Indian Ocean Tuna Commission
Commission des Thons de l'Ocean Indien

IOTC–2026–WPTmT_10-02_Rev1

Stock assessment of albacore tuna (*Thunnus alalunga*) in the Indian Ocean using Stock Synthesis.

10th IOTC Working Party on Temperate Tuna
20 - 27 July 2026
Virtual Meeting

Joel Rice[1], Genevieve Phillips[2]

[1] Joel Rice Consulting. (ricemarineanalytics@gmail.com)
[2] IOTC Secretariat

Changes made to this assessment report as IOTC–2026–WPTmT10-02_Rev1

The following revisions have been incorporated into this version of the assessment report:

- Expanded diagnostic analyses for both the northwest (NW) and southwest (SW) diagnostic models, including additional CPUE diagnostics based on *leave-one-out* and *let-one-in* analyses, together with jitter analyses to further evaluate model stability and the influence of individual abundance indices.
- Addition of Annex 4, which documents the evaluation of alternative time-varying selectivity formulations for both the NW and SW models, including comparisons of model fit, stock-status estimates, and residual diagnostics.
- Minor editorial revisions throughout the document to correct typographical errors and improve clarity, with no substantive changes to the assessment methodology beyond those described above.

Executive summary

This paper presents a stock assessment of albacore tuna in the Indian Ocean using Stock Synthesis (version 3.30.24.2 https://nmfs-ost.github.io/ss3-website/). The albacore tuna assessment model is an age structured (14 years), spatially aggregated (1 region) and two sex model. The catch, effort, and size composition of catch, are grouped into 23 fisheries covering the time period from 1950 through 2023. Fifteen indices of abundance, fourteen of which are from longline fisheries were considered for this analysis.

Albacore tuna are most often caught in long line fisheries in the Indian Ocean tuna fisheries, though some bycatch occurs in the purse seine fisheries as well as other mixed gear fisheries. This analysis was developed based on the 2019 ,2022 and 2025 assessments along with updates to the data and parameterization. Two diagnostic case models, are referred to in the main text when presenting the model parametrization and diagnostics. The upcoming 10th meeting of the Indian Ocean Tuna Commission Working Party on Temperate Tuna (WPTmT10) will recommend the final parameterization as a base case model for the provision of stock status.

This assessment is a continuation of assessment work conducted by the WPTmT in 2025, as endorsed by the IOTC SC (IOTC-2025-SC28-R) to progress and provide updated management advice to the SC in 2026. Although the stock status reference points are expected to change in relationship to the update indices of abundance, preliminary estimated abundance trends show a decreasing stock throughout the time frame of the model, and spawning stock abundance has decreased to approximately between 1-1.5 times SSBMSY. The fishing mortality in the terminal year of the assessment is dependent on the indices of abundance selected with F2023/FMSY ranging between 0.56 and 1.264.

Key Findings and Recommendations

Preliminary recommendation for the reference model

Two diagnostic model configurations were developed to evaluate alternative interpretations of the northwest and southwest longline CPUE indices. At the time of writing, no final reference model has been selected. The Working Party will review the diagnostic model performance, sensitivity analyses, and biological plausibility before recommending a reference model for each diagnostic case and for the final stock status assessment.

Suggested model configuration

Sensitivity analyses were performed to evaluate alternative assumptions regarding stock-recruitment steepness, growth variability, length-weight relationships, maturity, natural mortality, likelihood weighting of length-composition data, terminal selectivity, and selected forms of time-varying selectivity. The diagnostic models provide the basis for selecting the most appropriate configuration for the reference assessment.

The NW diagnostic model exhibits stable convergence, limited retrospective bias, and a well-defined likelihood profile for population scale, but the residual diagnostics indicate persistent temporal structure in the quarterly NW CPUE and LL1 length-composition series. Recent time variation in LL1 selectivity is strongly supported relative to the static-selectivity model. Models allowing either the selectivity peak or ascending width to vary produce nearly equivalent improvements in fit and similar estimates of spawning biomass and stock status. A model allowing both parameters to vary provides a slightly lower AIC and improves the residual pattern for one additional fleet, but the additional improvement is small, is not significant in nested likelihood-ratio tests, and produces a materially more pessimistic estimate of current stock status. The provisional recommendation is therefore to retain a single time-varying LL1 selectivity parameter in the reference configuration, while presenting the combined peak-and-ascending-width model as a key structural sensitivity for Working Party consideration.

The SW diagnostic model provides a generally stable and parsimonious basis for the reference assessment. Time-varying selectivity was evaluated for fleets 4–8, which exhibited the clearest temporal patterns in the mean-length residual runs tests. Allowing either the selectivity peak or ascending width to vary improved the runs-test outcome for some fleets and produced modest changes in estimated spawning biomass. However, the improvements in likelihood were small relative to the additional model complexity, and none of the nested likelihood-ratio tests supported inclusion of the additional time-varying parameters. Allowing both parameters to vary provided essentially no improvement over the ascending-width-only formulation. The provisional recommendation is therefore to retain constant selectivity in the SW reference model and present the ascending-width formulation as a key structural sensitivity illustrating the possible effect of temporally changing selectivity on biomass and stock-status estimates.

The NW and SW diagnostics provide different levels of support for time-varying selectivity. For the NW model, time variation in the selectivity parameters substantially improves the likelihood and is strongly supported relative to static selectivity. For the SW model, targeted time variation for fleets 4–8 improves selected residual diagnostics but is not supported by likelihood-ratio tests after accounting for the additional parameters. Accordingly, time-varying selectivity is recommended as part of the provisional NW configuration but retained as a structural sensitivity for the SW model.

Principal model sensitivities

The sensitivity analyses demonstrated that the assessment is generally robust to alternative assumptions regarding biological parameters, although some assumptions produced measurable changes in estimated stock status and population scale.

- Alternative assumptions regarding stock-recruitment steepness primarily affected estimates of virgin biomass and historical spawning biomass while having comparatively smaller effects on recent relative stock status.

- Alternative parameterizations of growth, length-weight relationships, maturity-at-length, and natural mortality produced moderate changes in biomass trajectories but did not substantially alter the overall interpretation of historical stock trends.

- Increasing the weight assigned to longline length-composition data improved the fit to observed size compositions, while highlighting trade-offs between fitting CPUE indices and length-composition information.

- Time-varying selectivity for selected longline fleets substantially improved the statistical fit to the length-composition data, indicating evidence for temporal changes in selectivity. However, further evaluation is required to determine the most appropriate parameterization and to assess the implications for biological quantities and stock status estimates.

Overall, preliminary results indicate a declining spawning biomass throughout the assessment period, although estimates of current biomass relative to reference points and fishing mortality relative to FMSY remain sensitive to the choice of abundance indices and selected model configuration.

Recommendations for future assessments (2028)

The present assessment identified several areas where additional research and data development could improve future assessments:

- Continue evaluation of alternative CPUE standardization methods and index selection for both the northwest and southwest fisheries.
- Improve characterization of temporal changes in fishery selectivity through evaluation of time-varying selectivity parameterizations and supporting fishery information.

- Continue investigation of conflicts between CPUE indices and length-composition data, including alternative likelihood weighting approaches.
- Improve biological parameter estimates, particularly growth, maturity, and natural mortality, through incorporation of new biological sampling and regional studies.
- Continue development and evaluation of spatially explicit and environmentally informed assessment approaches where supported by available data.
- Re-evaluate model structural assumptions and biological reference points as new data become available prior to the 2028 assessment.

1 Introduction

Commercial fisheries for albacore tunas have operated in the Indian Ocean since the early 1950s. The earliest known exploitation was by the Japanese longline fishery in the 1950s, followed by the Korean and Taiwanese longline fisheries in the mid and late 1950s (Kim et al. 2010; Chen 2009). Drift nets were employed in the albacore fishery from the early-1980s until 1992 when an international ban on drift net fishing came into force. Taiwanese and Indonesian longline catch has recently accounted for around 70% of the total catch. Between 2008 and 2011, following the onset of piracy in waters off Somalia, part of the longline fleets that had traditionally targeted tropical tunas or swordfish in those waters moved towards albacore fishing grounds in the southern part of the Eastern Indian Ocean.

Like albacore fisheries in other oceans, the Indian Ocean fishery is characterized by smaller fish at higher latitudes. Unlike other oceans however, there is no significant troll or pole and line fishery for albacore, and since the ban on the drift net fishery there has been no large-scale targeting of small fish.

Stock assessments of the Indian Ocean albacore stock have been conducted in the past using several different methods, including the non-equilibrium production model ASPIC (Chang et al. 2012, Matsumoto et al. 2012, Matsumoto et al. 2014, Matsumoto 2016), the age-structured production model ASPM (Nishida et al. 2012, Nishida et al. 2016), a Bayesian biomass dynamic model (Guan et al. 2016), and Stock Synthesis (SS) (Kitakado et al. 2012, Hoyle et al. 2014, Langley & Hoyle 2016, Langley 2019, Rice 2022, Rice and Phillips 2025).

The most recent (2025) assessment using Stock Synthesis incorporated data to 2020 (Rice 2022). The assessment incorporated longline CPUE indices for albacore from a collaborative project between Japan, Taiwan, Korea, and the IOTC, similar to the work done in 2016 (Hoyle et al 2019. Hoyle et al. 2015, Hoyle et al. 2016). The standardized CPUE indices were derived from operational-level longline data from the three fleets, incorporating cluster analyses to address the effects of target change and standardization models including vessel effects and spatial effects. The project built on a study applying similar methods to Indian Ocean bigeye and yellowfin tuna (Hoyle et al. 2015).

The 2019, 2022, and 2025 assessments investigated the interactions between the key data sets (CPUE indices and length composition data) and the uncertainty associated with the key biological parameters (natural mortality and SRR steepness). This assessment, similar to the 2022 and 2019 assessments used the fleets as areas approach. While the 2019 assessment used four longline fleets, and four corresponding CPUE series (LL1 – LL4), the 2022 and the current (2025) assessment further disaggregate the longline fisheries by quarter. The CPUE in the southwest Indian Ocean are thought to mostly likely to represent the abundance of the albacore tuna stock (Rice 2022, Langly 2019). This area represents the in main target fishery with more consistent fishing operations and a significant proportion of the albacore biomass in the Indian Ocean (Rice 2022, Langly 2019). The previous assessment (Rice 2025) included suggestions based on the WPTmT_07 meeting and incorporated changes to the fleet structure, use of the length frequency data and data weighting.

The 2025 WPTmT agreed to use the updated NW and SW models (with revised catch, length composition, and CPUE data) to estimate management quantities. However, the 2025 assessment meeting (WPTmT 09) and SC (SC28) noted that the delay in the provision of the CPUE index made it impossible to investigate some issues in the assessment during the WPTmT meeting (e.g., some instability of the model caused by changes in the configuration of selectivity parameters). The WPTmT and SC discussed in detail the outstanding issues in the updated NW and SW models. In summary the two outstanding issues were;

1) The SW model produced very high biomass estimates with large uncertainty when the selectivity for LL3 and LL4 was unconstrained (allowed to be domed-shaped).
2) The NW model showed bias in the predicted length composition for the LL1 fishery.

Despite several investigative model runs during the 2025 meeting, the exact causes of these issues and potential solutions remained unclear. Given the changes in input data and the late availability of the CPUE indices, the WPTmT agreed that, while the updated assessment model in its current configuration is sufficient for estimating stock status, further scrutiny is needed to improve its reliability and ensure robust management advice. As such, the WPTmT recommended that assessment work continue next year and that the SC convene another WPTmT assessment meeting in 2026 to review progress.

This report presents the development of runs to address these issues and the preliminary results of the 2026 stock assessment modelling of Indian Ocean albacore tuna using Stock Synthesis (Version 3.30.24.2). The assessment results will be finalized at the WPTmT10 meeting.

2 Methods

Data

There are many different fleets catching albacore tuna in the Indian Ocean, with the main fleets consisting of longline fleets from distant water fishing nations. The data used in the albacore tuna assessment consist of fishery specific catch and length composition data along with standardised longline CPUE indices. The details of the configuration of the fishery specific data sets are described below.

There is enough uncertainty about the selectivity assumptions with respect to time, and the interannual variability with respect to the numbers of size composition data, that the size composition data are not expected to be very informative about year-class strength. Hence, in the assessment presented here, the length-composition data are down weighted so as to inform the selectivity and recruitment but not alter the model fit to the abundance trend.

2.1 Spatial stratification

The 2016 assessment partitioned the Indian Ocean into four quadrants demarcated at 25°S latitude and 75°E longitude (Langley & Hoyle 2016). The spatial stratification was primarily adopted to partition the longline fisheries by the size of fish caught; the longline fisheries in the southern area tend to catch albacore that are considerably smaller than the fisheries in the northern area (Chen et al. 2004, Geehan & Hoyle 2013, Nikolic et al 2013). Higher longline CPUE for albacore has been associated with the North Subtropical Front in the southern Indian Ocean (30–35°S latitude) where SST was 15–19°C (Lan et al 2011). There is no indication of a longitudinal trend in the size of albacore caught by the longline fisheries (Geehan & Hoyle 2013) and the overall distribution of the southern longline fishery is continuous throughout the southern region (Figure 1). However, during the history of the fishery there were periods when there was an appreciable separation between the operation of the longline fishery in the southwestern and south-eastern quadrants of the Indian Ocean, most notably during the 1960s and 1970s (Figure 2). To account for potential longitudinal variation in the key fishery data sets, the northern and southern areas of the Indian Ocean were partitioned at 75°E longitude. An investigation of the stock structure of Indian Ocean albacore by morphometric and DNA sequence methods categorized samples into two major groups partitioned by 90°E longitude suggesting that there may be two albacore stocks in the Indian Ocean (Yeh et al. 1995). Thus, the spatial stratification of the assessment data sets can be applied to approximate the more complex stock structure within the Indian Ocean.

The four regions of the Indian Ocean were used to define the spatial domain of the model fisheries and define the region-specific longline data sets for the CPUE analysis (Hoyle et al 2016 and 2019, Kitado et al 2022). There are apparent differences in the trends in albacore CPUE indices between the two southern areas over the last decade.

2.2 Temporal stratification

The time period covered by the assessment is 1950-2022 representing the period for which catch data are available from the commercial fishing fleets. The model was further stratified by quarter of the calendar year (Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec) and the various data sets were compiled accordingly.

2.3 Definition of fisheries

The spatial stratification was applied to define model fisheries based on region and fishing gear. These “fisheries” are considered to represent relatively homogeneous fishing units, with similar selectivity and catchability characteristics. As an update for this assessment the single aggregate longline fisheries by region have been separated into quarterly fisheries, thus there are sixteen longline fisheries, as opposed the previous assessment which used four.

A total of twenty-three fisheries were defined, including 16 longline fisheries (one per quarter per region), two driftnet fisheries, one purse seine fishery and four ‘other’ fisheries that account for the troll, sport fish, and miscellaneous fisheries by region (Table 2).

Longline northwest (LL1). The composite longline fishery in the north-western region was initially developed by the Japanese fleet in the mid 1950s. Data for the Taiwanese fleet are available from the late 1960s and the fleet has operated continuously since then. Japanese fishing effort declined in the early 1970s and remained relatively low until the mid 1990s. Fishing effort by the Japanese fleet recovered to a moderate level during the late 1990s and 2000s but was at a relatively low level from 2010. The composite longline fishery included effort by the Korean longline fleet during the late 1970s and 1980s. Limited albacore catches have been taken by other fleets, most notably the Chinese longline fleet operating during the last decade.

Longline northeast (LL2). The composite longline fishery in the north-eastern region has a similar composition to the LL1 fishery. The Japanese fleet was the dominant component of the fishery during 1953-1970 and continued to operate at a lower level over the subsequent years. Data for the Taiwanese fleet’s operations in the area are available from the late 1960s, while the Korean fleet primarily operated in the fishery during 1976-1987. Longline southwest (LL3) . The composite longline fishery in the south-western region was dominated by the Japanese fleet from the inception of the fishery in the late 1950s until the introduction of the Taiwanese fleet in the mid-1960s. Since the early 1970s, most of the catch was taken by the Taiwanese fleet. There was a short period of higher Japanese catch during 2004-2008.

Longline southeast (LL4) . The composite longline fishery in the south-eastern region was dominated by the Japanese fleet from the inception of the fishery in the late 1950s until the introduction of the Taiwanese fleet in the early 1970s. During 1974-2006, most of the catch was taken by the Taiwanese fleet. Japanese longline catches increased from 2006 and in recent years considerable catches have also been taken by China and Korea.

Drift net fisheries. Drift net fisheries were defined for the south-western (DN3) and south-eastern (DN4) regions. These fisheries were comprised exclusively of drift net vessels flagged to Taiwan China, which operated in the southern waters of the Indian Ocean from 1982 until 1992 when the UN adopted a worldwide ban on drift nets.

Purse seine. A single purse seine fishery (PS1) was defined as virtually all purse seine catch was taken within the north-western region. The purse seine fishery is made up of various fleets although the majority of the catches of albacore are reported by purse seiners flagged to the European Union and other fleets under EU ownership, including the Seychelles (86% of the total catches of albacore over the time series). The purse seine catches of Iran, Japan, Mauritius, Thailand, and the Republic of Korea are also included in the fishery.

Other. A miscellaneous (“Other”) fishery was defined for each of the regions. The “Other” fisheries include various coastal longline, gillnet, trolling, hand lines and other minor artisanal gears, which are used in coastal countries of the Indian Ocean. Collectively, the “Other” fisheries account for a small proportion of the Indian Ocean albacore catch (2% of the entire catch). Most of the catch was reported by Indonesia with the remainder of the catches reported by Mauritius, Reunion and Mayotte (EU), Comoros, Australia, South Africa, and East Timor.

2.4 Total catch

Catch data were compiled by IOTC secretariat based on the fishery definitions (IOTC-2025-IOTC-2025-WPTmT09-AS-ALB_02_SS3.xlsx). All catches were expressed in metric tonnes (mt). There were minor changes in the gear specific catch histories from the previous assessment (Rice 2022). The longline fisheries for albacore developed from the mid 1950s and total annual catches averaged about 17,000 t during 1960-1980 (Figure 3). Catches increased during the period of drift net fishing in the late 1980s and early 1990s and continued to increase with the expansion of the longline fishery in the late 1990s to reach a peak in catch of 40-45,000 mt in 1998-2001. Total catches declined to about 30,000 t in 2003-2006 and have since fluctuated between 33-43,000 t per annum (Figure 3). During the last decade, the longline fisheries have accounted for 95% of the total albacore catch apportioned amongst the regions as follows: northwest 23%, northeast 13%, southwest 34% and southeast 30%. Most of the catch from the southern longline fisheries occurs during the 2nd and 3rd quarter of the year, while catches from the north-western longline fishery are predominantly taken in the 4th quarter.

Longline in the northwest region (LL1). Fishing Fleets 1-4. The longline catch was relatively low from the north-western region during the late 1980s and early 1990s immediately following the reduction in fishing by the Korean fleet and during a period of low catch by the Japanese fleet. The high catches in 2000 and 2001 were predominantly attributable to higher catches by the Taiwanese fleet in those years (Figure 3). Annual catches averaged about 8,000 t per annum over approximately the last decade (2010-2020), with the average catch by quarter being 2000, 287, 600, and 5030 MT for quarters 1-4 respectively.

Longline in the northeast region (LL2) (Fishing fleets 5-8) The higher catches from the fishery in the late 2000s (2007 and 2008) are primarily attributable to higher catches by the Taiwanese fleet in 2007 and 2008 and the Indonesian longline fleet during 2009 and 2010. Catches were generally declining from the early 2000’s to the approximately 2015, after which they increased in all quarters (Figure 2), and averaged 1235, 1257, 1912, and 1000 MT .by quarter (1-4 respectively) over 2010-2020.

Longline in the southwest region (LL3) Fishing fleets 9-12. The high catches from the fishery during 1996-2002 are primarily attributable to a considerable increase in catch by the Taiwanese fleet during that period (Figure 2). Annual catches dropped markedly in 2003 and the catch from the Taiwanese fleet remained relatively low during 2003-2005. Annual catches increased steadily since about 2006 , most notably in region 2. Quarterly catches averaged 854,7068, 5735 and 692 by quarter (1-4 respectively), with the highest level of catch of about 19,000 MT in 2019, followed by 17,000 MT in 2020.

Longline southeast region (LL4), Fishing Fleets 13-16. Longline catches from the south-eastern region reached an historically high level of about 10,000 t in 2014. The higher level of catch around this time was primarily attributable to an increase in catch by the Japanese fleet since 2006. Annual catches have declined in more recent years and were about 4,000 t in 2019 and 2020. Average quarterly catches are about 300, 4,000, 2360, and 193 MT for quarters 1-4 respectively.

2.5 Relative abundance indices

Similar to the 2019 assessment, for each of the four regions, standardised CPUE indices for albacore tuna were derived using generalized linear models (GLM) from operational longline catch and effort data provided by Japan, Korea, and Taiwan, China (Kitakado et al 2022). Contrary to the 2019 assessment data from the Seychelles was not used. Cluster analyses of species composition data by vessel-month for each fleet were used to separate datasets into fisheries that are believed to target different species or species compositions. Selected clusters were then combined and standardized using generalized linear models. In addition to the year-quarter, models included covariates for vessel identity, 5 square location, effort and cluster. The analysis is a refinement of the approach used to derive the CPUE indices included in the 2016 and 2019 albacore stock assessment (Hoyle et al 2016, Hoyle et al. 2019).

Four sets of CPUE indices by year-quarter were derived based on different treatment of the fishing vessel variable in the CPUE modelling, resulting in one LL fleet by region (Kitakado et al 2020). These four regional CPUE series were then dis-aggregated by quarter, resulting in 14 CPUE series pertaining to quarters 1-4 for regions 1-3 and quarters 2 & 3 for region 4. This Models were run for the period 1975-2020, the early time period included in the previous assessment was not considered as previous assessments have highlighted that the large decline in the CPUE indices during the early period (1960- 1970) was not consistent with the relatively low catches taken from each of these regions during the period (Langley & Hoyle 2016, IOTC 2019).

CPUE indices for regions 1-3 incorporated the effects of target and effort in the delta-component and all the effects including the quarter-space interactions in the positive lognormal component. Due to instability with same model with data from region 4, the log-normal model was used an alternative approach. The study authors note that the diagnostics plots show some deviation from the normal distribution.

The CPUE indices from region 1 are characterized by a decline in the magnitude and variability of the indices through the timeseries (Figure 4). The CPUE indices from region 2 are characterized by higher and more variable CPUE in quarter 2, compared to the other quarters, of which quarters 3 and 4 are the most stable. The CPUE series from region 3 are similar to region 2 with higher and more variable CPUE in quarter two, with lower and less variable standardized values in quarter 3, 4, and to an extent in region 1. The high variability in the region 3 CPUE indices during the later part of the 1970s may have been primarily attributable to a shift in targeting behavior (Figure 4).

A short time series of annual Taiwanese drift net CPUE indices is available from 1985–1992 (Chang & Liu 1995) (Figure 5). The indices exhibit a high degree of interannual variability with high CPUE indices for 1987 and 1990. The reliability of the indices as an index of albacore abundance is unknown, although the high variability may provide some indication of variation in year class strength during the period given that the drift net fisheries typically catch a relatively narrow length range of albacore (corresponding to fish of about 1–2 years old). The annual indices were assumed to represent the relative abundance in the first quarter of the year (corresponding to the peak season of DN catch).

2.6 Size composition data

Longline fishery

Size frequency data are available for the Japan longline fishery from 1965. Length and weight data were collected from sampling aboard Japanese commercial, research and training vessels. Weight frequency data collected from the fleet (as live weight) have been converted to length frequency data via a weight-length key. Levels of sampling aboard the Japanese composite longline fleet over time were uneven in terms of both the sampling platform (commercial and non-commercial vessels) and sampling source (fishermen, scientists, observers). While in recent years the majority of the samples available have come from scientific observers on commercial vessels, in the past samples came from training and research vessels (scientists), and commercial vessels (fishermen).

Length frequency data from the Taiwanese longline fleet are also available from 1980. In recent years, length data are also available from other fleets and periods (e.g. Indonesia fresh-tuna longline, Seychelles, etc.). Prior to the mid-2000s the length frequency data set is dominated by sampling from the Taiwanese deep-freezing longline fleet. Length samples from this component come from commercial vessels and include lengths recorded by fishermen and, to a lesser extent, lengths measured by scientific observers on some of those vessels in recent years. A review of the Taiwanese length frequency data identified major differences in the length frequencies of albacore recorded before and after 2003, with the majority of the smaller albacore missing from the length distributions since that year (Geehan and Hoyle 2013 and Hoyle et al 2021). Following the previous assessment concerns regarding the reliability of these data, all length samples collected from the Taiwanese longline fleet via logbooks from 2003–onwards were excluded from the length composition data sets.

Purse seine fishery

Purse seine fisheries catch adult albacore, as a bycatch (approximately 90-120cm), in the western central Indian Ocean. Albacore lengths are measured in port, by enumerators, during the unloading of purse seiners flagged in the EU and Seychelles. Length samples are available from the fishery from 1990-2020. The ESS of the individual samples was determined in the same manner as described for the longline fisheries.

Drift net fishery

Drift nets catch juvenile or sub-adult albacore (62-75 cm) (Figure 15). The 2019 assessment is the first time length composition data have been available from the Indian Ocean fishery. The data are from sampling of the Taiwanese drift net catch during 1985-1991. These data were recently provided to IOTC by T. Nishida (IOTC 2019).

Length Frequency Data Weighting

Given the recommendation of the WPTmT 08 that the length composition data weighting by fisheries should be given careful consideration the following protocol was adopted;

1) A minimum annual sample size of 30 fish measured by fleet and quarter was adopted
2) The initial effective sample size (ESS) was calculated as the square root of the number of fish measured following the methods outlined in Rice et al. 2025 (in prep).
3) The samples were assigned the relatively low ESS due to concerns regarding the reliability of the length samples from some key data sets.

3 Model Assumptions

The most important model assumptions are described in the following sections. Standard population dynamics and statistical terms are described below, while equations can be found in Methot (2000, 2009). Attachment 1 is the template specification file for all of the models, and includes additional information on secondary elements of model formulation which may be omitted in the description below. All of the specification files are archived with the IOTC Secretariat. Table 2 lists the assumptions for the sensitivity runs.

3.1 Software

The analysis was undertaken with Stock synthesis SS 3.30.24.2, 64 bit version (Methot 2000, 2009, executable available from http://nft.nefsc.noaa.gov/SS3.html), running on MS Windows™ 11). Typical function minimization of the fully disaggregated model on a 3.0 GHz personal computer required about 25 minutes. Additional simplifications and aggregations could probably reduce the minimization time further, without significant loss to the stock status inferences.

3.2 Population Dynamics

The assessment model was structured by sex and age with age classes of 0-13 years and an aggregate age class of 14+ fish. The model commences in 1950 at the start of the available catch history and continues to 2023. The initial population age structure was assumed to be in an unexploited, equilibrium state. Model years are partitioned into four quarterly seasons. A single spatial structures for the Indian Ocean albacore stock was used in conjunction with the quarterly disaggregated survey and fisheries, based on the partitioning of the Indian Ocean into four regions (Figure 1).

3.3 Biological inputs and assumptions

Sex Ratio

Across all oceans there are documented differences in the patterns of sex ratio at recruitment and at older ages and larger sizes, these patterns are likely to be caused by features of albacore biology. Males have been shown to grow considerably larger than females in the north Pacific (Chen et al. 2012), south Pacific (Williams et al. 2012), north Atlantic (Santiago and Arrizabalaga 2005), and Mediterranean (Megalofonou 2000). However there is no evidence for unbalanced sex ratio at the age of recruitment. Sex ratio at recruitment was assumed to be equivalent (1:1).

3.4 Growth

The standard assumptions made concerning age and growth in the SS model are (i) the lengths-at-age are assumed to be normally distributed for each age-class; (ii) the mean lengths-at-age are assumed to follow a von Bertalanffy growth curve. Following the previous assessment this study used the Farley et al. 2019 estimation of the age and growth of Indian Ocean albacore. The study sampled albacore from the western Indian Ocean, primarily from the longline fishery with smaller fish also sampled from pole-an-line and purse seine fisheries. The sampled fish ranged in length from 74 to 108 cm FL for females and 67 to 115 cm FL for males (Farley et al. 2019).

Reproductive potential of female albacore was assumed to be equivalent to south Pacific albacore maturity-at length (Farley et al. 2014) (Figure 7). This ogive takes into account sex ratio, sexual maturity, spawning fraction, and fecundity and so represents female reproductive output at length. The ogive has fish attaining sexual maturity at about 85 cm and 50% reproductive potential is reached at about 92 cm (Figure 7). When converted to age using the Indian Ocean albacore growth curve (Farley et al. 2019), sexual maturity is attained at about age 4 years and 50% reproductive potential is reached at about 5 years.

3.5 Natural mortality

Natural mortality was assumed to be 0.3 for both sexes, based on the previous assessment and the value applied in the north Pacific and the north Atlantic albacore stock assessments. Age specific natural mortality at age was investigated in a sensitivity analysis

3.6 Recruitment

The model partitions the population into 14 year-quarter age-classes in one region (Figure 1). The last age-class comprises a “plus group” in which mortality and other characteristics are assumed to be constant. The population is “monitored” in the model at year quarter time steps, extending through a time window of 1950-2023. The main population dynamics processes are as follows:

Recruitment in the model occurs in the fourth quarter of each year, reflecting the summer spawning season (and the ageing protocol that assumed a birthday of 1 December Farley et al. 2019). Recruitment was based on a BH stock recruitment relationship (SRR) and annual deviates were estimated for the period of the model where there was the most data (1975−2023). Deviates from the SRR were given a small penalty, so that recruitment estimates in periods with less data were estimated closer to the mean. The applied penalty was based on the assumption that the true standard deviation of recruitment deviates (σR) is 0.3, reflecting the upper range of the magnitude in the. variation of recruitment deviates estimated during preliminary modelling. Imperfections in models and lack of full information in the data cause models to underestimate recruitment variability. Recruitment variability is assumed to be lognormally distributed, therefore mean recruitment is higher than median recruitment. Equilibrium recruitment is meant to represent the average recruitment through time, so the median value in the recruitment function must be bias-corrected upwards. Following Methot and Taylor (2011), the bias correction was adjusted across the time series according to the relationship between the assumed and estimated recruitment variability. recruitment deviates were estimated for 1975−2021.

The final model options included three (fixed) values of steepness of the BH SRR (h 0.7, 0.8 and 0.9). These values are considered to encompass the plausible range of steepness values for tuna species such as albacore tuna and are routinely adopted in tuna assessments conducted by other tuna RFMOs.

3.7 Initial population state

In the previous assessment it was assumed that the albacore tuna population was at an unfished state of equilibrium at the start of the model (1950) with the beginning of longline fishing occurring in the following years (at least from the 1950s onwards), this assessment follows the same methodology.

3.8 Selectivity Curves and Fishing Mortality

Selectivity is fishery-specific and was assumed to be time-invariant. A double- normal functional (Methot 2015) form was assumed for all selectivity curves. No sex-specific length data was available, all length data were aggregated. Length composition data was split by quarter and region, fishery F16_LL4_Q4 was mirrored to F15_LL4_Q3 due to low sample size resulting in poor fits in the initial modelling.

A double normal selectivity was estimated for the purse seine fishery (PS1). The availability of length composition data from the drift net fishery enabled the estimation of a selectivity function for this method. The selectivity was parameterised as a double normal function and assumed to be equivalent for the two fisheries (DN3 & DN4).

No reliable length frequency data are available for the four “Other” fisheries. The selectivity of the “Other” fisheries was assumed to be equivalent to the selectivity of the drift net fishery. Initial modelling indicated that the model results were not sensitive to the selectivity assumed for the four “Other” fisheries due to the small magnitude of the catch associated with each of these fisheries. Length composition data was down weighted so that the data can inform removals by fisheries from the correct age class, and inform recruitment but not determine the scale or trend of the population.

Fishing mortality was modelled using the hybrid method that the harvest rate using the Pope’s approximation then converts it to an approximation of the corresponding F (Methot & Wetzel 2013).

3.9 Parameter estimation and uncertainty

Model parameters were estimated by maximizing the log-likelihoods of the data plus the log of the probability density functions of the priors, and the normalized sum of the recruitment deviates estimated in the model. For the catch and the CPUE series we assumed lognormal likelihood functions while a multinomial was assumed for the size data. The maximization was performed by an efficient optimization using exact numerical derivatives with respect to the model parameters (Fournier et al. 2012). The Hessian matrix computed at the mode of the posterior distribution was used to obtain estimates of the covariance matrix. This was used in combination with the Delta method to compute approximate confidence intervals for parameters of interest.

3.10 Profile Likelihood

An investigation of the information content in the data components was undertaken via the use of profile likelihood on the global scaling parameter (R0) (Lee et al 2014). The negative log likelihood of a specific parameter or data component should, in theory, decline to an obvious minimum. In situations where this does not happen, at least from one side, there may be insufficient information within the data to estimate other parameters. Virgin recruitment (R0) is an ideal scaling parameter because it is proportional to the unfished biomass. Profiles were run with the natural log of virgin recruitment, ln(R0), fixed at various values above and below the model estimated value; the corresponding likelihood profile quantified how much loss of fit was contributed by each data source. One of the primary uses of the likelihood profile is to identify conflicting data and provide a rationale for down weighting or excluding any data.

3.11 Selection of a diagnostic case.

During the WPTmT 09 meeting the group chose to use the model runs that included the NW CPUE and SW CPUE to provide stock status advice. This assessment continues with those runs, following the advice from the 2025 assessment meeting. Previous assessments investigated a wide range of structural assumptions and those results informed to structure of the most recent assessment. For the current assessment, there are no additional data to enable any further investigation of some of the main structural assumptions related to spatial structure, fishery selectivities, initial conditions and recruitment estimation. Sensitivities to the assumptions regarding certain model assumptions are noted in section 3.13.

3.12 Benchmark and Reference Point Methods

Benchmarks included estimates of absolute population levels and fishing mortality for the terminal year, 2023 (F_{2023}, SSB_{2023}, B_{2023}). These values are reported against reference points relative to MSY levels, and depletion estimates (relative to virgin levels).

3.13 Diagnostics and additional model runs

Residual analysis, retrospective analysis, likelihood profiles, age structured production models, leave one out CPUE, let one in CPUE, and jitter runs were used to investigate the selected diagnostic runs. Additional model runs that included alternative parameterizations of the stock recruitment relationship, variation in length-at-age for males and females, length-weight parameters, natura mortality and maturity at length were also run to explore the structural uncertainty in the assessment. An exploration of interactions/conflicts among the CPUE and size frequency was conducted by applying alternative weighting schemes on the different data sources was also conducted.

4 Results

In this section we focus on the results from the diagnostic case models and the key results and diagnostics for this model. We then comment on any important differences in both outputs and model diagnostics for the sensitivity analyses, and present preliminary results. Stock Synthesis 3 was implemented here as a length-based age-structured stock assessment model (Methot and Wetzel 2013; e.g., Wetzel and Punt 2011a, 2011b). Stock synthesis utilizes an integrated modeling approach (Maunder and Punt 2013) to take advantage of the many data sources available for the Indian Ocean stock of albacore tuna. An advantage of the integrated modeling approach is that the development of statistical methods that combine several sources of information into a single analysis allows for consistency in assumptions and permits the uncertainty associated with each data source to be propagated to final model outputs (Maunder and Punt 2013).

4.1 Diagnostic case model

The choice of model parameters and data inputs reflected the input of the WPTmT 09 assessment meeting and the available updated data for biology and life history to the extent possible. This will be further refined by input at the WPTmT assessment meeting.

Model Fits to Abundance Indices

The model was able to fit the general trends of the indices of abundance for the NW and SW diagnostic cases (Figures 9 & 10). With respect to the NW diagnostic case (Figure 9), the model provides a reasonable representation of the long-term trends in all four quarterly CPUE indices, although some systematic deviations remain. In particular, the model tends to underestimate the observed CPUE in the most recent years of the quarter 2 index, while slightly overestimating the observations during the terminal years of the quarter 4 index. The quarter 1 and quarter 3 indices exhibit substantial interannual variability together with a pronounced decline beginning around 2010 and continuing until approximately 2015–2016, followed by a partial recovery. A similar, though less pronounced, pattern is evident in the quarter 2 and quarter 4 indices. In the terminal years, however, the observed indices diverge among quarters. The quarter 1 index shows evidence of a modest increase beginning around 2018, whereas the quarter 2, quarter 3, and quarter 4 indices generally exhibit stable to declining trends, with quarters 3 and 4 showing a more gradual and variable decline. The model captures the broad long-term patterns but does not fully reproduce the magnitude and timing of these recent, quarter-specific changes. The estimated spawning biomass increased modestly during the late 1990s and early 2000s before declining thereafter, a pattern that coincides with increasing catches (Figure 2) and the subsequent decline observed in the CPUE indices.

The SW diagnostic model provides a reasonable representation of the overall patterns in the four quarterly CPUE indices from the southwest region, although some systematic residual patterns remain (Figure 10). In contrast to the NW indices, which exhibited a prolonged decline during the 2010s, the southwest indices are comparatively stable throughout the time series, with modest interannual variability and a period of elevated CPUE during the mid-2000s that is generally captured by the model. Following this period, the observed indices remain relatively stable, with only weak evidence of declining trends in the terminal years. The quarter 1 index shows a modest increase in the final years of the series, while the quarter 2, quarter 3, and quarter 4 indices exhibit relatively stable or only slightly increasing trends. Although the model reproduces the broad temporal patterns in each quarterly index, it does not capture all of the observed year-to-year fluctuations, particularly during years with relatively high CPUE observations.

Interpretation of the southwest CPUE indices should also consider the biological composition of the catch. Fish harvested in this region are generally smaller than those caught in the northwest region and are therefore more likely to comprise multiple adjacent year classes. Consequently, annual changes in CPUE may reflect variation in the relative abundance and availability of several cohorts rather than a single dominant cohort, resulting in smoother population signals and greater interannual variability that may not be fully reproduced by the assessment model.

Fits to the Length composition

The overall fit to the length data was generally good (Figure 13 & 14). Fleet specific annual length samples were often quite different, i.e. left skewed one year and bimodal the next, which accounts for the small amount of misfit in the aggregated samples. Pearson residuals of the fit to the length compositions were generally small and did not show any temporal trend (Figures 15-20).

Stock-recruitment Parameters

The predicted virgin recruitment (R0; number of age 0) was approximately 17,000,000 animals (for both the NW and SW). Recruitment deviations are estimated starting in the late 1970s, after which estimated recruitment experienced large fluctuations from about 1980-2020 (Figure 21). The bias correction estimated in the model is shown in Figure 22. The corresponding estimated stock recruitment relationship and annual deviations are also shown in Figure 23.

Fishing Mortality

Estimated F/F_{MSY} and fleet-specific instantaneous fishing mortality rates are presented in Figures 24 and 25 respectively. Fishing mortality was relatively low from the 1950 to the mid-1990s, which is in accordance with low catches and effort during that period. In the late 1980s fishing mortality increased with the advent of driftnet fishery. Starting in the late-1990s overall fishing mortality increased, with large fluctuations in the individual fisheries contribution to the overall fishing mortality. Since the early 2000s the overall fishing mortality has been increasing but below F_{MSY} (i.e. overfishing is not occurring) for the entire time series.

Estimated stock status and other quantities

The estimated equilibrium yield curve for the diagnostic case model is shown in Figure 21. The estimated MSY is approximately 41,000 and 47,000 MT (for the NW and SW respectively) and this is predicted to occur at 21.5% of the unfished biomass (Figure 21), which is less than the standard Schaefer production model ($0.5SSB_0$). The diagnostic case models estimates that the total biomass of the stock was at approximately 100% of the unfished level at the start of the model period (Figure 11) and steadily decreased to an estimate of SB_{2023}/SB_{MSY} = 1.016 and 1.6 (for the NW and SW respectively) which corresponds with and estimated F_{2023}/F_{MSY} = 1.26 and 0.556 (for the NW and SW respectively).

Recruitment is fairly well estimated throughout the model time period (Figure 23), with recent recruitment estimated to be lower than the implied stock recruitment curve due to deviations implied by the length data. The estimates of recruitment were quite tightly constrained to the stock recruitment curve for the initial period of the model when there was no length information to inform the model. The main trends in the population dynamics can be explained through the estimated fishing mortality which was greatly increased in the 1990s and early 2000s due to the increase in catch (Figures 24 and 25).

Model Uncertainty

Stock status uncertainty was evaluated delta-Multivariate lognormal (MVLN)approximation to generate joint error distributions for SSB_{2023}/SSB_{MSY} and F_{2023}/F_{MSY}. Figure 28 shows the estimated stock status based on the MVLN analysis for the base case model and Figure 23 shows the estimated timeseries based on the MVLN approximation.

Stock synthesis provides estimates of the MSY-related quantities and these and other quantities of interest for management are provided in Table 4.

Retrospective Analysis

As part of an analysis of model structure, retrospective analysis (sequentially deleting 1 year of data from the end of the model and re-running) was run using the diagnostic case formulation (the southwest CPUE series). The estimates of spawning depletion remain very similar across all the retrospective model runs considered (Figure 25) indicating that the changes in estimates of virgin spawning biomass are based on the total catch (Figure 25 right panel). The last retrospective run (-5 years) estimated a more depleted stock that corresponds to a slightly smaller virgin recruitment (Figure 25 right panel), this is associated with higher estimated total fishing mortalities in the last 4 years. In general, the retrospective analysis shows no large departures from the estimated scale, depletion, or overall trend based on the sequential deletion of the last 7 years of data.

4.2 Other model diagnostics

Annex 1 shows the results of the expanded analysis on the residuals, runs tests joint residual plots, likelihood profiles, age structured production model and hindcasting cross-validation are presented in Annex 1. Select details from Annex 1 are repeated here, the reader is encouraged to read Annex 1 in its entirety.

The runs test indicated that the CPUE from quarter 1 and two and length data did contain some patterns in the residuals, indicating a trend in the departure from the expected values the (Figures A7 and A8) the cause of this was the early time period in which the index had high contrast and little data. The joint residual plots (Figure A6) show that in the latter part of the time series the model fits the Japanese data fairly well.

An age structured production model (ASPM) can help evaluate whether the catch and CPUE data give evidence for a production function within the model (Carvalho 2017). Overall, the ASPM evaluates whether the effect of surplus production and observed catches alone could explain trends in the CPUE, in contrast to a more complex model (i.e. SS3) that incorporates annual recruitment deviations to improve the fit (Carvalho et al 2021). Maunder and Piner (2017) note that if the ASPM fits well to the indices of abundance with contrast the production function is likely to drive the stock dynamics and the indices will provide information about absolute abundance (Minte-Vera et al., 2017). Figure A13 shows that the biomass trajectories for both models (ASPM and the diagnostic) follow the same trend and that the estimates of *LN(R0)* are different. The fits to the indices are shown in Figure A14 and indicate an overall good fit, indicating that the information content in the data is sufficient.

5 Conclusion

The overall scale of the estimated stock biomass is similar to the previous assessment, though there is variation based on the model parameterization The change from the estimated selectivity in the NW to imposing the selectivity from the SW helped reduce the bias in expected numbers of individuals caught, along with revision of the size data for the LL1 fleet in the late 1990's and early 2000's. The declining trends in the CPUE indices during the early part of the model match the overall catch from the fisheries. The main drivers of this assessment are the trend in the catch and CPUE series. In particular the large increase in recent years of catch has different interpretations (within the model). based on which the CPUE series are included.

6 Acknowledgements

The author would like to thank the IOTC staff for their help in preparing the data as well as all of the member countries that submitted data and analyses. This assessment benefited from the discussion and contribution of the WPTmT attendees.

7 Reference

Bartoo, N., Holts. D. (1993). Estimated drift gillnet selectivity for albacore Thunnus alalunga. Fishery Bulletin92:371-378.

Bromhead, D., B., A. Williams and S. D. Hoyle (2009). Factors affecting size composition data from south Pacific albacore longline fisheries.

Cadrin, S.X., Vaughn, D.S., 1997. Retrospective analysis of virtual population estimates for Atlantic menhaden stock assessment. Fish. Bull. 95(3), 445-455.

Chang, F.-C., C.-Y. Chen, L.-K. Lee and S.-Y. Yeh (2012). Assessment on Indian albacore stock based mainly on Taiwanese longline data. Fourth Working Party on Temperate Tunas, Shanghai, China, 20–22 August 2012. . Indian Ocean Tuna Commission. IOTC-2012-WPTmT04-19.

Chang, S., H. Liu (1995). Adjusted Indian Ocean albacore CPUE series of Taiwanese longline and drift net fisheries. Sixth session of the Expert Consultations on Indian Ocean tunas, Colombo, Sri Lanka 25–29 September 1995. IOTC-1995-EC602-26.

Chang, S., H. Liu and C.-C. Hsu (1993). "Estimation of vital parameters for Indian albacore through length frequency data." Journal of the Fisheries Society of Taiwan 20(1): 1-17.

Chen, I. C., P. F. Lee and W. N. Tzeng (2005). "Distribution of albacore (Thunnus alalunga) in the Indian Ocean and its relation to environmental factors." Fisheries Oceanography 14(1): 71-80. Chen, K. S., T. Shimose, T. Tanabe, C. Y. Chen and C. C. Hsu (2012). "Age and growth of albacore Thunnus alalunga in the North Pacific Ocean." Journal of fish biology 80(6): 2328-2344.

Dhurmeea, Z.; Chassot, E.; Augustin, E.; Assan, C.; Nikolic, N.; Bourjea, J. et al. (2016). Morphometrics of albacore tuna (Thunnus alalunga) in the Western Indian Ocean. Seventh Working Party on Temperate Tunas: Data Preparatory Meeting, Kuala Lumpa, Malaysia, 14–17 January 2019. Indian Ocean Tuna Commission. IOTC-2019-WPTmT07(DP)-INF02.

Dhurmeea, Z.; Zudaire, I.; Chassot, E.; Cedras, M.; Nikolic, N.; Bourjea, J. et al. (2016) Reproductive Biology of Albacore Tuna (*Thunnus alalunga*) in the Western Indian Ocean. PLoS ONE 11(12): e0168605. doi:10.1371/journal.pone.0168605

Farley, J.H.; Eveson, J.P.; Bonhommeau, S.; Dhurmeea, Z.; West, W.; Bodin, N. (2019). Growth of albacore tuna (Thunnus alalunga) in the western Indian Ocean using direct age methods. Seventh Working Party on Temperate Tunas: Data Preparatory Meeting, Kuala Lumpa, Malaysia, 14–17 January 2019. Indian Ocean Tuna Commission. IOTC-2019-WPTmT07(DP)-21.

Farley, J. H., S. D. Hoyle, J. P. Eveson, A. J. Williams, C. R. Davies and S. J. Nicol (2014). "Maturity Ogives for South Pacific Albacore Tuna (Thunnus alalunga) That Account for Spatial and Seasonal Variation in the Distributions of Mature and Immature Fish." PloS one 9(1): e83017.

Farley, J., A. Williams, N. Clear, C. Davies and S. Nicol (2013). "Age estimation and validation for South Pacific albacore Thunnus alalunga." Journal of fish biology 82(5): 1523-1544.

Farley, J.H., A.J. Williams, S. D. Hoyle, C. R. Davies and S. J. Nicol (2013). "Reproductive dynamics and potential annual fecundity of South Pacific albacore tuna (Thunnus alalunga)." PLoS ONE 8(4): e60577.

Fonteneau, A. (2015). Indian Ocean albacore stock: review of its fishery, biological data and results of its 2014 stock assessment. IOTC/2015/XXX.

Fournier D A, Skaug HJ, Ancheta J, Ianelli J, Magnusson A, Maunder M, Nielsen A, Sibert J (2012) AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Methods Softw. 27:233-249.

Francis, R. I. C. C. (2011). "Data weighting in statistical fisheries stock assessment models." Canadian Journal of Fisheries and Aquatic Sciences 68(6): 1124-1138.

Francis RICC (2014) Replacing the multinomial in stock assessment models: A first step. Fisheries Research, 151, (2014), 70-84

Geehan, J. and S. Hoyle (2013). Review of length frequency data of the Taiwan,China distant water longline fleet, IOTC-2013-WPTT15-41 Rev_1. Indian Ocean Tuna Commission Working Party on Tropical Tunas, San Sebastian, Spain, 23–28 October, 2013.

Guan, W.; Zhu, J.; Xu, L.; Dai, X, (2016). Using a Bayesian biomass dynamics model to assess Indian Ocean albacore. Sixth Working Party on Temperate Tunas, Shanghai, China, 18–21 July 2016. Indian Ocean Tuna Commission. IOTC-2014-WPTmT06-22.

Gunn, J. S., N. P. Clear, T. I. Carter, A. J. Rees, C. A. Stanley, J. H. Farley and J. M. Kalish (2008). "Age and growth in southern bluefin tuna, *Thunnus maccoyii* (Castelnau): Direct estimation from otoliths, scales and vertebrae." Fisheries Research 92(2): 207-220.

Hoyle, S.D. (2019). Regional scaling factors for Indian Ocean albacore tuna. Seventh Working Party on Temperate Tunas: Data Preparatory Meeting, Kuala Lumpa, Malaysia, 14–17 January 2019. Indian Ocean Tuna Commission. IOTC-2019-WPTmT07(DP)-13.

Hoyle, S., J. Hampton and N. Davies (2012). Stock assessment of albacore tuna in the South Pacific ocean. Scientific Committee, Eighth Regular Session, 7-15 August 2012, Busan, Republic of Korea: 90.

Hoyle, S. D. (2008). Adjusted biological parameters and spawning biomass calculations for south Pacific albacore tuna, and their implications for stock assessments. WCPFC Scientific Committee. Nouméa, New Caledonia: 20.

Hoyle, S. D., A. D. Langley and R. A. Campbell (2014). "Recommended approaches for standardizing CPUE data from pelagic fisheries."

Hoyle, S.D., R. Sharma, M. Herrera (2014). Indian Ocean albacore assessment. Fifth Working Party on Temperate Tunas, Busan, Korea, 28–31 July 2014. Indian Ocean Tuna Commission. IOTC-2014-WPTmT05-24.

Hoyle, S.D. Yin Chang, Doo Nam Kim, Sung Il Lee, Takayuki Matsumoto, Kaisuke Satoh, and Yu-Min Yeh (2016). Collaborative study of albacore tuna CPUE from multiple Indian Ocean longline fleets. Sixth Working Party on Temperate Tunas, Shanghai, China, 18–21 July 2016. Indian Ocean Tuna Commission. IOTC-WPTmT06-19a.

Hoyle, S.D.; Chassot, E.; Fu, D.; Kim, D.N.; Lee, S.; Matsumoto, T.; Satoh, K.; Wang, S.; Kitakado, T. (2019). Collaborative study of albacore tuna CPUE from multiple Indian Ocean longline fleets in 2019. Seventh Working Party on Temperate Tunas: Data Preparatory Meeting, Kuala Lumpa, Malaysia, 14–17 January 2019. Indian Ocean Tuna Commission. IOTC-2019-WPTmT07(DP)-19.

Hoyle, S.D.; Yeh, Y.M.; Chang, S.T.; Wu, R.F. (2015). Descriptive analysis of the Taiwanese Indian Ocean longline fishery, focussing on tropical areas. Seventeenth Working Party on Tropical Tunas, Montpellier, France, 23–28 October, 2015. Indian Ocean Tuna Commission. IOTC-2015-WPTT17-INF09.

Hoyle, S.D. Chang, S-t. et al. (2021) Draft review of size data from Indian Ocean longline fleets, and its utility for stock assessment. IOTC-2021-WPTT23(DP)-08.

Huang, H.; Liu, K. (2010). Bycatch and discards by the Taiwanese large-scale tuna longline fleets in the IndianOcean. Fisheries Research 106: 261-270

Huang, H., C.-L. Wu, C. Kuo and W. Su (1991). "Age and growth of Indian Ocean albacore (Thunnus alalunga) by scales."

Ianelli, J., M. Maunder and A. E. Punt (2012). Independent review of 2011 WCPO bigeye tuna assessment. IOTC Secretariat (2013). Report and documentation of the Indian Ocean Tuna Fisheries of Indonesia Albacore

Catch Estimation Workshop: Review of Issues and Considerations. Bogor-Jakarta, 21-25 June 2013. IOTC Technical Report. No. 2013/01. Bogor-Jakarta, IOTC. 2013. 40 pp.

IOTC (2014). Report of the Fifth Session of the IOTC Working Party on Temperate Tunas. Fifth Working Party on Temperate Tunas, Busan, Korea, 28–31 July 2014. Indian Ocean Tuna Commission. IOTC-2014-WPTmT05-R[E].

IOTC (2016). Report of the Sixth Session of the IOTC Working Party on Temperate Tunas. Sixth Working Party on Temperate Tunas, Shanghai, China, 18–21 July 2016. Indian Ocean Tuna Commission. IOTC-2016-WPTmT06-R[E].

IOTC (2019). Report of the Seventh Session of the IOTC Working Party on Temperate Tunas (Data Preparatory Session). Seventh Working Party on Temperate Tunas: Data Preparatory Meeting, Kuala Lumpa, Malaysia, 14–17 January 2019. Indian Ocean Tuna Commission. IOTC-2019-WPTmT07(DP)- RE.

IOTC–WPTmT07(AS) 2019. Report of the Seventh Session of the IOTC Working Party on Temperate Tunas: Assessment Meeting. Shizuoka, Japan, 23–27 July 2019. IOTC–2019–WPTmT07(AS)–R[E]: 36 pp.

IOTC–WPTmT08(DP) 2022. Report of the Eighth Session of the IOTC Working Party on Temperate Tunas. Online, 13–15 April 2022. *IOTC–2022–WPTmT08(DP)–R[E]: 36 pp.*

ISSF (2011). Report of the 2011 ISSF Stock Assessment Workshop: Rome, Italy, March 14-17, 2011. ISSF Technical Report 2011-02, http://iss-foundation.org/wp-content/uploads/downloads/2011/08/ISSF-2011-02- Report-2011-ISSF-WS.pdf. McLean, Virginia, USA, International Seafood Sustainability Foundation.

Karakulak, F. S., E. Özgür, M. Gökoğlu, I. T. Emecan and A. Başkaya (2011). "Age and growth of albacore (Thunnus alalunga Bonnaterre, 1788) from the eastern Mediterranean." Turkish Journal of Zoology 35(6): 801-810.

Kitakado, T.; Fiorellato, F.; De Bruyn, P. (2019). Allometric curve for the Indian Ocean albacore. Seventh Working Party on Temperate Tunas: Data Preparatory Meeting, Kuala Lumpa, Malaysia, 14–17 January 2019. Indian Ocean Tuna Commission. IOTC-2019-WPTmT07XX.

Kitakado, T., E. Takashima, T. Matsumoto, T. Ijima and T. Nishida (2012). First attempt of stock assessment using Stock Synthesis III (SS3) for the Indian Ocean albacore tuna (Thunnus alalunga). IOTC Working Party on Temperate Tunas, Shanghai, China, 20–22 August 2012, Indian Ocean Tuna Commission.

Kitakado, T et al. (2022) Joint CPUE indices for the albacore Thunnus alalunga in the Indian Ocean based on Japanese, Korean and Taiwanese longline fisheries data. IOTC–2022–WPTmT08(DP)–15

Lan, K.; Kawamura, H.; Lee, M.; Lu, H.; Shimada, T.; Hosoda, K.; Saikaida, F. (2011). Relationship between albacore (Thunnus alalunga) fishing grounds in the Indian Ocean and the thermal environment revealed by cloud-free microwave sea surface temperature. Fisheries Research 113: 1-7

Langley, A.D.; Hoyle, S.D. (2016). Stock assessment of albacore tuna in the Indian Ocean using Stock Synthesis. Sixth Working Party on Temperate Tunas, Shanghai, China, 18–21 July 2016. Indian Ocean Tuna Commission. IOTC-WPTmT06-25.

Lee, H.-H., Piner, K.R., Methot R.D., Maunder, M.N. 2014. Use of likelihood profiling over a global scaling parameter to structure the population dynamics model: An example using blue marlin in the Pacific Ocean. Fish. Res.

Lee, H. H., M. N. Maunder, K. R. Piner and R. D. Methot (2012). "Can steepness of the stock-recruitment relationship be estimated in fishery stock assessment models?" Fisheries Research.

Lee, L.-K., C.-C. Hsu and F.-C. Chang (2014). Albacore (*Thunnus alalunga*) CPUE Trend from Indian Core Albacore Areas based on Taiwanese longline catch and effort statistics dating from 1980 to 2013.

Lee, Y. C. and H. C. Liu (1992). "Age determination, by vertebra reading, in Indian albacore, *Thunnus alalunga* (Bonnaterre)." Journal of the Fisheries Society of Taiwan 19(2): 89-102.

Lorenzen, K. (1996). "The relationship between body weight and natural mortality in juvenile and adult fish - a comparison of natural ecosystems and Aquaculture [Review]." Journal of Fish Biology 49(4): 627-647.

