A PROPOSAL FOR AN IOTC INTERACTIVE STATISTICAL DATA BROWSER

PREPARED BY: IOTC SECRETARIAT, LAST UPDATED: 23 NOVEMBER 2021

Purpose

To inform participants at the 17th Working Party on Data Collection and Statistics (WPDCS17) about the status of development of a new interactive statistical data browser that might replace the current IOTC Online Data Querying Services and improve access to the core IOTC datasets through filtering, display and analysis of the public records contained within.

Background

IOTC Online Data Querying Services

The Online Data Querying Services (OQS) have been available on a dedicated section of the IOTC website since its restructuring in 2013 (https://www.iotc.org/iotc-online-data-querying-service).

The OQS give public access to the raw nominal catch[1] and fishing craft datasets via a query panel that provides basic filtering features and the possibility of displaying the results of the filtering as:

- Column chart
- Bar chart
- Line chart
- Pie chart
- Tabular display

Filtered results can then be either viewed onscreen or downloaded locally, as a word document or as an image depending on the type of output selected by the user.

The possibility of interactively querying the core IOTC datasets and aggregate the results according to different criteria is an important service that IOTC must continue to provide to its target audience. Yet, based on the collected access statistics for the IOTC website, the current OQS enjoy moderate success with the IOTC scientific community.

The reasons for this are manifold, and include:

- Relative lack of awareness in the target user base
- Access to a limited number of datasets (nominal catch and fishing craft only)
- Limited number of filtering criteria and display options
- Difficult to maintain and evolve (the OQS underlying technical infrastructure is hosted outside of the IOTC Secretariat)
- Data is not always updated to the most recent information available to the Secretariat (the OQS is not directly linked to the IOTC databases)

The IOTC R-libraries

Starting in late 2020, the IOTC Secretariat has begun the development of a set of R[2] libraries to improve its internal data management workflow, standardize access to the core statistical datasets and provide the most common features

[1] Includes catch data for all species, as well as species and gear aggregates
[2] https://www.r-project.org/

for data access, manipulation and visualization both for off-line analysis and for the production of IOTC data papers and reports.

The IOTC R-libraries cover the majority of IOTC data assets, including the RAV[3], the ROS[4] regional database and the daily buoy position data (received by the Secretariat as per IOTC Res. 19/02) and are constantly under development by the data section of the Secretariat.

The IOTC R-libraries are not yet shared publicly, as they still require an active connection to the IOTC databases to function, and the Secretariat is actively working to fully decouple the IOTC data assets from the libraries, thus enabling their future dissemination to the IOTC scientific community.

The IOTC Data Browser

The R-Shiny[5] package greatly simplifies the task of building data dashboards on top of R scripts and libraries.

For this reason, the Secretariat started exploring the possibility of applying this technology on top of the IOTC R-libraries and datasets, to enable interactive data analysis and visualization and standardize with minimal effort the results and outputs of the analysis across different domains (IOTC official papers and data dashboards).

The IOTC Data Browser represents the current output of this exploration process and has now reached a stage of maturity adequate to consider its release to the public and eventually replace the official IOTC OQS.

Main access page

The entry page of the IOTC Data Browser is at http://data.iotc.org/browser/home and is currently password protected to prevent unauthorized access prior to its official release, with temporary credentials to be communicated during the meeting.

The main access page (**Fig. 2**) shows all available datasets, their metadata (**Fig. 1**), and the links to access their dedicated sections in the IOTC data browser.

Dataset metadata

Fig. 1: metadata for the raw nominal catch dataset

Each dataset listed in the IOTC data browser home page is presented with a list of metadata that describe the content and nature of its information (see Fig. 1).

The metadata provide the following information (from left to right, in **Fig. 1**):

Metadata field	Description
Number of records	Total number of records in the dataset
Date of last update	Date of last update for the dataset
Type of data	Raw / Disaggregated / Converted
Raising level	Raised to total catches / Unraised
Temporal resolution	Yearly / Quarterly
Type of measurement	Live weight / Estimated live weight / Numbers / Effort / Lengths / Fork lengths
Spatial resolution	IO main areas / Regular CWP grids
Species	All species / IOTC species / Albacore / Bigeye / Skipjack / Swordfish / Yellowfin

Table 1: Description of each dataset metadata

[3] https://www.iotc.org/vessels/current
[4] https://www.iotc.org/science/regional-observer-scheme-science
[5] https://shiny.rstudio.com/

Fig. 2: The IOTC data browser entry page, with the list of available datasets and their metadata

Available datasets

Currently the IOTC Data Browser gives access to the following datasets:

Nominal catches by year, IOTC area, fleet, fishery and species

- **Raw nominal catches** (**NC-RAW**): nominal catches in weight for all species, including bycatch ones. Reference dataset
- **Best scientific estimates of nominal catches** (**NC-SCI**): fully disaggregated, best scientific estimates of nominal catches in weight for IOTC species only. Reference dataset

Catch-and-effort by year, quarter, grid, fleet, fishery and species

- **Georeferenced effort data** (**CE-EF**)**:** reported georeferenced quarterly effort data. Reference datasets
- **Georeferenced catch data** (**CE-CA**): reported georeferenced quarterly catch data in weight and / or numbers. Reference datasets

Size-frequency by year, quarter, grid, fleet, fishery, species and measure

- **Raw size-frequency data** (**SF-RAW**): reported georeferenced size-frequency data
- **Standardized size-frequency data** (**SF-SCI**): georeferenced size-frequency data for IOTC and relevant shark species converted to fork length. Reference datasets: BET, SKJ, YFT, ALB, BILLFISH, NERITIC, SHARKS

Raised catches by year, quarter, 5x5 grid, fleet, fishery and species

This dataset is password-protected as the information included is not yet considered ready for public release to the IOTC scientific community, although it is used to produce stock assessment inputs for the five major IOTC species (Albacore tuna, Bigeye tuna, Skipjack tuna, Swordfish and Yellowfin tuna).

- **Raised georeferenced catch data** (**CA-RAISED**): estimated, raised georeferenced quarterly catches in weight and numbers for the five main IOTC species

Reference data and additional classification

The standard IOTC classifications and code lists are used throughout each section of the data browser, including the following data and filtering fields:

- Indian Ocean major area
- Fleet
- Species
- Fishery type
- Gear
- Species group
- Species working party
- Catch unit
- Effort unit
- Size measure type

Furthermore, and as part of the standardization exercise implemented by the IOTC R-libraries, additional categorization at species and gear level have been introduced, namely *IUCN status* and *Species category* for the species, and *Fishery group* and *Fishery* for the gears, with current extended categorizations adopted by the IOTC data browser provided in Appendix 1 as **Table A1** and **Table A2** respectively.

It is worth recalling that within IOTC the characterization of a gear as artisanal or industrial depends also by the reporting fleet, and the same gear can be considered differently in terms of its type (e.g., offshore gillnets - GIOF are considered artisanal when reported by Oman and industrial when reported by I.R. Iran).

Furthermore, the categorization of purse seine data by fishery (PSFS, PSLS or PSOT) is not available in the nominal catch datasets, where they all appear as PSOT.

User interface

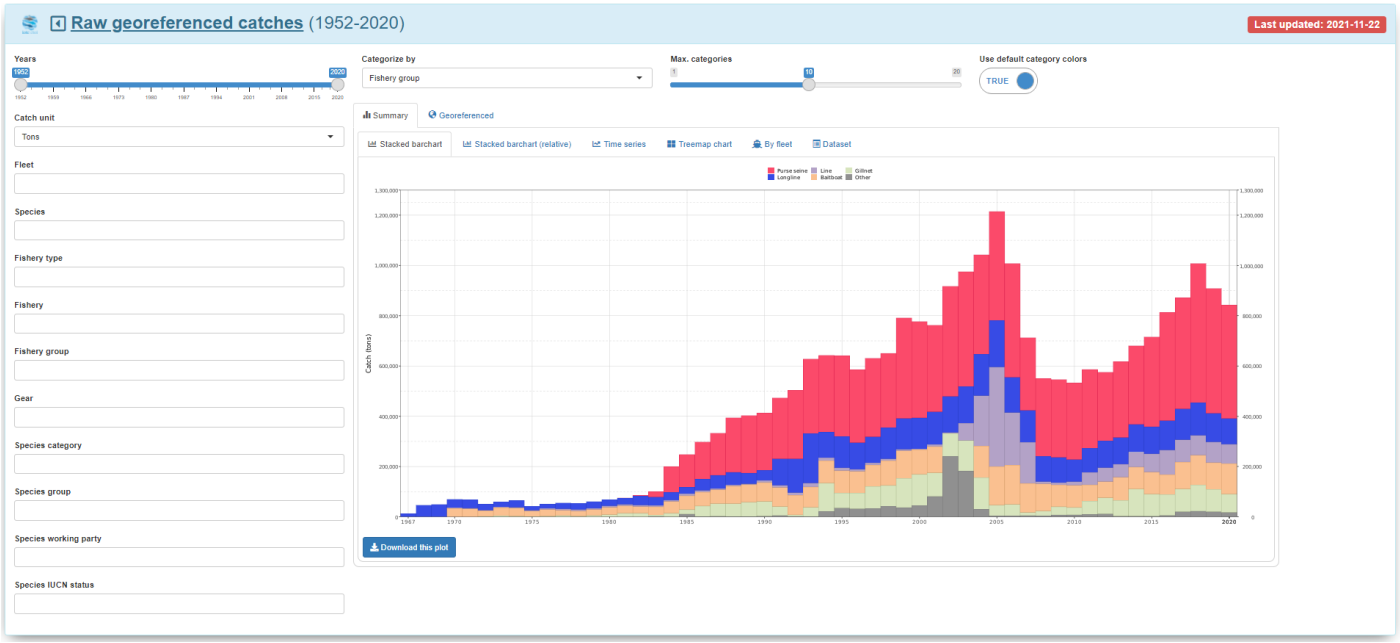


Fig. 3: example of user interface with all defaults selected (raw georeferenced catches)

The structure and behaviour of the user interface is standardized regardless of the selected dataset and is composed of:

- The **main filters panel** on the left
- The **categories' configurator panel** on the top right
- The **output panel** on the bottom right

With minor differences in the fields presented by the filtering panel or in the structure of the output panel depending on the dataset of interest.

