SCIENTIFIC COMMITTEE
TWENTY-FIRST REGULAR SESSION

Nuku'alofa, Tonga
13–21 August 2025

Revised 2024 Stock Assessment of Striped Marlin in the Southwest Pacific Ocean Part 1: Integrated Assessment in Stock Synthesis

WCPFC-SC21-2025/SA-WP-06 Rev-3
07 August 2025

C. Castillo-Jordán[1], N. Ducharme-Barth[2], J. Day[1], N. Davies[3], T. Peatman[4], K. Kim[1], P. Hamer[1]

---

[1] Oceanic Fisheries Programme of the Pacific Community
[2] NOAA, Pacific Islands Fisheries Science Centre
[3] TeTakina Ltd
[4] Shearwater Analytics LTD

1 Author contribution

Claudio Castillo-Jordan lead and conducted the analysis and contributed to results interpretation and the report production. **Nicholas Ducharme-Barth** supported analysis in a technical capacity and contributed to diagnosing model behavior and lead the Hawai'i workshp. **Paul Hamer** oversaw the assessment work, interpreted results and wrote the assessment paper. **Jemery Day** and **Nick Davies** provided technical support and guidance, and assistance with results interpretation. **Tom Peatman** conducted the work to provide the revised size data inputs. **Kyuhan Kim** provided technical advice and developed the code and workflow to run SS3 model grids on HTCondor.

2 Note on projections

SC20 requested that a range of stock projection scenarios be conducted to accompany the revised assessment (para 55, SC20 outcomes document). Owing to the issues raised during this revision, a meeting was held with the key delegations who proposed the projections; Australia and New Zealand, the SPC assessment team and NOAA collaborators, and the WCPFC Secretariat Science Manager. It was decided that projections may not provide reliable management advice and that the assessment should be reviewed by SC21 before considering any additional projections work. As such the projections have not been conducted, pending further advice from SC21.

3 Revisions

Rev 1: Figure of F/F_{MSY} trajectory for the grid of models added to Figure 40. **Rev 2:** Figure 38 correction to Y-axis label - should read t not 000s t. **Rev 3:** Additional details added to section 8.1.1 Convergence diagnostics.

Contents

4 Executive summary

This report describes the development of a revised integrated assessment for southwest Pacific ocean (SWPO) striped marlin (requested by SC20) that aimed to resolve issues raised at SC20 and subsequently at the 2025 Pre-Assessment Workshop. The assessment presented in this paper should be considered in combination with the companion assessment using a Bayesian Surplus Production Model (BSPM) (Ducharme-Barth (2025) SC21-SA-WP-07), as supporting information for SC21 to formulate management advice. The revised assessment involved collaboration with scientists from the US delegation based at the NOAA Pacific Islands Fisheries Sciences Center (PIFSC), Hawai'i, under an arrangement supported by SC20. This collaboration was initiated through an in-person workshop in Hawai'i in January 2025. An important part of this revision involved moving the assessment from Multifan-CL (MFCL) to Stock Synthesis (SS3), a collective decision that was made at the Hawai'i workshop. Along with the change in software, numerous other changes and data modifications were implemented in the revision based on the workshop in Hawai'i, discussions at the SPC Pre-Assessment Workshop (PAW) and OFP's regular stock assessment team meetings. Some of the more important changes included:

- Switch from MFCL to SS3.
- Start the model in 1952, and include early catch uncertainty.
- Implement size based selectivity and modifications to selectivity functions.
- Major revision of size data inputs and refined size data filtering, resulting in significant reduction in size data inputs.
- Separate NC and PF longline fleets from the mixed flag fisheries and create separate fisheries due to their different length data distributions.
- Use of ad hoc size data weighting to down weight the influential size data.
- Use CAAL data to estimate growth internally for the diagnostic model.
- Explored sensitivity to alternative CPUE indices, including Australian longline, NZ recreational and newly developed observer based longline indices for PICT fleets (Neubauer, 2025).

These changes did address the core concerns raised at SC20. However, despite the improvements the model results remained very similar to those from the previous 2024 MFCL assessment, and MFCL assessments prior to that (i.e. 2019 and 2012). In the course of the revision other issues were raised, most notably, concerns around the low population/biomass size (scale) that was estimated since the 1970s. These concerns were raised through the PAW process, and were not specifically noted as issues at SC20. The concerns are equally relevant to previous striped marlin assessments that estimated similar population scales. In this respect the review work on this assessment has relevance to previous SC accepted SWPO striped marlin assessments that were used for management advice by WCPFC.

The low population size estimates became a key focus of the assessment revision. Despite a range of modeling investigations, including one-off sensitivities on biological parameters, alternative CPUE and a factorial grid on biological parameters, the model estimates would not deviate substantially from the low population size. The SS3 model as configured was estimating the stock as being very small and highly productive, with the dynamics being driven by recruitment variability. Various diagnostic analysis pointed to issues the model was having to estimate a well determined production function. The development of the BSPM assessment was motivated to investigate these issues and see if a simplified more flexible Bayesian modeling approach could allow for the identification of a production function, and more effectively characterize uncertainty in estimates of population scale.

It is problematic to reconcile the small population scale in the context of the recent observed catches and the assessment spatial scale. Simple calculations indicated the population scale being estimated was implausibly low. None the less the assessment model passed a number of key diagnostics, provided reasonable fist to the data, and could produce all the quantities of management interest, such as; MSY, SB_{MSY}, SB/SB_{MSY}, F/F_{MSY}, $SB/SB_{F=0}$, and are presented herein. However, the reliability of the estimates of these reference point is questionable given the concerns over the low spawning biomass estimates, and uncertainty on how much higher the population biomass might actually be. More details on this issue are available in the discussion, however, MSY related quantities, as used by SC for billish management advice, will by highly uncertain if the absolute abundance scale is highly uncertain. Further model investigations also demonstrated that $SB/SB_{F=0}$ will also be biased low if population scale is biased low.

While we provide a summary of stock status relative to the estimated reference points, it is clear that uncertainty in population scale equates to uncertainty in the management reference points. This needs to be considered seriously by SC21 when considering how they might interpret the outcomes of this assessment for management advice. Assuming a higher 'true' population scale would result in more optimistic values of the key management reference points. The differences would depended on how much higher the true population scale is. The management reference points are more uncertain than presented here due to the population scale uncertainty, but it is reasonable to suggest that they may be overly pessimistic.

The revised integrated assessment should be considered in combination with the BSPM assessment when considering management advice, and also take into account trends in the empirical indicators such as the CPUE indices, that show recent increases in CPUE for several indices.

We invite the SC21 to:

- Note the considerable amount of work to revise this assessment including transitioning the assessment to SS3 (a request of SC20 for the 2029 assessment), and the collaborative work with NOAA PIFSC.
- Note the concerns raised on the uncertainty in population scale estimated by this assessment (and previous SWPO striped marlin assessments).

- Note the management reference points estimated by this assessment have greater uncertainty due to the uncertainty in population scale. Accepting that the population may be larger than estimated, the reference point estimates would be overly pessimistic.
- Consider the application of Close Kin Mark Recapture (CKMR) as a way to obtain a more reliable estimate of the population size.
- Support the increased collection of age-length data (otoliths) and genetic studies to improve understanding of population structure.
- Consider a workshop on assessment approaches and challenges for bycatch species, including billfish and sharks, to consolidate and share learnings and modeling approaches from this assessment revision and the multimodel approaches now used for shark assessments.

5 Introduction

Striped marlin (*Kajikia audax*) is one of six species of billfish commonly reported from commercial and recreational fisheries within the western and central Pacific Ocean (WCPO) (Whitelaw, 2001; Bromhead et al., 2004; Kopf, 2005; Molony, 2005). Nearly all commercial catches of striped marlin are made by longline fleets (Bromhead et al., 2004), although small catches of striped marlin have also been reported from purse-seine fisheries of the WCPO (Molony, 2005). Striped marlin is also an important recreational species throughout the region (Whitelaw, 2001; Bromhead et al., 2004; Kopf, 2005; Holdsworth, 2024).

In the southwestern Pacific Ocean (SWPO), there is a long history of striped marlin catches by longline fisheries, starting in the early 1950s with the expansion of the Japanese longline fleet into the waters off north east Australia when the MacArthur line restrictions were lifted (Williams, 2003; Ducharme-Barth et al., 2019; Honma and Kamimura, 1958; Okamoto, 2024). Recreational fisheries have also been important for a long time with sport fishing club records dating back to the 1920s in New Zealand (Kopf, 2005; Holdsworth, 2024). While striped marlin is typically considered a valuable byproduct species in tuna longline fisheries, in some areas of the SWPO they are opportunistically targeted (i.e. Australian East Coast Tuna and Billfish Fishery (ETBF) (Bromhead et al., 2004; Tremblay-Boyer and Williams, 2023)). Both sectors have seen declines in total catches over the last few decades, and long-term declines in fish size have been observed for recreational size data collected around New Zealand (Kopf et al., 2005; Ward and Myers, 2005; Ducharme-Barth et al., 2019; Holdsworth, 2024).

Compared to the tropical tuna species, which are regularly (i.e. every 3 years) assessed within the Western and Central Pacific Fisheries Commission (WCPFC) and Inter-American Tropical Tuna Commission (IATTC) convention areas, there have been relatively few assessments for striped marlin in the Pacific Ocean. One of the earlier assessments for striped marlin concluded that Pacific-wide longline effort (up to 1980) was below F_{MSY} (Skillman, 1989). A separate assessment (Suzuki, 1989), which considered northwest Pacific and SWPO populations as separate, concluded that both stocks were at healthy levels, although the SWPO stock appeared to be almost fully exploited (fishing mortality close to F_{MSY}).

The first attempt at a SWPO only assessment was in 2006 (Langley et al., 2006). This assessment used Multifan-CL (MFCL) (Fournier et al., 1998) and was considered more as a research assessment rather than for provision of management advice. The next assessment, also using MFCL, was in 2012 (Davies et al., 2012), and suggested the SWPO stock was close to being overfished but that overfishing was unlikely to be occurring with respect to maximum sustainable yield (MSY) based reference points. The most recent SC-accepted assessment for the WCPO, also using MFCL, indicated that the SWPO stock was likely overfished and close to undergoing overfishing according to MSY based reference points (Ducharme-Barth et al., 2019). The assessed poor condition of the SWPO stock was not dissimilar to the northern Pacific stock which according the last three assessments was overfished and undergoing overfishing relative to MSY based reference points (ISC-BWG, 2015, 2019, 2023).

In 2024 a new MFCL assessment of the SWPO striped marlin stock was conducted (Castillo-Jordan et al., 2024) and provided to SC20. This assessment, while showing very similar trends to the previous assessments, estimated higher fishing mortality levels and a worse stock status. However, there were several technical issues identified that meant the assessment was deemed not acceptable as a basis for management advice at SC20.

Detailed information on the **Biology and Fisheries** for SWPO striped marlin can be found in the 2024 assessment paper (Castillo-Jordan et al., 2024).

5.1 Revision of the 2024 assessment

SC20 tasked the Scientific Services Provider (SSP), the Oceanic Fisheries Programme of the Pacific Community (SPC) to conduct additional work and provide a revised SWPO striped marlin assessment to SC21 in 2025.

As noted by SC20 (Anon, 2024): **Agenda 4.4.1.1** Stock assessment of Southwest Pacific striped marlin

- **para 248.** SC20 thanked the SSP for conducting the stock assessment of the Southwest Pacific striped marlin (SC20-SA-WP-03) and acknowledged their transparency in outlining the issues of serious concern from the SC with the technical aspects of the assessment.
- **para 249.** Noting the above, SC20 recommended that further work, including resolving the conflict between the size composition data and the CPUE indices, should be undertaken as part of a revision to the assessment for consideration at SC21.
- **para 250.** SC20 acknowledged that further work on the assessment will place unplanned additional burden on the assessment team and welcomed the offer from the United States to provide in-kind support to SPC to moderate the impacts to other key SC21 deliverables.

While the SC20 concerns focused on the data conflict, other concerns raised by SC20 delegates included:

- A declining recruitment trend over the last decade.
- Annual selectivities which do not allow the model to capture fishing dynamics and lack the necessary flexibility to fit the size composition data, given the rapid growth in the new growth curve, resulting in poor fits to the length frequency data.
- The model's inability to estimate initial population conditions (i.e. starting the model in 1979 rather than 1952 when the catch history starts).

- Conflicts between length and weight frequency data in some fisheries.
- Lack of fit to the CPUE data in the most recent years.
- Fixed initial conditions (1979 start year) which affect uncertainty estimation and sensitivity analyses.
- Concerns the level of depletion may be overestimated because of the strong scaling effect from size data, resulting in a vulnerable biomass trend that appears to be more pessimistic than the CPUE trend indicates.

The US delegation's generous offer to support a revision of the assessment was accepted by the SSP and supported by SC20. This work was initiated through an in-person workshop in Hawai'i in January 2025, and the collaboration has continued through to SC21, including with the companion surplus production model assessment (Ducharme-Barth, 2025). The outcomes of the Hawai'i workshop are available at https://n-ducharmebarth-noaa.github.io/2025-swpo-mls-meeting/, and also provided in Ducharme-Barth et al. (2025).

A concern that was considered early in the assessment revision was the annual model structure and the coarse resolution of the selectivity at age (annual) used, given the revised growth information indicating faster early growth (Farley et al., 2021). The model struggled to select, and therefore, fit the smaller fish present in the fishery size data. While a quarterly age MFCL model helped to achieve a finer resolution to the selectivity ogive, a quarterly model also estimated quarterly recruitment, which is not consistent with the annual spawning cycle of the species (Farley et al., 2021; Kopf et al., 2012). While MFCL has since implemented a size based selectivity option and can deal with annual recruitment, at the time of the Hawai'i workshop this was not available and it was decided to explore the Stock Synthesis (SS3) platform given its size-based selectivity which could allow the desired resolution for the selectivity ogives and maintain an annual recruitment. The workshop participants made the decision to transfer the assessment from MFCL to SS3 and then focused on developing the SS3 model. SC20 made a request to shift the 2025 southwest Pacific swordfish assessment to SS3 and also expressed a desire to move the next (2029) striped marlin assessment to SS3. This general SC request to move billfish assessments to SS3 along with familiarity of the workshop team and the lead assessor to working with SS3, further justified the decision to collaborate on a revised assessment in SS3 rather than continue in MFCL. This decision was supported by the 2025 SPC pre-assessment workshop (PAW) (Hamer, 2025) as an efficient way forward, and in line with the MFCL succession plan (WCPFC project 123).

The work to revise the 2024 assessment in the SS3 platform is covered in this report and we do not present results from any MFCL model explorations that occurred post-SC20 but prior to the Hawai'i workshop. If people are interested in these, please contact the lead author (claudioc@spc.int).

5.1.1 Additional concerns raised at the 2025 SPC PAW

As the revision work progressed and was discussed in depth at the 2025 SPC PAW, further concerns were raised. As part of the revision (guided by the Hawai'i workshop), it was decided to revert back to starting the model at the beginning of the available catch history (1952), assuming a largely unfished state. This was in response to concerns raised on the fixed initial conditions that were required with the previous model starting in 1979. The preliminary SS3 model starting in 1952 estimated a very large decline in biomass that occurs over only a few years coincident with a single year (1954) spike in catches by the Japanese fleet recorded off the east coast of Australia towards New Caledonia (Figure 3). Irrespective of some uncertainty over the reliability of the magnitude of early catches, it may be questionable that such a short lived and regionally localised (relative to the entire SWPO) catch spike would drive the broader stock level down so rapidly and to the extent predicted by the model. Beyond this initial rapid depletion, the estimated stock status does not recover, and continues to gradually decline, despite apparent catch reductions over time, albeit more so in recent decades (Figure 3). The adult fishing mortality is estimated to therefore increase to relatively high levels. The PAW discussed the plausibility of the sustained estimated low spawning biomass and relatively high F, suggesting that the high F is required later in the time series to explain the size data, given that CPUE starts in 1979 and is relatively flat. The PAW suggested that the mechanisms driving the apparent stock trajectory and population scaling needed to be further investigated.

Growth was another key uncertainty that was subsequently raised post PAW as being potentially impactful on scaling and important to explore. Given the current information on growth is from a very limited age-length sample (around 80 samples) from otoliths samples restricted to the region west of Fiji (i.e. Fiji, New Caledonia, Australia and New Zealand) (Farley et al., 2021), the growth information may not be representative of the wider stock being modeled.

5.1.2 Scope

This report describes the development of a revised integrated assessment in SS3 that attempted to resolve the issues raised at SC20 and subsequently at the 2025 Pre-Assessment Workshop. The models presented in this paper should be considered in combination with the additional assessment using a Bayesian Surplus Production Model, as supporting information for SC21 to formulate management advice for the SWPO striped marlin stock (Ducharme-Barth, 2025).

6 Data and Methods

6.1 General notes

The raw data used for the revised SS3 assessment of SWPO striped marlin are more or less the same as those used in the 2024 MFCL assessment, except that the SS3 model begins in 1952 (compared to 1979) (Figure 2), which introduces earlier catch data and some earlier size data, notably weight data from the NZ recreational sport fishery records. The data have not been updated to include 2023 as this was not requested by SC20 and the expectation is that next assessment will go ahead in 2029 as planned (Brouwer and Hamer, 2024). The key changes to the data inputs for the SS3 assessment include (and discussed further below):

- **Catch data:** addition of catch data prior to 1979, mostly from the Japanese fleet, with some early catches from Korea and Chinese Taipei.
- **Size data:** addition of NZ recreational and some Japanese longline weight data available prior to 1979, revised size compositions for the extraction and index fisheries due to new filtering approaches to remove and reduce the influence of unrepresentative data (Peatman, 2025), removal of all length data from the Australian longline fishery as weight data are considered more representative, removal of aggregate Chinese Taipei longline size data that were collected by skippers or crew.
- **Abundance Indices:** additional CPUE abundance indices are included for comparison; Australian longline (Tremblay-Boyer and Williams, 2023), New Zealand recreational sport fishery (Holdsworth, 2024), and observer data form longline fisheries in New Caledonia, Fiji, Tonga and French Polynesia (Neubauer, 2025).

6.2 Spatial stratification

The spatial stratification for the SS3 assessment is the same as for the 2024 MFCL assessment. The stock assessment region covers the area from the equator to 40°S and from 140°E to 130°W (Figure 1). This represents the region of the SWPO where most striped marlin catches have been reported since 1952. The assessment region excludes areas to the north and east due to limited evidence for mixing between these regions (Bromhead et al., 2004). However, we acknowledge that spatial population structure for striped marlin in the south and central Pacific continues to be a major uncertainty, also noted in Ducharme-Barth (2025). The lack of clear evidence for population substructure in the SWPO means the default assumption of a single mixed population in the model region. Few striped marlin have been reported from Australian longline fisheries south of 40°S, or New Zealand longline fisheries south of 38°S. Overall, the assessment region is assumed to encompass a semi-independent stock of striped marlin and, given the spatial distribution of the catch, is thought to represent an appropriate spatial scale for assessment and management of the striped marlin resource in the SWPO.

Similar to the 2024 MFCL assessment, the SS3 assessment specifies four sub-regions (Figure 1) for a single region fleets-as-areas approach. The sub-regions have been defined based on qualitative and quantitative assessments of the distribution of fishing effort and catch for the major fisheries, the size composition of the catch and the qualitative assessment of available tagging and recapture data (Langley et al., 2006; Davies et al., 2012; Ducharme-Barth et al., 2019; Castillo-Jordan et al., 2024; Dormeier, 2006). These sub-regions were used to define the spatial boundaries for the grouping of the 'extraction' fisheries operating within the assessment region. Whereas the index (or survey) fisheries are considered as representative of the population across the entire model region.

6.3 Temporal stratification

Data used in the SS3 stock assessment covers the period 1952–2022, with different time periods available depending on the data sources and types (Figure 2), whereas the 2024 MFCL assessment started in 1979. Consistent with the 2024 MFCL models, the SS3 models have an annual age structure with quarterly population and fishery dynamics. Both models assumed a single annual recruitment in quarter 1, based on the spawning season of striped marlin in the SWPO being October-December (Kopf et al., 2012). In the 2024 MFCL model, the CPUE abundance index was separated into separate indices for each quarter, whereas in the SS3 models a single quarterly index corresponding to quarter 4 was used. Quarter 4 was selected as this is most likely to represent changes in the spawning component of the population, which is the predominant size selected in the index fisheries (Ducharme-Barth et al., 2025).

6.4 Definition of fisheries

Extraction fisheries: The sub-regions of the model were used to define the spatial boundaries for grouping fisheries in the SWPO. The model fisheries are defined in Table 1. The defined fisheries have changed slightly for the SS3 assessment. The changes involve:

- **Early Japanese longline F15.LL.JP.2.early:** an additional Japanese longline fishery was added for the early model period (1952-1978) in sub-region 2. This allowed inclusion of catch uncertainty in the model for the early Japanese catch data, which was important to consider given the switch back to starting the model in 1952 and the concerns raised about accuracy of the early high catch data.
- **2024 MFCL fishery 12 (LL.ALL.2):** this is a mixed flag fishery. Review of the size data suggested larger fish are landed by New Caledonia (NC) than the other flags, so a separate extraction fishery was created for NC in the SS3 model (F16.LL.NewCal.2).
- **2024 MFCL fishery 14 (LL.ALL.4):** this is a also mixed flag fishery. Review of the size data suggested larger fish are landed by French Polynesia (PF) than the other flags,

so a separate extraction fishery was created for PF in the SS3 model (F17.LL.FPoly.4).

The SS3 model now has 17 extraction fisheries. There are 15 longline fisheries and two recreational fisheries (Table 1). A more detailed description of the fisheries characteristics and history is available in the 2024 report (Castillo-Jordán et al., 2024). Briefly, the Japanese longline vessels historically dominated both total longline fishing effort and striped marlin catches within the model region and have had a continued presence since the beginning of the assessment period, so a separate Japanese fishery was defined in each of the four model sub-regions (Fisheries 1-4, 15), with the fishery in sub-region 2 being split into two time periods. Japanese longline fishing effort was inconsistent in sub-region 4, and by the 1980s was virtually non-existent, so a separate fishery was defined for Chinese Taipei longline vessels, as their effort and catches of striped marlin increased in sub-region 4 post-1970 (Fishery 5). Since the mid-1990s, major longline fisheries were developed off the Australian east coast (Fisheries 6-7). A longline fishery is defined for New Zealand (Fishery 8) in sub-regions 2 and 3, although this fishery had a non-retention policy for striped marlin introduced from 1987 and has very low historical catches. In order to account for the remaining mixed flag longline striped marlin catch, four additional longline fisheries were defined in each of the model sub-regions (Fisheries 11-14; Table 1). These fisheries grouped together the catch and composition data of the other DWFNs and PICTs operating in the assessment region, with the exception of the NC and PF fisheries 16 and 17 respectively, which have their own fisheries defined due to the observation of different size composition. These fisheries are also unique in that they are entirely localised fleets with no DWFN vessels operating in their EEZs.

In addition to the longline fisheries, individual fisheries were defined in sub-region 3 to account for the two major recreational fisheries targeting striped marlin in the assessment region: Australia (Fishery 9) and New Zealand (Fishery 10; Table 1). Though the magnitude of catches is dwarfed by total longline removals of striped marlin, these two fisheries are historically important with catch records dating back prior to 1940 for the Australian recreational fishery and weight composition records dating back to 1925 for the New Zealand recreational fishery (Holdsworth et al., 2019; Holdsworth, 2024). Other recreational fisheries capture striped marlin in the assessment region (Whitelaw, 2001), but these fisheries are relatively small, and catch and effort data are not readily available.

CPUE index (survey) fisheries: SS3 allows the specification of “survey fisheries”, which are referred to as “index fisheries” in MFCL assessments. The index (or survey) fisheries provide standardised CPUE indices of abundance. In this case they are fishery dependent abundance indices and technically not surveys as referred to in SS3. Plots however produced for this paper that used r4ss software may use the term “survey” to indicate the CPUE index fisheries. The full longline operational data-set (described in Davies et al., 2012; Ducharme-Barth et al., 2019) was used as the basis for developing the diagnostic model index fishery CPUE for the combined Japan/Chinese Taipei longline data that was used in the 2024

MFCL model. That index was developed using the sdmTMB package for implementing spatio-temporal delta-generalised linear mixed models and is described in detail in Castillo-Jordán et al. (2024). The Japan/Chinese Taipei sdmTMB longline index was used as the index fishery for the diagnostic model in SS3. However, in the revised SS3 assessment, based on the Hawai'i workshop, it was decided to explore several alternative CPUE abundance indices:

- **Australian longline CPUE:** standardised using Generalized Additive Models (GAMs) (1998-2022) (Tremblay-Boyer and Williams, 2023)
- **New Zealand recreational sport fishery CPUE:** standardised using Genalized Linear Models (GLM) (1975-2022) (Holdsworth, 2024)
- **Pacific Island Countries and Territories (PICT) observer longline CPUE:** standardised with Generalized Linear Mixed Models (GLMM) using spatio-temporal splines, oceanographic predictors (NINA4) and operational covariates including gear characteristics (2001-2022) Neubauer (2025)

The use of longline CPUE from the PICT regional observer (ROP) data, as has been done in shark assessments, was suggested at the 2025 PAW. This could take advantage of additional covariates, however, the time period of data is shorter. Two additional observer based longline CPUE indices are considered in the SS3 assessment. These are a combined observer longline index for Fiji (FJ), New Caledonia (NC) and Tonga (TO) and a separate index for the French Polynesia (PF) observer data. NC/FJ/TO were combined as they showed similar trends and a common recent increase in CPUE, whereas PF was different and had no recent increase. The development of the PICT observer based abundance indices is described in Neubauer (2025).

The revised SS3 assessment thus had 5 alternative CPUE index fisheries (F18-22, Table 1). The standardised CPUE abundance indices for each of these index fisheries are displayed in Figure 10.

6.5 Catch data

Catch data were compiled according to the fisheries defined in Table 1. Catch data in the assessment are input as numbers of fish for all fisheries. This is consistent with the form in which the catch data are recorded and reported for these fisheries. The catches are aggregated at 5° × 5° cells and quarter-year resolution, with the aggregation process either conducted by SPC, where operational data are available to inform this, or by the particular countries following statistical procedures that are reported to the WCPFC. The catch history is shown by major catch flags in Figure 3 and in numbers of fish by model fishery in Figure 4. The spatial distribution of recent catches across the model region are shown in Figure 5, for historical distributions see Ducharme-Barth et al. (2019). It is worth noting here the separation of the early period extraction fishery for Japan longline (Fishery 15). Moving to Stock Synthesis

(F_Method = 4) allowed some fisheries (i.e. extraction fisheries 1–14) to remain defined as catch-conditioned, where fishing mortality is calculated to fit the catch exactly, but for Fishery 15 (Japan longline, 1952-1978), to estimate the fishing mortality (F) to fit to the catch with observation error (Ducharme-Barth et al., 2025). More detailed descriptions of the catch history data for the model extraction fisheries can be found in the report from the 2024 MFCL assessment (Castillo-Jordan et al., 2024) and is not repeated here.

6.6 Size composition data

A key part of the assessment revision involved revising the size data inputs (Peatman, 2025), although the following aspects remain consistent with the 2024 analysis. Size composition data, comprising either length-frequency (LF) and/or weight-frequency (WF) data, are summarised in Figure 2. For most fisheries, temporal coverage of the size frequency data is limited or patchy (Castillo-Jordán et al., 2024; Peatman, 2025). Length data were provided based on three main length measurement methods: eye orbital–fork length (EO), lower jaw–fork length (LJ), or pelvic fin–fork length (PF) as is standard practice for billfish measurements (Hamer et al., 2025). Weight data are provided in various forms depending on the source, for example, Japanese processed weights (gilled, gutted, head and tail left on, fins trimmed, bill removed at a few cm in from the tip of the lower jaw), and gutted, finned and headed (Australia), or whole weights (NZ recreational) (Hamer et al., 2025). All length measurements were converted to EO, and weight measurements were converted to the equivalent whole (unprocessed) weight (WW) using the conversion factors listed in Hamer et al. (2025). More detailed information on the size data by fisheries is available in Castillo-Jordán et al. (2024); Peatman (2025). Size data were aggregated by fishery and year-quarter. Length data were aggregated into 63 5-cm size classes (20–335 cm EFL). Weight data were aggregated into 123 2-kg intervals (5–249 kg whole weight).

Working with Australian colleagues we were also able to obtain additional information to more accurately assign their longline weight composition data to sub-regions 2 and 3, which was a concerns raised in the 2024 analysis. Any data that could not be confidently assigned were not used. This has little impact on sample coverage given the large number of samples available for this data source.

As noted, size composition data were problematic in the 2024 MFCL assessment and underwent considerable review for the SS3 assessment, see Peatman (2025). A detailed scrutiny of the size data was also presented and discussed at the 2025 PAW (Hamer, 2025), where a range of suggestions were supported to improve the data. These included:

- Removing the length data from the Australian longline fishery as this was considered not as representative as the extensive port sampled weight data (on advice of Australian scientists), also the length data are no longer being collected and conflicted with the weight data.

- correcting the issue of accurately assigning weight composition data to sub-region 2 and 3 for the Australian longline weight data.
- Removing any data from the Chinese Taipei longline (TWLL) that were measured on vessels by skippers/crew as it was noted that these are mostly estimated by eye and accuracy/precision is questionable (advice from Chinese Taipei scientist).
- Removing data from port sampling in PICTs (i.e. SPLL - Regional (PICTs) longline port sampling data of lengths), due to evidence of biases, i.e. lack of small fish in port samples compared to observer data (Peatman, 2025).
- Filtering of the mixed flag fishery size data to improve the representativeness with respect to the catch proportions by the various flags, described further in Peatman (2025). Also separation of NC and PF size data into single flag fisheries owing to their different size compositions compared to the other flags in the original mixed flag fisheries.
- Truncate the NZ recreational weight data at 1988 after which voluntary guidelines to release fish less than 95kg were brought in which may introduce time series bias. Note the MFCL2024 assessment started in 1979 and used the NZ recreational weight data from 1979-2022.
- Finally in an exploration to reduce data conflict, for the revised SS3 diagnostic model we removed the size data altogether for the mixed longline fisheries, Fisheries 11 and 12 (Table 1). This had little impact on model estimates, but did reduce data conflict to a degree so we maintained this change.

6.6.1 Size composition data preparation: filtering and reweighting

Considerable time was placed on size data inputs for the assessment revision, as concerns were raised at SC20 and PAW around the perceived influence and representativeness of some of these data. This work involved an in-depth review of the size data inputs (document provided to the Hawai'i workshop) and presentation and discussion of this review at the PAW. We also made revisions and refinement to size data preparation methods, this included stronger procedures for filtering to remove unreliable data, and low sample sizes that are not likely to be representative of flag/fishery/strata catches. These filtering steps are built into the data preparation and reweighting process (below) described in detail in Peatman (2025). These revisions removed a lot of the patchy small sample size data (e.g. most recent Japan LL length data since the late 1990s, and all the data from Fisheries 13 and 14), and added to the reductions already noted above.

Statistical correction of size composition data is required as length and weight samples are often collected unevenly in space and time. The methods for filtering and re-weighting of the size composition data are detailed in Peatman (2025). For the extraction fisheries, re-weighting of composition data is required to ensure that sampling biases in space, time, and the fleets providing data, are minimised so that size composition data better reflect the composition of the overall removals. Strata-specific size data samples were therefore re-weighted by catch for the extraction fisheries.

For the index fisheries, re-weighting of composition data was required to ensure that the size composition of the abundance indices reflect the size component of the population that is being sampled by the index fisheries through time. Strata specific samples are therefore re-weighted by relative abundance using the CPUE data. Where the same composition data were used for both the extraction and index fisheries, the observed number of size-frequency samples input into the assessment was divided by two for both the extraction and index fisheries, noting that several index fisheries mirror selectivity of extraction fisheries (Table 1).

6.7 Model Description

The stock assessment was conducted using SS version 3.30.23.1-SAFE released 05/12/204 using ADMB 13.2 (Methot and Wetzel, 2013). The SS3 model was initially developed at the Hawai'i workshop and this development work is described in Ducharme-Barth et al. (2025). It initially aimed to preserve similarities to the MFCL model as much as possible as a starting point, applying consistent structural aspects, such as fisheries definitions and biological assumptions, such as M-at-age, maturity ogive, and stock recruitment steepness of the diagnostic model. The decision to move to SS3 was partly to take advantage of features available in SS3 to try to fit the data better, i.e. length based selectivity. The SS3 model can be viewed as a separate assessment and development of the SS3 model was somewhat free to take its own course. The goal was not try to replicate the MFCL model results, but to develop an assessment based on the same core datasets but in SS3.

In order to account for the observed spatio-temporal differences in the size structure of the catch an areas-as-fleets approached was applied with four sub-regions to allow the flexibility to estimate specific size based selectivity functions to match the pattern of observations in each sub-region (Figure 1). The population is modeled as a single sex, with annual age structure and quarterly population and fishery dynamics. The model partitions the population into 10 annual age-classes. The last age-class comprises a "plus group" above which mortality and other characteristics are assumed to be constant. Natural mortality (M) at age is described using the Lorenzen functional form for M-at-age, with an assumed average M of 0.36 yr^{-1} based on the Amax (Hamel, 2015; Cope and Hamel, 2022) method with a maximum age of 15 years for the diagnostic model. This is the oldest age from the limited otolith ages available for SWPO striped marlin, and would likely be an underestimate of the maximum age for this, apparently heavily and historically, fished population. The average M value is used by SS3 to 'scale' the Lorenzen M-at-age based on the growth curve. Recruitment was assumed to follow a Beverton and Holt stock-recruitment relationship (SRR) with a fixed value of steepness (h), of 0.8 for the diagnostic model and σ_R set at 0.6 (i.e. (ISC-BWG, 2023) to allow a reasonable amount of recruitment variation.

For the diagnostic model, following the Hawai'i workshop, the growth curve (von Bertalanffy) was estimated internally from the conditional-age-at-length (CAAL) data provided to the model and fixed external values of K, L1 and L2 were explored in sensitivities. CAAL data were associated with the appropriate fleet definitions based on sampling location and flag of vessel the samples came from. Though the aging data recorded decimal age, age was entered in the model as the nearest model age less than or equal to the decimal age. Given that model ages were annual, this produced integer age values. L1, L2 and K were freely estimated for the diagnostic model. An aging error matrix with standard deviations of aging error ranging from 0.35 at age zero to 2 at age 10 was included. This was done to account for both actual aging error and for converting from decimal age to integer age to match the model structure. The maturity ogive was logistically parameterized with length at 50% maturity at 181.21 cm and a slope parameter of -0.20, consistent with the 2024 MFCL model (Castillo-Jordan et al., 2024).

The SS3 diagnostic model estimates 234 parameters compared to the MFCL2024 diagnostic that estimates 76 parameters.

The model fits four data components: 1) total catch (early period 1952-1978); 2) relative abundance indices; 3) length composition; 4) weight composition; and 5) conditional-age-at-length data (diagnostic model only). The observed total catches were assumed to be accurate and F calculated to fit the catch exactly for the period 1979-2022.

The relative abundance indices were assumed to have log-normally distributed errors with SE in log-space (log(SE)) which was log(SE)=sqrt(log(1+CV2)), where CV is the standard error of the observation divided by the mean value of the observation. The error component was determined by first fitting a loess regression to the observed CPUE values by year and calculating observation error as the root mean square of the loess residuals. Extra observation error was computed as the difference between total observation error and the mean of the reported log-scale standard errors. The SS3 diagnostic model used the same Japanese/Chinese Taipei sdmTMB index as used in the 2024 MFCL diagnostic model (Figure 10).

