Western and Central Pacific Fisheries Commission

SCIENTIFIC COMMITTEE
TWENTIETH REGULAR SESSION

Manila, Philippines
14–21 August 2024

Stock Assessment of Striped Marlin in the Southwest Pacific Ocean: 2024

C. Castillo-Jordán[1]**, J. Day**[1]**, T. Teears** [1]**, N. Davies**[2]**, J. Hampton**[1]**, S. McKechnie**[1]**,**
A. Magnusson[1]**, T. Peatman**[3]**, T. Vidal**[1]**, P. Williams**[1]**, P. Hamer**[1]

---

[1]Oceanic Fisheries Programme of the Pacific Community
[2]TeTakina Ltd
[3]Private Consultant

Revision 1:

This version contains the full assessment report. There are some minor adjustments to some values for the reference points that include the estimation uncertainty at the bottom of Table 5.

Revision 2:

This version includes a follow-up note in relation to recent concerns we would like to raise on the estimation of the management quantities. With some more time to consider the model fits and diagnostics and discussions with other scientists, plus running some additional model explorations very late in the assessment process, we feel that the assessment would benefit from further work to improve the confidence in the management quantities. We also identified concerns around some data inputs that need further understanding. SC20 should consider the issues raised when evaluating whether the assessment results provide the best available science for management advice.

Contents

1   Executive summary

This paper describes the 2024 stock assessment of striped marlin (*Kajikia audax*) in the southwestern Pacific Ocean (SWPO) within the convention area of the Western and Central Pacific Fisheries Commission (WCPFC). An additional five years of data were available since the previous assessment was conducted in 2019, with the current assessment extending the model period through to 2022. The assessment maintained a single region areas-as-fleets approach and fisheries structure consistent with the 2019 assessment, except for some difference in the grouping of flags for the CPUE index fishery.

The assessment differed from the 2019 assessment by moving from a catch-errors to a catch-conditioned framework, with catch used as input data with no uncertainty (an approach used in the most recent WCPFC skipjack, yellowfin and bigeye assessments and recommended by the recent independent peer review of the WCPFC yellowfin tuna assessment). This assessment also moved from a traditional fixed factor orthogonal uncertainty grid for management advice to a Monte Carlo approach where uncertainty was characterised by randomly drawing parameters from prior distributions for average natural mortality (M) and steepness (h), and the inclusion of estimation uncertainty for the key management quantities.

New developments to the stock assessment include:

- Conversion from a catch-errors to a catch-conditioned modelling framework, and the inclusion of a likelihood component for the CPUE from the index fishery.

- Change from the use of VAST to sdmTMB to standardise the input CPUE series, the inclusion of additional covariates in the CPUE model, increased mesh resolution and coverage and grouping of Japan and Chinese Taipei fleets to provide sufficient spatial and temporal coverage for the major index fishery.

- Update input data through to the end of 2022.

- The size composition data was reweighted by catch for extraction fisheries and CPUE for index fisheries, as is standard practice for WCPFC tuna assessments, to address any biases due to possible unrepresentative fishery dependent sampling.

- Change from 6 cm to 5 cm length bins and filter and smooth length composition data before conversion to eye orbital fork length (EFL) used for length data in the assessment.

- Natural mortality at age described using the Lorenzen functional form for natural mortality at age, with average M based on the $A_{\max}$ method with a maximum age of 15 years for the diagnostic model.

- Update maturity ogive based on reanalysis of histology samples and recommendations by Farley et al. (2021a).

- Update the parameters for the length-weight relationship.

- Update the fixed growth parameters based on reanalysis of otolith age-at-length data Farley et al. (2021a) and recommendations made at the Pre-assessment Workshop.

- Implement the Francis weighting method to iteratively reweight size composition data allowing appropriate data weights to be applied to size composition data for each fishery, rather than choosing arbitrary weights and applying these uniformly across all fisheries.

- The model start date was changed from 1952 to 1979 when the model was updated to the catch-conditioned version and using new growth parameters, which featured much faster growth at young ages. With these two modifications to the model, several implausibly high recruitment events were estimated by the model for early years, when the data featured some high annual catch records with limited size data to help inform recruitment at the time. Changing the start date to 1979 avoided these problems.

- Starting the model in 1979, rather than 1952, also allows vessel ID to be included as a covariate in the CPUE standardisation, and avoids using lower confidence operational longline data for striped marlin prior to 1979.

- The diagnostic model was simplified from a model with 2648 parameters estimated in the 2019 diagnostic model to 76 parameters in the 2024 diagnostic model.

The assessment is supported by the analysis of catch and effort data to develop CPUE abundance indices, and size composition data (Castillo-Jordán et al., 2024; Potts et al., 2024). Other papers of relevance to the assessment are the analysis of NZ sport fishery size trends and CPUE (Holdsworth, 2024) and the analysis of Australian longline CPUE (Tremblay-Boyer and Williams, 2023).

In addition to the diagnostic model, a number of other models were investigated as sensitivities to assess the relative impacts of alternative data and model assumptions on the estimated assessment results and conclusions. These sensitivities included the use of additional CPUE indices (Australian longline and New Zealand sport fishery), alternative values for the initial fishing mortality assumed at the start of the model in 1979 and the variance on recruitment deviations. None of these sensitivities had important implications for stock status advice.

The uncertainty ensemble was based on two key uncertainties: average M for the Lorenzen M-at-age and steepness (h). In contrast to the 2019 assessment, size composition data weighting was not included in the uncertainty characterisation due to the use of Francis weighting in 2024. The use of alternative CPUE series, growth and recruitment variation were not included in the uncertainty characterisation, as the diagnostic model was not very sensitive to these changes, and/or these models displayed poor diagnostics, or there were single options clearly recommended, such as growth. While growth information continues to be limited for striped marlin in the southwestern Pacific Ocean, based on the advice from CSIRO analysts we have used the best available growth formulation given the limited otolith age-at-length samples available. We strongly encourage col-

lection of additional age-at-length data for striped marlin across the southwestern Pacific Ocean to further improve these growth estimates or to allow age-at-length data to be used as conditional age-at-length data to estimate growth internally within the assessment, as recommended by the recent WCPFC yellowfin assessment peer review.

A key improvement to the uncertainty characterisation was the implementation of an ensemble approach to incorporate uncertainty in natural mortality and steepness of the stock recruitment relationship. The method involved constructing prior distributions for average M and steepness and drawing independent replicates from these priors. One hundred pairs of average M and steepness were chosen to construct the uncertainty ensemble of 100 models. These models were then filtered based on convergence criteria, with any resulting models that failed to produce a positive definite Hessian rejected. This resulted in an ensemble of 59 models, each with positive definite Hessian. Rejected models had combinations of average M of generally less than about 0.35 and h greater than 0.75. For each of these 59 ensemble models, which comprise the model uncertainty, estimation uncertainty was also incorporated to produce the overall uncertainty for the key management quantities.

Based on the ensemble of models, the general conclusions of this assessment are as follows:

- Consistent with the findings of the previous SWPO striped marlin assessments (Langley et al., 2006; Davies et al., 2012; Ducharme-Barth et al., 2019), the spawning biomass and stock status, as indicated by the spawning biomass depletion, show a steady decline from 1979 onwards, although the stock status has stabilised somewhat in the last 10 years, albeit at a low level.

- The stock status shows an increase in the terminal year of the 2024 assessment. Although terminal year estimates are often uncertain and subject to potential bias and subsequent revisions with future assessments, as is shown consistently in the retrospective analyses, a very recent increase in abundance is at least consistent with several CPUE abundance indices.

- The declining trend in recruitment over time, identified in the previous two stock assessments, remains a feature of the current model, along with reduced recruitment variability over time, all while the spawning stock biomass has decreased over time.

- Overall, the median depletion from the uncertainty grid for the recent period (2019-2022; $SB_{\mathrm{recent}}/SB_{F=0}$) is estimated at 0.14 (80 percentile range including estimation and structural uncertainty 0.09–0.21, full range 0.05–0.26).

- Fishing mortality has continued to increase over time for both the juvenile and adult components. With respect to maximum sustainable yield (MSY) based reference points (the WCPFC default for all billfish stocks, in lieu of the absence of any WCPFC billfish specific limit reference points), the median recent fishing mortality from the uncertainty ensemble is estimated to be above the level for achieving MSY (median $F_{\mathrm{recent}}/F_{\mathrm{MSY}} = 1.27$, range 0.58–3.02). 71%

of the 59 ensemble models estimated $F_{\mathrm{recent}}/F_{\mathrm{MSY}} > 1$.

- The median recent spawning biomass from the uncertainty ensemble was well below the spawning biomass to achieve MSY (median $SB_{\mathrm{recent}}/SB_{\mathrm{MSY}} = 0.60$, range 0.18–1.24). 93% of the 59 ensemble models estimated $SB_{\mathrm{recent}}/SB_{\mathrm{MSY}} < 1$.

- Despite decreased catches since the early 2000s, based on the default MSY-based reference points, the 2024 assessment of striped marlin in the southwestern Pacific Ocean concludes that the stock is likely overfished, and undergoing overfishing.

- The outcomes of the 2024 stock assessment are broadly consistent with the previous 2019 stock assessment, and there are no signs of any sustained recent rebuilding of the striped marlin stock in the southwestern Pacific Ocean.

Concerns with the assessment in relation to management quantities:

- We have recognised some critical model issues late in the assessment process that would benefit from further work to improve the confidence in the management advice.

- These concerns include poor fits to some of the size data and the under fit to CPUE in the recent time period, data conflicts and data weighting impacts, and the results of the ASPM and Catch Curve models.

- Preliminary work to understand the implications of some of these issues suggest the current model estimates of stock status are likely biased towards more pessimistic management advice, the degree to which we cannot be certain without additional work.

- We also very recently identified some concerns around some data inputs that need further understanding.

- SC20 should consider the issues raised when evaluating whether the assessment results provide the best available science for management advice.

2 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).

Specifically in the southwestern Pacific Ocean (SWPO), there is a long history of striped marlin catches by longline fisheries (Williams, 2003) and some recreational fisheries (Kopf, 2005). However, both sectors have shown declines in total catches and long-term declines in fish size (Kopf et al., 2005; Ward and Myers, 2005; Holdsworth, 2024). In addition, longline vessels in some areas have opportunistically targeted striped marlin (and other billfishes) in the WCPO (i.e. Australia, (Bromhead et al., 2004; Tremblay-Boyer and Williams, 2023)).

Compared to the tropical tuna species, which are regularly 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 subsequent assessment (Suzuki, 1989), which considered northwest Pacific and SWPO populations as separate, concluded that both stocks were at healthy levels though the SWPO stock appeared to be almost fully exploited (fishing mortality close to F_{MSY}). The most recent assessment within the WCPO indicated that the SWPO stock is likely overfished and close to undergoing overfishing (Ducharme-Barth et al., 2019) while the northern stock is overfished and undergoing overfishing relative to maximum sustainable yield (MSY) based reference points, according to the last three stock assessments (ISC-BWG, 2015, 2019, 2023).

This report describes a new assessment for SWPO striped marlin which builds upon the previous stock assessment undertaken by Ducharme-Barth et al. (2019). As in previous assessments, the objectives of the 2024 striped marlin assessment are to estimate population level parameters which indicate the stock status and impacts of fishing, such as time series of recruitment, biomass, biomass depletion and fishing mortality. The methodology used for the assessment is based on the general approach of integrated modeling (Fournier and Archibald, 1982), which is carried out using the stock assessment framework MULTIFAN-CL[4] (MFCL version number 2.2.7.0; Fournier et al., 1998; Hampton and Fournier, 2001; Kleiber et al., 2019). MFCL implements a size-based, age- and spatially-structured population model. Model parameters are estimated by maximizing an objective function, consisting of both likelihood (data) and "prior"[5] information components (penalties).

---

[4]http://www.multifan-cl.org

[5]Note that any mention of a "prior" in this report does not refer to a prior in the Bayesian sense, though the effect on the parameter estimate is similar, but rather a penalty placed on the likelihood such that the estimated

Each new assessment of a WCPO stock typically involves updates to fishery catch, effort, and size composition data, implementation of new features in the MFCL modeling software, changes to preparatory data analysis, such as CPUE standardisations, and consideration of new information on biology, population structure and potentially other population parameters. These changes are an important part of efforts to continually improve the modeling procedures and more accurately estimate fishing impacts, biological and population processes and quantities used for management advice. Advice from the Scientific Committee (SC) on previous assessments, and the annual SPC pre-assessment workshop (PAW) (Hamer, 2024) guide this ongoing process. Furthermore, due to changes in assessment staff, new assessments often involve staff who did not participate in the previous versions and this may also influence differences in how assessments are conducted. Changes to aspects of an assessment can result in changes to the estimated status of the stock and fishing impacts, and resultant management advice. It is important to recognize that each new assessment represents a new estimation of the historical population dynamics and recent stock status, and each new assessment team strive to provide the best possible assessment with the time and data available.

The assessment uses an "uncertainty grid" of models as the basis for management advice. The uncertainty grid is a group of models that are run to explore the interactions among selected "axes" of uncertainty that relate to biological assumptions, data inputs and data treatment. The axes are generally selected from one-off sensitivity models of a diagnostic (or base case) model to indicate uncertainties that have notable effects on the estimates of key model parameters and management quantities. The variation in estimates of the key management quantities across the uncertainty grid represents the uncertainty in stock status and should be considered carefully by managers. This structural or "model" uncertainty is usually more important than the uncertainty in the estimation of parameters from individual models, referred to as "estimation uncertainty", however both are taken into account when documenting the uncertainty in the key management quantities provided by the assessment.

The assessment used a Monte Carlo ensemble approach, with prior distributions for key parameters, for estimating uncertainty as the basis for management advice, instead of the orthogonal "uncertainty grid" used in recent WCPFC assessments. The ensemble was structured to incorporate key sources of uncertainty identified by a series of sensitivity tests along with estimation uncertainty for the key management quantities.

Significant changes and improvements to the analysis used in this assessment include the following, which are discussed in more detail in relevant sections of this report.

- Conversion from a catch-errors to a catch-conditioned modelling framework, and the inclusion of a likelihood component for the CPUE from the index fishery.

---

parameter does not deviate too much from the specified "prior" value. The magnitude of the deviation from the "prior" is dependent on the information content of the data and the strength of the likelihood penalty applied.

- Change from the use of VAST to sdmTMB to standardise the input CPUE series, the inclusion of additional covariates in the CPUE model, increased mesh resolution and coverage and grouping of Japan and Chinese Taipei fleets to provide sufficient spatial and temporal coverage for the major index fishery.

- Update input data through to the end of 2022.

- The size composition data was reweighted by catch for extraction fisheries and CPUE for index fisheries, as is standard practice for WCPFC tuna assessments, to address any biases due to possible unrepresentative fishery dependent sampling.

- Change from 6 cm to 5 cm length bins and filter and smooth length composition data before conversion to eye orbital fork length (EFL) used for length data in the assessment.

- Natural mortality at age described using the Lorenzen functional form for natural mortality at age, with average M based on the $A_{\max}$ method with a maximum age of 15 years for the diagnostic model.

- Update maturity ogive based on reanalysis of histology samples and recommendations by Farley et al. (2021a).

- Update the parameters for the length-weight relationship.

- Update the fixed growth parameters based on reanalysis of otolith age-at-length data Farley et al. (2021a) and recommendations made at the Pre-assessment Workshop.

- Implement the Francis weighting method to iteratively reweight size composition data allowing appropriate data weights to be applied to size composition data for each fishery, rather than choosing arbitrary weights and applying these uniformly across all fisheries.

- The model start date was changed from 1952 to 1979 when the model was updated to the catch-conditioned version and using new growth parameters, which featured much faster growth at young ages. With these two modifications to the model, several implausibly high recruitment events were estimated by the model for early years, when the data featured some high annual catch records with limited size data to help inform recruitment at the time. Changing the start date to 1979 avoided these problems.

- Starting the model in 1979, rather than 1952, also allows vessel ID to be included as a covariate in the CPUE standardisation, and avoids using lower confidence operational longline data for striped marlin prior to 1979.

- The diagnostic model was simplified from a model with 2648 parameters estimated in the 2019 diagnostic model to 76 parameters in the 2024 diagnostic model.

3 Background

The background material in this report largely follows Ducharme-Barth et al. (2019) and Davies et al. (2012) since much of the biological information remains relevant and the underlying structure of the model is unchanged.

3.1 Biology

Striped marlin is a pelagic species with a distribution extending through equatorial to temperate waters, with the highest catches and catch rates occurring within sub-equatorial and sub-tropical areas, particularly in the Pacific Ocean (Nakamura, 1985). Most catches of striped marlin are reported from surface waters (less than 100 m deep) (Brill et al., 1993; Domeier et al., 2003). Archival tagging data suggest striped marlin spend most time in surface waters (less than 10 m deep), with most dives not exceeding 40 m. Occasional dives have been reported to depths of 40–100 m (Domeier et al., 2003). This surface oriented behaviour (Lam et al., 2015) makes striped marlin particularly vulnerable to surface fisheries (longline, recreational and purse seine method fisheries) from young ages.

Details of the biology and ecology of striped marlin are poorly known, mainly as a result of their relatively low abundances, low catch rates, highly mobile nature and low priority for research funding. Based on the observed distribution of larval striped marlin, spawning occurs between May and June in the northwest Pacific (10°N–30°N), June-–November in the central-eastern Pacific and between November and December in the SWPO (10°S–30°S) (Hanamoto, 1977; Nakamura, 1985). Histological samples taken from fish caught by longline fisheries in the SWPO suggest that spawning in the southern hemisphere occurs once per season in the last quarter of the calendar year. Spawning occurs in multiple concentrated aggregations (Kopf et al., 2012) within the broad longitudinal band described by Nakamura (1985). Length data from the Japanese distant water fleet (see Section 4.6) indicate that juvenile striped marlin are predominantly captured in the tropical regions of the SWPO between 0° and 20°S (Kopf and Davie, 2009) where they recruit to the longline fishery at approximately 80–100 cm in length (eye orbit–fork length, EFL) and around 1–2 years of age. Size and age of encountered striped marlin typically increase moving south away from the equator.

Striped marlin display very high initial growth rates, attaining up to 45% of their maximum size in the first year of life (Melo-Barrera et al., 2003). Recent studies in the SWPO have indicated faster initial growth with fish attaining 70-75% of their maximum size by the second year of life (Kopf et al., 2011; Farley et al., 2021a). Growth rates of striped marlin are lower following the onset of maturity (Melo-Barrera et al., 2003) with females attaining slightly larger maximum sizes in terms of length and weight (Kopf et al., 2011). Striped marlin were reported to mature at around 140 cm for males and 180 cm for females, (Kopf et al., 2012) and 27–40 kg corresponding to approximately 2 years of age (Skillman and Yong, 1976; Nakamura, 1985), but more recent work (Farley et al.,

2021a). Striped marlin live for at least 15 years (Farley et al., 2021a) and can exceed more than 260 cm EFL and 240 kg. We use 15 years as the maximum age, $A_{\max}$.

Large striped marlin tend to move further into temperate regions on a seasonal basis, especially in the SWPO. Relatively large fish are captured by the recreational fisheries in northern New Zealand (Kopf, 2005), and by recreational and commercial fisheries off south eastern Australia. While several large movements have been reported from fish tagged and released from northern New Zealand (e.g. several recaptures from waters of French Polynesia (Domeier, 2006; Holdsworth et al., 2009; Sippel et al., 2011)), clear migration pathways have not been established. Few large-scale movements have been recorded for marlin tagged off eastern Australia, creating some uncertainty about the extent of mixing of fish within the region. Current tag-recapture data suggest some level of broader sub-regional fidelity (Bromhead et al., 2004). In the southwestern Pacific Ocean, it is speculated that post-spawning striped marlin move south-eastwards from the Coral Sea into waters around northern New Zealand and south-eastern Australia to feed and recover, before returning to spawning grounds the following spawning season (Kopf, 2005; Kopf and Davie, 2009; Kopf et al., 2012).

Estimates of mortality rates of striped marlin are rare and have generally been generated from modes identified in length-frequency samples. Estimated natural mortality rates (M) are estimated to vary between sexes, being lower in males (0.57–0.79 year^{-1}) than females (0.82–1.33 year^{-1}) (Boggs, 1989; Pauly, 1980; Hinton and Bayliff, 2002). However, the natural mortality rates of striped marlin unable to be sexed or of unreported sex have been estimated to be lower (0.39–0.49 year^{-1}) (Boggs, 1989; Pauly, 1980; Hinton and Bayliff, 2002). Unsexed fish may be dominated by small, juvenile fish and thus the associated total mortality rates are likely to be lower as few striped marlin below 100 cm EFL are captured by longline fisheries. A meta-analysis of natural mortality estimates for striped marlin proposed a weighted mean value of 0.38 year^{-1} (Piner and Lee, 2011). The Center for the Advancement of Population Assessment Methodology (CAPAM) at the 2023 "Tuna Stock Assessment Good Practices Workshop" recommended applying an age-specific pattern in M by using the inverse mean length-at-age method developed by Lorenzen (1996).

A range of stock structures have been proposed for striped marlin in the Pacific Ocean (Graves and McDowell, 2003), including a horseshoe-shaped Pacific-wide stock (Nakamura, 1985) and a two-stock model (northern and southern Pacific) (Hinton and Bayliff, 2002). However, a number of studies have strengthened arguments for the occurrence of a semi-independent stock in the southwestern Pacific Ocean. Conventional tag recapture data indicated no trans-basin movements by striped marlin tagged in the southwestern Pacific, though a single individual tagged to the northeast of Hawaii with a pop-up satellite tag (PSAT) travelled to the east coast of Australia before the tag released on schedule (Lam et al., 2019). A summary of striped marlin tagging using pop-off satellite archival tags within a state-space model indicated complex spatiotemporal patterns in striped marlin distribution, with strong seasonal patterns by latitude (Chambers et al., 2013). Results of genetic analyses suggest the potential for a significant degree of stock structuring within the Pacific Ocean and support the assumption of a semi-independent stock in the SWPO (Graves

and McDowell, 1994). Recent work using genome-wide single nucleotide polymorphism (SNP) data identified a genetically distinct stock in the SWPO with a high degree of genetic connectivity among individuals sampled off of Australia and New Zealand (Mamoozadeh et al., 2018). Additionally, a few individuals sampled off of western Australia (eastern Indian Ocean), Ecuador (eastern central Pacific) and Hawaii (central north Pacific) were genetically similar to SWPO individuals indicating some level of spatial connectivity between these regions (Mamoozadeh et al., 2018).

Examination of the Pacific-wide spatial trends in Japanese longline CPUE over the past 50 years also supports an argument for stock structuring (Nakamura, 1985). Very low catch rates of striped marlin have been reported by longline fleets in equatorial regions of the Pacific Ocean (10°N–10°S), Nakamura (1985)) despite considerable longline effort. In contrast, high catch rates have been reported adjacent to the Baja coast of the eastern Pacific Ocean (EPO) (Nakamura, 1985; Hinton and Bayliff, 2002). In summary, current information suggests the potential for at least northern and southern stocks of striped marlin in the WCPO, with the potential for an additional stock in the eastern Pacific.

3.2 Fisheries

Striped marlin are captured mainly by longline fisheries, and to a lesser extent by recreational fisheries throughout their range in the SWPO (Figure 1). Relatively high catches of striped marlin were estimated during the 1950s and early 1960s, with a peak of more than 70,000 individuals ($\sim 6,000$ mt) estimated to be caught in 1954 (Ducharme-Barth et al., 2019). Since the mid 1960s, catches from the SWPO have varied between approximately 20,000 and 40,000 individuals with a decreasing trend since the late 1990s (Figure 2). Some of this decline is attributable to lower catches in the Japanese longline fishery, especially since 1999 (Figure 4), but also attributable to a decline in catches from the Australian longline fleet in sub-regions 2 and 3 since 2001 (Figure 5). Around 1999, there was a switch from Japanese longline catches generally taking the largest share of the catch to the catch being dominated by two flags: longline fisheries from Australia (in sub-regions 2 and 3) and longline fisheries from Chinese Taipei (in all subregions) (Figure 5).

Catches of striped marlin were dominated by the Japanese longline fleet until the early 1990s (Figures 4 and 5). Chinese Taipei and Korean fleets reported moderate to large catches of striped marlin since the mid-1960s and mid-1970s, respectively in the northern and eastern areas (Figures 4 and 5). Chinese Taipei catches have increased in recent years, mainly due to the effort of this fleet in the eastern temperate WCPO, targeting predominantly albacore tuna. Reductions in Japanese and Australian catches in the last 15 years have been partly offset by recent increases in catches in the Chinese longline fishery (Figures 6 and 7). Longline fleets of Pacific Island Countries and Territories (PICTs), and Australia have reported increasing catches since the early 1990s mainly due to the development of their domestic fleets (Figure 6). Catches by Australian longline fleets rapidly increased until 2003, partly due to specific targeting of striped marlin during some periods of the year. These catches have declined and stabilised since that time. Since 1987, longline fleets

operating in the New Zealand exclusive economic zone (EEZ) have been prohibited from landing striped marlin in an attempt to support recreational fisheries in the north of the country (Kopf et al., 2005).

Extensive recreational fisheries operate throughout the southwestern Pacific Ocean (Whitelaw, 2001; Bromhead et al., 2004; Kopf et al., 2005) although total catches by recreational fisheries are very small relative to commercial catches. In addition, a high proportion of striped marlin are released (after tagging) by recreational fisheries in the WCPO (up to 60%, Holdsworth et al., 2019). However, studies into the survival of marlin captured recreationally have estimated that up to 50% of marlin may suffer post-release mortality due to hook damage, stress or increased susceptibility to predation (Pepperell and Davis, 1999). In the limited number of studies, sample sizes were typically small and the duration of monitoring of post-released fish was relatively short (e.g. maximum of 93 days for striped marlin, Domeier et al. (2003)). Brodziak et al. (2012) calculated a mean post-release mortality rate of 25.4% for striped marlin from a meta-analysis of PSAT data.

3.3 Tagging data

A comprehensive summary of available striped marlin tagging data was provided by Langley et al. (2006). Though additional years of tagging data are available beyond the range of data considered in the 2006 stock assessment, the issues identified by Langley et al. (2006) persist in more recent data: low recovery rates, short times at liberty, and small movement distances (< 500 km). Despite the large number of conventional tags released by recreational fishers ($> 50{,}000$ combined releases in New Zealand and eastern Australia through 2018) less than 1% have been recaptured. This is likely due to the extremely high rates of tag shedding shown for striped marlin (Domeier et al., 2019). Of the recaptured individuals, between 80-90% are recaptured in less than two quarters. Such recaptures would be excluded from the model under the typical assumptions made in WCPO tuna assessments (Castillo Jordán et al., 2022; Day et al., 2023; Magnusson et al., 2023) for the time required for these tagged individuals to be considered well-mixed with the untagged individuals.

Although MULTIFAN-CL has the capacity to include tagging data in the assessment model, the raw tagging data from the striped marlin fishery is not included in the current assessment model, given the issues listed above. However, the tag data are useful to help inform the appropriate regional structure for the model. Individuals tagged off the east coast of Australia appear to show higher residence times and lower levels of dispersion relative to fish tagged off the North Island of New Zealand. However, there is not sufficient confidence in the available tagging data to use it to estimate sub-regional movement rates in a spatially explicit model. Additionally, available satellite tagging data summarised in a state-space model (Chambers et al., 2013) illustrated temporal and spatial complexity (including evidence of seasonal movement) in striped marlin distribution in the southwestern Pacific Ocean. However, that analysis failed to show a clear basis for defining discrete boundaries within the assessment region, and did not change the general view gained from conventional tagging data. While striped marlin are able to make rapid and/or long-distance

movements after tagging, almost all recoveries of fish tagged within the region were reported within the model region. This indicates a relatively high level of fidelity or probability of occurring within the assessment region (Langley et al., 2006), although it is worth noting that times-at-liberty were short. Further, most tagged fish recaptured in the second or third quarter following release had moved a considerable distance (500–1500 km) indicating strong seasonal movements (mainly south-north) and, thereby, suggestive of a relatively high level of regional-scale mixing of fish in the population. This observation, along with the recent genetic work described in Section 3.1 provides support for the adoption of a single model region (see Section 4).

3.4 Key changes from the last assessment

3.4.1 Catch conditioned approach

In previous MULTIFAN-CL (MFCL) assessments of striped marlin, catch was predicted by the model (termed a "catch-errors" model) with observation error allowed, and the standard deviation of the log-catch deviates assumed to be very small (equivalent to a CV of 0.002). This produced very accurate predictions of observed catches and therefore only a small contribution of the catch to the overall objective function. However, the cost of treating the catch in this way was that effort deviation coefficients had to be estimated as model parameters for each catch observation. Additionally, catchability deviation parameters were required for catch-effort observations for fisheries for which time-series changes in catchability were allowed. While these parameters were constrained by prior distributions and estimation was feasible, it resulted in very large numbers of parameters needing to be estimated by the function minimiser and many of these were effort deviation coefficients and parameters relating to catchability.

In an effort to reduce complexity and parameterisation, this assessment makes use of a relatively new feature of MFCL first applied to the 2022 skipjack tuna assessment in which catch is assumed to have no error, i.e., the model is "catch-conditioned" (Davies et al., 2022). This makes it possible to solve the catch equation for fishing mortality exactly, using a Newton-Raphson sub-iterative procedure. The main benefit of this approach is that effort deviation coefficients and catchability-related parameters do not require estimation as model parameters. Effort data for extraction fisheries is not required at all but can be used if available to estimate catchability through regressions of fishing mortality and effort, and this is important for making stock projections based on future effort scenarios. The reduction in parameters has enabled more rapid model convergence and Hessian matrix computation. The only cost of this approach is that missing catches, which could be accommodated in the catch-errors version if there was an accompanying effort observation, are no longer straight forward to account for. However, this is not an impediment for the key WCPO tuna assessments. The catch conditioned approach allows (but does not require) the specification of index fisheries to provide indices of relative abundance, these are discussed in section Section 4.4. In the stepwise model development runs conducted for this assessment, the transition from a "catch-errors" to a "catch-conditioned" model, without implementation of the survey fisheries, did not result in any

appreciable change in the estimated quantities of relevance to management advice.

3.4.2 Size composition treatment

Size composition filtering was modified to try to address regular spikes appearing at 5cm intervals in the raw length composition data, due to some measurements being recorded at 5cm accuracy. These records were smoothed, and then converted to EFL, and samples with less than 10 records per spatial and temporal unit were also filtered out. Catch reweighting for extraction fisheries and CPUE reweighting for index fisheries was then applied to size composition data, to try to minimise any effects from unrepresentative sampling (Castillo-Jordán et al., 2024). The final weighting of size composition data was to apply iterative reweighting using the Francis method (Francis, 2011), which enables the data weighting to be applied separately for each fishery, taking account of the information contained in the size composition data for each fishery

4 Data compilation

4.1 General notes

Data used in the striped marlin stock assessment using MFCL consist of catch, effort, length and weight composition data for the fisheries defined in this analysis, updated through to the end of 2022. Improvements in these data inputs are ongoing and readers should refer to the companion papers highlighted at the end of Section 2 for detailed descriptions of how the data and biological inputs were formulated as only brief overviews are provided below. A summary of the data available for the assessment is provided in (Castillo-Jordán et al., 2024), and in Figure 9.

4.2 Spatial stratification

The stock assessment of striped marlin in the SWPO 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, and also since 1979. The assessment region excluded areas to the north and east due to little evidence for mixing between these regions (Bromhead et al., 2004). 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 considered to encompass a semi-independent stock of striped marlin and, given the spatial distribution of the catch, represents an appropriate spatial scale for assessment and management of the striped marlin resource in the SWPO.

This assessment models a single population of striped marlin within the region, assuming individuals are well-mixed within the region. However, four sub-regions (Figure 1) were 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 (Figures 20 and 21) and the qualitative assessment of

available tagging and recapture data (Langley et al., 2006; Davies et al., 2012; Ducharme-Barth et al., 2019).

The catch from 1979-2022, recorded in numbers of fish per sub-region per year is shown in Figure 2, indicating a reasonable spread of catch between these four sub-regions. Regression tree analysis of size composition data (Potts et al., 2024) suggest there is insufficient evidence to suggest an alternative spatial structure be adopted for these fisheries. These sub-regions were used to define the spatial boundaries for the grouping of fisheries operating within the assessment region. However, it was noted the data coverage in space in time was probably too sparse to provide strong conclusions on meaningful spatial structure of size compositions. Separate selectivity functions were estimated for almost all the fisheries in each sub-region, with seasonal catchabilities being estimated for all fisheries. As such, these parameters may offer, to some extent, a description of age specific, temporal and spatial patterns in the availability of striped marlin to the various fisheries distributed within the model region.

Fisheries operating within the equatorial sub-region of the model (sub-region 1) generally target bigeye and yellowfin tunas, with striped marlin being a commercially important bycatch species for most fisheries. Catches starting in 1952 are shown by flag initially (Figure 3) to indicate the size variability of some of the early catch data prior to 1979, but subsequent catch plots are shown from 1979 onwards, given this was subsequently chosen as the start date for the diagnostic model. Japanese vessels were the dominant fleet up until the year 2000 (Figures 3 and 4), especially within sub-regions 2 and 3 (Figures 5 and 8). However, vessels from other distant water fishing nations (DWFNs) also operate within the region and domestic fisheries of PICTs developed during the 1990s (Figures 6 and 7).

Fisheries in sub-region 1 historically reported consistent catches of striped marlin, by number of fish caught, all with annual catch records greater than 5,000 individuals, but with only one year (2014) with over 15,000 fish recorded. Catches generally declining following this peak in 2014 to just over 5,000 fish caught in 2022 (Figure 7). Catches from sub-region 1 are typically characterized by smaller individuals than caught in other sub-regions. Fisheries in sub-region 2, recorded variable catches early in the time series. Prior to the year 2000, greater than 15,000 individuals were caught in four separate years, with less than 5,000 individuals caught on two separate years. Since 2000, catches in sub-region 3 have generally declined, with all catches below 10,000 individuals, and declining further, to below 5,000 individuals, since 2010 (Figure 7). From 1979 to 2000, catches in sub-region 3 have generally increased from around 2,000 individuals up to around 10,000 individuals in the year 2000, followed by a subsequent general decline in catches, with no catches over 5,000 in sub-region 3 since 2007, and with catches at around 3,000 individuals recorded annually since 2016 (Figure 7). With the exception of a few specific years with higher catches (8,000-13,000), catches in sub-region 4 generally started between 2,000 and 4,000 fish from 1979 up until 1991, followed by an increase in catch to around 5,000 individuals since the early 1990s (Figure 7). In 2019 and 2020, catches in sub-region 4 exceeded the catches in all other sub-regions, with sub-region 1 recording

the highest catches in all other recent years, i.e. in the period from 2003 to 2022.

In terms of biomass rather than numbers of fish removed, fisheries in sub-region 2 generally removed the greatest biomass, followed by sub-region 1, with much smaller quantities removed by fisheries in sub-regions 3 and 4 (Figure 5). Sub-region 1 has seen a decline in biomass removed since 2010. By contrast the biomass removed from sub-region 2 declined over the period 2000-2010, and has stabilised since 2010 with greater biomass per year removed than sub-region 1. Sub-region 3 has seen a slight decline in biomass removed in the last 10 years, while sub-region 4 has seen an increase in biomass removed in the same time period (Figure 5).

From 2003-2018 fisheries in sub-region 1 accounted for the highest catch by numbers of fish (Figure 7), although not by biomass (Figure 5).

The highest overall catches of striped marlin by biomass have been reported from sub-region 2 of the model region (Figures 5 and 7). Longline fisheries in sub-region 2 target bigeye and yellowfin tunas, or albacore tuna, with striped marlin being a commercially important bycatch species. Historically, Japanese vessels were the dominant fleet in this sub-region. However, Japanese catch has declined in this sub-region (Figures 5 and 8) since the early 1990s. The Australian longline fleet and longline fisheries of other PICTs within the sub-region have expanded since the mid-1990s. Some vessels within the Australian longline fleet have opportunistically targeted striped marlin in recent years within this sub-region and in sub-region 3. Total catches in this sub-region have generally declined since the late 1990s. Catches of striped marlin from sub-region 2 are dominated by relatively large fish.

Catches of striped marlin from sub-region 3 are also dominated by large fish. The Japanese fleet was the dominant fleet in this sub-region until the 1990s. The Australian domestic fleet accounts for most of the recent catches of striped marlin from sub-region 3. With the exception of the Australian fleet, longline fisheries in this region do not specifically target striped marlin. Significant recreational fishing effort also occurs in this sub-region (Figure 5), focused in northern New Zealand and the central-eastern coast of Australia, with striped marlin being a major target species (Bromhead et al., 2004; Holdsworth et al., 2019). Catches (in numbers of fish) from sub-region 3 are currently of a similar magnitude to those from sub-region 2 (Figure 5), although considerably lower in biomass in sub-region 3 compared to subregion 2 over the same period (Figure 7).

Catches of striped marlin in biomass are relatively low from sub-region 4 (Figure 5) but of greater significance in terms of numbers (Figure 7), especially since 2019. They are a commercially retained bycatch species for longline fisheries in this sub-region, which mainly target albacore. Only the distant water fleet of Chinese Taipei has reported high levels of effort from this sub-region. Longline fisheries of PICTs also operate in sub-region 4, again targeting albacore tuna. Recreational fisheries also exist within this sub-region but the total catch from these fisheries is relatively insignificant. Limited size data were available for this sub-region of the model.

4.3 Temporal stratification

Data used in the previous stock assessment covered the period 1952–2017. This assessment uses five additional years of new data, so covers the period 1979-2022, excluding the data from 1952-1978. By commencing the model in 1979, this allowed the use of Vessel ID as a factor in the CPUE standardisation and avoided the problems with the implausible model estimates of recruitments in the early years with some high catch records (Figure 3), discussed in more detail in (Section 6.1.1). Catches of striped marlin display strong seasonal variations (Bromhead et al., 2004; Kopf et al., 2005) particularly at the more southern latitudes of their range. Further, some fisheries show strong seasonal variations in effort (e.g. Chinese Taipei distant water fleet, Australian and New Zealand recreational fisheries). In contrast to the CPUE used in the 2019 assessment which used a quarterly time step in the model, we chose an annual time step to help smooth out seasonal fluctuations in the CPUE and the variation in data quality and availability by season. Instead of having a seasonal time step, we used an annual time step and we included season as a covariate in the CPUE standardisation.

4.4 Definition of fisheries

The sub-regions of the model region were used to define the spatial boundaries of the specific fisheries in the SWPO. A total of 12 longline fisheries and two recreational fisheries were defined (Table 1), based on sub-region boundaries, fishing method and flag. A more detailed breakdown of the catch and composition data available for each fishery can be found in Castillo-Jordán et al. (2024).

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). 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 for Australia (Fisheries 6-7) and New Zealand (Fishery 8) developed in sub-regions 2 and 3 off eastern Australia and extended south into the Tasman Sea. In order to account for the remaining longline fishing effort and 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 effort, catch, and composition data of the other DWFNs and PICTs operating in the assessment region.

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 significant 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 capturing striped marlin operate in the assessment region (Whitelaw, 2001), but these fisheries are relatively small, and catch and effort data are not readily available.

Index fisheries: The catch-conditioned approach (Section 3.4.1) allows the specification of "index fisheries" that are used to provide standardised CPUE indices of abundance. Index fisheries are akin to "survey fisheries" as described for other software such as Stock Synthesis, and may be the same fisheries as the extraction fisheries, but when used as index fisheries they do not take any catch, and must have effort data to allow modelling of CPUE. For this assessment one index fishery is defined for the whole model region as a composite fishery composed of longline fisheries from Japan and Chinese Taipei (Castillo-Jordán et al., 2024). 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 index fishery CPUE. The CPUE standardisation approach for the index fisheries used the sdmTMB package for implementing spatio-temporal delta-generalised linear mixed models and is described in detail in (Castillo-Jordán et al., 2024).

