# Stock assessment of yellowfin tuna in the Indian Ocean for 2024

Agurtzane Urtizberea[1], Giancarlo M. Correa[1], Adam Langley[2], Gorka Merino[1], Dan Fu[3], Emmanuel Chassot[3], Shiham Adam[4]

[1]AZTI, Marine Research, Basque Research and Technology Alliance (BRTA), Herrera Kaia, Portualdea z/g, Pasaia, Gipuzkoa, Spain
[2]Independent Consultant, Nelson, New Zealand
[3]Indian Ocean Tuna Commission Secretariat, Blend Building, Providence, Seychelles
[4]International Pole-and-Line Foundation, Maldives

## Contents

**Executive Summary** 2

**1 Introduction** 3

**2 Background** 4
2.1 Biology . . . 4
2.2 Stock structure . . . 4
2.3 Fisheries . . . 5

**3 Model structure** 6
3.1 Spatial stratification . . . 6
3.2 Temporal stratification . . . 7

**4 Model inputs** 7
4.1 Definition of fisheries . . . 7
4.2 Catch . . . 8
4.2.1 Processing . . . 8
4.2.2 Reassignment . . . 9
4.2.3 Aggregation . . . 9
4.3 Size data . . . 9
4.3.1 Processing . . . 11
4.3.2 Reassignment . . . 11
4.3.3 Filtering . . . 12
4.3.4 Aggregation . . . 13
4.4 Indices . . . 13
4.4.1 Longline (LL) CPUE . . . 13
4.4.2 Purse seine CPUE indices . . . 15
4.4.3 Effort creep . . . 16

4.5 Conditional age-at-length data . . . 17
4.6 Tagging . . . 18
4.6.1 Age assignment of tag release . . . 18
4.6.2 Initial tagging mortality . . . 19
4.6.3 Chronic tag loss . . . 19
4.6.4 Reporting rate . . . 19
4.6.5 Small-scale tagging programmes . . . 20
4.7 Environmental data . . . 20

**5 Model parameters** **21**
5.1 Population dynamics . . . 21
5.1.1 Recruitment . . . 21
5.1.2 Initial population . . . 22
5.1.3 Somatic growth . . . 22
5.1.4 Sexual maturity and fecundity . . . 23
5.1.5 Natural mortality . . . 23
5.1.6 Movement . . . 24
5.2 Fishery dynamics . . . 25
5.2.1 Fishing mortality . . . 25
5.2.2 Catchability . . . 25
5.2.3 Selectivity . . . 25
5.3 Dynamics of tagged fish . . . 26
5.4 Likelihood components . . . 27
5.4.1 Catch . . . 27
5.4.2 Indices of abundance . . . 27
5.4.3 Length frequency . . . 27
5.4.4 Conditional age-at-length . . . 27
5.4.5 Tagging data . . . 28
5.5 Parameter estimation and uncertainty . . . 28
5.5.1 Diagnostics . . . 28
5.6 Stock status . . . 29

**6 Model runs** **29**
6.1 Stepwise revisions . . . 29
6.1.1 Reference models . . . 30

**7 Model results** **31**
7.1 Fits . . . 31
7.2 Estimates . . . 32
7.3 Diagnostics . . . 33
7.3.1 Jitter analysis . . . 34
7.3.2 Retrospective analysis . . . 34
7.3.3 Run test . . . 34
7.3.4 Hindcast analysis . . . 34
7.3.5 Likelihood profiles . . . 34
7.4 Exploratory analyses . . . 35
7.5 Stock status . . . 36

**8 Projections** **37**

**9 Discussion** **38**

**10 Reproducibility and transparency** **40**

**11 Acknowledgements** **40**

**12 Tables** **41**

**13 Figures** **48**

**References** **117**

**14 Appendix** **125**
14.1 Clustering of size compositions . . . 125
14.1.1 Figures . . . 126
14.2 One-area model . . . 132
14.2.1 Results . . . 132
14.2.2 Tables . . . 132
14.2.3 Figures . . . 135
14.3 Two-area model . . . 139
14.3.1 Results . . . 139
14.3.2 Tables . . . 140
14.3.3 Figures . . . 143

## Executive Summary

This report presents the 2024 stock assessment for Indian Ocean yellowfin tuna (*Thunnus albacares*) using Stock Synthesis 3. The assessment uses an age-structured spatially-explicit population model and is fitted to catch, catch per unit effort (CPUE) indices, length compositions, tagging data, and conditional age-at-length. The assessment covers 1950 – 2023 and represents an update of the previous assessment model carried out in 2021, taking into account progress and improvements made since the previous assessment, including recommendations from the review of the previous assessment undertaken between 2018 and 2023. The assessment assumes that the Indian Ocean yellowfin tuna constitute a single spawning stock, modelled as spatially disaggregated four regions, with twenty-one fisheries. Key biological parameters were revised, specifically growth and natural mortality. Standardized CPUE series from the main longline fleets 1975 – 2023 were included in the models as the relative abundance index of exploitable abundance in each region, including alternative assumptions regarding changes in the efficiency of the longline fleet ("effort creep"). The CPUE indices from EU purse seine sets on free schools were included in a subset of models. Tag release and recovery data from the RTTP-IO program were included in the model to inform abundance, movement, and mortality rates. The inclusion of conditional age-at-length data from the GERUNDIO project was also explored in a subset of models.

Overall, the estimates of this stock assessment are principally driven by the increasing trend estimated for the longline CPUE indices and by the positive trends in recruitment deviates estimated by the model. The model estimates a large decline in biomass between the 1980s and early 1990s following the CPUE decrease. In that period, the catch increased from 54,000 mt to 404,000 mt in 1993, which was followed by a significant decrease in abundance. However, in the period 2003-2023, the catch has remained above 400,000 mt (average of 406,000 mt) and CPUE in the northwestern region has remained stable or increasing in recent years. The model seems to use the 1980-1993 data to scale long-term equilibrium productivity of the stock (stock-recruitment relationship, maximum sustainable yield -MSY-, and other benchmarks) and fit the more recent data with recruitment deviates. Without positive recruitment deviates in the recent period ($\sim$ 20 years), the observed catch would not be plausible. Therefore, there is a need to update reference points (MSY, unfished recruitment, and unfished spawning biomass) to recent conditions.

The stock status estimates of this assessment are qualitatively different from the 2021 assessment. The reasons for this being the significant increase in CPUE and a narrower range of biological parameters (growth and natural mortality). Spawning biomass in 2023 was estimated to be 32% higher than the level that supports the maximum sustainable yield ($SSB_{2023}/SSB_{MSY} = 1.32$). Current fishing mortality is estimated to be 25% lower than $F_{MSY}$ ($F_{2023}/F_{MSY} = 0.75$). The probability of the stock being in the green Kobe quadrant in 2023 is estimated to be 89%. Therefore, the yellowfin tuna stock is determined to be not overfished and not subject to overfishing.

The new model represents a marked improvement over the previous model, as demonstrated using a number of statistical diagnostic analyses. However, there are still important uncertainties in the data used for yellowfin and other tropical tuna stock assessments. There are uncertainties in relation to the CPUE standardization but also on reported catches from some Contracting Parties and Cooperating Non-Contracting Parties (CPCs), size frequency data, and model configuration such as the spatial structure. These will need to be further explored in the future.

# 1 Introduction

Before 2008, Indian Ocean (IO) yellowfin tuna (*Thunnus albacares*) was assessed using methods such as Virtual Population Analysis (VPA) and production models (Nishida and Shono, 2007, 2005). In 2008, a preliminary stock assessment of IO yellowfin tuna was conducted using MULTIFAN-CL (Langley et al., 2008), enabling the integration of the tag release/recovery data collected from the large-scale tagging programme conducted in the IO in the preceding years. The MULTIFAN-CL assessment was revised and updated in the following years (Langley et al., 2012, 2011, 2010, 2009).

The 2015 assessment (Langley, 2015) was implemented using the Stock Synthesis 3 (SS3) modelling platform (Methot and Wetzel, 2013). SS3 is conceptually similar to MULTIFAN-CL, and the two platforms have yielded similar results. On the basis of that assessment, the yellowfin tuna stock was determined to be overfished and subject to overfishing. At its 20th meeting, the Indian Ocean Tuna Commission (IOTC) adopted an Interim Plan for Rebuilding the Indian Ocean Yellowfin Tuna Stock (Res. 16/01). The interim rebuilding plan was revised in 2017 (Res 17/01), 2018 (Res 18/01), 2019 (Res 19/01) and most recently in 2021 (Res 21/01).

The SS3 assessment was updated in 2016 (Langley, 2016) and was revised and updated in 2018 (Fu et al., 2018). These assessments utilised new composite longline catch per unit effort (CPUE) indices derived from the main distant water longline fleets, replacing the Japanese longline CPUE indices used previously. The 2018 assessment also included a comprehensive analysis of the main assumptions of the stock assessment. A model ensemble covering major components of structural uncertainty was used to characterise the stock status. The assessment estimated that the spawning stock biomass in 2017 was below $SSB_{MSY}$, and that fishing mortality was above $F_{MSY}$. Therefore, the stock status was determined to remain overfished and undergoing overfishing.

An external review of the 2018 assessment provided recommendations to improve model parametrisations (Methot, 2019). An attempt was made to update the assessment in 2019 with extensive investigations of alternative spatial structures, data weighting, and biological parameters (Urtizberea et al., 2019). Further analyses were conducted in 2020 to refine the process of model selection through an objective scoring system based on diagnostic metrics (Urtizberea et al., 2020).

The most recent assessment was conducted in 2021 (Fu et al., 2021), which also used SS3 as the modelling platform and was based on the four-area spatial configuration as in 2018. This recent assessment included a standardised CPUE series from the main longline fleets as the main index but also tested the inclusion of EU purse seine indices, operating on free schools and floating objects, and an index from the Maldivian pole and line fishery. A range of exploratory models were presented to address issues in observational datasets, improve the stability of the assessment model, and explore the effects of alternative model assumptions. Overall, stock status estimates did not differ substantially from the 2018 assessment, estimating $SSB/SSB_{MSY} = 0.78$ and $F/F_{MSY} = 1.27$ for the terminal model year (2020), which suggests that the stock is overfished and experiencing overfishing. An external review of the 2021 assessment was carried out in 2023 (M. Maunder et al., 2023), which provided a set of recommendations for the next assessment implementation. Some of these recommendations were investigated by Langley et al. (2023).

In this report, we document the results of the stock assessment of IO yellowfin tuna accepted by the 26th WPTT meeting. The assessment included fishery and biological data up to the end of 2023 and its configuration is based on the 2021 assessment, although two additional spatial

configurations were also tested. It implements an age- and spatially-structured population model using SS3 (v3.30.22.1) and incorporates some recommendations made by the last external review (M. Maunder et al., 2023). The WPTT adopted a set of model options and sensitivities for key model parameters to derive estimates of stock status in 2023 and associated uncertainty.

# 2 Background

## 2.1 Biology

Yellowfin tuna is a cosmopolitan species distributed mainly in the tropical and subtropical oceanic waters of the three major oceans, forming large schools. Spawning occurs mainly from December to March in lower latitudes with warmer waters and mesoscale oceanographic activity (Muhling et al., 2017), with the main spawning grounds west of 75°E. However, spawning activity has also been reported in the Oman Sea (Hosseini and Kaymaram, 2016), Bay of Bengal (Kumar and Ghosh, 2022), off Sri Lanka and the Mozambique Channel, and in the eastern IO off Australia Nootmorn et al. (2005). The size at 50% maturity for this species in the IO was initially estimated at around 75 cm based on cortical alveolar stage (Zudaire et al., 2013). Still, an updated study suggests that it might be at a larger size ($\sim$ 101 cm) (Zudaire et al., 2022). Tag recoveries provide evidence of large movements of yellowfin tuna within the western equatorial region; however, few observations of large-scale transverse movements in the IO have been reported (Gaertner and Hallier, 2015). Yellowfin dwell preferentially in the surface mixed layer and the thermocline (Pecoraro et al., 2017), above 200 m approximately (Sabarros et al., 2015).

This species has a high metabolic rate and, therefore, requires large energy supplies to fulfil the bioenergetics demands for movement, growth, and reproduction (Artetxe-Arrate et al., 2021). Feeding behaviour is largely opportunistic, with a variety of prey species being consumed, including large concentrations of crustaceans that have occurred recently in tropical areas and small mesopelagic fishes (Duffy et al., 2017; Krishnan et al., 2024; Roger, 1994). Recent growth studies have generally supported a two-stanza growth curve, with a slow initial growth phase up to $\sim$ 60 cm followed by much faster growth (Farley et al., 2023). In addition, differences in mean length-at-age have been identified between males and females for fish older than four years. Environmental variability in the IO impacts the abundance and catch rates of this species. A significant negative association between the Indian Ocean Dipoles (IODs) and the catch rates of yellowfin tuna with a periodicity of approximately four years was observed (Lan et al., 2020, 2013). Likewise, Lan et al. (2020) also found that the El Niño Southern Oscillation (ENSO) impacted on catch rates near the Arabian Sea.

## 2.2 Stock structure

Fisheries information indicates that adult yellowfin are distributed continuously throughout the entire tropical IO, but some more detailed analysis of fisheries data suggests that the stock structure may be more complex. The tag recoveries may indicate that the western and eastern regions of the IO support relatively discrete sub-populations of yellowfin tuna. Studies of stock structure using DNA techniques have suggested that there may be genetically discrete subpopulations in the northwestern IO (Dammannagoda et al., 2008) and within Indian waters (Kunal et al., 2013). A recent study of stock structure using gene sequencing technology along with a basin-scale sampling design indicated genetic differentiation between north and south of the

equator within the IO, and possibly additional genetic structure within the locations north of the equator (Grewe et al., 2020). Parasite composition and abundance suggest limited movement between the Indonesian archipelago (eastern IO) and the Maldives (central IO) (Moore et al., 2019). Isotope studies have also suggested relatively limited movement, with resident behaviour at the temporal scale of their muscle turnover ($\sim$ 3 months) (Ménard et al., 2007). Otolith chemistry analyses concluded that fisheries operating in the western IO are mainly composed of fish of western origin, which suggests limited movement from east to west (Artetxe-Arrate et al., in review). These studies generally support the potential presence of population units of yellowfin tuna within the IO, despite the fact that considerable uncertainty remains on the sub-regional population structure in this region. This assessment assumes that the IO yellowfin tuna stock consists of several interconnected regional populations (Figure 1) that have the same biological characteristics; however, we acknowledge that more studies are needed to reveal the structure of this species.

## 2.3 Fisheries

Yellowfin tuna are harvested with a diverse variety of gear types, from small-scale artisanal fisheries (in the Arabian Sea, the Mozambique Channel and waters around Indonesia, Sri Lanka, the Maldives, and Lakshadweep Islands) to large gillnetters (from Oman, Iran and Pakistan, mainly operating but not exclusively in the Arabian Sea) and distant-water longliners and purse seiners that operate widely in equatorial and tropical waters (Figure 2). Purse seiners and gillnetters catch a wide size range of yellowfin tuna, whereas the longline fishery mostly catches adult fish (Figure 3).

Prior to 1980, annual catches of yellowfin tuna remained below 80,000 mt and were dominated by longline catches (Figure 4). Annual catches increased markedly during the 1980s and early 1990s, mainly due to the development of the purse-seine fishery as well as an expansion of the other established fisheries (fresh-tuna longline, gillnet, baitboat, handline and, to a lesser extent, troll). A peak in catches was recorded in 1993, with catches over 400,000 mt. The increase in catch is almost fully attributable to longline fleets, particularly longliners flagged in Taiwan, which reported exceptional catches of yellowfin tuna in the Arabian Sea. The Taiwanese longline fishery in the IO has been equipped with super-cold storage. Since around 1986, the fleet has fished more frequently with deep sets.

Catches declined in 1994 to about 350,000 mt, remaining at that level for the following decade, then increasing sharply to reach a peak of about 520,000 mt in 2004-2005, driven by a large increase in catch by all fisheries, especially the purse-seine (free school) fishery. Total annual catches declined sharply from 2004 to 2007 and remained at about 300,000 mt during 2007–2011. In 2012, total catches increased to about 400,000 mt and were maintained at about that level from 2013 to 2015. Total catches increased to an average of 430,000 mt between 2016 and 2019, and a maximum of close to 450,000 mt in 2019 (Figure 4), despite IOTC Resolution 17/01, which requested major fleets to substantially reduce their yellowfin catches below the 2014 or 2015 catch level. Furthermore, catch levels of about 440,000 mt reported for 2018 might be underestimated, to some extent, because of changes in data processing methodology by European Union-Spain for its purse seine fleet for that year (IOTC, 2021).

In recent years (2015–2023), purse seine has been the dominant fishing method, harvesting $\sim$ 35% of the total IO yellowfin tuna catch (by weight), with the gillnet and handline fisheries, principally in the Arabian Sea, comprising $\sim$ 20% and 18% of the catch, respectively. A smaller component of the catch was taken by industrial longline (5%) and the regionally important

baitboat (4%) and troll (4%) fisheries. The recent increase in the total catch has been mostly attributable to a rise in catch from the gillnet and handline fisheries (Figure 4), mostly from the Omani fleet.

The purse-seine catch is generally distributed equally between free-school and associated (log and FAD sets) schools, although the large catches in 2003–2005 were dominated by fishing on free-schools. Conversely, during 2015–2023, the purse-seine catch was dominated ($\sim$ 70%) by the associated fishery.

Historically, most of the yellowfin catch has been taken from the western equatorial region of the IO ($\sim$ 44%, Figure 2) and, to a lesser extent, the Arabian Sea ($\sim$ 26%), the eastern equatorial region ($\sim$ 24%) and the Mozambique Channel ($\sim$ 5%). The purse-seine and baitboat fisheries operate almost exclusively within the western equatorial region, while catches from the Arabian Sea are principally by handline, gillnet, and longline (Figure 2). Catches from the eastern equatorial region were dominated by longline and gillnet (around Sri Lanka and Indonesia). The southern IO accounts for a small proportion of the total yellowfin catch (1%) taken exclusively by longline.

During 2008–2012, due to the threat of piracy, the bulk of the industrial purse seine and longline fleets moved out of the western equatorial region to avoid the coastal and off-shore waters off Somalia, Kenya and Tanzania. The threat of piracy particularly affected the freezer longline fleet, and levels of effort and catch decreased markedly from 2007. The total catch by freezing longliners declined to about 2,000 mt in 2010, a 10-fold decrease in catch from the years before the onset of piracy. Purse seine catches also dropped in 2007–2009 and then started to recover. Piracy off the Somali coast was almost eliminated by 2013, but longline catches have not recovered.

The sizes caught in the IO range from 30 cm to 180 cm fork length (Figure 3). Intermediate-age yellowfin are seldom taken in industrial fisheries but are abundant in some artisanal fisheries, mainly in the Arabian Sea (Figure 5). Newly recruited fish are primarily caught by the purse seine fishery on floating objects and the pole-and-line fishery in the Maldives. Males are predominant in the catches of larger fish at sizes larger than 150 cm, which is also observed in other oceans. Medium-sized yellowfin concentrate for feeding in the Arabian Sea (Figure 5).

# 3 Model structure

## 3.1 Spatial stratification

The geographic area considered in the assessment is the IO, defined by the coordinates 40°S-25°N and 20°E-150°E. Earlier yellowfin stock assessments have adopted a five-area spatial structure (Langley et al., 2012), but several issues were identified for that structure. Since 2015, a four-area spatial structure is used for this stock (Figure 1). The Arabian Sea (region 1a) and western equatorial area (region 1b) make up the region 1 but kept the fishery information separated (i.e., areas-as-fleet approach) to account for differences in selectivity between these sub-areas (Punt, 2019). The spatial structure retains two regions that encompass the main year-round fisheries in the tropical area (regions 1 and 4) and two austral, subtropical regions where the longline fisheries occur more seasonally (regions 2 and 3).

The current spatial structure separates the purse-seine fishery in the northern Mozambique Channel (10-15°S) from the equatorial region, as the fishery in the northern Mozambique Channel

exhibits strong seasonal variation in effort and operates differently from the equatorial region (Langley, 2015). There is also a separation of the purse-seine fishery between the western and eastern tropical regions with the current boundary between regions 1b and 4. In addition to the four-area configuration, we also evaluated two additional spatial structures: one-area and two-area configurations. The 2021 assessment also evaluated a modified version of the four-area structure (Fu et al., 2021), but it produced similar results. We did not evaluate that modified four-area configuration in the current assessment.

### 3.2 Temporal stratification

The period covered by the assessment is 1950−2023, which represents the period for which catch data are available from the commercial fishing fleets. Langley (2015) suggested that the assessment results were not sensitive to the early catches from the model (pre-1972), and commencing the model in 1950 or 1972 (assuming unexploited equilibrium conditions) yielded very similar results as also found in the current assessment.

The time step in the assessment model was a quarter (i.e., three months duration, four quarters per year), representing 296 model time steps. The definition of these time steps enabled recruitment to be estimated for each quarter to approximate the continuous recruitment of yellowfin in the equatorial regions. However, the quarterly model time step precluded the estimation of seasonal model parameters, particularly the movement parameters. Fu et al. (2018) explored an alternative annual/seasonal model structure which explicitly estimated seasonal movement dynamics. However, the alternative temporal structure did not yield substantially different results.

## 4 Model inputs

Catch (1950-2023) and size (1952-2023) information was provided by the IOTC Secretariat in a comma-separated values (CSV) format. These datasets and the metadata can be found online at the IOTC website: https://iotc.org/documents/WPTT/26AS/Data/01. Four indices of abundance were also available: joint longline, purse seine fishing on free schools, purse seine fishing on associated schools, and the associative behaviour-based abundance index; however, we only used the joint longline index in the main set of models and the purse seine free school index as sensitivity analysis. These indices can also be found online at https://iotc.org/documents/standardised-cpue-yft-and-bet. Release and recovery data (2005-2015) from two tagging programs were also available, as well as age data from the GERUNDIO project (2009-2022). Raw tagging and age datasets are confidential and cannot be shared; however, processed tagging and age-length data for SS3 can be found at https://github.com/Fundacion-AZTI/IOTC_YFT_2024_Assessment/tree/reproducible.

### 4.1 Definition of fisheries

The current assessment adopted the equivalent fisheries definitions used in the previous stock assessments. First, nine *fishery groups* were defined based on fleet, gear, purse seine set type, and type of vessel in the case of the longline fleet (Table 1), representing relatively homogeneous fishing units with similar selectivity and catchability characteristics that do not vary greatly over time. Then, *fishery groups* were divided into regions, producing twenty-one *fisheries* in the

assessment model (Table 2). We also provide some initial analyses that might help to implement alternative fishery definitions in future assessments (see Section 14.1).

A brief description of each *fishery group* is provided below.

The longline fishery was partitioned into two main components:

- *Freezing longline fisheries (LL)*, or all those using drifting longlines for which one or more of the following three conditions apply: (i) the vessel hull is made up of steel; (ii) the vessel length overall of 30 m or greater; (iii) the majority of the catches of target species are preserved frozen or deep-frozen. A composite longline fishery was defined in each model area, aggregating the longline catch from all freezing longline fleets (principally Japan and Taiwan).
- *Fresh-tuna longline fisheries (LF)*, or all those using drifting longlines and made of vessels (i) having fibreglass, fibre-reinforced plastic, or wooden hull; (ii) having length overall less than 30 m; (iii) preserving the catches of target species fresh or in refrigerated seawater. A composite longline fishery was defined as aggregating the longline catch from all fresh-tuna longline fleets (principally Indonesia and Taiwan) in region 4, where the majority of the fresh-tuna longliners have traditionally operated.
- The purse-seine catch and effort data were apportioned into two separate method fisheries: catches from sets on associated schools of tuna (log and drifting FAD sets; *LS*) and sets on unassociated schools (free schools; *FS*).
- A single baitboat fishery (*BB*) was defined within region 1b (essentially the Maldives fishery).
- Gillnet fisheries (*GI*) were defined in the Arabian Sea (region 1a), including catches by Iran, Pakistan, and Oman, and in region 4 (Sri Lanka and Indonesia).
- Three troll fisheries (*TR*) were defined, representing separate fisheries in regions 1b (Maldives), 2 (Comoros and Madagascar) and 4 (Sri Lanka and Indonesia).
- A handline fishery (*HD*) was defined within region 1a, principally representing catches from Yemen, Oman, and Maldives.
- A miscellaneous "Other" fishery (*OT*) was defined as comprising catches from artisanal fisheries other than those specified above (e.g. trawlers, small purse seines or seine nets, sport fishing, and a range of small gears).

## 4.2 Catch

The catch dataset was composed of information about time (year and month), CPCs, gear type, type of association of the fish school, grid code at a $5^\circ \times 5^\circ$ resolution, and catch in weight (metric tons) and numbers. The grid code contained information on the grid resolution, quadrant, and longitude and latitude of the corner of the grid.

#### 4.2.1 Processing

We followed the next steps to produce the catch input for SS3:

- Month information was used to assign quarters (i.e., four quarters from January to December).
- The fishery group was assigned based on CPC, gear type, and the type of association of the fish school.
- Catch was summed by year, quarter, CPC, gear type, fishery group, and grid code.
- The grid code was used to calculate the longitude and latitude of the centre of the grid (called *centroid* hereafter).
- The centroid was used to assign regions used in the assessment model (Figure 6).
- Fishery was assigned based on fishery group and region.

#### 4.2.2 Reassignment

To simplify the fishery structure in the stock assessment model, we reassigned catches in regions with low fishing activity to main regions, as done in the 2021 assessment.

- *LF* fisheries: catch in regions 1 to 3, representing only $\sim 3\%$ of the total catches over the time series, was assigned to region 4.
- *FS* and *LS* fisheries: purse seine catches in region 1a and 3 were reassigned to region 1b and 4, respectively.
- *BB* fisheries: a small proportion of the total baitboat catch and effort occurs on the periphery of region 1b, within regions 1a and 4. Therefore, we assigned all *BB* catches to region 1b.
- *GI* fisheries: a very small proportion of the total gillnet catch and effort occurs in regions 1b and 2, which was reassigned to area 1a. Likewise, catch in region 3 was reassigned to region 4.
- *TR* fisheries: moderate troll catches are taken in regions 1a and 3, which were reassigned to regions 1b and 4, respectively.
- *HD* fisheries: moderate handline catches are taken in regions 1b, 2 and 4, which were reassigned to region 1a.
- *OT* fisheries: catch from region 1b and 2 was reassigned to region 1a, while catch from region 3 was reassigned to region 4.

#### 4.2.3 Aggregation

After catch reassignment, catch data (in metric tons) was summed by year, quarter, and fishery, and then organized in an SS3 format. Overall, the time series of catch per fleet were quite similar to the catch series included in the 2021 assessment (Figure 7). The largest differences were observed for *OT* and *TR* in region 4, especially during the last two decades. Also, current catch estimates for *LL* in region 3 are slightly larger than the previous assessment during 2008-2020. The changes are mostly attributed to revisions of catch estimation by the IOTC Secretariat.

## 4.3 Size data

The size data was composed of information about time (year and month), CPC, gear type, type of association of the fish school, grid code, number of fish sampled per fork length bin (cm), and the score of reporting quality (RQ). The RQ score is a proxy of the quality (e.g., sampling coverage, reporting details) of the size information provided to the IOTC Secretariat by CPCs (Herrera, 2010; IOTC, 2024). The length bin width was 2 cm and the length bins spanned from 10 to 340 cm. The data were collected from a variety of sampling programs, which can be summarised as follows:

- *FS* and *LS* fisheries: Length frequency samples from purse seiners have been collected from a variety of port sampling programmes since the mid-1980s. The samples are comprised of very large numbers of individual fish measurements. The length frequency samples are available by set type, with sets catches from associated sets typically composed of smaller fish than free school catches (Figure 3). The size composition of the catch from the free-school fishery is bimodal, being comprised of the smaller size range of yellowfin and a broad mode of larger fish. The bimodal distribution is likely to have reflected different types of schools in the catch composition (e.g., free schools of mostly large adult yellowfin or mixed species schools consisting of smaller yellowfin, M. Chassot, pers. comm.). Hence, the relative composition of large (>80cm) vs. small (<80cm) yellowfin in the purse seine free schools fluctuates considerably over time. Between 2010 and 2023, there was a dip in the average size of large fish caught in the FAD fishery and a temporary increase in the average sizes of large fish caught in the free school fishery (Figure 8). There is also a considerable catch of smaller fish taken during free school fishing operations in the Mozambique Channel area in region 2 (Chassot, 2014). The free-school fishery in region 4 appears to catch larger fish (Figure 5).
- *LL* fishery: Length and weight data have been collected from sampling at ports and aboard Japanese commercial, research vessels, and observer programmes. Weight frequency data collected from the fleet have been converted to length frequency data via a processed weight-whole weight conversion factor and a weight-length key. Length frequency data from the Taiwanese longline fleet from 1980−2003 were included in the 2018 assessment, although data from the more recent years were excluded due to concerns regarding their reliability (Geehan and Hoyle, 2013). Length data have also been available from other fleets (e.g., Seychelles, Korean, China, etc.) in more recent years. Analyses of size data show that the average lengths of yellowfin caught by the longline fleet are generally larger in the southern regions, particularly in the southwest (Hoyle, 2021a). There is considerable temporal variation in the length of fish caught (Figure 8), but some of this variation is inconsistent between datasets, such as temporal patterns of variation in the 1970s that differ between length and weight data from the Japanese fleet. For all longline fisheries, there was a marked decline in the size of fish caught by Japan during the 1950s and 1960s, while the size of fish caught stabilised during the 1970s and 1980s (Figure 8).
- *LF* fishery: Length and weight data were collected in port during the unloading of catches for several landing locations and time periods, especially on fresh-tuna longline vessels flagged in Indonesia and Taiwan/China (IOTC-OFCF sampling).
- *GI* fishery: Samples come from Iran, Pakistan, Sri Lanka, and Oman in the Arabian Sea from 1987 and from Indonesia and Sri Lanka in other tropical areas from 1975.

- *BB* fishery: Size data come principally from the Maldivian fleet from 1983 with a large proportion of juveniles. Also, samples from Indonesia and Sri Lanka have also been available for some years but with a low sample size.
- *TR* fishery: Samples come mainly from Comoros in the western IO from 2015, although some small samples are also available from EU (France, Mayotte) and Maldives. In the eastern IO, size data come from Indonesia (1985-1990 and after 2019) and Sri Lanka (1994-2018) fleets.
- *HD* fishery: Samples come from a high diversity of fleets, although the Maldivian fleet has been the most consistent over the years (2005-2023). Limited sampling was conducted over the last decade.
- *OT* fishery: Samples are available from 1983 in the eastern IO and from 1997 in the western IO. The main fleets are the Indian, Indonesian, Sri Lankan, and Maldivian. Limited samples have been available during the last few years.

The IOTC Secretariat provided two size datasets with two distinct grid resolutions:

- *original data*: the size dataset had six main types of grid dimensions (see Table 3 and Figure 9), although $\sim$ 97% of observations were category 5 or 6. A seventh category was also present that covered the Seychelles National Jurisdiction Area, but those observations were removed from the size database. This type of size dataset was used in the 2021 assessment.
- *cwp55 data*: the size dataset was provided at a $5^\circ \times 5^\circ$ grid resolution (Figure 6).

### 4.3.1 Processing

First, we removed gear types with unclear classification (*HOOK*, *HATR*, *PSOB*, and *PS* with unclassified school type *UNCL*). Then, we reduced the number of length bins in the data by summing the number of sampled fish $\geq$ 198 cm and assigning it to the 198 cm length bin. We followed these steps to produce the SS3 size inputs:

- Month information was used to assign quarters, with four quarters from January to December.
- The fishery group was assigned based on the CPC, gear type, and type of association of the fish school.
- The number of sampled fish per length bin was summed and the RQ was averaged by year, quarter, CPC, gear type, fishery group, and grid code.
- The grid code was used to calculate the grid centroid.
- The centroid was used to assign regions. Note that this region assignment varied depending on the type of dataset (see Figure 9 and Figure 6).
- Fishery was assigned based on fishery group and region.
- We converted the length bin width from 2 to 4 cm. To do so, we summed the number of sampled fish from pairs of length bins (e.g., 10 and 12 cm were summed and assigned to 10 cm, 14 and 16 cm were summed and assigned to 14 cm, and so on). After this conversion, we had a total of 48 length bins.

Only for the size dataset with regular grids, we then assigned the catch (in numbers) that corresponded to every observation in the size data (i.e., year, quarter, grid, CPC, and gear type). We found a perfect match for $\sim$ 76% of cases, but there were some size observations

without catch. In order to fill in these catch gaps, we followed an imputation procedure with four levels:

- *Level 1*: Fill in catch gaps with the average catch per grid for a given year, quarter, CPC, gear type, and fishery group.
- *Level 2*: Fill in catch gaps with the average catch per grid for a given year, CPC, gear type, and fishery group.
- *Level 3*: Fill in catch gaps with the average catch per grid for a given year, gear type, and fishery group.
- *Level 4*: Fill in catch gaps with the average catch per grid for a given gear type and fishery group.

Figure 10 shows the percentage of size observations that needed each level of imputation.

#### 4.3.2 Reassignment

We conducted the size data reassignment as done for the catch data in Section 4.2.2.

#### 4.3.3 Filtering

In order to remove inconsistent patterns in the length-frequency data, we carried out these filters:

- The first filter was to remove observations with less than 100 fish sampled and not considered best quality based on the RQ score.
- *LL* fishery: a review of the longline size data shows that the sampling behaviour of Taiwanese and Seychelles fleets (mostly reflagged Taiwanese vessels) have changed over time, with patterns in the logbook length data inconsistent with other fleets (Hoyle, 2021a), and as such the WPTT23 (Data Preparatory) recommended omitting all Taiwanese and Seychelles logbook length data from the 2021 assessment (IOTC, 2021). Following this advice, we removed length frequency data from the Taiwanese and Seychelles longline logbooks from the final length frequency data sets. Length data collected by observers from these fleets were retained in the final data set.
- *LL* fishery: longline length frequency data during 1970-1995 and 2010-2023 in region 1a was removed.
- *LL* fishery: attempts to fit the size data of this fishery in past assessments suggested that the large decline in mean size observed before 1960 is inconsistent with the yellowfin population dynamics. Hoyle (2021a) suggests that selectivity may have changed during this early period and recommends avoiding fitting to these data with the same selectivity. Therefore, we omitted longline size data before 1960 for regions 1b, 2, 3, and 4.
- *LL* fishery: longline length frequency data in 2001-2005, 2015, and 2019 in region 4 was removed.
- *LF* fishery: we removed size data before 2005.
- *GI* fishery: we removed size data from the Sri Lankan fleet in 2021.

- *HD* fishery: we only retained size information from the Maldivian fleet. We also removed Maldivian size data in 2003 and 2015 (quarter 1) due to inconsistent patterns. The contribution of the Omani fleet to the total catch has increased in recent years; however, length samples were almost absent (Figure 14). Lengths from the Maldivian fleet is assumed to be quite similar to the Omani fleet since both fisheries operate on dolphin-associated sets.
- *OT* fishery: we removed size data sampled during 2021-2022 in region 1a and data sampled in 2016 in region 4.
- *TR* fishery: we removed size data in regions 1b and 2. Also, we removed size data in region 4 from 2016 to 2019.

The filters applied to the *LL*, *LF*, and *TR* fisheries were also applied in the 2021 assessment. Other filters were exclusive to the current assessment.

#### 4.3.4 Aggregation

In order to aggregate the size data by year, quarter, and fishery for SS3, we followed two approaches: simple and catch-raised aggregation.

- *Simple aggregation*: this type of aggregation was performed only for the *original* size dataset. We summed the number of sampled fish per length bin and averaged the RQ values by year, quarter, and fishery. This aggregation approach was used in the 2021 assessment and assumed that the collection of samples was broadly representative of the operation of the fishery in each quarter.
- *Catch-raised aggregation*: this type of aggregation was performed only for the *cwp55* size dataset. We performed a catch-weighted sum of the number of sampled fish by length bin and a catch-weighted average of the RQ values.

A graphical representation of the availability of length samples is provided in Figure 11. The longline fishery provided size data from the 1960s but with relatively low quality during the first decades. Most of the size information started to be available in the 1980s, and the purse seine fisheries provided the most consistent and high quality size information (Figure 11). The RQ scores did not differ between aggregation methods.

The differences in size compositions between the simple and catch-raised aggregation methods were minimal for most fisheries (Figure 3). The largest differences were found for the *OT* fisheries in region 1a and 4, and for the handline fishery in region 1a. We observed an increase in the mean length for the *LL* fisheries in all regions (Figure 8). In the case of the free school purse seine fishery, we also noted an increase in mean length over time in region 1b. Conversely, we noted a decrease in the mean length for the log school purse seine fishery over the years in regions 1b, 2, and 4, especially from 1980 to $\sim$ 2005. For the handline fishery, we noted an increase in mean length from the 1990 until $\sim$ 2010, and then a decrease until recent years. These patterns were quite similar between the two aggregation methods (Figure 8).

In general, the size compositions used in the current assessment were comparable with the 2021 assessment (both using the simple aggregation approach), although small differences can be observed for the *OT* and *TR* fisheries in region 4 (Figure 12). Regarding mean length, most fisheries other than longline had similar tendencies over time when comparing 2021 and current values. For the *LL* fisheries, size compositions in 2021 had larger mean lengths before 1990

than the current size compositions (Figure 13). This difference is attributed to revisions of size estimation by the IOTC Secretariat regarding the conversion from fish weights to fish lengths.

## 4.4 Indices

### 4.4.1 Longline (LL) CPUE

Standardised LL CPUE indices (1975-2023) were available from a joint workshop held by Japan, Korea, and Taiwan (Matsumoto et al., 2024). The indices were derived following the methodology developed for previous stock assessments. Longline catch and effort data were analysed using a hurdle generalized linear modelling approach (GLM) utilizing either operational data (subsampled $\sim$ 10-30%) or aggregated data at 5°$\times$ 5° grid resolution. The main variables used in the GLM included operation date, fishing location, vessel ID, fishing effort (number of hooks per set), and catch in numbers of fish. Cluster analyses of species composition data for each fleet and model area were used to separate datasets into fisheries that target different species. Selected clusters were then combined and standardized using GLMs. The log-transformed yellowfin catch per number of hooks set was the dependent variable of the positive model component, while the probability of the catch rate being zero was the dependent variable in the binomial model component. In addition to the year and quarter variables, GLMs also included covariates for 5° square location, cluster, and vessel ID.

For the 2021 assessment, quarterly indices were provided using aggregated data only (COVID restrictions prevented the workshop from being held). For 2024, quarterly indices were provided for the aggregated data only, while annual indices were derived from the sub-sampled operational data (Table 4). Moreover, data from regions 1a and 1b were combined as a single region, whereas data from region 1a was not included in previous standardization processes.

For the regional longline fisheries, a common catchability coefficient (and selectivity) was estimated in the assessment model, thereby linking the respective CPUE indices among regions. This significantly increases the power of the model to estimate the relative (and absolute) level of biomass among regions. However, as CPUE indices are essentially density estimates, it is necessary to scale them to account for the relative abundance of the stock among regions. For example, a relatively small region with a very high average catch rate may have a lower level of total biomass than a large region with a moderate level of CPUE.

We determined regional scaling factors that incorporated both the size of the region and the relative catch rate to estimate the relative level of exploitable longline biomass among regions. This approach was also used in the 2021 assessment and is similar to that used in the Western and Central Pacific Fisheries Commission (WCPFC) regionally disaggregated tuna assessments. Hoyle and Langley (2020) proposed a set of regional weighing factors for IO yellowfin based on aggregated longline catch effort data. The authors recommended the estimates by method '8' for the period 1979–1994 (referred to as '7994m8', see Table 2 of Hoyle and Langley (2020)) to be included in the current assessment. The relative scaling factors calculated for regions 1–4 are 1.674, 0.623, 0.455 and 1, respectively.

For each of the principal longline fisheries, the standardized CPUE index was normalized to the mean of the period for which the region scaling factors were derived (1979–1994). The normalized GLM index was then scaled by the respective regional scaling factor to account for the regional differences in the relative level of exploitable longline abundance among regions (Figure 15).

A number of important trends are evident in the CPUE indices:

- The western tropical (region 1b) CPUE increased during the late 1970s and early 1980s, then suddenly declined from 1987 to 1990. After 1990, the CPUE in this region declined at a slower rate until the late 2000s, which coincided with a number of piracy incidents in the western Indian Ocean (2008–2011). After that time, it remained close to the lowest level observed in that region but showed very large seasonal and annual variations. From 2020 to 2023, there was a substantial ($\sim$ 50%) increase in the CPUE indices compared to the 2013-2018 period.
- The eastern tropical region 4 followed a similar pattern until 1990 but then declined steadily and, by 2016, was also close to the lowest level in the time series. The CPUE decline observed from 2007 to 2016 is consistent with a decline in the proportion of yellowfin in the combined tuna catch from the Japanese longline fleet in the eastern IO. It is unclear whether the change in species proportion is related to a decline in the abundance of yellowfin in the region (relative to the other species) or a regional change in the targeting of the fishing fleet. However, there is an indication that there has been a differential shift towards deeper longline gear (greater HBF) in the eastern IO since 2000 and this may indicate a shift in targeting toward bigeye tuna (*Thunnus obesus*) in this region (Hoyle pers. comm. additional JP LL analyses). Such factors may not be adequately accounted for in the standardisation of the yellowfin CPUE data. There is also a substantial increase in CPUE indices in 2022 and 2023.
- Since the 1980s, the CPUE indices for the western temperate region 2 fluctuated, typically with a 3-5 year cyclical trend. From 2019 to 2023, there was a general increase in the CPUE indices.
- The CPUE index values from the eastern temperate region (region 3) are the lowest compared to other regions, reflecting the low regional scaling factor. However, the overall trend in the CPUE indices is broadly comparable to the other regions. The temporal pattern was similar to the western temperate area before 1979. After 1979, catch rates decreased steadily until the mid-2010s and then slightly increased in recent years (2022 and 2023).
- There is an exceptionally high peak in CPUE indices 1976–78 (especially in regions 1 and 2), which is also associated with a high uncertainty. Hoyle et al. (2017) showed that this discontinuity exists in Japanese, Taiwanese and Korean data in different oceans for both bigeye and yellowfin tuna. Hoyle et al. (2017) suggested this is unlikely to be explained by changes to the population or catchability but may be associated with catch reporting and data management.
- The spike in the CPUE indices around 2012 in the west equatorial region (region 1) was evident for most fishing fleets. Several hypotheses have been proposed on what could have caused CPUE to have increased, including a return to fishing in areas that were most affected by piracy. However, further investigation is required.

The values and trends of LL CPUE used in the 2021 assessment and in the current assessment were quite similar for all model areas before 1990 (Figure 16). After 1990, we noted large differences for region 1, where the current CPUE showed consistently larger values ( $\sim$ 40%) than the 2021 CPUE, especially after 2005. For region 4 and after 2005, we also observed slightly larger current CPUE values compared to the CPUE series from the last assessment. On the other hand, this discrepancy was minimal for regions 2 and 3. The standardization procedure for the current and 2021 LL CPUE used aggregated data (Kitakado et al., 2021); however, the

difference between both series might be caused by the inclusion of information from region 1a in the current LL CPUE standardization and/or differences in the species clustering for the tropical fisheries.

The quarterly LL CPUE indices, which used aggregated data in the standardization, were the primary indices included in the 2024 stock assessment. The annual LL CPUE indices using operational data from regions 1 and 4 had a more pronounced decline during the 1990s compared to the quarterly aggregated indices (Figure 17). Further, the R1 annual operational indices did not exhibit a strong increase in recent years. A range of model options incorporating the annual indices was investigated, although the models failed to meet reasonable convergence criteria, indicating a conflict between key model inputs (or model miss-specification), particularly from region 1.

### 4.4.2 Purse seine CPUE indices

The European and associated flags purse seine fishing activities in the IO from 1981 to 2022 have been monitored through the collection of logbooks and observer sampling. Standardised indices of the biomass of yellowfin caught by European purse seiners (Spain and France) from sets on free-swimming schools (1991 – 2022) and sets on associated tuna schools (2010 – 2022) were developed (Figure 18). The free school index was based on the application of a general additive mixed effect model with three components to model (Kaplan et al., 2024): i) the detection rate of free swimming schools per unit search time, ii) the probability that adult yellowfin are present in a set, and iii) the adult biomass per set given presence assuming a lognormal distribution. The log school index was based on the application of two models: a generalized linear mixed model and a spatiotemporal model, both using a hurdle approach (Correa et al., 2024b). These standardizations considered a comprehensive list of candidate covariates, including the effect of the technological improvement related to the use of echosounder buoys and environmental variables. The predicted CPUE over time was obtained using the *predict-then-aggregate* approach, which is considered best practice (Hoyle et al., 2024).

The log school purse seine (*LS*) index mainly informs the biomass of juvenile yellowfin, while the free-school (*FS*) index informs the biomass of the adult portion of the population in region 1b. The *LS* index displays juvenile biomass fluctuations over the years, with larger values during 2013 and 2014, and a remarkable increase after 2020. On the other hand, the *FS* CPUE index showed an increase from the late 1990s until 2004 and then a dramatic decrease until 2009. From 2015, the free school index showed another dramatic decrease until 2018 and a slight recovery after that (Figure 18), possibly associated with the introduction of yellowfin catch limits during that period. Theoretically, the *LL* (region 1) and *FS* CPUE indices should display similar temporal patterns since both contain information from adults. However, we only noticed comparable trends between 2005 and 2017, and distinct trend after 2017 (Figure 19).

We evaluated the impact of incorporating the *FS* CPUE series as an auxiliary index (i.e., always in conjunction with the *LL* CPUE index) in the assessment model. We could not evaluate the impact of the *LS* CPUE index due to time constraints.

### 4.4.3 Effort creep

It is well recognised that the relationship between PS CPUE and abundance is unlikely to be proportional, as the improvement of catch efficiency due to technology development is challenging to quantify, and the changes in catchability are not fully accounted for in the standardisation

process. Effort creep can be defined as an unquantified increase in the average fishing power over time that disturbs the relationship of proportionality between the index and the stock trajectory (Hoyle, 2024). These changes in catchability over time can affect CPUE indices and, therefore, the outcomes of stock assessments. This is especially important for assessments that lack abundance indices from fishery-independent surveys, which include the majority of the fisheries managed by tuna regional fishery management organizations (RFMOs).

In the case of longline fleets, technological advances include electronic devices to help navigate, communicate, and find target species. Synthetic materials allowed fishers to improve hooks and lines, which increased the probabilities of both hooking and landing. Satellite imagery improved search efficiency. Freezers increased the proportion of time spent on fishing grounds, while equipment for faster longline retrieval increased hooks set without affecting soak time (Hoyle, 2024). Nonetheless, there may have been changes in the relative targeting of yellowfin by the longline fleet that are not adequately accounted for by the standardisation procedure, as evident from the decrease in the catch of yellowfin relative to bigeye tuna. Such changes may exaggerate the decline in the yellowfin tuna CPUE indices and counter any increases in fishing efficiency.

In the 2021 IO yellowfin assessment, a sensitivity analysis was run during the WPTT meeting that included 1% effort creep per year for the entire period of the index, which resulted in changes in the stock depletion level. In the current assessment, we evaluated two different levels of effort creep for the *LL* CPUE index: no effort creep and 0.5% per annum, assuming that the source is associated with vessel turnover.

The WCPFC assessments have often estimated substantial changes in PS FAD-associated fisheries (e.g., McKechnie et al. (2017)). Using a similar approach, Kolody (2018) estimated a catchability increase of approximately 1.25% per year for the standardised purse seine effort for yellowfin from sets on associated schools. Likewise, studies on the French fleet indicate a 10% increase in catch per set associated with echosounder use, equivalent to about 1% per annum, and a 1.7 – 4.0 % increase in efficiency (stable across time) arising from fishing their own floating objects (Wain et al., 2021). In the current assessment, since the inclusion of the *FS* CPUE index was done for model sensitivity, we assumed no effort creep for this index.

## 4.5 Conditional age-at-length data

Age and size information was available for fish sampled between 2009 and 2022 (Figure 20) from the *GERUNDIO* project that aimed the collection and analysis of biological samples of tropical tunas, swordfish, and blue sharks to improve age, growth, and reproduction data for the IOTC. In a first step, Farley et al. (2021) presented this data that contained information from otoliths from 253 yellowfin tuna sampled mainly in the western IO. Then, Farley et al. (2023) presented an updated dataset with age and length information from an additional 136 individuals. To calculate the decimal age of sampled fish, daily and annual ageing methods were used. Decimal age was calculated for each fish with an annual count based on the method developed for yellowfin and bigeye tuna in the western Pacific Ocean (Farley et al., 2020). To find more details on the ageing estimation method, see Farley et al. (2023).

Age and fork length information from a total of 389 individuals was provided to be used in the current stock assessment. This source of information can be included in SS3 as conditional age-at-length data, which is important to inform growth and stock age structure because they provide direct observations of the distribution of fish ages within length classes (Lee et al., 2019). The model fits the observed age-length data along with information from size mode

progression to influence the estimation of the growth curve. The inclusion of this data type was recommended by the yellowfin tuna assessment peer review (M. Maunder et al., 2023).

We followed these steps to produce the input CAAL data for SS3:

- We first removed observations with incomplete age and fleet data, retaining 375 observations.
- Month information was used to assign quarters.
- There was no discrimination between the set type for the purse seine fishery. So, we assigned observations $\leq$ 80 cm as *LS* (log school) and $>$ 80 cm as *FS* free school.
- Fork length was grouped into the length bins used in the assessment model.
- Fish older than seven years were grouped into a single group (age 7).
- Real ages were converted to model ages (i.e., quarters).
- Region was assigned based on geographical location.
- There were several observations ( $\sim$ 37%) with no geographical location, which were assigned to region 1b.

Figure 21 shows the CAAL information included in the assessment model per fishery. Adults were mostly sampled from the *LL*, *FS*, and *HD* fisheries. Conversely, juveniles were mainly present in the *LS* fishery. The lengths ranged from 18 to 182 cm, and all the age quarters were sampled. The largest and oldest fish were observed in the *GI*, *LL*, and *FS* fisheries. Due to the ageing error was not available, we assumed that the age estimation was precise (i.e., no error) in the assessment model.

## 4.6 Tagging

Tagging data was available for inclusion in the assessment model, which consisted of yellowfin tuna tag releases and returns from the Indian Ocean Tuna Tagging Programme (IOTTP) and the main phase of the Regional Tuna Tagging Project-Indian Ocean (RTTP-IO) conducted during 2005−2009. The IOTC has compiled all the release and recovery data from the RTTP-IO and the complementary small-scale programmes in a single database. A total of 54,688 yellowfin tuna were released by the RTTP-IO program. Most of the tag releases occurred within the western equatorial region (region 1b), and a high proportion of these releases occurred in the second and third quarters of 2006 (Figure 22). Limited tagging also occurred within regions 1a and 2. The model included all tag recoveries up to the end of 2014. The spatial distributions of tag releases and recoveries are presented in Figure 23.

A total of 9,916 tag recoveries (after removing tags with unknown recovery date or length) could be assigned to the fisheries included in the assessment model. Almost all of the tags released in region 1 were recovered in the home region, although some recoveries occurred in adjacent regions, particularly in region 2. A small number of tags were recovered in region 4 (from tags released in region 1b) and there were no tags recovered from region 3 (Figure 24). Most of the tag recoveries occurred between mid-2006 and mid-2008 (Figure 22). The number of tag recoveries started to attenuate in 2009, although small numbers of tags were still recovered up to the end of 2014.

Most of the tags were recovered by the purse seine fishery within region 1b (Figure 24). A significant proportion (35%) of the tag returns from purse seiners were not accompanied by information concerning the set type. These tag recoveries were assigned to either the free-school or log fishery based on the expected size of fish at the time of recapture; i.e. fish larger than 80 cm at release were assumed to be recaptured by the free-school fishery; fish smaller than 80

cm at release and recaptured within 18 months at liberty were assumed to be recovered by the floating object fishery; fish smaller than 80 cm at release and recaptured after 18 months at liberty were assumed to be recovered by the free-school fishery.

Tag releases were stratified by release region, time period of release (quarter), and age class for incorporation into the assessment model. The recaptures by fishery for each release group inform the assessment model on fishing mortality and abundance and fish movement. Therefore, factors that might have affected the interpretation of tag returns need to be accounted for to minimise potential bias. Fu (2020) provides a summary of how the tag data were incorporated into the assessments of IOTC tropical tuna species, and below is a description of the procedure applied to yellowfin tuna.

#### 4.6.1 Age assignment of tag release

The age at release was assumed based on the fish length at release and the average length-at-age from the yellowfin growth function (see Section 5.1.3). Fish aged 15 quarters and older were aggregated in a single age group. Tag releases in regions 1a and 1b were stratified in separate release groups due to the spatial separation of the individual release events. A total of 54,392 releases were classified into 131 tag release groups. Most of the tag releases were in the 5−8 quarter age classes (Figure 22).

#### 4.6.2 Initial tagging mortality

Hoyle et al. (2015) examined the effects of various covariates (e.g., individual tagger effect) on tag failures for the RTTP program and estimated a combined effect of 20% for all tropical tuna species relative to a base failure rate. No formal estimate was made for the base failure rate, but the WPTT suggested 7.5% in 2018 based on the assessment of the western and central Pacific tuna species. This equates to a total tag failure rate of 27.5%. For the current assessment, the number of tags in each release group was reduced by 27.5% to account for initial tag mortality.

#### 4.6.3 Chronic tag loss

Tag recoveries were also corrected for long-term tag loss (tag shedding) based on an update of the analysis of Gaertner and Hallier (2015). Tag loss for yellowfin was estimated to be approximately 20% at 2000 days at liberty. This was accounted for through the SS3 chronic tag loss parameter (an annual rate of 0.03).

#### 4.6.4 Reporting rate

The returns from tag release group were classified by recapture fishery and recapture time period (quarter). The results of associated tag seeding experiments conducted during 2005−2008 have revealed considerable temporal variability in tag reporting rates from the IO purse-seine fishery (Hillary et al., 2008b). Reporting rates were lower in 2005 (57%) compared to 2006 and 2007 (89% and 94%). Quarter estimates were also available but were similar in magnitude (Hillary et al., 2008a). This large increase over time was the result of the development of publicity campaign and tag recovery scheme raising the awareness of the stakeholders, i.e. stevedores and crew. SS3 assumes a constant fishery-specific reporting rate. To account for the temporal

change in reporting rate, the number of tag returns from the purse-seine fishery in each stratum (tag group, year/quarter, and length class) was corrected using the respective estimate of the reporting rate. Following Kolody et al. (2011), Fu (2020), and Fu (2017), a 100% reporting rate was assumed for at-sea recoveries, whereas tags recovered from Seychelles landings were corrected for reporting rates based on the quarterly estimates from Hillary et al. (2008a), and were also corrected for the portion of the total purse-seine catches examined for tags, based the proportions of EU purse seine catch landed in the Seychelles relative to the total EU purse seine catches (Kolody et al., 2011). For example, the adjusted number of observed recaptures for an LS fishery as input to the model, the reporting rate ($R'_L$) was calculated using the following equation:

$$R'_L = R_L^{sea} + \frac{R_L^{sez}}{P^{sez} r^{sez}}$$

where:

- $R_L^{sea}$ is the number of observed recaptures recovered at sea for the *LS* fishery
- $R_L^{sez}$ is the number of observed recaptures recovered in Seychelles for the *LS* fishery
- $r^{sez}$ is the reporting rates for purse seine tags removed from the Seychelles
- $P^{sez}$ is the scaling factor to account for the EU purse seine recaptures not landed in Seychelles

The adjusted number of recaptures for the *FS* fishery was calculated similarly. The SS3 reporting parameters for the purse seine fisheries were subsequently fixed at 100% in the model. Some of the other (no purse-seine) fisheries also returned a substantial number of tags. There are no direct estimates of fishery-specific reporting rates for these fisheries. The reporting rates for these fisheries are estimated within the assessment model.

### 4.6.5 Small-scale tagging programmes

Additional tag release/recovery data are available from a number of small-scale tagging programmes. The data set included a total of 7,828 tags released during 2002-08, primarily within regions 1b (70%) and 4 (28%). A total of 366 tag recoveries were reported, predominantly from the bait boat fishery in region 1a. There has been no comprehensive analysis of these data and there is no information available concerning the fishery-specific reporting rate of these tags. The tag release/recovery data from the SS tagging programmes were not incorporated in the current range of assessment models. Earlier analysis indicated that the stock assessment results were relatively insensitive to the inclusion of these data (Langley et al., 2012). Fu et al. (2018) investigated a range of alternative options for processing and incorporating the tagging data into the assessment model (see Table 5 of Fu et al. (2018)). These exploratory analyses are not repeated in the current assessment.

## 4.7 Environmental data

The 2018 assessment included a range of environmental data to investigate the potential for incorporating environmental covariates to inform the movement of fish. However, although there is evidence that there may be an association between the movement of yellowfin tuna and seasonal and temporal changes in ocean conditions in the IO, the potential relationship between environmental indices and fish movement is unclear. Langley (2016) and Fu et al. (2018)

suggested that these environmental indices had no influence on the estimation of yellowfin tuna movement rates of different life stages between adjacent model regions, and seasonal variation in movement may be better accounted for by models that can explicitly incorporate seasonal effects (Fu et al., 2018). Therefore, environmental information was not included in the 2021 assessment.