6.8 Size Data Weighting

The size composition data were assumed to have multinomial error distributions with the error variances determined by the effective sample sizes. Measurements of fish are usually not random samples from the entire population and are subject to considerable pseudoreplication due to being collected for individual longline sets or trips. The effective sample size should therefore be lower than the actual number of fish measured because the variance within each set or trip is substantially lower than the variance within the population. In the 2024 MFCL assessment the Francis weighting method was applied to iteratively adjust the effective sample sizes of the composition data in an attempt to allow appropriate data weights to be applied to size composition data for each fishery in the model fitting process (Castillo-Jordan et al., 2024). This approach was chosen to move away from the ad hoc size data divisors applied in previous assessments to reduce the effective sample sizes. Ultimately the Francis method did not produce a suitable level of down weighting of the size data which overwhelmed the abundance indices in the model fitting process. This is not desirable and was pointed out at SC20 as a critical issue.

For this SS3 assessment we applied the Francis weighting to the reduced length data, but moved back to applying fishery specific ad hoc size data divisors to the more influential weight data to achieve improved fit to the CPUE data. The weighting factors are shown in Table 1 in terms of the SS3 input variance adjustments factors.The input variance is in terms of an effective sample size and the variance adjustment factors are multiplied by the effective sample size. Due to concerns over influence of size data on population scaling, we also ran a diagnostic model with the effective sample adjustment multiplied x 10 to substantially further downweight all the size data (see Section 14).

Note that the additional filtering applied to the observed size data during the size data preparation and reweighting (Peatman, 2025) also substantially reduced the input sample sizes and removed noisy data with low sample sizes. Fitting CPUE indices was prioritised in the SS3 assessment.

6.9 Selectivity

Selectivity was estimated as a double-normal function of size for all fisheries except the NZ recreational sport fishery (F10.REC.NZ.3) which was assumed an asymptotic selectivity as it catches the largest fish, and to prevent accumulation of cryptic biomass. Several fisheries shared selectivity, summarised in Table 1. Additional models were explored where the right limb of the NZ recreational fishery selectivity was allowed to decline somewhat, but these had little impact on estimated quantities. Follow-up discussion on this issue with a NZ sport fishery expert (John Holdsworth, 5/7/2025) suggested that high loss rates of larger striped marlin were unlikely for the NZ recreational sport fishery as they are also targeting larger blue marlin and typically deploy heavier tackle. He also noted that all but one of the 14 International Game Fish Association (IGFA) line weight world records for striped marlin on breaking strains 6 Kg or heavier have been from NZ. An asymptotic selectivity at length assumption for this fishery seems reasonable. Selectivity curves estimated for the diagnostic model are shown in Figure 6.

6.10 Diagnostics

6.10.1 Model Convergence

Models in this assessment were considered for presentation if converged according to criteria below:

- 1. a maximum parameter gradient of <= 0.001 was achieved, with <= 0.0001 for the diagnostic model;
- 2. the Hessian was positive definite;
- 3. for the diagnostic model–the model fit is not be significantly improved by jittering the estimated parameters (jitter only performed on the diagnostic model);
- 4. estimated parameters should not be on their bounds.

Jitter involves changing the parameter start values by a small increment and rerunning the model to see if that adjustment causes the model to converge at a lower likelihood, and with changes to derived management quantities (Carvalho et al., 2021). 50 jitters were performed on the diagnostic model.

6.10.2 Fits to Data

Abundance Indices: Fits to CPUE time series data were examined from plots comparing the observed to estimated time series, and residual plots.

Size Data: Fits to size data by fishery were examined from plots comparing the aggregated observed to estimated size frequencies, comparisons of the time series of estimated and observed mean size and Pearson residual plots.

CAAL Data: Fits to the conditional age at length data were examined by plots comparing the observed (with standard deviation) and estimated mean ages by length, sampling times and fisheries from which the data were collected.

6.10.3 R_0 Likelihood Profiles

Likelihood profiling was performed on R_0 (number of recruits in an unfished population) to identify influential data sources on population scaling and the presence of data conflicts. R_0 is a useful parameter for profiling since it determines the absolute size of a population and as such provides a population scaling parameter. In MFCL assessments this diagnostic profiling is done on the total biomass.

6.10.4 Retrospective and hindcast analysis

Retrospective analysis (Hurtado-Ferro et al., 2015; Mohn, 1999) were conducted to the diagnostic model results by sequentially eliminating the five most recent years of data from the full stock assessment diagnostic model (i.e. 5 year "peel") and then re-estimating all stock assessment model parameters from each peel and from the full model. Then Mohn's rho was calculated for the biomass and fishing mortality peels, to measure the severity of the retrospective bias. Mohn's ρ values were interpreted using established thresholds: $|\rho| < 0.15$ indicates no significant bias, $0.15 \leq |\rho| < 0.20$ indicates minor bias, $0.20 \leq |\rho| < 0.30$ indicates moderate bias, and $|\rho| \geq 0.30$ indicates strong bias indicative of problematic retrospective patterns.

Hindcast analysis involves fitting the diagnostic model to truncated data sets and then forecasting the future values of the relative abundance index (JP/TW longline), using the fixed parameters from the truncated data model, and the observed catches for the full time series. The forecast index results are compared to the observed index. Ideally the recent data should not have much impact on the estimated parameters and the forecast estimates of the index should be similar to those for the observed index. If the hindcast estimates are inconsistent with observations it suggests the model is not stable and is sensitive to new data, the recent result may be unreliable or biased. Hindcasts were done with data truncation from 1 to 5 years, and cross validation performed against the full model using the MASE (Mean Absolute Scaled Error) statistic; Good (MASE <0.8), Acceptable (0.8-1), Poor (>1). Typically a MASE greater than 1 is interpreted as poor as it does not outperform the naive predictor,

6.10.5 Age Structured Production Model

An age-structured production model (ASPM) diagnostic for the diagnostic model was estimated by: 1. fixing growth and selectivity parameters at their estimated values; 2. removing the size composition data from the model, leaving only the CPUE indices as data to be fitted; 3. setting all log recruitment deviations 0, i.e. applying the mean recruitment from the stock recruitment curve; and 4. re-fitting the model estimating only the population scaling parameters.

A comparison of biomass and depletion scaling and trends estimated by the ASPM and the full model gives an indication of the extent to which these estimates are informed by the CPUE indices only. If the ASPM can produce a reasonable fit to the indices of abundance, assuming they have some contrasting trends in the time series, it suggests there is a production function, and the indices can provide information about absolute abundance (population scale) in the integrated assessment. If there is a poor fit to the indices, then the catch data alone does not explain the trends/dynamics in the indices of relative abundance (Carvalho et al., 2021). Poor fit of an ASPM to the abundance indices may indicate that: a) the stock is recruitment-driven; b) the stock has not yet declined to the point at which catch is a major factor influencing abundance; c) the model has mispecification issues; or d) the indices of relative abundance are poorly representative of abundance trends.

6.11 Sensitivity

The diagnostic model developed in SS3 improved on several diagnostic aspects that were of concern at SC20. Importantly the data conflict was reduced and the fit to the CPUE was improved. However, starting the model in 1952 produced a problematic rapid decline in biomass and early high recruitment and emphasized the issues around the lack of reliable early CPUE (relative abundance) information to estimate biomass scale. The subsequent low biomass throughout the remaining model period seemed unlikely in reality. Despite all the changes in the transition of the model to SS3 and improvements related to SC20 concerns, the SS3 diagnostic model starting in 1952 estimated population scale, stock status and historical trends that are very similar to the 2024 MFCL diagnostic (Figure 31) and the two previous MFCL assessments (Ducharme-Barth et al., 2019; Davies et al., 2012). Follow-up discussion with several technical experts from the PAW (Nicholas Ducharme-Barth, Simon Hoyle and Nick Davies) on the issues of the rapid early stock decline and sustained low biomass did not lead to a satisfactory resolution of these issues. An experiment on the diagnostic model where the size data were severely downweighted, while leading to better CPUE fits lead to only a minor increase in estimated biomass and very similar recruitment levels (Figure 49).

It was generally agreed that the rapid decline to low depletion levels warranted further investigation, given the localised and short-lived early catch spike. Furthermore, apart from the high catch year in 1954, landings in numbers have on average been quite similar from 1950s to mid-2000s, and only recently starting to show some declining trend (Figure 4). The plausibility of the biomass scale remained highly questionable in the SS3 model. Growth was suggested as potentially a part of the problem, with the current growth information being potentially biased, based on low age-length samples, and not representing the broader stock region. Growth and M are linked when using the Lorenzen M-at-age formulation, therefore if growth is miss-specified so to would M. Steepness is also highly uncertain, but might be expected to be higher than the 0.8 used in the diagnostic model (Brodziak et al., 2015). To explore the sensitivity to these biological uncertainties we ran a series of models on growth parameters (L1, L2 and K), fixed M-at-age simply scaled higher and lower than the diagnostic model M-at-age, and steepness (h). Other sensitivities were run on the alternative CPUE indices applied to the diagnostic model biology. The list of sensitivity models is provided in Table 2.

7 Grid models

One-off sensitivity models can often highlight influential uncertainties that then need to be captured in an uncertainty characterisation for management advice using a model ensemble or uncertainty grid. One-off sensitivities may not adequately capture the model uncertainty as interactions among biological assumptions could result in important changes to model estimation and performance not detected in one-off sensitivities. In this case, given the biomass scale plausibility concerns, running a model grid was conducted as an exploration to see if certain combinations of biological assumptions had interesting impacts on the model behaviour and estimation, and could perhaps provide more plausible results to explore further. A simple orthogonal grid on: growth parameters L2 (x3), K (x3) (noting the sensitivities suggested the model was not sensitive to changing L1 from 50 - 100 cm, so L1 was fixed at 50 cm for all models and Age 0 set to 0.5), M-at-age (x3) and Steepness (x3), resulting in 81 models (Table 3). The M-at-age options are displayed in Figure 11, the high and low M options are simply scaled up and down from the diagnostic model option. The growth options; K 0.2, 0.5, 0.8, and L2 195, 210, 225 cm bracket the range of values in published studies reviewed in the appendix associated with Brouwer et al. (2023) (see Figure 12). The diagnostic model with internal estimated growth is similar to K 0.5, and L2 210. Steepness options; 0.75, 0.80, 0.95 are based off the right skewed steepness distribution for the 2024 MFCL assessment (Castillo-Jordán et al., 2024).

8 Results

8.1 Diagnostic model

8.1.1 Convergence diagnostics

The diagnostic model had a maximum gradient of $4.02 \times 10e^{-6}$ with a positive definite Hessian. There were two soft-bounded parameters at bounds, these were F estimates from 1952-1953, and there were four NO-MOVE parameters. Rerunning the model with the NO-MOVE parameters fixed and extended bounds for the other two early F parameters had negligible influence. The 1952–53 F parameters hitting soft-bounds is not concerning since they are initialization parameters with no implications for the management quantities. A jitter analysis with 50 replicates was not able to find a model with a ‘notably’ better objective function (Figure 13). However, of the 50 jitter models, only 16 models were retained for plotting that were adequately converged. Multiple jitters found similar log likelihood values to the diagnostic model, 10 with slightly better log likelihood (<1 log likelihood points). The spawning biomass trajectories for these models were all very similar, with more divergence for the models converged to worse likelihood levels (Figure 13). Overall, these results suggested that the diagnostic model was close to the global minimum, but this minimum may not be well defined.

8.1.2 Fits to standardised CPUE: index fishery

The model fits to the index fishery standardised CPUE data were an improvement in the SS3 diagnostic model (Figure 14). The model followed the major trends in the observed CPUE, and fit the recent CPUE trends well, but it struggled to track some short term fluctuations, such as those around 1990 and the early 2000s. The distribution of residuals is reasonable, with no sustained bias, although there is a tendency for the model to overestimate relative to the observations from the late 1990s to around 2010 (Figure 15).

8.1.3 Fits to size composition data

Lengths: The length data, despite the additional processing, remained noisy with substantial spikiness in both the lower and higher length components, which the model will not be able to fit (Figure 16). There are also peaks in small fish for Japan longline in region two (F02) and the Chinese Taipei longline in region 4 (F05) that the model cannot fit well, which is evident from the residual plots (Figure 17), especially for the Japan longline (F02). The fits to the main components of the length distributions are however, reasonable, noting the limitations of the data. The fits to the mean lengths over time are really only worth considering for the New Zealand recreational data (F10) and New Caledonia longline (F16). As with the Japanese and Chinese Taipei data, there is a tendency to underestimate the frequency of smaller fish, with generally larger residuals, but estimated mean lengths are reasonably consistent with the observations, with better fit for the NZ recreational data (F10) (Figure 17).

Weights: With the exception of Japan longline (F01) the weight data has much smoother aggregated distributions due to the much larger sample sizes, and the aggregate fits are very good for all fisheries with weight data (Figure 18).

The fits to the mean weights over time for the Australian longline (F06, F07) and recreational fisheries (F09) are also reasonable, and there are no strong patterns in the residuals, although there is some underestimation in the frequencies of smaller fish for F06 (Figure 19). Similarly, fits to the mean weights over time for the NZ recreational fishery (F10) are quite good with no residual patterns. Fits to the the Japaneses data (F02, F03), while capturing the mean weight trends to a degree, are worse and show a tendency to underestimate the observed frequencies of the smaller fish (Figure 20).

8.1.4 Fits to CAAL data

There are limited CAAL data available, and most samples are from the two recreational fisheries, Australia F09.REC.AU.3 and New Zealand F10.REC.NZ.3. Given the limited sample sizes the fits are reasonable, but as would be expected with the low sample numbers, the observed standard errors are lower than the expected (Figure 21).

8.2 Additional Diagnostics results

8.2.1 Likelihood profiles

The likelihood profiles of the R0 population scaling parameter (Figure 22) indicates a substantial influence of the recruitment penalty (σ_R) over the other data sources (plot e). This is cause for concern and suggests that recruitment deviations are important in the overall model fit compared to the other data sources. However, they are not in conflict with other data components, which suggests that they are not driving the total likelihood away from the R0 preferred by the data components. None the less the likelihood profile indicates the recruitment deviations from the SRR are attracting too much of a likelihood penalty. This could be related to some mis-specification in stock productivity. If productivity is mis-specified the model needs more/larger recruitment deviations to make up the difference in the fit. This causes the recruitment component to have a larger likelihood weight which can overpower the other data components. There is a similar result presented in the R0 profile for the western and central north Pacific striped marlin assessment (Figure 15, in ISC-BWG (2023)), which also used a σ_R of 0.6. We did a simple test of sensitivity to σ_R by applying a higher (double) σ_R to reduce the influence of this recruitment penalty (Appendix: Figure 50). However, as noted the size data and the recruitment penalty both prefer a similar R0, and CPUE slightly higher, so we would not expect applying a higher σ_R to substantially influence the population scale, and this is what was observed from this test.

Profiles of the weight data by fishery (plot c) indicated that the NZ recreational and Australian longline data were most influential, and that there was no apparent conflict in the weight data. For the length data (plot d), again there was no conflict, with the NZ recreational data having the main influence, followed by F01 (Japan longline in sub-region 1) and F05 (Chinese Taipei longline in region 4).

The profile for the data sources (excluding the recruitment penalty) (plot b), showed the weight data is having by far the most influence on the R0. But unlike the MFCL2024 assessment, the length and weight data do not conflict. The index prefers higher R0 than the size data (see also plot a), so some data conflict between size data and CPUE remains. Overall the data conflict appear to have been moderated in the revised SS3 diagnostic model.

We explored a diagnostic model that more severally downweighted all the size data (x10 more than diagnostic) (Appendix: Figure 49). This did improve the fit to the CPUE, and did increase biomass, as expected, but only slightly, and recruitment levels did not change.

8.2.2 Age-structured production model and catch curve analysis

The age-structured production model (ASPM) was fitted (Figure 23), without using the estimated recruitment deviations from the diagnostic model, so that the population scale is being estimated, based on the catch and CPUE data. The ASPM could not fit the CPUE time series, and therefore could not produce a sensible biomass trajectory. This suggest that other information, such as the recruitment penalty is overwhelming the CPUE, i.e. in the ASPM recruitment is the mean of the stock recruitment relationship, so can not deviate.

Further experiments on the ASPM were run whereby the recruitment deviation was freed-up, by increasing σ_R to 1. This model was able to estimate a small population scale and fit the CPUE index. The conclusion being that the CPUE index is being overwhelmed in the likelihood by other components. There production function in the SS3 model appears to be poorly determined which may relate to the model being mis-specified and/or too constrained to produce a well determined production function.

8.2.3 Retrospectives and hindcasting

Each of the retrospective peels on spawning biomass (SB) shows some positive bias (Figure 24), suggesting the model may be underestimating the current stock size or overestimating the recent fishing impact. However, the average Mohn's ρ is 0.102 and indicates the bias is not a significant issue, with each peel remaining with the 95% asymptotic confidence interval of SB except for the 2020 peel that had a stronger bias ($\rho = 0.458$) indicating sensitivity to 2020-2021 data. In contrast there is negative bias in F, but with an average Mohn's ρ of -0.165, suggesting a minor bias for F estimates to decrease for each data peel. The strongest F effect is also for the 2020 peel (Mohn's $\rho = -0.337$). Similar to the retrospective analysis the hindcast analysis result (MASE 0.79) indicated that prediction accuracy is good (Figure 24).

8.3 Diagnostic model population dynamics estimates

8.3.1 Growth and natural mortality

The M-at-age in the SS3 diagnostic model is essentially the same as that used in the MFCL diagnostic model (Figure 11), based on a Lorenzen curve with an average M=0.36 year^{-1} (i.e. maximum age 15 years 5.4/Amax (Cope and Hamel, 2022)). Growth estimates from the diagnostic model, where growth is estimated internal to the model, are however, quite different from the external fitted growth curve that was applied in the MFCL diagnostic model which indicated a more rapid growth until age 3 years with a very high K=0.836. The internal estimated growth had a K=0.438, more similar to the growth used in the 2019 assessment Figure 12.

8.3.2 Recruitment and stock recruitment relationship (SRR)

The recruitment estimates for the diagnostic model show a downward trend for the first 10 years of the time series until the mid-1960s, when the trend flattens, but with high interannual variabilty (Figure 25). This initial downward trend appears influenced by the higher catches in the early period, and to a degree the declining trend in the NZ recreational weight data. We ran a model without the New Zealand recreational weight data, and the initial downward trend remains, but the variability is removed and recruitment uncertainty is increased, resulting in increased uncertainty in early spawning biomass (Figure 26). Apart from the positive bias in the first decade of the time series, recruitment deviates show no clear long-term trends. With the limited age composition data and lack of long-term time series of quality size data, the recruitment dynamics estimated by the model show a lagged correlation to the catch dynamics (i.e. recruitment lagged forward 2 years in relation to catch in numbers r^2=0.42) (Figure 27), suggesting that recruitment estimates are likely sensitive to the catch data. The average recruitment post 1965 is 44% less than the average recruitment prior to 1965, which may have an undue influence of the stock recruitment relationship. This model maintains the lower average recruitment throughout the 50+ year period since 1965.

For the diagnostic model (0.8 steepness), the estimated recruitment for most of the model period are associated with low spawning biomass on the descending limb of the stock recruitment relationship (Figure 28). There is high estimated recruitment variability at low spawning biomass, with some relatively high recruitment also estimated at low biomass. A high steepness is required to produce the recruitment to sustain the catches at low spawning biomass, and the model does not indicate a strong stock recruitment relationship.

8.3.3 Biomass and relative biomass depletion

Note this is a single sex model so spawning biomass (SB) refers to all fish of spawning age, i.e. it is not a specific female spawning biomass, which might be assumed to be roughly around 50% of the SB reported in this paper. The SB definition for the SS3 model is comparable to that for the single sex MFCL2024 model. SB and total biomass are estimated to decline rapidly at the beginning on the fishery (Figure 29), which started abruptly in 1952 when fishing restrictions (the MacArthur Line) were lifted allowing Japanese longline fleets to expand operations into the south Pacific. It is reasonable to assume that the stock was very lightly to unfished prior to 1952, but the catch increases to it's historical peaks within two years (Figure 3), unlike the gradual increases typically seen as most fisheries develop. The total biomass declines from around 30,000 t in 1952 to around 15,000 mt in 1960, and SB from around 20,000 t to about 6,000 t over that same period (Figure 29). The estimated biomass further declines until around the 1970s after which its remains low, with some increased periods during the early 1980s and late 1990s-2000s. By the end of the model period total biomass is estimated at around 7,500 t and SB is around 1300 t (perhaps around 600-700 t female SB).

The dynamic spawning biomass depletion $SB/SB_{F=0}$ shows a similar steep decline through to the 1970s, and after 1980 fluctuates between around 0.1 to 0.2 (Figure 30).

Comparison of the SS3 diagnostic model with the MFCL2024, 2019 and 2012 diagnostic models (Figure 31) shows virtually the same spawning biomass (including fished and unfished), depletion and recruitment dynamics for the overlapping time period from 1979. The recruitment quantities are slightly different but of similar scale, and appear more variable in the SS3, perhaps due to being calculated at different ages, however the mains trends, in particular the early period decline, are consistent. The absolute values of the biomass measures and spawning biomass depletion are remarkably similar given the changes made in revising the assessment, size data inputs and changing software.

8.3.4 Fishing mortality

F was calculated as the weighted (by numbers) mean across ages 2-10 years. F is estimated to be below 1 for the entire model period (Figure 32). The trajectory of F increased from the start of the model period until around 1990, after which it flattens, but is variable between 0.4–0.8. The dynamics of F over time are generally consistent with the 2024 MFCL model, including the periods of highest high F (Castillo-Jordan et al., 2024). We note that we still have some uncertainty in the approaches used to calculate F in SS3 in comparison to MFCL, and how other SS3 billfish assessment use these F settings. We understand that we have used the settings as per the western and north central Pacific striped marlin assessment, but the average is over ages 2-10 years, compared to age 3-12 years (ISC-BWG, 2023). We would caution against direct comparisons of absolute F values between the SS3 and MFCL2024 models until we fully understand any differences in calculation methods.

8.4 Sensitivity analyses

Sensitivities were run on the main biological uncertainties and alternative CPUE index times series. High level diagnostics (gradient and Hessian) indicated that some models did not converged (Table 2). We did not do additional work on the non-converged sensitivities, but these are noted in the table.

CPUE abundance indices: The sensitivity analyses for CPUE did show that the recent model period is sensitive to the choice of CPUE (Figure 33). The CPUE series start in 1979 (JP/TW longline), 1975 (NZ recreational), 1998 (Australian longline), 2001 and 2002 (PICT observer longline NC-FJ-TO and PF), and they have no real influence on the model estimation prior to their initiation. The model sensitivity to choice of CPUE time series occurs after 1995. Spawning biomass depletion $SB/SB_{F=0}$ is considerably more optimistic for the NC-FJ-TO index, followed by the Australian index, then the JP/TW index (diagnostic model). Similar patterns are seen for spawning biomass, and the model with the NC-FJ-TO index estimates estimated spawning biomass at least three time higher than the JP/TW diagnostic model (Table 2). The PF and NZ CPUE indices result in the most depleted status (highest depletion estimate) and lowest spawning biomass, noting both of these indices do not show the recent increase in CPUE. With exception of poor fit to early years of the AU longline index, fit to the various CPUE was reasonably good for most models (Figure 34). The AU and NC-FJ-TO indices are the most consistent with each other, although all indices except PF and NZ recreational appear to indicate a recent increase in CPUE that their respective models fit well.

Growth: The sensitivity models on growth parameters indicated that the L2 (i.e. max length at age 10 years) and K parameters were the most influential on the models estimates of $SB/SB_{F=0}$, which was highly sensitive to these growth parameters. The L1 (length at 0.5 year, 50cm v 100cm) parameter had negligible influence so we only include the L1=50cm models in the results (Figure 35). All models with L2 = 225cm failed to reach sufficiently low gradient and/or achieve a positive definite Hessian (PDH) to be considered converged, likewise the model with K=0.2 and L2=195cm (Table 2). Irrespective of convergence, spawning biomass was not strongly sensitive to the growth parameters and the general patterns are that higher L2 (225 cm) for a given K, results in lower spawning biomass, and a more depleted stock status and the opposite effect occurs for the lower L2 (195cm) (Table 2, Figure 35). In terms of K, spawning biomass depletion shows minor sensitivity, with slightly less depleted status for models with K=0.5 compared to 0.2 and 0.8. Based on these sensitivities the high L2=225cm produces the most unlikely scenarios of low spawning biomass (less than 1,000 t in 2022 for all K values) and is not supported by the observed age at length data, either the external estimate (L_{inf}=208cm) or the internal estimate (L2=215cm). The lowest L2 had poor convergence with K 0.2, but none the less, consistent with K 0.5 and 0.8 it estimated the highest spawning biomass, and least depletion (Figure 35, Table 2).

Natural mortality: $SB/SB_{F=0}$ showed sensitivity to the alternative M-at-age scenarios (Figure 36). Assuming a higher M-at-age, the estimated $SB/SB_{F=0}$ was noticeably less depleted, with very slightly higher spawning biomass in 2022 than the low M and diagnostic (intermediate) M-at-age (Table 2).

Steepness: $SB/SB_{F=0}$ showed some sensitivity to the steepness assumption with low steepness resulting in a more depleted status (Figure 36). Spawning biomass showed very little sensitivity to steepness (Table 2, Figure 36). The lack of sensitivity of spawning biomass to steepness was surprising given steepness is a key component of productivity.

8.5 Exploratory grid

Time series: Of the 81 models from the biology grid, 41 models with sufficient convergence (gradients ≤ 0.001 and a PDH) were retained for plots and the results summary table (Table 4).

Integrated across the 41 converged biology grid models, median $SB/SB_{F=0}$ declined rapidly until around the 1970s, after which it stabilises, with a more gradual decline to a median of 0.22 in 2022 (10%ile 0.06, 90%ile 0.65) (Figure 37, Table 4). However, there is large uncertainty across the grid models with $SB/SB_{F=0}$ ranging from less than 0.10 to around 0.80 at the end of the model period. However, the majority of models (i.e. 30 models or 71%) are grouped at more depleted $SB/SB_{F=0}$ from around 0.30 to 0.10 (Figure 37).

The spawning biomass had a median value for 2022 of 1,304 t (10%ile 670 t, 90%ile 6,611 t), which was similar to the diagnostic model (1360 t) (Figure 38). There is very large uncertainty for the initial spawning biomass across the grid.

Recruitment was mostly in the range of 100,000–500,000 individuals $year^{-1}$ (Figure 39). There is large uncertainty in recruitment estimates depending on the biological assumptions, with some annual estimates well over double to three times the median, but the distribution is skewed towards the lower part of the range.

F had a median value for 2022 of 0.11 (10%ile 0.05, 90%ile 0.37 mt) (Figure 40). Median F is lower than 0.3 across the time series, but there is a large range, for example F_{recent} ranges from 0.05–0.99. F is highly variable from year to year and shows a generally increasing trend, with a decline in the most recent years. F/F_{MSY} displays a similar trend as F, and shows that the upper ranges exceed 1 over much of the time series, but with the median value mostly less than 1 (Figure 40).

The plots of $SB/SB_{F=0}$ showing the individual models grouped according to the growth K parameter (Figure 41) show that models with K 0.2 stand out as those with the most depleted status.

The plots of $SB/SB_{F=0}$ showing the individual models grouped according to the growth L2 parameter (Figure 42) show that the models with L2 195cm stand out as less depleted than those for 210cm and 225cm, with the latter value indicating the most depleted status.

The plots of $SB/SB_{F=0}$ showing the individual models grouped according to the M-at-age (Figure 43) show some clustering of the base (diagnostic) and high M, but the low M models generally favour a more depleted status.

The plots of $SB/SB_{F=0}$ showing the individual models grouped according to steepness (Figure 44) show no clear clustering according to the assumed steepness value.

Influence of biological axes: Box plots of $SB_{recent}/SB_{F=0}$ show that across the grid models (Figure 45) the recent stock status is estimated as more depleted assuming K 0.2 and less depleted assuming L2 195cm. For M-at-age, the high M has a larger range of depletion with some clear outlier models, and favours a less depleted status than the base and low M. For steepness, the higher steepness tended to result in a more depleted recent status, which is unusual and may relate to model filtering rather that actual influence of steepness per se.

Box plots of SB_{recent}/SB_{MSY} show that across the grid models (Figure 46) the recent stock status is estimated as more depleted assuming K 0.2 and less depleted assuming L2 195cm, consistent with $SB_{recent}/SB_{F=0}$. For M-at-age, the pattern of lower M showing greater depletion is also consistent with $SB_{recent}/SB_{F=0}$. For steepness, the higher steepness included the models with the least depleted status, and overall high steepness resulted in slightly less depleted status.

Box plots of F/F_{MSY} for the grid models (Figure 47) show that estimates of F/F_{MSY} are most sensitive to the growth L2, with low L2 showing lower estimates. There is also sensitivity to M, with higher M having lower F/F_{MSY}.

8.6 Kobe Plot

The Kobe plot (noting it is based on equal possibility assumed for all models, which is likely not the case, i.e. L2 195 cm and 225cm seem unlikely) shows a wide spread of recent stock status estimates. The median point is in the over fished region and just outside the over fishing region (Figure 48).

However, we note there are strong caveats to the reliability of the management reference metrics when the biomass scale is very uncertain and likely to be considerably underestimated. These concerns are discussed further below. Furthermore, given the concerns around the biomass scale uncertainty, the model grid was not specifically designed as grid (or ensemble) for management advice. Models in the exploratory grid may need consideration of down weighting as some of the biological assumptions are less likely than others.

9 Discussion

9.1 General remarks

The revision of the 2024 southwest Pacific striped marlin assessment requested by SC20 has involved moving the assessment from MFCL to Stock Synthesis along with numerous other changes and data modifications in this process, various of which would have also likely occurred had the assessment been revised in MFCL. Some key changes (including the switch to SS3) were based on the workshop in Hawai'i with NOAA scientists, who continued to collaborate and support the SPC assessment scientists after that workshop. Other changes were the result of discussions at the SPC Pre-Assessment Workshop (PAW) (Hamer, 2025) and OFP's regular stock assessment team meetings. Some of the more important changes included:

- Switch from MFCL to SS3.
- Start the model in 1952, and include early catch uncertainty (i.e. designate early period Japan LL fishery).
- Implement size based selectivity and modifications to selectivity functions (i.e cubic splines in MFCL24, double normal or logistic in SS3).
- Major revision of size data inputs and refined size data filtering (apply stronger filters), resulting in significant reduction in size data inputs.
- Separate NC and PF longline fleets from the mixed flag fisheries and create separate fisheries due to their different length data distributions.
- Use of ad hoc size data weighting to down weight the influential size data.
- Use CAAL data to estimate growth internally for the diagnostic model, which produced

a lower K, similar to previous 2019 assessment, and considerably lower than the external growth curve used in the MFCL2024 model.

- Explored sensitivity to alternative CPUE indices, including Australian longline, NZ recreational and newly developed observer based longline indices for PICT fleets (Neubauer, 2025).

We suggest that the revised assessment has addressed several concerns from SC20, including:

- Improving fits to the CPUE abundance indices while also achieving good fits to the remaining higher quality size data through implementing size based selectivity.
- Reducing the influence of the size data on model estimation by reverting to ad hoc data weighting, plus detailed review of size date inputs and stronger filtering to remove unreliable size data.
- Reduced data conflicts.
- As recommended by SC20, starting the model in 1952 when the initial condition could be reliably assumed to be unfished.
- Explored alternative CPUE abundance indices.

However, others issue were raised through the revision process, most notably the concerns around the rapid drop in biomass when starting the model in 1952, and the perceived low population size (scale) since the 1970s. These concerns were raised through the PAW process, and were not specifically noted as issues at SC20. The concerns are equally relevant to previous SWPO striped marlin assessments that estimate similar population scales. In this respect the review work on this assessment has relevance to previous SC accepted SWPO striped marlin assessments that were used for management advice by WCPFC.

Ultimately despite the improvements to develop the revised assessment in SS3, the outcomes in terms of the key stock status and population metrics are very similar to those from the MCFL2024 assessment, and the previous MFCL assessments in 2019 and 2012. The population scale is consistent across all three assessments (Figure 31). While this at first seemed remarkable, given the extent of changes across three assessments, including the software change, the core data for the assessments that did not change between the SS3 and MFCL2024 assessment are the catch and diagnostic model CPUE abundance index (at least since 1979), although the JP/TW CPUE index was not the exactly same as the JP only index used in the MFCL2019 diagnostic model. Other consistences, at least between the MFCL2024 diagnostic model and the SS3 diagnostic model were the M-at-age and steepness. It appears that the recent model estimates of stock status and population scale were insensitive to other key differences between the different diagnostic models. The Bayesian Surplus Production Model (BSPM) in the companion assessment paper (Ducharme-Barth, 2025) provides a useful addition to the integrated assessment. It allows more flexibility in the estimation of a production function and removes all size data. The BSPM ensemble however also shows a tendency for estimating low population size, but has a larger uncertainty range towards higher levels compared to the SS3 models integrated across the biology grid. The spawning biomass (single sex) scale in the SS3 model was relatively insensitive to the biological uncertainties tested (i.e. all models were less that 6,000 t spawning biomass across recent period 2019–2022, with a median of 1,070 t). Starting the model in 1952 removed the need to consider uncertainty in the initial fishing mortality, however, the MFCL2024 model starting in 1979 showed only minor sensitivity to that assumption (Figure 45 in Castillo-Jordan et al. (2024)). Starting the model in 1952, while a preferred approach, does not appear to have had an important influence on the stock status estimates for the recent several decades (Figure 31). None the less the consistency in the outcomes across software packages is reassuring.

While this revision has not lead to the material changes in biomass scale or stock status estimates, that might have been expected, it has allowed a much deeper look into this assessment, which has raised concerns that are possibly more important than those raised at SC20. There was also significant added value from shifting the SWPO striped marlin assessment to SS3, as this was a request by SC20 for the next assessment (2029) and as part of WCPFC Project 123 to move all billfish assessments to this software. The transition to SS3 has required additional work, and we are grateful for the support from the US delegation (NOAA, Pacific Islands Fisheries Science Center) with this transition. Moving the assessment to SS3 has helped build capacity, knowledge and tools in the OFP assessment team for working with the SS3 platform and its various supporting packages for future assessments. It has also pointed out areas that require additional investigation and clarification from SS3 technical experts in relation to appropriate model settings. While SPC has staff with an incredible depth of knowledge on the inner workings of MFCL software to enable highly informed optimisation of the software settings, interpretation of outputs, issue recognition and resolution (i.e. Nick Davies and John Hampton), the same deep technical knowledge of the SS3 software is not available in-house. Transitioning software will inevitably have teething issues and a learning curve, and this needs to be factored in to the expectations of staff resources and time required to transfer assessments, even for a commonly used package such as SS3.

9.2 Key results summary

Diagnostic model: The main results from the revised diagnostic model in SS3 are broadly consistent with previous MFCL assessments (Langley et al., 2006; Davies et al., 2012; Ducharme-Barth et al., 2019; Castillo-Jordan et al., 2024), in that the recent estimated population size is low (i.e. 1,363 t sex aggregated spawning biomass, and approximately 7,000 t total biomass in 2022). The model estimates recruitment that is sufficient to maintain the catches at what seems a very low spawning stock size for the SWPO (i.e likely around 700 tonnes of female spawning biomass), implying high recruitment is required at low stock size. Given the model structure, population dynamics and observation model equations and biological information, it is able to fit the observations with a low stock size. One might question the validity of the statistical model. The high influence of the recruitment penalty in the likelihood profile was a concern in that it was driving the population scale and overwhelming the other data components. However, the penalty was not in conflict with the other data, and releasing the penalty did not result in notable changes to biomass levels. We also ran a model more severely downweighting the size date to give the CPUE more influence. This clearly improved the CPUE index fit, but the increase in biomass was still relatively low. The ASPM model diagnostic also suggested that the SS3 model production function was poorly determined and that population dynamics were recruitment driven. The BSPM model was able to obtain a production function and good fit to the CPUE with the same catch and CPUE index (Ducharme-Barth, 2025), which suggest there maybe some ongoing model misspecifications in the SS3 model or other features that overly restrict the flexibility of the model to explore the productivity parameter space and identify a well determined production function. The low population scale remains the key concern with interpreting the results of this revised striped marlin assessment.