Main filters panel

The main filters panel allows filtering the dataset by the most common criteria such as:

- ***Years*** - a period spanning between two years of the time series (1950-2020)
- ***IO major area*** (either Eastern or Western Indian Ocean) - only available for the nominal catch datasets
- ***Fleet***
- ***Catch unit*** (either metric tons or numbers) - only available for the georeferenced catch dataset
- ***Effort unit*** - only available for the georeferenced effort dataset
- ***Species***, ***Species category***, ***Species group***, ***Species working party***, and ***Species IUCN status*** - only available for the nominal catch and georeferenced catch datasets (see **Table 1**)
- ***Fishery type*** (either artisanal or industrial), ***Fishery group***, and ***Gear*** (see **Table 2**)
- ***Measure type*** and ***Raising*** (Original sample, Raised data or Unclassified) – only available for the raw georeferenced size-frequencies dataset

All filtering fields are presented as dropdown menus, with autocompletion enabled to filter out the available entries while typing, and the possibility of selecting multiple values for each filtering field (in general).

When filtering the underlying dataset, the IOTC data browsers combines all non-empty filtering criteria with an AND clause and all multiple filter values (when provided) with an OR clause.

Therefore, if the user has configured the following filtering criteria:

- ***Years*** = { `2010, …, 2020` }
- ***Fishery type*** = `ART - Artisanal fishery`
- ***Fishery group*** = { `PS - Purse seine, GN - Gillnet` }

- ***Species category code*** = { `TROPICAL – Tropical tuna species, BILLFISH – Billfish species` }

Then the system will return all records of the given dataset reported for the years between 2010 and 2020 by artisanal purse seine or gillnet fisheries and referring to tropical tunas or billfish species only.

Categories' configurator panel

This panel allows users to configure the categorization of the resulting outputs by a dataset-specific criteria chosen through the ***Categorize by*** dropdown menu (e.g., fishery group, species category, fleet, etc.).

The resulting data will be categorized by the chosen field, up to a certain number of categories set by the value of the ***Maximum categories*** slider. When the number of distinct categories available in the filtered data exceeds the number of maximum categories, then the system aggregates all records for the less common categories under a generic "*All other*" label.

Furthermore, through the ***Use default category colors*** switch the user can choose whether applying the standard colorization of each entry within the chosen category when displaying output charts and maps.

Output panel

This panel presents the results of the filtering criteria applied to the underlying dataset through one of many possible output charts (dataset dependent).

In general, for each dataset considered, the default output charts are the following:

- Stacked bar chart of absolute values by year and category
- Stacked bar chart of relative values by year and category
- Time series of absolute values by year and category
- Tree map of values by category
- Pareto-chart of values by reporting fleet and category
- Raw dataset (non-categorized)

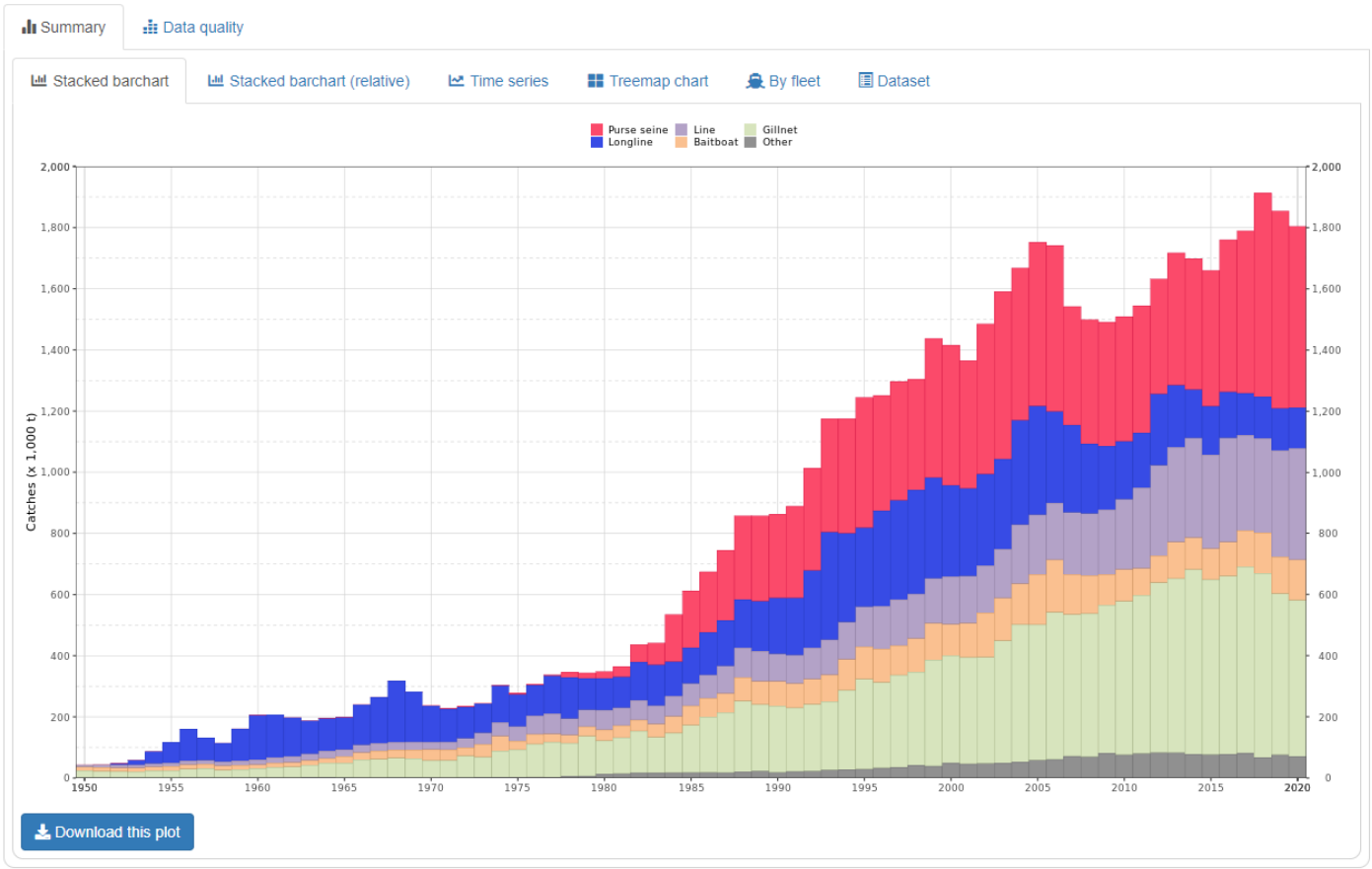


Fig. 4: NC-SCI dataset; stacked bar chart (absolute values) with no filtering criteria applied and categorized by fishery group