Matsumoto, T., T. Kitakado and T. Nishida (2014). Standardization of albacore CPUE by Japanese longline fishery in the Indian Ocean, IOTC–2014–WPTmT05–18 Indian Ocean Tuna Commission Working Party on Temperate Tuna (WPTmT): 16.

Matsumoto, T., T. Nishida and T. Kitakado (2012). Stock and risk assessments of albacore in the Indian Ocean based on ASPIC. Fourth Working Party on Temperate Tunas, Shanghai, China, 20–22 August 2012. Indian Ocean Tuna Commission. IOTC-2012-WPTmT04-20.

Matsumoto, T., T. Nishida and T. Kitakado (2014). Stock and risk assessments of albacore in the Indian Ocean based on ASPIC. Fifth Working Party on Temperate Tunas, Busan, Korea, 28–31 July 2014. Indian Ocean Tuna Commission. IOTC-2014-WPTmT05-22 Rev_1.

Matsumoto, T. (2016). Stock and risk assessments of albacore in the Indian Ocean based on ASPIC. Sixth Working Party on Temperate Tunas, Shanghai, China, 18–21 July 2016. Indian Ocean Tuna Commission. IOTC-2014-WPTmT06-20.

McAllister, M. K. and J. N. Ianelli (1997). "Bayesian stock assessment using catch-age data and the sampling importance resampling algorithm." Canadian Journal of Fisheries and Aquatic Sciences 54(2): 284-300.

Megalofonou, P. (2000). "Age and growth of Mediterranean albacore." Journal of Fish Biology 57(3): 700-715.

Methot, R.D. (2022). User Manual for Stock Synthesis model Version 330.19. 29 March2022.

Methot, R. D., Jr and I. G. Taylor (2011). "Adjusting for bias due to variability of estimated recruitments in fishery assessment models." Canadian Journal of Fisheries and Aquatic Sciences 68(10): 1744-1760.

Methot, R. D. and C. R. Wetzel (2013). "Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management." Fisheries Research 142: 86-99.

Methot, R. D. (2005) Technical description of the stock synthesis II assessment program: Version 1.17 (March, 2005), 54p.

Nikolic, N., Fonteneau, A., Hoarau, L., Morandeau, G., Puech, A., Bourjea, J. (2013). Short review on biology, structure, and migration of Thunnus alalunga in the Indian Ocean. Fifth Working Party on Temperate Tunas, Busan, Korea, 28–31 July 2014. Indian Ocean Tuna Commission. IOTC-2014-WPTmT05-13.

Nishida, T.; Dhurmeea, Z. (2019). Review of Indian Ocean albacore biological parameters for stock assessments. Seventh Working Party on Temperate Tunas: Data Preparatory Meeting, Kuala Lumpa, Malaysia, 14–17 January 2019. Indian Ocean Tuna Commission. IOTC-2019-WPTmT07(DP)-12

Nishida, T., T. Matsumoto and T. Kitakado (2012). Stock and risk assessments on albacore (*Thunnus alalunga*) in the Indian Ocean based on AD Model Builder implemented Age-Structured Production Model (ASPM).

Nishida, T.; Sato, K.; Chang, Y.; Matsumoto, T.; Lee, Y.C.; Kitakado, T. (2016). Stock assessments of albacore (*Thunnus alalunga*) in the Indian Ocean using Statistical-Catch-at-Age (SCAA). Age-Structured Production Model (ASPM). Sixth Working Party on Temperate Tunas, Shanghai, China, 18–21 July 2016. Indian Ocean Tuna Commission. IOTC-2014-WPTmT06-21.

Saji, N.H.; Yamagata, T. (2003). Possible impacts of Indian Ocean Dipole mode events on global climate. Climate Research 25(2) 151-169.

Santiago, J. and H. Arrizabalaga (2005). "An integrated growth study for North Atlantic albacore (*Thunnus alalunga*, Bonn. 1788)." ICES Journal of Marine Science 62(4): 740-749.

Schnute J.T., and R. Hilborn. 1993. Analysis of contradictory data sources in fish stock assessment. Canadian Journal of Fisheries and Aquatic Sciences, 50 (9): 1916-1923.

Wells, R. J. D., S. Kohin, S. L. H. Teo, O. E. Snodgrass and K. Uosaki (2013). "Age and growth of North Pacific albacore (*Thunnus alalunga*): Implications for stock assessment." Fisheries Research 147: 55-62.

Williams, A. J., J. H. Farley, S. D. Hoyle, C. R. Davies and S. J. Nicol (2012). "Spatial and sex-specific variation in growth of albacore tuna (Thunnus alalunga) across the South Pacific Ocean." PLoS One 7(6): e39318.

Xu, Y., T. Sippel, S. L. H. Teo, K. Piner, K.-S. Chen and R. J. Wells (2014). A comparison study of North Pacific albacore (Thunnus alalunga) age and growth among various sources, ISC/14/ALBWG/04 ISC Albacore Working Group Meeting, 14-28 April 2014, La Jolla, CA, 92037, USA.

Yeh, S.; Hui, C.; Treng, T.; Kuo, C. (1995). Indian Ocean albacore stock structure studies by morphometric and DNA sequence methods. Sixth Session of the Expert Consultations on Indian Ocean Tunas, Columbo, Sri Lanka, 25–29 September 1995. Indian Ocean Tuna Commission. IOTC-1995-EC602-25.

Zhu, J., Y. Chen, X. Dai, S. J. Harley, S. D. Hoyle, M. N. Maunder and A. M. Aires-da-Silva (2012). "Implications of uncertainty in the spawner–recruitment relationship for fisheries management: An illustration using bigeye tuna (*Thunnus obesus*) in the eastern Pacific Ocean." Fisheries Research 119-120: 89-93.

8 Tables

Table 1. Fishery and survey definitions for the Indian Ocean Albacore Assessment

ID Number	Fleet/ Survey Name	Gear(s)	Region	Quarter	Selectivty	Notes
1	F1_LL1_Q1	Longline	1	1	Estimated double normal	
2	F2_LL1_Q2	Longline	1	2	Estimated double normal	
3	F3_LL1_Q3	Longline	1	3	Estimated double normal	
4	F4_LL1_Q4	Longline	1	4	Estimated double normal	
5	F5_LL2_Q1	Longline	2	1	Estimated double normal	
6	F6_LL2_Q2	Longline	2	2	Estimated double normal	
7	F7_LL2_Q3	Longline	2	3	Estimated double normal	
8	F8_LL2_Q4	Longline	2	4	Estimated double normal	
9	F9_LL3_Q1	Longline	3	1	Estimated double normal[1]	
10	F10_LL3_Q2	Longline	3	2	Estimated double normal[1]	
11	F11_LL3_Q3	Longline	3	3	Estimated double normal[1]	
12	F12_LL3_Q4	Longline	3	4	Estimated double normal[1]	
13	F13_LL4_Q1	Longline	4	1	Estimated double normal[1]	
14	F14_LL4_Q2	Longline	4	2	Estimated double normal[1]	
15	F15_LL4_Q3	Longline	4	3	Estimated double normal[1]	
16	F16_LL4_Q4	Longline	4	4	Mirrored to Fleet 15	
17	F17_DN3	Driftnet	3	NA	Fixed Double normal	
18	F18_DN4	Driftnet	4	NA	Mirrored to fleet 17	
19	F19_PS1	Purse Seine	1	NA	Fixed Double normal	
20	F20_Other1	Other*	1	NA	Mirrored to fleet 17	
21	F21_Other2	Other*	2	NA	Mirrored to fleet 17	
22	F22_Other3	Other*	3	NA	Mirrored to fleet 17	
23	F23_Other4	Other*	4	NA	Mirrored to fleet 17	
24	LLCPUE1_Q1	Longline	1	1	Mirrored to fleet 1	
25	LLCPUE1_Q2	Longline	1	2	Mirrored to fleet 2	
26	LLCPUE1_Q3	Longline	1	3	Mirrored to fleet 3	
27	LLCPUE1_Q4	Longline	1	4	Mirrored to fleet 4	
28	LLCPUE2_Q1	Longline	2	1	Mirrored to fleet 5	
29	LLCPUE2_Q2	Longline	2	2	Mirrored to fleet 6	
30	LLCPUE2_Q3	Longline	2	3	Mirrored to fleet 7	
31	LLCPUE2_Q4	Longline	2	4	Mirrored to fleet 8	
32	LLCPUE3_Q1	Longline	3	1	Mirrored to fleet 9	
33	LLCPUE3_Q2	Longline	3	2	Mirrored to fleet 10	
34	LLCPUE3_Q3	Longline	3	3	Mirrored to fleet 11	
35	LLCPUE3_Q4	Longline	3	4	Mirrored to fleet 12	
36	LLCPUE4_Q2	Longline	4	2	Mirrored to fleet 14	
37	LLCPUE4_Q3	Longline	4	3	Mirrored to fleet 15	
38	DNCPUE4	Driftnet	4	NA	Mirroed to fleet 17	

*Other inclues: Coastal Longline, gillnet, trolling, handlines, and artisanal gear.

1. LL3 and LL4 are not constrained to approximate full selectivity for the largestlengths

Table 2. Recent catch data for Albacore in the Indian ocean, mean and max catch by region are calculated over the 2014-2023 period .

Fleet_Name	Region	Seas	MeanCatch	MaxCatch
F1_LL1_Q1	1	1	1,790	2,224
F2_LL1_Q2	1	2	425	760
F3_LL1_Q3	1	3	530	1,120
F4_LL1_Q4	1	4	3,998	5,983
F5_LL2_Q1	2	1	1,527	2,091
F6_LL2_Q2	2	2	1,603	2,646
F7_LL2_Q3	2	3	1,537	2,261
F8_LL2_Q4	2	4	972	1,736
F9_LL3_Q1	3	1	1,063	2,173
F10_LL3_Q2	3	2	8,920	11,264
F11_LL3_Q3	3	3	6,344	7,890
F12_LL3_Q4	3	4	668	1,079
F13_LL4_Q1	4	1	697	2,591
F14_LL4_Q2	4	2	4,016	7,298
F15_LL4_Q3	4	3	2,026	3,882
F16_LL4_Q4	4	4	84	174

Table 3. Summary of SS3 specification options for the Indian Ocean albacore tuna assessment models. Other assumptions were constant for all models.

Model	Description	Base Parameters	Alternative Parameterization
Steepness0.70	SRR steepness at 0.70	*h*=0.80	*h*=0.70
Steepness0.80	SRR steepness at 0.90	h=0.80	*h*=0.90
Length At Age	Increase variation in length-at-age for males and females.	CVs for length-at-age CV_young 0.06 CV_old 0.025	CV_young 0.10 CV_old 0.10
Length Weight Relationship	Use length-weight parameters from Kitakado et al. (2019)	a = 1.3718e-05 b = 3.0973	a = 6.9e-06 b = 3.2263
Natural Mortality	Age Specific Natural mortality	M=0.30 all ages	Lorenzen M, average of 0.3 for age 4+
Maturity	Use maturity at length from recent western Indian Ocean study (rather than South Pacific reproductive potential at length)	Length based Reproductive Potential from South Pacific albacore.	Length based maturity ogive from Dhrumeea et al (2016)
Model Weighting	Longline length data are up-weighted within model	Longline length data are weighted to 1/10 of the sample size after the Francis weighting	Longline length data are not weighted after the Francis weighting

Table 3 management quantities for the NW and SW diagnostic cases, all biomass is in mt.

Quantity	**NW Diagnostic**	**SW Diagnostic**
SSB_unfished_thousand_mt	116.59	132.85
B_MSY/SSB_unfished	0.212	0.218
Totbio_unfished	501,174	571,078
SmryBio_unfished	500,831	570,687
Recr_unfished_millions	14.8673	16.941
SSB_MSY_thousand_mt	24.696	28.937
SPR_MSY	0.261	0.267
annF_MSY	0.156	0.149
Catch_MSY	41,079	47,975
B_MSY/SSB_unfished	0.212	0.218
B_2023/BMSY	1.016	1.571
F_2023/FMSY	1.264	0.556

9 Figures

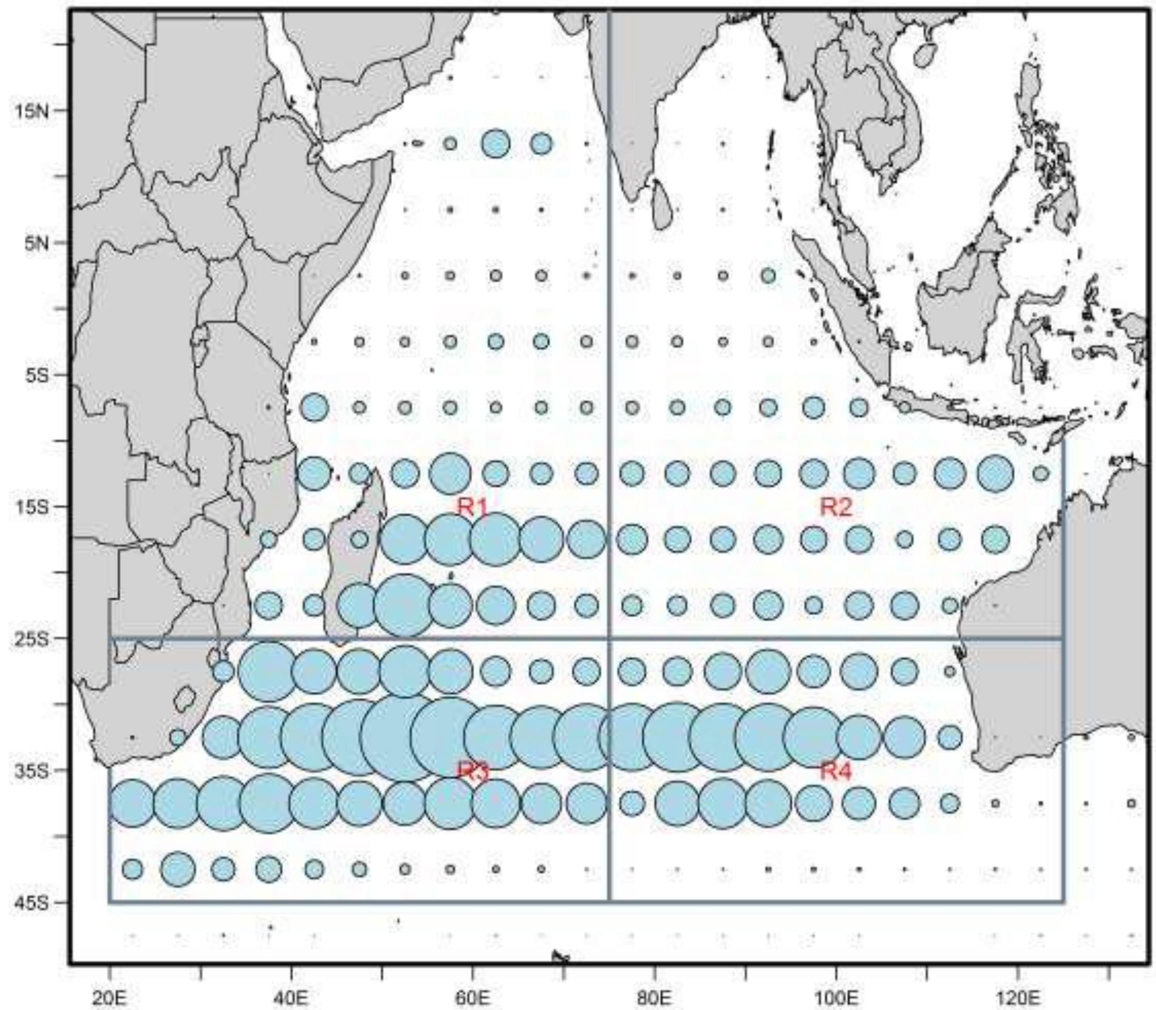


Figure 1. Spatial stratification of the Indian Ocean for the definition of the fisheries. The blue circles represent the aggregated Japanese and TW LL albacore catch (numbers of fish) by 5 degree cell from 1952-2017. The area of the circle is proportional to the magnitude of the catch (the largest circle represents a catch of 2.45 million fish).

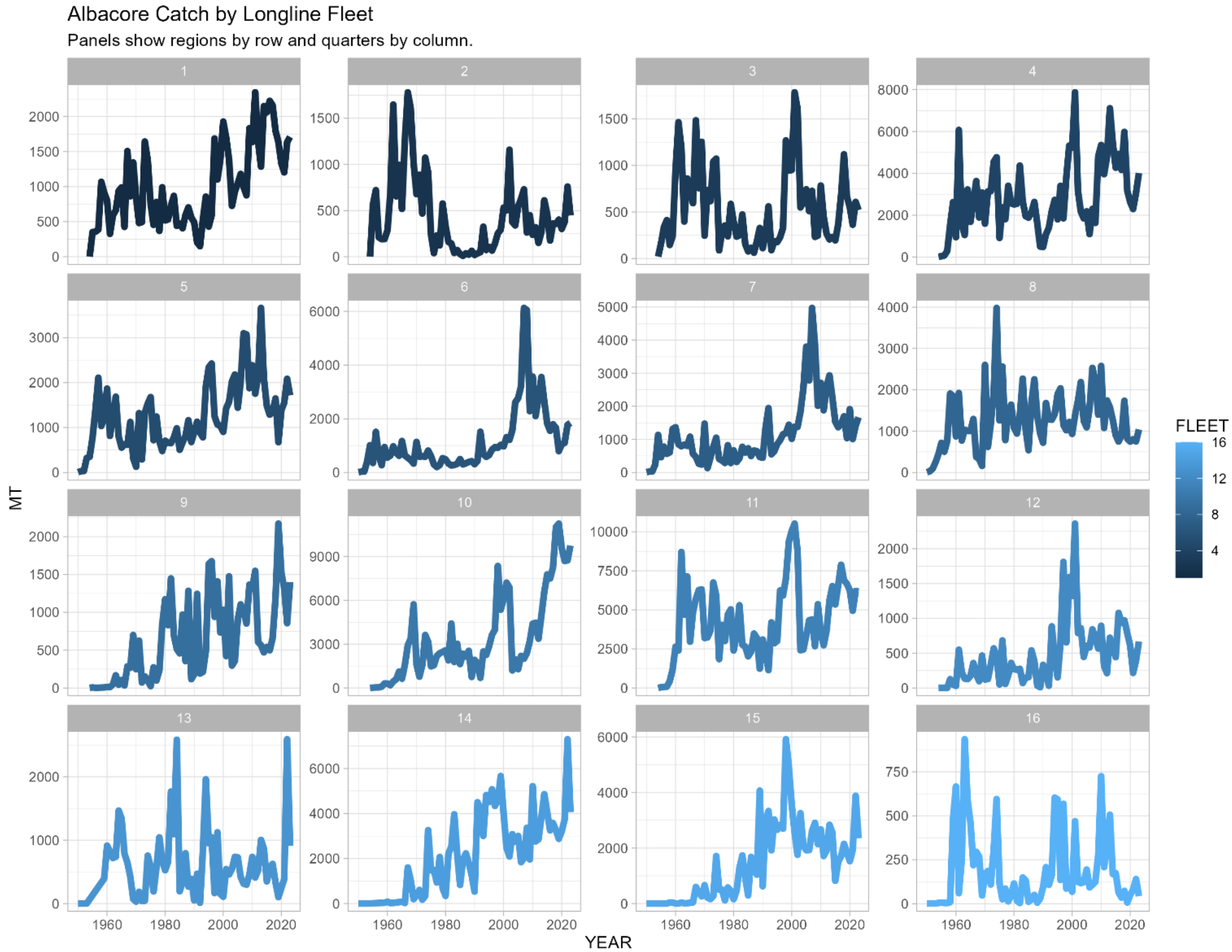


Figure 2. A comparison of Indian Ocean albacore longline catches. Rows indicate region (1-4) and columns indicate quarters (1-4).

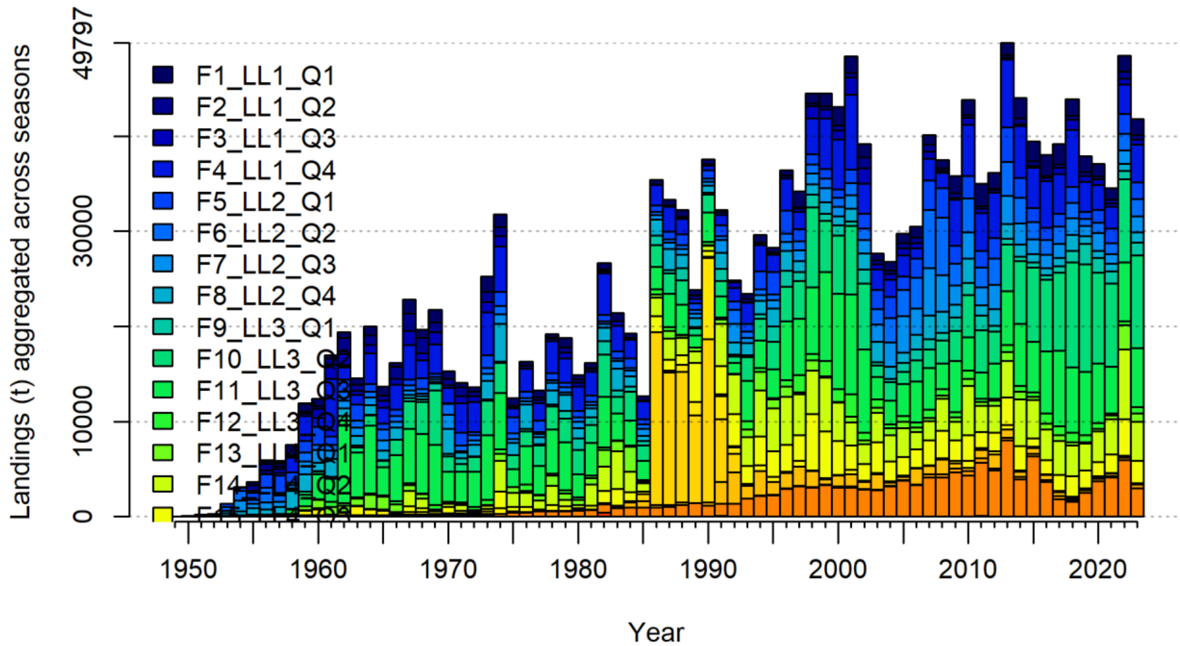


Figure 3. Total annual catch (1000s mt) of albacore tuna by fleet from 1950 to 2023. Fleet names indicate the fleet number, region and quarter, for example F1_LL1_Q1 is Fishing Fleet 1, in Region 1(LL1) and quarter1.

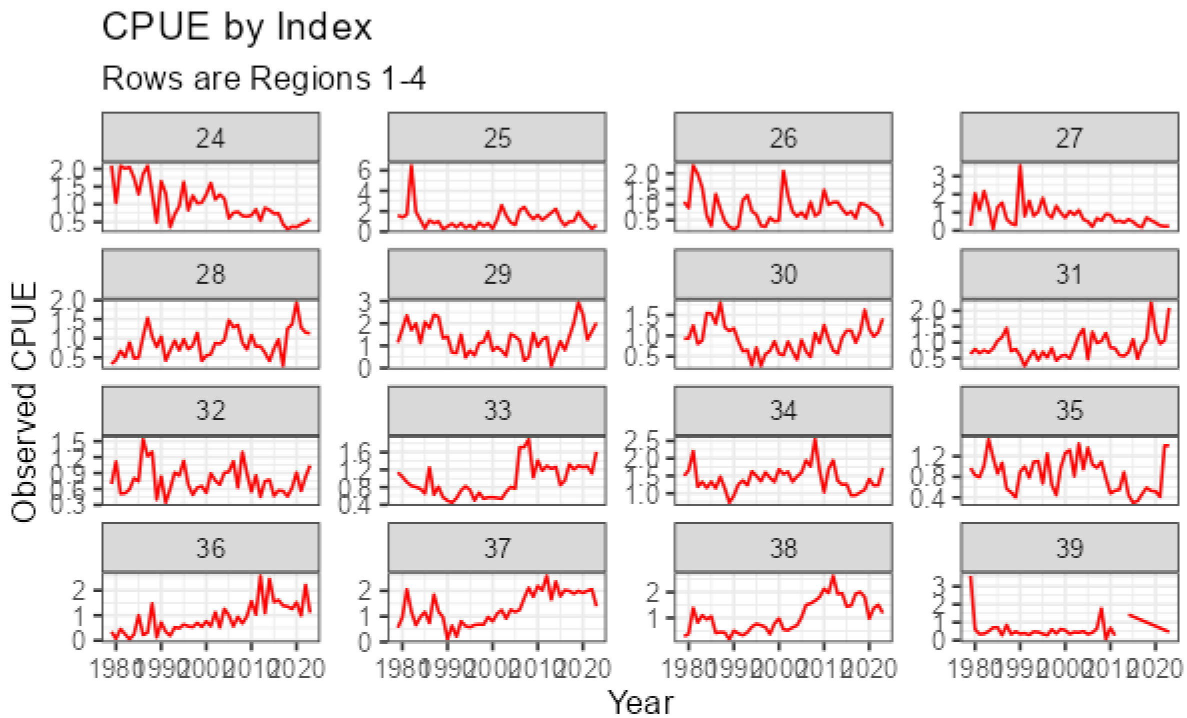


Figure 4. Standardized CPUE indices for the longline fisheries by region and quarter (LL 1-14) from 1975-2023. The figure shows the region in row, and individual quarterly CPUE series in each column.

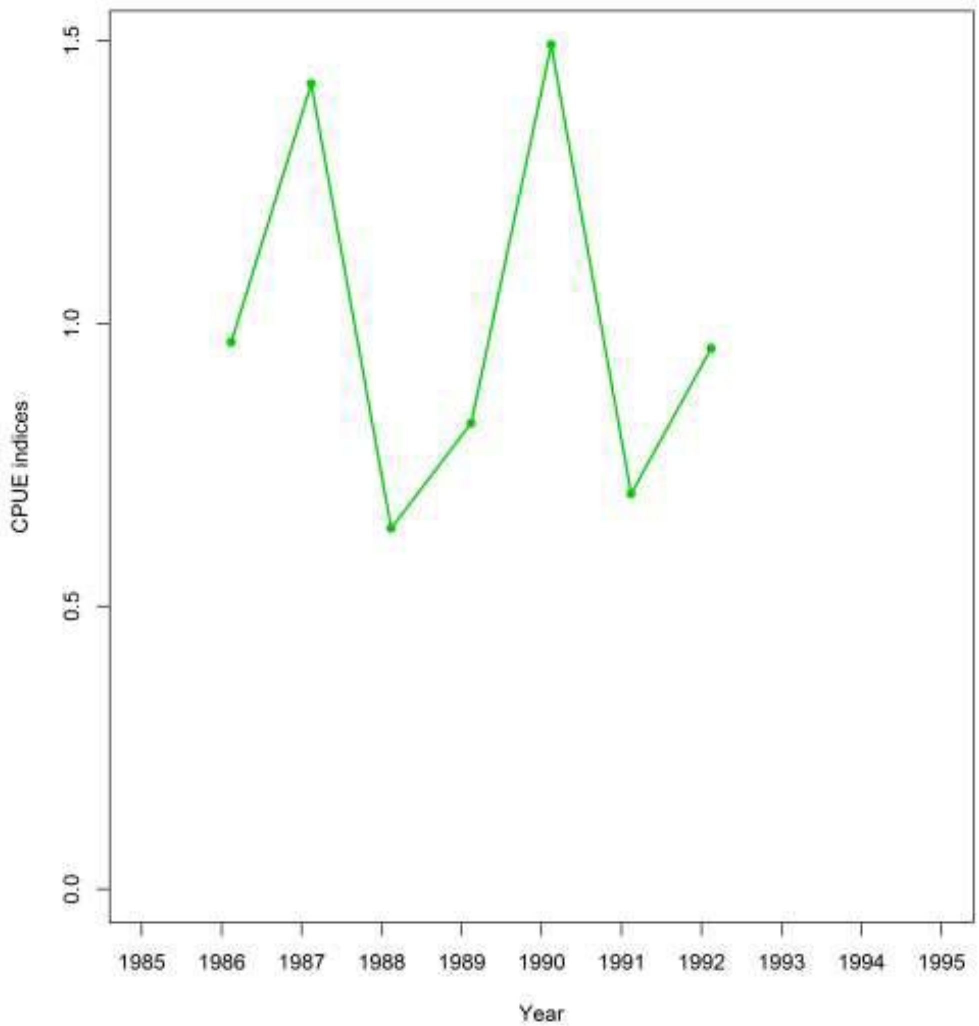


Figure 5 Annual drift net CPUE indices (source: Chang & Liu 1995).

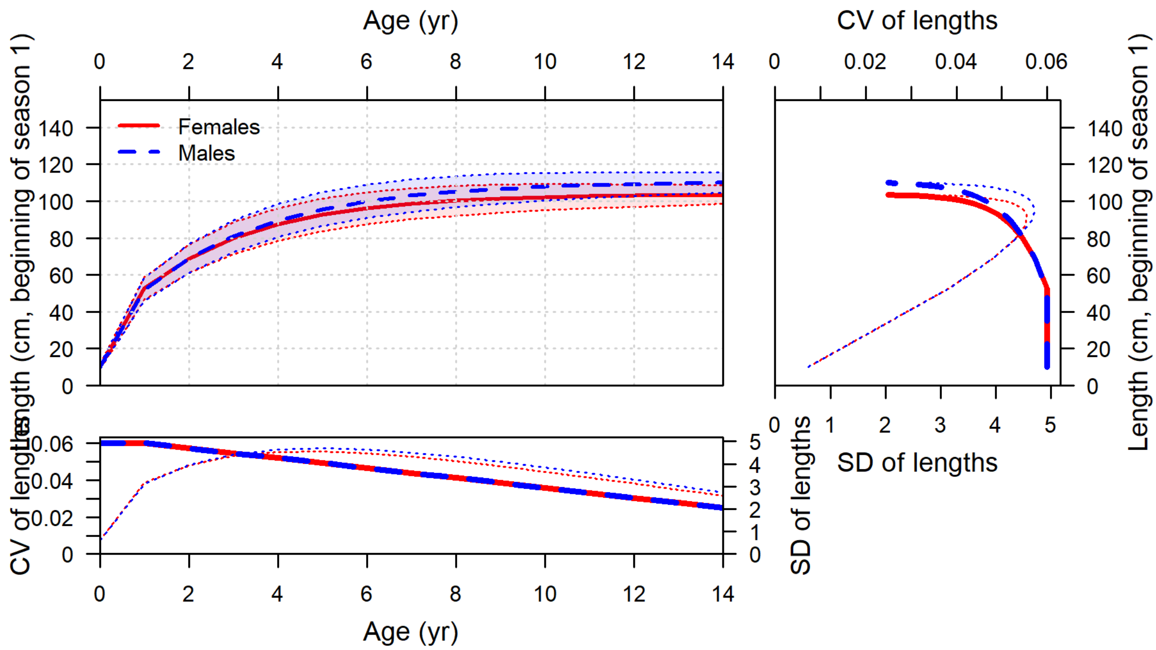


Figure 6. Growth function for female and male albacore.

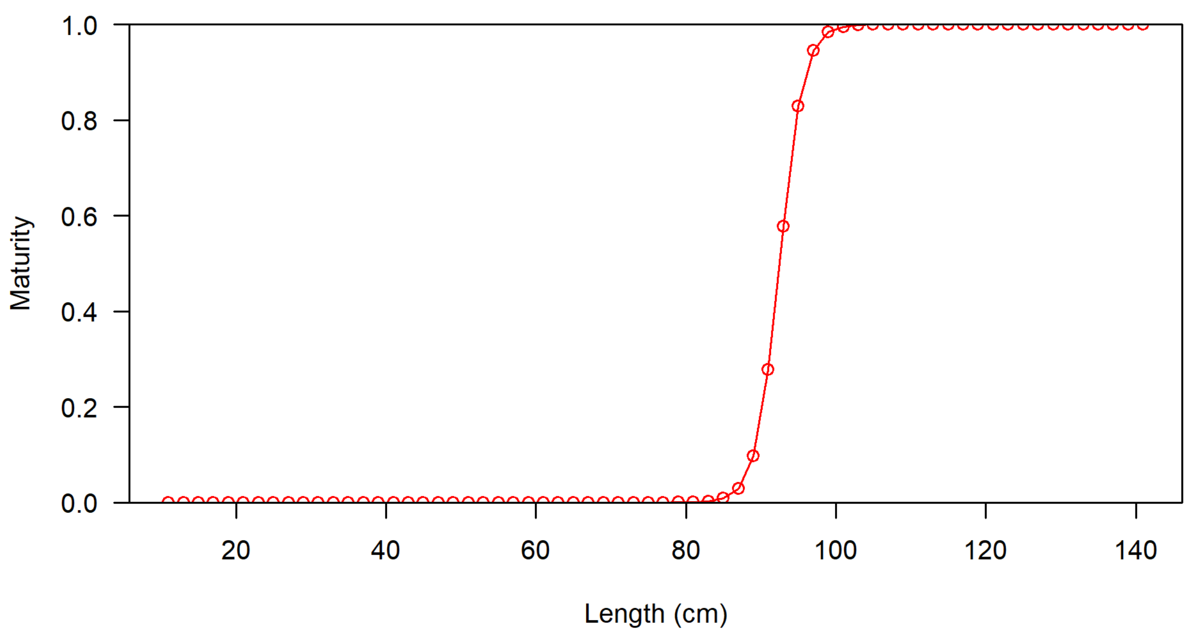


Figure 7 Maturity at length for female albacore in the Indian Ocean.

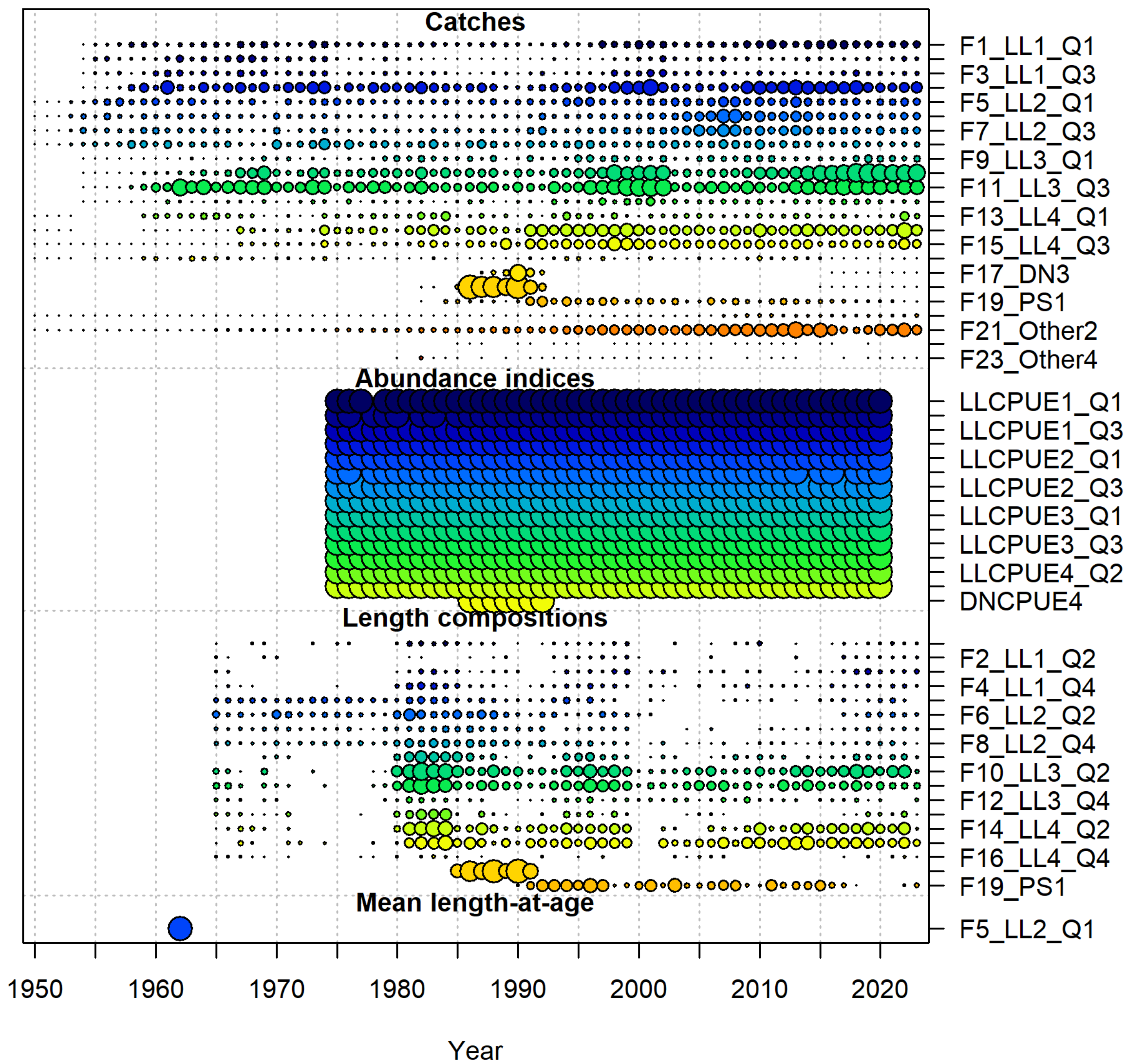


Figure 8: Temporal data coverage for the diagnostic case model for the assessment of albacore tuna.

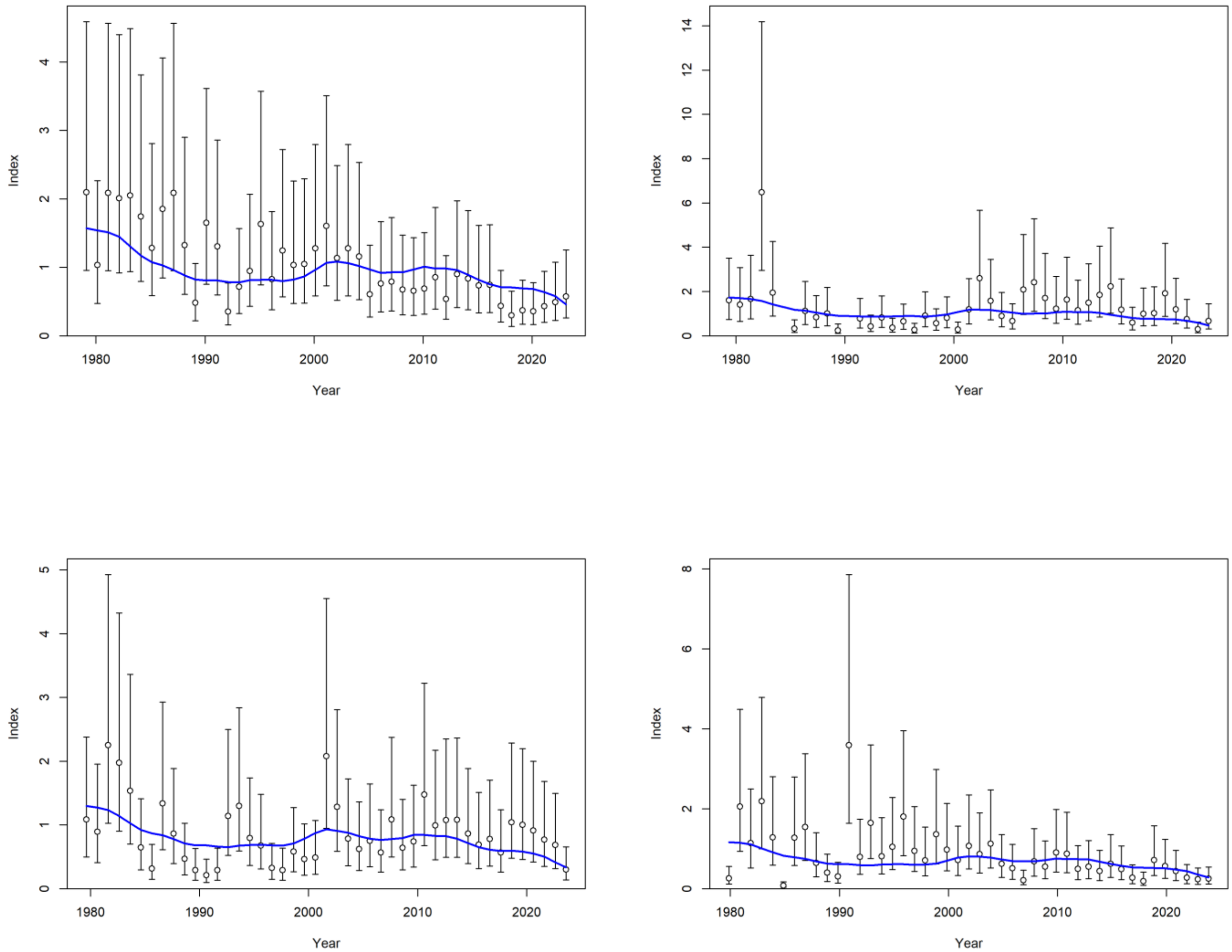


Figure 9: NW diagnostic case fit to the CPUE series (Fleets 24-27), from region 1. The top left panel is quarter 1, the top right is quarter 2, the bottom left is quarter 3, the bottom right is quarter 4.

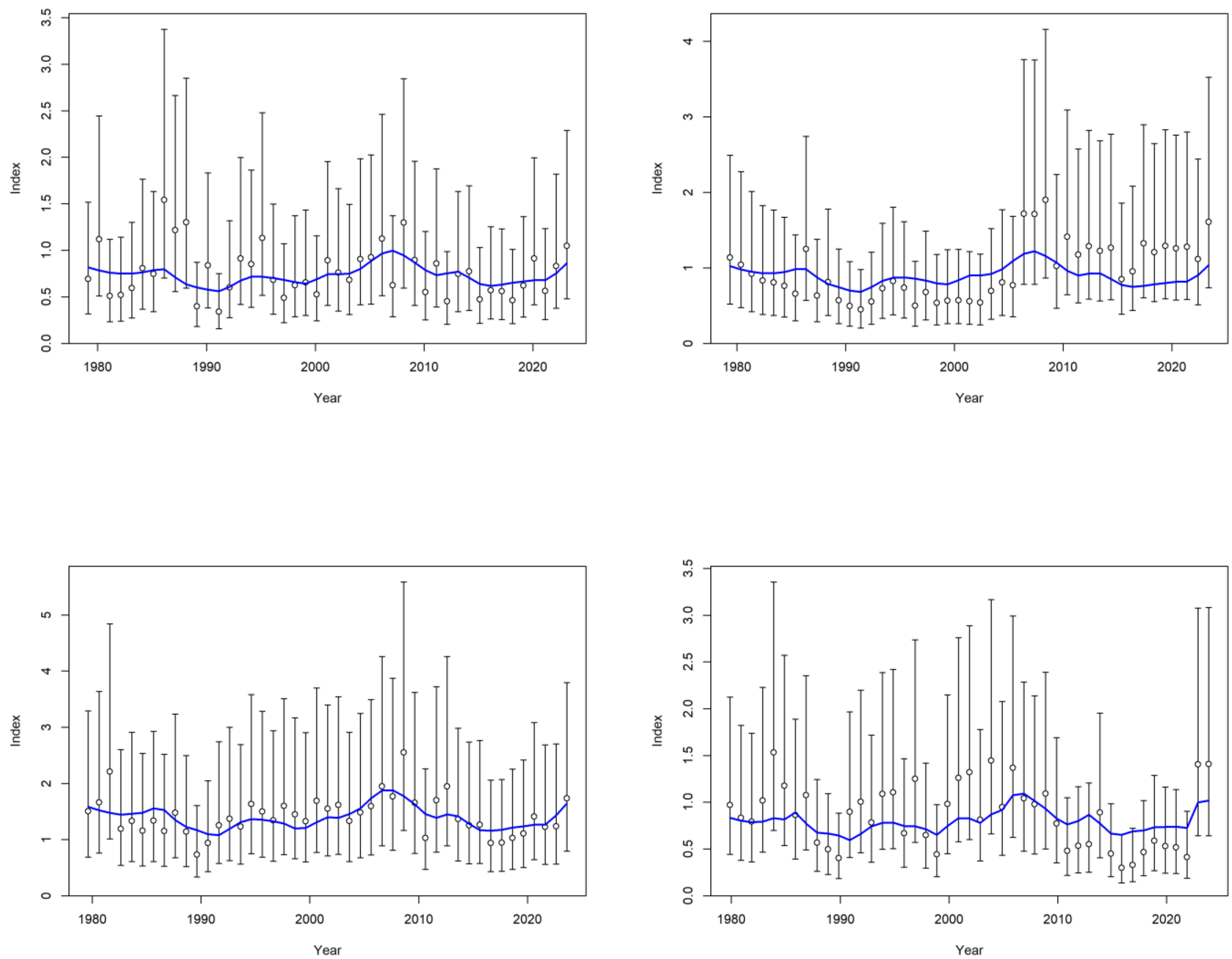
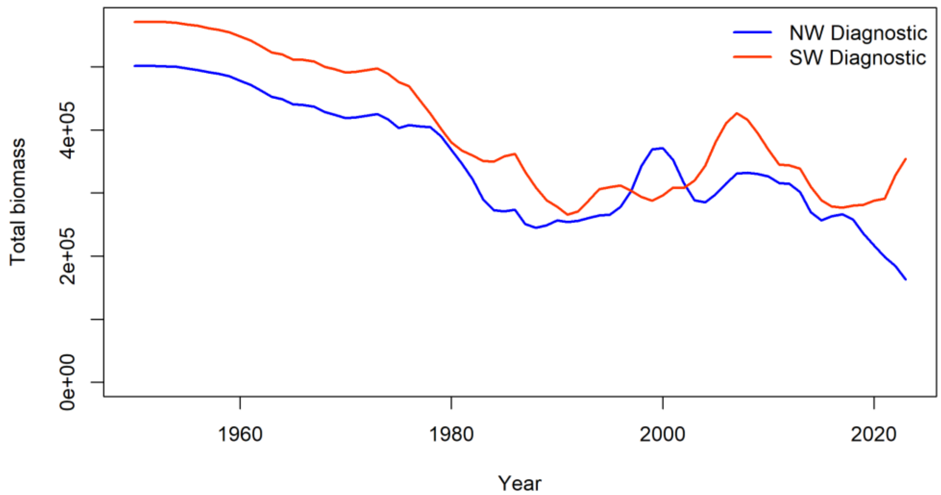


Figure 10: SW diagnostic case fit to the CPUE series (fleets 32-35), from region 3. The top left panel is quarter 1, the top right is quarter 2, the bottom left is quarter 3, the bottom right is quarter 4.



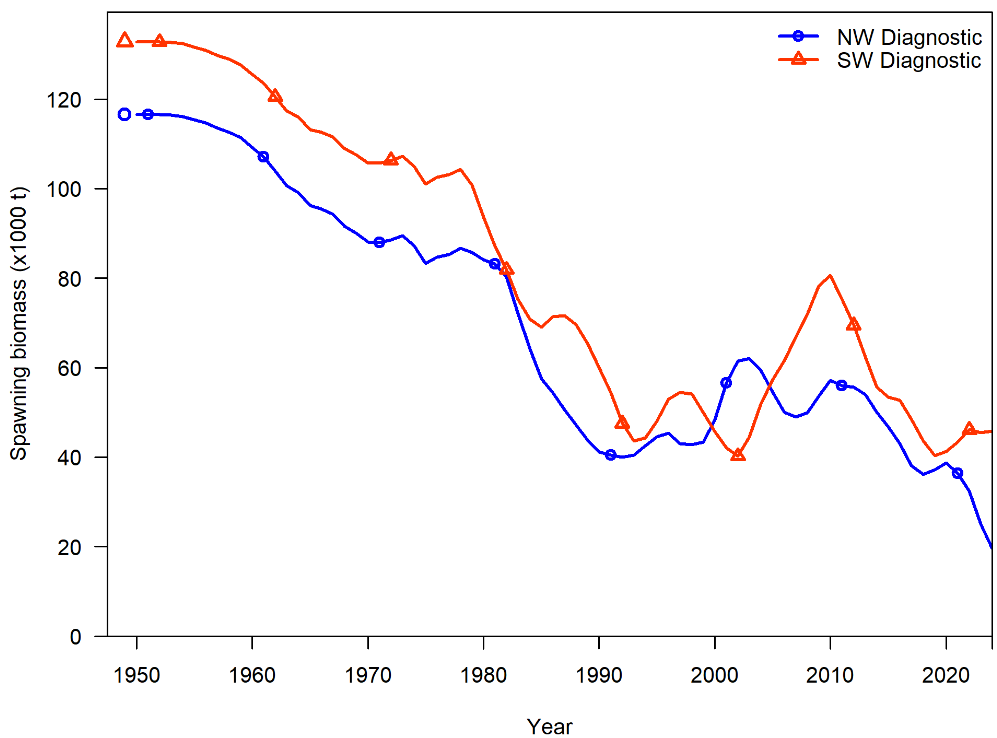


Figure 11: Total biomass (top) and spawning biomass for the diagnostic case parameterization models. The filled dot represents the pre-model estimate of unfished biomass.

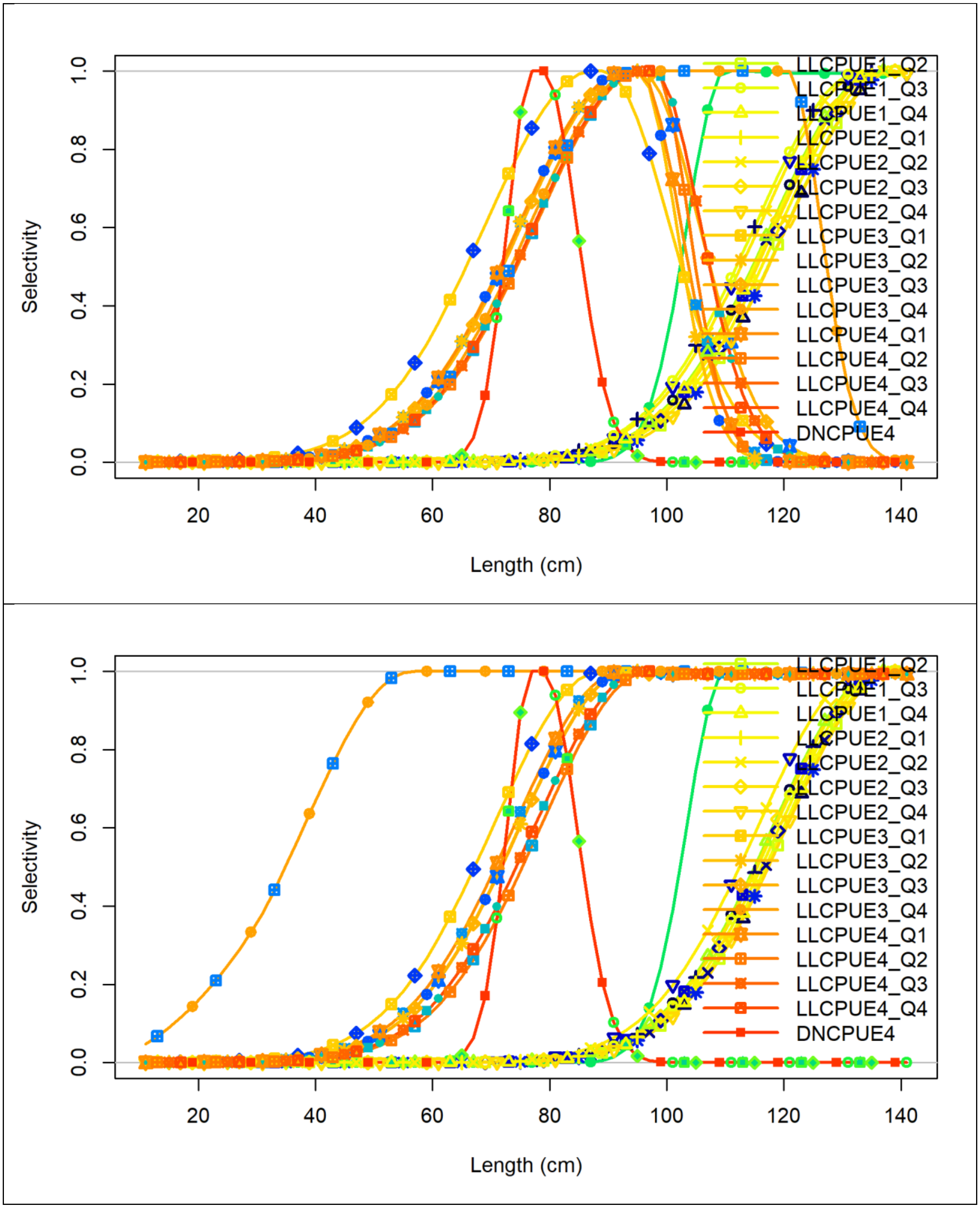


Figure 12: Selectivity curves estimated from the diagnostic case models (NW on top, SW on bottom) for the assessment of albacore tuna in the Indian Ocean.

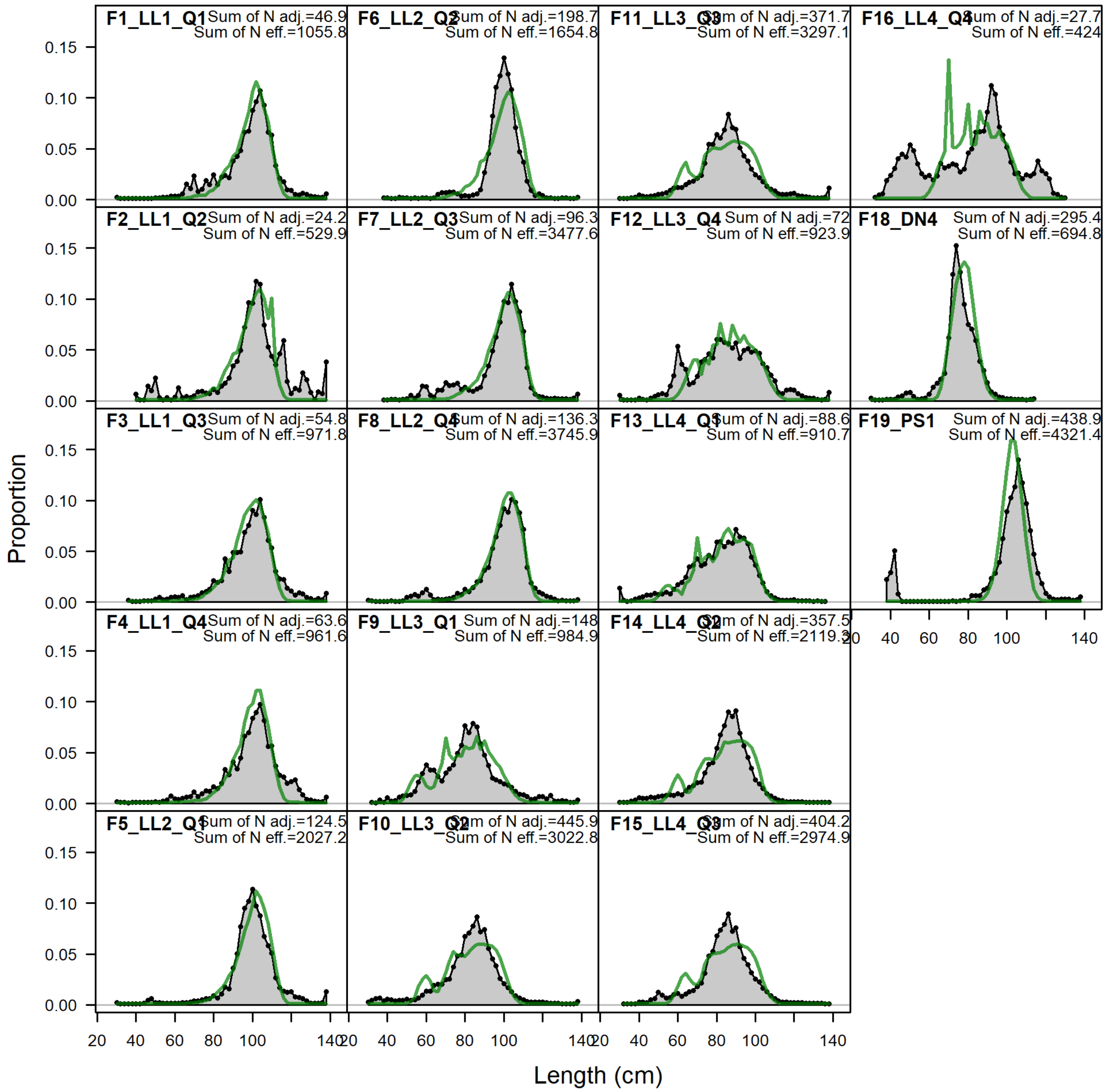


Figure 13. Fits to the length frequency data for the NW diagnostic case model for the assessment of albacore tuna in the Indian Ocean.