A significant negative association between the IODs and the catch rates of yellowfin tuna was observed (Lan et al., 2020, 2013), as well as an impact of the ENSO on catch rates near the Arabian Sea (Lan et al., 2020). Langley et al. (2023) found that periods of strong recruitment in regions 1 and 4 appear to correspond with oceanographic conditions indexed by Dipole Mode Index (DMI), but they did not find apparent correspondence between yearly trends in IO environmental conditions and the *LL* CPUE indices from each of the model regions. There is no strong indication that the catchability of the equatorial longline fisheries (region 1 and 4) is strongly influenced by the prevailing environmental conditions, but there is some indication that oceanographic conditions may influence short-term (1-2 yr) variation in longline catchability in the subtropical regions (region 2 and 3) (Langley et al., 2023). Variability in deviates of movement rates between region 1 and 4 is broadly consistent with the fluctuations in the DMI with higher movement estimated under positive IOD conditions.

Since no strong relationships between IO environmental conditions and the yellowfin dynamics have been identified, we did not incorporate any environmental variable in the current assessment; however, we suggest further studies on this topic.

# 5 Model parameters

## 5.1 Population dynamics

The main model configuration partitions the population into four spatial regions (Figure 1) and 29 or 41 quarterly age classes ($0-28+$ or $0-40+$), both sexes combined. The last age-class ($28+$ or $40+$) comprises a *plus group* in which mortality and other characteristics are assumed to be constant. Age quantities are partitioned into 48 4-cm length bins ranging from 10 to 198 cm, which covers the main size range observed for yellowfin in the IO. The population is monitored in the model at quarterly time steps, extending through a time window of 1950–2023. The main population dynamics processes are as follows.

### 5.1.1 Recruitment

Recruitment in SS3 is defined as the appearance of age class 0-quarter fish in the population. Yellowfin tuna spawning occurs all year round, with higher activity from November to February and higher batch fecundity estimates (i.e., number of oocytes released) in the largest females (Zudaire et al., 2022). The assessment model assumed that recruitment occurs instantaneously at the beginning of each quarter. This is a discrete approximation of continuous recruitment, but provides sufficient flexibility to allow a range of variability to be incorporated into the estimates as appropriate.

Global recruitment was assumed to be a function of spawning biomass via a Beverton and Holt stock-recruitment relationship (SRR) with a fixed value of steepness ($h$). Steepness is defined as the ratio of the equilibrium recruitment produced by 20% of the equilibrium unexploited spawning potential to that produced by the equilibrium unexploited spawning potential (Francis,

1992). Typically, fisheries data are not very informative about the steepness parameter of the SRR parameters (Lee et al., 2012); hence, the steepness parameter was fixed at a moderate value (0.80) and the sensitivity of the model results to the value of steepness was explored by setting it to lower (0.7) and higher (0.9) values as performed in other tuna RFMOs (Harley, 2011). Deviates from the SRR curve (*recruitment deviates*) were estimated from 1972 to 2021 (i.e., 200 deviates), which covers the period with more data in the assessment (Figure 25). The recruitment deviates were assumed to have a standard deviation of 0.6 ($\sigma_R$) in log scale, which was fixed during the model development process.

Global recruitment was apportioned to the tropical regions 1 and 4. The choice of these regions was based on the temperature preference for the spawning of yellowfin tuna and a minimum temperature for larval survival of about 24°C (Reglero et al., 2014; Suzuki, 1993). The overall proportion of the quarterly recruitment allocated to regions 1 and 4 was estimated and varied for quarters between 1977 and 2021 (180 deviates). A time block was imposed on the temporal deviates of the recruitment distribution parameters, which were divided into two periods: 1977 – 2008 and 2009 – 2021 (both assuming a standard deviation of 1.5 for the deviates). The time block makes it possible to use the average recruitment distribution of the most recent period instead of the long-term average in the model prediction. The selection of the time period is based on the estimated relative trend of regional recruitment distribution.

The 2021 assessment performed a sensitivity model in which recruitment was assumed to occur in all model regions, which allowed the model to have more flexibility in distributing fish in model regions. This might be preferable because although spawning and larvae require water of at least 24°C, the growing juveniles can move to other regions before they reach the size of recruitment to the fishery, when the model first needs to predict their distribution. Their pre-recruitment movement behaviour is likely to differ from older fish. However, there was a significant increase in the computational overhead associated with the additional temporal deviates parameters for the regional recruitment distribution (unless a stationary regional distribution is assumed). As such, this parametrization was not explored in the current assessment.

### 5.1.2 Initial population

The population age structure in the initial time period in each region was assumed to be in an unexploited, equilibrium state. As noted above, the population is partitioned into quarterly age classes with an aggregate class for the maximum age (plus-group). The aggregate age class makes it possible for the accumulation of old and large fish, which is likely in the early years of the fishery when exploitation rates were assumed to be absent.

### 5.1.3 Somatic growth

The 2021 assessment used growth parameters that replicated the growth curve derived by Fonteneau (2008) (Figure 26). This growth curve did not follow the traditional von Bertalanffy growth pattern, displaying slow growth between 30 and 60 cm fork length and faster growth between 60 and $\sim$ 120 cm, so the authors had to approximate it in SS3 by varying the growth rates ($k$ parameter) for age quarters from 2 to 13. Dortel et al. (2015) estimated growth parameters by integrating otolith readings, growth increments from mark-recapture data, and modal progressions from purse seine length frequency data. Mean length-at-age estimates were comparable to the values estimated by Fonteneau (2008) for young ages and then diverges, with larger length-at-age due to higher asymptotic length ($L_\infty$).

Eveson et al. (2012) used otolith and growth increment from tag data to estimate a mean asymptotic length of about 130 cm fork length, which was low compared to the maximum lengths historically reported for yellowfin in the IO. Sex-specific estimates from a small subset of samples supported the hypothesis that, on average, males grow to a larger size than females, with the mean asymptotic length estimate being 151 cm for males versus 140 cm for females. Similar differences between sexes have been observed in Atlantic yellowfin (Pacicco et al., 2021) and other *Thunnus* species. The sex-specific estimates were explored in a two-sex model in the 2018 assessment (Fu et al., 2018).

Farley et al. (2021) presented growth estimates using otolith information from the GERUNDIO project and found quite different growth patterns compared to previous studies, especially for young ages. In an updated study, Farley et al. (2023) provided new growth estimates based on daily and annual otolith readings validated with radiocarbon (Fraile et al., 2024), and used a two-stage von Bertalanffy curve to fit the age-length data. Farley et al. (2023) found a larger $L_\infty$ parameter compared to previous studies, and also found distinct mean length-at-age patterns between males and females for ages older than 4 years. Since Farley et al. (2023) is the most recent study on growth, we use it in the current assessment. We approximated the two-stage von Bertalanffy curve by varying the growth rate $k$ parameters for age quarters 2 to 13 (Figure 26). We also tested the modelling of different $L_\infty$ parameter for males and females (Figure 27) in a sensitivity run.

Farley et al. (2023) found a larger variation of lengths for older ages, as also seen in previous growth studies. In SS3, we modelled the variation of lengths at ages as a function of mean length-at-age. The parameters of the length-weight relationship were updated in 2016 using information from more than 20,000 fish sampled since 1987 (Chassot et al., 2016), which were used in the 2021 assessment. Zudaire et al. (2022) also estimated these parameters using data from the GERUNDIO and EMOTION programs, which were similar to the ones estimated in Chassot et al. (2016) despite the much smaller sample size. Therefore, we used the length-weight parameters from the 2021 assessment (Figure 28).

### 5.1.4 Sexual maturity and fecundity

The 2021 assessment used the length-based maturity ogive from Zudaire et al. (2013), who presented two alternative maturity ogive based on either the cortical alveolar or vitellogenic stages of ovarian development. Fu et al. (2021) converted the length-based ogive to age-based ogive, assuming an equilibrium population age-length structure. Fu et al. (2018) showed that the assessment estimates are not sensitive to whether age or length-based ogive was used. The maturity ogive based on cortical alveolar stage development indicated that the onset of maturity occurs at about age 5 quarters ($\sim$ 75 cm) and full maturity was attained at about 12 age quarters. The maturity ogive based on vitellogenic stage development was offset by about three quarters. The 2021 assessment included only the ogive based on cortical alveolar stage development.

Zudaire et al. (2022) provide the most recent review on the sexual maturity of yellowfin in the IO. They analyzed samples from a few data collection programs since 2009, principally from the purse seine fleet and western IO, and estimated the length at 50% maturity ($L_{50}$) by applying two different maturity thresholds depending on the oocyte development stage considered: physiological maturity (threshold established at cortical alveolar oocyte development stage) and functional maturity (threshold at initial vitellogenic oocyte development stage). The former method estimated $L_{50}$ at 75 cm fork length (FL) and was used in the 2021 assessment, while the

latter estimated $L_{50}$ 101.7 cm fork length. Using functional maturity as the threshold to determine sexually mature fish is preferable since it guarantees that the fish will inevitably reproduce in the very short term (Pacicco et al., 2023; Zudaire et al., 2022), so we used the functional maturity curve in the current assessment. In addition, we decided to use length-based maturity since some of the model runs estimated growth parameters, which can affect the length to age conversion. In SS3, we modelled maturity-at-length using a logistic function with two parameters (Figure 29) that remained fixed during the model development process.

Zudaire et al. (2022) also examined the sex ratio and found it to be close to 1:1 for most of the small and intermediate size classes (smaller than 115 cm). However, males dominated the samples for sizes larger than 150 cm, probably due to differences in growth between sexes (Farley et al., 2023). Based on the current evidence, we assumed a sex ratio of 0.5. Finally, we assumed that the egg production ($eggs/kg$) was a linear function of female body weight (kg).

### 5.1.5 Natural mortality

Hoyle (2021b) reviewed approaches for estimating natural mortality ($M$) and provided estimates of age-specific natural mortality based on methods described in M. N. Maunder et al. (2023). The new approach assumes a Lorenzen-type relationship (Lorenzen, 1996) between natural mortality and length or weight. This relationship models high $M$ for younger fish, which then declines as fish get older. Mortality increases again after individuals become mature, and this increase is linked to the proportion of mature fish following a logistic curve. This approach has also been applied in the 2021 South Pacific albacore stock assessment (Castillo-Jordan et al., 2021).

The 2019 yellowfin tuna assessment in the Atlantic Ocean (ICCAT, 2019) adopted natural mortality estimates based on the results of a study of the relationship between maximum observed age and natural mortality (Then et al., 2015), and maximum age estimates derived from an aging study using annual otolith increments (Andrews et al., 2020; Pacicco et al., 2021). A Lorenzen (1996) natural mortality form was developed with an $M$ of 0.35 for adult yellowfin, based on a validated maximum observed age of 18 years (Andrews et al., 2020). This level of $M$ was slightly lower than the low $M$ option considered in the 2018 assessment of IO yellowfin tuna.

In the 2021 assessment, two alternative estimates were provided: one based on a maximum age of 10.5 years observed for IO yellowfin tuna (Shih et al., 2014), and the other based on the validated maximum age of 18 years obtained in the Atlantic Ocean (as the *Mlow* option above). The range estimates by Hoyle (2021b) were somewhat lower than the natural mortality options in the basic and *Mlow* models for the sub-adults/adults, particularly for juveniles, and do not include a 'hump' to allow for higher female natural mortality 1.5 years after maturation. The hump is removed because the change in sex ratio at length observed for yellowfin tuna can be partly or completely explained by males growing larger than females (Pacicco et al., 2021). The estimates by Hoyle (2021b) were examined in sensitivity models in 2021.

Artetxe-Arrate et al. (2024) presented the most recent update of natural mortality estimates for yellowfin in the IO. They presented three $M$ estimates based on different methodologies that relied on biological features of the stock such as growth or maximum age. For the current assessment, we used the $M_a$ estimates following Hamel and Cope (2022) and then rescaled based on Lorenzen (2005). In SS3, the reference natural mortality is calculated from $M_{ref} = 5.4/A_{max}$, where $A_{max}$ is the assumed maximum age in the population equal to 11.7 years based on Farley et al. (2023). $M_{ref}$ is the natural mortality that corresponds to the age at the 95% maturity,

assumed to be 16 age quarters based on Zudaire et al. (2022). Then, the rescaling of natural mortality at age is performed as a function of the $L_\infty$ and $k$ growth parameters. The resulting $M_a$ is shown in Figure 30.

#### 5.1.6 Movement

Reciprocal movement was assumed to occur between adjacent model regions, specifically R1-R2, R1-R4, R3-R4 (Figure 31). Movement is parameterised as the proportional redistribution of fish among regions, including the proportion remaining in the home region. The redistribution of fish occurs instantaneously at the end of each model time step. Movement was parameterised to estimate differential movement for young (2–8 quarters) and old ($\geq$ 9 quarters) fish to approximate potential changes in movement dynamics associated with maturation. Thus, for each movement transition two separate movement parameters were estimated. Fish did not commence moving until the end of age 2 quarters.

It is not possible to estimate seasonal movements directly due to the temporal model configuration. The seasonal variation in the longline CPUE indices and the purse-seine catches, particularly in region 2, indicate that there are likely to be significant seasonal changes in the regional abundance of yellowfin. Seasonal movement dynamics were investigated in the 2018 assessment and Langley et al. (2023) by correlating movement parameters with the environmental indices or using an alternative model structure that can explicitly estimate seasonal movements. However, neither option was able to explain the magnitude of variability exhibited in the longline CPUE nor have any significant effect on the model results.

### 5.2 Fishery dynamics

#### 5.2.1 Fishing mortality

Fishing mortality was modelled using the hybrid method, which uses Pope's equation and an iterative method to approximate the fishing mortality ($F$).

#### 5.2.2 Catchability

Since we performed a regional scaling to the standardized LL CPUE (see Section 4.4.1), we only estimated the catchability parameter for LL CPUE in region 1b. Catchability in regions 2, 3, and 4 were mirrored to the catchability in region 1.

#### 5.2.3 Selectivity

- Longline ($LL$) fisheries: Assumed to be age-specific, time-invariant, and principally parameterised with a logistic function that constrains the older age classes to be fully selected ("flat top"). However, the resultant fits to the LL length compositions from the equatorial LL fisheries were poor: persistently over-estimating the lengths prior to 2000 and under-estimating the lengths for the subsequent period. Therefore, some configurations also included time-variant selectivity with two blocks: before 2000 (double-normal parametrization) and after 2000 (logistic parametrization) for *LL 1b* and *LL 4* to account for changes in fleet contribution to length data (Figure 32). The selectivity in each fishery

is shared by the corresponding set of *LL* CPUE indices. The LL CPUE index produced for region 1 was linked to the *LL 1b* selectivity.

- Longline fresh tuna (*LF*) fishery: The *LF 4* fishery was age-specific and parameterised using a logistic function.
- Purse seine (*FS* and *LS*) fisheries: Assumed to be length-based and formulated using a cubic spline with five nodes. The nodes were specified to approximate the main inflection points of the selectivity function. We only estimated selectivity parameters for fisheries in region 1b, while purse seine selectivities in other regions (2 and 4) were mirrored to the 1b selectivities.
- Handline (*HD*) fishery: Assumed to be age-specific and followed a logistic function.
- Gillnet (*GI*), baitboat (*BB*), and other (*OT*) fisheries: Assumed to be age-specific and used a double-normal parametrization.
- Troll (*TR*) fishery: Assumed to be age-specific and used a double-normal parametrization. Selectivity parameters were only estimated for region 1b, while selectivity in regions 2 and 4 was mirrored to region 1b.

## 5.3 Dynamics of tagged fish

The dynamics of the tagged and untagged fish are generally governed by the same model structures and parameters. An exception to this is recruitment, which for the tagged population is simply the release of tagged fish. The probability of recapturing a given tagged fish is the same as the probability of catching any given untagged fish in the same region. For this assumption to be valid, either the distribution of fishing effort must be random with respect to tagged and untagged fish and/or the tagged fish must be randomly mixed with the untagged fish. The former condition is unlikely to be met because fishing effort is almost never randomly distributed in space. The second condition is also unlikely to be met soon after release because of insufficient time for mixing to occur. Depending on the distribution of fishing effort in relation to tag release sites, the probability of capture of tagged fish soon after release may be different to that for the untagged fish. It is therefore desirable to designate one or more time periods after release as *pre-mixed* and compute fishing mortality for the tagged fish based on the actual recaptures, corrected for tag reporting (see below), rather than use fishing mortality from the general population parameters. This, in effect, desensitizes the likelihood function to tag recaptures in the pre-mixed periods while correctly discounting the tagged population for the recaptures that occurred.

Several analyses of the tag recovery data have been undertaken to determine an appropriate mixing period for the tagging programme (Kolody and Hoyle, 2013; Langley and Million, 2012). The analysis revealed that the tag recoveries from the FAD purse-seine fishery were not adequately mixed, at least during the first six months following release. Conversely, the free-school tag recoveries indicate a higher degree of mixing within the fished population. Most of the tagged yellowfin were in the length classes that are not immediately selected by the free-school fishery ($< 90$ cm). A mixing period of about 6−12 months is sufficient for most tagged fish to recruit to the free-school fishery ($> 90$ cm) and no longer be vulnerable to the FAD fishery. However, the maximum displacements of tags reached a plateau within a few weeks of release suggesting rapid movement of yellowfin within the tag release/recovery areas. On basis of the above, it was considered that a mixing period of 3 or 4 quarters was probably sufficient to allow a reasonable

degree of dispersal of tagged fish in the yellowfin tuna population within the primary region of release.

The release phase of the tagging programme was essentially restricted to the western equatorial region. Fu et al. (2018) showed that the recovery rate of tags after three quarters at liberty was similar both in trend and magnitude between the main latitude bands within the western equatorial region, which suggested a reasonable degree of mixing of tagged fish at the regional scale. The distribution of tags throughout the wider IO appears to have been relatively limited as is evident from the low number of tag recoveries from the fisheries beyond region 1b. Tag recoveries from beyond region 1b are unlikely to significantly inform the model regarding movement rates.

Estimates of tag reporting rates from the purse seine fishery were available from tag seeding trials. These estimates were applied to correct the number of tags included in the recovery dataset for the purse seine fisheries within region 1b and region 2 (see Section 4.6 for details). For the other fisheries, there was very limited information available to indicate the tag reporting rates. Fishery-specific reporting rates were estimated based on uninformative priors. All fishery reporting rates were assumed to be temporally invariant and were estimated within the model. The assumptions on tagging dynamics just described did not vary from the 2021 assessment.

## 5.4 Likelihood components

The total likelihood is composed of a number of components, including the fit to the catch data, indices of abundance (CPUE), length frequency, CAAL, and tagging data. There are also contributions to the total likelihood from the recruitment deviates and priors on the individual model parameters. Details of the formulation of the individual components of the likelihood are provided in Appendix A of Methot and Wetzel (2013).

### 5.4.1 Catch

The catch data assumed a lognormal error structure. There is no objective estimates of the degree in uncertainty in aggregated catch data, therefore, like in the 2021 assessment, we assumed a value of 0.01 for every observation.

### 5.4.2 Indices of abundance

The CPUE indices assumed a lognormal error structure. The 2021 assessment assumed a coefficient of variation for every *LL* CPUE observation of 0.2, which did not differentiate low or high uncertain CPUE estimates. In the current assessment, the coefficient of variation associated with every CPUE observation was derived from the standardization method (e.g., GLM) and then rescaled to a mean of 0.2 per region for the *LL* indices. A constant coefficient of variation of 0.2 was also evaluated. In a sensitivity run, the *FS* CPUE index assumed a constant CV of 0.3 for every observation. The *LS* CPUE index was not tested due to time constraints.

### 5.4.3 Length frequency

The length frequency assumed a multinomial error structure. The reliability of the length composition data is variable across fisheries and over time. For that reason, it was considered that the length composition data should not be allowed to dominate the model likelihood and directly influence the trends in stock abundance. In the 2018 and 2021 assessments, an overall input sample size (ESS) of 5 was assigned to all length composition observations (all fisheries, all time periods). That approach essentially gave the entire length composition data set a relatively low weighting in the overall likelihood, but did not differentiate data quality among fisheries and years. In order to incorporate a better proxy of size quality information, the current assessment treated the RQ values (see Section 4.3) as the input sample size.

### 5.4.4 Conditional age-at-length

For sensitivity analyses, a further log-likelihood component involves the CAAL dataset. The observed age composition within each length interval is assumed to be multinomially distributed, and this forms the basis of the likelihood component for this data source.

### 5.4.5 Tagging data

There are two components of the tag likelihood: the multinomial likelihood for the distribution of tag recoveries by fleets over time and the negative binomial distribution of expected total recaptures across all regions. The relative weighting of the tagging data was controlled by the magnitude of the over-dispersion parameters assigned to the individual tag release groups. Like in the 2018 and 2021 assessments, the overdispersion parameters for all tag release groups were set at 7 in the current assessment, determined iteratively from the residuals of the fit to the tag recovery data (observed – expected number of tags recovered). During the model implementation, we tested downweigthing the influence of the tagging likelihood component on the total likelihood by 90% (i.e., lambda of 0.1).

## 5.5 Parameter estimation and uncertainty

The parameters of the model were estimated by minimising the sum of the negative log-likelihood components associated with each of the data components plus the negative log of the probability density functions of the priors and recruitment deviates. Estimation was conducted in a series of phases, the first of which used relatively arbitrary starting values for most parameters. Some parameters had starting values consistent with available biological information. Models were run with a gradient criterion of $10^{-4}$. The Hessian matrix computed at the mode of the posterior distribution was used to obtain estimates of the covariance matrix, which was used in combination with the Delta method to compute approximate confidence intervals for parameters of interest.

The structural uncertainty grid attempts to describe the main sources of structural and data uncertainty in the assessment. For the current assessment, we have continued with a factorial grid of model runs which incorporates the following sources of uncertainties:

- Effort creep in the LL CPUE indices: no effort creep and a 0.5% annual rate.
- Steepness: 0.7, 0.8, and 0.9.

- *LL* selectivity: assuming time-invariant selectivity for *LL* fisheries, or modelling two time-blocks: before and after 2000.

### 5.5.1 Diagnostics

Misspecification of key parameters or assumptions in integrated stock assessment models such as SS3 can strongly impact model outcomes and the estimates of quantities of management interest (Mangel et al., 2013). Model misspecifications can include incorrect specifications of important biological parameters or selectivity functions, or not accounting for spatial stock structure (Punt, 2019) or temporal variation in recruitment (Thorson et al., 2019), growth (Correa et al., 2021), selectivity (Stewart and Monnahan, 2017), among others. Failing to account for an important process can also lead to conflicting information among data sets (Francis, 2011) and retrospective and forecast bias (Carvalho et al., 2017).

In order to evaluate model misspecification, we applied a series of diagnostics tools described in Carvalho et al. (2021) to candidate reference models. Regarding convergence, we examined the maximum final gradient, invertible Hessian, parameters stuck on bounds, and a jittering analysis to evaluate if models converged to a global solution. Regarding goodness-of-fit, we analyzed residuals patterns (*RunTest*) and the root mean square error (RMSE) for CPUE and mean length (Carvalho et al., 2017).

For highly complex population models fitted to large amounts of often conflicting data, it is common to have difficulties estimating total abundance. Therefore, a likelihood profile analysis was undertaken of the marginal posterior likelihood with respect to population scaling ($R_0$), variability in recruitment ($\sigma_R$), steepness ($h$), asymptotic length ($L_\infty$), and natural mortality ($M$). Retrospective analyses were conducted as a general test of the stability of the model, as a robust model should produce similar output when rerun with data for the terminal quarters sequentially excluded (Cadigan and Farrell, 2005). We used the Mohn's $\rho$ (Mohn, 1999) as an indicator of retrospective patterns for spawning biomass, recruitment, and $F/F_{MSY}$.

Providing fisheries management advice requires predicting a stock's response to management and checking that predictions are consistent with future reality (Kell et al., 2016). The accuracy and precision of the predictions depend on the validity of the model, the information in the data, and how far ahead of time predictions are made. We applied the hindcasting cross-validation technique (*HCXval*) to CPUE data. Likewise, we calculate the mean absolute scaled error (MASE) (Hyndman and Koehler, 2006) for CPUE and mean length. The MASE is built on the principle of evaluating the prediction skill of a model relative to a naive baseline prediction. Finally, we also evaluated possible misspecification of biological parameters by identifying trends in recruitment deviates (Merino et al., 2022).

## 5.6 Stock status

Maximum Sustainable Yield (MSY) based estimates of stock status were determined for the final candidate reference models, and those included in the uncertainty grid. The incorporation of both model and estimation uncertainty into management advice is necessary to accurately capture the current state of stock status as model uncertainty is not always greater than estimation uncertainty (Ducharme-Barth and Vincent, 2022). MSY based reference points were derived for the model options based on the average F-at-age matrix for 2023. The period was considered representative of the recent average pattern of exploitation from the fishery. An impact plot was also produced to display the influence of each fishery groups on spawning biomass.

# 6 Model runs

## 6.1 Stepwise revisions

All recommendations made by the 26th Working Party on Tropical Tunas data preparatory meeting were implemented in SS3 as stepwise, iterative model implementation to a proposed set of reference models (Table 5). The effect of each revision on spawning stock biomass estimates is highlighted in Figure 33 and the effect on stock spawning biomass relative to unfished biomass ($SSB/SSB_0$) in Figure 34. The stepwise process started with a base case, which was chosen from the uncertainty grid of 2021: assuming the regional structure adopted in the basic model (*io*), steepness of 0.8, catchability of LL CPUE in region 1 constant with time, growth from Fonteneau (2008), no downweighting the tagging data, and natural mortality based on values applied in the Pacific Ocean (*io_h80_q1_Gbase_Mbase_tlambda1*).

After all the recommendations from the data preparatory were included (run 17, Table 5), Figure 33 and Figure 34 show that the main changes in the trajectory and virgin biomass were due to the update of the CPUE, length, M, growth and consequently the update of the purse seiners selectivity. We noticed that the reference model 1 (i.e., run 17) overestimated the catch of large fish for longline fishery (Figure 35). The selectivity of longline CPUE of each region is linked to the longline fishery in the corresponding region, and therefore, this makes it especially important to understand the source of the overestimation of catch of large fish. Different alternatives were explored:

- *Higher natural mortality for ages older than 5 years*: We mirrored the natural mortality at ages older than 5 years with the previous values. However, the residual pattern was not improved and this option was no further explored.
- *Sex-specific growth*: Different growth for males and females. However, the residual pattern was not improved and this option was no further explored.
- *Estimate* $L_\infty$: Estimate it by introducing the CAAL data. The estimated $L_\infty$ was 163 cm but the residual pattern was not improved. Therefore, this option was no further explored.
- *Divide LL fisheries into two periods*: We divided the LL fisheries in regions 1, 2, and 4 into two parts: before and after 2000.

The length compositions of LL fisheries show differences in the mean length across time before and after 2000, observing larger fish, on average, in the second period. Prior to 2000, the length composition data came mainly from the Japanese fishery, but after 2000 the main source of length composition data was the Taiwanese fleet. The fits to the length composition data of LL fishery were substantially improved assuming two different fisheries previous and after the year 2000 for longlines in region 1, region 2 and region 4. The selectivity for regions 1 and 4 were estimated with a double normal shape in the first period and logistic in the second. For region 2 both periods were assumed with logistic selectivity. In region 3, the differences were not so obvious and thus the fishery *LL 3* was not splitted. Nevertheless, the selectivities of the *LL* CPUE indices were linked to the longlines' fishery and thus, different options were explored regarding the selectivity of the longline CPUE. It was difficult to define a unique reference model, and therefore, three reference models (RMs) are proposed with different assumptions regarding the selectivity of the *LL* CPUEs.

### 6.1.1 Reference models

- *RM1*: assuming the same selectivity the entire model period, estimated from *LL* fishery as one block (same as run 17 in Table 5).
- *RM2*: assuming the selectivity of the *LL* CPUEs did not change over time and is like the selectivity of *LL* fishery before 2000.
- *RM3*: assuming the selectivity of the *LL* CPUEs changed over time as the *LL* fisheries.

We also explored the effect of effort creep on the *LL* CPUE indices (0.5% per annum), and thus another reference model was also proposed for the assessment:

- *RM4*: apply an effort creep of 0.5% per year to the *LL* CPUE indices, based on the RM2.

In the uncertainty grid of the 2021 assessment, the tagging data was downweighted in half of the models selected for the reference ensemble, due to the bias of the data (on spatial distribution and the lack of mixing) and the general lack of fit to the data. The 2023 external review (M. Maunder et al., 2023) also suggested reducing the effect of the tagging data. Therefore, another scenario was developed as an additional candidate reference model:

- *RM5*: downweight tagging data by 90% (lambda of 0.1 in SS3). The model is based on the RM2.

In addition, based on the RM2, a range of exploratory analyses were carried out.

The temporal trajectory of the estimated spawning biomass was compared between the 5 reference models, as well as the trajectory of $SSB/SSB_0$ (Figure 36). The model with the largest virgin biomass is the RM1 and also the most optimistic in terms of depletion ($SSB/SSB_0$) during the last years.

The estimated $SSB_0$ of three reference models is larger than the base case from the 2021 assessment. The RM4 is the most pessimistic and RM1 the most optimistic. The spawning biomass levels in recent years is similar to the RM1 model (Figure 36). It is estimated that in the 1950s, 1960s, and early 1970s, the spawning biomass was still relatively high, reflecting the relatively low catch and the assumption of equilibrium recruitment during this period. Total spawning biomass declined rapidly from the late 1980s to the mid-1990s, recovered slightly in the late 1990s and early 2000s, and then fell to low levels in 2008–2009. The spawning biomass rebounded slightly from 2009 to 2011 and then increased to the current year with fluctuations, while in the base case from 2021 it showed a decreasing trend.

# 7 Model results

## 7.1 Fits

The model provides a reasonable fit to the overall trend in the CPUE indices for each region (Figure 37). The CPUE indices exhibit a high degree of seasonal variability that is not estimated by the model. There is no discernible temporal trend in the residuals from the fit to the CPUE indices for regions 1, 2 and 3, but for region 4 between 2014 and 2018 the residuals are negative while for the last 2 years the residuals are positive, showing difficulties to estimate the large increase of the LL CPUE in region 4 for the last two years (Figure 38). The large decline in the

CPUE index for the tropical regions over the data period appeared to be consistent with the exploitation history in the regions.

For most fisheries, there is a reasonable overall fit to the length composition data (Figure 39 and Figure 40). For the main longline fisheries (*LL 1b*, *LL 2*, and *LL 4*), the model fit the long-term trends in the average length substantially better (Figure 41 and Figure 42) as a consequence of the split in the *LL* fisheries in two fisheries (Figure 43).

For the main purse seine fisheries (particularly the *FS* fishery), the relative proportion of small ($\leq$ 80cm) and large ($>$ 80cm) fish is variable over time, probably due to size-related schooling behavior of adult yellowfin tuna, resulting in strong residual patterns in the fits (Figure 44). The recent trends in the predicted average fish size for the *FS 1b* and *FS 2* fisheries are broadly consistent with the sampling data with larger fish caught during the mid-2000s and smaller fish from 2010 onwards. There is a marked decline in the average size of fish sampled from the purse seine *LS* fisheries in both region 1b and region 2 (Figure 44), particularly during the mid-1990s. This trend is not evident in the predicted average fish size derived from the model for region 2. There is an improvement in the fits to the length data from the handline and gillnet fisheries in region 1a.

A comparison of the observed and predicted numbers of tags recovered (excluding recoveries during the four-quarter mixing period) by quarterly time period aggregated across tag groups are presented in Figure 45. Overall, the model underestimate the recovered tags. The model cannot fit the high reported tag data during the main recovery period (2007–2009). Most of the tag returns were from the purse-seine fishery in region 1b, to a lesser extent, region 2 (Figure 46). In region 1b, there are several quarters when the model substantially underestimates the number of tag recoveries from both regional purse seine fisheries. These quarters correspond to the first quarter following the four-quarter mixing period for the large releases of tags in 2006 (quarters 2, 3 and 4) and 2007 quarter 3 (see Figure 47). The lack of fit to the recoveries in those quarters suggests that even the four-quarter mixing period may not be sufficient to allow for adequate dispersal of tagged fish in the population. The lack of fit is also spread though time which may indicate that the fishing mortality estimate may be too low and biomass too high, and/or the natural mortality may be too high. The tag recoveries from the non purse seine fisheries are not considered to be very informative and the model has the flexibility to freely estimate reporting rates for these fisheries. Of these fisheries, only the LL fisheries in regions 1b and 2 recovered moderate numbers of tags during the period following the four-quarter mixing phase. The numbers of tags recovered from these fisheries was low relative to the purse-seine fishery and the fishery specific tag reporting rates were estimated to be very low.

## 7.2 Estimates

The estimated parameters in the basic model include: the overall population scale parameter $R_0$, the time series of recruitment deviates, the distribution of recruitment among regions, age specific movement parameters, the fishery selectivity parameters, fishery tag reporting rates, and the catchability parameters for the CPUE indices.

The age-based and size-based (for purse seine fisheries) selectivity functions are presented in Figure 48 and Figure 49. Independent selectivity functions are estimated for the principal longline fisheries (*LL* 1a, 1b, 2–4). In regions 1b and 4, double normal selectivity is assumed prior to 2000 with full selectivity between 10 and 19 age quarters in region 1b and a bit smaller from age 8 to 17 quarters in region 4. After 2000, the logistic shape is assumed for both fishery with full selectivity in age 18 and age 12. In region 2, in both periods is assumed logistic

selectivity but the full selectivity age changes from 12 to 22 age quarters. The fresh tuna fishery (*LF 4*) is estimated to have a relatively similar selectivity to the principal longline fisheries, with logistic selectivity and with full selectivity at age 12 quarters. The logistic selectivity of the handline (*HD 1a*) fishery is estimated to have a full selectivity at age 21 quarters. The associated purse-seine fisheries have a high selectivity for small fish, while the *FS* fishery catches substantially larger fish. For all regions, the selectivity of the purse seine fisheries were held constant through time. The selectivity of the *LS* fishery is relatively broad compared to the modal structure of the length frequency data. The pole-and-line fishery is also highly selective for juvenile fish. Limited or no size data were available for several fisheries, specifically the artisanal fisheries (*OT* 1a and 4) and the troll fishery in regions 1b and 2 (*TR* 1b and 2). Consequently, the selectivity for these fisheries is poorly estimated or, in the absence of size data, assumed equivalent to a fishery with the same gear code in another region. The model did not estimate a significant change of selectivity for gillnet fishery in region 1a, despite the fishery appeared to have caught smaller fish after the 2000s than the early period.

The quarterly recruitment deviates indicate that recruitment varies seasonally (Figure 50). Recruitment deviates were positive from 1996 to 2003 and negative afterwards during 2004–2006, especially during 2005. This low recruitment occurred shortly before the tagging program and may be related to the intention of the model to achieve the estimation of a lower biomass (and a higher fishing mortality) to better predict the tag returns. The low recruitment estimate may also be due to the subsequent decline in CPUE rates in the 2007–2011 piracy period. However, the pattern persists in models where either the tagging data or *LL* 1b CPUE indices (after 2007) were removed, but disappeared when both datasets were removed (see Figure C22 in Fu et al. (2018)), suggesting that the estimation of low recruitment in 2004–2006 was likely related to both factors. After 2012, there is an overall increase of positive recruitment deviates.

Recruitment is parameterised to occur in region 1 and 4 only. The model estimates about 70% and 30% of the total annual recruitment is assigned to regions 1 and 4 in the initial period 1950–1977, respectively. The proportion of total recruitment assigned to either region varies temporally during the estimation period (1977–2021) and the proportion allocated to region 1 has increased to be above 80% since mid-2000s (and viceversa for region 4, Figure 51). The large increase of the recruitment to region 1 coincided with the exceptionally high catches that occurred in the western tropical region between 2003–2006. In a hypothetical model which assumed that the sharp increase in catches in region 1 occurred in region 4 instead, the regional recruitment trend is reversed (Fu et al., 2021).

The model estimates that there is a relatively low degree of connectivity between the two western regions (R1 and R2) and between the eastern regions (R3 and R4), and no longitudinal movement between regions 1 and 4 (Figure 52). This contrasts with the estimates from the early assessment which indicated that the movement between R1 and R2, and between R3 and R4 is relatively high, especially for the juveniles (Fu et al., 2018).

The relative trends of the four model regions are largely comparable (Figure 53), although the overall magnitude of decline is substantially higher in regions 1 and 4. The biomass in region 4 declined steadily throughout the 1990s and 2000s following the trend in the LL CPUE index. For the most recent years, biomass in regions 4 and 3 is estimated to be at a very low level.

Fishing mortality rates for the *HD 1a* fishery increased sharply since 2010, corresponding to relatively high catches from that fishery in 2020 (Figure 54). Estimates of fishing mortality for the *LS 1b* fishery appeared to be low by comparison, considering that this fishery mostly catches juvenile fish (in contrast to the *HD 1a* fishery which catches mostly adults). The highest fishing

mortality rates in area 1 in 2020 were from *HD 1a*, *LS 1b*, but since 2015 both fishing mortality rates are decreasing and *GI 1a* and *BB 1b* increasing.

In region 4, recent fishing mortality rates from the *LF 4* fishery were increasing until 2019 but after start to decrease (Figure 54). Although there remains great uncertainty in annual catches from the fishery during last years (Geehan, 2018). The high fishing mortality rates correspond to the sharp decline in model biomass from the late 2000s and are also related to the selectivity of the fishery, with full selection occurring at age 22 quarters. The *GI 4* fisheries represent the other main sources of fishing mortality in Region 4 until 2014, when it started to decrease (Figure 54). *LS 4* fishing mortality is increasing the last years with the highest fishing mortality in this region the last year. Fishing mortality rates are estimated to be very low in both regions 2 and 3 (Figure 54).

Spatially aggregated, age-specific fishing mortality rates are derived for each model time period (Methot and Wetzel, 2013). Average total fishing mortality rates by region were derived for the last years of the assessment model (2021-2023), with highest fishing mortality for age 20-40 quarters (Figure 55).

## 7.3 Diagnostics

A summary of the diagnostic analysis is shown in Table 7.

### 7.3.1 Jitter analysis

Jitter analysis suggests that RM2 model converged at parameters with the maximum negative log-likelihood, although it also shows sensitivity to the starting values (Figure 56). However, the models converging in the minimum log-likelihood give the same results. From the five reference models proposed in this study, RM4 and RM5 are the only models that have convergence level $< 1e^{-4}$, however, the five reference models inverted the Hessian matrix.

### 7.3.2 Retrospective analysis

The reference models temporal structure treat quarter as years, and this makes to be difficult the use of packages as *ss3diags* for diagnosis of the models. However, with some ad-hoc functions it was possible to estimate Mohn's $\rho$ for SSB, F and recruitment, but we were only able to produce figures for SSB and F. Figure 57 and Table 7 show values of Mohn's $\rho$ for SSB within the acceptable range proposed by Carvalho et al. (2021); however, the Mohn's $\rho$ for $F/F_{MSY}$ is larger than 0.2, but this estimate gets smaller when we add new data (Figure 58). This pattern may be explained due to the revised unfished biomass between the runs, the increase of the estimated $R_0$ in the last years, and the positive recruitment deviates of the last years. For recruitment, the analysis was performed for the RM2, and estimated a Mohn's $\rho$ of -0.08.

### 7.3.3 Run test

Most of the models only passed the runs test for the LL CPUE in region 1, but the RM3, which includes seven LL CPUE indices, also passed the runs test for LL CPUE in region 2 and 4 before year 2000 (Figure 59, Table 7). The best performance on runs test of the mean length of the

fleets also is RM3 model, seven fleets passed the runs test, while the worst performance is again RM1 model (Table 7).

#### 7.3.4 Hindcast analysis

Prediction skill of the model was tested using the hindcasting cross-validation approach of Kell et al. (2021). A skilled model would improve the model forecast compared to the baseline (i.e., random walk), with a mean absolute scaled error value of 0.5 indicative of a forecast being twice as accurate as the baseline and values $> 1$ indicative of average model forecasts worse than the baseline (Carvalho et al., 2021; Kell et al., 2016). The results suggest that the RM3 has the best LL CPUE predictive ability, with prediction skills of the CPUE in region 1 for every season. In the case of RM1 model shows poor LL CPUE prediction skills, only one season has MASE value <1. Figure 60 shows the prediction skills of RM2 in regions 1 and 3 of four seasons show predictive skills of the LL CPUE (MASE $< 1$).

#### 7.3.5 Likelihood profiles

Profiling of $R_0$ was studied with the RM2 model (Figure 61). Results suggest that the minimum log likelihood is close to the estimated value by the model ($ln(R_0) = 11.61$), although the model also show some convergency issues. The profiling also shows some conflict between tagging data, which prefer smaller $R_0$ while the length composition and recruitment data seems to have better fit at higher $R_0$ values. Profiling of steepness suggest to have the minimum log-likelihood at the largest steepness value. Profiling of $\sigma_R$ shows conflict between recruitment and the rest of the parameters. The minimum log-likelihood is close to $\sigma_R = 0.4$, while the model assumes $\sigma_R = 0.6$. Profiling of $L_\infty$ shows the minimum log-likelihood around 165 cm, which is very close to the value assumed in the model 167 cm. Profiling of natural mortality at age 4 was performed assuming steepness of 0.7, 0.8 and 0.9 (Figure 62). The minimum log-likelihood for all of them is $\sim 0.46$, which is the value assumed in the model.

### 7.4 Exploratory analyses

The next sensitivity analyses were explored during the 26th WPTT meeting.

- *Lower natural mortality*: A sensitivity analysis was run assuming M=0.3 at age 4 in RM2 model (Figure 63). The results show a lower R0 was estimated ($Ln(R_0) = 10.45$) with a very clear trend in the recruitment deviates and very high fishing mortality values reaching values of $F > 3$.
- *Natural mortality derived assuming differences by gender*: The same natural mortality as in 2021 was assumed for sensitivity in RM2 model (Figure 64). The estimated $ln(R_0)$ was lower than 11.14, $F/F_{MSY}$ values were lower than 1 the all time series and the spawning biomass in region 1 at the end of the time series reach the SSB at the beginning of the time series.
- *Last years of recruitment deviates*: Sensitivity analysis was performed based on RM2 model to understand the driver of the last year recruitment deviates. The length composition data from quarter 296 (i.e., last quarter of 2021) onward were removed step by step by gear until the last fishery was left Other fishery (Figure 65). So the results suggest that

the model is driven by the information of the other fishery to estimate the high and the low recruitment deviates.

- *Variable CV in the CPUE*: The variability on the CV of the CPUE was estimated scaled on the values of the CPUE standardization but assuming a mean of 0.2 across years (Figure 66). The residual pattern on the eastern part of the region suggest some conflicts between regions (Figure 67).
- *Operational LL CPUE*: different analysis were performed with the operational LL CPUE based on RM2 model. The operational LL CPUE indices were estimated by quarter following the same trend as in the aggregated LL CPUE (Figure 68). However, the model did not converge and the residual pattern show the conflicts on the trend of the CPUE between regions (Figure 69).
- *LL CPUE 2021*: sensitivity analysis performed adding the LL CPUE used in the 2021 assessment to the RM3 model. This analysis indicates that the relative stock biomass in 2021 estimated using the LL CPUE index available in 2021 would be -23% ($SSB/SSB_{MSY}$ 1.01 vs 0.77) than the estimated by the RM3 (Figure 70). For that analysis, the estimation period was terminated in 2021, as in the stock assessment developed in 2021. In order to update this exploration for this report, we carried out the analysis using the *6_SplitCPUE_tag01_EC0_h08* included in the uncertainty grid (see Table 6), which is very similar to the RM3, except that tagging data was downweighed. We run the same exploration with the CPUE data available in 2021 but finalizing the estimation period in 2023 (Figure 71). In other words, the new information for 2021-2023 was included except the new CPUE (i.e., catch and length data). In relative terms, the results are almost identical as the model run previously (difference between the reference model and the model with the old CPUE being -23% ($SSB/SSB_{MSY}$ 1.07 vs 0.78) in 2021 and -24% ($SSB/SSB_{MSY}$ -1.39 vs 1.05) in 2023. A salient aspect of this analysis is that the new information on catch and length data also seems to suggest a change in stock status (at least with this reference model) by 2023, and it is not only the new CPUE that is driving the recovery estimated by this new stock assessment. The combination of the four models used for this exploration: RM3, RM3 using the 2021 CPUE (terminated in 2021), *6_SplitCPUE_tag01_EC0_h08* and *6_SplitCPUE_tag01_EC0_h08* using the 2021 CPUE (terminated in 2023) are shown in Figure 72.
- *Downweighted tagging data and effort creep of 0.5% per year*: was run on RM3. The diagnostics of the three models were similar and also the trends, with the most pessimistic model when effort creep is assumed in the model and the more optimistic with tagging data downweighted (Figure 73).
- *FS CPUE*: The purse seine *FS* CPUE was introduced in the RM2 model assuming a constant CV of 0.3 linked with the selectivity of the *FS* fishery in region 1 (Figure 74). However, the residual pattern show the conflict between this index and *LL* CPUE.
- *Indonesian revised catches*: Indonesian revised catches were introduced in the RM2 model and the time series of SSB shows very small difference with the catches assumed in the reference model (Figure 75).

## 7.5 Stock status

Reference points and estimates of stock status in 2023 were calculated for the five candidate reference models Table 8. MSY-based reference points were derived from the average F-at-

age matrix that represents the most recent pattern of exploitation for the fishery. Note that the biomass at MSY is the benchmark used to evaluate stock status ($SSB_{MSY}$) and it was calculated using two methods: (i) estimated by SS3 from the SRR relationship and (ii) scaled to the recent 20-year recruitment deviates average. The first method is what has been generally used in the IOTC and is based on long-term equilibrium conditions, while the second method aims to update the benchmark to current conditions (see Section 8 for details). The new benchmark will also be used to evaluate the impact of alternative catch limits in the *K2SM*. This adaptation aims to reflect the current conditions and productivity of yellowfin, including the apparent increase in recruitment, which should be followed by an increase in $SSB_{MSY}$ and $MSY$ as well.

Figure 76 shows the stock status for the candidate reference models. For all models, the unfished biomass has been increasing in recent years. This is a consequence of the estimated increase in recruitment (and recruitment deviates), which seem to be necessary for the model to explain the large catch observed since 2003 and, therefore, the recent productivity ($MSY$).

In general, the observed trends are very similar for the five candidate models with an appreciable difference between the models with the *LL* selectivity splitted (RMs 2, 4 and 5) and not splitted (RMs 1 and 3) estimated for the recent years. Overall, the five models estimate a steep decline after 1980s and recovery after 2008, and a steep increase between 2020 and 2023. Fishing mortality trends display a period of overfishing between 1995 and 2020, with a notable reduction in the last three years.

The 26th WPTT decided to build the final uncertainty grid for management advice based on RM1 (*NoSplitCPUE*) and RM3 (*SplitCPUE*), two options for effort-creep (EC0 = 0% and EC1 = 0.5% per year), and three options for steepness ($h$): 0.7, 0.8 and 0.9. With regards to tagging data, the 26th WPTT agreed to use only models with these data downweighed (lambda of 0.1). This resulted in an ensemble of 12 models for management advice (Table 6).

The model *6_SplitCPUE_tag01_EC0_h0.8* was used to illustrate each fishery's impact on the spawning stock biomass using an impact plot (Figure 77).

The estimated trajectories, stock status in 2023 for the 12 models in the final uncertainty grid are shown in Figure 78, Figure 79, and Figure 80. For each scenario, the probability of the biomass being below the $SSB_{MSY}$ level and the probability of fishing mortality being above $F_{MSY}$ were determined using the delta-MVLN estimator (Walter and Winker, 2020), based on the variance-covariance derived from estimates of $SSB/SSB_{MSY}$ and $F/F_{MSY}$ across the model grid. These results indicate that in 2023, the yellowfin tuna stock is determined to be not overfished and not subject to overfishing.

# 8 Projections

In order to evaluate the impact of alternative catch limits on the stock's sustainability and to develop management advice in the form of Kobe 2 Strategy Matrix, we ran ten-year projections from each of the models in the final uncertainty grid.

The projections had the following biological and selectivity specifications:

- Biological features (e.g., natural mortality, growth, maturity) used in the assessment was kept during the projection period.

- The predicted quarterly recruitment was estimated from the deterministic SRR (i.e., no recruitment deviates). The predicted recruitment was then multiplied by a scalar (see below) to represent the recent recruitment levels.
- The distribution of recruitment between regions 1 and 4 for the projection period was based on the recent distribution (2008-2023) estimated by the assessment model. This approach reflects current conditions in recruitment patterns.
- The fleets' selectivity at the terminal year (2023) was used during the projection period.

Regarding the catch per fleet in the projection period, different projected catch scenarios were evaluated. First, The quarterly catch per fleet was calculated using the 2023 catch information. Then, that quarterly catch was multiplied by a factor (*catch multiplier*), which ranged from 0.6 to 1.2, resulting in the final quarterly catch per fleet used for all years in the projection period.

The recruitment deviates showed a strong temporal trend, with higher frequency of negative and positive values before and after $\sim$ 1997, respectively. The use of a scalar to multiply the projected recruitment aimed to represent better the higher frequency of above-average quarterly recruitment estimated in recent years. Three different scalars were tested during the 26th WPTT:

- Scalar of 1: this approach do not vary the recruitment calculated from the SRR, making it representative of the entire assessment period (1950-2023).
- Scalar calculated from the last twelve years: the average recruitment deviates from the last twelve years ($\bar{r}_{12}$) is calculated. Then, the scalar is $exp(\bar{r}_{12})$, which multiplies the recruitment from the SRR in the projection period.
- Scalar calculated from the last twenty years: the average recruitment deviates from the last twenty years ($\bar{r}_{20}$) is calculated. Then, the scalar is $exp(\bar{r}_{20})$, which multiplies the recruitment from the SRR in the projection period.

Figure 81 shows the impact of the three scalars on the projected recruitment for RM 3. The 26th WPTT agreed to use the scalar calculated from the last 20 years due to the stability of catch levels and size compositions observed during that period for the development of management advice and the *K2SM*.

Likewise, to evaluate the impact of the projected catch scenarios on the stock status, the $SSB_{MSY}$ calculated by SS3 was multiplied by the chosen scalar. The stock status was calculated for each model in the final uncertainty grid, and then averaged by quarter. We noticed that all catch multipliers would maintain the stock status above 1 in 2026, and catch multipliers smaller than 1.2 would maintain the stock status above 1 in 2033 (Figure 82).

# 9 Discussion

The update of the model was done using the stepwise process and this facilitates the evaluation of the impact that each element has in the model outputs. So, the updates of the CPUE, the length composition data, natural mortality, growth and the updates of the PS selectivity had the biggest impact in the model outcomes.

One of the most influential factors in this assessment was the LL CPUE. In 2021, the joint LL CPUE was obtained using aggregated data over 1° square grid by month and vessel. In the current assessment, we also had to use the LL CPUE standardized from aggregated data due to it was available at a quarterly time step. Operational data at a yearly time step was also

available and tested in a sensitivity run. There are large differences between the current LL CPUE time series and the ones used in 2021, especially for region 1 and after 1990. The main difference between both standardization processes is the inclusion of region 1a, which used to be excluded in previous standardizations.

Length composition data were revised but mainly the revision on LL length composition data had an impact in the scale of stock spawning biomass and unfished biomass, where now the revised mean lengths previous to the year 2000 are smaller. This had an impact also on the fit to the longline length composition data and also on the selectivity where the model suggest that LL catch smaller fish than in the base case of 2021.

The updated M-at-age value decreases with age following the Lorenzen curve, while in 2021 the M-at-age increased after age 10 quarter up to age 16 and decreased after that. However, the revision of 2021 mentioned that the higher proportion of males at large size was probably due to the differences in growth and, therefore, the WPTT adopted the Lorenzen curve for M. The update in M implied a change in scale of stock spawning biomass with a higher unfished biomass. The fit to the longline fleets were worst and the model starts to overestimate the catch of large fish in the beginning of the time series.

The updated growth model estimated by Farley et al. (2023), estimates higher growth rate for smaller fish than the growth estimated by Fonteneau (2008). In addition, the average maximum length is also higher from 145 to 167 cm. This update had a significant impact in the model and the stock spawning biomass changed downwards with a lower unfished biomass. Due to this update the fits to the longline length composition data got worst and the model overestimates the mean length of the catches in the beginning of the time series and underestimates the mean length at the end of the time series.

All the selectivity estimates are based on age, except for the purse seine fisheries, which are estimated assuming a length-based spline function. The fits to the *FS* were not good from the beginning but with the update of growth the model overestimates the catches of large fish because the model assume that *FS* fleet has full selectivity after 150 cm.

Due to the overestimated mean length of LL fleets and the differences in mean length for LL fleets in region 1,2 and 4 by RM1, it was decided to add as another reference model a model splitting the LL fleets (regions 1, 2, and 4) in two periods: previous to 2000 and after 2000. The best model assumed double normal selectivity for LL in region 1 and region 4 and logistic for both periods in region 2, where bigger fish are present. However, the LL CPUE selectivity is linked to the LL fisheries therefore, another two options were considered about the selectivity of the LL CPUE where the LL CPUE selectivity is linked to the selectivity of the first period (RM2) and another where the LL CPUE selectivity follows the selectivity of LL fishery 2 periods, so splitting the selectivity in two periods (RM3). RM3 is the model with the lowest virgin biomass but lower depletion rate than RM2.

The lack of fit of tagging data as well as the possible bias of tagging data, due to the spatial distribution of the tagging data as well as the possible issues about the mixing were also discussed in previous WPTT, and still the model underestimate the reported tagging data of purse seiners, therefore, the tagging data downweighted by 90% (lambda of 0.1) is presented as another reference model RM4 for the working party. This model estimates higher unfished biomass, and more optimistic status respect to $SSB_0$ than the RM2 model.

An increase of effort creep of 0.5% per year was discussed during the data preparatory meeting. This model estimates the highest depletion rate of the five proposed reference models, although the trend is very similar to the RM2 model or the RM4 model.

Another point of discussion during the 26th WPTT was the estimation of stock status. Due to the remarkable trend in recruitment over time, we decided to rescale the stock status quantities ($MSY$, $SSB_{MSY}$) to recent conditions. There is no standard way to do this, so we decided to use the recruitment deviates estimated in the last 20 years to obtain a scaling factor, which was used to rescale the $MSY$ and $SSB_{MSY}$ estimated by SS3. We acknowledge that this approach needs to be further explored, and future studies could perform simulation analyses to investigate different alternatives to rescale stock status when trend in recruitment is present.

Finally, the change in stock status between the 2021 and current assessment is remarkable. The stock was assumed to be overfished and subject to overfishing in 2021, and currently it is assumed not to be overfished and not subject to overfishing. There is some evidence that support this conclusion. First, the catches from the longline fleet have increase during recent years, as well as their CPUE. Second, the *LS* CPUE, although not used in the current assessment, displayed a remarkable increasing trend from 2020. Third, the total catches have slightly decreased during the last years, which may help with the recovery of the stock. Fourth, the strong modes of young yellowfin found in the *LS* fishery during the last years. The management actions implemented during the recent period could also be another factor that may be helping to improve the stock status of this stock.

# 10 Reproducibility and transparency

A number of authors have recently advocated for a culture of open science and reproducible research (i.e., a change in the transparency and reproducibility of science) (Hampton et al., 2015; Hampton et al., 2013). Proponents of open science and reproducible research highlight a number of benefits, including a more productive and responsible scientific culture, an ability to address larger and more complex questions, as well as a more efficient workflow and ability to reproduce one's own work (Fomel and Claerbout, 2009). Reproducibility of the implementation of stock assessment models has barely explored and published for tuna stocks, but some progress has been made in other RFMOs (e.g., ICES) (Millar et al., 2023).

Magnusson et al. (2022) listed several advantages of make a stock assessment open and reproducible:

- Easy to review the assessment process.
- Easy to pick up a stock assessment from a previous year and run an update, which is particularly important when a new scientist is leading the assessment.
- Easy to modify data treatment or model settings and rerun the entire workflow, which allows more analyses, exploration, and potential improvements.
- Improve traceability and credibility of the assessment process.

For the current assessment, we used Github to store the scripts to analyze and process the raw data, prepare the inputs for SS3, run all the SS3 models in the stepwise implementation, and produce summary figures and tables. The R scripts and instructions to rerun the entire assessment process presented in this document can be found at: https://github.com/Fundacion-AZTI/IOTC_YFT_2024_Assessment/tree/reproducible. In order to reproduce this assessment, the user needs to have some knowledge on R programming, Github, and Stock Synthesis and associated packages such as *r4ss* (Taylor et al., 2021). Future assessments may improve transparency and reproducibility by developing interactive tools to analyze assessment results (Regular et al., 2020), especially for people with no programming skills.

# 11 Acknowledgements

We thank the various fisheries agencies and regional fisheries observers for their support with data collection, as well as the IOTC Secretariat for tyding, analyzing, and preparing the data for the assessment. We thank participants at the 26th data preparatory meeting WPTT26(DP) for their contributions to the assessment. We particularly thank the independent peer review experts for the insightful comments that helped advance and improve this assessment. Finally, we highlight the lifetime of work that Dave Fournier put into ADMB and Richard Methot for leading the development of SS3, without which we could not have performed this assessment.