Sensitivities and biology uncertainty grid: Given the production function concerns, the fixed productivity assumptions (i.e. steepness) and uncertainty in biological assumptions, the analysis of sensitivities and the biological uncertainty grid were conducted to explore how alternative assumptions influenced model estimates including population scale. While there was some sensitivity to the uncertainties explored, none of them moved the population scale an order of magnitude into a different, perhaps more plausible realm, with the highest estimate of spawning biomass across models in 2022 being 7,571 tonnes, 90^{th}% 6,611 tonnes. This would still seem low at the scale of WCPO (i.e. <4,000 t female spawning biomass). Fishing mortality, recruitment levels and depletion, $SB/SB_{F=0}$, were however quite sensitive to the alternative biological assumptions.

The growth parameters, especially K, were considered a problematic biological uncertainty in the MFCL2024 assessment (which assumed faster earlier growth). However, this concern was initially around model structure not being able to estimate the small size classes in the size data using age based selectivity and an annual model. Moving to SS3 length based selectivity, while also maintaining an annual population/recruitment model helped with size data fits. However, although the growth curve was considered the best available growth information in 2024, it was likely biased due to samples coming from fishery catches and targeted sampling of otoliths from length bins. Estimating growth internally using the age at length data in the model was seen as a way to account for selectivity effects on the data and mitigate some of the likely biases. This did change the K to a lower value similar to the previous assessments. None the less in terms of biomass estimation the model estimates were more sensitive to choice of the L2 parameter (length at model max age) than K.

Although biomass scale did not change dramatically with different growth parameters, MSY based reference points and $SB/SB_{F=0}$ were highly sensitive. Reliable and representative growth information is essential, as these reference points will still be sensitive to growth assumptions if the population scale is different. Improved growth information is essential for reducing uncertainty in striped marlin assessments, noting we still only have around 80 age length samples for the species in the SWPO.

It is likely that by assuming a single region and single sex with consistent biology, the model is too simple and does not capture the underlying spatial complexities of the stock and the sex specific biology that appears to be a feature of striped marlin (Farley et al., 2021). However, this is the default approach when biological data (including sex specific biology) and population structure information is limited across the assessment region. It was interesting that the PF CPUE index was not consistent with the Australia and NC-FJ-TO observer CPUE indices, particularly in not showing the recent increase in CPUE. This may indicate different dynamics in that region, potentially related to an alternative recruitment source. In the south Pacific, the historic Japanese larval survey data suggest there are two spawning areas, one in the Coral Sea region off north east Australia and one just to the north of French Polynesia (Ijima and Jusup, 2023; McDowell and Graves, 2008).

The most recent comprehensive study of striped marlin genetics found similarity between samples from New Zealand and Australia but unfortunately did not include samples from French Polynesia or other regions of the southwest Pacific (Mamoozadeh et al., 2020). There may be aspects of sub-regional fidelity and biological variability of striped marlin that are not currently well described (Domeier, 2006), or perhaps even more wider meta-population structure and connectivity with other with areas outside the SWPO model region as suggested in recent work by Martinez (2021). A spatially structured model might be more appropriate, also because most of the early fishing pressure was in sub-region 2, the region of the Coral Sea and the southern PICTs, identified as a spawning area (McDowell and Graves, 2008). The 2019 assessment (Ducharme-Barth et al., 2019) conducted a preliminary attempt at a spatially structured two region model and noted the model estimates were highly sensitive to the assumptions made for movement and recruitment partitioning between regions.

The issue of plausibility: The population scale plausibility concern has not been raised by SC for the previous assessments but now presents a key discussion point for this assessment, particularly in relation to management reference points and related advice. The estimated population scale is consistent with previous integrated model estimates, which fit to similar indices and assumed similar catch and fixed-levels of productivity (e.g., natural mortality and steepness). It is challenging to reconcile the small population scale in the context of recent observed catches (approx. 1,400mt) and the assessment spatial scale. The area of the assessment is roughly 40 million km^2. The estimated level of total biomass (i.e. assuming median from the grid models of around 9,000 t in 2022, which is higher than the diagnostic model) would imply around 0.225 kg of biomass per km^2 of ocean surface area. At an average harvests weight of around 80 kg per fish, this is about 1 fish per 350 km^2. Given a typical catch rate of 1-2 fish per thousand hooks (and noting most of these hooks are depth where striped marlin are not very vulnerable), which could be spread over 100 km, it is reasonable to conclude that the biomass estimates are implausibly low to achieve the catch rates, even if the longline caught every striped within 3 km of the main line.

Though the biomass scale does appear low, the model does what it can to explain the catch dynamic in the face of a relative long-term stationary trend of the CPUE, through the estimated recruitments. The correlation of the recruitment with catch, the failure of the ASPM diagnostic, the qualitative similarity of CPUE and total catch dynamics (Figure 52) indicate the model population dynamics are driven by recruitment variability, and the stock-recruit function is poorly determined.

Development of the Bayesian surplus production modeling (BSPM) approach (Ducharme-Barth, 2025) was motivated to investigate these issues and see if a simplified modeling approach could allow for the identification of a production function, and more effectively characterize uncertainty in estimates of population scale. By distilling productivity and fishing assumptions into a restricted parameter subset, BSPMs facilitate a tractable approach for more extensive exploration of model assumptions. Additionally, incorporation of priors in a Bayesian framework assists estimation by constraining the parameter space to more plausible outcomes, allowing for a more explicit characterization of uncertainty while maintaining computational efficiency and interpretability. While overall results across the SS3 model sensitivity grid appear broadly consistent to what is indicated by the BSPM approach (Ducharme-Barth, 2025), the inability of the SS3 model to achieve a well determined production function, when it is indicated by the BSPM, is cause for concern. It is possible that the SS3 model is too constrained. In the BSPM all productivity and scale parameters are estimated with priors. This gives the model the flexibility to capture general trends in the data where the SS3 model might struggle, especially since stock productivity is fixed (e.g. natural mortality and steepness are not estimated). If productivity is fixed at the wrong level then it will not be able to reconcile changes in the catch and CPUE index. Accordingly, with fixed productivity the SS3 model's main flexibility in fitting the data is by adjusting the recruitment which may contribute to the appearance of a recruitment driven stock.

The SS3 model as currently configured appears to be estimating the stock as being small and highly productive. This result may be driven by some of the fixed underlying model productivity assumptions, particularly related to how recruitment is generated through the stock recruitment curve. Though potential mis-specification in the fixed productivity assumptions were investigated through the sensitivity analysis, it is possible that the range of uncertainty in productivity considered was insufficiently broad, or that penalties on recruitment deviations from the stock recruit relationship overwhelm the model likelihood. Indeed, an exploratory ASPM of a simplified SS3 model fitting to the same catch and index data without recruitment deviates, was able to fit the index while estimating a larger population scale and only estimating R0, natural mortality, and steepness (all constrained with priors). This exploration also highlighted the negative correlation between scale (R0) and productivity (steepness), indicating that constraints placed on productivity will impact the estimate of scale.

Determining the cause of the low biomass scale has been a frustrating process, and though the exact driver(s) remains undetermined the evidence appears to indicate that mis-specification in the SS3 model productivity assumptions may play an important role. On this point it is worth noting some interesting similarities between the issue discussed above and the most recent western and central north Pacific striped marlin assessment, including; ASPM failure, influence of recruitment penalties, recent low total biomass <10,000 tonnes, female spawning biomass around 1,000 t (ISC-BWG, 2023). We suggest the ISC and SPC assessment teams have informal discussion to compare the issues and similarities between these two assessments, now that both are being conducted in SS3. While the causes of these issues need further investigation, it appears evident that building intuition of model dynamics by progressively increasing model complexity in a principled Bayesian workflow can be a useful approach applied to both stocks.

9.3 Assessment conclusions

The assessment conclusions are problematic in relation to the reliability of the management reference points. The estimated Beverton-Holt stock recruitment relationship (SRR) and subsequently, the spawner per recruit and equilibrium population calculations that underpin the estimation of MSY, are scaled by the estimated absolute recruitments. Uncertainty in the parameters determining the total population scaling, usually an estimated mean recruitment level, will directly affect the uncertainty in the estimated MSY. This means that high uncertainty in absolute abundance scale is typically accompanied by similar levels of uncertainty in MSY and SB_{MSY}. If we accept the low population scale estimated, MSY and SB_{MSY} would be considered highly uncertain, and could be quite different depending on how much higher the true population scale actually is. Therefore the SB/SB_{MSY} reference point must be treated with caution in this case where the population scale is likely higher than the assessment models estimate.

In terms of F/F_{MSY} and $SB/SB_{F=0}$, the implications of biased population scale on these reference point was less clear at first. Intuitively, increasing biomass scale, but maintaining the same catch, will mean reduced fishing impacts. To test this we ran experiments in SS3 and MFCL where we increased the population scale by increasing the mean recruitment and refitted the model with the same catches. As expected the ratio reference points of F/F_{MSY} and $SB/SB_{F=0}$ change considerably indicate a less exploited stock status with an assumed higher population scale (Section 14, Figure 53, Figure 54).

It is clear that uncertainty in population scale will mean uncertainty in the management reference points. This needs to be considered seriously by SC20 when considering how they might interpret outcome of this assessment for management advice. Noting that population scale is suggested to be low, and based on the considerations above, assuming a higher 'true' population scale would result in more optimistic values of the key management reference points. The differences would depended on how much higher the true population scale is. The management reference points are more uncertain than presented here due to the population scale uncertainty, but it is reasonable to suggest that they may be overly pessimistic.

Finally, it is worthwhile to consider the empirical data in relation to the recent trends in stock abundance. Most notable here is the consistent indications of a recent increase in CPUE for the PICT observer (excluding PF), the Australian and JP/TW longline indices. The recent increases in CPUE are confirmed anecdotally (Ashley Williams, CSIRO, Australia) to have continued for the more recent years not included in the index for the Australian fishery. Observations of lots of smaller fish being captured suggest the recent CPUE increase may be indicative of recent strong recruitment.

Notwithstanding the issues and caveats raised over the estimates of population scale, we provide the stock status outcomes for completeness based upon Table 4, using both the MSY-related metrics as standard for billfish stock assessments, and depletion ($SB/SB_{F=0}$). We note the uncertainty analysis (the grid) presented in this assessment was designed to explore the influence of biological parameter uncertainty on population scale, rather than as an uncertainty characterisation for management advice. Importantly, the uncertainty in the population scale that has been recognised as a key uncertainty in this assessment is not adequately captured. Future assessment will benefit from developing approaches to capture a plausible range (or prior) for the population scale if this is not adequately captured within the existing biological and structural uncertainty characterisation.

The result of the assessment can be summarised as follows:

- Spawning biomass relative to the spawning biomass at MSY is highly uncertain, median level of $SB_{recent}/SB_{MSY} = 0.73$ (10th and 90th%iles = 0.29 and 2.29 respectively), and 63% (26) of the 41 models indicated SB_{recent} was $< SB_{MSY}$.
- $SB_{recent}/SB_{F=0}$ had a median = 0.18, but was highly uncertainty (10th and 90th%iles = 0.06 and 0.65 respectively).
- F_{recent}/F_{MSY} had a median = 0.84, but was highly uncertain (10th and 90th%iles = 0.24 and 1.57 respectively), and 42% (17) of the 41 models were > 1.

According to the MSY reference points, the results would imply a higher probability that the stock is not undergoing overfishing, but a higher probability that the stock is overfished.

9.4 Recommendations

- 1. Implement other methods to try and estimate the size of the striped marlin population in the SWPO, the most appropriate proven method would be Close Kin Mark Recapture, building on the work of WCPFC Project 100c. Implementing a CKMR study could also provide improved information on population structure. Explore other methods that might provide a basis for developing a prior on the absolute population size.
- 2. As is routinely recommended from striped marlin assessments, there is a need for improved understanding of critical biological processes such as growth, movement, early life history and population structure across the entire south Pacific, and especially at finer resolution within the southwest Pacific region of this assessment. Importantly, given these models are age structured, and with the fast growth of striped marlin to maturity and maximum size, it is very difficult to obtain reliable age structure from the length or weight data. Age structured models really need age data for species with growth characteristics like marlin and indeed tunas. **Investing in collecting regular age composition data is critical to improving the confidence in the current age stuctured model estimations**. Noting otolith collections are minimal for this species due the difficulty in accessing fish (often headed at sea) and extraction of the otoliths (tiny otoliths in a very big head), exploring epigenetic aging is recommended, not only for striped marlin, but also swordfish.
- 3. Explore a Bayesian approach to the integrated stock assessment.
- 4. Continue to utilize the observer data which has the advantage of additional covariates to develop abundance indices, as these will become more valuable over time.
- 5. Conduct a workshop on assessment approaches and challenges for bycatch species, including billfish and sharks, to consolidate and share learnings and modeling approaches from this assessment revision and the multimodel approaches now used for shark assessments.

10 Acknowledgments

We thank the various fisheries agencies for the provision of the catch, effort and size frequency data used in this analysis, and are grateful for the work done by fisheries observers throughout the region to collect observations and measurements at sea. We are grateful for the support of the team from NOAA Pacific Island Fisheries Science Centre in Hawai'i with the Hawai'i workshop (Felipe Carvalho, Rob Ahrens, Michelle Sculley). We thank Laura Tremblay-Boyer for providing standardised CPUE for the Australian longline fisheries, John Holdsworth for providing the standardized CPUE index for the New Zealand sport fishery, and Phil Neubauer (DragonFly) for the observer CPUE indices. We thank Ashely Williams and Scott Cooper for assistance with the Australian weight data. We thank Meg Oshima (NOAA) for assistance with SS3diags. We greatly appreciate the advice from Simon Hoyle on various aspects of the assessment. Additionally, we thank the data management team at SPC for their work compiling the data inputs for the assessment. We thank the other members of the OFP stock assessment team not in main author list for their support and advice in the regular assessment team discussions. We thank Graham Pilling for review of the paper.

11 References

Anon (2024). Summary Report: Twentieth regular session of the Scientific Committee. Technical report, The Commission for the Conservation and Management of Highly Migratory Fish Stocks in the Western and Central Pacific Ocean, Manila, Philippines, 14–21 August 2024.

Brodziak, J., Mangel, M., and Sun, C.-L. (2015). Stock-recruitment resilience of North Pacific striped marlin based on reproductive ecology. *Fisheries Research*, 166:140–150.

Bromhead, D., Pepperell, J., Wise, B., and Findlay, J. (2004). Striped marlin: biology and ecology. Technical report, Bureau of Rural Sciences, Canberra.

Brouwer, S., Farthing, M., and Hamer, P. (2023). Billfish research plan 2023-2027. Technical Report WCPFC-SC19-SA-WP-16_rev1.

Brouwer, S. and Hamer, P. (2024). Progress against the 2023-2030 Billfish Research Plan - 2024- Rev02. Technical Report WCPFC-SC20-SA-IP-09.

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 Methot, R. D. (2021). A cookbook for using model diagnostics in integrated stock assessments. *Fisheries Research*, 240:105959.

Castillo-Jordan, C., Day, J., Teears, T., Davies, N., Hampton, J., McKechnie, S., Magnusson, A., Peatman, T., Vidal, T., Williams, P., and Hamer, P. (2024). Stock Assessment of Striped Marlin in the Southwest Pacific Ocean: 2024. Technical Report WCPFC-SC20-2024/SA-WP-03_Rev1.

Castillo-Jordán, C., Day, J., Teears, T., Davies, N., Hampton, J., McKechnie, S., Magnusson, A., Peatman, T., Vidal, T., Williams, P., and Hamer, P. (2024). Background analyses for the 2024 stock assessment of SW Pacific striped marlin. Technical Report WCPFC-SC20-SA-IP-06.

Cope, J. M. and Hamel, O. S. (2022). Upgrading from M version 0.2: An application-based method for practical estimation, evaluation and uncertainty characterization of natural mortality. *Fisheries Research*, 256:106493. Publisher: Elsevier BV.

Davies, N., Hoyle, S., and Hampton, J. (2012). Stock assessment of striped marlin (Kajikia audax) in the Southwest Pacific Ocean. Technical Report WCPFC-SC8-2012/SA-WP-05, WCPFC Scientific Committee, Busan, Republic of Korea, 7–15 August 2012.

Domeier, M. (2006). An analysis of Pacific striped marlin (Tetrapturus audax) horizontal movement patterns using pop-up satellite archival tags. *Bulletin of Marine Science*, 79(3):811–825.

Dormeier, M. (2006). An analysis of Pacific striped marlin (Tetrapturus audax) horizontal movement patterns using pop-up satellite archival tags. *Bulletin of Marine Science*, (79):8811–825.

Ducharme-Barth, N. (2025). 2025 Stock Assessment of Striped Marlin in the Southwest Pacific Ocean: Part II -Bayesian Surplus Production Model. Technical Report WCPFC-SC21-SA-WP-07.

Ducharme-Barth, N., Castillo-Jordan, C., Sculley, M., Ahrens, R., and Carvalho, F. (2025). Summary report from the NOAA-SPC assessment model meeting on SWPO striped marlin. Technical Report WCPFC-SC21-SA-IP-15.

Ducharme-Barth, N., Pilling, G., and Hampton, J. (2019). Stock assessment of SW Pacific striped marlin in the WCPO. Technical Report WCPFC-SC15-2019/SA-WP-07, Pohnpei, Federated States of Micronesia.

Farley, J., Eveson, P., Krusic-Golub, K., and Kopf, K. (2021). Southwest Pacific striped marlin population biology (Project 99). Technical Report WCPFC-SC17-SA-IP-11.

Fournier, D., Hampton, J., and Sibert, J. (1998). MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, Thunnus alalunga. *Canadian Journal of Fisheries and Aquatic Sciences*, 55:2105–2116.

Hamel, O. S. (2015). A method for calculating a meta-analytical prior for the natural mortality rate using multiple life history correlates. *Ices Journal of Marine Science*, 72(1):62–69.

Hamer, P. (2025). Summary report from the SPC Pre-assessment Workshop – April 2025. Technical Report WCPFC-SC21-SA-IP-01.

Hamer, P., Schneiter, E., Vidal, T., and Williams, P. (2025). Progress report: Review and reconciliation of size data collected in the WCPFC-CA for stock assessment purposes (WCPFC Project: 127). Technical Report WCPFC-SC21-ST-WP-02.

Holdsworth, J. (2024). Striped marlin catch and CPUE in the New Zealand sport fishery, 2019-20 to 2021-22. Technical Report WCPFC-SC20-SA-IP-17.

Holdsworth, J., Kendrick, T., and Domeier, M. (2019). Characterisation of New Zealand striped marlin fisheries. Technical Report WCPFC-SC-15/SA-IP-16, Pohnpei, Federated States of Micronesia.

Honma, M. and Kamimura, T. (1958). A Population Study on the so-called Makajiki (Striped Marlin) of both Northern and Southern Hemispheres of the Pacific: II. Fishing Conditions in the Southern Hemisphere. Technical Report Report of Nankai Regional Fisheries Research Laboratory, No. 8; pp. 12–21.

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., and Punt, A. E. (2015). Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. *Ices Journal of Marine Science*, 72(1):99–110.

Ijima, H. and Jusup, M. (2023). Tuna and billfish larval distributions in a warming ocean.

ISC-BWG (2015). Stock assessment update for striped marlin (Kajikia audax) in the western and central North Pacific Ocean through 2013. Technical Report WCPFC-SC11-2015/SA-WP-10, Pohnpei, Federated States of Micronesia, 5–13 August 2015.

ISC-BWG (2019). Stock Assessment Report for Striped Marlin (Kajikia audax) in the Western and Central North Pacific Ocean through 2017. Technical Report WCPFC-SC15-2019/SA-WP-09, Pohnpei, Federated States of Micronesia.

ISC-BWG (2023). Stock Assessment Report for Striped Marlin (Kajikia audax) in the Western and Central North Pacific Ocean through 2020. Technical Report WCPFC-SC19_2023/SA-WP-11, Koror, Palau.

Kopf, R. (2005). Population characteristics of striped marlin, Tetrapturus audax in the New Zealand fishery. Master's thesis, Massey University, Palmerston North, New Zealand.

Kopf, R., Davie, P., Bromhead, D., and Young, J. (2012). Reproductive biology and spatiotemporal patterns of spawning in striped marlin Kajikia audax. *Journal of Fish Biology*, (81):1834–1858.

Kopf, R. K., Davie, P. S., and Holdsworth, J. C. (2005). Size trends and population characteristics of striped marlin, *Tetrapturus* audax caught in the New Zealand recreational fishery. *New Zealand Journal of Marine and Freshwater Research*, 39(5):1145–1156.

Langley, A., Molony, B., Bromhead, D., Yokawa, K., and Wise, B. (2006). Stock assessment of striped marlin (Tetrapturus audax) in the southwest Pacific ocean. Technical Report WCPFC-SC2-2006/SA-WP-6, Manila, Philippines, 7–18 August 2006.

Mamoozadeh, N. R., Graves, J. E., and McDowell, J. R. (2020). Genome-wide SNPs resolve spatiotemporal patterns of connectivity within striped marlin (*Kajikia audax*), a broadly distributed and highly migratory pelagic species. *Evolutionary Applications*, 13(4):677–698. Publisher: Wiley.

Martinez, J. (2021). *Stock Composition Of Striped Marlin (Kajikia Audax) In The Central North Pacific Ocean Inferred By Analyses Of Genome-Wide Molecular Markers.* PhD Thesis, Virginia Institute of Marine Science.

McDowell, J. R. and Graves, J. E. (2008). Population structure of striped marlin (Kajikia audax) in the Pacific Ocean based on analysis of microsatellite and mitochondrial DNA. *Canadian Journal of Fisheries and Aquatic Sciences*, 65(7):1307–1320.

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

Mohn, R. (1999). The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. *Ices Journal of Marine Science*, 56(4):473–488.

Molony, B. (2005). Summary of the biology, ecology and stock status of billfishes in the WCPFC, with a review of major variables influencing longline fishery performance. Technical Report WCPFC-SC1-2005/Working Paper EB WP–2, Noumea, New Caledonia, 8–19 August 2005.

Neubauer, P. (2025). Exploring the potential for observer CPUE for southwest Pacific swordfish (Xiphias gladius) and striped marlin (Kajikia audax). Technical Report WCPFC-SC21-SA-IP-13.

Okamoto, H. (2024). Japanese longline fishery under the occupation after World War II: Compilation and primary analyses of fishing data of single longline vessels. *International Fisheries Research*, 2.

Peatman, T. (2025). Analysis of longline size frequency data for the 2025 southwest Pacific swordfish and striped marlin assessments. Technical Report WCPFC-SC21-SA-IP-14.

Skillman, R. (1989). Status of Pacific billfish stocks. In *Planning the Future of Billfishes: Research and Management in the 90s and Beyond. Part 1: Fishery and Stock Synopses, Data Needs and Management*, pages 179–195, Kailua-Kona, Hawaii, August 1-5, 1988.

Suzuki, Z. (1989). Catch and fishing effort relationships for striped marlin, blue marlin and black marlin in the Pacific Ocean, 1952–1985. In *Planning the Future of Billfishes: Research and Management in the 90s and Beyond. Part 1: Fishery and Stock Synopses, Data Needs and Management*, pages 165–178, Kailua-Kona, Hawaii, August 1-5, 1988.

Tremblay-Boyer, L. and Williams, A. (2023). Standardised CPUE indices for the target species in the eastern tuna and billfish fishery–1998 to 2022. Technical Report Technical Report Working Paper presented to the 38th meeting of the Tropical Tuna Resource Assessment Group held 11-13 July 2023, Mooloolaba, CSIRO, Hobart., CSIRO, contact Ashley.Williams@csiro.au.

Ward, P. and Myers, R. A. (2005). Shifts in open-ocean fish communities coinciding with the commencement of commercial fishing. *Ecology*, 86(4):835–847.

Whitelaw, W. (2001). Country Guide to Gamefishing in the Western and Central Pacific. Technical report, SPC-OFP, Noumea.

Williams, P. (2003). Estimates of annual catches for billfish species taken in commercial fisheries of the western and central Pacific Ocean. Technical Report Working Paper SWG–3., Majuro, Republic of the Marshall Islands.

12 Tables

Table 1: Definition of fisheries and selectivity information for the striped marlin SS3 stock assessment. L=length data, W=weight data

Fishery	Flag	Gear	Sub-region	Catch	Years	Selectivity	Size data adjustment factors
F01.LL.JP.1	JP	Longline	1	Number	1952-2022	Double Normal	L 0.4688, W 0.25
F02.LL.JP.2.late	JP	Longline	2	Number	1979-2022	Double Normal	L 0.104, W 0.35
F03.LL.JP.3	JP	Longline	3	Number	1952-2022	Double Normal	L 0.1314, W 0.30
F04.LL.JP.4	JP	Longline	4	Number	1954-2022	Double Normal	NA
F05.LL.TW.4	TW	Longline	4	Number	1967-2022	Mirror 4	L 0.3537
F06.LL.AU.2	AU	Longline	2	Number	1987-2022	Double Normal	W 0.2
F07.LL.AU.3	AU	Longline	3	Number	1987-2022	Double Normal	W 0.3
F08.LL.NZ.4	NZ	Longline	3	Number	1990-2022	Mirror 7	NA
F09.REC.AU.3	AU	Recreational	3	Number	1952-2022	Double Normal	W 0.25
F10.REC.NZ.3	NZ	Recreational	3	Number	1952-2022	Logistic	L 0.2561, W 0.35
F11.LL.ALL.1	DWFN/PICT	Longline	1	Number	1967-2022	Mirror 1	L 0.0943
F12.LL.ALL.2	DWFN/PICT	Longline	2	Number	1967-2022	Mirror 2	L 0.0197
F13.LL.ALL.3	DWFN/PICT	Longline	3	Number	1968-2022	Mirror 3	NA
F14.LL.ALL.4	DWFN/PICT	Longline	4	Number	1976-2022	Mirror 4	NA
F15.LL.JP.2.early	JP	Longline	2	Number	1952-1978	Mirror 2	NA
F16.LL.NewCal.2	NC	Longline	2	Number	1983-2022	Double Normal	L 0.054
F17.LL.FPoly.4	PF	Longline	4	Number	1992-2022	Double Normal	L 0.0401
F18.S01.1-4	JP & TW	Longline	1–4	-	1979-2022	Double Normal	NA
F19.S02.NZ	NZ	Recreational	3	-	1975-2022	Double Normal	NA
F20.S03.AU	AU	Longline	2& 3	-	1998-2022	Mirror 6	NA
F21.S04.PICT.NCFJTO	NC-FJ-TO	Longline	2	-	2001-2022	Mirror 16	NA
F22.S04.PICT.PF	PF	Longline	1 & 4	-	2002-2022	Mirror 17	NA

Table 2: Description of the sensitivity models and convergence. O1.Diagnostic model (diag) has JP/TW CPUE index, growth estimated internal, M-at-age Lorenzen with ave M 0.36, and h=0.8. All the fixed growth sensitivities have A0=0.5, L1=50cm, h=0.8 and the JP/TW CPUE index. M and h sensitives are off the diagnostic with internal growth and JP/TW index.

Model	Description	Gradient	PDH	Converged	SB2022 (t)	$SB_{2022}/SB_{F=0}$
01.Diagnostic	see caption	4.02E-06	Y	Y	1363	0.14
03.CPUE AU LL	diagn biol	7.26419E-05	Y	Y	2386	0.24
04.CPUE NZ rec	diag biol	9.07E-05	Y	Y	693	0.08
05.CPUE ob NC-FJ-TO	diag biol	1.66E-04	Y	Y	5126	0.45
06.CPUE ob PF	diag biol	5.57E-05	Y	Y	1140	0.13
07.Growth-1-M-diag	L2=195cm, K=0.2	9.66E-03	Y	**N**	1988	0.28
08.Growth-2-M-diag	L2=210cm, K=0.2	1.37E-04	Y	Y	1014	0.12
09.Growth-3-M-diag	L2=225cm, K=0.2	1.85E-03	Y	**N**	620	0.05
10.Growth-4-M-diag	L2=195cm, K=0.5	8.69E-05	Y	Y	5150	0.54
11.Growth-5-M-diag	L2=210cm, K=0.5	3.35E-05	Y	Y	1275	0.15
12.Growth-6-M-diag	L2=225cm, K=0.5	>1	**N**	**N**	734	0.06
13.Growth-7-M-diag	L2=195cm, K=0.8	1.69E-04	Y	Y	4284	0.48
14.Growth-8-M-diag	L2=210cm, K=0.8	5.23E-04	Y	Y	898	0.10
15.Growth-9-M-diag	L2=225cm, K=0.8	>1	N	**N**	703	0.04
16.Diag-high M	1.35 x diag M-at-age	3.93E-05	Y	Y	1503	0.23
17.Diag-low M	0.65 x diag M-at-age	2.33E-03	Y	**N**	1332	0.10
18.Diag-high h	h=0.95	6.57E-06	Y	Y	1341	0.19
19.Diag-low h	h=0.75	4.69E-04	Y	Y	1375	0.12
20.Diag-exc-NZ-rec-size	diag biol, JP/TW index	1.847E-05	Y	Y	1427	0.15

Table 3: Description of the exploratory grid focused on the biological uncertianties.

Axis	Option 1	Option 2	Option 3
Growth K	0.2	0.5	0.8
Growth L2	195	210	215
M-at-age	low (diag x 0.65)	diag	high (diag x 1.35)
Steepness	0.75	80	0.95

Table 4: Stock assessment summary statistics for the 41 model grid

Metric	**Mean**	**Median**	**Min**	**10%ile**	**90%ile**	**Max**
SB_{latest}	2527	1304	469	670	6611	7571
SB_{recent}	1964	1070	488	627	4809	5780
$TotalBiomass_{latest}$	11258	8702	2600	3936	20544	33243
$TotalBiomass_{recent}$	9891	7393	2266	3280	16989	30188
F_{latest}	0.16	0.11	0.03	0.05	0.37	0.60
F_{recent}	0.26	0.15	0.05	0.05	0.68	0.99
SB_{MSY}	2071	2081	758	946	2940	4463
F_{MSY}	0.30	0.25	0.05	0.09	0.55	0.86
F_{recent}/F_{MSY}	0.90	0.84	0.12	0.24	1.57	1.97
F_{latest}/F_{MSY}	0.61	0.56	0.08	0.16	1.11	1.53
SB_{recent}/SB_{MSY}	1.13	0.73	0.17	0.29	2.29	5.19
SB_{latest}/SB_{MSY}	1.45	0.90	0.16	0.29	3.10	6.80
$SB_{recent}/SB_{F=0}$	0.25	0.18	0.03	0.06	0.65	0.75
$SB_{latest}/SB_{F=0}$	0.32	0.22	0.03	0.07	0.89	0.99

13 Figures

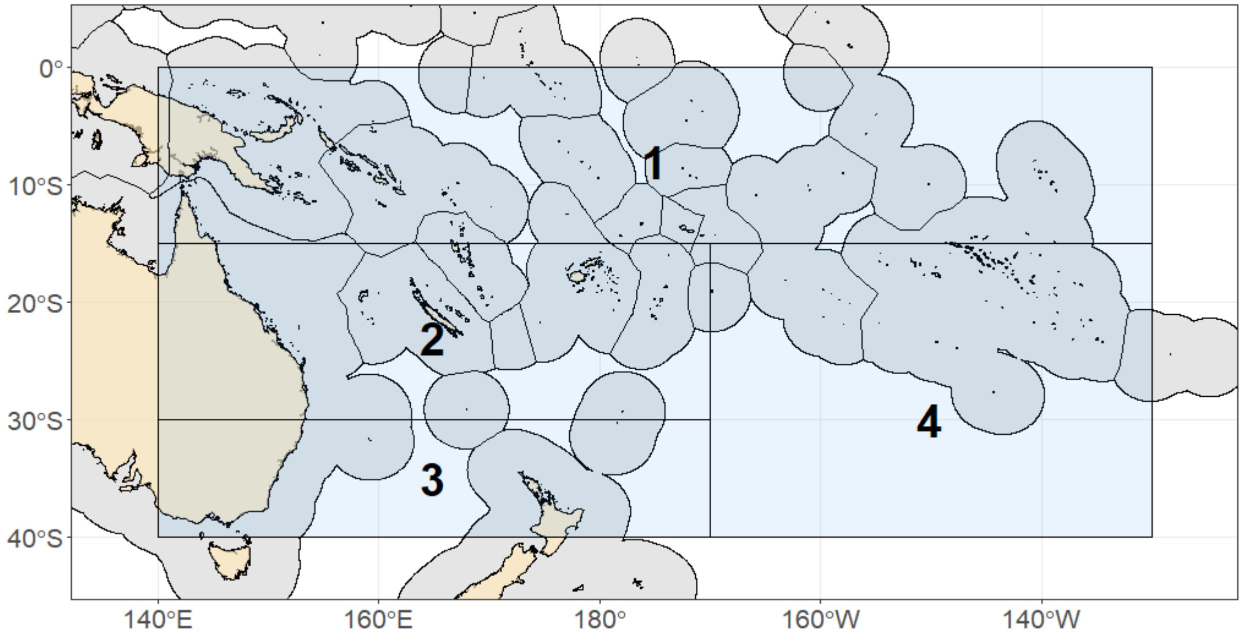


Figure 1: The geographical area covered by the stock assessment and the boundaries of the four model sub-regions used for 2024 SWPO striped marlin assessment.

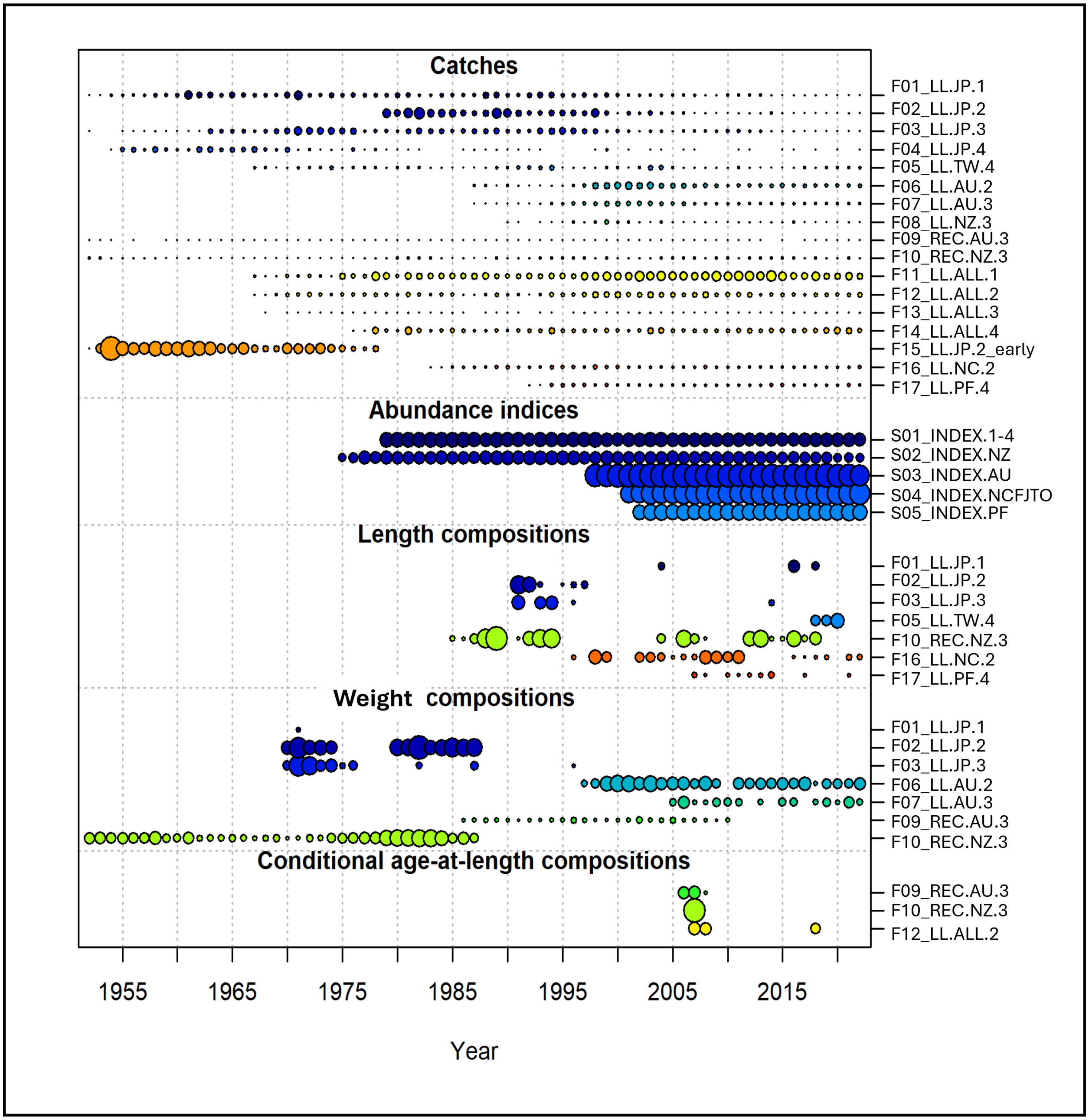


Figure 2: Summary of data coverage by fishery for the SWPO 2024 striped marlin assessment performed in SS3. LL=longline, REC=recreational gears.