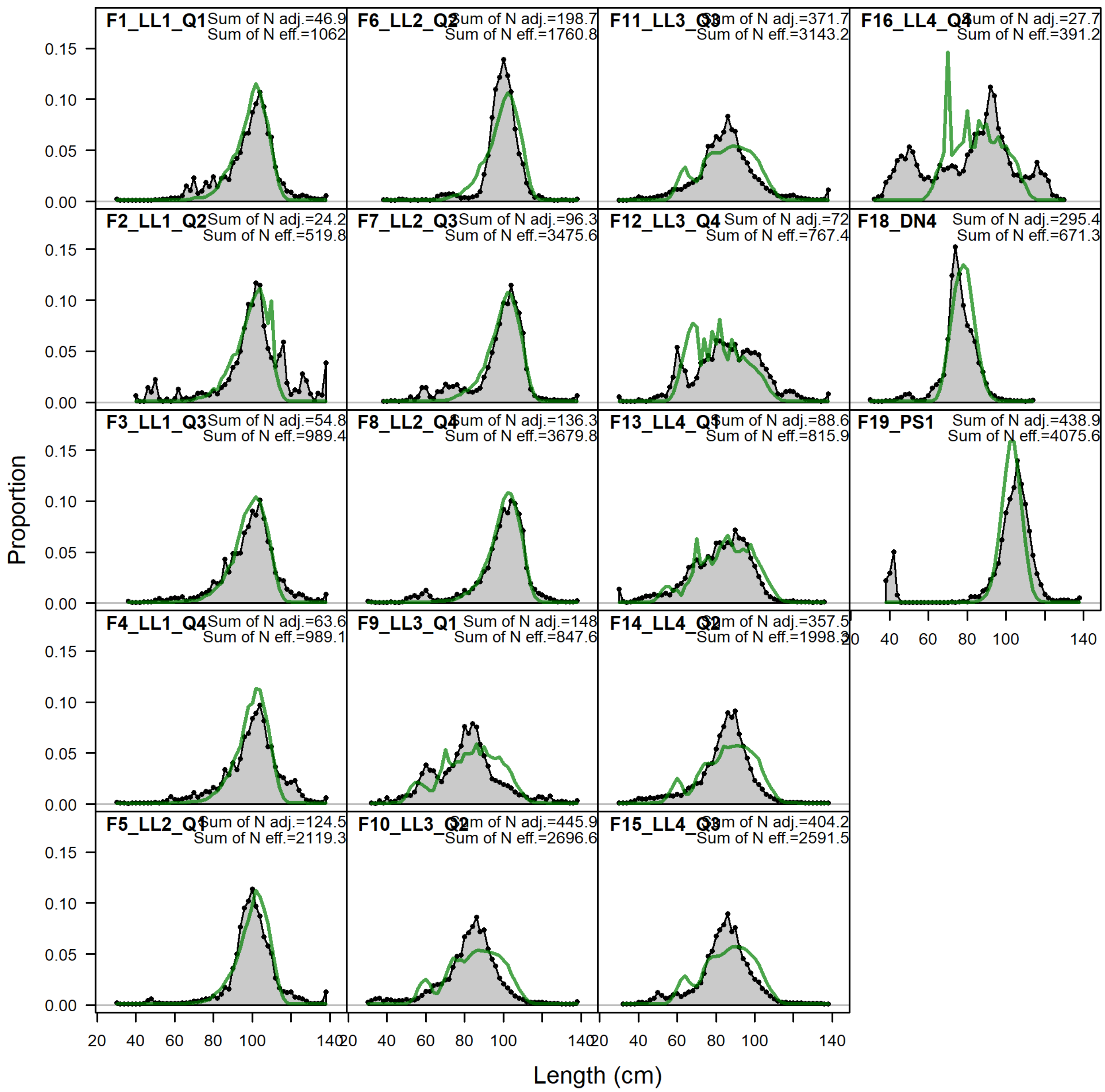


Figure 14. Fit to the length frequency data for the SW diagnostic case model for the assessment of albacore tuna in the Indian Ocean.

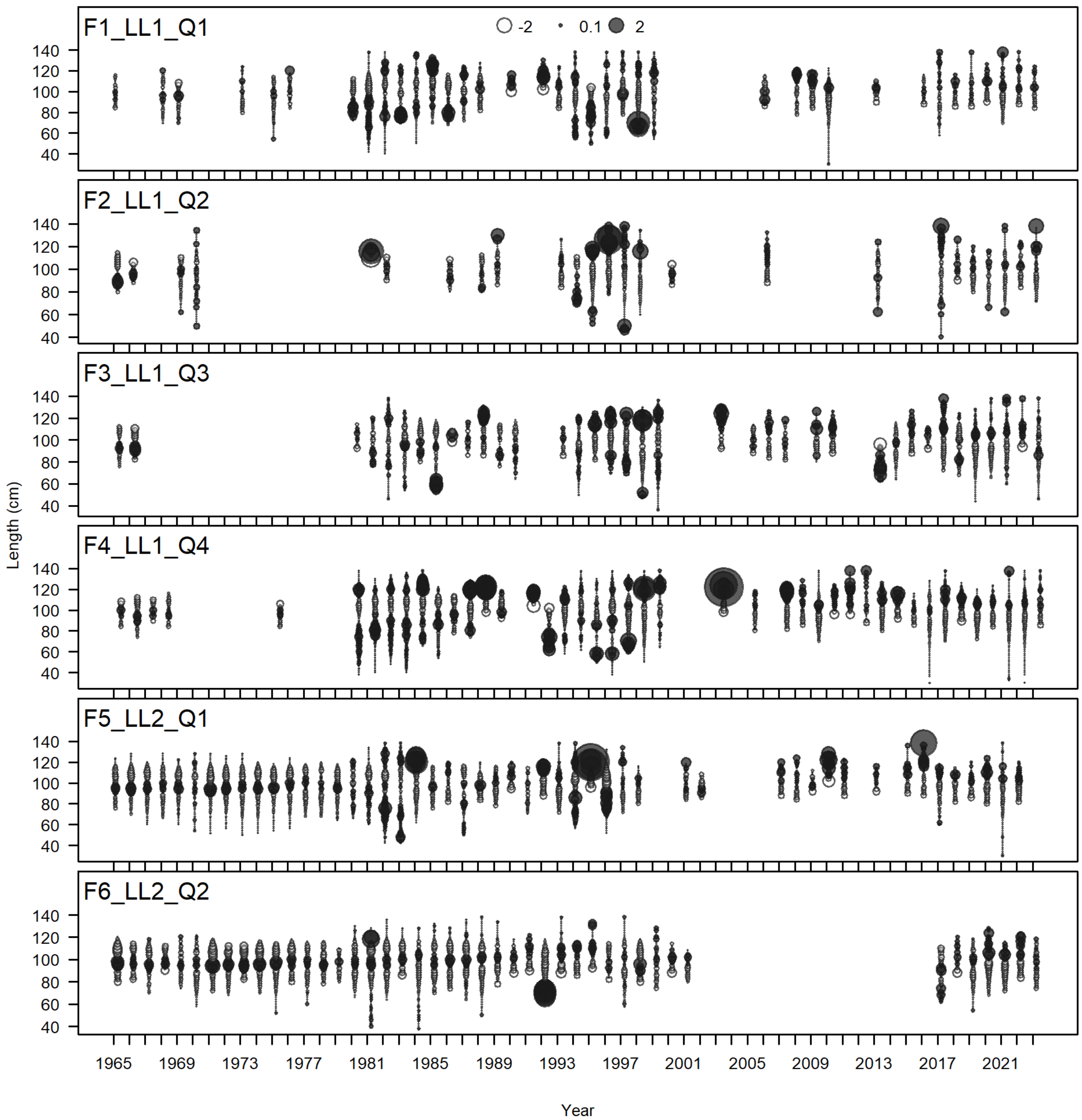


Figure 15. Residuals from the fit to the length frequency data for the NW diagnostic case model for the assessment of albacore in the Indian Ocean, Fleets 1-6. Closed bubbles are positive residuals and open bubbles are negative residuals, bubble sizes are scaled to maximum within each panel. Thus, comparisons across panels should focus on patterns, not bubble sizes.

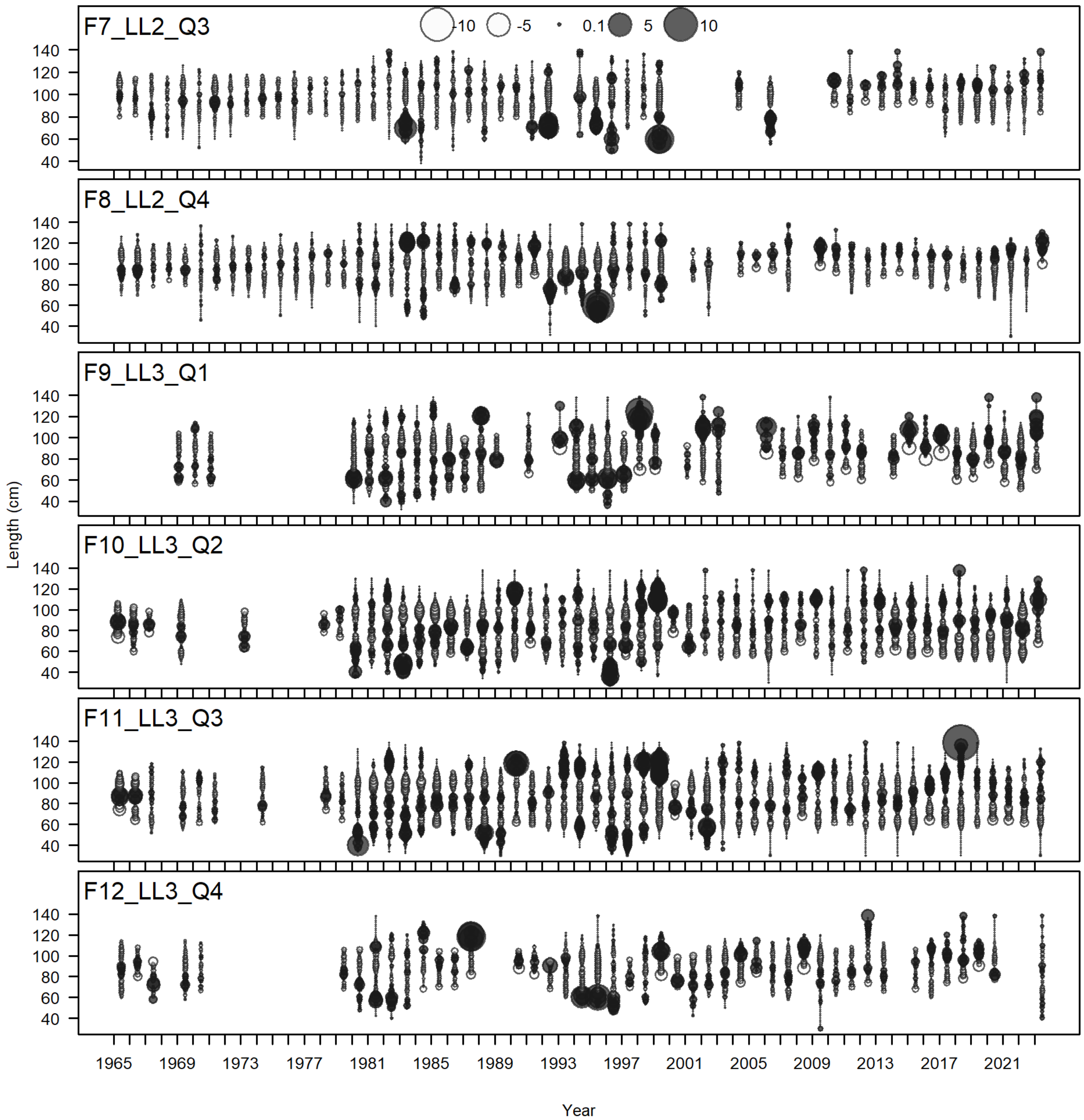


Figure 16. Residuals from the fit to the length frequency data for the NW diagnostic case model for the assessment of albacore in the Indian Ocean, Fleets 7-12. Closed bubbles are positive residuals and open bubbles are negative residuals, bubble sizes are scaled to maximum within each panel. Thus, comparisons across panels should focus on patterns, not bubble sizes.

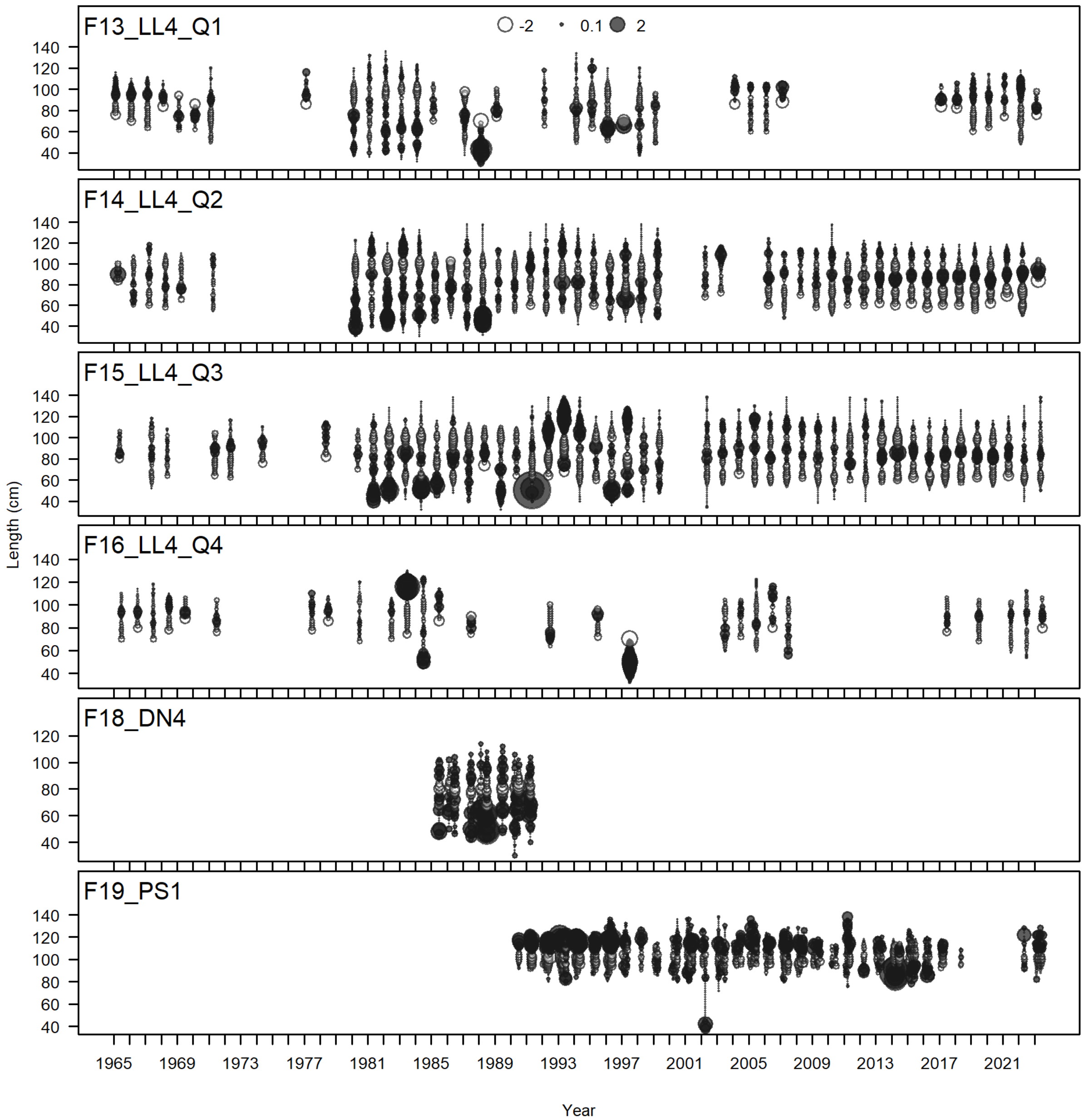


Figure 17 Residuals from the fit to the length frequency data for the NW diagnostic case model for the assessment of albacore n the Indian Ocean, Fleets 13-19. Closed bubbles are positive residuals and open bubbles are negative residuals, bubble sizes are scaled to maximum within each panel. Thus, comparisons across panels should focus on patterns, not bubble sizes.

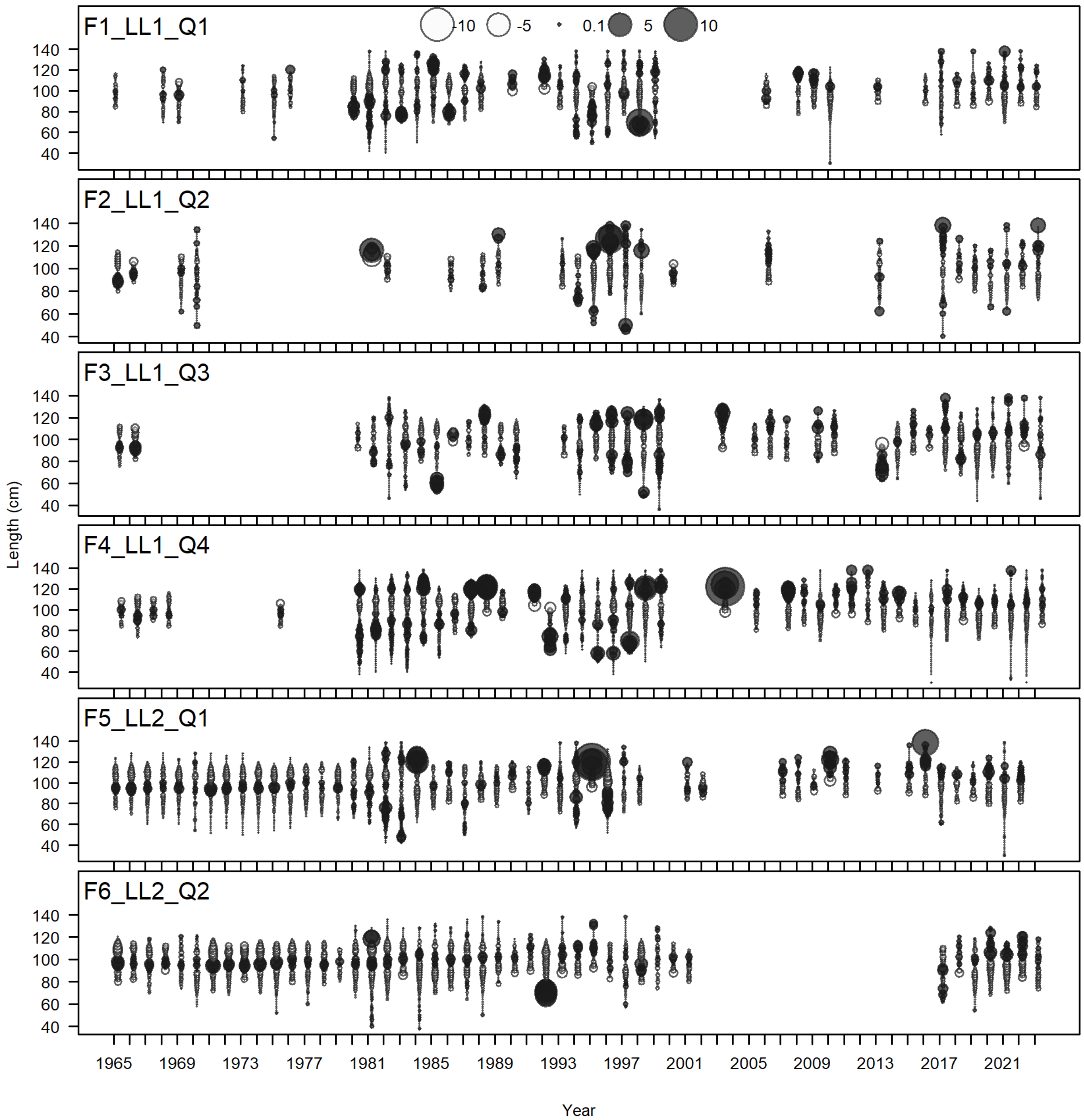


Figure 18. Residuals from the fit to the length frequency data for the SW diagnostic case model for the assessment of albacore in the Indian Ocean, Fleets 1-6. Closed bubbles are positive residuals and open bubbles are negative residuals, bubble sizes are scaled to maximum within each panel. Thus, comparisons across panels should focus on patterns, not bubble sizes.

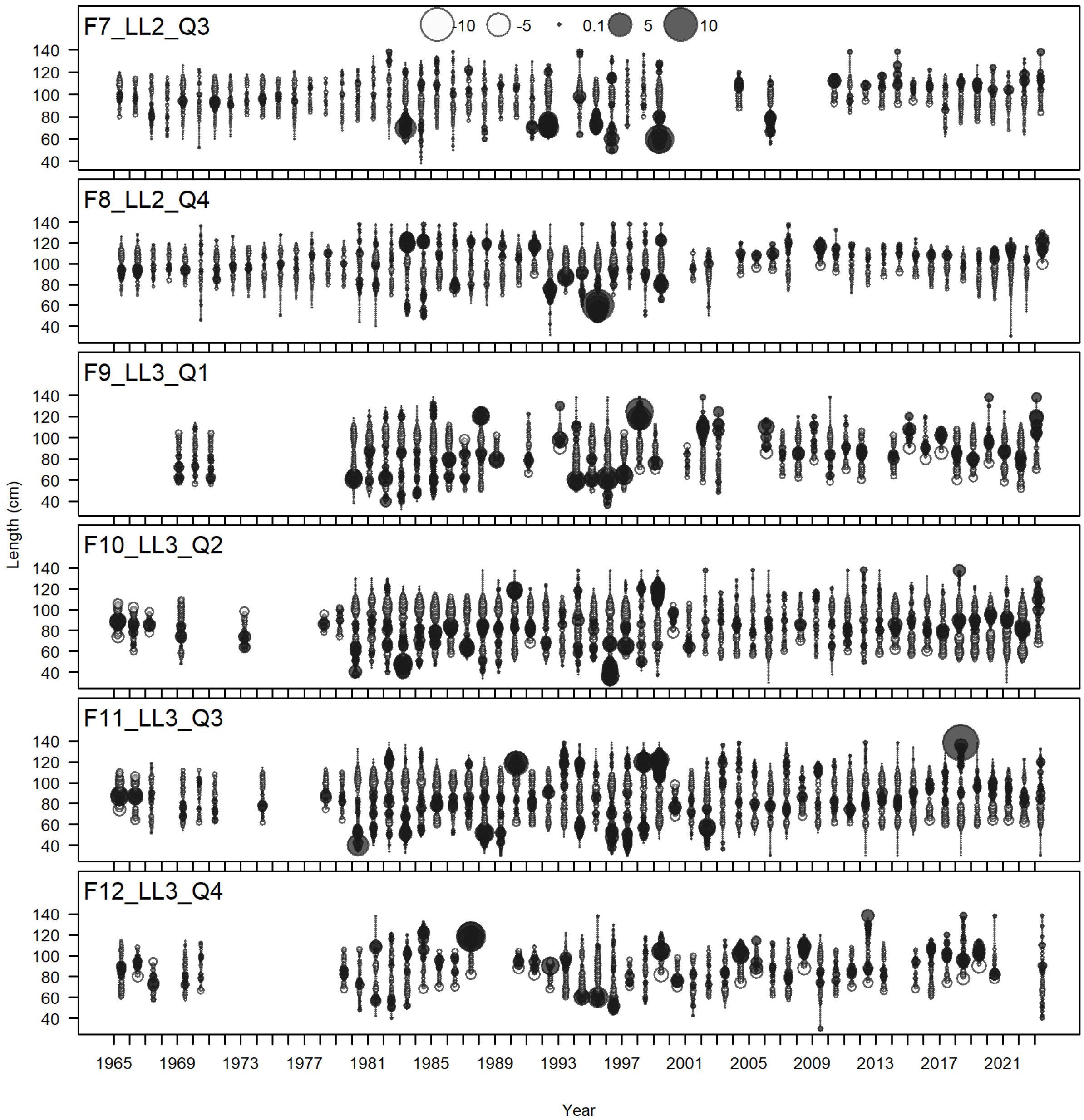


Figure 19. Residuals from the fit to the length frequency data for the SW diagnostic case model for the assessment of albacore in the Indian Ocean, Fleets 7-12. Closed bubbles are positive residuals and open bubbles are negative residuals, bubble sizes are scaled to maximum within each panel. Thus, comparisons across panels should focus on patterns, not bubble sizes.

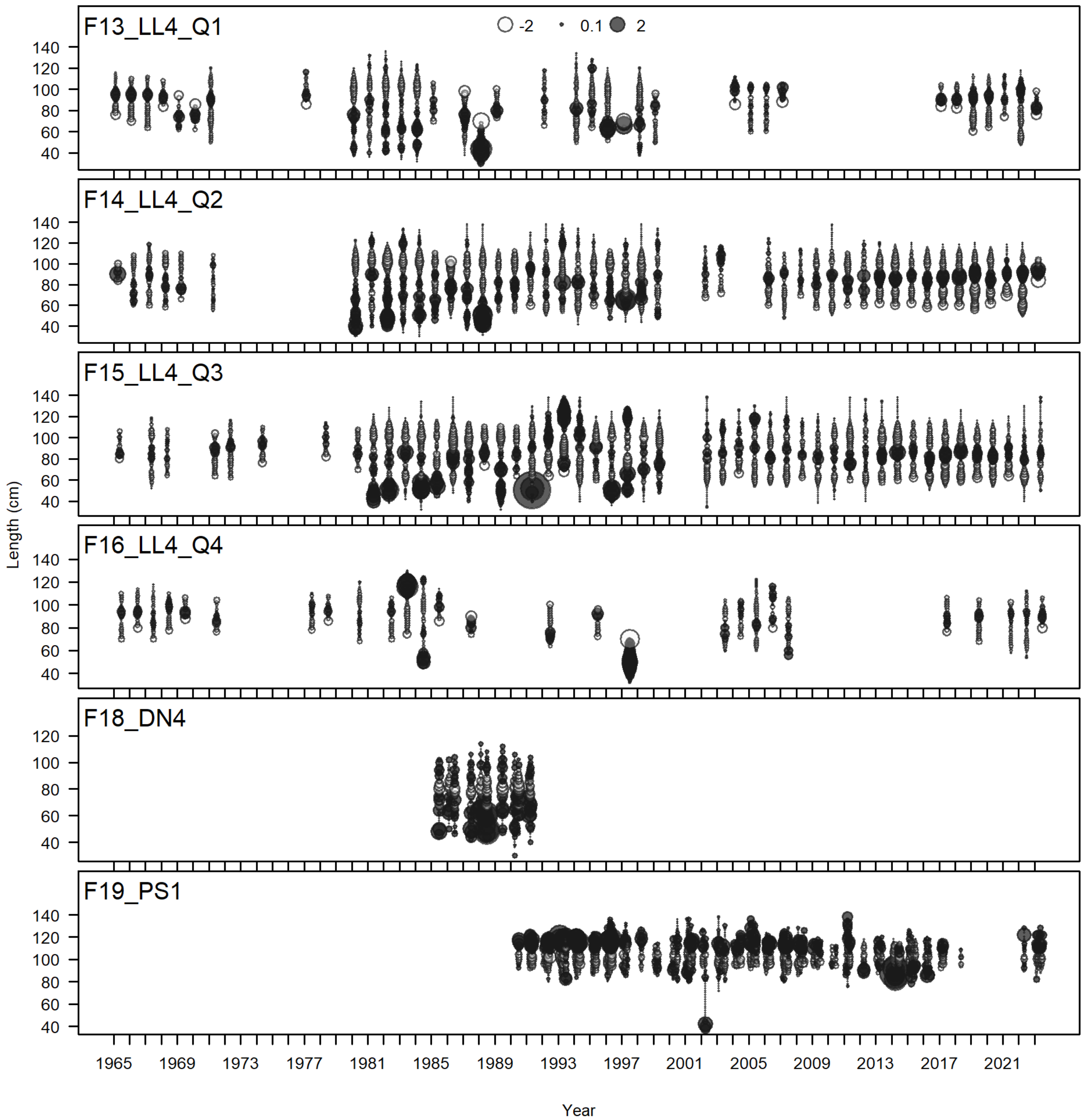


Figure 20. Residuals from the fit to the length frequency data for the SW diagnostic case model for the assessment of albacore in the Indian Ocean, Fleets 13-19. Closed bubbles are positive residuals and open bubbles are negative residuals, bubble sizes are scaled to maximum within each panel. Thus, comparisons across panels should focus on patterns, not bubble sizes.

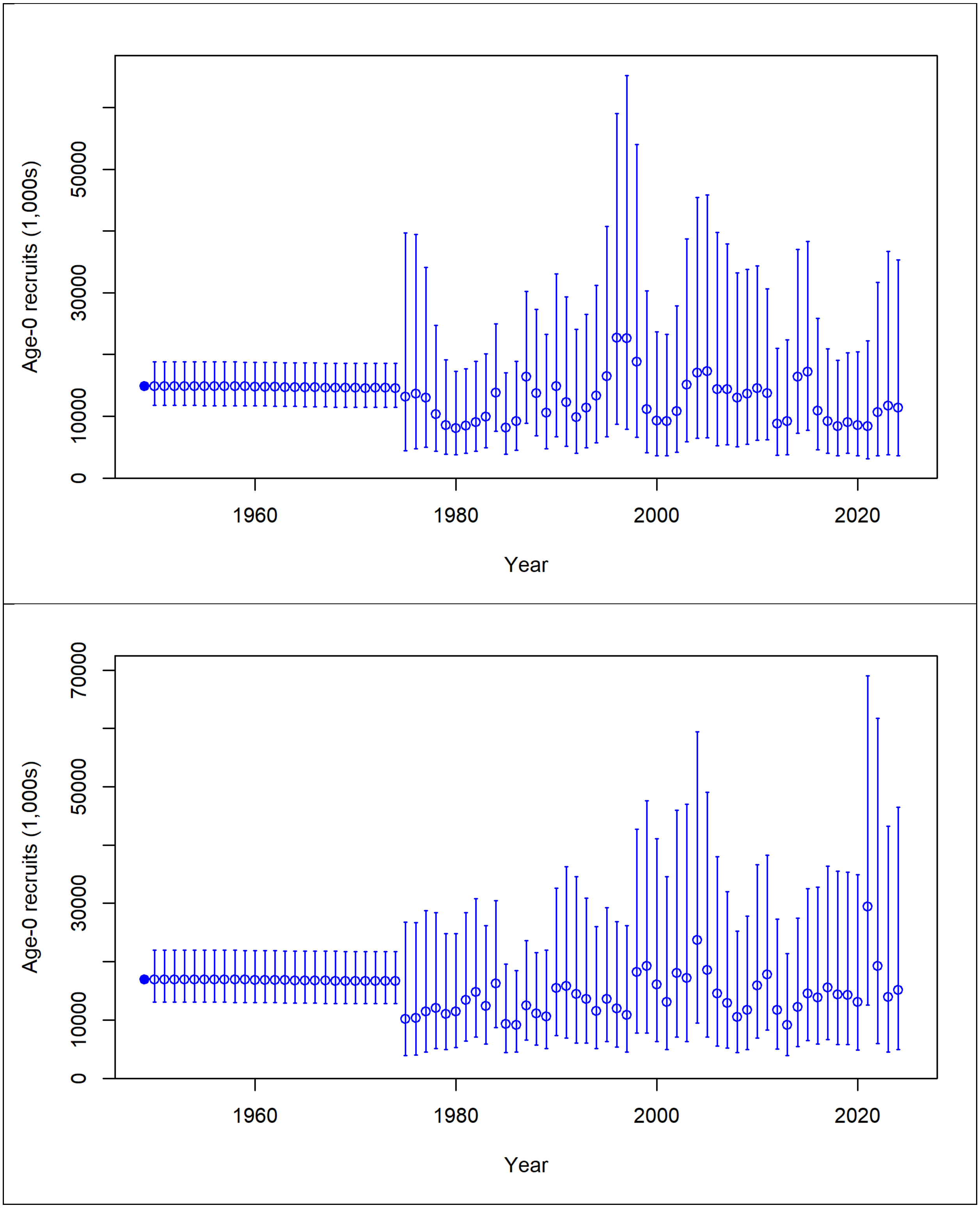


Figure 21. Estimated recruitment including the estimate of virgin recruitment (filled circle at the start of the time series) for the diagnostic case models (NW on top) for the assessment of albacore tuna in the Indian Ocean.

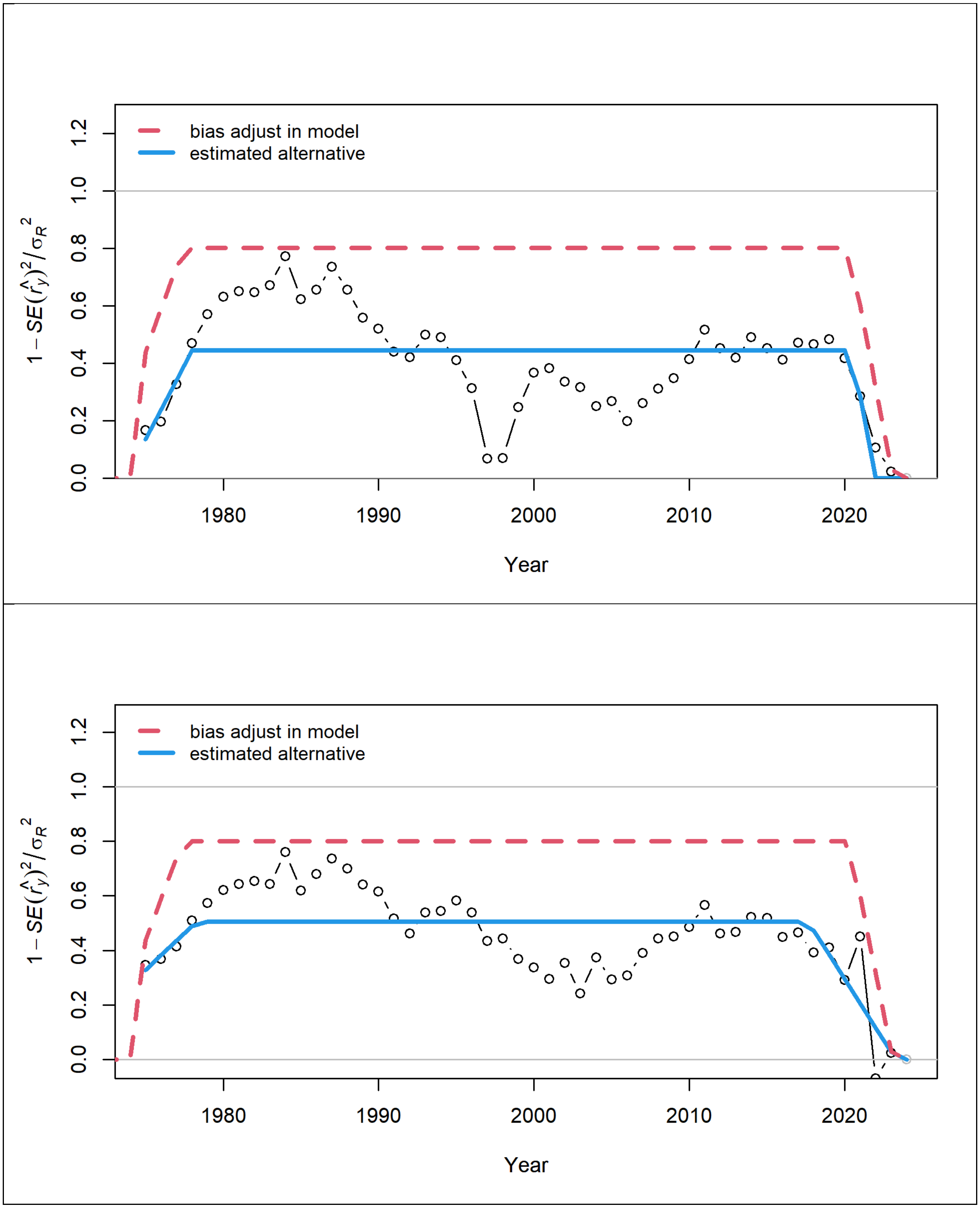


Figure 22. Estimated bias adjustment in the model, NW diagnostic case is on top.

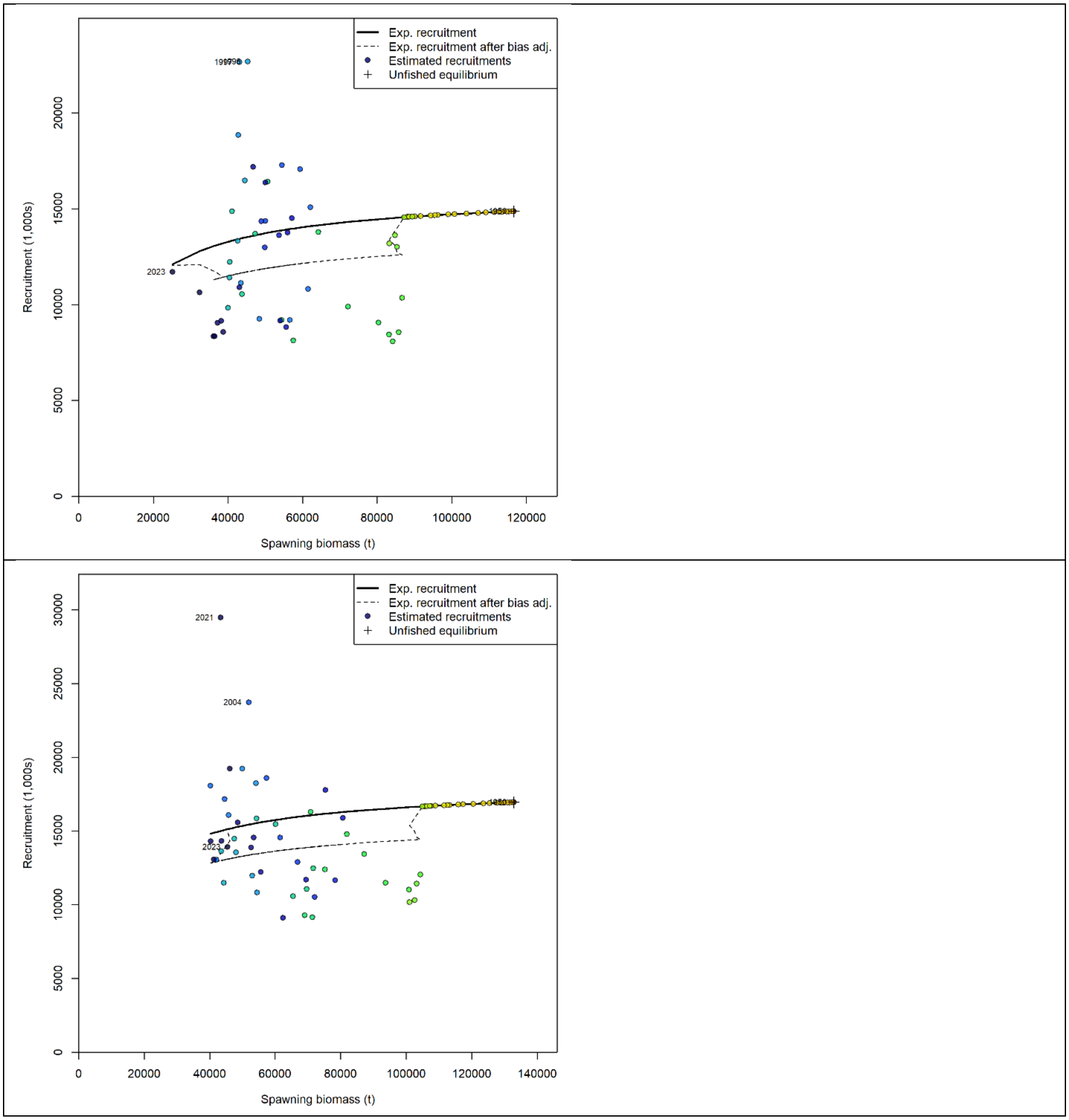


Figure 23. Stock recruitment curve used in the assessment and time series of estimates of recruitment deviations (colored points).

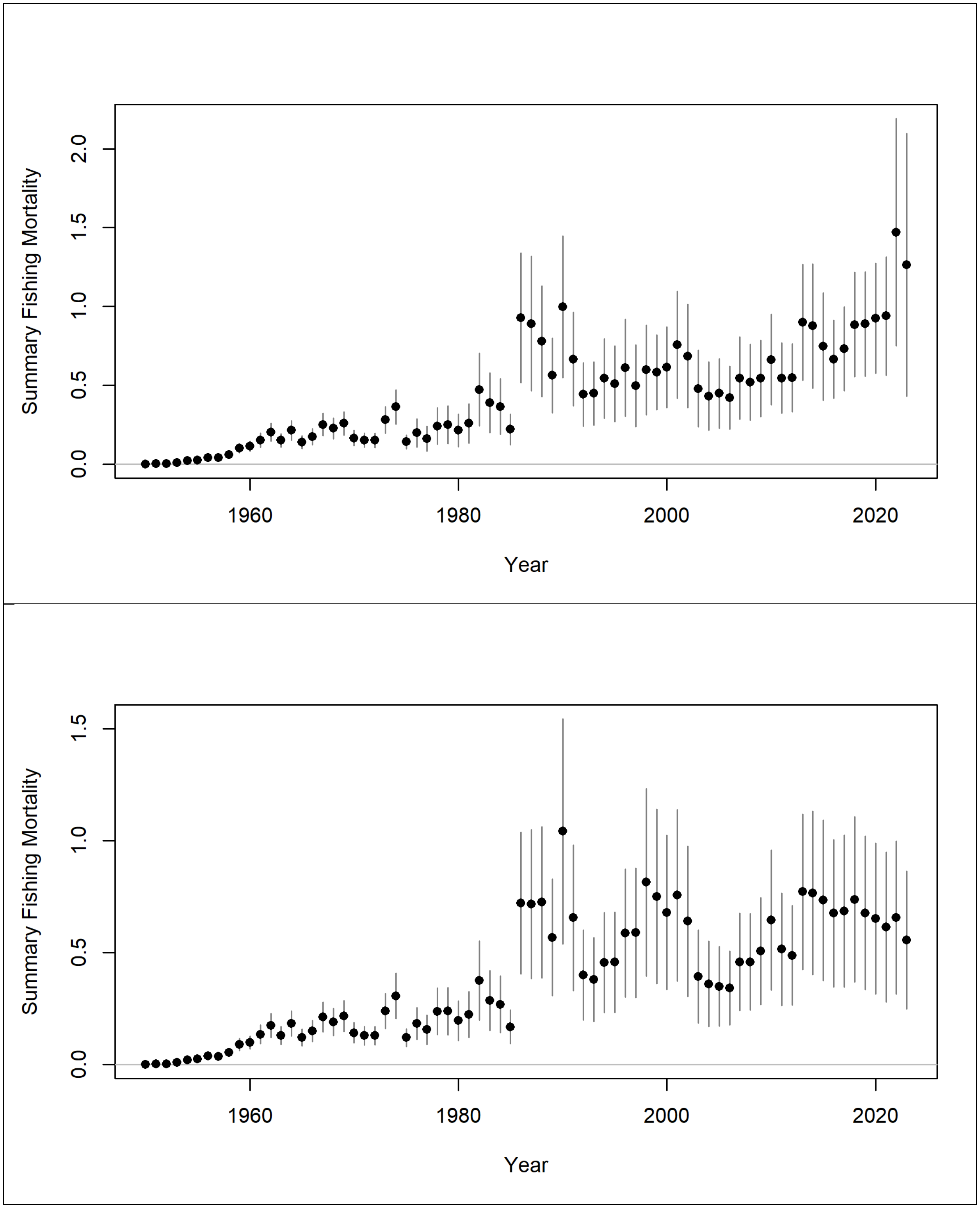


Figure 24. Estimated total fishing mortality. The top panel depicts the estimated values from the NW diagnostic case, the SW diagnostic case is shown on the bottom.

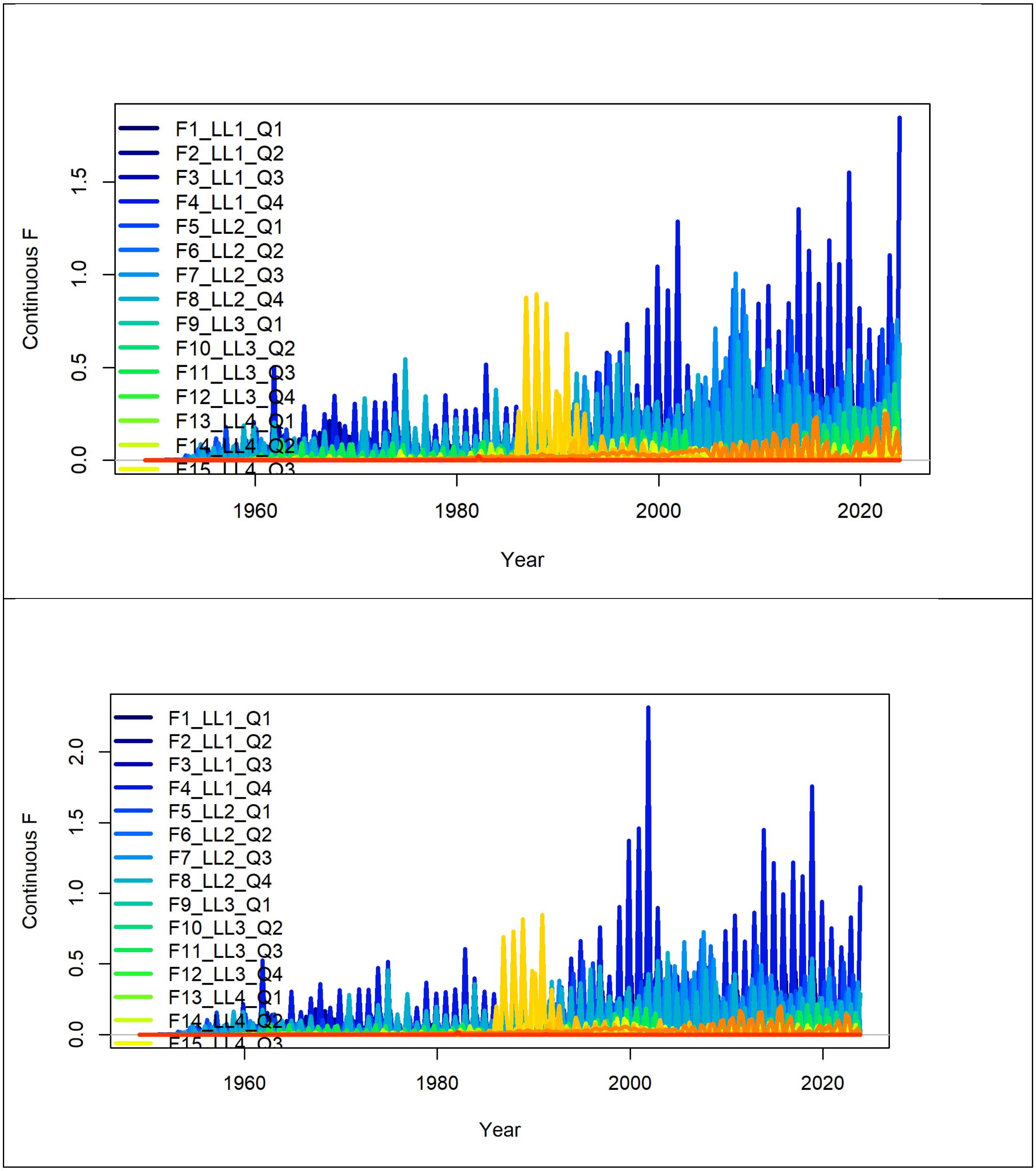


Figure 25. Estimated fleet specific fishing mortality by year for the base case model configurations. The top panel depicts the estimated values from the NW diagnostic case, the SW diagnostic case is shown on the bottom.

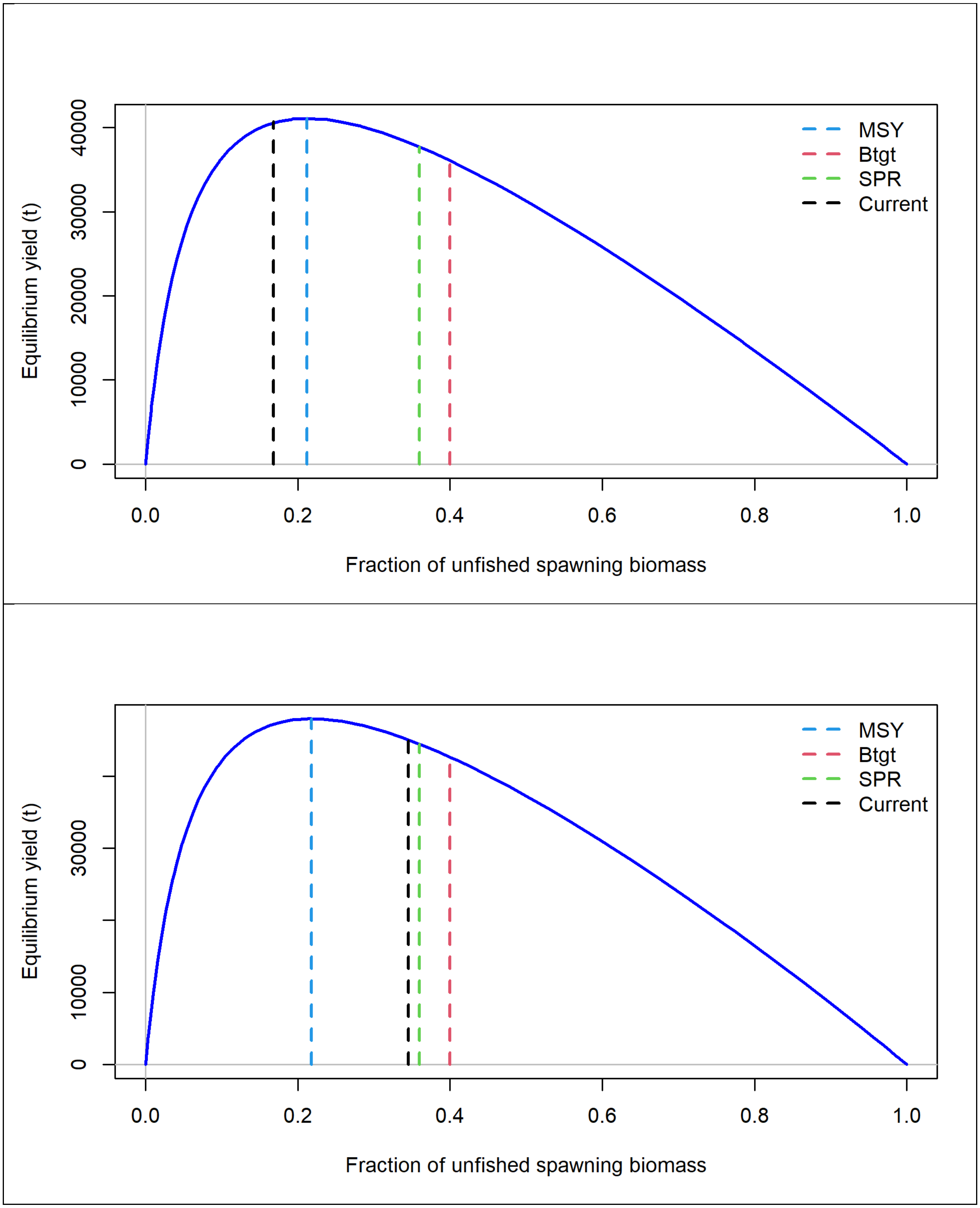


Figure 26. Equilibrium yield curve for the diagnostic case models for the assessment of albacore tuna in the Indian Ocean The top panel depicts the estimated values from the NW diagnostic case, the SW diagnostic case is shown on the bottom.

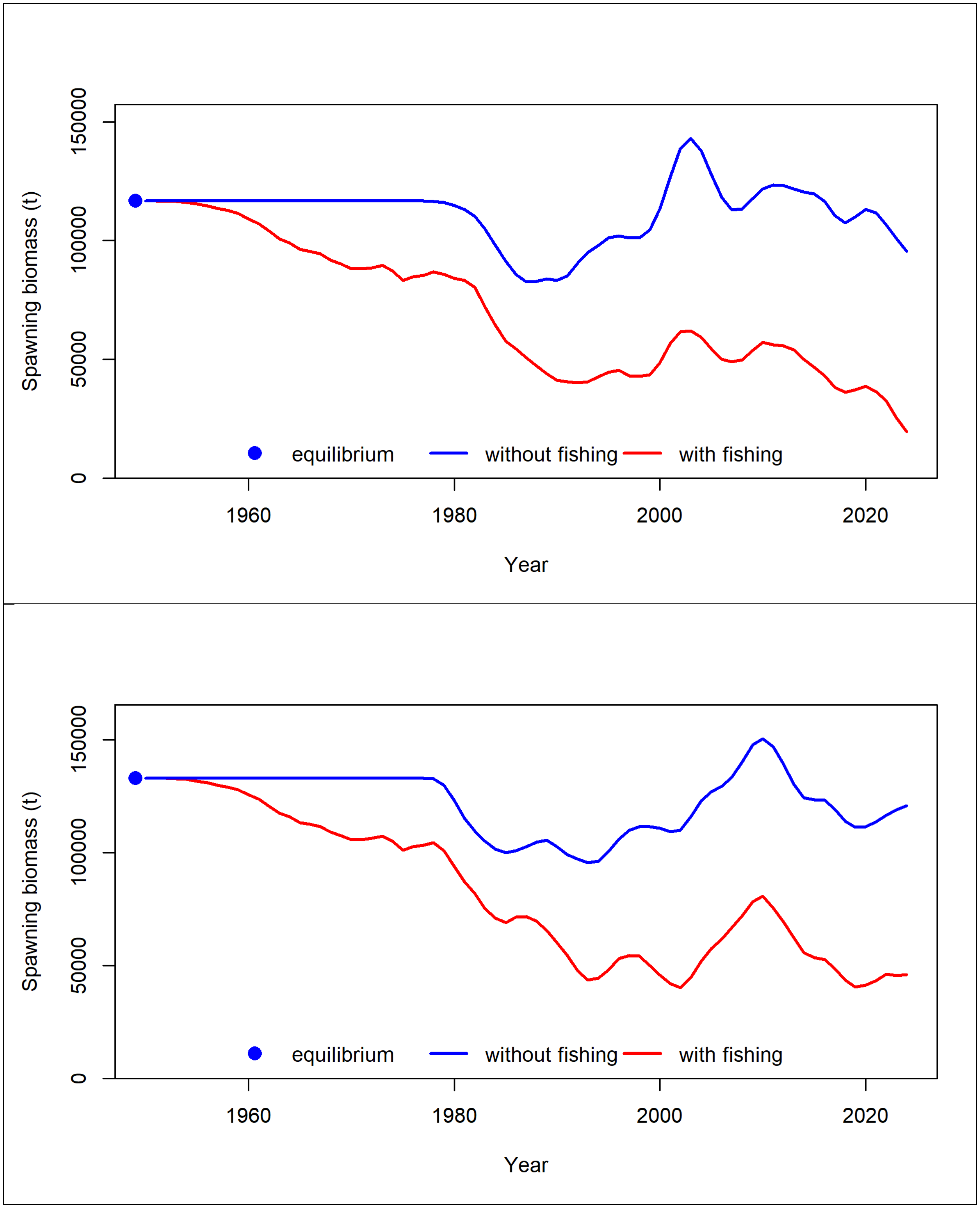


Figure 27. Dynamic B0 plot showing the spawning biomass under conditions no fishing and assuming the input catch series. The top panel depicts the estimated values from the NW diagnostic case, the SW diagnostic case is shown on the bottom.

