + internal ring	Opaque
	N of Ring + internal ring	Trasparent

All data were extracted from SmartDOT and analysed using the GuusEltink spreadsheet (Eltink, 2000). The spreadsheet (Eltink, 2000) was completed according to the instructions contained in Guidelines and Tools for Age Reading Comparisons by Eltink et al. (2000). Modal ages were calculated for each spine red, with percentage agreement (PA), coefficient of variation (CV) and average percent error (APE), as a definition (for each spines):

$$PA = \frac{\sum |n_{diff} \leq 1|}{n}$$

$$CVj(\%) = 100.\frac{\sqrt{\sum_{i=1}^{R} \frac{(X_{ij}-X_j)^2}{R-1}}}{x_j}$$

Where R is the number of times each fish is aged, Xij the i(th) age determination of the j(th) fish, Xj is the mean age calculated for the j(th) fish, and ndiff is the difference in age determination between the readings of two readers.

$$APE\text{j}(\%) = 100.\frac{1}{R}\sum_{i=1}^{R} \frac{|\text{Xij} + \text{Xj}|}{\text{Xj}}$$

Where xij is the ith age determination of the jth fish, x j is the average age calculated for the jth fish and R is the number of times each fish was aged.

3.3 Results

In the analysis were utilized the data from all readers and the precision analyse with CV, APE and percent of agreement to modal age for *X. gladius* spines sets was presented in the Table 3.3.1. All data showed the low precision with the percent agreement between 52.7 and 67.2%, the CV from 33.9 to 17.8% and the APE from 22.7 and

24.4%. For the all samples together the CV, APE and percent of agreement to modal age were respectively: 30.8%, 23 and 64.4%.

Table 3.3.1 - Reading's precision for *X. gladius* by sampling area

Species	Geographical area	Otoliths Spine	Length LJFL Range (cm)	Age range (year)	Percentage of Agreement	CV	APE
X. gladius	Ligurian Sea ITA	64	69/177	0/7	67.2%	33.9%	22.7%
	Alboran Sea SPA	15	160/213	1/9	52.7%	17.8%	24.4%
	TOTAL	**79**	**69/213**	**1/9**	**64.4%**	**30.8%**	**23%**

Moreover the precision indices (PA, CV and APE) not showed significant differences (Kruskal–Wallis test; p>0.05) if they were stratified by readers' experience (Expert >500 spines read; Basic < 500 spines read) (Tab. 3.3.2).

Table 3.3.2 - Reading's precision for *X. gladius* by sampling area

Species	Expertise	Percentage of Agreement	CV	APE
X. gladius	Expert	65.1%	30.9%	21.6%
	Basic	64.2%	30.5%	24.7%
	TOTAL	**64.4%**	**30.8%**	**23%**

The coefficient of variation (CV), percent agreement and the standard deviation (STDEV) are plotted against MODAL age (Fig. 3.3.1). The results show a decrease trend from the lower age groups to the higher one for PA and STDEV and the opposite trend for the CV. These results could be explained by the position of the first growth increment (Quelle et al., 2014) and the overlapping the growth increments in the older specimens (Lanteri and Garibaldi, 2019). In general, after the first age groups was observed a decrease of the agreement, the increment of STDEV and a constant CV around the 20%.

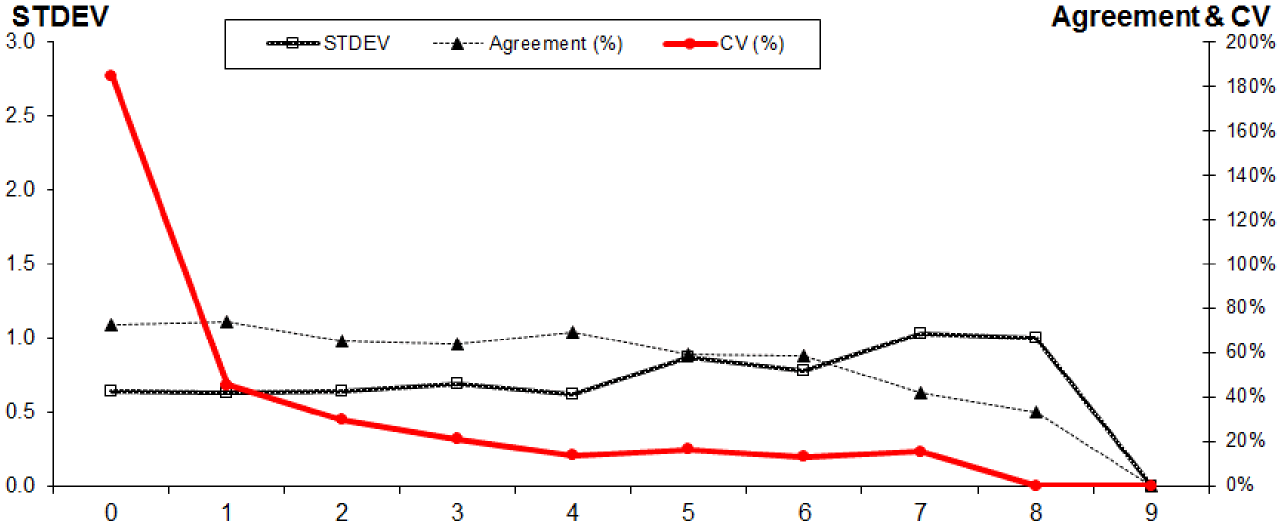


Figure 3.3.1 - The coefficient of variation (CV), percent agreement and the standard deviation (STDEV) are plotted against MODAL age.

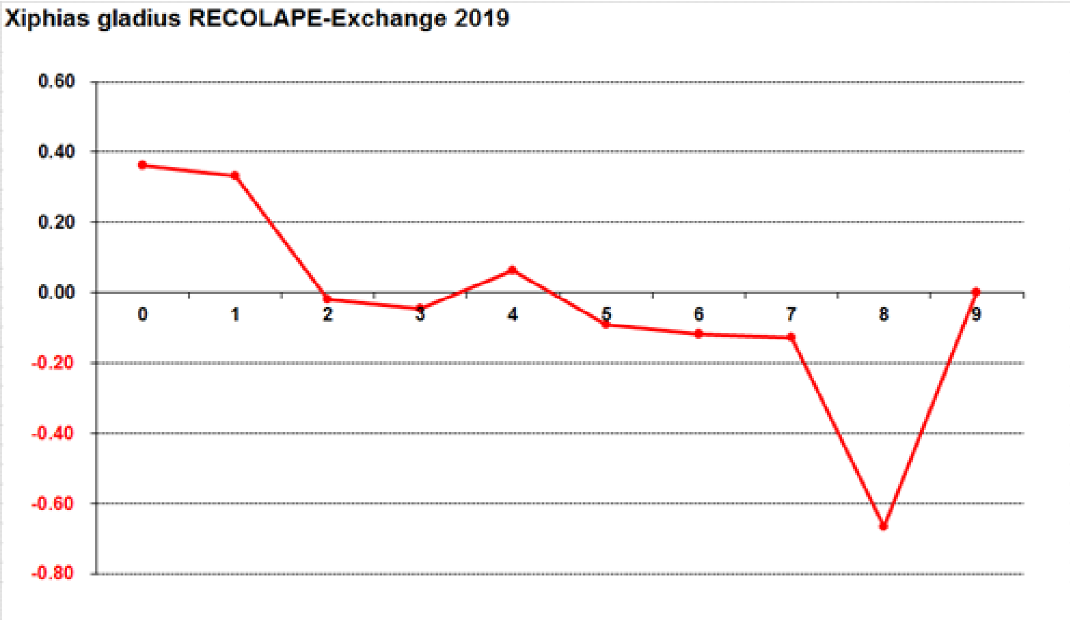
The percentage of agreement by readers weighed by the number of samples read are included between 38% to 79.5% (Table 3.3.3). Moreover, the PA by age group shows a negative trend passing from 72% for the age 0 to 33% for the age 8.

Table 3.3.3 Percentage of agreement by readers and age group.

	PERCENTAGE AGREEMENT										
	MODAL age	France AG	Italy FG	Italy SB	Portugal DR	Italy LL	Portugal RC	Greece GT	Cyprus IT	Spain PQ	ALL
	0	0%	100%	75%	100%	100%	0%	100%	75%	100%	72%
	1	0%	100%	100%	67%	78%	56%	89%	78%	100%	74%
	2	50%	33%	45%	91%	75%	83%	100%	73%	36%	65%
	3	36%	68%	57%	77%	68%	77%	74%	68%	50%	64%
	4	33%	60%	57%	87%	93%	87%	70%	77%	60%	69%
	5	50%	60%	40%	80%	80%	80%	63%	50%	30%	59%
	6	100%	50%	100%	50%	50%	100%	0%	50%	0%	59%
	7	75%	25%	0%	50%	25%	25%	0%	67%	75%	42%
	8	100%	0%	0%	100%	0%	0%	0%	100%	0%	33%
	9	-	-	-	-	-	-	-	-	-	-
Weighted mean	0-9	38.0%	62.0%	58.1%	79.5%	74.7%	70.9%	75.4%	69.3%	55.1%	64.6%

Relative bias can be defined as a systematic over- or underestimation of age compared to the modal age. In the results of the exchange the bias is higher in the first two age groups (age 0 and age 1) reaching about 0.4 year and in the last age group where the bias reach about 0.6 year (Fig. 3.3.2).

Figure 3.3.2 - The RELATIVE bias by MODAL age as estimated by all age readers combined



The hypothesis of an absence of bias between two readers or between a reader and the modal age estimated was tested non-parametrically with a one-sample Wilcoxon signed rank test. The results of the test (Fig. 3.3.3) highlighted that there is a group of readers that not show significant difference among them and with modal age.

Inter-reader bias test and reader against MODAL age bias test

	France AG	Italy FG	Italy SB	Portugal DR	Italy LL	Portugal RC	Greece GT	Cyprus IT	Spain PQ
Reader AG									
Reader FG	**								
Reader SB	**	—							
Reader DR	**	*	**						
Reader LL	**	**	**	—					
Reader RC	**	*	**	—	—				
Reader GT	**	*	—	—	—	*			
Reader IT	**	**	**	—	—	—	*		
Reader PQ	**	—	—	**	**	**	**	**	
MODAL age	**	*	**	—	*	—	—	—	**

— = no sign of bias (p>0.05)
* = possibility of bias (0.01<p<0.05)
** = certainty of bias (p<0.01)

Figure 3.3.3 - Inter-reader bias test and reader against modal age bias test of *X. gladius* spines.-: no sign of bias (p>0.05); *: possibility of bias (0.01<p<0.05); **: certainty of bias (p<0.01)

11 images of the all sample (79 images) presented an agreement ≥ 80% (Tab. 3.3.4). These are from the lower age groups (age group 1, 3 and 4) and they could be represented the base for the age reference collection of the swordfish spines.

Table 3.3.4 – The number of images with an agreement ≥ 80% by modal age.

Criterion 80% agreement	
MODAL AGE	n
0	0
1	4
2	0
3	3
4	4
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
	11

Plotting the mean length by age group and readers (Fig. 3.3.4) seems clear that the mean length of the first 6 age groups (from age 0 to age 5 years) are comparable for the mostly of the readers. So this could be explained by the relative easiness to recognize the first growth increments.

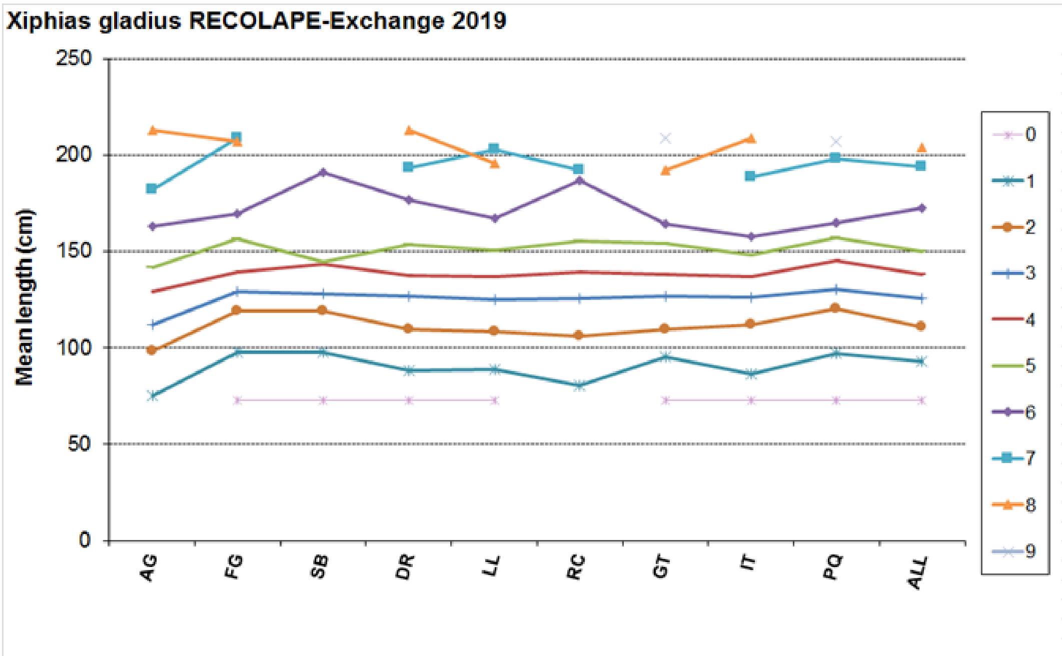


Figure 3.3.4 - The mean length at age as estimated by each age reader.

3.4 Remarks

The exchange exercise was based on a total of 79 fish sampled from 2003 to 2017 in Mediterranean from 2 sites sample: Ligurian sea and Alboran Sea. The pictures of HS (spines of the anal fin) tin section were prepared in the same way (Quelle et al., 2014; Lanteri and Garibaldi, 2019). The overall precision are PA, CV and APE respectively of 64.4%, 30.8% and 23%. These values are respectively lower and higher than those considered acceptable: 80% PA and 20% CV (PGCCDBS 2011). Moreover, they were not significantly different if they were stratified by readers' experience, so this factor not explained fully the low PA and high CV reach in this exchange exercise. The analysis of the precision indices by age groups showed a negative trend from the first age group to the older one. In addition, the bias analysis on the all data seems highlight an under-estimation for the older age group, while an overestimation for the first age group (0 and 1 year). These results could be explained by the difficult to recognize the first growth increment and mostly growth increments (overlapping of the rings) in the older fish (age > 5 years).

The comparison of the age readings among the readers and each reader with modal age highlighted that a group of readers follow a same age criteria. These results are confirmed also of the mean length at age as estimated by each age reader. Indeed, in the first 6 age groups (from age 0 to age 5 years) the mean length at age are comparable for the mostly of readers. All these results were discussed during the next workshop.

4. Workshop on age reading of *Xiphias gladius*

The Workshop on age reading of *Xiphias gladius* expected in the context of the RECOLAPE was held in Olhão (Portugal) from the 2th to the 4th April 2019. The meeting was host by IPMA Institute. Eight age readers from 4 countries and 5 laboratories (IPMA, IEO Santander, Genoa University, Unimar, IRD) participated in this workshop (Tab. 4.1). The agenda is presented in the Annex 2.