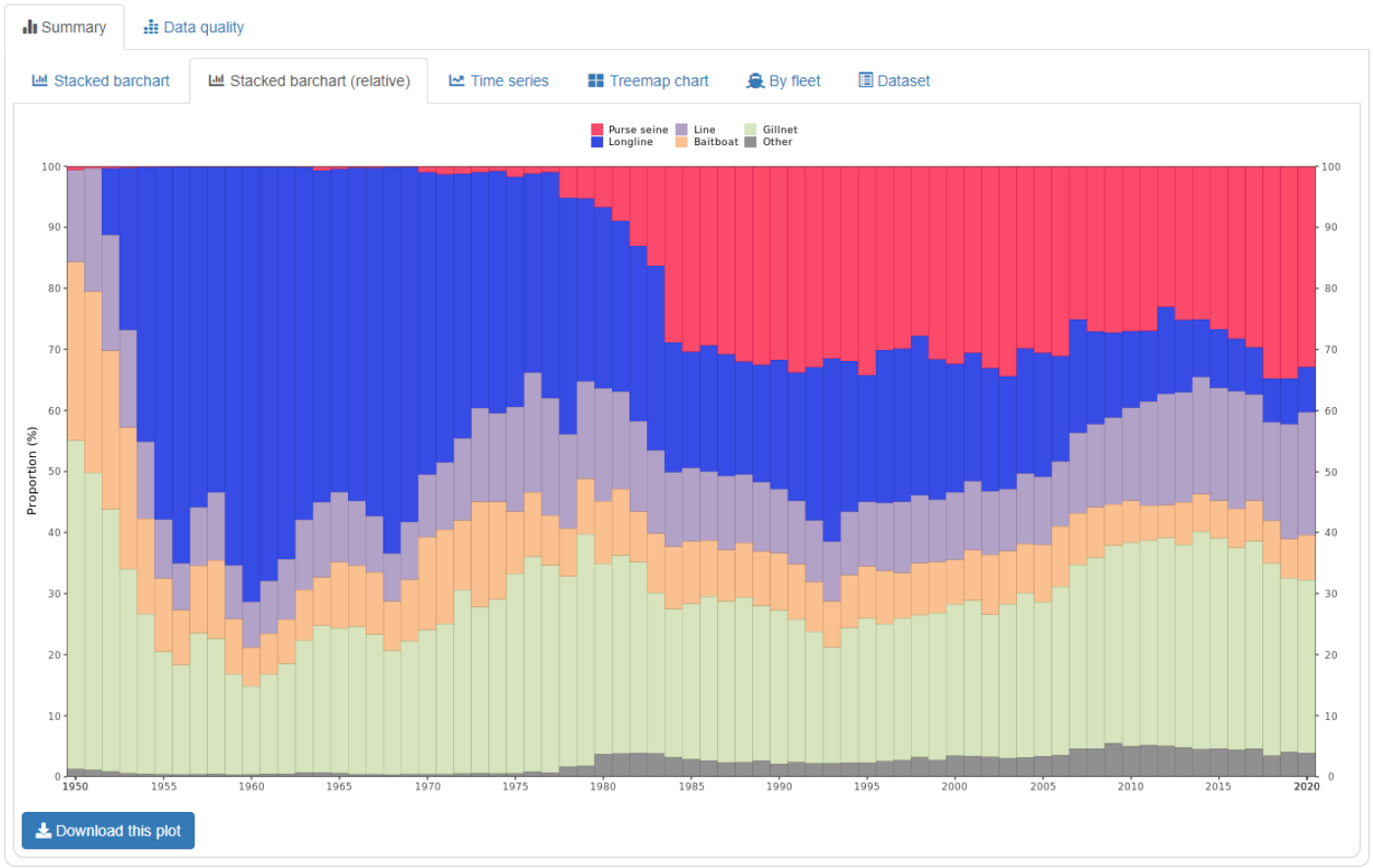


Fig. 5: NC-SCI dataset; stacked bar chart (relative values) with no filtering criteria applied and categorized by fishery group

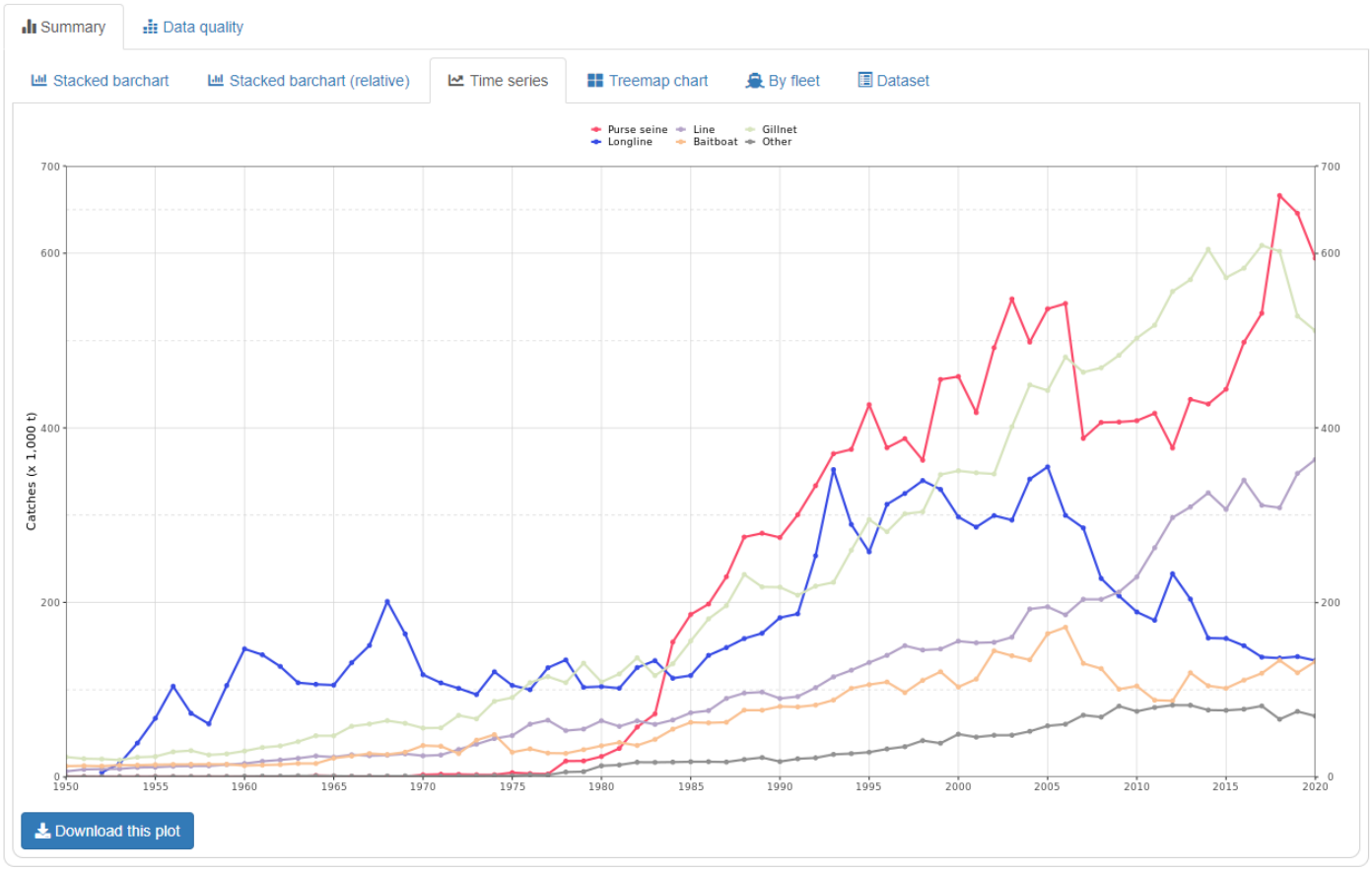


Fig. 6: NC-SCI dataset; time series with no filtering criteria applied and categorized by fishery group