4.5 Catch data

4.5.1 General characteristics

Catch data were compiled according to the fisheries defined in Table 1. Catch data are expressed in numbers of fish (Table 1) 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 $5° \times 5°$ cells and annual resolution, with the aggregation process either conducted by SPC, where operational data is available to inform this, or by the particular countries following statistical procedures that are reported to the WCPFC.

Total annual catches by region for the SWPO are shown in Figure 2, by flag in Figure 4, by both flag and region in Figure 5, by major gear category in Figure 6 and by both major gear category and region in Figure 7. The spatial distribution of catches by major gear type averaged over the first ten years and over the most recent ten years is provided in Figure 8. Discarded catches are estimated to be minor for all longline fisheries and were not included in the analysis.

Data were supplied at a variety of spatial and temporal resolutions. For example, longline catch and effort data were generally available aggregated by month and $5° \times 5°$ spatial resolution early in the model period, and more recently operational-level logsheet data were available for many of the longline fisheries. Recreational data were supplied for individual sub-regions of the model.

Japanese distant-water longline fisheries (Fisheries 1, 2, 3 and 4; sub-regions 1–4): Catch and effort data from the Japanese fleet were supplied by the National Research Institute for Far Seas Fisheries (NRIFSF) stratified by spatial cell ($5° \times 5°$), month, and gear configuration (number of hooks between floats, HBF). The spatial extent of the Japanese longline fleet has declined over the

last 30 years. For example, since 1992, limited longline effort was reported by the Japanese fleet in sub-region 4 of the region (Fishery 4).

Chinese Taipei distant-water longline fishery (Fishery 5; sub-region 4): Catch and effort data for this fleet were available aggregated by $5° × 5°$ cell and month. Data were supplied by the Fisheries Agency - Chinese Taipei.

Australian longline fisheries (Fisheries 6 and 7; sub-regions 2 and 3): Longline catch and effort data were provided on a quarterly basis for each sub-region by the Australian Fisheries Management Authority (AFMA). Data were raised to provide estimates of total catches applying the scaling factors used by Campbell et al. (2002).

CPUE indices were provided by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) using General Additive Models (GAMs) and a metier approach described in Tremblay-Boyer and Williams (2023). This approach was applied to operational-level data to derive a standardised effort series for the Australian longline fisheries in sub-regions 2 and 3.

New Zealand longline fishery (Fishery 8; sub-region 3): Longline data, in both numbers and tons of striped marlin, were provided by the Ministry of Fisheries (MFish), New Zealand. However, the landing of striped marlin by commercial longline vessels fishing within the New Zealand EEZ has been prohibited since 1987 (Kopf et al., 2005). While records of retained and discarded striped marlin do occur in the logsheet data for this fishery, it is likely total catches for this fishery are under-estimated because operational data provided does not represent 100% of the vessels.

Other longline fisheries (Fisheries 11, 12, 13 and 14; sub-regions 1–4): Other longline fleets have also operated within the model region since 1952, aside from the fisheries identified above. These "other" mixed-longline fisheries were pooled into the relevant model sub-regions on a quarterly basis. Fishery 11 (sub-region 1) included fleets from PICTs (e.g. Papua New Guinea, Solomon Islands, Vanuatu, Cook Islands and French Polynesia), plus fleets from distant water fishing nations other than Japan. Fishery 12 (sub-region 2) included catch and effort data from PICTs (e.g. New Caledonia, Vanuatu, Fiji and Tonga) plus fleets from distant water fishing nations other than Japan and Australia. Fishery 13 (sub-region 3) included fleets from all flags other than Australia, Japan and New Zealand. Fishery 14 (sub-region 4) included fleets from some PICTs (e.g. Cook Islands, French Polynesia) and all fleets from other DWFNs other than Japan and Chinese Taipei.

Australian recreational fishery (Fishery 9, sub-region 3): A summary of the recreational fishery catch data (1952-2011) was available from Ghosn et al. (2012). The data were collected during game fish tournaments from operators fishing off south-eastern Australia. Unfortunately, many of the major recreational fishing clubs changed their record keeping format in 2012 so catch records beyond the 2012 assessment were unavailable. This catch data was complemented by more recent Australian recreational catch data to 2022 (J Pepperell pers. comm, 2024).

New Zealand recreational fishery (Fishery 10, sub-region 3): Information for this fishery was

supplied by Holdsworth et al. (2019); Holdsworth (2024) with catch data for the period 1952–2022, with this data obtained from extensive fishing club records.

4.5.2 CPUE

Given the dynamic and patchy effort of longline fleets in the South Pacific over time, there are advantages to developing index fisheries from a combined-fleets dataset. This makes maximum use of the fully integrated, multi-fleet standardised CPUE analyses by providing the best possible spatial and temporal coverage for the indices of relative abundance in the assessment, and avoids assigning the multi-fleet standardised CPUE time series to only one fleet component within the assessment. The multi-fleet regional abundance indices were calculated using the sdmTMB package for implementing spatial delta-generalized linear mixed models (Anderson et al., 2022) and is discussed further below and in Castillo-Jordán et al. (2024).

We denoted one index fishery for the SWPO, characterised by all the longline fisheries from Japan and Chinese Taipei in all four sub-regions, with catchability assumed constant across years and fisheries. In the 2019 assessment, only the Japanese longline fisheries were included in the standardisation of the SWPO longline index. However, in addition to changes in apparent targeting among the Japanese fleet over time, the relative influence of the fisheries from Chinese Taipei increased in over more recent years which led to the decision to include these additional fisheries in the analysis. Data from the longline index fisheries from Japan and Chinese Taipei spanned the period 1979-2022 and the fisheries from Chinese Tapei are important to provide good temporal coverage due to the reduction in catch of striped marlin by the Japanese fisheries in recent years. The longline operational data prior to 1979 is suggested to be of lower quality than more recent data by Japanese scientists.

In the previous assessment (Ducharme-Barth et al., 2019), an index of relative abundance based on Japanese longline data was developed using a spatio-temporal modelling approach implemented in the VAST R package (Thorson et al., 2015). The current assessment uses a similar spatio-temporal modelling approach implemented in the sdmTMB R package (Anderson et al., 2022) as was used in 2023 for the estimation of bigeye and yellowfin tuna (Teears et al., 2023). The sdmTMB geostatistical software was selected as it has been developed to be computationally efficient, flexible, and user-friendly with online community support (Anderson et al., 2022) and thus, represents a reasonable alternative for improving reproducibility and efficiency. The modelling process is described in detail in Castillo-Jordán et al. (2024).

The sdmTMB framework was used to implement a spatio-temporal delta generalized linear mixed model (GLMM), from which area-weighted abundance indices were generated after "standardising out" the influence of the catchability covariates. The modelling approach explicitly addresses the spatial structure in the response variable, that is, the fact that observations closer in space are more likely to be similar. This allows the spatial autocorrelation to be accounted for, which increases the precision in estimates and in some instances makes it easier to identify a relationship between the

response and candidate explanatory variables.

In the 2019 assessment, the CPUE standardization model used targeting cluster as a catchability covariate, which was derived using longline catch species composition (i.e., yellowfin, bigeye, albacore, and other species) and hooks between floats (HBF). This approach follows the assumption that changes in individual species' composition represent changes in targeting behaviour. However, species composition may be confounded with changes in individual species' abundance. Additionally, vessel flag was predicted to be an important covariate as there can be fleet effects and differences in fishing strategies employed. Therefore, the final model used to generate the CPUE indices included HBF in place of targeting cluster and vessel flag as catchability covariates.

In the SWPO, an annual index was developed for striped marlin that included HBF, flag, vessel ID and season (i.e., month) as a cyclic spline; all of which were treated as catchability covariates, with vessel ID an additional covariate to the list used in the 2019 assessment, and the nominal and standardised CPUE index is shown in Figure 10, with the standardised CPUE with 95% confidence shown in Figure 11.

4.6   Size composition data

Available length data for each of the defined fisheries were compiled into 5 cm size classes, rather than the 6 cm size classes used in previous striped marlin assessments. An investigation into the length composition data indicated frequent misreporting of bin size resolution in data submissions.

There were clear peaks in the raw data in length bins that were multiples of five, so, efforts were made to improve the quality of the data, including removing samples with less than 10 records, while preserving as much of the "uncontaminated" data as possible (see Castillo-Jordán et al. (2024) for further details). Data were collected from a number of sources and can be summarized as follows.

Size composition data, comprising either length-frequency (LF) and/or weight-frequency (WF) data, were available from many of the fisheries defined in the assessment (Table 2), although data were provided in a number of different formats depending on the specific fishery. For most fisheries, temporal coverage of the size frequency data was relatively limited (Castillo-Jordán et al., 2024).

Length data were provided based on four different length measurement methods: eye orbital–fork length (EFL), lower jaw–fork length (LJFL), bill tip-fork length (BFL) or pelvic fin–fork length (PFFL). A range of weights were supplied including whole weight, Japanese processed weights (gilled, gutted, head and tail left on, bill removed at a point level with the tip of the lower jaw), and gilled, gutted and headed (i.e. trunked) weights. All length measurements were standardized to EFL and weight measurements were standardized to the equivalent whole (unprocessed) weight using the conversion factors listed in Williams and Smith (2018).

Japanese longline fisheries (fisheries 1–4; sub-regions 1–4): Data were supplied by the NRIFSF and length and/or weight data were available from 1970–2005. Coverage varied between sub-regions,

and limited data exist after 1998. Length data were recorded as EFL while weight data were supplied as Japanese processed weights.

Chinese Taipei longline fisheries (fishery 5; sub-regions 4): Data were supplied by the Fisheries Agency - Chinese Taipei. Very limited length composition data was available for fishery 5. Length data were recorded as LJFL while weight data (used when Chinese Taipei was included in Fisheries 11 -13) were supplied as trunked weights.

Australian longline fisheries (fisheries 6 and 7; sub-regions 2 and 3): Data provided by AFMA represented the most extensive collection of size composition data used in this analysis (nearly 3,000 length measurements and over 75,000 weight measurements). These data are available from the Australian longline fisheries beginning in the mid-1990s. The weight data were originally sourced from the main fish processors receiving striped marlin from Australian longline vessels and represent a comprehensive sample of the entire catch. Weights were supplied as processed (trunked) weights to the nearest 0.1 kg.

New Zealand longline fishery (fishery 8; sub-region 3): Eleven quarters of length data were available from observers on board New Zealand longline vessels during 1997–2011. Data were supplied by MFish with lengths measured as LJFL.

Australian recreational fishery (fishery 9; sub-region 3): Weight frequency data were available for the Australian recreational fishery for the full model period. All weights were supplied as whole weights and were from landed fish (therefore assumed accurately measured). Only landed weights were incorporated into the analysis because the tagged component of the catch was not included as fishing mortality, and the weight of released individuals was not reliably measured.

New Zealand recreational fishery (fishery 10; sub-region 3): Weight data were supplied for the New Zealand recreational fishery for the entire model period (Holdsworth et al., 2019). All weights were recorded as whole weights and were from landed fish (therefore accurately measured). Only landed weights were incorporated into the analysis because the tagged component of the catch was not included as fishing mortality, and the weight of released individuals was not reliably measured.

Other longline fisheries (fisheries 11, 12 and 14; sub-regions 1, 2 and 4): Length data were available from 1996 for fishery 11; from 1993 for fishery 12, and from 1996 for fishery 14. Data from these fisheries were supplied from a combination of regional observer programs, regional port-sampling programs and/or from research institutes of distant water fleets. Very limited length and weight data were available for fishery 13.

Size data were aggregated by fishery and year. Length data were aggregated into 63 5-cm size classes (20–335 cm EFL), 5 cm size classes, rather than the 6 cm size classes used in the 2019 assessment, to help ensure that that data recorded at 5 cm accuracy was only assigned to give one 5 cm measurement entry per size class Castillo-Jordán et al. (2024)

Weight data were aggregated into 123 2-kg intervals (5–249 kg whole weight), as used in the

2019 assessment (246 1-kg intervals). Length or weight data were not available for all quarters for the period of data supplied for each fishery. The exception was fishery 10 (New Zealand recreational fishery; sub-region 3) for which there were landed weight data in the first quarter for all years. Fisheries for which comprehensive weight-frequency data were available included 02.JP.LL.2, 03.JP.LL.3, 06.AU.LL.2, 07.AU.LL.3 and 10.NZ.REC.3 fisheries; and those with comprehensive length-frequency data available included 11.LL.ALL.1 and 12.LL.ALL.2 (Table 2).

Overall, smaller fish were more commonly captured by longline fisheries in sub-region 1 (equatorial areas) with sequentially larger fish tending to be captured by longline fisheries in more temperate waters of sub-regions 2 and 3 (Figures 20 and 21). Recreational fisheries tended to capture larger striped marlin than the longline fisheries.

Instances of erroneous samples for small fish were identified during analyses of the size frequency data in previous assessments (Langley et al., 2006; Davies et al., 2012; Ducharme-Barth et al., 2019), and subsequently removed. These samples were also excluded from the current assessment. For fishery 1 in 1995, a single sample of atypical size composition and exceptionally large sample size (10- to 100-fold higher than in other years) was removed. Samples consisted of only small fish in the period 1995–2001 for fishery 14.LL.ALL.4, which was clearly very different relative to samples in subsequent years. These instances were suggestive of sampling bias. It is likely that these exceptional samples were not representative of the fishery catch composition because they were very different to other years, or they may reflect spatial heterogeneity in size compositions and are most likely not representative. Excluding these potential outliers is unlikely to bias model estimates because growth and natural mortality are assumed at fixed values, and fish in these size classes are only partially recruited to the fishery. Weight-frequency observations for fishery 09.REC.AU.3 collected prior to 1986 had fewer than 30 observations per year, and were collected from a period when the fishery was operating in the northern range of the fishery that is atypical of its core operations since 1986 when most of the catch was taken (Ghosn et al., 2012). These observations were also removed as they had proved detrimental to model fit in the previous assessments.

4.7 Reweighting size compositions

Statistical correction of size composition data is required as length and weight samples are often collected unevenly in space and time. The methods for re-weighting of the size composition data are detailed in Castillo-Jordán et al. (2024) and are based on those developed by McKechnie (2014) for longline extraction fisheries, and Tremblay-Boyer et al. (2018) for longline index fisheries. 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 is 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. Given that 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 where these are the same fisheries.

5 Model description

5.1 General characteristics

The model can be considered to consist of several components, (i) the dynamics of the fish population; (ii) the fishery dynamics; (iii) the observation models for the data; (iv) the parameter estimation procedure; and (v) stock assessment interpretations. Detailed technical descriptions of components (i)–(iii) are given in Hampton and Fournier (2001) and Kleiber et al. (2019), and summarised below. In addition, we describe the procedures followed for estimating the parameters of the model and the way in which stock status conclusions are drawn relative to a series of reference points.

5.2 Population dynamics

The model partitions the population into 10 annual age-classes. The last age-class comprises a "plus group" in which mortality and other characteristics are assumed to be constant. The population is "monitored" in the model at annual time steps, extending through a time window of 1979–2022.

Given the limited data available to estimate and define the movement dynamics of striped marlin, a simple model structure was adopted with a single model region. This structure assumes a well-mixed population throughout the spatial extent of the model. In order to account for the observed spatio-temporal differences in the size structure of the catch, a sub-regional structure was used to define the fisheries. This areas-as-fleets approach allowed the flexibility to estimate specific size-based selectivity functions to match the pattern of observations in each sub-region.

While there is some evidence to suggest that a degree of sexual dimorphism exists for striped marlin (Kopf et al., 2011, 2012), a single sex model was used in the assessment as the biological and sex-specific fishery data was considered insufficient, plus the two-sex catch-conditioned model in MFCL has not yet been fully tested for production assessments.

Following research conducted prior to the 2012 assessment (Piner and Lee, 2011; Kopf et al., 2011, 2012), very little progress was made to address the key biological uncertainties for SWPO striped marlin raised in the 2019 assessment. However, Farley et al. (2021a) revised current knowledge of both maturity and growth with resulting changes in this assessment (Figures 12 to 15).

The main population dynamics processes are as follows.

5.2.1 Recruitment

Recruitment is defined as the appearance of fish from age-class one in the population (Fournier et al., 1998). Striped marlin are known to spawn during the final quarter of the calendar year (October - December) in the sub-tropical latitudes of the SWPO (Kopf et al., 2012). On this basis, recruitment to the model population was assumed to occur annually in the middle of the first quarter (February) following a one year lag between when fished spawned and the ensuing recruits entered into the population.

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. Steepness is defined as the ratio of the equilibrium recruitment produced by 20% of the equilibrium unexploited spawning biomass to that produced by the equilibrium unexploited spawning biomass (Francis, 1992; Harley, 2011). Typically, fisheries data are not very informative about the steepness parameter of the SRR. Therefore h was fixed at a moderate value (0.80) for the diagnostic model, with a prior for h used to select values in the Monte Carlo model ensemble approach, with another independent prior used to select values of M-at-age, rather than the approach used in the 2019 assessment of systematically choosing from a restricted set of fixed values from these uncertainty axes. This is consistent with a meta-analysis based on reproductive ecology (Brodziak et al., 2015) which found the median steepness for striped marlin to be 0.87 (0.38-0.98: $10^{th} - 90^{th}$ percentile range). The SRR was calculated over the entire model period except for the last two years as the terminal recruitments are often poorly estimated by the model. These were fixed at the geometric mean recruitment level which is consistent with recent WCPO tuna stock assessments.

The SRR was incorporated so that yield analyses could be undertaken for stock assessment purposes, and for the determination of equilibrium and depletion based reference points. Typically in WCPO tuna and billfish stock assessments (Castillo Jordán et al., 2022; Day et al., 2023; Magnusson et al., 2023) a weak penalty (equivalent to a CV of 2.2) is applied for the deviation from the SRR so that it would have a negligible effect on recruitment and other model estimates (Hampton and Fournier, 2001), but still allow the estimation of asymptotic recruitment. However, previous stock assessments for striped marlin (Davies et al., 2012; Ducharme-Barth et al., 2019) noted a very strong trend in the recruitment deviates when the weak penalty was applied, and thus adopted a moderately restrictive penalty (equivalent to a CV of 0.2) in order to mediate this effect. The current assessment maintains the penalty on recruits used in the 2012 and 2019 diagnostic models, however a one off sensitivity to a weaker recruitment penalty was carried out.

5.2.2 Initial population

The population age structure in 1979 was assumed to have been influenced by years of fishing prior to 1979 and there are catch records going back to 1952 (Figure 3). The initial fishing mortality was assumed to be 1.7 times natural mortality ($Z = 1.7M$), which resulted in an initial depletion of around 0.35 in 1979. This is a similar depletion level to the diagnostic model in the 2019 stock

assessment and the depletion level in models run with the full catch history (Figure 3). Sensitivities were conducted to this assumed initial exploitation rate in 1979 (Section 7.5).

5.2.3 Growth

The standard assumptions for WCPO assessments fitted in MFCL were made concerning age and growth: i) the lengths-at-age are normally distributed for each age-class; ii) the standard deviations of length for each age-class are a log-linear function of the mean lengths-at-age; iii) the probability distributions of weights-at-age are a deterministic function of the lengths-at-age and a specified length-weight relationship and mean lengths-at-age follow a von Bertalanffy (VB) growth curve. These processes are assumed to be regionally and temporally invariant.

Growth for striped marlin was updated from the 2019 assessment, using an externally derived von Bertalanffy growth curve (Farley et al., 2021b). This updated growth model predicts much faster growth for younger aged individuals than the growth assumed in the diagnostic model from the 2019 assessment (Ducharme-Barth et al., 2019). The growth used in the diagnostic model in 2019 used the spine-based "back-calculated 1" growth model from Kopf et al. (2011), and the uncertainty grid included a preliminary growth model developed by CSIRO, combining daily age estimates from Kopf et al. (2011) with a limited number of annual age estimates from otoliths. Farley et al. (2021b) subsequently produced updated growth parameters based on a larger sample of otoliths for annual ageing and data for calculating decimal age, avoiding inherent biases in back calculation methods, and the resulting updated otolith-based growth closely aligned with the "observe" spine-based growth from Kopf et al. (2011).

This new growth model (Farley et al., 2021b) was recommended as the best available growth given the limited quantity of age-length samples, and was adopted in the 2024 assessment of striped marlin. This updated growth model is preferred because otolith analysis is generally considered more reliable for ageing and spines are subject to resorption making them more difficult to read accurately in older fish. The updated growth model suggests striped marlin have a faster growth rate and potentially longer lifespan (15 years) than previously indicated by Kopf et al. (2011).

Von Bertalanffy growth is parameterized in MFCL using the length-at-first-age (L_1), length-at-last-age (L_2), and K as leading parameters. Following the previous assessment, growth was fixed in the 2019 diagnostic model using to the values ($L_1 = 142.8446$ cm EFL; $L_2 = 208.0635$ cm EFL; $K = 0.8365812$). While the mean lengths-at-age were fixed at these values, the standard deviations of lengths-at-age were estimated when fitting the population model.

Weight-at-age is a function of length-at-age ($W_a = \alpha L_a^\beta$) where the parameters α and β from the length-weight relationship were obtained from an updated analysis of current and historical port sampling data under WCPFC Project 90 (Macdonald et al., 2024): $\alpha = 5.39942 \times 10^{-7}$ and $\beta = 3.583776$.

5.2.4 Natural mortality

The 2019 stock assessment assumed age-specific natural mortality values from (Piner and Lee, 2011). In this assessment, we chose to use the Lorenzen functional form for M-at-age (with M-at-age being inversely proportional to the mean length at age). The Center for the Advancement of Population Assessment Methodology (CAPAM) "Tuna Stock Assessment Good Practices Workshop" held in 2023 recommended applying an age-specific pattern in M by using the inverse mean length-at-age method developed by Lorenzen (1996). Furthermore, average M should be scaled using a maximum age ($A_{\max}$) approach when estimating M-at-age internally is not possible, as was the case with the current assessment. For this assessment, we have adopted this approach for specifying M-at-age and have followed the method of Hamel and Cope (2022a) ($5.4/A_{\max}$). We applied an assumed $A_{\max}$ of 15 years based mainly on the observations of the oldest fish observed (Farley et al., 2021a). An $A_{\max}$ of 15 years results in an expected value for average M of 0.36. Following the CAPAM Workshop advice, we allowed M-at-age to be inversely proportional to mean length-at-age following Lorenzen (1996). We assumed that the value of 0.36 could be applied as a simple average M ($\bar{M}$) for the mature age classes (2–10+) in the model (Hamel and Cope, 2022a; Hoyle et al., 2023) and configured the Lorenzen function to meet this assumption.

5.2.5 Maturity

Reproductive output (maturity) at age, which is used to derive spawning biomass, attempts to provide a measure of the relative contribution of fish at different ages to the next generation. The age-at-maturity ogive in the 2019 assessment was based on studies from Kopf et al. (2012). To account for potential sex-specific differences in growth at larger sizes, the maturity-at-length ogive was computed as the product of the sex-ratio at length and the proportions of females mature-at-length from Farley et al. (2021a).

5.3 Fishery dynamics

Interaction of fisheries with the population occurs through fishing mortality. Fishing mortality is assumed to be a composite of several separable processes: selectivity which describes the age-specific pattern of fishing mortality, catchability which scales fishing effort to fishing mortality and fishing effort. In contrast to the population dynamics, which occur at an annual time step, the fishery dynamics are assumed to interact with the stock at a quarterly time step. Given the limited amounts of composition data available for some of the defined fisheries, some of these components are shared between fisheries. These groupings and the fleet specific definitions of the different components are defined in Table 3.

5.3.1 Selectivity

Selectivity is fishery-specific and assumed to be time-invariant. Selectivity parameters are age-based but estimated in a way that smooths the estimates according to the overlap in length-at-age

([Kleiber et al., 2019](#)). Differences in selectivities among fisheries using the same methods (longline or recreational gears) in different sub-regions of the model region may be proxies for spatial structuring of the striped marlin population by size.

In the diagnostic model, the different selectivity functions were implemented as cubic splines. This is a form of smoothing where the number of parameters for each fishery is the number of cubic spline "nodes" that characterize the selectivity over the age range of the population. Consistent with the 2019 stock assessment, four nodes were used to parameterize the the selectivity patterns. This allowed a considerable flexibility in the estimated shape of the curve while minimizing the number of parameters estimated. These selectivities were further constrained, following [Ducharme-Barth et al. (2019)](#), by assuming that selectivity was invariant beyond a given age (6–7) since there was limited change in growth beyond this point.

Limited length frequency data were available for a number of fisheries and the selectivities for these fisheries were assumed to be equivalent to other fisheries of the same gear operating within the same sub-region. Specifically, the aggregate longline fisheries in sub-regions 1, 3 and 4 (i.e. fisheries 11.LL.ALL.1, 13.LL.ALL.3 and 14.LL.ALL.4) were assumed to have equivalent selectivities to the corresponding Japanese longline fisheries. Fishery 12.LL.ALL.2 had enough composition data to be estimated independently. Very limited size data were available from fishery 05.TW.LL.4 and the selectivity was assumed to be equivalent to fisheries 04.JP.LL.4 and 14.LL.ALL.4 , although limited size data were available from either fishery. Separate cubic spline selectivities were estimated for the two Australian longline fisheries (06.LL.AU.2 and 07.LL.AU.3). The selectivity of the New Zealand longline fishery, 08.LL.NZ.3, was assumed the same as the Australian longline fishery in sub-region 3, 07.LL.AU.3. The selectivity for the Australian recreational fishery, 09.Rec.AU.3, was not grouped with any of the other fisheries and was estimated using a cubic spline.

The selectivity for the New Zealand recreational fishery, 10.REC.NZ.3, was penalized to be non-decreasing as a function of age (i.e. asymptotic selectivity). This assumption appears reasonable as this fishery consistently caught the largest individuals across all fisheries.

5.4   Likelihood components

There are three data components that contribute to the log-likelihood function for the striped marlin stock assessment: the index fishery CPUE data and the size and weight composition data.

5.4.1   Index fishery CPUE likelihood

In previous catch-errors models, abundance indices were constructed for extraction fisheries by assuming that catchability remained constant over time. In catch-conditioned models, a new approach has been implemented to directly model the CPUE for "index" fisheries. While these index fisheries exist in the model and in the data inputs as defined "fisheries", they differ from the regular "extraction" fisheries in that no catch is extracted and their CPUE is modelled directly as a lognormal

likelihood contribution. Each CPUE index series is assigned a likelihood weight in the form of an index-specific σ, representing a variance term, describing the magnitude of observation error, based on the output from the mean of annual σ values obtained from the CPUE standardisation using sdmTMB.

5.4.2 Length and weight frequency

The distributions for the length- and weight-frequency compositional data are assumed to be approximated by robust normal distributions, with the variance determined by the input sample size and the observed length or weight-frequency proportion. Size composition distributions for each time period are typically assigned input sample sizes lower than the number of fish measured. Lower input sample sizes recognize that (i) length- and weight-frequency samples are not truly random (because of non-independence in the population with respect to size) and would have higher variance as a result; and (ii) the model does not include all possible process error, resulting in further under-estimation of variances. The input sample sizes used by the model are capped at 1,000 within MFCL.

In previous striped marlin and WCPFC assessments, a down-weighting factor was arbitrarily applied to the input sample sizes from every fishery, with this down-weighting factor often used as an axis in an uncertainty grid. In the current assessment, the input size frequency sample sizes were modified using the data-weighting method developed by Francis (2011), referred to here as the "Francis method". The Francis method involves calculating a multiplier (not greater than one) for the LF data for each fishery from a previous model run (either using the initial input values, or those from a previous iteration) with the aim of matching the input variance (from the observed data) to the output variance (from the standardized residuals). This new multiplier for each fishery is applied to produce modified input sample sizes for each fishery for the next iteration. Note that this process is applied after the input sample size has been capped at 1000 samples. The model is rerun with the modified input sample sizes, and repeated iteratively until the modified sample sizes stabilise, producing approximately matching input and output variances for each fishery.

Some of the WCPO tuna assessments fit to either the weight or the length frequency data but typically not both, as there is frequently conflict between the two types of data. Given the comparatively few samples available for striped marlin, compared to the tropical tunas, both length and weight composition data were included.

5.5 Parameter estimation and uncertainty

The parameters of the model were estimated by minimising the sum of the negative log-likelihoods associated with each of the data components plus the negative log of the probability density functions of the priors and penalties specified in the model. The optimisation is continued either until a pre-specified minimum gradient is achieved or until a maximum number of iterations is reached. This optimisation is performed by an efficient algorithm using exact derivatives with respect to

the model parameters (auto-differentiation, Fournier et al. (2012)). Estimation was conducted in a sequence of phases, the first of which uses somewhat arbitrary starting parameter values. A bash shell script, or "doitall", implements the phased sequence for gradually estimating more and more parameters for the model. Some parameters were assigned starting values consistent with available biological information. The values of these parameters are provided in the .ini input file or within the MFCL code.

Generally, models were run with a gradient criterion of 10^{-5} and a with the maximum iteration criterion of greater than 20 000.

Estimation uncertainty A positive definite Hessian is desirable to calculate parameter or estimation uncertainty for individual models, although as noted, MFCL has advanced procedures for computing estimation uncertainty in the absence of a positive definite Hessian, with negative eigenvalues. SC18 requested the Hessian status be reported for all models used for management advice (which includes all grid models). A positive definite Hessian was one criteria for accepting or rejecting candidate models from the uncertainty ensemble used in this model.

Model uncertainty Previous WCPFC stock assessments referred to a "structural uncertainty grid" describe the main sources of structural and data uncertainty in the assessment. This now referred to as model uncertainty and this assessment uses a Monte Carlo ensemble uncertainty. Previous experience has shown that overall uncertainty is dominated by the model uncertainty, rather than estimation uncertainty. For this assessment, the model uncertainty no longer uses a factorial grid of model runs, as in the 2019 assessment (Ducharme-Barth et al., 2019), which incorporated selected uncertainties explored in one-off sensitivity analyses. Instead, we follow a similar approach used by Ducharme-Barth et al. (2021) for swordfish, repeatedly selecting a value of steepness from a prior distribution for steepness, and pairing this with an independently drawing from a prior distribution on the average value of natural mortality.

The combined model and estimation uncertainty is recommended to form the basis for assessing uncertainty and risk for the key stock status indicators.

5.6 Diagnostics methods

Diagnostic tools are vital for evaluating the quality of integrated stock assessment models for informing management advice (Carvalho et al., 2017, 2021) and there is no single diagnostic tool that is capable of comprehensive evaluation for all models. As such, a suite of tools from the "diagnostic toolbox" was applied to the current model assessment.

5.6.1 Convergence diagnostics

All models in this assessment were considered to be converged if:

1. a maximum parameter gradient of 10^{-5} was achieved;

2. the Hessian was positive definite;

3. the model fit could not be significantly improved by jittering the estimated parameters (Carvalho et al., 2021);

4. estimated parameters should not be on their bounds; or

5. estimated parameters should not be highly correlated with each other ($-0.9 > r > 0.9$).

5.6.2   Model fit

Plots of observed and predicted CPUE and size composition data were examined, including residual plots.

5.6.3   Age-Structured Production Model

An age-structured production model (ASPM) diagnostic (Carvalho et al., 2017, 2021; Maunder and Piner, 2015; Minte-Vera et al., 2017) 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 to a fixed value of 0; 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.

5.6.4   Catch Curve Analysis

A catch curve analysis (CCA) is essentially the reverse of the ASPM, whereby the CPUE index data are removed from the model and all parameter estimation retained. The CCA indicates the information on population trends and scaling provided by the size composition data.

5.6.5   Likelihood profile

For highly complex population models fitted to large amounts of often conflicting data, absolute estimates of total abundance can be unstable. Therefore, a likelihood profile analysis was undertaken of the marginal penalized likelihood in respect of population scaling, following the procedure outlined by McKechnie et al. (2017) and Tremblay-Boyer et al. (2017). Likelihood profiles and Piner plots are only presented for the diagnostic model.

5.6.6 Retrospective analysis

Retrospective analyses were undertaken as a general test of the stability of the model. A robust model, when rerun with data for the terminal year sequentially excluded (Cadigan and Farrell, 2005), should produce outputs that are variable across runs, and without a systematic pattern in either the scaling or time-series trends. The Mohn's rho statistic (Hurtado-Ferro et al., 2015), a measure of the average relative bias of retrospective estimates, was computed for spawning biomass (SB) and spawning biomass depletion ($SB/SB_{F=0}$) to indicate whether significant retrospective bias was present in the model.

5.6.7 Sensitivity analyses

Comparisons of the spawning depletion and spawning potential trajectories for the diagnostic model and the related one-off sensitivity models were explored during model development, in particular to explore the initial start year for the model, the initial fishing mortality assumed for a model start year of 1979, recruitment variability, the number of age classes used in the model and sensitivities to alternative CPUE series from Australian longline data, as a replacement index, and New Zealand recreational data, as an additional index.

5.6.8 Monte–Carlo model ensemble uncertainty estimation methods

Typically, three types of uncertainty should be incorporated into the estimates of stock status used for management advice. One involves the statistical uncertainty of the estimates produced by individual models, often referred to as "estimation uncertainty" (described above). The second involves "model uncertainty", which is the uncertainty in the structural and fixed-parameter assumptions underpinning individual models, such as fixed M and h. The third involves data inputs, such as alternative abundance indices or other data. Stock assessments of tuna for the WCPFC have often included an approach to assess the model uncertainty in the assessment model by running a factorial "grid" of models to explore the interactions among selected "axes of uncertainty". The grid contains all combinations of parameter settings or assumptions for each uncertainty axis and this was commonly referred to as the "structural uncertainty grid".

In the current assessment, the characterization of uncertainty in management reference points and quantities of interest was accomplished by applying a Monte Carlo model ensemble approach following the methods introduced by Ducharme-Barth and Vincent (2021), implemented in the 2021 stock assessment of south-west Pacific swordfish (Ducharme-Barth et al., 2021) and recommended as best practice by Neubauer et al. (2023). Building from the "model uncertainty grid", the model ensemble approach continues to consider the effects of model uncertainty and extending this to also account for estimation uncertainty from each model in the ensemble. This allows for a more holistic and transparent description of the uncertainty in estimates of stock status. Another key difference between the model ensemble and the model uncertainty grid is the relaxation of the full factorial design. Instead of choosing set levels for certain fixed parameters (e.g. $h \in 0.65, 0.8, 0.95$),

a random set of the fixed parameters is drawn from a prior distribution for each model in the ensemble. This approach has the advantage of implicitly weighting the ensemble to the most likely parameter combinations given the shape of the prior.

For aspects of model uncertainty that cannot be parametrized using prior assumed distributions (e.g., fitting to alternate CPUE indices), a full factorial approach could still be used and then overlaid on the parameter draws from the priors in a hybrid factorial ensemble

The estimation uncertainty for each model in the ensemble was determined as described in Section However, for the ensemble models, we computed the estimation uncertainty (i.e., standard deviations of the estimates) for only the key reference point variables ($SB_{\text{recent}}/SB_{F=0}$, $F_{\text{recent}}/F_{\text{MSY}}$ and $SB_{\text{recent}}/SB_{\text{MSY}}$) and not the full set of time-series estimates as an efficiency measure. This was done for all models in the model ensemble and the estimation uncertainty was combined across models in a parametric bootstrap similar to the approach used in stock assessments conducted by the International Pacific Halibut Commission (Stewart and Martell, 2014) and applied in the 2023 yellowfin and bigeye assessments (Day et al., 2023; Magnusson et al., 2023).

Natural mortality and steepness emerged from the sensitivity analyses as the key potential sources of model uncertainty impacting stock-assessment-related estimates (see Section 7.5). Therefore, prior distributions for these parameters were constructed as described below.

Natural mortality prior distribution: As described in Section 5.2.4, $\bar{M}$ of 0.36 yr^{-1} was specified using the A_{max} approach assuming an A_{max} of 15 years. Hamel and Cope (2022b) recommended a CV of 0.31 for $\bar{M}$. However, we found that sampling from this distribution produced a significant number of low (<0.25) and high (>0.7) values of $\bar{M}$ that were considered outside the range. Therefore, the CV was reduced to 0.2 to focus the replicates on a more plausible range of $\sim$0.25–0.55 . Each value of $\bar{M}$ sampled from the prior was converted to M-at-age using the Lorenzen approach described in Section 5.2.4.

Steepness prior distribution: In order to develop the prior on h, the approach of Brodziak et al. (2011) was considered, which uses various life history parameters, including M-at-age, in characterising a prior for h. Following the approach used for South Pacific albacore in 2024 (Teears et al., 2024), to avoid unrealistic values for h, a censored (0.2–1) prior with a mean of 0.845 and $\sigma = 0.082$ was developed resulting in the prior distribution used for h.

The approach adopted was similar to an approach used by Ducharme-Barth et al. (2021). Firstly, we wanted to recognise the application of Brodziak et al. (2011) that incorporated life-history criteria and indicated that h, on average, was likely to be considerably higher than the values typically assumed in previous assessments. Secondly, we also wanted to respect the approach used in the previous assessment, including meta-analyses, that indicated a reasonable range of h was likely to be $\sim$0.65–0.95, as applied in the factorial grid in 2019.

5.7 Stock assessment interpretation methods

Several ancillary analyses using the fitted models were conducted in order to interpret the results for stock assessment purposes. The methods involved are summarized below and further details can be found in Kleiber et al. (2019).

5.7.1 Depletion and fishery impact

Many assessments estimate the ratio of recent to initial biomass (usually spawning biomass) as an index of fishery depletion. The problem with this approach is that recruitment may vary considerably over the time series, and if either the initial or recent biomass estimates (or both) are "non-representative" because of recruitment variability or uncertainty, then the ratio may not measure fishery depletion, but simply reflect recruitment variability.

This problem is better approached by computing the spawning potential time series (at the model region level) using the estimated model parameters, but assuming that fishing mortality was zero. Because both the estimated spawning potential SB_t (with fishing), and the unexploited spawning potential $SB_{F=0[t]}$, incorporate recruitment variability, their ratio at each quarterly time step (t) of the analysis, $SB_t/SB_{F=0[t]}$, can be interpreted as an index of fishery depletion. The computation of unexploited biomass includes an adjustment in recruitment to acknowledge the possibility of reduction of recruitment in exploited populations through stock-recruitment effects. To achieve this the estimated recruitment deviations are multiplied by a scalar based on the difference in the SRR between the estimated fished and unfished spawning potential estimates.

A similar approach can be used to estimate depletion associated with specific fisheries (fisheries impact analysis) or groups of fisheries. Here, fishery groups of interest - six fisheries impact groups were defined: four joint DWFN-PICT longline groups defined in each of the 4 sub-regions, a single group for the major "domestic" longline fleets (Australia and New Zealand) in sub-regions 2 and 3 combined, and the combined recreational fisheries for Australia and New Zealand, are removed in-turn in separate simulations. The changes in depletion observed in these runs are then indicative of the depletion caused by the removed fisheries.

5.7.2 Reference points

Unlike the tropical tuna species assessed in the WCPO which have target and limit reference points based on the level of recent spawning biomass (SB_{recent}) relative to the unfished spawning biomass ($SB_{F=0}$), striped marlin does not have formally defined target and limit reference points. Assessment results are presented relative to both MSY-based reference points and depletion based reference points.

The key MSY-based reference points for striped marlin are $SB_{\text{recent}}/SB_{\text{MSY}}$, $SB_{\text{latest}}/SB_{\text{MSY}}$, and $F_{\text{recent}}/F_{\text{MSY}}$ (Table 4). SB_{recent} and SB_{latest} are the mean of estimated spawning biomass over the period 2019-2022 and the estimate in the last year of the model (2022), respectively. The other key

reference point, $F_{\text{recent}}/F_{\text{MSY}}$, is the estimated average fishing mortality over a recent period of time (F_{recent} : 2018-2021 for the current assessment) divided by the fishing mortality producing MSY (as described in Section 5.7.3).

Depletion based reference points, as used for the tropical tuna species, are also provided. Stock status is referenced against these points by calculating $SB_{\text{recent}}/SB_{F=0}$ and $SB_{\text{latest}}/SB_{F=0}$, where $SB_{F=0}$ is calculated as the average over the period 2012-2021 (Table 4).