# 12 Tables

Table 1: Nine fishery groups and codes used in the current assessment.

| Fishery code | Fishery group |
|---|---|
| GI | Gillnet |
| HD | Handline |
| LL | Longline |
| OT | Others |
| BB | Baitboat |
| FS | Purse seine, free school |
| LS | Purse seine, log school |
| TR | Troll |
| LF | Longline (fresh tuna) |

Table 2: Fishery definition in the four-areas assessment configuration. The fishery label is derived from the fishery group and the model region.

| Fishery number | Fishery label |
|---|---|
| 1 | GI_1a |
| 2 | HD_1a |
| 3 | LL_1a |
| 4 | OT_1a |
| 5 | BB_1b |
| 6 | FS_1b |
| 7 | LL_1b |
| 8 | LS_1b |
| 9 | TR_1b |
| 10 | LL_2 |
| 11 | LL_3 |
| 12 | GI_4 |
| 13 | LL_4 |
| 14 | OT_4 |
| 15 | TR_4 |
| 16 | FS_2 |
| 17 | LS_2 |
| 18 | TR_2 |
| 19 | FS_4 |
| 20 | LS_4 |
| 21 | LF_4 |

Table 3: Grid size categories in the *original* size dataset.

| Grid category | Resolution (latitude × longitude) |
|---|---|
| 9 | $30^\circ \times 30^\circ$ |
| A | $10^\circ \times 20^\circ$ |
| 7 | $10^\circ \times 10^\circ$ |
| 8 | $20^\circ \times 20^\circ$ |
| 5 | $1^\circ \times 1^\circ$ |
| 6 | $5^\circ \times 5^\circ$ |
| NJA_SYC | Seychelles National Jurisdiction Area |

Table 4: Features of provided LL CPUE time series. The column *Regions* shows the model area combinations. The main LL CPUE time series used in the current assessment was derived using aggregate data.

| Regions | Data source | Time resolution |
|---|---|---|
| R1/R2/R3/R4 | Operational | Yearly |
| R1/R2/R3/R4 | Aggregated | Quarterly |
| R1+R2/R3+R4 | Operational | Yearly |
| R1+R2+R3/R4 | Operational | Yearly |
| R1+R2+R3/R4 | Operational | Quarterly |

Table 5: Model description during the stepwise implementation.

| Run | Description |
|---|---|
| 1 | Base case from the 2021 assessment. |
| 2 | Update catch. |
| 3 | Update LL CPUE indices. |
| 4 | Update length compositions. |
| 5 | Remove some warnings. |
| 6 | Natural mortality at age 4.07 years of 0.467. |
| 7 | Use growth curve from Farley et al. (2023). |
| 8 | Update age information in the tagging dataset using the new growth curve. |
| 9 | Update maturity from Zudaire et al. (2022). |
| 10 | Update purse seine selectivities. |
| 11 | Update some parameter boundaries. |
| 12 | Update recruitment deviates and bias correction period. |
| 13 | Use RQ as input sample size for length compositions. |
| 14 | Use length compositions from the cwp5x5 dataset. No RQ. |
| 15 | Use length compositions from the cwp5x5 dataset and RQ. |
| 16 | Free up second selectivity parameter for LL 3. |
| 17 | RM1: Apply bias correction ramp. |
| 18 | RM2: Splitted LL fisheries in regions 1, 2 and 4 (RM2). Double-normal parametrization before 2000 and logistic after 2000. |
| 19 | RM3: Splitted LL CPUE selectivities following RM2. |
| 20 | RM4: Apply effort creep of 0.5% per year from RM2. |
| 21 | RM5: downweight tagging data by 90% (lambda of 0.1) from RM2. |

Table 6: Model labels included in the final uncertainty grid. Labels are related to splitting the LL CPUE selectivity, effort creep rate on the LL CPUE, and the steepness value.

| Model label | LL CPUE splitted? | Tagging data downweighted? | Effort creep (% per year) | Steepness |
|---|---|---|---|---|
| 1_NoSplit_tag01_EC0_h0.7 | No | Yes | 0.0 | 0.7 |
| 2_SplitCPUE_tag01_EC0_h0.7 | Yes | Yes | 0.0 | 0.7 |
| 3_NoSplit_tag01_EC1_h0.7 | No | Yes | 0.5 | 0.7 |
| 4_SplitCPUE_tag01_EC1_h0.7 | Yes | Yes | 0.5 | 0.7 |
| 5_NoSplit_tag01_EC0_h0.8 | No | Yes | 0.0 | 0.8 |
| 6_SplitCPUE_tag01_EC0_h0.8 | Yes | Yes | 0.0 | 0.8 |
| 7_NoSplit_tag01_EC1_h0.8 | No | Yes | 0.5 | 0.8 |
| 8_SplitCPUE_tag01_EC1_h0.8 | Yes | Yes | 0.5 | 0.8 |
| 9_NoSplit_tag01_EC0_h0.9 | No | Yes | 0.0 | 0.9 |
| 10_SplitCPUE_tag01_EC0_h0.9 | Yes | Yes | 0.0 | 0.9 |
| 11_NoSplit_tag01_EC1_h0.9 | No | Yes | 0.5 | 0.9 |
| 12_SplitCPUE_tag01_EC1_h0.9 | Yes | Yes | 0.5 | 0.9 |

Table 7: Diagnostics metrics for the five candidate reference models.

| Models | 1_OneBlock_LLsel | 2_TwoBlock_LLsel | 3_TwoBlockCPUE | 4_Dwtag01 | 5_EffortCreep |
|---|---|---|---|---|---|
| Converged | TRUE | TRUE | TRUE | TRUE | TRUE |
| No. pars | 476 | 489 | 489 | 429 | 429 |
| Max gradient | 0.0197 | 0.002 | 0.0047 | < 1e-04 | < 1e-04 |
| Hessian invertible? | TRUE | TRUE | TRUE | TRUE | TRUE |
| No. pars on bounds | 0 | 0 | 0 | 0 | 0 |
| Runs test passed (LL 1) | TRUE | TRUE | TRUE | TRUE | TRUE |
| No. runs test passed (indices) | 1 | 1 | 4 | 1 | 1 |
| No. runs test passed (mean length) | 1 | 5 | 7 | 6 | 5 |
| Hind R1 (No. seasons MASE < 1) | 1 | 3 | 4 | 3 | 3 |
| Hind R2 (No. seasons MASE < 1) | 2 | 2 | 3 | 2 | 2 |
| Hind R3 (No. seasons MASE < 1) | 1 | 0 | 0 | 0 | 0 |
| Hind R4 (No. seasons MASE < 1) | 1 | 1 | 2 | 0 | 1 |
| Mohns' rho SSB | -0.1 | -0.07 | -0.07 | -0.12 | -0.09 |
| Mohns' rho F/Fmsy | 0.36 | 0.28 | 0.25 | 0.28 | 0.24 |
| Trends in rec devs | TRUE | TRUE | TRUE | TRUE | TRUE |

Table 8: Stock status for the candidate five reference models. The asterisk represents the adapted stock status to recent conditions (last 20 years).

| model | 1_OneBlock_LLsel | 2_TwoBlock_LLsel | 3_TwoBlockCPUE | 4_Dwtag01 | 5_EffortCreep |
|---|---|---|---|---|---|
| SSB/SSBmsy* | 1.31 | 0.90 | 1.06 | 0.99 | 0.85 |
| F/Fmsy | 0.76 | 0.11 | 0.96 | 1.01 | 1.17 |
| SSBmsy (mt) | 962351.00 | 902119.00 | 854973.00 | 949367.00 | 942205.00 |
| Fmsy | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |
| MSY (mt) | 369967.60 | 373226.80 | 345797.20 | 379660.80 | 390348.40 |
| SSB0 (mt) | 3185560.00 | 2975810.00 | 2785600.00 | 3122820.00 | 3109390.00 |
| SSBmsy* (mt) | 1205774.15 | 1131688.92 | 1172727.45 | 1163598.20 | 1138698.70 |
| MSY* (mt) | 463549.55 | 468205.01 | 474314.24 | 465333.87 | 471754.25 |
| SSB0* (mt) | 3561099.98 | 3406995.10 | 3419884.50 | 3434993.23 | 3424243.67 |
| SSBmsy/SSB0 | 0.30 | 0.30 | 0.31 | 0.30 | 0.30 |

## 13 Figures

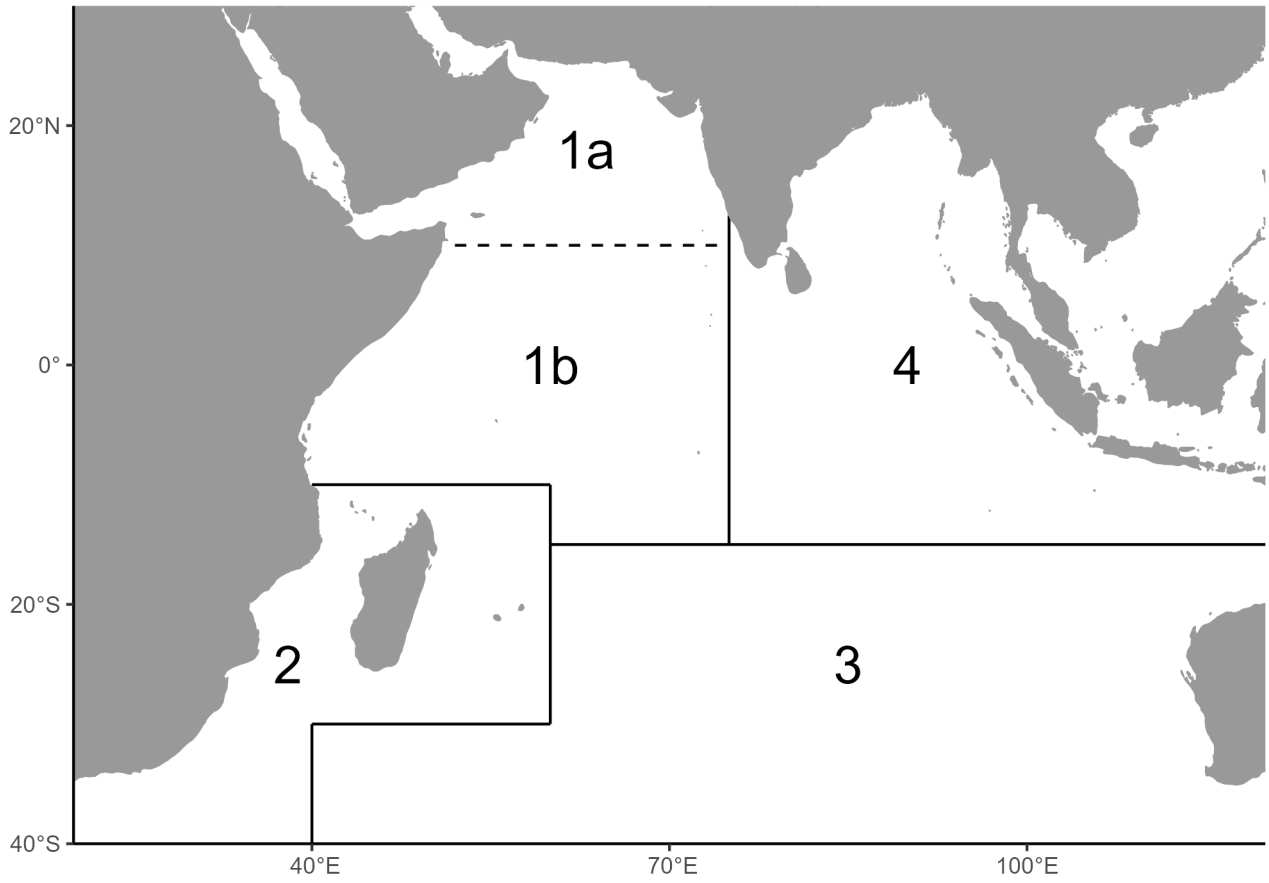


Figure 1: Model areas or regions used in the four-areas assessment configuration. The region 1 is divided into two sub-regions implicitly modelled in the assessment model using the areas-as-fleets approach.

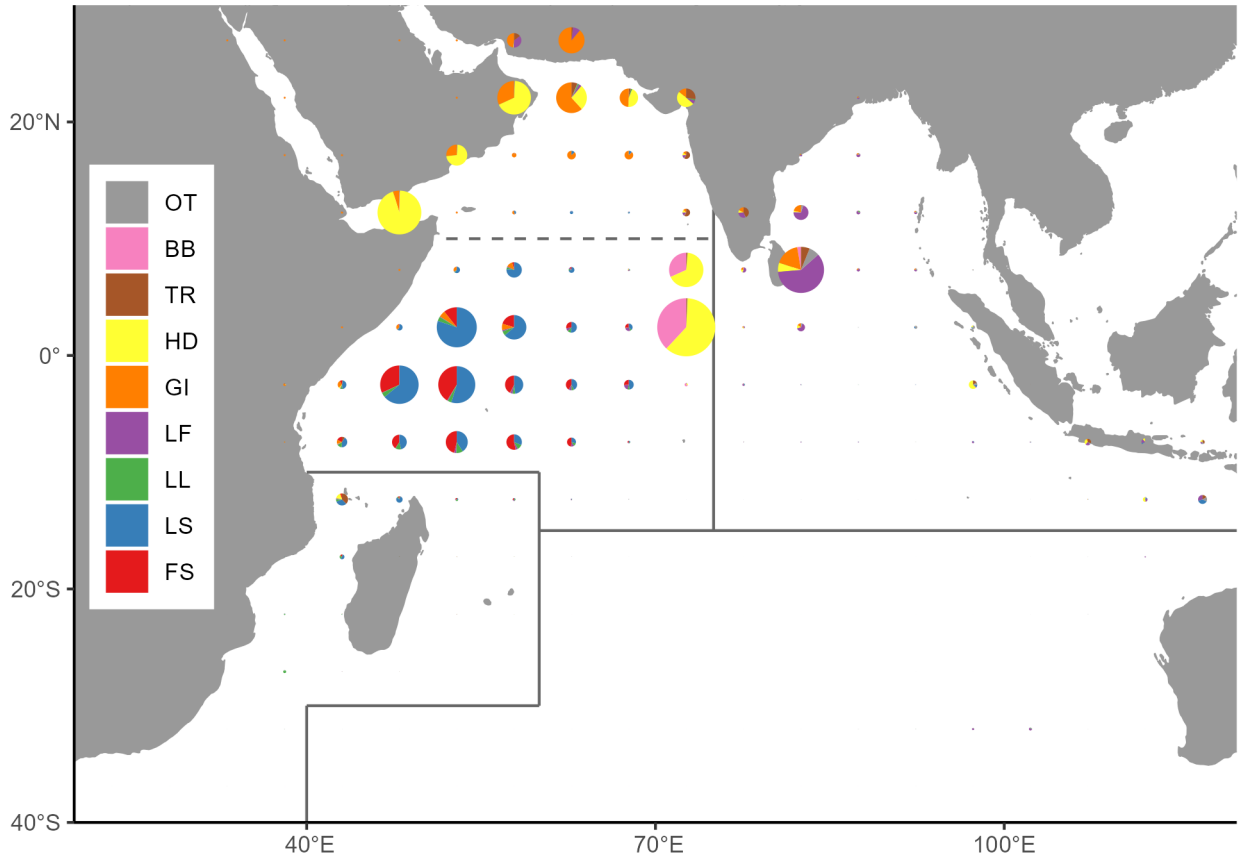


Figure 2: Spatial distribution of yellowfin catches per fishery group. The pie radius represents the aggregated catch from 2010 to 2023. Fishery group codes are described in Table 1.

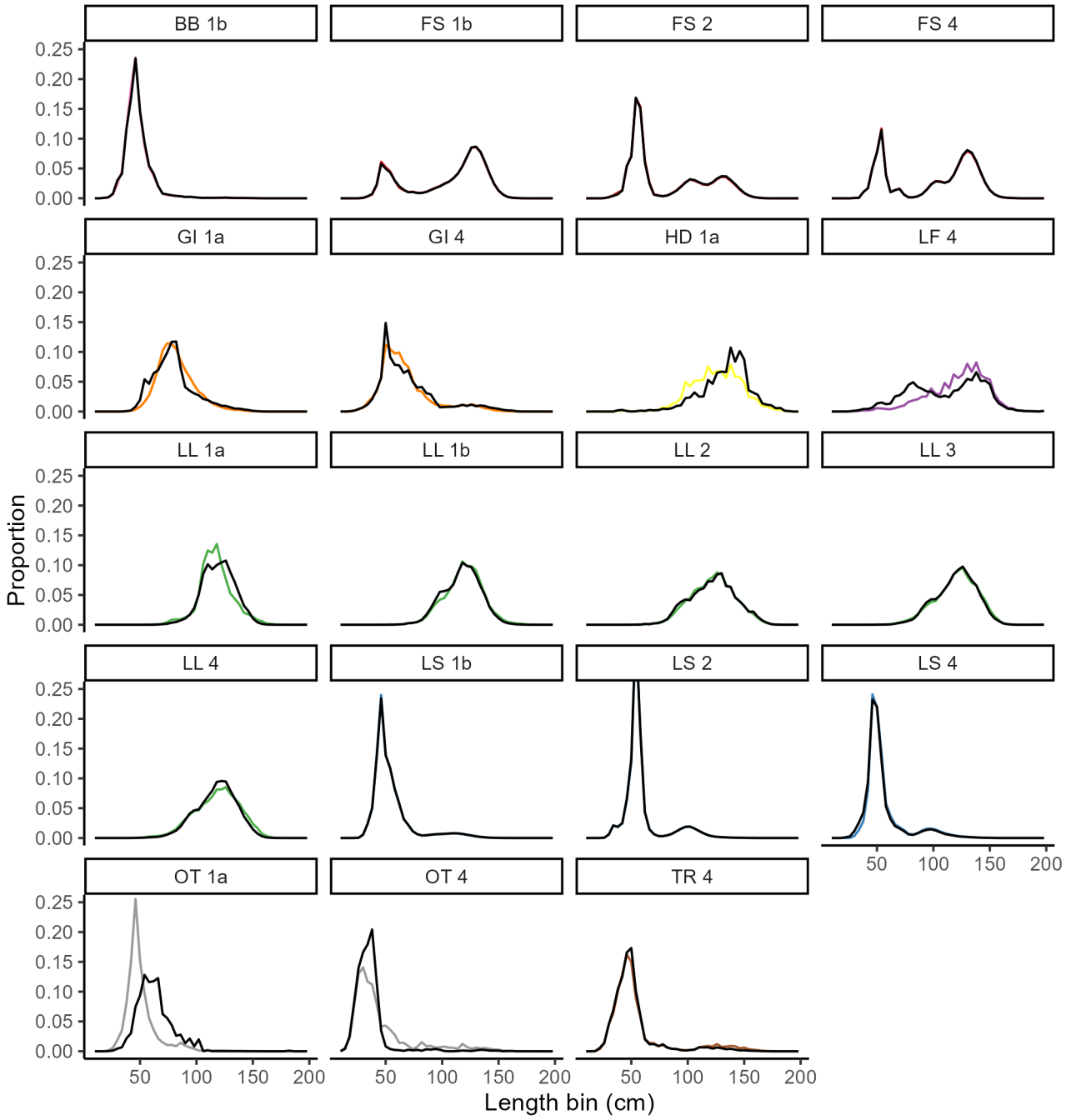


Figure 3: Size compositions per fishery included in the assessment model. Colored lines are size compositions obtained using simple aggregation while black lines used the catch-raised aggregation. Size compositions were aggregated over time.

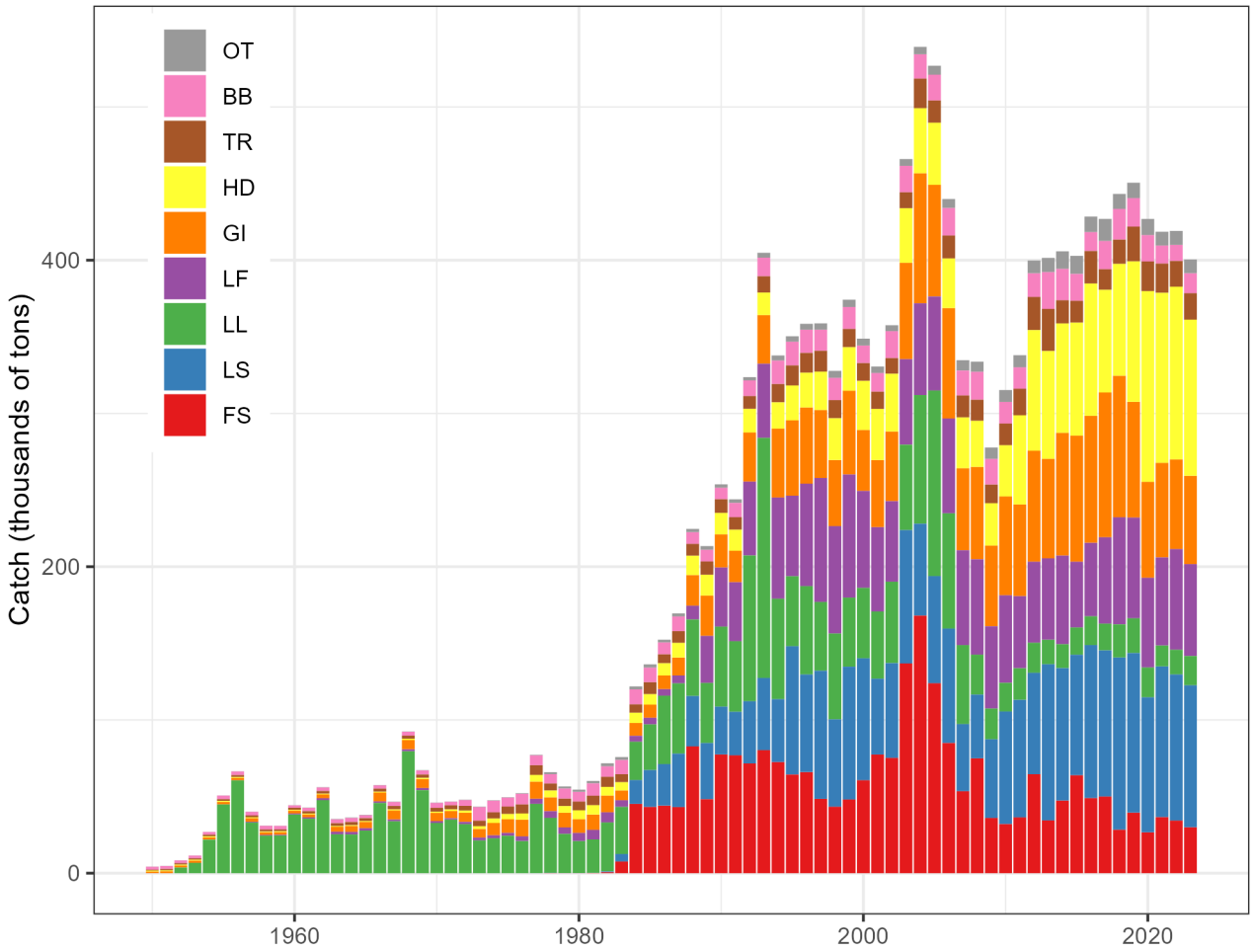


Figure 4: Total annual catch of yellowfin tuna by fishery group from 1950 to 2023. Fishery group codes are described in Table 1.

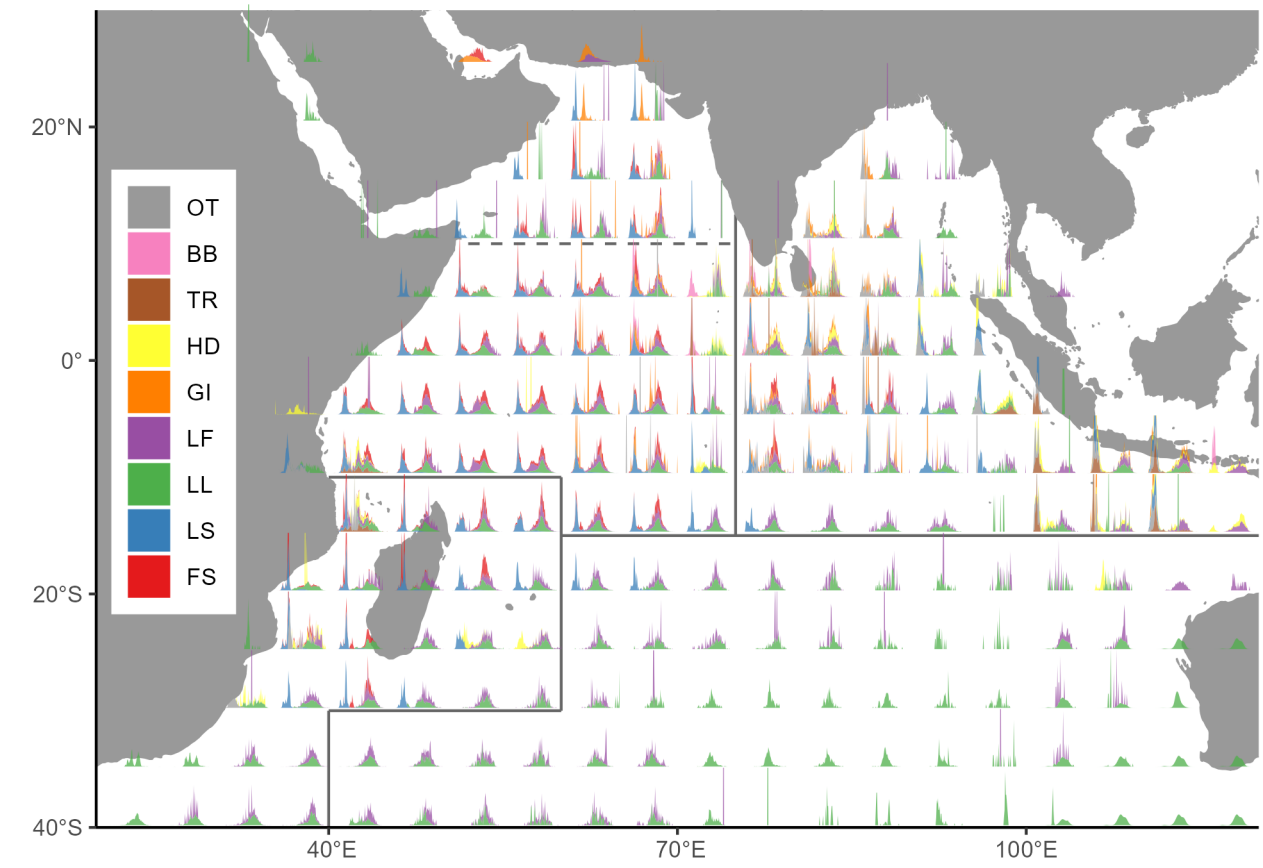


Figure 5: Spatial distribution of size compositions per fishery group. The size compositions were aggregated over time from 2010 to 2023. Fishery group codes are described in Table 1.

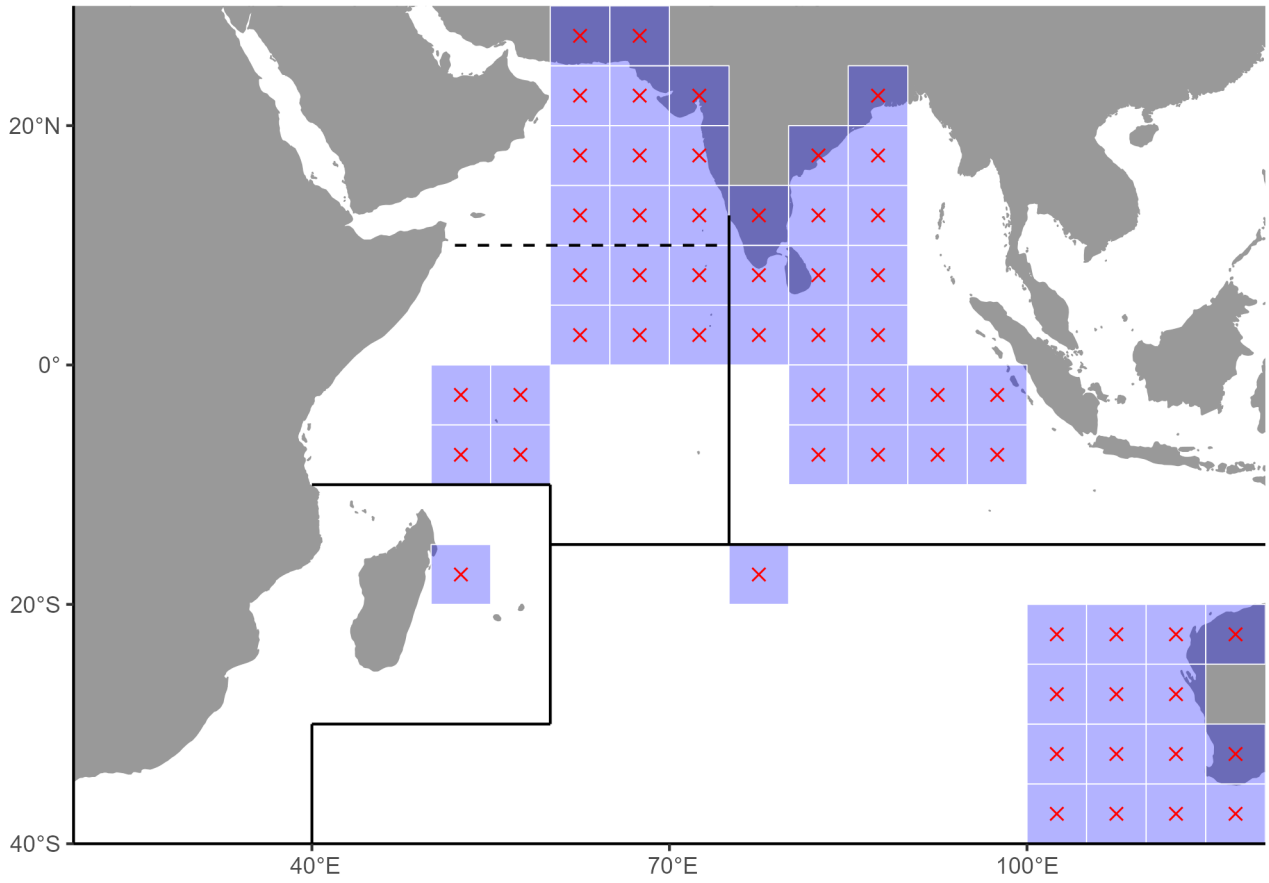


Figure 6: Example of grids in the cwp55 size dataset. The region assignment was done based on the grid centroid.

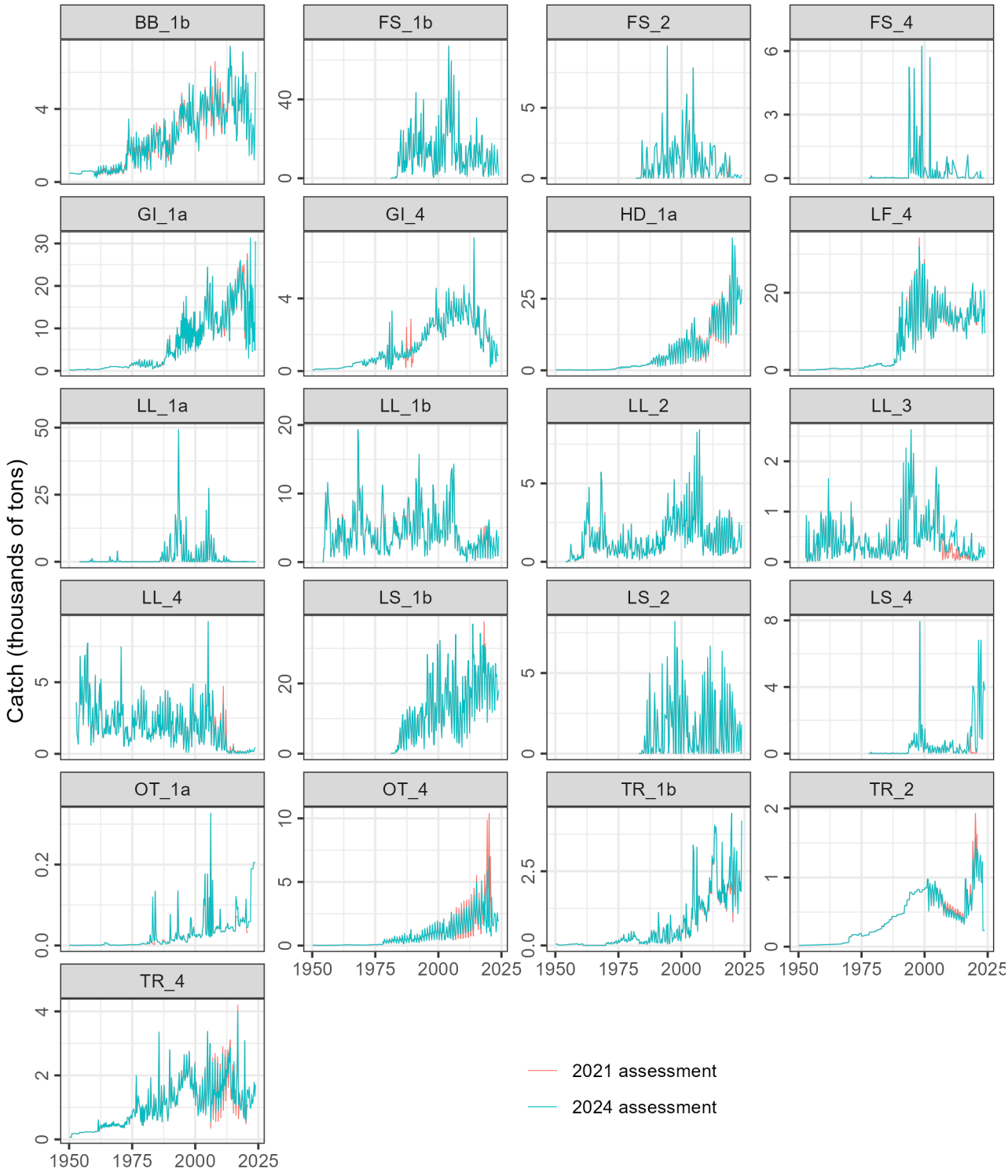


Figure 7: Comparison between the catch values used in the 2021 assessment and the current (2024) values.

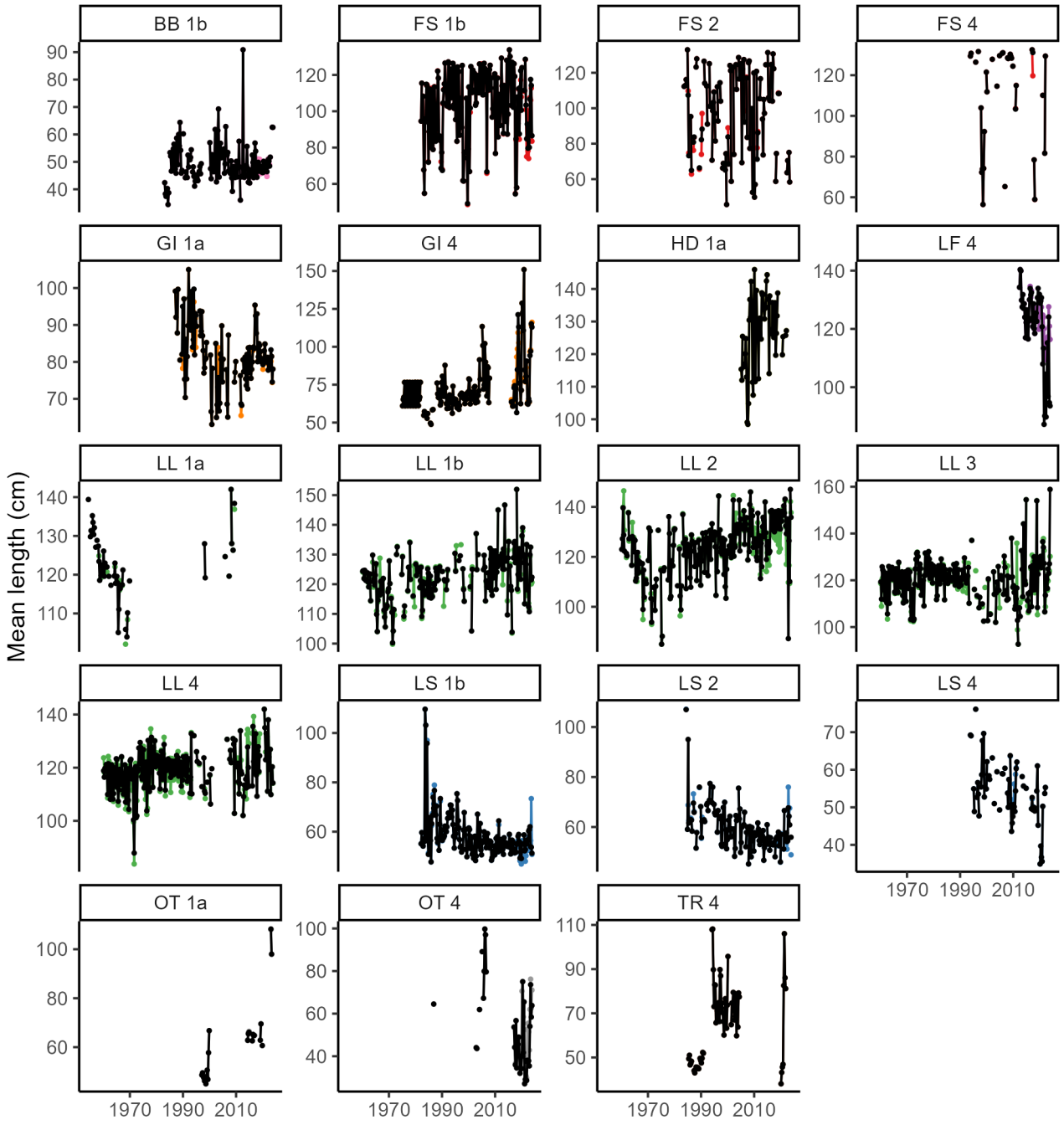


Figure 8: Mean length per quarter per fishery. Colored lines used the simple aggregation while black lines used the catch-raised aggregation.

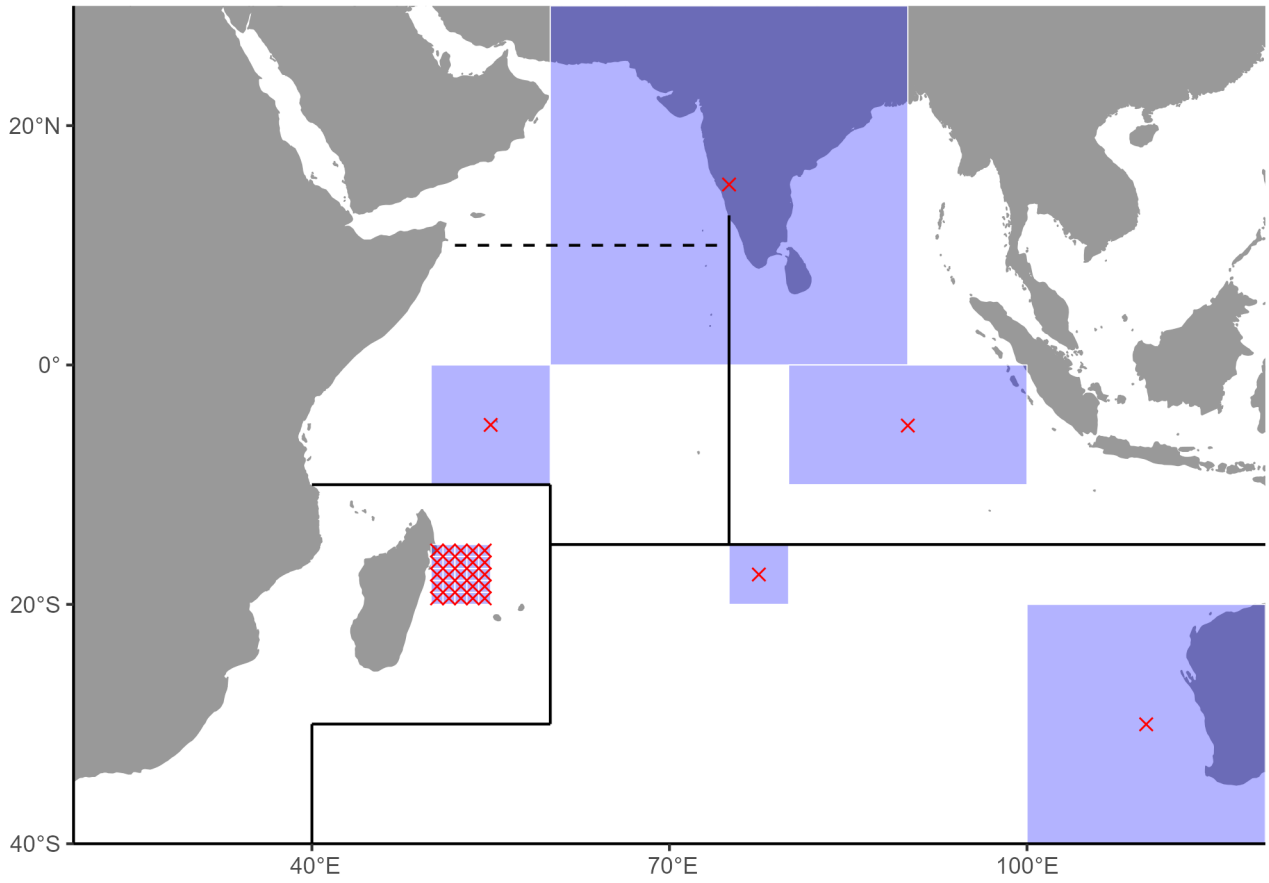


Figure 9: Example of grid categories in the original size dataset. The region assignment was done based on the grid centroid.

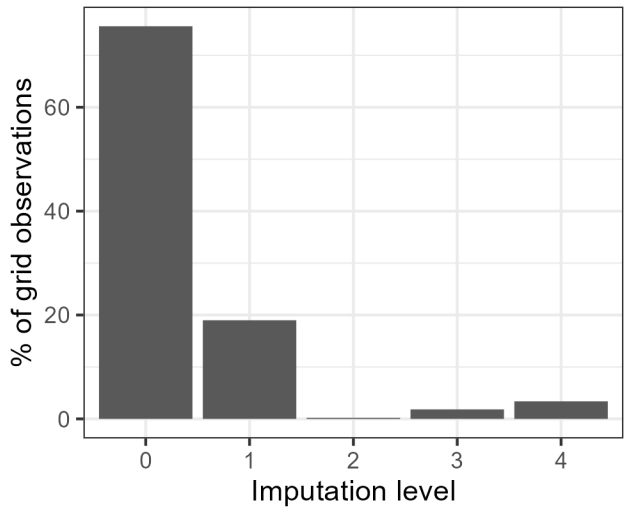


Figure 10: Percentage of size observations by imputation level. Level 0 means perfect match, so no imputation was required.

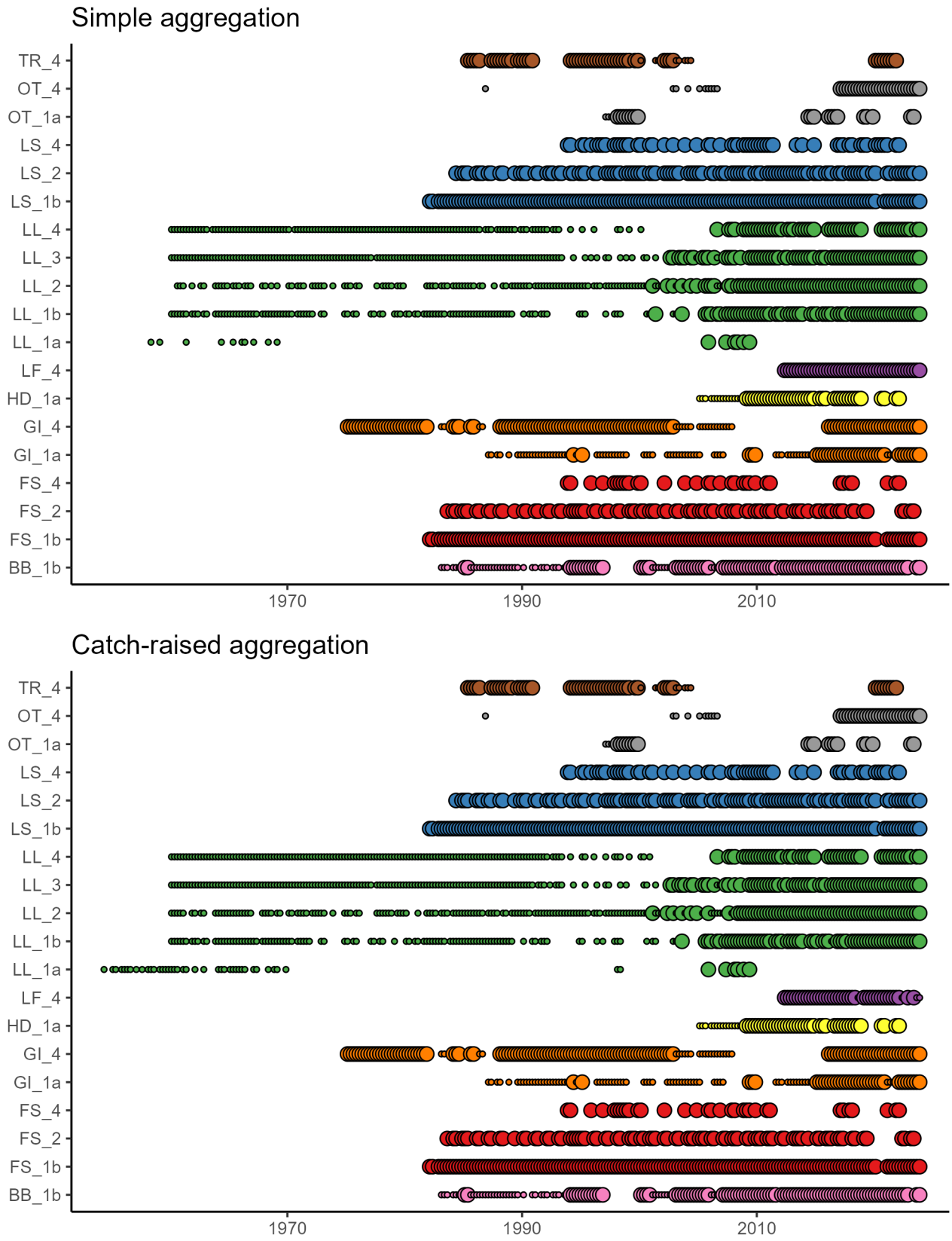


Figure 11: The availability of size composition data from each fishery by quarter. The size of the bubble indicates the input sample size calculated from reporting quality scores.

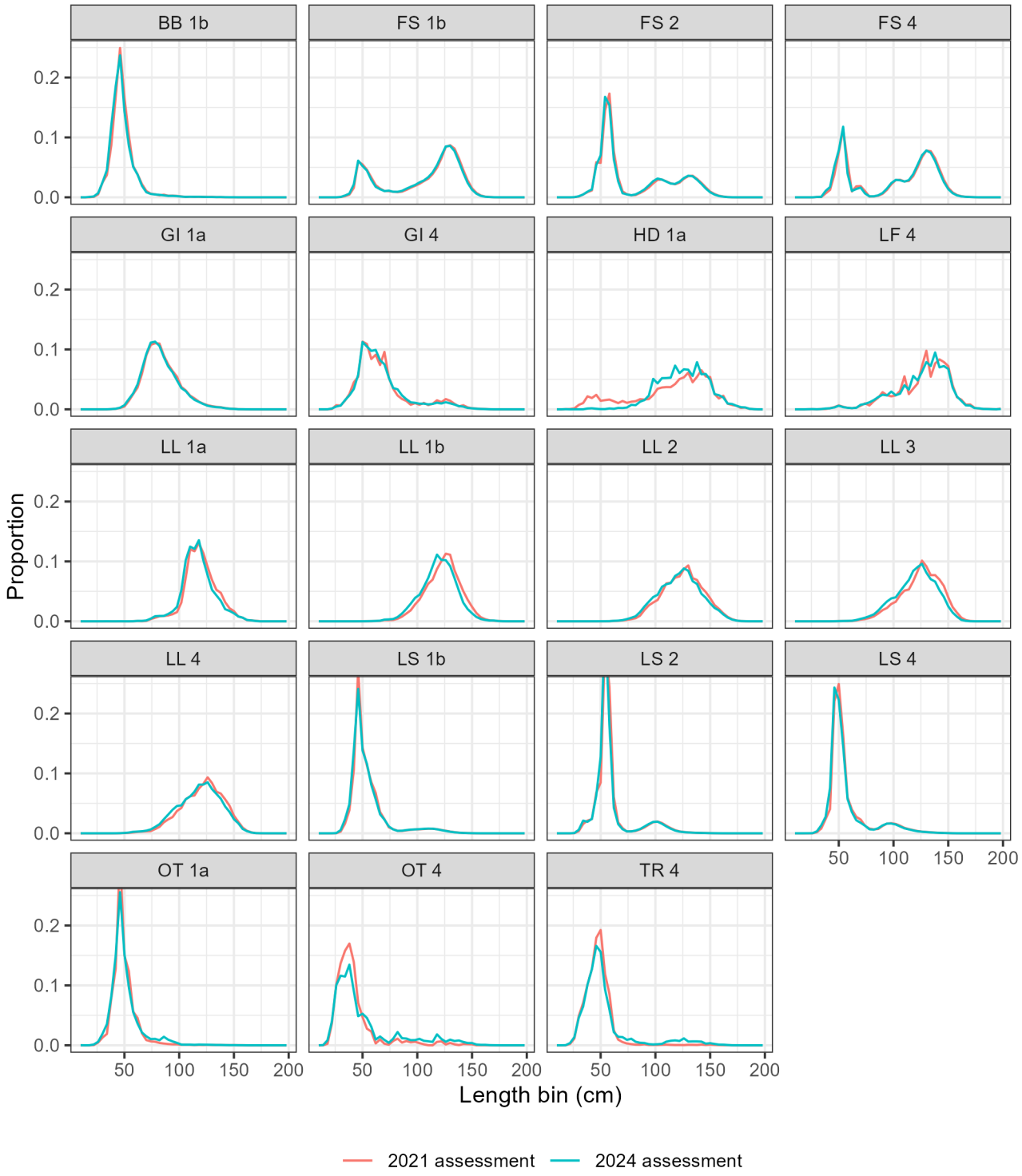


Figure 12: Comparison between the aggregated size compositions per fishery used in the 2021 and the current (2024) assessment. Both used the simple aggregation approach.

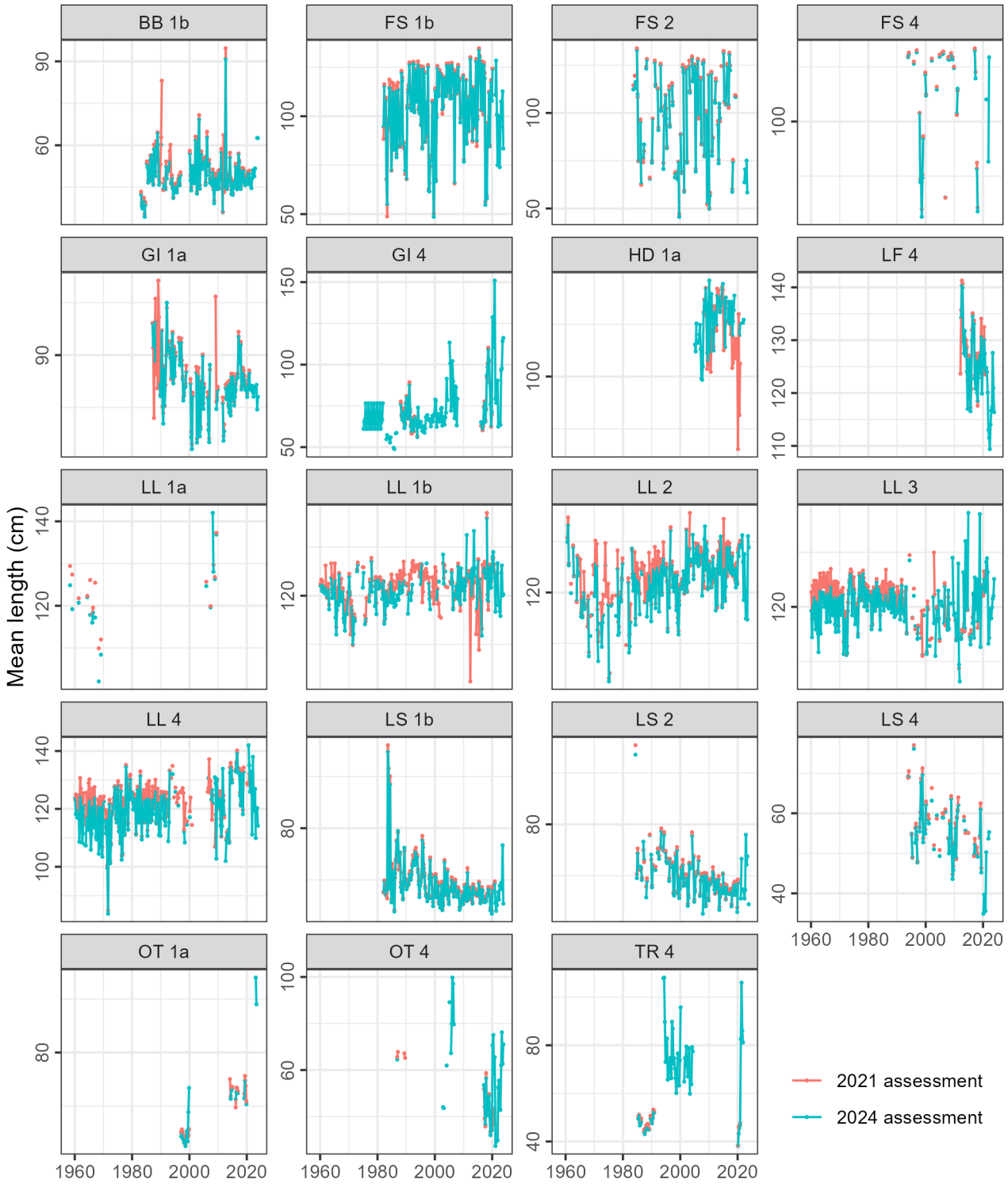


Figure 13: Comparison between the mean length per quarter per fishery group used in the 2021 and in the current (2024) assessment. Both used the simple aggregation approach.

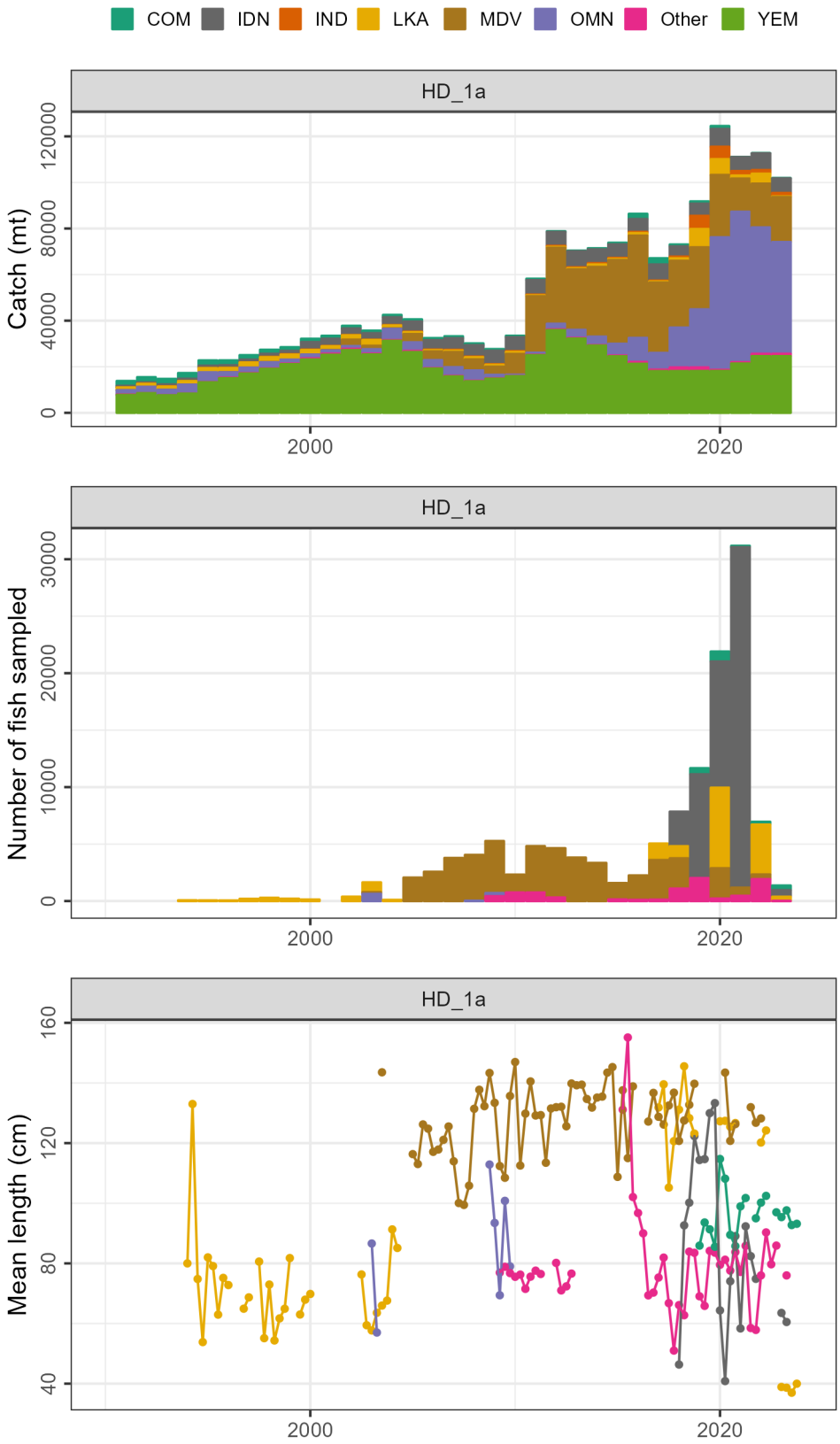


Figure 14: Annual catch (first row), number of fish sampled in the size data (second row), and mean length (third row) by CPC before filtering for the HD 1a fishery. Only data from 1990 is shown.

