# REVIEW OF DETECTED ANOMALIES IN SIZE FREQUENCY DATA SUBMITTED TO THE SECRETARIAT

**PREPARED BY: IOTC SECRETARIAT LAST UPDATED: 25[TH] NOVEMBER 2020**

## Abstract

*This document constitutes a review of the processing applied to the size data submitted to the Secretariat, including the current species-specific measurement types, and provides compelling examples of the major issues encountered with some data sets available in the IOTC database. It illustrates four major cases: data inconsistencies, truncation of size distributions, duplication of size data, and discrepancies in size data between sources. Through these examples, the paper aims to recall the need for each CPC to (i) provide a thorough and up-to-date description of the data collection and curation systems in place and (ii) scrupulously follow IOTC reporting guidelines so as to reduce as much as possible the exclusion of data not compliant with IOTC standards.*

## Purpose

To provide participants at the 16th Working Party on Data Collection and Statistics (WPDCS16) with:

1. An overview of the types of size measurement available in the IOTC database;
2. A description of the processing performed on the size frequency data submitted to the Secretariat;
3. Some compelling examples of apparent anomalies detected in size frequency data sets submitted to the Secretariat.

## Background

IOTC Resolution 15/02 on *"Mandatory statistical requirements for IOTC Members and Cooperating Non-Contracting Parties (CPC's)"* states that size data shall be submitted to the Secretariat for all species under the IOTC mandate as well as for the most commonly caught elasmobranch species, in accordance with IOTC reporting guidelines and, possibly, through Form 4SF in agreement with the IOTC reporting guidelines. The resolution also states that the size sampling shall be representative of all periods and areas fished, and that it shall cover *at least one fish by ton caught, by species and type of fishery*. Finally, documents describing in details the sampling and raising procedures adopted by CPCs shall regularly be provided together with the data sets.

## Materials & Methods

### Form 4SF

IOTC Form 4SF is designed to accommodate several metadata describing the processing applied by the source (e.g. validation and raising level), together with the measurement type and measuring tool which can be entered through drop-down menus built on top of standard code lists (see reporting guidelines). The "SIZE INTERVAL" cell is an unconstrained, numeric cell which is expected to include the unit of measure for the chosen measurement type and the value of the size intervals, assumed to be constant[1] for the whole sample, and the form does not enforce any

[1] The actual value of the constant size interval can also be inferred from the actual data, which include both low and high size classes

control on the minimum and maximum sizes of the individuals sampled. Each individual length (or weight) measurement shall be rounded to the lowest measurement unit in the size class.

Both original and raised size data can be submitted to the Secretariat, along with the description of the sampling design and processing steps: a general *data quality* flag (assuming one of the values among *good*, *fair*, *poor* and *unclassified*) is assigned to each individual size data set according to the underlying knowledge of the data collection and verification system in place, and to the consistency of the data submitted when compared to previous years (see paper IOTC-2020-WPDCS16-10).

Size data submitted to the Secretariat include a large diversity of measurement types that depend on: i) the sampling location and platform (fishing boat, landing site, factory, etc.), ii) the type of processing performed on the fish (e.g. beheaded, gilled, gutted, etc.), iii) the sampling protocol (e.g. measurement of fork length vs. total length), and iv) the measuring tool (e.g. curved length) (**Table 1**). Conversions of the raw size data may have been performed prior to submission and there is a general lack of information available from the CPCs on the conversion methods used, and whether they are consistent with the IOTC reference morphometric relationships (see below). In addition, measurements can be made with different measuring tools (e.g. tape measure, measuring board, calliper) although the IOTC reporting guidelines strongly recommend the use of callipers.