5.7.3   Yield analysis

The yield analysis consists of computing equilibrium catch (or yield) and spawning potential, conditional on a specified basal level of age-specific fishing mortality (F_a) for the entire model domain, a series of fishing mortality multipliers (f_{mult}), the natural mortality-at-age (M_a), the mean weight-at-age (w_a) and the SRR parameters. All of these parameters, apart from f_{mult}, which is arbitrarily specified over a range of 0–50 (in increments of 0.1), are available from the parameter estimates of the model. The maximum yield with respect to f_{mult} can be determined using the formulae given in Kleiber et al. (2019), and is equivalent to the MSY. Similarly, the spawning potential at MSY SB_{MSY} can be determined from this analysis. The ratios of the recent average levels of fishing mortality and spawning potential to their respective levels at MSY are determined for all models of interest. This analysis was conducted for all models in the Monte Carlo ensemble and thus includes alternative values of steepness assumed for the SRR.

Fishing mortality-at-age (F_a) for the yield analysis was determined as the mean over a recent period of time (2018–2021). We do not include 2022 in the average as fishing mortality tends to have high uncertainty for the terminal data year of the analysis and the catch and effort data for this terminal year are potentially incomplete. Additionally, recruitments for the terminal year of the model are constrained to be the geometric mean across the entire time series, which affects the F for the youngest age classes.

MSY was also computed using the average annual F_a from each year included in the model (1979–2022). This enabled temporal trends in MSY to be assessed and a consideration of the differences in MSY levels under historical patterns of age-specific exploitation.

5.7.4   Kobe analysis and Majuro plots

For the standard yield analysis (Section 5.7.3), the fishing mortality-at-age, F_a, is determined as the average over some recent period of time (2018–2021). In addition to this approach the MSY-based reference points (F_t/F_{MSY}), and SB_t/SB_{MSY}) are computed by repeating the yield analysis for each year in turn. This enabled temporal trends in the reference points to be estimated and a consideration of the differences in MSY levels under historical patterns of age-specific exploitation. This analysis is presented in the form of dynamic Kobe plots and "Majuro plots", which have been presented for all stock assessments in recent years.

6 Model runs

6.1 Model developments from the last assessment

6.1.1 Stepwise model development

The progression of the model development (referred to as the "stepwise") from the 2019 diagnostic model to the 2024 diagnostic model is described here. These stepwise changes were undertaken in order to identify the impact of each modification to the assessment outcomes. The changes incorporated at each step in the diagnostic model development are summarised below, including the model names used in the figures listed to describe each step. Each step builds from the previous step, and retains all the previous changes.

1. [**00_Diag2019**] The 2019 striped marlin diagnostic model.

2. [**01_NewMFCL**] This step updates the MULTIFAN-CL (MFCL) executable to version 2.2.7.0.

3. [**02_CatchCond**] This step implements a catch-conditioned model framework instead of using the catch-errors framework, as used in the 2019 assessment, by removing the catch-errors specific MFCL flags and introducing grouped index fisheries with common catchability.

4. [**03_NewCPUE**] This step updates the CPUE.

5. [**04_NewData**] This step incorporates all the new data and historical data revisions up to the end of 2022. This included revisions to data from 1952-2018, and five years of new data from 2018-2022, new length and weight composition data and new CPUE modified slightly with the additional five years of data. There are also an additional five years of recruitment deviations estimated as part of this step. The objective function is not comparable between this step and the previous step because all components of the data have changed.

6. [**05_CatchWeighting**] This step incorporates catch-weighting of composition data for extraction fleets and CPUE-weighting of composition data for index fleets, to ensure that more representative data is given greater weight than less representative data. Previous striped marlin assessments did not catch-weight the composition data and simply used raw size composition data, assuming that the samples taken were representative of the fishery.

7. [**06_5cm_bins**] This step changes the length compositions from 6 cm length-bins, as used in previous assessments, to 5 cm length-bins. This is complicated slightly by the use of different measurement protocols and the need to convert all raw length data to eye orbital fork length (EFL) (Castillo-Jordán et al., 2024), but the choice of 5 cm length bins was considered to be the best use of the raw data available.

8. [**07_NatMortality**] This step implements the Lorenzen form of natural mortality. The use of Lorenzen mortality and estimating the scale, if possible, was recommended as good prac-

tice in the CAPAM tuna stock assessment good practices workshop held in March 2023, https://www.capamresearch.org/Tuna_Stock_Assessment_Good_Practices_Workshop.

9. [**08_Maturity**] Maturity was updated using new information (Farley et al., 2021a).

10. [**09_LW**] The fixed length-weight conversion parameters were updated using new information (Macdonald et al., 2024).

11. [**10_NewGrowth**] This step updates the fixed growth, using the latest recommendations on growth from CSIRO (Farley et al., 2021a).

12. [**11_Data1979**] Start the model in 1979, rather than 1952.

13. [**12_Diagnostic2024**] Implement Francis weighting for size composition data.

6.2 Sensitivity analyses

One-off sensitivity models were explored to understand the sensitivity of the diagnostic model estimations to structural and data uncertainties. Each one-off sensitivity model was created by making a single change to the 2023 diagnostic model.

Key one-off sensitivity models explored during model development include exploring the initial start year for the model, the initial fishing mortality assumed for a model start year of 1979, recruitment variability, the number of age classes used in the model and sensitivities to alternative CPUE series from Australian longline data, as a replacement index, and New Zealand recreational data, as an additional index.

No one-off sensitivity was conducted to size composition data weighting becuase this data was objectively weighted using the Francis iterative reweighting method.

7 Results

7.1 Consequences of key model developments

The progression of model development from the 2019 diagnostic model to the 2024 diagnostic model is described in Section 6.1.1 and the results are displayed in Figures 16 and 17 . The stepwise analysis is presented by running MFCL at each step, plotting the results and attributing any substantial change in results to changes in the modelling, the data or structural changes involved.

A total of 13 steps are described here, with steps which involve changes in the model structure. A brief description of the consequences of this progression through the major steps is provided below, focusing on the key management quantities of spawning potential SB and dynamic spawning potential depletion $SB/SB_{F=0}$.

1. [**00_Diag2019**] The 2019 striped marlin diagnostic model.

2. [**01_NewMFCL**] This step results in essentially no change in either $SB/SB_{F=0}$ or SB, and a small deterioration in the value of the objective function.

3. [**02_CatchCond**] This step represents a major change to the model structure incorporating CPUE likelihood from the index fisheries and a reduction from 2,647 to 104 parameters estimated, compared to the 01_NewMFCL step. This step results in a small decrease in $SB/SB_{F=0}$ and minor adjustments to SB. Changes to the objective function are not meaningful at this step, given the structural changes to the model.

4. [**03_NewCPUE**] This step results in essentially no change in either $SB/SB_{F=0}$ or SB. Changes to the objective function are not meaningful at this step, given the structural changes to the model.

5. [**04_NewData**] This step results in minimal changes to either $SB/SB_{F=0}$ or SB. The objective function is not comparable between this step and the previous step because many components of the data have changed.

6. [**05_CatchWeighting**] This step results in a small decrease in both $SB/SB_{F=0}$ and SB. The total objective function is not comparable between this step and the previous step because the input size composition data has changed, although the CPUE component of the likelihood has deteriorated slightly.

7. [**06_5cm_bins**] This step results in essentially no change in either $SB/SB_{F=0}$ or SB. The objective function is not comparable between this step and the previous step because the input length composition data has changed, although the CPUE component of the likelihood has improved slightly, while the weight component of the likelihood has deteriorated.

8. [**07_NatMortality**] This step results in essentially no change in either $SB/SB_{F=0}$ or SB. The objective function is also essentially unchanged, with a very small deterioration, largely due to the weight data.

9. [**08_Maturity**] This step results in essentially no change to either $SB/SB_{F=0}$, SB or the objective function.

10. [**09_LW**] This step results in an increase in $SB/SB_{F=0}$ or SB, with a very small improvement to the objective function.

11. [**10_NewGrowth**] This step results in an increase in $SB/SB_{F=0}$, due to some large changes in SB in the 1950s, with some deterioration in the objective function. The changes to SB result in implausibly large early values. These implausibly high estimates of SB seem to be linked to the changes in growth in combination with catch-conditioning.

12. [**11_Data1979**] Changes to the objective function are not meaningful at this step. Importantly, this step removes the high SB values estimated in the 1980s in the previous step, and gives a slightly higher final value for $SB/SB_{F=0}$.

13. [**12_Diag2024**] Changes to the objective function are not meaningful at this step. This step moderates the increases seen in the last few years in SB and $SB/SB_{F=0}$ seen at the new growth step, with values closer to those seen earlier in the stepwise.

Modifying the growth resulted in estimates of unrealistically large recruitment events early in the time series, which appear to be biologically unlikely, and appear to be a model artifact allowing the likelihood to be improved due to large catches and large catch variation in this period, but with no size composition data to support such large recruitment deviations. While these large early recruitment estimates have a relatively small impact on depletion estimates towards the end of the time series, this produces a model that seems implausible. To avoid producing a model which gives implausible recruitment early in the time series, with consequences on the estimated average recruitment for any projected average recruitment late in the series, and for future projections, alternative start dates for the model were explored, with an assumed initial fishing mortality, prior to the model commencing.

7.2 Model fit for the diagnostic model

This section discusses the model results for the diagnostic model, defined by the final step in the stepwise model development (Figures 16 and 17), in particular the fit to the various sources of data, and the biological and fisheries-related parameters.

7.2.1 Standardised CPUE: index fishery

The model fits to the index fishery standardised CPUE data were moderate to poor (Figure 18). The model roughly followed the major oscillations in the CPUE, but with some poor residual patterns (Figure 19), notably at the beginning (over-predicting abundance for seven successive years compared to the CPUE, negative residuals) and the end of the time series (under-predicting abundance for eleven successive years compared to the CPUE, positive residuals).

7.2.2 Size composition data

This discussion is focused on aggregated fits to the weight and length composition, aggregated over all time periods, rather than the individual quarterly based size composition fits that contribute directly to the likelihood. The number of samples available is quite variable between fisheries for both length and weight composition data (Table 2).

The aggregated length composition fits are of variable quality (Figure 20), with some fits quite good (for fisheries 10.REC.NZ.3, 03.LL.JP.3), and others quite poor (for fisheries 12.LL.ALL.2 and 15.INDEX.1_4), and with considerable variation in samples sizes between fisheries. The time series fits (Figure 22) also show these fits, but with more temporal information, and showing consistently poor fits to fishery 12.LL.ALL.2, despite the large sample size.

The aggregated weight composition fits (Figure 21) are generally better than the length fits, and

generally with larger sample sizes as well. In particular there are good aggregated fits (for fisheries 10.REC.NZ.3 and 07.LL.AU.3) and reasonable aggregated fits for the fishery with the most samples (06.LL.AU.2). The time series fits (Figure 23) also show these fits, but with more temporal information, and showing generally much better fits than the length data, noting that there are considerably more weight samples than length samples. Despite having good aggregate fits, Fishery 07.LL.AU.3 appears to have the worst of the time series fits, over-predicting the fish weights in most time period, partly due to conflict in the length data from the same fishery. The lower weight distributions observed in 2021 and 2022 for fisheries 06.LL.AU.2 and 07.LL.AU.3 were not able to be well fitted by the model.

Notably there were also some poor fits for some fisheries in the time series plots for length and weight composition data in the 2019 assessment, so this indicates there may be some structural problems with the model or with the data, which have still not been addressed adequately.

7.3 Population dynamics estimates

7.3.1 Selectivity

A range of selectivity curves are estimated for the different fisheries in the model and are classified by gear type.

For the longline fleets, the age-, length- and weight-specific selectivities are shown in Figures 24 to 26 and generally peaking at around age 2 or 3 years, and declining for all long line fleets after age 3 and to zero by age 4 for the Australian and New Zealand longline fleets.

For the recreational fleets, the age-, length- and weight-specific selectivities are shown in Figures 27 to 29 with the Australian fleet estimated to only catch individuals aged 3 and 4 with the New Zealand recreational fishery the only fishery with a penalty imposed on non-decreasing selectivity and as a result this is the only fishery with an asymptote at full selectivity for the oldest fish.

7.3.2 Growth and natural mortality

The Lorenzen form of natural mortality is used in the 2024 assessment. The comparison between the estimated natural mortality-at-age in the 2024 diagnostic model and the 2019 diagnostic model is shown in Figure 13. Compared to the mortality used in the 2019 diagnostic model, the Lorenzen form of natural mortality features higher levels of natural mortality for one year old fish, and a slightly lower mortality-at-age for all age classes from age 2 years and above.

Growth estimates from the diagnostic model indicated rapid growth until age 3 years, at which age striped marlin approach their maximum size, with the comparison to the growth used in the 2019 assessment shown in Figure 12. The estimated growth parameters were 142.84 cm, 208.06 cm, and 0.83658 for $L1$, $L2$, and k, respectively. The M-at-age curve derived from the growth estimates are shown in Figure 13.

7.3.3 Maturity

Maturity-at-age is derived from the fixed maturity-at-length (Figure 15) and applying the growth curve to derive the maturity-at-age for the 2024 diagnostic model. This maturity-at-age curve differs from the 2019 diagnostic model (Figure 14) due to the differences in growth curves in these two models.

7.3.4 Recruitment

The time series of estimated recruitments is shown in Figure 30. The recruitment estimates are quite variable until about 2010, with some indication of a series of below average recruitment estimates and reduced recruitment variability since 2008, noting that the last two recruitments shown in this series are not estimated, and are fixed at the mean of the time series.

The stock recruit relationship for the diagnostic model is shown in Figure 31. Recruitment deviations around the stock recruitment relationship are similar (above and below the line) across the higher biomass levels, but there appears to be some bias at the lower biomass levels.

7.3.5 Convergence diagnostics

The diagnostic model resulted in an objective function of -167076 with 76 parameters and a maximum gradient of 4.70458×10^{-6} with a positive definite Hessian achieved. A jitter analysis with 30 replicates was not able to find a model with a better objective function (Figure 32) and the spawning potential and depletion time series was very similar for all jittered replicates (Figure 33). Overall, these results provide an indication of good model convergence.

7.3.6 Biomass and biomass depletion

SB shows a gradual declining trend (Figure 34), with some variability, from the mid 1980s to around 2015, with some relative stability in SB at a low level from 2015 onwards. $SB/SB_{F=0}$ has generally shown a gradual decline since 1979, with some variation (Figure 34), with the lowest value in 2021.

7.3.7 Fishing mortality and age-specific exploitation

A general increase in fishing mortality of adult and juvenile age-classes is estimated to have occurred over the entire assessment period (Figure 35), albeit with some notable periodic declines (e.g. around 1982, 2005, 2015 and 2021), notably in the last year of the assessment, although the recent estimates are likely to be less reliable.

The estimated relative frequency at age by decade (Figure 36) indicates the dominance of the first two age classes and the low numbers of older fish throughout the time series. The estimated instantaneous mortality at age and decade indicated that prior to 2000, mortality was lower on fish aged 2 and 3, and a slight increase in mortality for fish over 5 years old over time. The estimated

annual instantaneous fishing mortality by age (Figure 37) indicates that the highest fishing mortality applies to age classes 2 and 3, and that this fishing mortality varied by year.

7.3.8 Fishery impact

It is possible to attribute the fishery impact with respect to depletion levels to specific fishery components (i.e., grouped by gear-type), in order to estimate which types of fishing activity have the most impact on the spawning potential.

Overall fishing impacts were estimated to increase gradually since 1979 (Figure 38), with an initial decline in the AU-NZ recreational fisheries and a notable increase in the AU-NZ longline fisheries starting in the late 1990s (grouped together for visual clarity in (Figure 38), although this is actually attributable wholly to increases in the Australian longline fishery, given restrictions on the NZ longline fishery during that time. The estimated fishery impacts are low for the recreational fisheries.

7.4 Diagnostics results

In the following section, we present diagnostics for the diagnostic model.

7.4.1 Age-structured production model and catch curve analysis

An age-structured production model (ASPM) was fitted, without using the estimated recruitment deviations from the diagnostic model, so that the population scale is being estimated, based only on the CPUE data. In contrast a catch curve analysis (CCA) fits the model to the composition data, without fitting to the CPUE. The CPUE has more influence on informing the population scale than the composition data up until the year 2013, when the composition data becomes more influential (Figures 39 and 40). Further investigation of the influence of the recent composition data may be worthwhile, to identify the cause of this change.

7.4.2 Likelihood profiles

This likelihood profile of the total population scaling parameter (Figure 41) shows the contribution to the total likelihood from each major data source with the total likelihood shown in black and components of the total likelihood from different data sources shown in a range of colours. This likelihood profile indicates considerable conflict in the data with the length composition data supporting a lower population scale and the CPUE supporting a higher population scale, with the weight composition data supporting some intermediate value.

Piner plots are useful for separating the individual components contributing to the total likelihood into smaller components, allowing the effect of data from different fisheries and regions on the estimated total population to be distinguished. These plots can be used to indicate influential data sources by fishery, which could lead to further investigation or verification, to ensure that these

data sources are reliable, or to direct research as to whether different processes may be important in different fisheries.

The Piner plot for length (Figure 42) shows that the length data from three fisheries support a larger biomass (11.LL.ALL.1, 01.LL.JP.1 and 05.LL.TW.4) in declining order of influence, and one fishery (12.LL.ALL.2) supports a smaller biomass and has the largest influence of all of the fisheries with length composition data.

The Piner plot for weight (Figure 43) shows that the weight data from three fisheries support a larger biomass (10.REC.NZ.5, 02.LL.JP.2 and 15.INDEX.1-4) in declining order of influence, and one fishery (07.LL.AU.4) supports a smaller biomass (with the largest influence of all the fisheries with weight composition data), with a minor contribution (favouring smaller biomass) from 08.LL.NZ.3.

7.4.3 Retrospectives

The retrospective analysis indicated the diagnostic model showed a significant retrospective pattern (Figure 44) as evidenced by a value of Mohn's rho greater than 0.3 for both SB and $SB_{F=0}$. This retrospective pattern is clearly related to new recruits entering the biomass through the fixed recruitment in the last two time steps, fixed at the mean recruitment level over the whole recruitment time series. This mean recruitment level is much higher than recent recruitment, and with individuals entering the spawning biomass aged one or two years old, with the new faster growth used in this assessment, this high late recruitment is having a clear impact. As additional data is added to the model, through retrospective peels, the previous fixed recruitment at the end of the series, for a model with less data used, are revised downwards as additional data is available to the model, leading to this concerning retrospective pattern. This pattern could likely be addressed by modifying the values of the fixed recruitment in the terminal years. Unfortunately this feature is not yet available in MFCL, but hopefully this will be an addition in the near future, to address this problem, related to the assumption that recruitment in the final year is fixed at the historical average.

7.5 Sensitivity analyses

Sensitivities to the initial start year were conducted, but the decision to start the model in 1979 was influenced by the availability of vessel ID information, for inclusion in the CPUE standardisation, and also by the implausible recruitment estimated when starting the model in 1952 with the catch conditioned formulation of the model with updated growth. A sensitivity was also conducted into the number of age classes in the model, using 15 (to match the maximum age) rather than 10 for the plus group in the model. This sensitivity made no discernible difference to the results.

7.5.1 Initial fishing mortality

Sensitivities were carried out using a range of assumed initial fishing mortality (F) when the model started in 1979, with initial F set to values of 0, M and $3M$, with the diagnostic model with $F = 0.7M$. This can also be expressed in initial total mortality (Z), where $Z = F + M$, so the three sensitivities had values for initial Z with values of M, $2M$ and $4M$, with the diagnostic model with $Z = 1.7M$. While these sensitivities all showed different values for SB and $SB/SB_{F=0}$ at the start of the time series, they all achieved very similar values at the end of the time series (Figure 45), even with quite extreme values assumed for initial fishing mortality.

7.5.2 CPUE indices

Two sensitivities were carried out with variations on the CPUE series used in the model. The first sensitivity used only the standardised CPUE series from the Australian longline fisheries operating in sub-regions 2 and 3 (Tremblay-Boyer and Williams, 2023), replacing the standardised CPUE series used in the diagnostic model, which uses data from the longline fisheries from Japan and Chinese Taipei in all subregions. The second sensitivity included an additional standardised CPUE series from the New Zealand recreational fleet (Holdsworth, 2024), operating only in sub-region 3, as well as retaining the standardised CPUE series used in the diagnostic model. The Australian CPUE series started in 1998, which explains the differences in SB and $SB/SB_{F=0}$ up until the late 1990s, so the model is unconstrained in fitting CPUE in this period. From the late 1990s onwards, this sensitivity is very similar to the diagnostic model. Using the NZ recreational series as an additional CPUE series gives slightly more optimistic results for SB and $SB/SB_{F=0}$, but it is not clear how well this CPUE series indexes the abundance of the total population, coming from a section of a single sub-region.

7.5.3 Recruitment variation

The sensitivity to the variance on recruitment was examined by using two alternative values of 0.7 and 2.2. The model was largely insensitive to this value (Figure 47). Further, models with higher recruitment variation had convergence issues, so the value for the diagnostic model was set to 0.2, which is the same value used in the 2019 assessment.

7.6 Monte-Carlo model ensemble uncertainty estimation

Analysis of the 100 models selected as candidates for the ensemble indicated 41 models did not have a positive definite Hessian and were excluded from the ensemble as a result. Only results form the remaining 59 ensemble members were included in the final results.

The results of the ensemble uncertainty analysis are summarised in several forms.

1. histograms of model uncertainty estimated from Monte-Carlo draws from the model ensemble of SB_{recent}/SB_{MSY}, and F_{recent}/F_{MSY} coloured by values of $\bar{M}$ and h (Figures 48 and 49).

2. estimates of $SB_{\text{recent}}/SB_{\text{MSY}}$, and $F_{\text{recent}}/F_{\text{MSY}}$ by values of $\bar{M}$ and h for each model in the ensemble

3. quantile trajectories (90% and 75%) are provided for $SB_{\text{recent}}/SB_{F=0}$ and SB in (Figures 50 and 51);

4. a table of summary statistics of reference points for the model ensemble is included in Table 5;

5. Majuro and Kobe plots are shown for estimates from the model ensemble and the dynamic MSY analysis (Figure 52)

The median depletion from the uncertainty grid for the recent period (2019-2022; $SB_{\text{recent}}/SB_{F=0}$) is estimated at 0.14 (80 percentile range including estimation and model uncertainty 0.09–0.21, full range 0.05–0.26). Fishing mortality has continued to increase over time for both the juvenile and adult components. With respect to maximum sustainable yield (MSY) based reference points (the WCPFC default for all billfish stocks, in lieu of the absence of any WCPFC billfish specific limit reference points), the median recent fishing mortality from the uncertainty ensemble is estimated to be above the level for achieving MSY (median $F_{\text{recent}}/F_{\text{MSY}} = 1.27$, range 0.58–3.02). 71% of the 59 ensemble models estimated $F_{\text{recent}}/F_{\text{MSY}} > 1$. The median recent spawning biomass from the uncertainty ensemble was well below the spawning biomass to achieve MSY (median $SB_{\text{recent}}/SB_{\text{MSY}} = 0.60$, range 0.18–1.24). 93% of the 59 ensemble models estimated $SB_{\text{recent}}/SB_{\text{MSY}} < 1$. Despite decreased catches since the early 2000s, based on the default MSY-based reference points, the 2024 assessment of striped marlin in the southwestern Pacific Ocean concludes that the stock is likely overfished, and undergoing overfishing.

The yield curves, as a function of F, are displayed in Figure 53. Estimated MSY has remained stable since 1980 at between 1,700 - 1,800 t (Figure 54).

8 Discussion

8.1 General remarks

While many of the core structural aspects of the 2024 assessment of striped marlin in the southwestern Pacific Ocean were consistent with the 2019 assessment (and earlier assessments), there were some key changes implemented both with regards to data inputs and biology, and also the modelling approaches and the approach to characterising uncertainty in the key management reference points. Perhaps the most noteworthy of the changes was to truncate the start year of the model from 1952 to 1979. This change came about during the stepwise model bridging when implementing the new growth recommended by Farley et al. (2021a), and the Pre-assessment workshop (Hamer, 2024). The implausibly high recruitment estimates associated with the early high catches and subsequent implausible biomass estimates were unsatisfactory. We are still not certain of the processes in the estimation process which drive this, but it appears to be linked to a combination of the new growth and catch-conditioning. We explored some possible solutions to address this, such as starting the

model just after the anomalous single year high catch record from the Japanese fleet, but this still resulted in unacceptable high recruitment and biomass fluctuations. More recent efforts have used a relatively new feature of MFCL that has not been implemented in a production assessment to date, the orthogonal-polynomial recruitment approach, to force zero recruitment deviations for the period prior to 1979. This looked promising in constraining the early recruitments without major implications for model estimation in the later periods. If there is a strong desire to include the early data, this is an option that warrants further investigation. While there may be other advantages to including this contrasting early data, ultimately, the estimation of the key management quantities and the resultant management advice is not influenced to any important extent by removing the early data prior to 1979.

The new growth parameters were the major change in the biological assumptions. The new growth estimates a faster early growth than that used in the previous assessments, with implications for earlier age at maturity and vulnerability to the fisheries compared to the previous assessments. The highest relative fishing mortality for the new growth is ages 2 and 3, whereas in the previous assessment instantaneous fishing mortality was more similar across ages 3-10 and much lower for age 2. The earlier vulnerability and increased fishing mortality on juvenile striped marlin was also likely linked to the high recruitment required to explain the very high early catches in the data prior to 1979. We did not provide an alternative set of external growth parameters, as these would have been counter to the best available information and not based on an observed data set for the southwestern Pacific Ocean. However, the recent depletion and spawning biomass were relatively insensitive to the addition of the new growth, although the recruitment scale was reduced. We acknowledge the very limited age-length samples to inform the external growth parameters and more sampling is important to better understand the possibility of spatial growth patterns and sex specific growth, which might provide alternative observed options for growth uncertainty (noting variable growth is not possible to model in MFCL).

Most of the other changes involved improvements to the data inputs, such as rectifying issues with size composition data bins, reweighting the input composition data with respect to catches and CPUE, and updating both the length-weight conversion factors and the maturity at length ogive based on the recommendations in Farley et al. (2021a). Unfortunately, there is negligible new biological data on striped marlin since the last assessment, despite the recommendations in that assessment to increase the level of biological sampling.

The application of the catch-conditioned approach has substantially reduced the number of parameters, and the new model is well converged with positive definite Hessian and is stable to jittering. While there are concerns with the quality of fits to some of size data, that further finessing might improve with more time, it would be unlikely that improving these fits would have marked effect on the management advice. Probably of more concern is the model under fit to the CPUE in the recent decade. In relation to this under fit of CPUE, we note there is conflict between the weight and length frequency data (preferring lower biomass) and the CPUE (preferring higher biomass),

with the overall likelihood being pushed towards a lower biomass scale. The size composition data, as also indicated by the Catch Curve and ASPM analyses, is having a strong influence on the model estimate of population size compared to CPUE, especially over the last decade. This may not be a desired outcome in assessments, where fitting the CPUE better would generally be preferable. Given the influence of size composition data it is worthwhile to consider a review of these inputs. While we applied the Francis iterative weighting as an improvement over the traditional arbitrary fixed divisors, it could be worth further exploration of alternative data weighting given the data conflicts.

We had some concerns on the recent size data (2021, 2022) from the Australian longline fishery (fisheries 06.LL.AU.2, 07.LL.AU.3), with a large proportion of lower weight fish, but few higher weight fish, and large sample numbers compared to other fisheries. The Piner plot constructed on the weight composition component of the likelihood profile indicated that these data were also in conflict with the weight data from the other fisheries. We looked at the impacts of removing these samples for 2022 and it appears they strongly influence the recent estimates of depletion, and spawning biomass, and the model estimates lower recent fishing mortality if these samples are removed. However, irrespective of their inclusion, while the terminal depletion might be slightly more optimistic, the management advice would not differ, and we have no current justification to remove or severely down weight these data. Nonetheless it will be important to seek some further clarity from the Australian delegation on the validity of their recent weight composition date.

In relation to the data weighting, we ran another exploration model where we applied the 2019 diagnostic model fixed divisor for size data weighting, this had a similar effect on spawning biomass and depletion as discussed above, producing a slightly more optimist trend in the recent depletion. While several key sensitivities are reported in this paper, various models (such as the above) were run to better understand the model behaviour. These are not reported in detail here but can be followed up with the authors. Ultimately, the model was not highly responsive to these various explorations and maintained the low recent depletion. We would have liked more time to thoroughly document these various additional models, and these could potentially be incorporated into some follow work, should the SC deem this necessary.

An important consideration that came up during our work on CPUE modelling was the surface associated habit for striped marlin. Since the mid 1990's, the longline fishery methods have developed to target tuna at greater depths. Striped marlin are mostly caught on hooks that are closer to the surface. Despite HBF being a catchability covariate in the CPUE analysis (Castillo-Jordán et al., 2024), and having the effect of adjusting the standardised CPUE upwards in the recent 10-20 years, it may not account sufficiently for the true reduction of effective hook effort for striped marlin in the tuna target longline fishery. We suggest the possibility of negative effort creep in the longline fisheries used for the CPUE indices (i.e. Japan and Chinese Taipei). Longline CPUE may be overestimating the decline in abundance in the case of striped marlin.

The uncertainty characterisation using the Monte Carlo approach was an improvement for this

assessment. Although there is more work required to improve approaches for generating the prior assumed distributions, this approach provides a more objective/self-weighting method for the uncertainty ensemble. Our original plan was to sample the average M values from the $A_{\max}$ prior then input these sampled values into the steepness (h) estimator (Brodziak et al., 2011). We would have then retained the pairings of the M used for the Lorenzen M-at-age with those used to generate the steepness values, which would have better dealt with any correlation. When we decided not to use the Brodziak et al. (2011) prior distribution for h, and create a beta distribution for h, the M and h priors had to be independent random draws from their respective distributions. This led to some combinations of M and h that the model struggled to reconcile with the other inputs, and thus failed the convergence criteria and were rejected. We suggest more work is necessary to develop standard SC accepted approaches to generating prior assumed distributions for common uncertainties such as M and h that can be applied consistently across assessments.

8.2 Main assessment conclusions

The 2024 assessment did not provide any clear indications of sustained improvement in stock status of striped marlin in the southwestern Pacific since the last assessment in 2019. On balance the assessment provided a slightly more pessimistic outcome. The various diagnostics indicate the model is reasonably well fit and converged to a stable solution, although we note the concerns on the CPUE fit and the strong influence of the size composition data. We are not concerned about the change to starting the model in 1979, the various sensitivities to the initial fishing mortality showed the model estimates and trajectories all quickly converged to similar stock status. The uncertainty ensemble presents a spread of stock status, but with the majority of models, and the median values of the key management reference points indicating the stock remains in a heavily depleted state and below the levels estimated to produce MSY.

Based on the ensemble of models, the general conclusions of this assessment are as follows:

- Consistent with the findings of the previous SWPO striped marlin assessments (Langley et al., 2006; Davies et al., 2012; Ducharme-Barth et al., 2019), the spawning biomass and stock status, as indicated by the spawning biomass depletion, show a steady decline from 1979 onwards, although the stock status has stabilised somewhat in the last 10 years, albeit at a low level.

- The stock status shows an increase in the terminal year of the 2024 assessment. Although terminal year estimates are often uncertain and subject to potential bias and subsequent revisions with future assessments, as is shown consistently in the retrospective analyses, a very recent increase in abundance is at least consistent with several CPUE abundance indices.

- The declining trend in recruitment over time, identified in the previous two stock assessments, remains a feature of the current model, along with reduced recruitment variability over time, all while the spawning stock biomass has decreased over time.

- Overall, the median depletion from the uncertainty grid for the recent period (2019-2022; $SB_{\mathrm{recent}}/SB_{F=0}$) is estimated at 0.14 (80 percentile range including estimation and structural uncertainty 0.09–0.21, full range 0.05–0.26).

- Fishing mortality has continued to increase over time for both the juvenile and adult components. With respect to maximum sustainable yield (MSY) based reference points (the WCPFC default for all billfish stocks, in lieu of the absence of any WCPFC billfish specific limit reference points), the median recent fishing mortality from the uncertainty ensemble is estimated to be above the level for achieving MSY (median $F_{\mathrm{recent}}/F_{\mathrm{MSY}} = 1.27$, range 0.58-3.02). 71% of the 59 ensemble models estimated $F_{\mathrm{recent}}/F_{\mathrm{MSY}} > 1$.

- The median recent spawning biomass from the uncertainty ensemble was well below the spawning biomass to achieve MSY (median $SB_{\mathrm{recent}}/SB_{\mathrm{MSY}} = 0.60$, range 0.18-1.24). 93% of the 59 ensemble models estimated $SB_{\mathrm{recent}}/SB_{\mathrm{MSY}} < 1$.

- Despite decreased catches since the early 2000s, based on the default MSY-based reference points, the 2024 assessment of striped marlin in the southwestern Pacific Ocean concludes that the stock is likely overfished, and undergoing overfishing.

- The outcomes of the 2024 stock assessment are broadly consistent with the previous 2019 stock assessment, and there are no signs of any sustained recent rebuilding of the striped marlin stock in the southwestern Pacific Ocean.

Concerns with the assessment in relation to management quantities:

- We have recognised some critical model issues late in the assessment process that would benefit from further work to improve the confidence in the management advice.

- These concerns include poor fits to some of the size data and the under fit to CPUE in the recent time period, data conflicts and data weighting impacts, and the results of the ASPM and Catch Curve models.

- Preliminary work to understand the implications of some of these issues suggest the current model estimates of stock status are likely biased towards more pessimistic management advice, the degree to which we cannot be certain without additional work.

- We also very recently identified some concerns around some data inputs that need further understanding.

- SC20 should consider the issues raised when evaluating whether the assessment results provide the best available science for management advice.

8.3   Recommendations for further research

- Increase the biological sampling of otoliths for building greater coverage of age-length data and application or otolith-based methods to explore population structure (i.e. otolith chemistry).

Increase the quantity of tissue samples collected for expanded genetic studies.

- The assessment should incorporate robust age validation procedures to enhance accuracy.

- Further research into growth parameters across different Pacific regions is essential to refine the assessment.

- Increased collection of representative size composition data could help improve the model outcomes.

- Explore the influence of environmental factors on recruitment. By exploring these connections, we can gain valuable insights into population dynamics and identify potential vulnerabilities to environmental changes.

- Review the length and weight composition data and explore further models with alternative data weighting to explore the range of possible stock status under greater influence of specific data such as CPUE.

- Consider the issue of effective hook effort changes (reductions) over time for surface dwelling billfish, and the possibility of negative effort creep due to modern tuna target sets deploying hooks at deeper depths than in the pre-1990s. There is likely a case for developing methods for adjusting hook numbers to better represent hook effort that is at the depth range where striped marlin are most vulnerable, and or estimating negative effort creep scenarios.

- Continue to develop the approaches for creating prior assumed distributions for natural mortality and steepness.

- Consider using a quarterly time step for the model, to improve the model resolution of the early life stages when growth is very fast.

- To ensure the most up-to-date methodologies are employed, consideration should be given to utilizing alternative modelling platforms that offer advanced capabilities for billfish stock assessments.

9   Acknowledgments

10 References

Anderson, S. C., Ward, E. J., English, P. A., and Barnett, L. A. K. (2022). sdmTMB: an R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields. *bioRxiv*, 2022.03.24.485545.

Boggs, C. (1989). Vital rate statistics for billfish stock assessment. In Stroud, H., editor, *Planning the Future of Billfishes: Research and Management in the 90s and Beyond. Part 1: Fishery and Stock Synopses, Data Needs and Management*, pages 225–233, Kailua-Kona, Hawaii, August 1-5, 1988.

Brill, R. W., Holts, D. B., Chang, R. K. C., Sullivan, S., Dewar, H., and Carey, F. G. (1993). Vertical and horizontal movements of striped marlin (Tetrapturus audax) near the Hawaiian Islands, determined by ultrasonic telemetry, with simultaneous measurement of oceanic currents. *Marine Biology*, 117(4):567–574.

Brodziak, J., Dai, X., and Katahira, L. (2012). Report of the Billfish Working Group Workshop. Technical Report ISC/12/BILLWG-1/REPORT.

Brodziak, J., Lee, H.-H., and Mangel, M. (2011). Probable values of stock-recruitment steepness for north pacific albacore tuna. Technical report, ISC/11/BILLWG-2/11 Probable.

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.

Cadigan, N. G. and Farrell, P. J. (2005). Local influence diagnostics for the retrospective problem in sequential population analysis. *ICES Journal of Marine Science*, 62(2):256–265.

Campbell, R., Davis, T., Edwards, B., Henry, G., Kalish, J., Lamason, B., Pepperell, J., and Ward, P. (2002). Assessment of black marlin and blue marlin in the Australian fishing zone: Report of the Black and Blue Marlin Working Group. Technical report, Department of Agriculture, Fisheries and Forestry, Canberra.

Carvalho, F., Punt, A. E., Chang, Y.-J., Maunder, M. N., and Piner, K. R. (2017). Can diagnostic tests help identify model misspecification in integrated stock assessments? *Fisheries Research*, 192:28–40.

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-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-SC20-2024/SA-IP-06, Manila, Philippines.

Castillo Jordán, C., Teears, T., Hampton, J., Davies, N., Scutt Phillips, J., McKechnie, S., Peatman, T., Macdonald, J., Day, J., Magnusson, A., Scott, R., Scott, F., Pilling, G., and Hamer, P. (2022). Stock assessment of skipjack tuna in the western and central Pacific Ocean: 2022. Technical Report WCPFC-SC18-SA-WP-01.

Chambers, M., Sippel, T., Domeier, M., and Holdsworth, J. (2013). The spatial distribution of striped marlin in the SW Pacific Ocean Estimates from PSAT tagging data. Technical Report WCPFC-SC9-2013/ SA-IP-09, Pohnpei, Federated States of Micronesia, 6-14 August 2013.

Davies, N., Fournier, D., Bouyé, F., and Hampton, J. (2022). Developments in the MULTIFAN-CL software 2021–2022. Technical Report WCPFC-SC18-2022/SA-IP-03.

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.

Day, J., Magnusson, A., Teears, T., Hampton, J., Davies, N., Castillo Jordán, C., Peatman, T., Scott, R., Scutt Phillips, J., McKechnie, S., Scott, F., Yao, N., Natadra, R., Pilling, G., Williams, P., and Hamer, P. (2023). Stock assessment of yellowfin tuna in the western and central Pacific Ocean: 2023. WCPFC-SC19-2023/SA-WP-05, Koror, Palau.

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.

Domeier, M. L., Dewar, H., and Nasby-Lucas, N. (2003). Mortality rate of striped marlin (Tetrapturus audax) caught with recreational tackle. *Marine and Freshwater Research*, 54(4):435.

Domeier, M. L., Ortega-Garcia, S., Nasby-Lucas, N., and Offield, P. (2019). First marlin archival tagging study suggests new direction for research. *Marine and Freshwater Research*, 70(4):603.

Ducharme-Barth, N., Castillo-Jordan, C., Hampton, J., Williams, P., Pilling, G., and Hamer, P. (2021). Stock assessment of swordfish in the southwest Pacific Ocean. Technical Report WCPFC-SC17-2021/SA-WP-04.

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.

Ducharme-Barth, N. and Vincent, M. (2021). Focusing on the front end: A framework for incorporating in uncertainty in biological parameters in model ensembles of integrated stock assessments. Technical Report WCPFC-SC17-2021-SA-WP-05.

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

Farley, J., Krusic-Golub, K., and Eveson, P. (2021b). Ageing of South Pacific Albacore: (Project 106). Technical Report WCPFC-SC17-SA-IP-10.

Fournier, D. and Archibald, C. P. (1982). A general-theory for analyzing catch at age data. *Canadian Journal of Fisheries and Aquatic Sciences*, 39(8):1195–1207.

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.

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A., and Sibert, J. (2012). AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, 27(2):233–249.

Francis, R. I. C. C. (1992). Use of risk analysis to assess fishery management strategies: A case study using orange roughy (*Hoplostethus atlanticus*) on the Chatham Rise, New Zealand. *Canadian Journal of Fisheries and Aquatic Science*, 49:922–930.