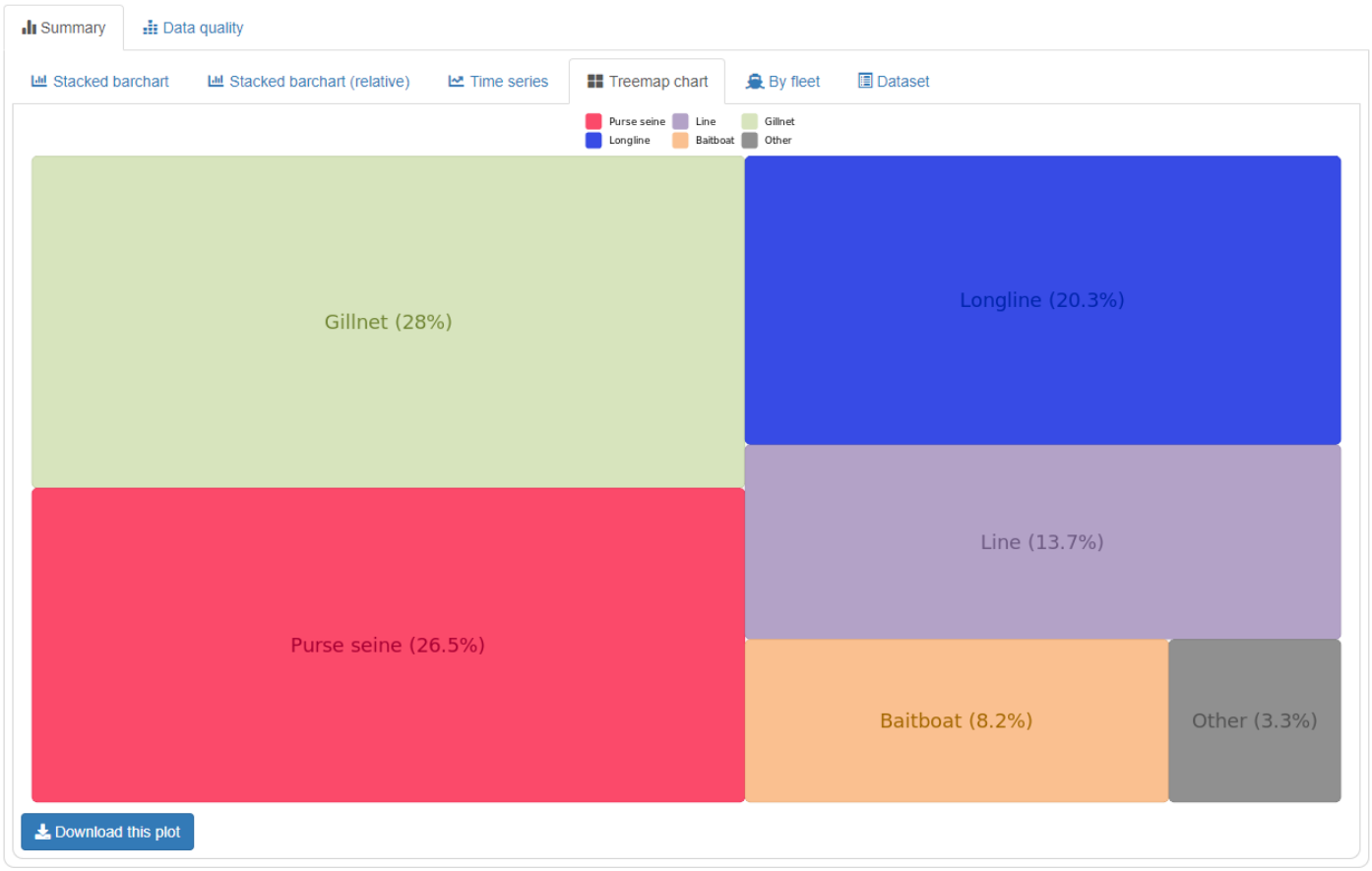


Fig. 7: NC-SCI dataset; tree map with no filtering criteria applied and categorized by fishery group

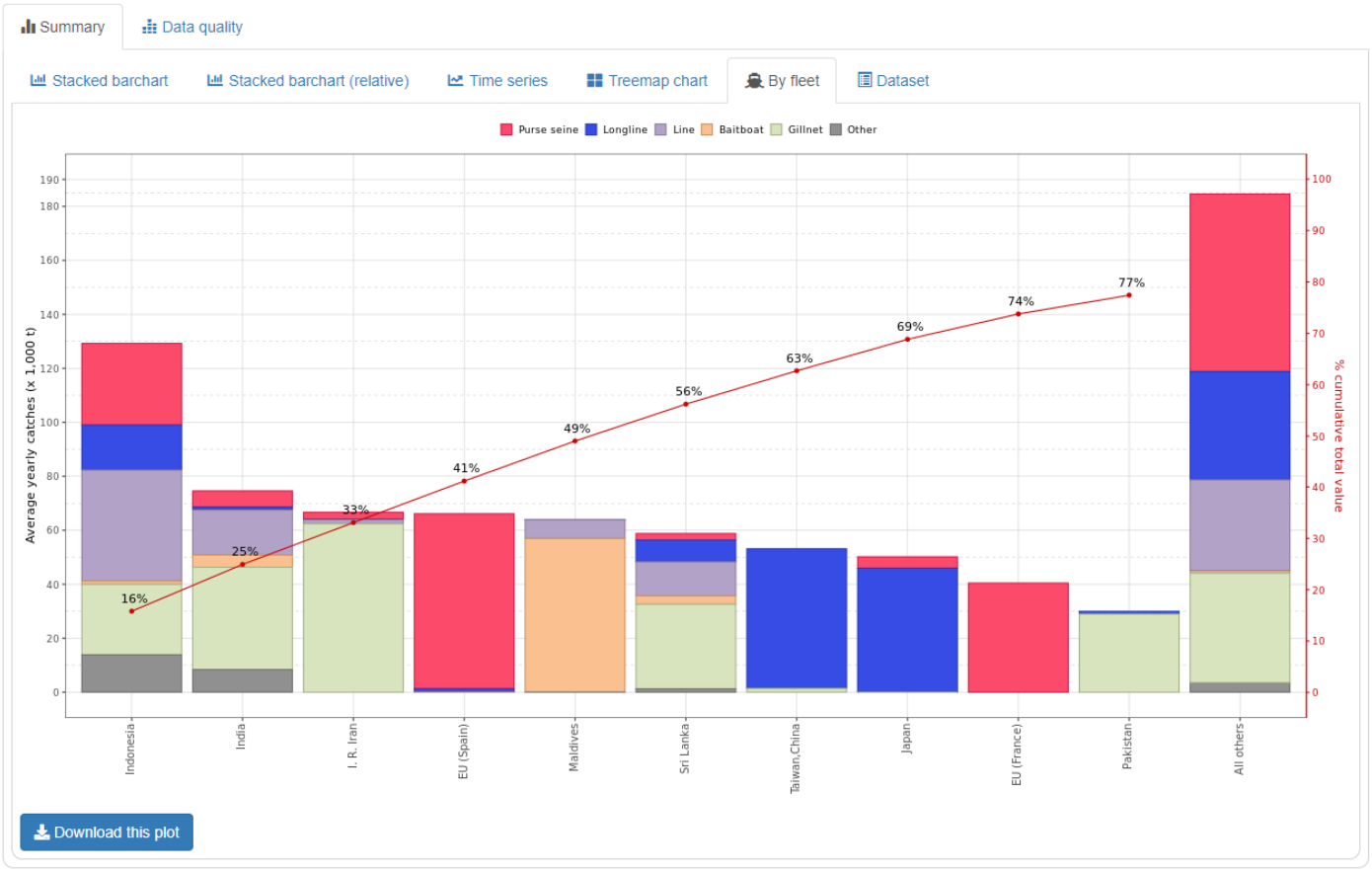
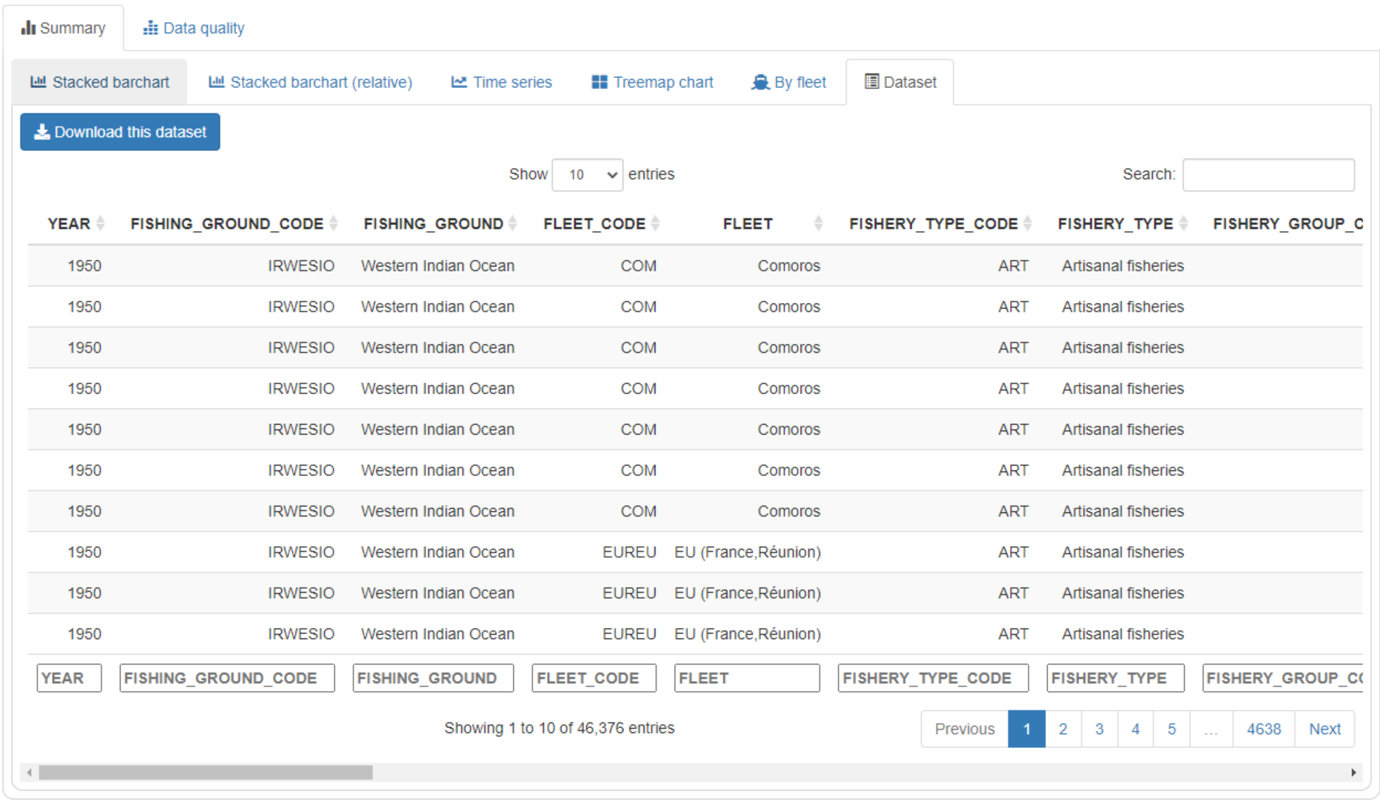