**Table 1.** Number of fish sampled for size (N) by type of measurement available in the IOTC database as of November 2020 for the data sets submitted as original samples. For most raised size data sets, the numbers of fish originally sampled were not submitted to the Secretariat

| Code | Description | N |
|---|---|---|
| CF | Cleithrum to caudal fork length | 12,926 |
| CKL | Cleithrum-keel length | 10,452 |
| CKUT | Cleithrum-keel length (unconverted tape m lengths) | 22 |
| CL | CL-Carapace length (turtles) (curved length) | 12 |
| DML | Dorsal mantle length (squid) | 1 |
| EFL | Eye-Fork Length | 186,240 |
| EFUT | Eye-Fork Length (unconverted tape measure lengths) | 91 |
| FL | Fork length (lower jaw fork length for BIL) | 25,636,506 |
| FLB | Fork length (by using a Board) | 364,571 |
| FLC | Fork length (converted from weight/length) | 441,945 |
| FLCT | Fork length (converted tape measure lengths) | 757,937 |
| FLUT | Fork length (unconverted tape measure lengths) | 2,155,705 |
| GGT | Gilled and gutted (bill is off for billfish) | 705,468 |
| GIL | Gilled weight | 7,728 |
| HDD | Headed and gutted weight | 7,246 |
| LDF | First dorsal fin-fork length | 3,022 |
| LDFT | First dorsal fin-fork length (Tape measure length) | 75 |
| PAL | Pectoral-anal length (by using a calliper) | 5,431 |
| PALT | Pectoral-anal length (by using a tape measure) | 2,788 |
| PCL | Pectoral-caudal(fork) length | 32,722 |
| PCLT | Pectoral-fork length (unconverted tape length) | 27 |
| PPD | Headed and peduncle-off | 16 |
| RND | Round Weight | 288,787 |
| TL | Total length | 12,316 |
| TW | Disc width | 166 |
| WR | Round weight | 70 |

## Processing steps

It is important to recall here that several size-frequency data sets submitted to the Secretariat cannot be processed due to the lack of information on key variables such as month, grid or species (e.g. aggregated species such as BILL): in some cases, raw size data are not included in the submission which only provides summary statistics for the fishery and species concerned.

When the data submitted are consistent with the requirements of Res. 15/02 and in agreement with the structure of Form 4-SF, the first processing step consists in applying a filter to the provided size data in order to remove all fish sizes that are reported with an interval larger than a maximum (species-specific) value. This filter aims to remove the size samples for which the precision is not compliant with the recommendations of the IOTC reporting guidelines, and that are therefore considered too poor for most scientific analyses (**Table A1**).

Over the period 1952-2018, the total number of samples removed from the size-frequency data sets made available to the Working Parties due to low precision in reporting was very small for the unraised size data sets (0.8%). By contrast, most size samples submitted without information on the data processing were associated with larger size bins, exceeding the recommended values (95.1%). This concerned neritic tunas, and Narrow-barred Spanish mackerel in particular, for which over 1.5 million size measurements were reported by Saudi Arabia in 5 cm size intervals between 1985 and 1992 (**Table 2**).

The analysis confirmed that all size samples already submitted by the original data providers in raised format were described by adequate precision, likely because the data processing applied by the source already included similar filtering methods. The raised data sets submitted since 1982 to the Secretariat are dominated by European and associated purse seine fisheries, but include also other CPCs such as Australia and Indonesia.

**Table 2.** Number of size samples by species submitted to the Secretariat with a size interval larger than recommended in the IOTC reporting guidelines and considered necessary for scientific analyses. **OS** = original sample (i.e. unraised data); **UNCL** = unclassified (i.e. no information available on the data processing)