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

Ghosn, D., Collins, D., Baiada, C., and Steffe, A. (2012). Catch per unit effort and size composition of striped marlin caught by recreational fisheries in southeast Australian waters. Technical Report WCPFC-SC8-SA-IP-07, NSW Department of Primary Industries.

Graves, J. E. and McDowell, J. R. (1994). Genetic Analysis of Striped Marlin (Tetrapturus audax) Population Structure in the Pacific Ocean. *Canadian Journal of Fisheries and Aquatic Sciences*, 51(8):1762–1768.

Graves, J. E. and McDowell, J. R. (2003). Stock structure of the world's istiophorid billfishes: a genetic perspective. *Marine and Freshwater Research*, 54(4):287.

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

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

Hamer, P. (2024). Report from the SPC Pre-assessment Workshop – March 2024. WCPFC Scientific Committee paper WCPFC-SC20-2024/SA-IP-01, Manila, Philippines.

Hampton, J. and Fournier, D. (2001). A spatially-disaggregated, length-based, age-structured population model of yellowfin tuna (*Thunnus albacares*) in the western and central Pacific Ocean. *Marine and Freshwater Research*, 52:937–963.

Hanamoto, E. (1977). Fishery Oceanography of Striped Marlin-II. *Nippon Suisan Gakksishi*, 43(11):1279–1286.

Harley, S. J. (2011). A preliminary investigation of steepness in tunas based on stock assessment results. Technical Report WCPFC-SC7-2011/SA-IP-08, Pohnpei, Federated States of Micronesia, 9–17 August 2011.

Hinton, M. and Bayliff, W. (2002). Status of striped marlin in the Eastern Tropical Pacific Ocean in 2001 and outlook for 2002. Honolulu, Hawaii, United States of America, 22-27 July 2002.

Holdsworth, J. (2024). Striped marlin catch and CPUE in the New Zealand sport fishery, 2019-20 to 2021-22. Technical Report WCPFC-SC20-2024/SA-IP-17, Manill a, Phillippines, 14–21 August 2024.

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.

Holdsworth, J. C., Sippel, T. J., and Block, B. A. (2009). Near real time satellite tracking of striped marlin (Kajikia audax) movements in the Pacific Ocean. *Marine Biology*, 156(3):505–514.

Hoyle, S. D., Williams, A. J., Minte-Vera, C. V., and Maunder, M. N. (2023). Approaches for estimating natural mortality in tuna stock assessments: Application to global yellowfin tuna stocks. *Fisheries Research*, 257:106498.

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.

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, Palau.

Kleiber, P., Fournier, D., Hampton, J., Davies, N., Bouyé, F., and Hoyle, S. (2019). MULTIFAN-CL User's Guide. Technical report.

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. and Davie, P. (2009). *Population biology and habitat preferences of striped marlin, Kajikia audax in the southwest Pacific Ocean.* PhD, Charles Sturt University.

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., Bromhead, D., and Pepperell, J. G. (2011). Age and growth of striped marlin (Kajikia audax) in the Southwest Pacific Ocean. *ICES Journal of Marine Science*, 68(9):1884–1895.

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.

Lam, C., Tam, C., and Lutcavage, M. (2019). Preliminary results on popup satellite archival tagging of striped marlin in the Central North Pacific. Technical Report ISC19-BILLWG2-WP1.

Lam, C. H., Kiefer, D. A., and Domeier, M. L. (2015). Habitat characterization for striped marlin in the Pacific Ocean. *Fisheries Research*, 166:80–91.

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.

Lorenzen, K. (1996). The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystem and aquaculture. *Journal of Fish Biology*, 42:627–647.

Macdonald, J., Bell, L., Schneiter, E., Ghergariu, M., Sanchez, C., Contreras, R., Judd, P., Hosken, M., Potts, J., Day, J., Vidal, T., and Williams, P. (2024). Project 90: Better data on fish weights and lengths for scientific analysis. WCPFC Scientific Committee paper WCPFC-SC20-2024/ST-IP-04, Manila, Philippines.

Magnusson, A., Day, J., Teears, T., Hampton, J., Davies, N., Castillo Jordán, C., Peatman, T., Scott, R., Scutt Phillips, J., McKechnie, S., Scott, F., Yao, N., Natadra, R., Pilling, G., Williams, P., and Hamer, P. (2023). Stock assessment of yellowfin tuna in the western and central Pacific Ocean: 2023. WCPFC-SC19-2023/SA-WP-04, Koror, Palau.

Mamoozadeh, N., McDowell, J., and Graves, J. (2018). Genetic population structure of striped marlin (Kajikia audax) in the Indian Ocean, with relationship to Pacific Ocean populations. Technical Report IOTC-2018-WPB16-20, Cape Town, South Africa, September 4-8, 2018.

Maunder, M. N. and Piner, K. R. (2015). Contemporary fisheries stock assessment: many issues still remain. *ICES Journal of Marine Science*, 72(1):7–18.

McKechnie, S. (2014). Analysis of longline size frequency data for bigeye and yellowfin tunas in the WCPO. Technical Report WCPFC-SC10-2014/SA-IP-04, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

McKechnie, S., Pilling, G., and Hampton, J. (2017). Stock assessment of bigeye tuna in the western and central Pacific Ocean. Technical Report WCPFC-SC13-2017/SA-WP-05, Rarotonga, Cook Islands, 9–17 August 2017.

Melo-Barrera, F., Felix-Uraqa, R., and Quinonez-Velazquez, C. (2003). Growth and length-weight relationship of the striped marlin, Tetrapturus audax (Pisces: Istiophoridae), in Cabo San Lucas, Baja California Sur, Mexico. *Ciencias Marinas*, 29(3):305–313.

Minte-Vera, C. V., Maunder, M. N., Aires-da Silva, A. M., Satoh, K., and Uosaki, K. (2017). Get the biology right, or use size-composition data at your own risk. *Fisheries research*, 192:114–125.

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.

Nakamura, I. (1985). FAO species catalogue. Volume 5. Billfishes of the World. An annotated and illustrated catalogue of marlins, sailfishes, spearfishes and swordfishes known to date. Technical Report Volume 5. FAO Fisheries Synopsis. No. 125 (5).

Neubauer, P., Kim, K., A'mar, T., and Large, K. (2023). Addressing uncertainty in wcpfc stock assessments: Review and recommendations from WCPFC project 113. Technical Report WCPFC/SC19-SA-WP-10, Koror, Palau, 16–24 August 2023.

Pauly, D. (1980). On the interrelationships between natural mortality, growth-parameters, and mean environmental-temperature in 175 fish stocks. *Journal Du Conseil*, 39(2):175–192.

Pepperell, J. G. and Davis, T. L. O. (1999). Post-release behaviour of black marlin, Makaira indica , caught off the Great Barrier Reef with sportfishing gear. *Marine Biology*, 135(2):369–380.

Piner, K. and Lee, H. (2011). Meta-analysis of striped marlin natural mortality. Technical Report ISC/11/BILLWG-1/10.

Potts, J., Castillo-Jordan, C., Day, J., Hamer, P., and Teears, T. (2024). Analysis of longline size frequency data for the 2024 south pacific albacore and wcpo striped marlin assessments. Technical Report WCPFC-SC20-2024/SA-IP-03, Manila, Phillippines, 14–21 August 2024.

Sippel, T., Holdsworth, J., Dennis, T., and Montgomery, J. (2011). Investigating Behaviour and Population Dynamics of Striped Marlin (Kajikia audax) from the Southwest Pacific Ocean with Satellite Tags. *PLoS ONE*, 6(6):e21087.

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.

Skillman, R. and Yong, M. (1976). Von Bertalanffy growth curves for striped marlin, Tetrapturus audax, and blue marlin, Makaira nigricans, in the central north Pacific Ocean. *Fishery Bulletin*, 74(3):553–566.

Stewart, I. J. and Martell, S. J. (2014). A historical review of selectivity approaches and retrospective patterns in the Pacific halibut stock assessment. *Fisheries Research*, 158:40–49.

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.

Teears, T., Castillo Jordán, C., Davies, N., Day, J., Hampton, J., Magnusson, A., Peatman, T., Pilling, G., Xu, H., Vidal, T., Williams, P., and Hamer, P. (2024). Stock assessment of South Pacific albacore: 2024. WCPFC-SC20-2024/SA-WP-01, Manila, Philippines, 14–21 August 2024.

Teears, T., Peatman, T., Ducharme-Barth, N., Williams, P., Scutt Phillips, J., Magnusson, A., Day, J., Hampton, J., McKechnie, S., and Hamer, P. (2023). CPUE analysis and data inputs for the 2023 bigeye and yellowfin tuna assessments in the WCPO. WCPFC Scientific Committee paper SC19-SA-WP-03, Palau.

Thorson, J. T., Shelton, A. O., Ward, E. J., and Skaug, H. J. (2015). Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. *ICES Journal of Marine Science*, 72(5):1297–1310.

Tremblay-Boyer, L., Hampton, J., McKechnie, S., and Pilling, G. (2018). Stock assessment of South Pacific albacore tuna. Technical Report WCPFC-SC-14-2018/SA-WP-05, Busan, South Korea, 8-16 August 2018.

Tremblay-Boyer, L., McKechnie, S., Pilling, G., and Hampton, J. (2017). Stock assessment of yellowfin tuna in the Western and Central Pacific Ocean. Technical Report WCPFC-SC13-2017/SA-WP-06, Rarotonga, Cook Islands, 9-17 August 2017.

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 Working Paper presented to the 38th meeting of the Tropical Tuna Resource Assessment Group held 11-13 July 2023, Mooloolaba, CSIRO, Hobart.

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.

Williams, P. and Smith, N. (2018). Requirements for enhancing conversion factor information. Technical Report WCPFC-SC14-2018/ST-WP-05, Busan, South Korea, 8-16 August 2018.

11   Tables

Table 1: Definition of fisheries for the striped marlin stock assessment.

Fishery	Nationality	Gear	Region	Catch	Years
01.LL.JP.1	JP	Longline	1	Number	1979-2022
02.LL.JP.2	JP	Longline	2	Number	1979-2022
03.LL.JP.3	JP	Longline	3	Number	1979-2022
04.LL.JP.4	JP	Longline	4	Number	1979-2022
05.LL.TW.4	TW	Longline	4	Number	1979-2022
06.LL.AU.2	AU	Longline	2	Number	1987-2022
07.LL.AU.3	AU	Longline	3	Number	1987-2022
08.LL.NZ.3	NZ	Longline	3	Number	1989-2022
09.REC.AU.3	AU	Recreational	3	Number	1979-2022
10.REC.NZ.3	NZ	Recreational	3	Number	1979-2022
11.LL.ALL.1	DWFN/PICT	Longline	1	Number	1979-2022
12.LL.ALL.2	DWFN/PICT	Longline	2	Number	1979-2022
13.LL.ALL.3	DWFN/PICT	Longline	3	Number	1979-2022
14.LL.ALL.4	DWFN/PICT	Longline	4	Number	1979-2022
15.INDEX.1-4	JP & TW	Longline	1–4	-	1979-2022

Table 2: Total number of striped marlin length- and weight-frequency samples for each of the defined fisheries in the 2024 diagnostic model.

Fishery	Length.Frequency	Weight.frequency
01.LL.JP.1	242	454
02.LL.JP.2	658	13,504
03.LL.JP.3	438	6,047
04.LL.JP.4	0	70
05.LL.TW.4	280	0
06.LL.AU.2	1,500	59,105
07.LL.AU.3	1,396	17,159
08.LL.NZ.3	21	0
09.REC.AU.3	0	2.044
10.REC.NZ.3	1,165	15,632
11.LL.ALL.1	12,022	0
12.LL.ALL.2	21,021	0
13.LL.ALL.3	44	0
14.LL.ALL.4	1,414	0

Table 3: Summary of the groupings of fisheries used within the diagnostic model for estimation of selectivity. See Table 1 for further details on each fishery.

Fishery	Sub.region	Sel.	Sel.Term.Age	Non.Decrease.Pen
01.LL.JP.1	1	1	6	N
02.LL.JP.2	2	2	7	N
03.LL.JP.3	3	3	7	N
04.LL.JP.4	4	4	6	N
05.LL.TW.4	4	4	6	N
06.LL.AU.2	2	5	7	N
07.LL.AU.3	3	6	7	N
08.LL.NZ.3	3	6	7	N
09.REC.AU.3	3	7	7	N
10.REC.NZ.3	3	8	9	Y
11.LL.ALL.1	1	1	6	N
12.LL.ALL.2	2	9	7	N
13.LL.ALL.3	3	3	7	N
14.LL.ALL.4	4	4	6	N

Table 4: Description of symbols used in the yield and stock status analyses. For the purpose of this assessment, "recent" for F is the average over the period 2018–2021 and for SB is the average over the period 2019–2022 and "latest" is 2022.

Symbol	Description
C_{latest}	Catch in the last year of the assessment (2022)
F_{recent}	Average fishing mortality-at-age for a recent period (2018–2021)
$Y_{Frecent}$	Equilibrium yield at average fishing mortality for a recent period (2018–2021)
f_{mult}	Fishing mortality multiplier at maximum sustainable yield (MSY)
F_{MSY}	Fishing mortality-at-age producing the maximum sustainable yield (MSY)
MSY	Equilibrium yield at F_{MSY}
$F_{\text{recent}}/F_{\text{MSY}}$	Average fishing mortality-at-age for a recent period (2018–2021) relative to F_{MSY}
SB_0	Equilibrium unexploited spawning potential
SB_{latest}	Spawning biomass in the latest time period (2022)
SB_{recent}	Spawning biomass for a recent period (2019–2022)
$SB_{F=0}$	Average spawning biomass predicted in the absence of fishing for the period 2012–2021
SB_{MSY}	Spawning biomass that will produce the maximum sustainable yield (MSY)
$SB_{\text{MSY}}/SB_{F=0}$	Spawning biomass that produces maximum sustainable yield (MSY) relative to the average spawning biomass predicted to occur in the absence of fishing for the period 2012–2021
SB_{MSY}/SB_0	Spawning biomass that produces the maximum sustainable yield (MSY) relative to the equilibrium unexploited spawning potential
SB_{latest}/SB_0	Spawning biomass in the latest time period (2022) relative to the equilibrium unexploited spawning potential
$SB_{\text{latest}}/SB_{F=0}$	Spawning biomass in the latest time period (2022) relative to the average spawning biomass predicted to occur in the absence of fishing for the period 2012–2021
$SB_{\text{latest}}/SB_{\text{MSY}}$	Spawning biomass in the latest time period (2022) relative to that which will produce the maximum sustainable yield (MSY)
$SB_{\text{recent}}/SB_{F=0}$	Spawning biomass for a recent period (2019–2022) relative to the average spawning biomass predicted to occur in the absence of fishing for the period 2012–2021
$SB_{\text{recent}}/SB_{\text{MSY}}$	Spawning biomass for a recent period (2019–2022) relative to the spawning biomass that produces maximum sustainable yield (MSY)

Table 5: Summary of reference points over the 59 individual models in the uncertainty model ensemble, along with results incorporating estimation uncertainty. Note that these values do not include estimation uncertainty, except for the key management quantities (bottom three rows of the table).

Metric	Mean	Median	Min	10%ile	90%ile	Max
C_{latest}	1241	1241	1234	1237	1241	1248
F_{MSY}	0.34	0.35	0.21	0.25	0.43	0.45
f_{mult}	0.82	0.78	0.34	0.45	1.26	1.63
$F_{\text{recent}}/F_{\text{MSY}}$	1.43	1.28	0.61	0.79	2.24	2.95
MSY	1789	1760	1698	1708	1907	2067
SB_0	10145	9457	7250	7850	13153	18950
$SB_{F=0}$	8527	7725	5517	6121	11743	18384
SB_{latest}/SB_0	0.12	0.12	0.06	0.08	0.17	0.19
$SB_{\text{latest}}/SB_{F=0}$	0.15	0.15	0.06	0.09	0.21	0.25
$SB_{\text{latest}}/SB_{\text{MSY}}$	0.63	0.62	0.19	0.31	0.98	1.19
SB_{MSY}	2224	1897	1138	1283	3574	5599
SB_{MSY}/SB_0	0.21	0.20	0.14	0.16	0.26	0.30
$SB_{\text{MSY}}/SB_{F=0}$	0.25	0.25	0.18	0.21	0.30	0.34
$SB_{\text{recent}}/SB_{F=0}$	0.14	0.14	0.06	0.09	0.20	0.23
$SB_{\text{recent}}/SB_{\text{MSY}}$	0.60	0.59	0.18	0.30	0.91	1.10
$Y_{Frecent}$	1563	1704	137	1079	1861	1931
Including estimation uncertainty						
	Mean	Median	Min	10%ile	90%ile	Max
$SB_{\text{recent}}/SB_{F=0}$	0.14	0.14	0.05	0.09	0.20	0.26
$F_{\text{recent}}/F_{\text{MSY}}$	1.42	1.27	0.58	0.74	2.30	3.02
$SB_{\text{recent}}/SB_{\text{MSY}}$	0.60	0.60	0.18	0.29	0.95	1.24