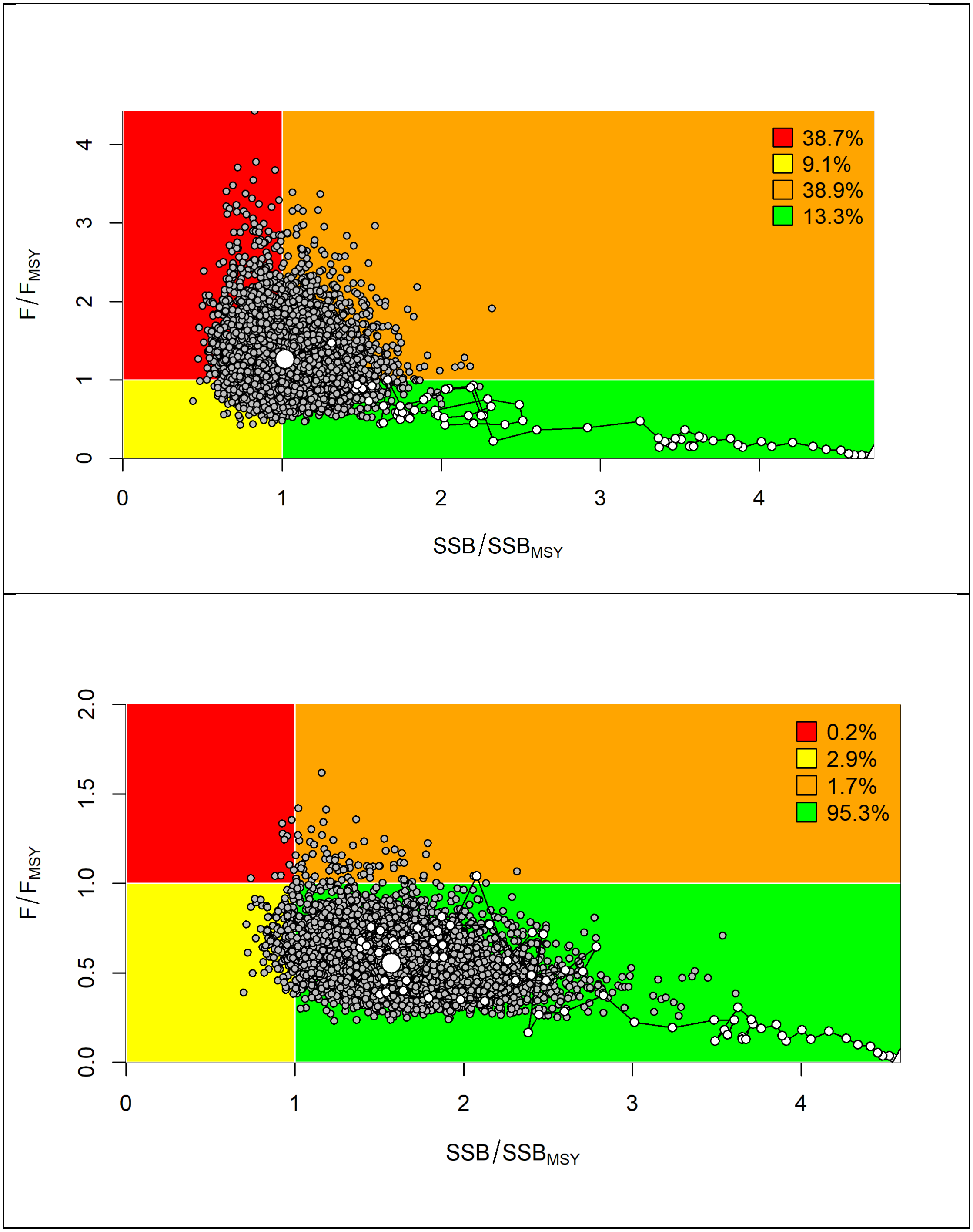


Figure 28. Estimated stock status trajectory (black line and white circles) with associated uncertainty estimated via MVLN approximation. The top panel depicts the estimated values from the NW diagnostic case, the SW diagnostic case is shown on the bottom.

10 Annex 1. Model Diagnostics for the NW diagnostic case models.

Diagnostics for the 2026 Indian Ocean Albacore Stock Assessment

NW Diagnostic case model

10.1 Introduction

The supplementary section provides additional analyses and diagnostics that complement the primary stock assessment report. Diagnostic tests are important in determining the robustness of estimates for management advice in integrated stock assessment models. The diagnostics tests included in are based on diagnostics prepared for the previous assessment as well as recently developed methods (Carvallo et al. 2021). Here we present the model diagnostics for the diagnostic model run presented in the main assessment analysis.

10.2 Goodness of fit

10.2.1 Runs test and joint residual plots.

The goodness of fit of the model can be used as an indication of whether there is presence of significant model misspecification. Models that do not fit the data should be considered suspect, and further investigated. Here we use residual plots to investigate the trends and patterns in the data over time. Temporal correlation (autocorrelation) can drive bias and drift in the model estimates over time. A runs test (Wald and Wolfowitz, 1940) can test for randomness in a data sequence, such as model residuals (Carvalho et al., 2021) (Figure 4). Residuals can also be investigated along side the root mean square error (RMSE, Carvalho et al., 2017), and a joint residual plot (Winker et al., 2018), which can highlight the systematically auto-correlated residual patterns (Figure 1, Figure 2).

10.3 Model consistency

10.3.1 R0 Profile

Use of a likelihood component profile on the a global scaling parameter (or other parameter) has been identified as a key model diagnostic to identify the influence of information sources on model estimates (Carvalho et al. 2017, Ichinokawa et al., 2014; Lee et al., 2014; Wang et al., 2014). Here the equilibrium recruitment parameter, R0, is used because it represents an ideal global scaling parameter given that unfished (virgin) recruitment is proportional to unfished biomass (Carvalho et al 2021, Lee et al., 2014; Maunder and Piner, 2015; Wang et al., 2014).

A relatively large change in negative log-likelihood units along the profile suggests a relatively informative data source for that particular model. Close association in the location of the minimum negative log-likelihood along the profile between data sources suggest that model consistency, and lack of conflict in the data. Figures Figure 5, Figure 6, Figure 7 show the profile likelihoods of R0 for the overall, CPUE and length components of the model. Figure A12 shows the fit to the CPUE series for the range of LN(R0) assumed. The likelihood profiles show that overall the LN(R0) parameter is well estimated, led by the length likelihood then the index likelihood and then the recruitment. Interestingly the lower edge of the likelihood is better defined (steeper) for all three components than the upper (Figure 5). The fleet indices are generally in agreement (Figure 6), however the length data shows different minimums (Figure 7), which is consistent with the fleets that encounter different components of the stock. The fits to the CPUE series at different values of LN(R0) show that values of approximately 9.7 fit the series (Figure 6).

10.4 Age-Structured Production Model (ASPM)

This diagnostic can help evaluate whether the catch and CPUE data give evidence for a production function within the model (Carvalho 2017). Overall the ASPM evaluates whether the effect of surplus production and observed catches alone could explain trends in the CPUE, in contrast to a more complex model (i.e. SS3) that incorporates annual recruitment deviations to improve the fit (Carvalho et al 2021). Maunder and Piner (2017) note that if the ASPM fits well to the indices of abundance with contrast the production function is likely to drive the stock dynamics and the indices will provide information about absolute abundance (Minte-Vera et al., 2017). Figure A13 shows that the biomass trajectories for both models (ASPM and the diagnostic) follow the same trend and that the estimates of LN(R0) are comparable. The fits to the indices are shown in Figure 8 and indicate an overall good fit, indicating that the information content in the data is sufficient.

10.5 Retrospective analysis

Retrospective analysis is common in fisheries stock assessment to check the consistency of model estimates (Brooks and Legault, 2016; Carvalho et al., 2017; Hurtado-Ferro et al., 2015; Miller and Legault, 2017). A retrospective analysis is carried out by sequentially deleting a number of years of day (i.e. from 0 to 7) and re-running the model. Comparisons of model estimates from the full time-series and the truncated time-series can illuminate the bias and accuracy of the modelled quantities. Statistical analysis in the form of calculating the retrospective bias, rho (ρM, Mohn 1999), is common, with values between -0.15 and 0.2 being considered indicative of no bias. Figure Figure 14 shows the analysis of spawning stock biomass (SSB) and fishing mortality estimates for Indian Ocean albacore tunaalong with the Mohn's rho which indicates no retrospective bias. Forecasting the next year based on the retrospective analysis shows similar analysis (Figure 14, Figure 15).

10.6 Prediction Skill

Kell et al. (2016) proposed the hindcasting cross-validation technique (HCXval) where observations are compared to their predicted future values. The key concept behind the HCXval approach is 'prediction skill', which is defined as any measure of the accuracy of a forecasted value to the actual observed that is not known by the model (Kell et al., 2021). The difference, which is referred to as the 'prediction residual' (Michaelsen, 1987) can be evaluated by the mean absolute scaled error (MASE; Hyndman and Koehler, 2006). Carvalho et al note that a MASE score > 1 indicates that the average model forecasts are worse than a random walk. Conversely, a MASE score of 0.5 indicates that the model forecasts twice as accurately as a naïve baseline prediction; i.e. The model has prediction skill. For the CPUE series that constitute the diagnostic case the MASE values are greater than 1 for the length and CPUE series (respectively, Figure **?@fig-HCXval**), this indicates a mix of poor, good and decent prediction skill.

10.1 Convergence

The jitter (Figure 18) test for global convergence was implemented in Stock Synthesis, utilizing the jitter feature described in detail within the Stock Synthesis manual (e.g., for version 3.30.15 see Methot et al., 2020). The jitter feature is implemented in R using a function in the r4ss package (see https://github.com/jabbamodel/ss3diags for example R code). The

10.2 Data Influence

The *leave one out* and *let one in* diagnostic for data influece shows the effect of dropping (Figure 20) or using (Figure 21) one CPUE series at a time.

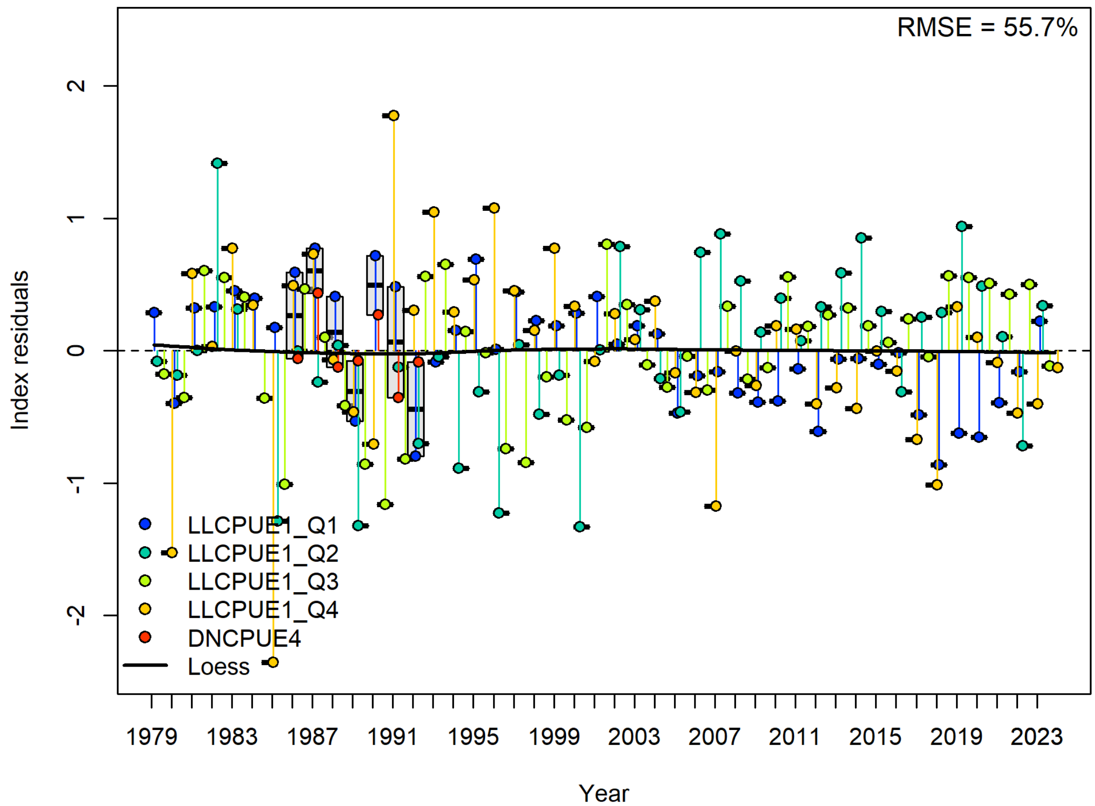


Figure 1: Joint residual plots for CPUE fits. Vertical lines with points show the residuals (in colors by index), and solid black lines show loess smoother through all residuals. Boxplots indicate the median and quantiles in cases where residuals from the multiple indices are available for any given year. Root-mean squared errors (RMSE) are included in the upper right-hand corner of the plot.

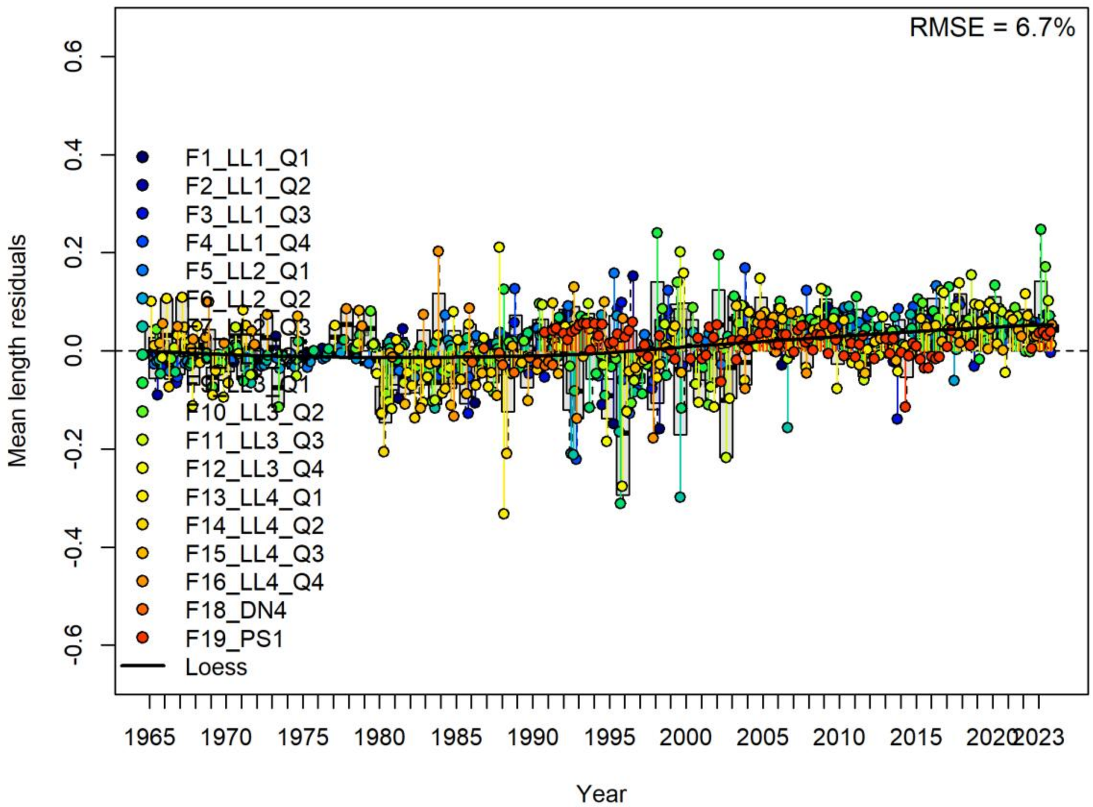


Figure 2: Joint residual plots for annual mean length estimates for multiple fishing fleets . Vertical lines with points show the residuals (in colors by index), and solid black lines show loess smoother through all residuals. Boxplots indicate the median and quantiles in cases where residuals from the multiple indices are available for any given year. Root-mean squared errors (RMSE) are included in the upper right-hand corner of the plot.

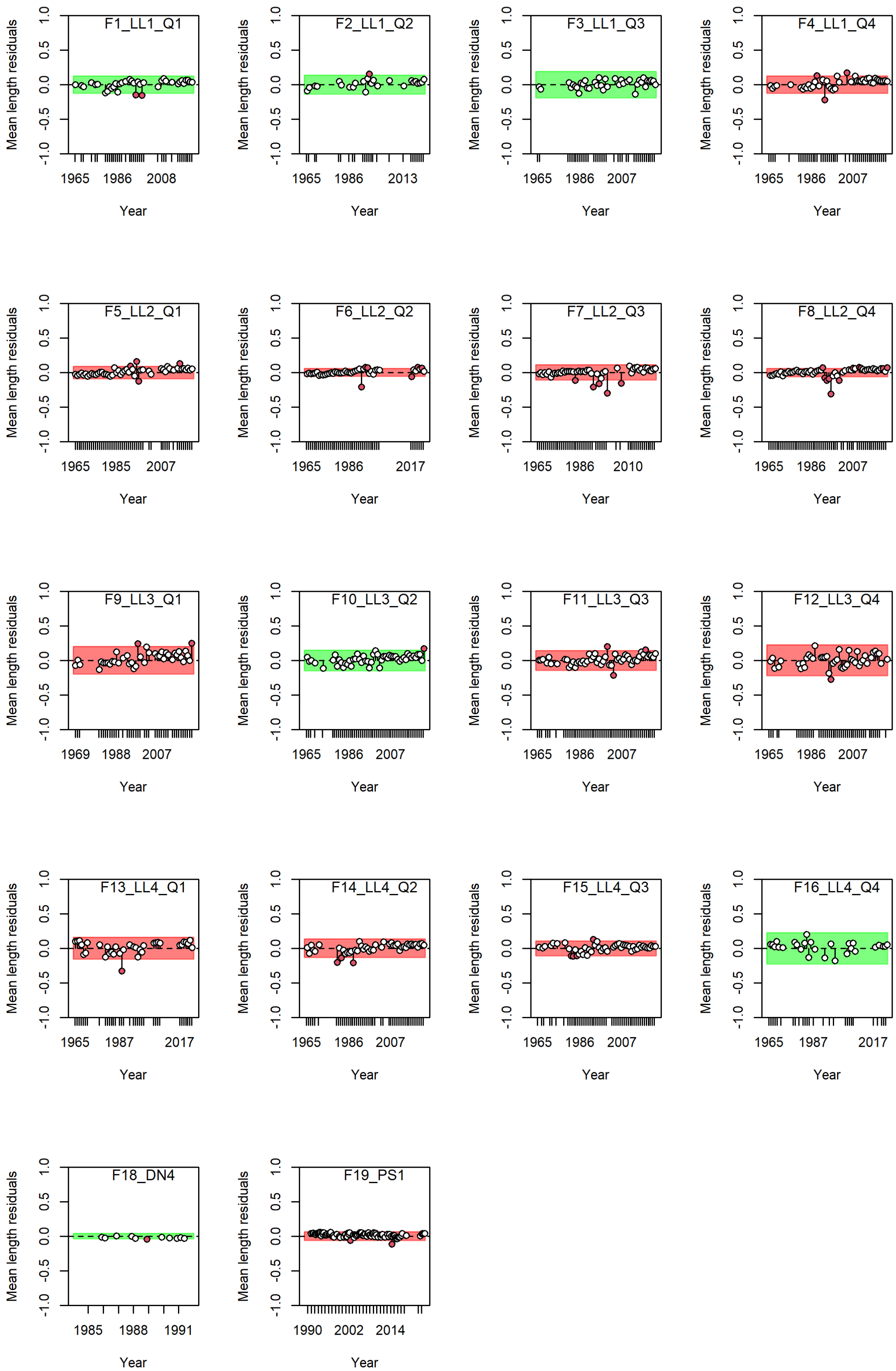

F1_LL1_Q1
F2_LL1_Q2
F3_LL1_Q3
F4_LL1_Q4
F5_LL2_Q1
F6_LL2_Q2
F7_LL2_Q3
F8_LL2_Q4
F9_LL3_Q1
F10_LL3_Q2
F11_LL3_Q3
F12_LL3_Q4
F13_LL4_Q1
F14_LL4_Q2
F15_LL4_Q3
F16_LL4_Q4
F18_DN4
F19_PS1
Mean length residuals
-1.0 -0.5 0.0 0.5 1.0
Year
1965 1986 2008
1965 1986 2013
1965 1986 2007
1965 1986 2007
1965 1985 2007
1965 1986 2017
1965 1986 2010
1965 1986 2007
1969 1988 2007
1965 1986 2007
1965 1986 2007
1965 1986 2007
1965 1987 2017
1965 1986 2007
1965 1986 2007
1965 1987 2017
1985 1988 1991
1990 2002 2014

Figure 3: Runs tests results illustrated for the length composition fits (with names and fleet numbers on top). Green shading indicates no evidence (p ≥ 0.05) and red shading evidence (p < 0.05) to reject the hypothesis of a randomly distributed time-series of residuals, respectively. The shaded (green/red) area spans three residual standard deviations to either side from zero, and the red points outside of the shading violate the 'three-sigma limit' for that series.

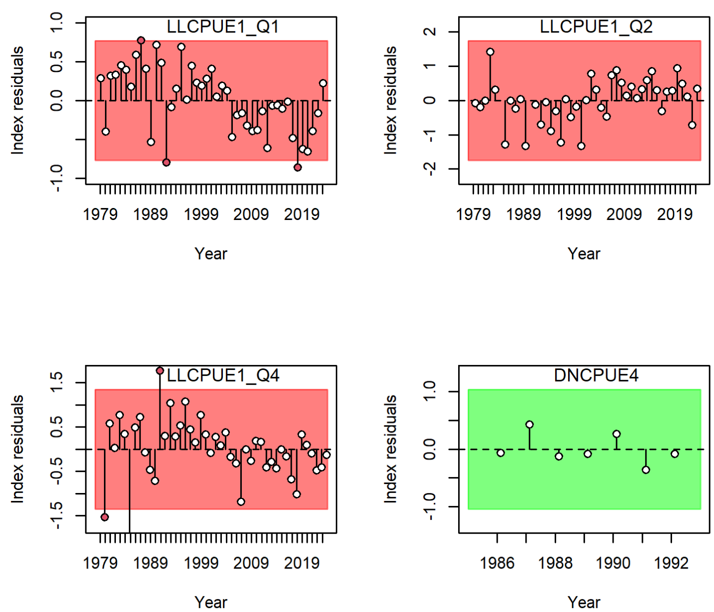


Figure 4: Runs tests results illustrated for the catch-per-unit-effort (CPUE) fits, panels indicate separate fleets. Green shading indicates no evidence (p ≥ 0.05) and red shading evidence (p < 0.05) to reject the hypothesis of a randomly distributed time-series of residuals, respectively. The shaded (green/red) area spans three residual standard deviations to either side from zero, and the red points outside of the shading violate the 'three-sigma limit' for that series.

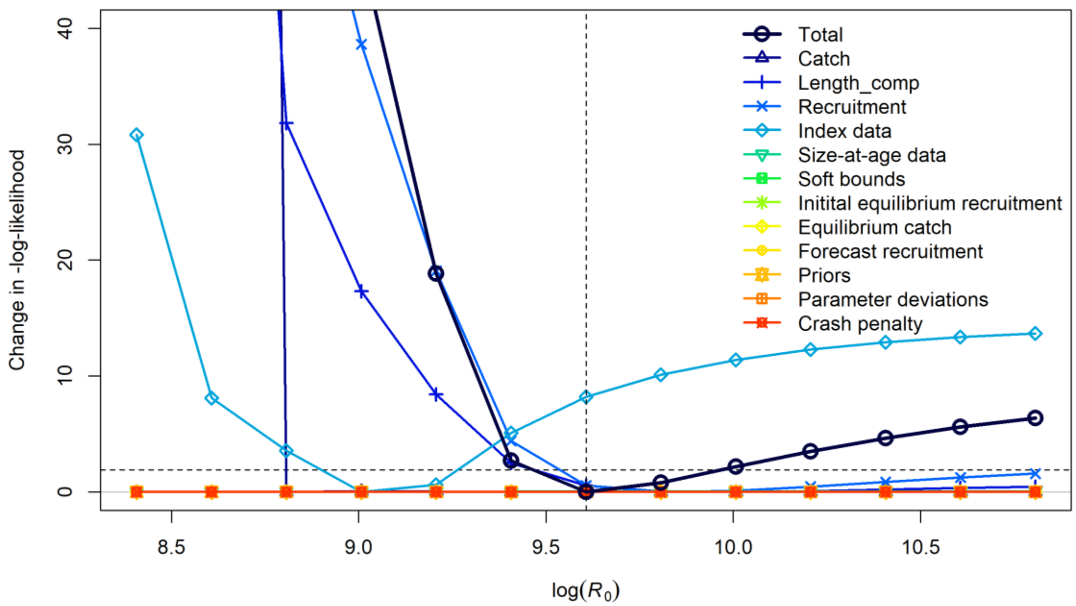


Figure 5: Log-likelihood profiles for R0 for the various data components included in the Indian Ocean albacore Stock Synthesis models, showing the contribution of (a) - (b) all data likelihood components., (c) - (d) among abundance indices and (e) - (f) among lengthand age-composition data, respectively. Total likelihood and component profiles (recruitment, length and index (CPUE) components).

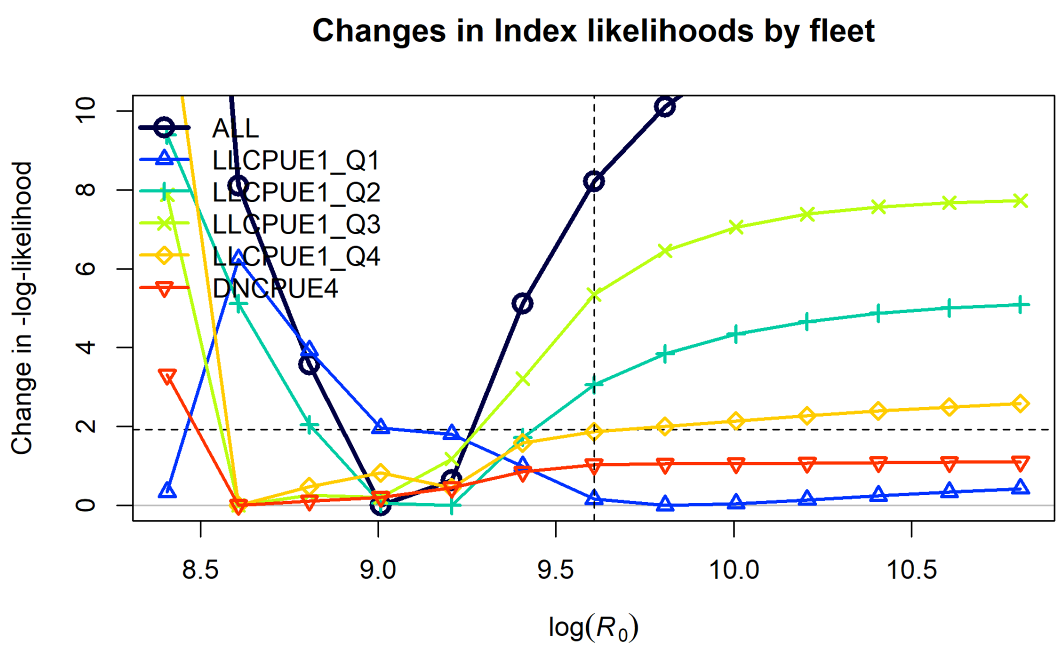


Figure 6: Log-likelihood profiles for R0 for the various data components included in the Indian Ocean albacore Stock Synthesis models, showing the contribution among abundance indices.

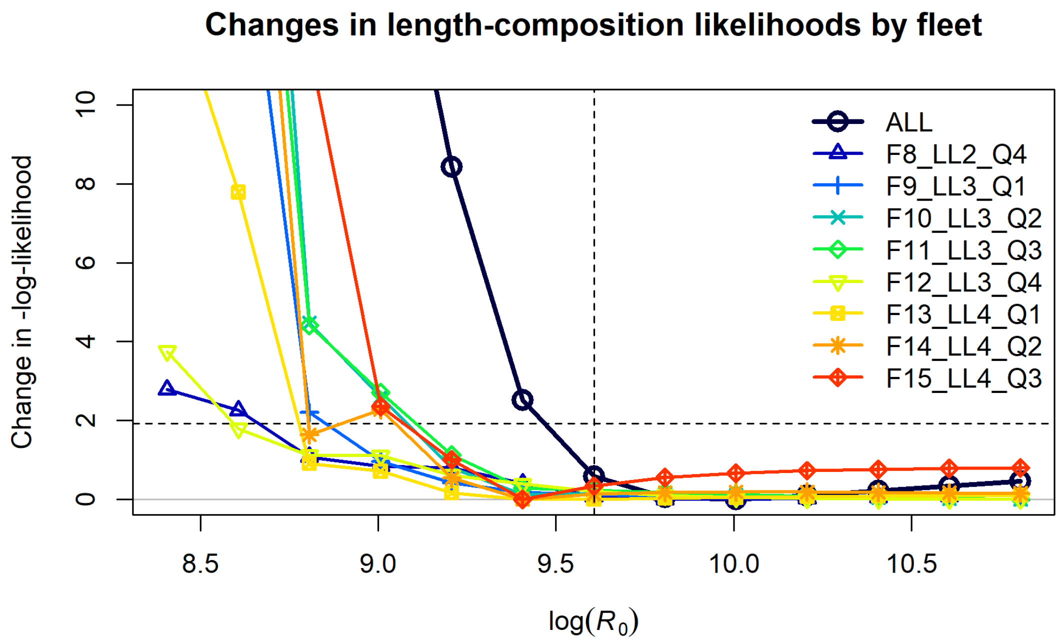


Figure 7: Log-likelihood profiles for R0 for the various data components included in the Indian Ocean albacore Stock Synthesis models, showing the contribution among length data.

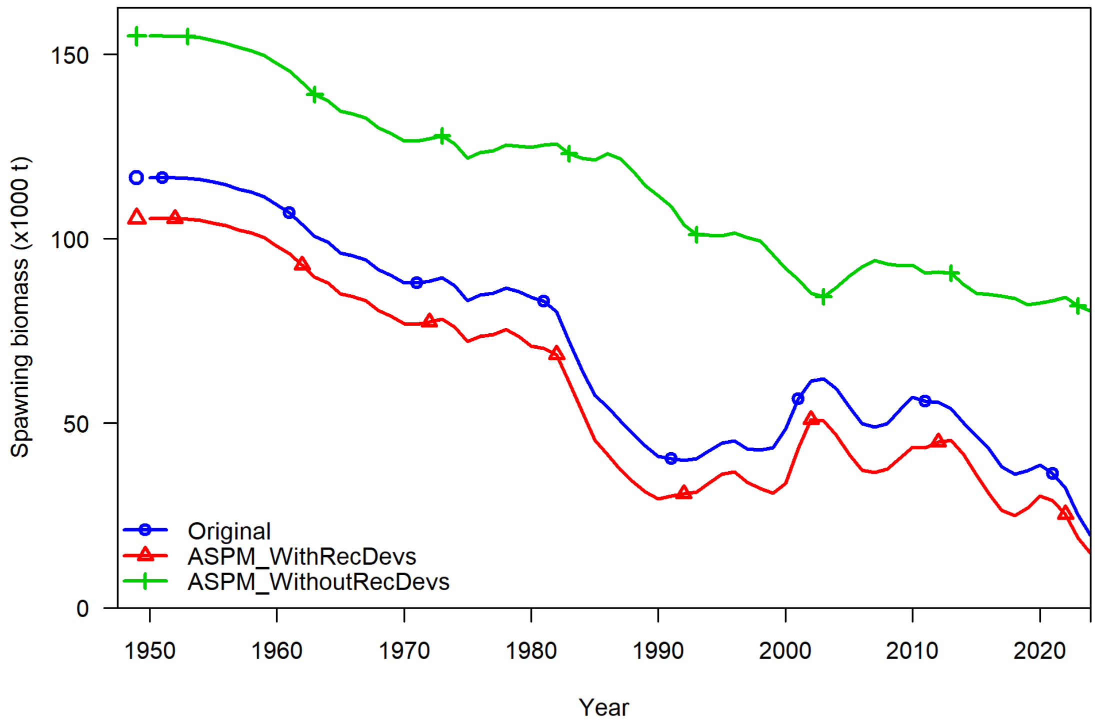


Figure 8: Comparison of spawning biomass trajectories for the ASPM and the diagnostic case of the assessment model carried out in stock synthesis.

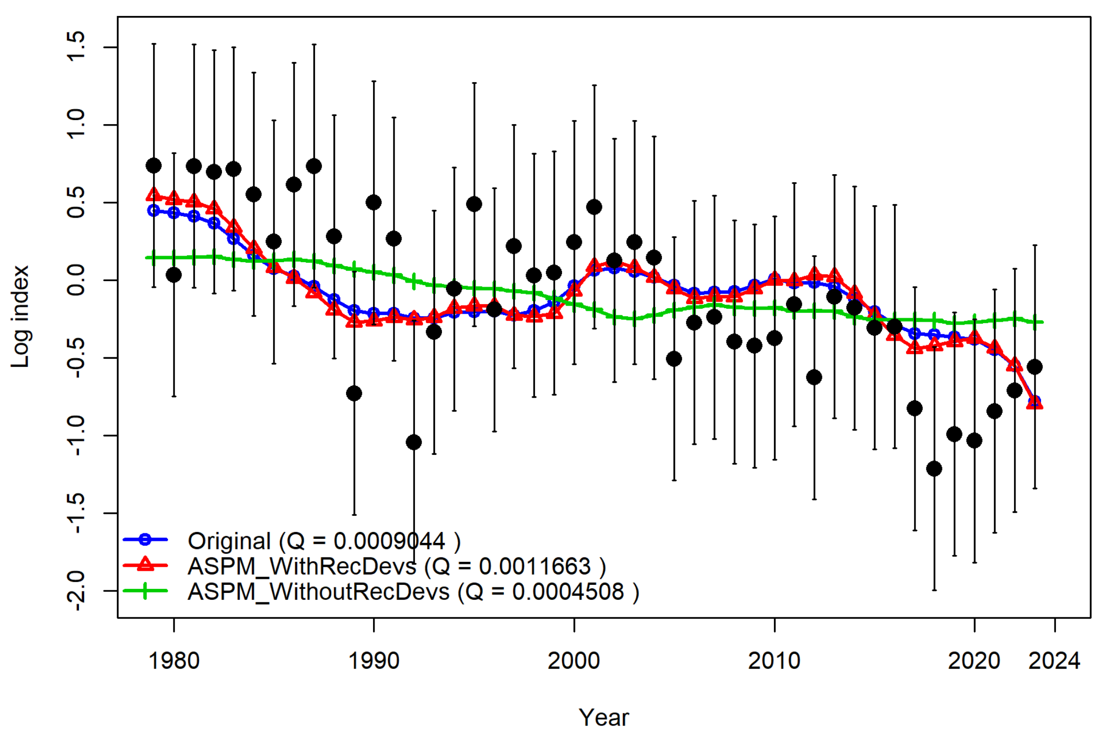


Figure 9: Fits to the indices for the ASPM anddiagnostic case. The panels indicate the fits to the CPUE.

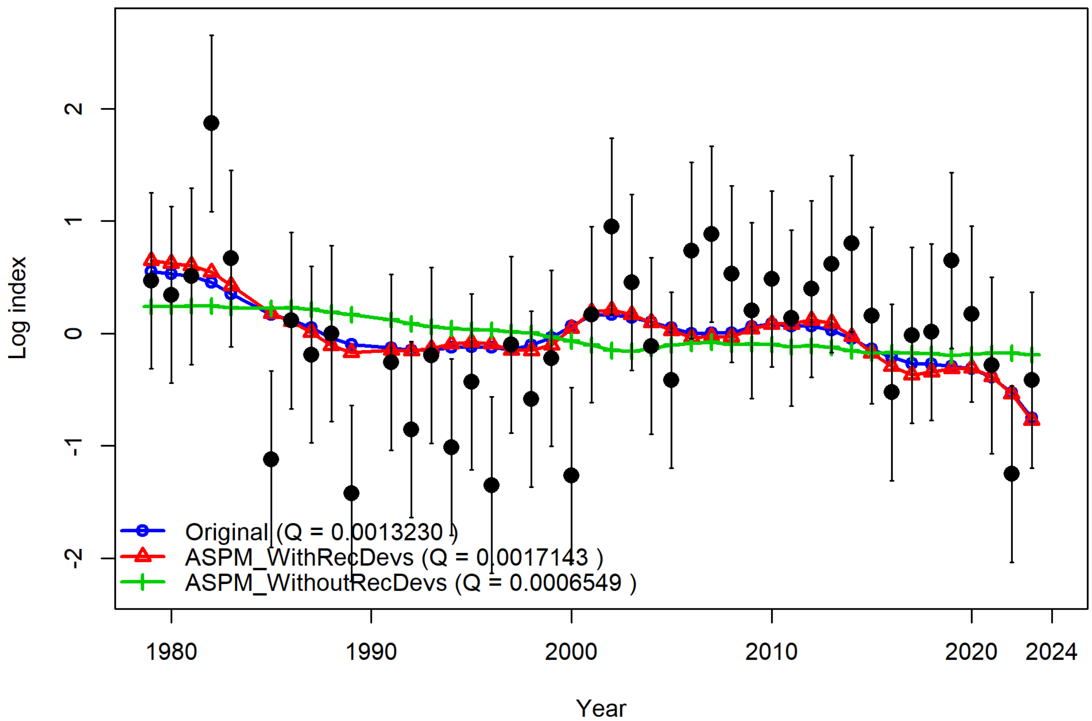


Figure 10: Fits to the indices for the ASPM anddiagnostic case. The panels indicate the fits to the CPUE.

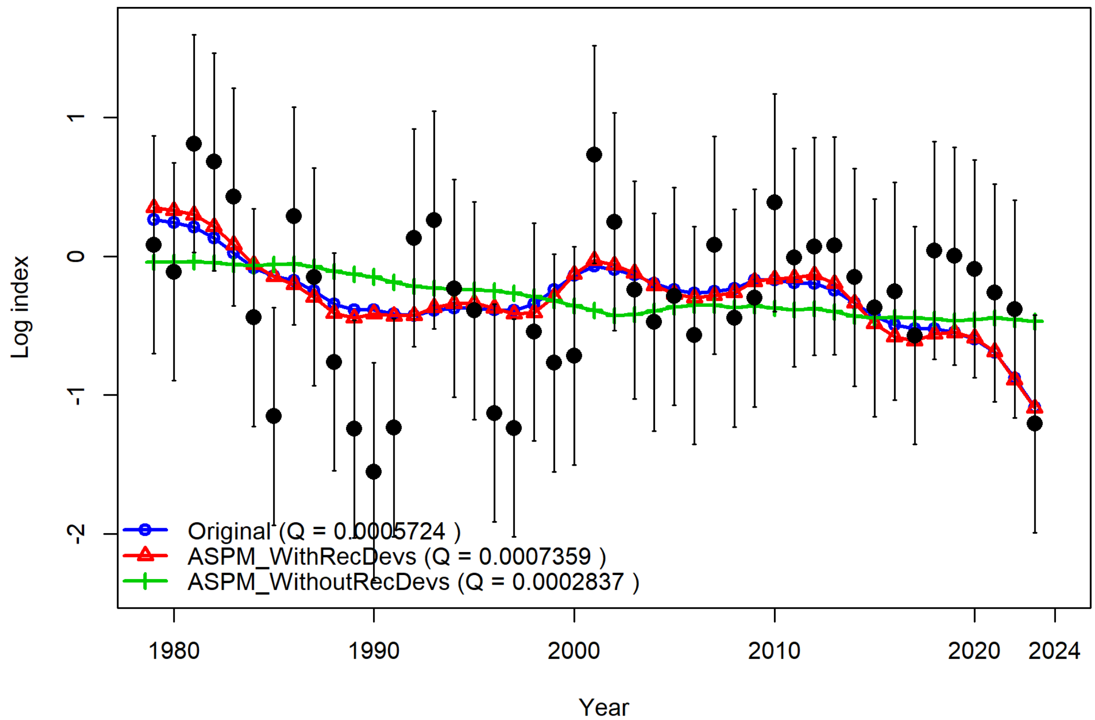


Figure 11: Fits to the indices for the ASPM anddiagnostic case. The panels indicate the fits to the CPUE.

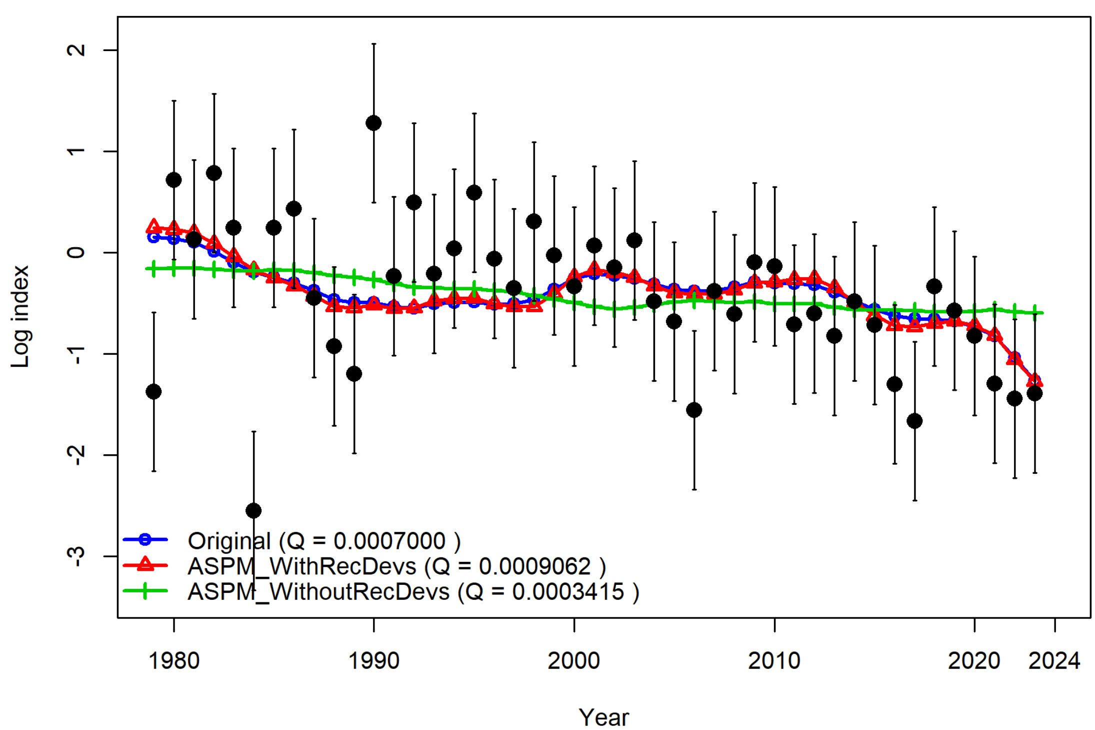


Figure 12: Fits to the indices for the ASPM anddiagnostic case. The panels indicate the fits to the CPUE.

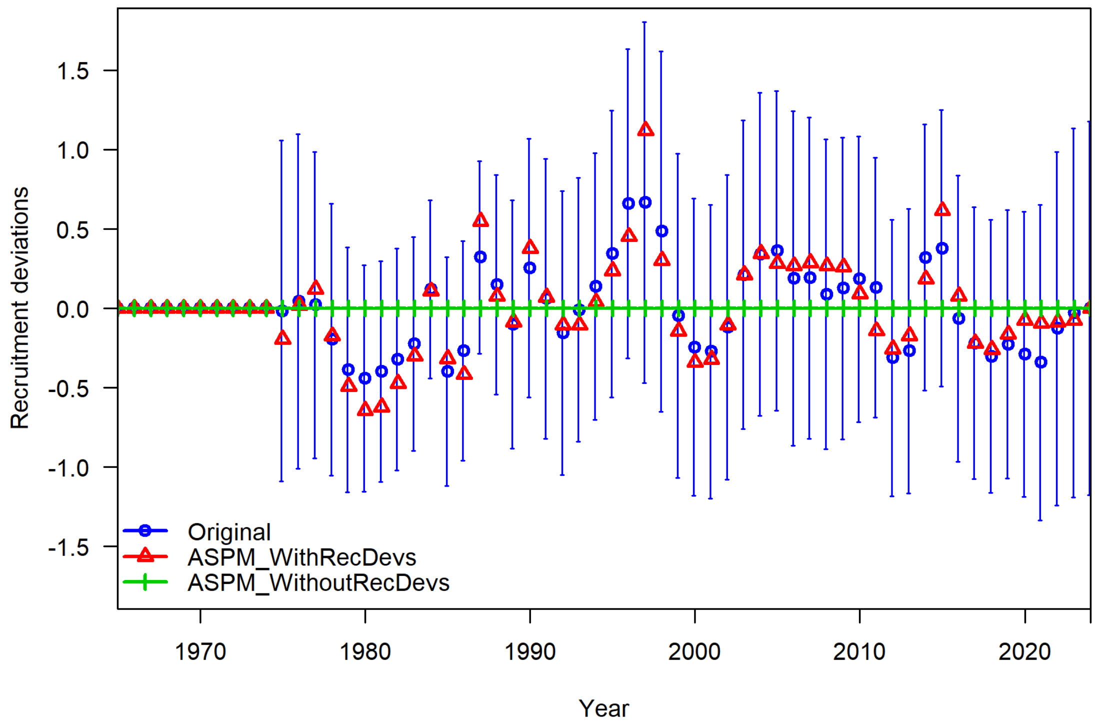


Figure 13: Comparison between the fully integrated base-case and the Age-Structured-Production Model (ASPM) results recruitment deviation estimates.

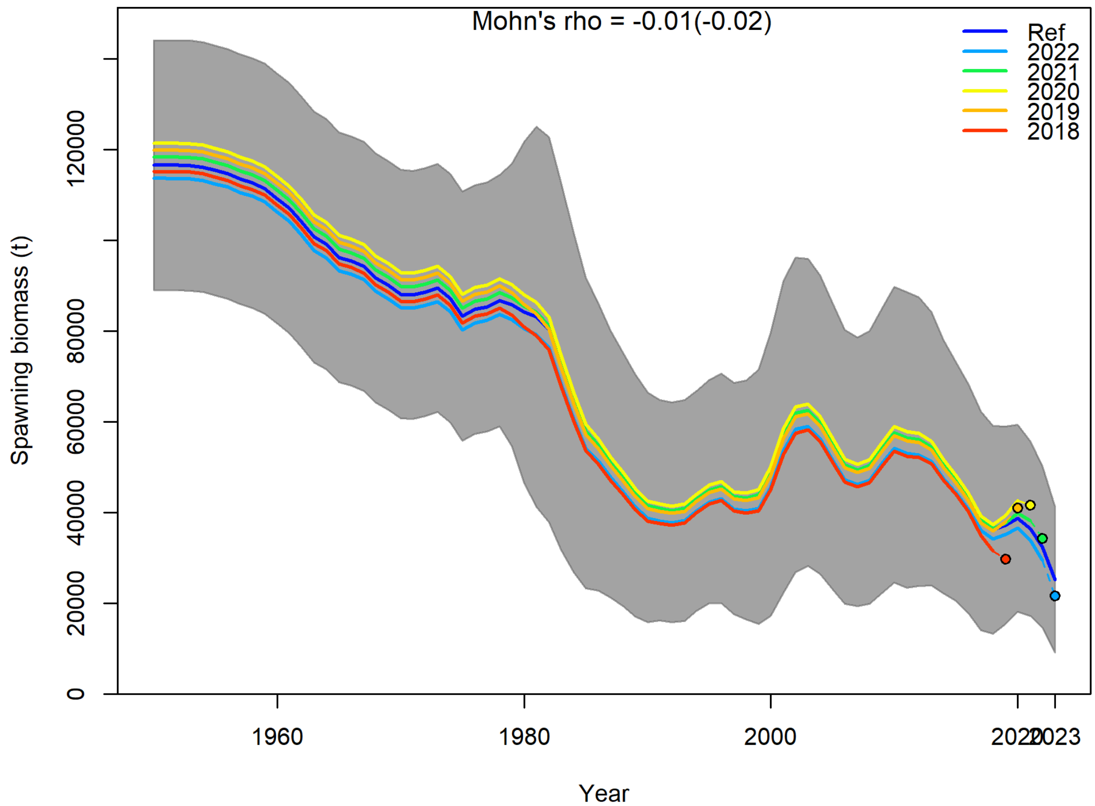


Figure 14: Retrospective analysis of spawning stock biomass (SSB) estimates for Indian Ocean albacore models conducted by re-fitting the reference model (Ref) after removing five years of observations The retrospective results are shown for (a) the entire time series and – (b) for the most recent years only. Mohn's rho statistic and the corresponding 'hindcast rho' values (in brackets) are printed at the top of the panel (b) One-year-ahead projections denoted by color-coded dashed lines with terminal points are shown for each model. Grey shaded areas are the 95 % confidence intervals from the reference model.

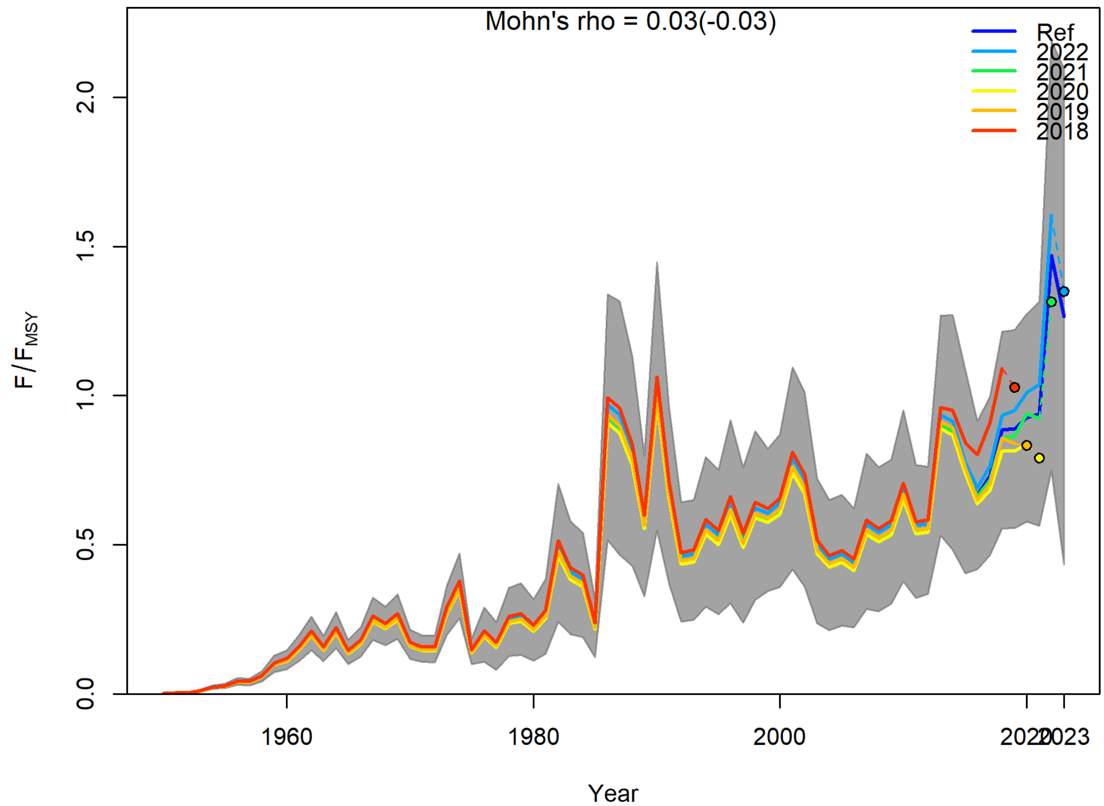
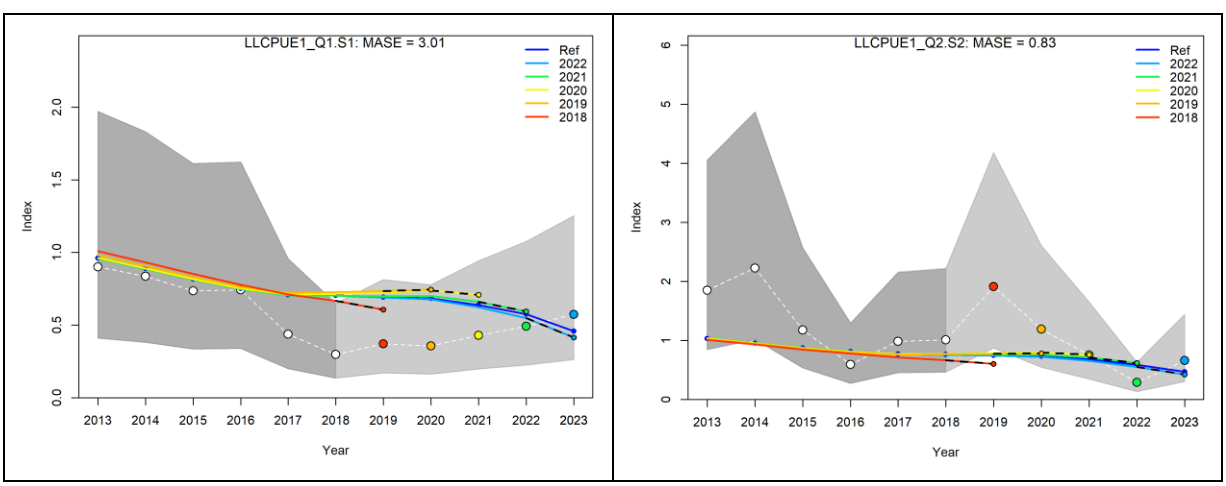


Figure 15: Retrospective analysis of fishing mortality (F) estimates for Indian Ocean albacore models conducted by re-fitting the reference model (Ref) after removing five years of observations The retrospective results are shown for (a) the entire time series and – (b) for the most recent years only. Mohn's rho statistic and the corresponding 'hindcast rho' values (in brackets) are printed at the top of the panel (b) One-year-ahead projections denoted by color-coded dashed lines with terminal points are shown for each model. Grey shaded areas are the 95 % confidence intervals from the reference model.



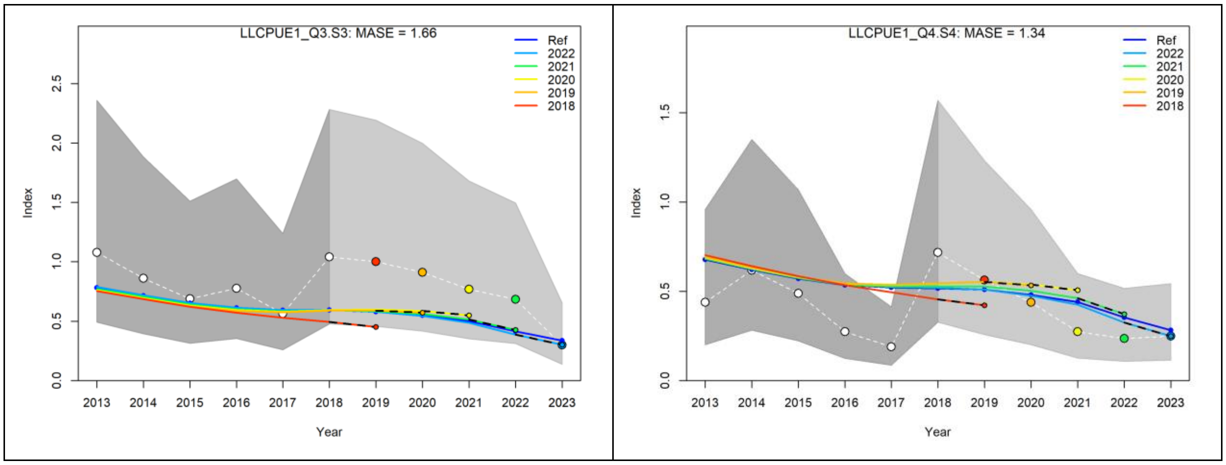


Figure 16: Hindcasting cross-validation (HCxval) results from CPUE fits, showing observed (large points connected with dashed line), fitted (solid lines) and one-year ahead forecast values (small terminal points). HCxval was performed using one reference model (Ref) and seven hindcast model runs (solid lines) relative to the expected CPUE. The observations used for cross-validation are highlighted as color-coded solid circles with associated 95 % confidence intervals. The model reference year refers to the endpoints of each one-year-ahead forecast and the corresponding observation (i.e., year of retrospective + 1). The mean absolute scaled error (MASE) score associated with each CPUE."

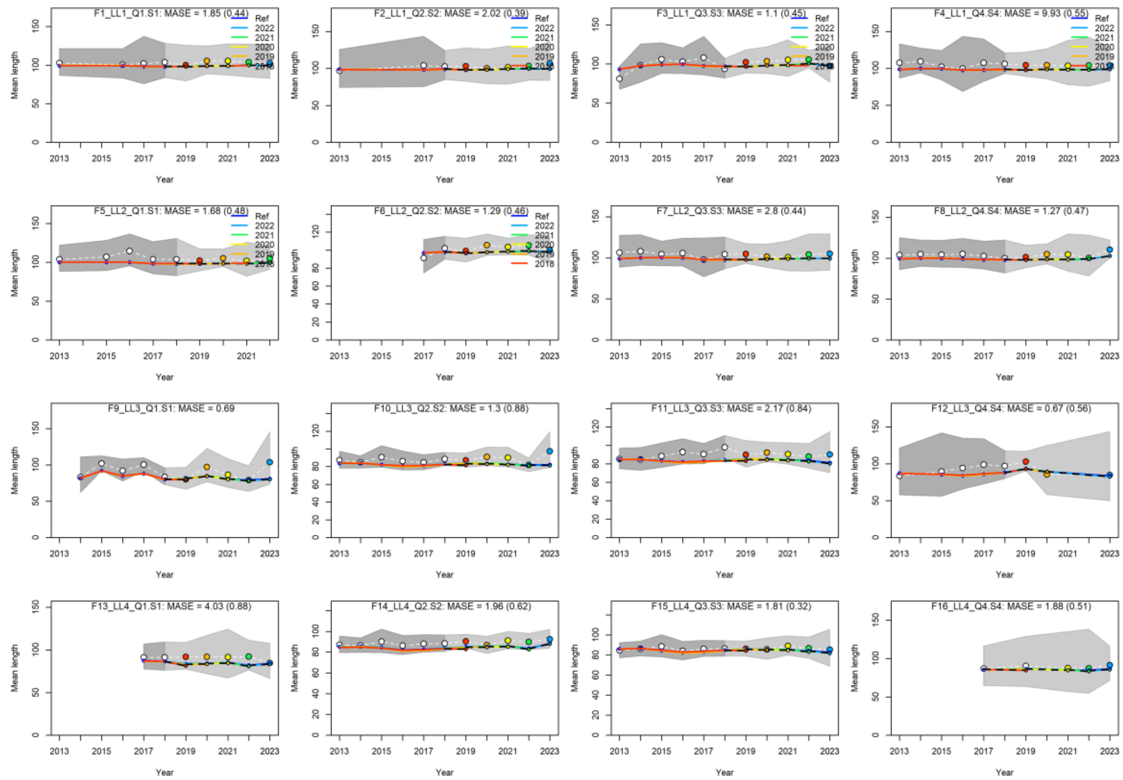


Figure 17: Hindcasting cross-validation (HCxval) results from length fits, showing observed (large points connected with dashed line), fitted (solid lines) and one-year ahead forecast values (small terminal points). HCxval was performed using one reference model (Ref) and seven hindcast model runs (solid lines) relative to the expected length. The observations used for cross-validation are highlighted as color-coded solid circles with associated 95 % confidence intervals. The model reference year refers to the endpoints of each one-year-ahead forecast and the corresponding observation (i.e., year of retrospective + 1). The mean absolute scaled error (MASE) score associated with each CPUE."