| Processing | Species group | Species | English name | Size Interval | N |
|---|---|---|---|---|---|
| OS | BILL | SWO | Swordfish | 10 | 1,792 |
| OS | NERI | BLT | Bullet tuna | 5 | 125 |
| OS | NERI | COM | Narrow-barred Spanish mackerel | 5 | 16,143 |
| OS | NERI | FRI | Frigate tuna | 3 | 690 |
| OS | NERI | FRI | Frigate tuna | 5 | 521 |
| OS | NERI | GUT | Indo-Pacific king mackerel | 5 | 1,231 |
| OS | NERI | KAW | Kawakawa | 3 | 3,879 |
| OS | NERI | KAW | Kawakawa | 5 | 12,518 |
| OS | NERI | LOT | Longtail tuna | 5 | 1,716 |
| OS | TEMP | ALB | Albacore | 5 | 1,365 |
| OS | TEMP | ALB | Albacore | 10 | 190 |
| OS | TEMP | ALB | Albacore | 5 | 2,300 |
| OS | TROP | BET | Bigeye tuna | 5 | 1,921 |
| OS | TROP | BET | Bigeye tuna | 10 | 478 |
| OS | TROP | SKJ | Skipjack tuna | 3 | 26,479 |
| OS | TROP | SKJ | Skipjack tuna | 5 | 91,614 |
| OS | TROP | SKJ | Skipjack tuna | 5 | 159 |
| OS | TROP | SKJ | Skipjack tuna | 5 | 678 |
| OS | TROP | YFT | Yellowfin tuna | 5 | 58,878 |
| OS | TROP | YFT | Yellowfin tuna | 10 | 924 |
| OS | TROP | YFT | Yellowfin tuna | 5 | 1,948 |
| OS | TROP | YFT | Yellowfin tuna | 5 | 1,905 |
| UNCL | NERI | COM | Narrow-barred Spanish mackerel | 5 | 1,530,391 |
| UNCL | NERI | FRI | Frigate tuna | 3 | 15,986 |
| UNCL | NERI | KAW | Kawakawa | 3 | 23,524 |
| UNCL | TROP | SKJ | Skipjack tuna | 3 | 7,159 |

The second step of the internal IOTC data processing consists in converting non-standard into standard measurement types, which are typically fork length[2] (FL) for tuna and sharks, and eye fork length[3] (EFL) for black and blue marlins (**Table A1**). The conversions are performed through conversion reference methods composed of both deterministic relationships and proportion keys (see IOTC-2020-WPTT22(AS)-DATA13). It is worth recalling that while some of the relationships adopted are inferred from a large sample collected in the Indian Ocean, several of them are based on a small sample size or borrowed from other oceans instead: the effects of uncertainties in conversion methods on final size distribution of the catch have not been fully explored, but are suspected to be the cause explaining some inconsistencies observed in the length-frequency data sets of some longline fleets (Geehan and Hoyle 2013).

The third processing step applied to the measurements, once these are transformed into standard lengths, consists in further converting all measurements reported in size intervals larger than the standard ones (yet sufficiently small to remain valid, according to the data reporting guidelines) in standard length intervals, by proportionally assigning the initial number of fish to equally spaced smaller intervals. For instance, the standard size interval for albacore tuna is 1 cm fork length while the filter defined above allows to retain size data reported in both 1 and 2 cm size intervals: in this case, if 1000 fish are reported in the fork length interval 80-82 cm, 500 fish will be assigned to the class 80-81 cm and 500 fish will be assigned to the class 81-82 cm.

In the fourth processing step, all the fish whose initial size converted to standard length turns out to be smaller than a species-specific minimum size (as defined by the "first size class" column) are aggregated within the minimum allowed size class for the species (**Table A1**).

Similarly, a final filter is applied to the standard length data to remove all those size measurements that are larger than a species-specific maximum size (data are considered inconsistent with the species biology) (**Table A1**).

When considering the unraised size samples submitted in standard length, a total of 17,292 size measurements were found to be larger than the species-specific maximum sizes (**Table 3**): overall, this only represented 0.1% of all fishes submitted in standard length between 1952 and 2018. Several of the inconsistent records were reported by I.R. Iran for kawakawa caught with gillnet during 2008-2018 and of fork length >70 cm (**Table A1**). However, it is noteworthy that this maximum size limit is lower than the maximum size of 100 cm FL which is instead available in FishBase: hence, we recommend that all size limits currently implemented in the IOTC database are carefully scrutinized, and updated to scientifically sounder values (when required) in a near future.