12   Figures
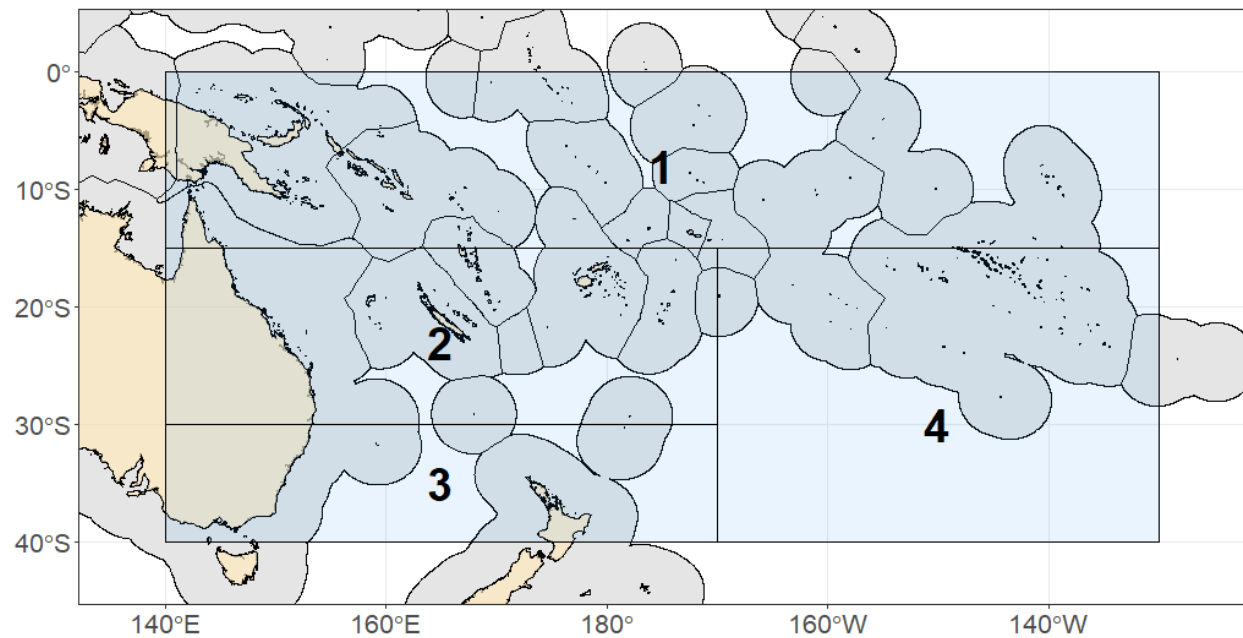
Figure 1: The geographical area covered by the stock assessment and the boundaries of the four model sub-regions used for 2024 SWPO striped marlin assessment.
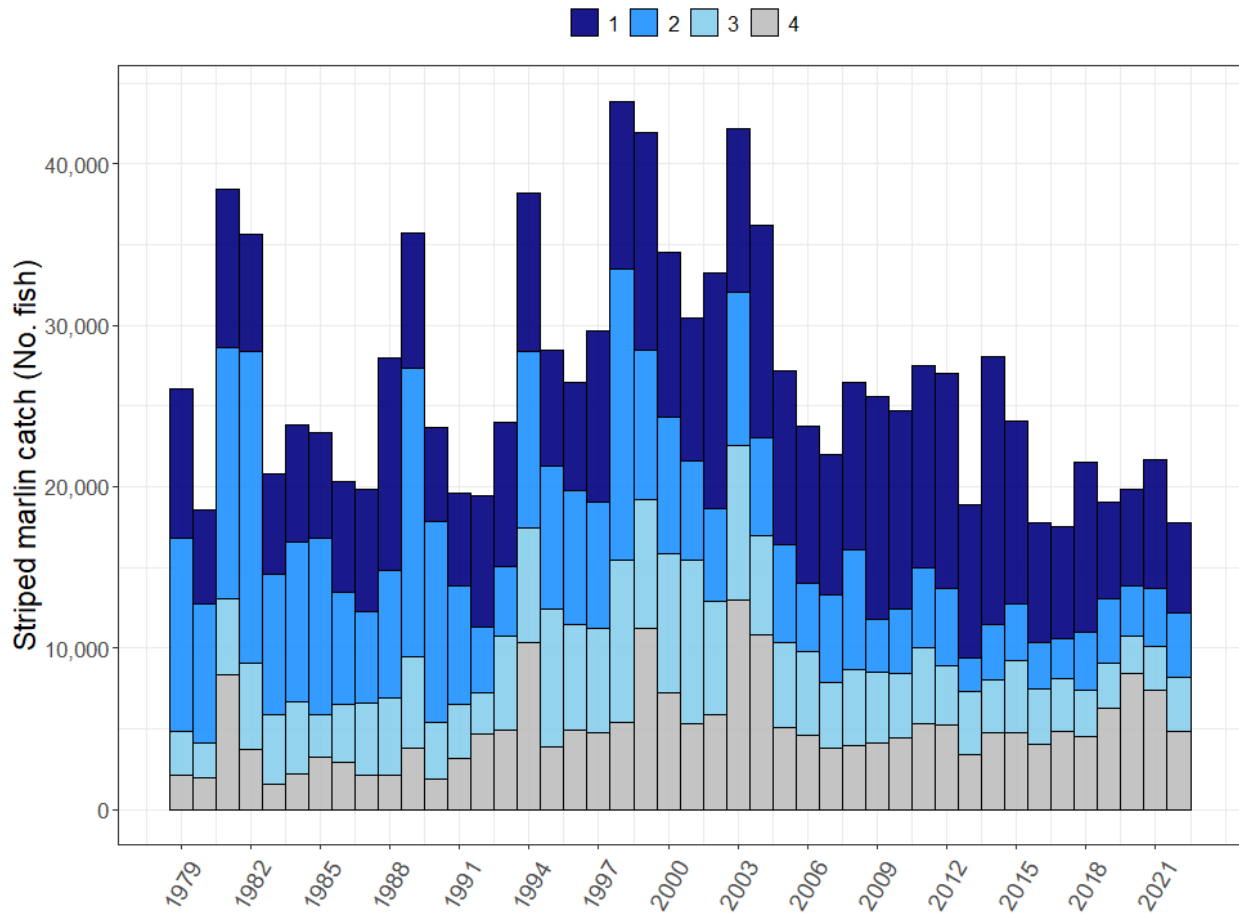
Figure 2: Annual catch (numbers of fish) of striped marlin by sub-region in the area covered by the assessment.
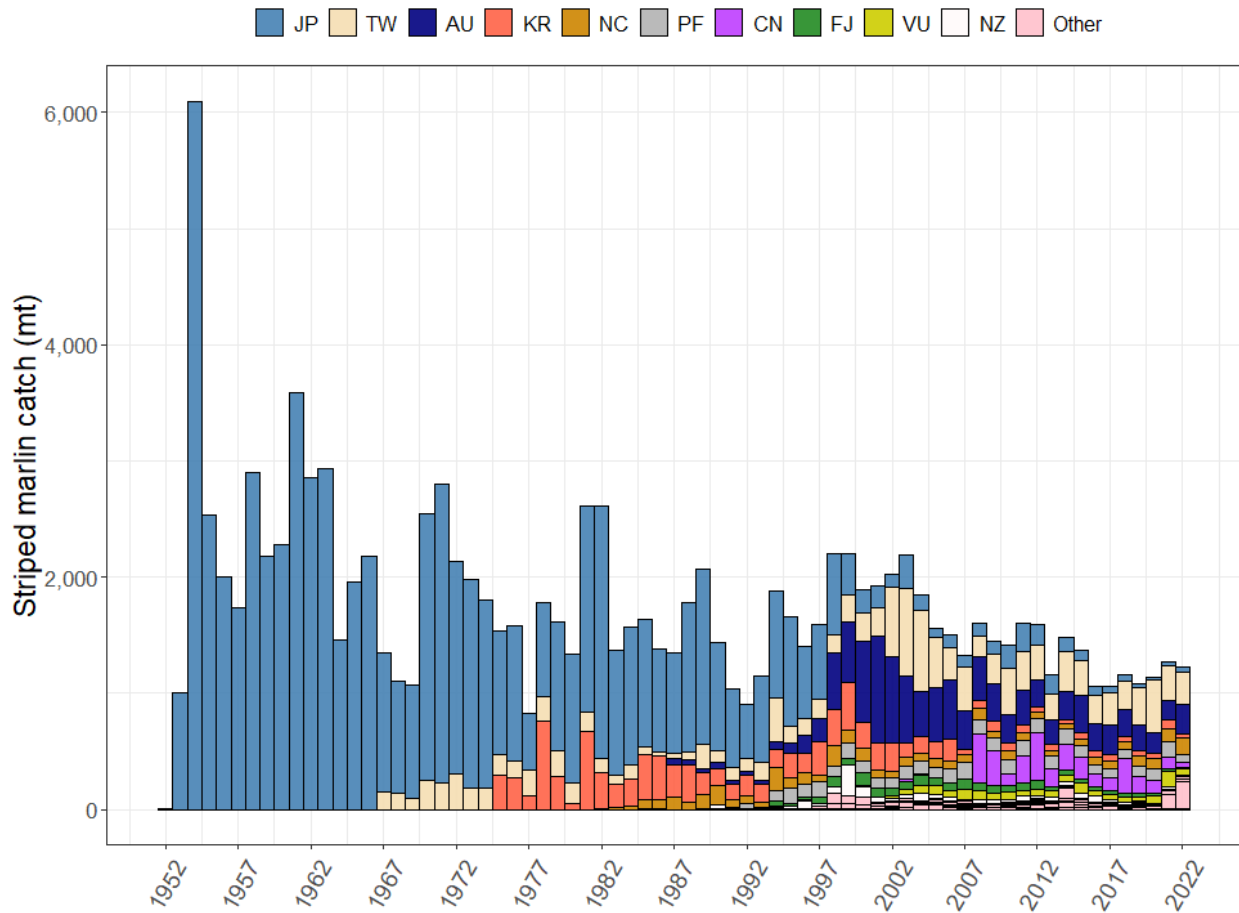
Figure 3: Annual catch (mt) of striped marlin by flag starting in 1952 in the SWPO area covered by the assessment.
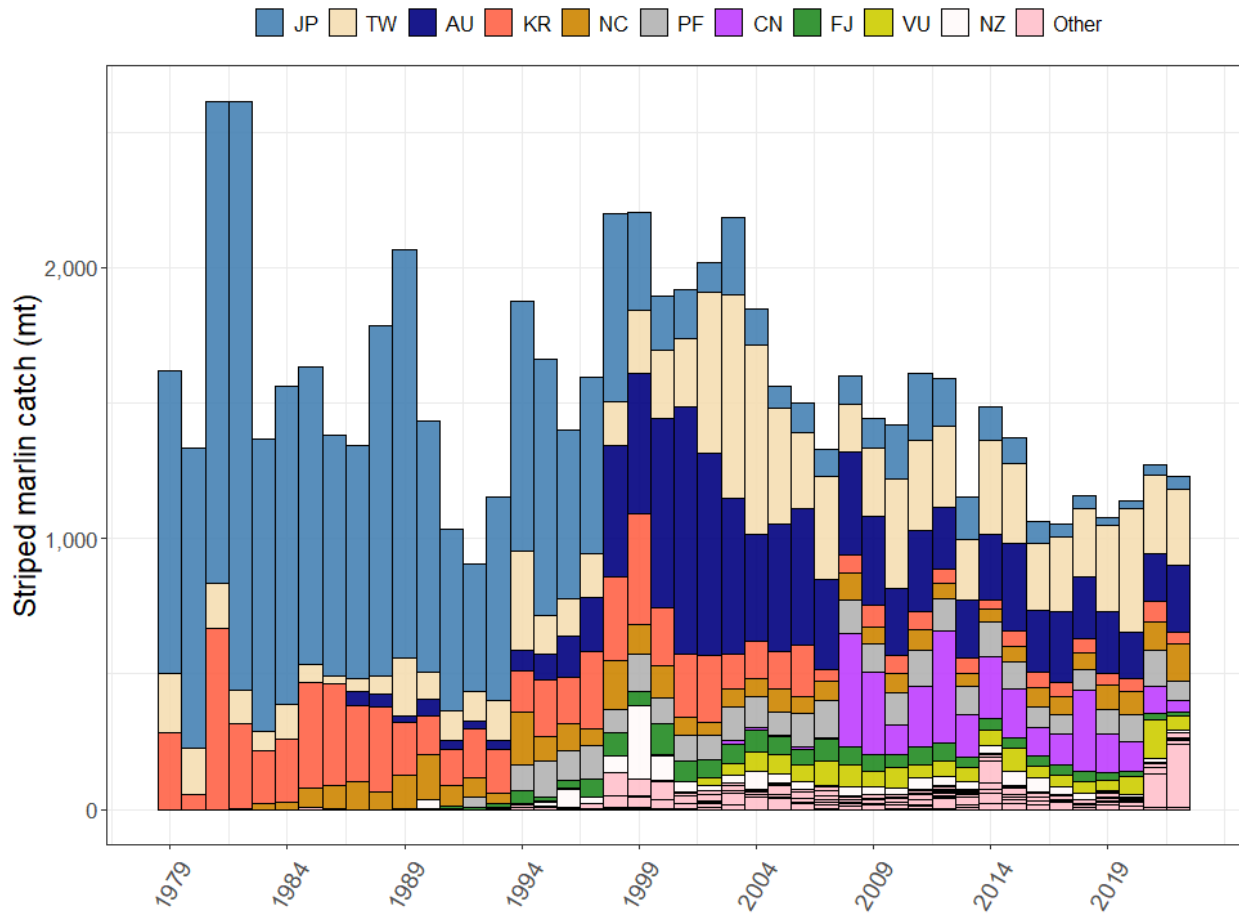
Figure 4: Annual catch (mt) of striped marlin by flag starting in 1979 in the SWPO area covered by the assessment.
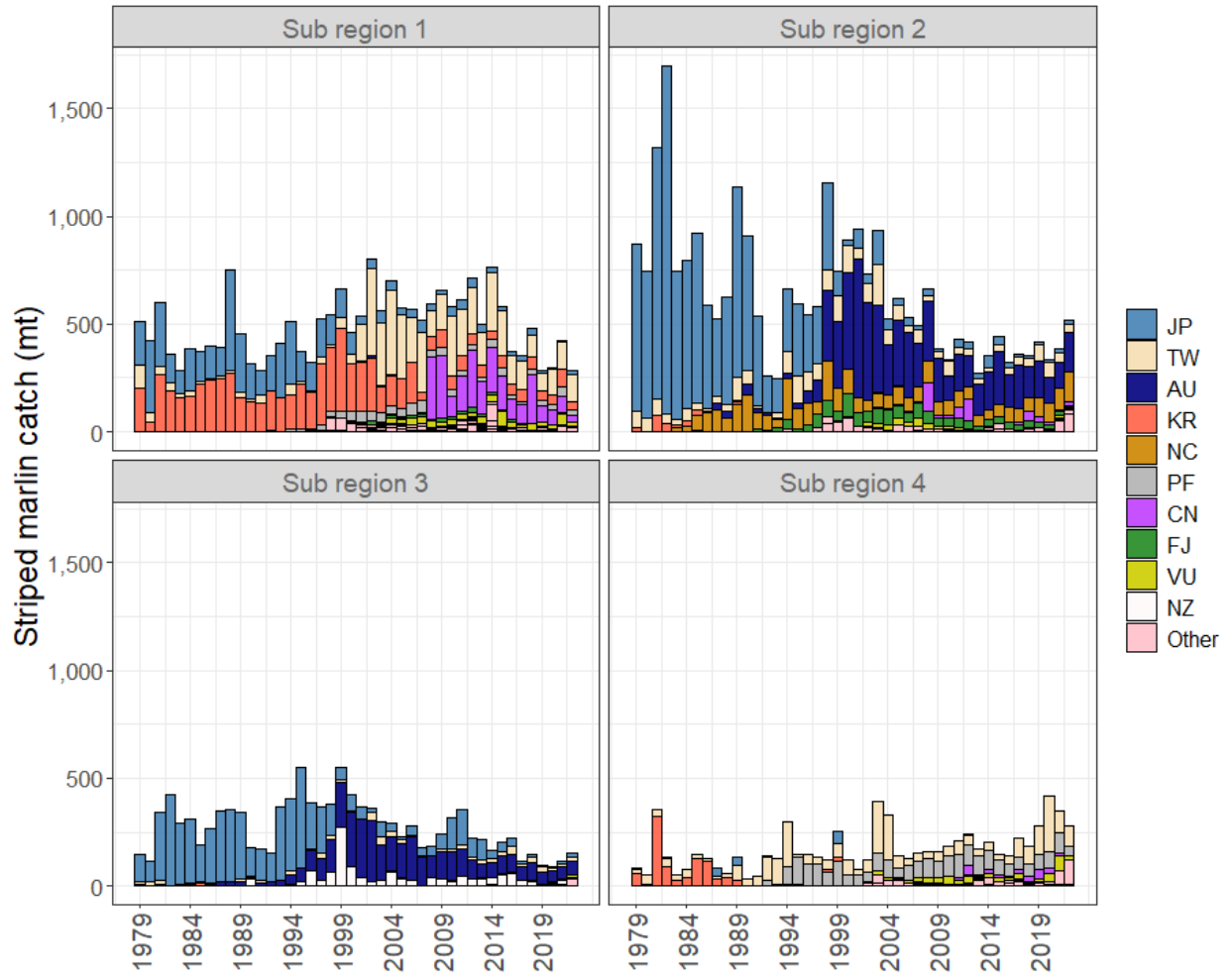
Figure 5: Annual catch (mt) of striped marlin by flag for each of the four model sub-regions.
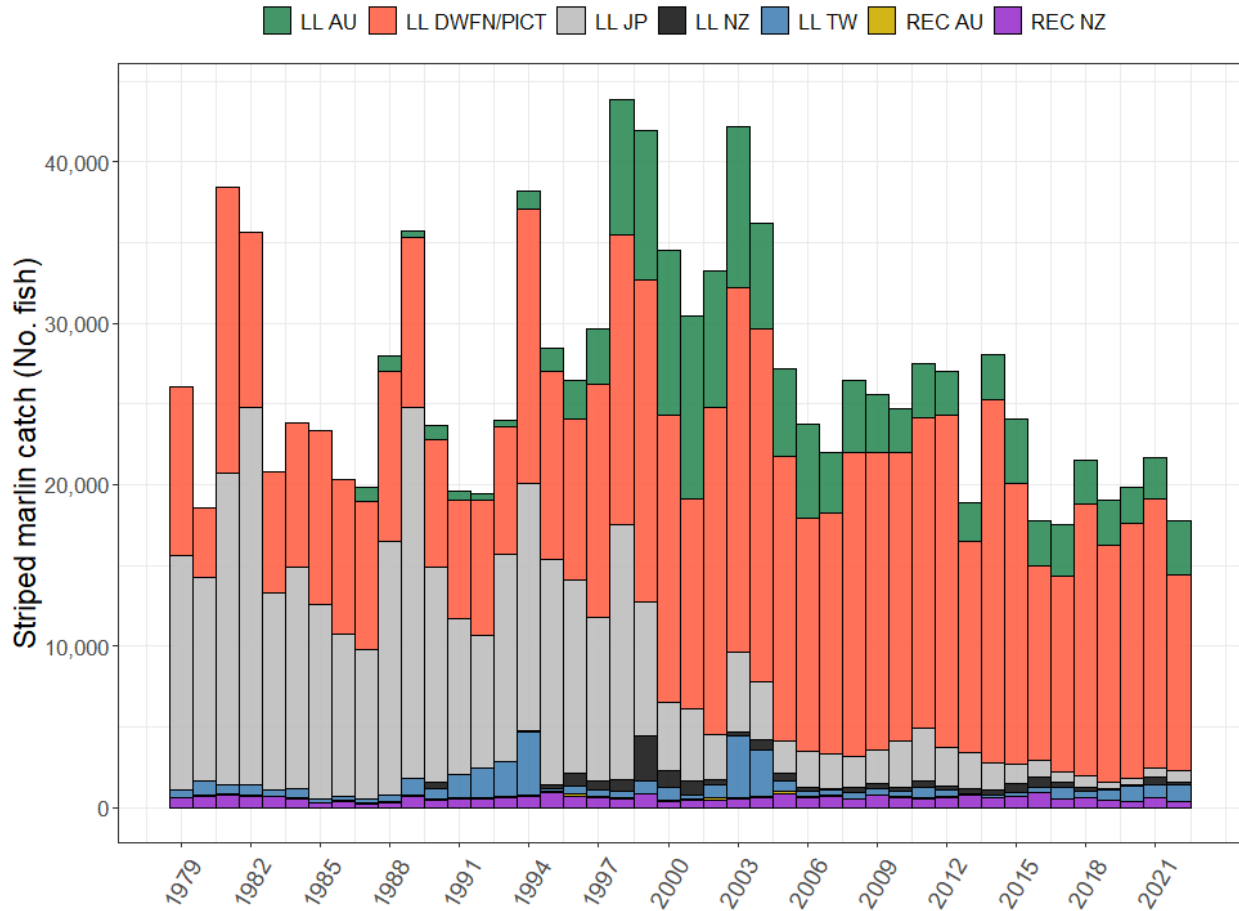
Figure 6: Annual catch (numbers of fish) of striped marlin by gear type and flag in the SWPO area covered by the assessment.
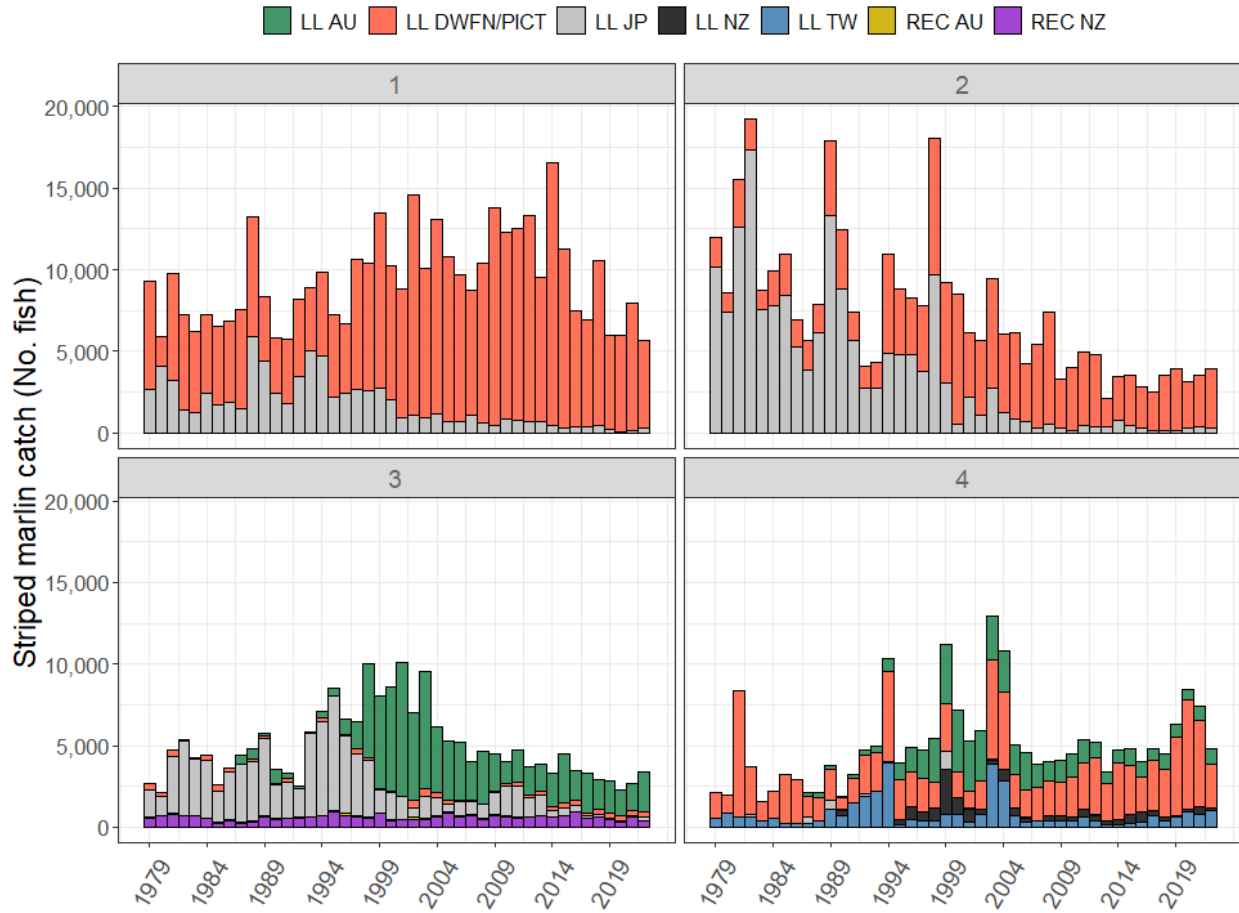
Figure 7: Annual catch (numbers of fish) of striped marlin by gear type and flag for each of the four model sub-regions.
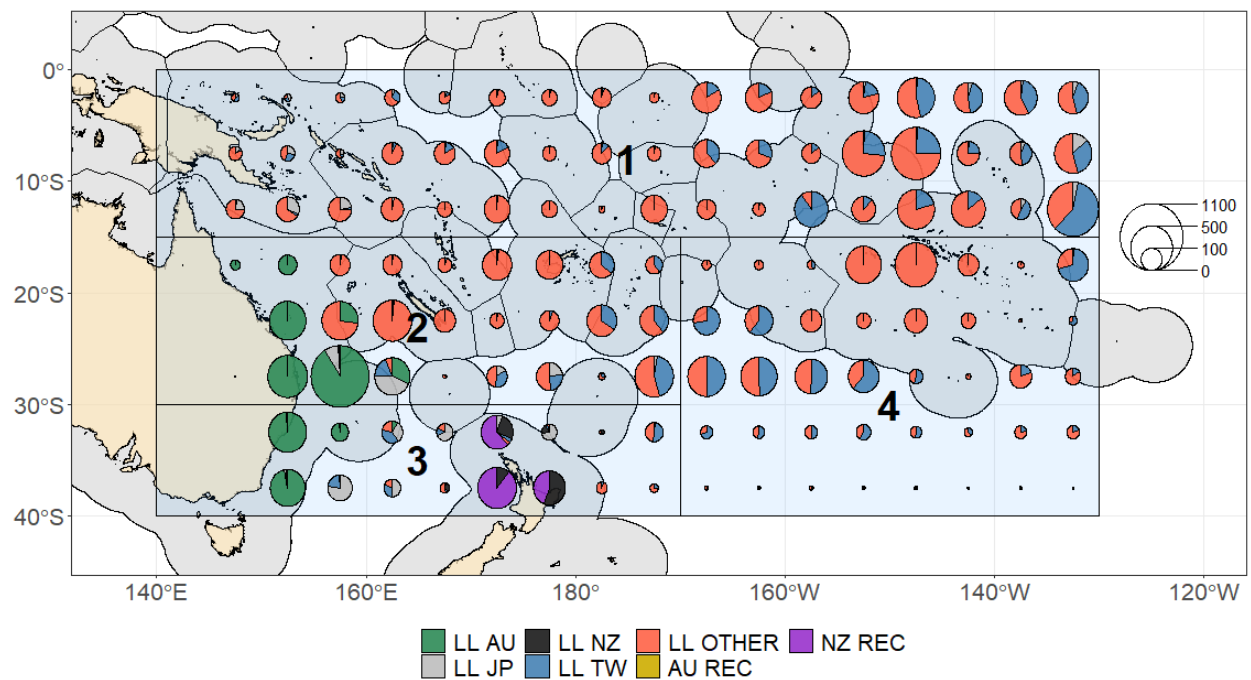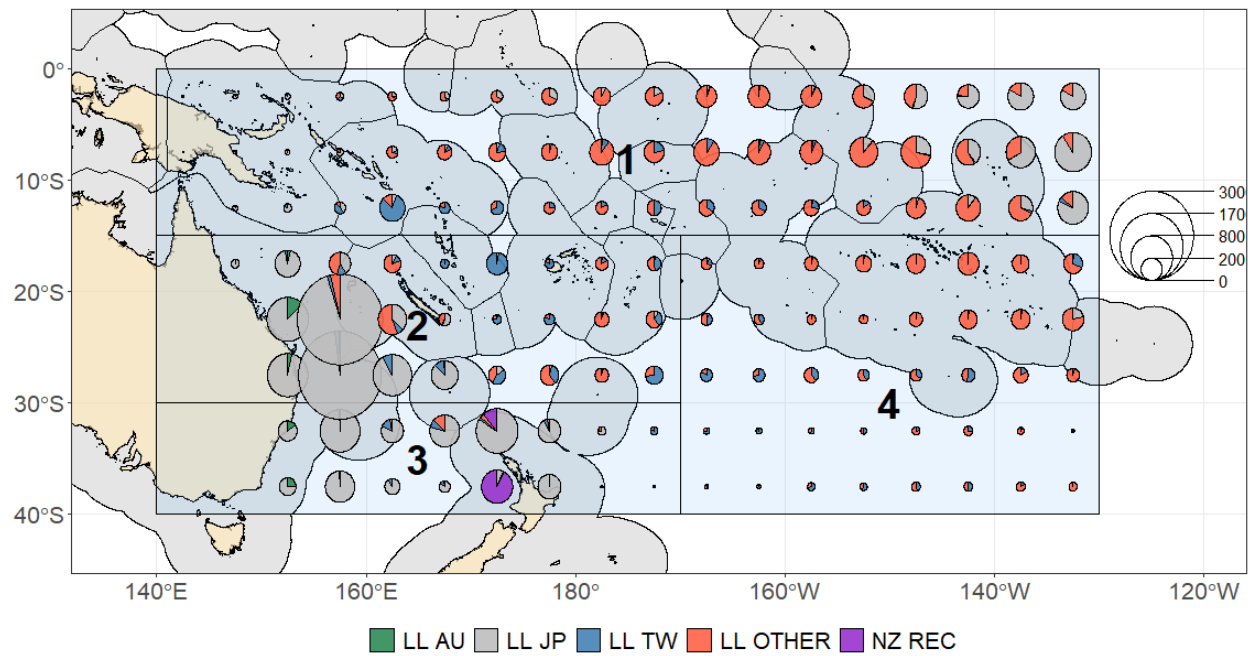
Figure 8: Distribution and magnitude of striped marlin catch (mt) by gear type summed over 10 year periods from 1979-1988 (upper plot) and from 2013-2022 (lower plot) for $5° \times 5°$ cells.
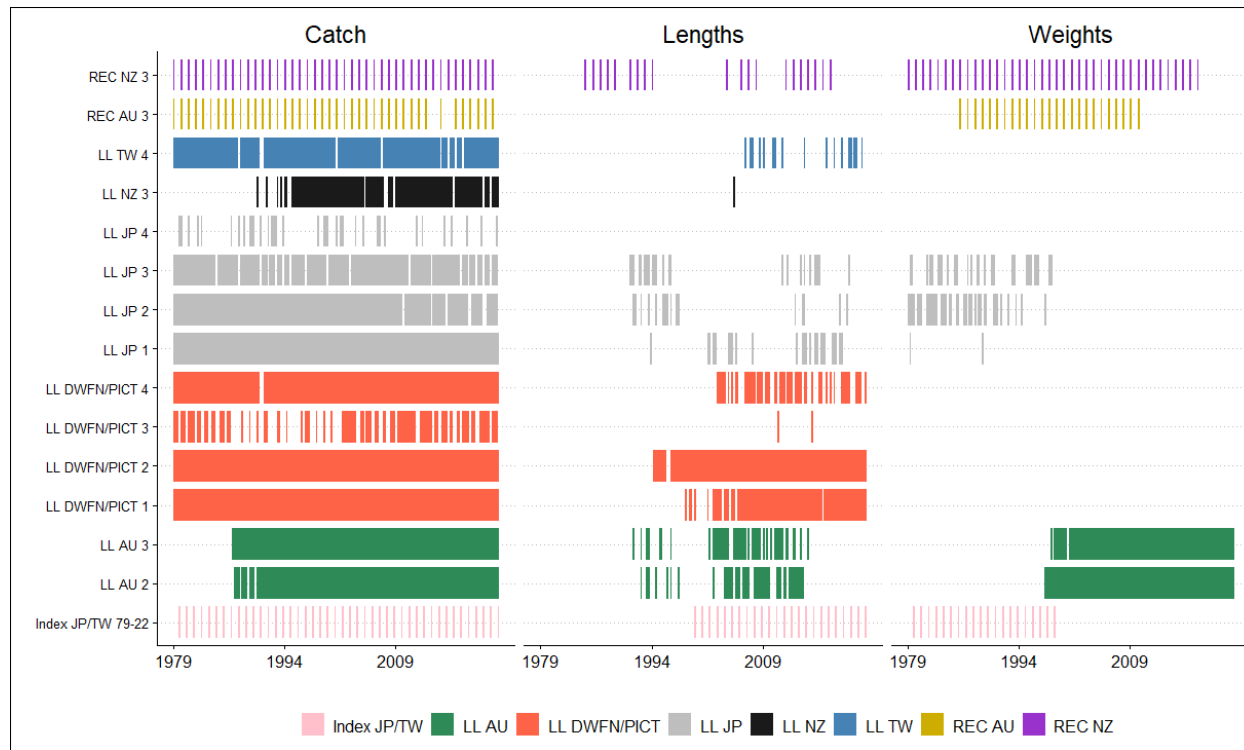
Figure 9: Summary of data coverage by fishery for the SWPO 2024 striped marlin assessment. LL=longline, REC=recreational gears.
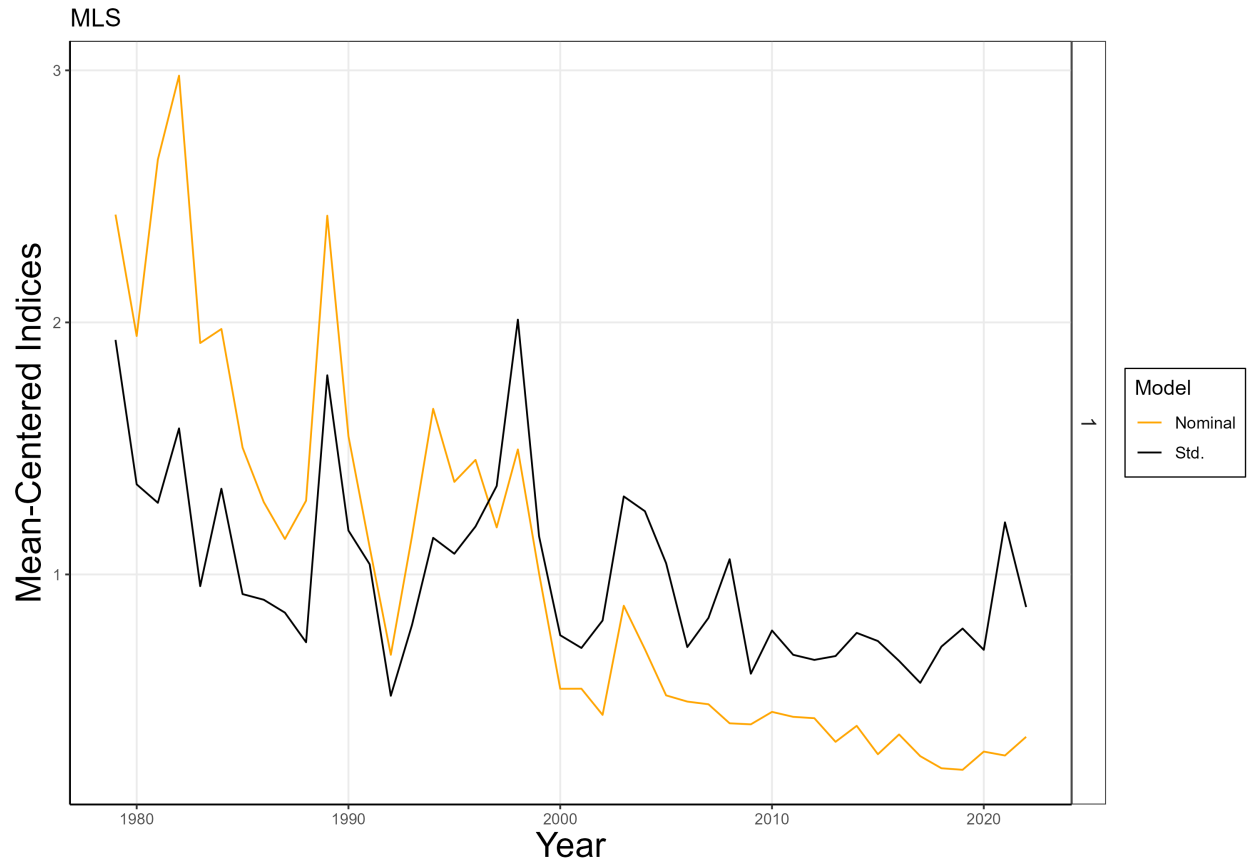
Figure 10: Nominal (orange line) and standardised (black line) CPUE for the combined longline index fisheries from Japan and Chinese Taipei in model sub-regions 1-4.
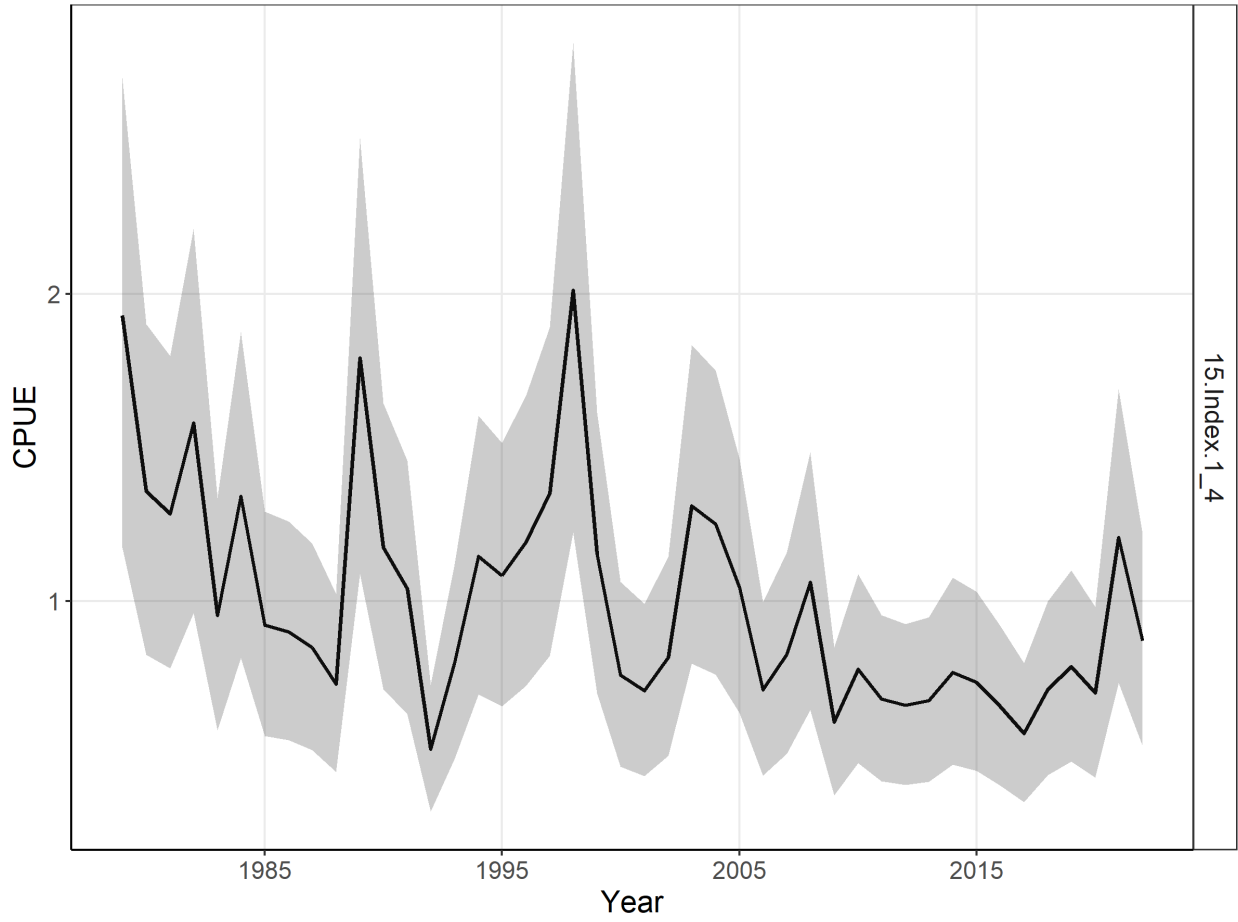
Figure 11: Standardised CPUE (black line) for the combined longline index fisheries from Japan and Chinese Taipei in model sub-regions 1-4, with 95% CI (grey band).
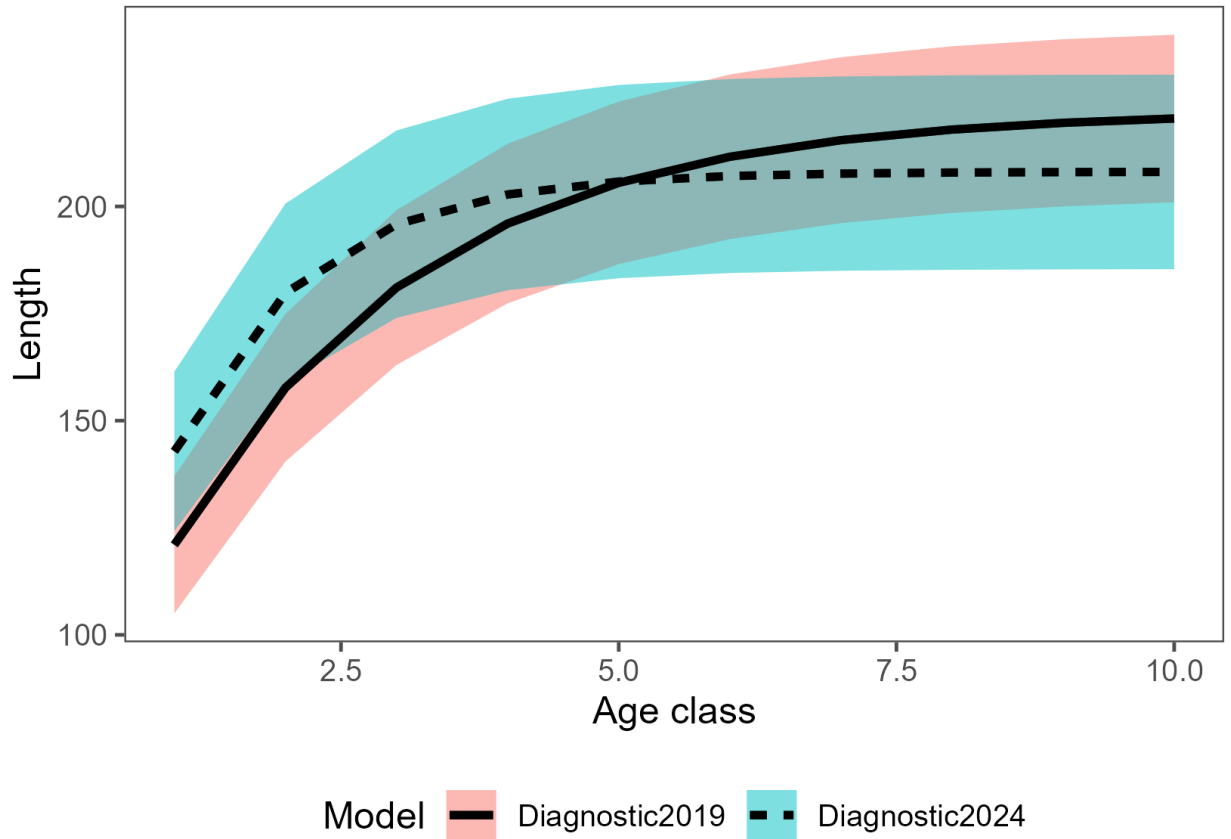
Figure 12: Growth comparison for the 2019 and 2024 diagnostic model. The growth used in the diagnostic model 2019 (Kopf et al., 2011) is shown in red and the growth function used for the diagnostic model 2024 developed by CSIRO using otolith aging is shown in green. The estimated uncertainty in growth shown as ± 1 standard deviations from the mean.
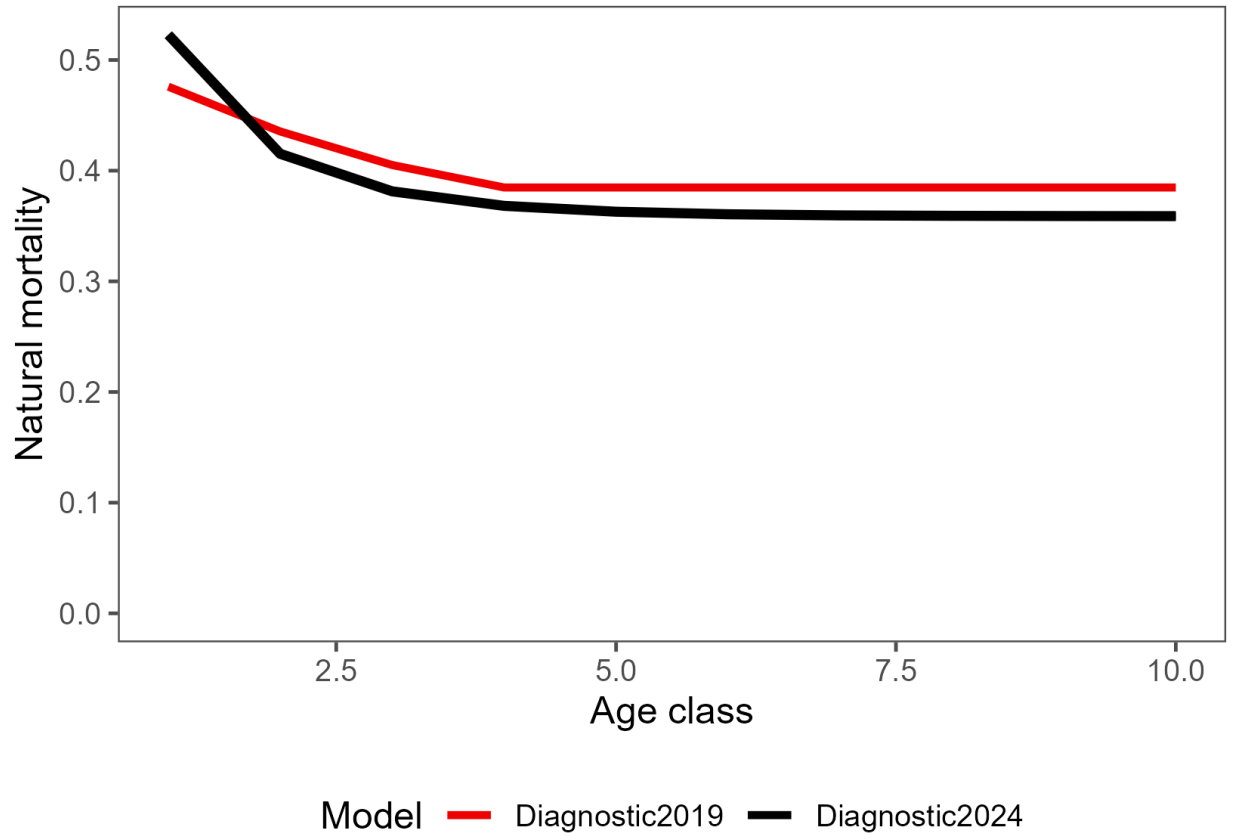
Figure 13: Natural mortality at-age comparison for the 2019 and 2024 diagnostic models.
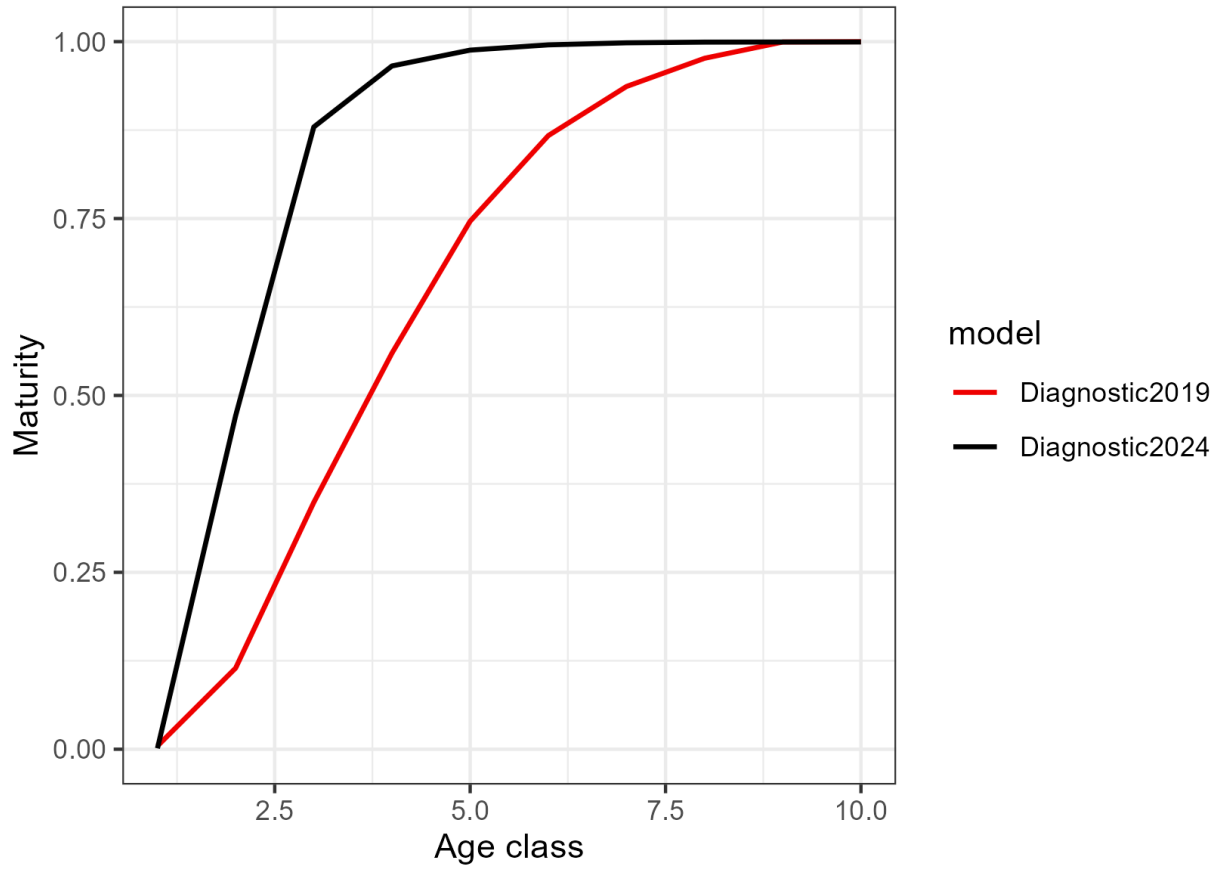
Figure 14: Maturity-at-age comparison for the 2019 and 2024 diagnostic model. The maturity-at-age used in the diagnostic model 2019 is shown in red and the maturity-at-age used for the diagnostic model 2024 is shown in black.
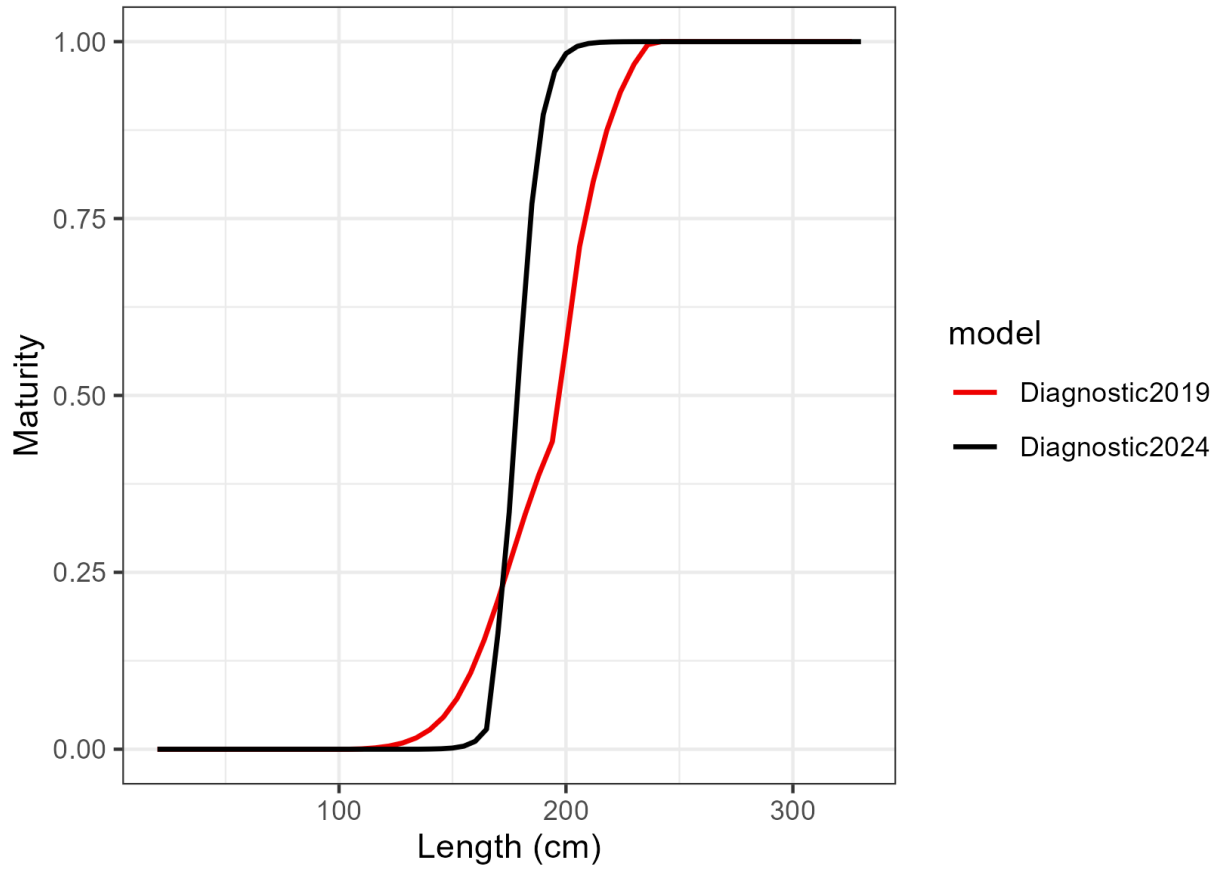
Figure 15: Maturity-at-length comparison for the 2019 and 2024 diagnostic model. The maturity-at-length used in the diagnostic model 2019 is shown in red and the maturity-at-length used for the diagnostic model 2024 is shown in black.