Table 4.1 – List of participants

Name	Affiliation Country	e-mail	Exchange
Pablo Quelle	IEO - Spain	pablo.quelle@ieo.es	YES
Sergio Bizzarri	Unimar - Italy	sergiobizza@alice.it	YES
Rui Coelho	IPMA - Portugal	rpcoelho@ipma.pt	YES
Fulvio Garibaldi	Genoa University - Italy	largepel@unige.it	YES
Daniela Rosa	IPMA - Portugal	daniela.rosa@ipma.pt	YES
Aurelie Guillou	IRD - France	aurelie.guillou@ird.fr	YES
Pedro Lino	IPMA - Portugal	plino@ipma.pt	NO
Pierluigi Carbonara	COISPA - Italy	carbonara@coispa.it	NO

Figure 4.1 Participants to the workshop from left to right: Pablo Quelle, Rui Coelho, Sergio Bizzarri, Fulvio Garibaldi, Daniela Rosa, Aurelie Guillou, Pierluigi Carbonara, Pedro Lino.

The Term of Reference (ToR) of the meeting are following listed:

ToR a: Preparation method;
ToR b: Age scheme and Age Criteria;
ToR c: Analysis of the Exchange exercise;

ToR d: Develop a reference collection of spines4.1 Preparation method

During the meeting all laboratories involved in the sampling and ageing analysis presented the preparation methods used routinely in each laboratory. Here they are described the agreed methods after the discussion the pros and cons for each lab procedures.

4.1.1 Sampling the anal fin

The anal fin must be cut close to the body trying to keep the condyle base in the fin part. Once the fin is collected, we must conserve freezer until the processing (Fig. 4.1.1.1).

Figure 4.1.1.1 - Extraction of the 2nd anal fin spine of *X. gladius* (Lanteri and Garibaldi, 2019)

If the fin is too big, or there is some problem to keep the whole sample, it is possible to cut it to reduce the size, but we must collect at least the 4th first radios. LJFL, sex, date of capture, fishing gear and situation are recommended as additional information. If it is possible macro maturity must be taken too.
The recommendation is using the second ray of the anal fin when the fin type morphology is the A type (Fig. 4.1.1.1).

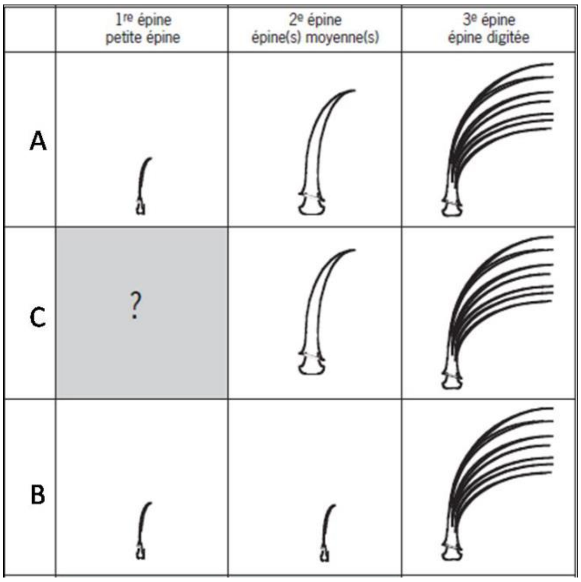


Figure 4.1.1.1 - Fin type configuration according to the first three radii by order of frequency appearance (Quelle et al., 2014)

After putting into boiled water, for short periods, to help to remove all the flesh and cleaned the second anal fin ray can be dried and stored in paper bags until the cut.

4.1.2 Section location

Two possible cuts were discussing. Finally, according to the results showed that the best one and recommended by the group is 1d, optionally 0.5d can be used to analysis too. As shows the Figure 4.1.2.1.

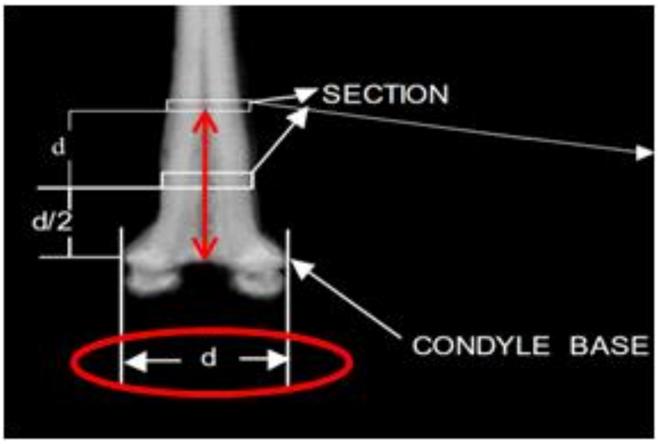


Figure 4.1.2.1 - Detail of the second anal fin ray of a swordfish, showing the maximum diameter of the condyle base (d) as the measurement to locate the sections. (Adapted from Sun et al., 2002).

The sections can be prepared in different thickness, but the most common, and the recommendation of the group is from 0.4 to 0.6mm. Usually the spines can be cut without resin inclusion.

4.1.3 Measurements

The definition of the section focus was discussed about to continue with the defined by Erhardt et al. (1996), the point where the radial striations appear to converge, or change to the line proposed to use by Quelle et al. 2004. The line was finally recommended as the best option.

The focus line is the reference line which connects the two innermost ends of the lobes (margins) of the structure (Fig. 4.1.3.1).

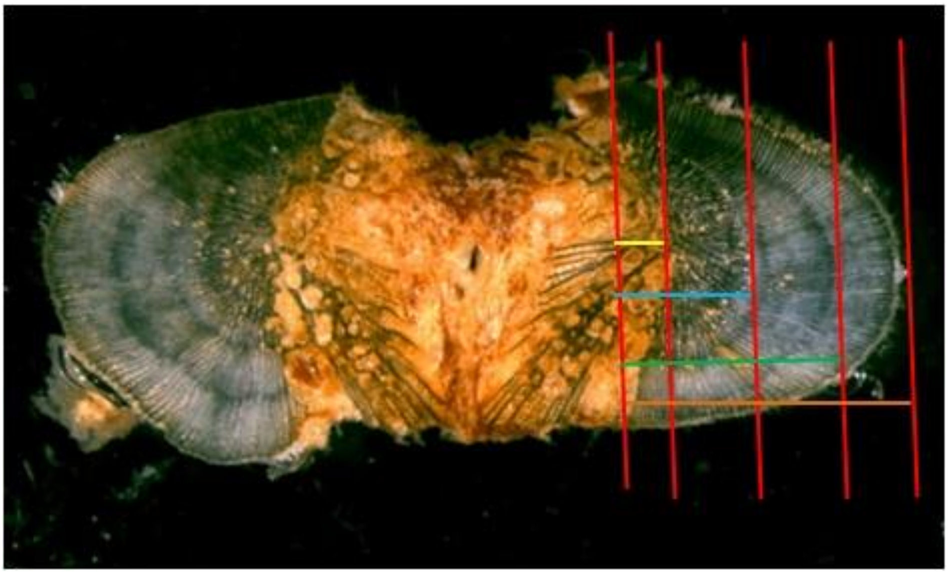


Figure 4.1.3.1 - Section of the second anal fin ray of Swordfish. The red vertical line in the inner part of the lobe is the focus line. The line from where all the measurements are taken. The yellow distance is the pre-growth structure. The blue distance is the inner resorption distance. The green distance is an example of the 2 years old annulus measurement. The orange distance is the maximum width of the lobe.

For measurements, we create parallel lines on the focus and put them in the end of the structure that we want to take the measurement. Parallel lines are created from the focus. In the Figure 4.1.3.1 the are reported an example how take the measurements in standardised way. To help in the creation of these measurements, any software can be used. At this time, a private software, Niss elements, was used. The group recommend look for some free option. The group enhance too that the Smartdot platform be adapted to this protocol and be useful for tuna and tuna like species spines too.

In aim to understand better the process that happens in the lobe, other measurements are required (Fig. 4.1.3.1):

- Pre-growth structure. Distance between the focus and the area where the growth process starts.
- Inner resorption distance. Maximum distance between the focus and the end of the vascularised area (the area damaged by the effect of the growth of the structure), where the growth tissue starts to be visible.
- Maximum width of the lobe. The maximum distances until the end of the selected lobe.

- Annulus measurements. The line to take measurement must be placed in the end of the translucent band.

In order to avoid the variability between the structures, all measurements must be adjusted to the distances only in the growth area. Due to this, we can talk about:

- Vascularisation. It was quantified as the inner resorption minus the pre-growth structure distance.
- Adjusted annulus measurement. It is the annulus size minus the pre-growth structure distance.

Regarding the light type the sections can be observed and photographed using reflected or transmitted light. As a general suggestion, transmitted light seems to be the best option.

4.2 Age scheme and Age criteria

The age scheme should take into account several elements. During the workshop each of them were assessed and following they were reported the conclusions of the discussion.

- Date of capture - For each sample it is fundamental to know the exact date of capture of the specimen, in order to better define the age.
- Date of birthday - For each sample it is fundamental to know the birthday in order to better define the age. However, usually a common reference date is used for all samples (assumed birthdate). Specifically, for the Mediterranean the birthdate has been described as occurring in the 1st of July, as the spawning period goes from June to late August/September, and the peak of spawning occurs in July (Palco et al., 1981). In the Atlantic the spawning period is prolonged the whole year (Neilson et al. 2013) and it might be more difficult to define a birthday for the Atlantic stocks (North and South).
- Ageing resolution - Given the longevity of the species, a resolution of one year seems to be adequate.
- Band enhancement - Readings should be conducted on images, photographed using a digital camera mounted on a microscope, testing different light levels is recommended to achieve the best band contrast possible. Enhancing images after photographing is also possible using an imaging analysis software, however, it was noted that enhancing also needs to be standardized so that different people can enhance the images in a similar manner (Rodriguez-Marin et al. 2019).

- Given that annual rings are composed by one opaque and one translucent band, dot/measurements should be put/taken at the end of the translucent bands.
- Identification of the 1st annulus/band - this is one of the most important issues, given the very well-known problems created by the reabsorption of the inner bands in the vascularized area. For this purpose, it is important to take measurements of the 1st visible band, following the criteria described above. It was recommended that it could be useful to create a reference stick/distance, in order to have an average width of the first annulus/band. This should be made using sections collected from juveniles, to be sure of the age and distance. After that, applying this reference distance, it will be possible to have a better idea of the rings that could be have been reabsorbed.
- Annual bands should be measured (measurements shall be taken from the focus line as explained above). It was discussed the possibility to create a template or at least a measurements reference table for a certain number of years (e.g. from age 1-4), as made for bluefin tuna, but it was underlined that this procedure could be more difficult to apply to swordfish spines, given their high variability in band deposition pattern and non-symmetrical spine growth. It is also important, that for vascularized spines, to identify the first translucent band that is clearly defined and from which measurements can be taken.
- Edge type - Edge can be described as translucent or opaque, or in more detail by characterizing the amount of translucent or opaque zones (e.g. wide/narrow).
- Age Scheme - age from band counts sometimes does not correspond to the biological age of the fish, different age adjustments were discussed that could be used for the Mediterranean swordfish. For these adjustments, it is necessary to know at least the date of capture and the birth date, but also the edge type and time of translucent band deposition (in this case translucent bands are being used for counts). For swordfish in the Mediterranean the translucent band is formed in the earl spring months (Garcia et al., 2016;). Several ageing schemes that could be used are presented below.

1st of January birthday

Date of Capture	Age	Edge
1st Semester	Nr of translucent rings +1	Translucent
	Nr of translucent rings	Opaque
2nd Semester	Nr of translucent rings	Translucent
	Nr of translucent rings	Opaque

Date of Capture	Age	Edge
1st Semester	Nr of translucent rings +1	Translucent
	Nr of translucent rings	Narrow Opaque
	Nr of translucent rings +1	Wide Opaque
2nd Semester	Nr of translucent rings	Translucent
	Nr of translucent rings	Narrow Opaque
	Nr of translucent rings	Wide Opaque

1st of July birthday

Date of Capture	Age	Edge
1st Semester	Nr of translucent rings	Translucent
	Nr of translucent rings -1	Opaque
2nd Semester	Nr of translucent rings	Translucent
	Nr of translucent rings	Opaque

Date of Capture	Age	Edge
1st Semester	Nr of translucent rings	Translucent
	Nr of translucent rings - 1	Narrow opaque
	Nr of translucent rings	Wide opaque
2nd Semester	Nr of translucent rings 1	Translucent
	Nr of translucent rings	Narrow opaque
	Nr of translucent rings	Wide opaque

After adjusting for edge type, it is also possible to adjust for capture date, by adding or subtracting (depending on the date of capture) to the age the proportion of time that has passed since the birth date. A more detailed adjustment taking into consideration the width of the band is also possible (e.g. see Farley et al., 2013).

Ages can also be adjusted to capture date without consideration of the edge type, for example when the birthdate of the species is not known or has a very wide spawning season. The age counts are incremented or decreased based on the proportion from birthdate to the capture date (e.g assuming a fish is born in the 1st of July 2018, if it caught in the 1st of June 2019 will be 0.9 while if it caught in the 1st of august it will be 1.08 years.)

The appropriateness of using one adjustment scheme rather than other was also discussed, if the objective is to construct an age-length-key the adjustment to the 1st of January was considered more useful, while for growth curves the adjustment to the 1st of July was considered more suitable as it takes into account the biology of the species.

Regarding the distinction and recognition of the winter rings and therefore of the count of the annuli for the age determination an agreement was reached to adopt two basic criteria:

- the presence of the rings on both sides of the section;
- the distance between the transparent ring (winter)should decrease proportionately.

4.3 Analysis of the Exchange exercise and Smartdot notes

The analysis of the results of exchange exercise is reported in the section 3 of the present document. Anyway, some notes regarding the use of the new tools for the exchange as Smartdot are reported here as a document of the discussion.
The Smartdot is a very useful tool in order to develop exchanges and intercalibrations between laboratories. From our point of view, the use of this tool must be supported and encourage.
In our experience, Smartdot has not been the best tool. It must be adapted for its use in spines, particularly tuna and like-tuna species. The special morphology of these structures has to be taken into account and some new utilities should be developed.

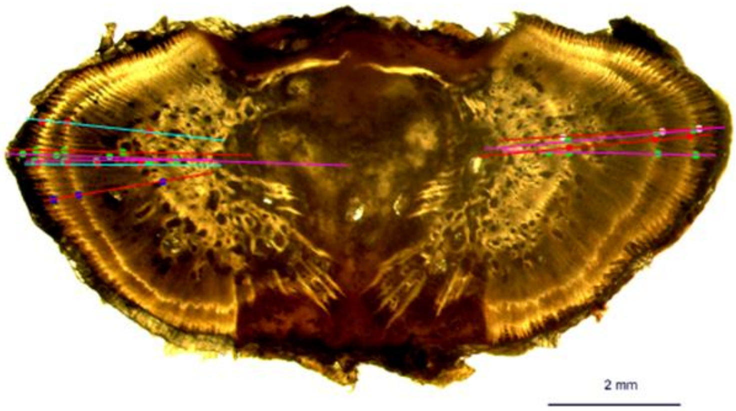
- It would be very useful a new measurement system. The reader must be able to create a line and replicate parallel lines to this first one. The measurements must be the distances between these two lines.
- It would be very useful create items (measurements) that are not considered like annuli. It would give the possibility to mark different structures that can be important in the readings (e.g. false annulus, end of vascularisation).
- The readers must be able to know the measurements. Due to the resorption in the structure, some rings can be masked. To avoid the problem of these lost rings, if the reader knows the measurement, he can be able to know how much annulus have been lost.
- Consequently, with the previous point. Reader can have an option to assign the "real" age to the first ring that is read/marked.