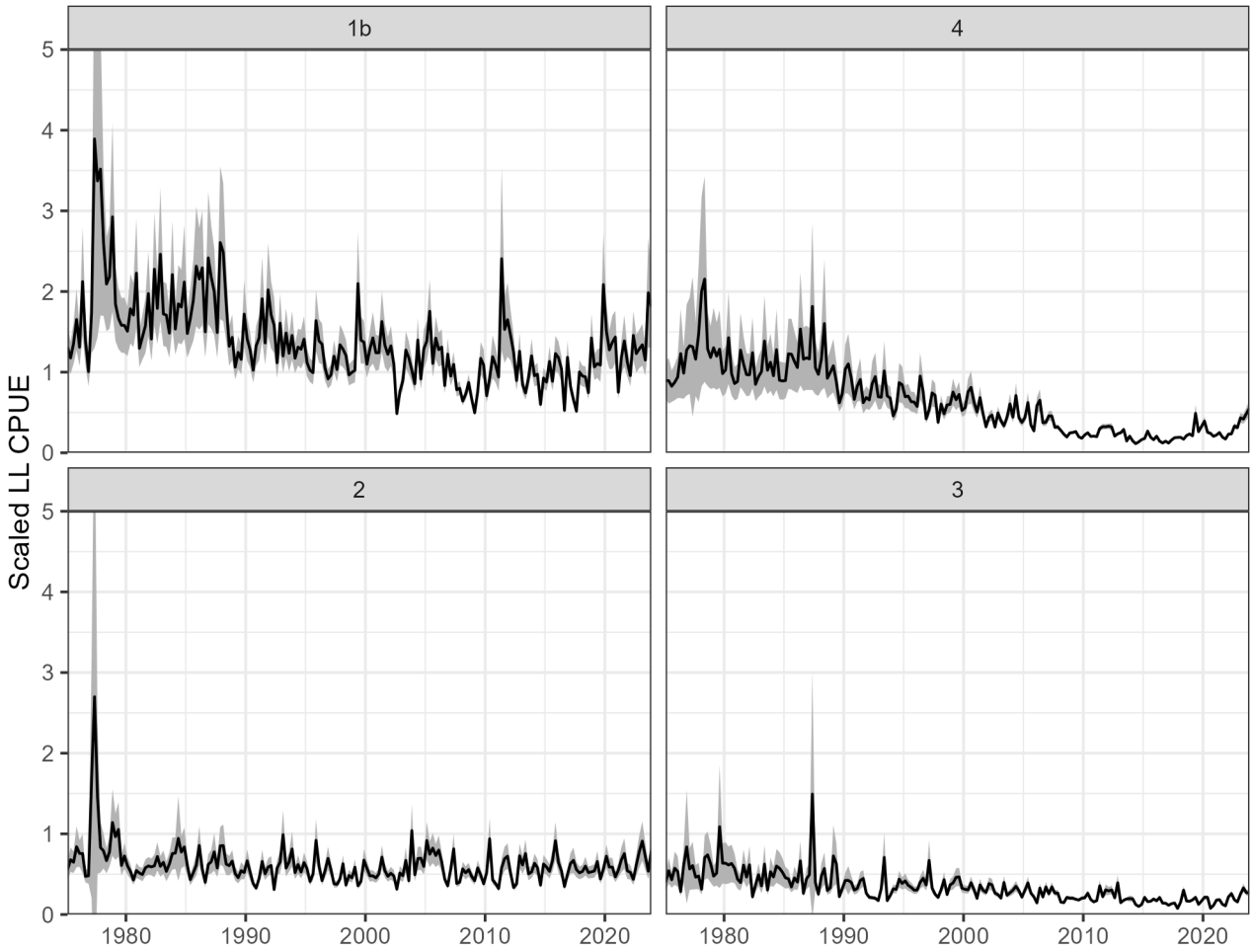


Figure 15: Scaled LL CPUE time serie per region. The shaded area represents the 95% confidence interval.

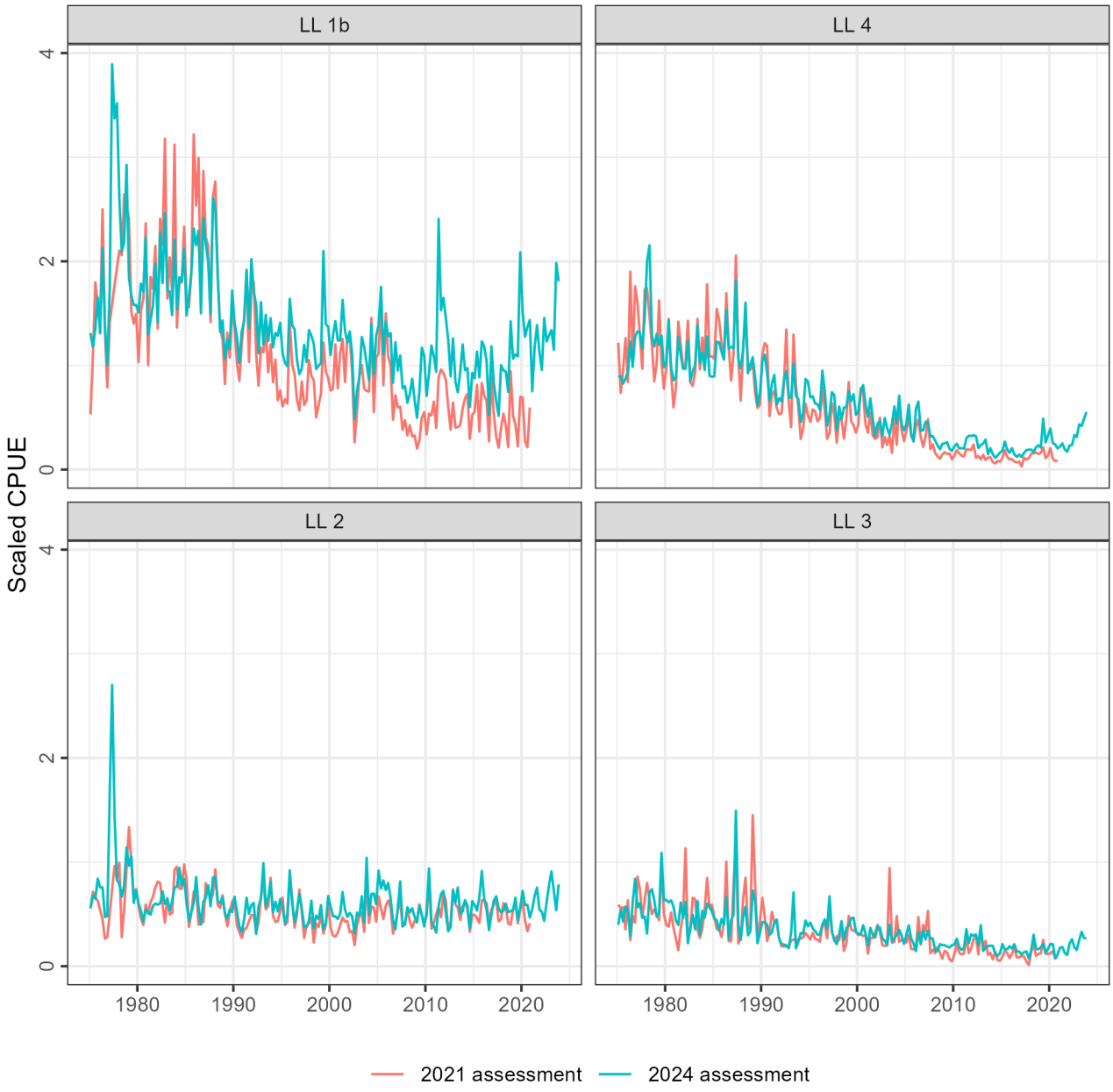


Figure 16: Comparison between the scaled LL CPUE values used in the 2021 assessment and the current (2024) values per region.

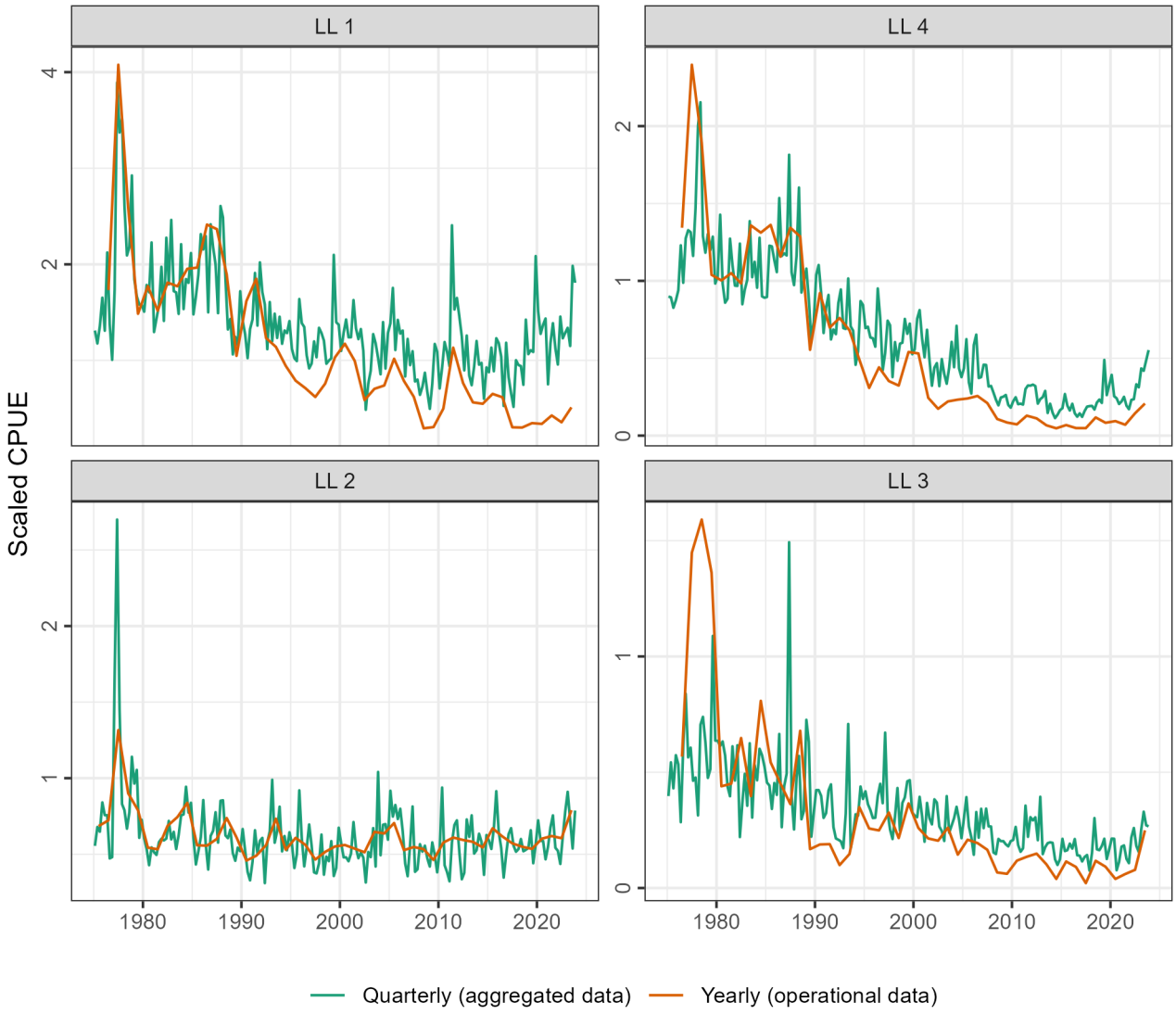


Figure 17: Comparison between the scaled LL CPUE values using aggregated data (quarterly time step) or operational data (yearly time step) for standardization.

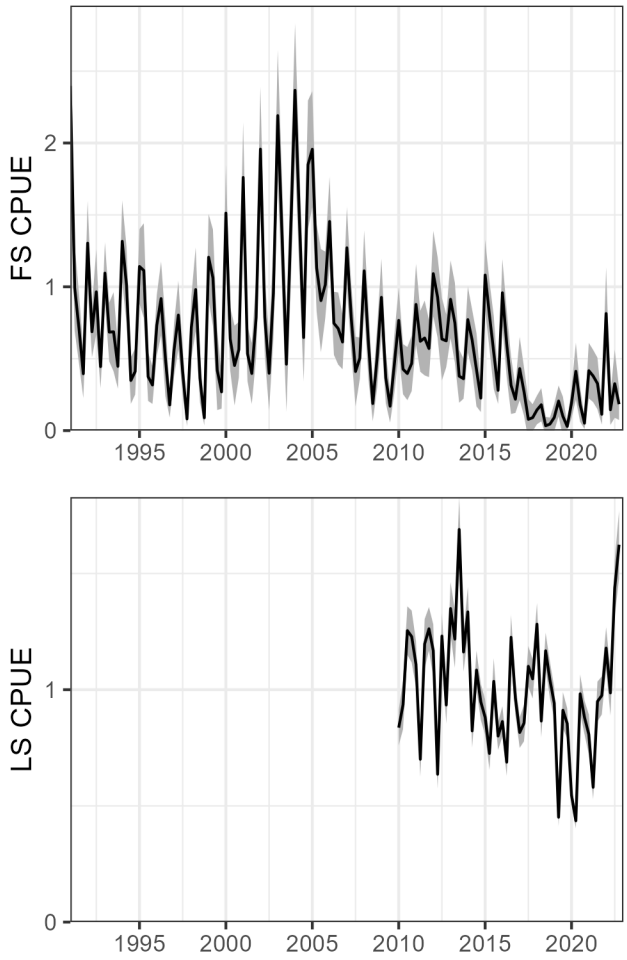


Figure 18: Purse seine free school (FS) and log school (LS) CPUE time series. Both series were used for region 1b in the stock assessment model. The shaded area represents the 95% confidence interval.

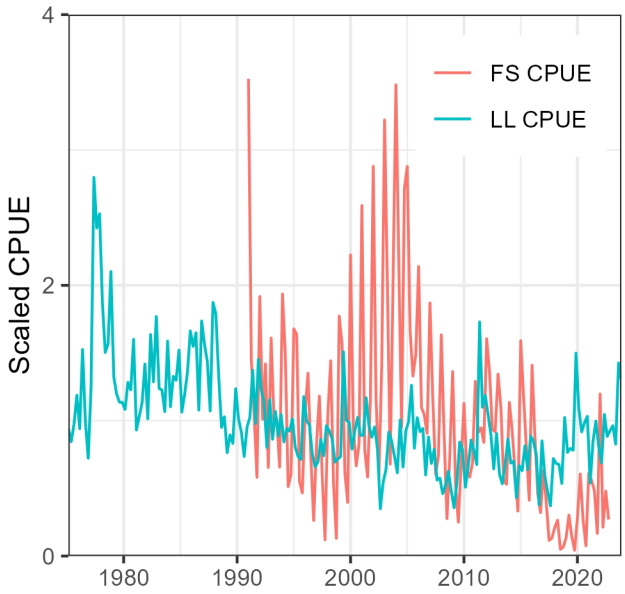


Figure 19: Time series of FS and LL CPUE in region 1b. Both series were rescaled to a mean of 1 for comparison.

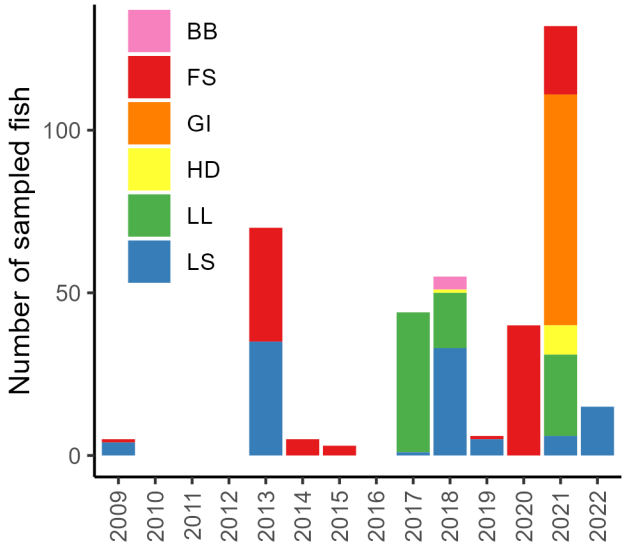


Figure 20: Number of sampled fish per year and fishery group in the age-length dataset from the GERUNDIO project.

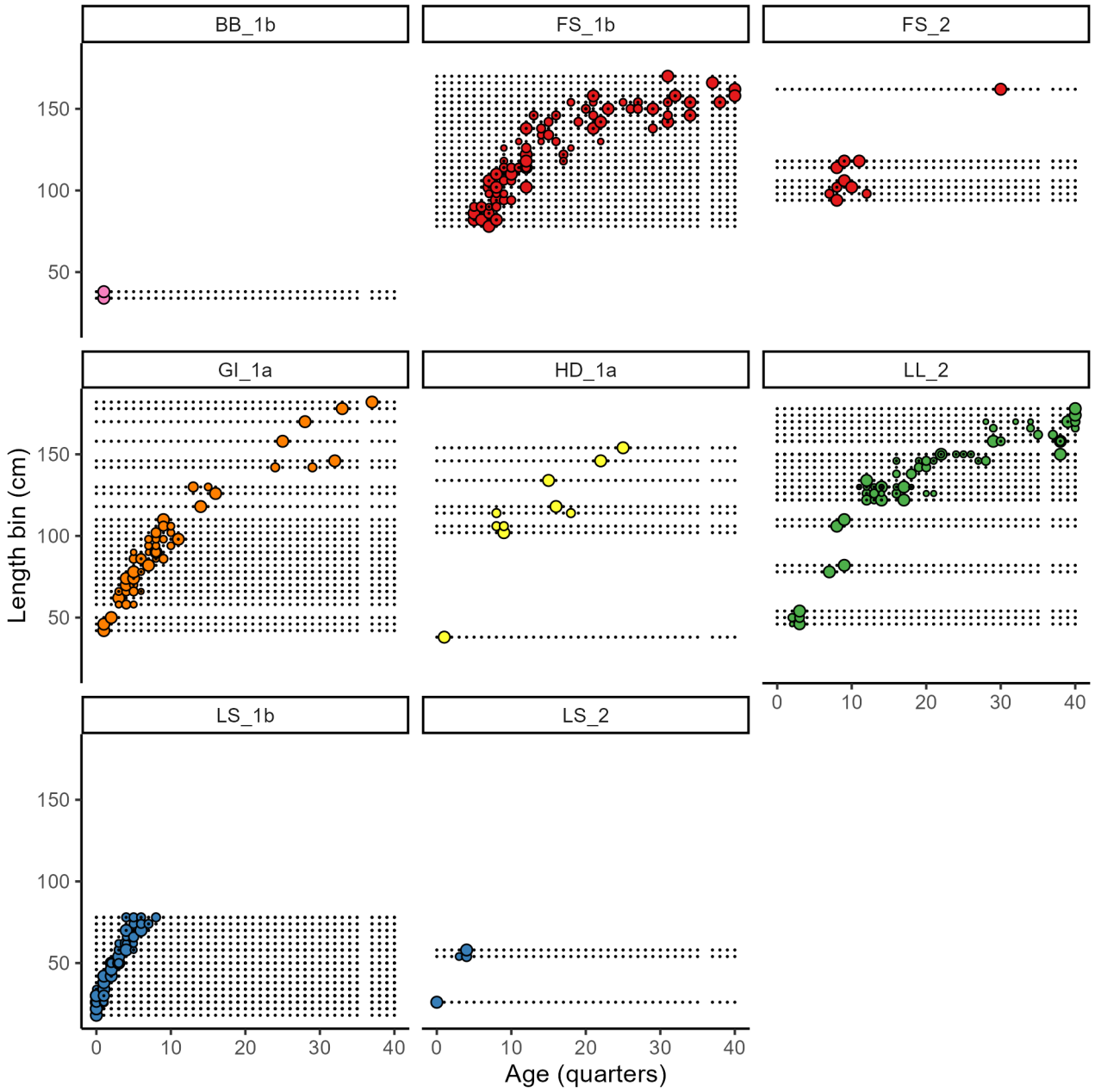


Figure 21: Conditional age-at-length (CAAL) data included in the assessment model. The bubble size represents the proportion of ages at a given length.

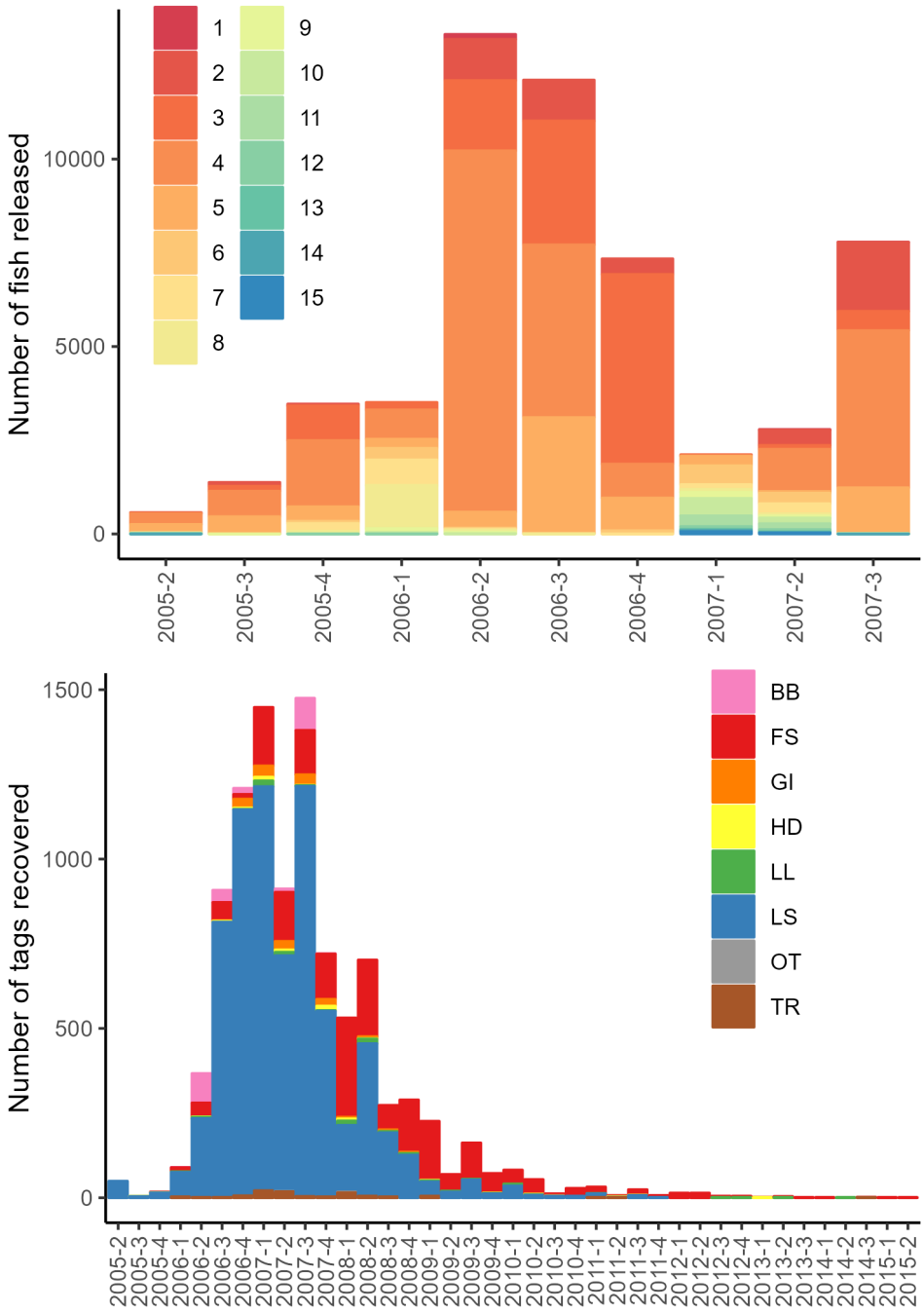


Figure 22: Number of tag releases by year-quarter and age class (in quarters, upper panel), and tag recoveries by year-quarter and fishery group (lower panel). Ages were assigned based on length. Purse seine tag recoveries are not corrected for reporting rate.

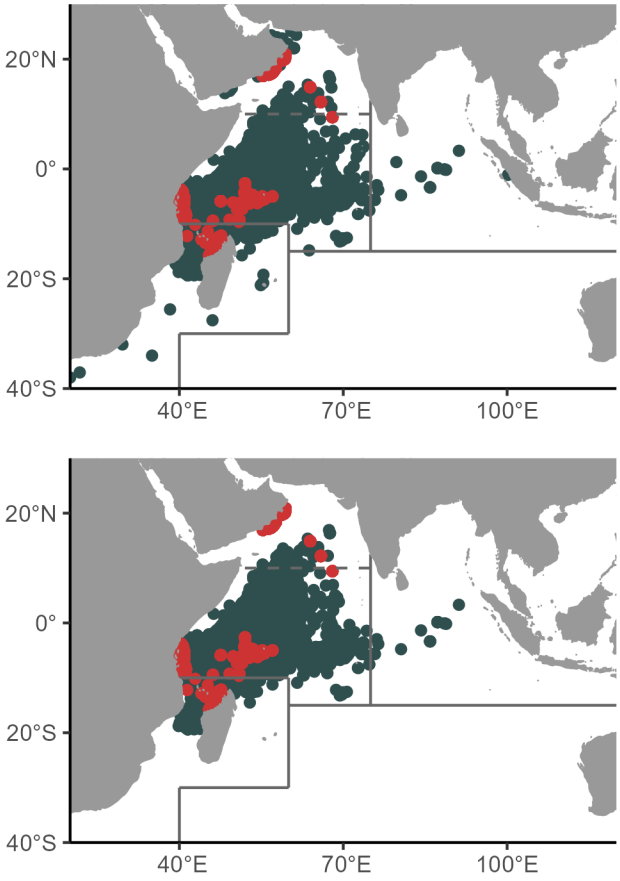


Figure 23: Locations of releases (red points) and recoveries (gray points) reported in the yellowfin tuna RTTO-IO tag Program. The upper and lower panel shows the recoveries by all fisheries and only by the purse seine fishery, respectively. The limits of regions in the four-area configuration are also shown.

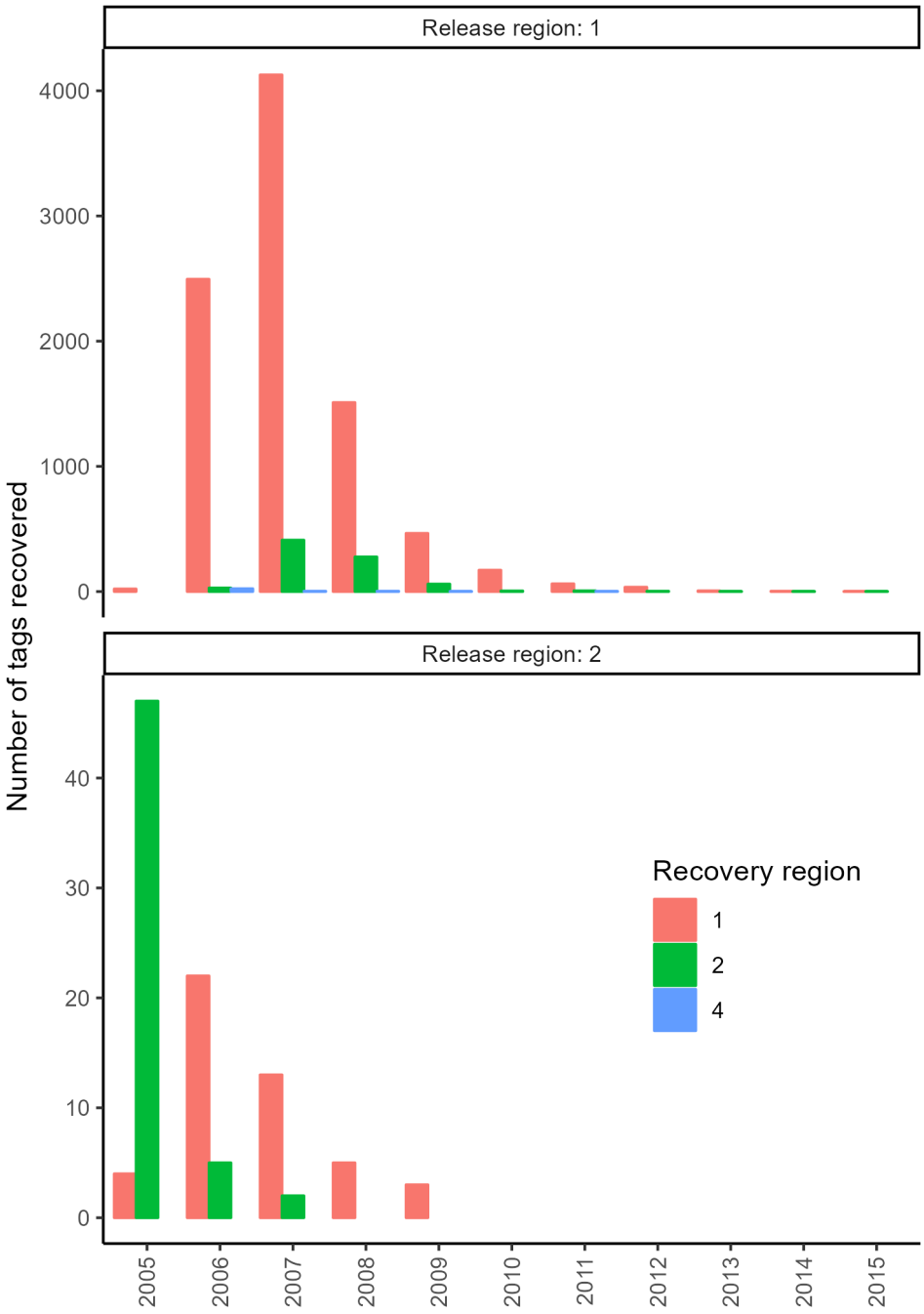


Figure 24: Tag recoveries by year of recovery, region of release, and region of recovery. Regions are defined by the four-area model configuration (see Figure 1).

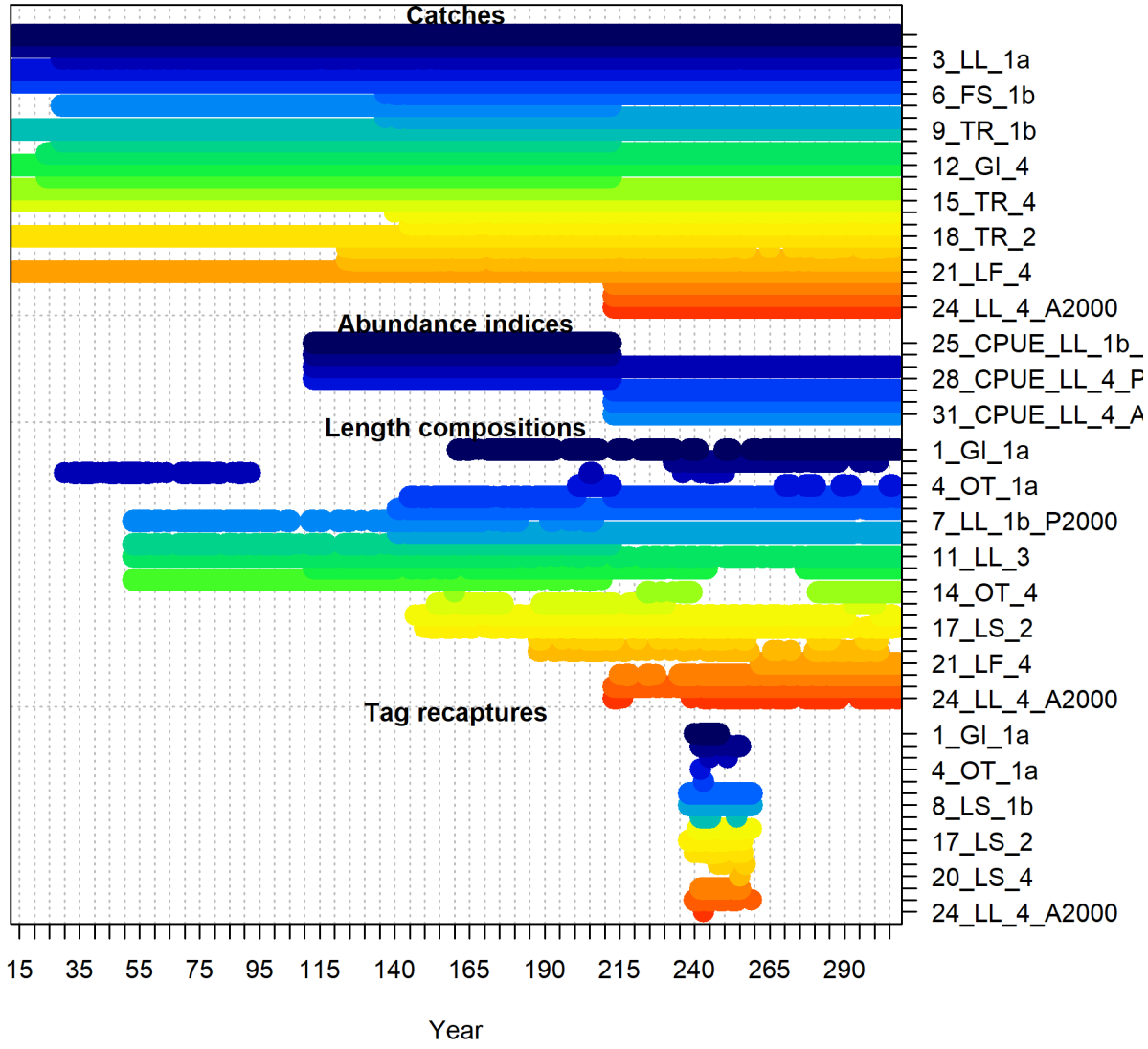


Figure 25: Temporal coverage of data types included in the candidate reference models. The x-axis represents model years (i.e., quarters).

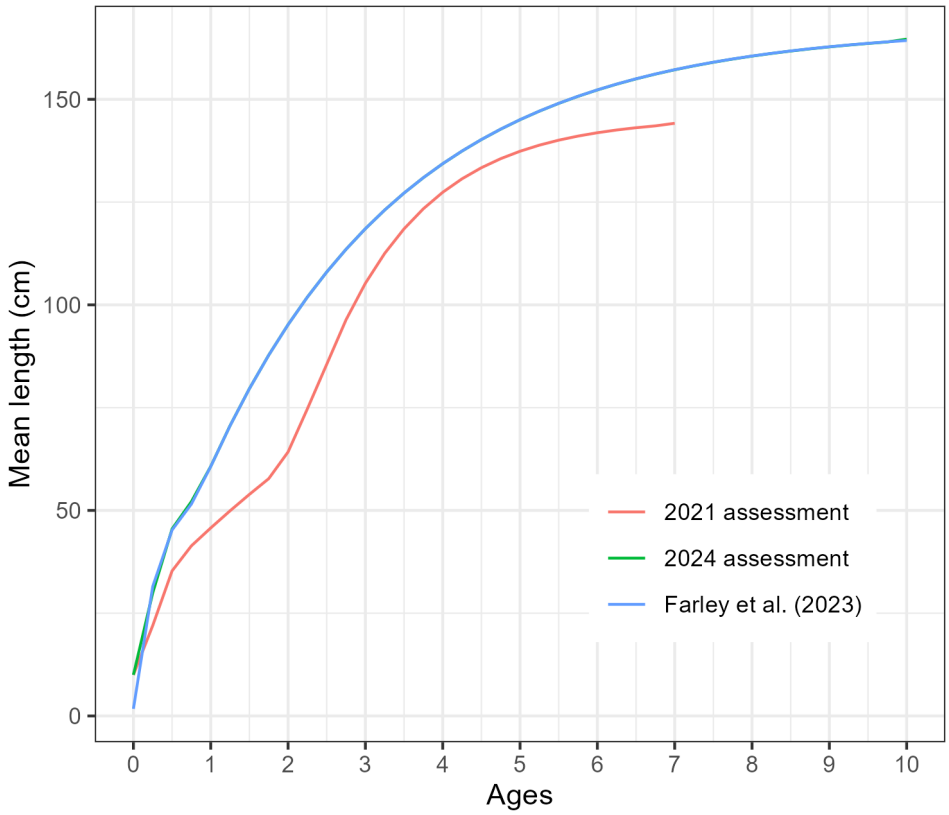


Figure 26: Mean length-at-age calculated using the two-stage growth function in Farley et al. (2023), and its approximation in SS3. The mean length-at-age from the growth parametrization used in the 2021 assessment is also shown.

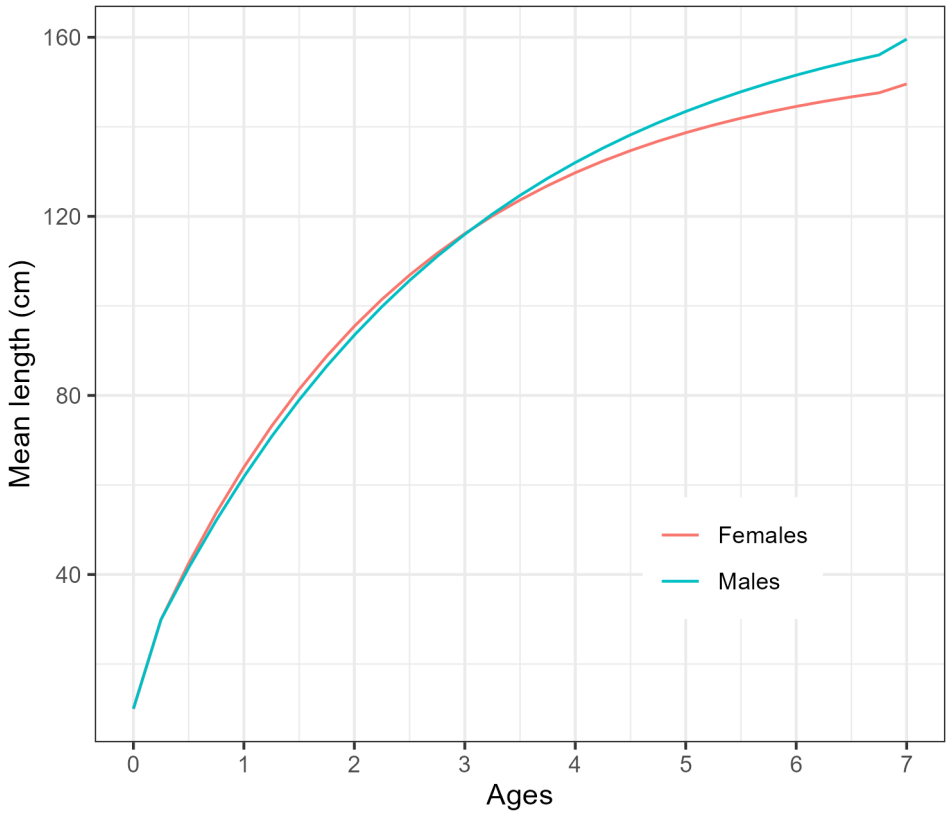


Figure 27: Mean length-at-age by sex approximated in SS3 based on the two-stage growth function in Farley et al. (2023).

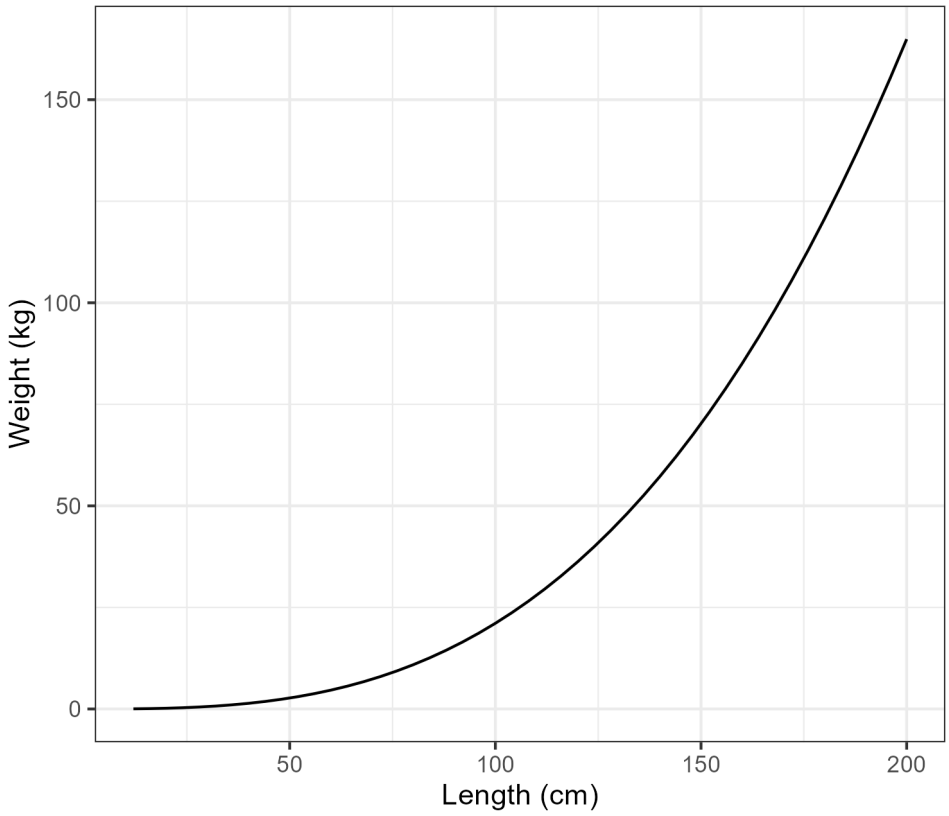


Figure 28: Length-weight relationship used in the current assessment.

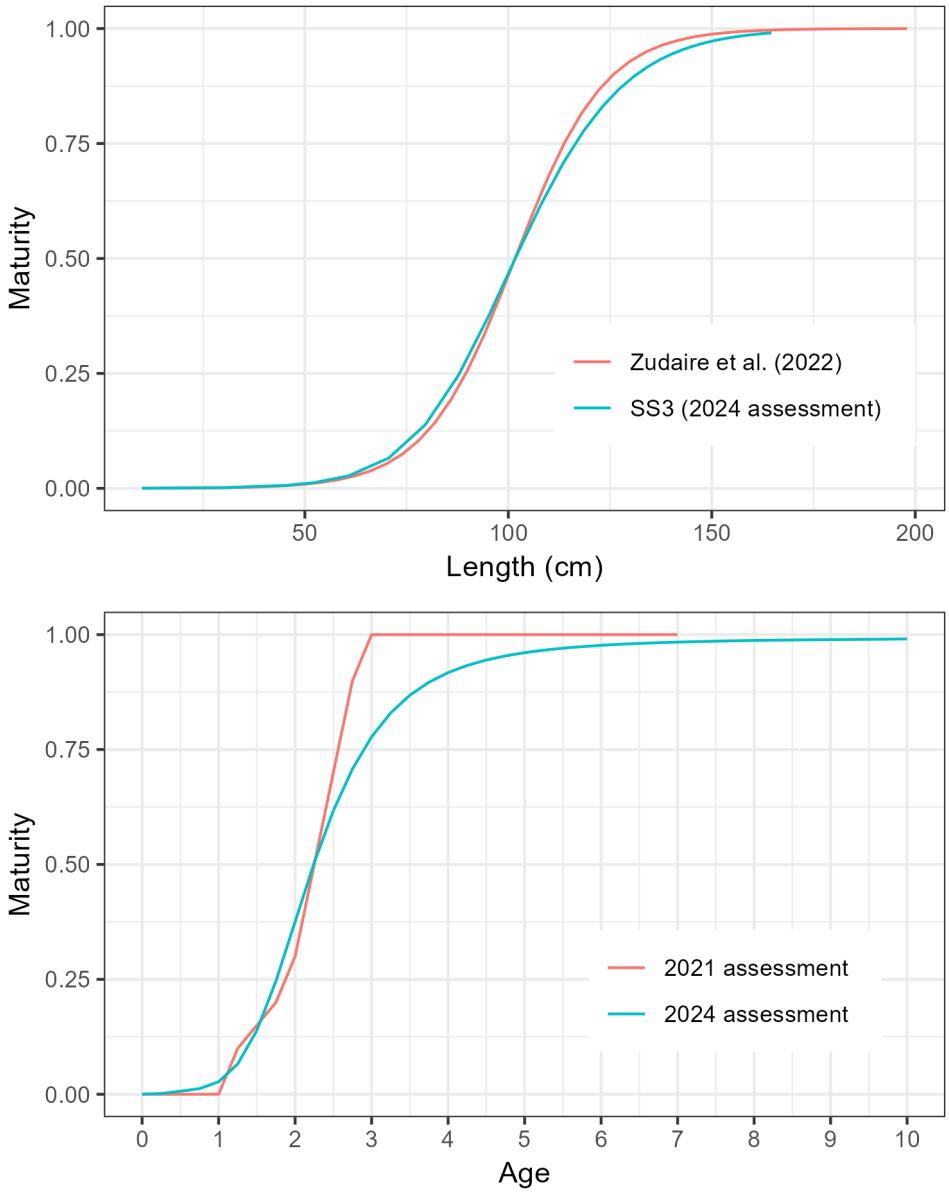


Figure 29: Upper panel: Maturity-at-length specified in the current assessment, based on Zudaire et al. (2022). Lower panel: Comparison between maturity-at-age specified in the 2021 and current assessment. For the current assessment, age-based maturity was converted from length-based maturity.

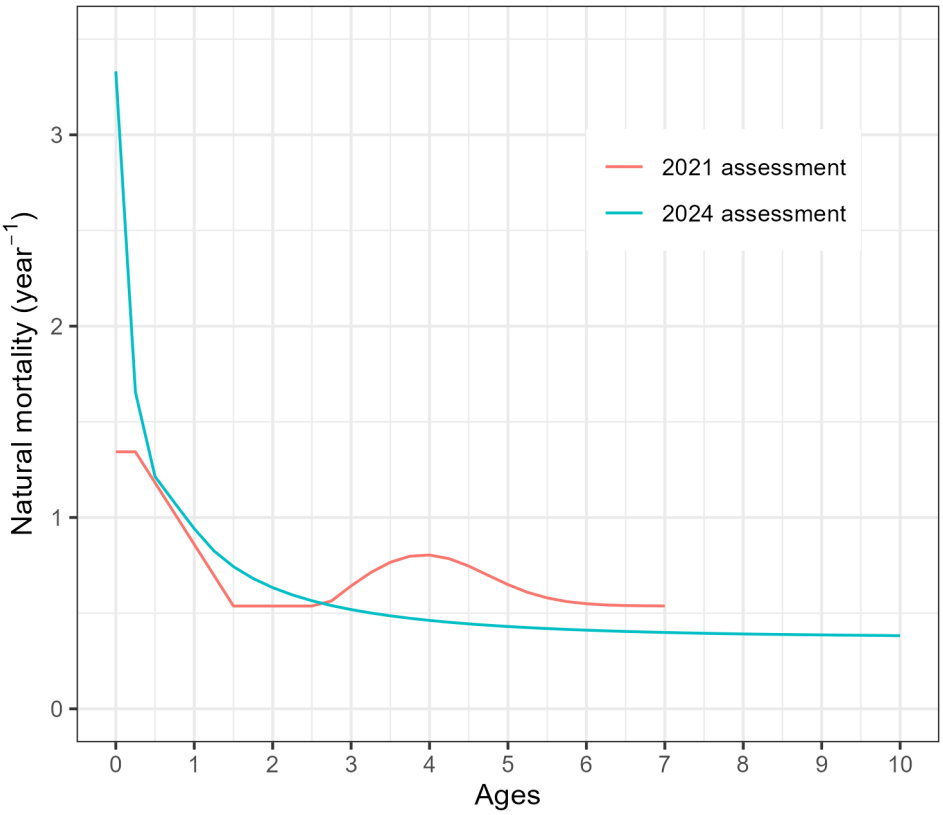


Figure 30: Natural mortality at age used in the 2021 and current assessment.

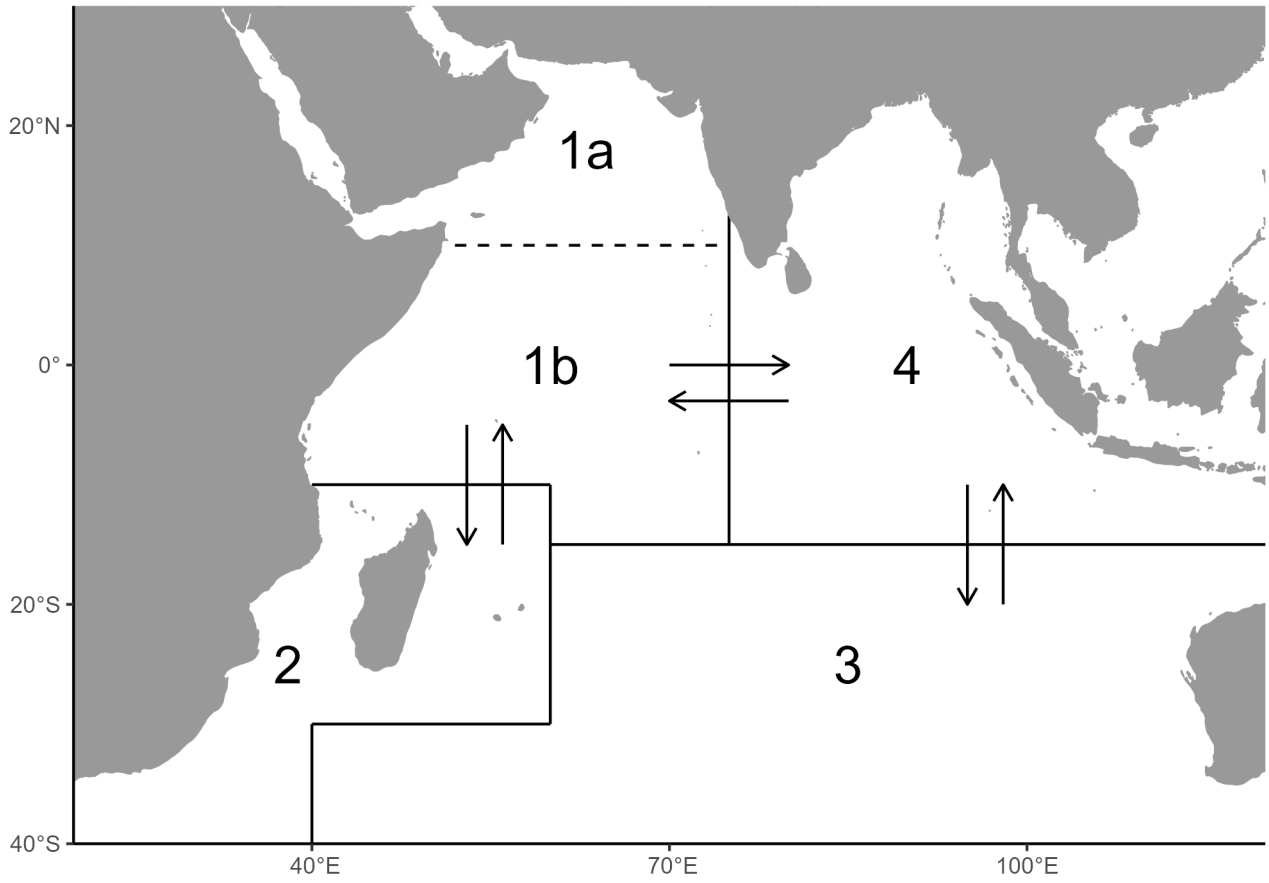


Figure 31: Movement parametrization in the current assessment model (four-areas configuration).

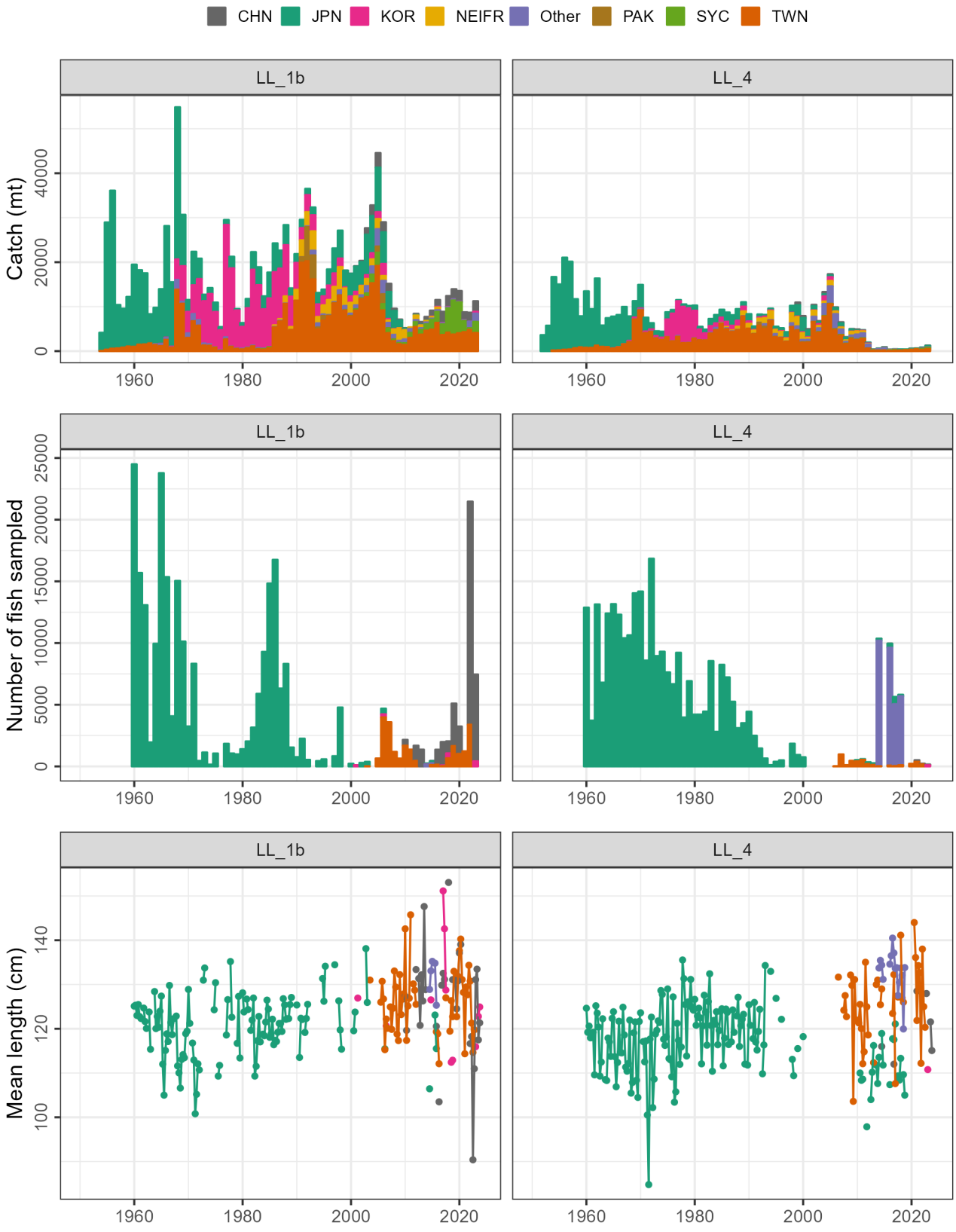


Figure 32: Annual catch (first row), number of fish sampled in the size data (second row), and mean length (third row) by CPC after filtering for the LL 1b and LL 4 fisheries.