**Table 3.** Number of size samples by species submitted to the Secretariat with a size larger than the maximum size consistent with the biology of the species. Data only include samples submitted in fork length in unraised format

| Species group | Species | English name | N |
|---|---|---|---|
| NERI | BLT | Bullet tuna | 121 |
| NERI | FRI | Frigate tuna | 615 |
| NERI | GUT | Indo-Pacific king mackerel | 845 |
| NERI | KAW | Kawakawa | 13,377 |
| NERI | LOT | Longtail tuna | 2 |
| SKH | PSK | Crocodile shark | 7 |
| TEMP | ALB | Albacore | 2,278 |
| TROP | BET | Bigeye tuna | 1 |
| TROP | SKJ | Skipjack tuna | 45 |
| TROP | YFT | Yellowfin tuna | 1 |

## Some examples of anomalies

We provide hereafter some compelling examples of issues in the size data submitted to the Secretariat which suggests a lack of quality control procedures implemented in the data collection and management systems of some CPCs. In

[2] Defined as the straight distance from the tip of the upper jaw to the fork of the tail
[3] Defined as the straight distance from the tip of the lower jaw to the fork of the tail

addition to the extra work required by the Secretariat to carefully check and filter the data, such issues may result in some data sets not being disseminated to the IOTC Scientific Community despite the monetary costs and research efforts spent on the collection and management of these data.

## Data inconsistencies

According to the EU and Seychelles observer protocols, all the measurements of sharks must be made in total length[4]. The distribution of silky shark (*Carcharhinus falciformis*) size data collected by human observers at sea illustrates some major inconsistencies encountered in the data received at the Secretariat (**Fig. 1**). First, several size measurements are below 65 cm which is the minimum size at birth for silky shark (e.g. Varghese et al. 2016). Second, and despite the removal of measurements below the birth threshold, the distribution appears to be skewed to the left, i.e. some values below 65 cm would be consistent with the overall size distribution: this may suggest that many of the size measurements were originally taken as *fork length* which is – by definition – lower than *total length*. Third, the distribution presents several equally spaced spikes, suggesting that many values were rounded or approximated at the time of measurement, possibly in consequence of a visual exam.

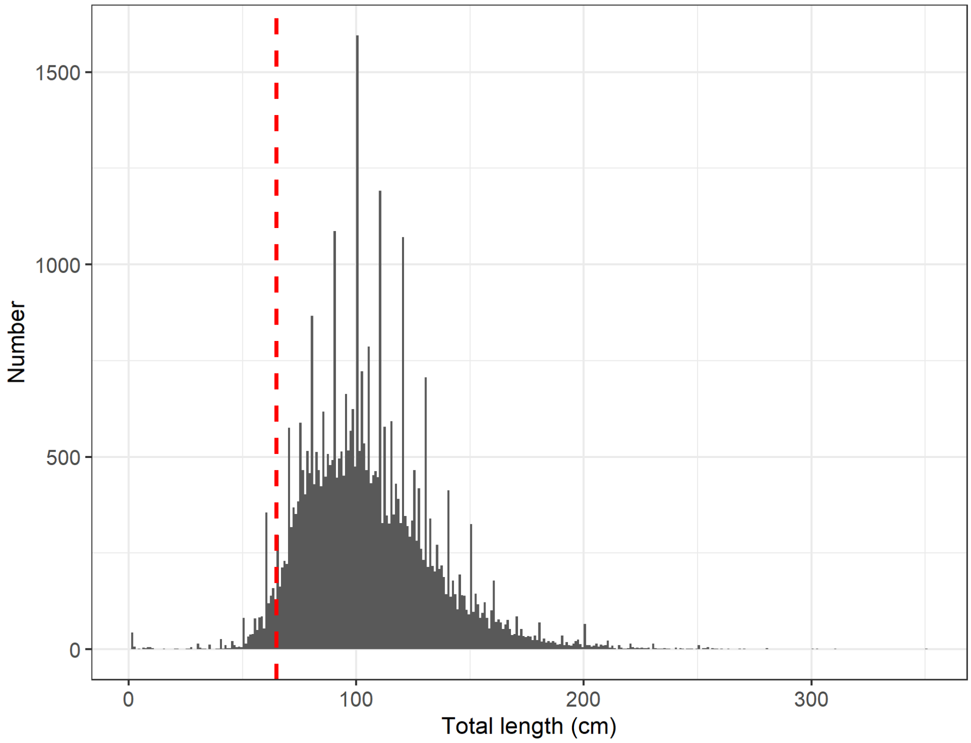


**Fig. 1.** Total length distribution (1-cm size classes) of silky shark caught by EU and Seychelles purse seiners and measured by at-sea observers during 2005-2019. Red vertical dashed line indicates the total length at birth according to IUCN

## Size distribution truncation

Data sets for other species and fisheries show major truncations in the final distribution, which suggest that the sampling operations were made at a later stage, following a size-based selection process (for instance for market or processing reasons), or that some kind of filter was applied to the data during the curation process.