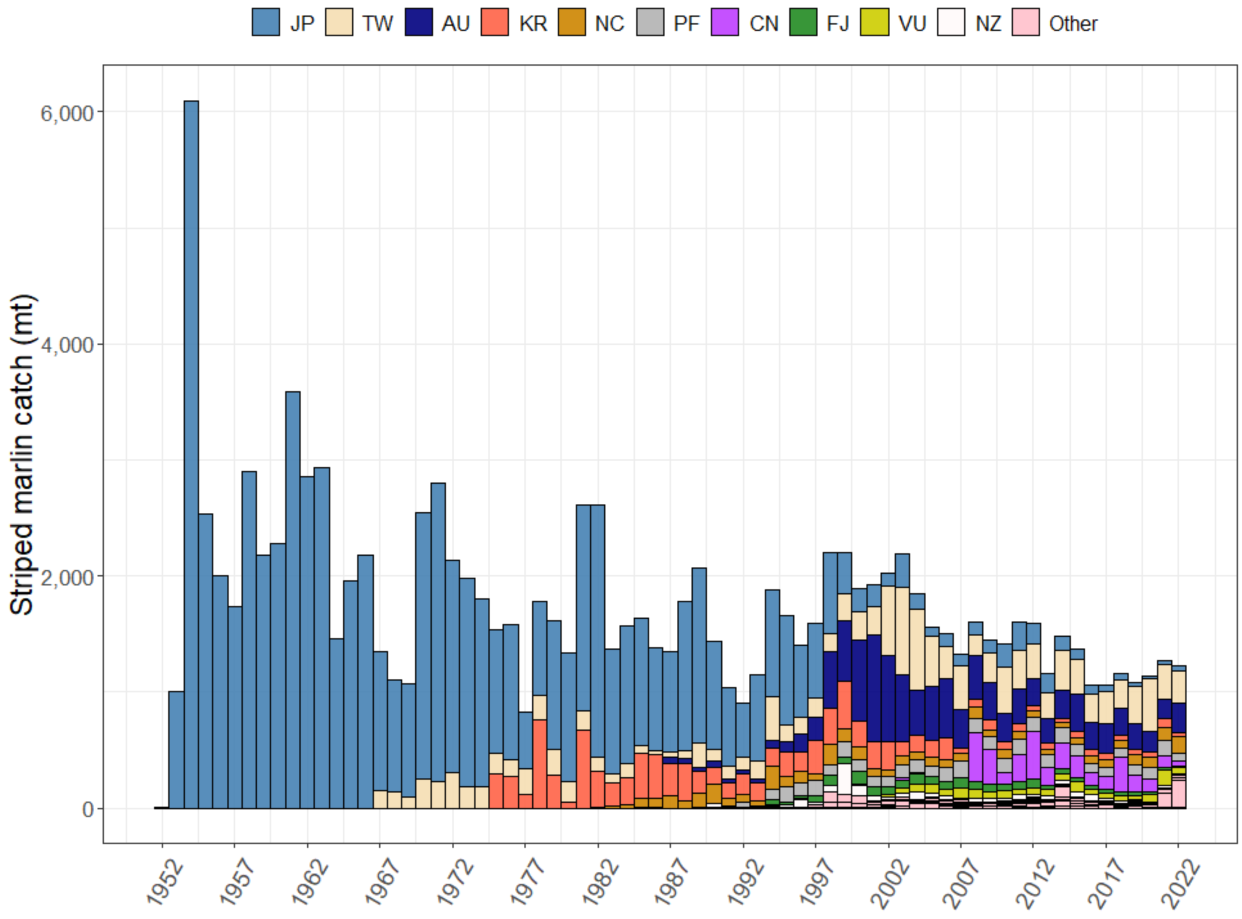


Figure 3: Annual catch (mt) of striped marlin by flag from 1952-2022 in the SWPO area covered by the assessment.

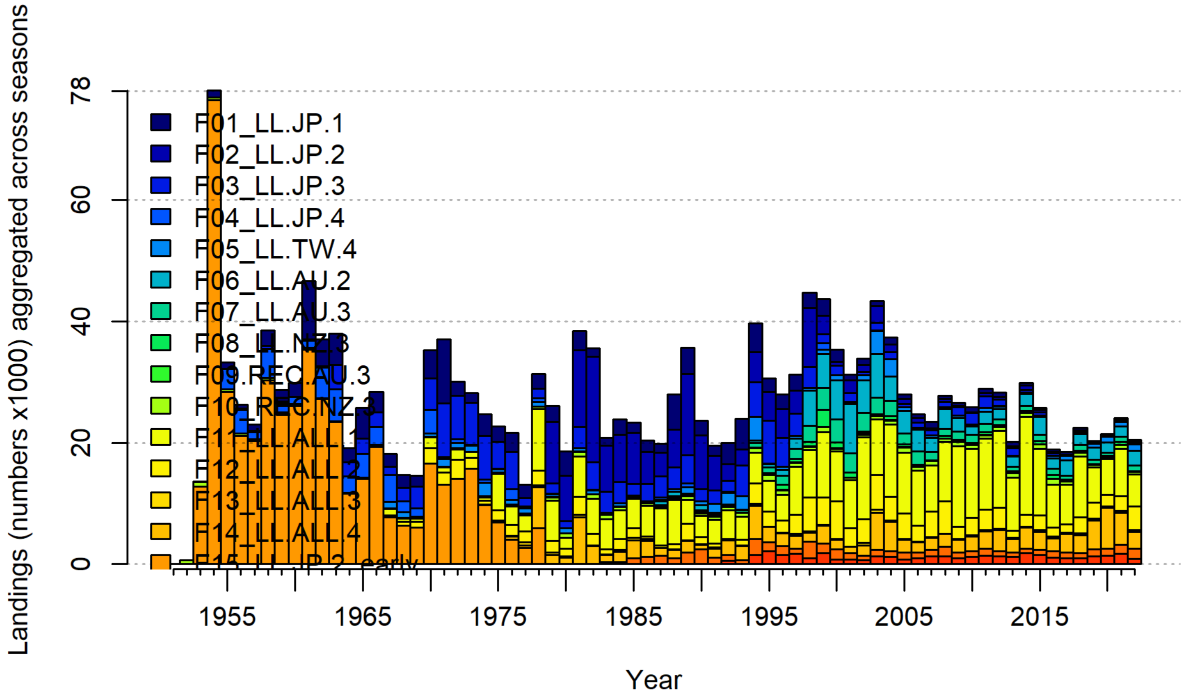


Figure 4: Annual catch (numbers) of striped marlin by model fishery from 1952-2022 in the SWPO area covered by the assessment.

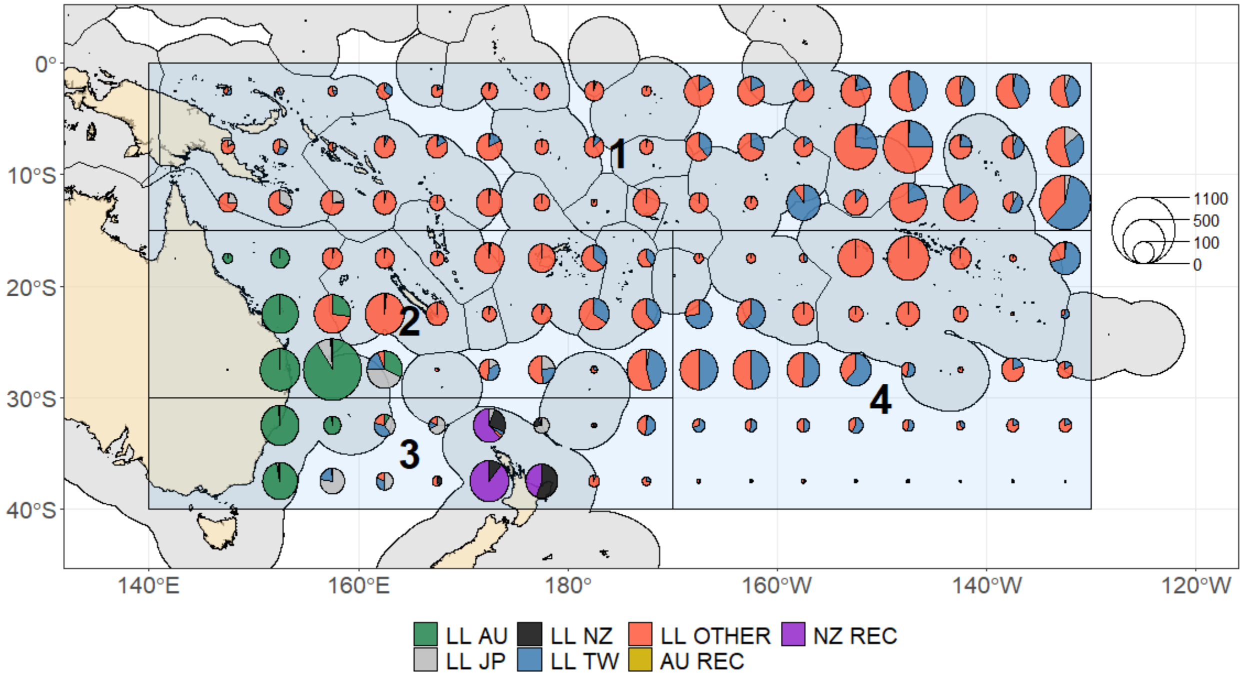


Figure 5: Distribution and magnitude of striped marlin catch (mt) by gear type summed over the recent 10 year period from 2013-2022 for 5° × 5° cells.

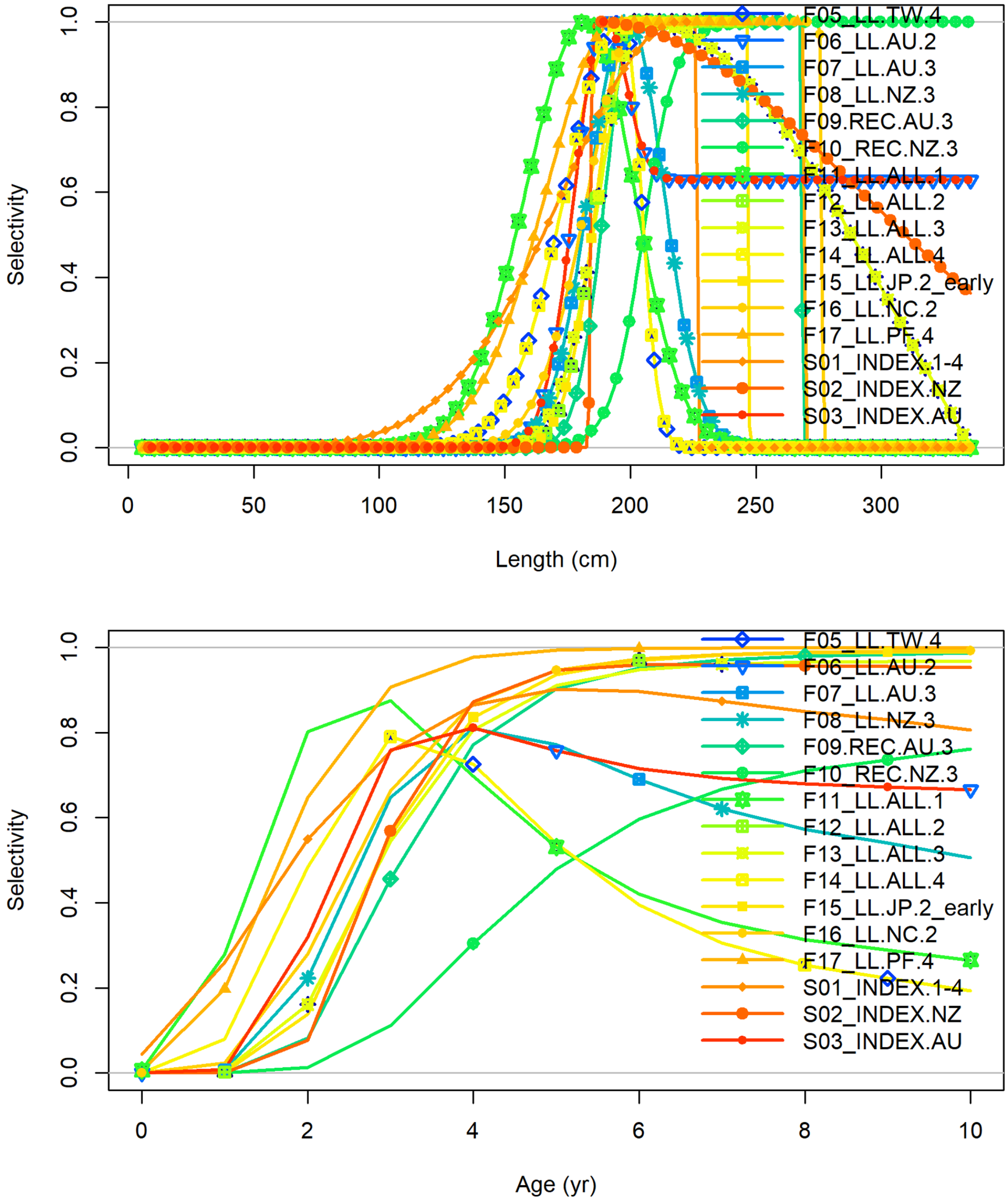


Figure 6: Estimated length-specific and age specific selectivity curves by fishery combined for the diagnostic model.

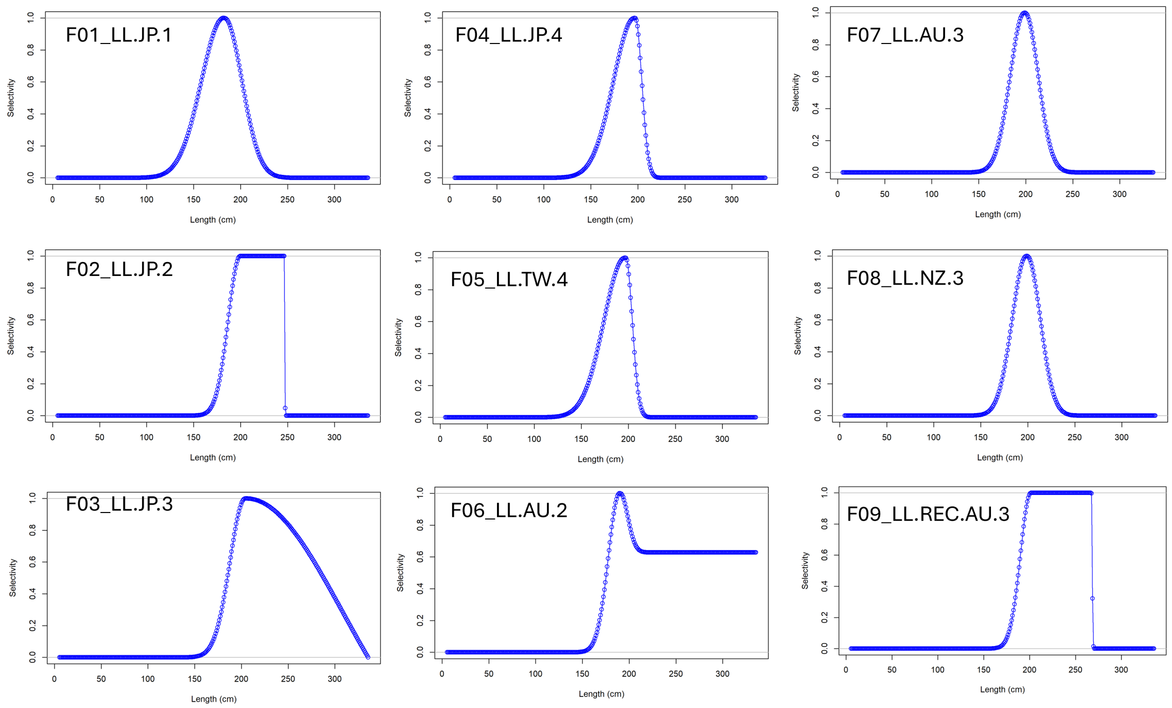


Figure 7: Extraction fisheries 1-9: Estimated length-specific selectivity curves shown individually by fishery for the diagnostic model.

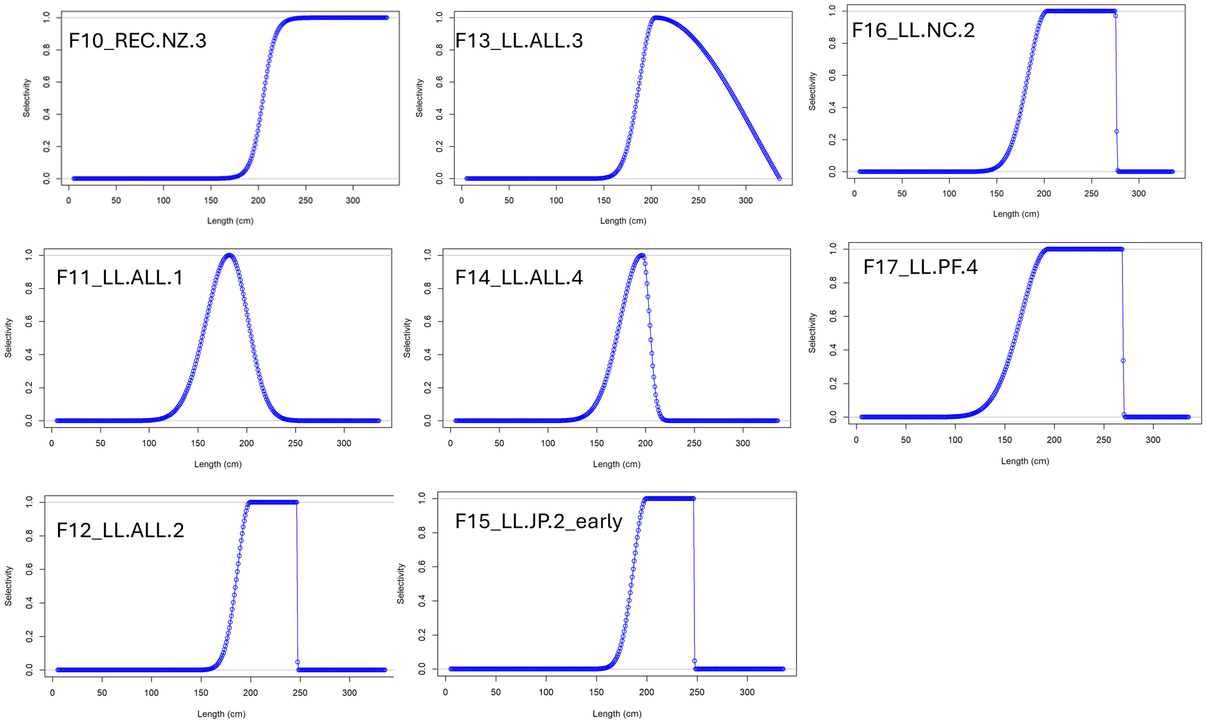


Figure 8: Extraction fisheries 10-17: Estimated length-specific selectivity curves shown individually by fishery for the diagnostic model.

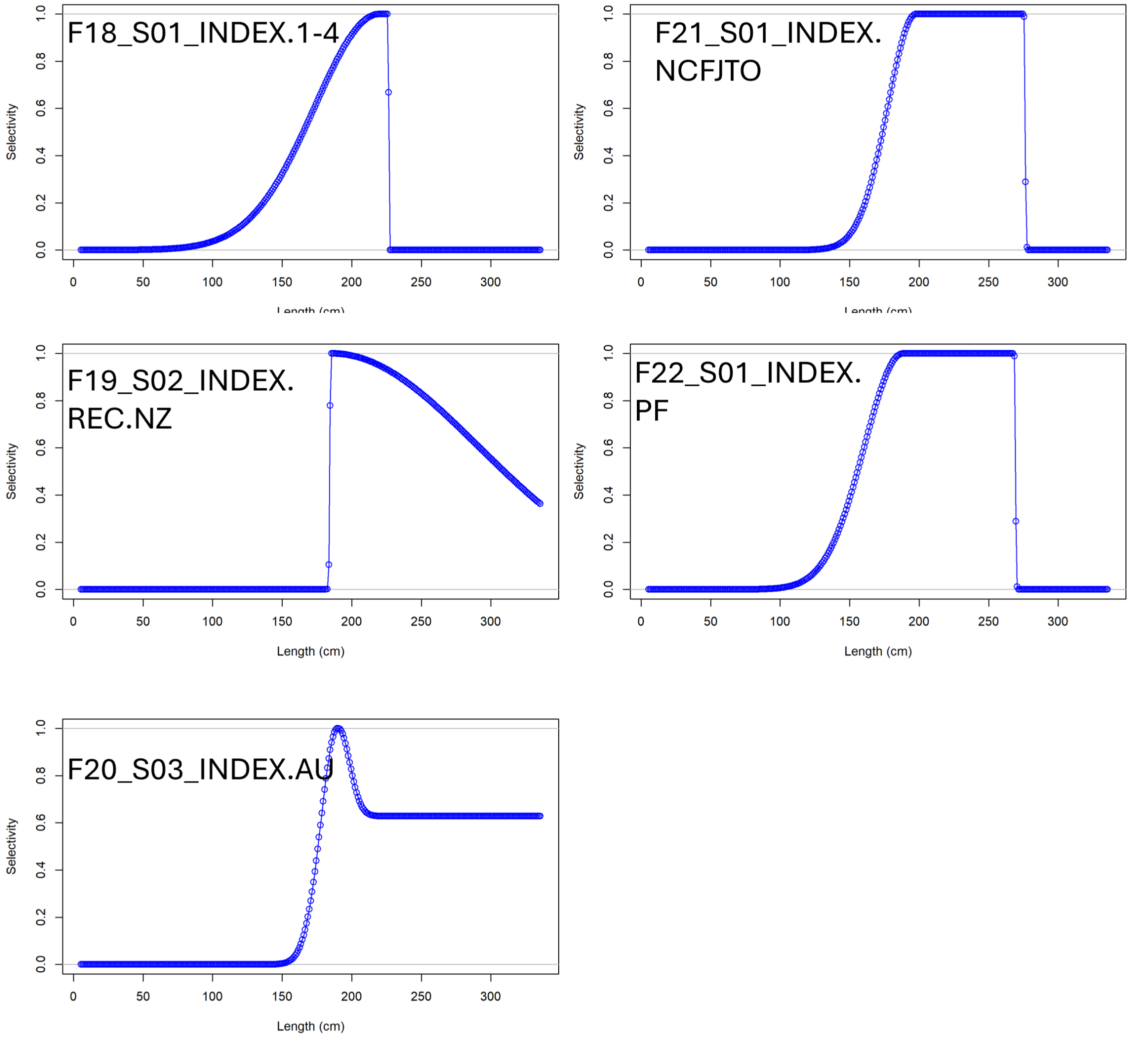


Figure 9: Index fisheries 18-22: Estimated length-specific selectivity curves shown individually by fishery.

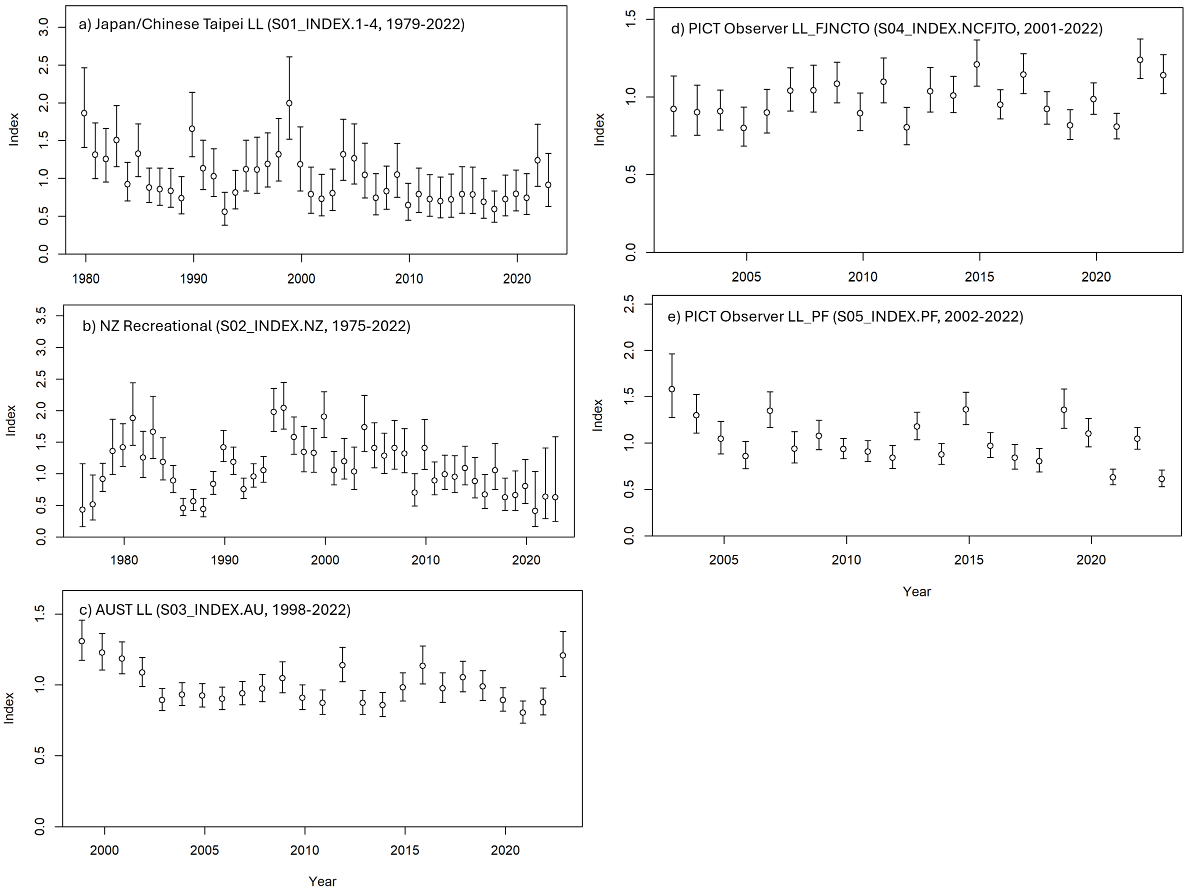


Figure 10: Comparison of alternative standardised CPUE abundance indices. Error bars indicate 95% uncertainty interval around index values based on the model assumption of lognormal error.

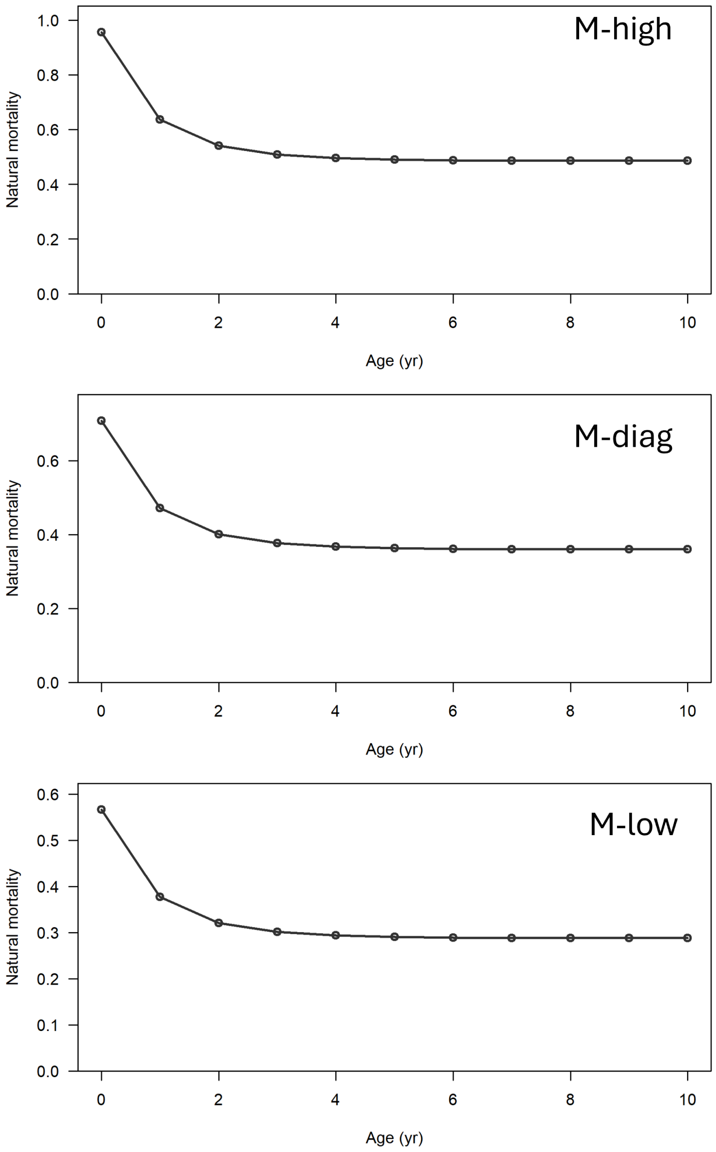


Figure 11: Comparison of the high, diagnostic and low M-at-age scenarios applied in sensitivity models and the exploratory grid

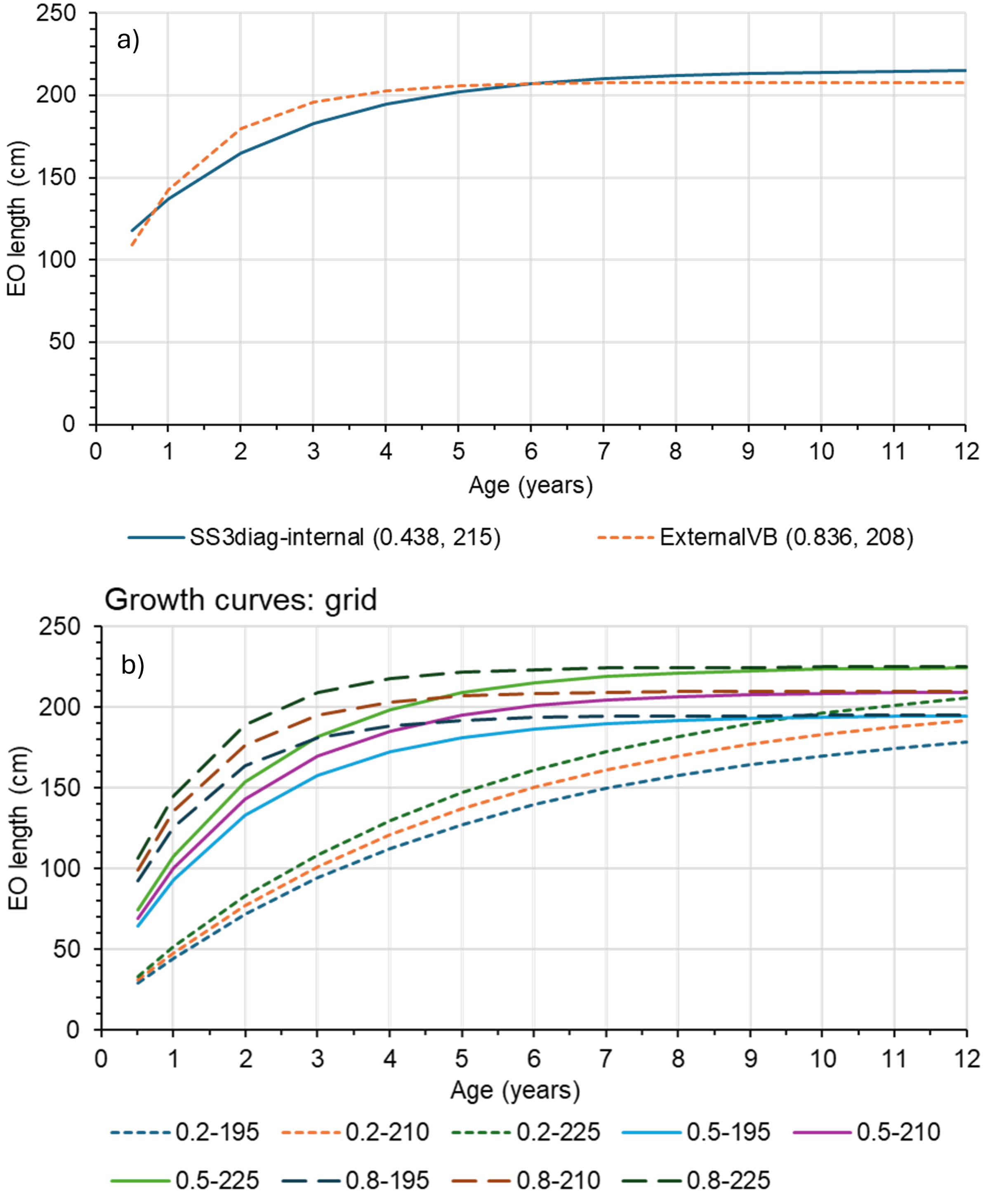


Figure 12: Comparison of growth curves; a) SS3 diagnostic model growth estimated internally from CAAL data (SS3diag-internal, and the external vB growth curve fitted to the age-length data as used in the 2024 MFCL diagnostic model (ExternalVB), b) the nine growth curves applied in the exploratory grid. Numbers in parenthesis are K, L2.