Fig. 8: NC-SCI dataset; pareto chart by fleet with no filtering criteria applied and categorized by fishery group



YEAR	FISHING_GROUND_CODE	FISHING_GROUND	FLEET_CODE	FLEET	FISHERY_TYPE_CODE	FISHERY_TYPE	FISHERY_GROUP_C
1950	IRWESIO	Western Indian Ocean	COM	Comoros	ART	Artisanal fisheries	
1950	IRWESIO	Western Indian Ocean	COM	Comoros	ART	Artisanal fisheries	
1950	IRWESIO	Western Indian Ocean	COM	Comoros	ART	Artisanal fisheries	
1950	IRWESIO	Western Indian Ocean	COM	Comoros	ART	Artisanal fisheries	
1950	IRWESIO	Western Indian Ocean	COM	Comoros	ART	Artisanal fisheries	
1950	IRWESIO	Western Indian Ocean	COM	Comoros	ART	Artisanal fisheries	
1950	IRWESIO	Western Indian Ocean	COM	Comoros	ART	Artisanal fisheries	
1950	IRWESIO	Western Indian Ocean	EUREU	EU (France,Réunion)	ART	Artisanal fisheries	
1950	IRWESIO	Western Indian Ocean	EUREU	EU (France,Réunion)	ART	Artisanal fisheries	
1950	IRWESIO	Western Indian Ocean	EUREU	EU (France,Réunion)	ART	Artisanal fisheries	
YEAR	FISHING_GROUND_CODE	FISHING_GROUND	FLEET_CODE	FLEET	FISHERY_TYPE_CODE	FISHERY_TYPE	FISHERY_GROUP_C

Fig. 9: NC-SCI dataset; raw data with no filtering criteria applied

Geo-referenced datasets that also allow the production of:

- Pie maps of average annual values by category
- Heatmaps of average annual values

by a specified CWP grid resolution (from 5x5 to 30x30 degrees grids) and for a specific IO area only.

Size-frequency datasets that also allow the production of samples by size bin as:

- Stacked bar chart of absolute number of samples by size bin and category
- Stacked bar chart of relative number of samples by size bin and category
- Time series of absolute number of samples by size bin and category
- Time series of relative number of samples by size bin and category

all with a selectable size bin of 1, 2, 5 or 10 cm, and of:

- size distributions by year, size bin and category

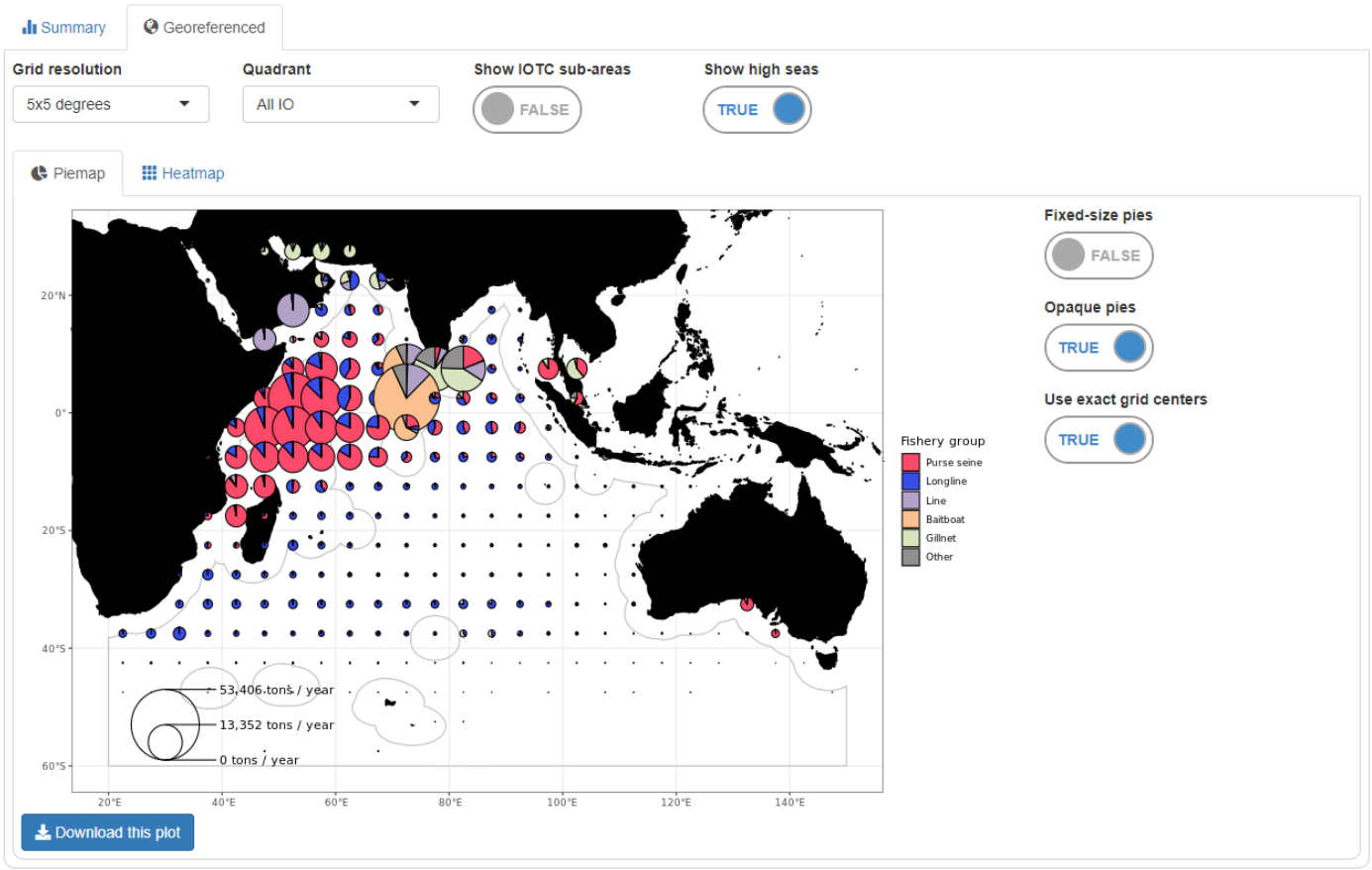


Fig. 10: CE-CA dataset; georeferenced piemap (average annual catches in weight) with no filtering criteria applied and categorized by fishery group