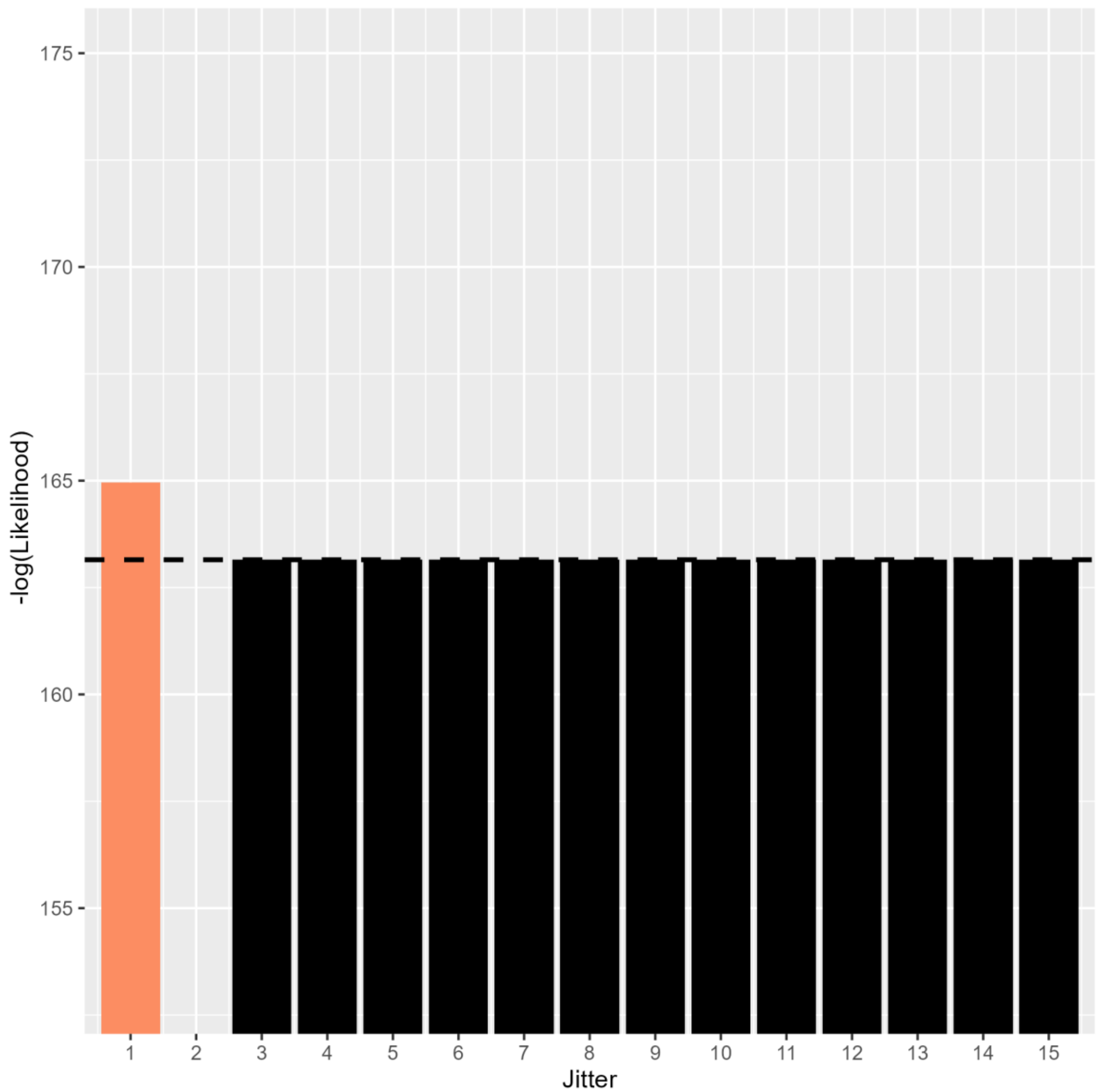


Figure 18: The jitter diagnostic for global convergence conducted for the Indian Ocean albacore model. The vertical bars represent the total likelihood obtained from jittered model runs, black bars indicate where the model found the same total likelihood value from the base-case model, missing bars indicate a failure to converge. The black horizontal dashed line represents the total likelihood value from the base-case model.

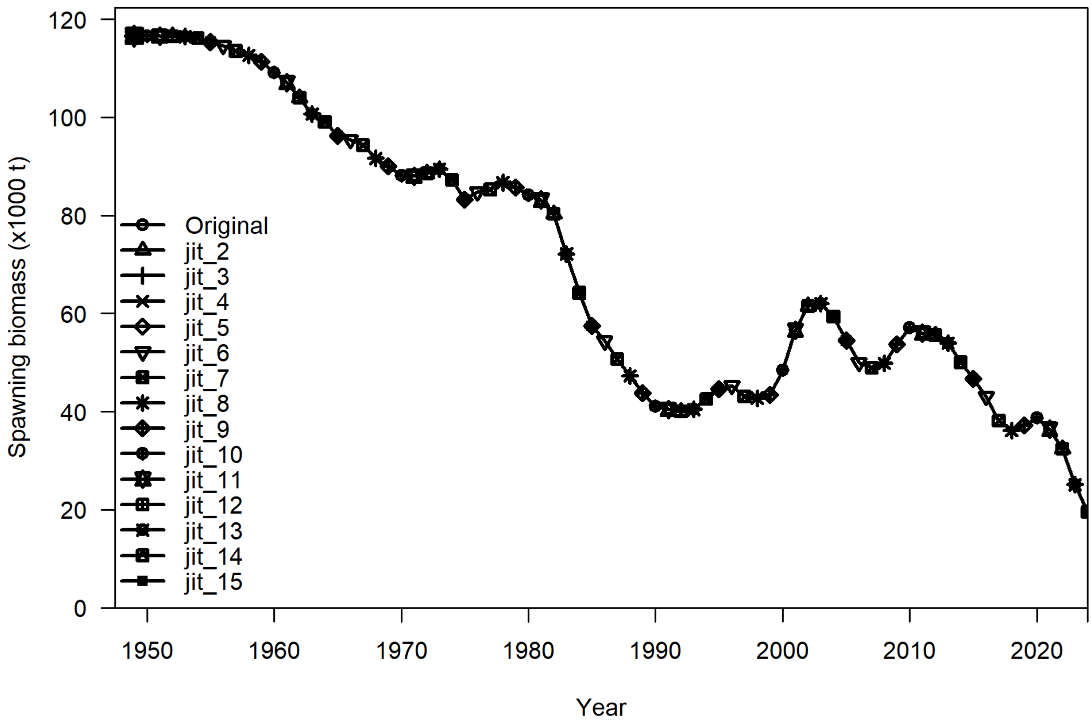


Figure 19: The jitter diagnostic for global convergence conducted for the Indian Ocean albacore model. The figure shows the spawning stock biomass (SSB) from jittered model runs. Grey shaded areas are the 95 % confidence intervals from the base-case model."

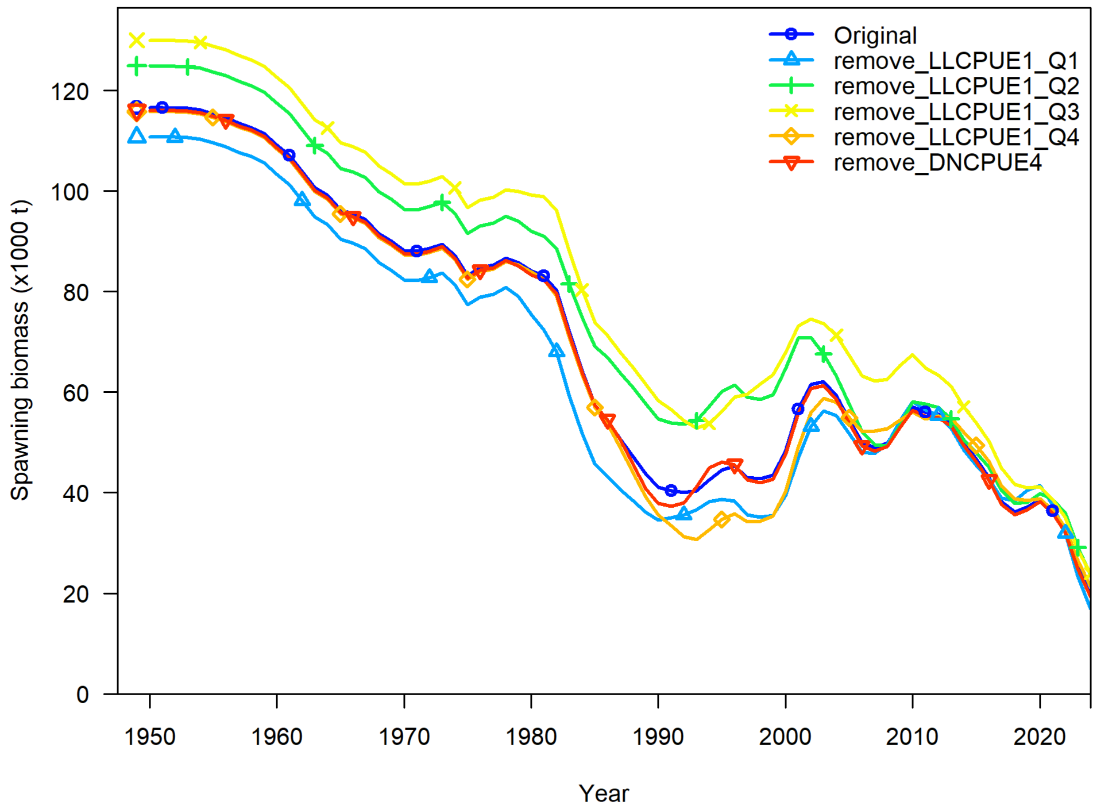


Figure 20: The leave one out diagnostic for CPUE indices for the Indian Ocean albacore SS3 model. This shows the effect of dropping one CPUE series at a time.”

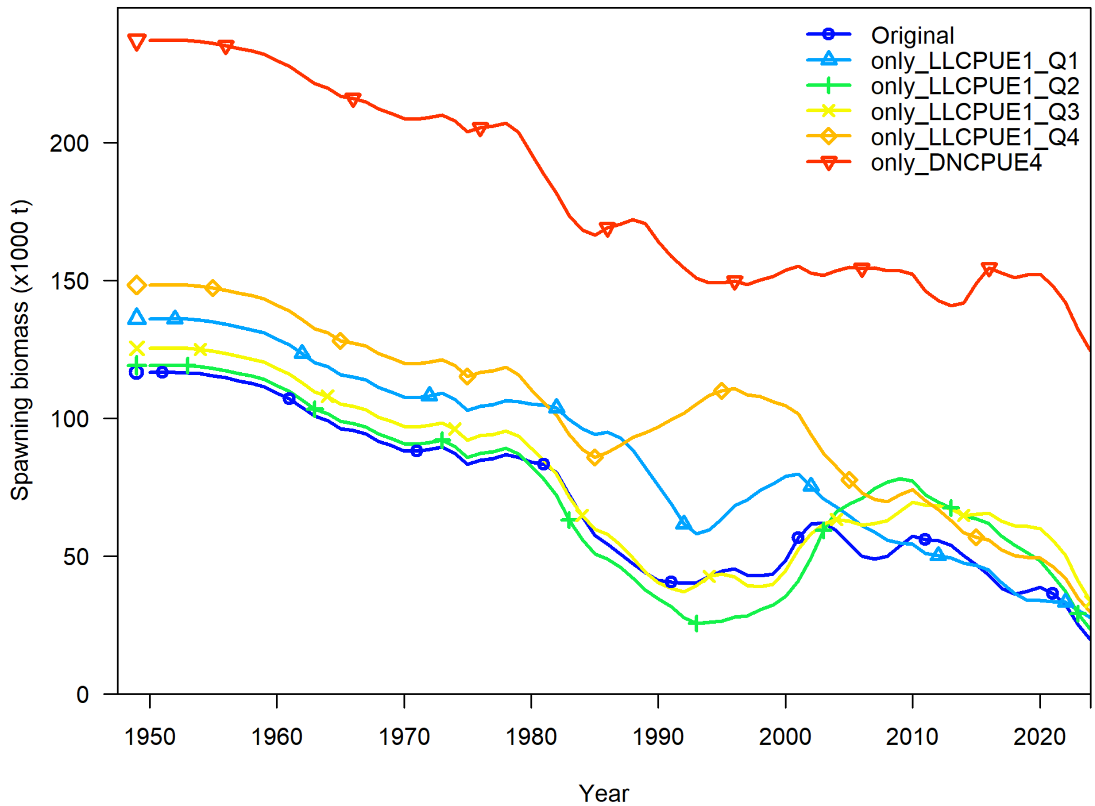


Figure 21: The let one in diagnostic for CPUE indices for the Indian Ocean albacore SS3 model. This shows the effect of using one CPUE series at a time."

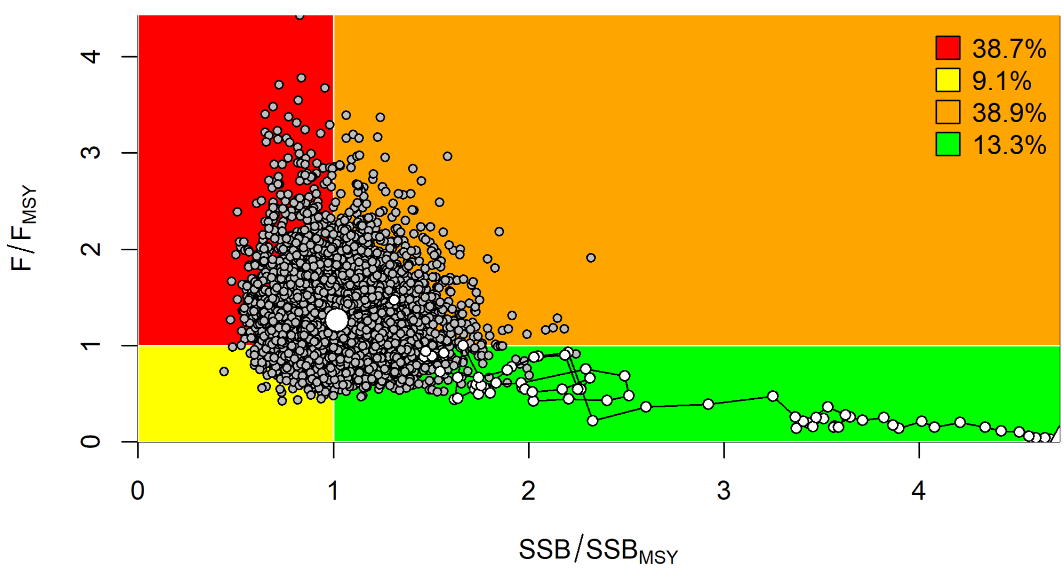


Figure 22: The multi-variate log-normal Kobe Plot.”

10.3 References

Brooks, E.N., Legault, C.M., 2016. Retrospective forecasting — evaluating performance of stock projections for New England groundfish stocks. Can. J. Fish. Aquat. Sci. 73, 935–950.

Carvalho, F., Punt, A. E., Chang, Y. J., Maunder, M. N., and K. R. Piner. 2017. Can diagnostic tests help identify model misspecification in integrated stock assessments? Fisheries Research 192:28–40. Available: https://doi.org/10.1016/j.fishres.2016.09.018 (3/7/2023).

Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., Kitakado, T., Yemane, D, Piner, K. R., Maunder, M. N., Taylor, I., Wetzel, C. R., Doering, K., Johnson, K. F., and R. D. Methot. 2021. A cookbook for using model diagnostics in integrated stock assessments. Fish. Res. 240:105959. Available: https://doi.org/10.1016/j.fishres.2021.105959 (3/7/2023).

Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L., Ono, K., Vert-Pre, K.A., Whitten, A.R., Punt, A.E., 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. Ices J. Mar. Sci. 72, 99–110. https://doi.org/10.1093/icesjms/fsu198.

International Commission for the Conservation of Atlantic Tunas (ICCAT). 2017. Report of the 2017 ICCAT Shortfin Mako Data Preparatory Meeting (Madrid, Spain 28-31 March, 2017).

Kell, L. T., I. Mosqueira, P. Grosjean, J-M. Fromentin, D. Garcia, R. Hillary, E. Jardim, S. Mardle, M. A. Pastoors, J. J. Poos, F. Scott, R. D. Scott. 2007. FLR: an open-source framework for the evaluation and development of management strategies. ICES J Mar Sci, 64 (4): 640-646. doi: 10.1093/icesjms/fsm012

Lee, H.-H., Piner, K.R., Methot, R.D., Maunder, M.N., 2014. Use of likelihood profiling over a global scaling parameter to structure the population dynamics model: An example using blue marlin in the Pacific Ocean. Fish. Res. 158, 138–146. https://doi.org/10.1016/j.fishres.2013.12.017.

Maunder, M.N., Piner, K.R., 2015. Contemporary fisheries stock assessment: many issues still remain. ICES J. Mar. Sci. 72, 7–18. https://doi.org/10.1093/icesjms/fsu015.

Michaelsen, J., 1987. Cross-validation in statistical climate forecast models. J. Clim. Appl. Meteorol. 26, 1589–1600. https://doi.org/10.1175/1520-0450(1987)026<1589:CVISCF>2.0.CO;2.

Miller, T.J., Legault, C.M., 2017. Statistical behavior of retrospective patterns and their effects on estimation of stock and harvest status. Fish. Res. 186, 109–120. https://doi.org/10.1016/j.fishres.2016.08.002.

Minte-Vera, C.V., Maunder, M.N., Aires-da-Silva, A.M., Satoh, K., Uosaki, K., 2017. Get the biology right, or use size-composition data at your own risk. Fish. Res. 192, 114–125. https://doi.org/10.1016/j.fishres.2017.01.014.

Mohn, R., 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES J. Mar. Sci. 56, 473–488. https://doi.org/10.1006/jmsc.1999.0481

Wald, A., Wolfowitz, J., 1940. On a test whether two samples are from the same population. Ann. Math. Stat. 11, 147–162. http://www.jstor.org/stable/2235872.

Winker, H., Carvalho, F., Kerwath, S., 2020. Age-structured biomass dynamics of North Atlantic shortfin mako with implications for the interpretation of surplus production models. Col. Vol. Sci. Pap. ICCAT 76, 316–336. https://www.iccat.int/Documents/CVSP/CV076_2019/colvol76.html.

11 Annex 2 Diagnostics for the SW Diagnostic case run

Diagnostics for the 2026 Indian Ocean Albacore Stock Assessment

SW Diagnostic case

11.1 Introduction

The supplementary section provides additional analyses and diagnostics that complement the primary stock assessment report. Diagnostic tests are important in determining the robustness of estimates for management advice in integrated stock assessment models. The diagnostics tests included in are based on diagnostics prepared for the previous assessment as well as recently developed methods (Carvallo et al. 2021). Here we present the model diagnostics for the diagnostic model run presented in the main assessment analysis.

11.2 Goodness of fit

11.2.1 Runs test and joint residual plots.

The goodness of fit of the model can be used as an indication of whether there is presence of significant model misspecification. Models that do not fit the data should be considered suspect, and further investigated. Here we use residual plots to investigate the trends and patterns in the data over time. Temporal correlation (autocorrelation) can drive bias and drift in the model estimates over time. A runs test (Wald and Wolfowitz, 1940) can test for randomness in a data sequence, such as model residuals (Carvalho et al., 2021) (Figure 4). Residuals can also be investigated along side the root mean square error (RMSE, Carvalho et al., 2017), and a joint residual plot (Winker et al., 2018), which can highlight the systematically auto-correlated residual patterns (Figure 1, Figure 2).

11.3 Model consistency

11.3.1 R0 Profile

Use of a likelihood component profile on the a global scaling parameter (or other parameter) has been identified as a key model diagnostic to identify the influence of information sources on model estimates (Carvalho et al. 2017, Ichinokawa et al., 2014; Lee et al., 2014; Wang et al., 2014). Here the equilibrium recruitment parameter, R0, is used because it represents an ideal global scaling parameter given that unfished (virgin) recruitment is proportional to unfished biomass (Carvalho et al 2021, Lee et al., 2014; Maunder and Piner, 2015; Wang et al., 2014).

A relatively large change in negative log-likelihood units along the profile suggests a relatively informative data source for that particular model. Close association in the location of the minimum negative log-likelihood along the profile between data sources suggest that model consistency, and lack of conflict in the data. Figures Figure 5, Figure 6, Figure 7 show the profile likelihoods of R0 for the overall, CPUE and length components of the model. Figure A12 shows the fit to the CPUE series for the range of LN(R0) assumed. The likelihood profiles show that overall the LN(R0) parameter is well estimated, led by the length likelihood then the index likelihood and then the recruitment. Interestingly the lower edge of the likelihood is better defined (steeper) for all three components than the upper (Figure 5). The fleet indices are generally in agreement (Figure 6), however the length data shows different minimums (Figure 7), which is consistent with the fleets that encounter different components of the stock. The fits to the CPUE series at different values of LN(R0) show that values of approximately 9.7 fit the series (Figure 6).

11.4 Age-Structured Production Model (ASPM)

This diagnostic can help evaluate whether the catch and CPUE data give evidence for a production function within the model (Carvalho 2017). Overall the ASPM evaluates whether the effect of surplus production and observed catches alone could explain trends in the CPUE, in contrast to a more complex model (i.e. SS3) that incorporates annual recruitment deviations to improve the fit (Carvalho et al 2021). Maunder and Piner (2017) note that if the ASPM fits well to the indices of abundance with contrast the production function is likely to drive the stock dynamics and the indices will provide information about absolute abundance (Minte-Vera et al., 2017). Figure A13 shows that the biomass trajectories for both models (ASPM and the diagnostic) follow the same trend and that the estimates of LN(R0) are comparable. The fits to the indices are shown in Figure 8 and indicate an overall good fit, indicating that the information content in the data is sufficient.

11.5 Retrospective analysis

Retrospective analysis is common in fisheries stock assessment to check the consistency of model estimates (Brooks and Legault, 2016; Carvalho et al., 2017; Hurtado-Ferro et al., 2015; Miller and Legault, 2017). A retrospective analysis is carried out by sequentially deleting a number of years of day (i.e. from 0 to 7) and re-running the model. Comparisons of model estimates from the full time-series and the truncated time-series can illuminate the bias and accuracy of the modelled quantities. Statistical analysis in the form of calculating the retrospective bias, rho (ρM, Mohn 1999), is common, with values between -0.15 and 0.2 being considered indicative of no bias. Figure Figure 14 shows the analysis of spawning stock biomass (SSB) and fishing mortality estimates for Indian Ocean albacore tunaalong with the Mohn's rho which indicates no retrospective bias. Forecasting the next year based on the retrospective analysis shows similar analysis (Figure 14, Figure 15).

11.6 Prediction Skill

Kell et al. (2016) proposed the hindcasting cross-validation technique (HCXval) where observations are compared to their predicted future values. The key concept behind the HCXval approach is 'prediction skill', which is defined as any measure of the accuracy of a forecasted value to the actual observed that is not known by the model (Kell et al., 2021). The difference, which is referred to as the 'prediction residual' (Michaelsen, 1987) can be evaluated by the mean absolute scaled error (MASE; Hyndman and Koehler, 2006). Carvalho et al note that a MASE score > 1 indicates that the average model forecasts are worse than a random walk. Conversely, a MASE score of 0.5 indicates that the model forecasts twice as accurately as a naïve baseline prediction; i.e. the model has prediction skill. These are presented in Figure 16 and Figure 17.

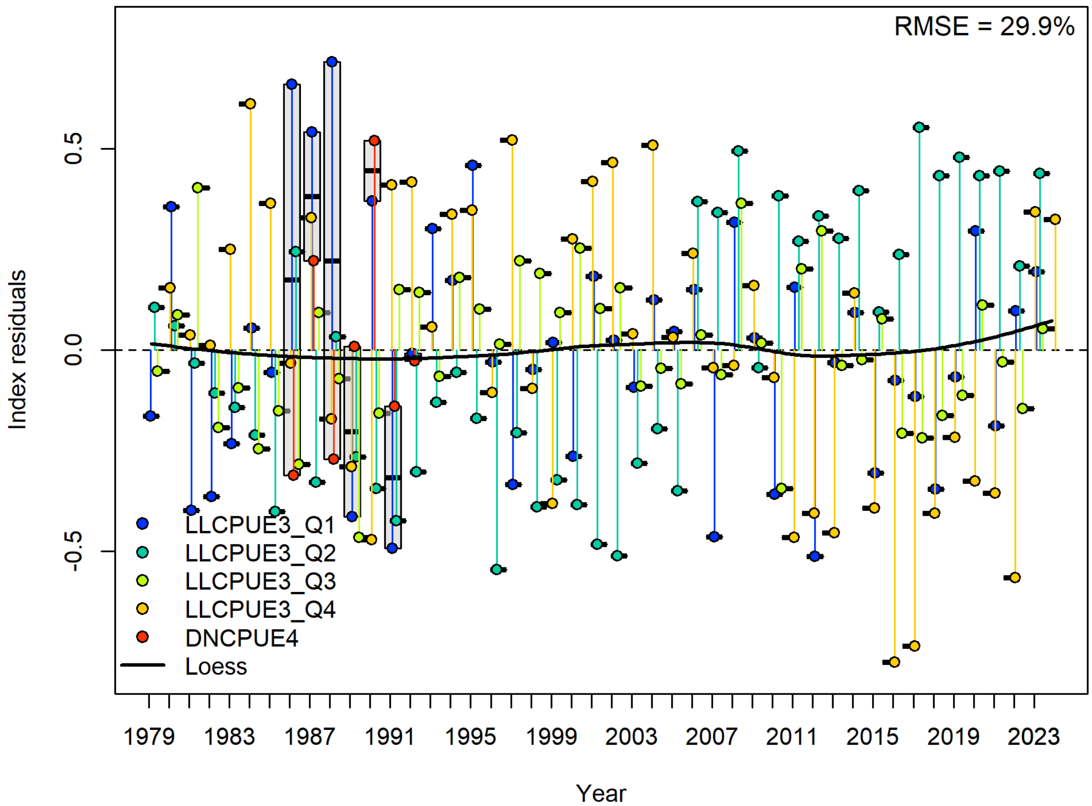


Figure 1: Joint residual plots for CPUE fits. Vertical lines with points show the residuals (in colors by index), and solid black lines show loess smoother through all residuals. Boxplots indicate the median and quantiles in cases where residuals from the multiple indices are available for any given year. Root-mean squared errors (RMSE) are included in the upper right-hand corner of the plot.

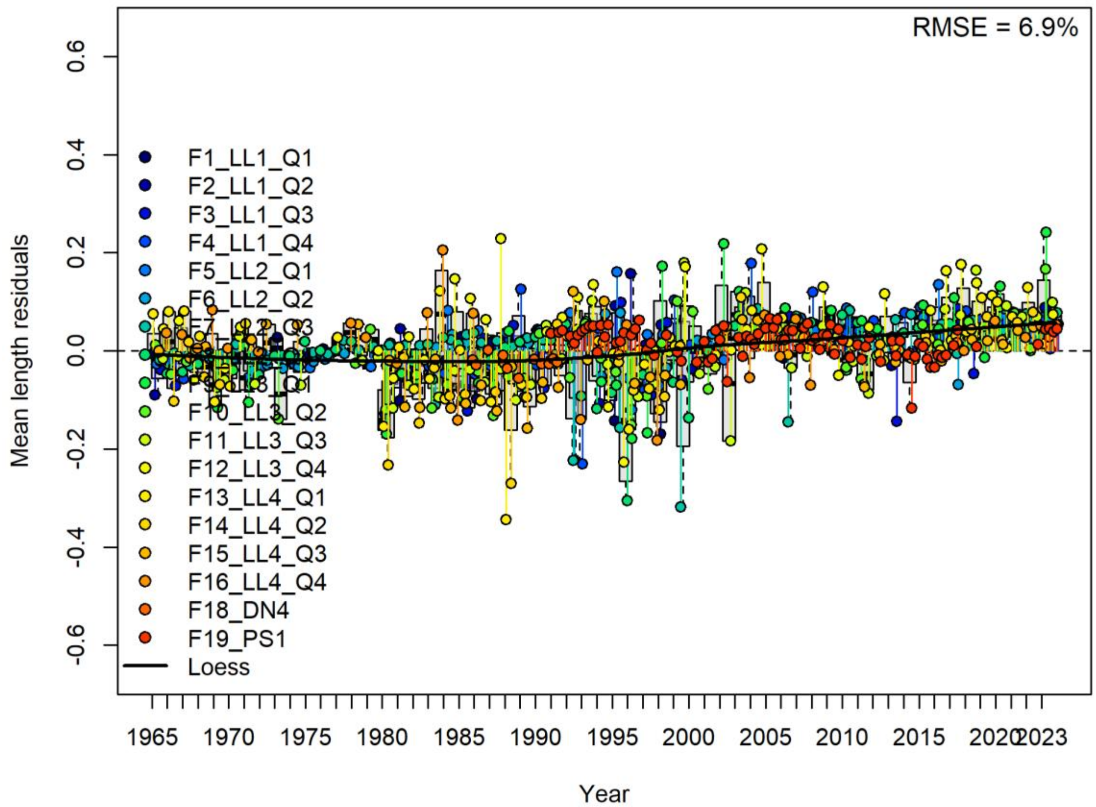


Figure 2: Joint residual plots for annual mean length estimates for multiple fishing fleets . Vertical lines with points show the residuals (in colors by index), and solid black lines show loess smoother through all residuals. Boxplots indicate the median and quantiles in cases where residuals from the multiple indices are available for any given year. Root-mean squared errors (RMSE) are included in the upper right-hand corner of the plot.

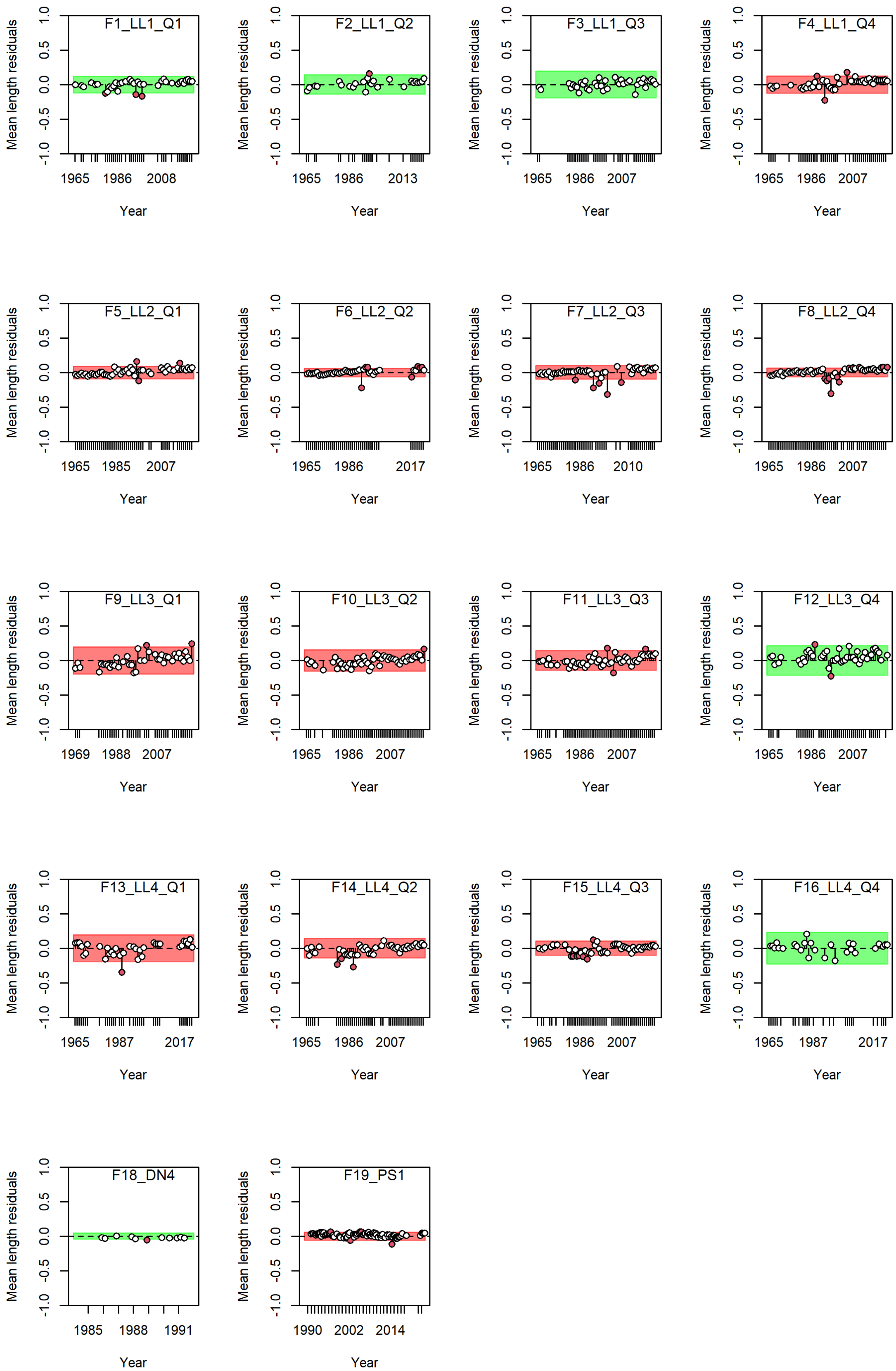

F1_LL1_Q1
F2_LL1_Q2
F3_LL1_Q3
F4_LL1_Q4
F5_LL2_Q1
F6_LL2_Q2
F7_LL2_Q3
F8_LL2_Q4
F9_LL3_Q1
F10_LL3_Q2
F11_LL3_Q3
F12_LL3_Q4
F13_LL4_Q1
F14_LL4_Q2
F15_LL4_Q3
F16_LL4_Q4
F18_DN4
F19_PS1
Mean length residuals
Year
-1.0 -0.5 0.0 0.5 1.0
1965 1986 2008
1965 1986 2013
1965 1986 2007
1965 1985 2007
1965 1986 2017
1965 1986 2010
1969 1988 2007
1965 1987 2017
1985 1988 1991
1990 2002 2014

Figure 3: Runs tests results illustrated for the length composition fits (with names and fleet numbers on top). Green shading indicates no evidence (p ≥ 0.05) and red shading evidence (p < 0.05) to reject the hypothesis of a randomly distributed time-series of residuals, respectively. The shaded (green/red) area spans three residual standard deviations to either side from zero, and the red points outside of the shading violate the 'three-sigma limit' for that series.

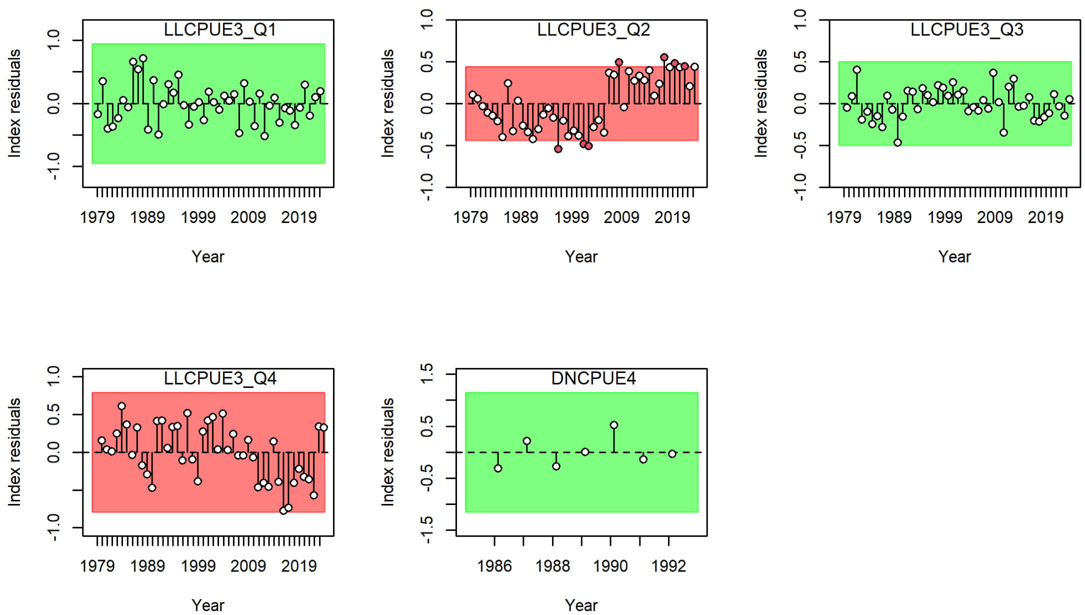


Figure 4: Runs tests results illustrated for the catch-per-unit-effort (CPUE) fits, panels indicate separate fleets. Green shading indicates no evidence (p ≥ 0.05) and red shading evidence (p < 0.05) to reject the hypothesis of a randomly distributed time-series of residuals, respectively. The shaded (green/red) area spans three residual standard deviations to either side from zero, and the red points outside of the shading violate the 'three-sigma limit' for that series.

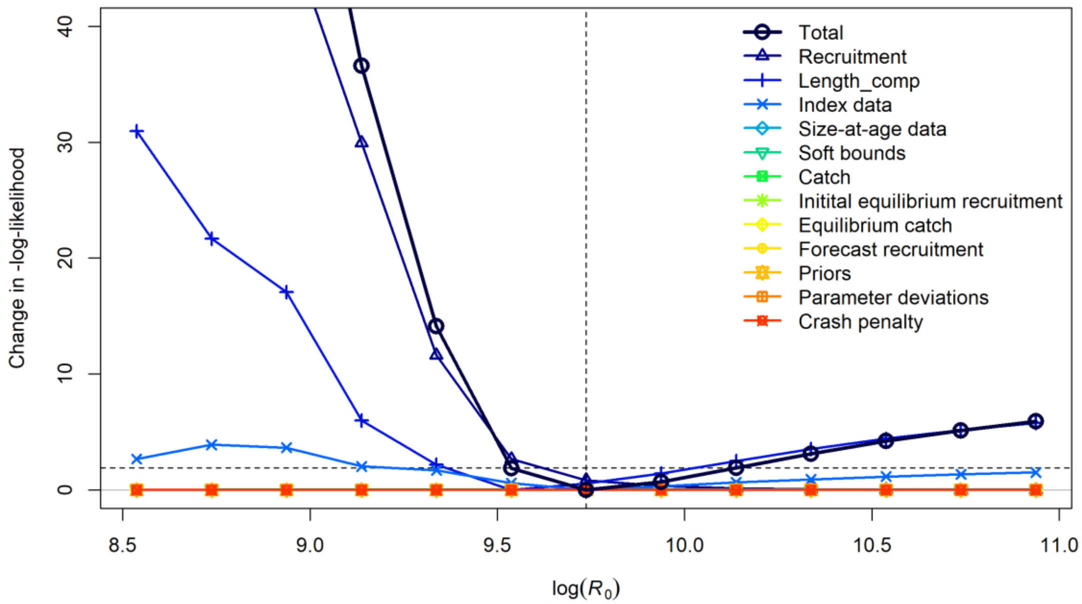


Figure 5: Log-likelihood profiles for R0 for the various data components included in the Indian Ocean albacore Stock Synthesis model , showing the contribution of (a) - (b) all data likelihood components., (c) - (d) among abundance indices and (e) - (f) among lengthand age-composition data, respectively. Total likelihood and component profiles (recruitment, length and index (CPUE) components).

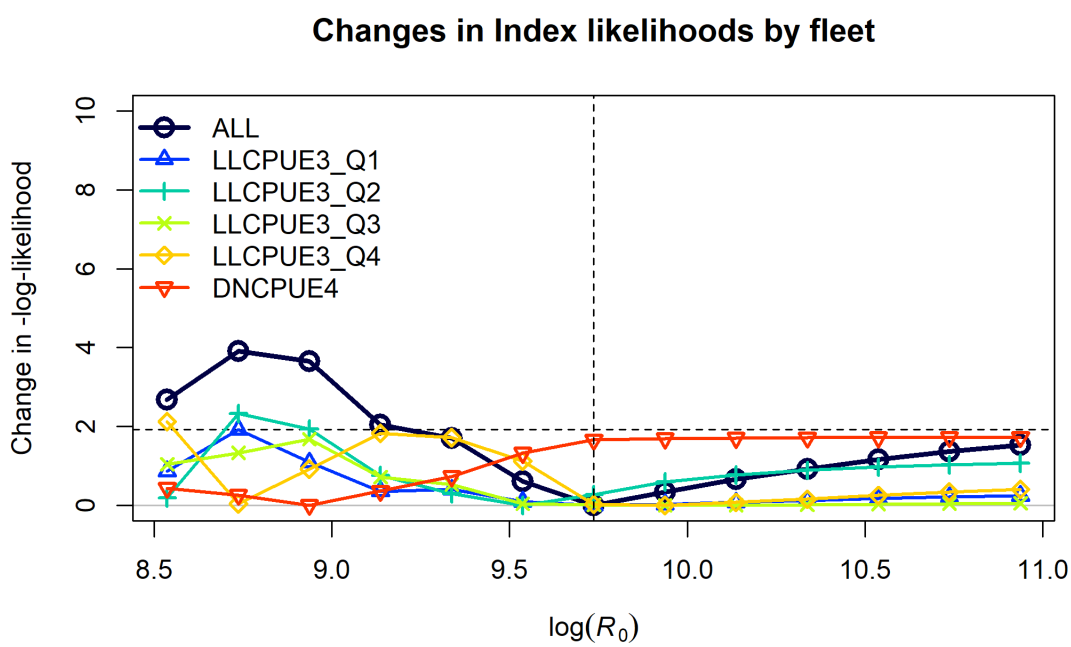


Figure 6: Log-likelihood profiles for R0 for the various data components included in the Indian Ocean albacore Stock Synthesis models, showing the contribution among abundance indices.

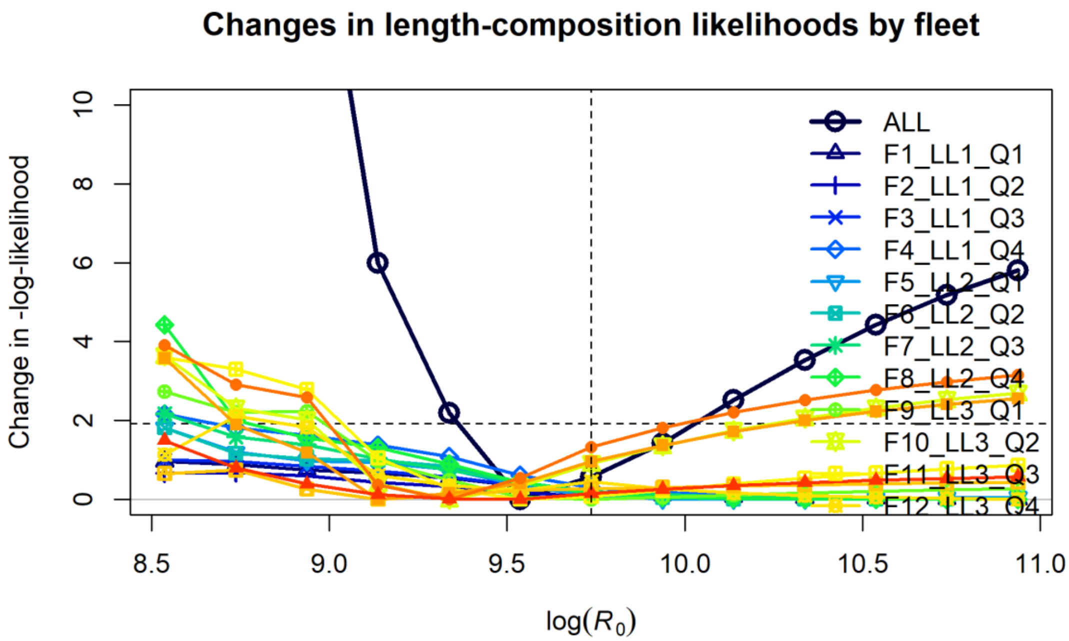


Figure 7: Log-likelihood profiles for R0 for the various data components included in the Indian Ocean albacore Stock Synthesis models, showing the contribution among length data.

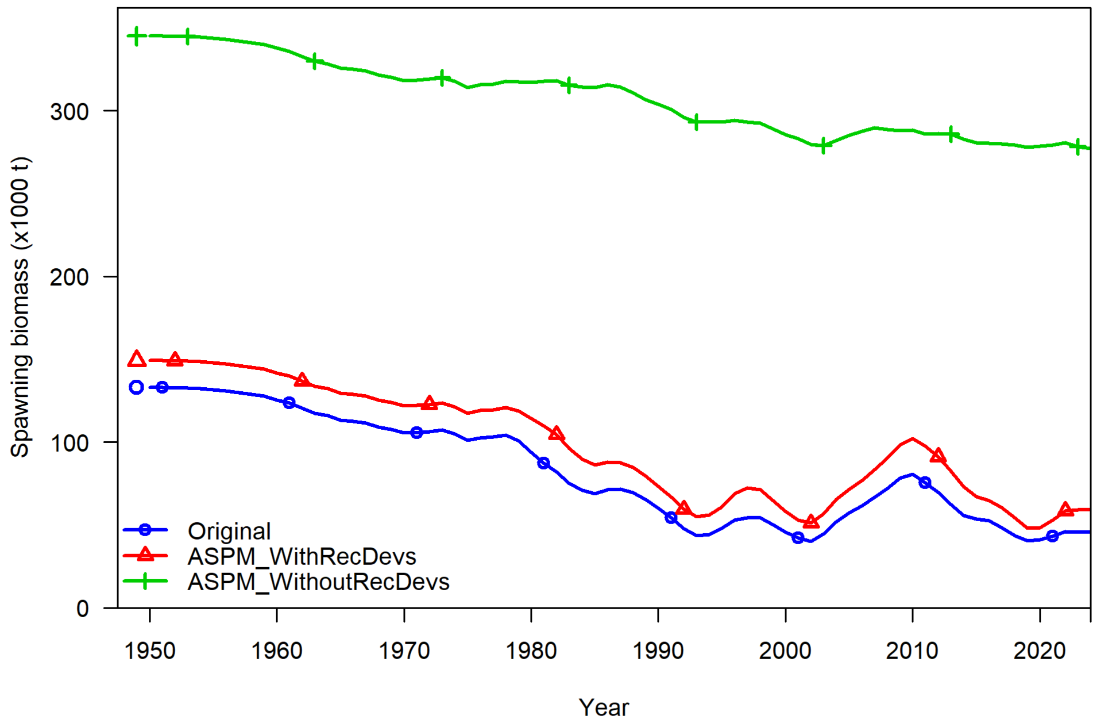


Figure 8: Comparison of spawning biomass trajectories for the ASPM and the diagnostic case of the assessment model carried out in stock synthesis.

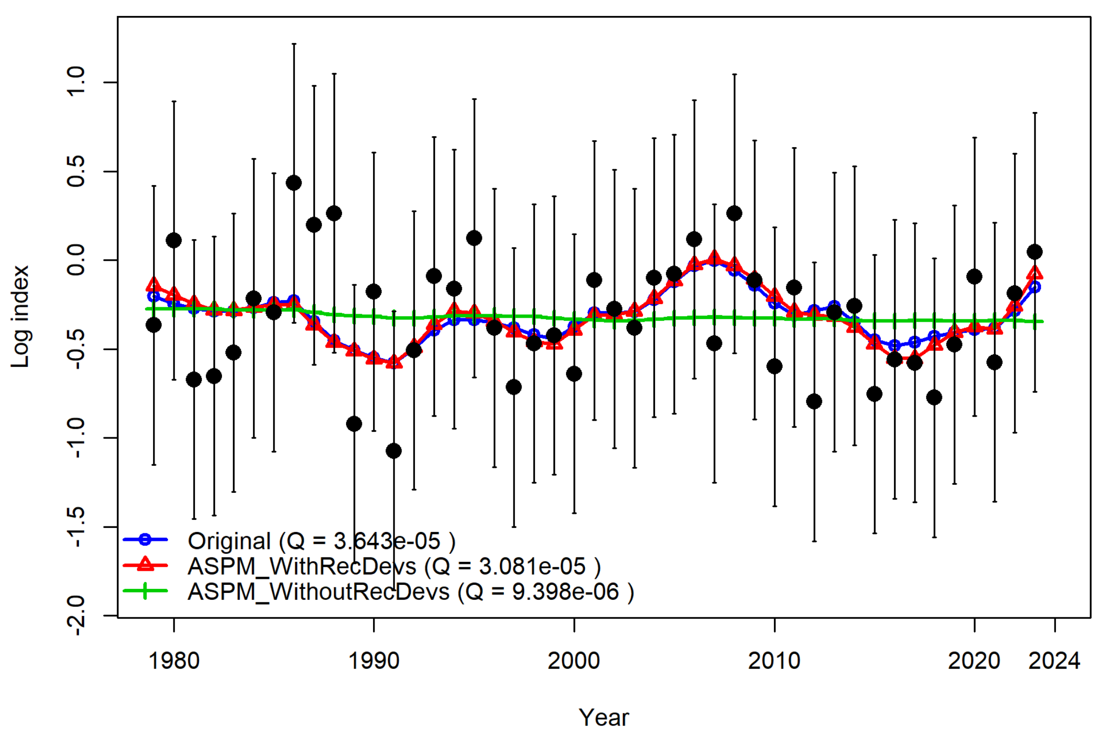


Figure 9: Fits to the indices for the ASPM diagnostic. The panels indicate the fits to the CPUE, on a log scale.

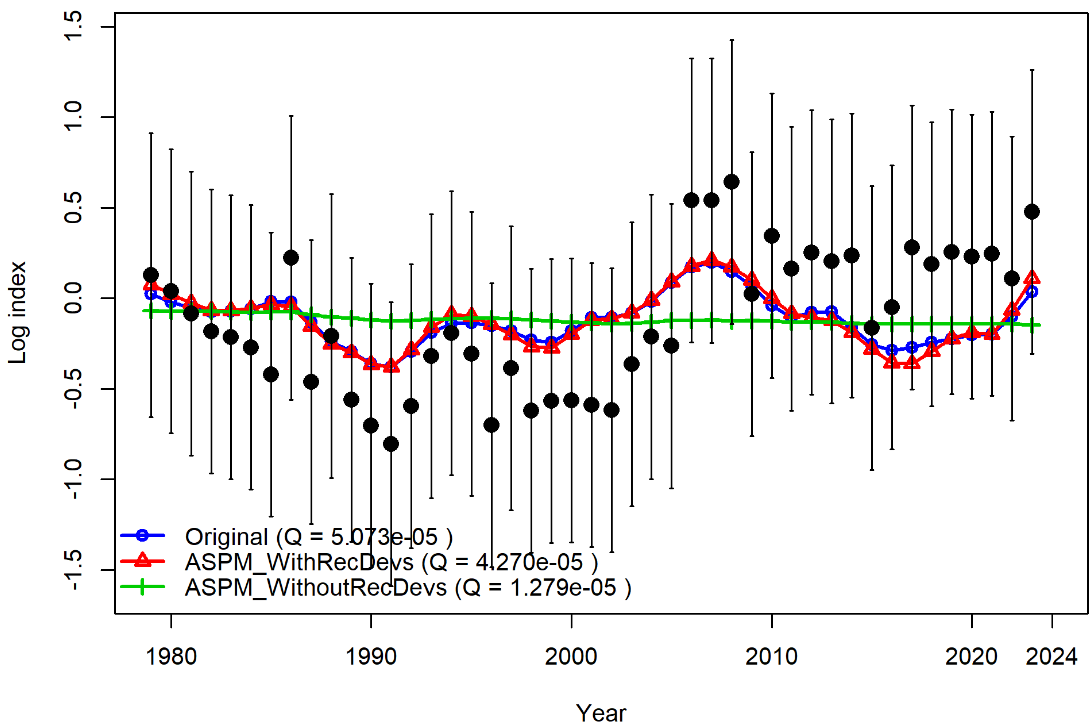


Figure 10: Fits to the indices for the ASPM diagnostic. The panels indicate the fits to the CPUE, on a log scale.

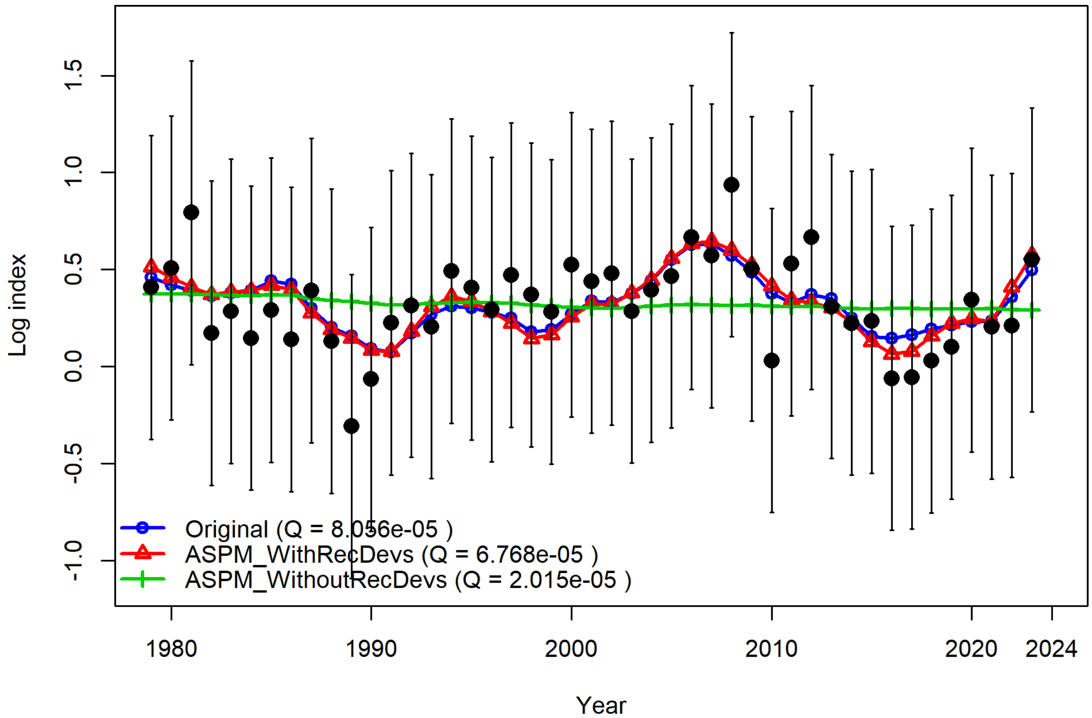


Figure 11: Fits to the indices for the ASPM diagnostic. The panels indicate the fits to the CPUE, on a log scale.

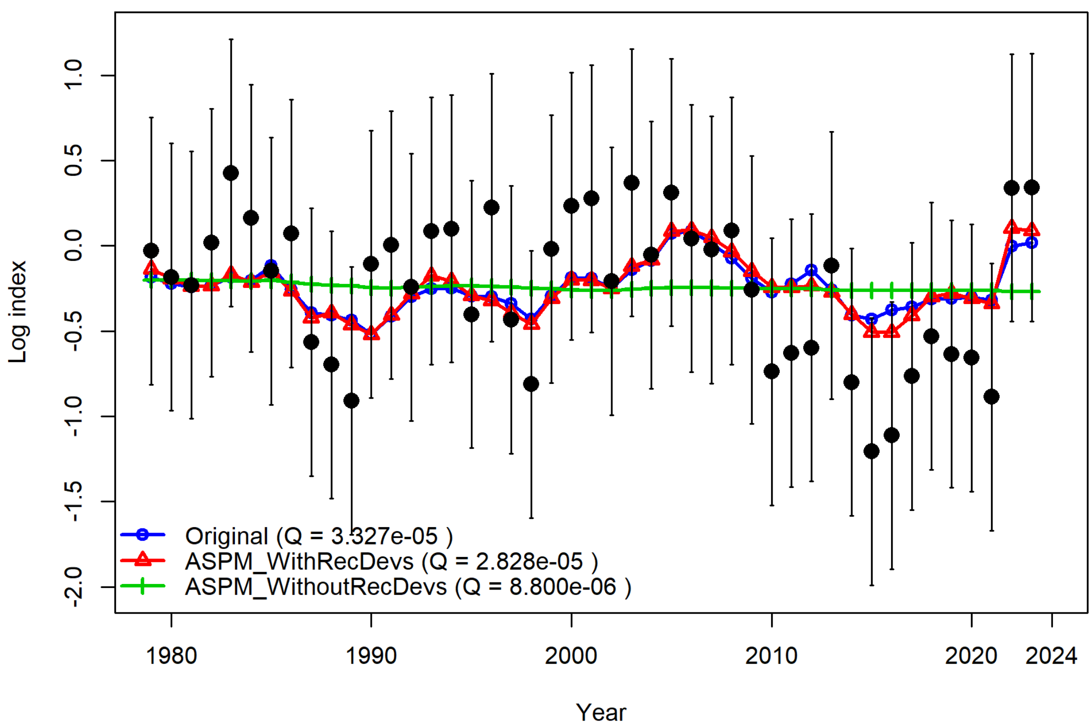


Figure 12: Fits to the indices for the ASPM diagnostic. The panels indicate the fits to the CPUE, on a log scale.