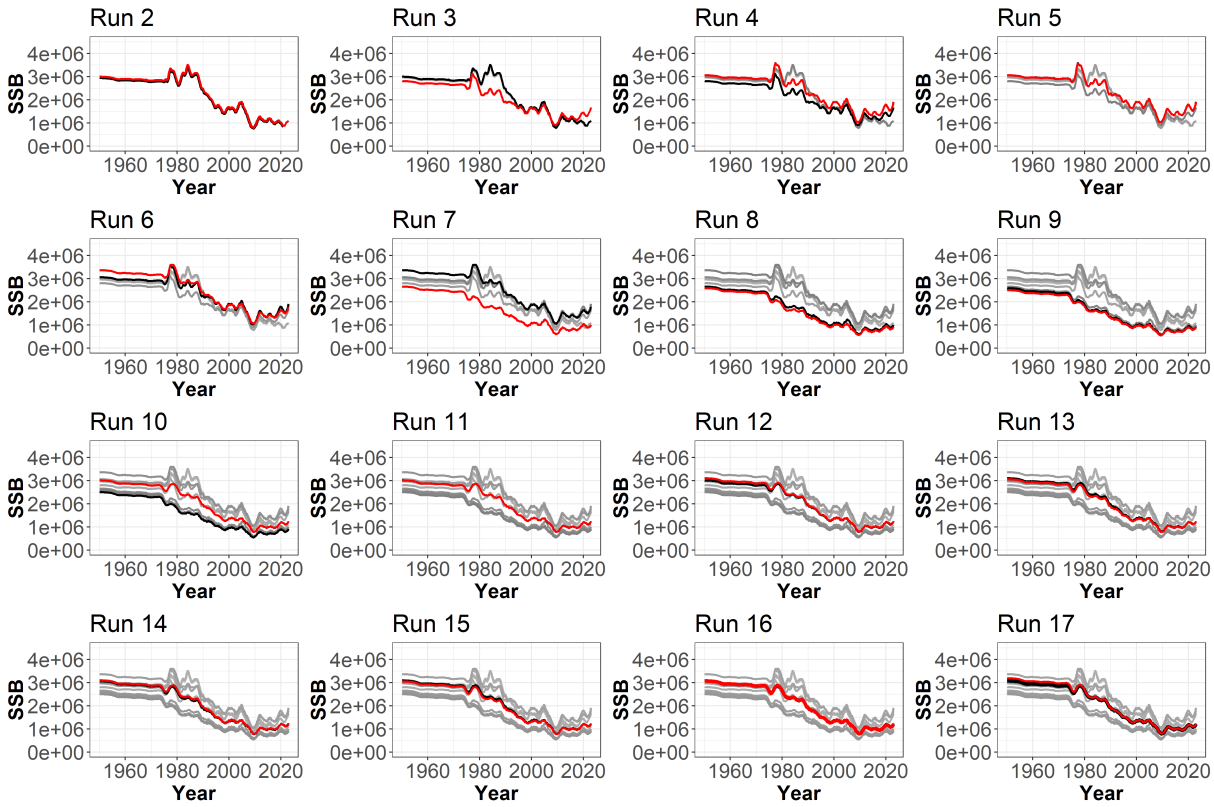


Figure 33: Plots of spawning stock biomass across iterative Stock Synthesis model runs. The red line shows the estimates from the current model run listed, the black line shows the estimates from the prior run, and the gray lines show the estimates from all previous runs in the stepwise build of the reference case model.

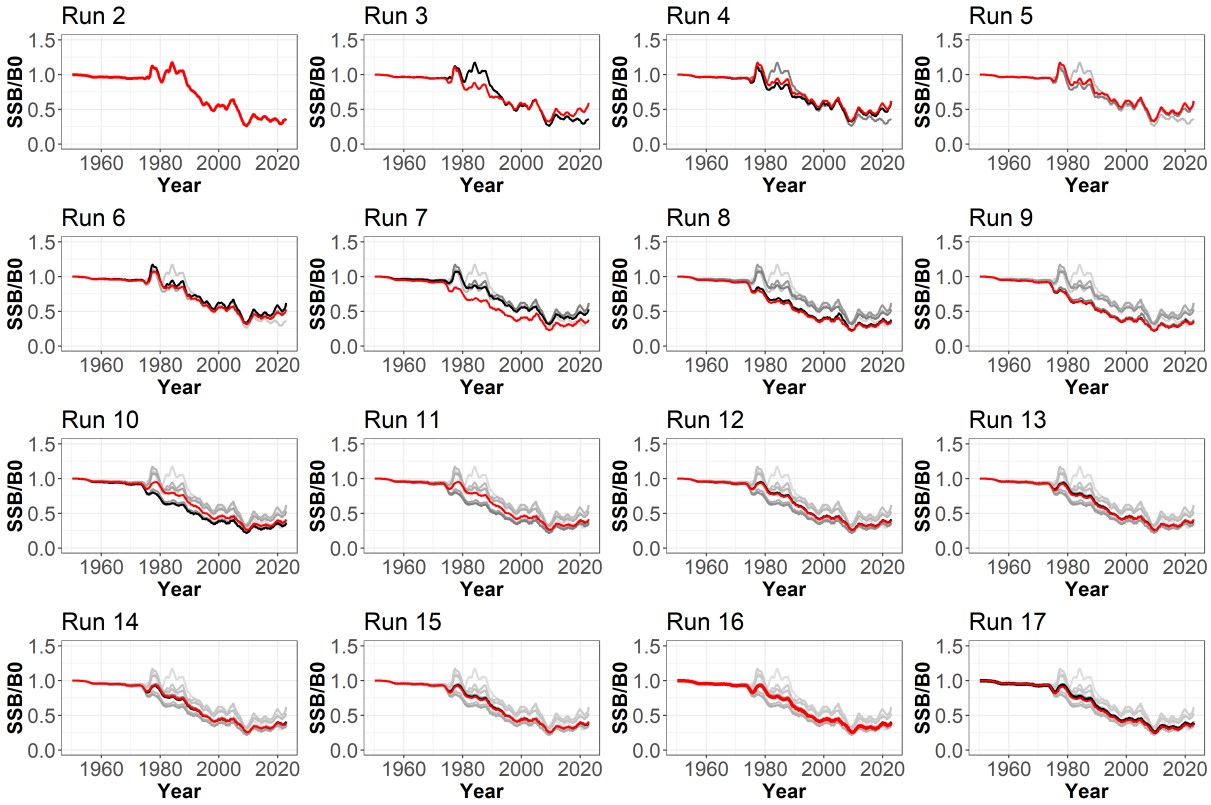


Figure 34: Spawning stock biomass relative to unfished biomass ($SSB/SSB_0$) across iterative model runs. The red line shows the estimates from the current model run listed, the black line shows the estimates from the prior run, and the gray lines show the estimates from all previous runs in the stepwise build of the reference case model.

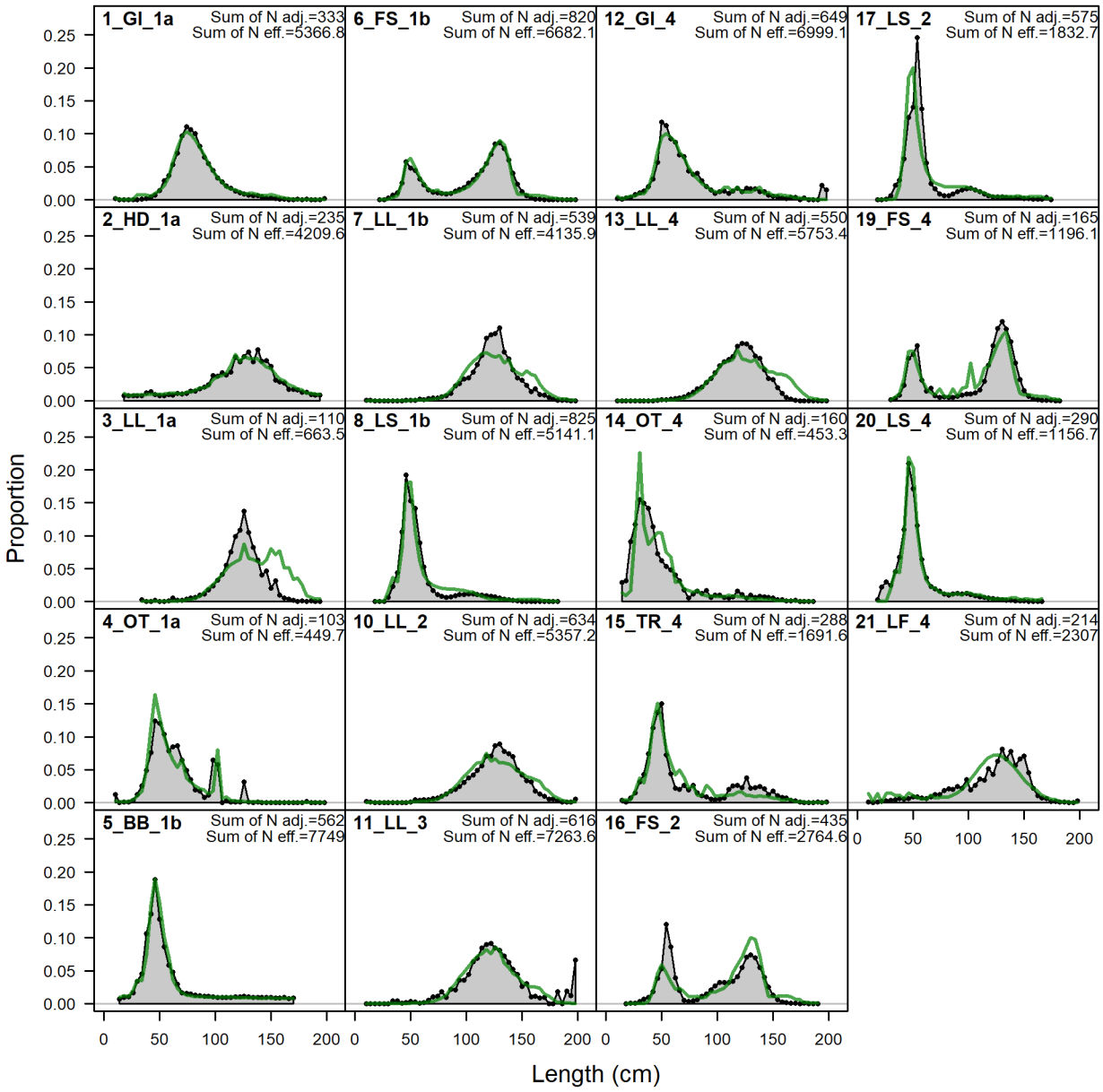


Figure 35: Observed (grey shadow) and predicted (green line) proportion of length compositions for each fishery aggregated over time for the RM1 model.

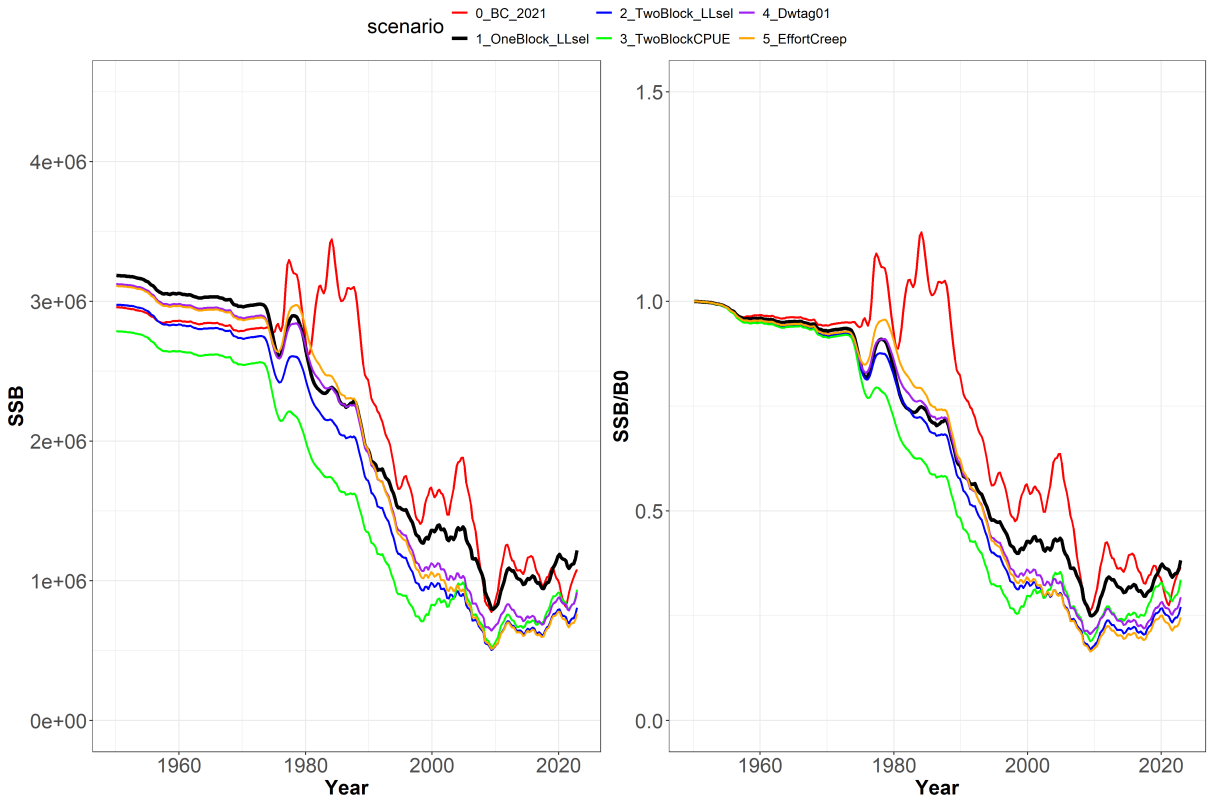


Figure 36: Comparison of $SSB/SSB_0$ trajectories from the base case from 2021 versus the alternative proposed reference models in 2024.

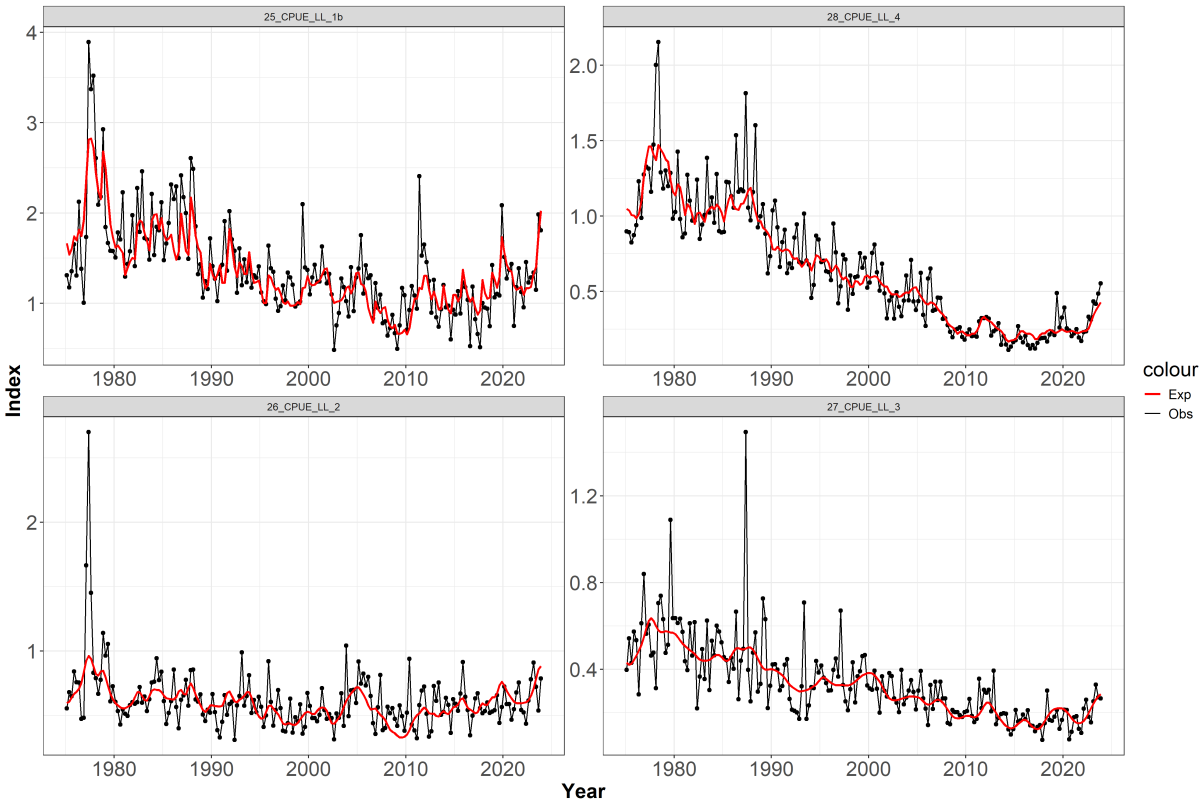


Figure 37: Fit to the regional longline CPUE indices, 1975–2023 from RM2.

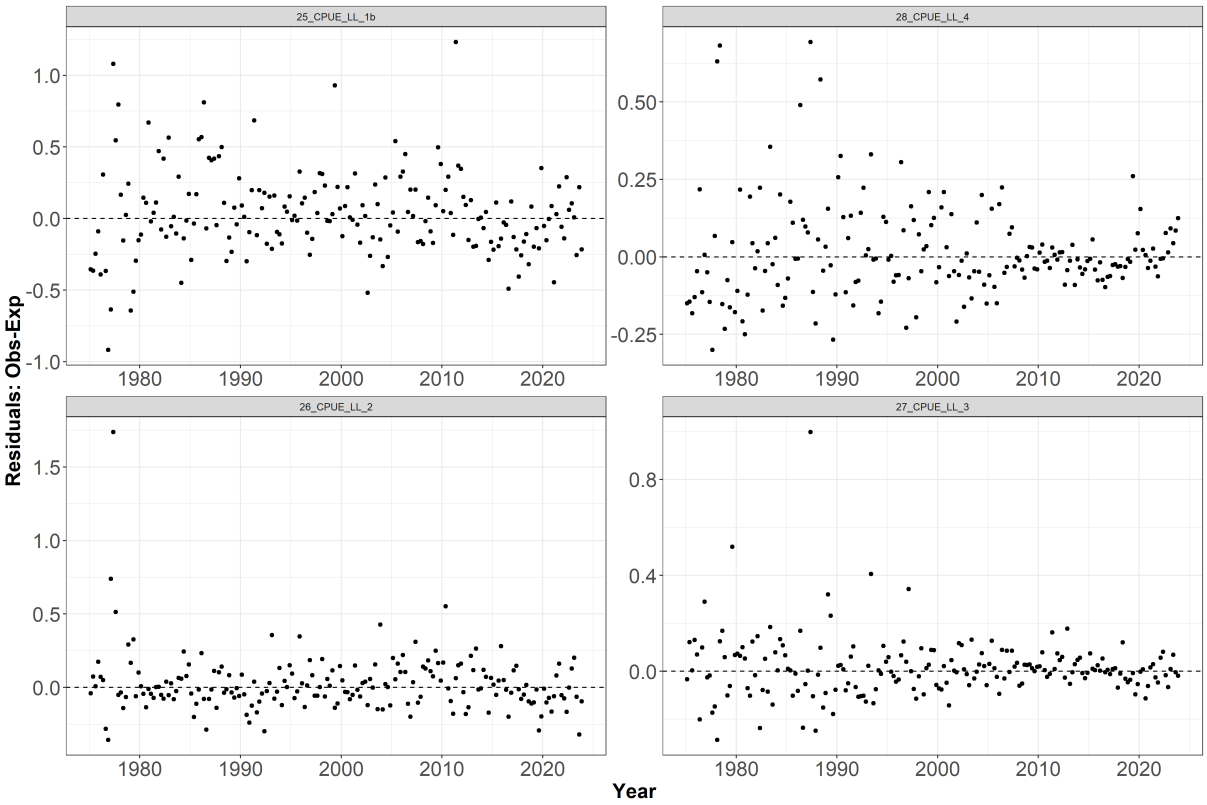


Figure 38: Standardised residuals from the fits to the CPUE indices from RM2.

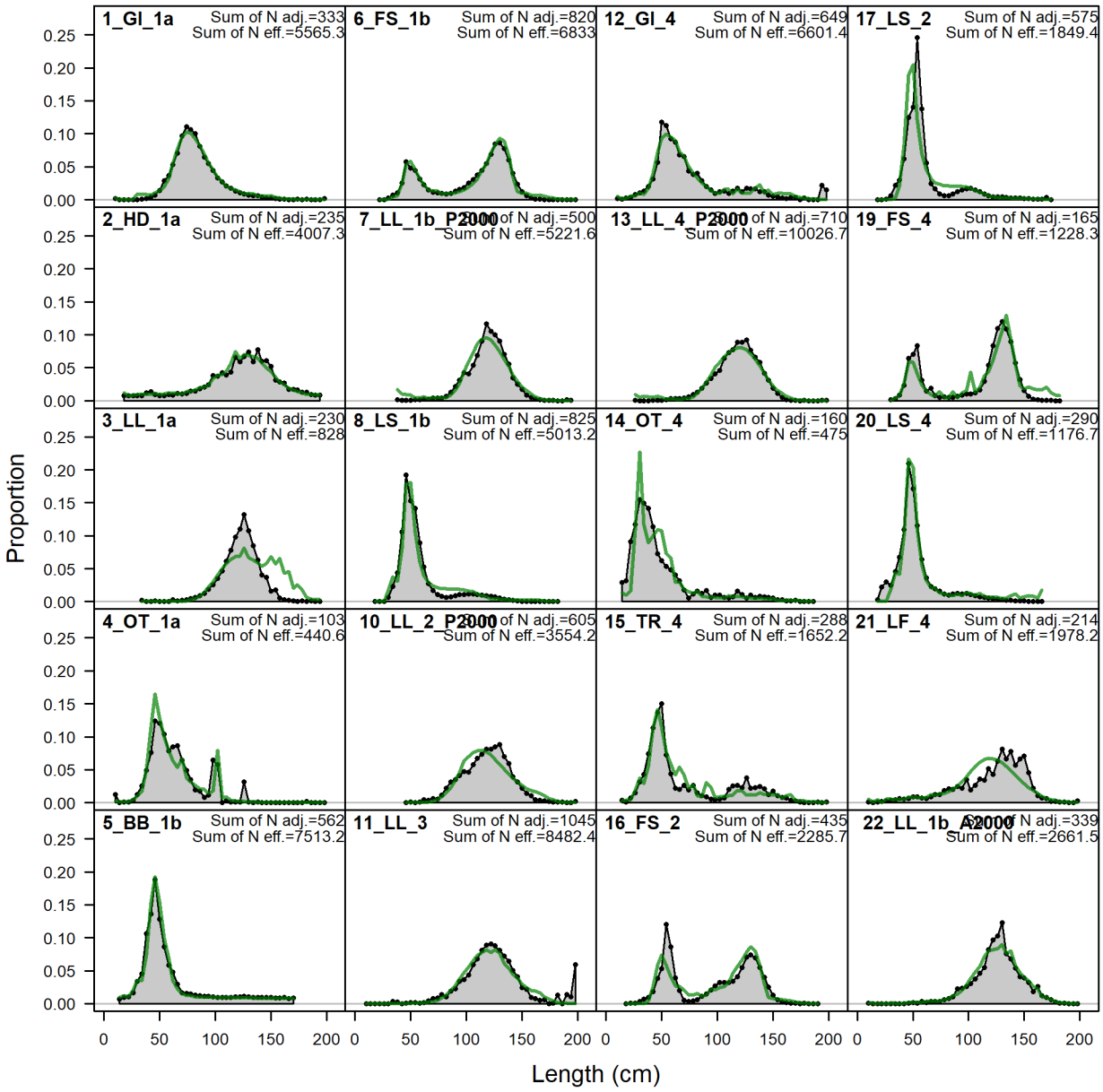


Figure 39: Observed (grey shadow) and predicted (green line) proportion of length compositions for each fishery aggregated over time for RM2.

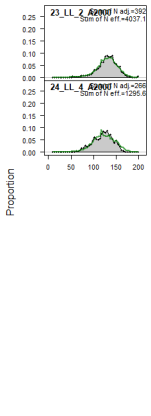


Figure 40: Observed (grey shadow) and predicted (green line) proportion of length compositions for each fishery aggregated over time for RM2.

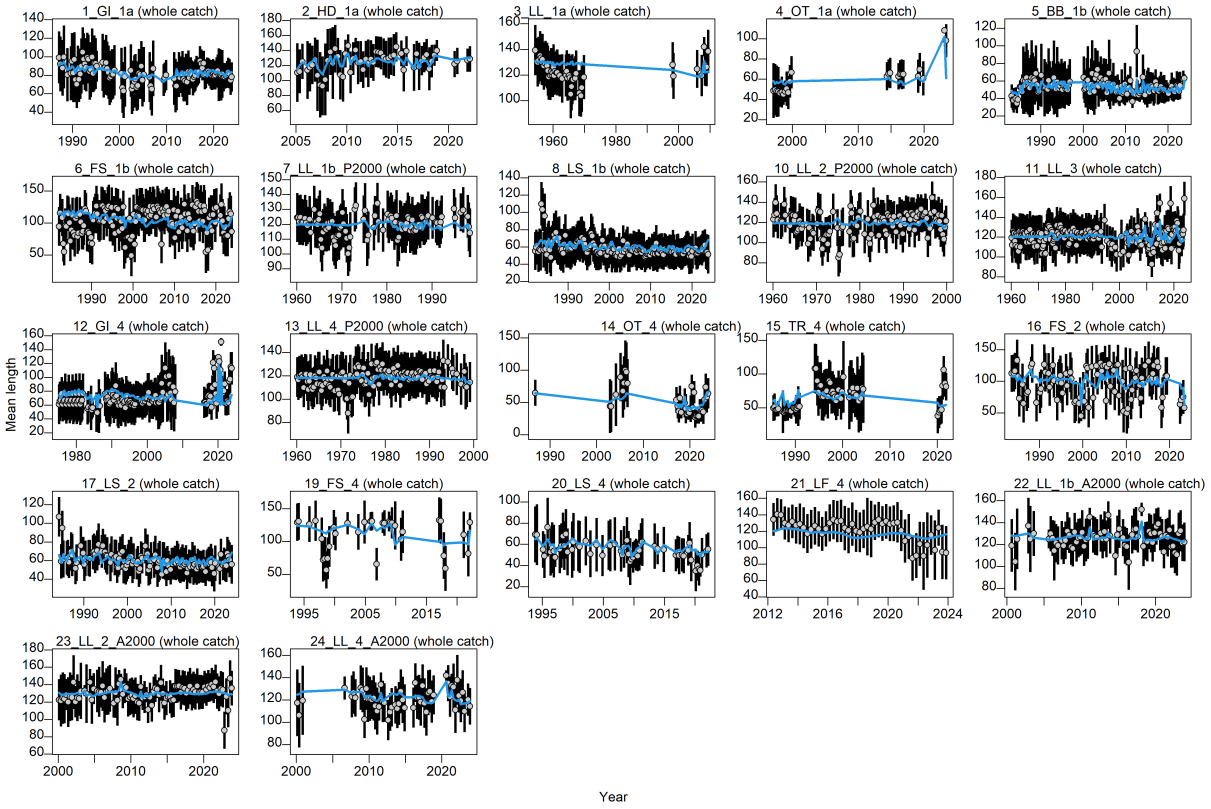


Figure 41: A comparison of the observed (grey points) and predicted (red points and line) average fish length (FL, cm) by fishery for RM2.

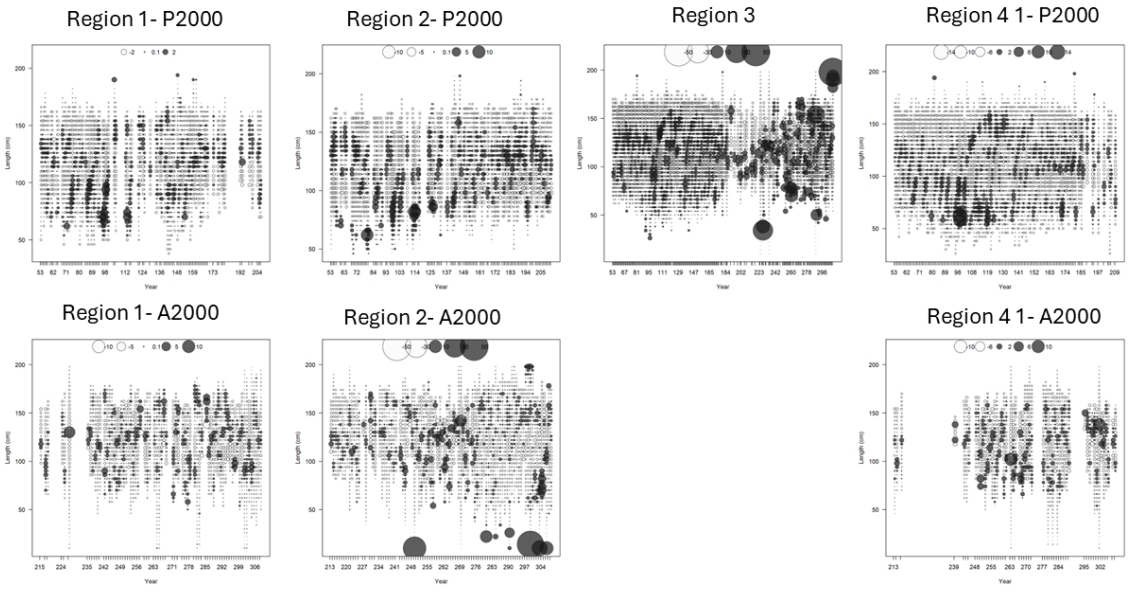


Figure 42: Relative residuals from the fits to the length compositions for LL regions 1b, 2, 3, and 4, for previous and after 2000 for RM2.

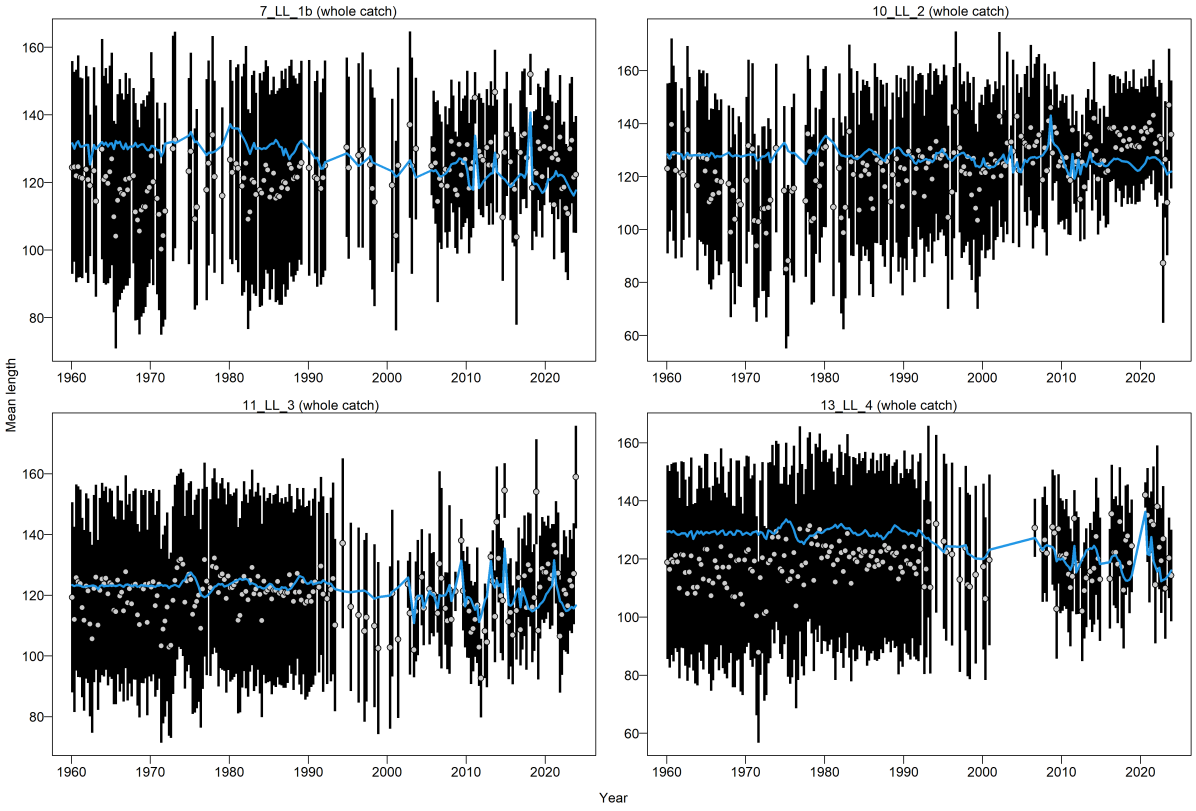


Figure 43: A comparison of the observed (grey points) and predicted (red points and line) average fish length (FL, cm) by fishery for RM1.

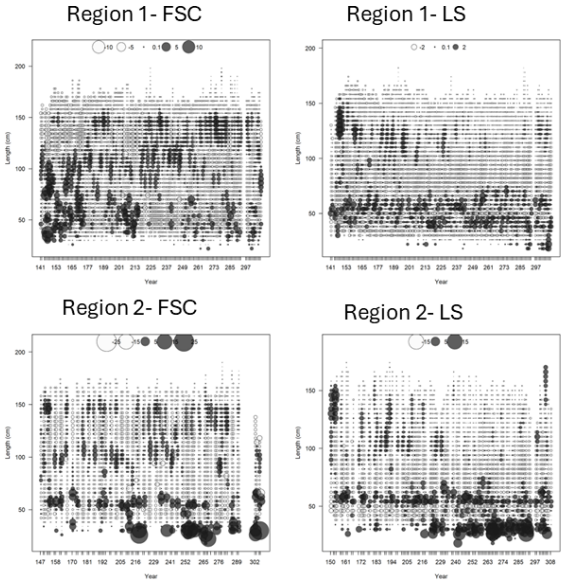


Figure 44: Relative residuals from the fits to the length compositions for FS in regions 1b and 2, and LS fisheries in regions 1b and 2, for the RM2 model.

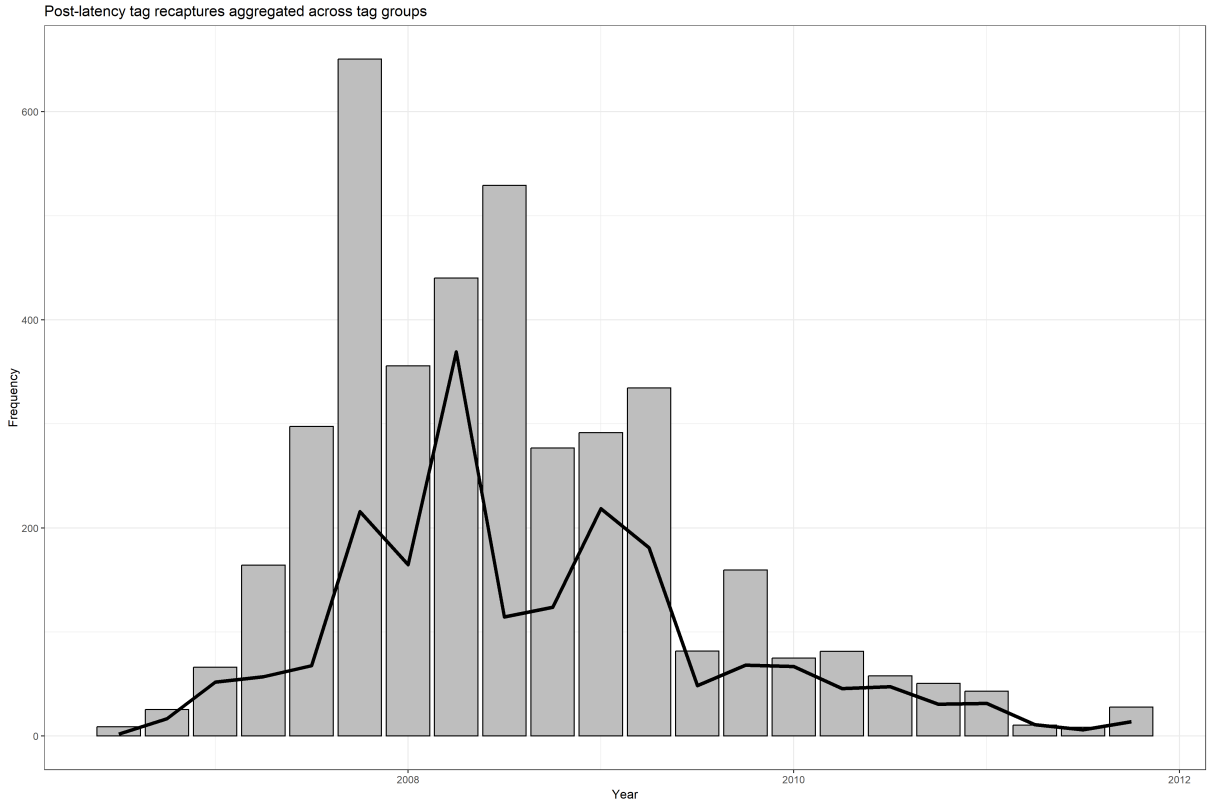


Figure 45: The grey bars show the number of tag recapture aggregated across tag groups with time, and the line shows the prediction by RM2.

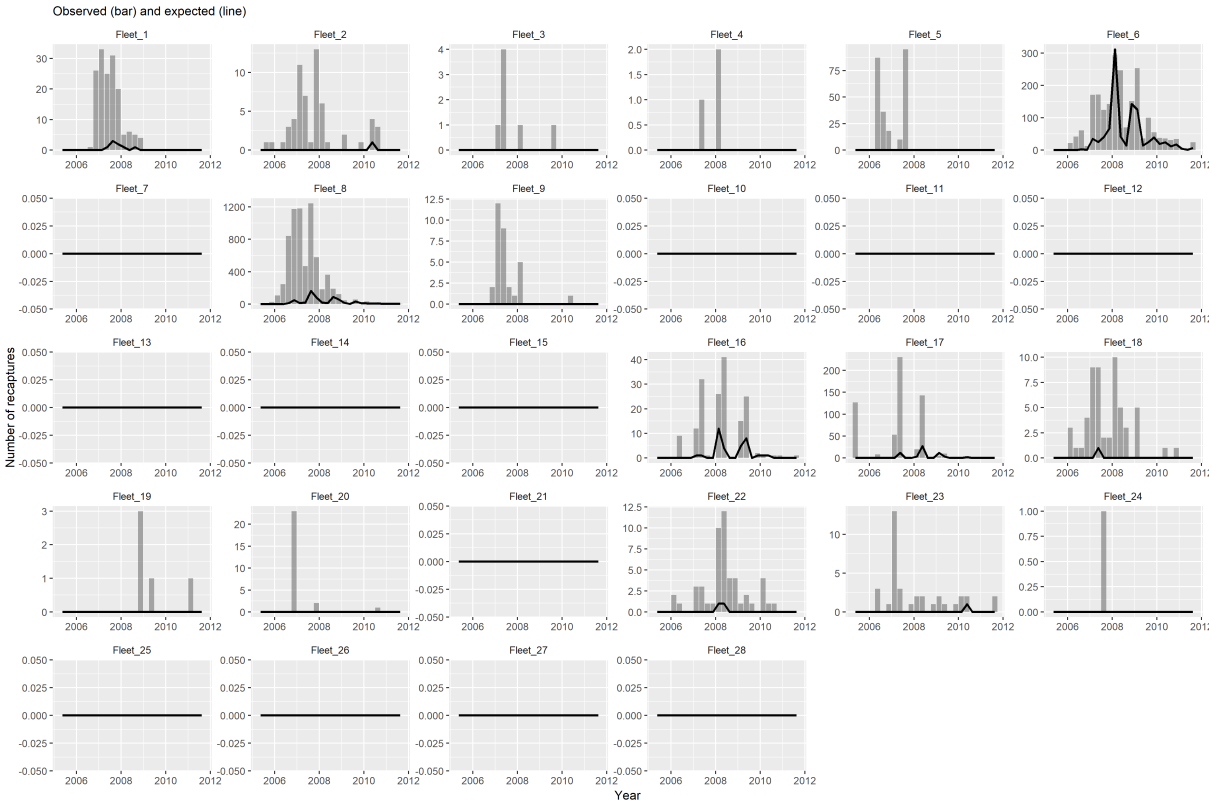


Figure 46: The grey bars the number of tag recaptures by fleet with time and the line the predicted by the RM2.

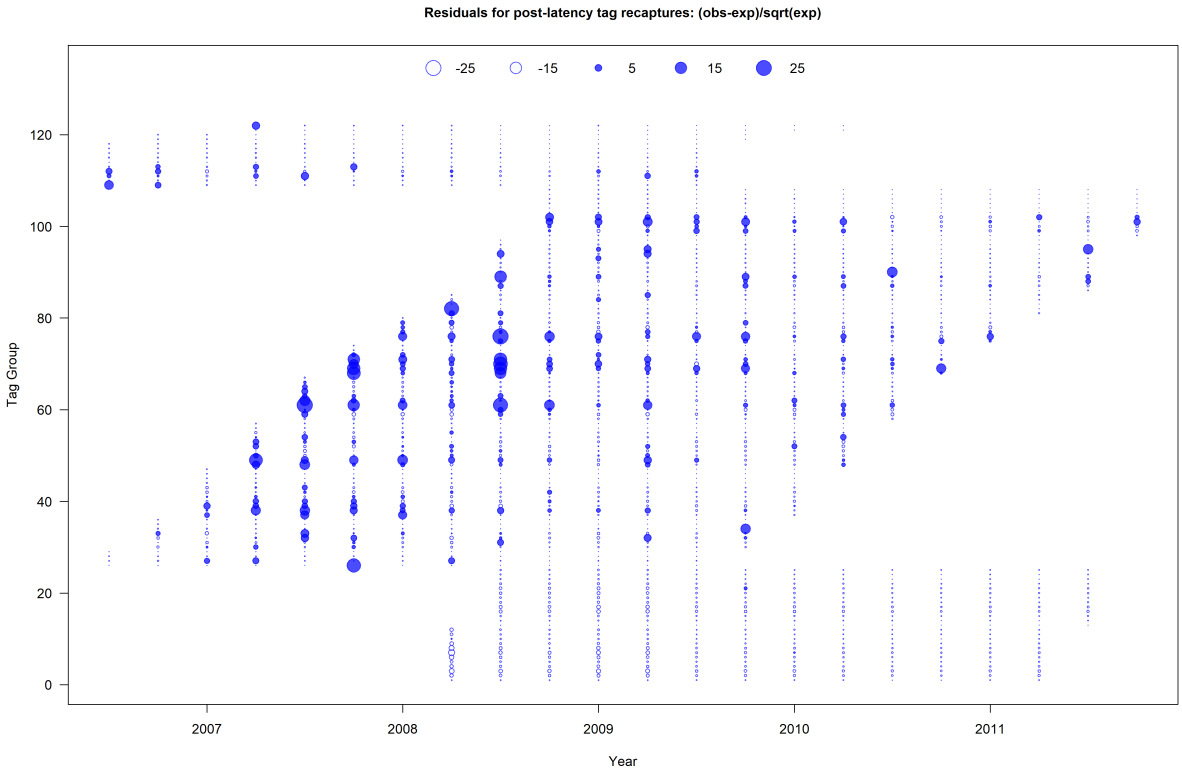


Figure 47: Residuals for post-latency tag recaptures: (obs-exp)/sqrt(exp) by RM2.

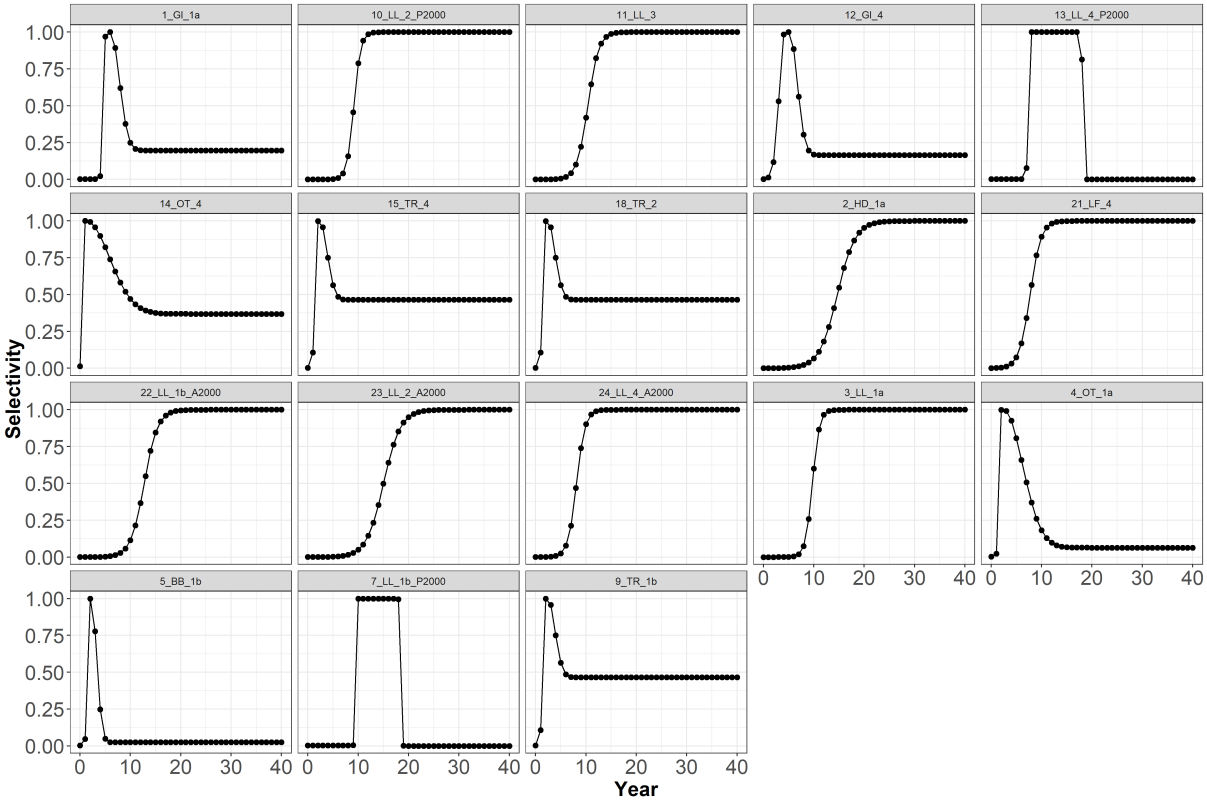


Figure 48: The estimated age-based selectivity functions (except for purse seine fisheries which are length-based) by RM2.

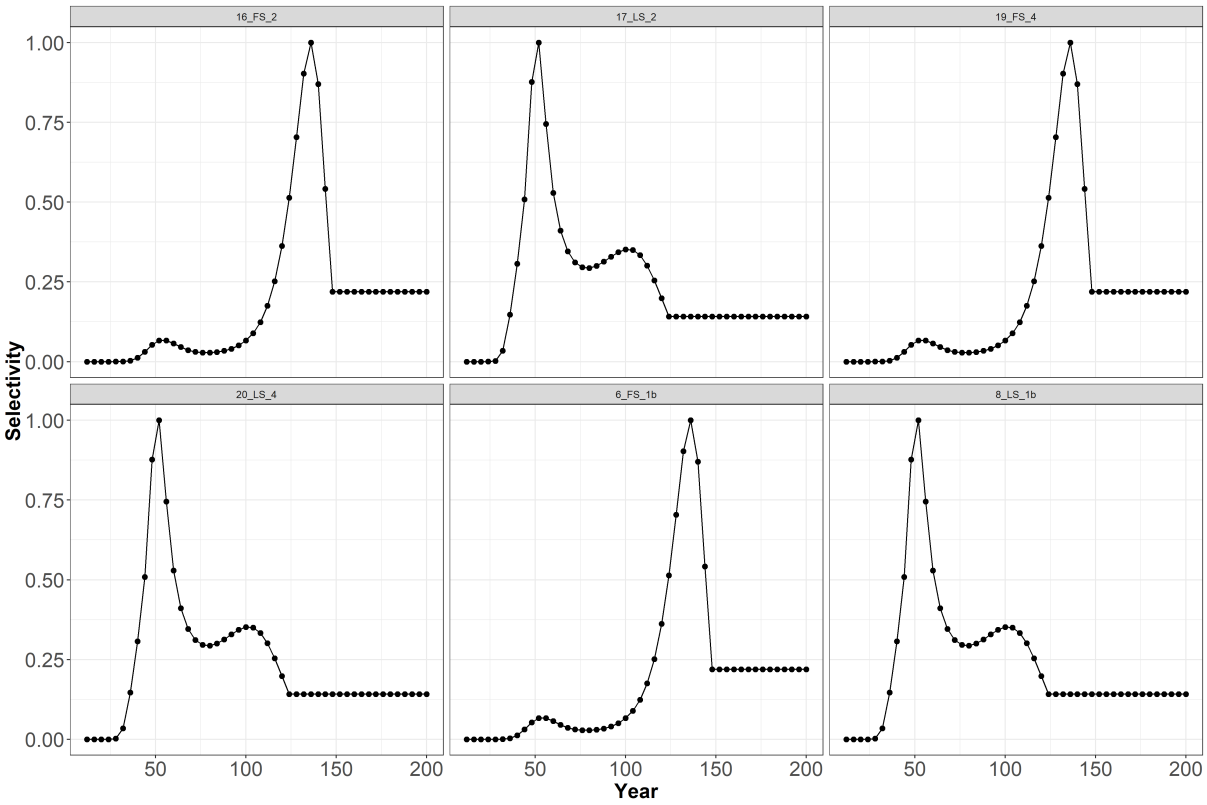


Figure 49: The estimated size based selectivity functions (except for purse seine fisheries which are length-based) by RM2.

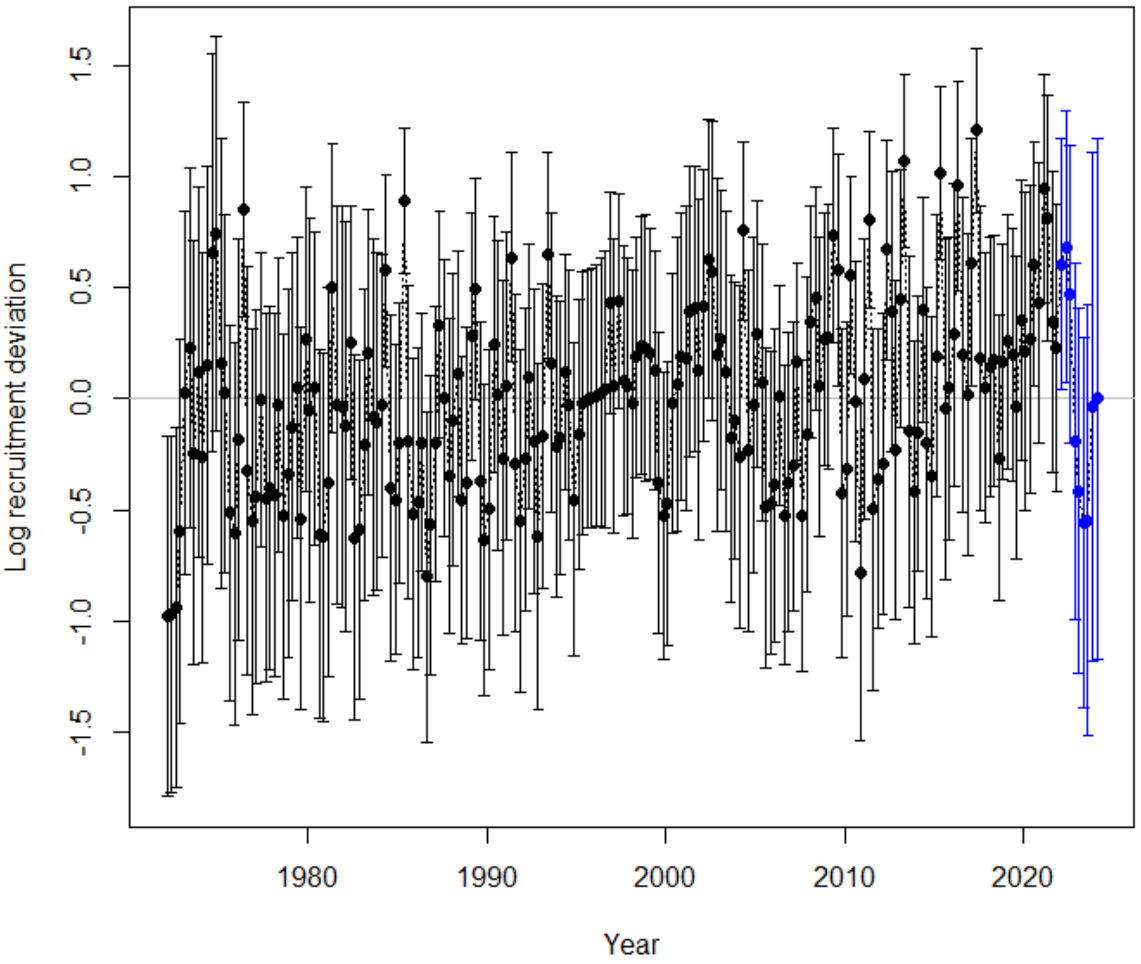


Figure 50: Recruitment deviates from the SRR with 95% confidence interval from the RM2.

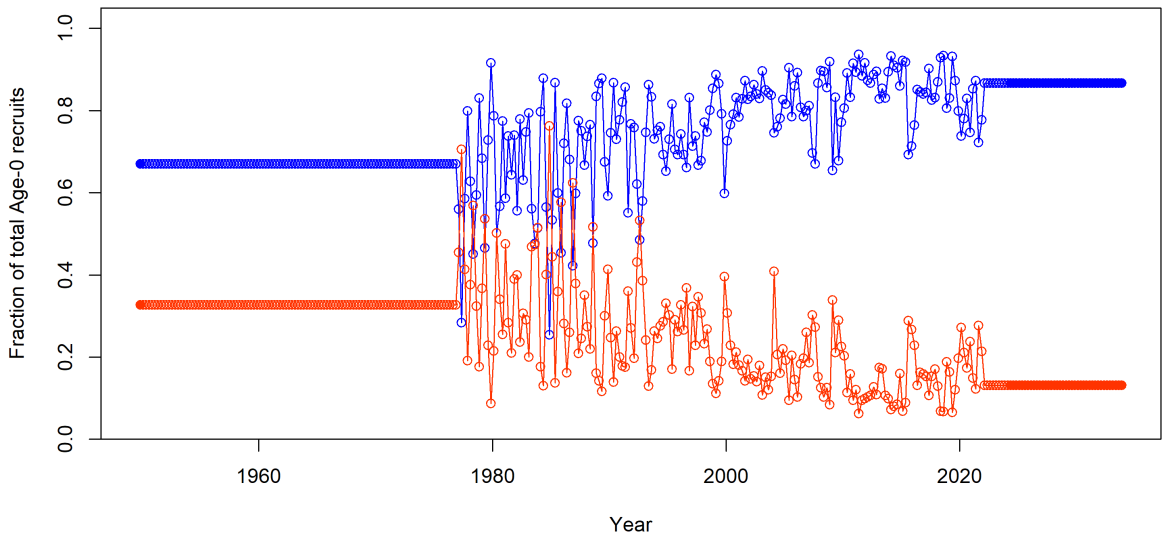


Figure 51: Proportion of the total quarterly recruitment assigned to region 1 (red) and region 4 (blue).

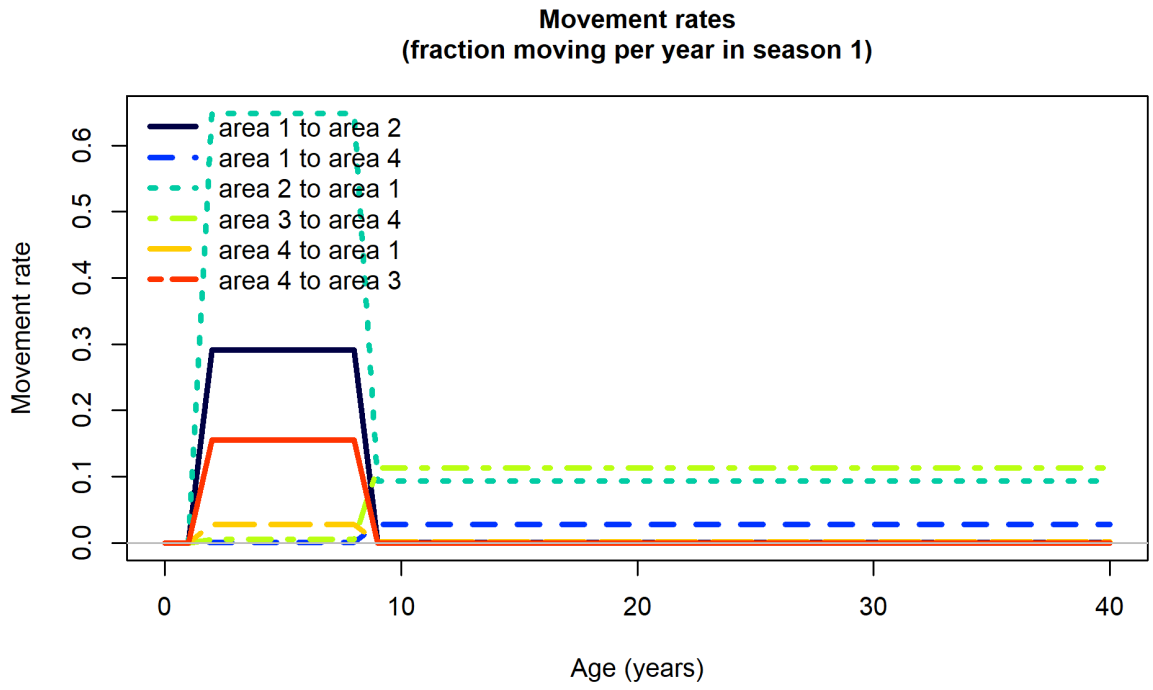


Figure 52: Estimated age specific movement parameters for the RM2.