For instance, fork length data of skipjack tuna (*Katsuwonus pelamis*) submitted by Sri Lanka for some of their fisheries show a distribution truncated to the left at about 70 cm and to the right at 25 cm for gillnetters in 2017 and ring-netters in 2016, respectively (**Fig. 2**).

[4] Defined as the projected straight distance from the most forward point of the head to the tip of the tail when the tail is left in the 'natural position', i.e. unsqueezed

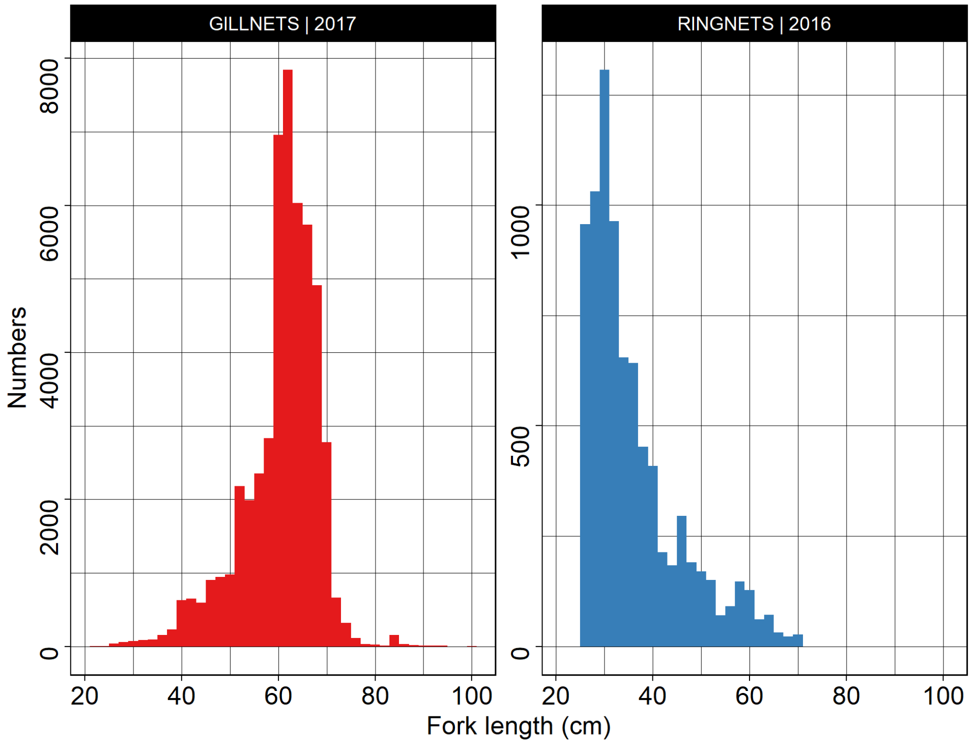


**Fig. 2.** Fork length distribution (1-cm size classes) of skipjack tuna caught by Sri Lankan (left panel) gillnetters in 2017 and (right panel) ring-netters in 2016 and measured at landing site

## Size data duplication

Another issue encountered with some size data sets is the apparent duplication of data from one year to the other, which results in the distributions to show very similar, when not identical, peaks and modes at the same size intervals. This is for instance the case detected with the annual size distributions of skipjack tuna caught by Sri Lankan gillnetters that show very similar patterns and suggest potential duplication of the data from 2016 to 2017 and 2019 (**Fig. 3**).