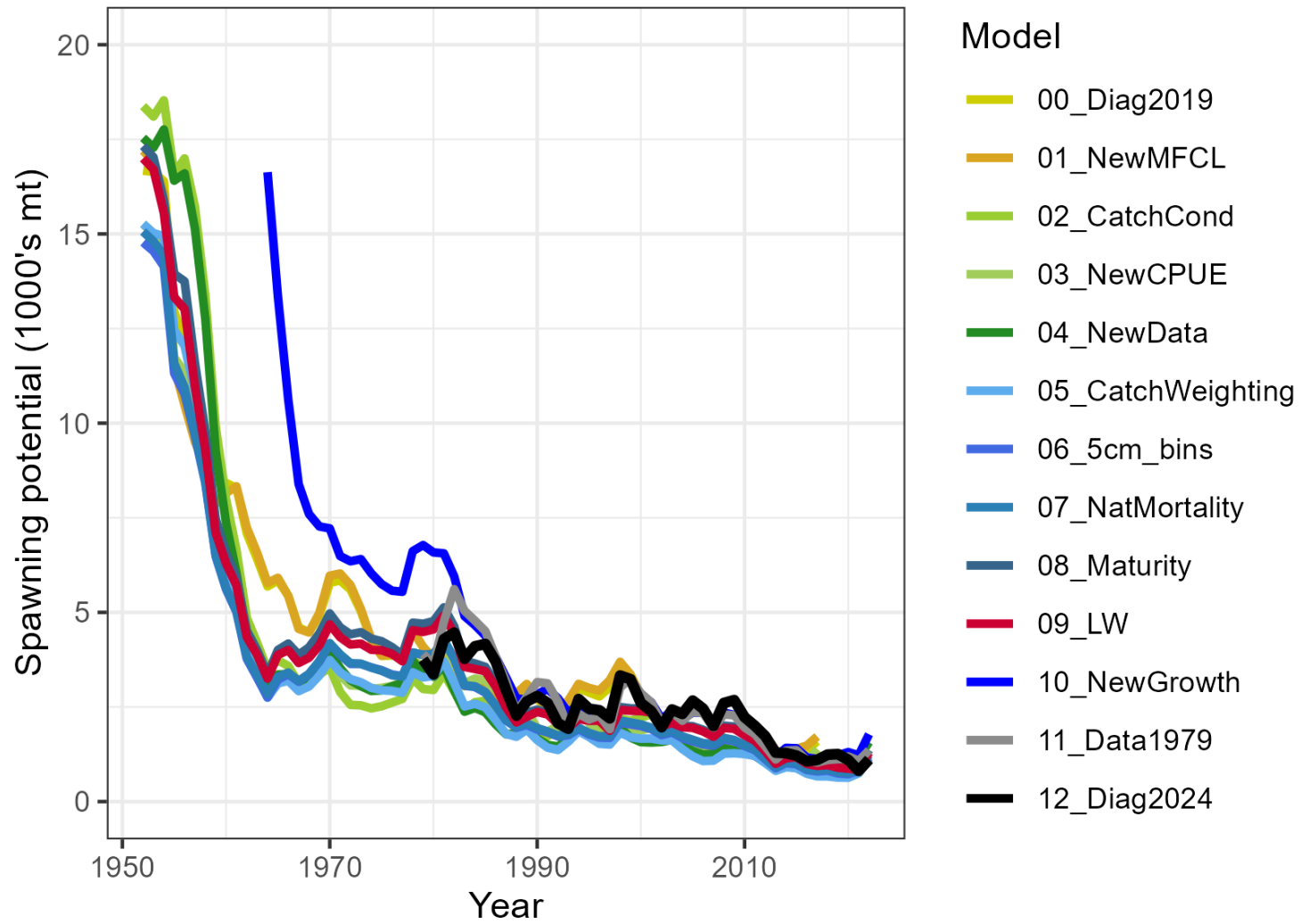
Figure 16: Estimated spawning potential, SB_t, trajectories for each of the main steps in the stepwise model development, with the 2024 diagnostic model shown in black. Note: The time series of SB for step 10 is truncated in the mid 1960's, to show the detail for the other steps shown in this plot, as SB increased to very high levels early in the time series for step 10.
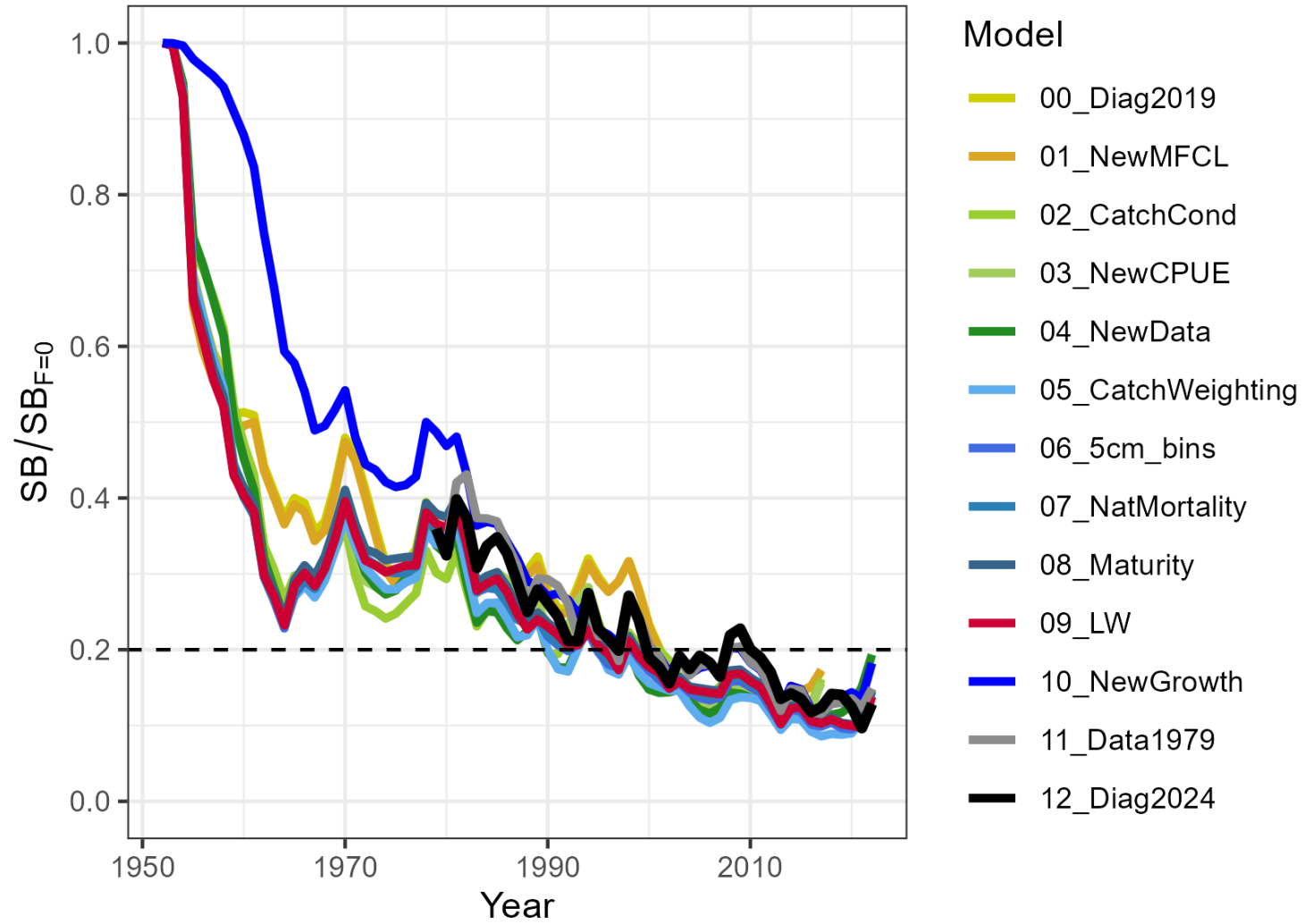
Figure 17: Estimated dynamic spawning depletion, $SB_t/SB_{t,F=0}$, trajectories for each of the main steps in the stepwise model development, with the 2024 diagnostic model shown black.
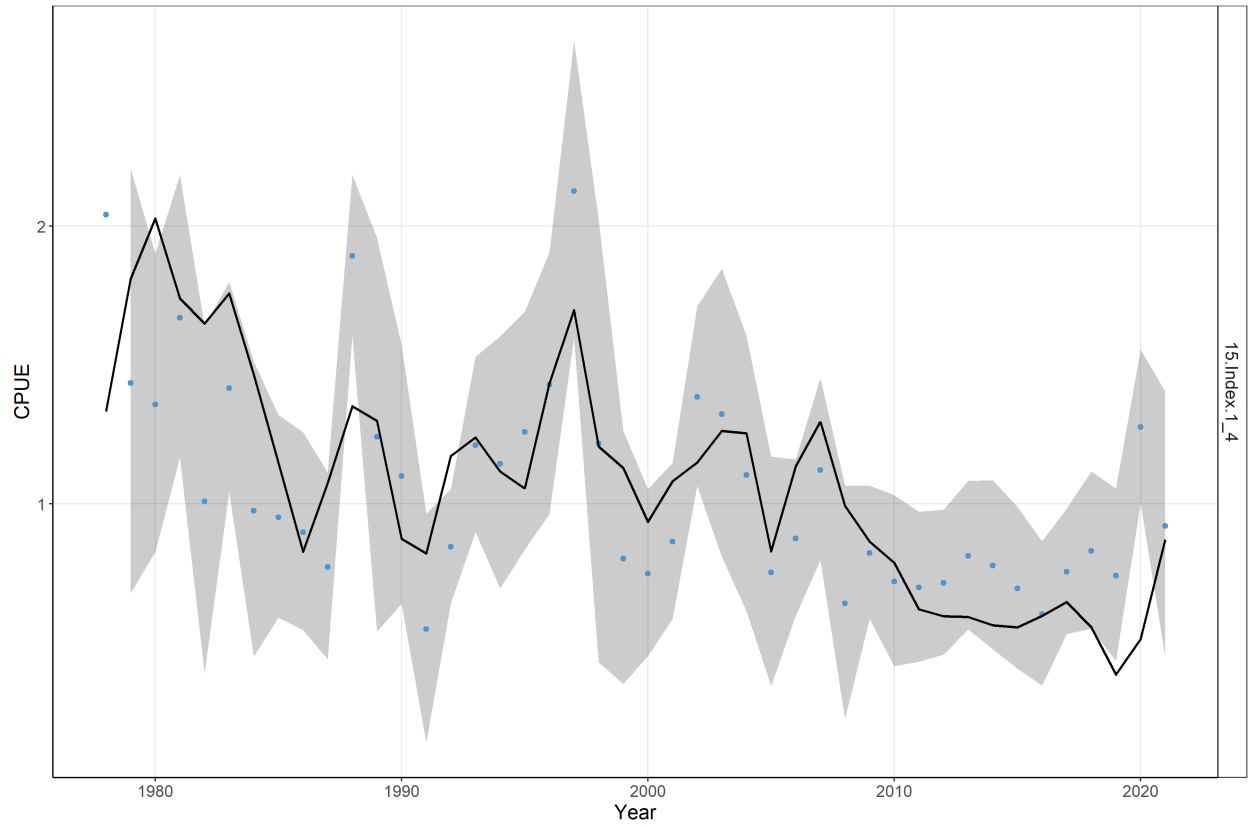
Figure 18: Fits (black line) to the standardised CPUE (blue dots) for the longline index fishery in sub-regions 1-4. The grey band indicates the 95% CI on the standardised CPUE.
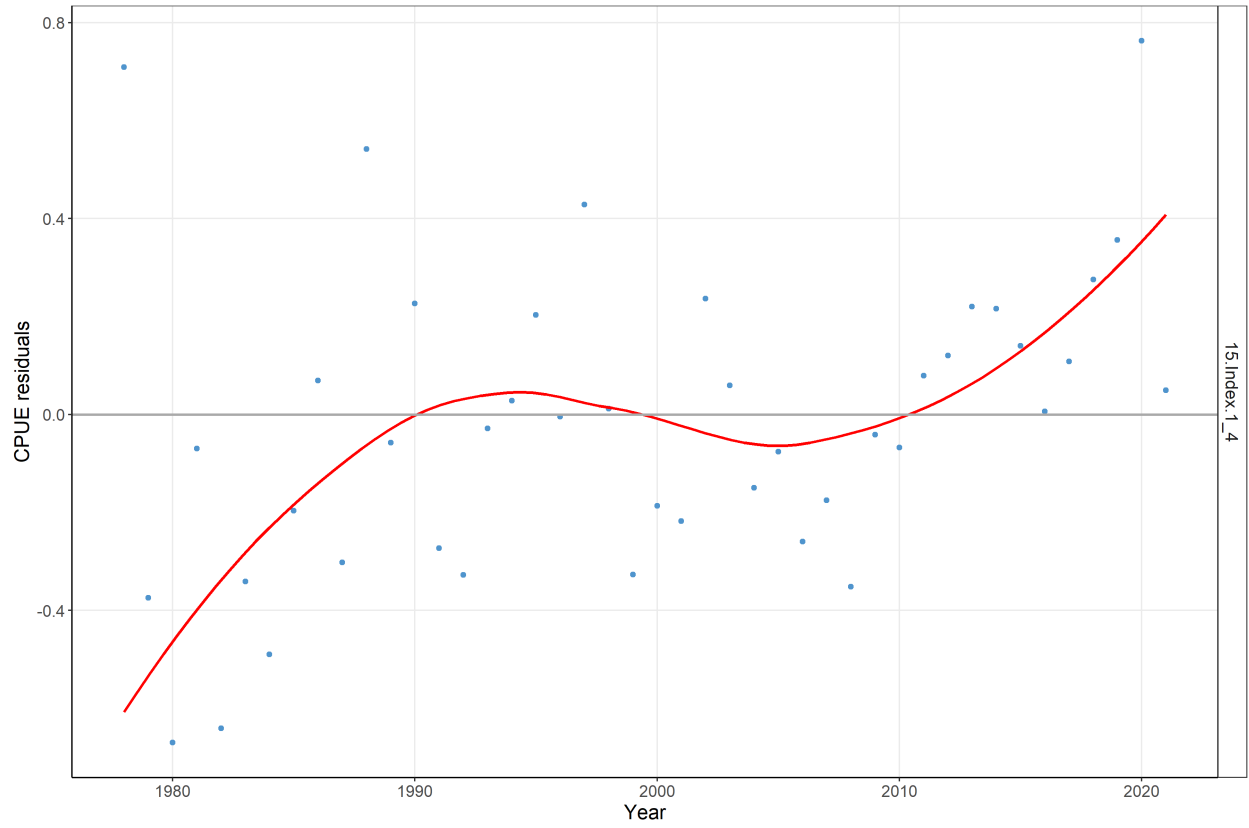
Figure 19: Residuals for the standardised CPUE for the the longline index fishery in sub-regions 1-4.
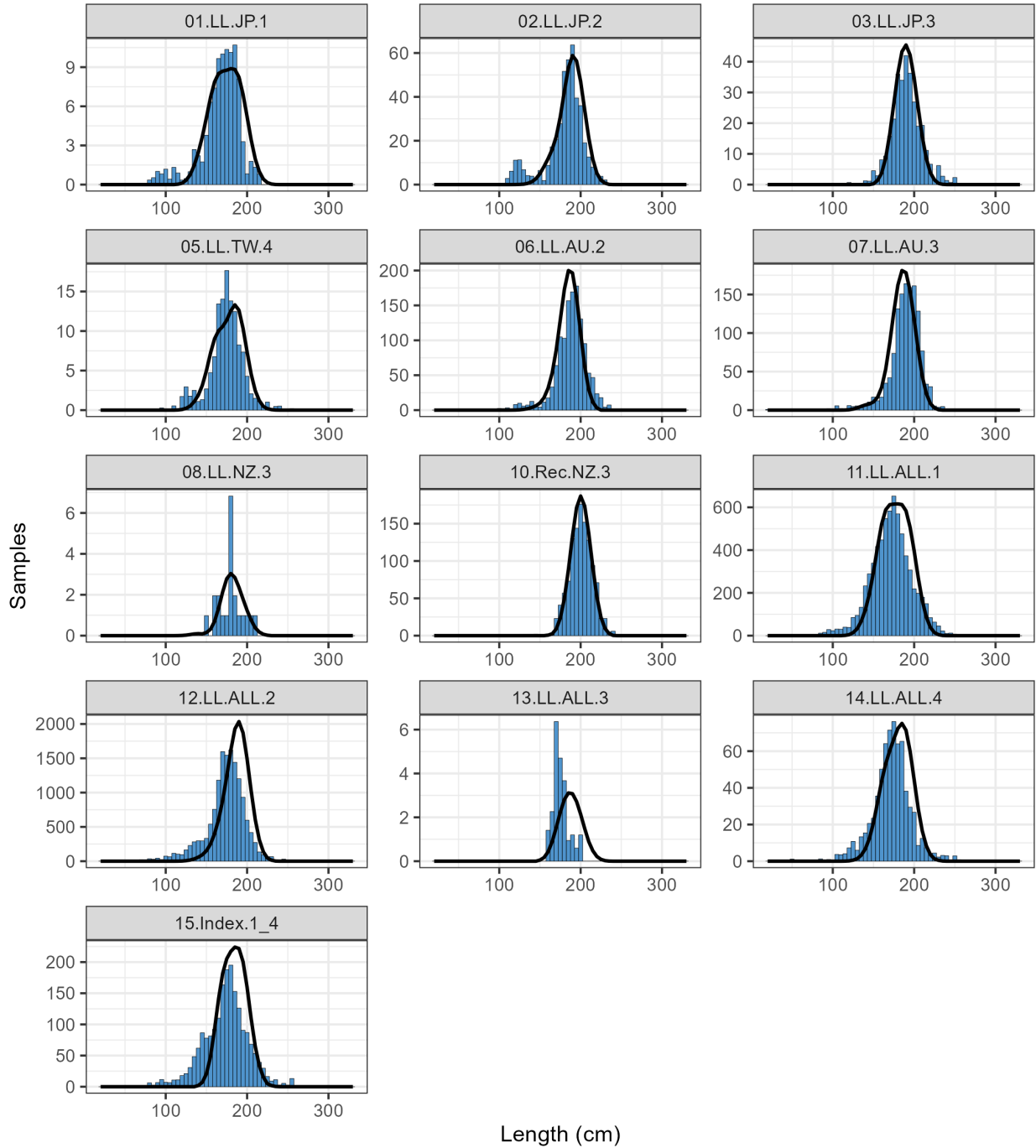
Figure 20: Composite (all time periods combined) observed (blue histograms) and predicted (black line) length frequency for fisheries with length frequency data for the 2024 diagnostic model.
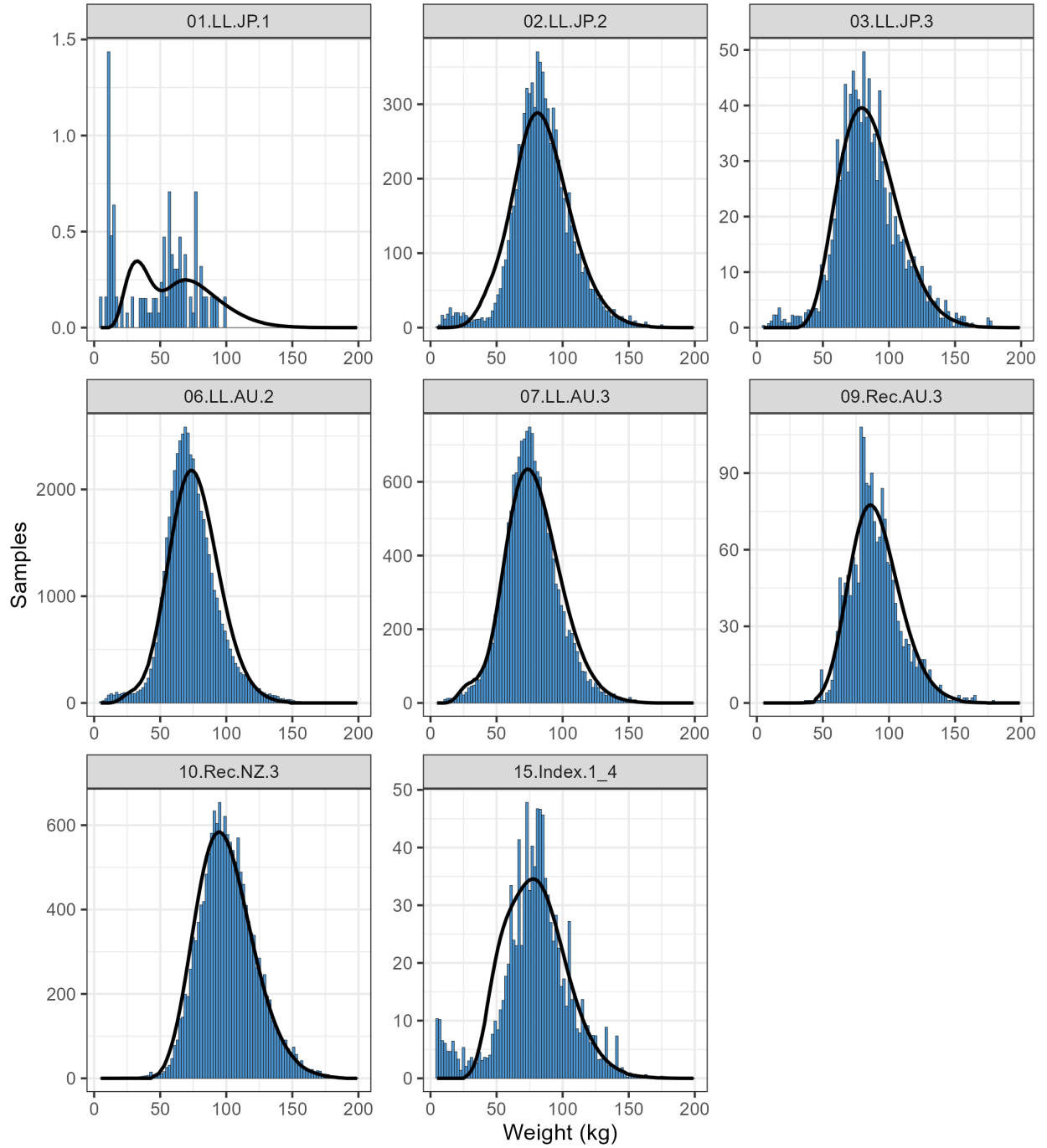
Figure 21: Composite (all time periods combined) observed (blue histograms) and predicted (black line) weight frequency for fisheries with weight frequency data for the 2024 diagnostic model.
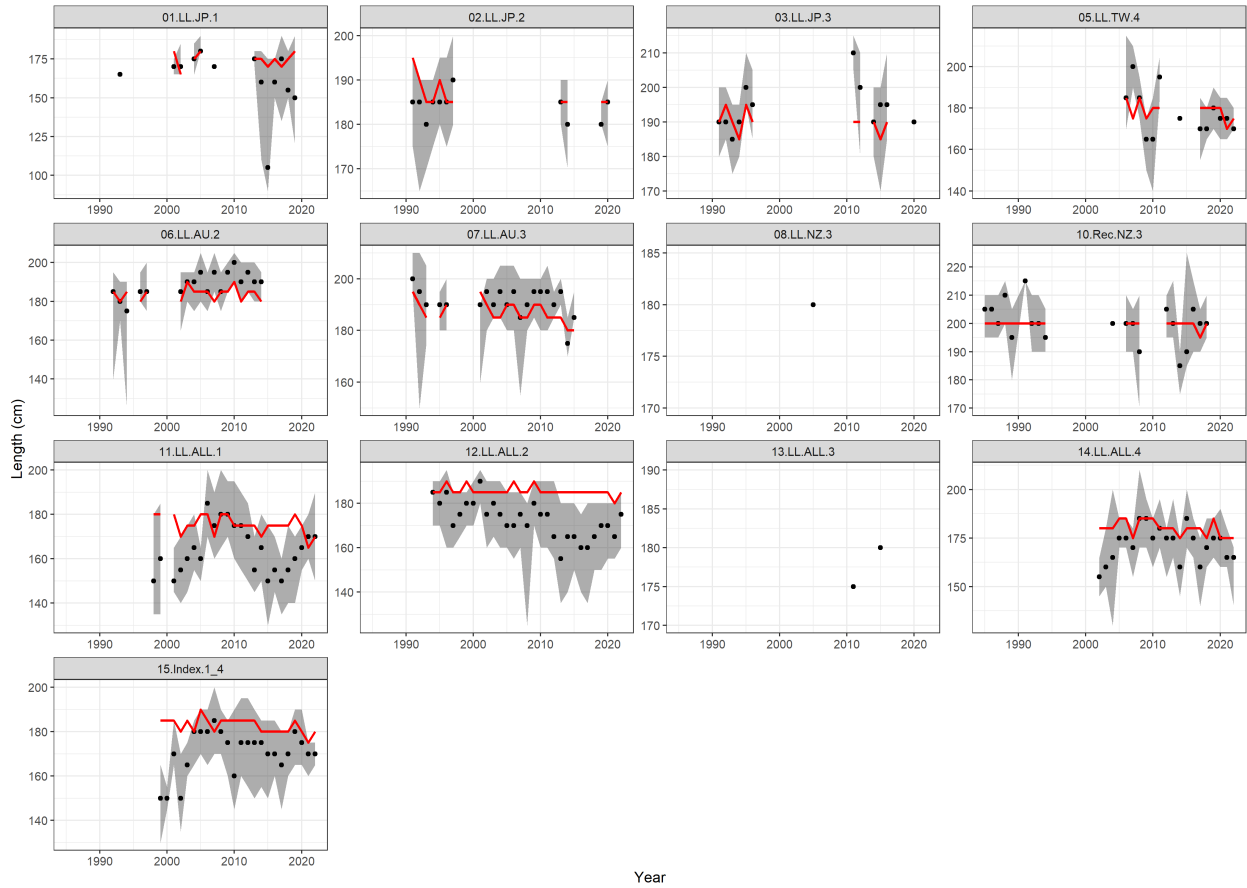
Figure 22: Observed (black points) and predicted (red line) median fish lengths (EFL, cm) for the fisheries with length data for the 2023 diagnostic model. The uncertainty intervals (grey shading) represent the values encompassed by the 25% and 75% quantiles. Sampling data are by year and only length samples with more than 10 fish per quarter are plotted.
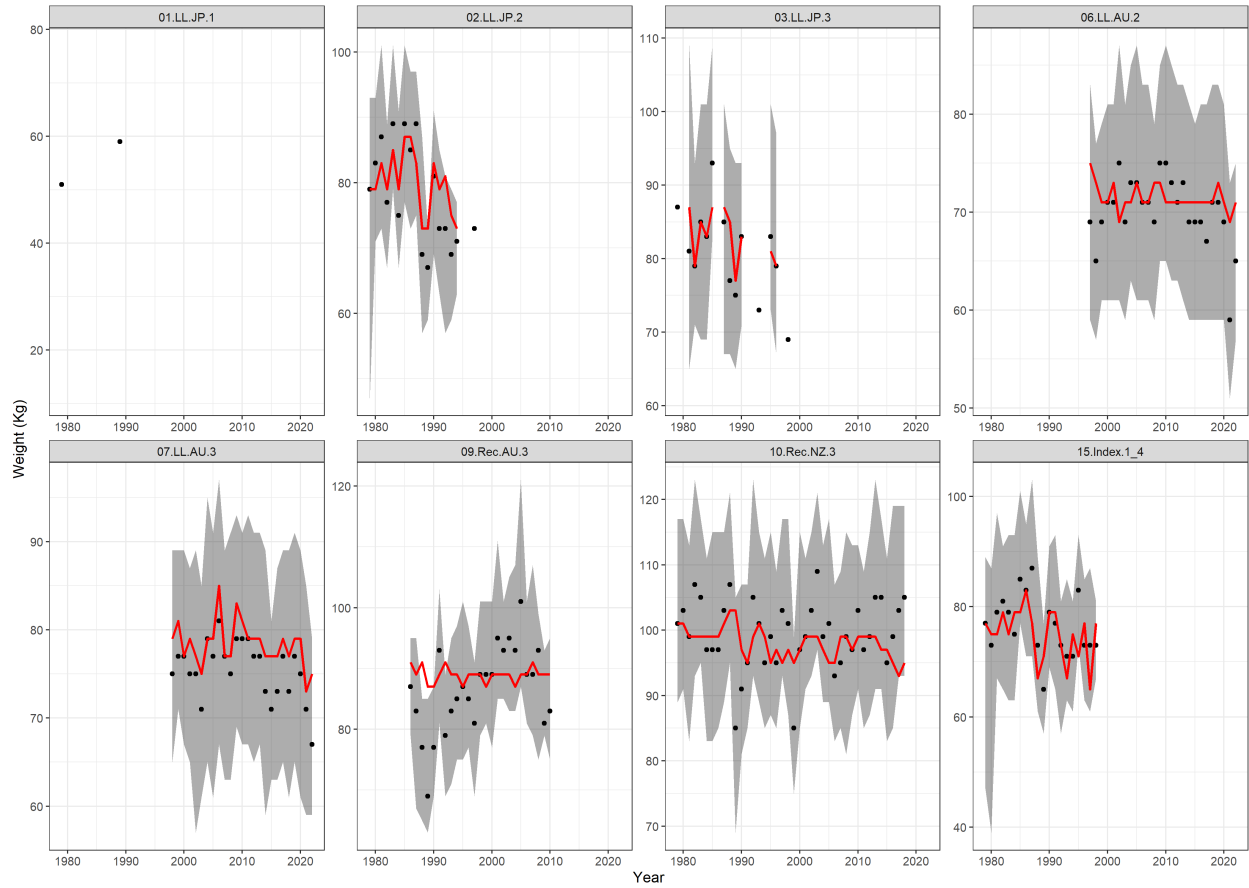
Figure 23: Observed (black points) and predicted (red line) median fish weights (kg) for the fisheries with weight data for the 2023 diagnostic model. The uncertainty intervals (grey shading) represent the values encompassed by the 25% and 75% quantiles. Sampling data are by year and only weight samples with more than 10 fish per quarter are plotted.
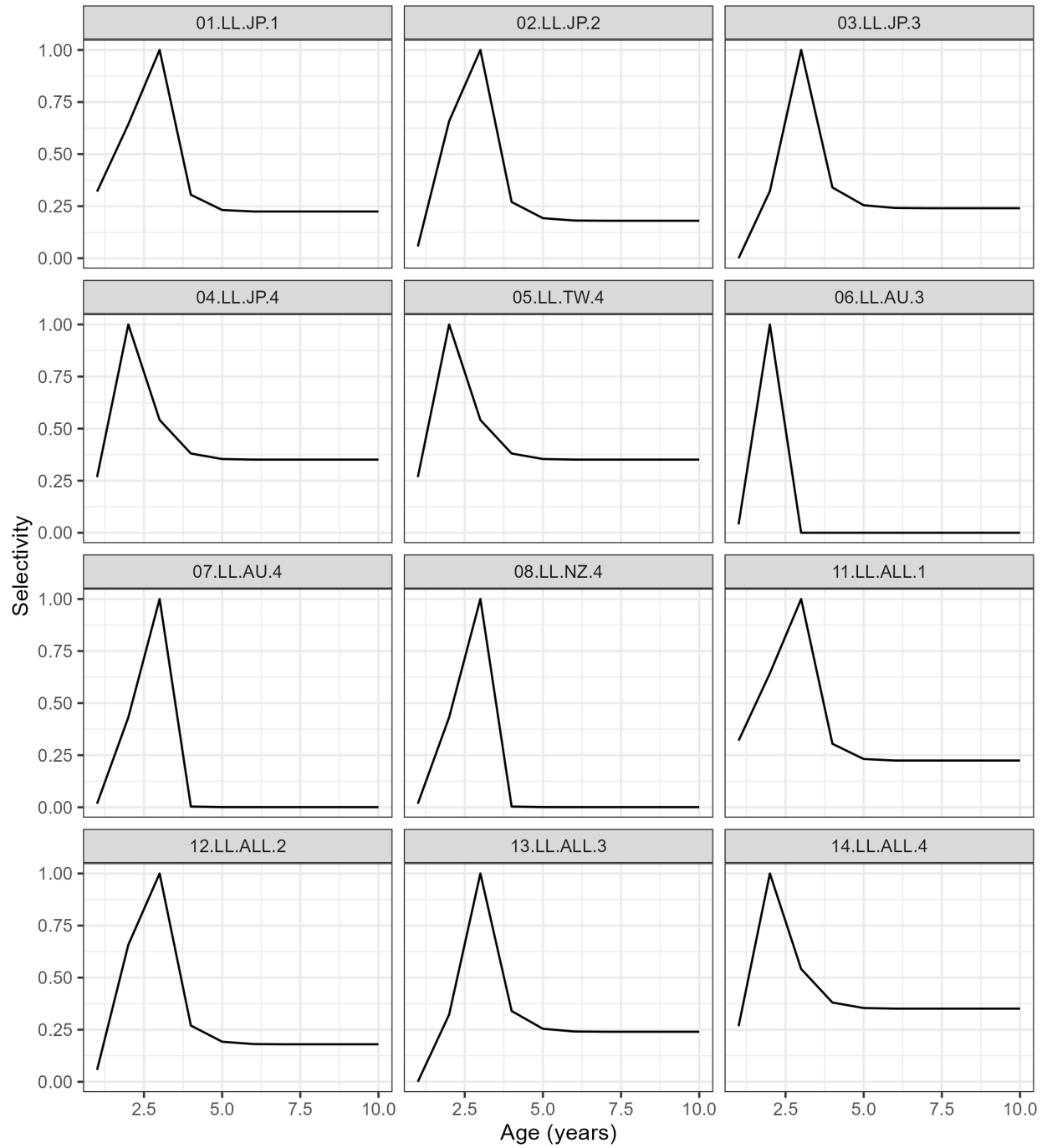
Figure 24: Estimated age-specific selectivity curves by long-line fishery for the diagnostic model.
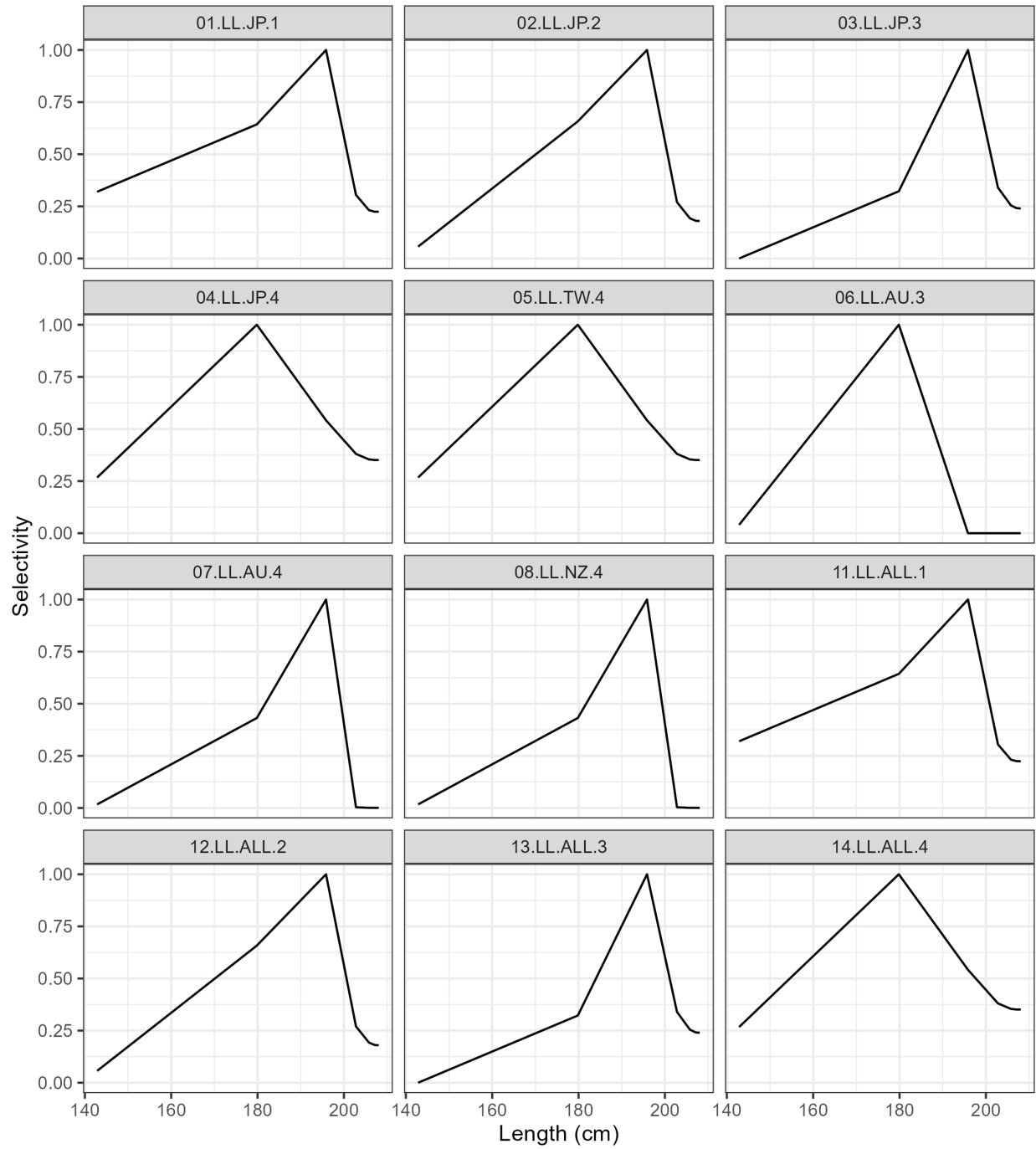
Figure 25: Estimated length-specific selectivity curves by long-line fishery for the diagnostic model.
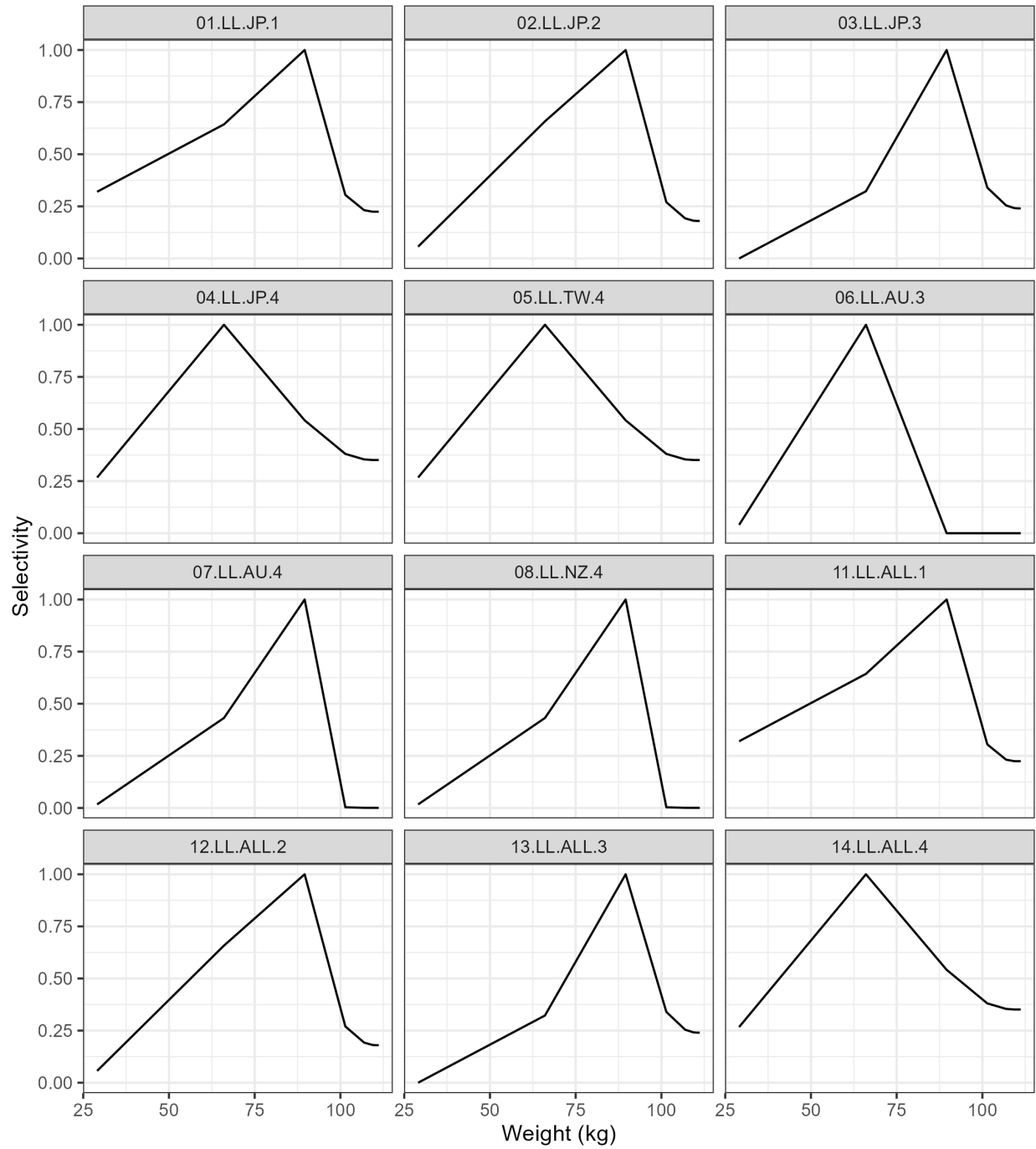
Figure 26: Estimated weight-specific selectivity curves by long-line fishery for the diagnostic model.
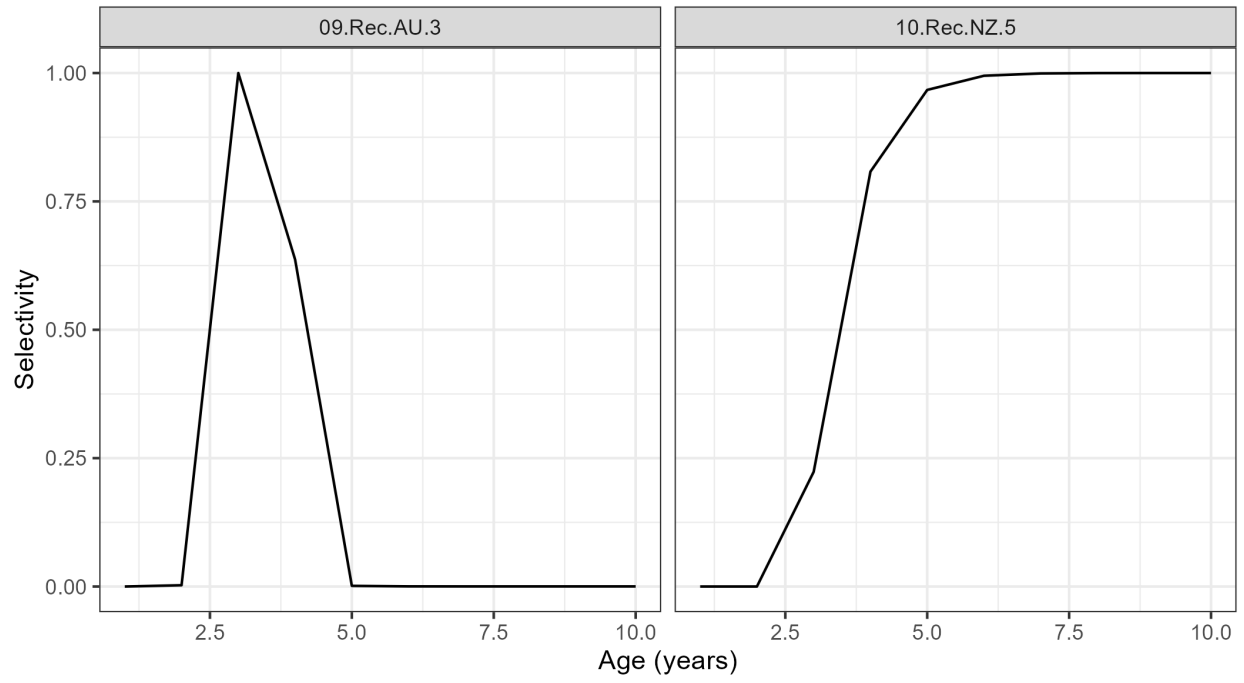
Figure 27: Estimated age-specific selectivity curves by recreational fishery for the diagnostic model.
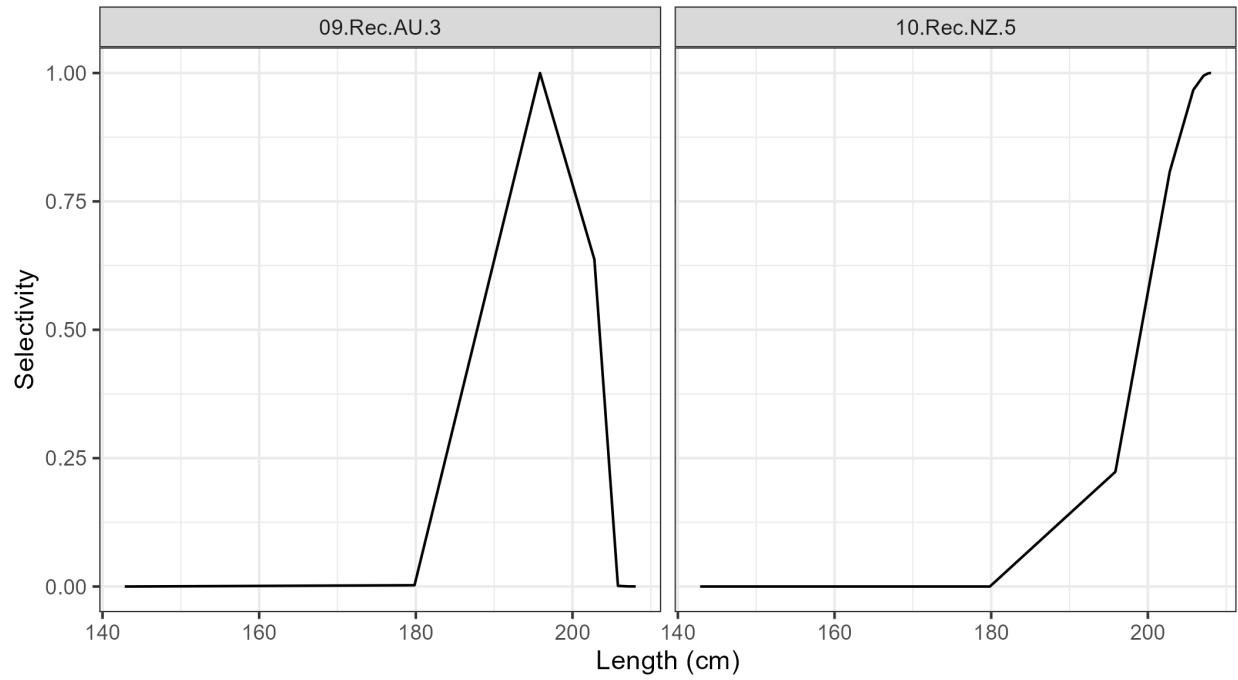
Figure 28: Estimated length-specific selectivity curves by recreational fishery for the diagnostic model.
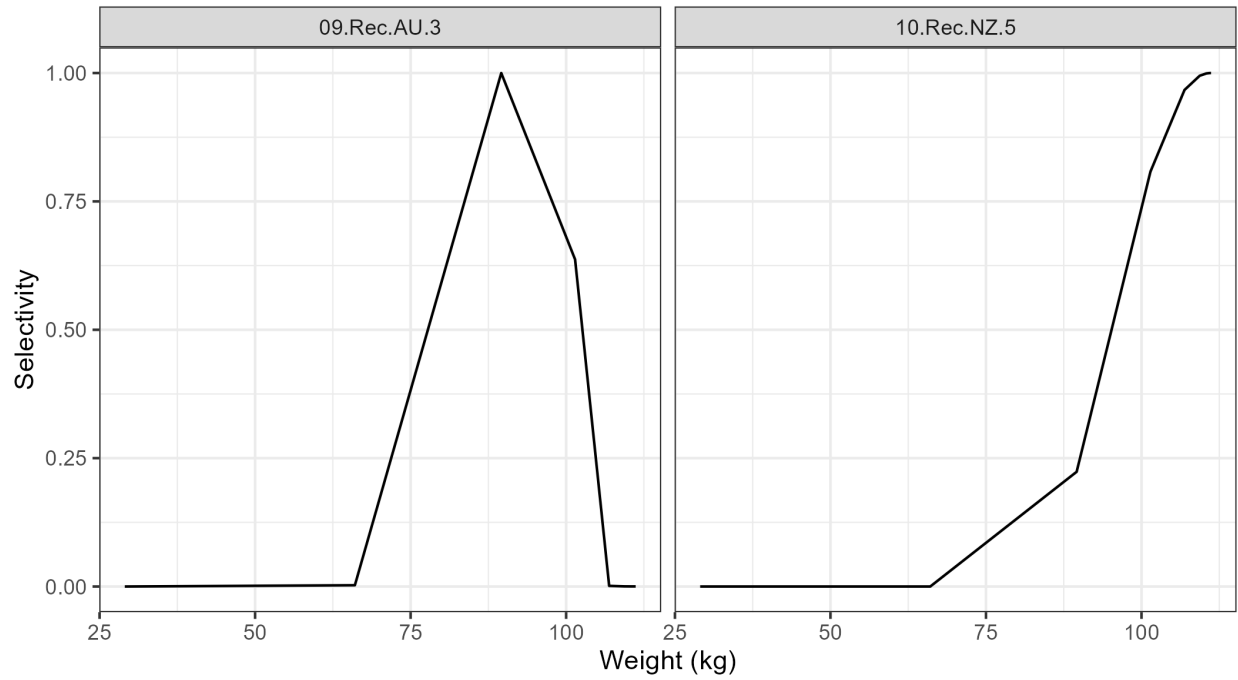