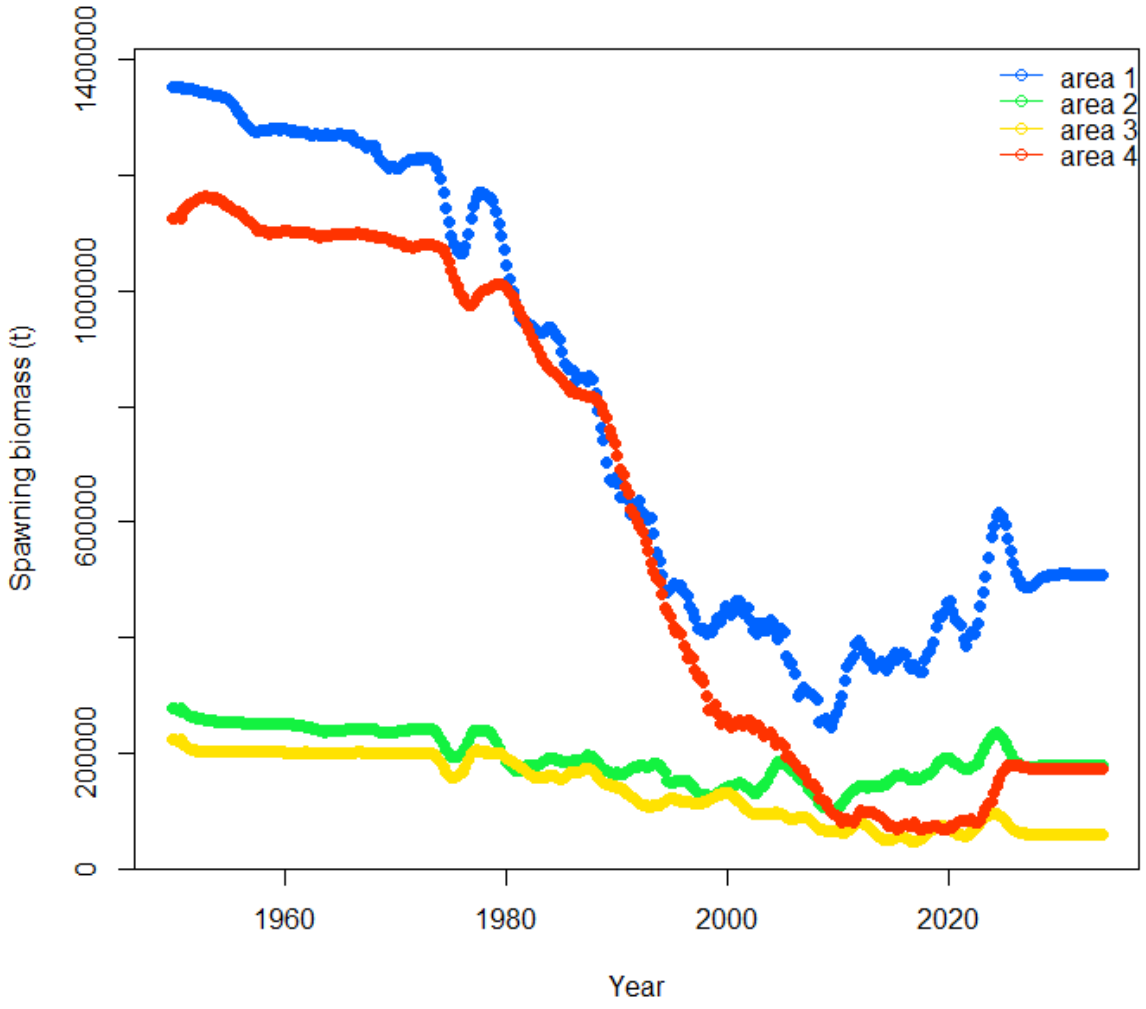


Figure 53: Estimated spawning biomass trajectories for the individual model regions from the RM2.

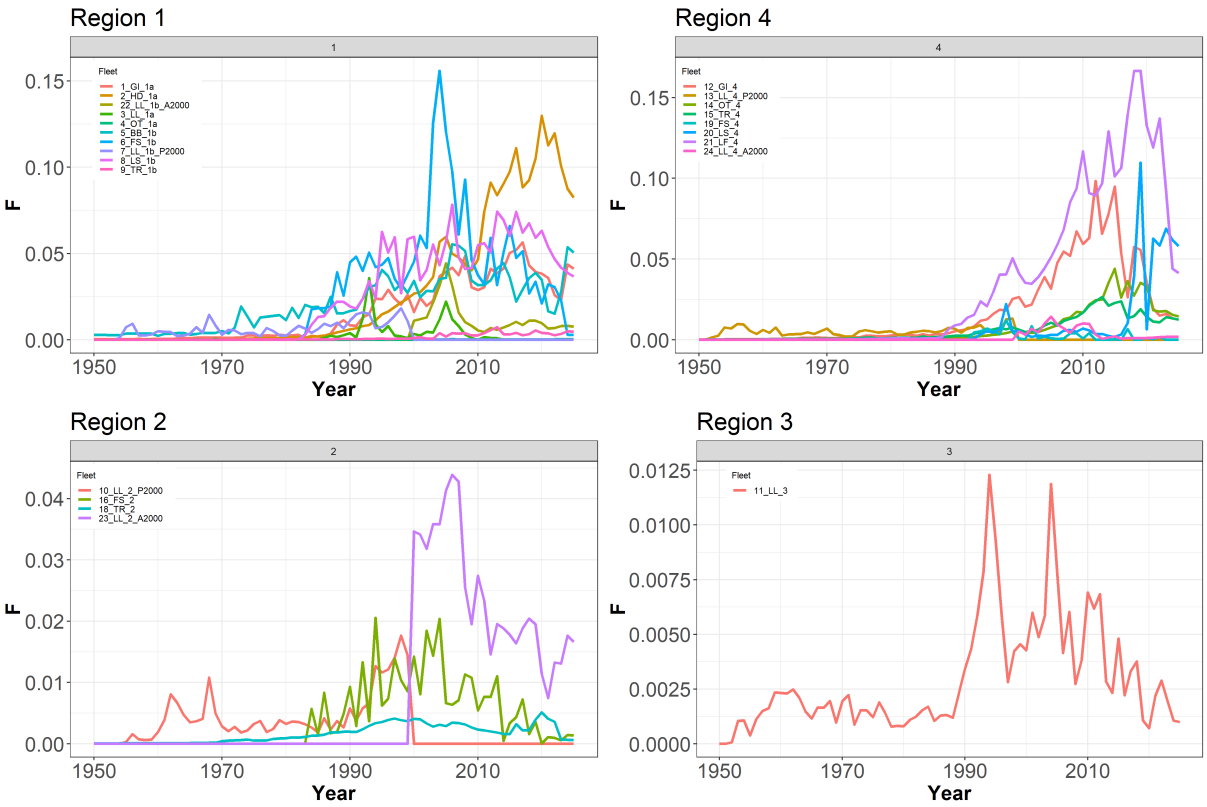


Figure 54: Trends in annual fishing mortality by fleet and area.

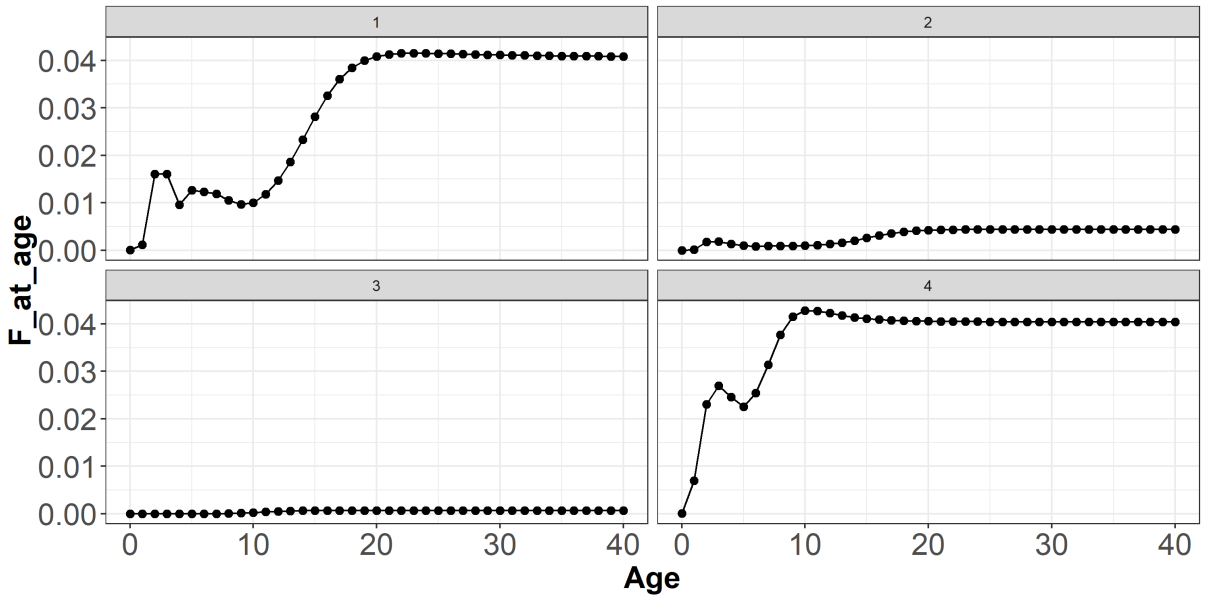


Figure 55: Fishing mortality (quarterly, average) by age class and region estimated by the RM2.

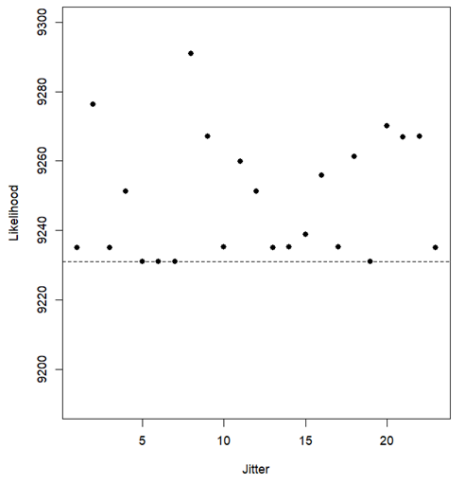


Figure 56: Jitter analysis from 30 runs the log-likelihood of 23 runs with convergency level <0.001. The runs at the minimum log-likelihood converged with the same results.

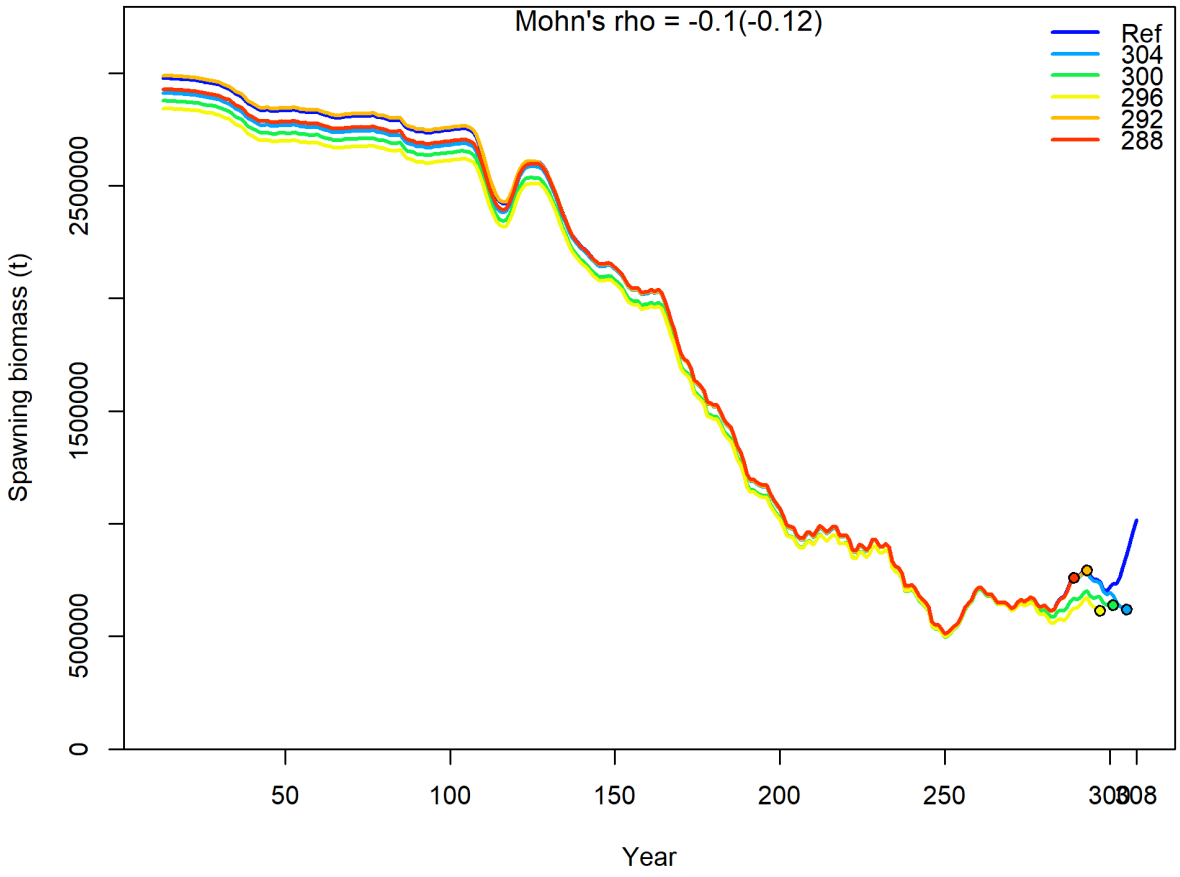


Figure 57: Retrospective analysis for SSB from RM2.

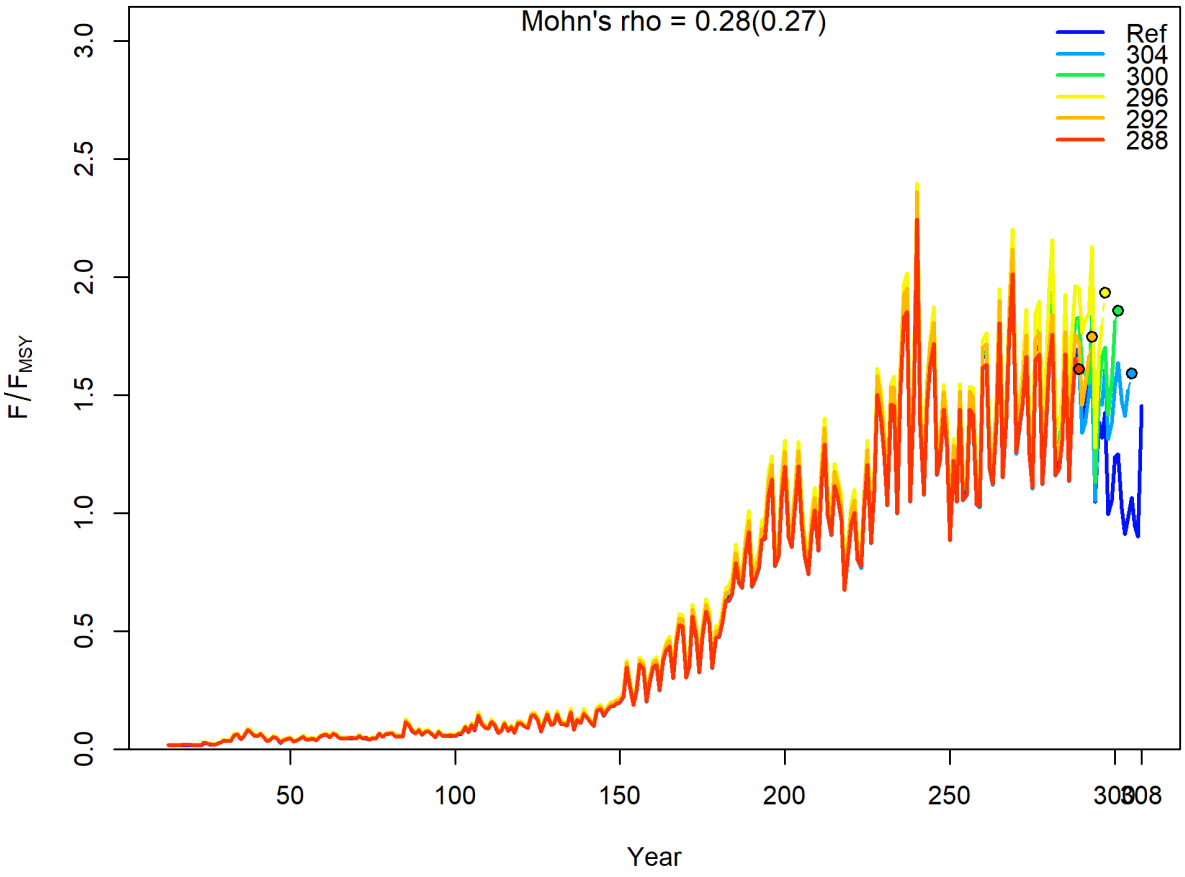


Figure 58: Retrospective analysis for $F/F_{MSY}$ from RM2.

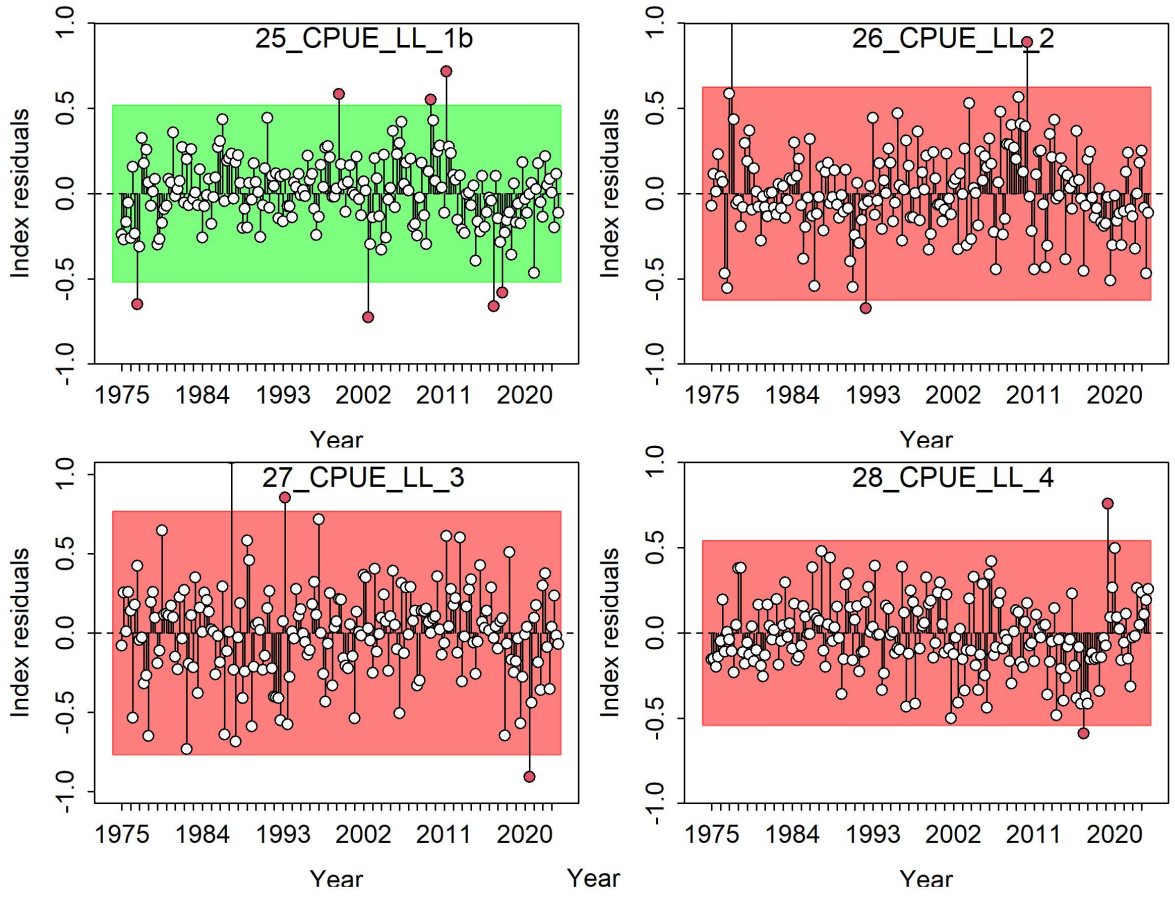


Figure 59: Analysis of the residual pattern of the fits to the LL CPUE by region of RM2 with runs test.

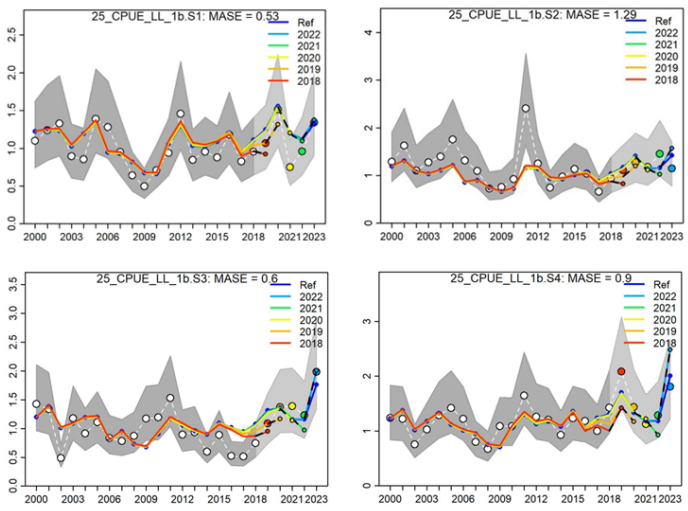


Figure 60: Hindcast analysis to evaluate the skills of the model to predict the CPUE in region 1 by season, by the RM2.

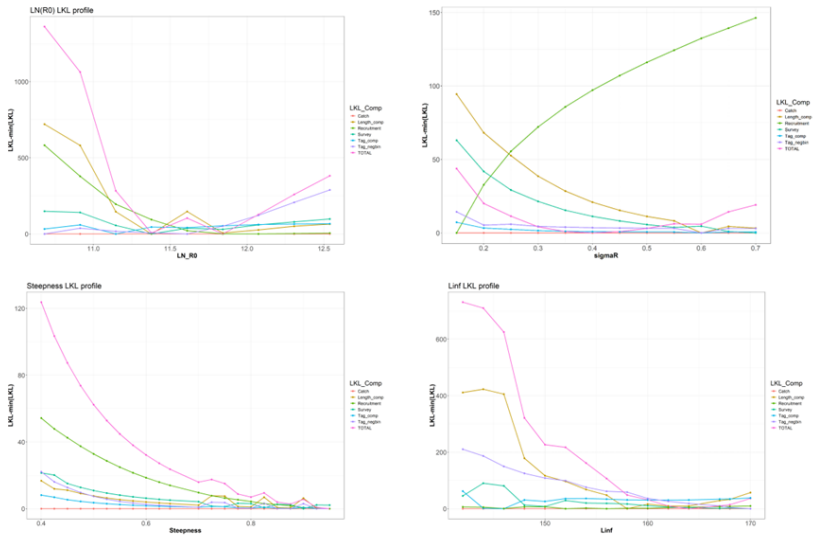


Figure 61: Profiling of steepness, unfished recruitment, recruitment variability, and asymptotic length using RM2.

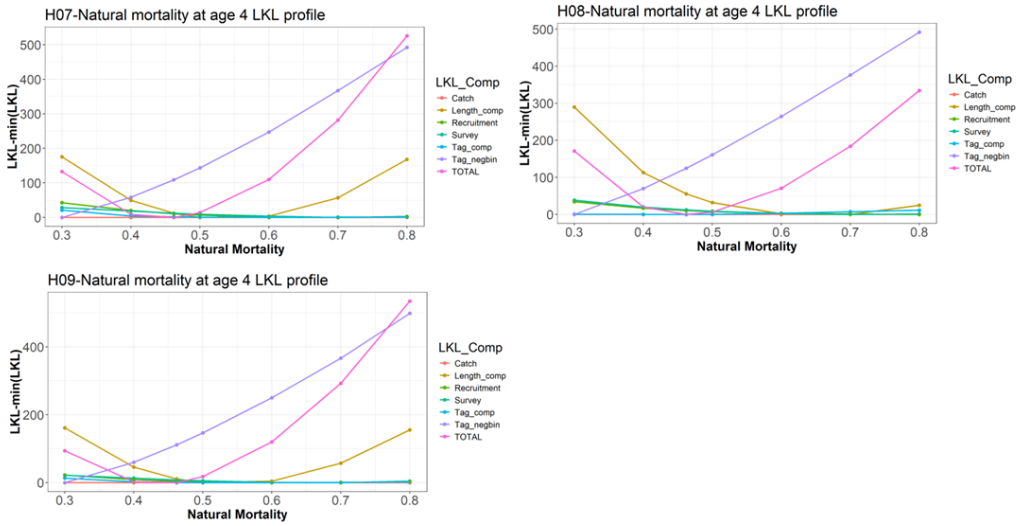


Figure 62: Profiling of natural mortality (M) in RM2 model assuming different values of steepness: 0.7, 0.8, and 0.9.

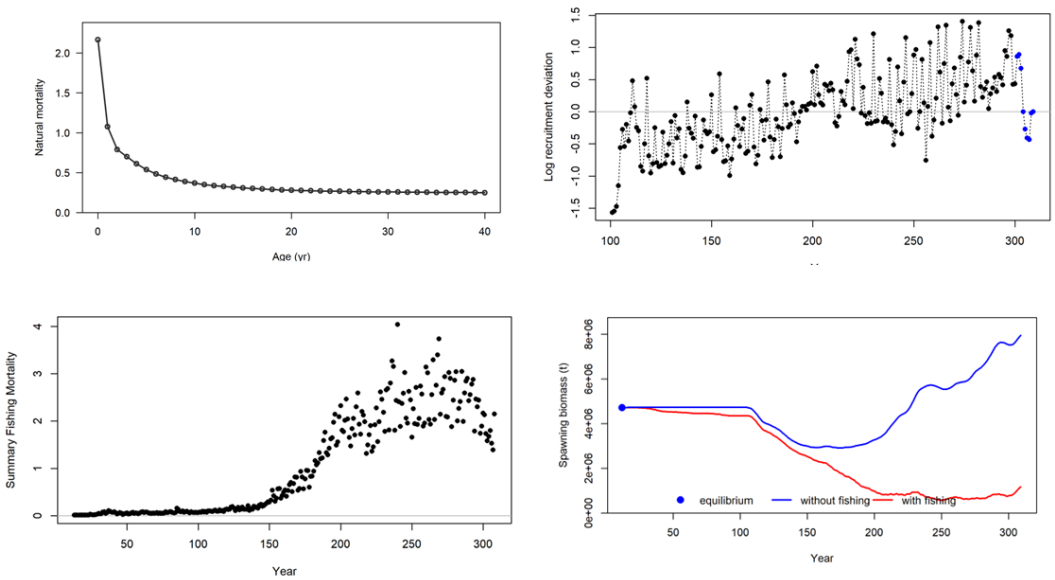


Figure 63: Sensitivity analysis assuming M=0.3: natural mortality at age, recruitment deviates, biomass without fishing, and $F/F_{MSY}$.

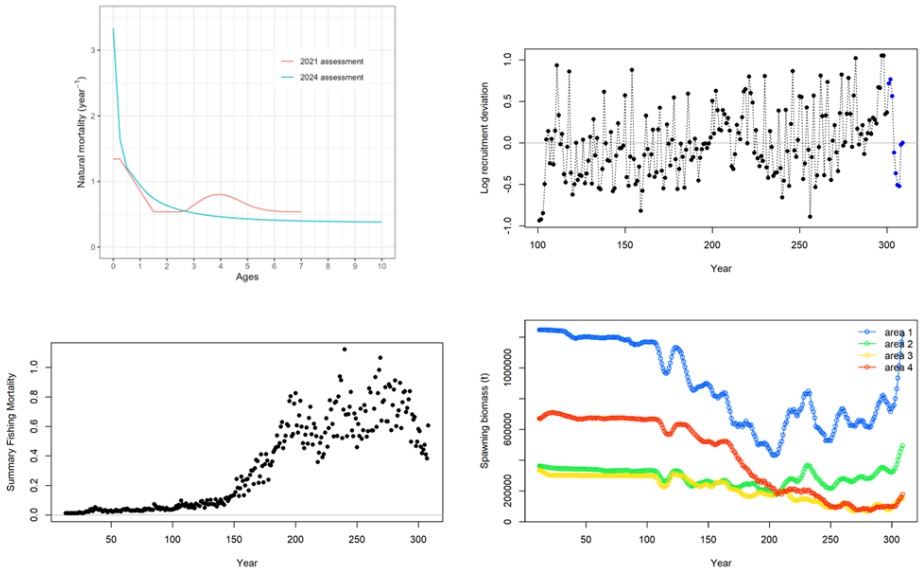


Figure 64: Sensitivity analysis assuming differences on natural mortality by gender (same M at age values as in the 2021 assessment): natural mortality at age, recruitment deviates, $F/F_{MSY}$ and spawning biomass by area.

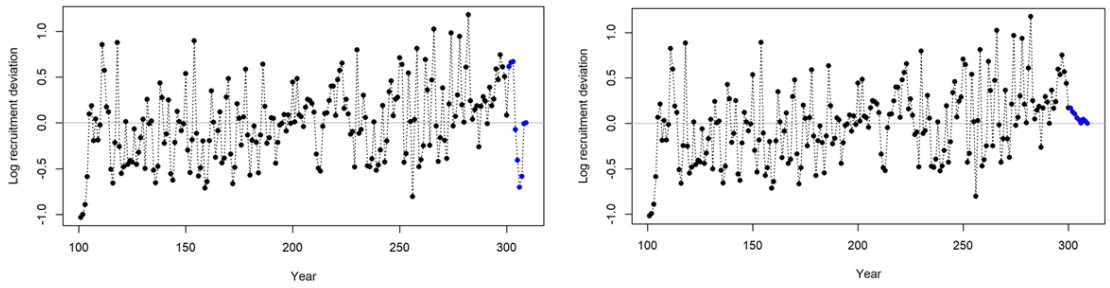


Figure 65: Sensitivity analysis of the recruitment deviates of the last years. The left figure shows the recruitment deviates after removing all the fleets length composition data from 300 (year 2021) onwards but no Other fisheries fishing in regions 1 and 4. The right figure shows the recruitment deviates when all the length composition data are removed from 300 onwards.

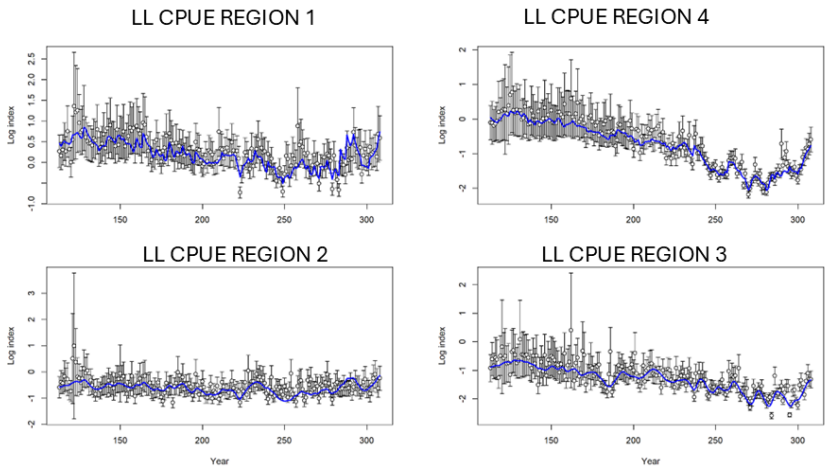


Figure 66: Sensitivity analysis of assuming variable CV based on the estimated values in the standardization of the CPUE but assuming a mean CV of 0.2: the observed CPUE LL, the 95% CI by region and the fits of the RM2.

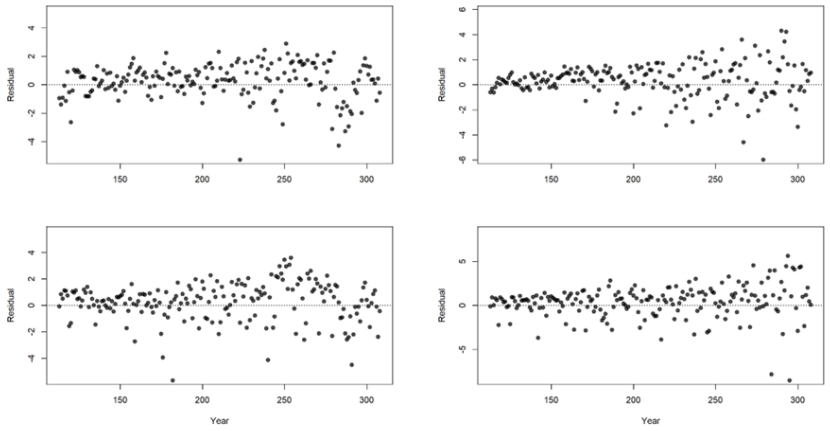


Figure 67: Sensitivity analysis of assuming variable CV based on the estimated values in the standardization but assuming a mean CV of 0.2: residuals by region, of the sensitivity analysis considering LL CPUE with variable CV, of the RM2.

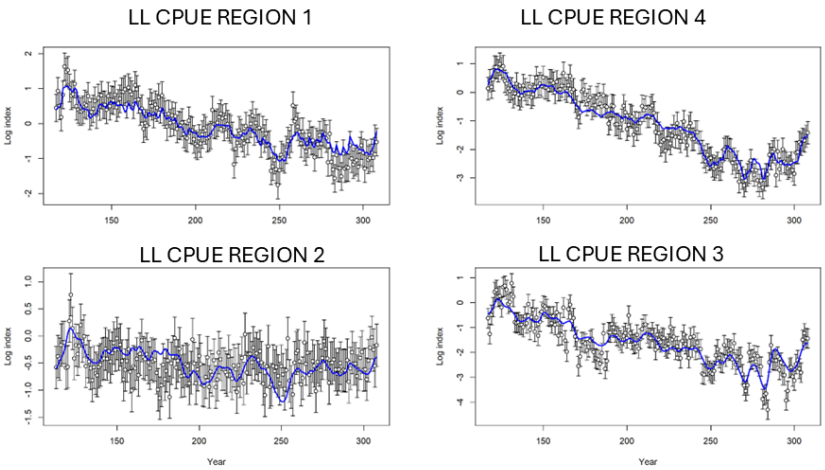


Figure 68: Sensitivity analysis considering operational LL CPUE as index; the observed data and the fits to the LL operational CPUE by region in the RM2.

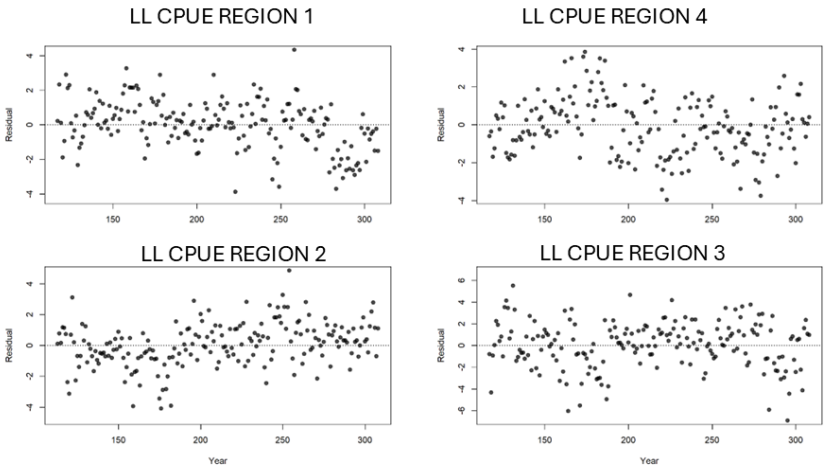


Figure 69: Sensitivity analysis considering operational LL CPUE as index; the residuals to the fits of LL operational CPUE by region in the RM2.

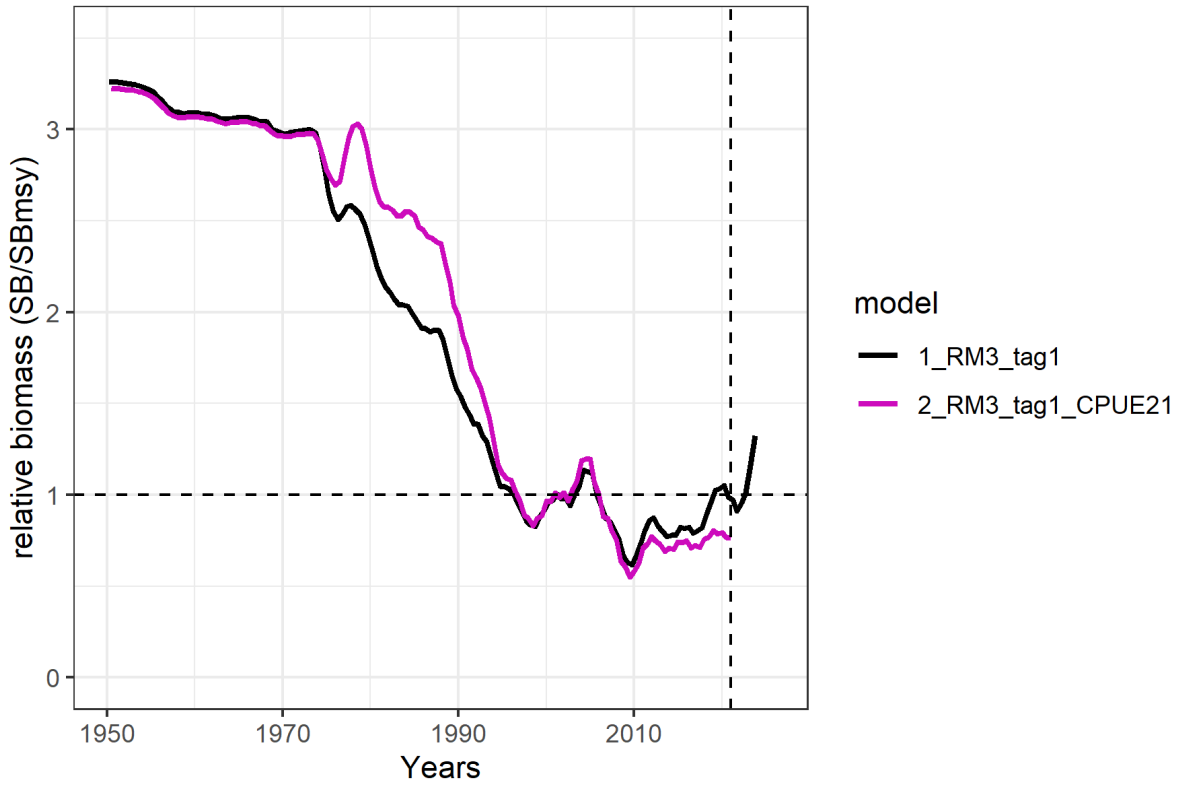


Figure 70: Comparison of the estimated relative SSB time series introducing the 2021 LL CPUE in the RM3 and running the model until 2021. This exploration was run and discussed during the 26th WPTT.

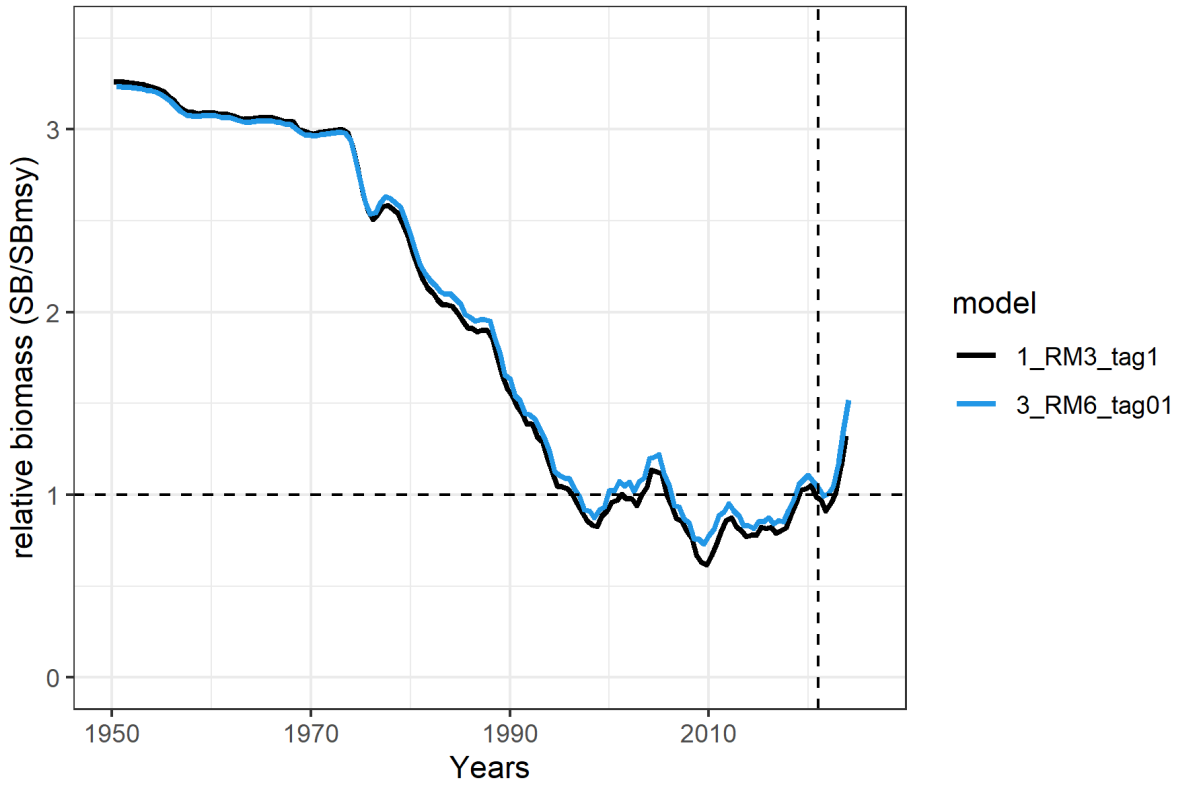


Figure 71: Comparison of the estimated relative SSB time series using RM3 (not included in the final uncertainty grid) and the *6_SplitCPUE_tag01_EC0_h08* model (blue line). Both series use the 2024 LL CPUE.

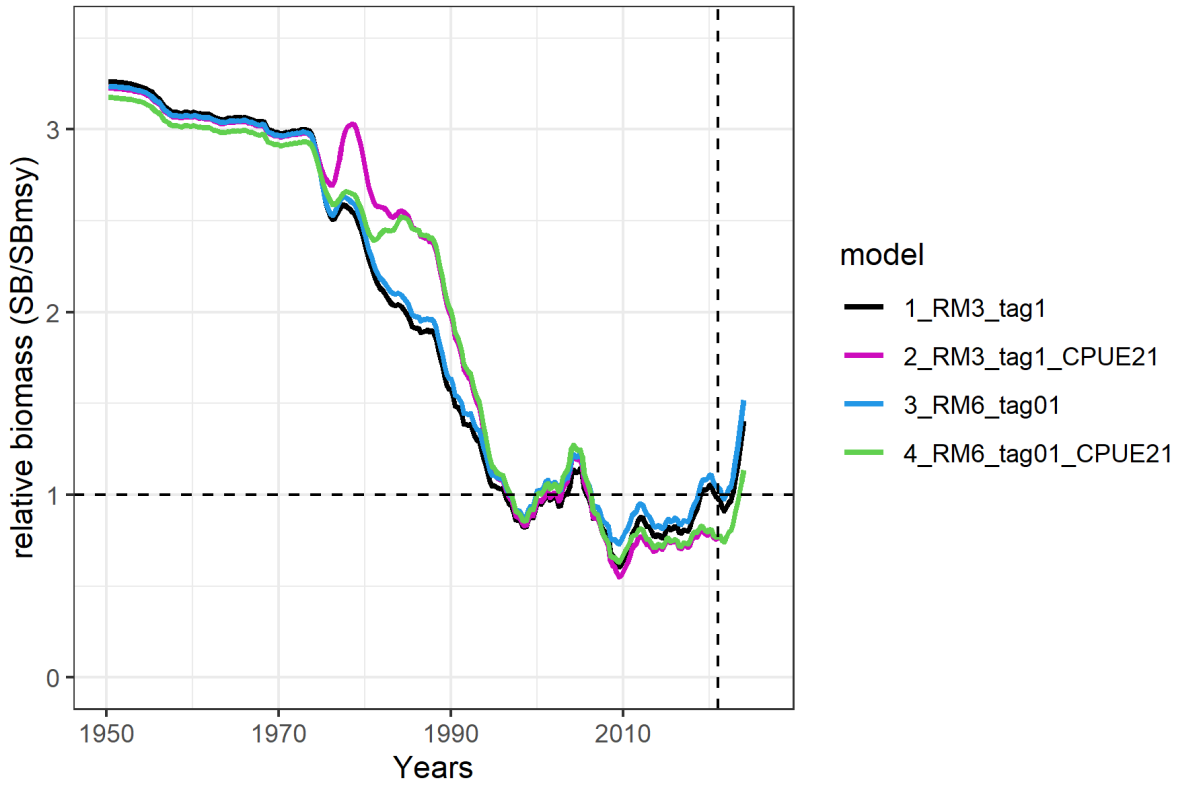


Figure 72: Comparison of the estimated relative SSB time series using the RM3 (black line), RM3 with the LL CPUE used in 2021 and estimation period terminated in 2021 (purple line), *6_SplitCPUE_tag01_EC0_h08* (blue line), and *6_SplitCPUE_tag01_EC0_h08* using the 2021 CPUE and including catch and length frequency data for the period 2021-2023 (green line).

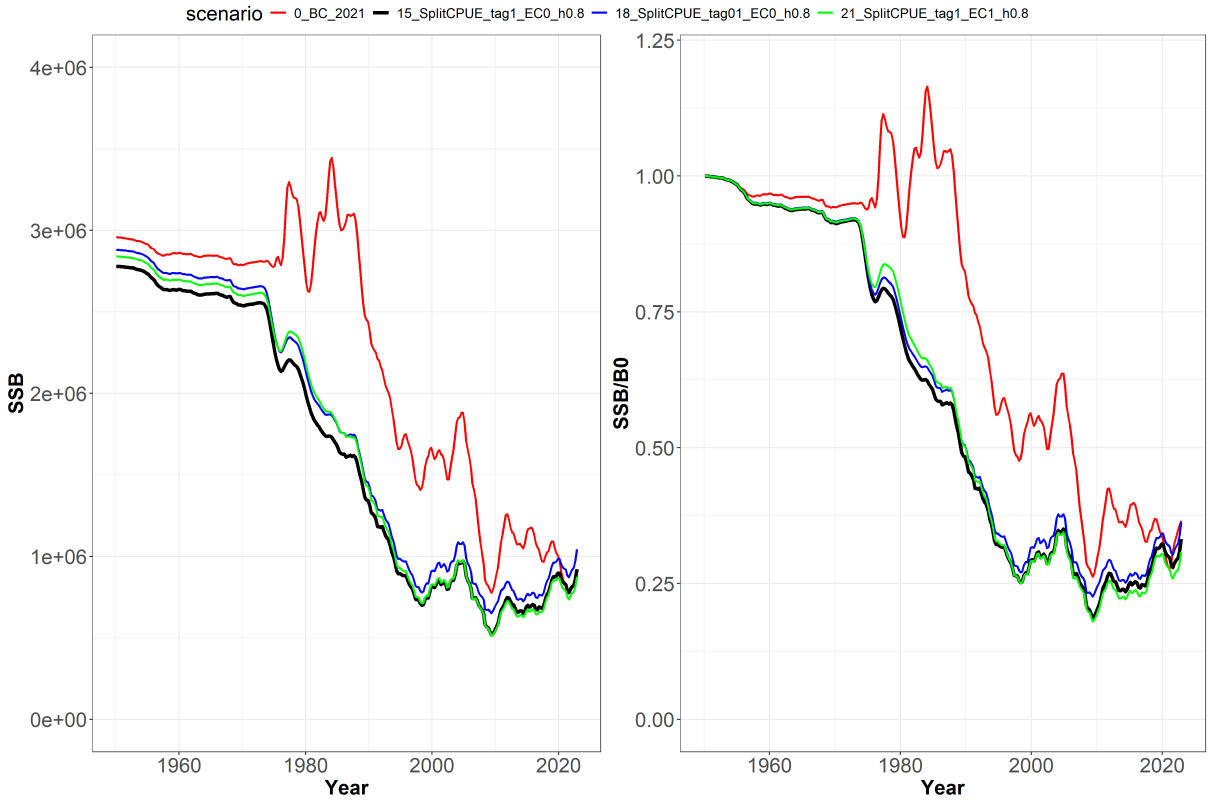


Figure 73: Comparison of the time series of the RM3 model (LL CPUE selectivity the same as the splitted LL fishery) on $SSB/SSB_0$ with tagging data downweighted and applying effort creep of 0.5% per year.

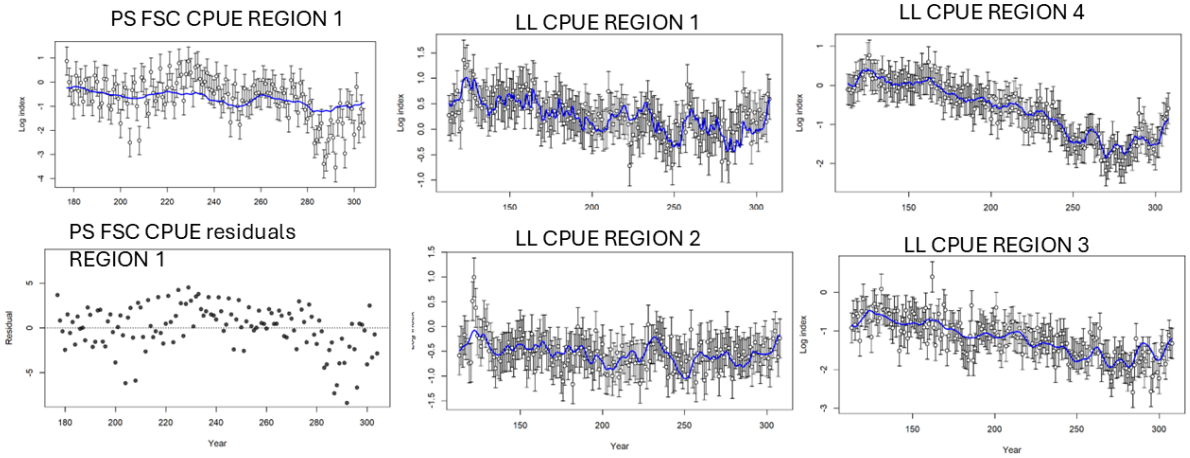


Figure 74: Sensitivity analysis including FS CPUE in region 1, assuming a constant CV of 0.3.

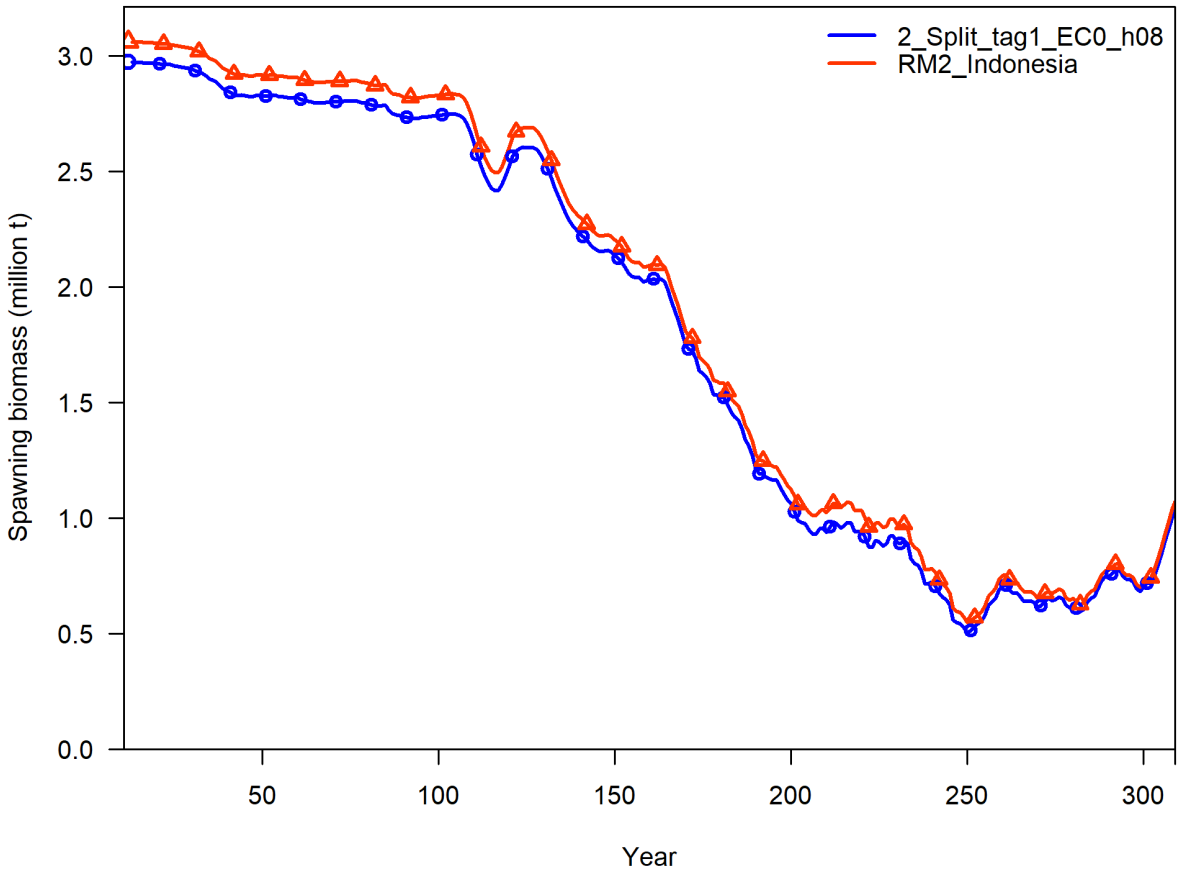


Figure 75: Comparison of the time series of SSB of RM2 with the model including Indonesian revised catches.

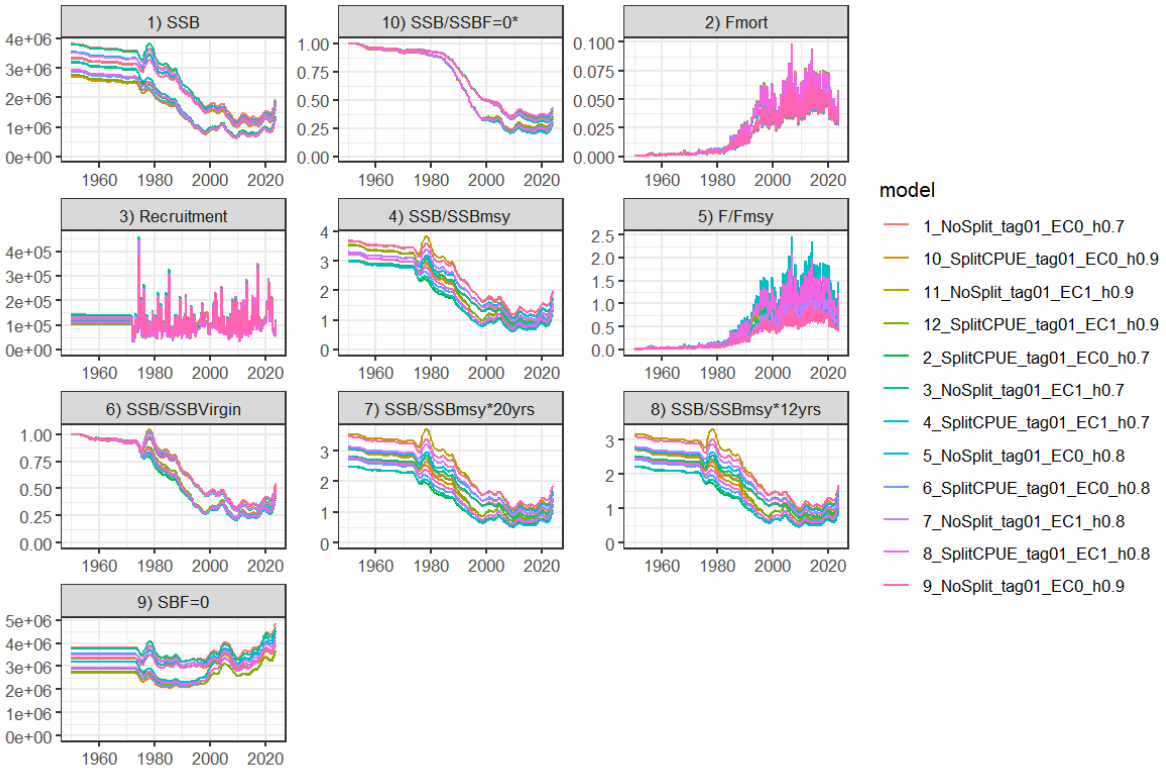


Figure 76: Stock status for the final five candidate reference models.