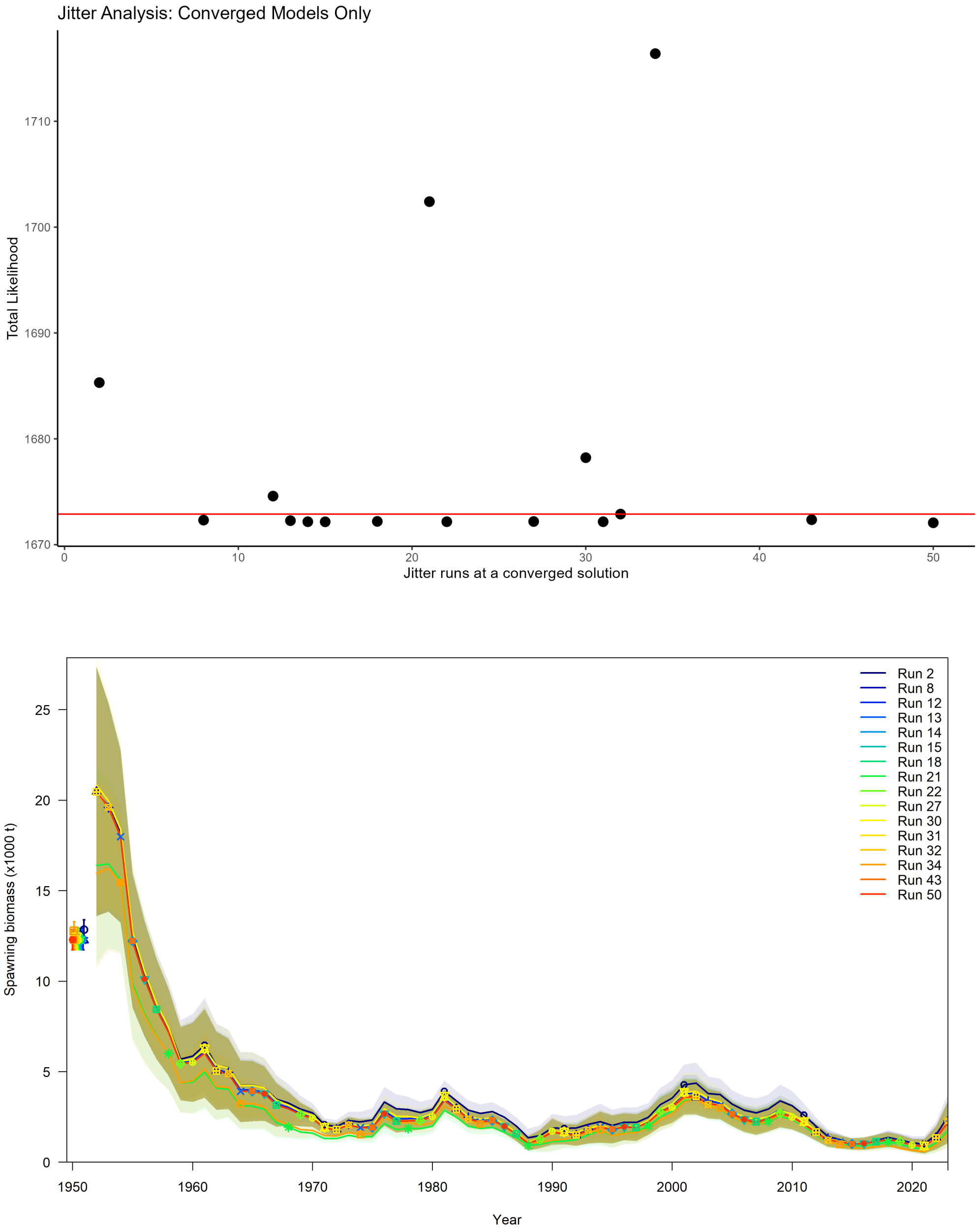


Figure 13: Top–Log likelihood plotted against jitter number, with the diagnostic model indicated by the red line. Only models with gradient ≤0.001 and positive definite Hessian plotted. Models above red line are worse fits than diagnostic, below red line are better fits. Bottom–spawning biomass trajectories compared among the suitably converged jitter models.

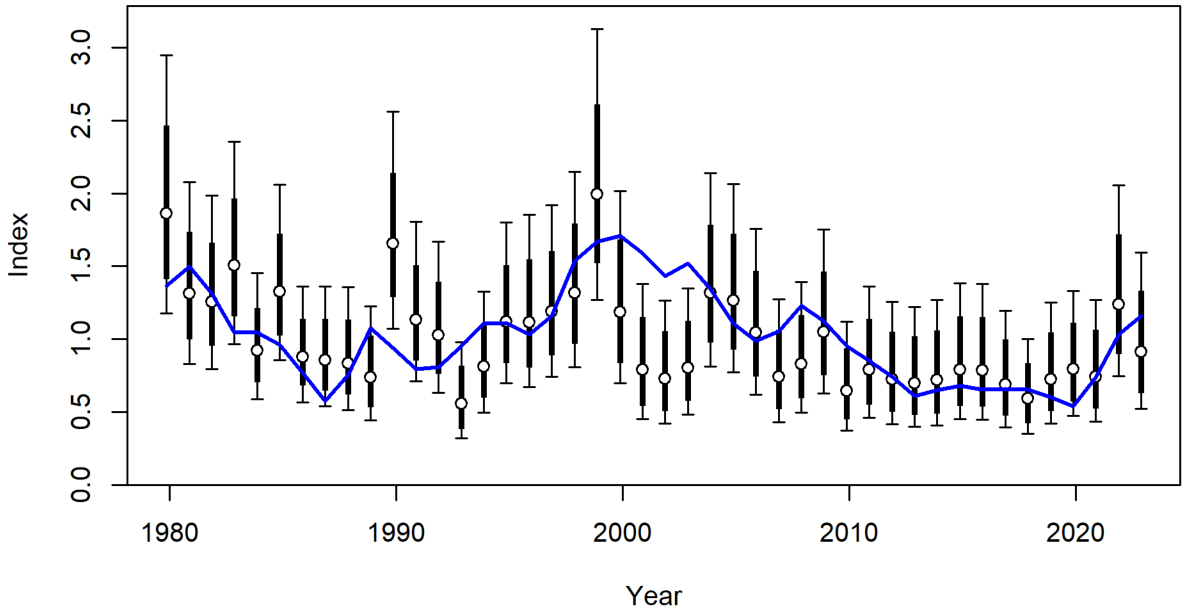


Figure 14: Fits (blue line) to the standardised CPUE (open circles/black bars) for the JP/TW longline index fishery (sub-regions 1-4). Lines indicate 95% uncertainty interval around index values based on the model assumption of lognormal error. Thicker lines indicate input uncertainty before addition of estimated additional uncertainty parameter.

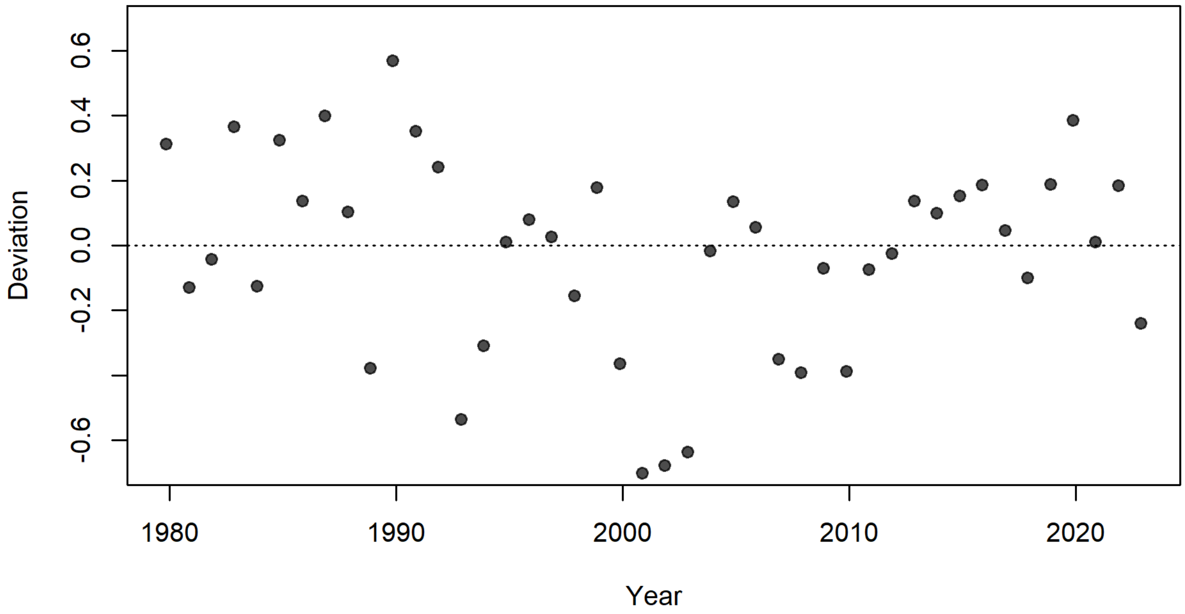


Figure 15: Residuals for the standardised CPUE for the the JPTW longline index fishery in sub-regions 1-4. Values are log(Obs) - log(Exp). Negative residuals indicate the model overestimates the observations.

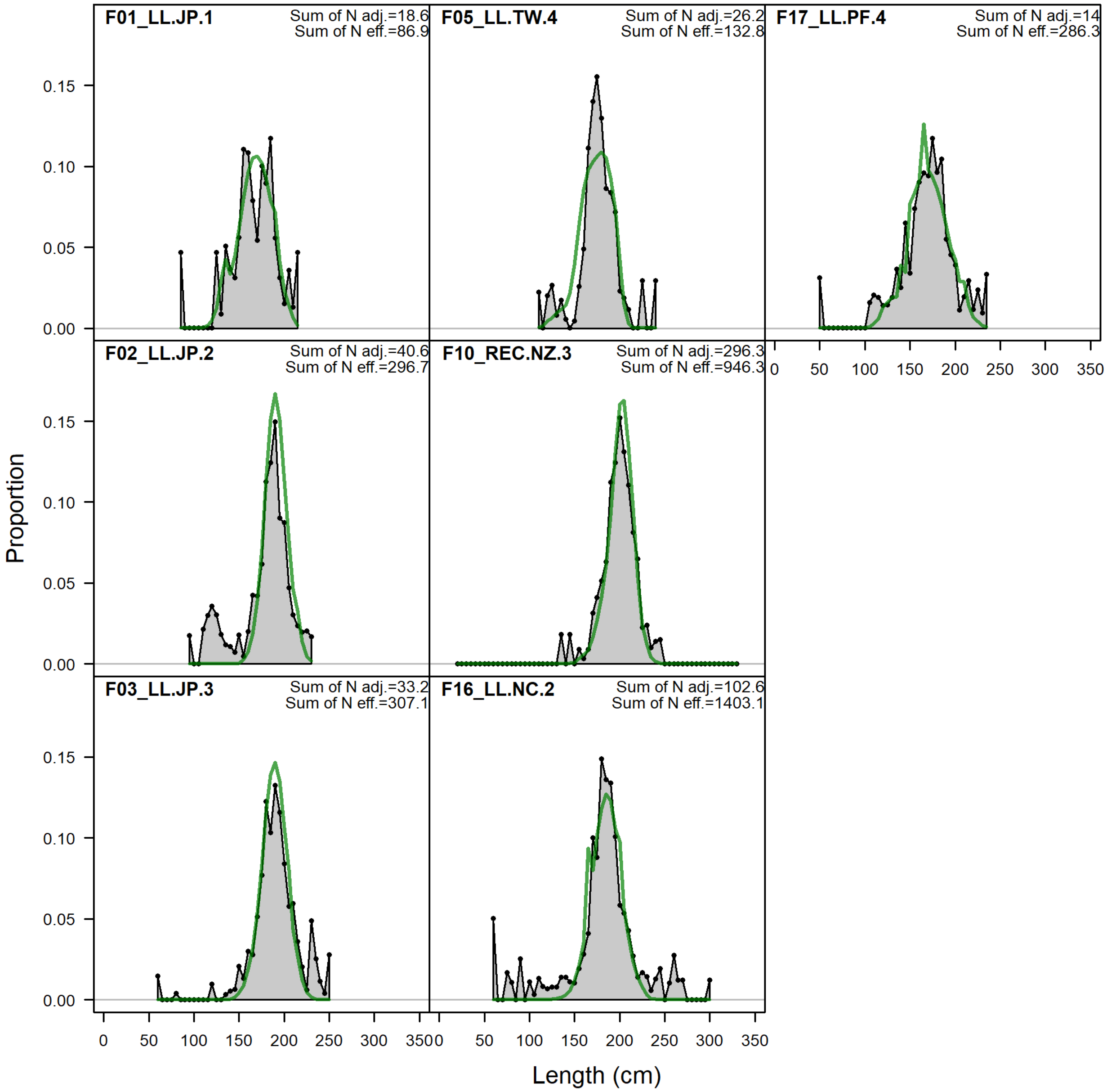


Figure 16: Composite (all time periods combined) observed (grey) and predicted (green line) length frequency for fisheries with length frequency data.

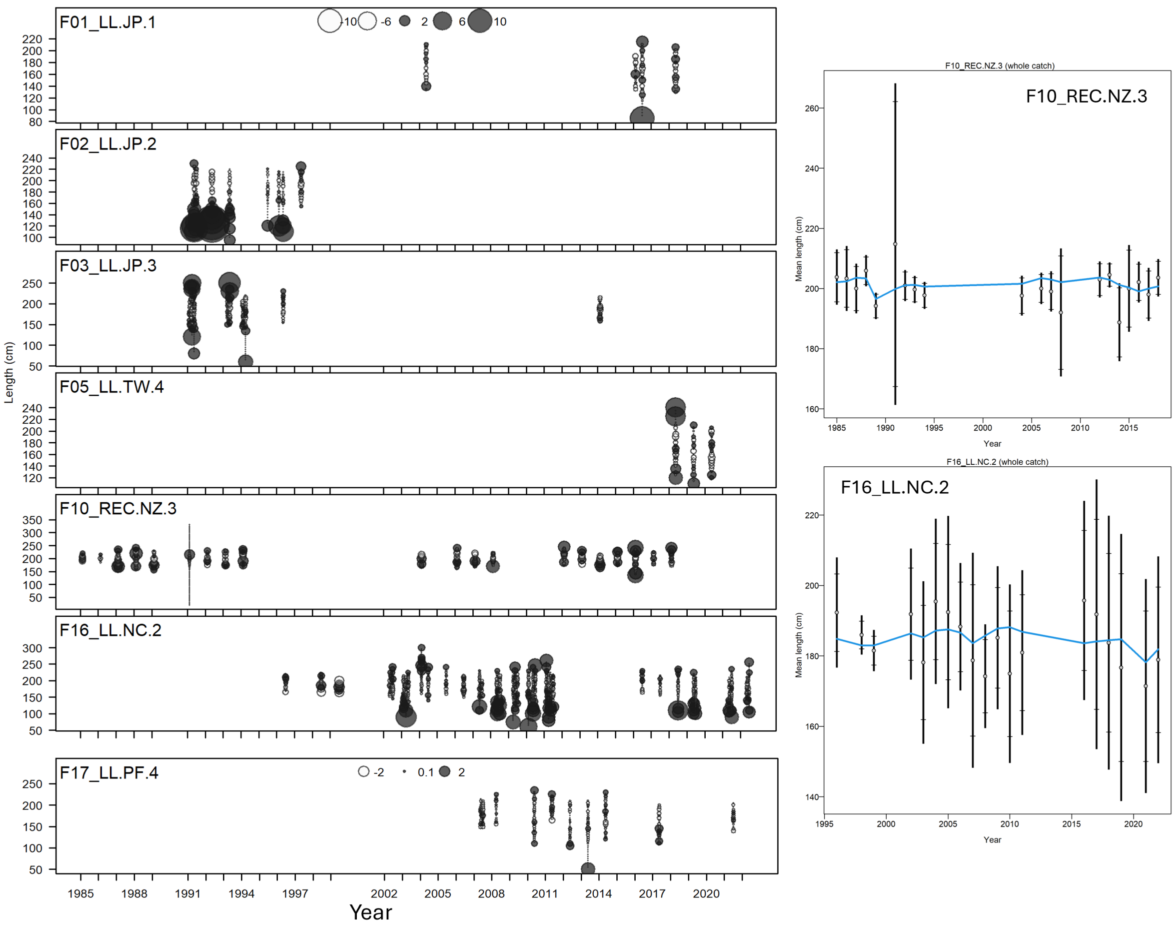


Figure 17: Plots of Pearson residuals for length composition fits for all fisheries with length data (left), closed bubbles are positive residuals (observed > expected) and open bubbles are negative residuals (observed < expected); and fits to meant length time series (right) for F10.Rec.NZ.3 and F16.LL.NC.2 only, observations are mean with 95% confidence intervals based on current sample sizes.

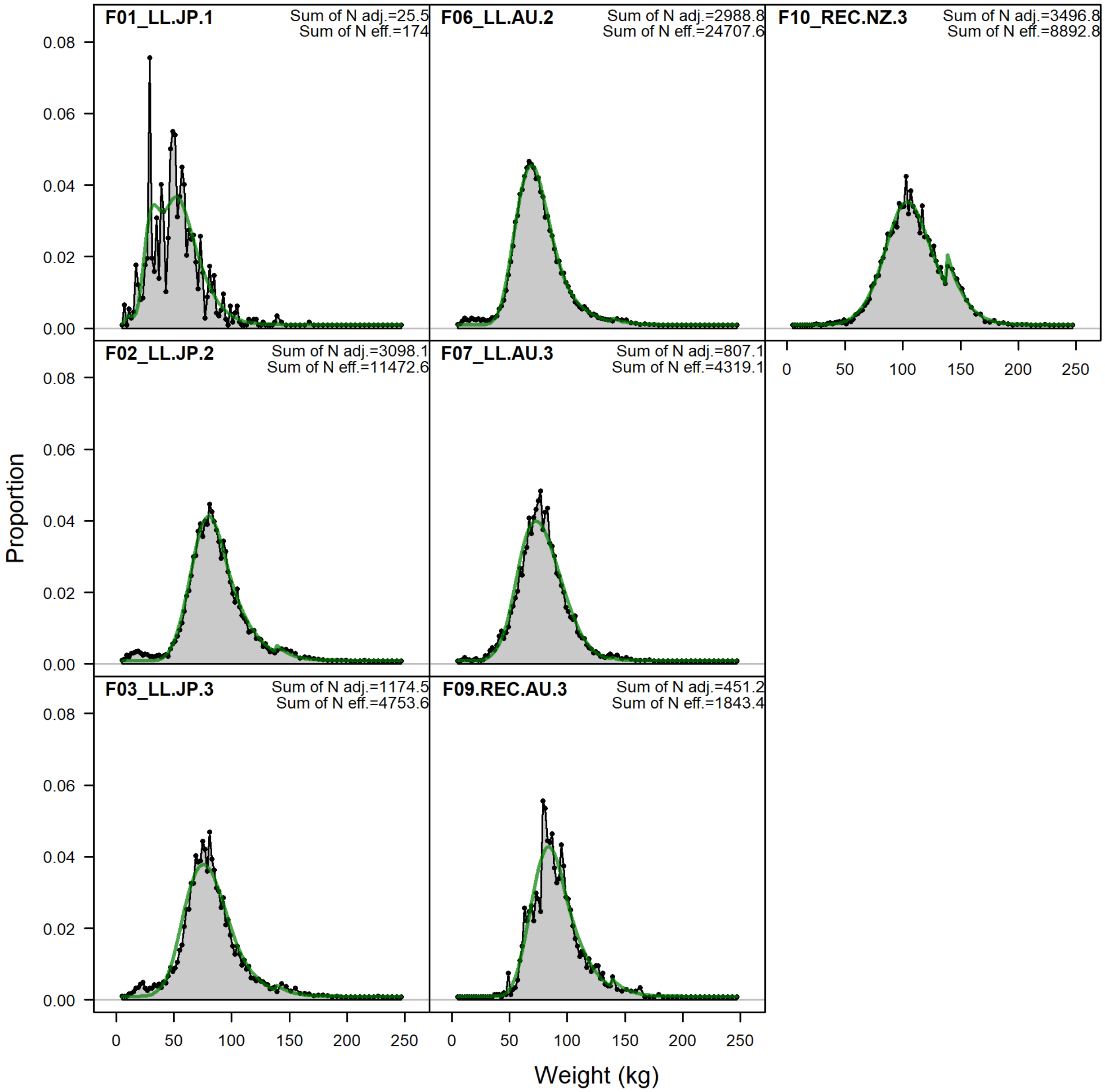


Figure 18: Composite (all time periods combined) observed (grey) and predicted (green line) weight frequency for fisheries with weight frequency data.

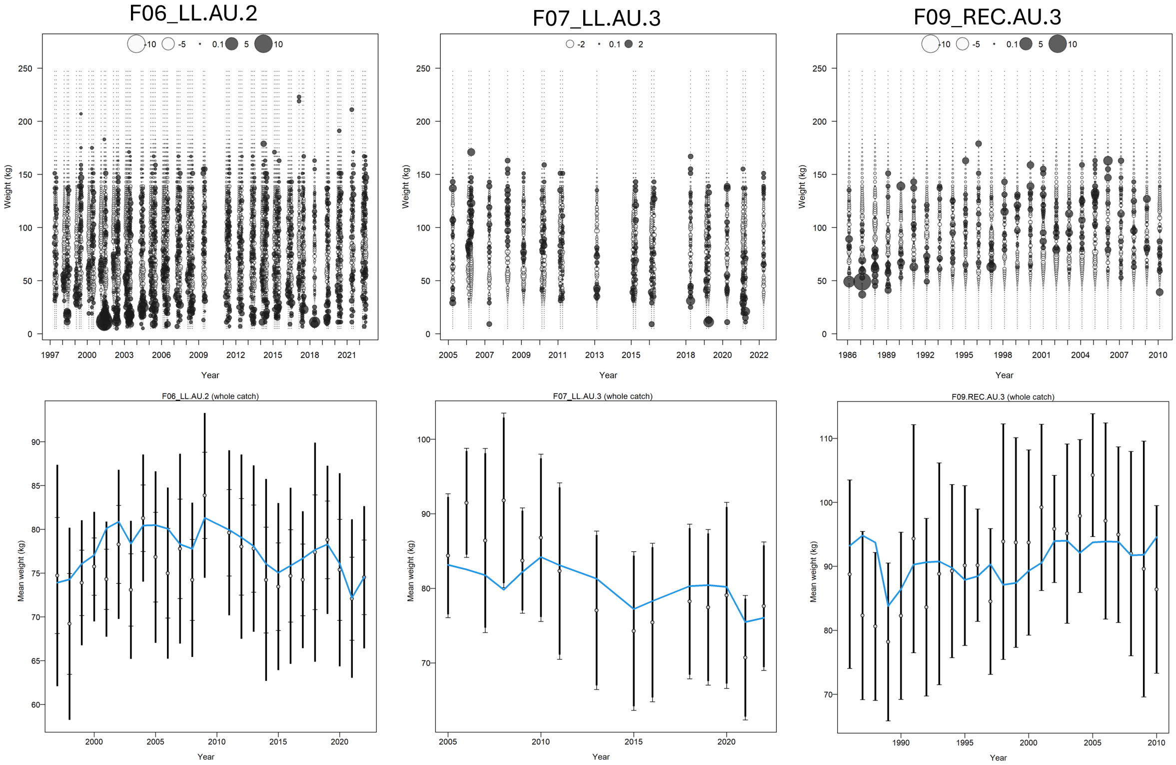


Figure 19: Plots of Pearson residuals for weight composition fits for Australian longline (F06, F07) and recreational fishery (F09) data (top), closed bubbles are positive residuals (observed > expected) and open bubbles are negative residuals (observed < expected); and fits to meant length time series (bottom), observations are mean with 95% confidence intervals based on current sample sizes.

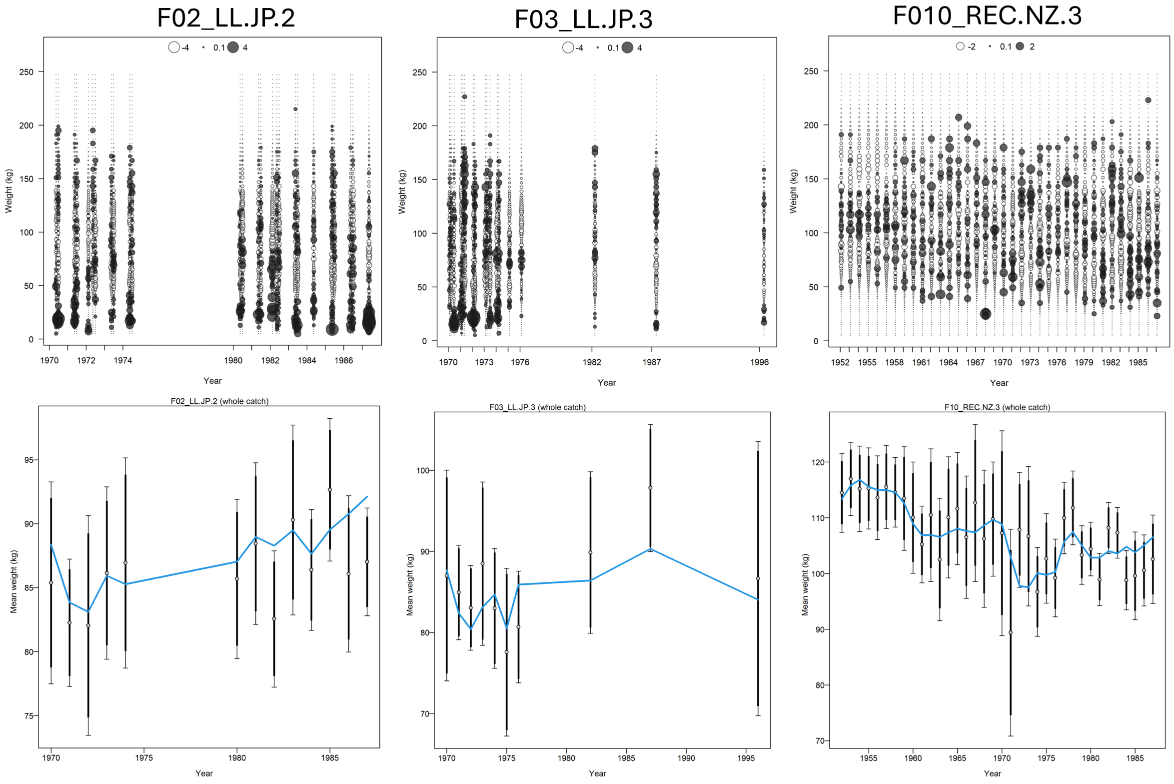


Figure 20: Plots of Pearson residuals for weight composition fits for Japanese longline (F02, F03) and NZ recreational fishery (F10) weight data (top), closed bubbles are positive residuals (observed > expected) and open bubbles are negative residuals (observed < expected); and fits to meant length time series (bottom), observations are mean with 95% confidence intervals based on current sample sizes.

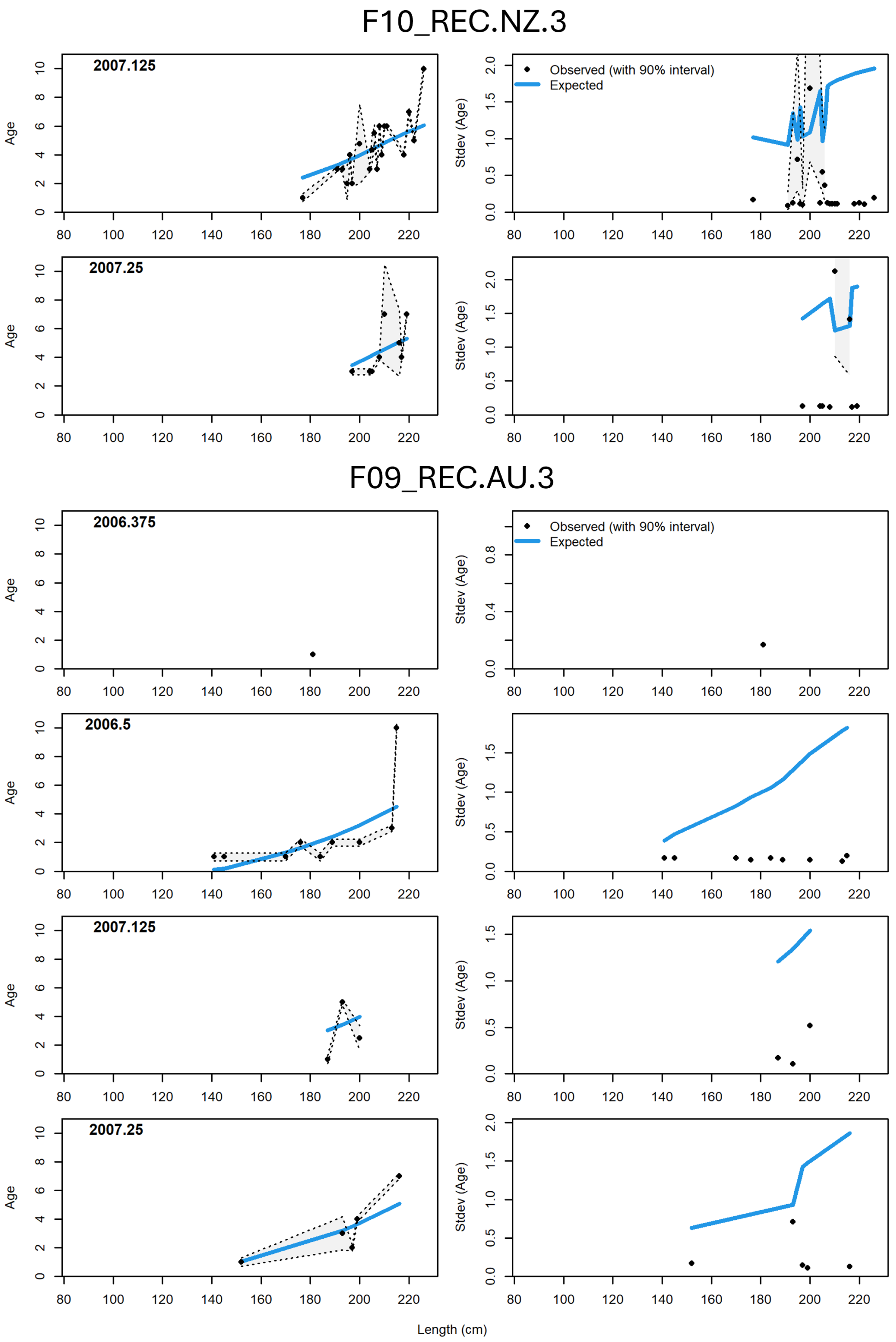


Figure 21: Conditional age at length plots of the catch for F10.REC.NZ.3 and F09.REC.AU.3. These plots show mean age and standard deviation in conditional age at length. Left plots are mean age at length by size-class (obs. and exp.) with 90% CIs based on adding 1.64 SE of mean to the data. Right plots in each pair are SE of mean age at length (obs. and exp.) with 90% CIs based on the chi-square distribution.

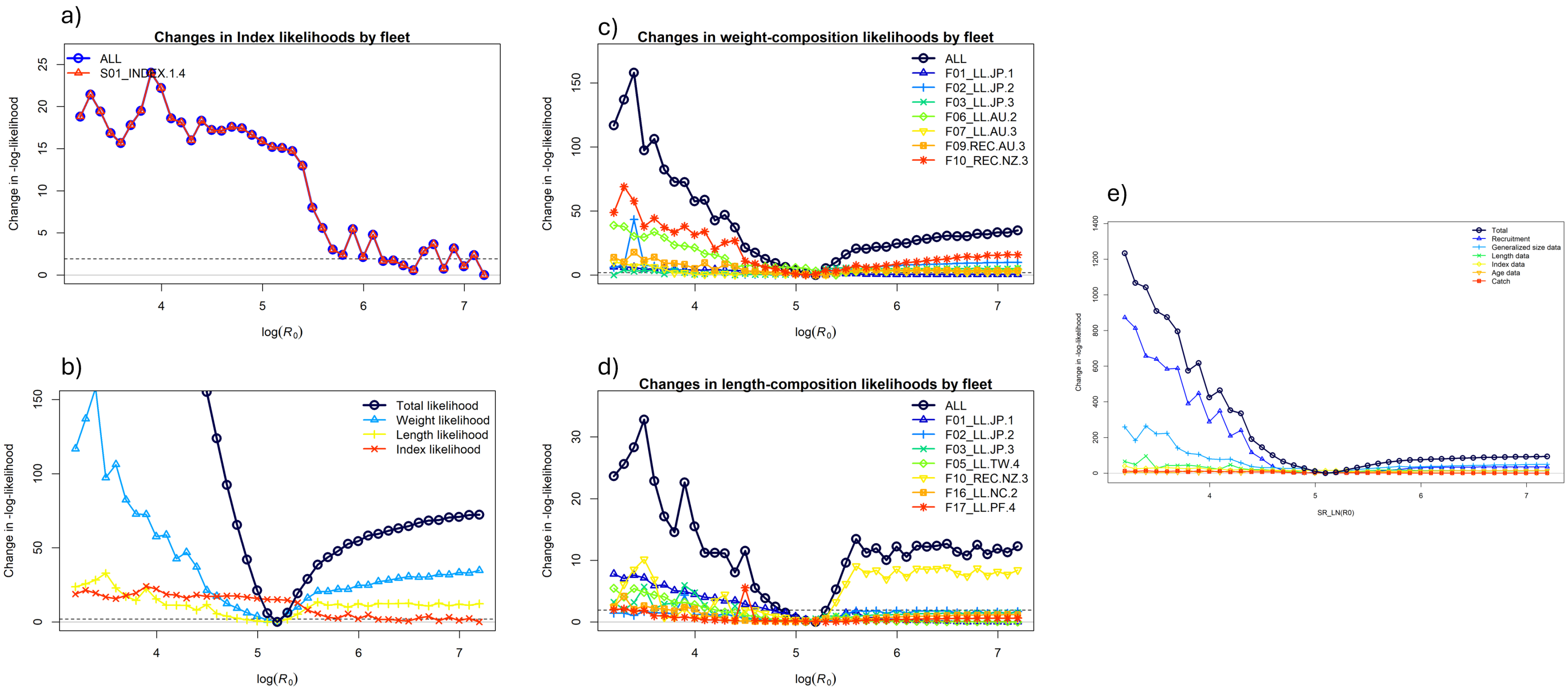


Figure 22: Likelihood profiles on R0 for different data components. a) JP/TW CPUE abundance index, b) length, weight and CPUE index data, c) weight data by fishery, d) length data by fishery, e) all data components and the recruitment (σ_R, deviates), generalised size date refers to the weight composition. The black line indicates the total likelihood with the coloured lines representing data components of the total likelihood.

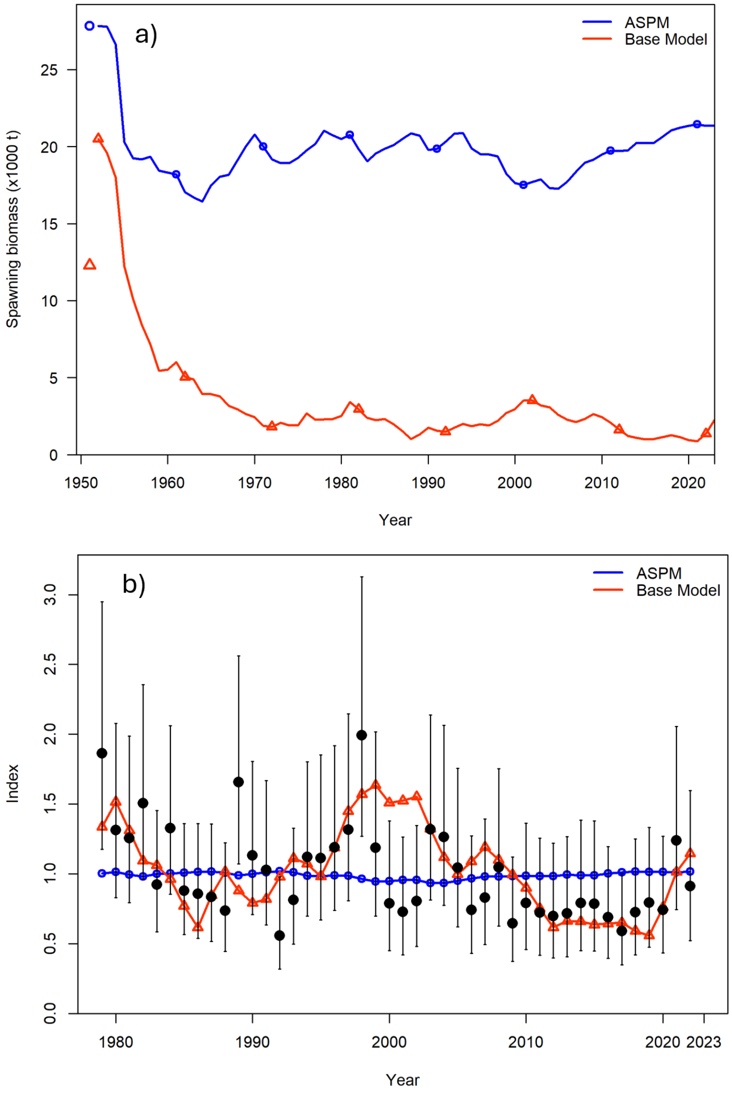


Figure 23: a) Spawning biomass trajectories compared between the ASPM and the diagnostic (base) model, and b) ASPM fit to the JP/TW CPUE index.