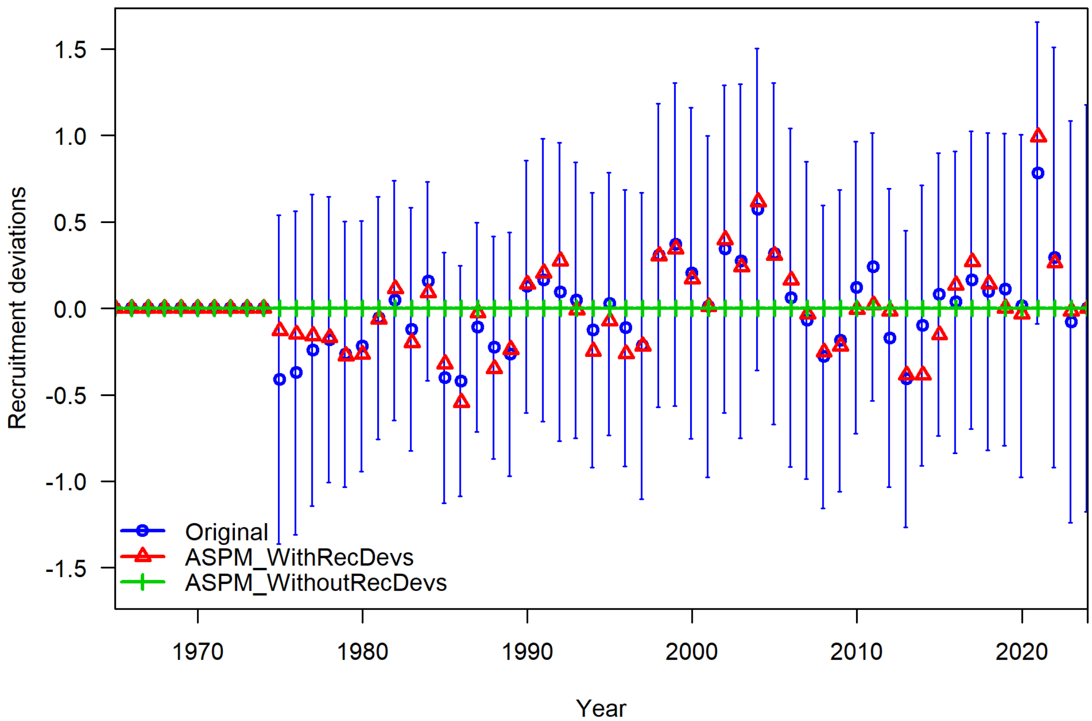


Figure 13: Comparison between the fully integrated base-case and the Age-Structured-Production Model (ASPM) results recruitment deviation estimates.

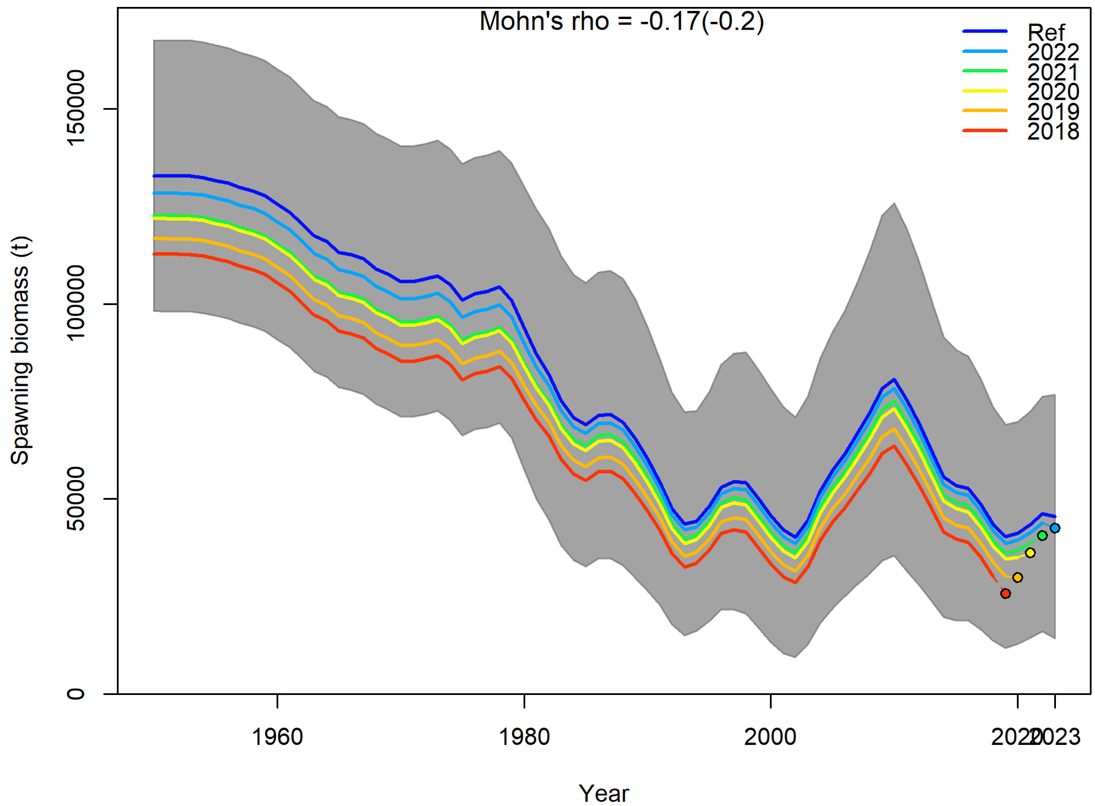


Figure 14: Retrospective analysis of spawning stock biomass (SSB) estimates for Indian Ocean albacore model conducted by re-fitting the reference model (Ref) after removing five years of observations The retrospective results are shown for (a) the entire time series and – (b) for the most recent years only. Mohn's rho statistic and the corresponding 'hindcast rho' values (in brackets) are printed at the top of the panel (b) One-year-ahead projections denoted by color-coded dashed lines with terminal points are shown for each model. Grey shaded areas are the 95 % confidence intervals from the reference model.

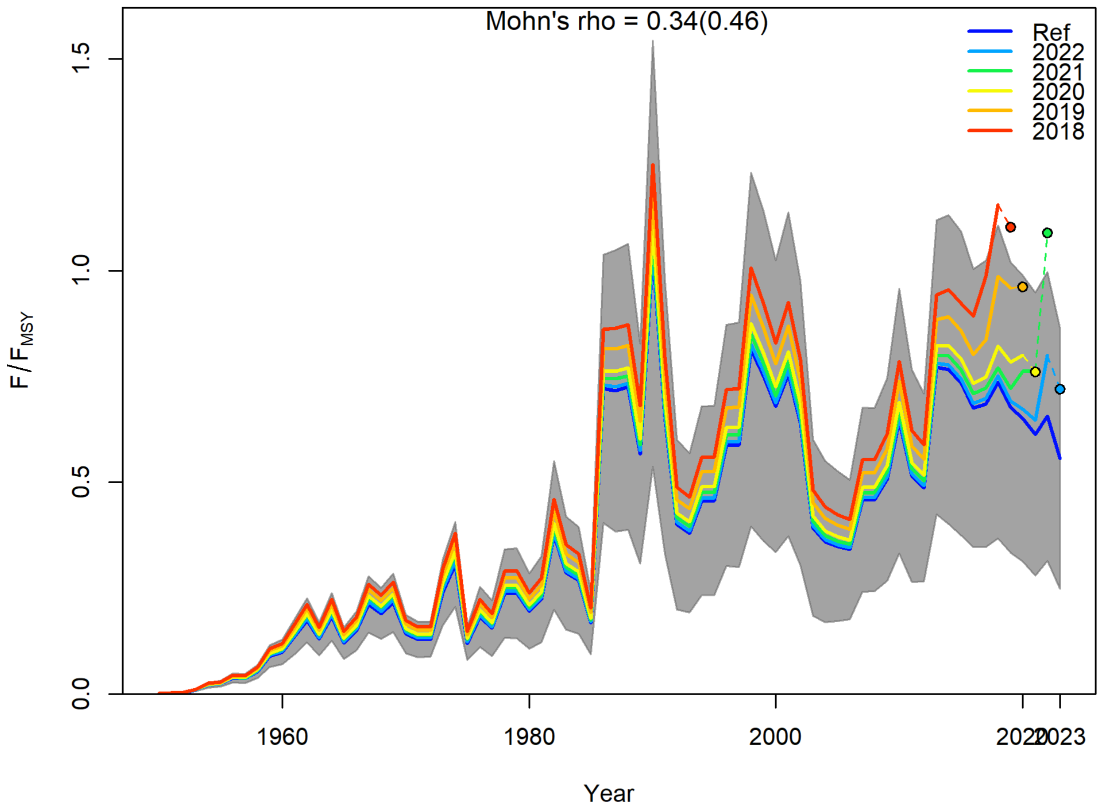


Figure 15: Retrospective analysis of fishing mortality (F) estimates for Indian Ocean albacore model conducted by re-fitting the reference model (Ref) after removing five years of observations The retrospective results are shown for (a) the entire time series and – (b) for the most recent years only. Mohn's rho statistic and the corresponding 'hindcast rho' values (in brackets) are printed at the top of the panel (b) One-year-ahead projections denoted by color-coded dashed lines with terminal points are shown for each model. Grey shaded areas are the 95 % confidence intervals from the reference model.

Figure 16: Hindcasting cross-validation (HCxval) results from CPUE fits, showing observed (large points connected with dashed line), fitted (solid lines) and one-year ahead forecast values (small terminal points). HCxval was performed using one reference model (Ref) and

seven hindcast model runs (solid lines) relative to the expected CPUE. The observations used for cross-validation are highlighted as color-coded solid circles with associated 95 % confidence intervals. The model reference year refers to the endpoints of each one-year-ahead forecast and the corresponding observation (i.e., year of retrospective + 1). The mean absolute scaled error (MASE) score associated with each CPUE."

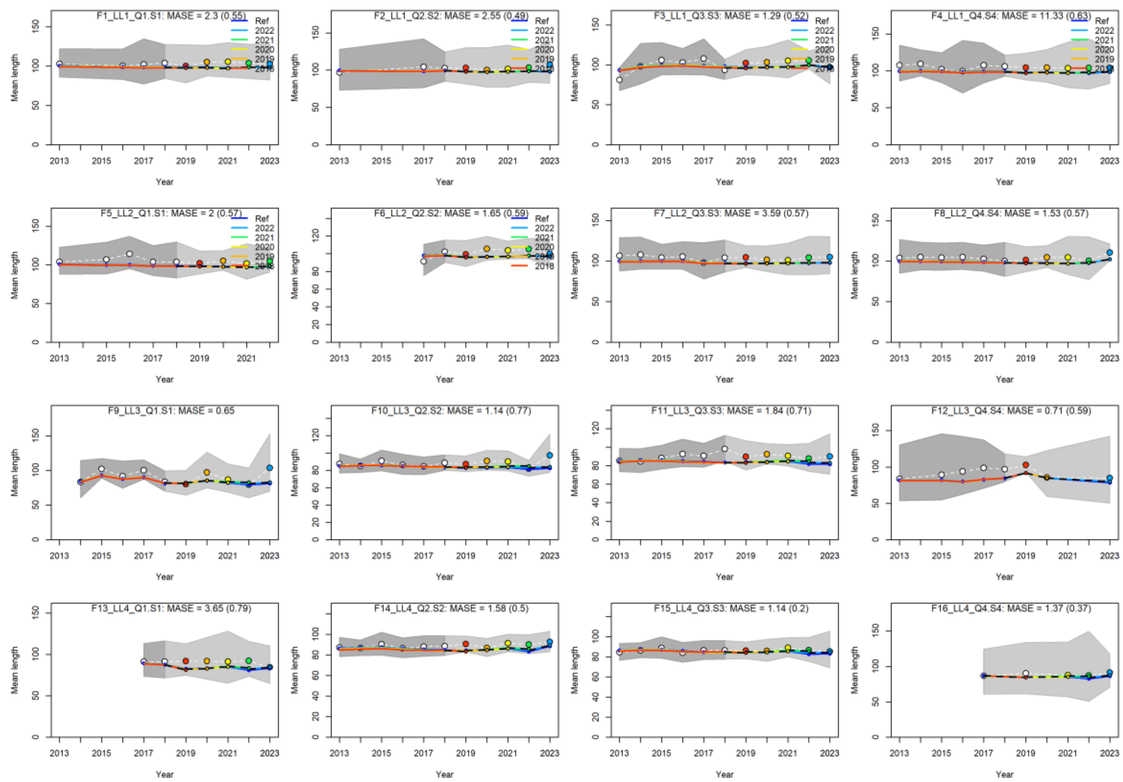


Figure 17: Hindcasting cross-validation (HCxval) results from length fits, showing observed (large points connected with dashed line), fitted (solid lines) and one-year ahead forecast values (small terminal points). HCxval was performed using one reference model (Ref) and seven hindcast model runs (solid lines) relative to the expected length. The observations used for cross-validation are highlighted as color-coded solid circles with associated 95 % confidence intervals. The model reference year refers to the endpoints of each one-year-ahead forecast and the corresponding observation (i.e., year of retrospective + 1). The mean absolute scaled error (MASE) score associated with each length composition is shown in the plot"

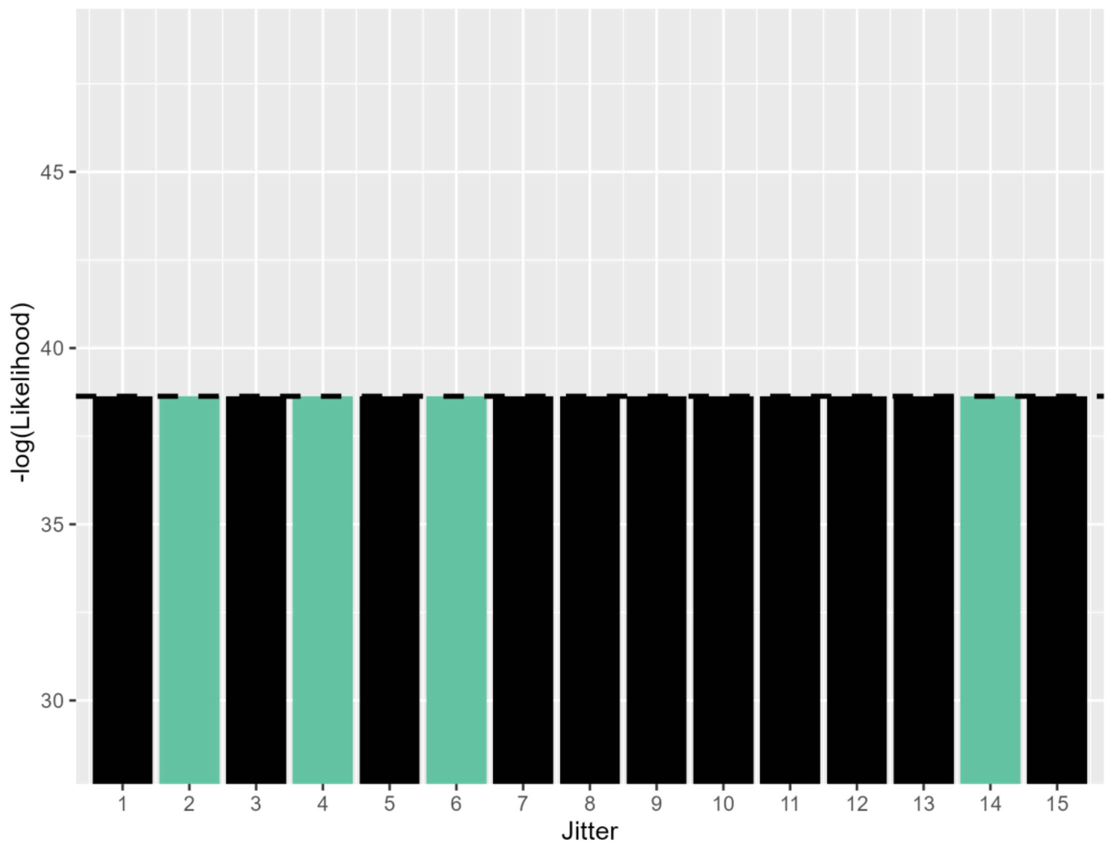


Figure 18: The jitter diagnostic for global convergence conducted for the Indian Ocean albacore model. The vertical bars represent the total likelihood obtained from jittered model runs, black bars indicate where the model found the same total likelihood value from the base-case model, missing bars indicate a failure to converge. The black horizontal dashed line represents the total likelihood value from the base-case model.

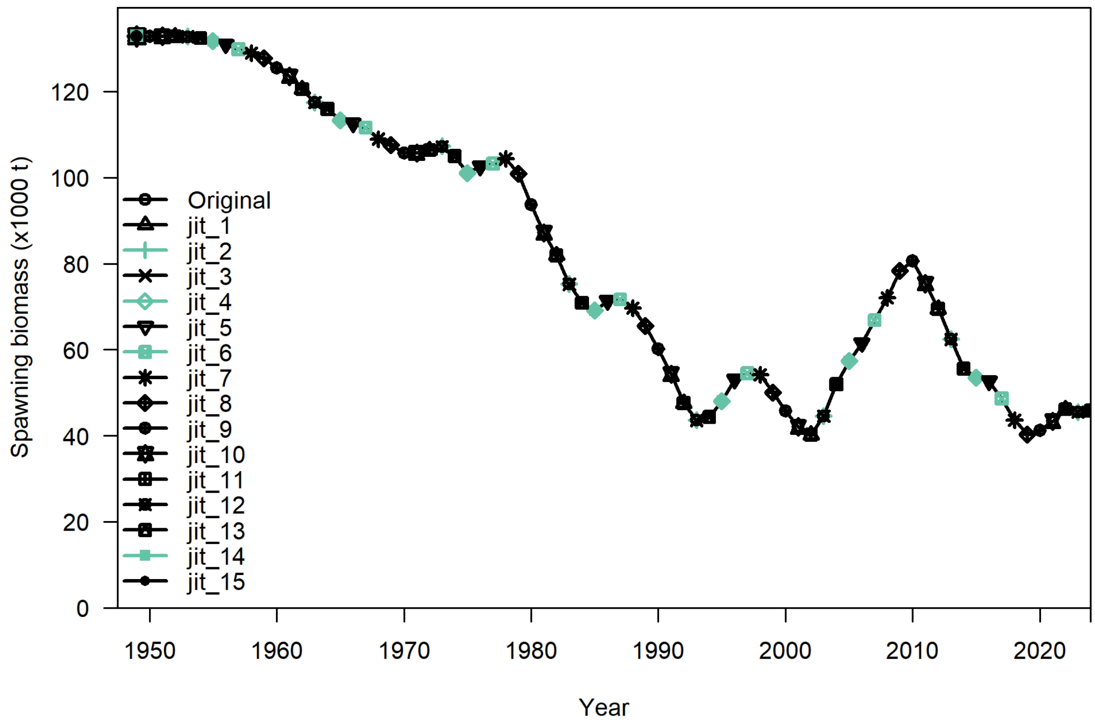


Figure 19: The jitter diagnostic for global convergence conducted for the Indian Ocean albacore model. The figure shows the spawning stock biomass (SSB) from jittered model runs. Grey shaded areas are the 95 % confidence intervals from the base-case model."

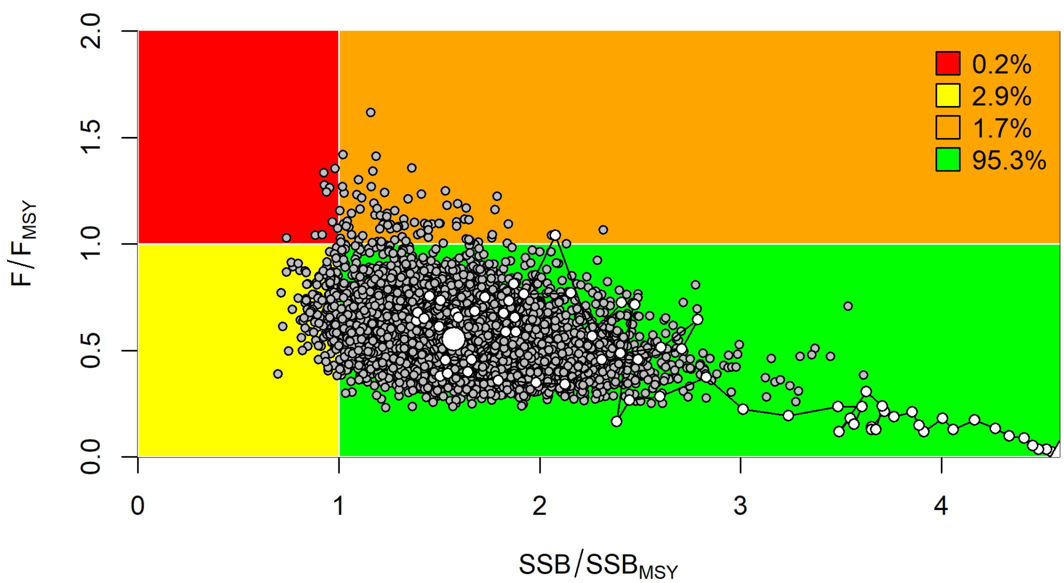


Figure 20: The multi-variate log-normal Kobe Plot."

11.7 References

Brooks, E.N., Legault, C.M., 2016. Retrospective forecasting — evaluating performance of stock projections for New England groundfish stocks. Can. J. Fish. Aquat. Sci. 73, 935–950.

Carvalho, F., Punt, A. E., Chang, Y. J., Maunder, M. N., and K. R. Piner. 2017. Can diagnostic tests help identify model misspecification in integrated stock assessments? Fisheries Research 192:28–40. Available: https://doi.org/10.1016/j.fishres.2016.09.018 (3/7/2023).

Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., Kitakado, T., Yemane, D, Piner, K. R., Maunder, M. N., Taylor, I., Wetzel, C. R., Doering, K., Johnson, K. F., and R. D. Methot. 2021. A cookbook for using model diagnostics in integrated stock assessments. Fish. Res. 240:105959. Available: https://doi.org/10.1016/j.fishres.2021.105959 (3/7/2023).

Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L., Ono, K., Vert-Pre, K.A., Whitten, A.R., Punt, A.E., 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. Ices J. Mar. Sci. 72, 99–110. https://doi.org/10.1093/icesjms/fsu198.

International Commission for the Conservation of Atlantic Tunas (ICCAT). 2017. Report of the 2017 ICCAT Shortfin Mako Data Preparatory Meeting (Madrid, Spain 28-31 March, 2017).

Kell, L. T., I. Mosqueira, P. Grosjean, J-M. Fromentin, D. Garcia, R. Hillary, E. Jardim, S. Mardle, M. A. Pastoors, J. J. Poos, F. Scott, R. D. Scott. 2007. FLR: an open-source framework for the evaluation and development of management strategies. ICES J Mar Sci, 64 (4): 640-646. doi: 10.1093/icesjms/fsm012

Lee, H.-H., Piner, K.R., Methot, R.D., Maunder, M.N., 2014. Use of likelihood profiling over a global scaling parameter to structure the population dynamics model: An example using blue marlin in the Pacific Ocean. Fish. Res. 158, 138–146. https://doi.org/10.1016/j.fishres.2013.12.017.

Maunder, M.N., Piner, K.R., 2015. Contemporary fisheries stock assessment: many issues still remain. ICES J. Mar. Sci. 72, 7–18. https://doi.org/10.1093/icesjms/fsu015.

Michaelsen, J., 1987. Cross-validation in statistical climate forecast models. J. Clim. Appl. Meteorol. 26, 1589–1600. https://doi.org/10.1175/1520-0450(1987) 026<1589:CVISCF>2.0.CO;2.

Miller, T.J., Legault, C.M., 2017. Statistical behavior of retrospective patterns and their effects on estimation of stock and harvest status. Fish. Res. 186, 109–120. https:// doi.org/10.1016/j.fishres.2016.08.002.

Minte-Vera, C.V., Maunder, M.N., Aires-da-Silva, A.M., Satoh, K., Uosaki, K., 2017. Get the biology right, or use size-composition data at your own risk. Fish. Res. 192, 114–125. https://doi.org/10.1016/j.fishres.2017.01.014.

Mohn, R., 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES J. Mar. Sci. 56, 473–488. https://doi.org/10.1006/jmsc.1999.0481

Wald, A., Wolfowitz, J., 1940. On a test whether two samples are from the same population. Ann. Math. Stat. 11, 147–162. http://www.jstor.org/stable/2235872.
Winker, H., Carvalho, F., Kerwath, S., 2020. Age-structured biomass dynamics of North Atlantic shortfin mako with implications for the interpretation of surplus production models. Col. Vol. Sci. Pap. ICCAT 76, 316–336. https://www.iccat.int/Documents/CVSP/CV076_2019/colvol76.html.

12 Annex3 Sensitivity and model exploration runs

Sensitivity and model exploration runs for the 2026 Indian Ocean Albacore Stock Assessment

12.1 Overview

Sensitivity analyses were conducted to characterize structural uncertainty in the assessment and evaluate the robustness of model results to alternative biological assumptions, stock–recruitment dynamics, data weighting, and model structure. The analyses included alternative fixed values of stock–recruitment steepness, increased variation in sex-specific length-at-age, alternative length–weight relationships, age-specific natural mortality, and an alternative maturity-at-length relationship based on recent observations from the western Indian Ocean. Additional analyses examined the influence of the relative weighting assigned to CPUE indices and size-composition data by increasing the contribution of longline length-composition information to the likelihood. Finally, likelihood profile analyses of the terminal selectivity parameters for selected fleets were undertaken to examine parameter identifiability, evaluate the extent to which these parameters were informed by the available data, and quantify the sensitivity of model fit and derived quantities to alternative assumptions regarding terminal selectivity.

The sensitivity results were evaluated by comparing spawning biomass trajectories, biomass relative to unfished levels, fishing mortality relative to reference levels, recruitment deviations, fits to the quarterly CPUE indices, and the estimated stock–recruitment scale parameter (SR_LN(R0)). Together, these comparisons were used to identify which structural assumptions had the greatest influence on estimated stock status and population scale, and to assess whether alternative biological parameterizations or data-weighting schemes resulted in materially different interpretations of historical biomass trends, fishing pressure, or recent stock dynamics.

12.2 NW Diagnostic model

The estimated and derived quantities for the diagnostic and sensitivity models related to the NW diagnostic model are summarized in Table Table 1. Most alternative biological parameterizations resulted in modest changes to equilibrium quantities, including unfished spawning biomass, unfished recruitment, spawning biomass at MSY, and equilibrium yield. As expected, changes to the assumed stock–recruitment steepness, growth variability, natural mortality, maturity-at-length, and length–weight relationship primarily affected the overall scale of the population and associated reference points, while maintaining relatively consistent estimates of recent stock status. Across these sensitivity runs, the estimated spawning biomass in 2023 remained at or above the estimated biomass reference point (B_2023/B_MSY ≈ 1.0–

1.1), whereas fishing mortality remained above the reference level (F/F_MSY ≈ 1.1–1.3), indicating that the qualitative assessment conclusions were robust to these alternative biological assumptions.

The largest departure from the diagnostic model was observed in the sensitivity that removed the post-Francis down-weighting of the longline length-composition data. Increasing the influence of the length-composition information substantially increased the contribution of the length-composition likelihood to the objective function and produced appreciable changes in estimated unfished biomass, recruitment, MSY quantities, and current stock status. This result is consistent with the increased influence of size-composition data on parameter estimation and highlights the importance of appropriately balancing the relative contributions of CPUE and size-composition data within the assessment. Overall, however, the remaining structural sensitivity analyses produced comparatively small differences in model outputs, suggesting that the assessment is relatively robust to the range of alternative biological parameterizations considered.

The likelihood profile over terminal selectivity for fleets 9–15 indicated that the model fit was sensitive to the assumed terminal selectivity value(Figure 9). The profile likelihood was minimized at intermediate terminal selectivity values, whereas both low and high terminal selectivity assumptions resulted in poorer fits. Most of the change in the likelihood was attributable to the length-composition component, indicating that the size-composition data contain information about the descending limb and terminal portion of the selectivity curves for these fleets.

The terminal selectivity assumptions also affected derived stock quantities. Lower terminal selectivity values resulted in higher estimates of spawning biomass and depletion, and lower estimates of fishing mortality relative to the reference level. In contrast, higher terminal selectivity values produced lower spawning biomass and depletion estimates and higher relative fishing mortality. This result indicates that terminal selectivity for these fleets is an influential structural assumption and should be considered when evaluating uncertainty in stock status.

12.3 Southwest diagnostic case

The southwest (SW) diagnostic case was used to evaluate how the model responded to the same set of alternative biological assumptions and data-weighting configurations examined for the northwest (NW) diagnostic case. These sensitivities included alternative stock–recruitment steepness values, increased variation in length-at-age, alternative length–weight parameters, Lorenzen natural mortality, an alternative maturity-at-length relationship, and increased weighting of the longline length-composition data.

Overall, the SW sensitivity runs produced broadly similar temporal patterns in spawning biomass and fishing mortality, although the absolute scale of the population differed among biological assumptions. The maturity-at-length sensitivity produced the highest spawning biomass estimates, while the increased length-composition weighting case produced lower spawning biomass and a distinct recruitment-deviation pattern. Despite these differences in scale, the main temporal signal was consistent across most sensitivities: spawning biomass declined from the unfished period through the early 1990s, increased during the 2000s, and declined again in the most recent period.

The fits to the quarterly SW CPUE indices were generally similar across most sensitivity runs, indicating that the alternative biological parameterizations had limited influence on the predicted CPUE trajectories after re-estimation of catchability. The increased length-composition weighting sensitivity differed more noticeably from the other runs in some years, reflecting the stronger influence of size-composition data on the estimated population trajectory and recruitment deviations. The SW indices also show relatively broad uncertainty intervals, particularly during some earlier and mid-series years, which reduces the ability of the CPUE data alone to distinguish strongly among the sensitivity cases.

The estimated scale parameter, SR_LN(R0), varied among sensitivity runs (Figure 20), indicating that alternative biological assumptions affected the inferred scale of the population. The Lorenzen natural mortality sensitivity produced a substantially higher estimate of SR_LN(R0), while the increased length-composition weighting case produced a lower and more precise estimate relative to most other runs. This pattern is consistent with the interpretation that biological assumptions primarily influence population scale, whereas changes in data weighting can alter both scale and the relative fit to composition and CPUE data.

A likelihood profile was conducted for the terminal selectivity parameters of the southwest diagnostic case to evaluate the extent to which these parameters were informed by the available data and to assess their influence on derived stock quantities. As in the northwest diagnostic case, terminal selectivity values were fixed across a range of assumed values and the model was re-estimated for each value. The likelihood profile indicates that the model fit is sensitive to terminal selectivity, with the length-composition likelihood accounting for most of the change in the objective function (Figure 21).

The terminal selectivity assumptions had a clear influence on estimated population scale and stock status. Lower terminal selectivity values generally produced higher estimates of spawning biomass, unfished spawning biomass, and relative biomass, whereas higher terminal selectivity values resulted in lower biomass estimates and higher relative fishing mortality. This pattern indicates that terminal selectivity is an influential structural assumption for the southwest diagnostic case and should be considered when interpreting uncertainty in stock status.

12.4 Figures and Tables NW Diagnostic model.

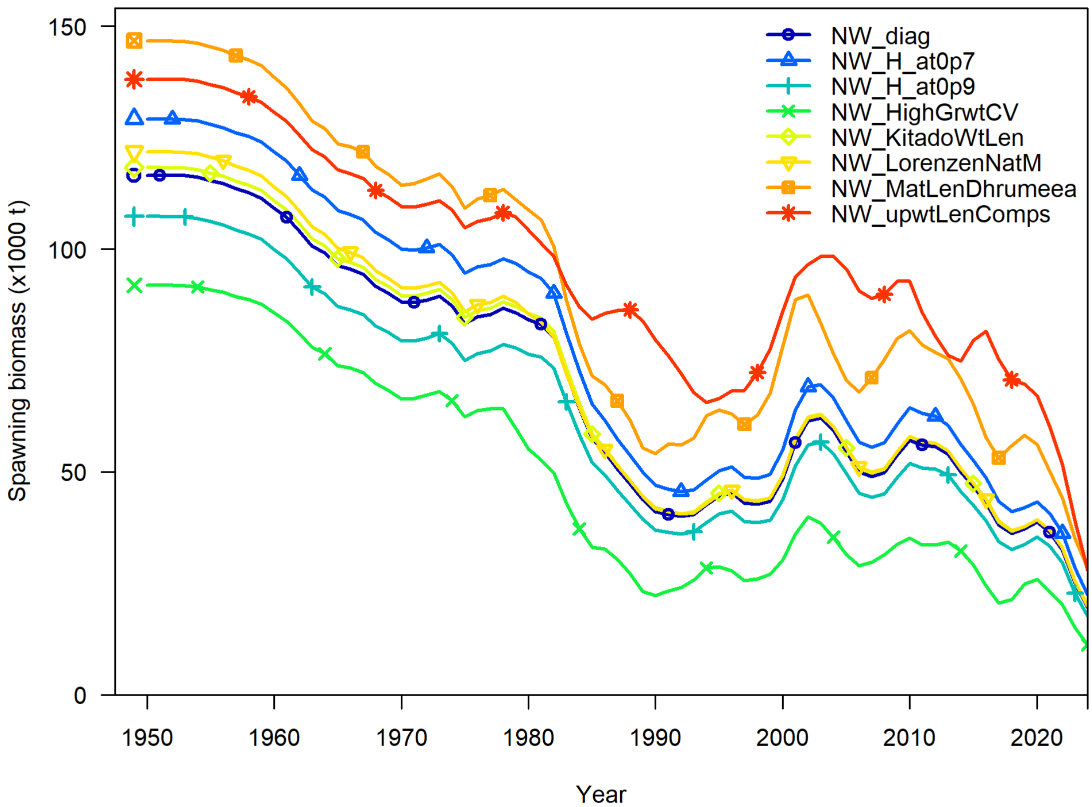


Figure 1: Comparison of estimated spawning biomass trajectories across northwest diagnostic and sensitivity model runs. Spawning biomass is shown in thousands of tonnes. The sensitivity models include alternative stock–recruitment steepness values, growth variation, length–weight relationships, natural mortality, maturity-at-length, and length-composition weighting assumptions.

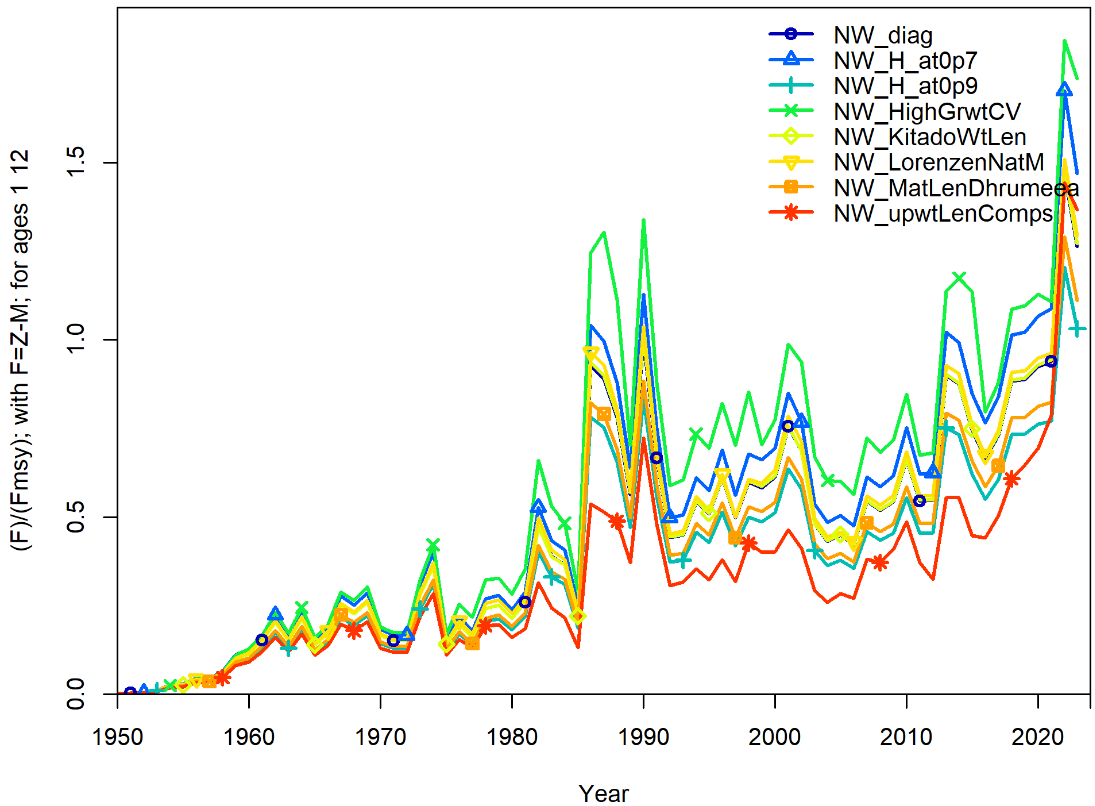


Figure 2: Comparison of fishing mortality relative to the reference level across northwest diagnostic and sensitivity model runs. Lines show annual estimates of F relative to the reference fishing mortality for ages 1–12. Differences among models reflect the influence of alternative biological assumptions and data-weighting schemes on estimated fishing pressure.

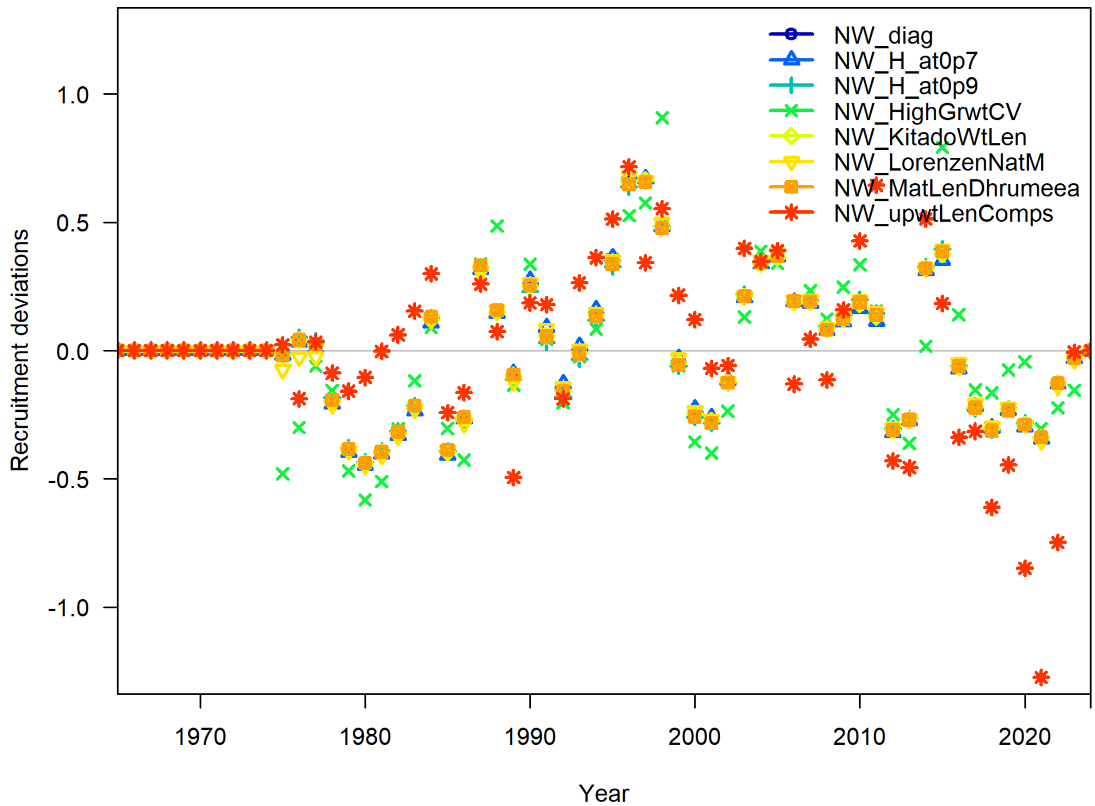


Figure 3: Comparison of estimated recruitment deviations across northwest diagnostic and sensitivity model runs. Points show annual recruitment deviations by model, highlighting differences in recruitment patterns implied by alternative biological assumptions and data-weighting configurations.

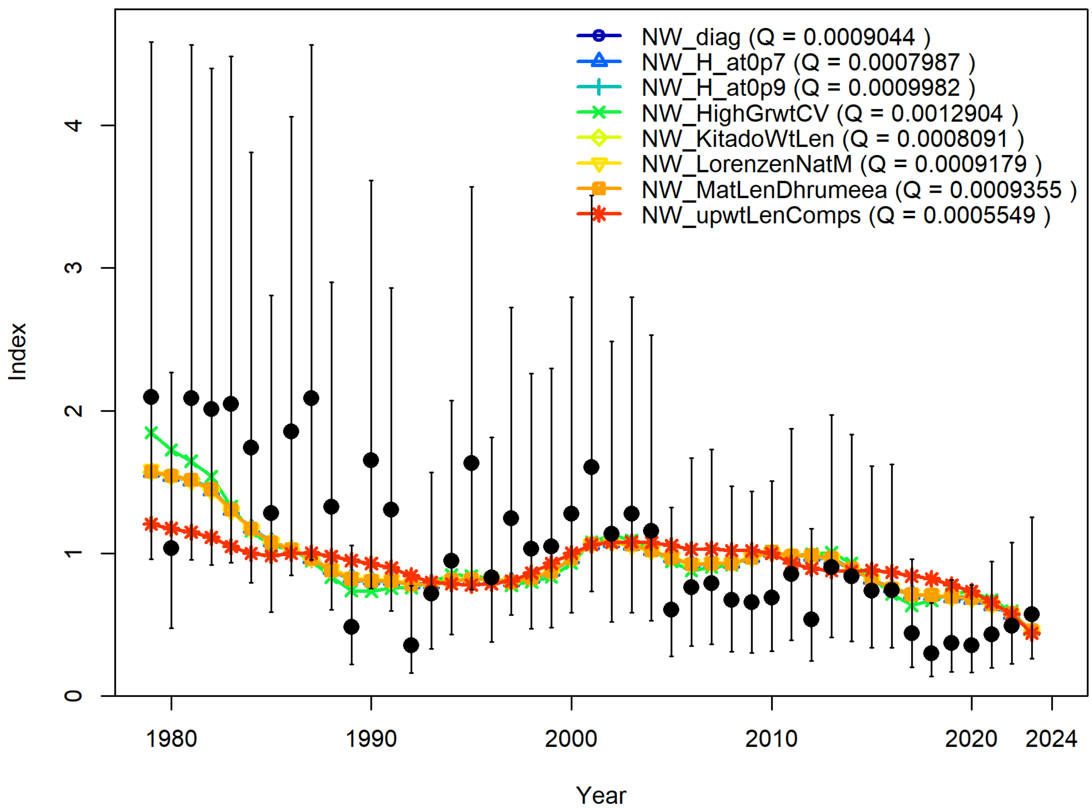


Figure 4: Comparison of model fits to the northwest quarter 1 CPUE index across diagnostic and sensitivity model runs. Observed index values are shown with uncertainty intervals, and colored lines show the predicted index trajectories from each model. Estimated catchability values are shown in the legend.

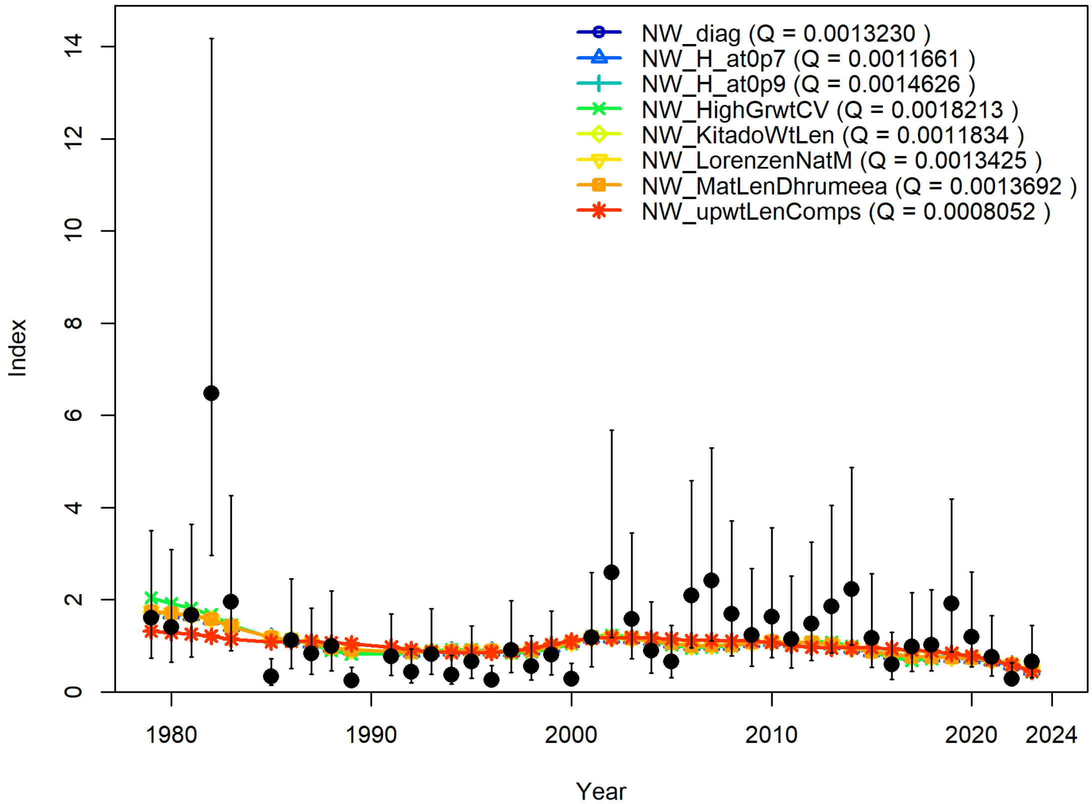


Figure 5: Comparison of model fits to the northwest quarter 2 CPUE index across diagnostic and sensitivity model runs. Observed index values are shown with uncertainty intervals, and colored lines show the predicted index trajectories from each model. Estimated catchability values are shown in the legend.

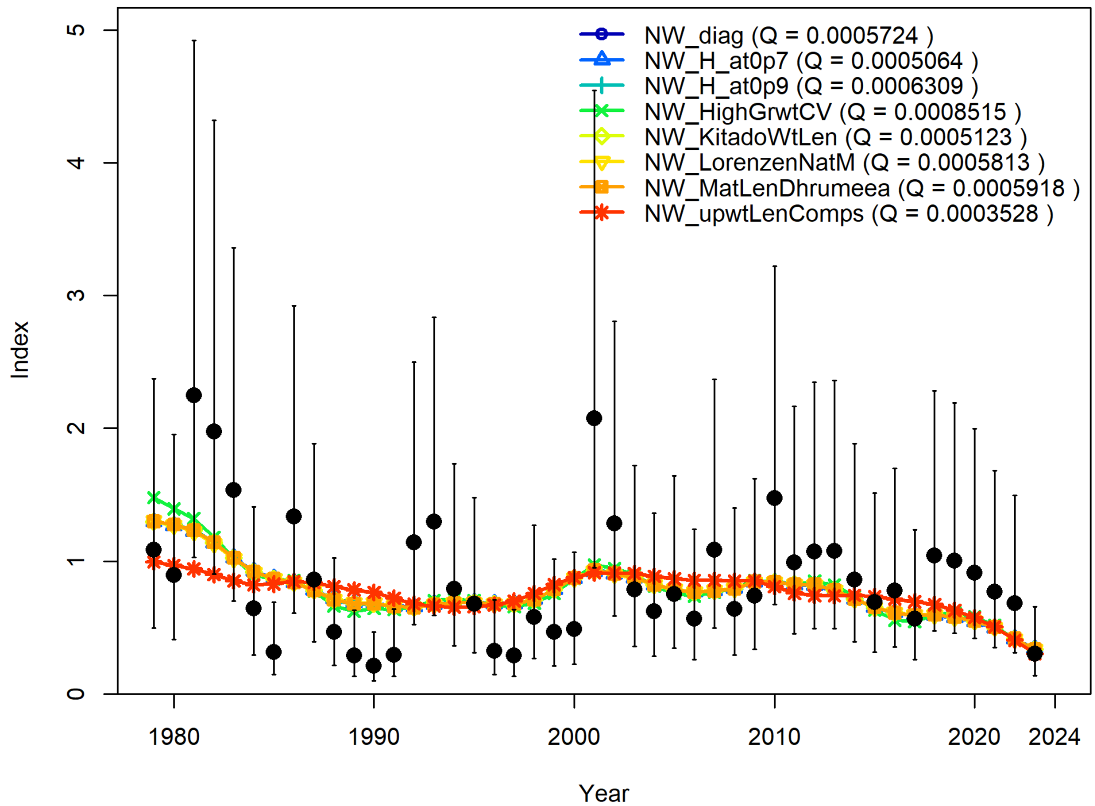


Figure 6: Comparison of model fits to the northwest quarter 3 CPUE index across diagnostic and sensitivity model runs. Observed index values are shown with uncertainty intervals, and colored lines show the predicted index trajectories from each model. Estimated catchability values are shown in the legend.

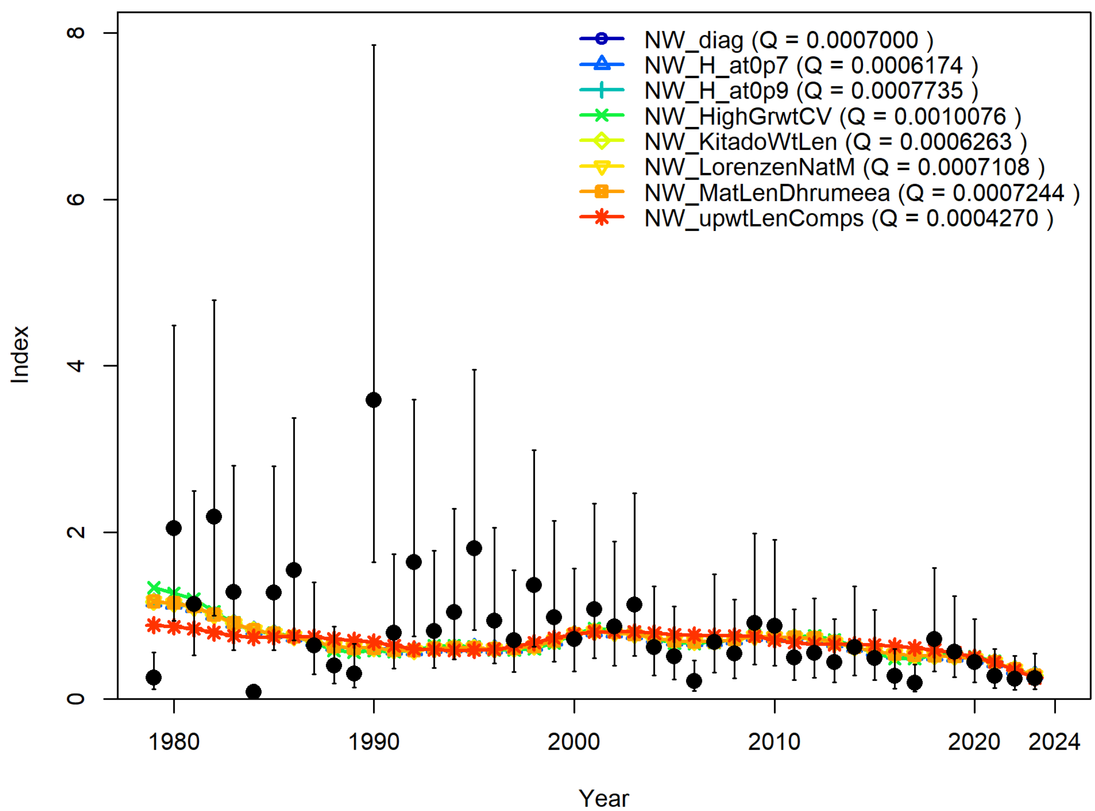


Figure 7: Comparison of model fits to the northwest quarter 4 CPUE index across diagnostic and sensitivity model runs. Observed index values are shown with uncertainty intervals, and colored lines show the predicted index trajectories from each model. Estimated catchability values are shown in the legend.

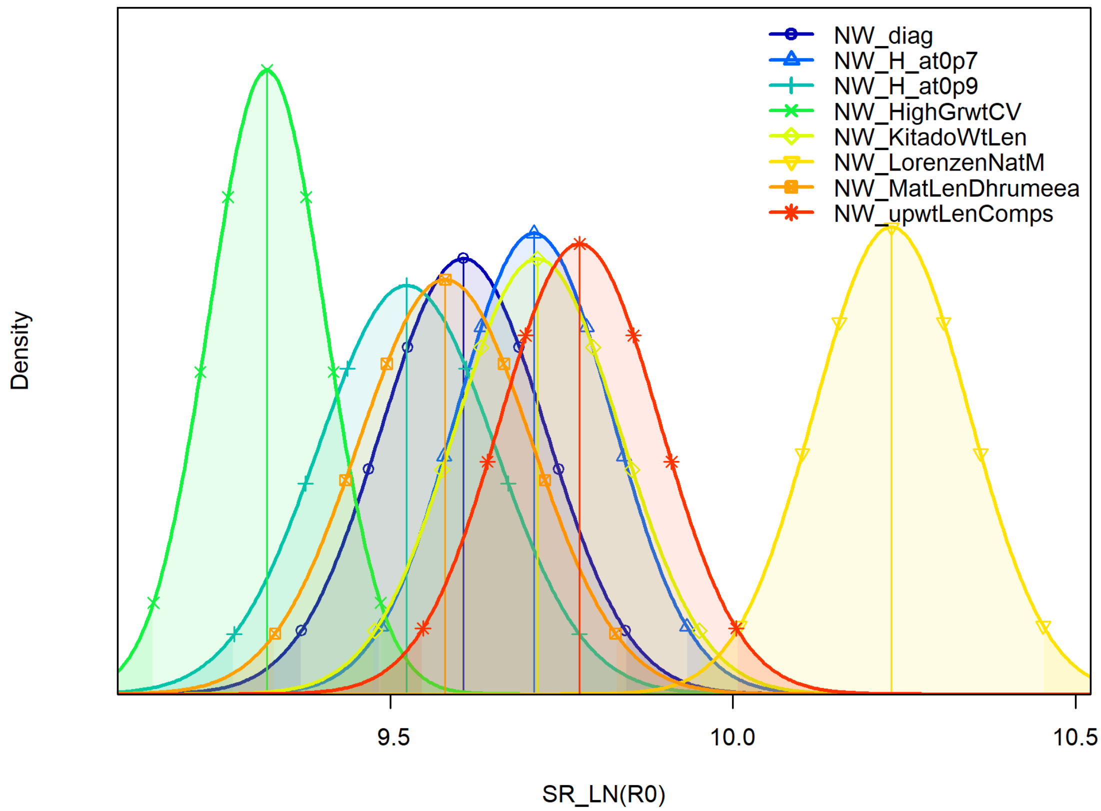


Figure 8: Comparison of approximate distributions for the estimated stock–recruitment scale parameter, SR_LN(R0), across northwest diagnostic and sensitivity model runs. Vertical lines indicate point estimates for each model, and density curves illustrate relative differences in estimated population scale among sensitivity runs.

Table 1: Estimated and derived quantities for the northwest diagnostic model and structural sensitivity analyses. Sensitivity runs include alternative stock–recruitment steepness, growth variability, length–weight relationships, natural mortality, maturity-at-length, and longline length-composition weighting.