4.4 Reference collection of spines

A key outcome of the fish ageing workshop is the need to develop an annotated reference set of representative Swordfish spines samples for training purposes and to maintain consistency within and among laboratories (Secor et al., 2013; ICES 2011). Ideally, a reference set would be composed of known-age samples (Campana2001), however due to the inherent difficulty in acquiring known age samples of wild fish, many groups have developed reference sets based on consensus ages accomplished by a group of experts. Accordingly, a reference set of 11spines was developed among the spines that they reached a PA ≥ 80%. The objectives of this reference collection (1) to evaluate age estimates from the Swordfish reference set for bias and precision (2) compare age estimates and deposition pattern from the reference collection to each laboratories and (3) suggest future work for production ageing of Swordfish.

SWO024 (EventID:210)

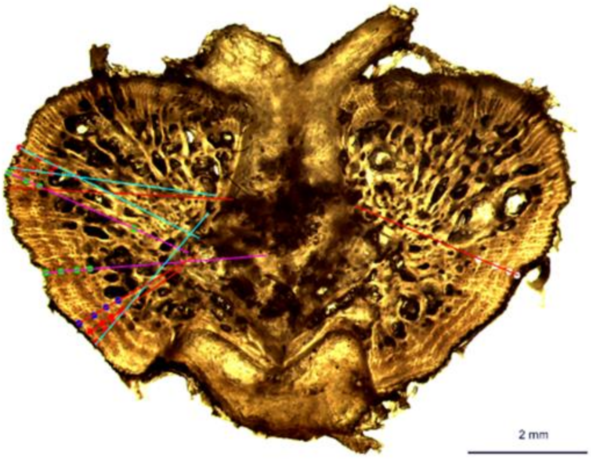
FishID: F
Fish Lenght: 120
Sex:
Area:
Fish Weight:
Catch date: 18/08/2017



Modal Age: 3 years; Agreement: 88.9%

SWO027 (EventID:210)

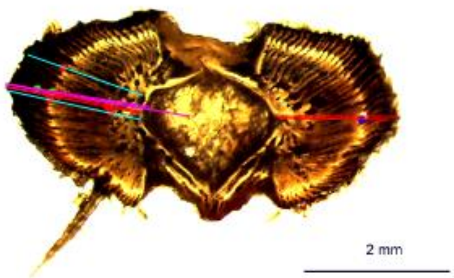
FishID: M	Sex:	Fish Weight:	
Fish Lenght: 133	Area:	Catch date: 25/08/2017	



Modal Age: 4 years; Agreement: 88.9%

SWO004 (EventID:210)

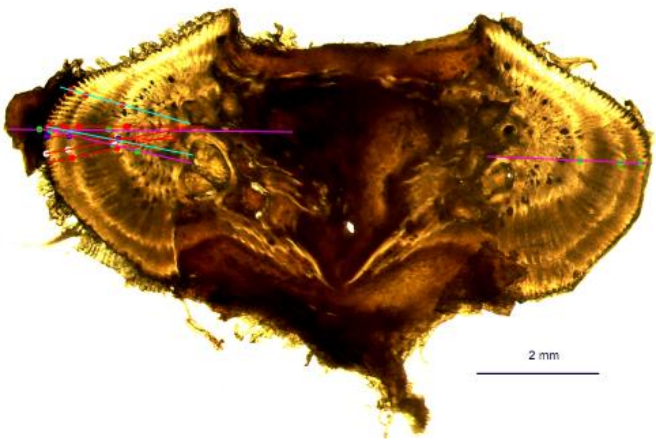
FishID: F	Sex:	Fish Weight:
Fish Lenght: 87	Area:	Catch date: 09/07/2017



Modal Age: 1 years; Agreement: 88.9%

SWO005 (EventID:210)

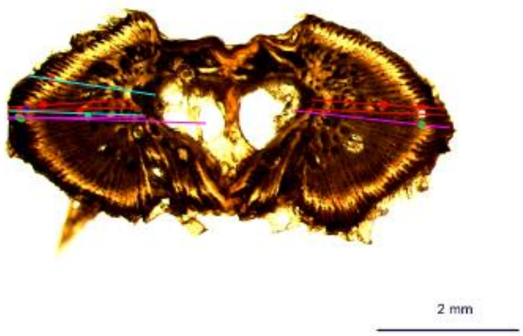
FishID: F
Fish Lenght: 120
Sex:
Area:
Fish Weight:
Catch date: 17/07/2017



Modal Age: 3 years; Agreement: 88.9%

SWO006 (EventID:210)

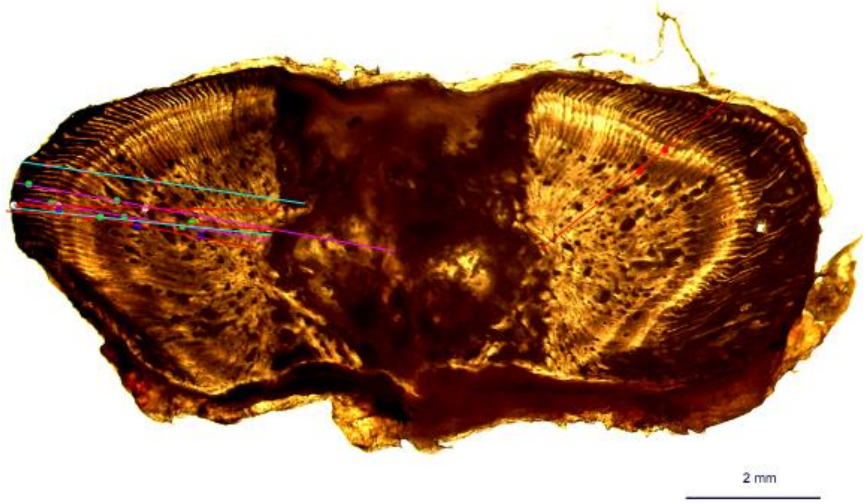
FishID: F
Fish Lenght: 81
Sex:
Area:
Fish Weight:
Catch date: 17/07/2017



Modal Age: 1 years; Agreement: 88.9%

SWO032 (EventID:210)

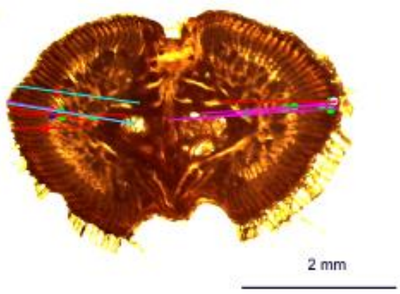
FishID: M	**Sex:**	**Fish Weight:**
Fish Lenght: 130	**Area:**	**Catch date:** 22/08/2017



Modal Age: 3 years; Agreement: 87.5%

SWO036 (EventID:210)

FishID: F	**Sex:**	**Fish Weight:**
Fish Lenght: 83	**Area:**	**Catch date:** 22/08/2017



Modal Age: 1 years; Agreement: 88.9%

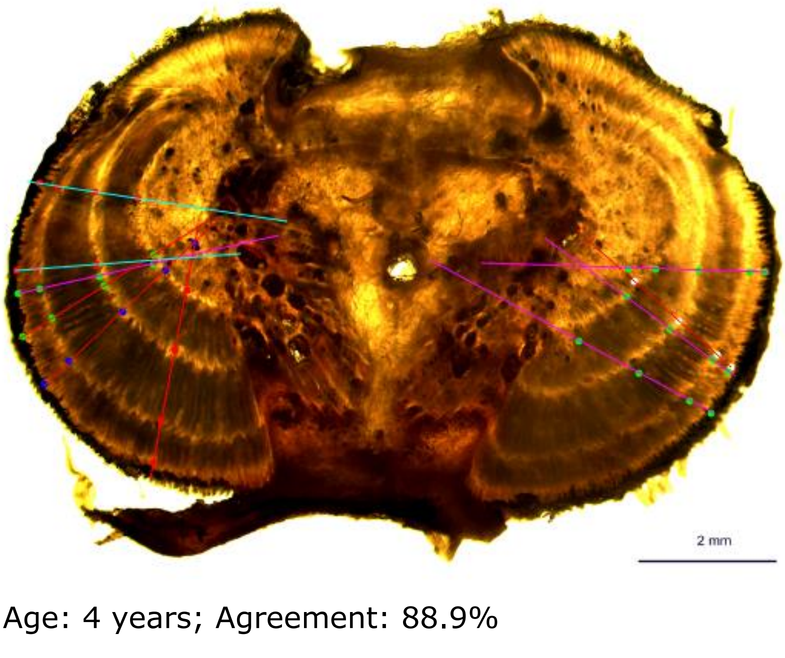
SWO037 (EventID:210)

FishID: F		**Sex:**	**Fish Weight:**
Fish Lenght: 142		**Area:**	**Catch date:** 28/08/2017

Modal Age: 4 years; Agreement: 87.5%

SWO043 (EventID:210)

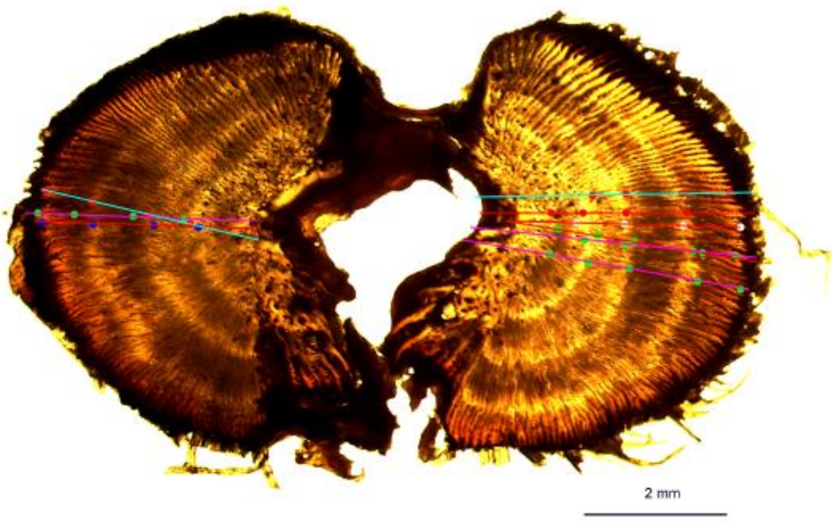
FishID: F		**Sex:**	**Fish Weight:**
Fish Lenght: 142		**Area:**	**Catch date:** 07/09/2017



Modal Age: 4 years; Agreement: 88.9%

SWO064 (EventID:210)

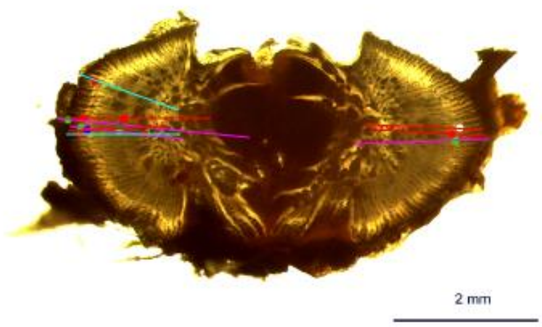
FishID: M	**Sex:**	**Fish Weight:**
Fish Lenght: 135	**Area:**	**Catch date:** 13/06/2011



Modal Age: 4 years; Agreement: 88.9%

SWO008 (EventID:210)

FishID: F	**Sex:**	**Fish Weight:**
Fish Lenght: 87	**Area:**	**Catch date:** 29/07/2017



Modal Age: 1 years; Agreement: 88.9%

References

Abdi H., Williams L. J. 2010. Principal component analysis. Wiley interdisciplinary reviews: computational statistics, 2(4), 433-459.

Abid N., Bakkali M., Tserpes G., Idrissi 2014. Swordfish growth pattern in the strait of Gibraltar; implications for Atlantic and Mediterranean stock mixing Medit. Mar. Sci., 15/1, 135-144

Anticamara J.A., Watson R., Gelchu A., Pauly D. 2011. Global fishing effort (1950–2010): Trends, gaps, and implications. Fish. Res. 107: 131–136.

Appelberg M., Formigo N., Geffen A.J., Hammer C., McCurdy W., Modin J., Moksnessf E., Mosegaard H.,. Morales-Nin B., Troadec H., Wright P. 2005. A cooperative effort to exchange age reading experience and protocols between European fish institutes. Fish. Res. 76: 167–173

Arocha F., Moreno C., Beerkircher L.,. Lee D.W and MarcanL. 2003. Update on growth estimates for swordfish, Xiphias gladius, in the northwestern Atlantic Col. Vol. Sci. Pap. ICCAT, 55 (4): 1416-1429

AA.VV. 2011 Manual for the Baltic International Trawl Surveys, ADDENDUM 1: WGBIFS BITS Manual 2011. ICES web site:
http://datras.ices.dk/Documents/Manuals/Manuals.aspx
AA.VV. NSIBTS indices calculation procedure ICES web site:
http://datras.ices.dk/Documents/Manuals/Manuals.aspx

AAVV, 2014. Italian National Program Under Council Regulation (EC) N° 199/2008 and Commission Regulation (EC) N° 665/2008 National Program 2011-2013 December 2010 extended 2014-2016 (EU decision n. 5568 30/08/2013 and directorial Decree n. 0011911 22/05/2014)

Beamish R. J., McFarlane G. A. 1995. Recent developments in fish otolith research (edsSecor, D. H., Dean, J. M. &Campana, S. E.) (University of South Carolina Press, Columbia, 1995) 545–565

Baird J.W., 1983 - A method to select optimum numbers for aging in a stratified random approach. In Sampling commercial catches of marine fish and invertebrates. Edited by W.G. Doubleday and D. Rivard. Can. Spec. Publ. Fish. Aquat. Sci. 66: 161–164.

Campana, S. E. 2001. Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. J Fish Biol, 59(2):197-242.

Coggins L. G., Gwinn D. C., Allen M. S. 2013. Evaluation of Age–Length Key Sample Sizes Required to Estimate Fish Total Mortality and Growth. Trans. Am. Fish. Soc. 142: 832–840

Carbonara, P., Follesa, M.C., eds. 2019. Handbook on fish age determination: a Mediterranean experience. Studies and Reviews. No. 98.
Rome, FAO. 2019. 180 pp.