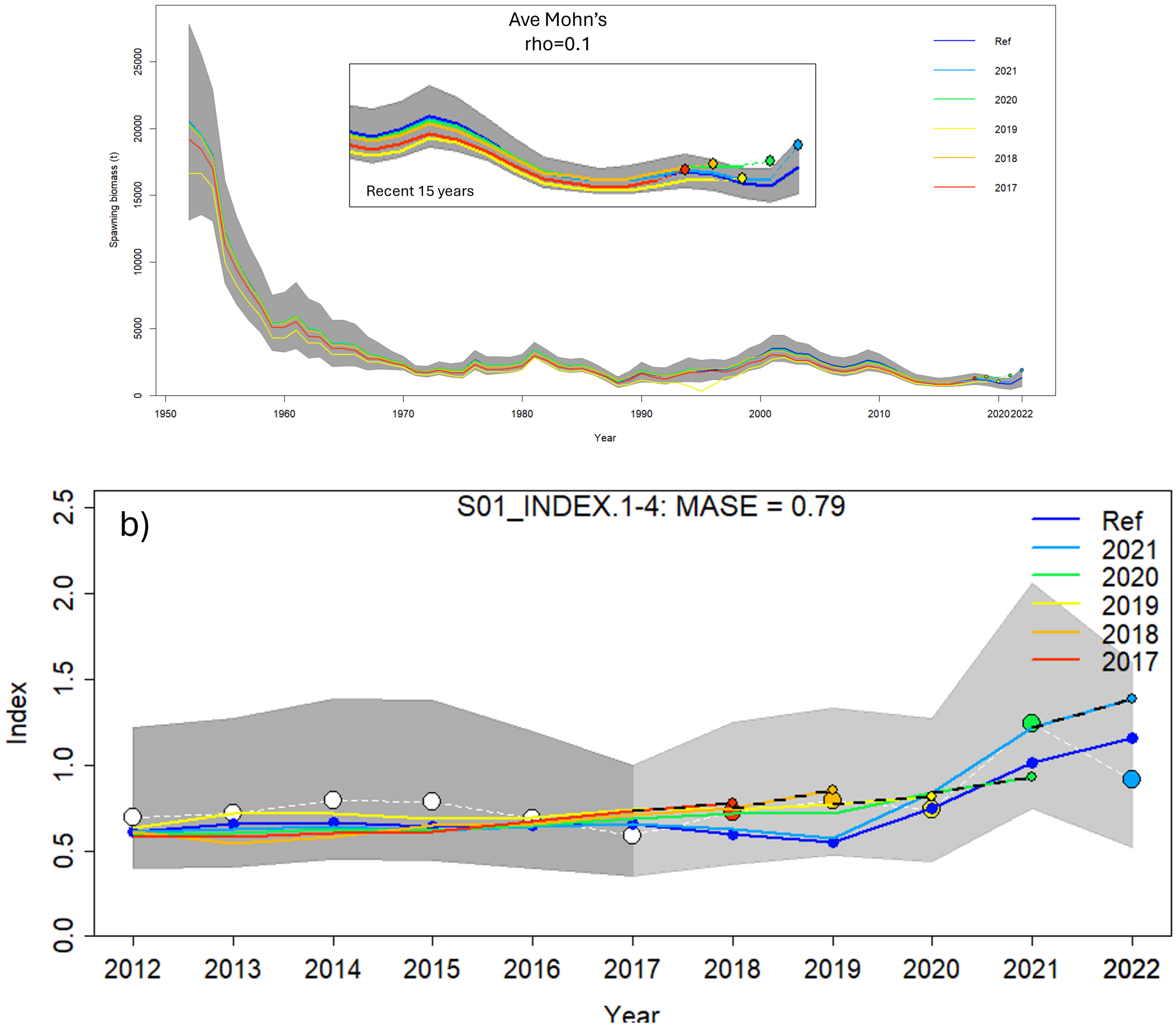


Figure 24: Diagnostic model: a) Retrospective analysis results for spawning biomass, and b) hindcast results for the JPTW CPUE abundance index.

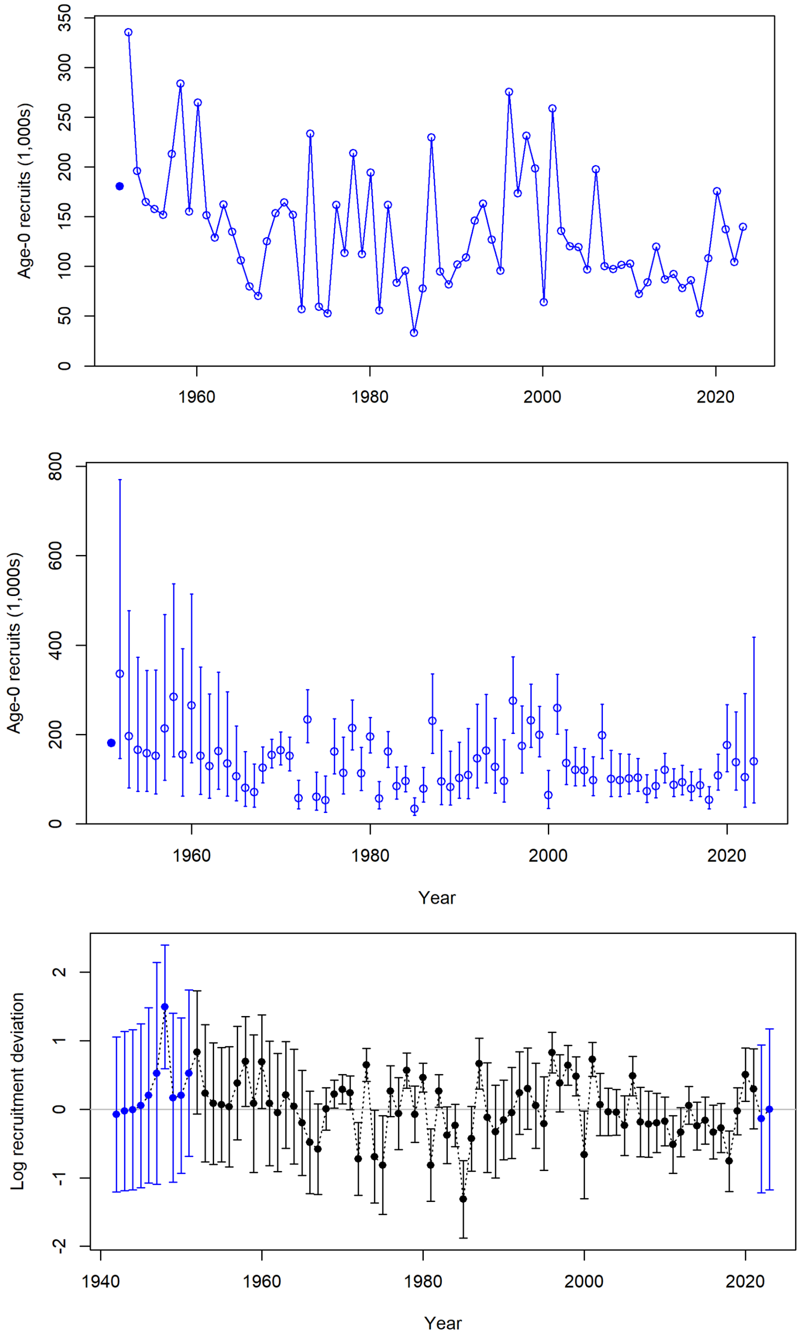


Figure 25: The estimated recruitment from the diagnostic model (top two panels) and log recruitment deviates (bottom panel) with 95 % confidence intervals.

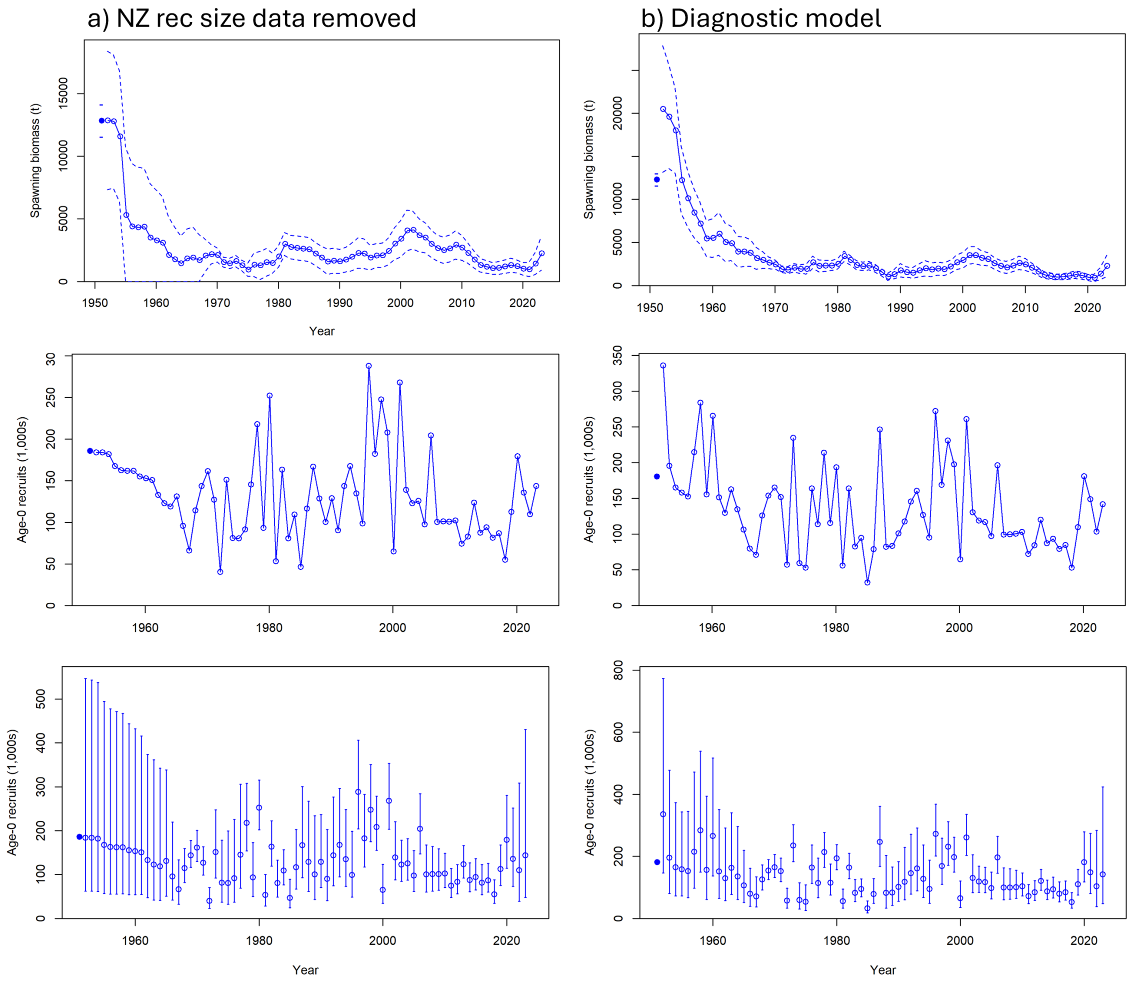


Figure 26: a) The estimated recruitment (numbers, and with 95% asymptotic intervals) and spawning biomass (t) (with 95% asymptotic intervals) from a model without the NZ recreational size data (left panels); and, b) the same plots for the diagnostic model (right panel). % confidence intervals.

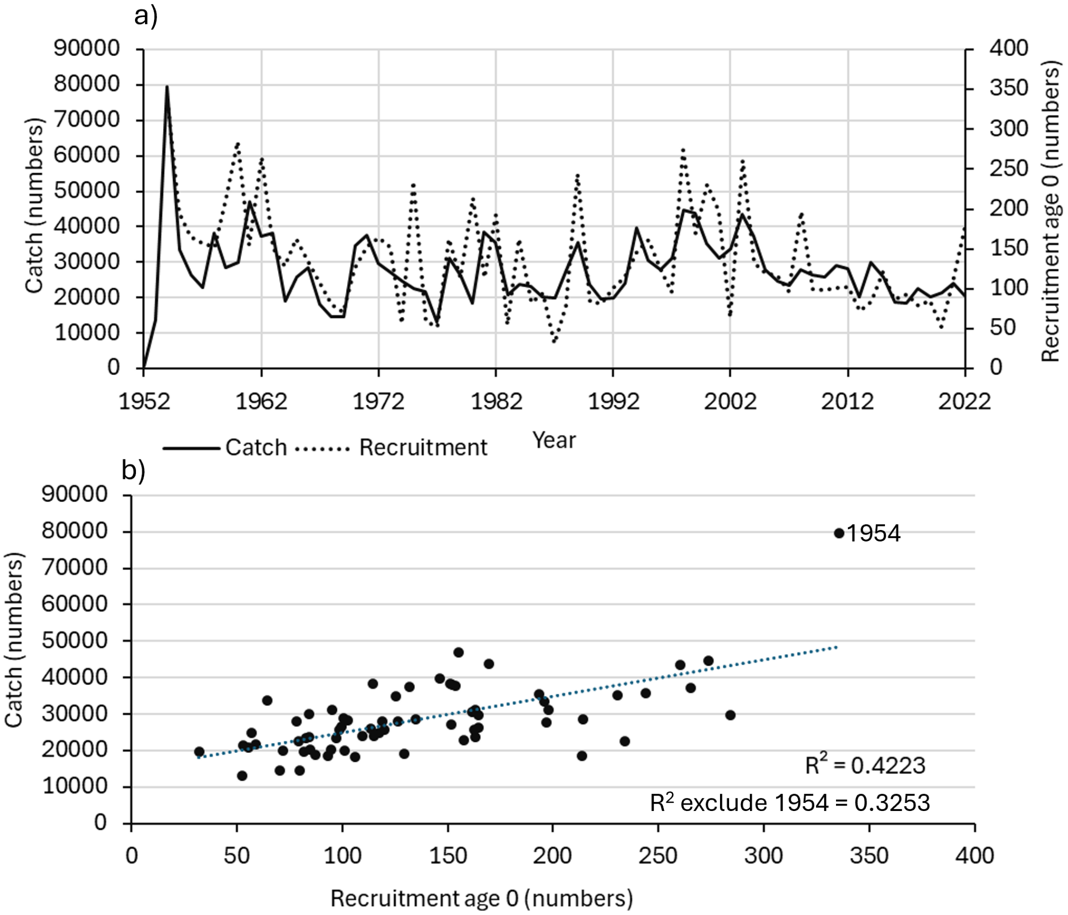


Figure 27: a) Relationship between model estimated recruitment (lagged forward 2 years) and catch in number time series, and b) linear regression, for the diagnostic model.

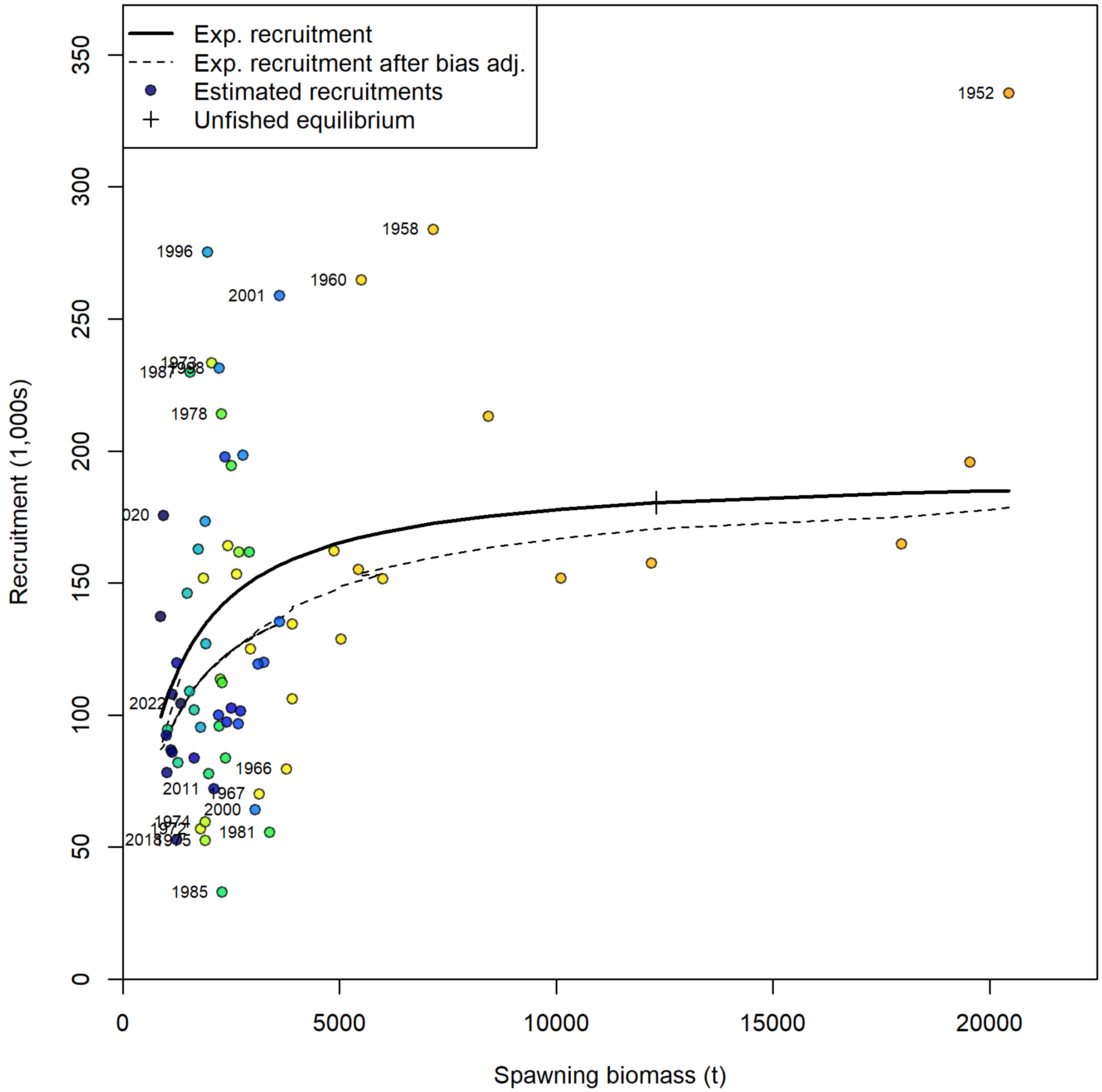


Figure 28: Estimated relationship between recruitment and spawning biomass based on annual values for the diagnostic model. Darker colours are more recent years.

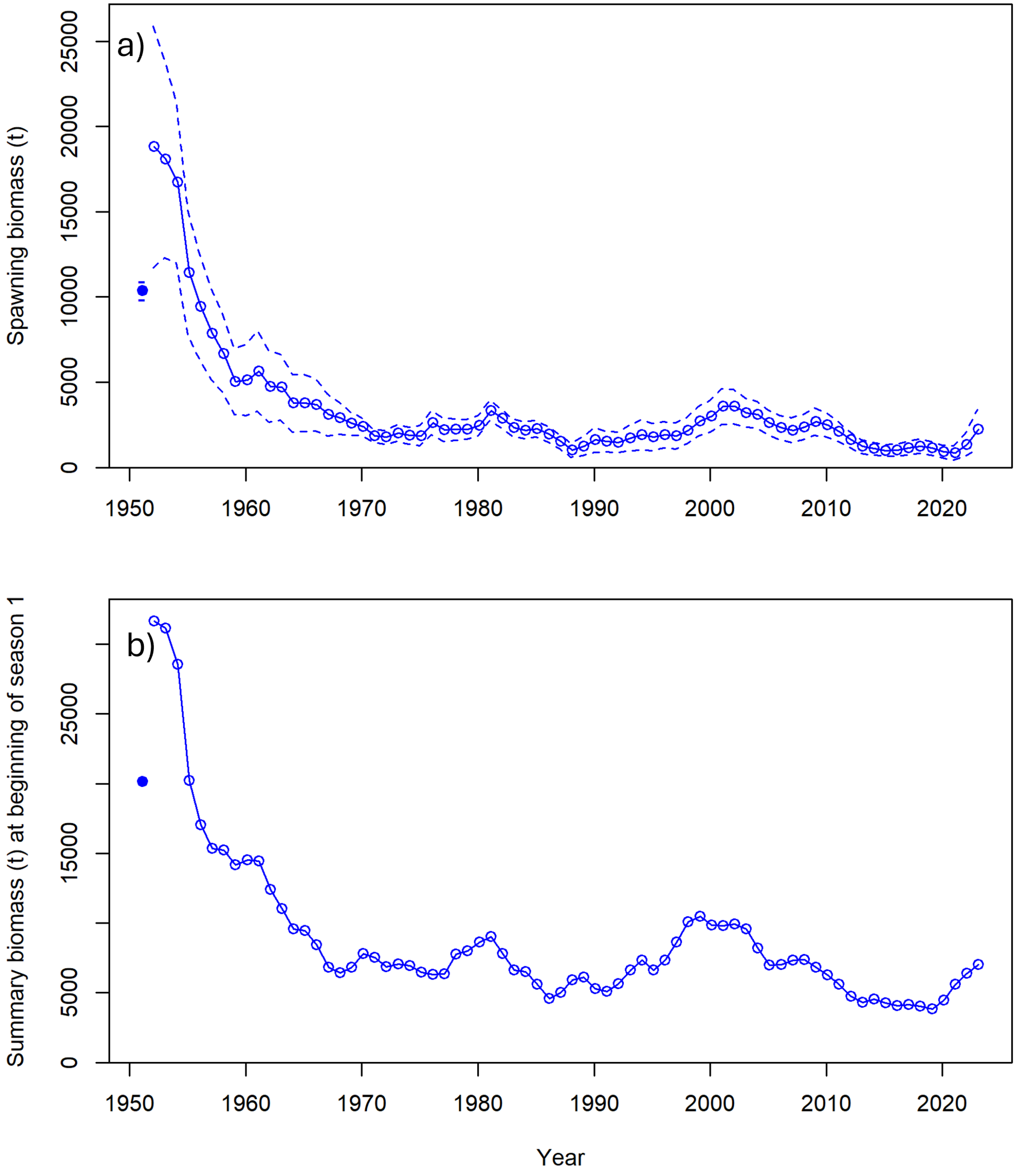


Figure 29: Spawning biomass (SB) (with 95% asymptotic intervals) and total (summary) biomass (TB) time series for the diagnostic model. Note spawning biomass for a single sex model is for the model population it is not ‘female’ spawning biomass.

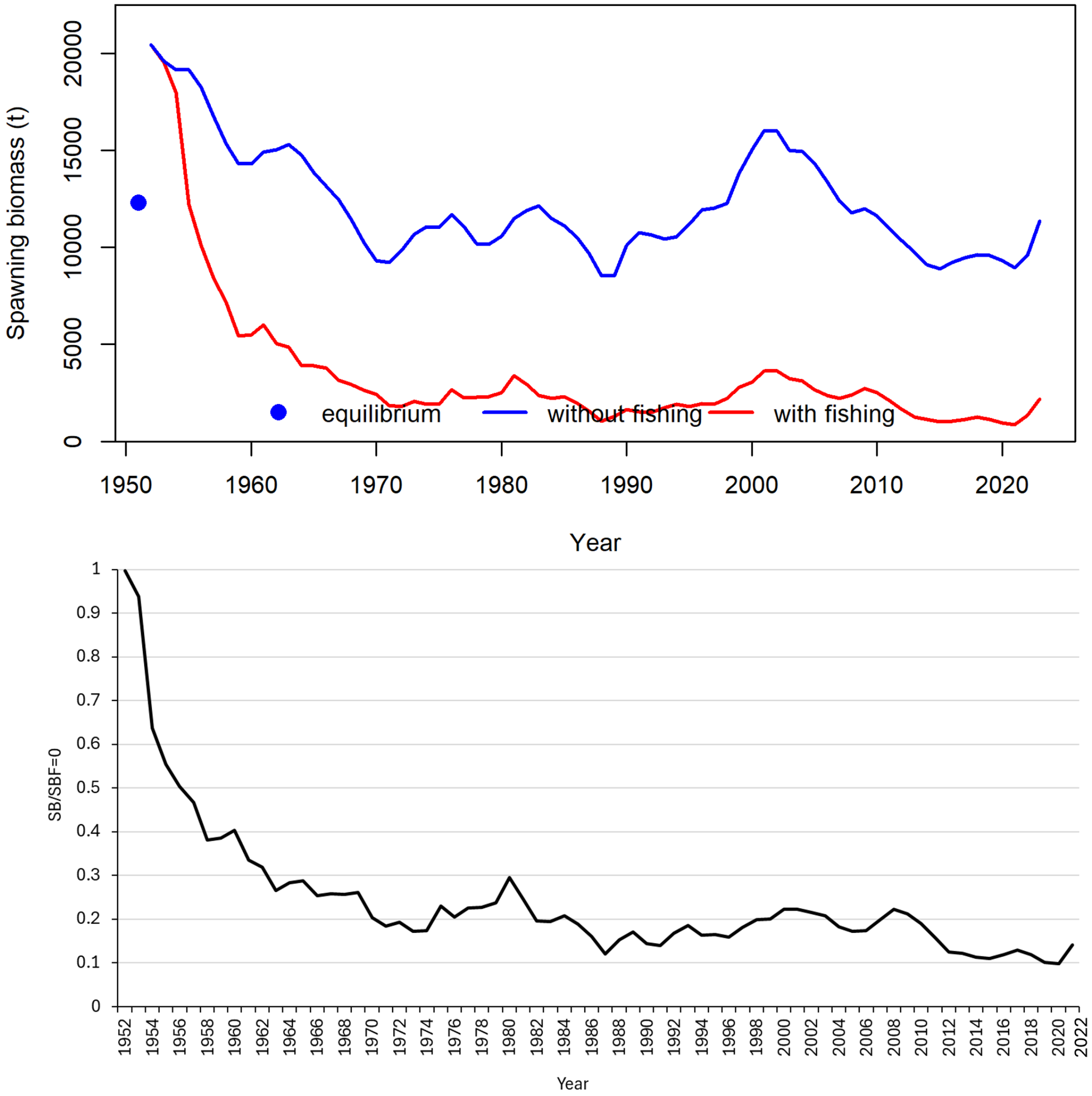


Figure 30: Estimated unfished ($SB_{F=0}$) and fished spawning biomass (SB) (top) and the annual dynamic spawning biomass depletion $SB/SB_{F=0}$ (bottom) for the diagnostic model.

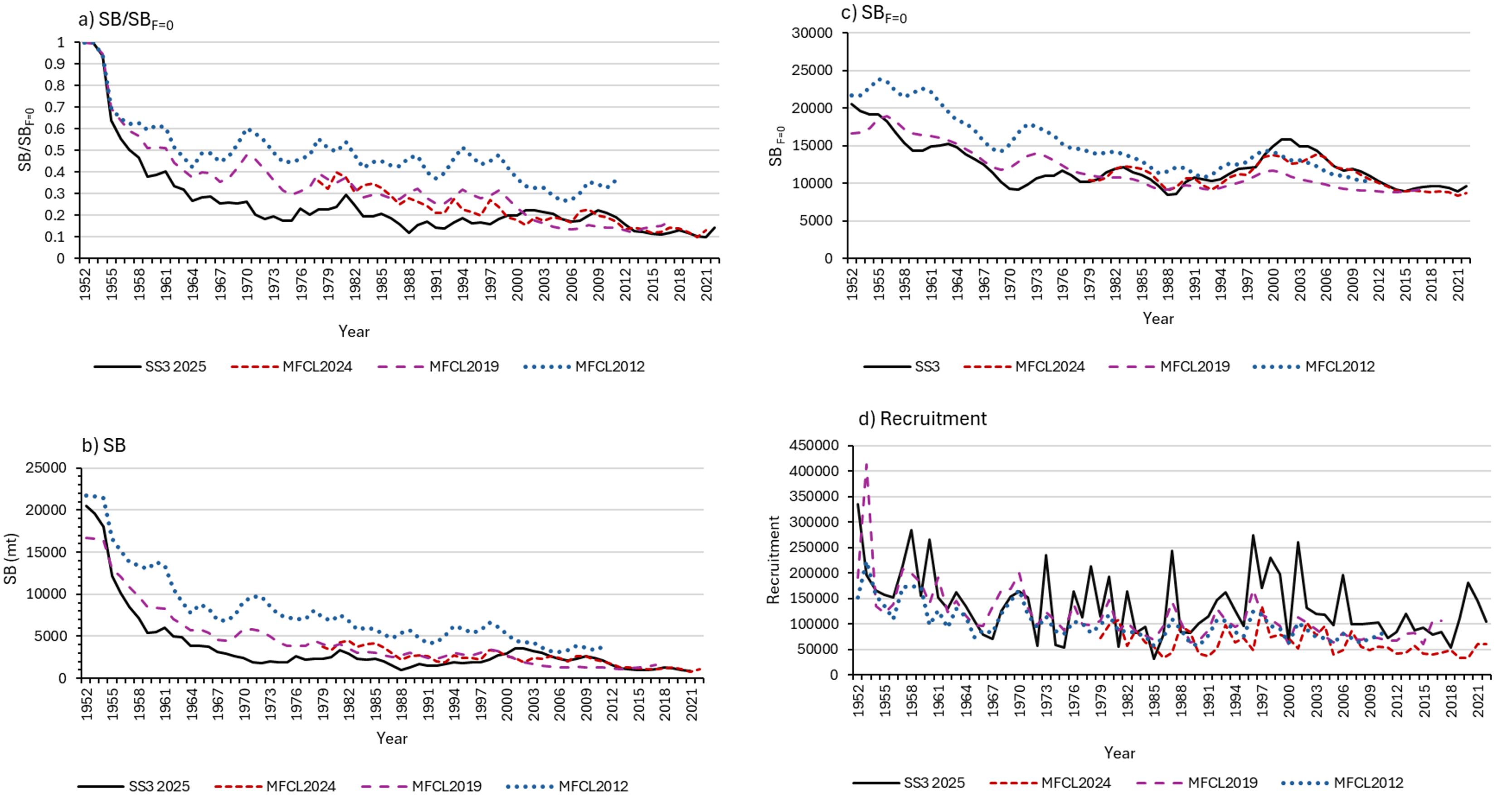


Figure 31: Comparison of spawning biomass (unfished and fished), dynamic spawning biomass depletion, $SB/SB_{F=0}$, and recruitmemt trajectories for the MFCL2024 and 2019 diagnostic models (dashed lines) and the SS3 diagnostic model (dark solid line).

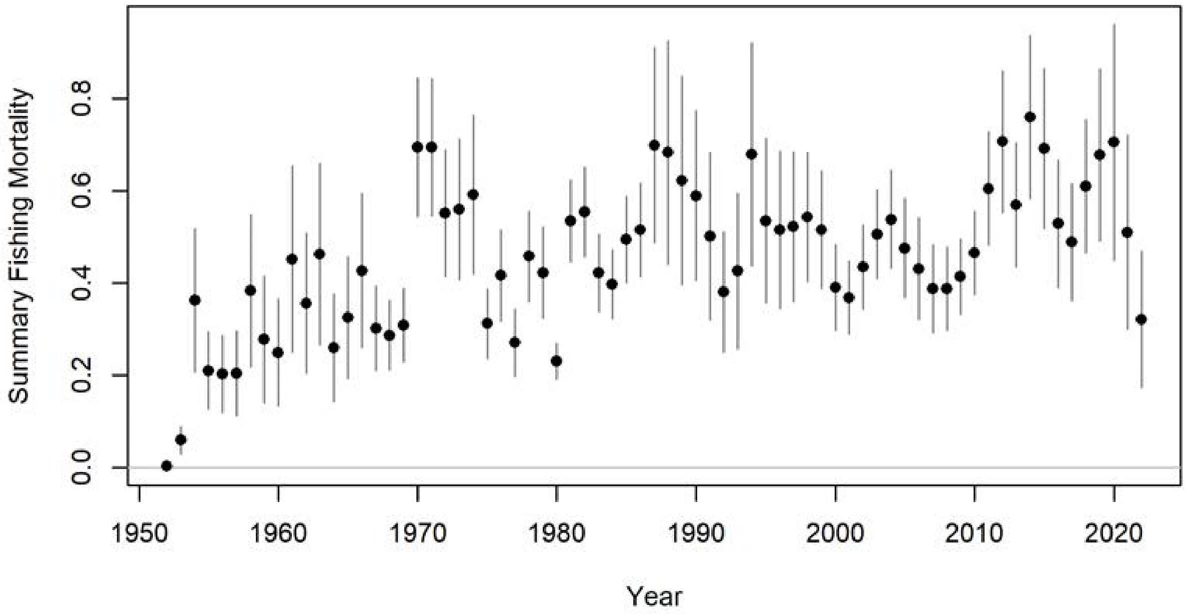


Figure 32: Estimated annual F (F calculated mean across age classes 2-10 year, numbers weighted, 95% asymptotic intervals) for the diagnostic model.