Figure 29: Estimated weight-specific selectivity curves by recreational fishery for the diagnostic model.
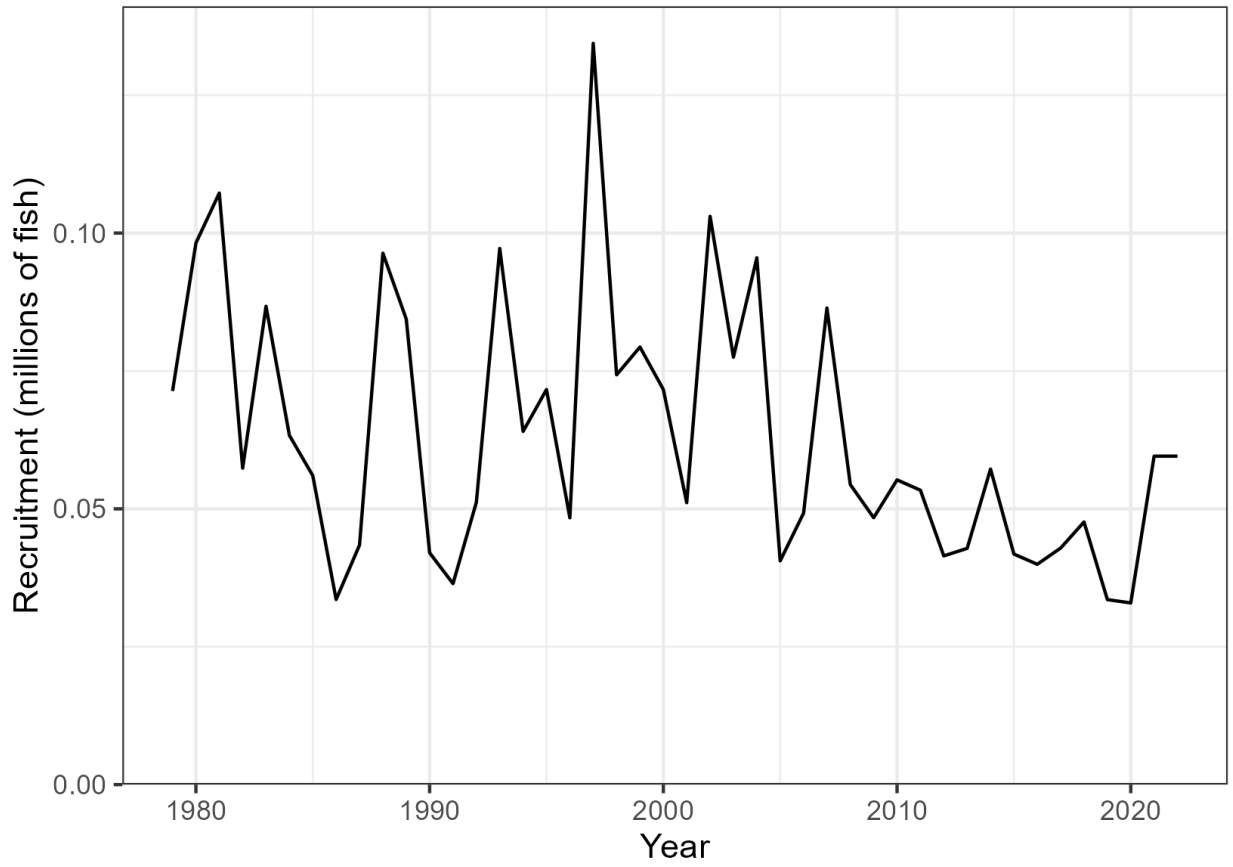
Figure 30: The estimated recruitment from the diagnostic model.
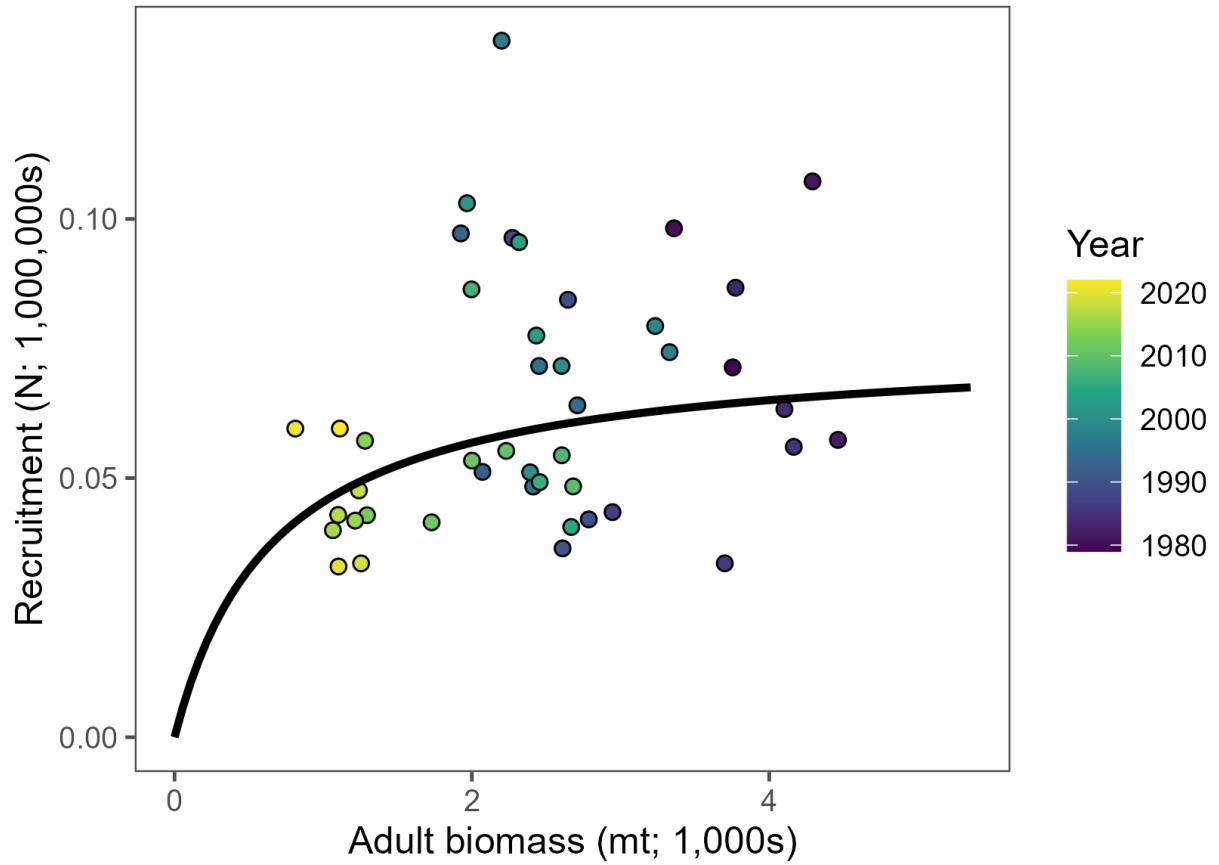
Figure 31: Estimated relationship between recruitment and spawning biomass based on annual values for the diagnostic model. The colour of the circle corresponds to the year: darker-cooler colours are early in the model period, and lighter-warmer colours are more recent years.
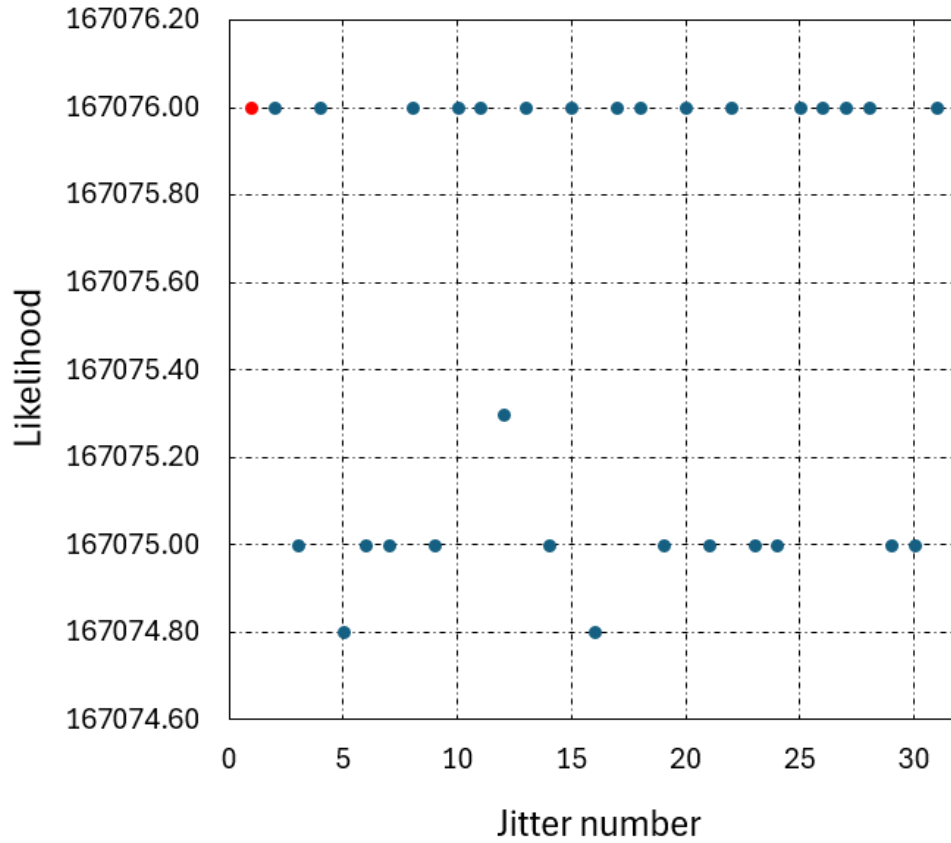
Figure 32: Log likelihood plotted against jitter number, with the diagnostic model plotted in red. Multiple jitters found the same value for log likelihood as the diagnostic model and all others jittered iterations had lower values for the log likelihood, noting that we aim to final a model which minimises the negative log likelihood.
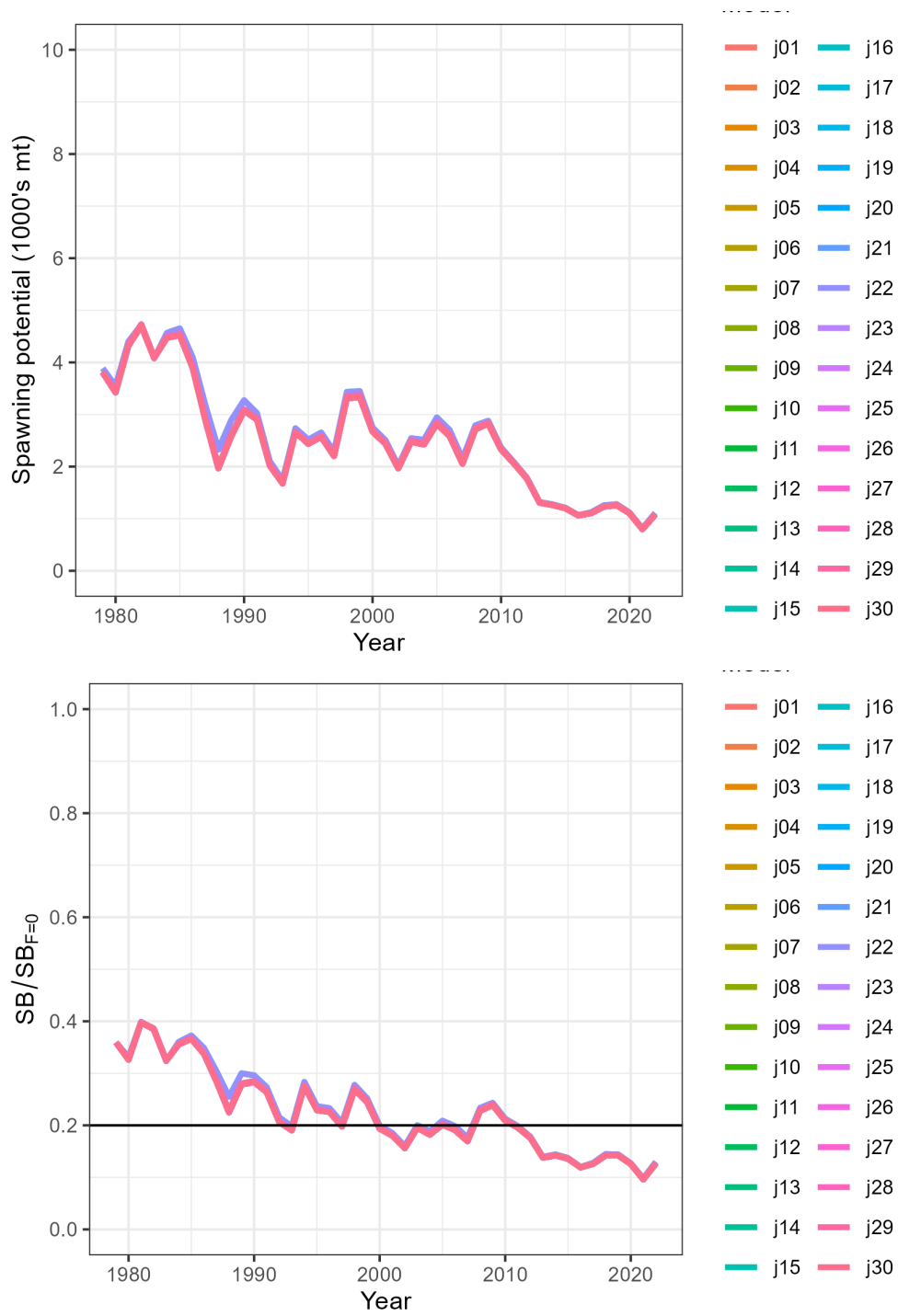
Figure 33: Spawning potential (top) and depletion (bottom) time series for the diagnostic model and for each of the jittered replicates.
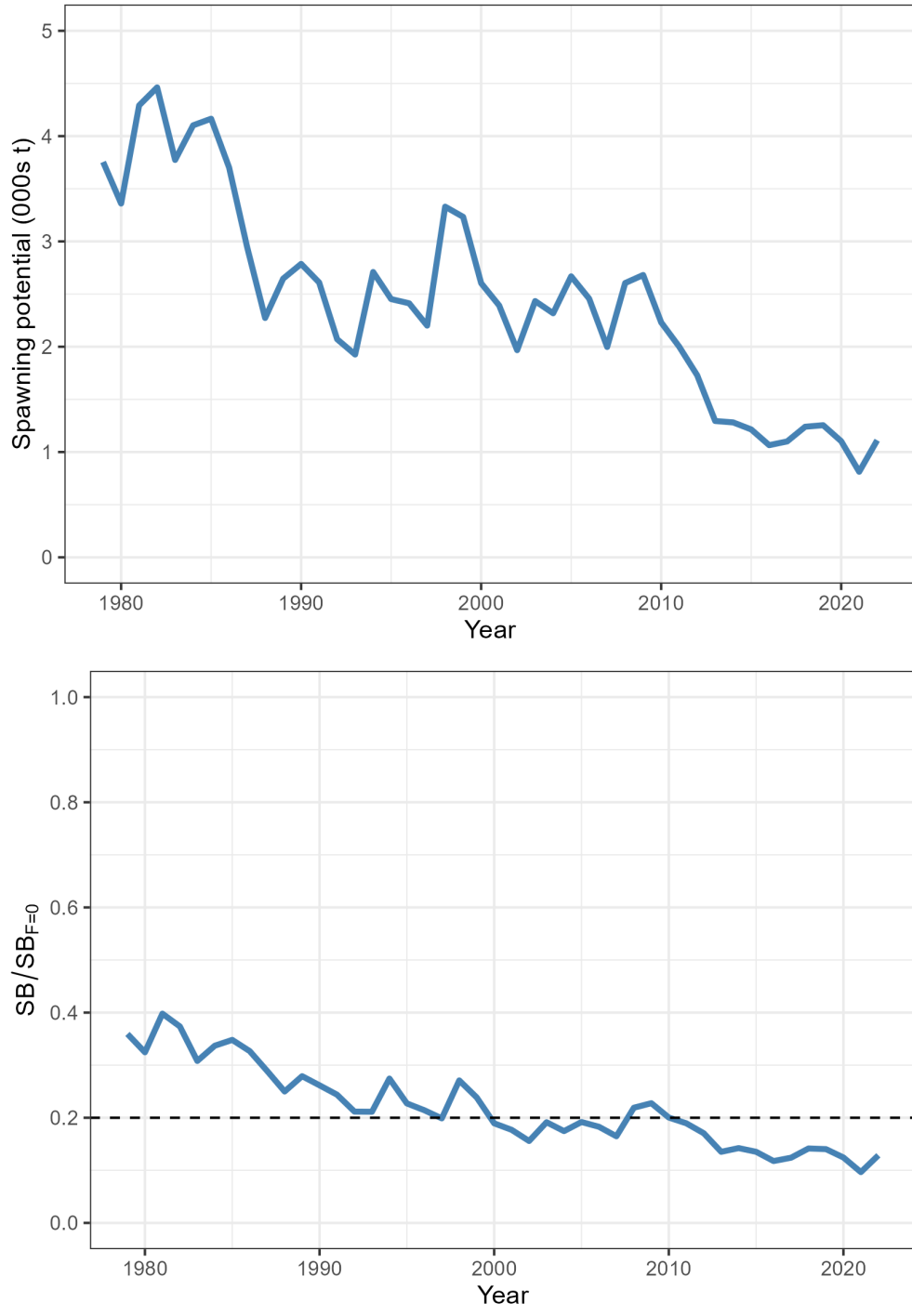
Figure 34: Spawning potential (top) and depletion (bottom) time series for the diagnostic model.
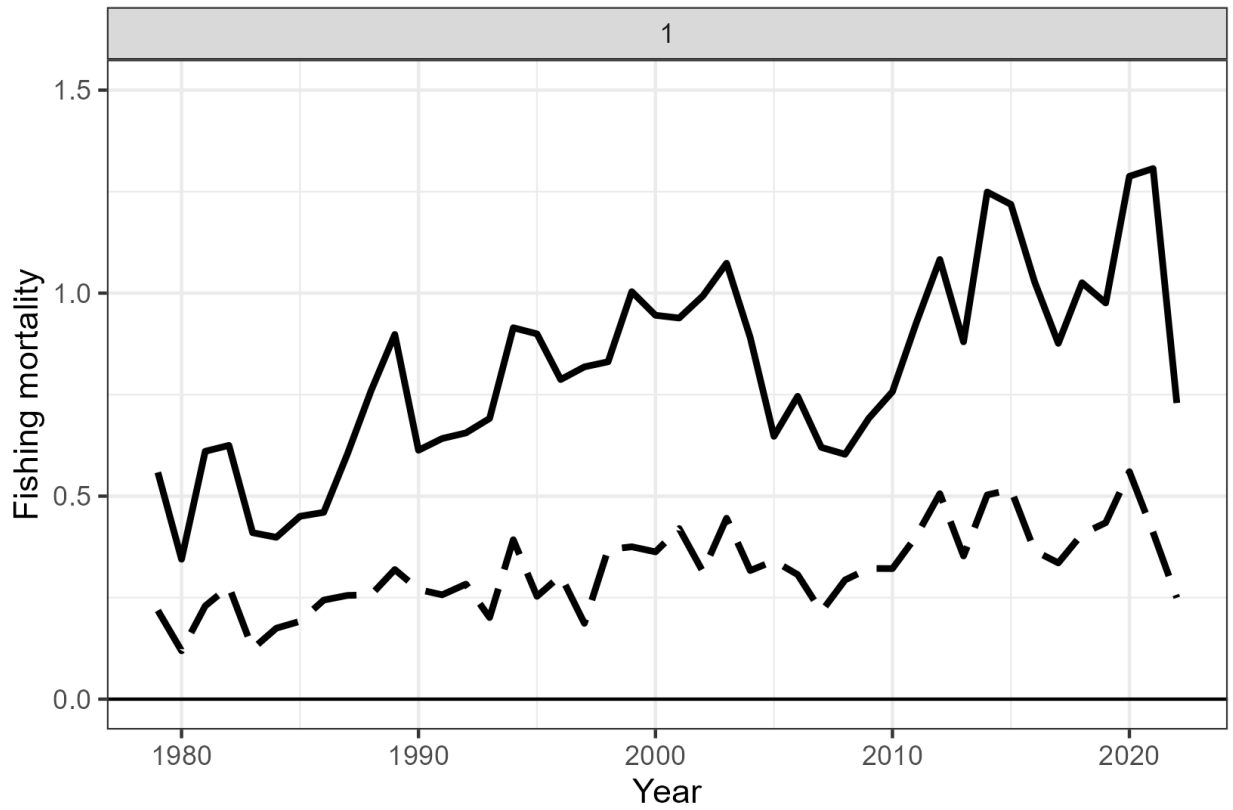
Figure 35: Estimated average annual juvenile (dash line) and adult (solid line) fishing mortality for the diagnostic model.
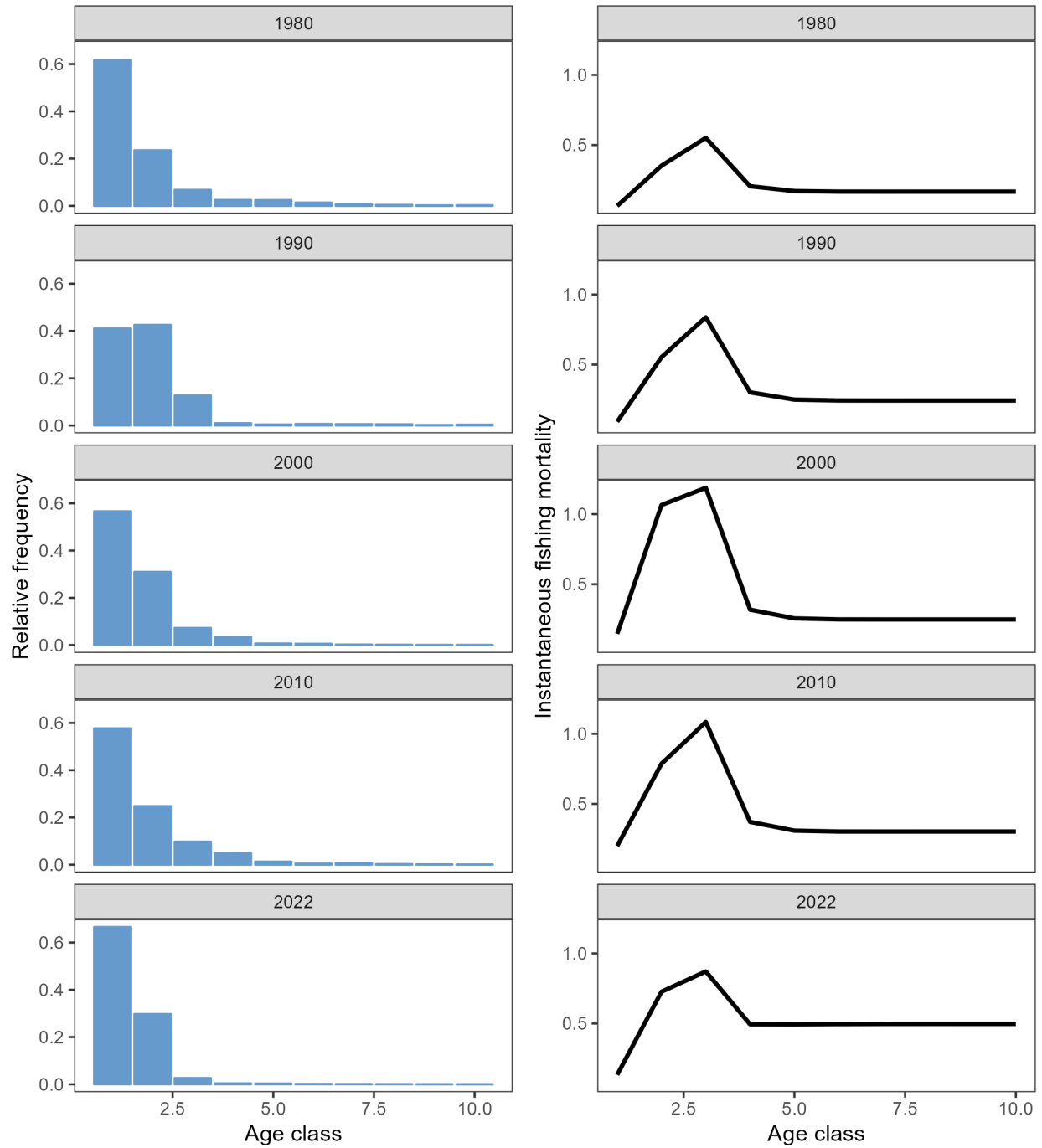
Figure 36: Estimated relative frequency by age class and decade (numbers-at-age; left) and estimated instantaneous fishing mortality by age class and decade (right) for the diagnostic model.
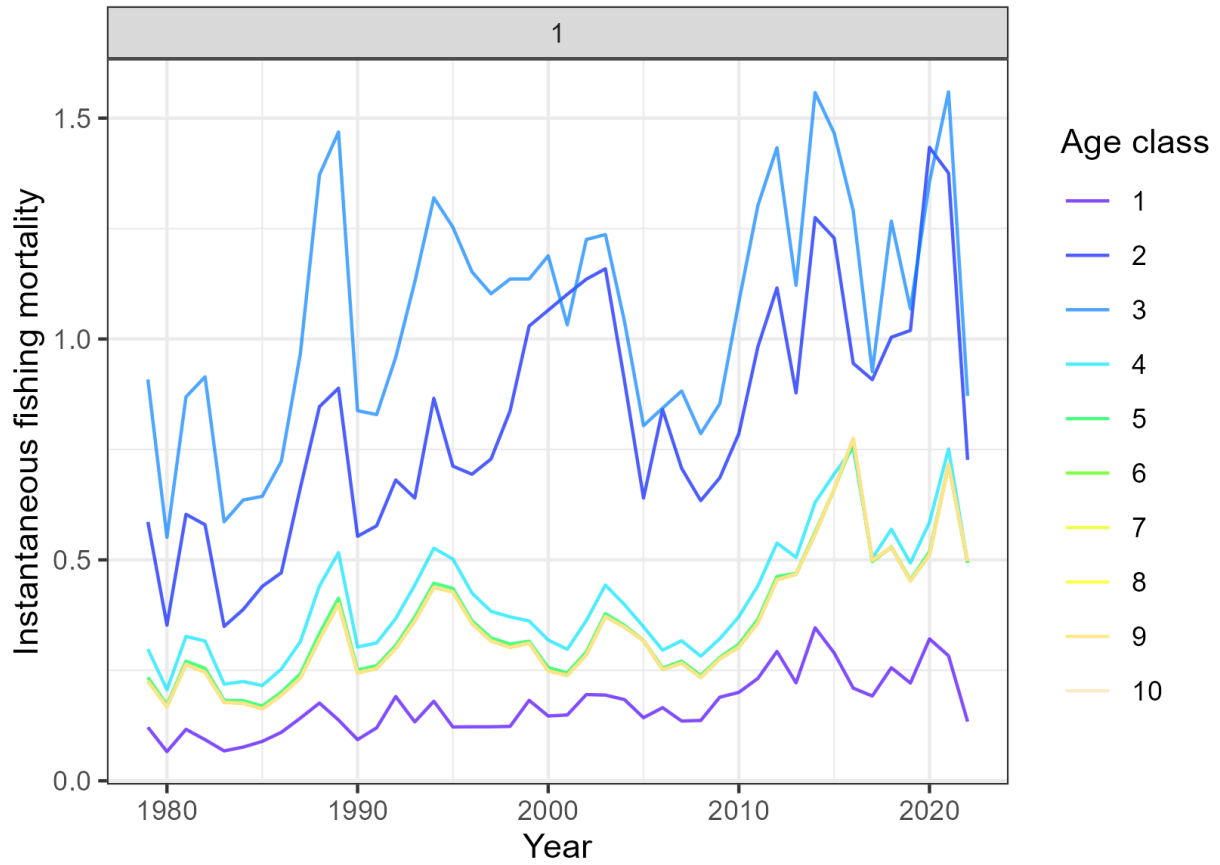
Figure 37: Estimated annual instantaneous fishing mortality by age class for the diagnostic model.
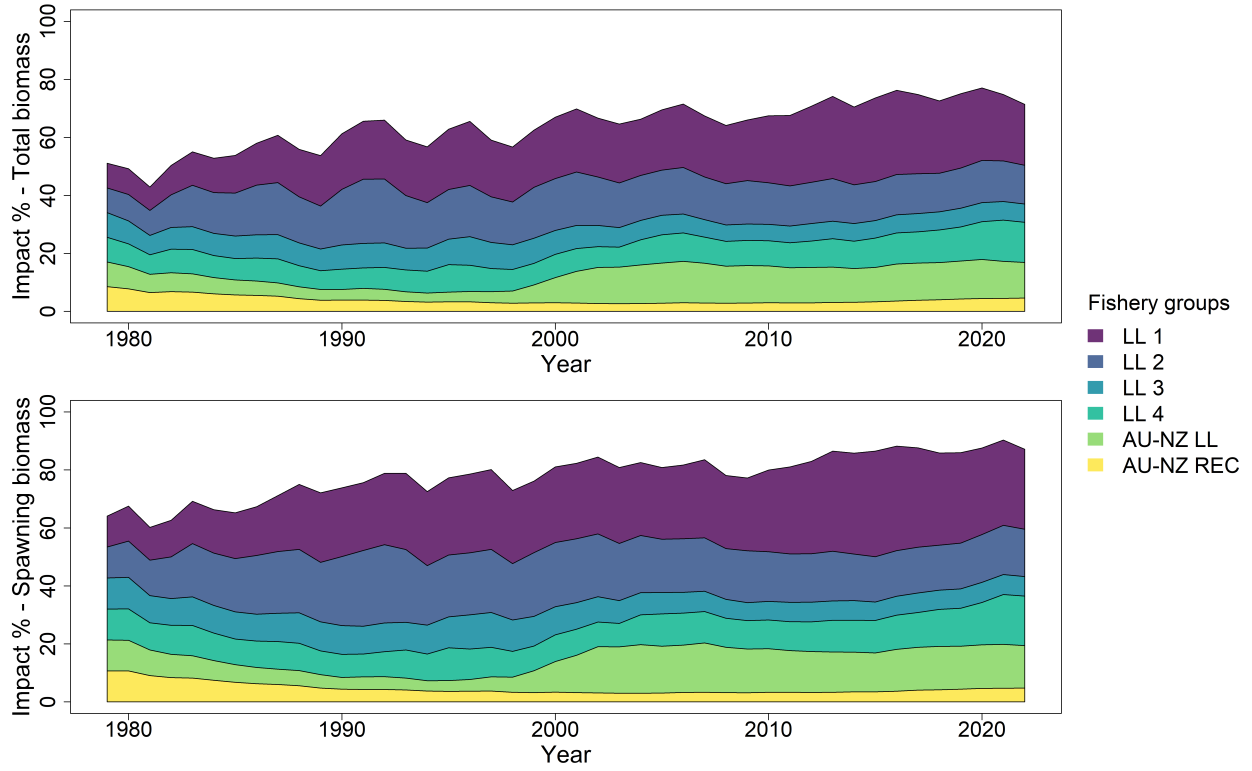
Figure 38: Estimates of reduction in total biomass (top) and spawning biomass (bottom) due to fishing (fishery impact $= 1 - B_t/B_{t,F=0}$) attributed to various fishery groups for the diagnostic model.
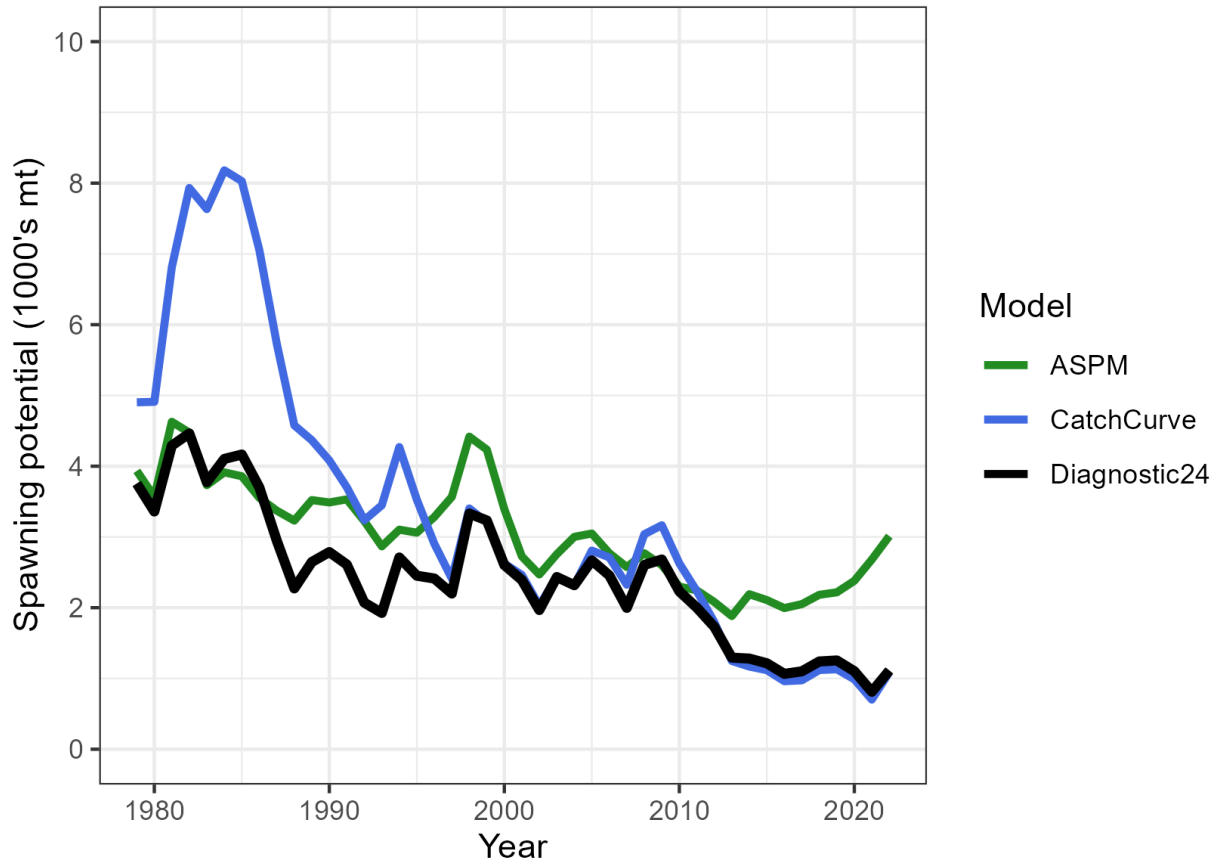
Figure 39: *SB* for the age-structured production model (ASPM) without recruitment deviations, catchcurve analysis (CatchCurve), and the diagnostic model (Diagnostic24).
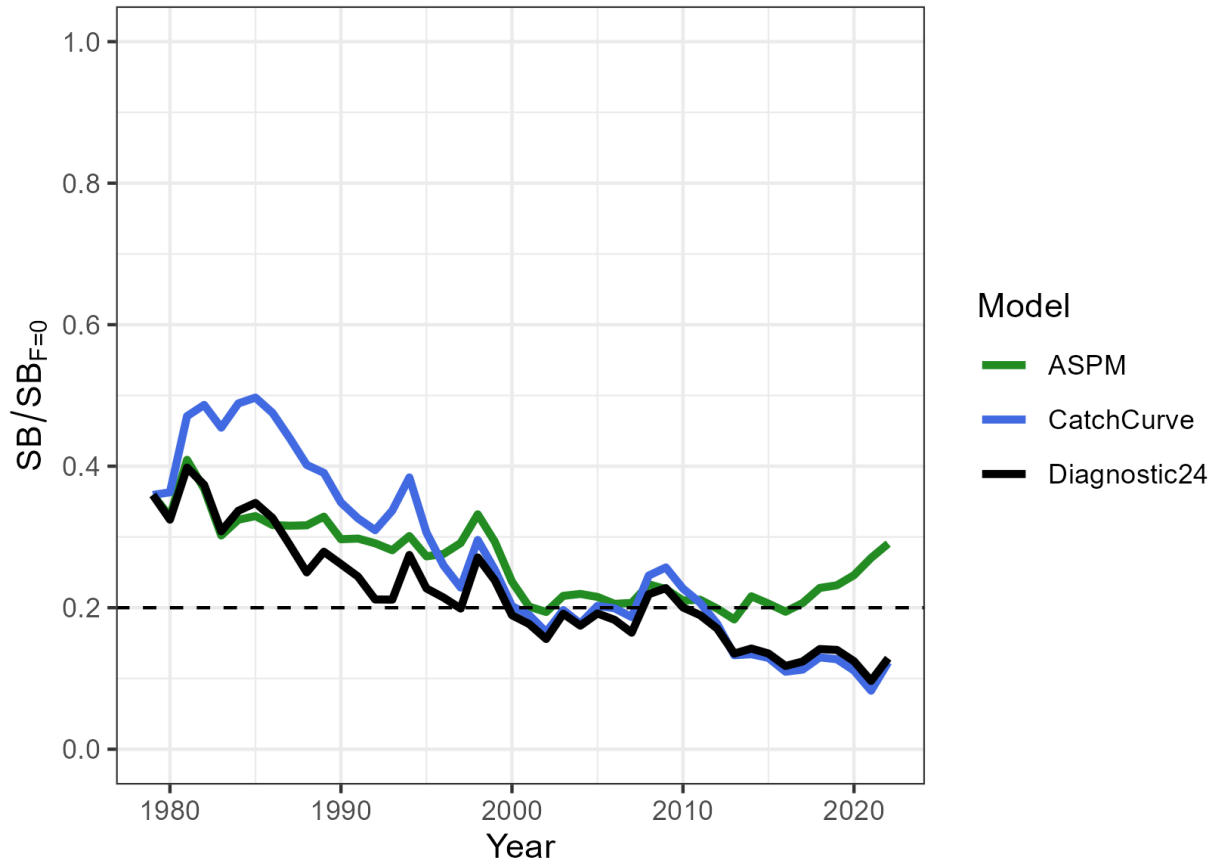
Figure 40: $SB_{F=0}$ for the age-structured production model (ASPM) without recruitment deviations, catchcurve analysis (CatchCurve), and the diagnostic model (Diagnostic24).
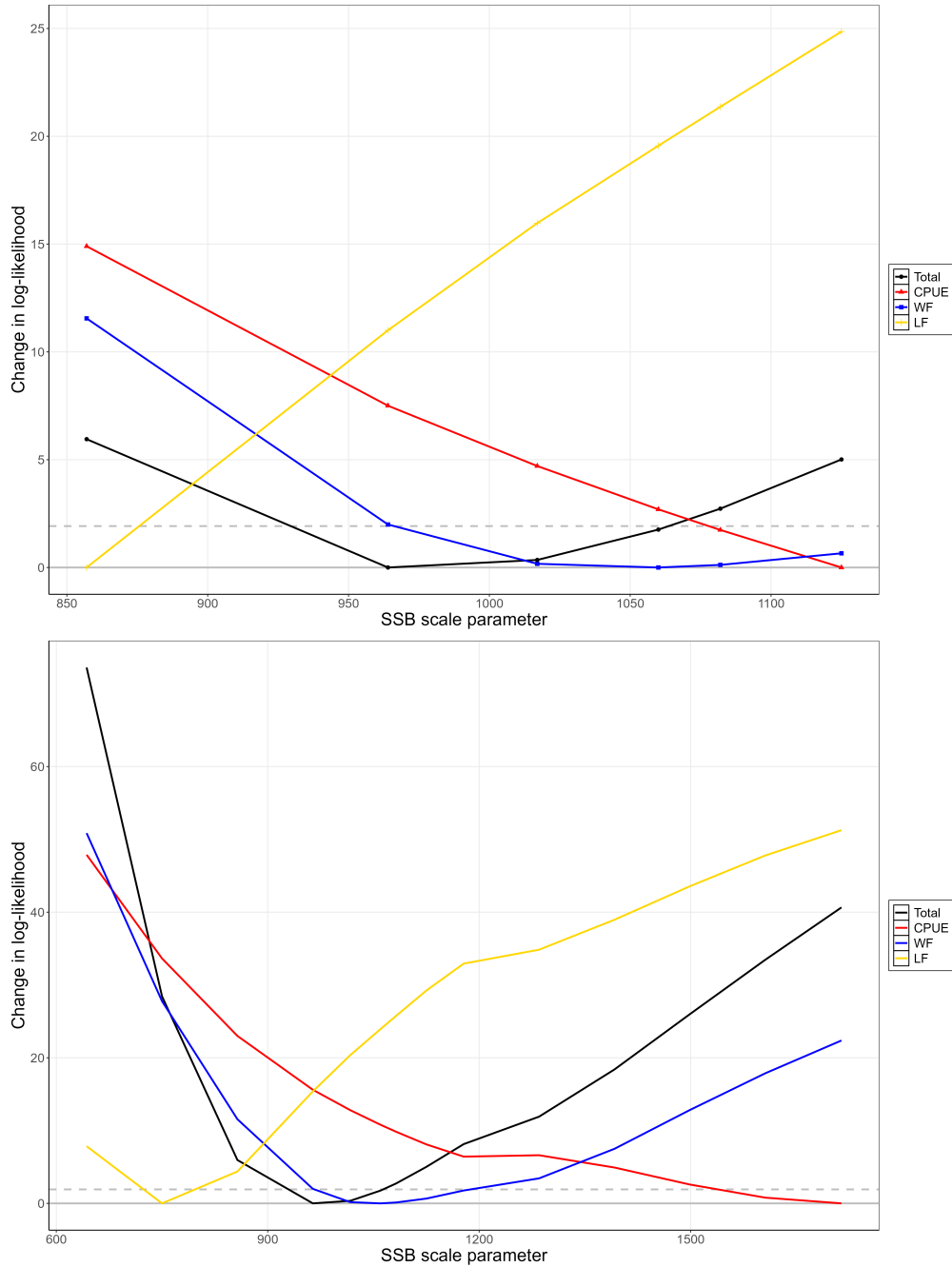
Figure 41: Likelihood profile of *SB*. The black line indicates the total likelihood with the colours representing components of the total likelihood coming from different data sources, CPUE (red), length (yellow) and weight (blue). The upper plot focuses on the range of total biomass values where the change in total likelihood is not statistically significant, with less than 1.92 likelihood units of change, indicated by the horizontal dashed line. The lower plot shows a broader range indicating the minima for most of the individual data sources.
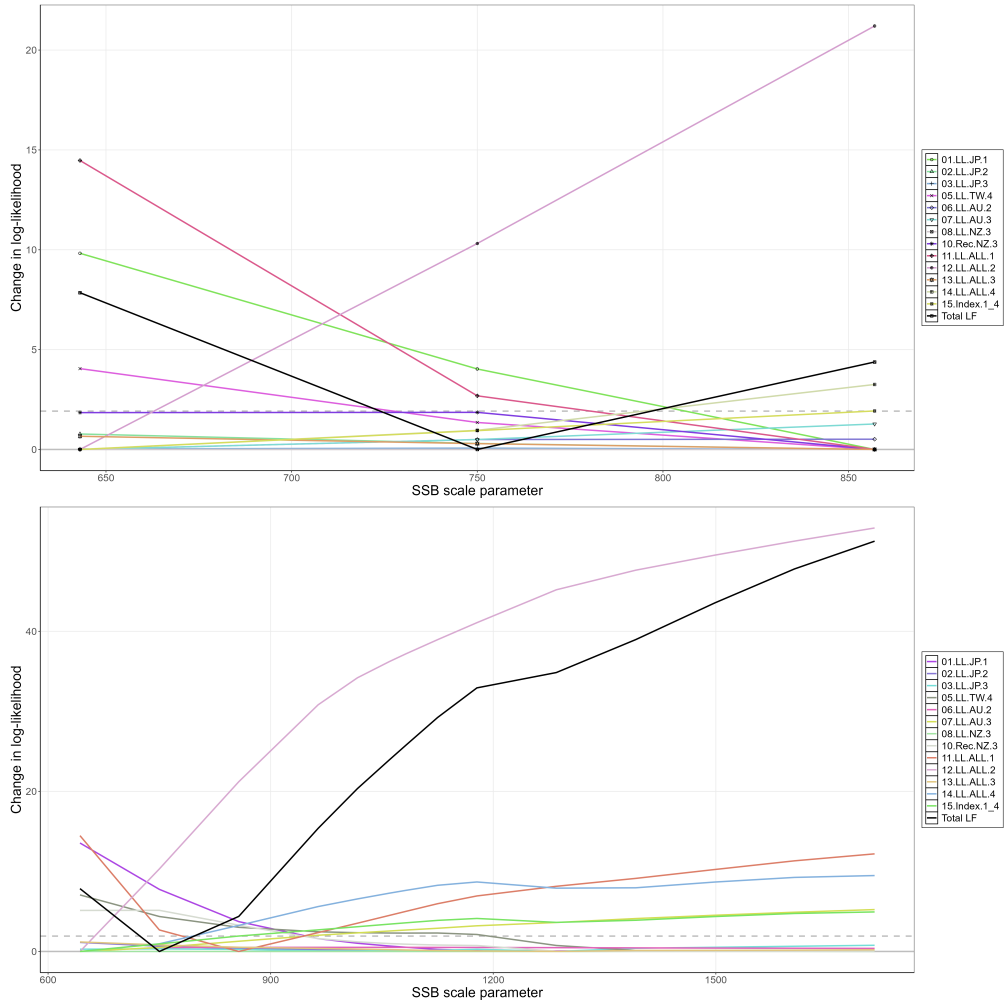