Eero M., HjelmJ.,Behrens J., Buchmann K., Cardinale M., Casini M., Gasyukov P., Holmgren N., Horbowy J., Hüssy K., Kirkegaard E., Kornilovs G., Krumme U., Köster F. W., Oeberst R., Plikshs M., Radtke K., Raid T., Schmidt J., Tomczak M. T., Vinther M., Zimmermann C., Storr-Paulsen M. 2015. Eastern Baltic cod in distress: biological changes and challenges for stock assessment. ICES J. Mar. Sci. 72: 2180–2186.

Ehrhardt, N.M., Robbins R.J and Arocha, F. 1996. Age validation and growth of swordfish, Xiphias gladius, in the northwest Atlantic. Collect. Vol. Sci. Pap. ICCAT 45(2): 358-367.

Eltink, A.T.G.W. 2000.Age reading comparisons. (MS Excel workbook version 1.0 October 2000) Internet: http://efan.no.

Farley, J.H., Williams, A.J., Clear, N.P., Davies, C.R., Nicol, S.J. 2013. Age estimation and validation for South Pacific albacore Thunnusalalunga. Journal of Fish Biology, 82:1523-1544.

Farley J., Clear N., KolodyD., Krusic-Golub K., EvesonP., Young J. 2016 Determination of swordfish growth and maturity relevant to the southwest Pacific stock. CSIRO Oceans & Atmosphere, Hobart, 114 pp.

Gulland, J.A. (1966). Manual of sampling and statistical methods for fisheries biology. Part 1. Sampling methods. FAO Man.Fish.Sci., (3): 87 p

Garcia A., Tserpes G., Santos N. M. 2016. Validation of annulus formation and growth estimation of South Atlantic swordfish. Journal of the Marine Biological Association of the UK -1(07):1-8.

Hüssy K., Radtke K., Plikshs M., Oeberst R., Baranova T., Krumme U., Sjöberg R., Walther Y., Mosegaard H. 2016. Challenging ICES age estimation protocols: lessons learned from the eastern Baltic cod stock. ICES J. Mar. Sci. 73(9): 2138–2149.

ICCAT. 2006-2016. ICCAT Manual. International Commission for the Conservation of Atlantic Tuna. In: ICCAT Publications [on-line]. Updated 2016.

ICES. 2011. Report of the Workshop of National Age Readings Coordinators (WKNARC), 5-9 September 2011, Boulogne-sur-Mer, France. ICES CM 2011/ACOM:45. 175 pp.

ICES. 2013. Report of the Second Workshop of National Age Readings Coordinators (WKNARC2), 13 - 17 May 2013, Horta, Azores. ICES CM 2013/ACOM:52. 65 pp.

ICES 2015. First Interim Report of the Working Group on Working Group on Biological Parameters (WGBIOP), 7-11 September 2015, Malaga, Spain. ICES CM 2015/SSGIEOM:08. 67 pp.

ICES. 2016. Report of the Working Group on Biological Parameters (WGBIOP), 10–14 October, 2016, Monopoli, Italy. ICES CM 2016/SSGIEOM:08. 106 pp

ICES. 2017. Report of the Working Group on Biological Parameters (WGBIOP), 2–6 October 2017, Sardinia, Italy. ICES CM 2017/SSGIEOM:08. 129 pp.

ICES. 2018. Working Group on Biological Parameters (WGBIOP), 1–5 October 2018. Ghent, Belgium. ICES CM 2018/EOSG:07. 186pp

Jolliffe I. T. 2002. Principal Component Analysis. Second Edition, Springer Series in Statistics.

Jolliffe I. T. 2002. Principal Component Analysis. Second Edition, Springer Series in Statistics.

Williams A. J., Leroy B. M., Nicol S. J., Farley J. H., Clear N. P., Krusic-Golub K., and Davies C. R., 2013. Comparison of daily- and annual-increment counts in otoliths of bigeye (Thunnusobesus), yellowfin (T. albacares), southern bluefin (T. maccoyii) and albacore (T. alalunga) tuna. ICES J. Mar. Sci., 70: 1439–1450

Kimura D. K., Lyons J. J. 1991 Between-Reader Bias and Variability in the Age-Determination Process. Fish. Bull. 89: 53-60

Kimura D. K., Anderl D. M. 2005 Quality control of age data at Alaska Science Center. Mar. Freshwater Res. 56: 783-789

Kimura, D. K. 1977. Statistical assessment of the age–length key. Journal of theFisheries Research Board of Canada 34:317–324.

Lanteri L., Garibaldi F. 2019 Large Pelagic Fish. In Carbonara, P., Follesa, M.C., eds. 2019. Handbook on fish age determination: a Mediterranean experience. Studies and Reviews. No. 98. Rome, FAO. 2019 111-145.

Liao H., Sharov A. F., Jones C. M., Nelson G. A. 2013. Quantifying the Effects of Aging Bias in Atlantic Striped Bass Stock Assessment. Trans. Am. Fish. Soc. 142: 193–207.

Lê S., Josse J. and Husson F. 2008. FactoMineR : An R Package for Multivariate Analysis, Journal of Statistical Software.

Mandado M., Vázquez A. 2011. On otoliths sampling. NAFO SCR Doc. 11/023: 9pp.

Morales-Nin B., Panfili J. 2002. Specific age estimation. In: Manual of fish sclerochronologyPanfili, J., Troadec, H., Pontual, H. D., Wright, P. J. eds. Ifremer-IRD coedition, Brest, France

Neilson, J., Arocha, F., Calay, S., Mejuto, J., Ortiz, M., Scott, G., Smith, C., Travassos, P., Tserpes, G. and Andrushchenko, I. 2013. The Recovery of Atlantic Swordfish: The Comparative Roles of the Regional Fisheries Management Organization and Species Biology. Rev. Fish. Sci. 21(2): 59-97.

Oeberst R. 2000. An universal cost function for the optimization of the number of age readings and length measurements for Age-Length-Key-Tables (ALKT). Arch. Fish. Mar. Res. 48(1): 43–60.

Oele D. L., Lawson Z. J., McIntyre P. B. 2015. Precision and Bias in Aging Northern Pike: Comparisons among Four Calcified Structures. N. Am. J. Fish. Manag. 35 (6): 1177-1184

Palco, B. J., Beardsley, G. L. and Richards, W. J. 1981. Synopsis of the biology of the swordfish Xiphias gladius (L). U.S. Dep. Commer., NOAA Tech. Rep. NMFS. Circ. 441, 21 p.

Panfili J., de Pontual H., Troadec J.-P., Wright P.J. (eds.) (2002). Manual of fish sclerochronology. Brest, France: IFREMER-IRD co-edition, 464pp. ISBN 2-7099-1490-5

PGCCDBS 2011 Guidelines for Otolith Exchanges, Vienna 9pp

Punt AE, Smith DC, KrusicGolub K, Robertson S. 2008. Quantifying age-reading error for use in fisheries stock assessments, with application to species in Australia's southern and eastern scalefish and shark fishery. Can. J. Fish. Aquat. Sci. 65(9): 1991–2005

Quelle, P., González, F., Ruiz, M., Valeiras, X., Gutierrez, O., Rodríguez-Marín, E. and Mejuto, J. 2014. An approach to age and growth of South Atlantic Swordfish (Xiphias gladius) stock. ICCAT Coll. Vol. Sci. Pap. Vol. 70(4): 1927-1944.

R Development Core Team 2018. R Software, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.r-project.org/.

Reeves S. A. 2003. A simulation study of the implications of age-reading errors for stock assessment and management advice. ICES J. Mar. Sci. 60: 314–328

RCM MED&BS-LP 2016. Report of the 2016 Regional Co-ordination Meeting Mediterranean and Black Sea, Large Pelagic Fisheries, 19-23/09/2016 Pasaia (Donostia – San Sebastian), Spain, 244 pp

Ricker, W. E. 1975. Computation and interpretation of biological statistics offish populations. Fisheries Research Board of Canada Bulletin 191.

Rodriguez-Marin, E., Quelle, P., Adis, P., Alemany, F., Bellodi, A., Busawon, D., Carnevali, O., Cort, J.L., Di Natale, A., Farley, J., Garibaldi, F., Karakulak, S., Krusic-Golub, K., Luque, P.L., Ruiz, M. 2019. Protocol for the preparation of Atlantic bluefin tuna otoliths for direct ageing. In: Report of the ICCAT GBYP International Workshop on Atlantic bluefin tuna growth (Santander, Spain, February 2019). ICCAT/SCRS/Doc. In press.

Rodríguez-Marín E., N. Clear, J.L. Cort, P. Megalofonou, J.D. Neilson, M. Neves dos Santos, D. Olafsdottir, C. Rodriguez-Cabello, M. Ruiz, J. Valeiras 2007 Report of the 2006 ICCAT workshop for Bluefin Tuna direct ageing. Col. Vol. Sci. Pap. ICCAT, 60(4): 1349-1392.

Rude N. P., Hintz W. D., Norman J. D., Kanczuzewski K. L., Yung A. J., Hofer K. D., Whitledge G. W. 2013. Using pectoral fin rays as a nonlethal aging structure for smallmouth bass: precision with otolith age estimates and the importance of reader experience. J. Freshwater Ecol. 28: 2-199

Scientific Technical and Economic Committee for Fisheries (STECF) 2017. 54th Plenary Meeting Report (PLEN-17-01); Publications Office of the European Union, Luxembourg; EUR 28569 EN

Sun, C.L., Wang, S.P., and Yeh, S.Z. 2002. Age and growth of the swordfish (Xiphias gladius L.) in the waters around Taiwan determined from anal-fin rays. Fish. Bull. 100(4): 822-835.

Secor, D.H., Allman, R., Busawon, D., Gahagan, B., Golet, W., Koob, E., Luque, P.L., and Siskey, M. 2013. Standardization of otolith-based ageing protocols for Atlantic bluefin tuna. SCRS/2013/084.

Smith B., J., Dembkowski D. J., James D. A., and Wuellner, Melissa R. 2016. A Simple Method to Reduce Interpretation Error of Ages Estimated from Otoliths" 2016). Open Fish Sci. J. 9: 1-7

Young J., Drake A. and Langridge A. 2004. Age and growth of broadbill swordfish, Xiphias gladius, inthe domestic longline fishery off eastern Australia. CSIRO Marine ResearchProject No. 2001/014.

Annex1. Smartdots user manual

2018 USER MANUAL

smartdots

SMARTDOTS

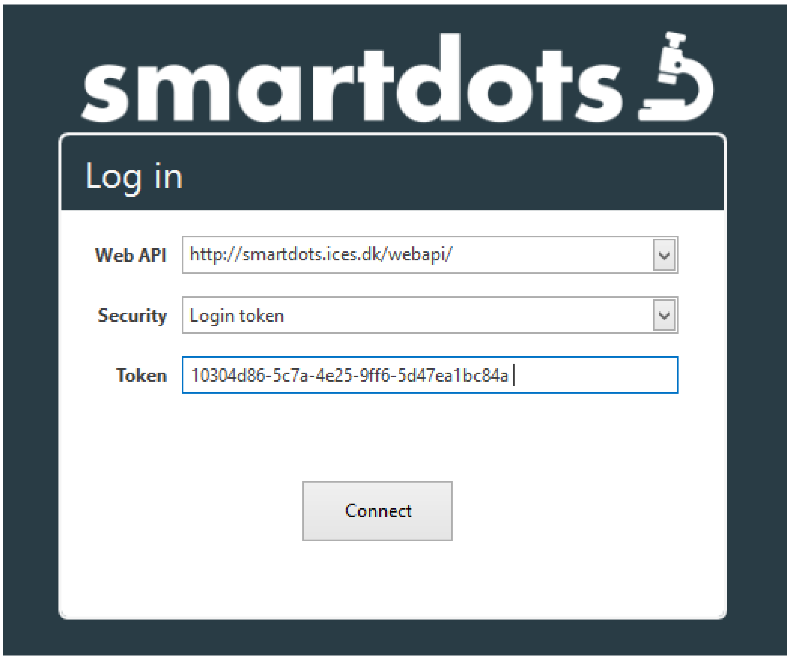
http://www.ices.dk/marine-data/tools/Pages/smartdots.aspx

CONTACT

carlos@ices.dk
Kevin.DeCoster@ilvo.vlaanderen.be
joco@aqua.dtu.dk

Login

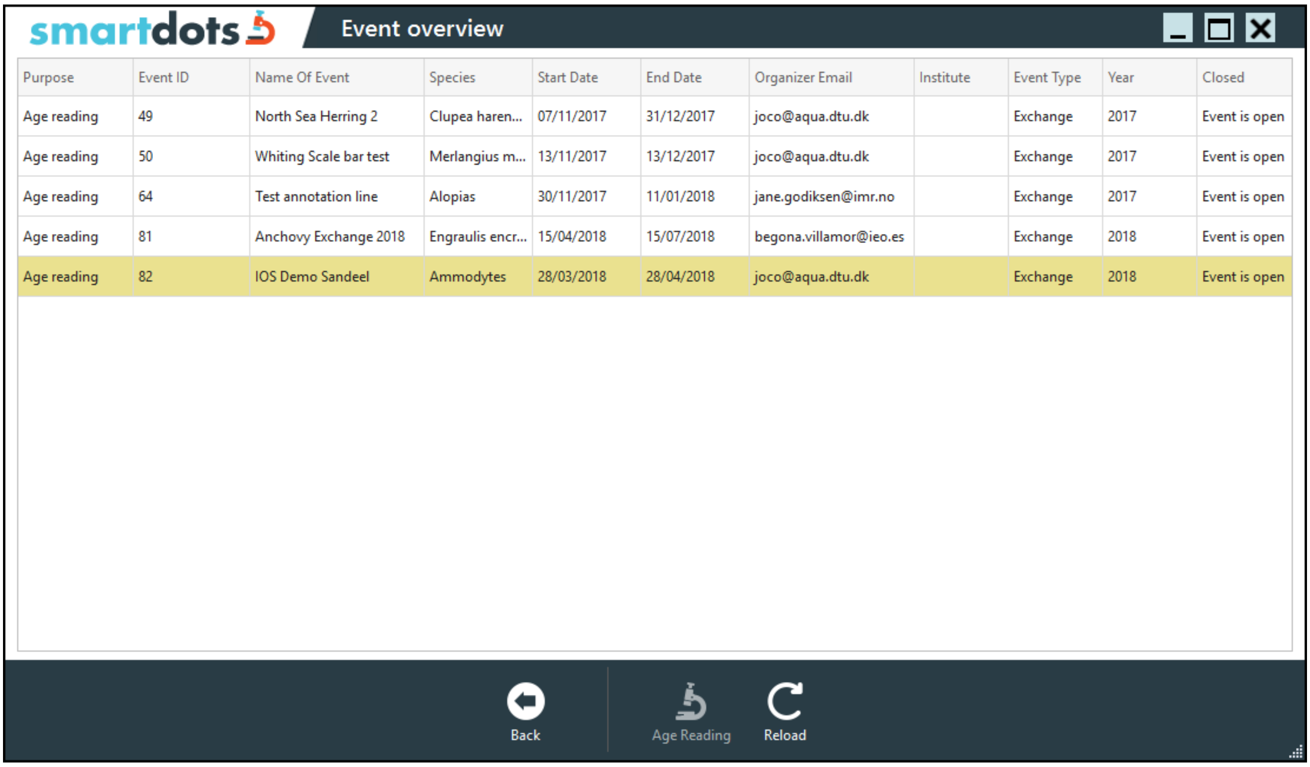
You need to connect to the ICES web api using the Token authentication. A token can be obtained here http://ices.dk/marine-data/tools/Pages/smartdots. The token provided should be copied and pasted into the login screen shown in the image below. Press connect to continue.