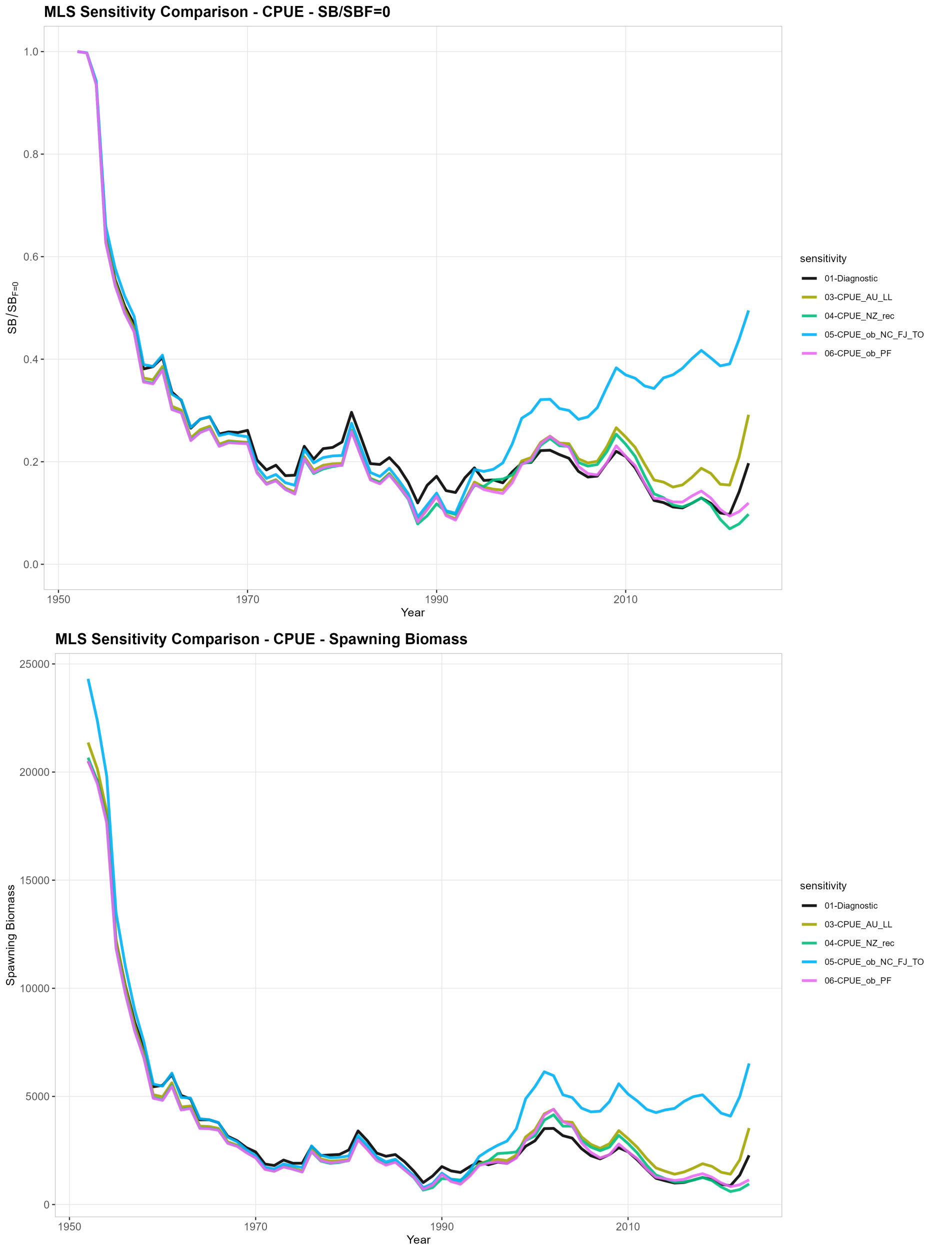
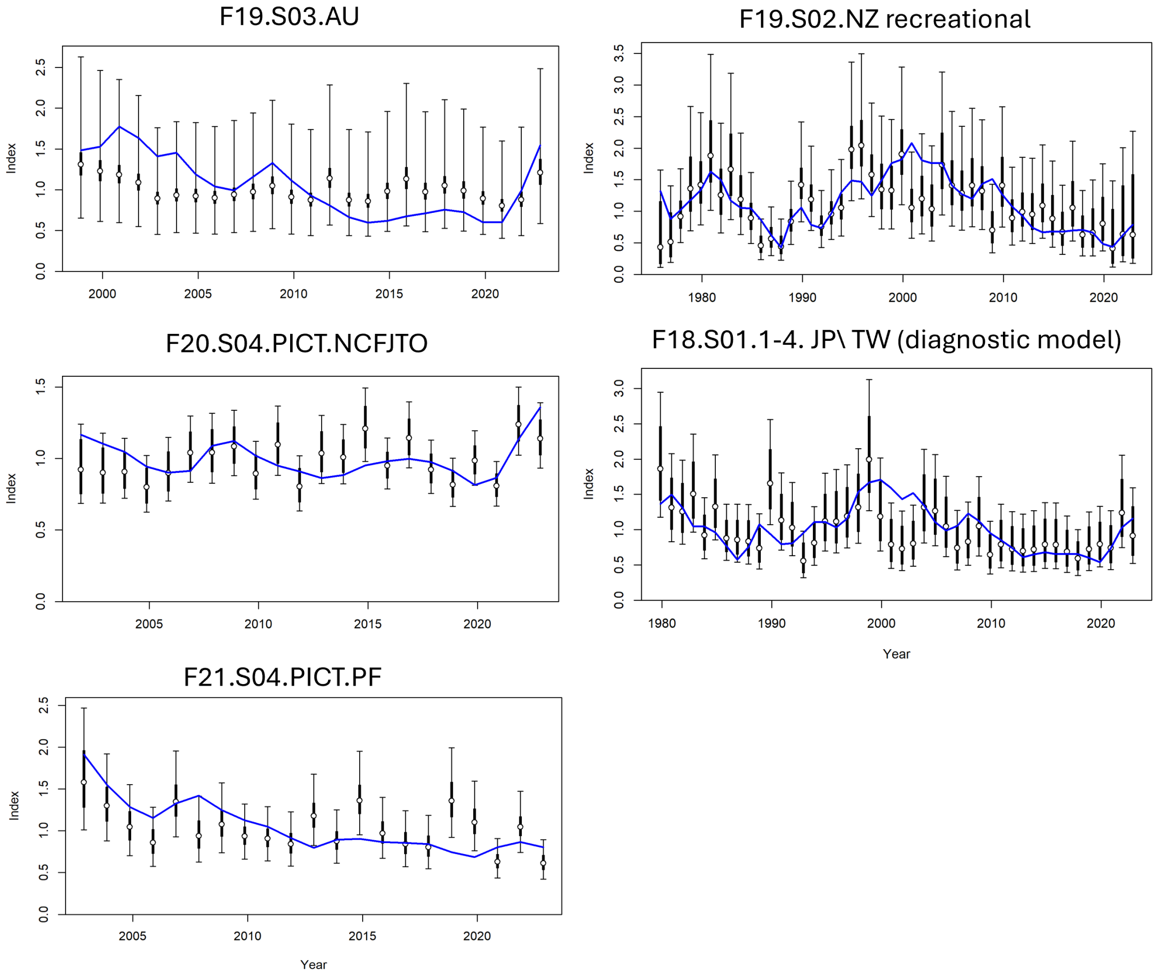


Figure 33: Comparison of spawning biomass depletion, $SB/SB_{F=0}$ (top), and spawning biomass (bottom) trajectories for the alternative CPUE models.



F19.S03.AU

F19.S02.NZ recreational

F20.S04.PICT.NCFJTO

F18.S01.1-4. JP\ TW (diagnostic model)

F21.S04.PICT.PF

Figure 34: Comparison of CPUE fits for alternative CPUE indices run as one off sensitivities.

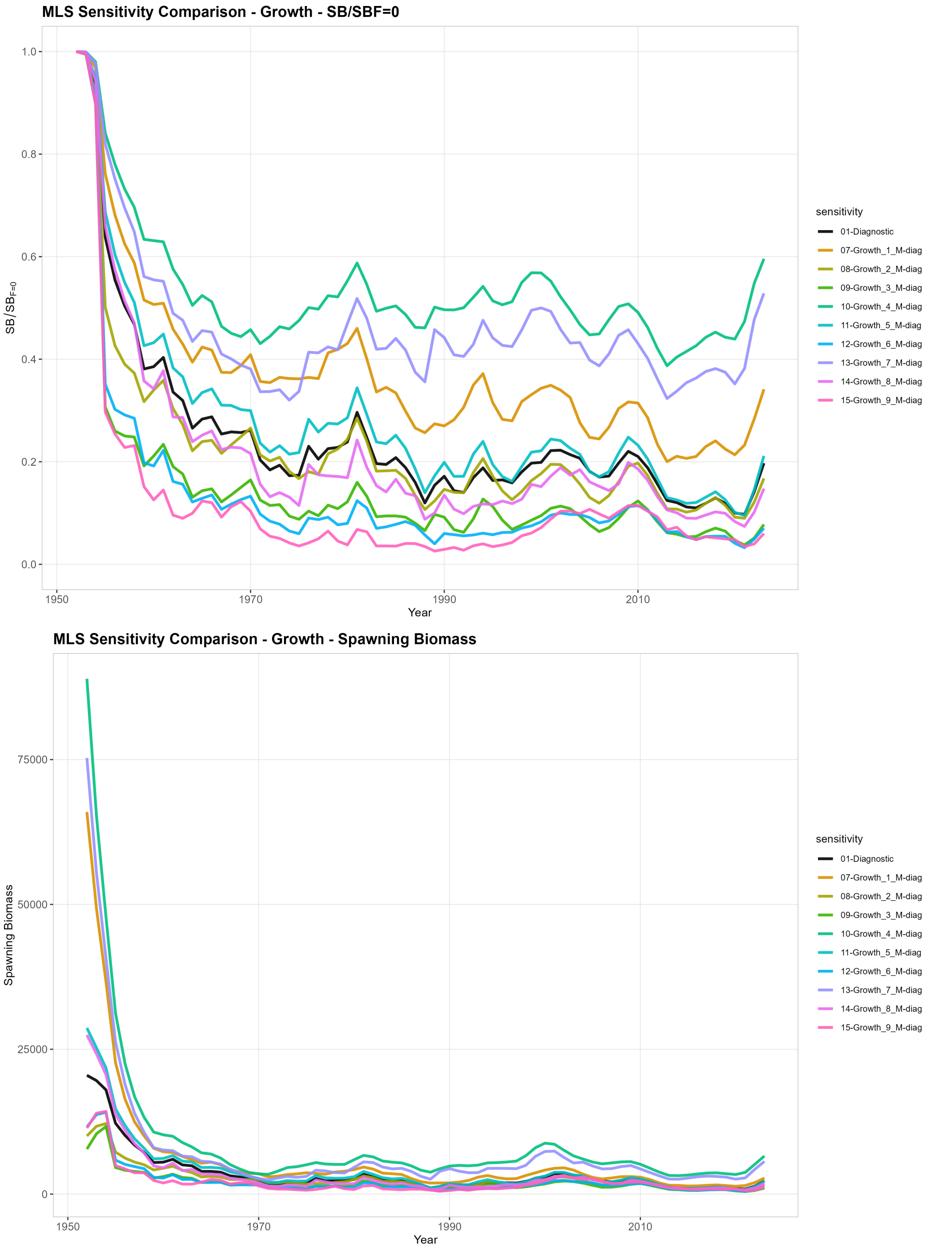


Figure 35: Comparison of spawning biomass depletion, $SB/SB_{F=0}$ (top), and spawning biomass (bottom) trajectories for the alternative growth models, all use the same M-at-age from the diagnostic model. **01**.Diagnostic - internal. **07**.Growth-1: L2=195cm, K=0.2. **08**.Growth-2: L2=210cm, K=0.2. **09**.Growth-3: L2=225cm, K=0.2. **10**.Growth-4: L2=195cm, K=0.5. **11**.Growth-5: L2=210cm, K=0.5. **12**.Growth-6: L2=225cm, K=0.5. **13**.Growth-7: L2=195cm, K=0.8. **14**.Growth-8: L2=210cm, K=0.8. **15**.Growth-9: L2=225cm, K=0.8

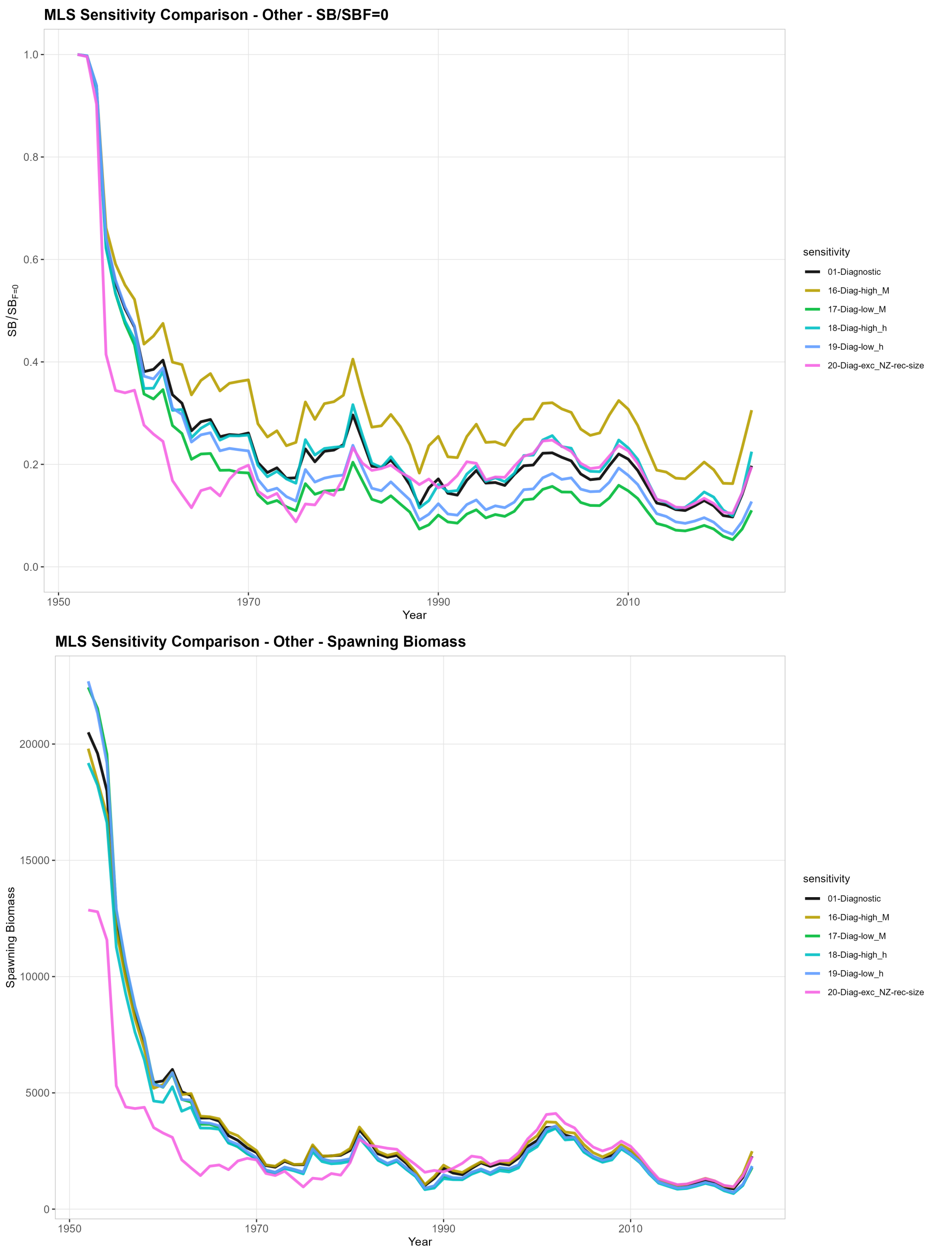
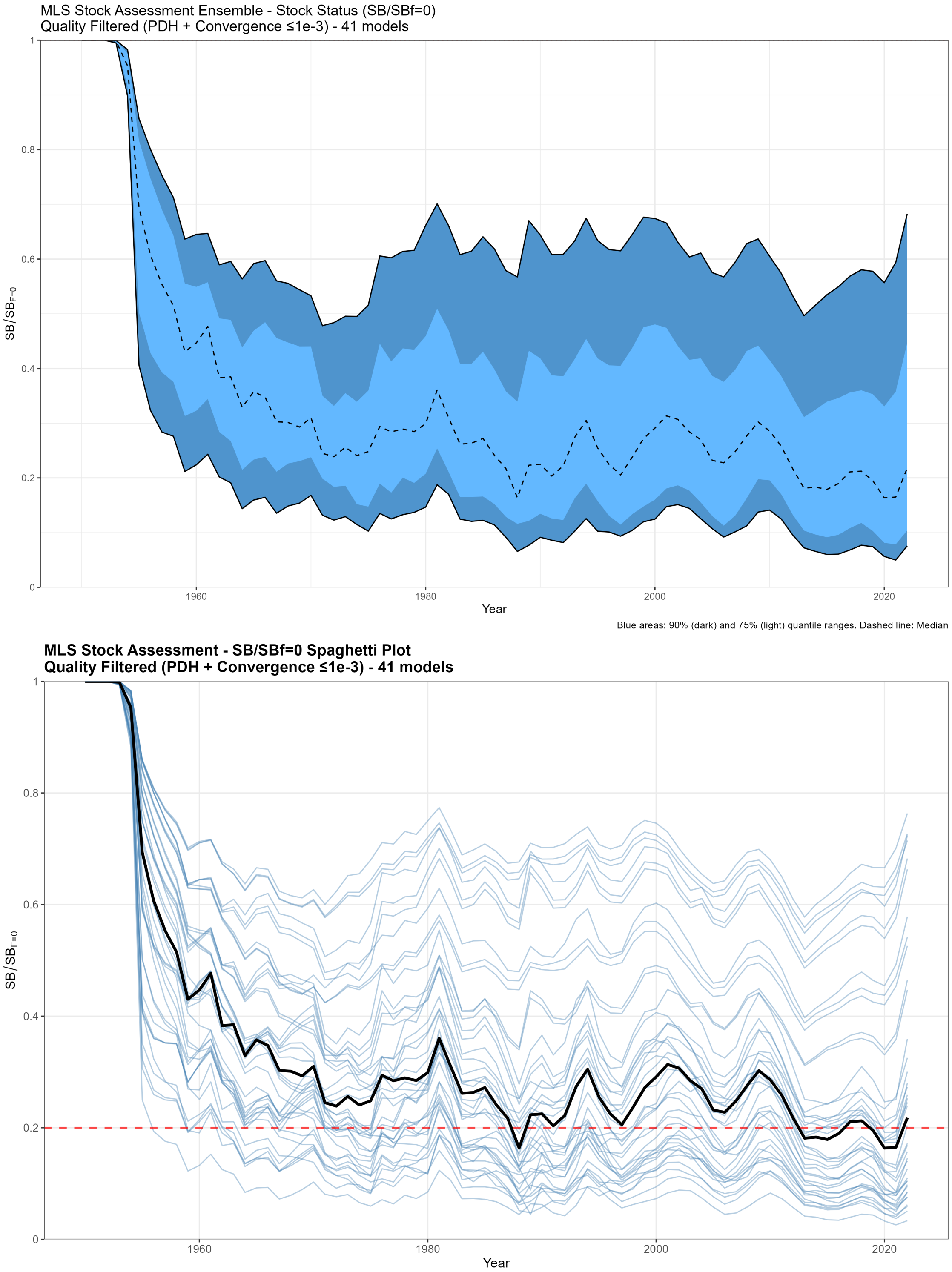


Figure 36: Comparison of spawning biomass depletion, $SB/SB_{F=0}$ (top), and spawning biomass (bottom) trajectories for the alternative M-at-age, steepness, and the model were the New Zealand recreational weight data were removed. **01**.Diagnostic h=0.8. **16.**Diag-high: diag M x 1.35. **17**.Diag-low: diag M x 0.65. **18.**Diag-high h=0.95. **19**.Diag-low h=0.75 **20**.Diag-exclude-NZ-rec-size



MLS Stock Assessment Ensemble - Stock Status (SB/SBf=0)
Quality Filtered (PDH + Convergence ≤1e-3) - 41 models

Blue areas: 90% (dark) and 75% (light) quantile ranges. Dashed line: Median

MLS Stock Assessment - SB/SBf=0 Spaghetti Plot
Quality Filtered (PDH + Convergence ≤1e-3) - 41 models

Figure 37: Ribbon plot (top) and spaghetti plot (bottom, lines are individual models) of $SB/SB_{F=0}$ across 41 grid models, dashed line is median, and for the ribbon plot, dark blue is the 90% quantile and light blue is the 75% quantile.

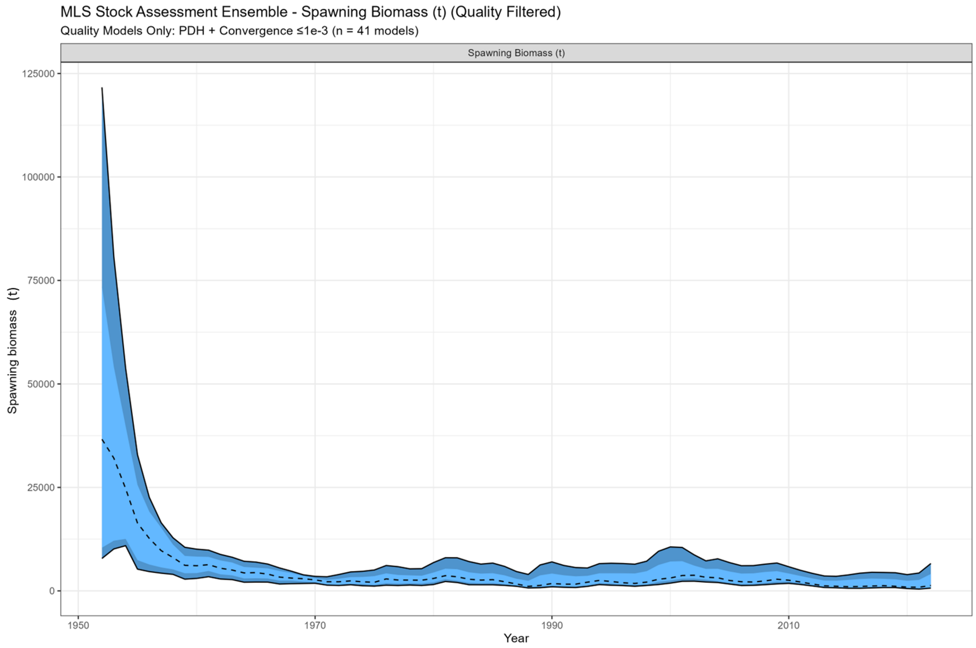


Figure 38: Ribbon plot of spawning biomass trajectory across 41 grid models, dashed line is median, dark blue is the 90% quantile and light blue is the 75% quantile.

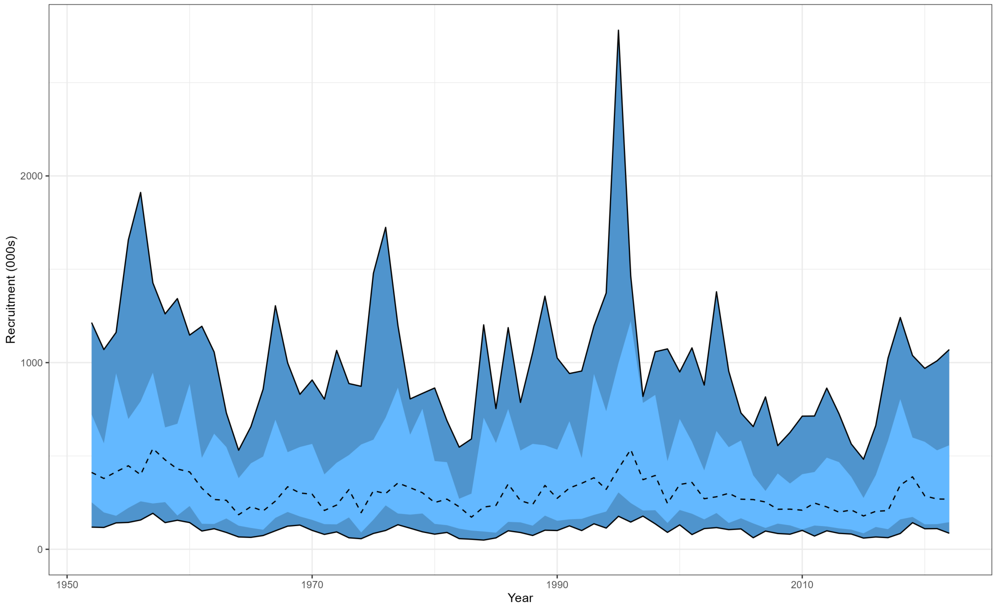


Figure 39: Ribbon plot of recruitment trajectory across 41 grid models, dashed line is median, dark blue is the 90% quantile and light blue is the 75% quantile.

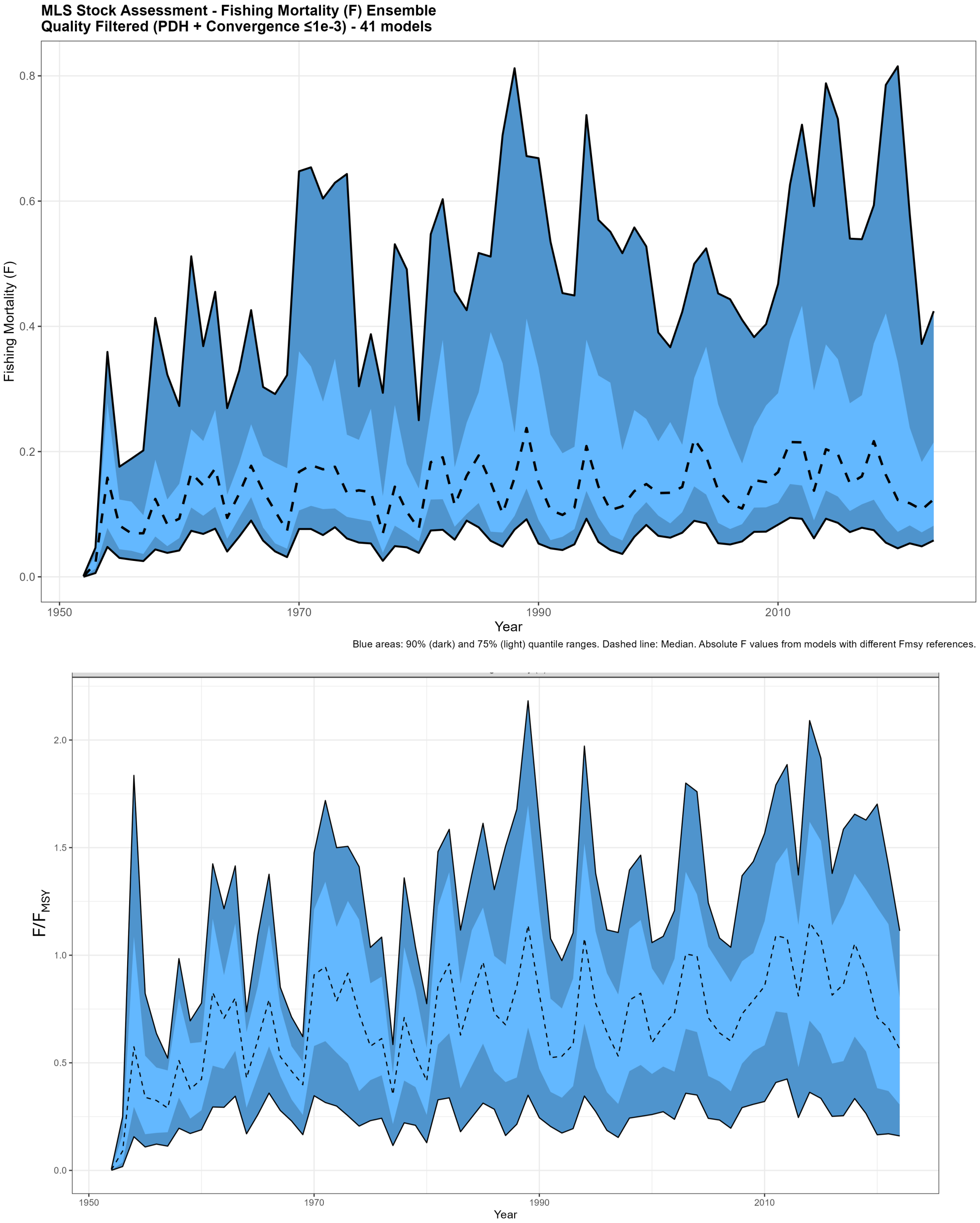


Figure 40: Ribbon plot of instantaneous fishing mortality (F) (top) and F/F_{MSY} (bottom) across 41 grid models, dashed line is median, dark blue is the 90% quantile and light blue is the 75% quantile.

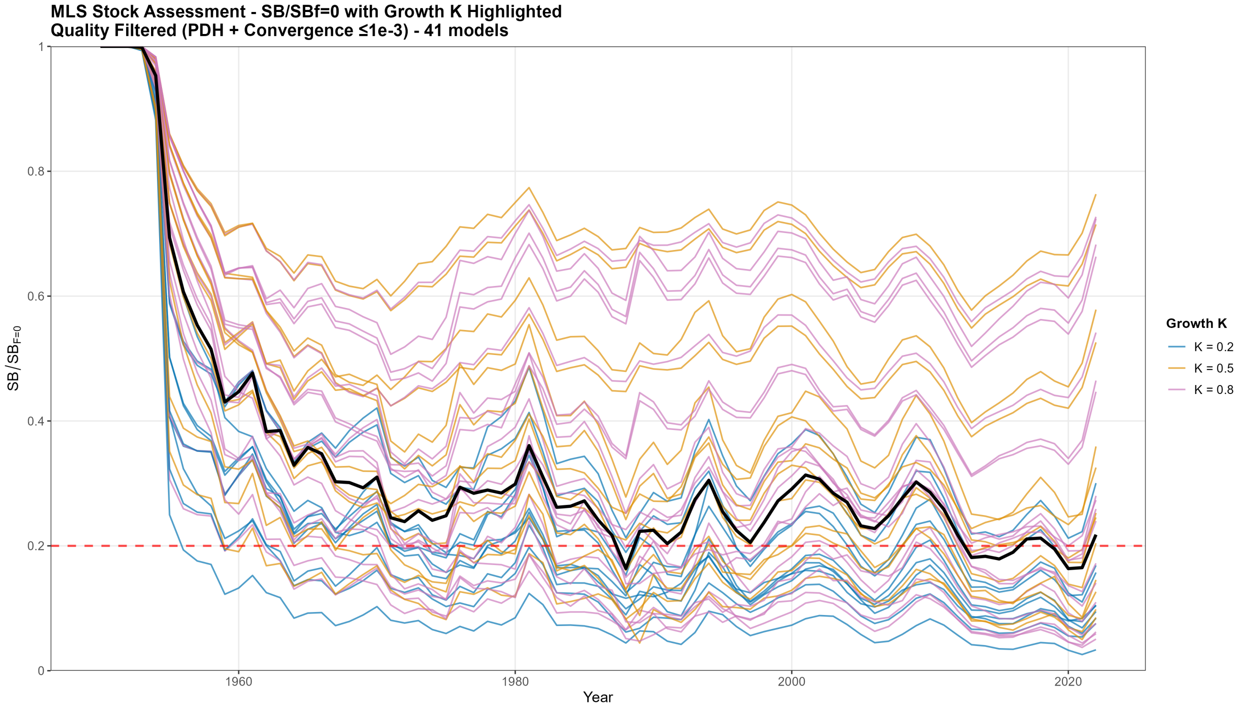


Figure 41: Spaghetti plot of $SB/SB_{F=0}$ across 41 grid models grouped according to the growth parameter K.

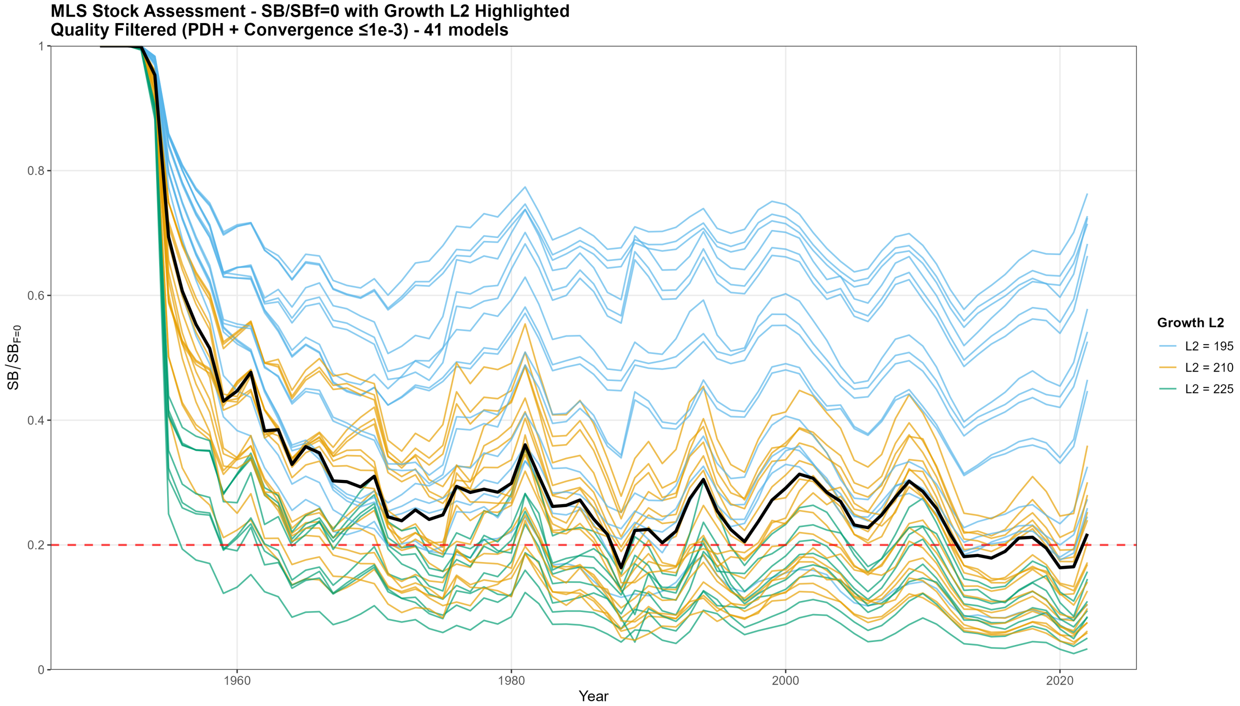


Figure 42: Spaghetti plot of $SB/SB_{F=0}$ across 41 grid models grouped according to the growth parameter L2.

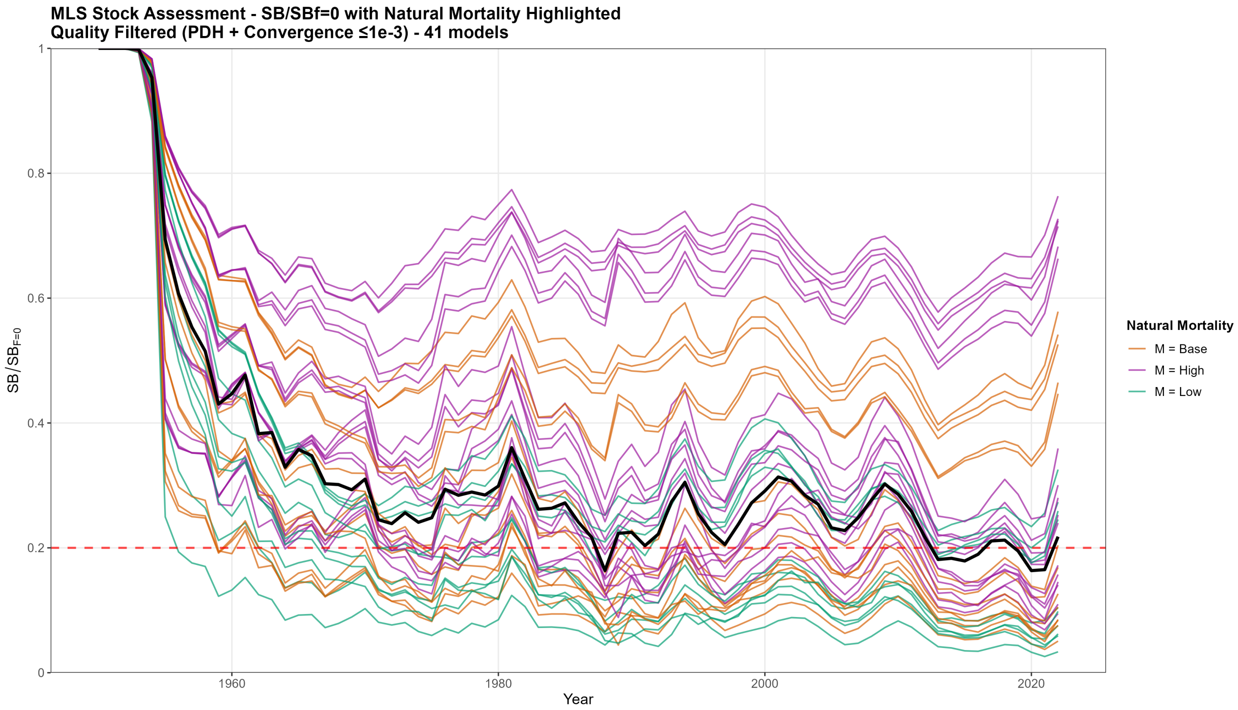


Figure 43: Spaghetti plot of $SB/SB_{F=0}$ across 41 grid models grouped according to the M-at-age.