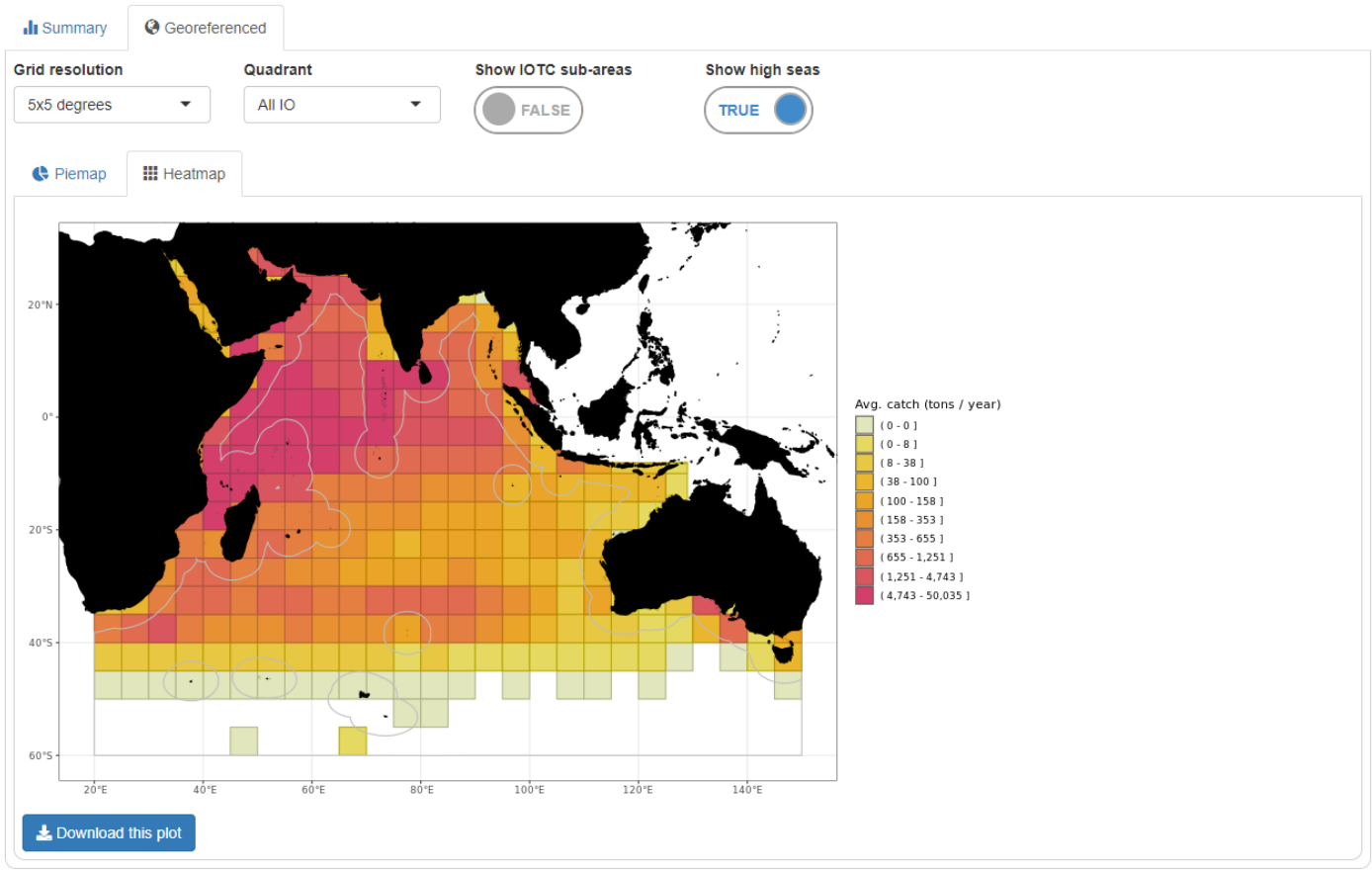


Fig. 11: CE-CA dataset; georeferenced heatmap (average annual catches in weight) with no filtering criteria applied

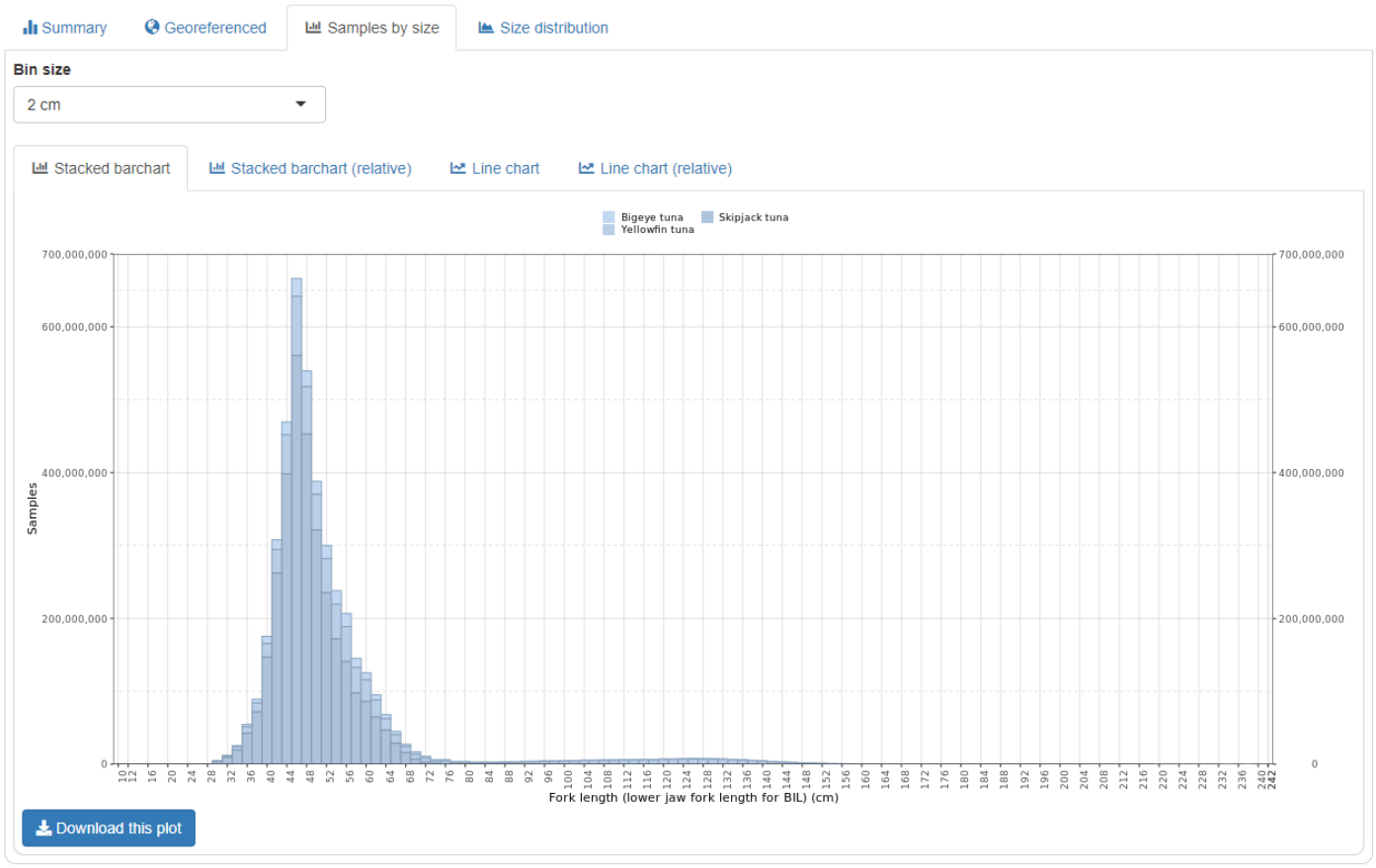


Fig. 13: SF-SCI dataset in fork length; stacked bar chart (absolute number of samples by 2 cm size bin) of tropical tunas by species

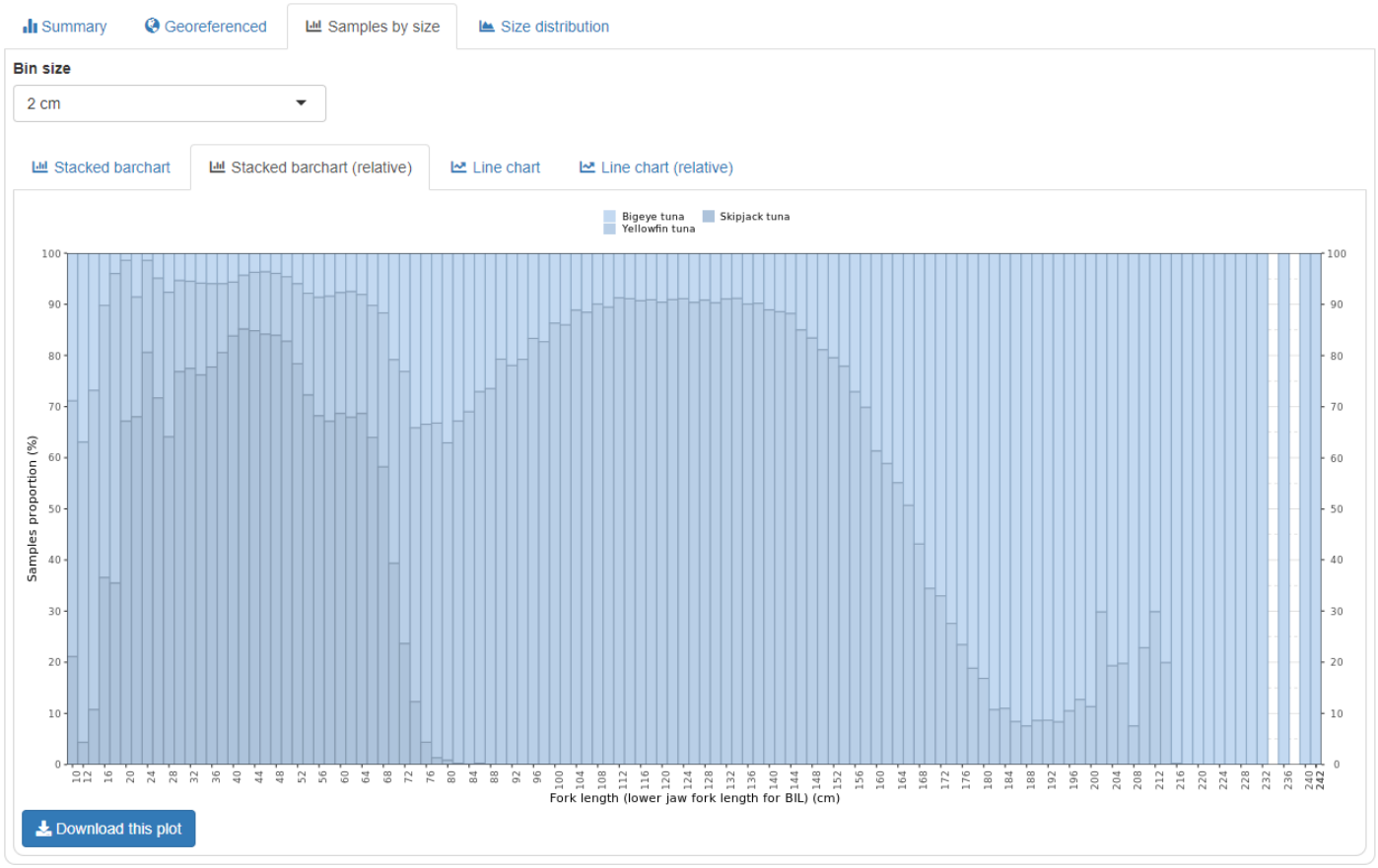


Fig. 14: SF-SCI dataset in fork length; stacked bar chart (relative number of samples by 2 cm size bin) of tropical tunas by species