Label	model 1	model 2	model 3	model 4	model 5	model 6	model 7	model 8
TOTAL_like	163.150	163.493	162.892	212.112	163.186	165.105	163.084	1035.880
Survey_like	9.080	9.665	8.581	4.728	9.143	9.716	8.993	24.962
Length_comp_like	108.466	108.154	108.765	113.272	108.451	109.617	108.522	960.037
Parm_priors_like	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Size_at_age_like	59.239	59.240	59.238	105.236	59.240	59.240	59.238	59.241

SSB_unfished_thousand_mt	116.590	129.207	107.336	91.914	118.282	121.864	146.766	138.126
B_MSY/SSB_unfished	0.212	0.257	0.155	0.221	0.214	0.207	0.229	0.208
Totbio_unfished	501174.000	555407.000	461395.000	385545.000	500900.000	512962.000	488012.000	593746.000
SmryBio_unfished	500831.000	555027.000	461079.000	385284.000	500639.000	512322.000	487678.000	593339.000
Recr_unfished_millions	14.867	16.476	13.687	11.161	16.557	27.763	14.477	17.613
SSB_MSY_thousand_mt	24.696	33.248	16.673	20.354	25.276	25.214	33.647	28.672
SPR_MSY	0.261	0.337	0.179	0.270	0.263	0.256	0.277	0.257
annF_MSY	0.156	0.128	0.197	0.168	0.155	0.148	0.180	0.167
Dead_Catch_MSY	41079.200	38651.000	44570.400	35902.500	40966.900	40274.500	43233.400	47330.100
Ret_Catch_MSY	41079.200	38651.000	44570.400	35902.500	40966.900	40274.500	43233.400	47330.100
B_MSY/SSB_unfished	0.212	0.257	0.155	0.221	0.214	0.207	0.229	0.208
Bratio_2023	1.016	0.855	1.363	0.736	1.009	1.012	1.041	1.362
F_2023	1.264	1.470	1.031	1.738	1.271	1.295	1.112	1.368

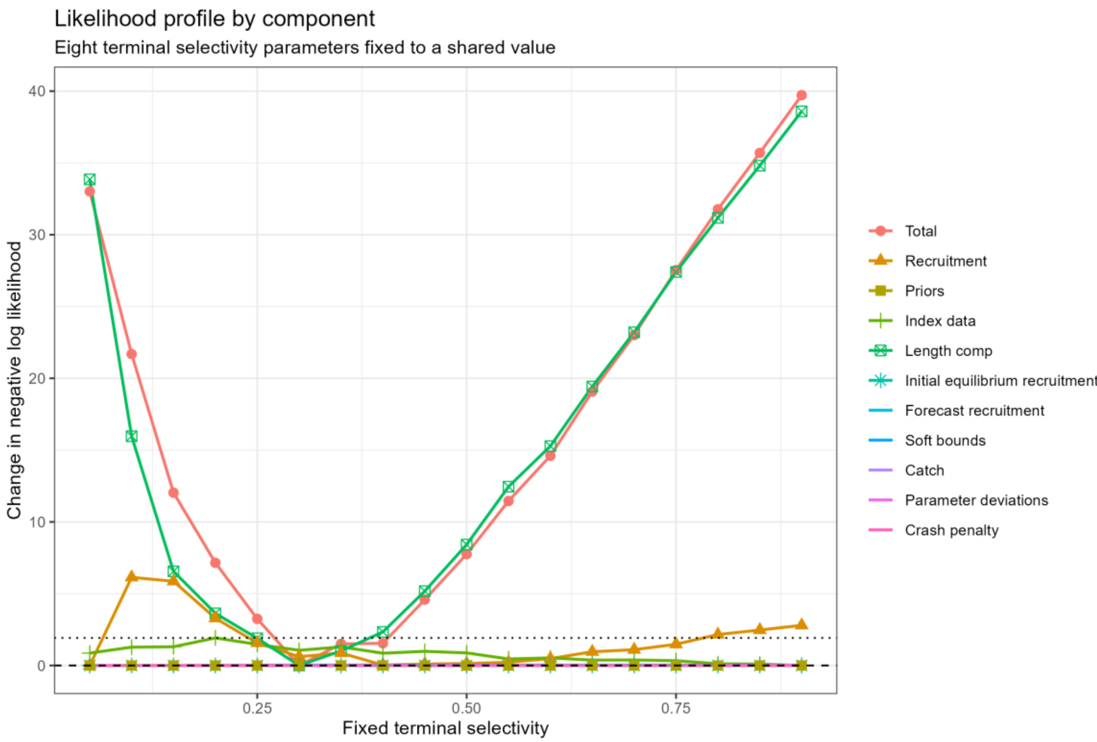


Figure 9: Likelihood profile by component for the terminal selectivity sensitivity applied to northwest fleets 9–15. Terminal selectivity values were fixed across a range of values, and changes in negative log likelihood are shown by likelihood component. Most of the change in model fit is associated with the length-composition likelihood.

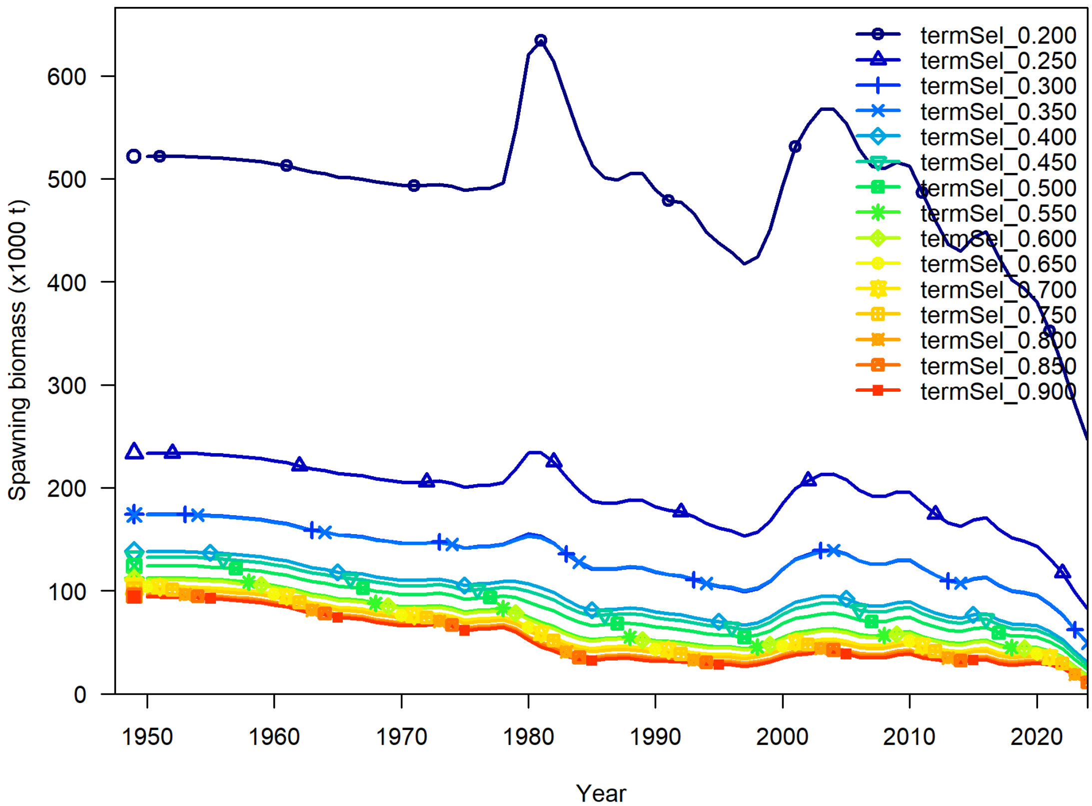


Figure 10: Estimated spawning biomass trajectories for terminal selectivity sensitivity runs applied to northwest fleets 9–15. Lines show spawning biomass in thousands of tonnes for alternative fixed terminal selectivity values.

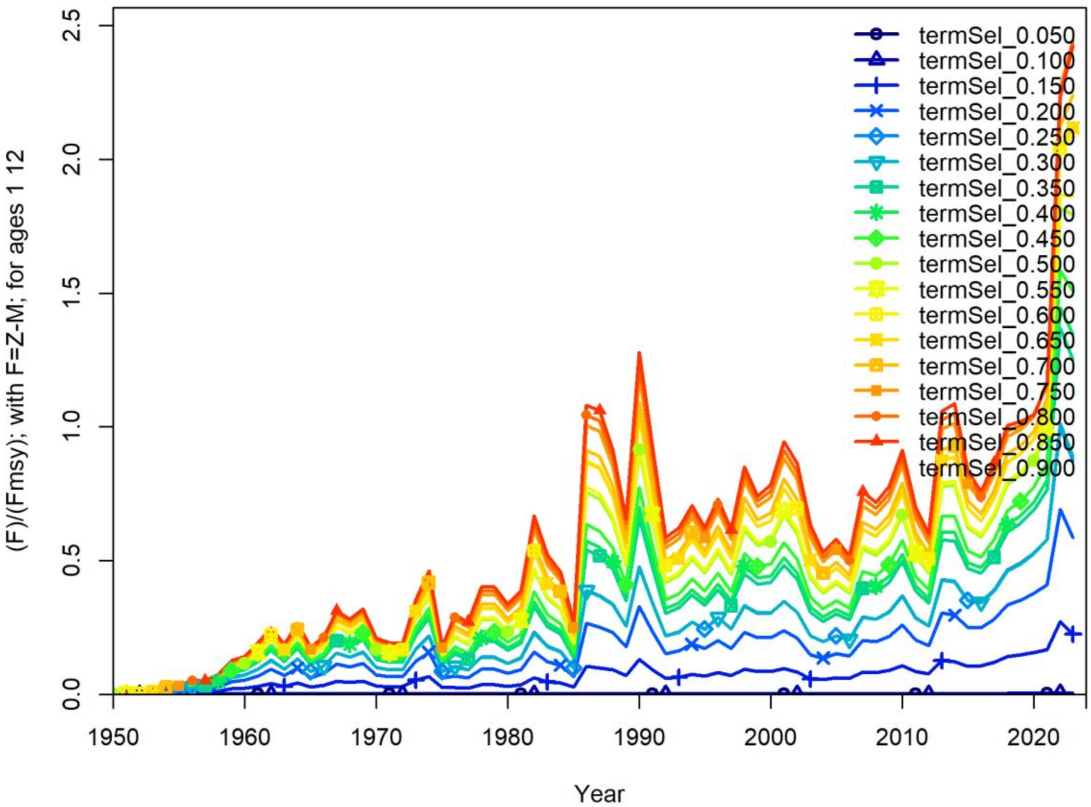


Figure 11: Estimated fishing mortality relative to the reference level for terminal selectivity sensitivity runs applied to northwest fleets 9–15. Lines show annual estimates of F relative to the reference fishing mortality for ages 1–12 under alternative fixed terminal selectivity values.

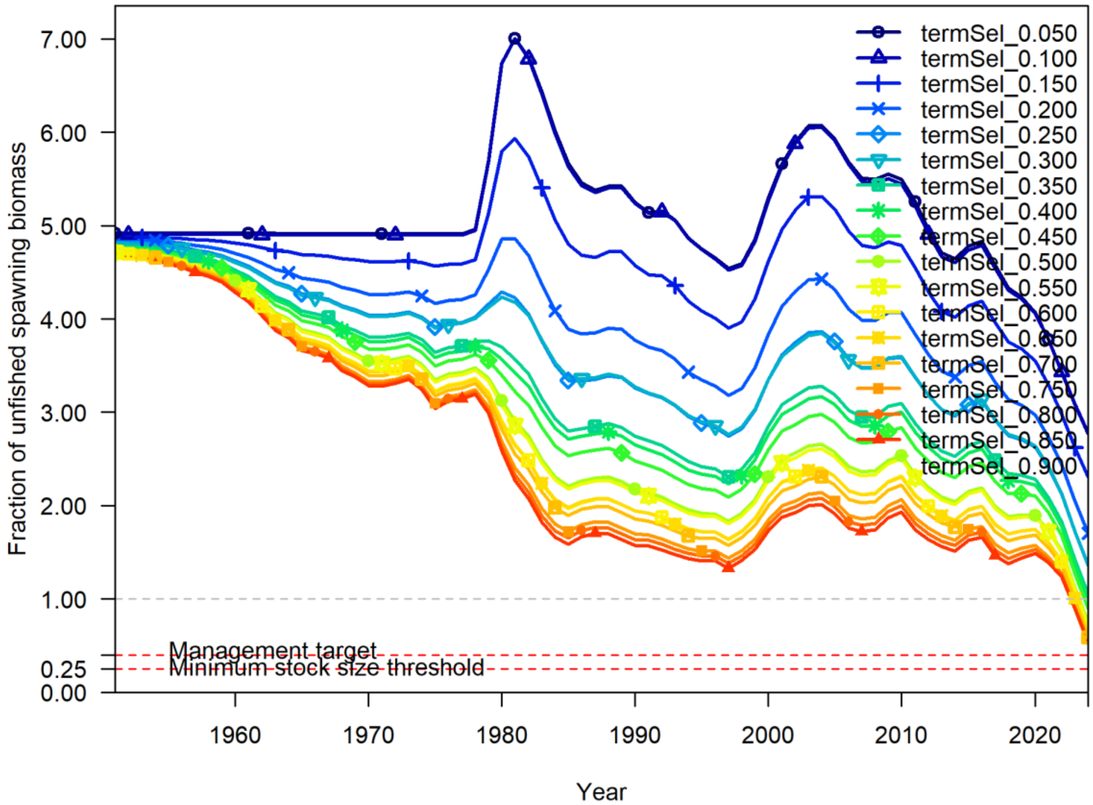


Figure 12: Estimated spawning biomass relative to unfished spawning biomass for terminal selectivity sensitivity runs applied to northwest fleets 9–15. Horizontal dashed lines show the management target and minimum stock size threshold.

12.5 Tables and Figures for the SW diagnostic case

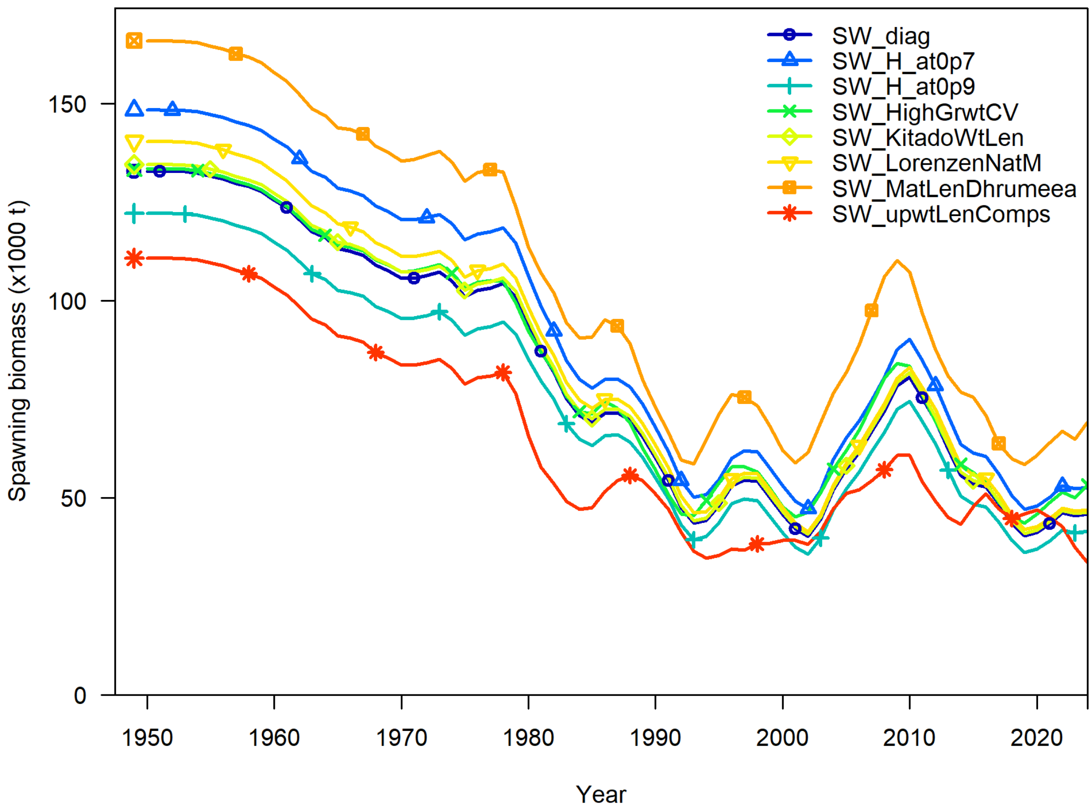


Figure 13: Comparison of estimated spawning biomass trajectories across southwest diagnostic and sensitivity model runs. Spawning biomass is shown in thousands of tonnes. The sensitivity models include alternative stock–recruitment steepness values, growth variation, length–weight relationships, natural mortality, maturity-at-length, and length-composition weighting assumptions.

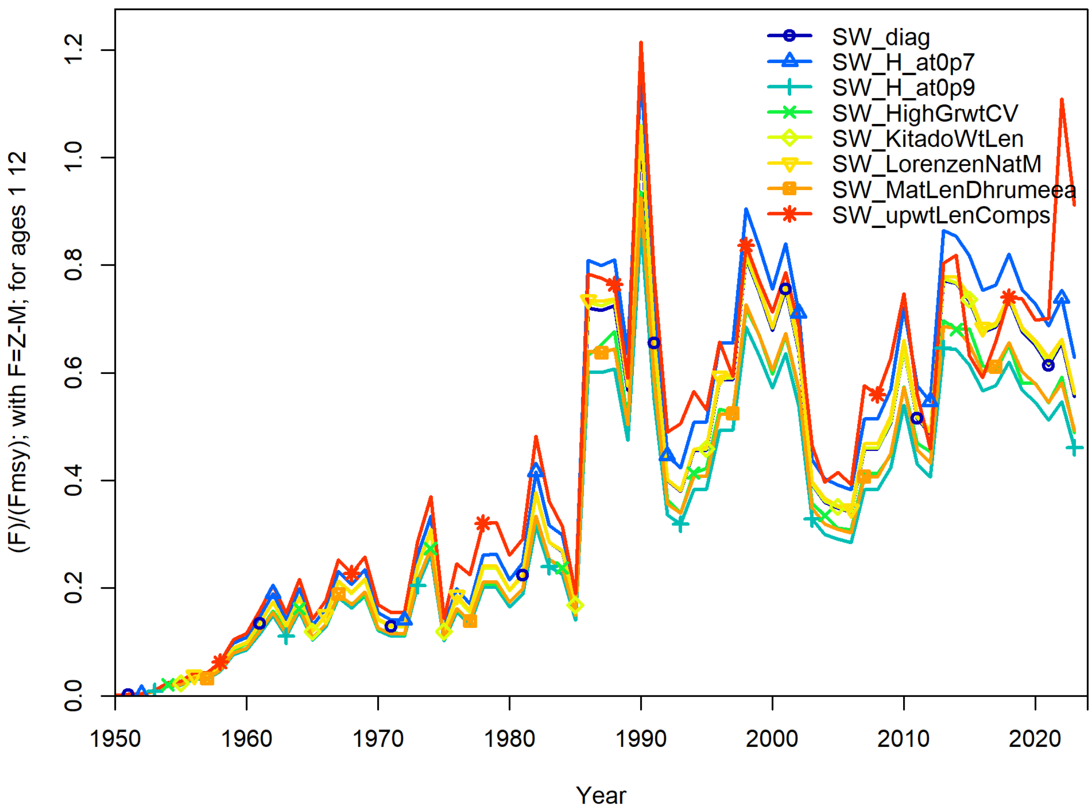


Figure 14: Comparison of fishing mortality relative to the reference level across southwest diagnostic and sensitivity model runs. Lines show annual estimates of F relative to the reference fishing mortality for ages 1–12 under alternative biological assumptions and data-weighting schemes.

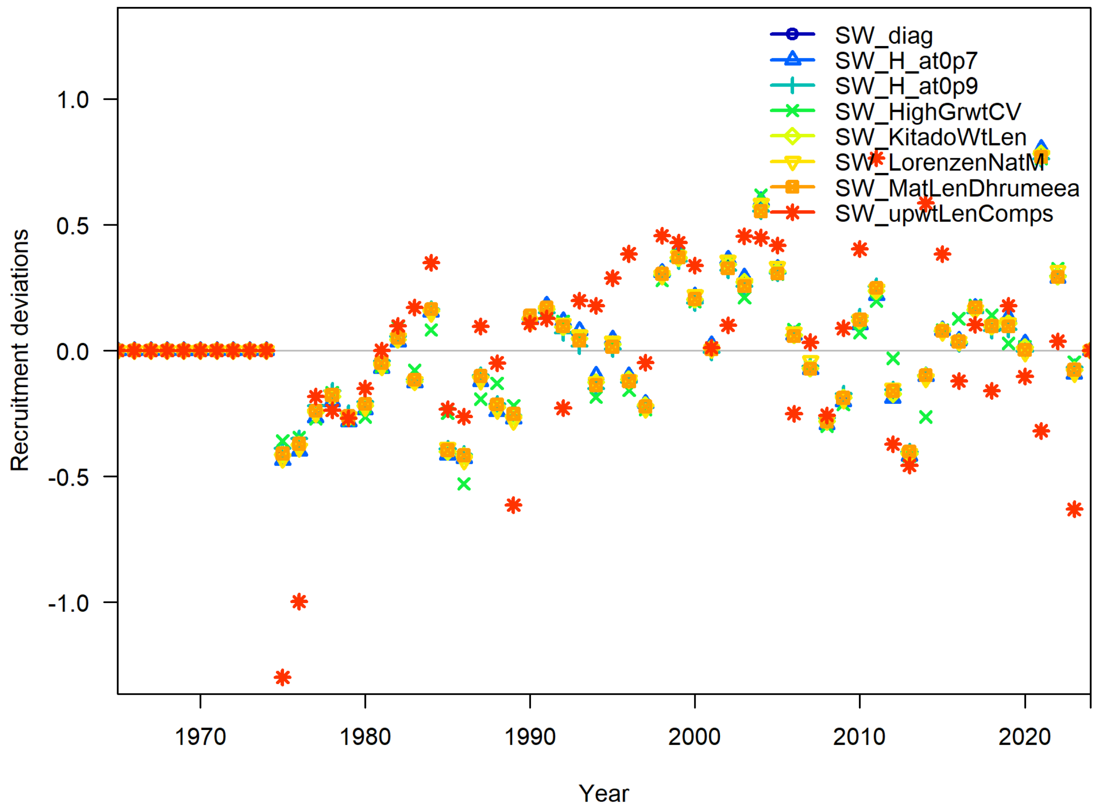


Figure 15: Comparison of estimated recruitment deviations across southwest diagnostic and sensitivity model runs. Points show annual recruitment deviations by model, highlighting differences in recruitment patterns implied by alternative biological assumptions and data-weighting configurations.

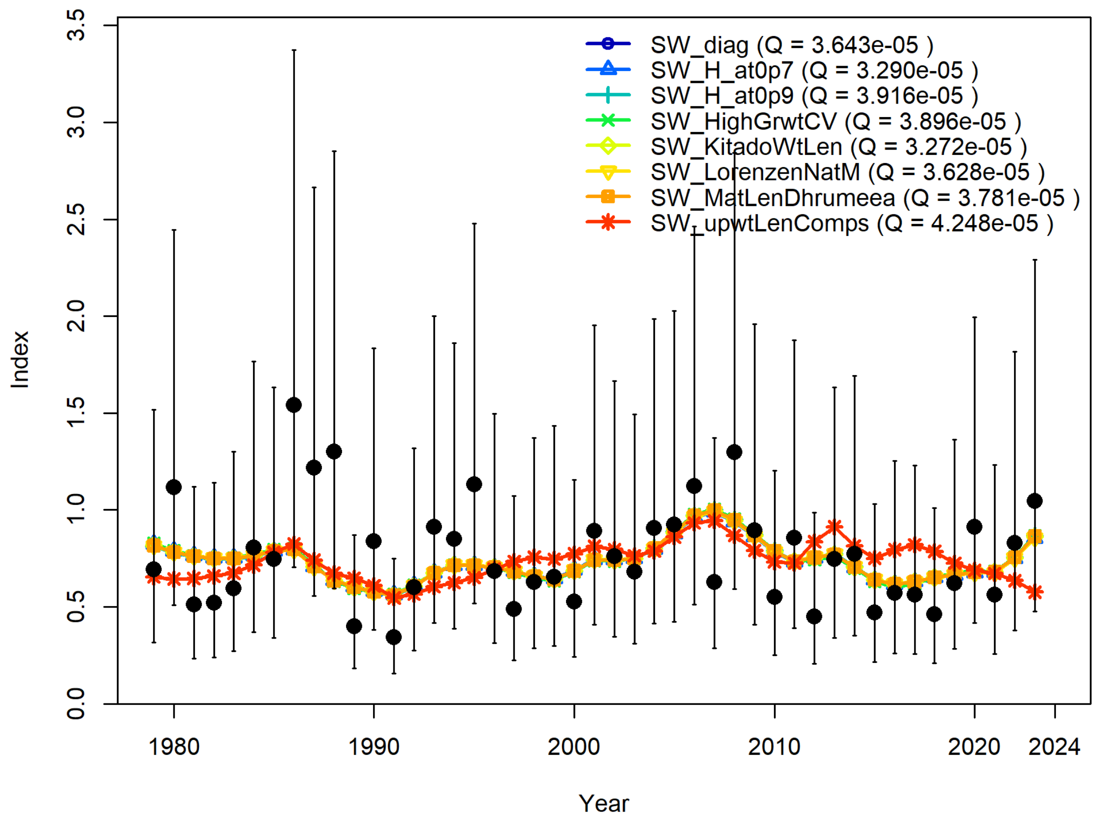


Figure 16: Comparison of model fits to the southwest quarter 1 CPUE index across diagnostic and sensitivity model runs. Observed index values are shown with uncertainty intervals, and colored lines show predicted index trajectories from each model. Estimated catchability values are shown in the legend.

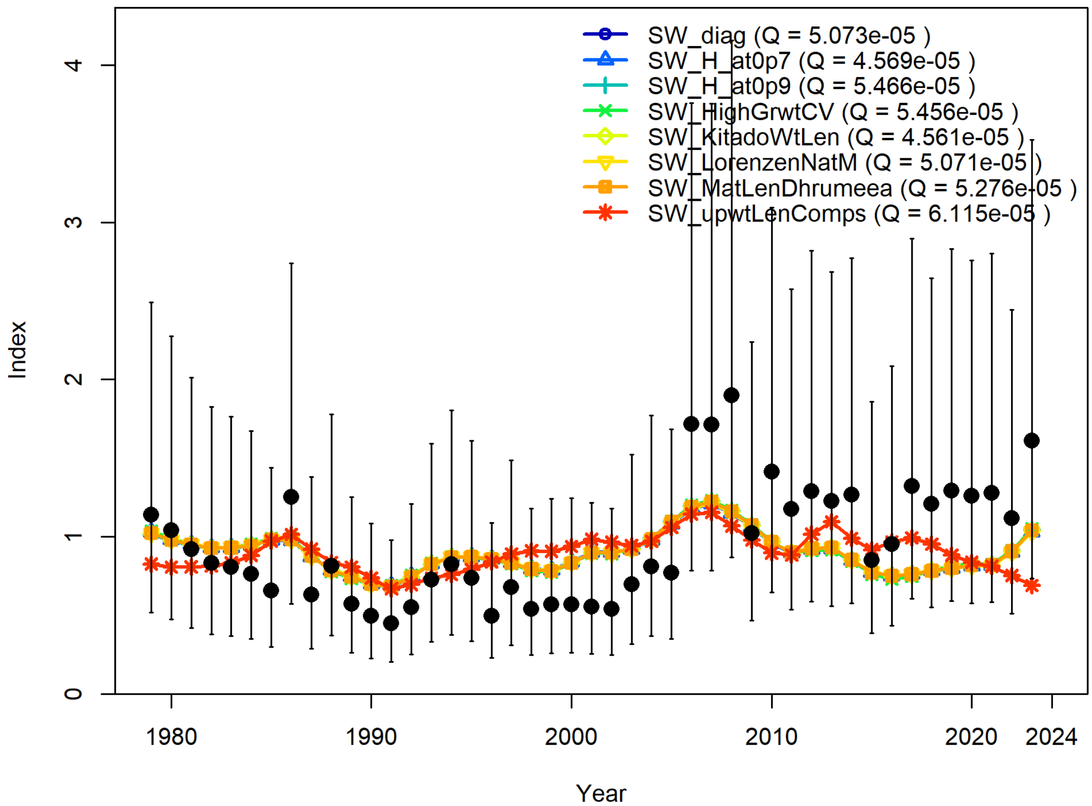


Figure 17: Comparison of model fits to the southwest quarter 2 CPUE index across diagnostic and sensitivity model runs. Observed index values are shown with uncertainty intervals, and colored lines show predicted index trajectories from each model. Estimated catchability values are shown in the legend.

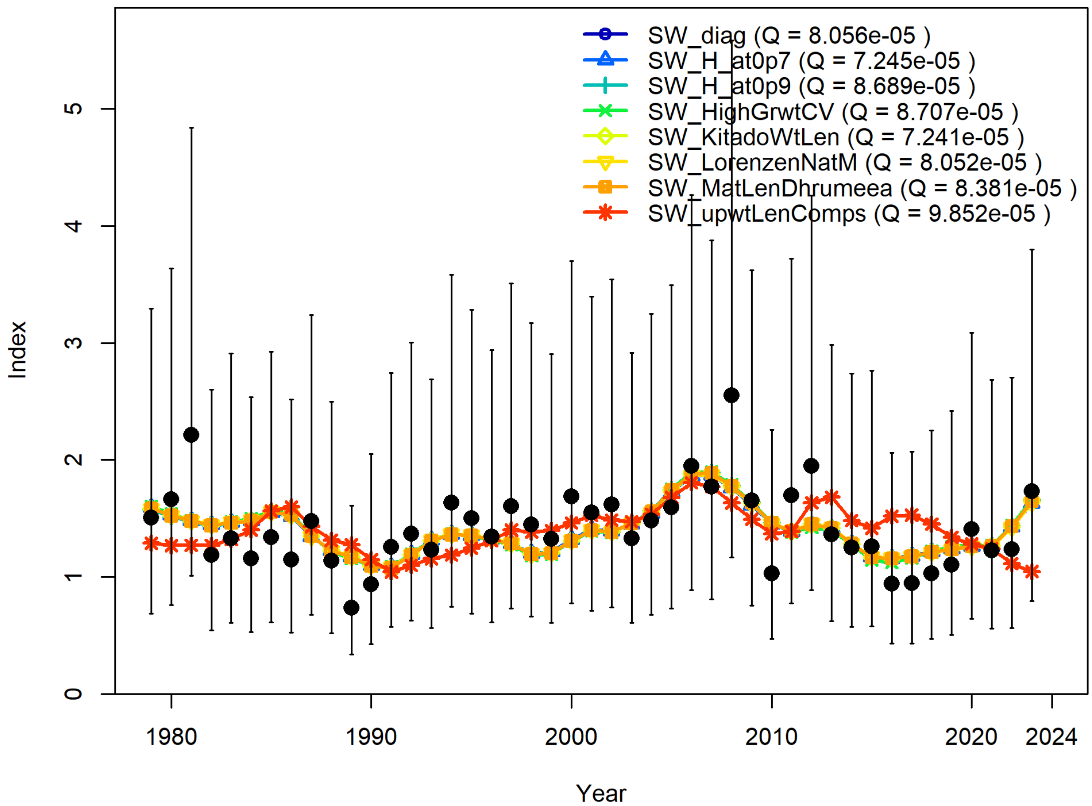


Figure 18: Comparison of model fits to the southwest quarter 3 CPUE index across diagnostic and sensitivity model runs. Observed index values are shown with uncertainty intervals, and colored lines show predicted index trajectories from each model. Estimated catchability values are shown in the legend.

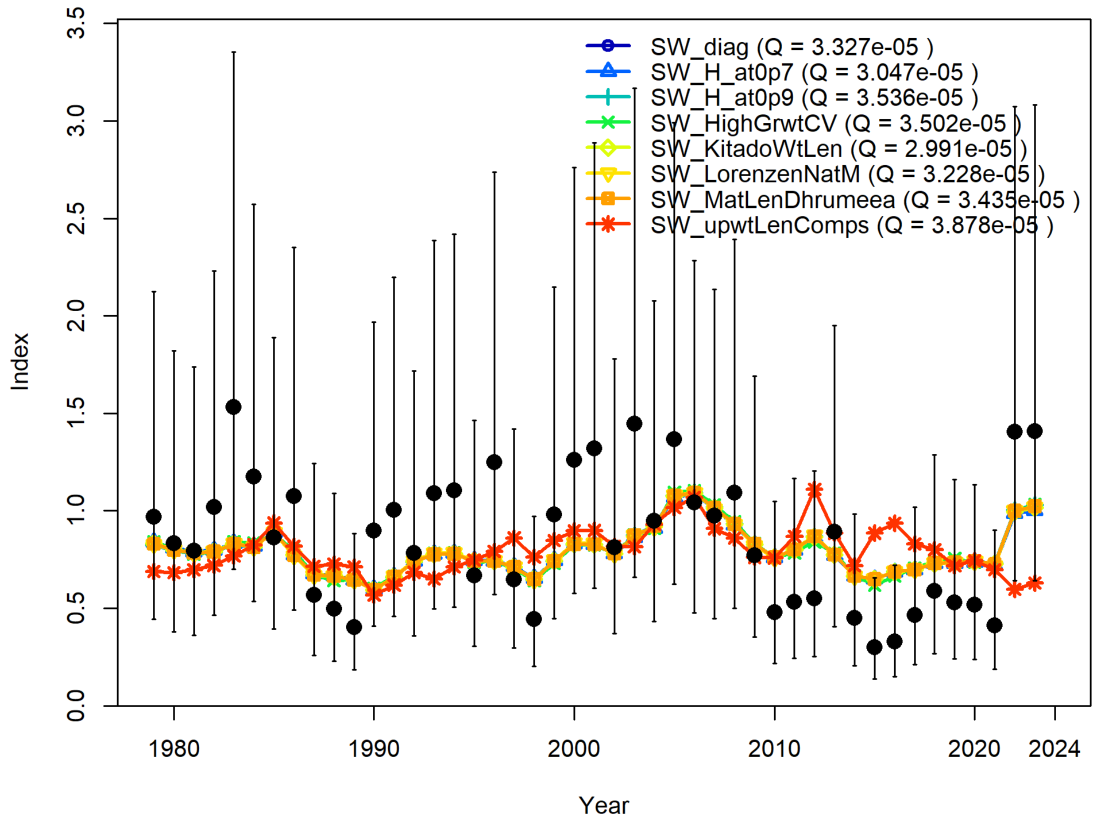


Figure 19: Comparison of model fits to the southwest quarter 4 CPUE index across diagnostic and sensitivity model runs. Observed index values are shown with uncertainty intervals, and colored lines show predicted index trajectories from each model. Estimated catchability values are shown in the legend.

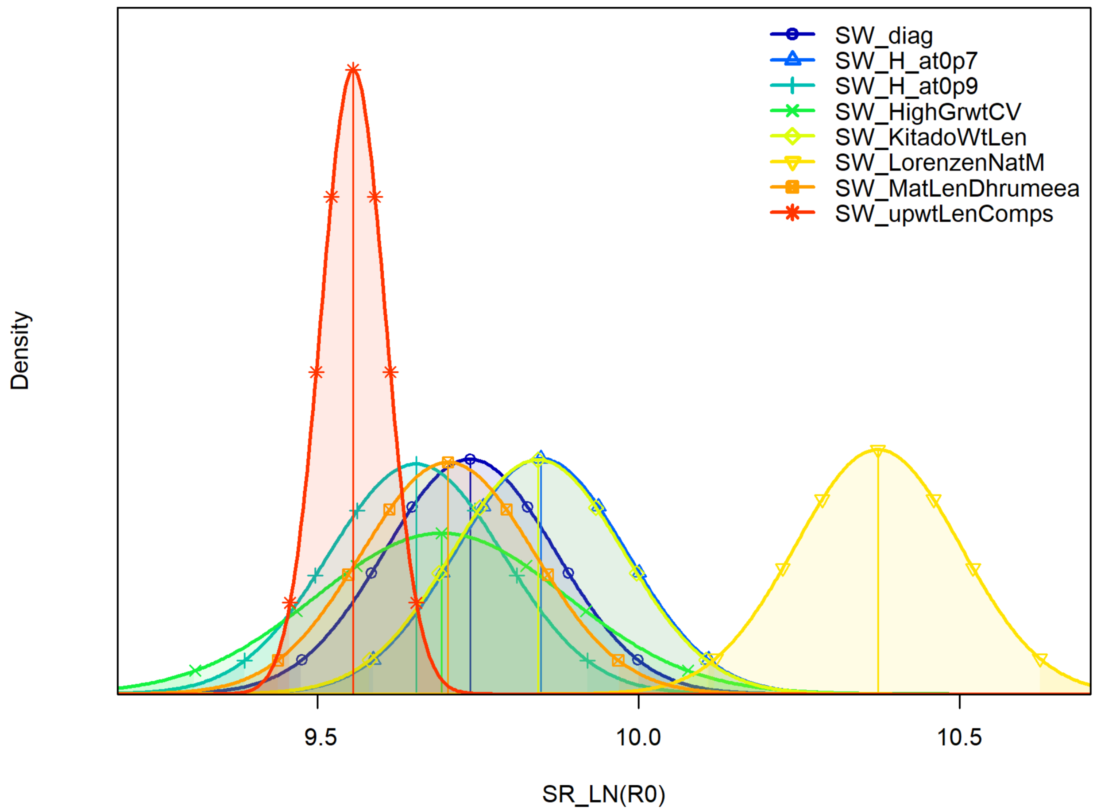


Figure 20: Comparison of approximate distributions for the estimated stock–recruitment scale parameter, SR_LN(R0), across southwest diagnostic and sensitivity model runs. Vertical lines indicate point estimates for each model, and density curves illustrate relative differences in estimated population scale among sensitivity runs.

Table 2: Estimated and derived quantities for the southwest diagnostic model and structural sensitivity analyses. Sensitivity runs include alternative stock–recruitment steepness, growth variability, length–weight relationships, natural mortality, maturity-at-length, and longline length-composition weighting.

Label	SW_diag	SW_H_at0p7	SW_H_at0p9	SW_HighGrwtCV	SW_KitadoWtLen	SW_LorenzenNatM	SW_MatLenDhrumeea	SW_upwtLenComps
TOTAL_like	38.633	39.596	37.907	39.284	38.686	40.337	38.341	957.157
Survey_like	-119.865	-119.681	-119.992	-120.476	-119.860	-119.866	-119.926	-98.407
Length_comp_like	113.728	114.169	113.441	128.845	113.781	115.294	113.624	1005.040

Parm_priors_like	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Size_at_age_like	59.241	59.244	59.239	45.364	59.242	59.244	59.240	59.244
SSB_unfished_thousand_mt	132.852	148.355	122.138	133.440	134.554	140.363	165.952	110.750
B_MSY/SSB_unfished	0.218	0.264	0.160	0.221	0.220	0.212	0.237	0.216
Totbio_unfished	571078.000	637716.000	525022.000	559733.000	569813.000	590832.000	551807.000	476069.000
SmryBio_unfished	570687.000	637280.000	524662.000	559354.000	569515.000	590095.000	551430.000	475743.000
Recr_unfished_millions	16.941	18.918	15.575	16.204	18.835	31.977	16.369	14.123
SSB_MSY_thousand_mt	28.937	39.186	19.564	29.556	29.613	29.779	39.391	23.973
SPR_MSY	0.267	0.343	0.184	0.270	0.269	0.261	0.285	0.265
annF_MSY	0.149	0.122	0.190	0.174	0.148	0.143	0.174	0.153
Dead_Catch_MSY	47974.500	45826.000	51602.800	50541.000	47867.300	47311.800	50143.000	39695.300
Ret_Catch_MSY	47974.500	45826.000	51602.800	50541.000	47867.300	47311.800	50143.000	39695.300
B_MSY/SSB_unfished	0.218	0.264	0.160	0.221	0.220	0.212	0.237	0.216
Bratio_2023	1.571	1.334	2.099	1.690	1.554	1.565	1.646	1.560
F_2023	0.556	0.629	0.461	0.489	0.561	0.563	0.492	0.912

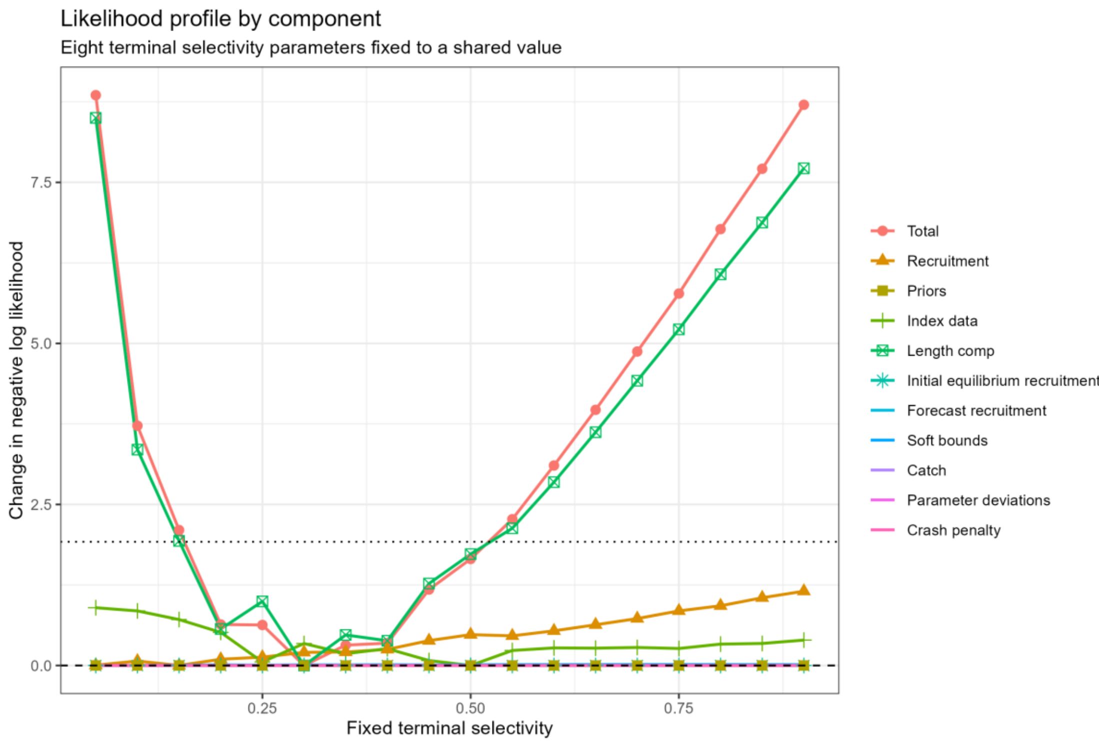


Figure 21: Likelihood profile by component for the terminal selectivity sensitivity applied to the southwest diagnostic case. Terminal selectivity values were fixed across a range of values, and changes in negative log likelihood are shown by likelihood component. Most of the change in model fit is associated with the length-composition likelihood.

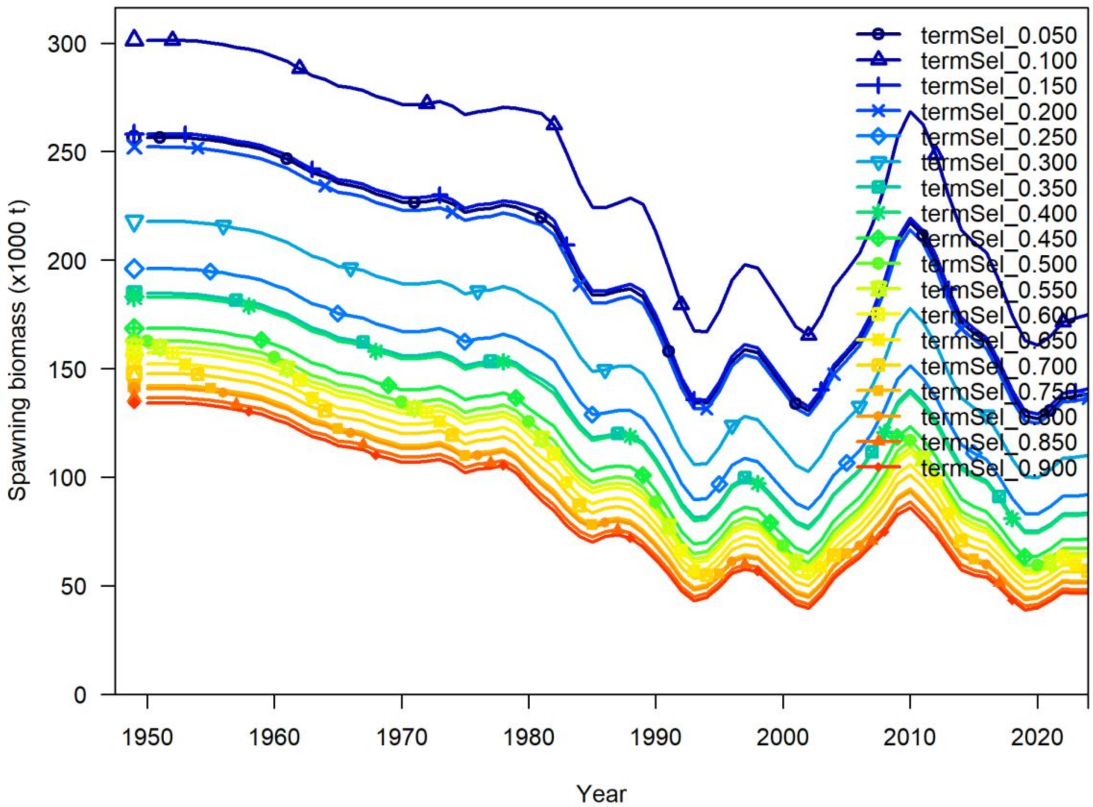


Figure 22: Estimated spawning biomass trajectories for terminal selectivity sensitivity runs applied to the southwest diagnostic case. Lines show spawning biomass in thousands of tonnes for alternative fixed terminal selectivity values.

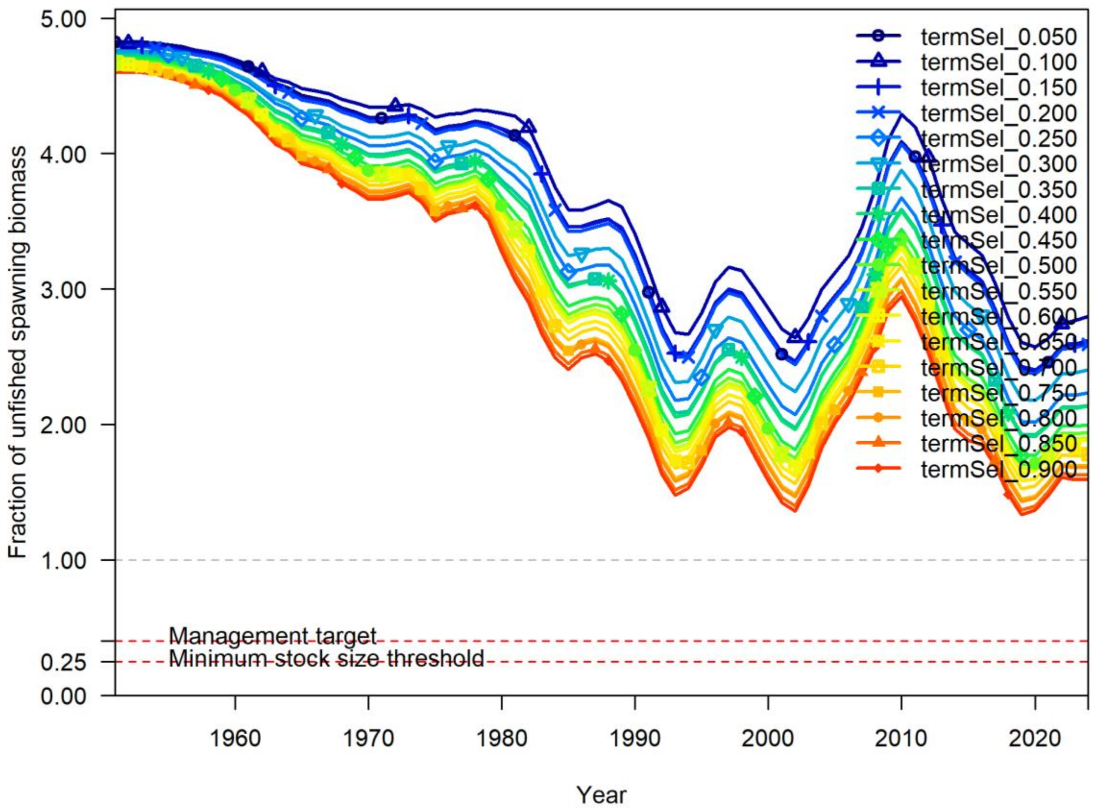


Figure 23: Estimated spawning biomass relative to unfished spawning biomass for terminal selectivity sensitivity runs applied to the southwest diagnostic case. Horizontal dashed lines show the management target and minimum stock size threshold.

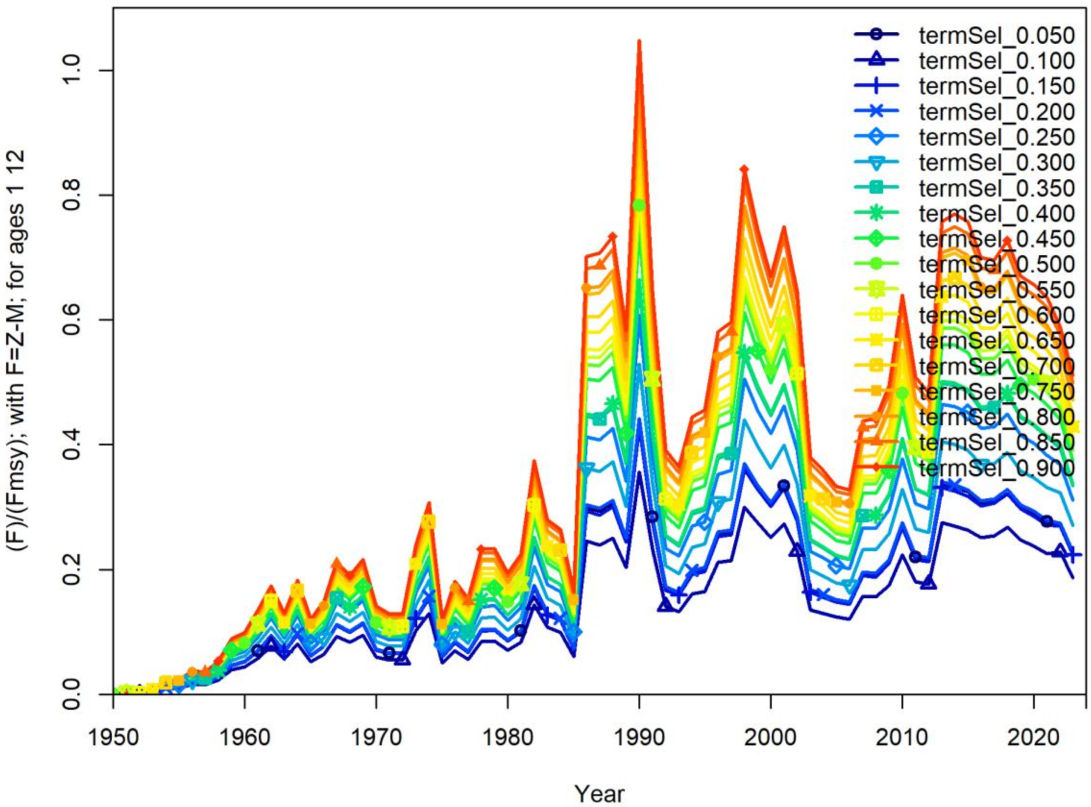


Figure 24: Estimated fishing mortality relative to the reference level for terminal selectivity sensitivity runs applied to the southwest diagnostic case. Lines show annual estimates of F relative to the reference fishing mortality for ages 1–12 under alternative fixed terminal selectivity values.

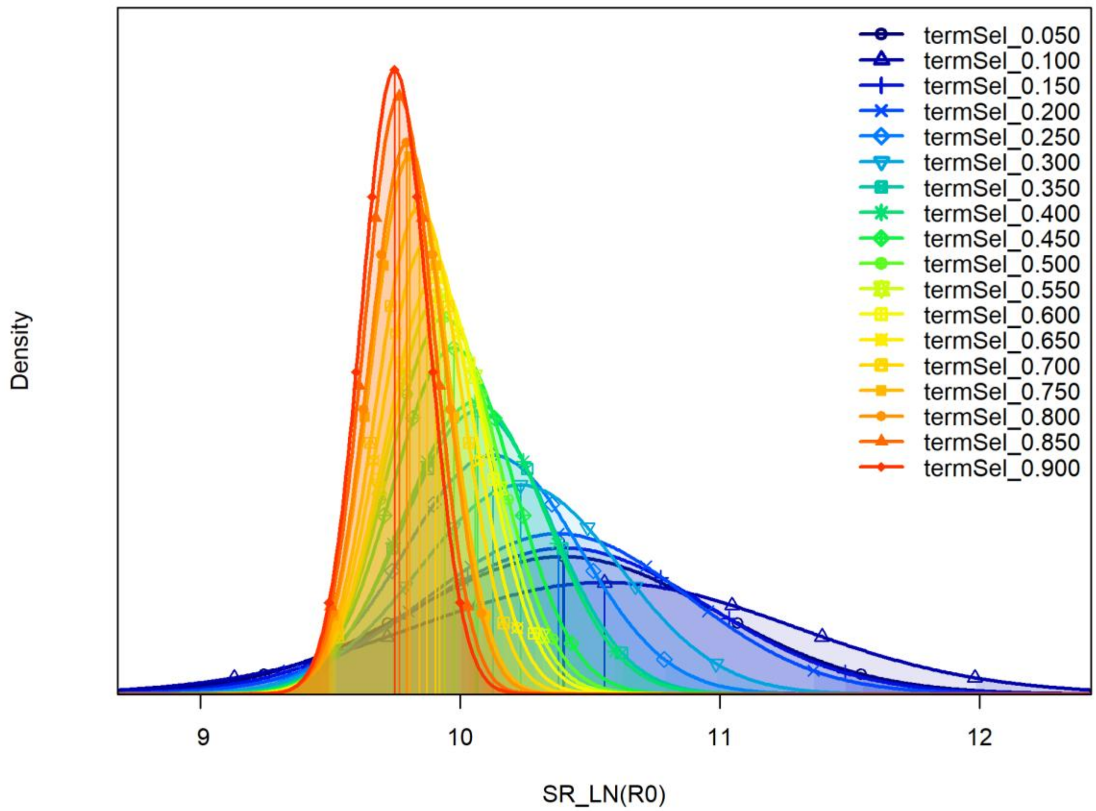


Figure 25: Comparison of approximate distributions for the estimated stock–recruitment scale parameter, SR_LN(R0), across terminal selectivity sensitivity runs in the southwest diagnostic case. Vertical lines indicate point estimates for each fixed terminal selectivity value, and density curves illustrate the effect of terminal selectivity on estimated population scale.

Evaluation of time-varying selectivity 2026 Indian Ocean Albacore Stock Assessment

Joel rice

2026-07-16

13 Annex 4. Evaluation of time-varying selectivity

13.1 Introduction

Time-varying selectivity was evaluated to determine whether persistent temporal patterns in the length-composition residuals could be attributed to changes in the size-selection characteristics of the longline fisheries. The diagnostic plots indicated that the assumption of constant selectivity across the full assessment period was potentially restrictive for several fleets. Separate analyses were therefore conducted for the northwest and southwest diagnostic models, with the fleets selected for time variation based on the corresponding mean-length residual runs tests.

For the northwest diagnostic model, recent time blocks were applied to the selectivity parameters for fleets 1–4, representing the four seasonal LL1 fisheries. Three alternative formulations were examined: a time block on the selectivity peak, a time block on the ascending-width parameter, and time blocks on both parameters. For the southwest diagnostic model, the same three formulations were applied to fleets 4–8, which exhibited the clearest evidence of non-random temporal patterns in the mean-length residuals. In each case, the time-varying parameters were estimated independently by fleet.
The alternative formulations were evaluated using total and component negative log-likelihoods, Akaike information criterion (AIC), nested likelihood-ratio tests, mean-length residual runs tests, fits to the abundance indices, estimates of population scale, and spawning-biomass trajectories. Particular attention was given to whether improvements in fit were sufficient to justify the additional parameters and whether the alternative formulations materially affected estimated biomass and stock status.

13.2 Northwest diagnostic model

13.2.1 Model comparison

The northwest time-varying selectivity analyses provide strong evidence that allowing recent selectivity to differ from the historical configuration improves the model. All three time-varying formulations had substantially lower AIC values than the static-selectivity diagnostic model. The combined peak-and-ascending-width formulation had the lowest AIC, although the differences among the three time-varying formulations were less than one AIC unit and therefore provide little basis for selecting among them using AIC alone.