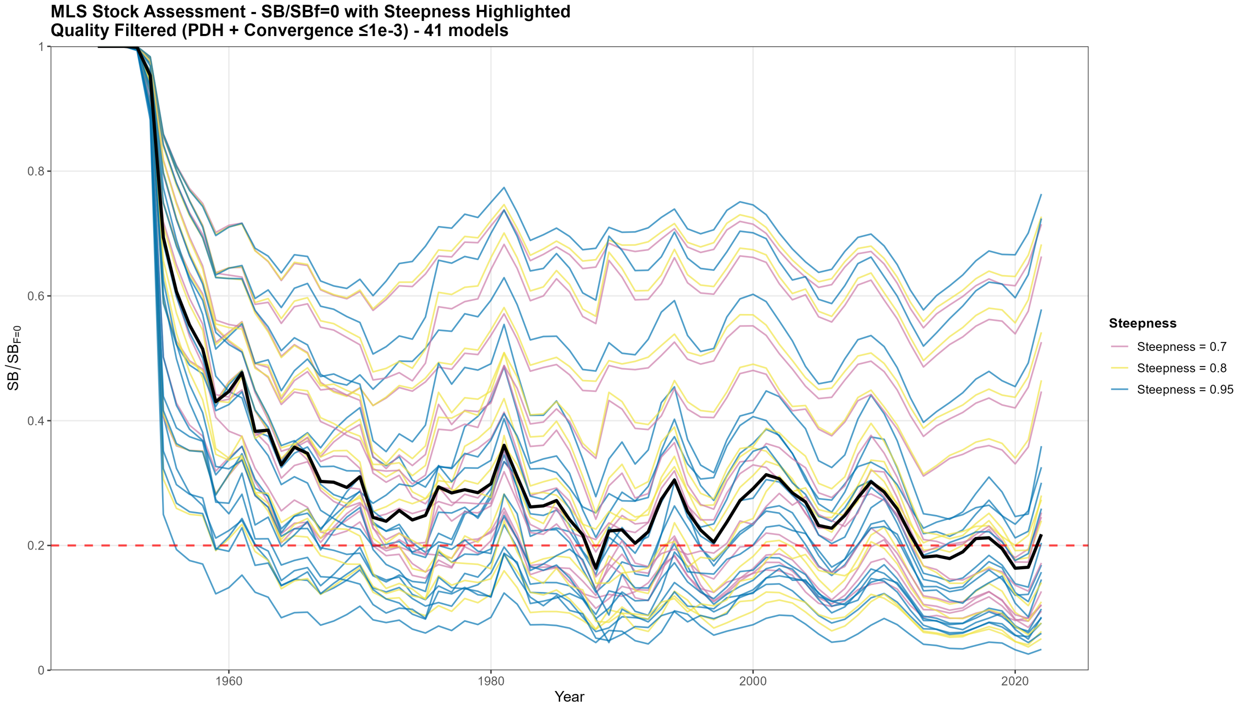


Figure 44: Spaghetti plot of $SB/SB_{F=0}$ across 41 grid models grouped according to the steepness parameter.

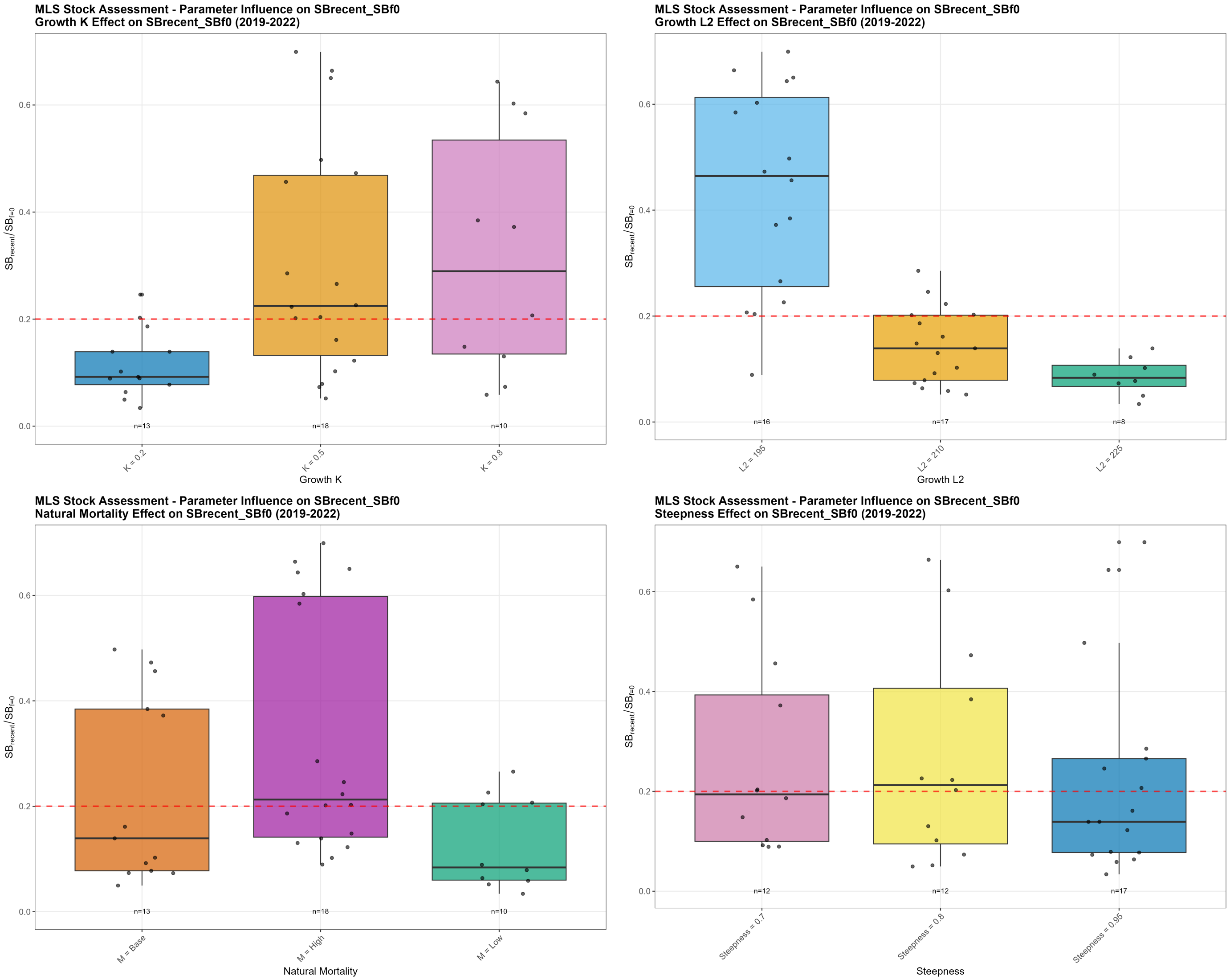


Figure 45: Box plots of $SB_{recent}/SB_{F=0}$ across 41 grid models grouped according to the four biological grid axes. Recent=2019–2022, dots are the individual models. Recent=2019–2022. Note the M=base is the diagnostic model level and is intermediate between M=high and M=low.

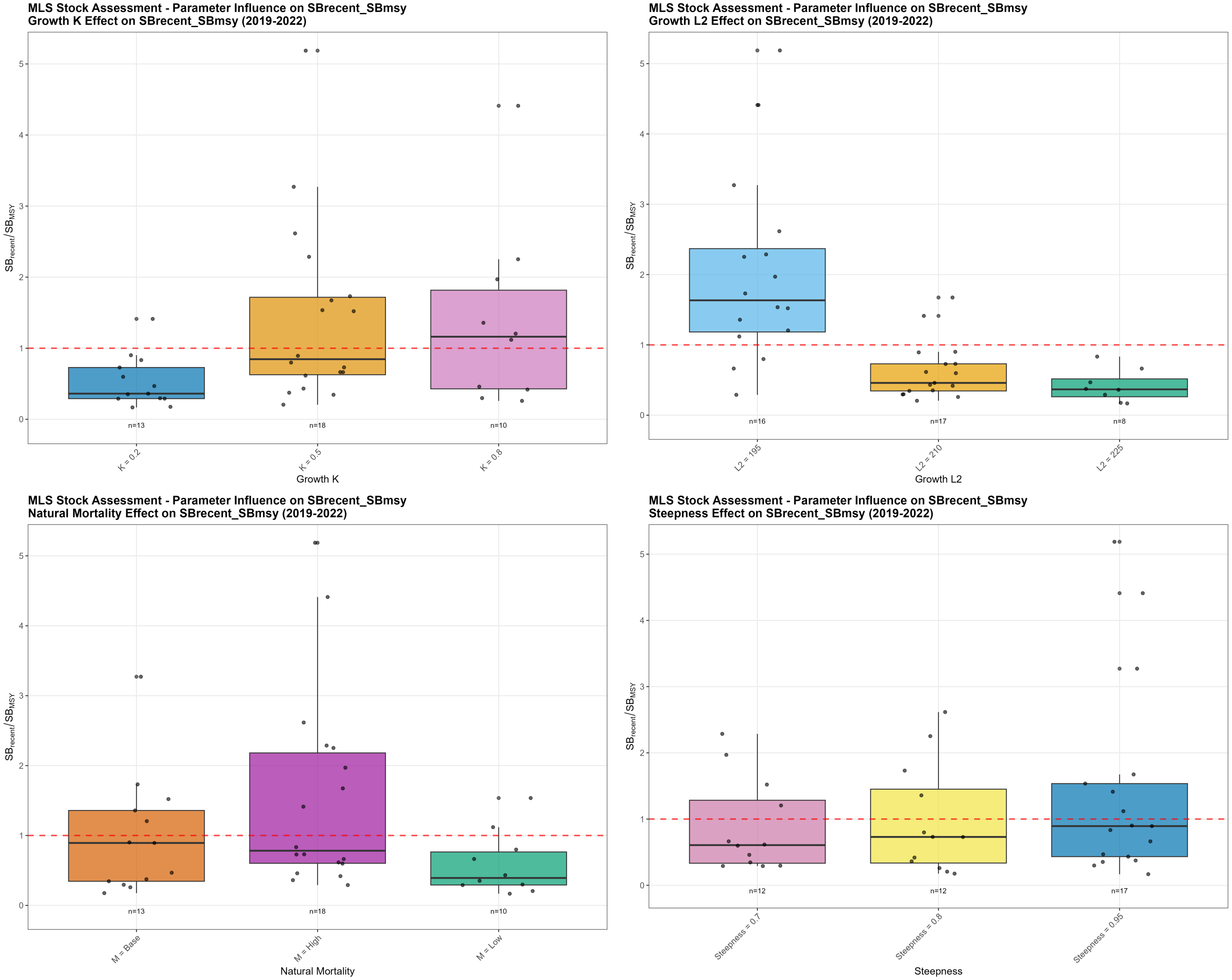
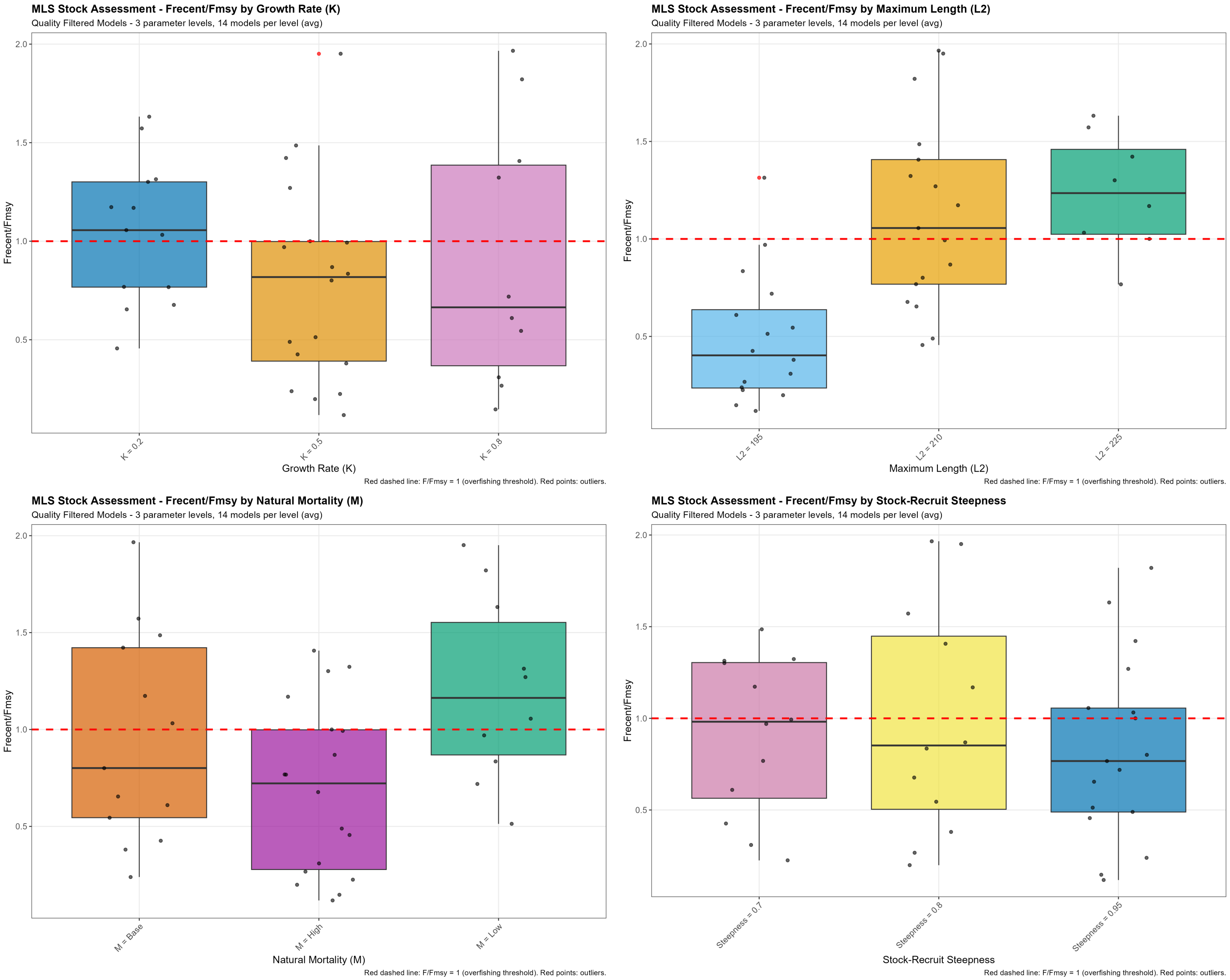


Figure 46: Box plots of SB_{recent}/SB_{MSY} across 41 grid models grouped according to the four biological grid axes. Recent=2019–2022, dots are individual models. Note the M=base is the diagnostic model level and is intermediate between M=high and M=low.



MLS Stock Assessment - Frecent/Fmsy by Growth Rate (K)

Quality Filtered Models - 3 parameter levels, 14 models per level (avg)

MLS Stock Assessment - Frecent/Fmsy by Maximum Length (L2)

Quality Filtered Models - 3 parameter levels, 14 models per level (avg)

MLS Stock Assessment - Frecent/Fmsy by Natural Mortality (M)

Quality Filtered Models - 3 parameter levels, 14 models per level (avg)

MLS Stock Assessment - Frecent/Fmsy by Stock-Recruit Steepness

Quality Filtered Models - 3 parameter levels, 14 models per level (avg)

Red dashed line: F/Fmsy = 1 (overfishing threshold). Red points: outliers.

Figure 47: Box plots of F_{recent}/F_{MSY} across 41 grid models grouped according to the four biological grid axes. Recent=2018–2021. Note the M=base is the diagnostic model level and is intermediate between M=high and M=low.

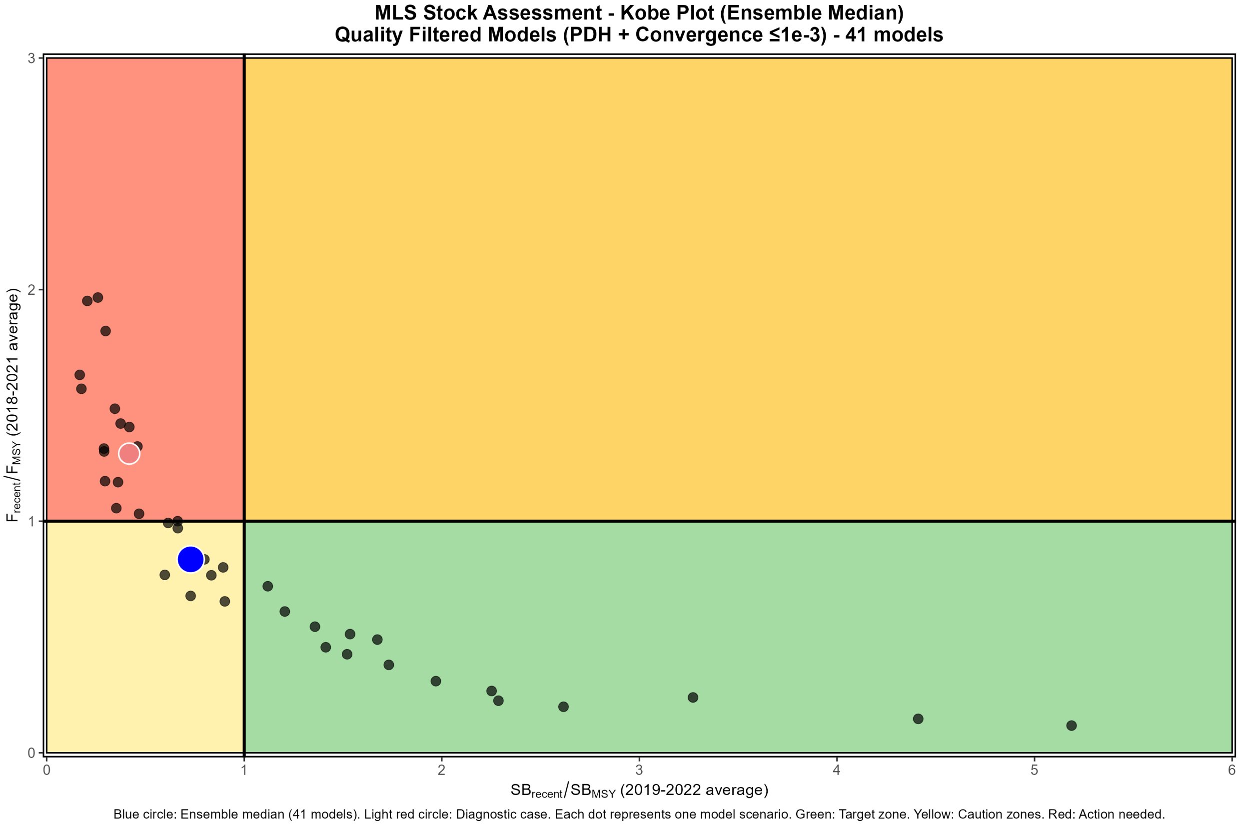


Figure 48: Kobe plot summarising the results across each of the 41 retained models in the model grid and the diagnostic model for the recent period (2019–2022). The blue point represents the median and the white circle is the diagnostic model.

14 Appendix

14.1 Extreme downweight of size data

A diagnostic model was run with all size data effective sample sizes reduce by 10 times those in the diagnostic model (that are already downweighted). This model was aimed at further improving the fits to the CPUE index and downweighting the influence of size data on population estimates (scale). The results showed as expected, improved fit to the CPUE, but only a minor increase in the biomass scale.

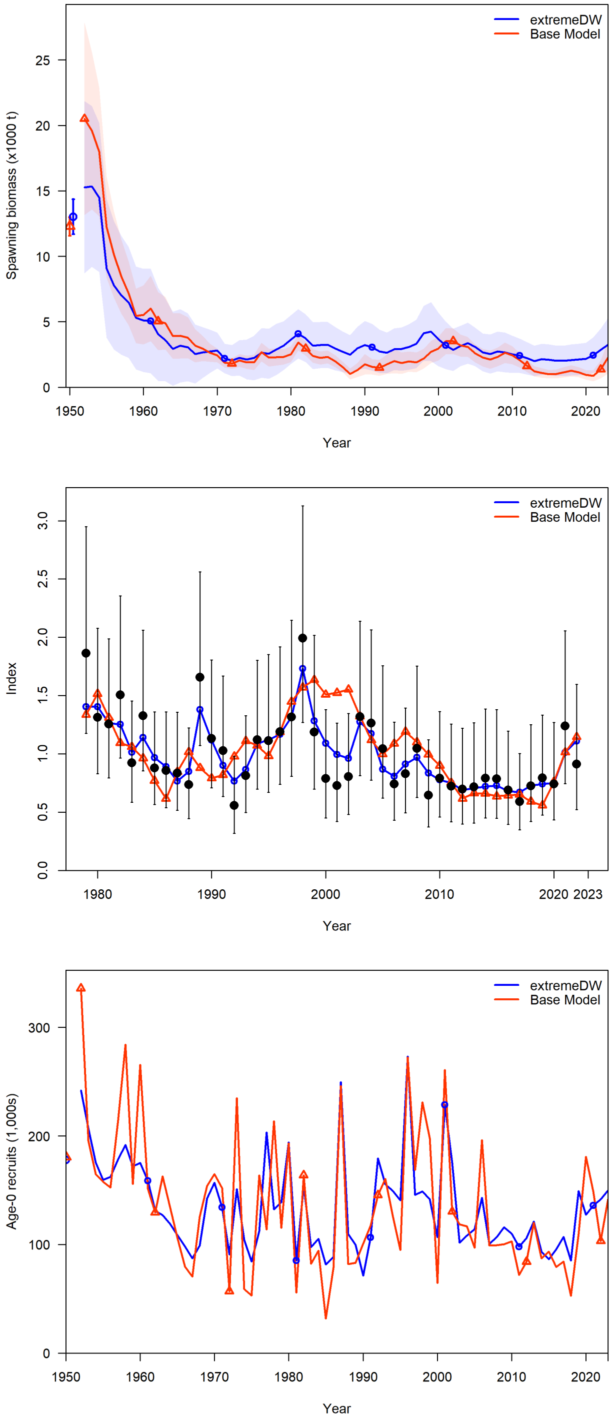


Figure 49: Plots of spawning biomass trajectories (top), fit to CPUE index (mid) and recruitment trajectories, compared between the diagnostic model (base) and the model (red) with extreme downweighting of size date.

14.2 High σ_R

A diagnostic model was run with the (σ_R) doubled to 1.2, to see what influence reducing the penalty on recruitment deviations and allowing more influence of the other data components would have on population scale. This experiment showed minor impact on population scale when a higher σ_R was used.

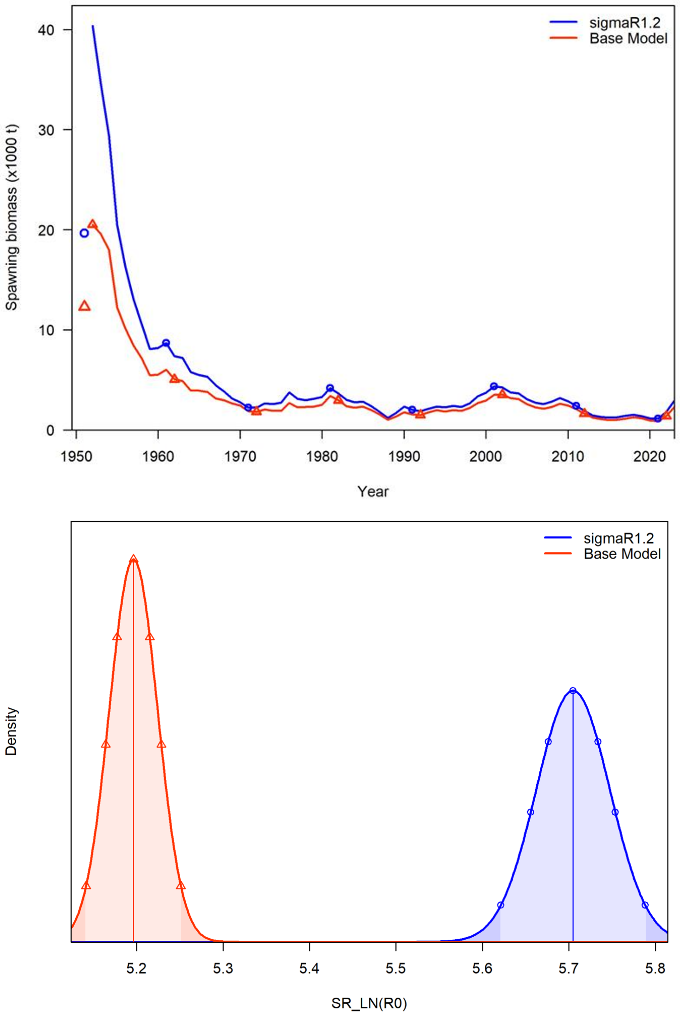


Figure 50: Plots of spawning biomass trajectories (top), and LN(R0) distribution (bottom) for the diagnostic model with double σ_R (1.2).

14.3 ASPM with relax R0 penalty

ASPM model experiment was run with a model with a simple prior constraining R0 and a model with no R0 prior but a high (σ_R)=1 allowing a low penalty on recruitment deviation to fit the data. The later model was able to fit the CPUE, and produced a similarly low spawning biomass and recruitment scale as the integrated assessment. The production function likely exists but the existing Stock Synthesis 3 model may be mis-specified and too constrained to estimate a production function from the data.

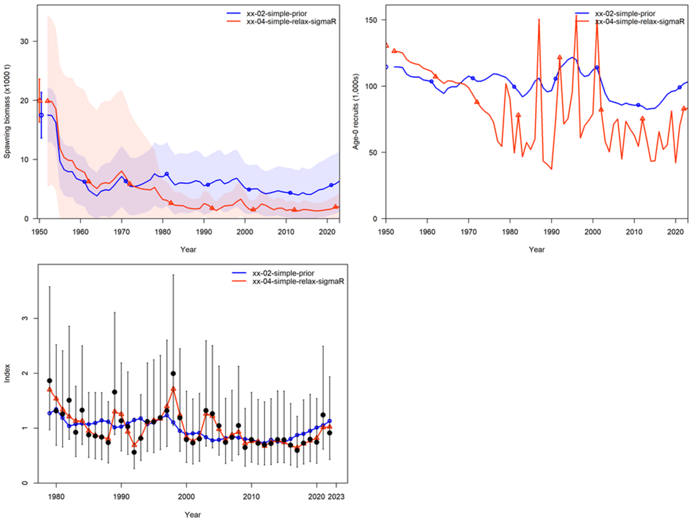


Figure 51: Plots of spawning biomass (top left), recruitment (top right) and fits to the JP/TW CPUE index (bottom left) for two alternatve ASPM models, one with a simple production prior (blue) and one with a reduces penalty of the recruitment deviation (σ_R)=1 (red).

14.4 Catch v CPUE indices

The plots below of total catch in numbers (all fisheries) versus the two longest CPUE abundance indices available (JP/TW and NZ recreational) are included in relation to the interpretation of the ASPM model which could not fit JP/TW CPUE or estimate a production function from the catch and CPUE contrast. The plots show that the CPUE and catch follow similar trends, with no strong contrasts. While similar trends in catch and CPUE might be expected when effort is stable, longline effort in hooks fished across the model region has increased by 3 x over the time period for which the CPUE indices cover (see figure 1 in Ducharme-Barth (2025)). This may suggest the CPUE is not a reliable stock abundance indicator of long-term abundance trends but may be indicative of recruitment variability.

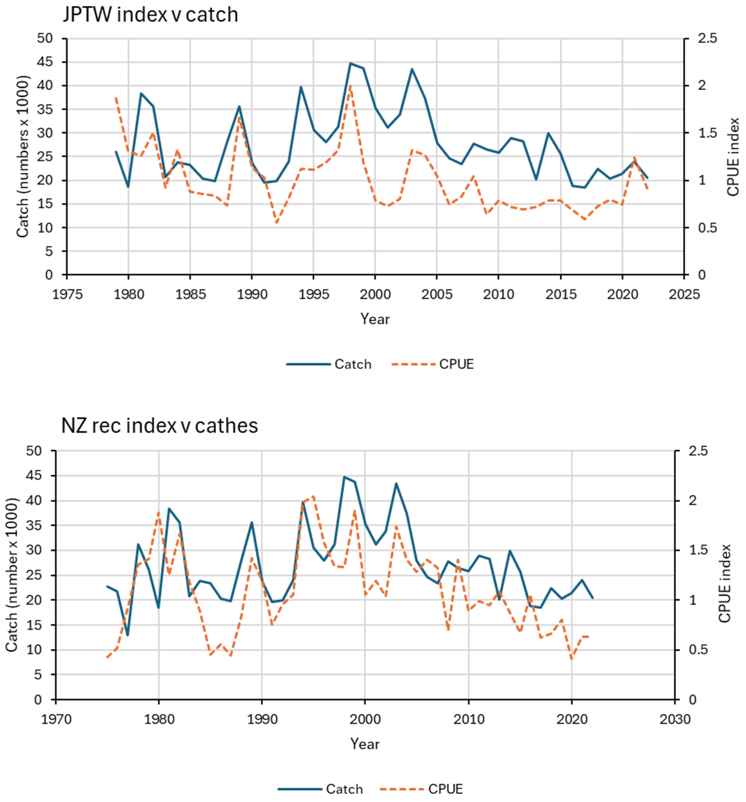
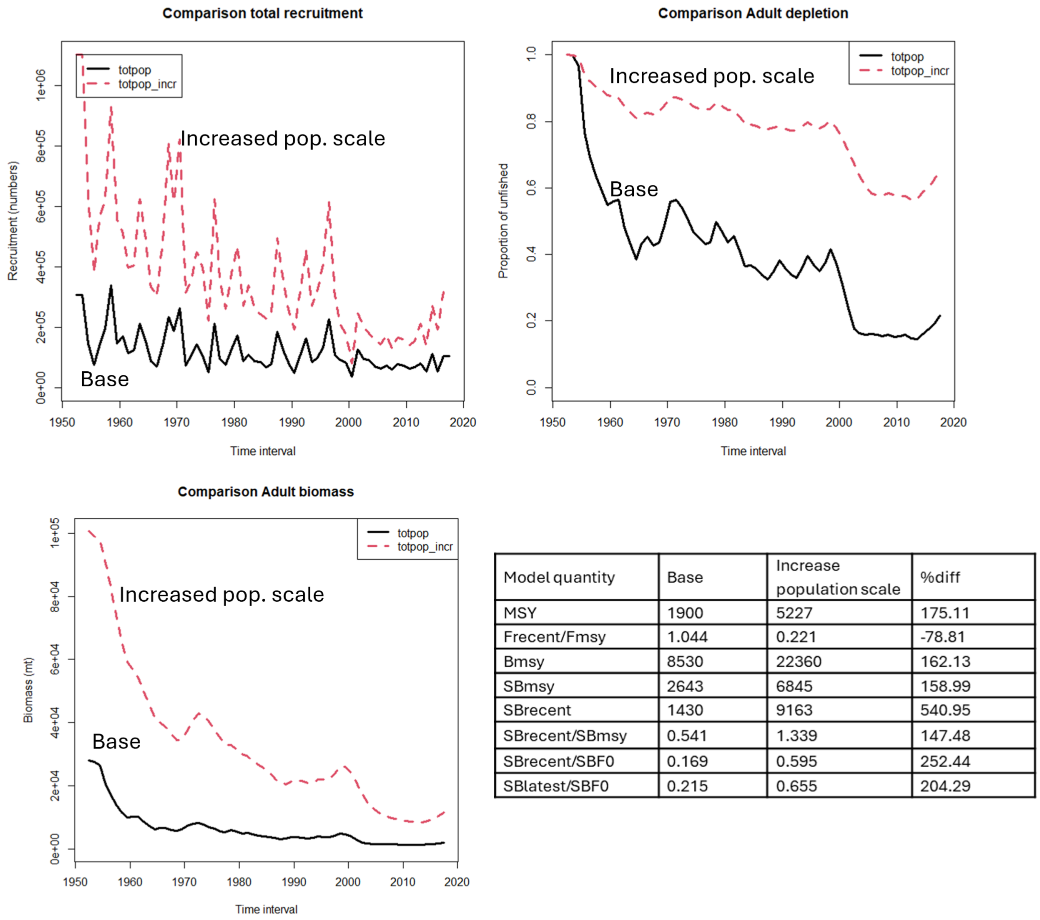


Figure 52: Plots of catch (numbers, all fisheries) versus CPUE indices.

14.5 Reference point sensitivity to increase population scale

Experiments were run in SS3 and MFCL with striped marlin models where the mean recruitment was scaled up, and the model refitted (i.e. the catch/fishing levels are the same).

In the MFCL model the mean recruitment was scaled up by 3 times, and the resultant biomass scaled up accordingly. The results below show the strong effects on the management reference points of assuming a higher population scale but the same catch levels. F/F_{MSY} was reduced by almost 78% and, $SB/SB_{F=0}$ was increased (less depleted) by 252%.



Model quantity	Base	Increase population scale	%diff
MSY	1900	5227	175.11
Frecent/Fmsy	1.044	0.221	-78.81
Bmsy	8530	22360	162.13
SBmsy	2643	6845	158.99
SBrecent	1430	9163	540.95
SBrecent/SBmsy	0.541	1.339	147.48
SBrecent/SBF0	0.169	0.595	252.44
SBlatest/SBF0	0.215	0.655	204.29

Figure 53: Plots of recruitment (top left), unfished spawning biomass (top right) and spawning biomass (bottom left) for a base case model (black) and the same model refitted with an assumed 3 x higher mean recruitment (Increased pop. scale - red dashed). Table shows the comparison of the reference points.

In the SS3 example the R0 was increased x 4 in normal space. Consistent with MFCL, the F/F_{MSY} was substantially reduced (i.e. around 90% lower) and $SB/SB_{F=0}$ was increased (less depleted) compared to the diagnostic case when the higher population scale was assumed (below).

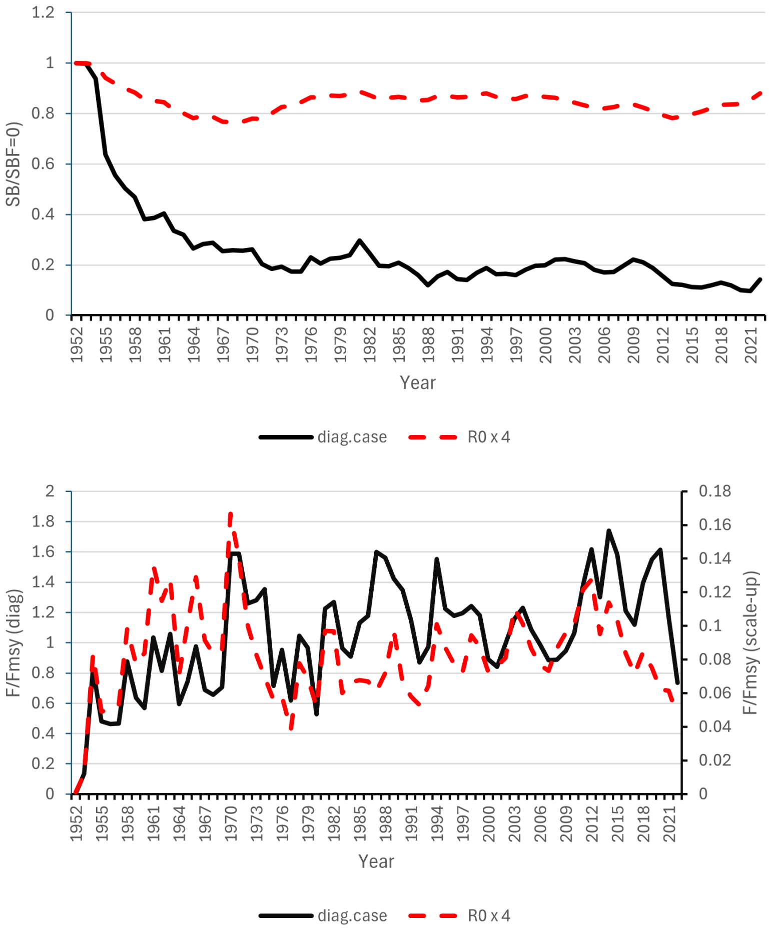


Figure 54: Time series of $SB/SB_{F=0}$ (top) F/F_{MSY} (bottom) compared between the SS3 diagnostic model (black) a model fitted with an increased population scale assumption (dashed red) and the same catches applied.