1. Event screen

2

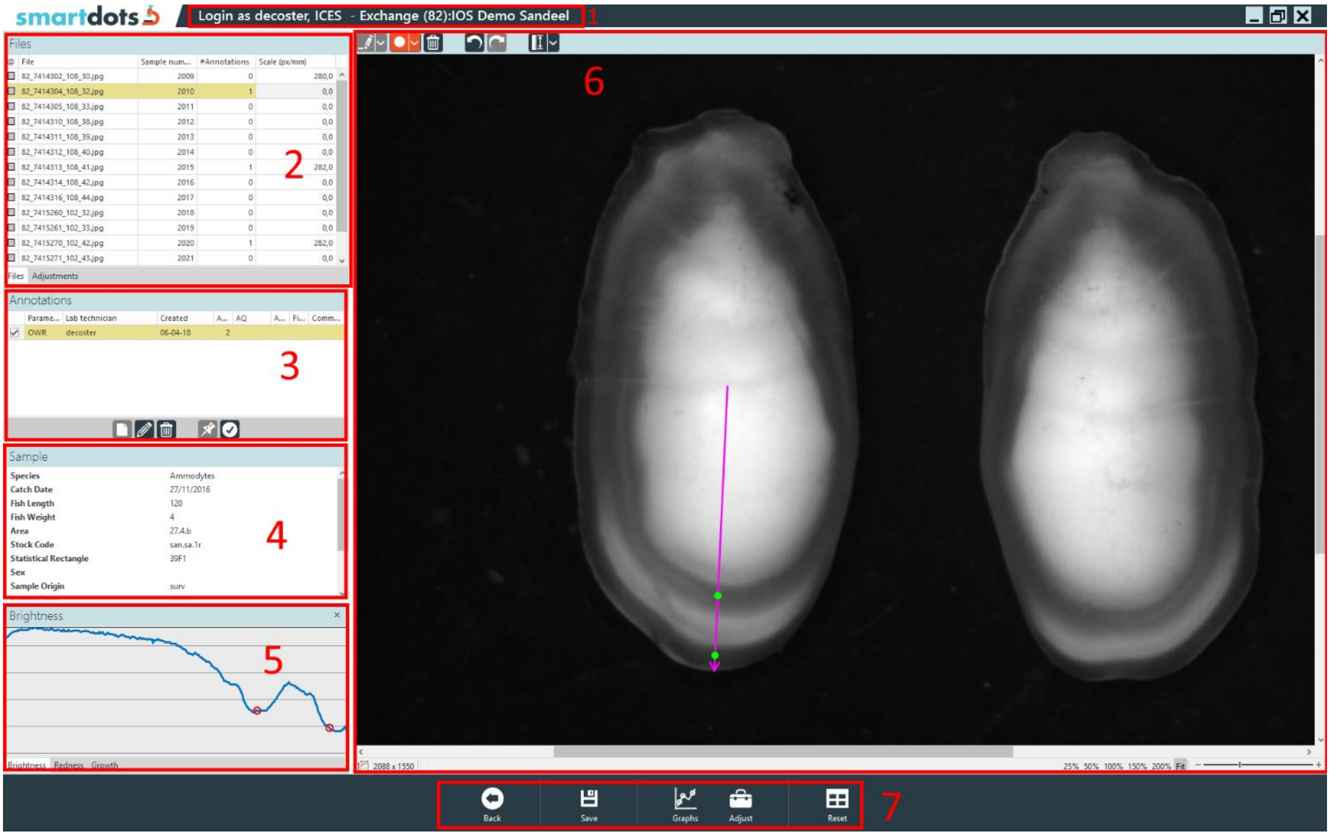
On this screen we get a list of all available workshops/exchanges to make age readings on. Double-Click on your exchange/workshop to start the age reading.



Purpose	Event ID	Name Of Event	Species	Start Date	End Date	Organizer Email	Institute	Event Type	Year	Closed
Age reading	49	North Sea Herring 2	Clupea haren...	07/11/2017	31/12/2017	joco@aqua.dtu.dk		Exchange	2017	Event is open
Age reading	50	Whiting Scale bar test	Merlangius m...	13/11/2017	13/12/2017	joco@aqua.dtu.dk		Exchange	2017	Event is open
Age reading	64	Test annotation line	Alopias	30/11/2017	11/01/2018	jane.godiksen@imr.no		Exchange	2017	Event is open
Age reading	81	Anchovy Exchange 2018	Engraulis encr...	15/04/2018	15/07/2018	begona.villamor@ieo.es		Exchange	2018	Event is open
Age reading	82	IOS Demo Sandeel	Ammodytes	28/03/2018	28/04/2018	joco@aqua.dtu.dk		Exchange	2018	Event is open

2. Age reading

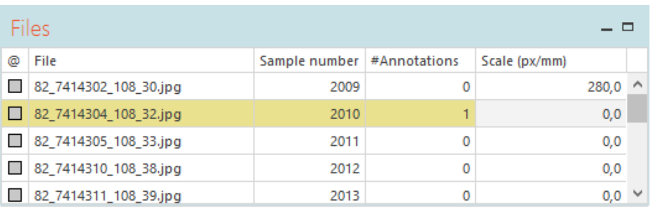
This is the screen where everything related to age reading happens. This screen consists of multiple components.



2.1. Title

Shows info about the logged in user and the current workshop/exchange he/she is working on.

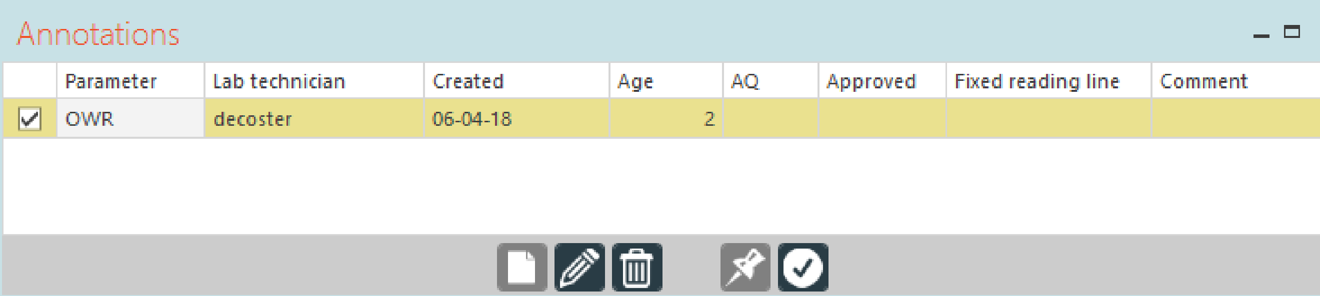
2.2. Files


Files

@	File	Sample number	#Annotations	Scale (px/mm)
☐	82_7414302_108_30.jpg	2009	0	280,0
☐	82_7414304_108_32.jpg	2010	1	0,0
☐	82_7414305_108_33.jpg	2011	0	0,0
☐	82_7414310_108_38.jpg	2012	0	0,0
☐	82_7414311_108_39.jpg	2013	0	0,0

Contains information about all image-files that are loaded. The file name is the physical file name the image has. The sample number is the link between the image and the sample (= fish) the otolith originates from. The #Annotations is the number of readings the user can see. The scale is the amount of pixels that represent 1 millimeter when the image is zoomed 100%.

2.3. Annotations



Shows information about the age readings for the selected file. Parameter describes what type of annotations are made. This can be winter rings, summer rings, day rings, … Lab technician is the reader. The AQ indicates the quality of the age reading. Approved means the user has approved the reading so that it can be included in reporting. The fixed reading line indicates that the line for this age reading will be used for everyone who creates an age reading. Only administrators can create fixed reading lines. The comment is optional but can be used by the user to add remarks.

There are several buttons here which need explaining. From left to right:

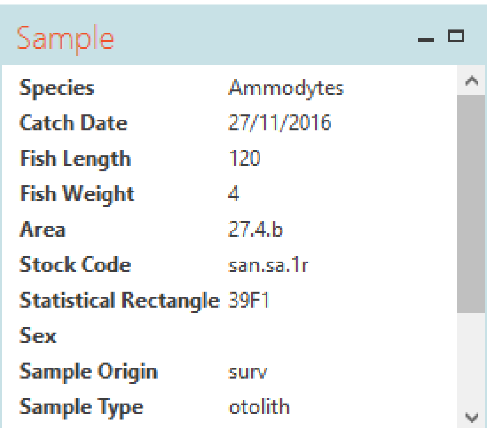
- Create: Creates a new record for age reading. The user can now start making annotations.
- Edit: A popup shows up with the properties that the user can change. Here you can also add comments.
- Delete: delete the reading
- Pin: This is only for administrators. It creates the fixed reading line. All users will have to annotate on this line.
- Approve/Disapprove: Toggles the approval of the reading.

SmartDots will automatically give the age of the fish based on the number of annotations that you make.

You do <u>NOT</u> need to mark the centre or the outermost edge of the otolith. You <u>MUST</u> place your mark at the end of the winter ring and only on the winter rings which are counted as a year when estimating the age.

Please provide an AQ score for each age that you give: AQ1: Easy to age with high precision; AQ2: Difficult to age with acceptable precision; AQ3: Unreadable or very difficult to age with acceptable precision. You <u>MUST</u> approve the final annotation for each otolith that you wish to have included in the reporting.

2.4. Sample



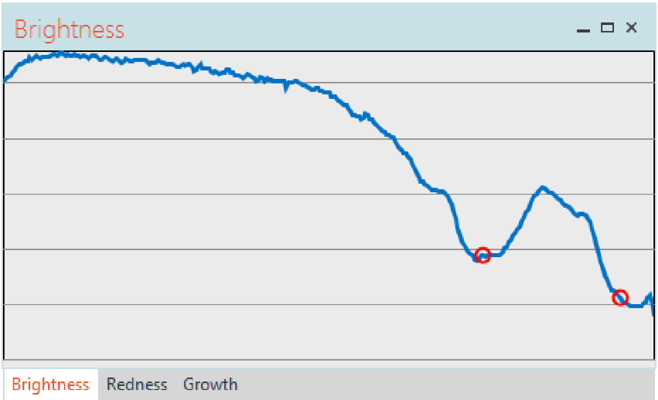
This window shows all information about the sample fish linked with the image. Everything here is read-only.

2.5. Graphs

Graphs show information about the selected age reading's line and annotations. There are currently 3 graphs in SmartDots.

2.5.1. Brightness

Displays the brightness of the image under the line from start to end. This can be used to detect rings. The graph will display the location of the cursor if hovered near the line. When Dot-mode is active, the user can click the graph on a specific location to add the dot.

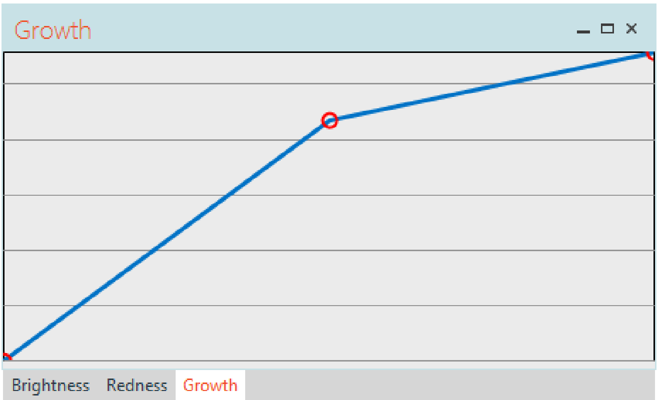


2.5.2. Redness

The same for Brightness applies to redness, but instead of the brightness of pixels it displays the red-value. This is only used for otoliths that are imbedded in a red colouring agent.

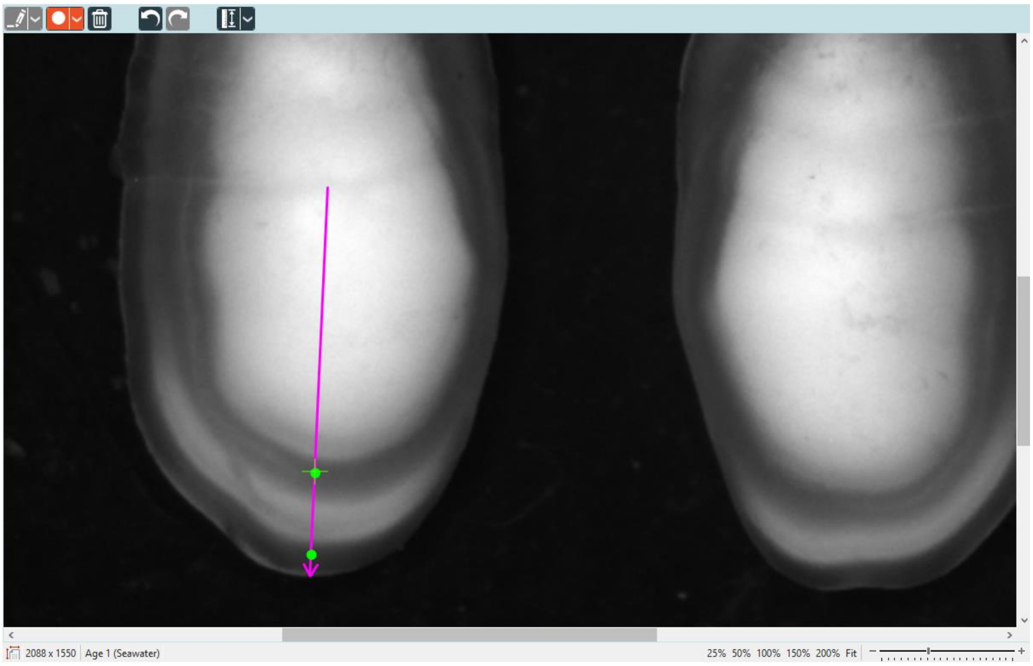
2.5.3. Growth

Shows the distance between annotations.



2.6. Editor

Here is where the drawing takes place.

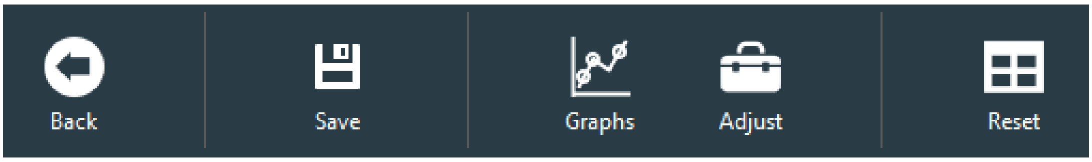


At the top we have the 3 modes (line-mode to draw lines, dot-mode to draw dots and delete-mode to delete lines and/or dots). There are also buttons for undo/redo. The last button is used to measure the scale. All buttons with an arrow next to them have additional options.

In the middle is the image and all drawn shapes.