The likelihood improvements were driven primarily by improved fits to the abundance indices. The length-composition likelihood became slightly poorer in the time-varying models, indicating that changes in selectivity affected the broader balance among the CPUE and size-composition data rather than simply improving the length-composition likelihood. The resulting spawning-biomass trajectories were lower than those from the diagnostic model, particularly for the combined formulation, demonstrating that the treatment of selectivity has an important influence on estimated recent biomass and stock status.

Table 1: Estimated and derived quantities for the northwest diagnostic model and alternative time-varying selectivity configurations. Model 1 is the static-selectivity diagnostic model, Model 2 includes a recent time block on the selectivity peak, Model 3 includes a recent time block on the ascending-width parameter, and Model 4 includes recent time blocks on both parameters.

Quantity	NW diagnostic	TV peak	TV ascending width	TV peak and ascending width
TOTAL_like	163.2	129.3	129.7	125.2
Survey_like	9.08	-26.62	-26.84	-31.55
Length_comp_like	108.5	111.2	111.5	112
Parm_priors_like	0	0	0	0
Size_at_age_like	59.24	59.24	59.24	59.24
SSB_unfished_thousand_mt	116.6	113	113.4	108.4
B_MSY/SSB_unfished	0.2118	0.213	0.2129	0.2192
Totbio_unfished	501174	485796	487548	466030
SmryBio_unfished	500831	485464	487215	465711
Recr_unfished_millions	14.87	14.41	14.46	13.82
SSB_MSY_thousand_mt	24.7	24.07	24.15	23.76
SPR_MSY	0.2611	0.2622	0.2621	0.268
annF_MSY	0.1557	0.1528	0.1544	0.1488
Dead_Catch_MSY	41079	40141	40158	39274
Ret_Catch_MSY	41079	40141	40158	39274

B_MSY/SSB_unfished	0.2118	0.213	0.2129	0.2192
Bratio_2023	1.016	0.760 1	0.8112	0.6552
F_2023	1.264	1.504	1.492	1.639

Table 2: Akaike information criterion comparison for the northwest diagnostic model and alternative time-varying selectivity configurations. Delta AIC is calculated relative to the model with the lowest AIC.

Model	AIC	Delta AIC
NW diagnostic	470.300	59.848
TV peak	410.594	0.142
TV ascending width	411.432	0.980
TV peak and ascending width	410.452	0.000

Table 3: Likelihood-ratio tests comparing nested northwest selectivity models. The peak-only and ascending-width-only models are not directly compared because neither model is nested within the other.

Reduced model	Full model	NLL reduced	NLL full	df	LR statistic	p-value
NW diagnostic	TV peak	163.150	129.297	4	67.706	0.0000
NW diagnostic	TV ascending width	163.150	129.716	4	66.868	0.0000
TV peak	TV peak and ascending width	129.297	125.226	4	8.142	0.0865
TV ascending width	TV peak and ascending width	129.716	125.226	4	8.980	0.0616

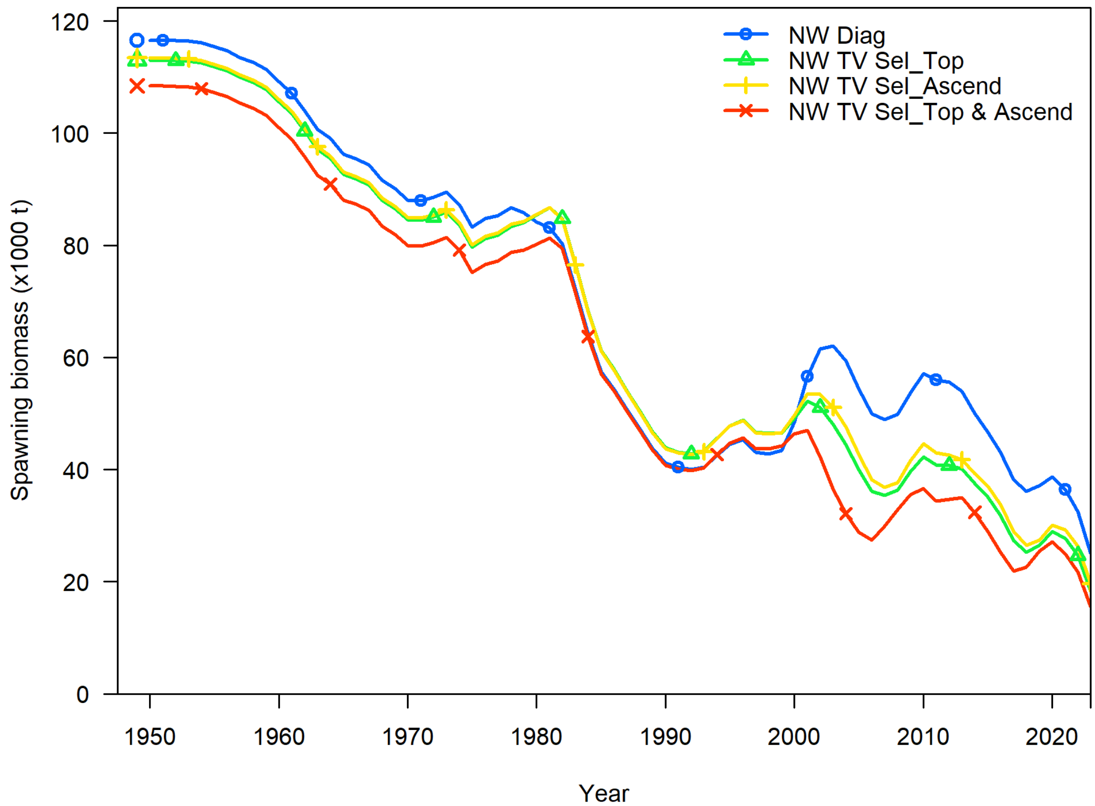


Figure 1: Estimated spawning-biomass trajectories for the northwest diagnostic model and alternative time-varying selectivity configurations. The alternative models include recent time blocks on the selectivity peak, ascending width, or both parameters for fleets 1–4.

The single-parameter time-varying models produced similar estimates of population scale and spawning biomass. Both resulted in lower recent spawning biomass than the static diagnostic model. The combined model produced a further reduction in population scale and recent biomass, with a correspondingly more pessimistic estimate of stock status. Because the AIC differences among the three time-varying formulations were small, these differences in biological output should be considered when selecting the reference configuration.

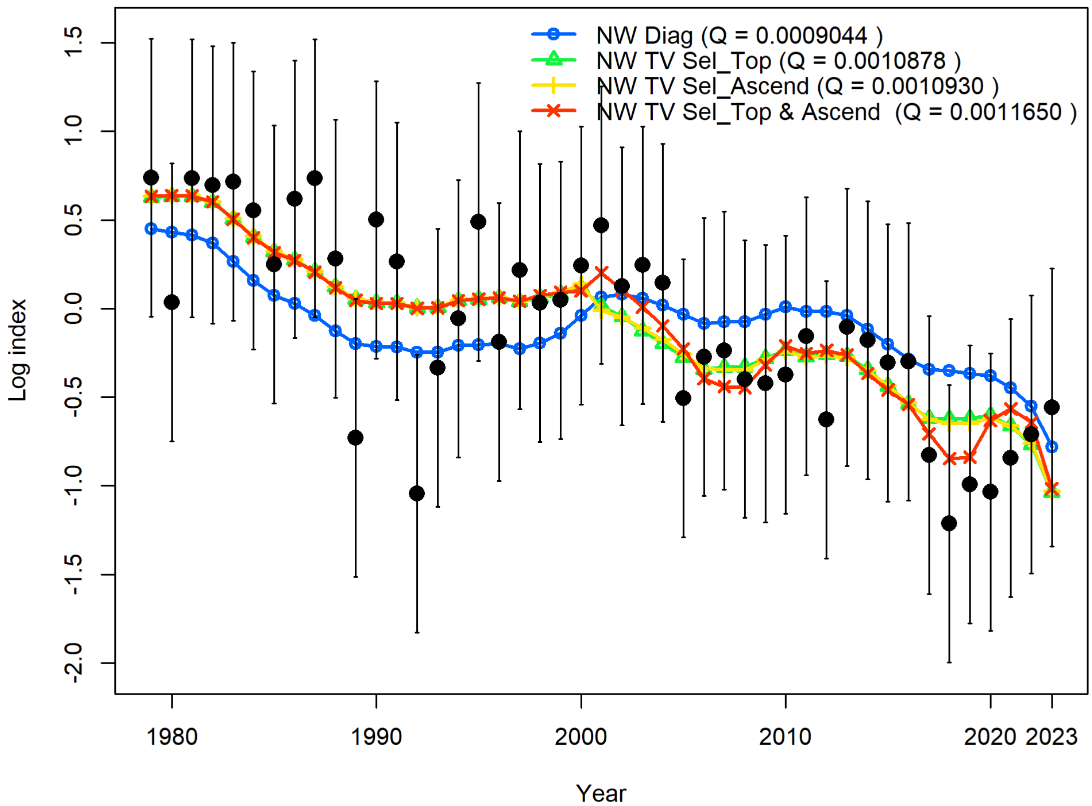


Figure 2: Observed and predicted northwest CPUE index for quarter 1 under the diagnostic model and alternative time-varying selectivity configurations. The reported Q values are the estimated catchability coefficients for each model.

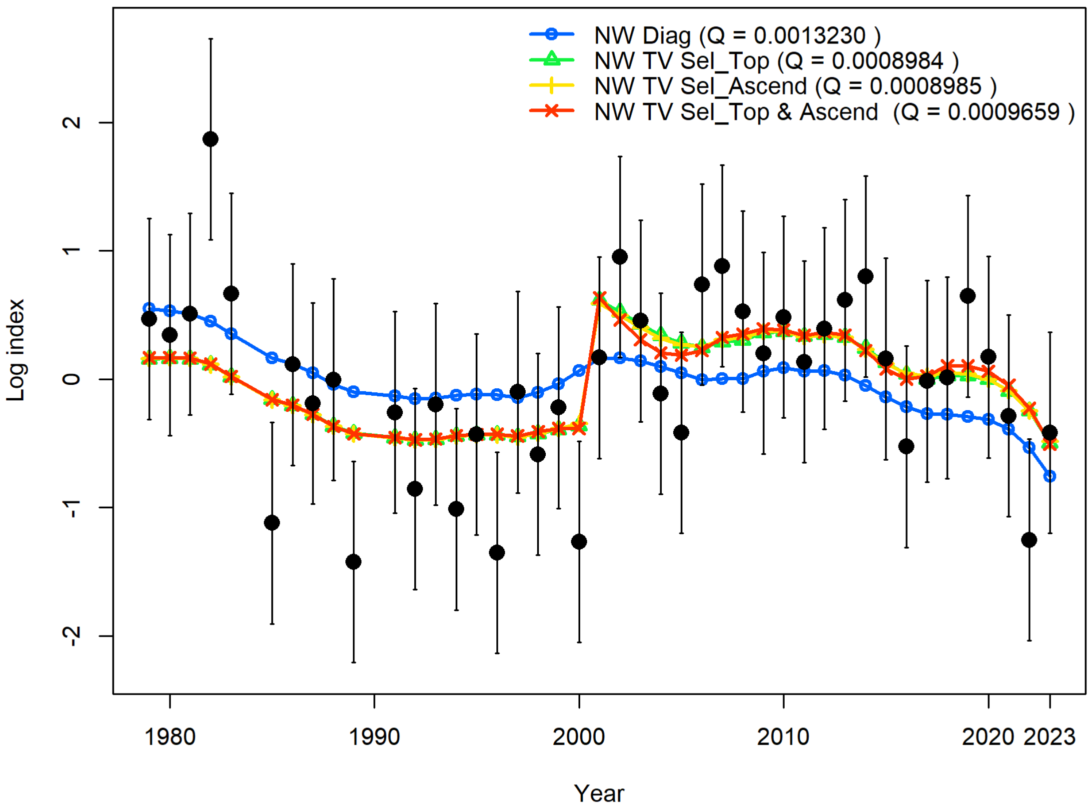


Figure 3: Observed and predicted northwest CPUE index for quarter 2 under the diagnostic model and alternative time-varying selectivity configurations. The reported Q values are the estimated catchability coefficients for each model.

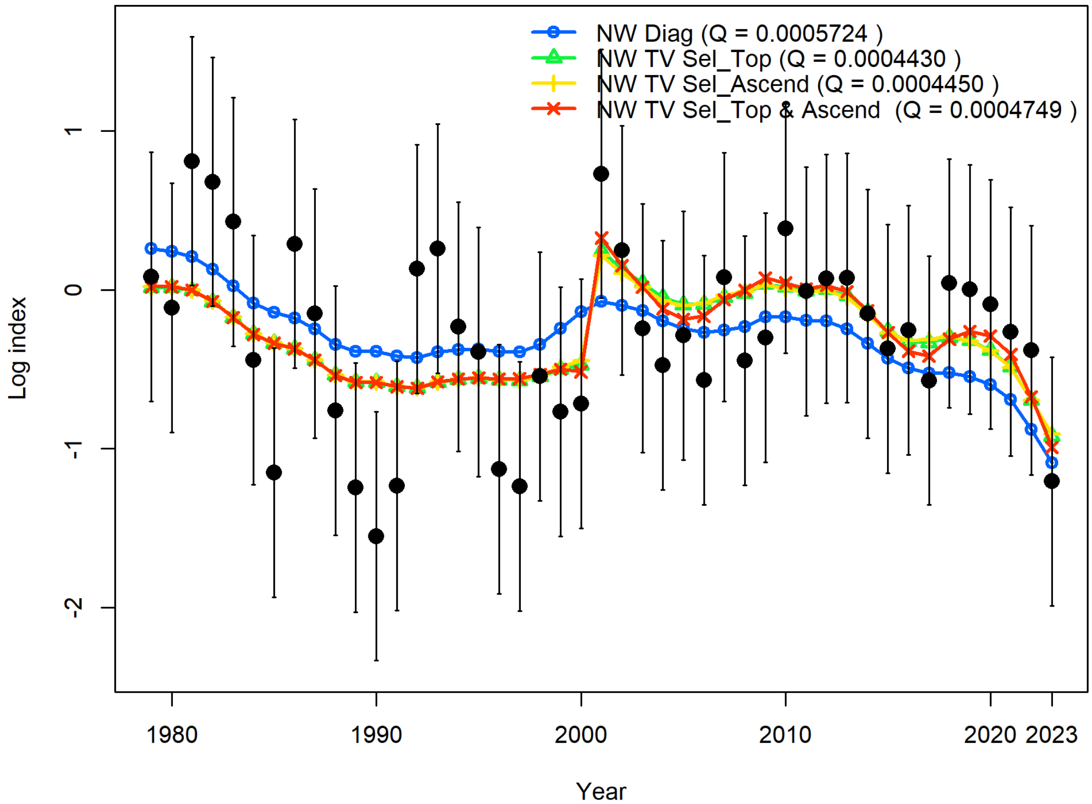


Figure 4: Observed and predicted northwest CPUE index for quarter 3 under the diagnostic model and alternative time-varying selectivity configurations. The reported Q values are the estimated catchability coefficients for each model.

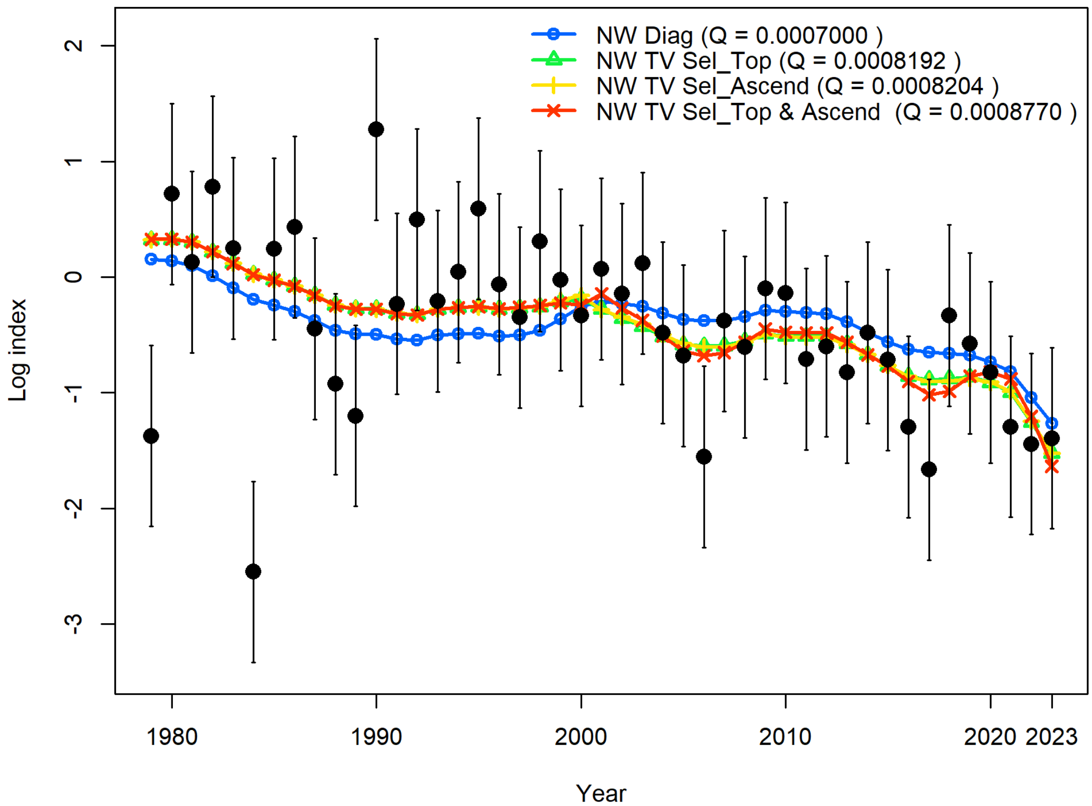


Figure 5: Observed and predicted northwest CPUE index for quarter 4 under the diagnostic model and alternative time-varying selectivity configurations. The reported Q values are the estimated catchability coefficients for each model.

The time-varying selectivity formulations produced similar fitted trends for the four quarterly indices and generally shifted the predicted values relative to the static diagnostic model. The combined model provided the greatest improvement in the survey likelihood and removed the statistically significant runs-test result for one additional LL1 fleet. However, persistent residual structure remained for some fleets, indicating that selectivity changes alone do not explain all temporal variation in the data.

13.3 Mean-length residual runs tests

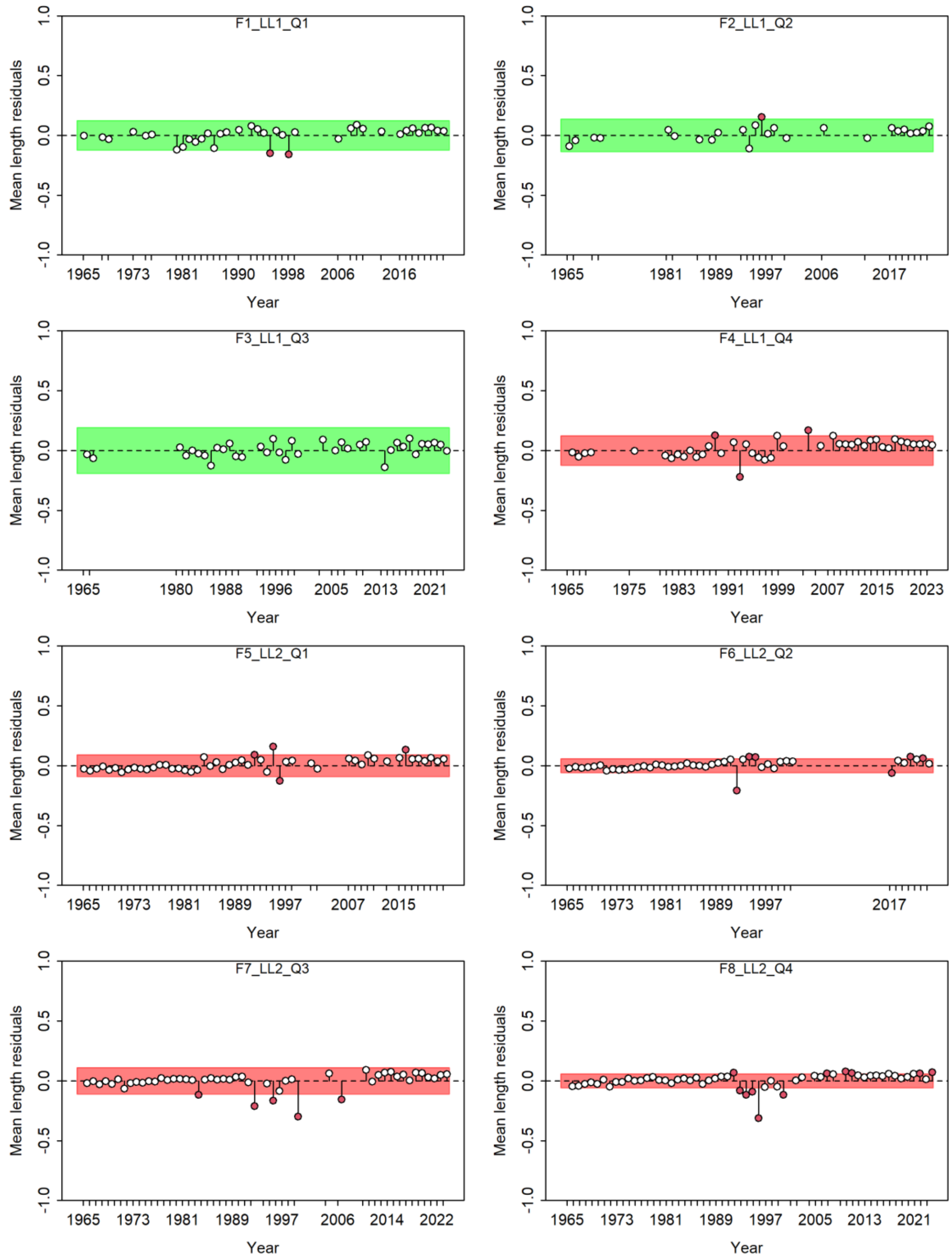


Figure 6: Runs tests applied to mean-length residuals for fleets 1–8 in the northwest diagnostic model with constant selectivity. Green backgrounds indicate that the null hypothesis of randomly ordered residuals was not rejected, whereas red backgrounds indicate significant temporal structure.

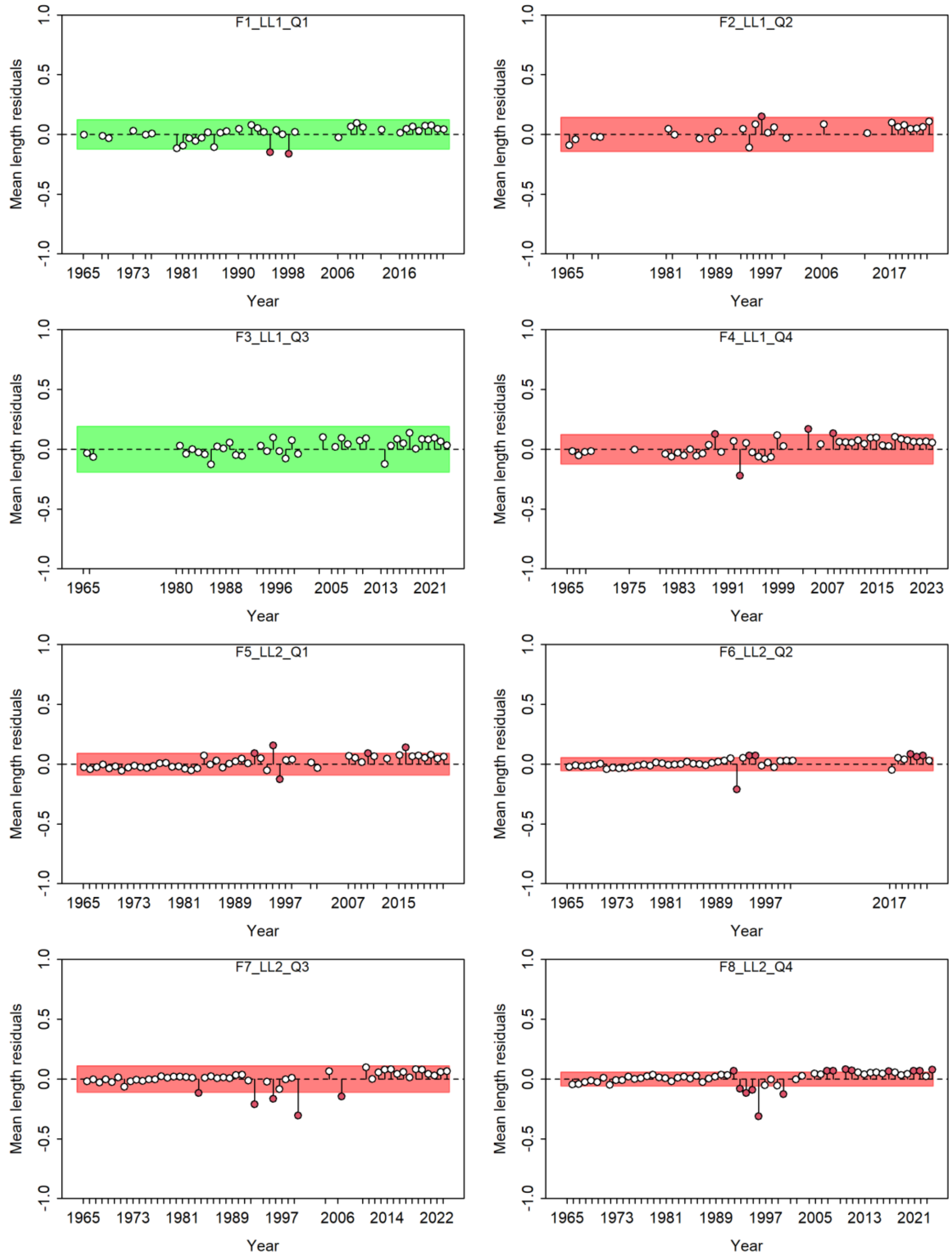


Figure 7: Runs tests applied to mean-length residuals for fleets 1–8 in the northwest model with recent time blocks on the selectivity peak for fleets 1–4.

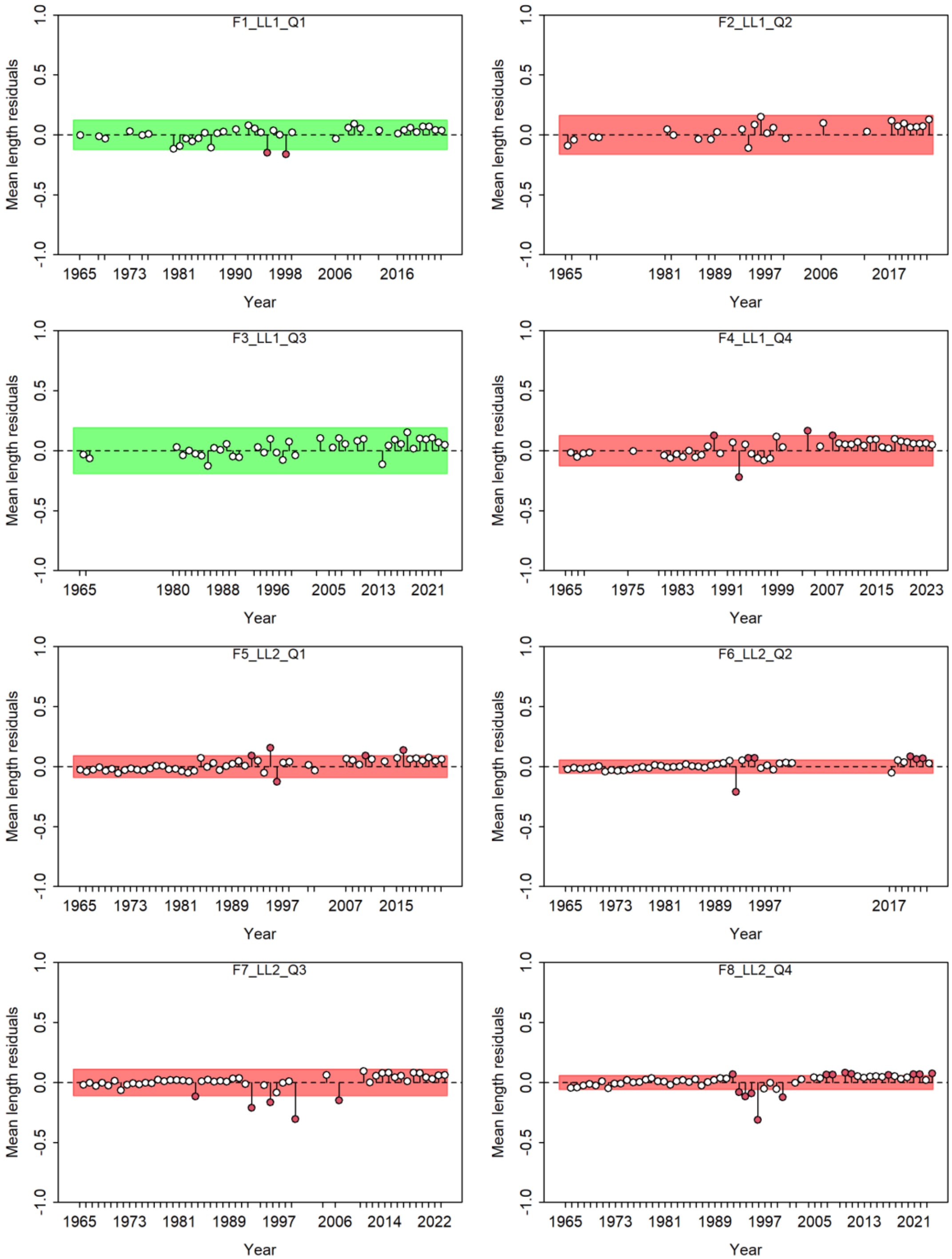


Figure 8: Runs tests applied to mean-length residuals for fleets 1–8 in the northwest model with recent time blocks on the ascending-width selectivity parameter for fleets 1–4.

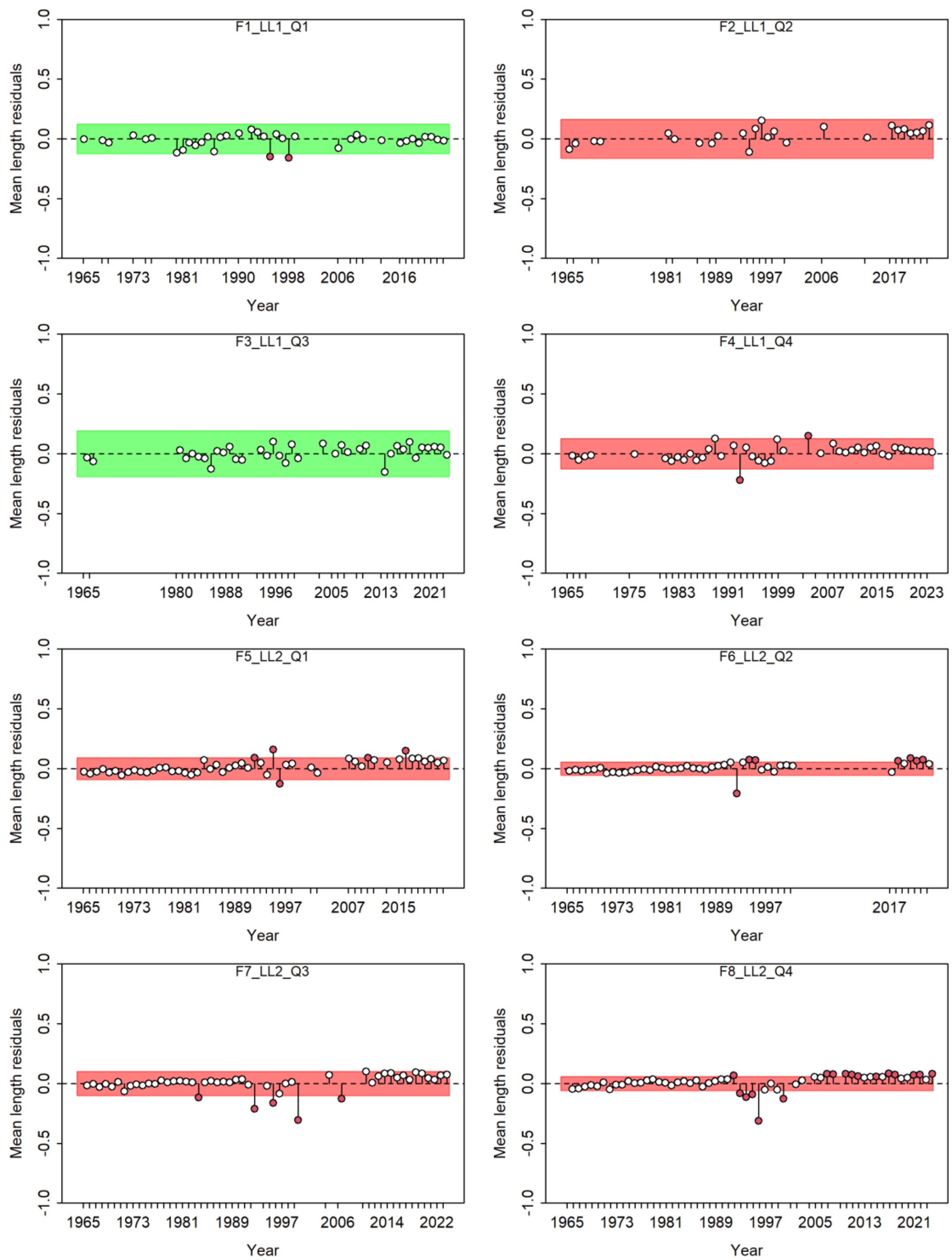


Figure 9: Runs tests applied to mean-length residuals for fleets 1–8 in the northwest model with recent time blocks on both the selectivity peak and ascending-width parameters for fleets 1–4.

Overall, the northwest analyses strongly support allowing some form of recent time variation in LL1 selectivity. The peak-only and ascending-width-only formulations provide nearly equivalent improvements and produce similar biological results. The combined model has the lowest AIC and improves the runs-test result for one additional fleet, but it also produces lower recent biomass and higher fishing pressure. The combined formulation should therefore be evaluated as an important structural sensitivity rather than selected solely on the basis of its marginally lower AIC.

14 Southwest diagnostic model

14.1 Model comparison

For the southwest model, time-varying selectivity was applied to fleets 4–8 because these fleets exhibited the clearest problems in the mean-length residual runs tests. The changes produced modest improvements in the length-composition likelihood and resolved the runs-test result for several individual fleets. However, the improvements were small relative to the five additional parameters introduced by each single-parameter formulation.

None of the likelihood-ratio tests supported the additional time-varying parameters at the 5% significance level. The combined model provided virtually no improvement over the ascending-width-only model, indicating that the data do not support estimating both forms of time variation simultaneously.

Table 4: Estimated and derived quantities for the southwest diagnostic model and alternative time-varying selectivity configurations. Time variation was applied to fleets 4–8, which exhibited the clearest temporal structure in their mean-length residuals.

Quantity	SW diagnostic	TV peak	TV ascending width	TV peak and ascending width
TOTAL_like	38.63	37.86	35.25	35.21
Survey_like	-119.9	-119.9	-120.1	-120.1
Length_comp_like	113.7	113	110.5	110.4
Parm_priors_like	0	0	0	0
Size_at_age_like	59.24	59.26	59.26	59.26
SSB_unfished_thousand_mt	132.9	135	125.2	123.4
B_MSY/SSB_unfished	0.2178	0.2182	0.2276	0.2283
Totbio_unfished	571078	580236	538113	530480
SmryBio_unfished	570687	579839	537745	530117
Recr_unfished_millions	16.94	17.21	15.96	15.74

SSB_MSY_thousand_mt	28.94	29.46	28.5	28.17
SPR_MSY	0.2667	0.2671	0.2759	0.2765
annF_MSY	0.1495	0.1496	0.1467	0.1463
Dead_Catch_MSY	47975	48761	46231	45666
Ret_Catch_MSY	47975	48761	46231	45666
B_MSY/SSB_unfished	0.2178	0.2182	0.2276	0.2283
Bratio_2023	1.571	1.599	1.427	1.393
F_2023	0.5561	0.5411	0.6011	0.6163

Table 5: Akaike information criterion comparison for the southwest diagnostic model and alternative time-varying selectivity configurations. Delta AIC is calculated relative to the model with the lowest AIC.

Model	AIC	Delta AIC
SW diagnostic	269.266	0.000
TV peak	277.724	8.458
TV ascending width	272.500	3.234
TV peak and ascending width	282.413	13.147

Table 6: Likelihood-ratio tests comparing nested southwest selectivity models. The tests do not provide significant support for the additional time-varying parameters.

Reduced model	Full model	NLL reduced	NLL full	Parameters reduced	Parameters full	df	LR statistic	p-value
SW diagnostic	TV peak	38.6330	37.8620	96	101	5	1.5420	0.9082

SW diagnostic	TV ascending width	38.6330	35.2498	96	101	5	6.7664	0.2386
TV peak	TV peak and ascending width	37.8620	35.2064	101	106	5	5.3112	0.3791
TV ascending width	TV peak and ascending width	35.2498	35.2064	101	106	5	0.0868	0.9999

14.1.1 Spawning biomass

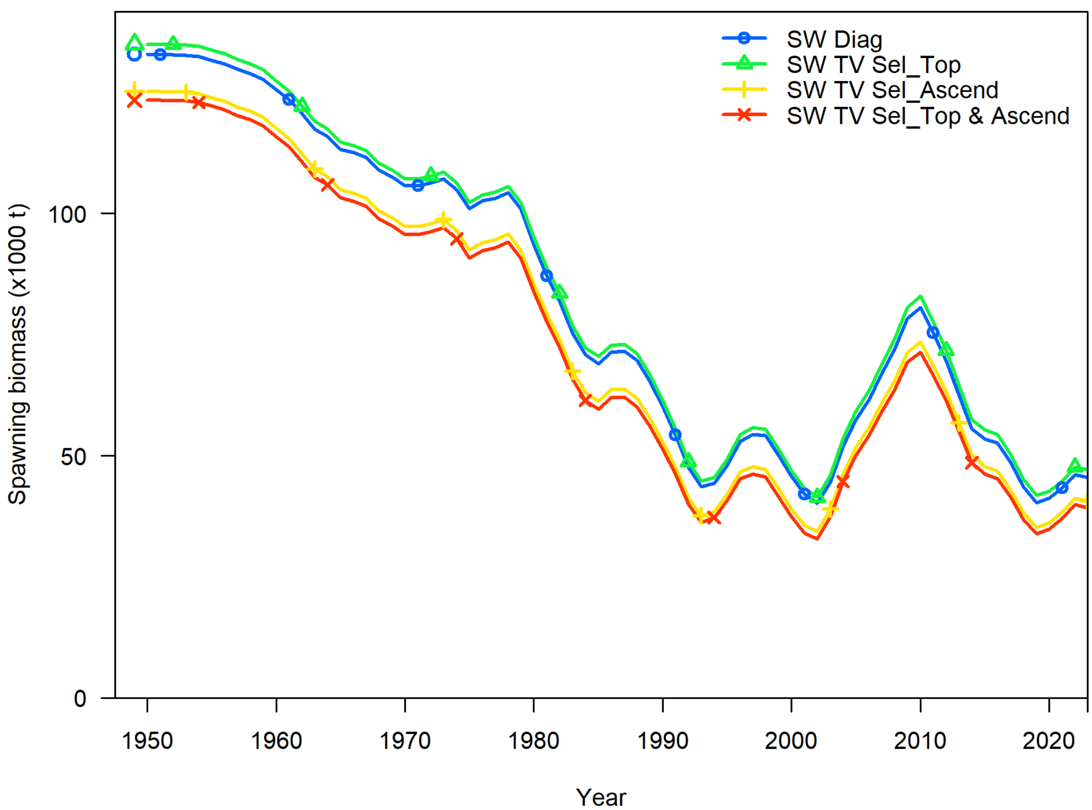


Figure 10: Estimated spawning-biomass trajectories for the southwest diagnostic model and alternative time-varying selectivity configurations. Time variation was applied to fleets 4–8.

The peak-only formulation produced a spawning-biomass trajectory close to, and at times slightly above, that of the diagnostic model. The ascending-width and combined formulations produced lower biomass throughout much of the assessment period. Although these changes illustrate that selectivity assumptions can affect estimated population scale and recent biomass, the statistical evidence does not demonstrate that the additional parameters are required.

14.1.2 Mean-length residual runs tests

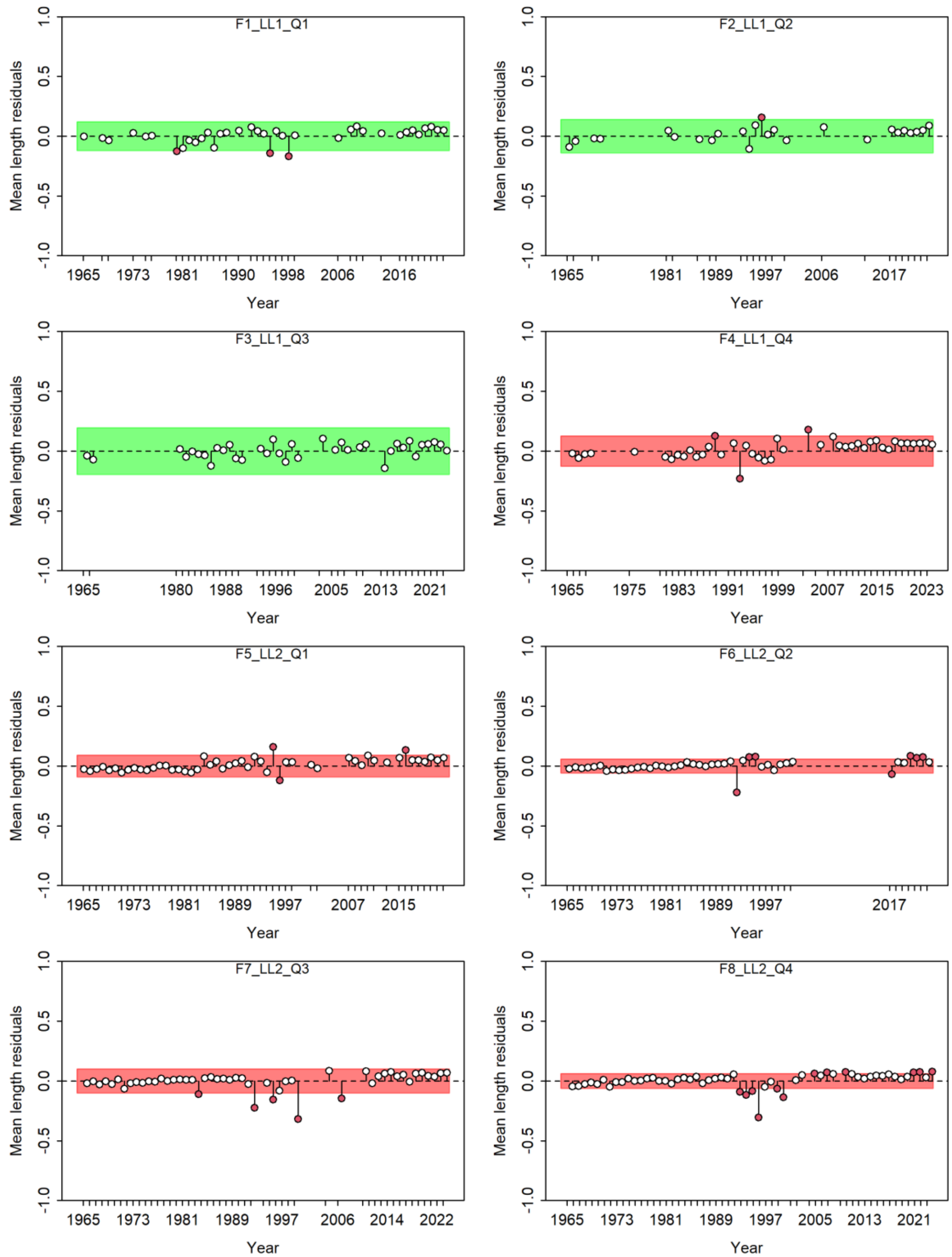


Figure 11: Runs tests applied to mean-length residuals for fleets 1–8 in the southwest diagnostic model with constant selectivity. Fleets 4–8 exhibited the clearest evidence of non-random temporal structure and were selected for the time-varying selectivity analyses.

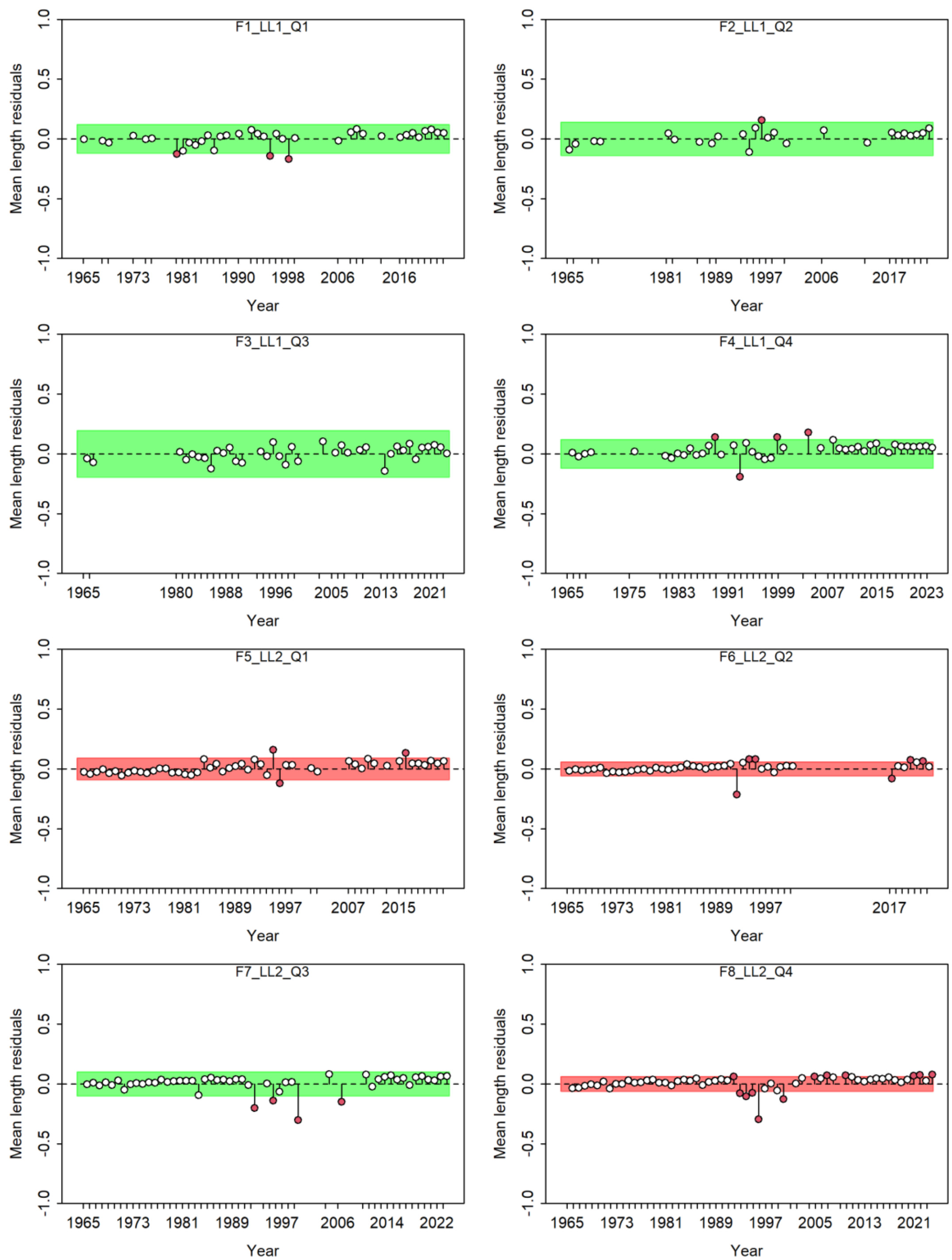


Figure 12: Runs tests applied to mean-length residuals for fleets 1–8 in the southwest model with time variation in the selectivity peak for fleets 4–8.

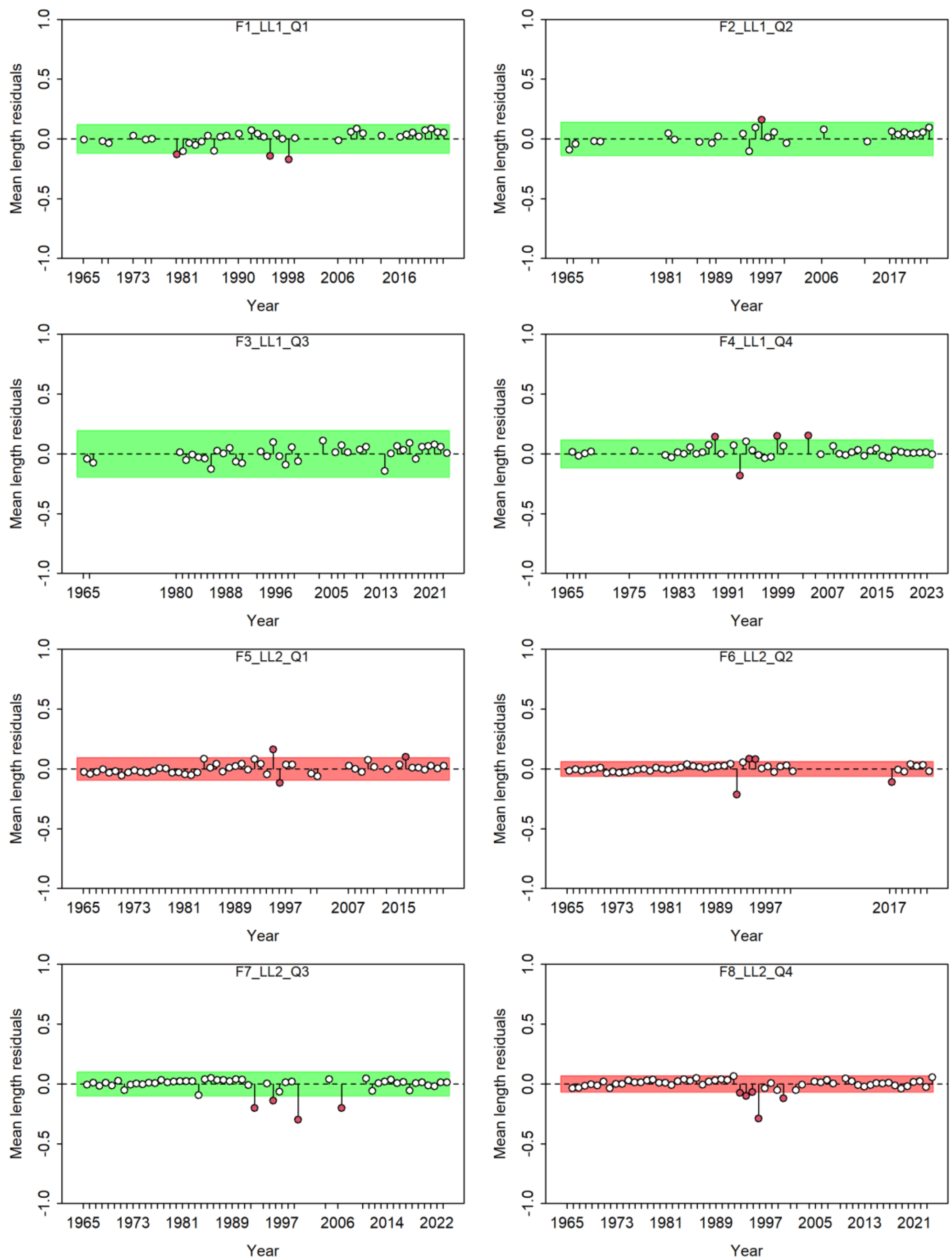


Figure 13: Runs tests applied to mean-length residuals for fleets 1–8 in the southwest model with time variation in the ascending-width selectivity parameter for fleets 4–8.

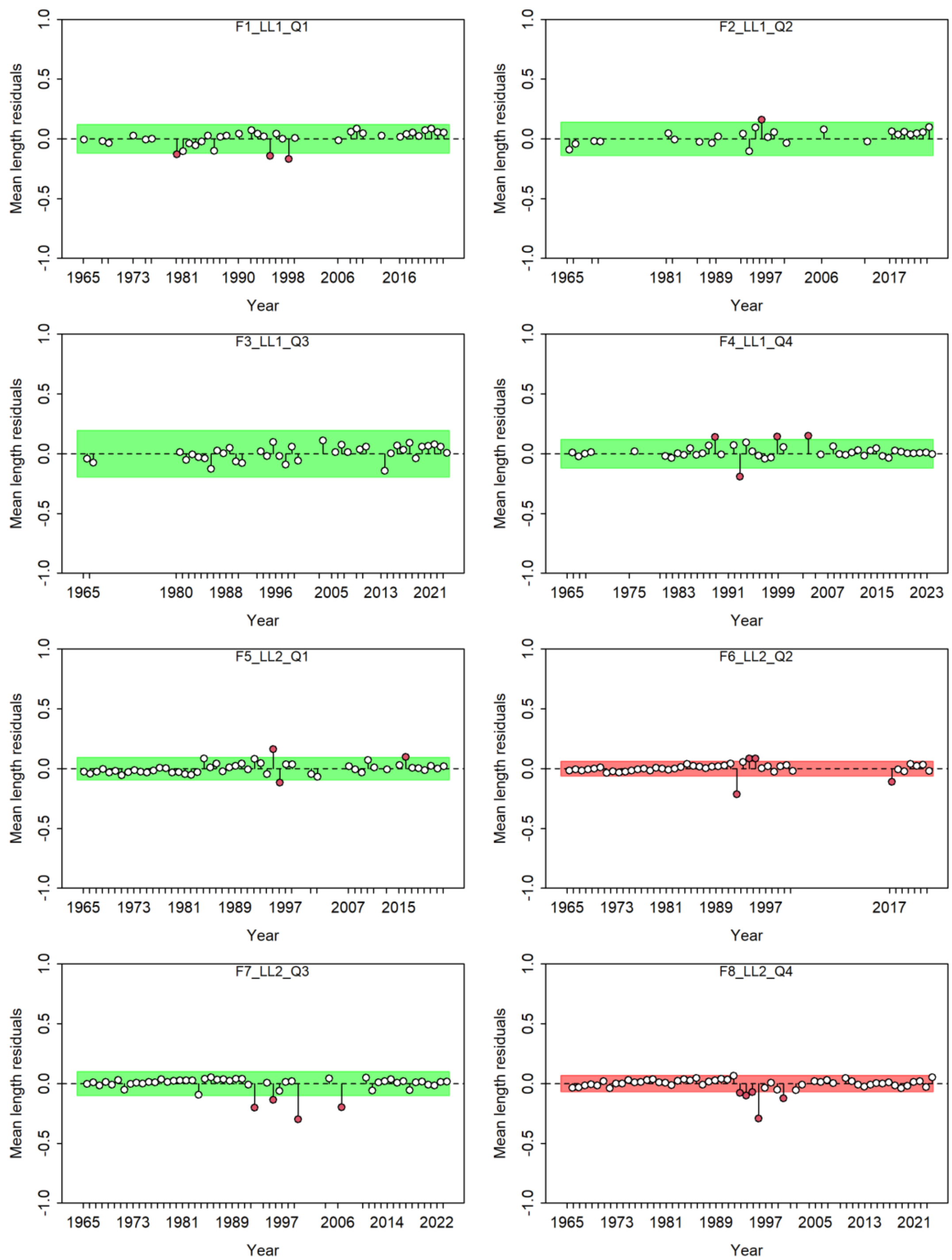


Figure 14: Runs tests applied to mean-length residuals for fleets 1–8 in the southwest model with time variation in both the selectivity peak and ascending-width parameters for fleets 4–8.

Time-varying selectivity improved the runs-test outcomes for some of the targeted southwest fleets, particularly under the ascending-width and combined formulations. Nevertheless, significant temporal structure remained for other fleets, and the improvements were not consistently sufficient to justify the additional parameters. The combined formulation provided essentially no improvement in likelihood over the ascending-width model.

14.2 Conclusions

The time-varying selectivity analyses produced different conclusions for the two diagnostic models. For the northwest model, the likelihood, AIC, and residual diagnostics strongly support allowing recent selectivity to differ from the historical configuration. The data support at least one time-varying selectivity parameter, although they provide limited discrimination among the peak-only, ascending-width-only, and combined formulations. The single-parameter models provide a more parsimonious representation, while the combined model represents an important sensitivity because it produces a more pessimistic estimate of recent stock status.

For the southwest model, targeted time-varying selectivity for fleets 4–8 improved selected mean-length residual diagnostics but was not supported by the likelihood-ratio tests after accounting for the additional parameters. The static-selectivity diagnostic model therefore remains the provisional reference configuration for the southwest case, with the ascending-width model retained as a structural sensitivity illustrating the possible effects of changing selectivity on estimated biomass and stock status.