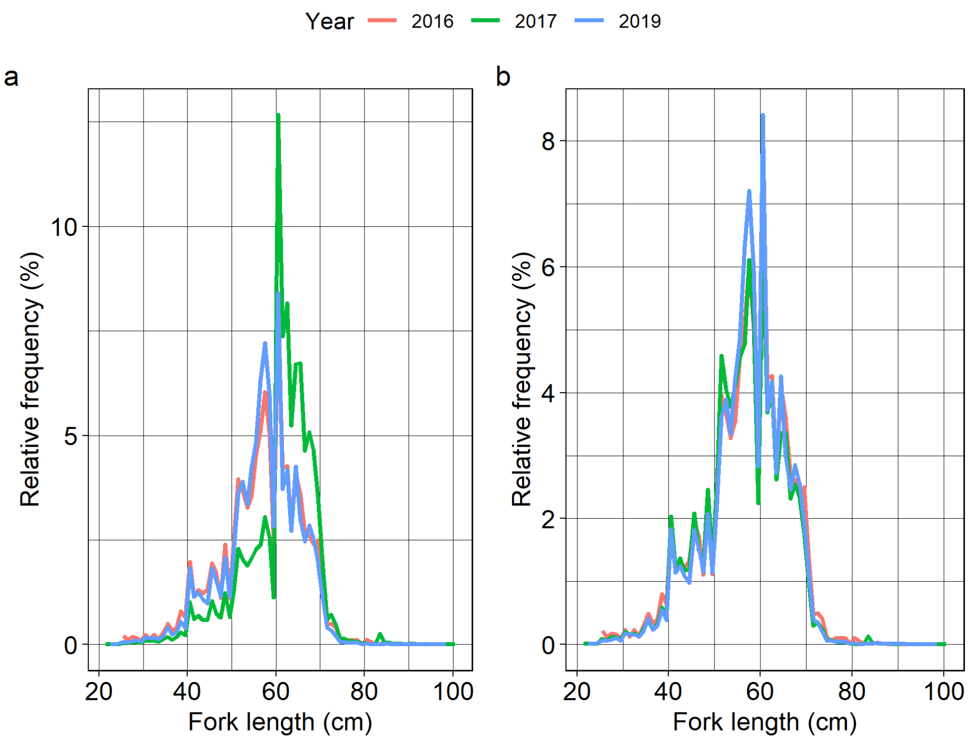


**Fig. 3.** Annual relative size-frequency distributions of skipjack tuna caught by Sri Lankan gillnetters in 2016, 2017 and 2019. (a) Raw data, (b) Same data after multiplying relative numbers of fish sampled in 2017 by 2 for fish <60 cm and dividing by 2 the numbers of fish >60 cm

## Discrepancies between data sources

As part of the improvement of its quality control process (see paper IOTC-2020-WPDCS16-10), the Secretariat is developing cross-check procedures between official data sources. In particular, the implementation of national observer programs under Res. 11/04 provides an additional source of information on the size composition of retained catch from several industrial longline fisheries.

As the size frequency data for yellowfin tuna, bigeye tuna, and albacore tuna collected by fishermen on Taiwanese longliners have been excluded in recent assessments for the period from 2000 onward, size data from observers may constitute a valuable alternative data source, as well as a way for exploring the reasons for the detected inconsistencies (Geehan and Hoyle 2013).

In the case of yellowfin tuna, the comparison of annual size histograms shows some similar patterns (e.g. 2011 or 2019) but also illustrates some major discrepancies between logbook and observer data sources in some years, with more yellowfin tuna in the size-range 150-175 cm reported during 2012, 2013, 2016 and 2018 (**Fig. 4**).