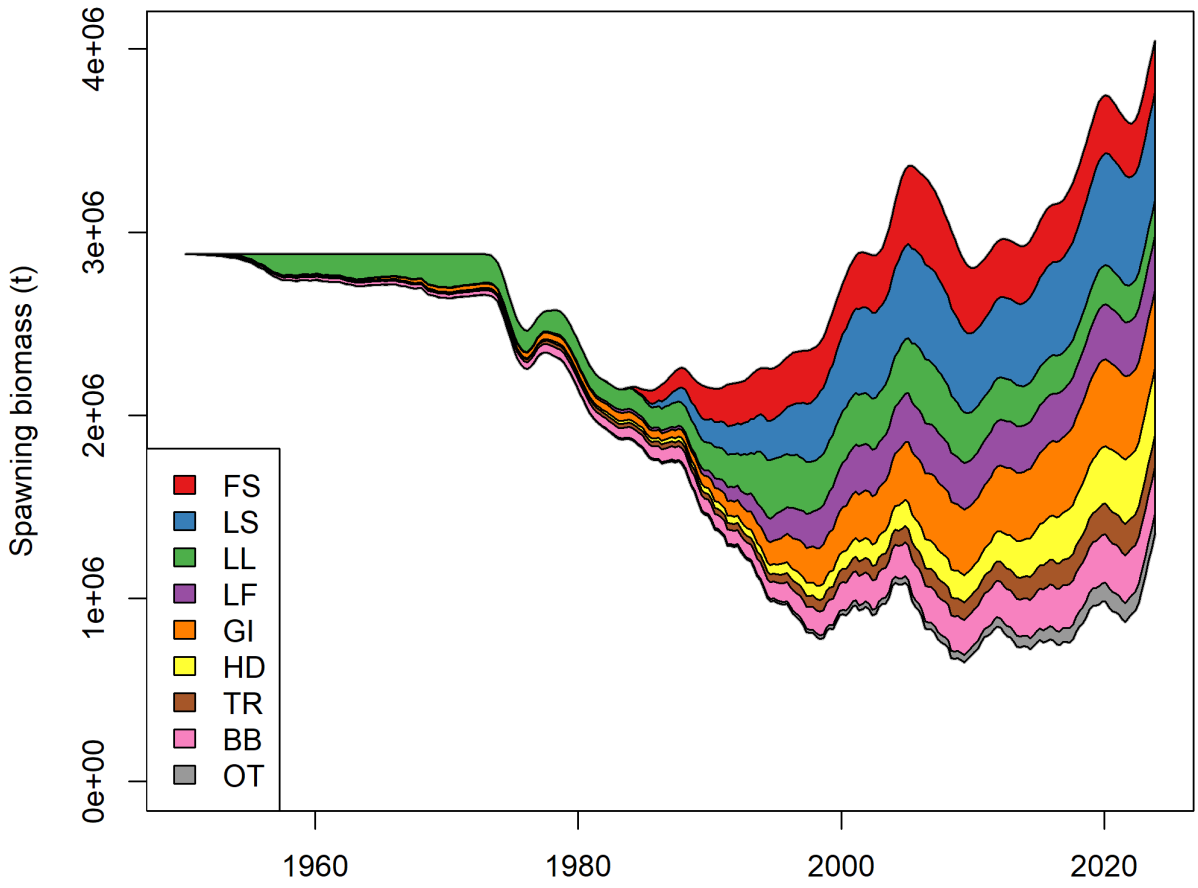


Figure 77: Impact plot from *6_SplitCPUE_tag01_EC0_h0.8*. The coloured area repsents the effect of removing each fishery group.

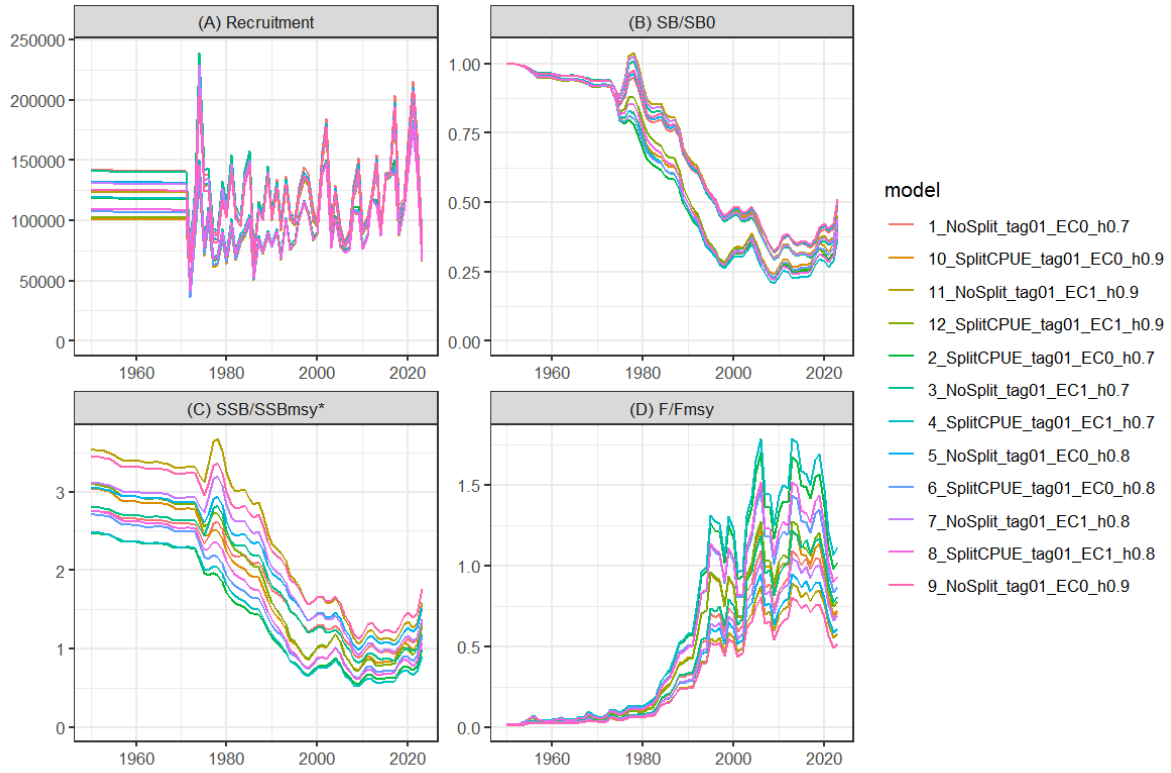


Figure 78: Temporal series of relevant model estimates for all models in the uncertainty grid.

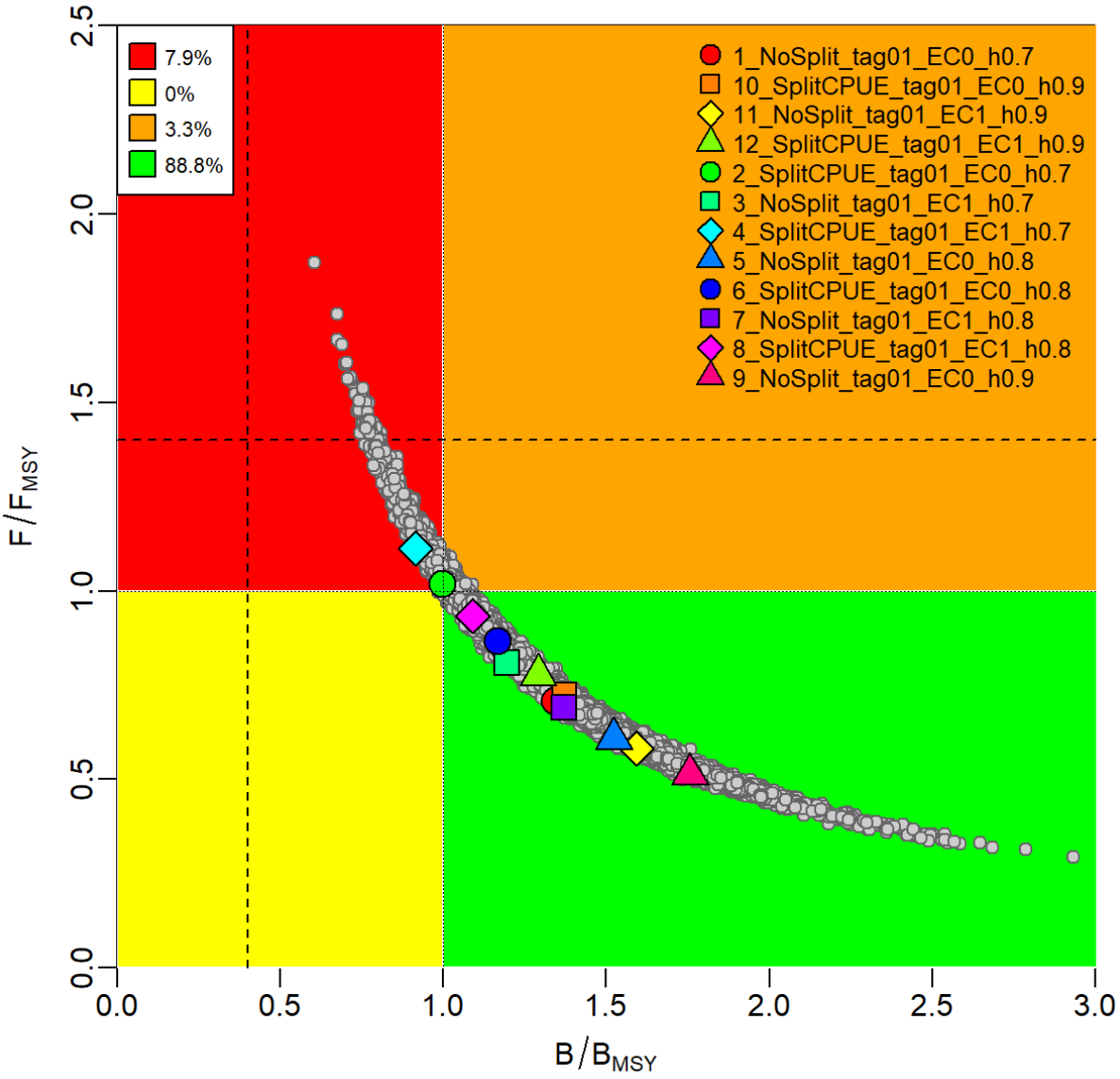


Figure 79: Kobe plot. The colored symbols represent the estimated stock status ($SSB/SSB_{MSY}$ and $F/F_{MSY}$) in 2023 for each model in the final uncertainty grid. The gray dots represent the uncertainty in the stock status estimates. The probability of being in each quadrant of the Kobe plot in 2023 is also shown.

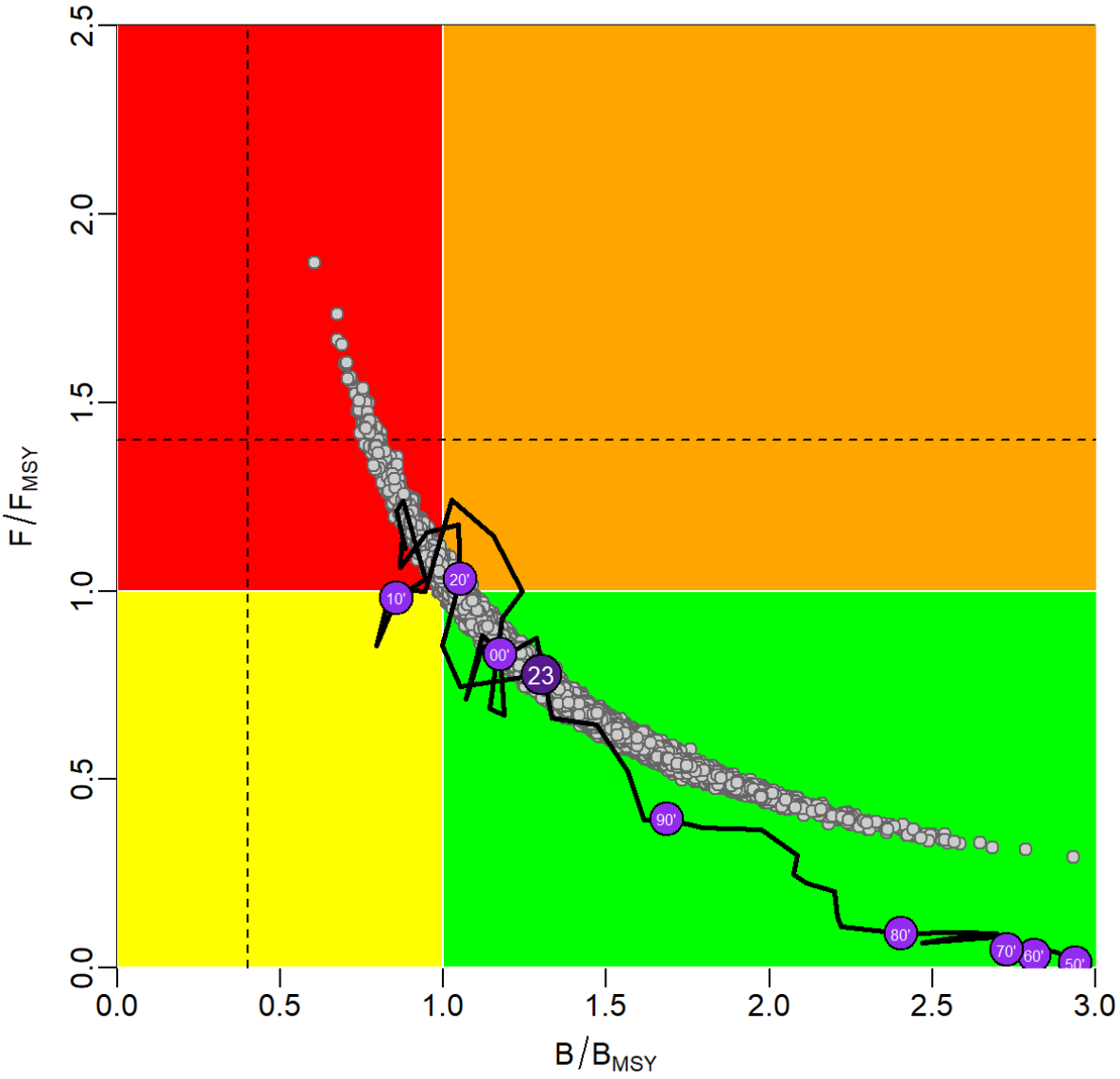


Figure 80: Kobe plot. Decadal trajectory of stock status based on all models in the uncertainty grid.

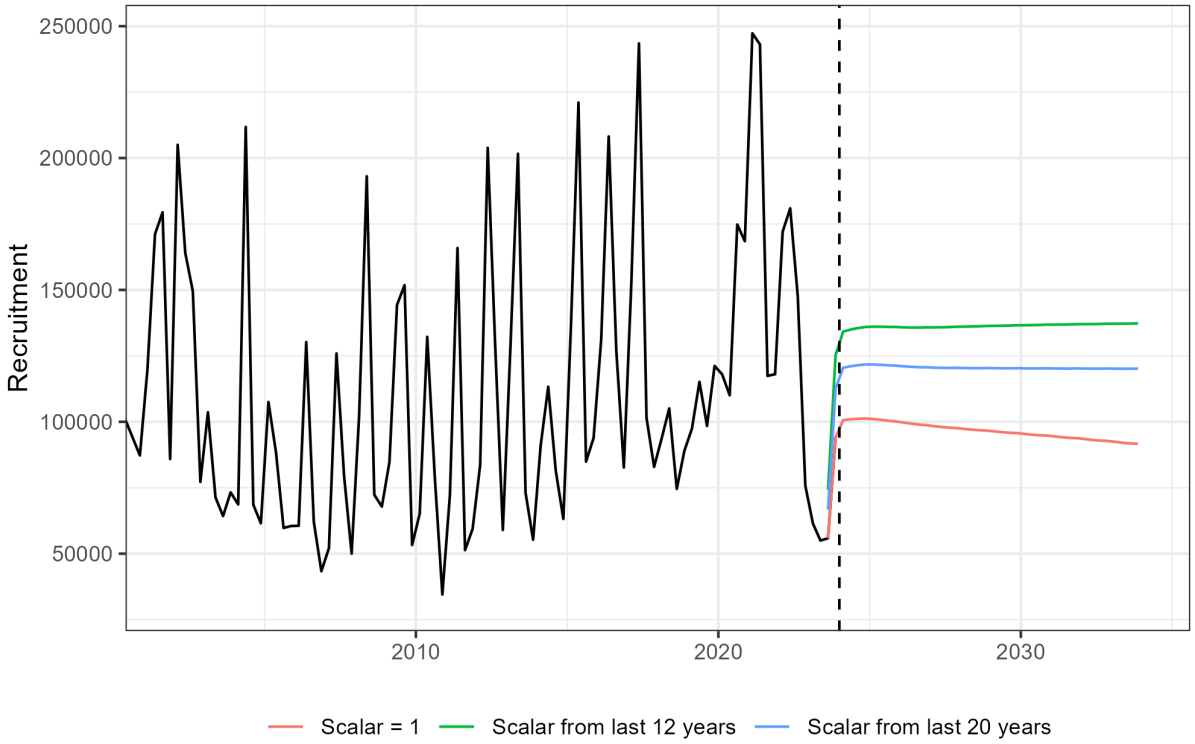


Figure 81: Impacts of different scalars on projected recruitment (colored lines) for RM3. The black continuous line is the estimated recruitment during the model period.

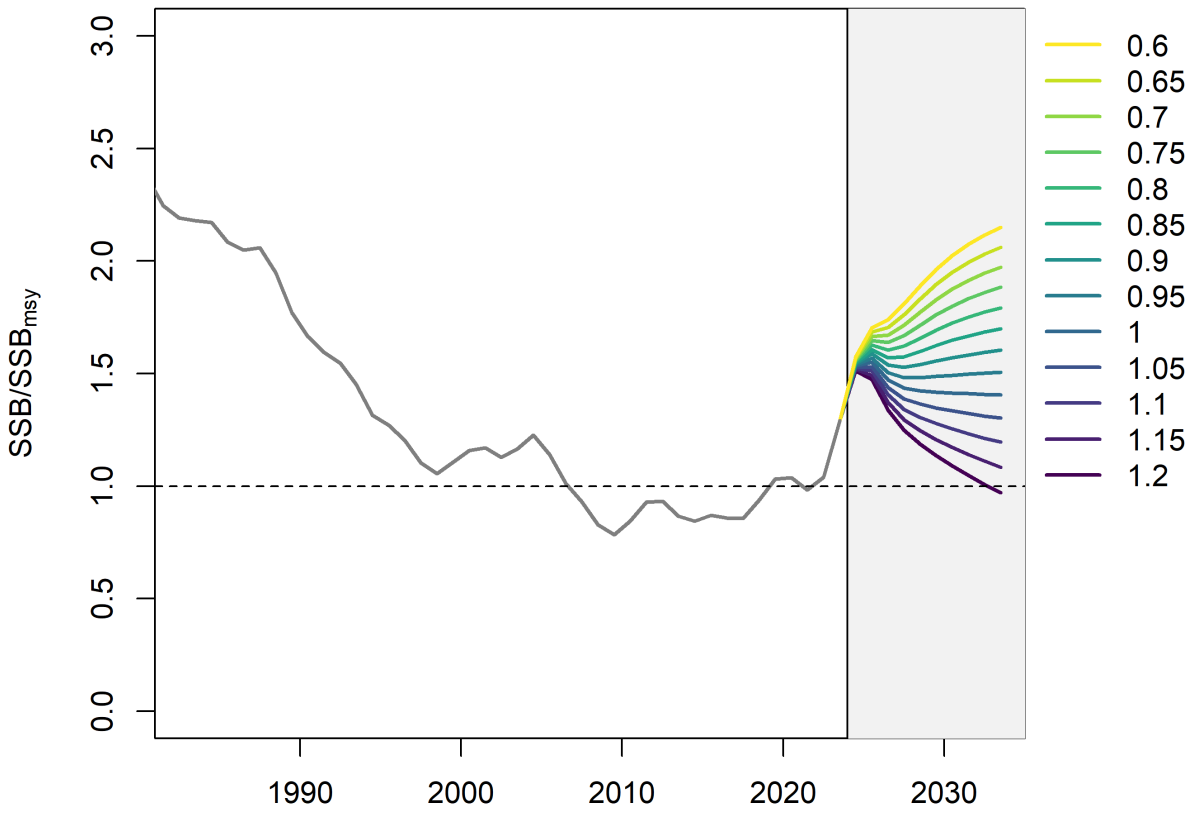


Figure 82: Changes in stock status ($SSB/SSB_{MSY}$) over the years produced by different catch levels expressed as multipliers of catch in 2023. These projections are developed scaling MSY, $SSB_{MSY}$, and recruitment to the recent 20 year average recruitment deviates. The white and gray boxes represent the model and projection period, respectively. Values are yearly averages.

## References

Andrews, A.H., Pacicco, A., Allman, R., Falterman, B.J., Lang, E.T., Golet, W., 2020. Age validation of yellowfin ( *Thunnus Albacares* ) and bigeye ( *Thunnus Obesus* ) tuna of the northwestern Atlantic Ocean. Canadian Journal of Fisheries and Aquatic Sciences 77, 637–643. https://doi.org/10.1139/cjfas-2019-0328

Artetxe-Arrate, I., Fraile, I., Lastra-Luque, P., Farley, J., Clear, N., Shahid, U., Adbul-Razzaue, S., Ahusan, M., Vidot, A., Parker, D., Marsac, F., Murua, H., Merino, G., Zudaire, I., in review. Otolith stable isotopes highlight the importance of local nursery areas as the origin of recruits to yellowfin tuna (Thunnus albacares) fisheries in the western Indian Ocean. Fisheries Research.

Artetxe-Arrate, I., Fraile, I., Marsac, F., Farley, J.H., Rodriguez-Ezpeleta, N., Davies, C.R., Clear, N.P., Grewe, P., Murua, H., 2021. A review of the fisheries, life history and stock structure of tropical tuna (skipjack Katsuwonus pelamis, yellowfin Thunnus albacares and bigeye Thunnus obesus) in the Indian Ocean, in: Advances in Marine Biology. Elsevier, pp. 39–89. https://doi.org/10.1016/bs.amb.2020.09.002

Artetxe-Arrate, I., Lastra-Luque, P., Fraile, I., Zudaire, I., Correa, G.M., Merino, G., Urtizberea, A., 2024. Natural mortality estimates of yellowfin tuna (Thunnus albacares) in the Indian Ocean (No. IOTC-2024-WPTT26(DP)-09). Indian Ocean Tuna Comission.

Cadigan, N.G., Farrell, P.J., 2005. Local influence diagnostics for the retrospective problem in sequential population analysis. ICES Journal of Marine Science 62, 256–265. https://doi.org/10.1016/j.icesjms.2004.11.015

Carvalho, F., Punt, A.E., Chang, Y.-J., Maunder, M.N., Piner, K.R., 2017. Can diagnostic tests help identify model misspecification in integrated stock assessments? Fisheries Research 192, 28–40. https://doi.org/10.1016/j.fishres.2016.09.018

Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., Kitakado, T., Yemane, D., Piner, K.R., 2021. A cookbook for using model diagnostics in integrated stock assessments. Fisheries Research 240, 105959.

Castillo-Jordan, C., Hampton, J., Ducharme-Barth, N.D., Xu, H., Vidal, T., Williams, P., Scott, F., Pilling, G., Hamer, P., 2021. Stock assessment of South Pacific albacore tuna (No. WCPFC-SC17-2021/SA-WP-02). Western and Central Pacific Fisheries Commission.

Chassot, E., 2014. Are there small yellowfin caught by purse seiners in free-swimming schools? (No. IOTC-2014-WPDCS10-INF05). Indian Ocean Tuna Comission.

Chassot, E., Assan, C., Esparon, J., Tirant, A., Delgado de Molina, A., Dewals, P., Augustin, E., Bodin, N., 2016. Length-weight relationships for tropical tunas caught with purse seine in the Indian Ocean: Update and lessons learned (No. IOTC-2016-WPDCS12-INF05). Indian Ocean Tuna Comission.

Correa, G.M., Artetxe-Arrate, I., Urtizberea, A., Merino, G., Zudaire, I., 2024a. Towards a conceptual model for yellowfin tuna in the Indian Ocean (No. IOTC-2024-WPTT26(DP)-17rev1). Indian Ocean Tuna Comission.

Correa, G.M., McGilliard, C.R., Ciannelli, L., Fuentes, C., 2021. Spatial and temporal variability in somatic growth in fisheries stock assessment models: Evaluating the consequences of misspecification. ICES Journal of Marine Science 78, 1900–1908. https://doi.org/10.1093/icesjms/fsab096

Correa, G.M., Uranga, J., Kaplan, D., Merino, G., Ramos, L., 2024b. Standardized catch per unit effort of yellowfin tuna in the Indian Ocean for the European purse seine fleet operating on floating objects (No. IOTC-2024-WPTT26(DP)-11rev1). Indian Ocean Tuna Comission.

Dammannagoda, S.T., Hurwood, D.A., Mather, P.B., 2008. Evidence for fine geographical scale heterogeneity in gene frequencies in yellowfin tuna (Thunnus albacares) from the north

Indian Ocean around Sri Lanka. Fisheries Research 90, 147–157. https://doi.org/10.1016/j.fishres.2007.10.006

Dortel, E., Sardenne, F., Bousquet, N., Rivot, E., Million, J., Le Croizier, G., Chassot, E., 2015. An integrated Bayesian modeling approach for the growth of Indian Ocean yellowfin tuna. Fisheries Research 163, 69–84. https://doi.org/10.1016/j.fishres.2014.07.006

Ducharme-Barth, N.D., Vincent, M.T., 2022. Focusing on the front end: A framework for incorporating uncertainty in biological parameters in model ensembles of integrated stock assessments. Fisheries Research 255, 106452. https://doi.org/10.1016/j.fishres.2022.106452

Duffy, L.M., Kuhnert, P.M., Pethybridge, H.R., Young, J.W., Olson, R.J., Logan, J.M., Goñi, N., Romanov, E., Allain, V., Staudinger, M.D., Abecassis, M., Choy, C.A., Hobday, A.J., Simier, M., Galván-Magaña, F., Potier, M., Ménard, F., 2017. Global trophic ecology of yellowfin, bigeye, and albacore tunas: Understanding predation on micronekton communities at ocean-basin scales. Deep Sea Research Part II: Topical Studies in Oceanography 140, 55–73. https://doi.org/10.1016/j.dsr2.2017.03.003

Eveson, P., Million, J., Sardenne, F., Le Croizier, G., 2012. Updated growth estimates for skipjack, yellowfin and bigeye tuna in the Indian Ocean using the most recent tag-recapture and otolith data (No. IOTC-2012-WPTT14-23). Indian Ocean Tuna Comission.

Farley, J.H., KrusicGolub, K., Eveson, P., Luque, P., Fraile, I., Artetxe-Arrate, I., Zudaire, I., Romanov, E., Shahid, U., Razzaque, S., Parker, D., Clear, N., Murua, H., Marsac, F., Merino, G., 2023. Updating the estimation of age and growth of yellowfin tuna (Thunnus albacares) in the Indian Ocean using otoliths (No. IOTC-2023-WPTT25-20). Indian Ocean Tuna Comission.

Farley, J., Krusic-Golub, K., Eveson, P., Luque, P., Clear, N., Fraile, I., Artetxe-Arrate, I., Zudaire, I., Vidot, A., Govinden, R., Ebrahim, A., Ahusan, M., Romanov, E., Shahid, U., Chassot, E., Bodin, N., Parker, D., Murua, H., Marsac, F., Merino, G., 2021. Estimating the age and growth of yellowfin tuna (Thunnus albacares) in the Indian Ocean from counts of daily and annual increments in otoliths (No. IOTC-2021-WPTT23-05_Rev1). Indian Ocean Tuna Comission.

Farley, J., Krusic-Golub, Eveson, P., Clear, N., Rouspard, F., Sanchez, C., Nicol, S., Hampton, J., 2020. Age and growth of yellowfin and bigeye tuna in the western and central Pacific Ocean from otoliths (No. WCPFC-SC16-2020/SC16-SA-WP-02). Western and Central Pacific Fisheries Commission.

Fomel, S., Claerbout, J.F., 2009. Guest Editors' Introduction: Reproducible Research. Computing in Science & Engineering 11, 5–7. https://doi.org/10.1109/MCSE.2009.14

Fonteneau, 2008. A working proposal for a Yellowfin growth curve to be used during the 2008 yellowfin stock assessment (No. IOTC-2008-WPTT-4). Indian Ocean Tuna Comission.

Fraile, I., Luque, P., Campana, S., Farley, J., Krusic-Golub, K., Clear, N., Eveson, J., Artetxe-Arrate, I., Zudaire, I., Murua, H., Merino, G., 2024. Age validation of yellowfin tuna Thunnus albacares in the Indian Ocean using post-peak bomb radiocarbon chronologies. Marine Ecology Progress Series 734, 91–104. https://doi.org/10.3354/meps14555

Francis, R.I.C.C., 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences 68, 1124–1138. https://doi.org/10.1139/f2011-025

Francis, R.I.C.C., 1992. Use of Risk Analysis to Assess Fishery Management Strategies: A Case Study using Orange Roughy ( *Hoplostethus Atlanticus* ) on the Chatham Rise, New Zealand. Canadian Journal of Fisheries and Aquatic Sciences 49, 922–930. https://doi.org/10.1139/f92-102

Froese, R., Pauly, D., 2024. FishBase.

Fu, D., 2020. Tag data processing for IOTC tropical tuna assessments (No. IOTC-2020-

WPTT22(DP)-10). Indian Ocean Tuna Comission.

Fu, D., 2017. Indian ocean skipjack tuna stock assessment 1950-2016 (Stock Synthesis) (No. IOTC-2017-WPTT19-47). Indian Ocean Tuna Comission.

Fu, D., Langley, A.D., Merino, G., Urtizberea, A., 2018. Preliminary Indian Ocean Yellowfin Tuna Stock Assessment 1950-2017 (Stock Synthesis) (No. IOTC–2018–WPTT20–33). Indian Ocean Tuna Comission.

Fu, D., Urtizberea Ijurco, A., Cardinale, M., Methot, R.D., Hoyle, S.D., Merino, G., 2021. Preliminary Indian Yellowfin tuna stock assessment 1950-2020 (Stock Synthesis) (No. IOTC-2021-WPTT23-12). Indian Ocean Tuna Comission.

Gaertner, D., Hallier, J.P., 2015. Tag shedding by tropical tunas in the Indian Ocean and other factors affecting the shedding rate. Fisheries Research 163, 98–105. https://doi.org/10.1016/j.fishres.2014.02.025

Geehan, J., 2018. Revision to the IOTC scientific estimates of Indonesia's fresh longline catches (No. IOTC-2018-WPDCS14-23). Indian Ocean Tuna Comission.

Geehan, J., Hoyle, S.D., 2013. Review of length frequency data of the Taiwanese Distant Water Longline Fleet (No. IOTC-2013-WPDCS09-12). Indian Ocean Tuna Comission.

Grewe, P., Feutry, P., Foster, S., Aulich, Lansdell, M., Cooper, S., Clear, N., Farley, J., Nikolic, N., Krug, I., Mendibil, I., Ahusan, M., Parker, D., Wudianto, Ruchimat, T., Satria, F., Lestari, P., Taufik, M., Fernando, D., Priatna, A., Zamroni, Rodriguez-Ezpeleta, N., Artetxe-Arrate, I., Fahmi, Z., Murua, H., Marsac, F., Davies, C., 2020. Genetic population connectivity of yellowfin tuna in the Indian Ocean from the PSTBS-IO Project (No. IOTC-2020-WPTT22(AS)12_REV1). Indian Ocean Tuna Comission, Seychelles.

Hamel, O.S., Cope, J.M., 2022. Development and considerations for application of a longevity-based prior for the natural mortality rate. Fisheries Research 256, 106477.

Hampton, S.E., Anderson, S.S., Bagby, S.C., Gries, C., Han, X., Hart, E.M., Jones, M.B., Lenhardt, W.C., MacDonald, A., Michener, W.K., Mudge, J., Pourmokhtarian, A., Schildhauer, M.P., Woo, K.H., Zimmerman, N., 2015. The Tao of open science for ecology. Ecosphere 6, 1–13. https://doi.org/10.1890/ES14-00402.1

Hampton, S.E., Strasser, C.A., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A.L., Duke, C.S., Porter, J.H., 2013. Big data and the future of ecology. Frontiers in Ecology and the Environment 11, 156–162. https://doi.org/10.1890/120103

Harley, S., 2011. Preliminary examination of steepness in tunas based on stock assessment results (No. WCPFC-SC7-2011/IP-08). Western and Central Pacific Fisheries Commission.

Herrera, M., 2010. Proposal for a system to assess the quality of fisheries statistics at the IOTC (No. IOTC-2010-WPDCS-06). Indian Ocean Tuna Comission.

Hillary, R., IOTC, S., Areso, J., 2008a. Reporting rate analyses for recaptures from Seychelles port for yellowfin, bigeye and skipjack tuna (No. IOTC-2008-WPTT-18). Indian Ocean Tuna Comission.

Hillary, R., Million, J., Anganuzzi, A., Areso, J., 2008b. Tag shedding and reporting rate estimates for Indian Ocean tuna using double-tagging and tag-seeding experiments (No. IOTC-2008-WPTDA-04). Indian Ocean Tuna Comission.

Hosseini, S.A., Kaymaram, F., 2016. Investigations on the reproductive biology and diet of yellowfin tuna, Thunnus albacares, (Bonnaterre, 1788) in the Oman Sea. Journal of Applied Icthyology 32, 310–317. https://doi.org/10.1111/jai.12907

Hoyle, S.D., 2024. Effort creep in longline and purse seine CPUE and its application in tropical tuna assessments (No. IOTC-2024-WPTT26(DP)-16). Indian Ocean Tuna Comission.

Hoyle, S.D., 2021b. Approaches for estimating natural mortality in tuna stock assessments: Application to Indian Ocean yellowfin tuna (No. IOTC-2021-WPTT23-08). Indian Ocean Tuna Comission.

Hoyle, S.D., 2021a. Review of size data from Indian Ocean longline fleets, and its utility for stock assessment (No. IOTC-2021-WPTT23-07). Indian Ocean Tuna Comission.

Hoyle, S.D., Campbell, R.A., Ducharme-Barth, N.D., Grüss, A., Moore, B.R., Thorson, J.T., Tremblay-Boyer, L., Winker, H., Zhou, S., Maunder, M.N., 2024. Catch per unit effort modelling for stock assessment: A summary of good practices. Fisheries Research 269, 106860. https://doi.org/10.1016/j.fishres.2023.106860

Hoyle, S.D., Langley, A.D., 2020. Scaling factors for multi-region stock assessments, with an application to Indian Ocean tropical tunas. Fisheries Research 228, 105586. https://doi.org/10.1016/j.fishres.2020.105586

Hoyle, S.D., Leroy, B.M., Nicol, S.J., Hampton, W.J., 2015. Covariates of release mortality and tag loss in large-scale tuna tagging experiments. Fisheries Research 163, 106–118. https://doi.org/10.1016/j.fishres.2014.02.023

Hoyle, S.D., Satoh, K., Matsumoto, T., 2017. Exploring possible causes of historical discontinuities in Japanese longline CPUE (No. IOTC-2017-WPTT19-33). Indian Ocean Tuna Comission.

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. International journal of forecasting 22, 679–688.

ICCAT, 2019. Report of the 2019 ICCAT yellowfin tuna stock assessment meeting. ICCAT (International Commission for the Conservation of Atlantic Tunas), Grand-Bassam, Cote d'Ivoire.

IOTC, S., 2024. Review of the statistical data available for yellowfin tuna (1950-2022) (No. IOTC-2024-WPTT26(DP)-07). Indian Ocean Tuna Comission.

IOTC, S., 2021. Review of Yellowfin Tuna Statistical Data (No. IOTC-2021-WPTT23(DP)-07_Rev1). Indian Ocean Tuna Comission.

Kaplan, D., Correa, G.M., Ramos, L., Duparc, A., Uranga, J., Santiago, J., Floch, L., Baez, J.C., Rojo, V., Pascual, P., Merino, G., 2024. Standardized CPUE abundance indices for adult yellowfin tuna caught in free-swimming school sets by the European purse-seine fleet in the Indian Ocean, 1991-2022 (No. IOTC-2024-WPTT26(DP)-13rev2).

Kell, L.T., Kimoto, A., Kitakado, T., 2016. Evaluation of the prediction skill of stock assessment using hindcasting. Fisheries Research 183, 119–127. https://doi.org/10.1016/j.fishres.2016.05.017

Kell, L.T., Sharma, R., Kitakado, T., Winker, H., Mosqueira, I., Cardinale, M., Fu, D., 2021. Validation of stock assessment methods: Is it me or my model talking? ICES Journal of Marine Science 78, 2244–2255. https://doi.org/10.1093/icesjms/fsab104

Kitakado, T., Wang, S.-P., Satoh, K., Lee, S.I., Tsai, W.-P., Matsumoto, T., Yokoi, H., Okamoto, D.K., Lee, M.K., Lim, J.-H., Kwon, Y., Su, N.-J., Chang, S.-T., Chang, F.-C., 2021. Report of trilateral collaborative study among Japan, Korea and Taiwan for producing joint abundance indices for the yellowfin tunas in the Indian Ocean using longline fisheries data up to 2019 (No. IOTC–2021-WPTT23(DP)-14). Indian Ocean Tuna Comission.

Kolody, D., 2018. Estimation of Indian Ocean Skipjack Purse Seine Catchability Trends from Bigeye and Yellowfin Assessments (No. IOTC–2018–WPTT20–32). Indian Ocean Tuna Comission.

Kolody, D., Herrera, M., Million, J., 2011. Indian Ocean Skipjack Tuna Stock Assessment 1950-2009 (Stock Synthesis) (No. IOTC-2011-WPTT13-31). Indian Ocean Tuna Comission.

Kolody, D., Hoyle, S.D., 2013. Evaluation of Tag Mixing Assumptions for Skipjack, Yellowfin and Bigeye Tuna Stock Assessments in the Western Pacific and Indian Oceans (No. WCPFC-SC9-2013/ SA-IP-11). Western and Central Pacific Fisheries Commission.

Krishnan, S., Antony Pillai, T., Chembian Antony Rayappan, J., Yagappan, T., Rajapandian, J., 2024. Diet composition and feeding habits of yellowfin tuna *Thunnus Albacares* (Bonnaterre,

1788) from the Bay of Bengal. Aquatic Living Resources 37, 10. https://doi.org/10.1051/alr/2024008

Kumar, M.S., Ghosh, S., 2022. Reproductive Dynamics of Yellowfin Tuna, Thunnus albacares (Bonnaterre 1788) Exploited from Western Bay of Bengal. Thalassas: An International Journal of Marine Sciences 38, 1003–1012. https://doi.org/10.1007/s41208-022-00429-1

Kunal, S.P., Kumar, G., Menezes, M.R., Meena, R.M., 2013. Mitochondrial DNA analysis reveals three stocks of yellowfin tuna Thunnus albacares (Bonnaterre, 1788) in Indian waters. Conservation Genetics 14, 205–213. https://doi.org/10.1007/s10592-013-0445-3

Lan, K.-W., Chang, Y.-J., Wu, Y.-L., 2020. Influence of oceanographic and climatic variability on the catch rate of yellowfin tuna (Thunnus albacares) cohorts in the Indian Ocean. Deep Sea Research Part II: Topical Studies in Oceanography 175, 104681. https://doi.org/10.1016/j.dsr2.2019.104681

Lan, K.-W., Evans, K., Lee, M.-A., 2013. Effects of climate variability on the distribution and fishing conditions of yellowfin tuna (Thunnus albacares) in the western Indian Ocean. Climatic Change 119, 63–77. https://doi.org/10.1007/s10584-012-0637-8

Langley, A.D., 2016. An update of the 2015 Indian Ocean Yellowfin Tuna stock assessment for 2016 (No. IOTC-2016-WPTT18-27). Indian Ocean Tuna Comission.

Langley, A.D., 2015. Stock assessment of yellowfin tuna in the Indian Ocean using Stock Synthesis (No. IOTC–2015–WPTT17–30). Indian Ocean Tuna Comission.

Langley, A.D., Fu, D., Maunder, M., 2023. An investigation of the recruitment dynamics of Indian Ocean yellowfin tuna (No. IOTC-2023-WPTT25-12). Indian Ocean Tuna Comission.

Langley, A.D., Hampton, J., Herrera, M., Million, J., 2008. Preliminary stock assessment of yellowfin tuna in the Indian Ocean using MULTIFAN-CL (No. IOTC-2008-WPTT-10). Indian Ocean Tuna Comission.

Langley, A.D., Herrera, M., Hallier, J.-P., Million, J., 2009. Stock assessment of yellowfin tuna in the Indian Ocean using MULTIFAN-CL (No. IOTC-2009-WPTT-10). Indian Ocean Tuna Comission.

Langley, A.D., Herrera, M., Million, J., 2012. Stock assessment of yellowfin tuna in the Indian Ocean using MULTIFAN-CL (No. IOTC-2012-WPTT-14-38 Rev_1). Indian Ocean Tuna Comission.

Langley, A.D., Herrera, M., Million, J., 2011. Stock assessment of yellowfin tuna in the Indian Ocean using MULTIFAN-CL (No. IOTC-2011-WPTT-13). Indian Ocean Tuna Comission.

Langley, A.D., Herrera, M., Million, J., 2010. Stock assessment of yellowfin tuna in the Indian Ocean using MULTIFAN-CL (No. IOTC-2010-WPTT-23). Indian Ocean Tuna Comission.

Langley, A.D., Million, J., 2012. Determining an appropriate tag mixing period for the Indian Ocean yellowfin tuna stock assessment (No. IOTC-2012-WPTT14-31). Indian Ocean Tuna Comission.

Lee, H.-H., Maunder, M.N., Piner, K.R., Methot, R.D., 2012. Can steepness of the stock–recruitment relationship be estimated in fishery stock assessment models? Fisheries Research 125–126, 254–261. https://doi.org/10.1016/j.fishres.2012.03.001

Lee, H., Piner, K.R., Taylor, I.G., Kitakado, T., 2019. On the use of conditional age at length data as a likelihood component in integrated population dynamics models. Fisheries Research 216, 204–211. https://doi.org/10.1016/j.fishres.2019.04.007

Lennert-Cody, C.E., Maunder, M.N., Aires-da-Silva, A., Minami, M., 2013. Defining population spatial units: Simultaneous analysis of frequency distributions and time series. Fisheries Research 139, 85–92. https://doi.org/10.1016/j.fishres.2012.10.001

Lennert-Cody, C.E., Minami, M., Tomlinson, P.K., Maunder, M.N., 2010. Exploratory analysis of spatial–temporal patterns in length–frequency data: An example of distributional regression trees. Fisheries Research 102, 323–326. https://doi.org/10.1016/j.fishres.2009.11.014

Lorenzen, K., 2005. Population dynamics and potential of fisheries stock enhancement: Practical theory for assessment and policy analysis. Philosophical Transactions of the Royal Society B: Biological Sciences 360, 171–189. https://doi.org/10.1098/rstb.2004.1570

Lorenzen, K., 1996. The relationship between body weight and natural mortality in juvenile and adult fish: A comparison of natural ecosystems and aquaculture. Journal of Fish Biology 49, 627–642. https://doi.org/10.1111/j.1095-8649.1996.tb00060.x

Magnusson, A., Millar, C.P., Sharma, R., 2022. Open and Reproducible Fisheries Science: Standardized workflows at ICES and FAO.

Mangel, M., MacCall, A.D., Brodziak, J., Dick, E.J., Forrest, R.E., Pourzand, R., Ralston, S., 2013. A perspective on steepness, reference points, and stock assessment. Canadian Journal of Fisheries and Aquatic Sciences 70, 930–940. https://doi.org/10.1139/cjfas-2012-0372

Matsumoto, T., Satoh, K., Tsai, W.-P., Wang, S.-P., Lim, J.-H., Park, H., Lee, S.I., 2024. Joint longline CPUE for yellowfin tuna in the Indian Ocean by the Japanese, Korean and Taiwanese longline fishery (No. IOTC-2024-WPTT26(DP)-14). Indian Ocean Tuna Comission.

Maunder, M., Minte-Vera, C.V., Langley, A.D., Howell, D., 2023. Independent review of recent IOTC yellowfin tuna assessment (No. IOTC-2023-WPTT25-13_Rev1). Indian Ocean Tuna Comission.

Maunder, M.N., Hamel, O.S., Lee, H.-H., Piner, K.R., Cope, J.M., Punt, A.E., Ianelli, J.N., Castillo-Jordán, C., Kapur, M.S., Methot, R.D., 2023. A review of estimation methods for natural mortality and their performance in the context of fishery stock assessment. Fisheries Research 257, 106489. https://doi.org/10.1016/j.fishres.2022.106489

McKechnie, S., Pilling, G., Hampton, J., 2017. Stock assessment of bigeye tuna in the western and central Pacific Ocean (No. WCPFC-SC13-2017/SA-WP-05). Western and Central Pacific Fisheries Commission.

Ménard, F., Lorrain, A., Potier, M., Marsac, F., 2007. Isotopic evidence of distinct feeding ecologies and movement patterns in two migratory predators (yellowfin tuna and swordfish) of the western Indian Ocean. Marine Biology 153, 141–152. https://doi.org/10.1007/s00227-007-0789-7

Merino, G., Urtizberea, A., Fu, D., Winker, H., Cardinale, M., Lauretta, M.V., Murua, H., Kitakado, T., Arrizabalaga, H., Scott, R., 2022. Investigating trends in process error as a diagnostic for integrated fisheries stock assessments. Fisheries Research 256, 106478.

Methot, R.D., 2019. Recommendations on the configuration of the Indian Ocean yellowfin tuna stock assessment model.

Methot, R.D., Wetzel, C.R., 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research 142, 86–99. https://doi.org/10.1016/j.fishres.2012.10.012

Millar, C.P., Magnusson, A., Kokkalis, A., Mosqueira, I., Umar, I., Parner, H., 2023. IcesTAF: Functions to support the ices transparent assessment framework.

Mohn, R., 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES Journal of Marine Science 56, 473–488.

Moore, B.R., Lestari, P., Cutmore, S.C., Proctor, C., Lester, R.J.G., 2019. Movement of juvenile tuna deduced from parasite data. ICES Journal of Marine Science 76, 1678–1689. https://doi.org/10.1093/icesjms/fsz022

Muhling, B.A., Lamkin, J.T., Alemany, F., García, A., Farley, J., Ingram, G.W., Berastegui, D.A., Reglero, P., Carrion, R.L., 2017. Reproduction and larval biology in tunas, and the importance of restricted area spawning grounds. Reviews in Fish Biology and Fisheries 27, 697–732. https://doi.org/10.1007/s11160-017-9471-4

Nishida, T., Shono, H., 2007. Stock assessment of yellowfin tuna (Thunnus albacares) in the Indian Ocean by the age structured production model(ASPM) analyses (No. IOTC-2007-

WPTT-12). Indian Ocean Tuna Comission.
Nishida, T., Shono, H., 2005. Stock assessment of yellowfin tuna (Thunnus albacares) resources in the Indian Ocean by the age structured production model (ASPM) analyses (No. IOTC-2005-WPTT-09). Indian Ocean Tuna Comission.
Nootmorn, P., Yakoh, A., Kawises, K., 2005. Reproductive biology of yellowfin tuna in the eastern Indian Ocean (No. IOTC-2005-WPTT-14). Indian Ocean Tuna Comission.
Pacicco, A.E., Allman, R.J., Lang, E.T., Murie, D.J., Falterman, B.J., Ahrens, R., Walter, J.F., 2021. Age and Growth of Yellowfin Tuna in the U.S. Gulf of Mexico and Western Atlantic. Marine and Coastal Fisheries 13, 345–361. https://doi.org/10.1002/mcf2.10158
Pacicco, A.E., Brown-Peterson, N.J., Murie, D.J., Allman, R.J., Snodgrass, D., Franks, J.S., 2023. Reproductive biology of yellowfin tuna (Thunnus albacares) in the northcentral U.S. Gulf of Mexico. Fisheries Research 261, 106620. https://doi.org/10.1016/j.fishres.2023.106620
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
Punt, A.E., 2019. Spatial stock assessment methods: A viewpoint on current issues and assumptions. Fisheries Research 213, 132–143. https://doi.org/10.1016/j.fishres.2019.01.014
Reglero, P., Tittensor, D.P., Álvarez-Berastegui, D., Aparicio-González, A., Worm, B., 2014. Worldwide distributions of tuna larvae: Revisiting hypotheses on environmental requirements for spawning habitats. Marine Ecology Progress Series 501, 207–224.
Regular, P.M., Robertson, G.J., Rogers, R., Lewis, K.P., 2020. Improving the communication and accessibility of stock assessment using interactive visualization tools. Canadian Journal of Fisheries and Aquatic Sciences 77, 1592–1600. https://doi.org/10.1139/cjfas-2019-0424
Roger, C., 1994. Relationships among yellowfin and skipjack tuna, their prey-fish and plankton in the tropical western Indian Ocean. Fisheries Oceanography 3, 133–141. https://doi.org/10.1111/j.1365-2419.1994.tb00055.x
Sabarros, P., Romanov, E., Bach, P., 2015. Vertical behavior and habitat preferences of yellowfin and bigeye tuna in the South West Indian Ocean inferred from PSAT tagging data (No. IOTC–2015–WPTT17–42 Rev_1). Indian Ocean Tuna Comission.
Shih, C.-L., Hsu, C.-C., Chen, C.-Y., 2014. First attempt to age yellowfin tuna, Thunnus albacares, in the Indian Ocean, based on sectioned otoliths. Fisheries Research 149, 19–23. https://doi.org/10.1016/j.fishres.2013.09.009
Stewart, I.J., Monnahan, C.C., 2017. Implications of process error in selectivity for approaches to weighting compositional data in fisheries stock assessments. Fisheries Research 192, 126–134. https://doi.org/10.1016/j.fishres.2016.06.018
Suzuki, Z., 1993. A review of the biology and fisheries for yellowfin tuna (Thunnus albacares) in the Western and Central Pacific Ocean (No. 36). FAO, Rome.
Taylor, I.G., Doering, K.L., Johnson, K.F., Wetzel, C.R., Stewart, I.J., 2021. Beyond visualizing catch-at-age models: Lessons learned from the R4ss package about software to support stock assessments. Fisheries Research 239, 105924. https://doi.org/10.1016/j.fishres.2021.105924
Then, A.Y., Hoenig, J.M., Hall, N.G., Hewitt, D.A., Handling editor: Ernesto Jardim, 2015. Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. ICES Journal of Marine Science 72, 82–92. https://doi.org/10.1093/icesjms/fsu136
Thorson, J.T., Rudd, M.B., Winker, H., 2019. The case for estimating recruitment variation in data-moderate and data-poor age-structured models. Fisheries Research 217, 87–97. https:

//doi.org/10.1016/j.fishres.2018.07.007

Urtizberea, A., Cardinale, M., Winker, H., Methot, R.D., Fu, D., Kitakado, T., Fernandez, C., Merino, G., 2020. Towards providing scientific advice for Indian Ocean yellowfin in 2020 (No. IOTC-2020-WPTT22(AS)-21). Indian Ocean Tuna Comission.

Urtizberea, A., Fu, D., Merino, G., Methot, R.D., Cardinale, M., Winker, H., Walter, J., Murua, H., 2019. Preliminary assessment of Indian Ocean yellowfin tuna 1950-2018 (Stock Synthesis, v3.30) (No. IOTC-2019-WPTT21-50). Indian Ocean Tuna Comission.

Wain, G., Guéry, L., Kaplan, D.M., Gaertner, D., 2021. Quantifying the increase in fishing efficiency due to the use of drifting FADs equipped with echosounders in tropical tuna purse seine fisheries. ICES Journal of Marine Science 78, 235–245. https://doi.org/10.1093/icesjms/fsaa216

Walter, J., Winker, H., 2020. Projections to create Kobe 2 Strategy Matrix using the multivariate log-normal approximation for Atlantic yellowfin tuna (No. SCRS/2019/145). ICCAT (International Commission for the Conservation of Atlantic Tunas).

Xu, H., Lennert-Cody, C.E., 2024. FishFreqTree: IATTC's regression tree R package for analyzing size frequency data.

Xu, H., Maunder, M., Minte-Vera, C., Valero, J.L., Lennert-Cody, C.E., 2024. Stock assessment of bigeye tuna in the eastern Pacific Ocean: 2024 benchmark assessment (No. SAC-15-02). Inter-American Tropical Tuna Commission, La Jolla, CA.

Zudaire, I., Artetxe-Arrate, I., Farley, J.H., Murua, H., Kukul, D., Vidot, A., Razzaque, S., Ahusan, M., Romanov, E., Eveson, P., Clear, N., Luque, P., Fraile, I., Bodin, N., Chassot, E., Govinden, R., Ebrahim, A., Shahid, U., Fily, T., Marsac, F., Merino, G., 2022. Preliminary estimates of sex ratio, spawning season, batch fecundity and length at maturity for Indian Ocean yellowfin tuna (No. IOTC-2022-WPTT24(DP)-09). Indian Ocean Tuna Comission.

Zudaire, I., Murua, H., Grande, M., Bodin, N., 2013. Reproductive potential of Yellowfin Tuna (Thunnus albacares) in the western Indian Ocean. Fishery Bulletin 111, 252–264. https://doi.org/10.7755/FB.111.3.4

# 14 Appendix

## 14.1 Clustering of size compositions

Three-based methods are useful techniques for studying structure in different data types and exploratory analyses, because minimal assumptions are made about the processes that generated the data. In fisheries science, these methods have been traditionally used for exploring large-scale spatial CPUE patterns. The regression tree algorithm (Lennert-Cody et al., 2013, 2010) uses recursive partitioning to search for hierarchical binary decision rules that divide the data into more homogeneous subgroups. The binary decision rules are selected to provide the greatest decrease in the heterogeneity of length composition data, which is measured based on the Kullback–Leibler divergence. The regression tree algorithm has become popular in stock assessment for defining fleets based on size frequency patterns over time and space (e.g., Xu et al. (2024)). Correa et al. (2024a) did an initial exploration of the IO yellowfin length frequency data from the longline and purse seine fisheries using this method. Below, we present an update of these analyses, which might be relevant for future revisions of the fleet structure used in the current assessment model as suggested by the review panel in 2023 (M. Maunder et al., 2023). These analyses were carried out using the *FishFreqTree* R package (Xu and Lennert-Cody, 2024).

We found seasonal patterns in the clustering analysis for the *LS* fishery (purse seine log school). The size structure in quarter 3 and 4 was similar, with a main mode at $\sim$ 45 cm and a second mode at $\sim$ 60 cm (Figure 83). In quarter 1, the mode was between 45 and 55 cm. In quarter 2, two clusters were identified, where one of them had a larger mode ($\sim$ 60 cm) than the other ($\sim$ 45 cm). Spatially, we did not identified spatial differences in the length structure for quarter 1, 3, and 4 (Figure 84). In quarter 2, we found a break at 0°, finding a smaller mode in the northern area.

In the *FS* fishery (purse seine free school), two different bimodal size structures were identified for quarter 1 and 2 (Figure 85), with larger fish in quarter 1. Two clusters were also identified in quarter 3, which were also found in quarter 4. One of these clusters had a large dominance of smaller fish with mode $\sim$ 45 cm. Spatially, no differences were found in quarter 1 and 2, but the large dominance of smaller fish was found in northern areas in quarter 3 and 4 (Figure 86).

The cluster analysis for the *LL* fishery (longline) explained less than 10% of the variance in the length frequency data (Figure 87). Two clusters were identified in quarter 1, which were also found in quarter 4. These clusters were similar, with a mode $\sim$ 125 cm and lengths from 75 to 160 cm. In quarter 2 and 3, three clusters were identified in each of them. One of these clusters had a dominant mode at $\sim$ 110 cm. A wide range of sizes were also observed for these quarters. Spatially, there was a clear break at 55°E (Figure 88).

### 14.1.1 Figures

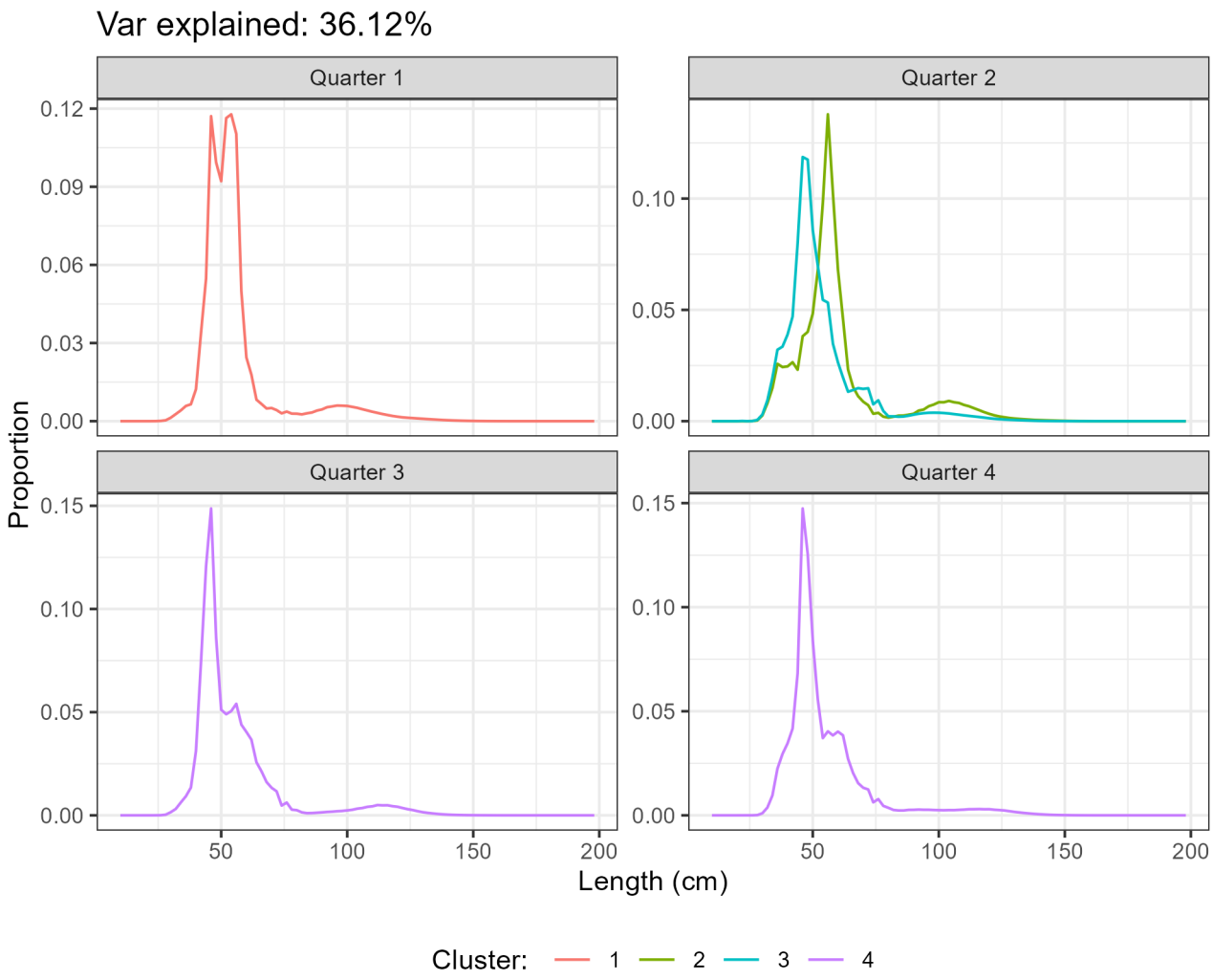


Figure 83: Clustering of length compositions for the LS fishery. Panels correspond to quarters. Four clusters were identified (colors).