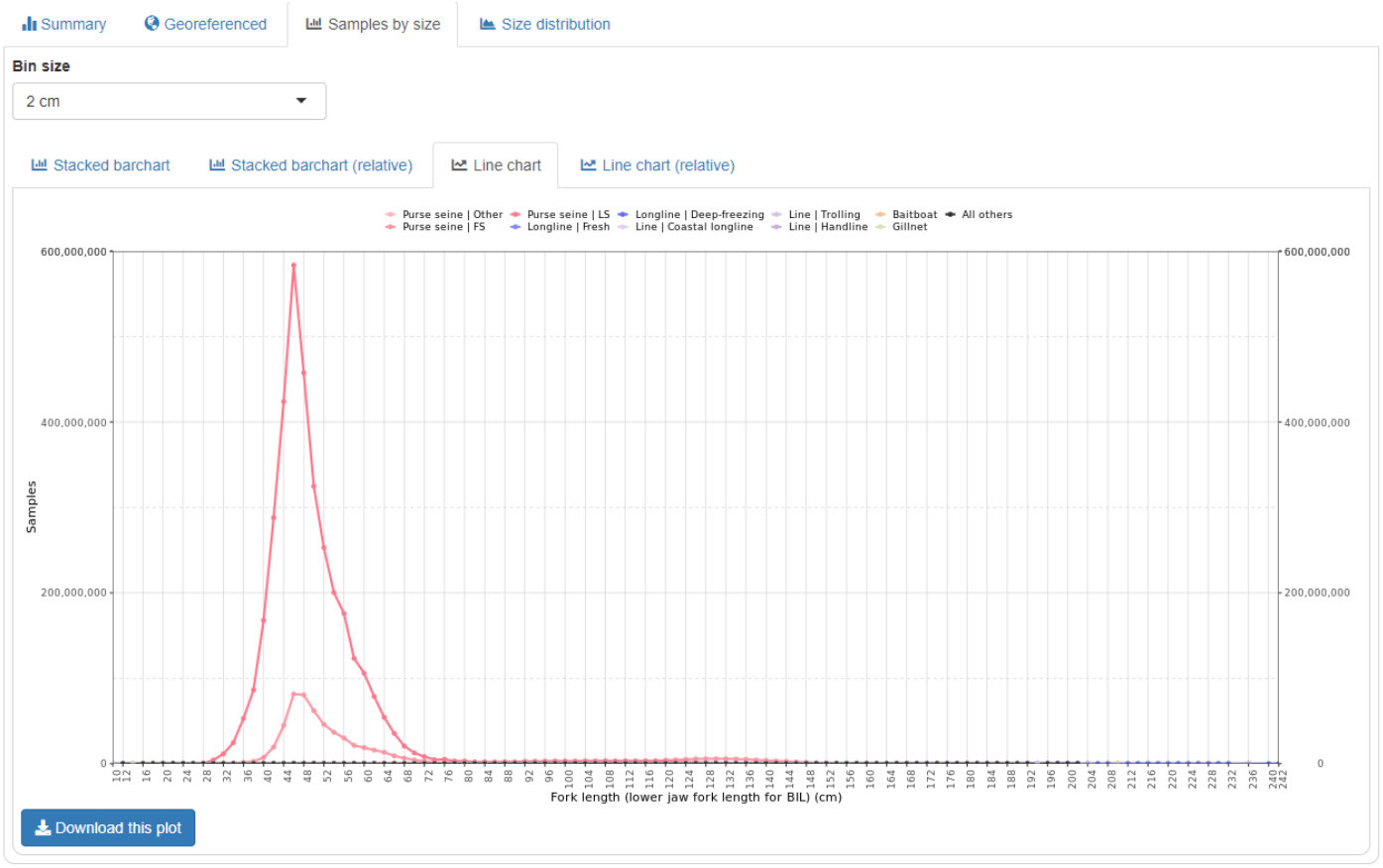


Fig. 15: SF-SCI dataset in fork length; line chart (absolute number of samples by 2 cm size bin) of tropical tunas by fishery

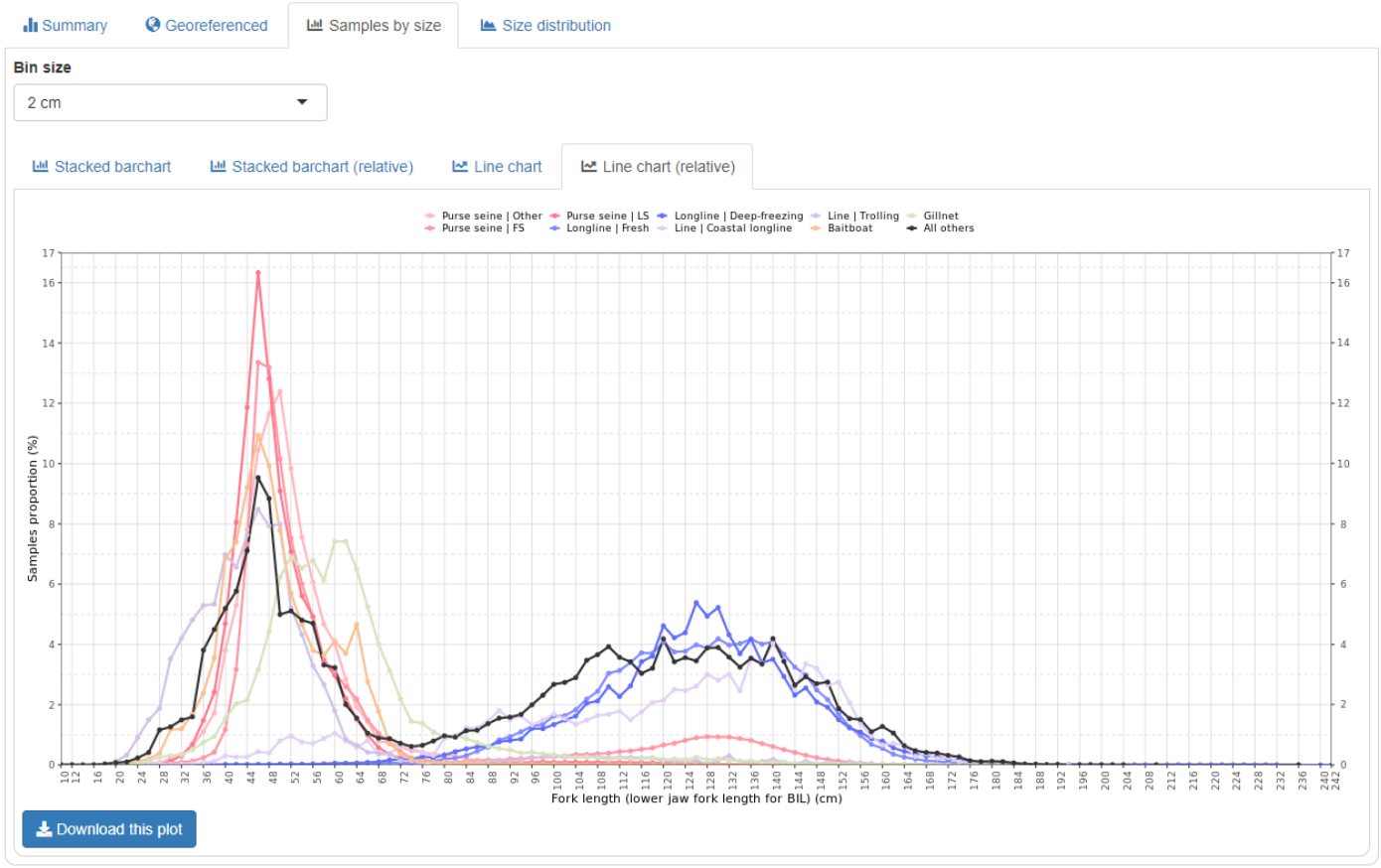


Fig. 16: SF-SCI dataset in fork length; line chart (relative number of samples by 2 cm size bin) of tropical tunas by fishery