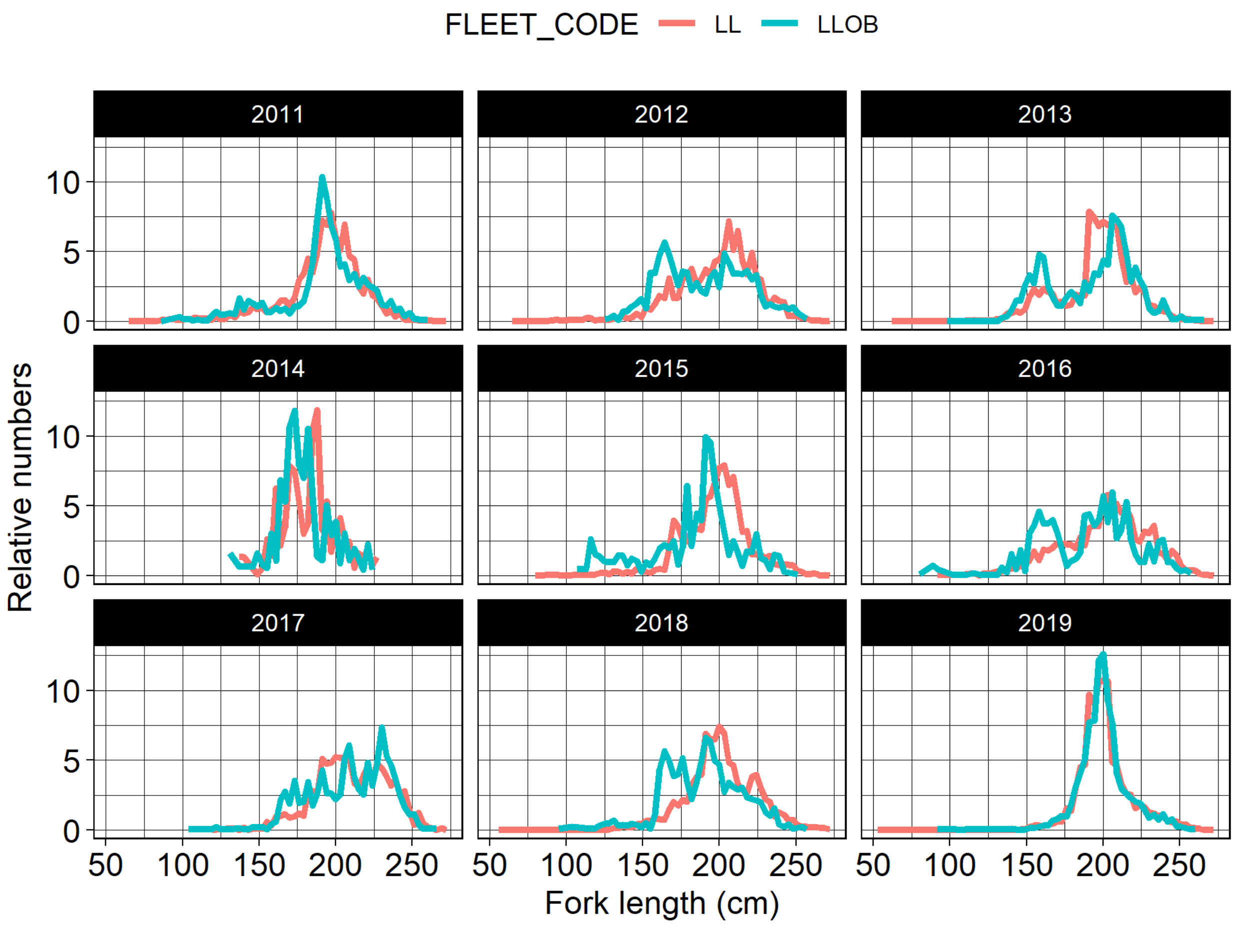


**Fig. 4.** Annual distribution of fork length (3-cm size classes) of yellowfin tuna caught by Taiwanese large-scale longliners and measured at sea by fishermen (LL; red) and observers (LLOB; blue) during 2011-2019

# Conclusions

The checking and processing steps performed by the Secretariat are based on a set of simple, reproducible rules which aim to select and prepare the size frequency data in a way that is suitable for the support of scientific analyses dealing with tuna and tuna-like fisheries. In absence of description of the data collection and curation required by Res. 15/02 and regarding the highly aggregated character of size frequency data available at the Secretariat (i.e. aggregated at a spatial resolution of 5°x5°), it is generally difficult to *separate the wheat from the chaff* and some large data sets may be put aside while they include some valuable information which might not be available elsewhere (e.g. size-frequency data from Taiwanese driftnetters operating in the Northern Arabian sea during the '90s, that were initially lacking the geo-spatial information required by Res. 15/02). In this context, it is essential for each CPC to develop and implement well described and transparent quality control procedures to verify their size data sets. Although some control can be conducted by the Secretariat on the aggregated data, it is always better to perform the checks on the original data available at the operational level in order to identify and correct the errors (e.g. data entry errors).