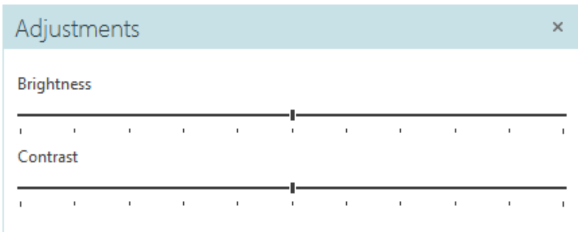
At the bottom is the information bar and the zoom options.

2.7. Toolbar



These buttons contain logic not specifically related to an age reading but can be helpful for the user.

- Back: return to the previous screen
- Save: save current progress
- Graphs: show/hid the graphs
- Adjust: show/hide the adjustments. Adjustments can be found as a tab under Files. It allows the user to temporarily adjust the brightness/contrast of the image.
- Reset: resets all layout



Annex 2. Agenda of the RECOLAPE Workshop on swordfish age reading

Strengthening Regional cooperation in the area of large pelagic fisheries data collection (Acronym: RECOLAPE)

Task 5.3.1 Comparison of age-length keys between Member States and exploratory analysis

Agenda

The meeting will start at 10.30 of 2th April 2019

and will end on 4th April (~13.00)

Meeting place:

Portuguese Institute for the Ocean and Atmosphere,

Av. 5 de Outubro s/n

8700-305 Olhão, Portugal

Tuesday 2th April 2019

10.30-11.00

Welcome of the participants

Round Table of participants

Approval of the agenda

11.00-11.30

Presentation from each lab/participant about biology, otolith preparation methods, ageing criteria and ageing scheme (IPMA, UNIMAR, IEO; IRD, AZTI, HCMR, DFMR)

11.30-12.00 Coffee break

12.30-13.00

Presentation from each lab/participant about biology, otolith preparation method, ageing criteria and ageing scheme (IPMA, UNIMAR, IEO; IRD, AZTI, HCMR, DFMR)

13.00-14.30 Lunch breack

14.30-15.30

Presentation from each lab/participant about biology, otolith preparation method, ageing criteria and ageing scheme (IPMA, UNIMAR, IEO; IRD, AZTI, HCMR, DFMR)

15.30-16.00

Presentation of the "HANDBOOK ON FISH AGE DETERMINATION: a Mediterranean experience – Large Pelagic Fish" - Fulvio Garibaldi

16.00-16.30 Coffee break

16.30-17.00

Presentation of Exchange results – Pierluigi Carbonara

Wednesday 3th April 2019

9.30-10.30

Discussion on the Exchange results

10.30-11.00 Coffee break

11.00-12.00

Preparation method

12.00-13.00

Clarify the position of the first annulus with the images analysis for the three species

13.00-14.30 Lunch break

14.30-15.00

Continue the guidelines and common ageing criteria

15.00-16.00 Coffee break

16.30 – 17.00

Comparison of age-length keys (ALK) between Member States and exploratory analysis – Pierluigi Carbonara

Thursday 4th April 2019

9.30-10.30

Discussion on Agreed Age reading Protocol and Develop reference collections of Image

10.30-11.00 Coffee break

11.00-13.00

Plenary session

Recommendations and conclusion

Planning of future work

Any Other Business and meeting closure

FRAMEWORK CONTRACT - MARE/2016/22 « Strengthening regional cooperation in the area of fisheries data collection » Annex III « Biological data collection for fisheries on highly migratory species »

Strengthening Regional cooperation in the area of large pelagic fisheries data collection (Acronym: RECOLAPE)

T.5.4- Calendar for national and regional data quality checking process

D.5.4.1 - Proposal for a detailed annual calendar for the national and regional quality checking process

Responsables: Mathieu Depetris (IRD) & Pierluigi Carbonara (COISPA)

Index

Executive summary 3

1. Introduction..4

    1.1. Aims of this document 4

    1.2. Data used 4

2. Results 6

3. Interpretation and advices 6

Executive summary

This document is in relation to the work package 5 of the project MARE/2016/22: "Strengthening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species" (acronym RECOLAPE). The main objective of this package is to propose an annual calendar for quality assessment procedures.

Results and interpretation have to be considerate as global trends and it's very important to keep in mind that all proposals and advice should be managing case by case.

For now, two proposals were made:

- if all human and material resources were available, the best option was to run data quality checks even new data are available,
- otherwise, according to regular annual data call of Regional Fisheries Management Organisations and European Union, data quality checks should be run in February or March (without any modification of data provided after running processes).

1. Introduction

1.1. Aims of this document

This document was established under the European project MARE/2016/22 « Strengthening regional cooperation in the area of fisheries data collection » Annex III « Biological data collection for fisheries on highly migratory species » (acronym: RECOLAPE). He is a part of the WP5 (procedures to assess the quality of biological data collected at regional level) and aim to answer about the last question on data quality processes: When do you need to apply data checks to make available the best data on large pelagic fisheries?

1.2. Data used

For these analyses, we considered all the meetings, working groups, reports deadlines or any events when possibly data could be provided. Data and script used were available GitHub through the following link: https://github.com/OB7-IRD/RECOLAPE/tree/master/WP5/T5.4.

Furthermore, we are not only focusing on the two studies cases of the RECOLAPE project (swordfish in the Mediterranean targeted by longline fisheries and major tropical tunas in the Atlantic Ocean targeted by purse seiner fishery). The idea is to have global trends and provide advises for all European country and all fisheries associated with large pelagic. It's important to keep in mind that theses advice should be adapted to specifications of each country and just give a global framework.

For a better understanding of results, and especially for adaptation to each country, number of events were presented by structure or organization:

- by Regional Fisheries Management Organisations (RFMOs) with a focus on the 5 which manage highly migratory species (tunas and associated species) and linked to the European Union (figure 1),
- by Tropical Tunas Treatments (T3[1]) meetings,
- by European data call, like the Fisheries Dependent Information (FDI).

[1] The T3 processing was built about 30 years ago in order to correct biases of the logbook data on species composition and to provide more accurate catch estimates per species for the European purse seine fleet.

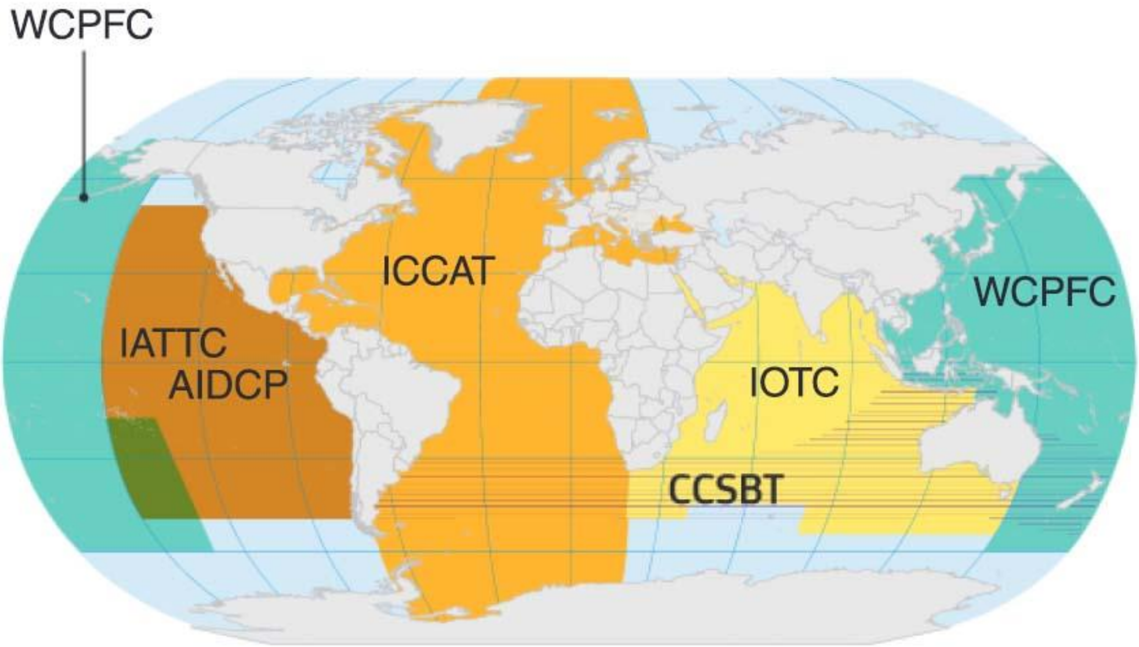
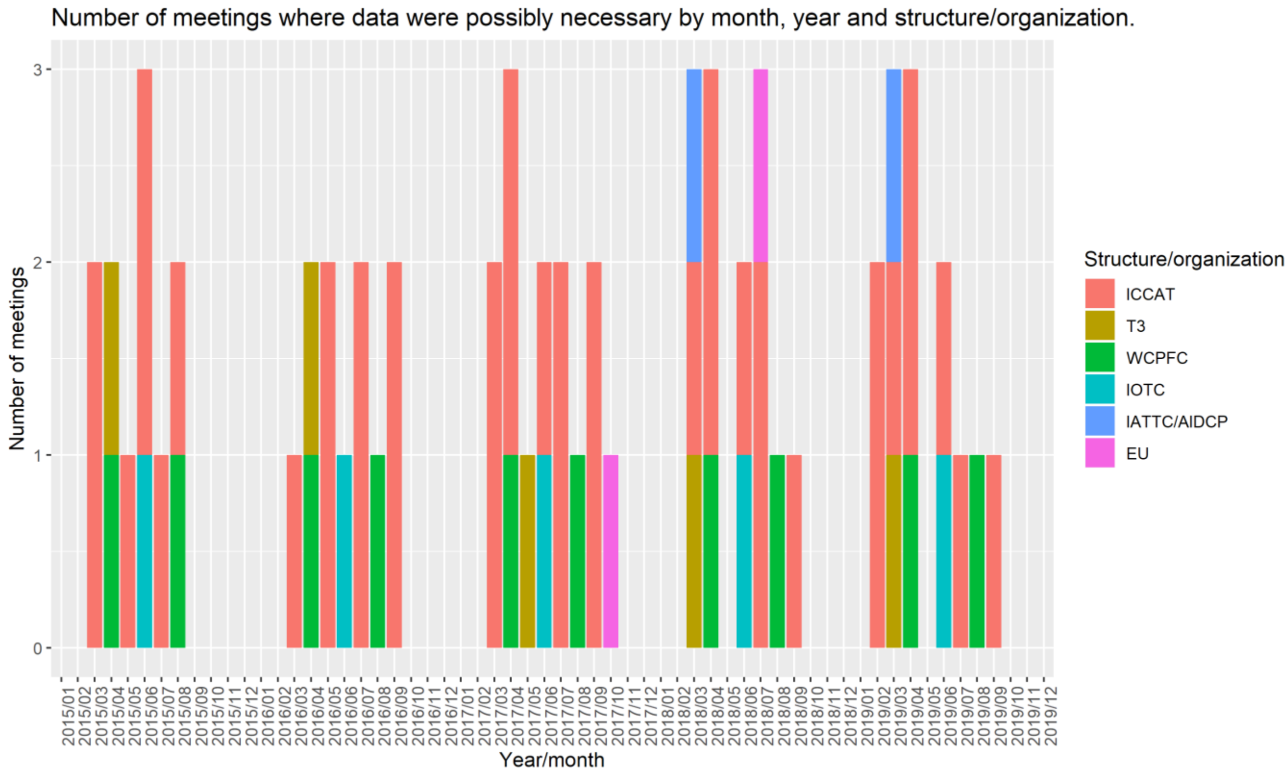


Figure 1: RFMOs[2] which manages highly migratory species, mainly tuna.

[2] ICCAT: International Commission for the Conservation of Atlantic Tunas
IOTC: Indian Ocean Tuna Commission
WCPFC: Western and Central Pacific Fisheries Commission
IATTC: Inter-American Tropical Tuna Commission
AIDCP: Agreement on the International Dolphin Conservation Program
CCSBT: Commission for the Conservation of Southern Bluefin Tuna

2. Results

For covered a large panel, we look events append during the last 5 years (with events expected for 2019). The results are in the figure 2 below.



As expected, we can globally the same pattern between each year, with at least one event since March to September-October.

3. Interpretation and advice

It’s difficult to make a global advice or framework because it depends on each specification. In a perfect world with fewer limitations of resources (human and material), data quality process should be run every time that new data were available. This case is restrictive but assure less work at one time (you split the effort) and makes revised data available faster (especially regarding non-regular data calls). With our example, we should run data quality processes every month since March and until September-October (with a country involved in all the structure/organization). Even if this advice seems difficult to reach, it could be easier understand and naturally accepted in the case where you use a database for storage. Data quality processes associated are often incremented before putting data in the database. For example, France and Spain use several controls, especially through the AKaDo[3] software, for checking row data and underline possibly mistakes. Nevertheless, this kind of calendar request available of human resources and in several cases (sometimes most) it's impossible to reach this purpose. A compromise could be to run data quality checks before regular annual data call of RFMOs, at least two months before June-July. For example, several countries work together (Spain, Seychelles, Senegal and France) to prepare and validate data before submitting it to ICCAT and IOTC. In the case, data quality procedures should be run before the T3 technical meeting, which means around February-March. The following table 1 summarizes advice below:

	Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
In a perfect world	Even new data are available											
Minimum[4]												

Otherwise, it's very important to talk about global trends only and keep in mind that there could be some potential bias:

- for now, we only discuss regular data call which append almost every year. Each non-regular data (for a research project) should be integrated in this calendar case by case.
- data expected for one event could be the same for another. For example, data provided for an ICCAT data preparatory (focusing on one specie) could be the same (or a part) of annual data transmission. This specification could lighten the calendar (if there are no modifications or corrections across the year).

[3] AKaDo software: quality data analysis for AVDTH database. This software automatically performs set of tests and checks on data and provides summary tables with potential anomalies (with percentages of errors that remain to be corrected).

[4] Without any modification of data provided after running of data quality process

FRAMEWORK CONTRACT – MARE/2016/22 “Strengthening regional cooperation in the area of fisheries data collection” Annex III “Biological data collection for fisheries on highly migratory species”

Project acronym: RECOLAPE

WP6 – Regional consultation of Member States

Participating partners:

D.6.1- List of qualitative and quantitative consultation outputs

Deliverable coordinated by Jon Ruiz (AZTI)

Index

Executive summary 3

1. Objective 4
2. Methodology 4
3. Results 5

Executive summary

This deliverable is part of the Work Package (WP) 6 within the project MARE/2016/22: "Strengthening regional cooperation in the area of fisheries data collection, Annex III Biological data collection for fisheries on highly migratory species" (acronym RECOLAPE).

WP6 was designed to collect inputs from regional consultations with all the MSs dealing with the collection of data on large pelagic fisheries, as well as from end-users (tuna RFMOs and RCG). Furthermore, this WP dealt with dissemination of the results, taking into account that the audience for the different WPs could be different (such as tuna RFMO Working Group meetings or RCG meetings).

To make the project truly successful, it is essential that the results and experience gained in the different WPs be well received, understood and implemented by all organisations involved in the data collection for highly migratory species. WP6 identified possible impediments that may come to light during this process.