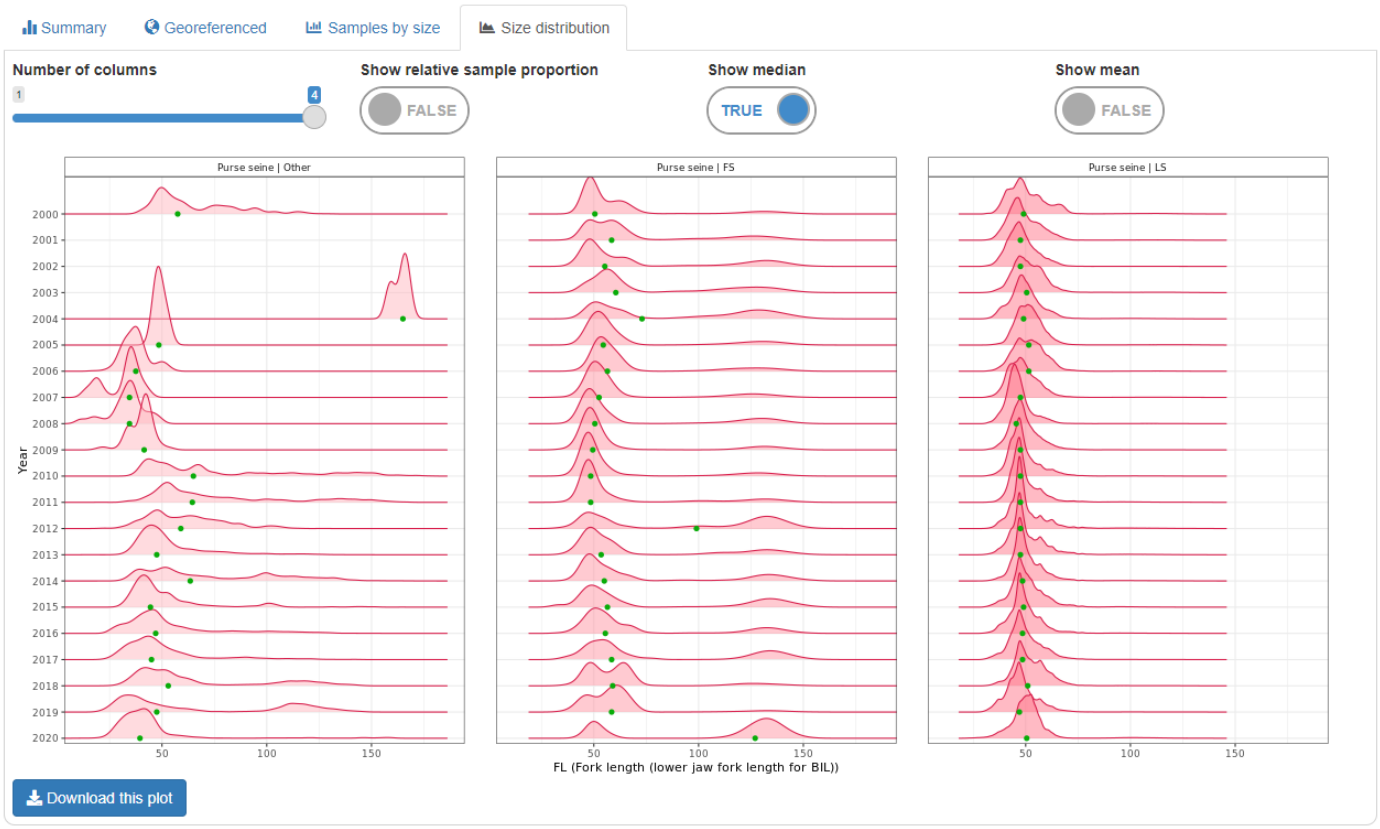


Fig. 17: SF-SCI dataset in fork length (2000-2020); size distribution (relative number of samples by size bin) of tropical tunas by purse seine fishery

All charts and maps can be downloaded as watermarked images (in PNG format) while tabular data can be downloaded as CSV files.

Conclusions

The proposed IOTC data browser can currently replace all functionalities of the IOTC OQS, except for the possibility of providing access to the Fishing Craft statistics which are not yet included in the list of its available datasets.

In terms of user interface, responsiveness, and overall querying and charting capabilities, the IOTC data browser is a definite step forward compared to the OQS, and it also brings in the advantage of leveraging the same R-libraries used by the Secretariat to produce papers and reports, thus fostering uniformity and coherence throughout all IOTC data dissemination products.

Currently, the IOTC data browser is hosted by the IOTC Secretariat on a local shared server: for this reason, the IOTC data browser might not be able to sustain heavy concurrent loads of several users accessing it at the same time.

Furthermore, the IOTC data browser is built on top of the current stable version of the IOTC R-libraries which can sometimes incur in performance issues when producing georeferenced maps or size-distribution of large size-frequency datasets.

Participants to the WPDCS17 are therefore invited to test the proposed IOTC data browser and encouraged to report back to the Secretariat on slow response times, detected bugs, malfunctioning or inaccuracies in the outputs which will be of particular interest in view of a future public release of the tool.

The Secretariat is also interested in receiving suggestions on potential ways to improve the IOTC data browser and extend its functionalities and the list of datasets of interest beyond those currently available in this pre-production release.

Appendix 1 – Species and gear categorizations

Species categorization

Species working party	BYCT						UNCL		TROP	TEMP	NERI		BILL
Species group	SHARKS		OTHERS				OTHERS	TUNAS	TUNAS	TUNAS	TUNAS	SEERFISH	BILLFISH
Species category	SHARKS	RAYS	CETACEANS	SEABIRDS	TURTLES	OTHERS	OTHERS	TUNAS NEI	TROPICAL	TEMPERATE	NERITIC	SEERFISH	BILLFISH
Species	*SKH*	*SRX*	…	…	…	…	***TUX***	**TUN**	**BET**	**ALB**	**BLT**	**COM**	**BLM**
	BSH	RMJ	…	…	…	…	…	**TUS**	**SKJ**	**SBF**	**FRI**	**GUT**	**BUM**
	FAL	RMM	…	…	…	…	…	***AG10***	**YFT**		**KAW**	**KGX**	**MLS**
	SPN	RMB	…	…	…	…	…	***AG39***	***AG34***		**LOT**		**SFA**
	THR	*EAG*	…	…	…	…	…	***AG40***	***AG35***		***FRZ***		**SWO**
	UNCL	*RMV*	…	…	…	…	…	***AG45***			***AG06***		***BIL***
	MAK	*RAJ*	…	…	…	…	…				***AG07***		***BXQ***
	SMA	PLS	…	…	…	…	…						***AG01***
	OCS		…	…	…	…	…						***AG02***
	ALV		…	…	…	…	…						***AG02***
	PTH		…	…	…	…	…						***AG14***
	SPZ		…	…	…	…	…						
	POR		…	…	…	…	…						
	SPL		…	…	…	…	…						
	LMA		…	…	…	…	…						
	BTH		…	…	…	…	…						
	SPK		…	…	…	…	…						
	RHN		…	…	…	…	…						

Table A1: classification of species by working party, group, and category. BYCT = Working Party on Ecosystems and Bycatch, TROP = Working Party on Tropical Tunas, TEMP = Working Party on Temperate Tuna, NERI = Working Party on Neritic Tunas, BILL = Working Party on Billfish; species codes in **boldface** = IOTC species or species aggregates; species codes in *italic* = species aggregates; *AGxy* = species aggregates not presently included in the FAO ASFIS list[6].

[6] https://www.fao.org/fishery/collection/asfis/en

Gear categorization

Fishery group	PS			LL			LI			BB	GN	OT
Fishery	PSFS	PSLS	PSOT	LLO	LLF	LLD	LIC	LIH	LIT	BB	GN	OT
School type	FS	LS	UNCL	UNCL			UNCL			UNCL	UNCL	UNCL
Gear	PS	PS	PS	ELL	FLL	LL	LLCO	DL	*HABBM*	BB	*G/L*	BS
			PSOB	ELLOB		LLEX		HAND	*HABBNM*	BBM	*GIHA*	CN
			PSP	LG		LLOB		HANP	*HABBTR*	BBN	*GIHT*	DDSEI
			PSS	SLL				HLOF	*HATR*	BBOF	GILL	DGN
			RIN	*LLHA*				LLF	*HOOK*	*BBPS*	GIOF	DSEI
			RINP	*LLTR*					SPOR	*RRBB*	GL	FN
			RNOF						TROL			*GHLI*
			SUPP						TROLM			HARP
									TROLN			LIFT
												LIGB
												OTHER
												RR
												TRAP
												TRAW
												TROT
												TWLHT
												UNCL

Table A2: classification of gears by fishery group, fishery, and school type. **PS** = Purse seine fisheries, **LL** = Longline fisheries, **LI** = Line fisheries, **BB** = Pole-and-line fisheries, **GN** = Gillnet fisheries, **OT** = all other fisheries; **PSFS** = Purse seine | Free school, **PSLS** = Purse seine | Log school, **PSOT** = Purse seine | Other, **LLO** = Longline | Other, **LLF** = Longline | Fresh, **LLD** = Longline | Deep-freezing, **LIC** = Line | Coastal longline, **LIH** = Line | Handline, **LIT** = Line | Trolling, **FS** = Free-swimming School, **LS** = Log-associated School, **UNCL** = Unclassified school type; gear codes in *italic* = gear aggregates.