The development of data fact sheets by the Secretariat is a first step to guarantee almost immediate feedback to each CPC following data submission, and is a future guarantee towards the improvement of size data submissions.

In addition, the Secretariat aims to further develop guidelines for the systematic verification of each data set (including other information than size frequency) and encourages all CPCs to accurately follow the IOTC data reporting guidelines, and describe all steps of their data collection and curation processes.

In a near future, the Secretariat is expected to focus its efforts towards improving the processing of size data (e.g. raising procedure) to better account for the sampling coverage and spatial-temporal fishery patterns whenever this type of data processing is not already performed at the operational level.

The Secretariat is also committed to provide full support to all CPCs interested in improving their data curation workflow, under the confidentiality rules set forth by Res. 12/02.

# References

Geehan J, Hoyle S (2013) Review of length frequency data of the Taiwanese distant water longline fleet. IOTC–2013–WPTT15–41_Rev1. IOTC, San Sebastian, Spain, 23-28 October 2013, 30p

Varghese SP, Gulati DK, Unnikrishnan N, Ayoob AE (2016) Biological aspects of silky shark *Carcharhinus falciformis* in the eastern Arabian Sea. Journal of the Marine Biological Association of the United Kingdom 96:1437–1447

# Appendix

**Table A1.** Standard length code, maximum and minimum standard lengths, and standard length class and maximum length class interval for the 16 IOTC species and sharks for which size frequency data have to be reported to the Secretariat following IOTC Res. 15/02. FL = straight fork length; EFL = straight eye fork length. All lengths and intervals are given in cm

| Species group | Species | English name | Standard length code | Maximum standard length | Minimum standard length | Standard length class interval | Standard maximum length class interval |
|---|---|---|---|---|---|---|---|
| BILL | BLM | Black Marlin | EFL | 465 | 15 | 3 | 5 |
| BILL | BUM | Blue Marlin | EFL | 500 | 15 | 3 | 5 |
| BILL | MLS | Striped marlin | FL | 420 | 15 | 3 | 5 |
| BILL | SFA | Indo-Pacific sailfish | FL | 300 | 15 | 3 | 5 |
| BILL | SWO | Swordfish | FL | 450 | 15 | 3 | 5 |
| NERI | BLT | Bullet tuna | FL | 50 | 10 | 1 | 2 |
| NERI | COM | Narrow-barred Spanish mackerel | FL | 240 | 10 | 1 | 3 |
| NERI | FRI | Frigate tuna | FL | 65 | 10 | 1 | 2 |
| NERI | GUT | Indo-Pacific king mackerel | FL | 76 | 10 | 1 | 3 |
| NERI | KAW | Kawakawa | FL | 70 | 10 | 1 | 2 |
| NERI | LOT | Longtail tuna | FL | 145 | 10 | 1 | 3 |
| SKH | ALV | Thresher Shark | FL | 760 | 30 | 5 | 10 |
| SKH | BSH | Blue shark | FL | 400 | 30 | 5 | 10 |
| SKH | BTH | Bigeye thresher | FL | 760 | 30 | 5 | 10 |
| SKH | FAL | Silky shark | FL | 350 | 30 | 5 | 10 |
| SKH | LMA | Longfin mako | FL | 417 | 30 | 5 | 10 |
| SKH | OCS | Oceanic whitetip shark | FL | 396 | 30 | 5 | 10 |
| SKH | POR | Porbeagle | FL | 350 | 30 | 5 | 10 |
| SKH | PSK | Crocodile shark | FL | 110 | 30 | 5 | 10 |
| SKH | SMA | Shortfin mako | FL | 400 | 30 | 5 | 10 |
| SKH | SPL | Scalloped hammerhead | FL | 430 | 30 | 5 | 10 |
| SKH | SPZ | Smooth hammerhead | FL | 500 | 30 | 5 | 10 |
| TEMP | ALB | Albacore | FL | 140 | 10 | 1 | 2 |
| TROP | BET | Bigeye tuna | FL | 250 | 10 | 2 | 4 |
| TROP | SKJ | Skipjack tuna | FL | 110 | 10 | 1 | 2 |
| TROP | YFT | Yellowfin tuna | FL | 239 | 10 | 2 | 4 |