All Member States dealing with large pelagic fisheries were consulted and invited to discuss the project results. This consultation was initially conducted in writing and results were presented during the RCG-LP 2019 meeting. Seven responses were obtained, from France, Cyprus, Spain (partially), Portugal (partially), Malta, Greece and from the Chair of the RCG-LP. The participation rate exceeded 50% and included some of the most relevant countries with large pelagic captures.

The main points of agreement and disagreement concerning WP1, WP2 and WP5 are listed below.

- The general proposal to structure the RCG-LP in 3 stages, as well as the number of subgroups and meetings, achieved a strong consensus.
- Some MSs identified the shortage of human resources as a limiting factor that might hamper their participation in certain subgroups.
- Pan-regional subgroups: The priority subgroups for the MSs coincide with those proposed by the contractors ("data management" and "Regional Sampling Plans").
- It seems that there is no clear consensus on the need for a regional database to host the LP data (RDBES or any other). However, this requirement is explicitly mentioned in point 8 of Article 9 of the recast regulation.
- There are no notable disagreements among the participants regarding results obtained in WP2: All MS agree with the data requirements and priorities proposed to design an RSP for Mediterranean swordfish and tropical tunas. All MSs concur that the sampling protocol should be unique and agreed-on at the RCG level.
- There are no notable disagreements among the participants regarding results obtained in WP5: it seems that the scripts in the R language for checking the data quality are a valid option, favoured by the participants. MSs confirm the need to standardise age-reading protocols and agree on the benefits of establishing a working group with this aim.

1. Objective

The purpose of this WP was to collect inputs from regional consultations from all Member States participating in the large pelagic data collection and from the end-users (such as tuna RFMOs or the RCG-LP). The results were also discussed with the tuna industry, mainly concerning WP3.1.- *Development of data collection protocols and tools for fad management plans* and WP4.- *Data collection strategy for standardisation of catch per unit effort (CPUR) or for alternative abundance indices in tropical tuna fisheries.*

The main activity under this WP was a regional consultation process involving the Member States taking part in the large pelagic data collection (irrespective of their participation in the project). The results of the project were discussed to identify points of consensus and/or disagreement.

2. Methodology

All Member States dealing with large pelagic fisheries were consulted and invited to discuss the project results. This consultation was initially conducted in writing and results were presented during the RCG-LP 2019 meeting, which coincided with the end of the project (May 13, Madrid). The written consultation (survey) was conducted through the national correspondents. This survey only refers to the WPs of common interest (WP1, WP2 & WP5) that concern all the countries, irrespective of the fisheries in which they are involved. The survey was conducted by e-mail using the SurveyMonkey online tool (https://es.surveymonkey.com/). It could also be answered in Word format; the questionnaire included in Annexe 1 was attached. The survey consists of 21 questions in three blocks (i.e., WP1, WP2 and WP5). In most of the cases, it presented a choice of a number in the range from 1 (completely disagree) to 6 (completely agree). This type of answer choice increases participation and facilitates the later analysis. The scale of 1 to 6 does not allow choosing the middle number; every answer has to be at least slightly agreeing or disagreeing. The consultation drew up a list of qualitative and quantitative outputs, where points of consensus and/or disagreement were identified.

Additionally, WP6 dealt with the dissemination of the results. There is no doubt that the meeting of the RCG-LP is the main forum, where these results must be presented and discussed as there is a clear and direct interest. Thus, the results were partially presented during the annual RCG-LP meeting in 2018 (June 26–28, Heraklion), and the final results, during the 2019 meeting (May 13-14, Madrid). There is also no doubt that the feedback and views of the different tuna RFMOs are of paramount importance, as they constitute the end-users of the collected fishery data. Therefore, in cases of those WPs for which the participation and feedback of the RFMOs are essential (such as WP3 and WP4), the results were presented at the RFMO level (in specifically selected working groups).

3. Results

Survey results

The survey had been sent to 10 national correspondents representing MSs with LP fisheries included in their National Work Programmes (Spain, France, Portugal, Greece, Malta, Cyprus, UK, Italy, Ireland and Croatia). The current Chair of the RCG-LP had also been included as a survey recipient. Seven responses were obtained, from France, Cyprus, Spain (partially), Portugal (partially), Malta, Greece and from the Chair of the RCG-LP. The participation rate exceeded 50% and included some of the most relevant countries with large pelagic captures.

The main points of agreement and disagreement concerning WP1, WP2 and WP5 are listed below. Tables 1, 2 and 3 show the detailed answer by MS.

WP1 – Proposal for a future RCG-LP structure

- There seems to be a broad consensus on the need for RCG-LP to be autonomous.
- The general proposal to structure the RCG-LP in 3 stages, as well as the number of subgroups and meetings, achieved a strong consensus. All MSs showed their interest in participating in the subgroups associated with stage 1 (data needs), stage 3 (RCG-LP main meeting), and stage 2 (data analysis) for the fisheries in which they are involved.
- Some MSs identified the shortage of human resources as a limiting factor that might hamper their participation in certain subgroups.
- Pan-regional subgroups: The priority subgroups for the MSs coincide with those proposed by the contractors ("data management" and "Regional Sampling Plans"). However, most MSs consider that the RCG-LP should be somehow represented in all subgroups, including "end user", "governance" and "implication of landing obligation for data collection".
- It seems that there is no clear consensus on the need for a regional database to host the LP data (RDBES or any other). However, this requirement is explicitly mentioned in point 8 of Article 9 of the recast regulation.

WP2 – Development of Regional Sampling Plans

- Neither Spain nor Portugal has responded to this section of the survey. Lack of response could reduce the robustness of the results.
- There are no notable disagreements among the participants: All MS agree with the data requirements and priorities proposed by the WP2 to design an RSP for Mediterranean swordfish and tropical tunas. All MSs concur that the sampling protocol should be unique and agreed-on at the RCG level. Finally, all (except for Greece) support the idea of uploading LP data to the current RDBES (Regional Database and Estimation System currently used by the northern RCGs).

WP5 – Procedures to assess the quality of biological data

- Neither Spain nor Portugal has responded to this survey section, which might reduce the robustness of the results.
- There are no notable disagreements among the participants: it seems that the scripts in the R language for checking the data quality are a valid option, favoured by the participants. They also agree on the focus on data checks. Finally, the participants confirm the need to standardise age-reading protocols and agree on the benefits of establishing a working group with this aim. The current Chair of the RCG-LP proposes to expand this standardisation to maturity scales.

RFMO feedback on WP3.1 (ICCAT & IOTC)

The general scope of the project has been already presented in several ICCAT and IOTC data-collection working groups. The results from WP3.1 were presented during the ICCAT species group (Madrid, September 2018) and discussed with the ICCAT Standing Committee on Research and Statistics. The Committee reviewed results from WP3.1 (presented as *docs. SCRS 2018/159 & SCRS 2018/158*), which proposed the Best Standards for Data Collection and Reporting Requirement on FADs as a response to Annex 8 of ICCAT Recommendation 16-01. These documents also proposed new forms (ST08a and ST08b) for data reporting on FADs and buoys, to replace the form currently used by ICCAT. The SCRS adopted the new proposed ST08a and ST08b forms. They proposed that the Best Standards for Data Collection included in WP3.1 should be considered a minimum standard for data collection in the ICCAT framework.

The same results were presented during the IOTC WPDCS (Working Party on Data Collection and Statistics) (Seychelles, December 2018). This IOTC working group acknowledged the effort put into the harmonisation of terminology and data collection and reporting requirements for FOB. However, due to the differences in classification and reporting requirements between this proposal and the existing IOTC classifications, IOTC suggested the joint tuna RFMOs FAD working group (May 2019, La Joya) as the appropriate forum for harmonising FAD classifications across RFMOs. Thus, the main outputs from WP3.1 and WP4 were presented again during the joint tuna RFMO FAD Working Group. This last meeting coincided with the end of the project so that the feedback from this group has not been included in this report. However, the output from this working group will need to be considered and further discussed by the IOTC/ICCAT Secretariat and the scientific community.

Finally, although the project ended in May 2019, there are plans to present the results of some WPs in later ICCAT / IOTC working groups. The main outputs from WP3.2– "EMS feasibility study in longline fisheries"–will be presented at the next IOTC WPDCS (November 2019, Seychelles).

RCG 2019 feedback

The results from WP1, WP2 & WP5 were presented and discussed during the RCG-LP 2019 meeting (Madrid, May 2019). At the time of writing this report, the report and final list of recommendations done by the RCG – LP 2019 was not available. However, many of the results obtained through the written consultation were confirmed. The proposed structure of the RCG-LP, as well as the number of stages and subgroups was adopted. In addition, the RCG -LP recommended to add a fourth technical subgroup focused on the coordination of the bait boat fisheries. As for WP2 and WP5 results, there were no notable disagreements among the RCG-LP participants. However, it is worth highlighting the doubts that persist in some MS in relation to the RDBES. In this sense, the RCG-LP raised the possibility to organize a practical session on the RDBES, where the RDBES Steering Committee could clarify any doubt.

Table 1. Survey result by MS (WP1). 1 (completely disagree), 2 (disagree), 3 (slightly disagree), 4 (slightly agree), 5 (agree), 6, (completely agree) na (not applicable), empty (no answer).

Questions Block 1: Large Pelagic Regional Coordination Group (RCG-LP) structure		MEMBER STATE						OTHER
*Please answer to the following questions where **1 is completely disagree** and **6 completely agree***	Mean value	Malta	Spain	Portugal	France	Cyprus	Greece	RCG-LP chair
1- Do you consider that the RCG LP should be a group independent from the rest of RCGs?	4.57	6	4	5	5	4	5	3
i. If you answered the first question with a 3 or less, do you consider that it should be part of the RCG MED & BS?	1.00	na	na	na	na	na	na	1
ii. If you answered the first question with a 3 or less, do you consider that it should be part of the RCG LDF?	3.00	na	na	na	na	na	na	3
2- The global proposal to structure the RCG LP in 3 stages (i.e. identification of data needs/design of RSP/decision making) is appropriate	5.43	5	5	6	6	5	5	6
3- The number of subgroups or fisheries in stage 2 is adequate (1) Tropical tuna. (2) Longline outside the MED. (3) Longline in the MED. (4) Bluefin tuna	5.14	5	5	6	4	5	5	6
4- Globally the number of meetings (subgroups) proposed for the internal structure of the RCG LP is appropriate	5.00	5	4	5	6	4	5	6
5- I consider that the Member State I represent must attend or be represented at the following meetings / subgroups (RCG LP internal subgroups)								
i. (Stage 1) Data need and data gaps	5.57	5	6	6	6	5	5	6
ii. (Stage 2) Design of RSP for Tropical tunas	3.86	1	6	6	6	1	1	6
iii. (Stage 2) Design of RSP for Longline fisheries outside the MED	3.57	1	6	6	3	1	2	6
iv. (Stage 2) Design of RSP for Longline fisheries in the MED	3.86	5	6	1	1	5	6	3
v. (Stage 2) Design of RSP for bluefin tuna	4.57	6	6	6	1	3	4	6
vi. (Stage 3) RCG LP main meeting	5.86	5	6	6	6	6	6	6
If you answered the with 3 or less to questions 5, which would you consider the main reason?								
i. Lack of human resources	4.33		na		4	5	4	
ii. Lack of expertise	3.33		na		5	2	3	
iii. I am not interested in some specific groups	5.75	6	na		6	6	5	
iv. Overall, the proposed RCG structure is too complex	2.33		na		1	4	2	
v. Other (Please specify)								(1)
6- I consider that RCG LP must attend or be represented at the following Pan Regional meetings / subgroups (Inter RCG subgroups)								
i. Data analysis and quality	5.57	4	5	6	6	6	6	6
ii. End users and RCGs	5.14	6	5	6	5	4	4	6
iii. Governance	4.43	3	5	6	5	3	3	6
iv. Regional sampling plans	5.29	5	5	6	6	4	5	6
v. Implications of management measures in data collection (Landing Obligation)	4.00	3	5	2	6	1	5	6
vi. Other				LM				
7- I support the idea of providing LP data to a Regional Data Base, for DCF requirements (Current RDBES or any other)	4.83	5	3		6	6	3	6
General comments to the proposed structure for the Large Pelagic Regional Coordination:		RCG Chair: As a comment to Question 5, although Large Pelagics have specific issues due to the worldwide distribution of stocks, splitting the NR and Scientists in more Groups and meetings seems to be an obstacle for CPC participation						

(1) Some EU.PRT vessels have licenses to fish in the Med but currently no vessel is using it.

Table 2. Survey result by MS (WP2). 1 (completely disagree), 2 (disagree), 3 (slightly disagree), 4 (slightly agree), 5 (agree), 6, (completely agree) na (not applicable), empty (no answer).

Questions Block 2: Design of Regional Sampling Plan on Large Pelagic (RSP-LP)		MEMBER STATE						OTHER
Please answer to the following questions where 1 is completely disagree and 6 completely agree	Mean value	Malta	Spain	Portugal	France	Cyprus	Greece	RCG-LP chair
8- Several data were identified (see table1) essential to design a RSP (in relation to the two study cases).								
i. The variables included in table 1 cover all the needs of the tropical tuna stocks in the Atlantic. (if not please specify)	6.00	na			6	na	na	na
ii. The variables included in table 1 cover all the needs of the Mediterranean swordfish stock (if not please specify)	5.67	6			na	5	6	na
iii. I agree with the variables selected as priority for the tropical tuna stocks in the Atlantic. (if not please specify)	6.00	na			6	na	na	na
iv. I agree with the variables selected as priority for the Mediterranean swordfish stock (if not please specify)	5.00	5			na	5	5	na
9- SharePoint (Project or RCG-LP official SharePoint) seems to be appropriate for the data sharing, until there is an RBD implemented that permits the exchange of LP data.	4.80	4			6	5	3	6
10- Do you think feasible that all Mediterranean countries share a commonly agreed protocol for the biological sampling of swordfish catches/landings?	5.33	5			na	6	5	na
11- Do you think that the sampling protocol should be common at the RCG level? or different scale for each country	5.40	6			4	6	5	6
12- The minimum sampling intensity on swordfish landings by major Mediterranean area has been defined on a monthly/quarterly basis, based on predetermined precision levels. Would you agree to follow such a sampling protocol?	5.00	5			na	5	5	na
13- Concerning the data storage, provided that LP national data fit to the RDBES data model (Regional Data Base and Estimation System) would you support the idea of uploading LP data to this platform?	5.20	5			6	6	3	6
General comments to the design of the Regional Sampling Plans		RCG CHAIR: RSPs are an obvious goal of any RCG. But it requires that all (or at least the most relevant) CPCs are present and the representatives have the knowledge and the position to make decisions						

Table 3. Survey result by MS (WP3). 1 (completely disagree), 2 (disagree), 3 (slightly disagree), 4 (slightly agree), 5 (agree), 6, (completely agree) na (not applicable), empty (no answer).