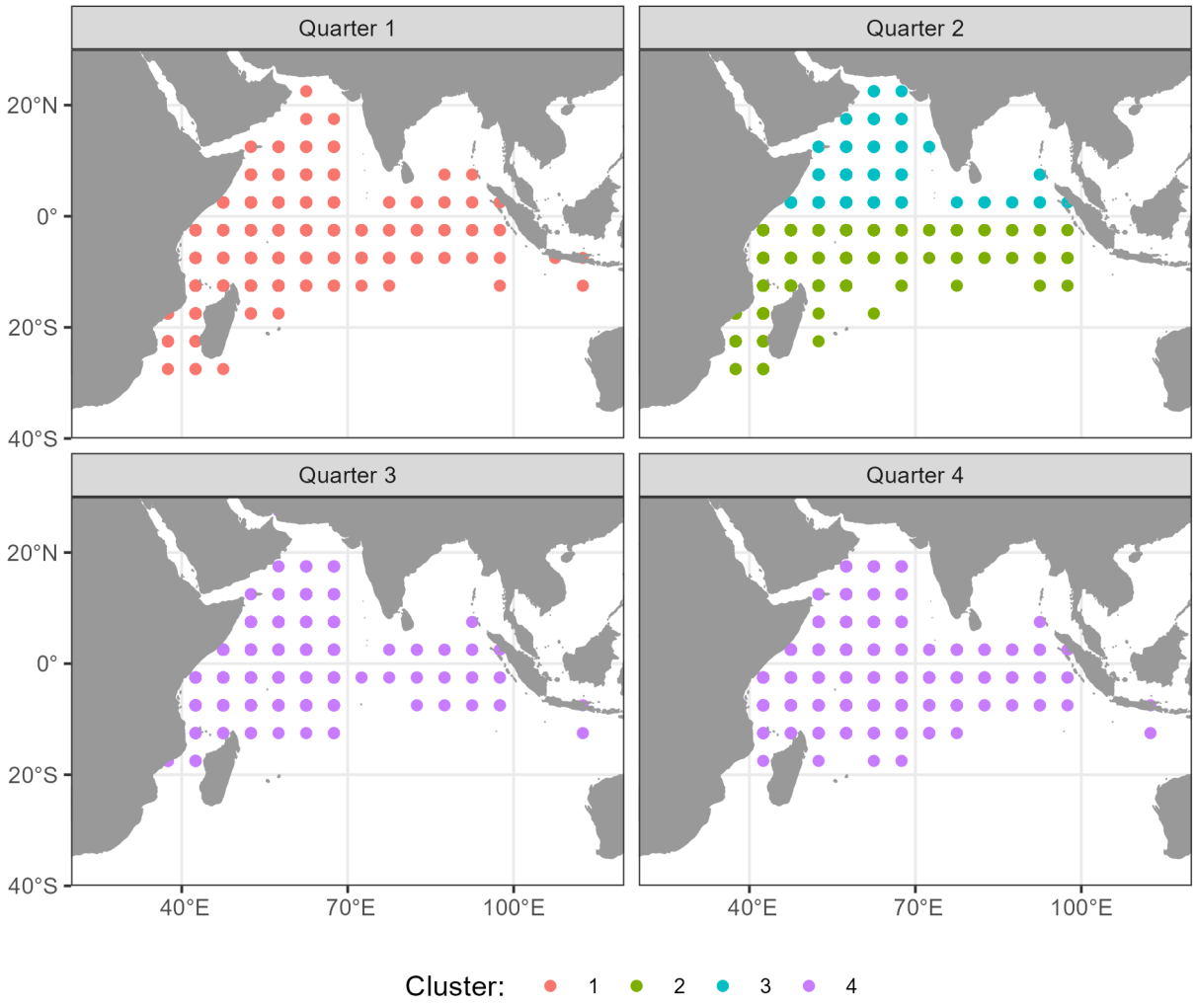


Figure 84: Clustering of length compositions for the LS fishery shown over space. Panels correspond to quarters.

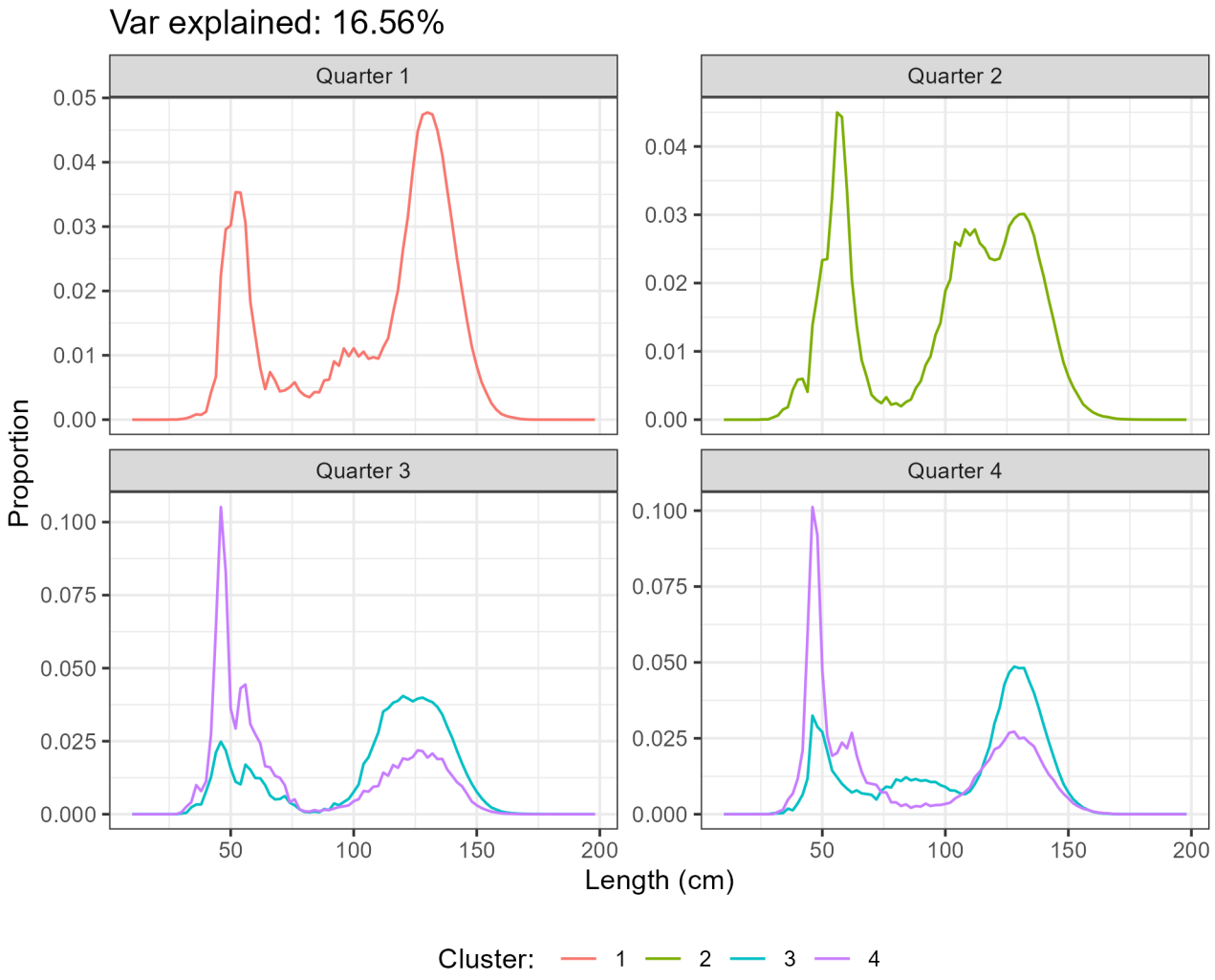


Figure 85: Clustering of length compositions for the FS fishery. Panels correspond to quarters. Four clusters were identified (colors).

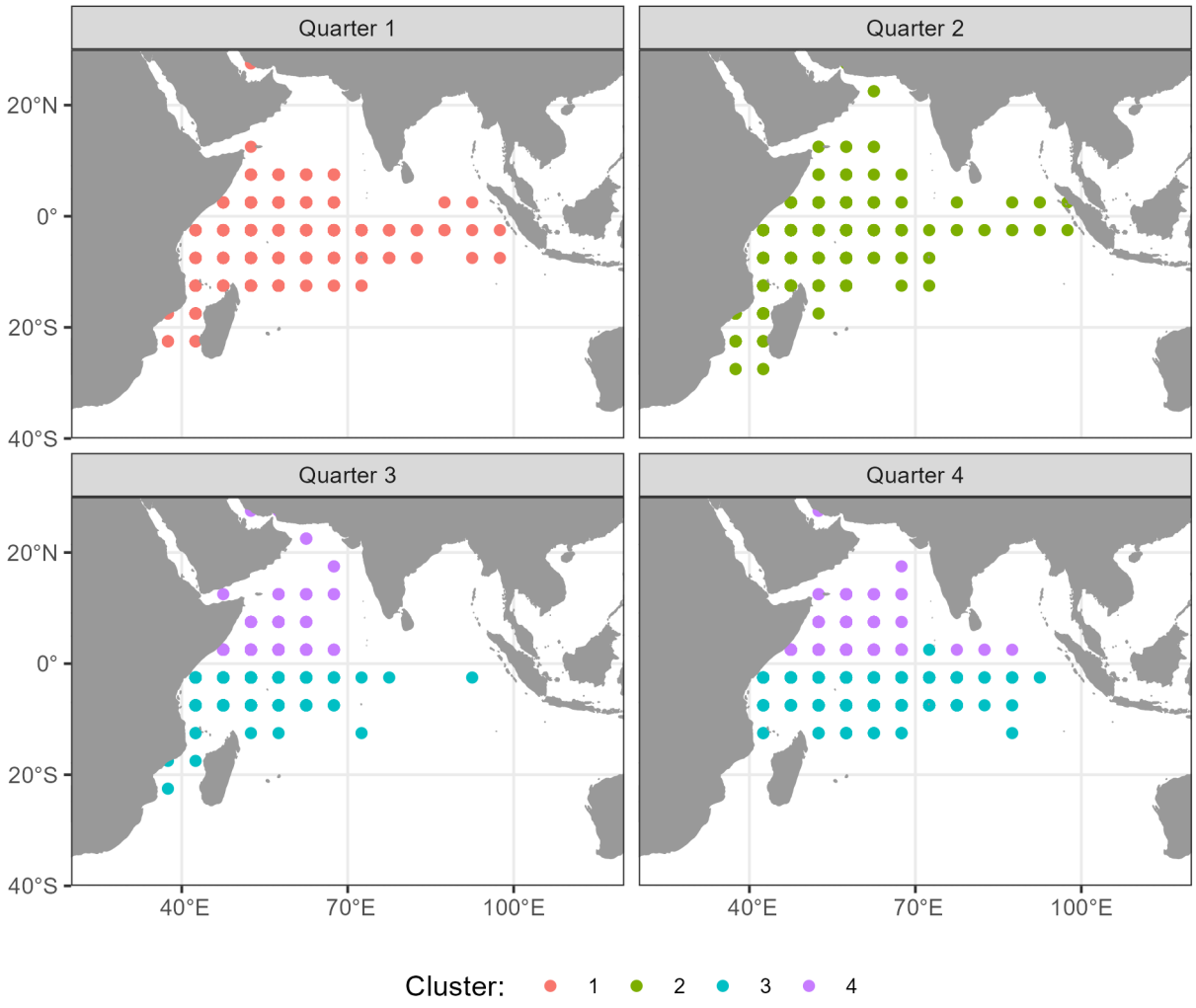


Figure 86: Clustering of length compositions for the FS fishery shown over space. Panels correspond to quarters.

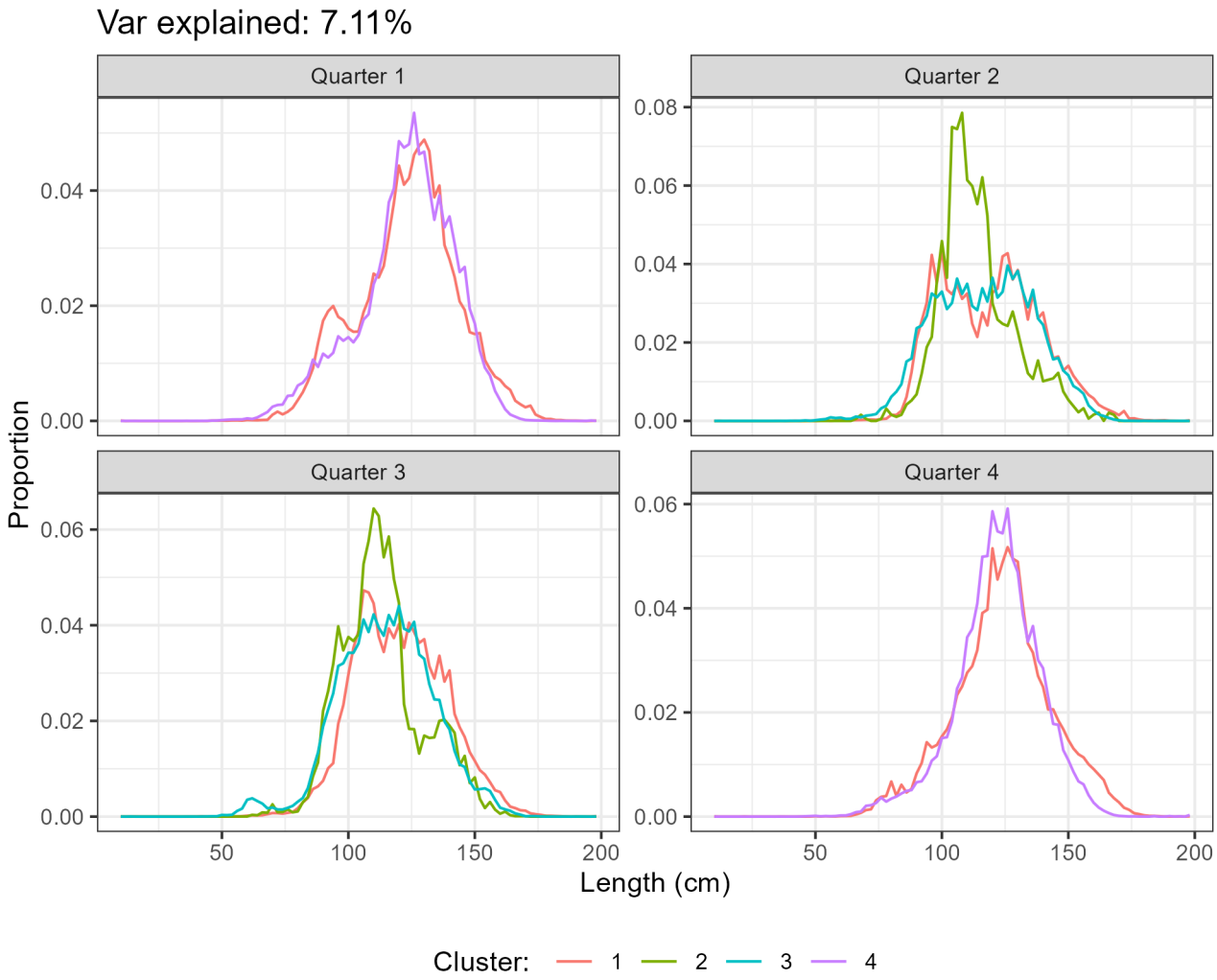


Figure 87: Clustering of length compositions for the LL fishery. Panels correspond to quarters. Four clusters were identified (colors).

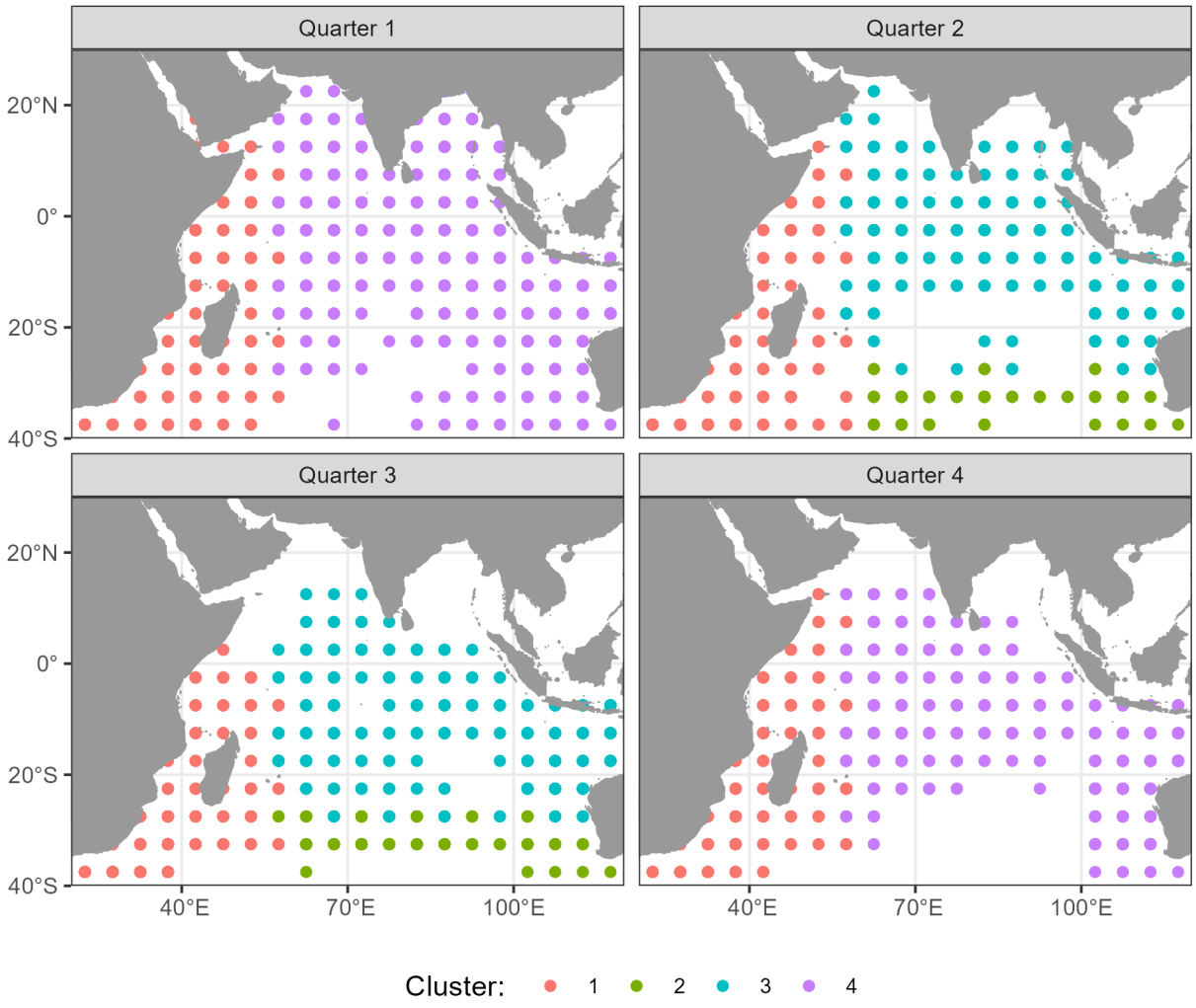


Figure 88: Clustering of length compositions for the LL fishery shown over space. Panels correspond to quarters.

## 14.2 One-area model

The data inputs were generated following the steps described for the four-area model. These models were implemented based on the assumptions of reference model 3, which models distinct LL fishery selectivity before and after 2000 (see Section 6.1 for details). An unique LL CPUE index was derived from the four LL indices after regional scaling, and then linked to the LL 1b selectivity.

We always modelled a single area in SS3 but evaluated two types of data aggregation (Figure 89):

- Aggregated (*agg*): Information per fishery group was aggregated, except for the Arabian sea (region 1a). The areas-as-fleets approach was implemented between region 1a and 1b. We had 13 fisheries (Table 9) and two indices of abundance: LL CPUE before and after 2000.
- Areas-as-fleets (*aaf*): Information per fishery group was kept separated as in the four-area configuration. The areas-as-fleets approach was implemented in the entire assessment area as shown in Figure 89. We had 24 fisheries (Table 10) and two indices of abundance: LL CPUE before and after 2000.

We compared three model configurations regarding the data types:

- *base*: include catch, CPUE, and length compositions.
- *addTag*: same as *base*, but also include tagging data downweighted by 90% (i.e., lambda of 0.1).
- *addCAAL*: same as *addTag*, but also include conditional age-at-length data.

### 14.2.1 Results

The time series of spawning biomass showed contrasting values among *agg* models, and were higher than the SSB estimated by the four-area model (Figure 90). Models that included tagging data tended to estimate lower SSB for the entire period, while models that included CAAL data estimated larger values. The three models showed a dramatic increase in SSB during the last few years. Regarding stock status ($SSB/SSB_{msy}$), the three models showed similar trends, starting at a value of $\sim$ 3.5 for the first model year, and ending at $\sim$ 1.45-1.75 for the last model year (Figure 91). Recruitment deviates did not display remarkable temporal trend during the model period; however, we did observe an increasing trend during the last decade (Figure 92).

The *aaf* models showed similar SSB time series before 2000; however, after 2000, the *addCAAL* and *addTag* models showed larger and lower SSB, respectively (Figure 90). Moreover, SSB was more similar but still higher than SSB estimates by the four-area model. Likewise, the time series of stock status showed similar values before 2000, and large stock status for the *addCAAL* model after 2000 (Figure 91). The stock status for the final model year ranged between 0.95 and 1.4. We noticed a temporal trend in recruitment deviates over the model period for the three *aaf* models (Figure 92), similar to the four-area models.

### 14.2.2 Tables

Table 9: Fishery definition in the one-area aggregated assessment configuration (agg).

| Fishery number | Fishery label |
|---|---|
| 1 | GI_1a |
| 2 | HD_1a |
| 3 | LL_1a |
| 4 | OT_1a |
| 5 | BB_1b |
| 6 | FS_1b |
| 7 | LL_1b_pre2000 |
| 8 | LS_1b |
| 9 | TR_1b |
| 10 | GI_1b |
| 11 | OT_1b |
| 12 | LF_1b |
| 13 | LL_1b_post2000 |

Table 10: Fishery definition in the one-area areas-as-fleets assessment configuration (aaf).

| Fishery number | Fishery label |
|---|---|
| 1 | GI_1a |
| 2 | HD_1a |
| 3 | LL_1a |
| 4 | OT_1a |
| 5 | BB_1b |
| 6 | FS_1b |
| 7 | LL_1b_pre2000 |
| 8 | LS_1b |
| 9 | TR_1b |
| 10 | LL_2_pre2000 |
| 11 | LL_3 |
| 12 | GI_4 |
| 13 | LL_4_pre2000 |
| 14 | OT_4 |
| 15 | TR_4 |
| 16 | FS_2 |
| 17 | LS_2 |
| 18 | TR_2 |
| 19 | FS_4 |
| 20 | LS_4 |
| 21 | LF_4 |
| 22 | LL_1b_post2000 |
| 23 | LL_2_post2000 |
| 24 | LL_4_post2000 |

### 14.2.3 Figures

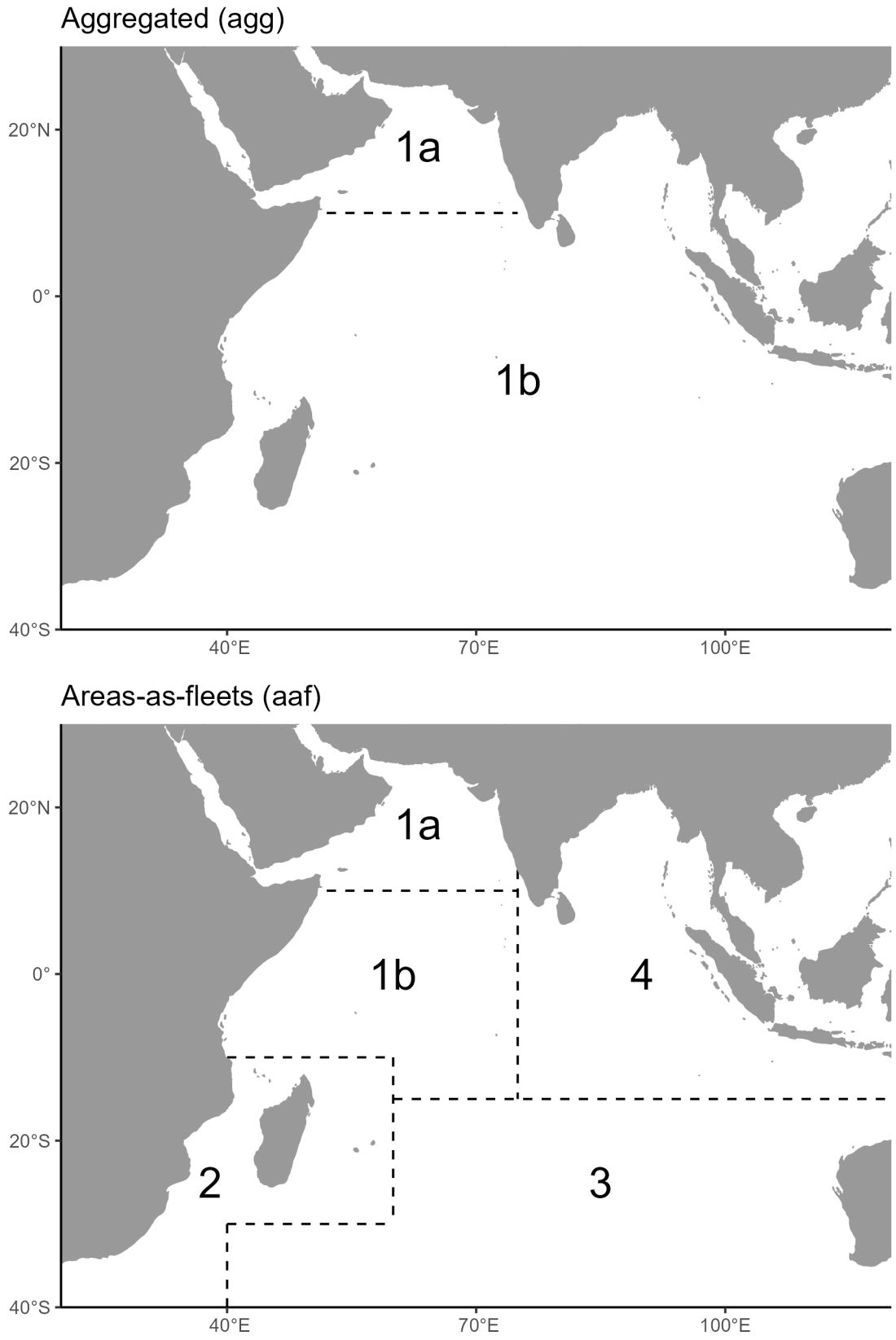


Figure 89: Types of aggregation explored for the one-area model configuration. The dashed line delimit the subregions where the areas-as-fleets approach was modelled.

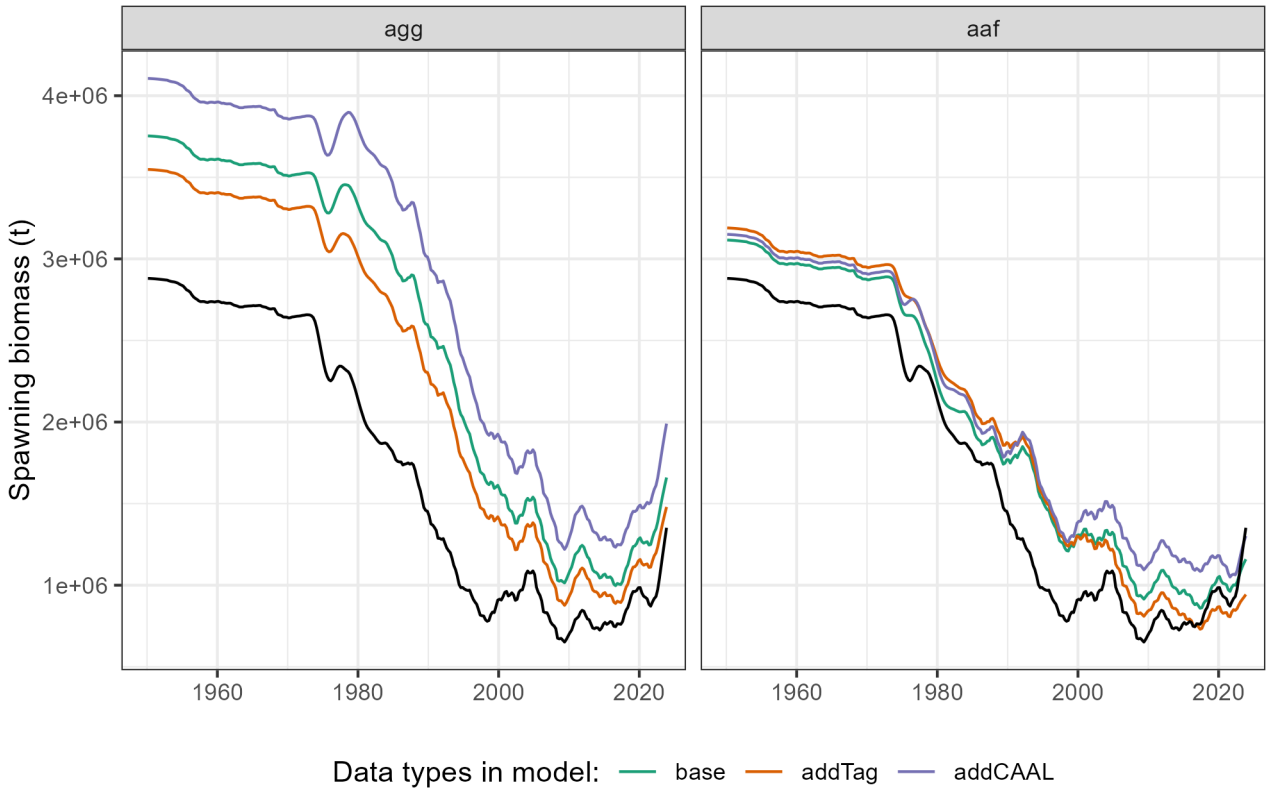


Figure 90: Time series of SSB estimated by the one-area model configurations. Estimates from the four-area configuration is shown in black.

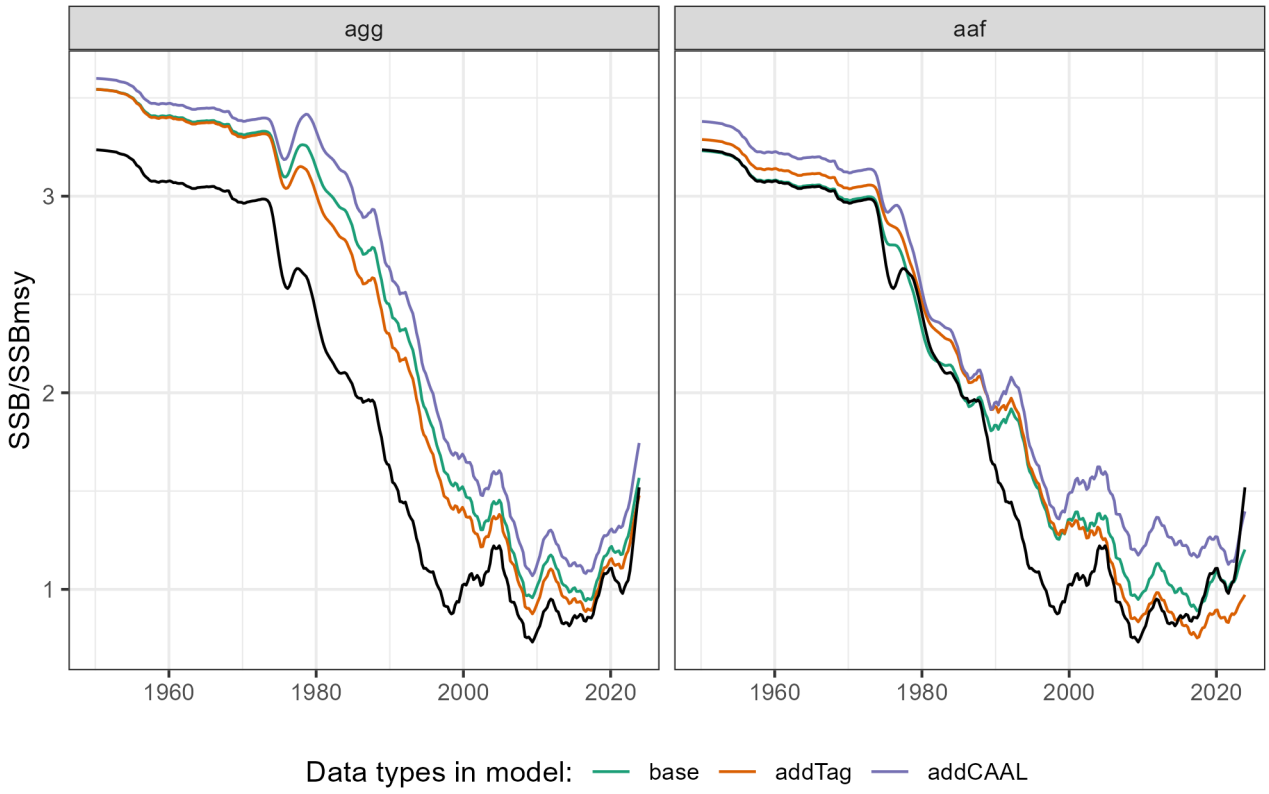


Figure 91: Time series of stock status (SSB/SSBmsy) estimated by the one-area model configurations. Estimates from the four-area configuration is shown in black.

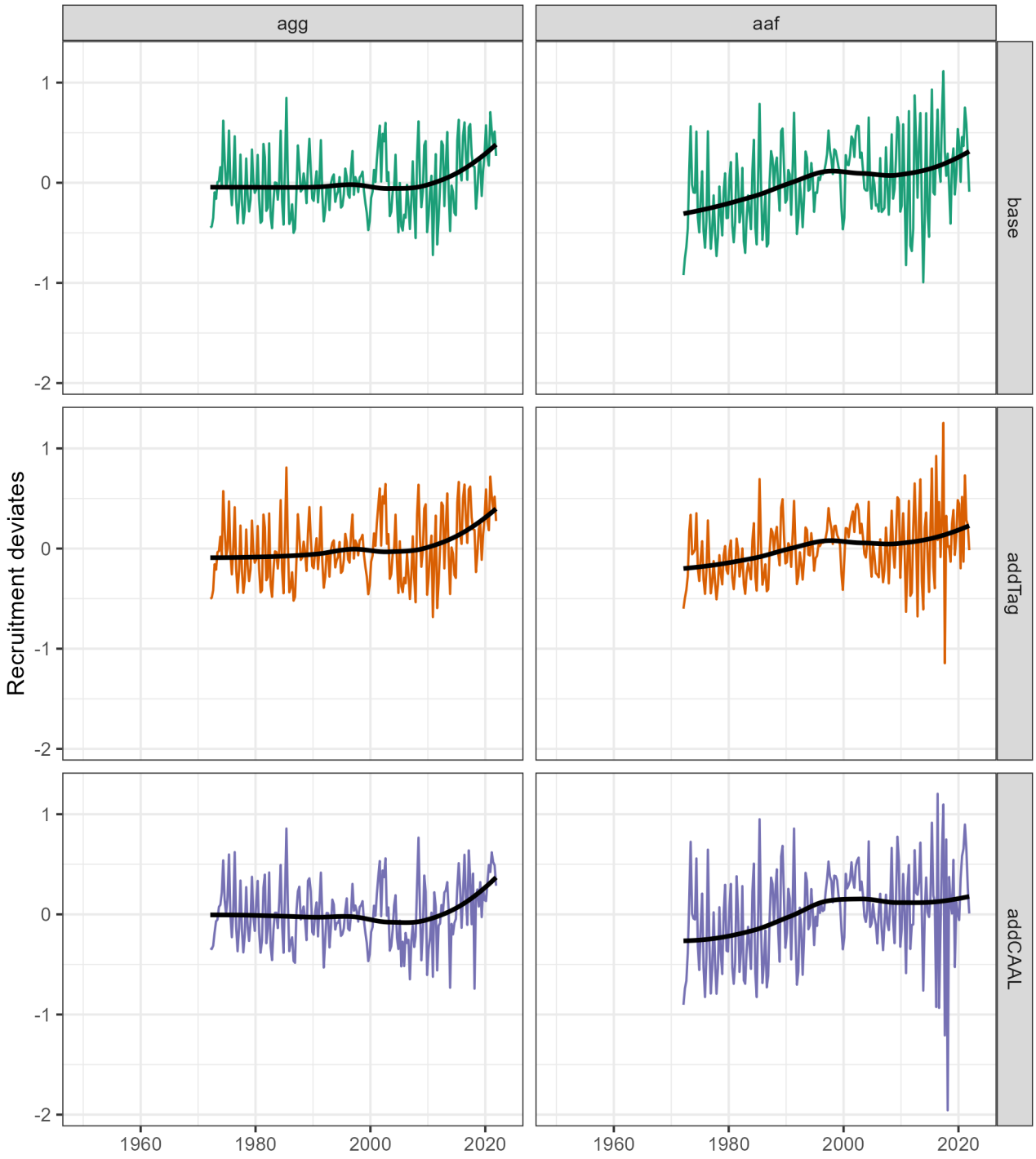


Figure 92: Time series of recruitment deviates estimated by the one-area model configurations. The black line is the smooth trend.

## 14.3 Two-area model

The data inputs were generated following the steps described for the four-area model. These models were implemented based on the assumptions of reference model 3, which models distinct LL fishery selectivity before and after 2000 (see Section 6.1 for details). Two LL CPUE indices ($LL1$ and $LL2$) were derived from the four LL indices after regional scaling: $LL1$ was the sum of LL 1 and LL 2 in the four-area model, and $LL2$ was the sum of LL 3 and LL 4 in the four-area model. $LL1$ was linked to the selectivity of fishery LL 1b, while $LL2$ was linked to the selectivity of LL 2 or LL 4, depending on the data aggregation type (see below). Movement between area 1 and 2 was modelled as in the four-area model.

We always modelled two areas in SS3 but evaluated two types of data aggregation (Figure 93):

- Aggregated (*agg*): Information per fishery group was aggregated, except for the Arabian sea (region 1a). The areas-as-fleets approach was implemented between region 1a and 1b. We had 18 fisheries (Table 11) and four indices of abundance: $LL1$ and $LL2$ CPUE before and after 2000.
- Areas-as-fleets (*aaf*): Information per fishery group was kept separated as in the four-area configuration. The areas-as-fleets approach was implemented in the entire assessment area as shown in Figure 93. We had 24 fisheries (Table 12) and four indices of abundance: $LL1$ and $LL2$ CPUE before and after 2000.

We compared three model configurations regarding the data types:

- *base*: include catch, CPUE, and length compositions.
- *addTag*: same as *base*, but also include tagging data downweighted by 90% (i.e., lambda of 0.1).
- *addCAAL*: same as *addTag*, but also include conditional age-at-length data.

### 14.3.1 Results

For *agg* models, the time series of spawning biomass showed quite similar values between *base* and *addTag* models, which were slightly higher than the SSB estimated by the four-area model (Figure 94). Models that included CAAL data tended to estimate much larger SSB for the entire period. The three models showed a dramatic increase in SSB from 2020. Regarding stock status ($SSB/SSB_{msy}$), the three models showed similar trends, starting at a value of $\sim$ 3.5 for the first model year, and ending at $\sim$ 1.65-1.8 for the last model year (Figure 95), which was slightly higher than the four-area model. Recruitment deviates for *base* and *addTag* models displayed a remarkable temporal trend during the model period; however, this trend was less evident for the *addCAAL* model. The three models showed an increasing trend during the last decade (Figure 96).

The *aaf* models showed similar SSB temporal trends. The *addCAAL* and *addTag* models showed larger and lower SSB, respectively (Figure 94), and the three models displayed larger values than the four-area model. Likewise, the time series of stock status showed similar trend and values. The stock status for the final model year ranged between 1.6 and 1.8, which was slightly larger than the four-area model (Figure 95). We noticed a temporal trend in recruitment deviates over the model period for the three *aaf* models (Figure 96), similar to the four-area models, although this trend was less pronounced for the *addCAAL* model.

### 14.3.2 Tables

Table 11: Fishery definition in the two-area aggregated assessment configuration (agg).

| Fishery number | Fishery label |
|---|---|
| 1 | GI_1a |
| 2 | HD_1a |
| 3 | LL_1a |
| 4 | OT_1a |
| 5 | BB_1b |
| 6 | FS_1b |
| 7 | LL_1b_pre2000 |
| 8 | LS_1b |
| 9 | TR_1b |
| 10 | LL_2_pre2000 |
| 11 | GI_2 |
| 12 | OT_2 |
| 13 | TR_2 |
| 14 | FS_2 |
| 15 | LS_2 |
| 16 | LF_2 |
| 17 | LL_1b_post2000 |
| 18 | LL_2_post2000 |

Table 12: Fishery definition in the two-area areas-as-fleets assessment configuration (aaf).

| Fishery number | Fishery label |
|---|---|
| 1 | GI_1a |
| 2 | HD_1a |
| 3 | LL_1a |
| 4 | OT_1a |
| 5 | BB_1b |
| 6 | FS_1b |
| 7 | LL_1b_pre2000 |
| 8 | LS_1b |
| 9 | TR_1b |
| 10 | LL_2_pre2000 |
| 11 | LL_3 |
| 12 | GI_4 |
| 13 | LL_4_pre2000 |
| 14 | OT_4 |
| 15 | TR_4 |
| 16 | FS_2 |
| 17 | LS_2 |
| 18 | TR_2 |
| 19 | FS_4 |
| 20 | LS_4 |

Table 12: Fishery definition in the two-area areas-as-fleets assessment configuration (aaf).

| Fishery number | Fishery label |
|---|---|
| 21 | LF_4 |
| 22 | LL_1b_post2000 |
| 23 | LL_2_post2000 |
| 24 | LL_4_post2000 |

### 14.3.3 Figures

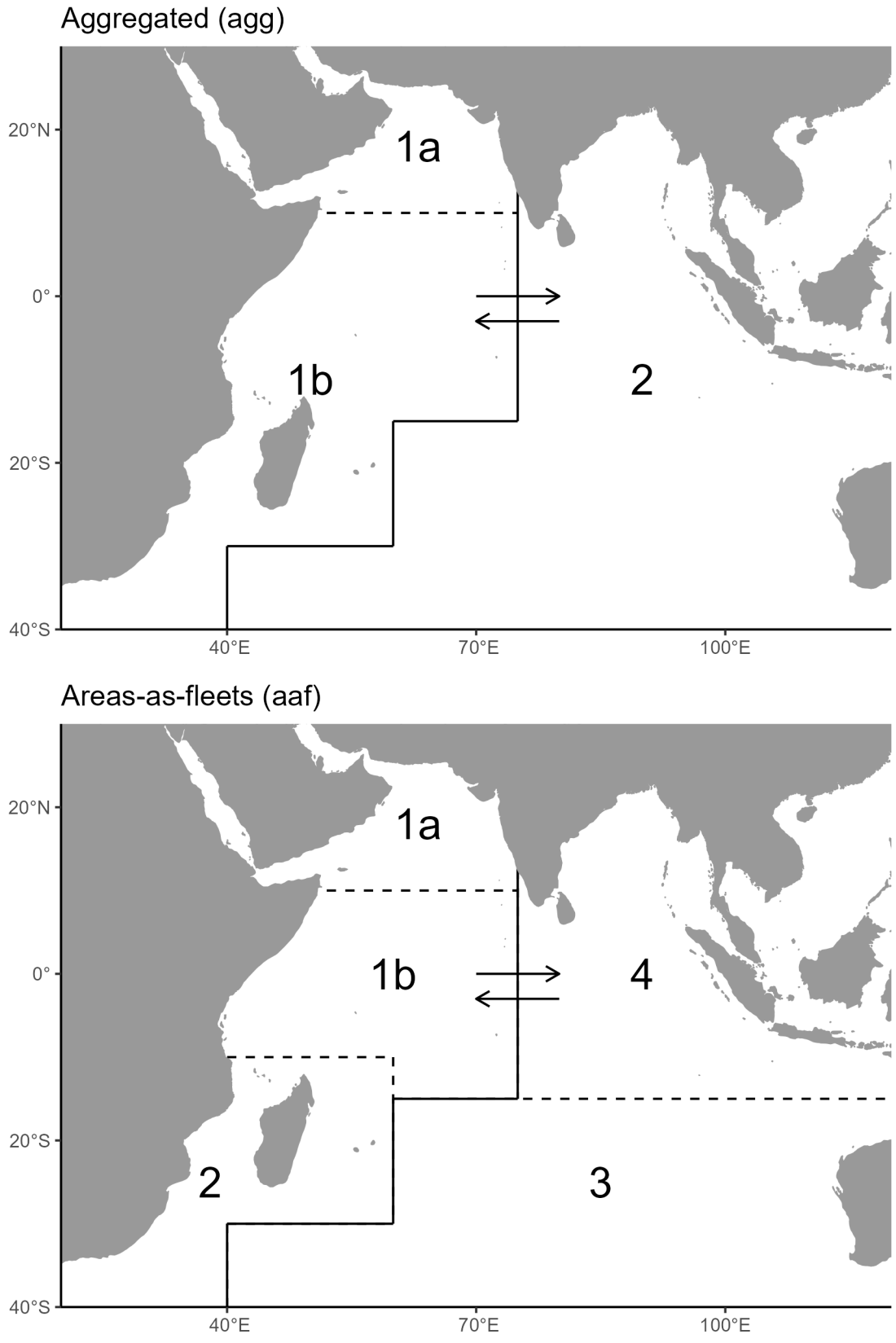


Figure 93: Types of aggregation explored for the two-area model configuration. The dashed line delimit the subregions where the areas-as-fleets approach was modelled.

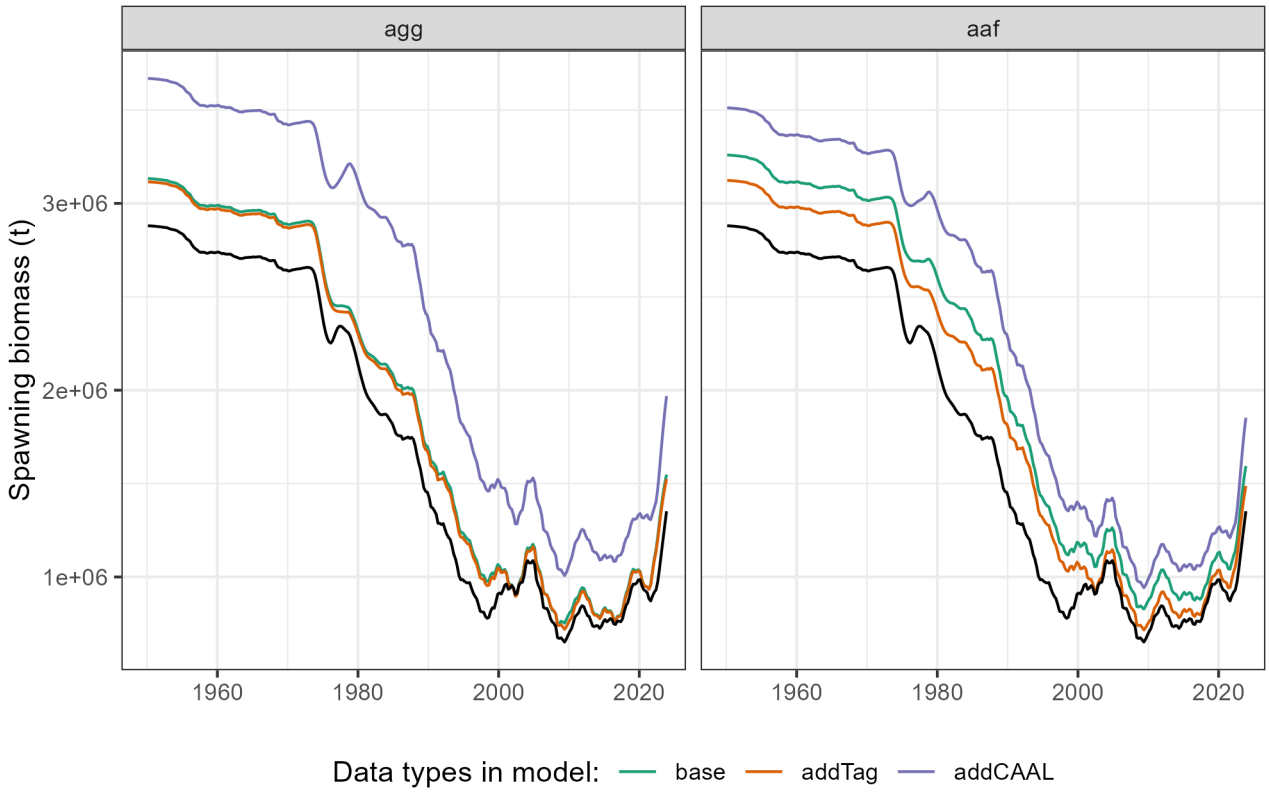


Figure 94: Time series of SSB estimated by the two-area model configurations. Estimates from the four-area configuration is shown in black.

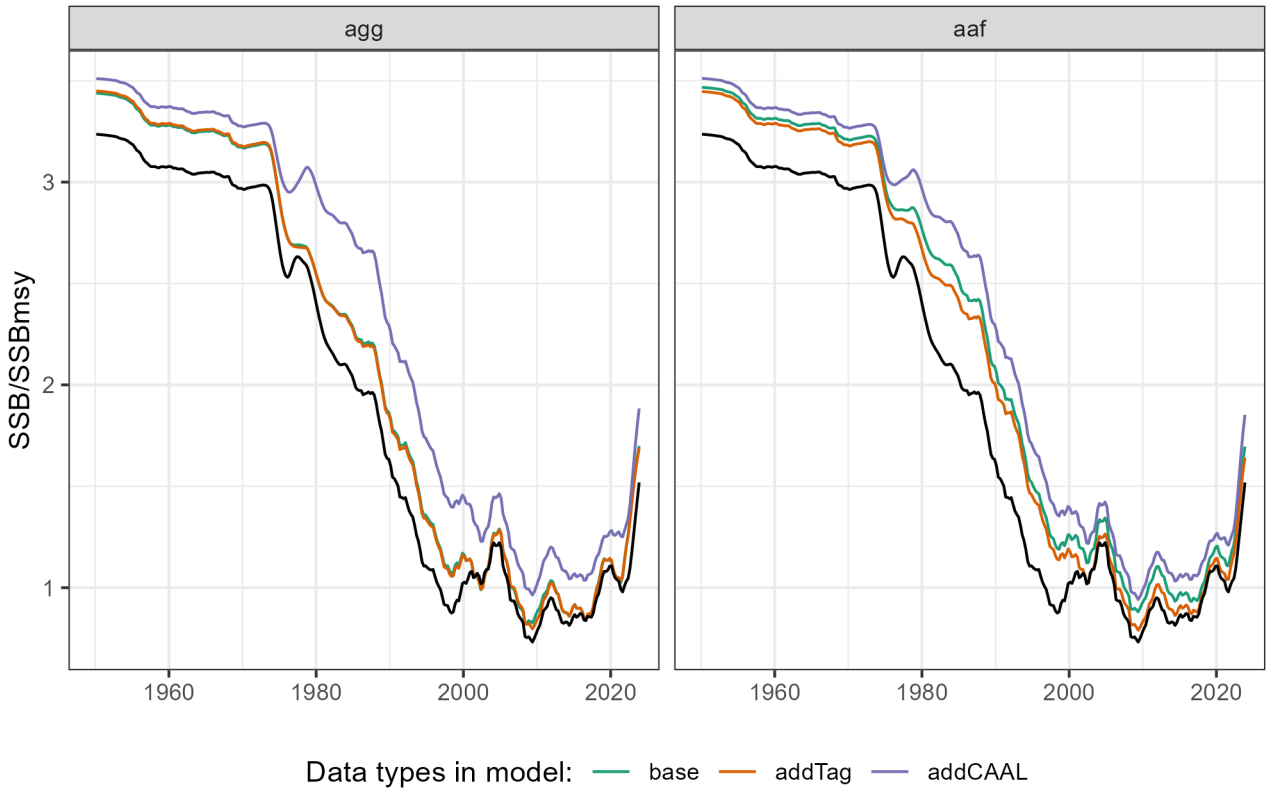


Figure 95: Time series of stock status (SSB/SSBmsy) estimated by the two-area model configurations. Estimates from the four-area configuration is shown in black.

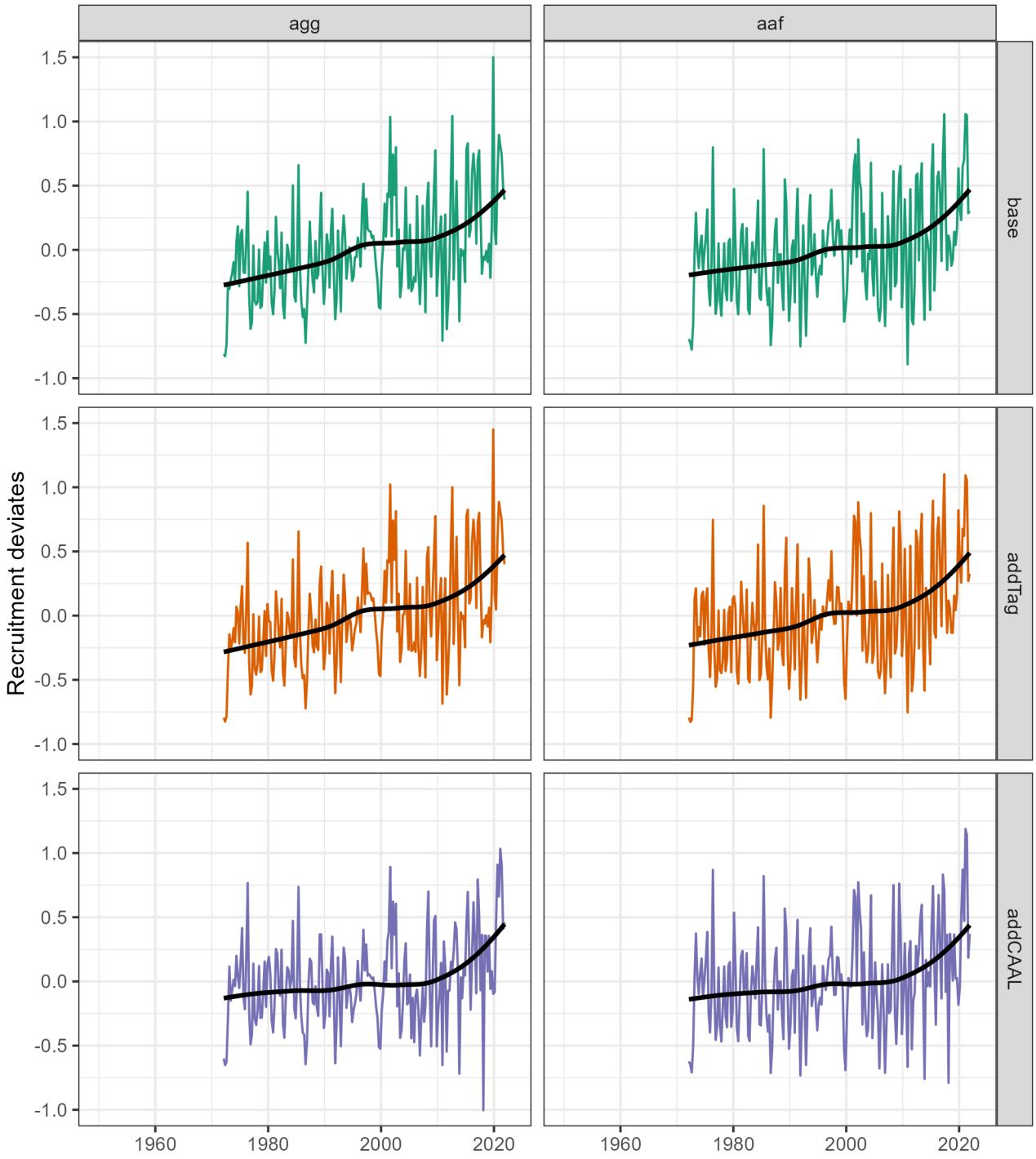


Figure 96: Time series of recruitment deviates estimated by the two-area model configurations. The black line is the smooth trend.