Block 3 Procedures to assess the quality of biological data collected at regional level.		MEMBER STATE						OTHER
Please answer to the following questions where 1 is completely disagree and 6 completely agree	Mean	Malta	Spain	Portugal	France	Cyprus	Greece	RCG-LP chair
14- R packages are useful tools to evaluate the quality of the data.	6.00	6			6	6	6	6
15- I prefer other languages or platforms different from R. (if yes, please specify)	1.60	3			2	1	1	1
16- The R Quality assessment package developed in RECOLAPE uses two kind of inputs data: a format definition and a data file to be checked. The format definition is an R object or a xls/xlsx file, and the data file could be an R object or a csv/xls/xlsx file. This format satisfies my needs.	5.40	5			5	6	5	6
17- We currently have/use tools and procedures at national level to detect errors in the data collected under the DCF.	5.40	4			6	6	5	6
18- Procedures to assess the quality of biological data should be mainly focused on								
i. Data structure	5.60	5			6	6	5	6
ii. Spatial/temporal errors	5.60	5			6	6	5	6
iii. Species ID	5.40	5			6	6	4	6
iv. Other (please specify)								(1)
19- I consider that the Member State I represent has sufficient experts in age reading ages for the different LP species we are targeting.	2.40	3			1	2	5	1
20- At national level, age reading procedures are standardize in a formalized protocol/handbook? (If yes, could you please provide reference)	1.75	2				2	2	1
21- I consider of major relevance to establish a working group that through exchange-exercise/workshop improve the quality of age data by species	5.40	6			6	5	4	6
General comments to the design of the Regional Sampling Plans		RCG CHAIR: Common maturity scales for each species must be developed and workshops for calibration should be organized						

(1) Invalid values for each field like sizes below or above the admissible for a given species

Annexe 1. RECOLAPE SURVEY

MARE/2016/22 :Strengthening Regional cooperation in the area of large pelagic fisheries data collection (Acronym: RECOLAPE)

IDENTIFICATION

COUNTRY	
NAME	
PROFESSIONAL BACKGROUND (Biologist, Economist, Statistician, Administrative manager, ETC.)	
EXPERIENCE (Number of years you have been involved in the DCF)	

Annexe 1. RECOLAPE SURVEY

MARE/2016/22 :Strengthening Regional cooperation in the area of large pelagic fisheries data collection (Acronym: RECOLAPE)

European Commission

Questions Block 1: Large Pelagic Regional Coordination Group (RCG-LP) structure

RECOLAPE project makes a proposal for the future internal organization and structure of the RCG-LP (figure 1), exploring at the same time synergies and mechanism for coordination of this group among the various defined RCGs/regions.

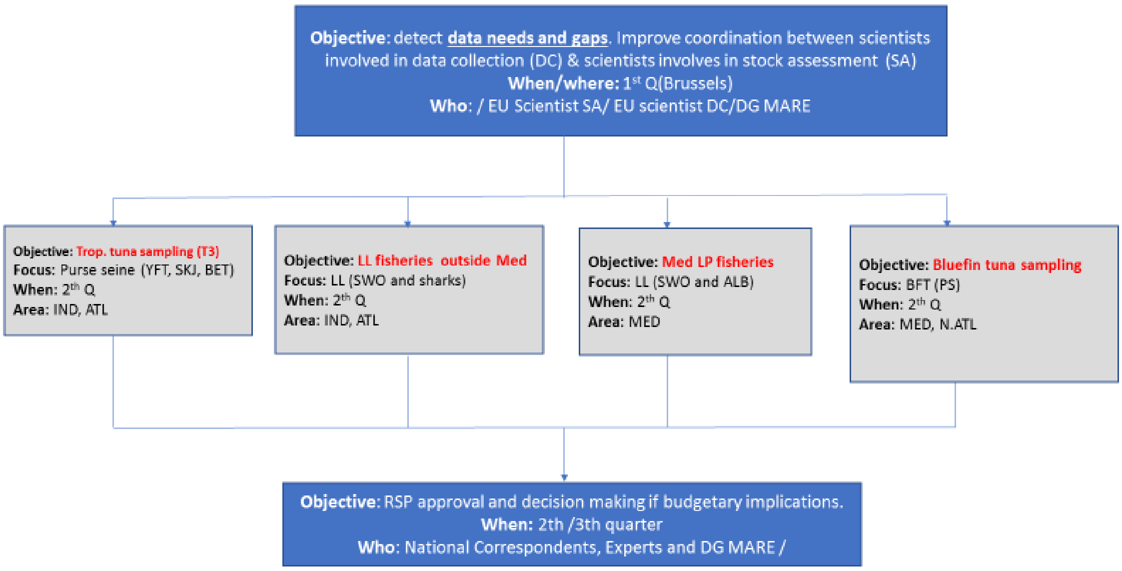


Figure1. Proposal for the future internal structure of the RCG-LP

This proposal considers 3 phases or stages for the internal RCG-LP functioning: Two preparatory meetings (1st and 2nd stages) prior to the RCG-LP main meeting (3rd stage).

- The **first stage** would have the purpose to identify **data gaps** and prioritize **LP data needs**, including tuna RFMO data requirements and data transmission.
- **Second stage** will be much more gear/stock specific. This second stage would **design Regional Sampling Plans** (RSP) by coordinating at the EU level dockside/onboard sampling for the different stocks. Ideally this coordination should be done in independent methodological subgroups dealing with specific fisheries: **1) Tropical tuna (this subgroup already exists). 2) Longline outside the MED. 3) Longline in the MED. 4) Bluefin tuna.**
- Finally, the **main RCG meeting (third stage)** will evaluate the outputs from the previous two stages, and it would be where **the final decisions** of greater importance and approval of Regional Sampling Plan should be made.

In a second step, the main common fields identified for cooperation among the RCGs are the once related with the need to put in common tools to design a Statistically Sound Sampling Regional Sampling Plans (RSP), and tools to evaluate the quality of the data (i.e. development of guidelines to evaluate the quality of the data, development of common software tools in R etc.). This requires (**Pan-regional) intersessional work among the different RCGs**. This intersessional work will facilitate a cooperation on a supra regional level and develop the work that is needed to fulfil future coordination tasks in a broad sense. Once the subgroups are created, it´s the responsibility of each of the RCGs to identify the experts with the needed skills by region that will work together to these issues.

A key issue to work in these common fields is to have a **Regional Data Base**, with data stored in common formats, ensuring transparency and consistent standards for data processing and dissemination. This is the current situation for the 3 northern RCGs (RCG NA, RCG NS&EA and RCG Baltic)

Annexe 1. RECOLAPE SURVEY

MARE/2016/22 :Strengthening Regional cooperation in the area of large pelagic fisheries data collection (Acronym: RECOLAPE)

European Commission

*Please answer to the following questions where **1 is completely disagree** and **6 completely agree***	
1- Do you consider that the RCG LP should be a group independent from the rest of RCGs?	1 2 3 4 5 6
i. If you answered the first question with a 3 or less, do you consider that it should be part of the RCG MED & BS?	1 2 3 4 5 6
ii. If you answered the first question with a 3 or less, do you consider that it should be part of the RCG LDF?	1 2 3 4 5 6
2- The global proposal to structure the RCG LP in 3 stages (i.e. identification of data needs/design of RSP/decision making) is appropriate	1 2 3 4 5 6
3- The number of subgroups or fisheries in stage 2 is adequate (1) Tropical tuna. (2) Longline outside the MED. (3) Longline in the MED. (4) Bluefin tuna	1 2 3 4 5 6
4- Globally the number of meetings (subgroups) proposed for the internal structure of the RCG LP is appropriate	1 2 3 4 5 6
5- I consider that the Member State I represent must attend or be represented at the following meetings / subgroups (RCG LP internal subgroups)	
i. (Stage 1) Data need and data gaps	1 2 3 4 5 6
ii. (Stage 2) Design of RSP for Tropical tunas	1 2 3 4 5 6
iii. (Stage 2) Design of RSP for Longline fisheries outside the MED	1 2 3 4 5 6
iv. (Stage 2) Design of RSP for Longline fisheries in the MED	1 2 3 4 5 6
v. (Stage 2) Design of RSP for bluefin tuna	1 2 3 4 5 6
vi. (Stage 3) RCG LP main meeting	1 2 3 4 5 6
If you answered the with 3 or less to questions 5, which would you consider the main reason?	
Lack of human resources	1 2 3 4 5 6
Lack of expertise	1 2 3 4 5 6
I am not interested in some specific groups	1 2 3 4 5 6
Overall, the proposed RCG structure is too complex	1 2 3 4 5 6
Other (Please specify)	
6- I consider that RCG LP must attend or be represented at the following Pan Regional meetings / subgroups (Inter RCG subgroups)	1 2 3 4 5 6
i. Data analysis and quality	1 2 3 4 5 6
ii. End users and RCGs	1 2 3 4 5 6
iii. Governance	1 2 3 4 5 6
iv. Regional sampling plans	1 2 3 4 5 6
v. Implications of management measures in data collection (Landing Obligation)	1 2 3 4 5 6
vi. Other	Please specify
7- I support the idea of providing LP data to a Regional Data Base, for DCF requirements (Current RDBES or any other)	1 2 3 4 5 6
General comments to the proposed structure for the Large Pelagic Regional Coordination:	

Annexe 1. RECOLAPE SURVEY

MARE/2016/22 :Strengthening Regional cooperation in the area of large pelagic fisheries data collection (Acronym: RECOLAPE)

Questions Block 2: Design of Regional Sampling Plan on Large Pelagic (RSP-LP)

The design of the RSP of the different LP stocks will be a primary task for the future RCGs, which, as indicated in the previous section, could be carried out in fisheries specific subgroups. Thus, RECOLAPE project makes a proposal for an RSP for two of the main LP fisheries; **Mediterranean swordfish targeted by longlines** and **tropical tuna purse seine fleet operating in the Atlantic Ocean**. In both study cases the following tasks have been carried out:

a) **Identification of data needs and priorities:** This is essential to design an RSP-LP (table1) and to perform robust estimates of the catch levels and size composition

b) **Data sharing:** Data exchange is essential for the RSP design. However, until there is a common mechanism such as the RDB (Regional Data Base) that allows data to be shared quickly and safely, data sharing can be supported through specific data call by the RCG-LP or conducted by projects such RECOLAPE in agreement with the RCG-LP, and data can be stored in SharePoint (RCG LP SharePoint or specific project SharePoint such as RECOLAPE's SharePoint). In addition, some other data sharing mechanisms have been tested for some specific biological data (i.e. SmartDots online platform for the exchange of samples of hard structures for fish ageing)

c) A **common sampling protocol** has been proposed, and regional **sampling design simulated** to **check its feasibility**. Through this simulation, some changes to the current sampling scheme are proposed: such as the definition of new sampling stratum for tropical tunas in the Atlantic, or a proposal for new sampling effort by stratum in order to increase representativeness of the size distribution and the accuracy of the species composition).

d) Finally, a proposal for **data storage and management solution** has been made through the test of the integration of large pelagic data in the Regional Data Base and Estimation System (RDBES, hosted and maintained by the International Council for the Exploration of the Sea) This is the current platform used by the 3 northern RCGs (RCG NA, RCG NS&EA and RCG Baltic)

*Please answer to the following questions where **1 is completely disagree** and **6 completely agree***

8- Several data were identified (see table1) essential to design a RSP (in relation to the two study cases).	
i. The variables included in table 1 cover all the needs of the tropical tuna stocks in the Atlantic. (if not please specify)	1 2 3 4 5 6
ii. The variables included in table 1 cover all the needs of the Mediterranean swordfish stock (if not please specify)	1 2 3 4 5 6
iii. I agree with the variables selected as priority for the tropical tuna stocks in the Atlantic. (if not please specify)	1 2 3 4 5 6
iv. I agree with the variables selected as priority for the Mediterranean swordfish stock (if not please specify)	1 2 3 4 5 6
9- SharePoint (Project or RCG-LP official SharePoint) seems to be appropriate for the data sharing, until there is an RBD implemented that permits the exchange of LP data.	1 2 3 4 5 6
10- Do you think feasible that all Mediterranean countries share a commonly agreed protocol for the biological sampling of swordfish catches/landings?	1 2 3 4 5 6
11- Do you think that the sampling protocol should be common at the RCG level? or different scale for each country	1 2 3 4 5 6
12- The minimum sampling intensity on swordfish landings by major Mediterranean area has been defined on a monthly/quarterly basis,	1 2 3 4 5 6

based on predetermined precision levels. Would you agree to follow such a sampling protocol?	
13- Concerning the data storage, provided that LP national data fit to the RDBES data model (Regional Data Base and Estimation System) would you support the idea of uploading LP data to this platform?	1 2 3 4 5 6

General comments to the design of the Regional Sampling Plans

Annexe 1. RECOLAPE SURVEY

MARE/2016/22 :Strengthening Regional cooperation in the area of large pelagic fisheries data collection (Acronym: RECOLAPE)

Block 3 Procedures to assess the quality of biological data collected at regional level.

Development of procedures to evaluate the quality of the data is a key issue for the RCGs. Thus, **RECOLAPE project proposes some data quality assessment procedures, both at national and at regional level**.

In this context, an **R package** has been developed to ensure the **compatibility of any dataset produced, compared with a defined format of our interest** (such as the SDEF format used several times by the RCG LP for data calls, RDBES format, etc.). This package performs some format quality assessments such as the checking of the concordance of the codes used with respect to a reference list, or values if intervals for numeric type). In addition, the project proposes a detailed annual calendar for the implementation of these quality checks. This calendar was established based in the dates of the last 4-5 years data calls and working groups.

Finally, especial attention has been put on the quality procedures and tools to improve age data, focusing on the comparison and exploratory analysis of age-length keys (ALK) and the development of tools to coordinate age reading between Member States. The exploratory analysis highlighted that there are still several issues to be clarified. Indeed, despite the number of meetings and exchange exercises done, the precision in fish ageing, in terms of percentage of agreement and coefficient of variation, are still out of the acceptable limits. In particular; reader experience, ageing protocol (e.g. preparation methods, age scheme) and ageing criteria (e.g. definition of first winter ring).

Please answer to the following questions where __1 is completely disagree__ and __6 completely agree__

14- R packages are useful tools to evaluate the quality of the data.	1 2 3 4 5 6
15- I prefer other languages or platforms different from R. (if yes, please specify)	1 2 3 4 5 6
16- The R Quality assessment package developed in RECOLAPE uses two kind of inputs data: a format definition and a data file to be checked. The format definition is an R object or a xls/xlsx file, and the data file could be an R object or a csv/xls/xlsx file. This format satisfies my needs.	1 2 3 4 5 6
17- We currently have/use tools and procedures at national level to detect errors in the data collected under the DCF.	1 2 3 4 5 6
18- Procedures to assess the quality of biological data should be mainly focused on	
i. Data structure	1 2 3 4 5 6
ii. Spatial/temporal errors	1 2 3 4 5 6
iii. Species ID	1 2 3 4 5 6
iv. Other (please specify)	
19- I consider that the Member State I represent has sufficient experts in age reading ages for the different LP species we are targeting.	1 2 3 4 5 6
20- At national level, age reading procedures are standardize in a formalized protocol/handbook? (If yes, could you please provide reference)	1 2 3 4 5 6
21- I consider of major relevance to establish a working group that through exchange-exercise/workshop improve the quality of age data by species	1 2 3 4 5 6