Figure 42: Likelihood profile of *SB*, showing only the length component of the total log-likelihood separated by individual fishery. The black line indicates the total likelihood with the colours representing components of the total likelihood coming from different fisheries. The upper plot focuses on the range of total biomass values where the change in total likelihood is not statistically significant, with less than 1.92 likelihood units of change, indicated by the horizontal dashed line. The lower plot shows a broader range indicating the minima for more of the individual data sources.
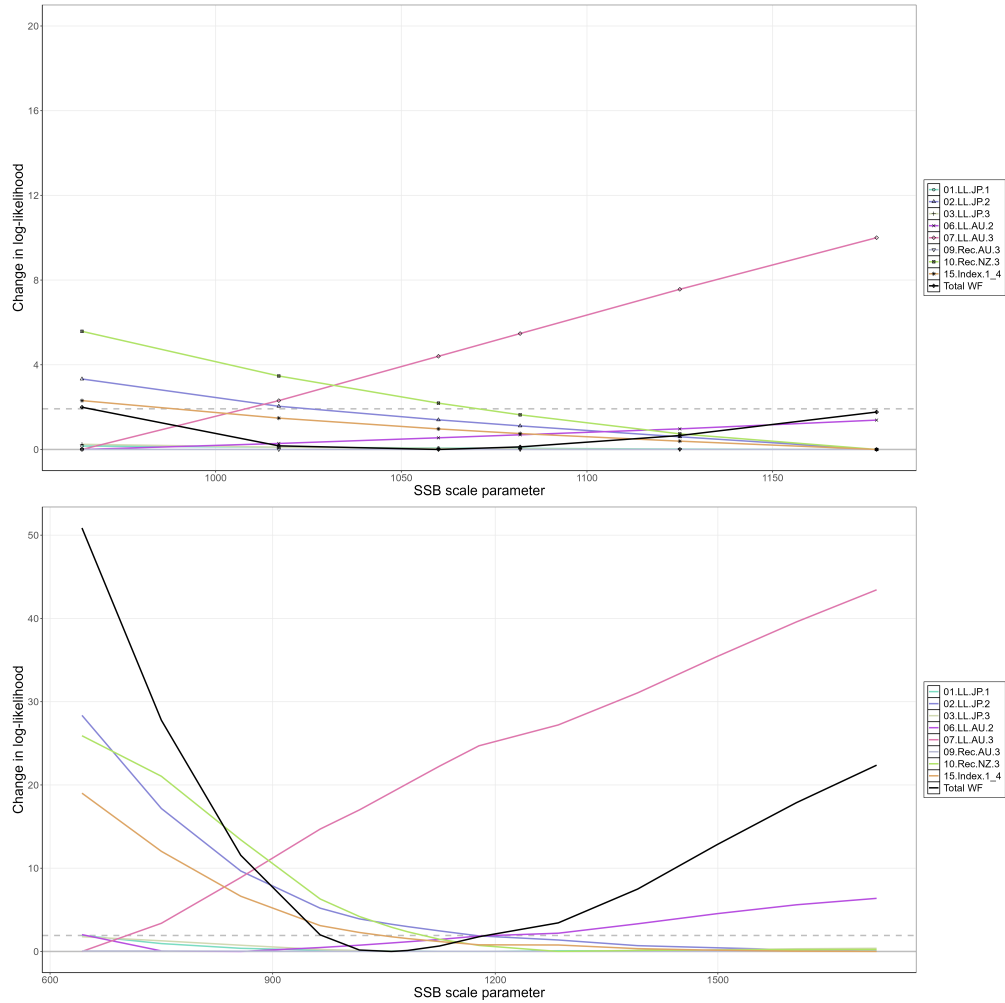
Figure 43: Likelihood profile of *SB*, showing only the weight component of the total log-likelihood separated by individual fishery. The black line indicates the total likelihood with the colours representing components of the total likelihood coming from different fisheries. The upper plot focuses on the range of total biomass values where the change in total likelihood is not statistically significant, with less than 1.92 likelihood units of change, indicated by the horizontal dashed line. The lower plot shows a broader range indicating the minima for more of the individual data sources.
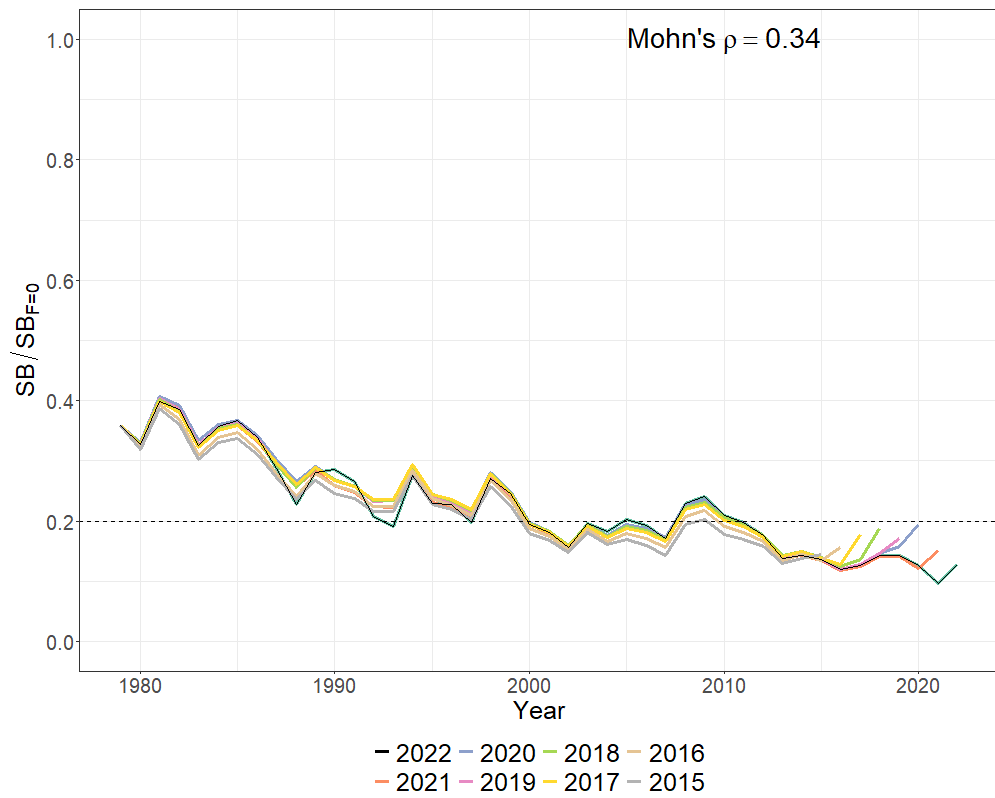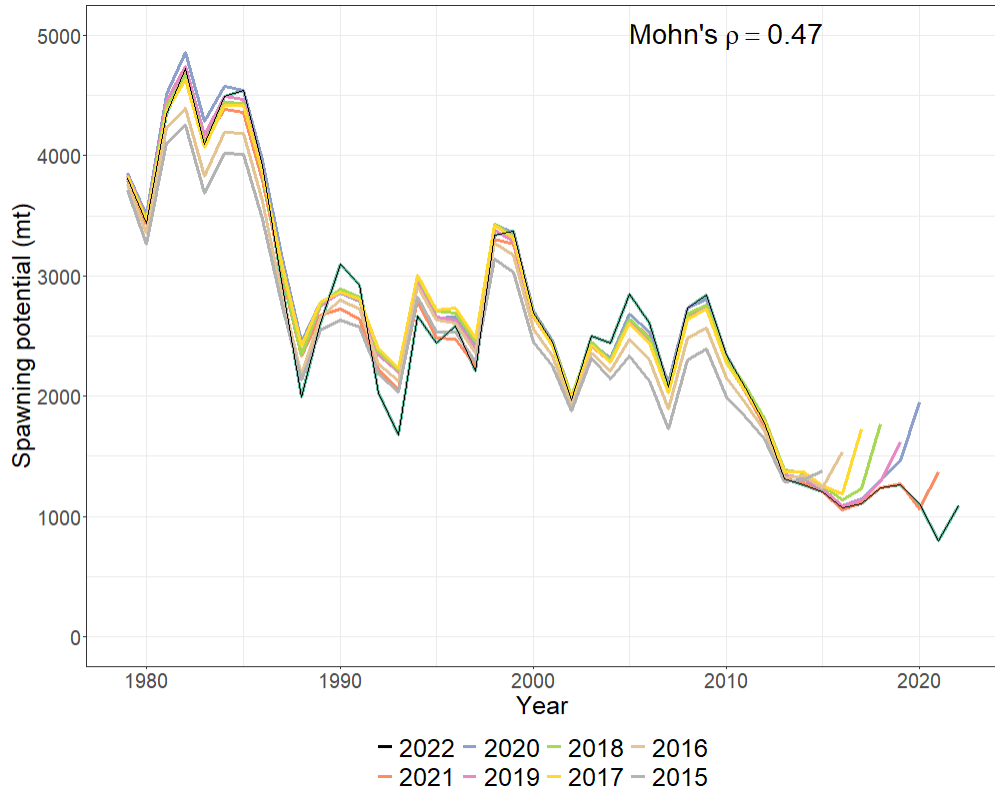
Figure 44: Retrospective analysis results for SB (top) and for $SB_{F=0}$ (bottom).
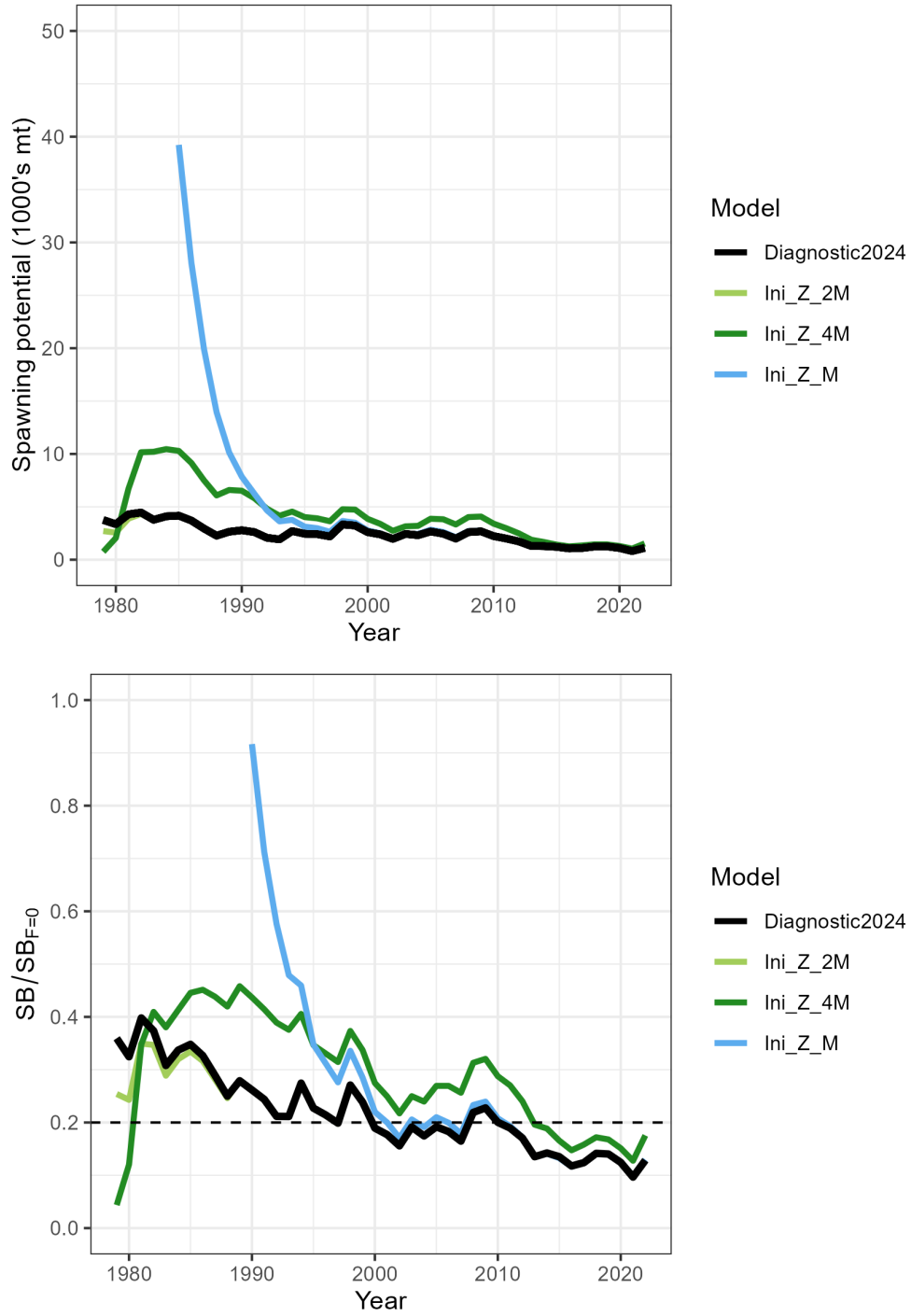
Figure 45: Sensitivity analysis to assumed initial fishing mortality for SB (top) and for $SB_{F=0}$ (bottom).
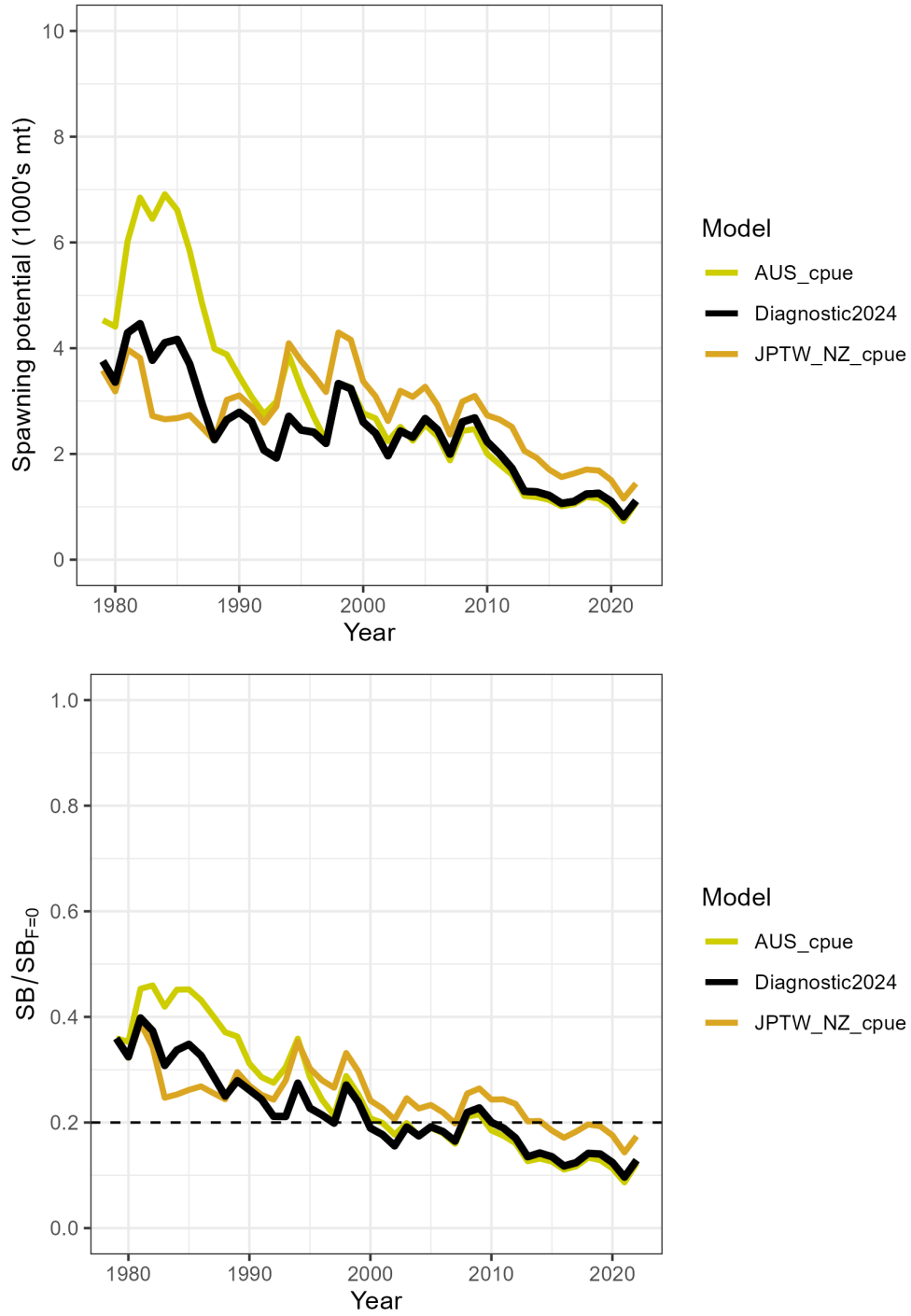
Figure 46: Sensitivity analysis to alternative CPUE series for SB (top) and for $SB_{F=0}$ (bottom).
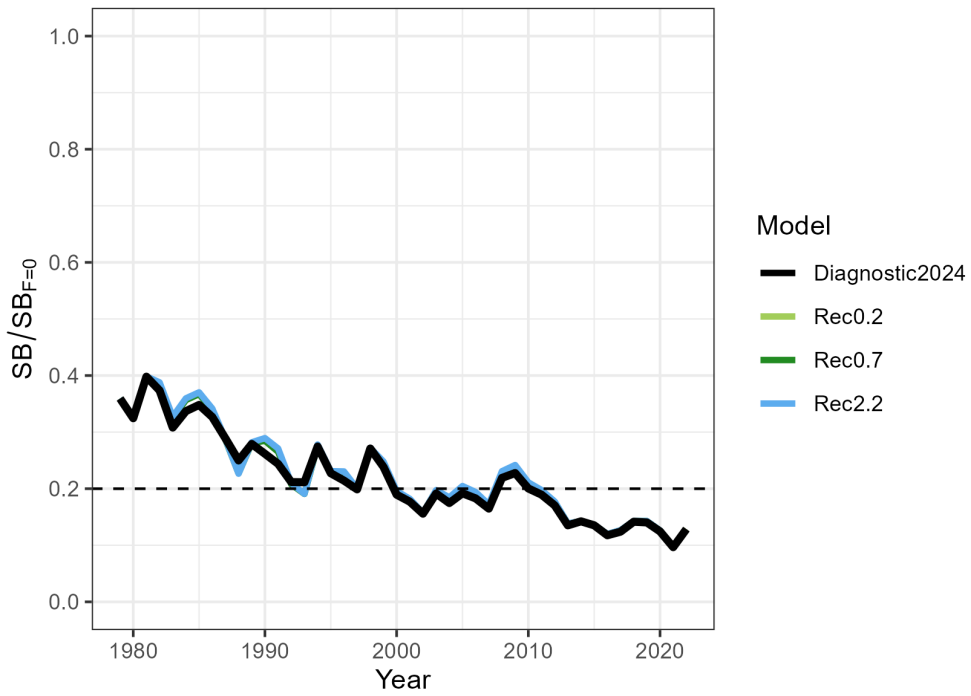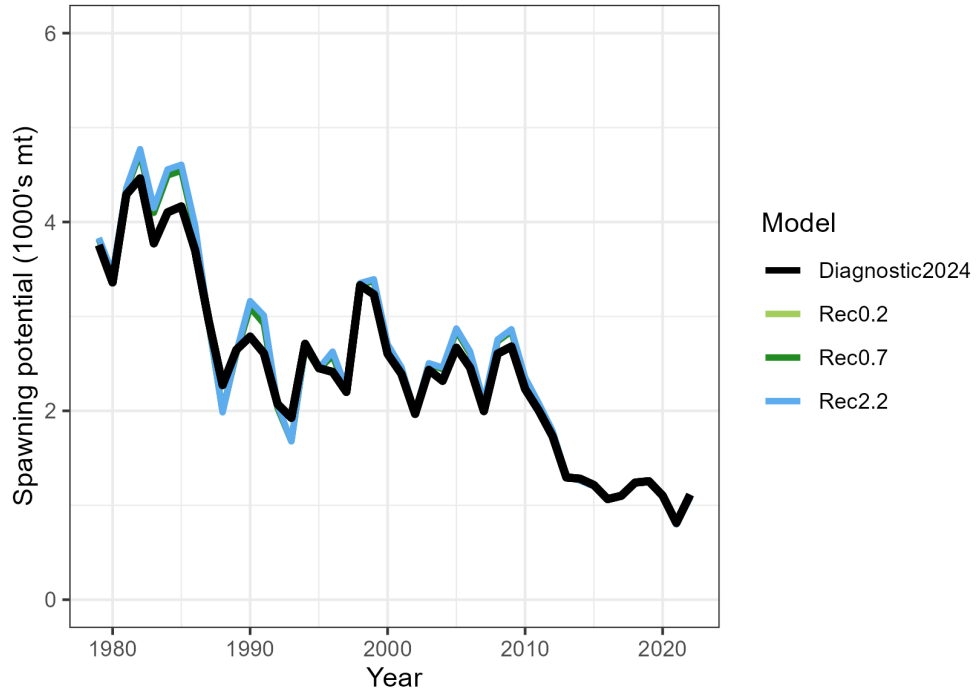
Figure 47: Sensitivity analysis to alternative recruitment variation values for SB (top) and for $SB_{F=0}$ (bottom).
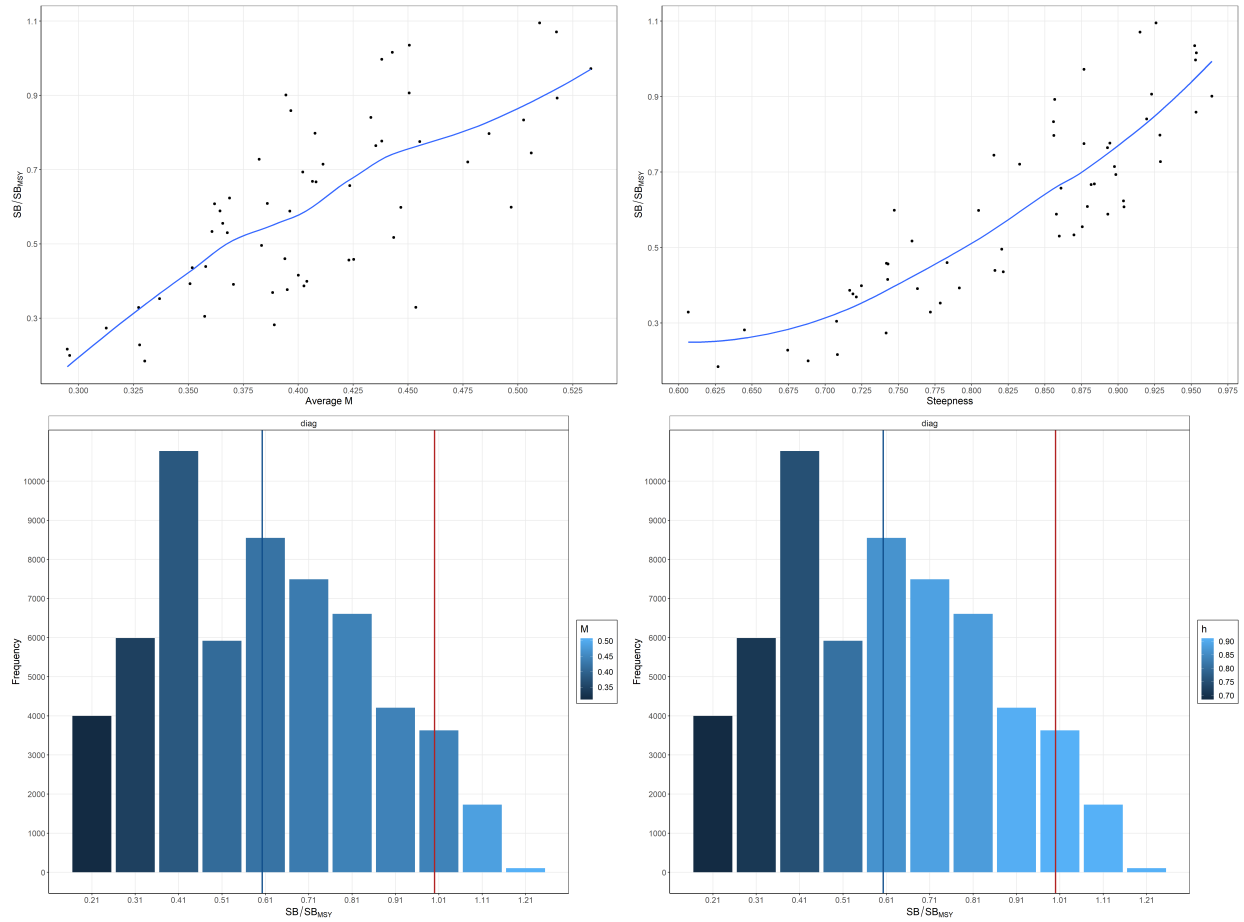
Figure 48: Histograms of Monte-Carlo estimated model uncertainty for $SB_{\text{recent}}/SB_{\text{MSY}}$ by $\bar{M}$ (top-left) and h (top-right) with mean line (blue) and $SB_{\text{recent}}/SB_{\text{MSY}} = 1$ line (red). Also includes estimated $SB_{\text{recent}}/SB_{\text{MSY}}$ by $\bar{M}$ (bottom-left) and h (bottom-right) for each model in the ensemble with a loess smoother.
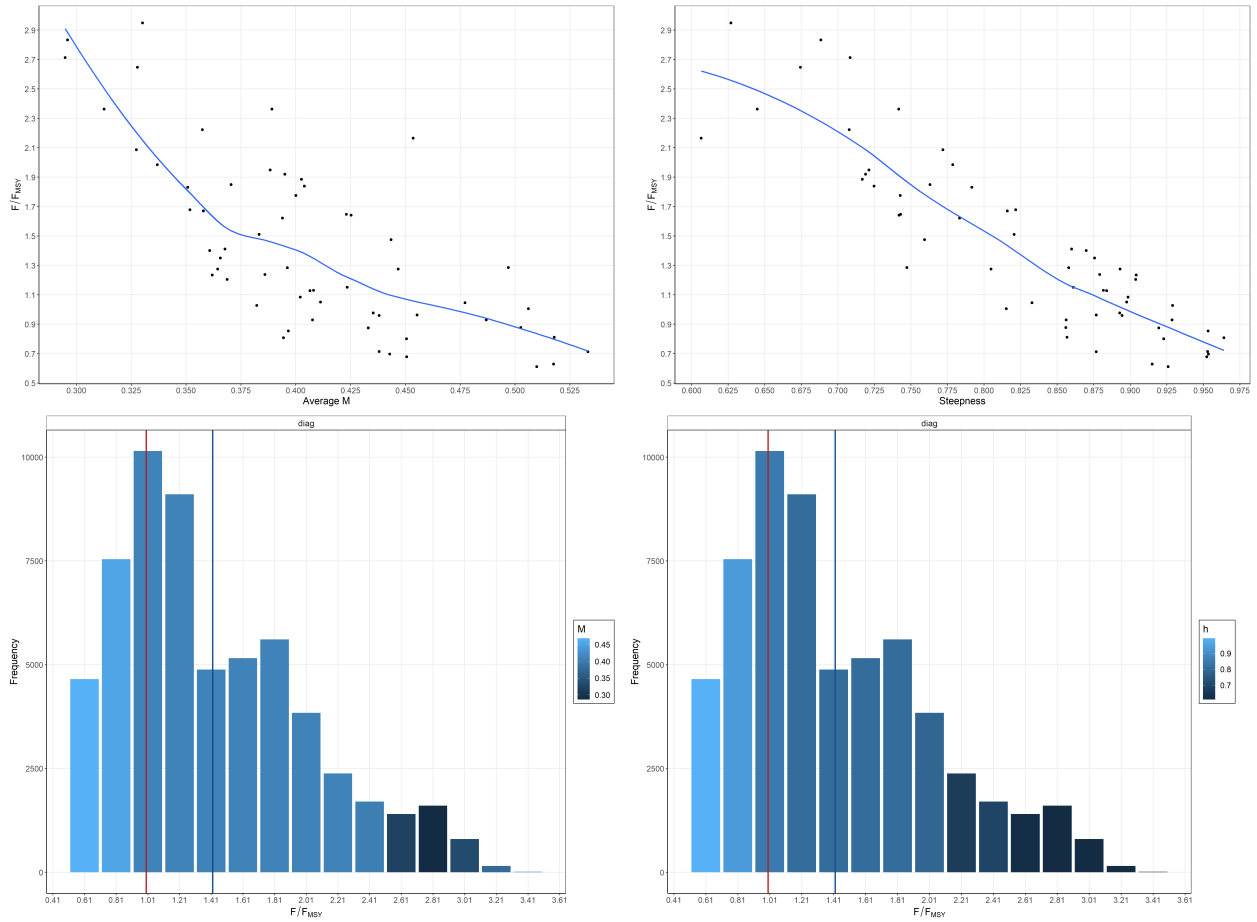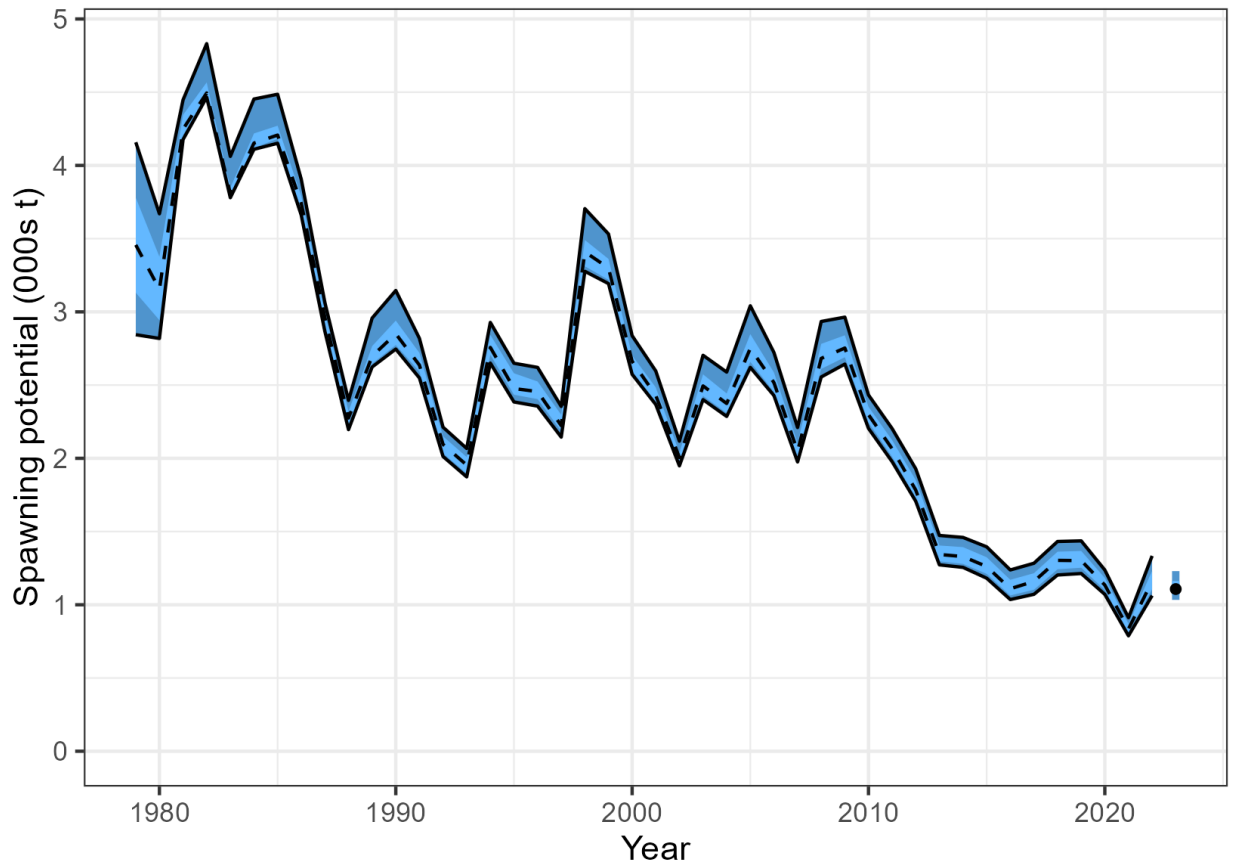
Figure 49: Histograms of Monte-Carlo estimated model uncertainty for $F_{\text{recent}}/F_{\text{MSY}}$ by $\bar{M}$ (top-left) and h (top-right) with mean line (blue) and $F_{\text{recent}}/F_{\text{MSY}} = 1$ line (red). Also includes estimated $F_{\text{recent}}/F_{\text{MSY}}$ by $\bar{M}$ (bottom-left) and h (bottom-right) for each model in the ensemble with a loess smoother.

Figure 50: Annual estimated 90% (dark blue) and 75% (light blue) quantiles of *SB* by region from the model ensemble. The dashed line indicates the median. The bar at the right of the ribbon indicates the median (black dots) and 80th percentile range for *SB*.
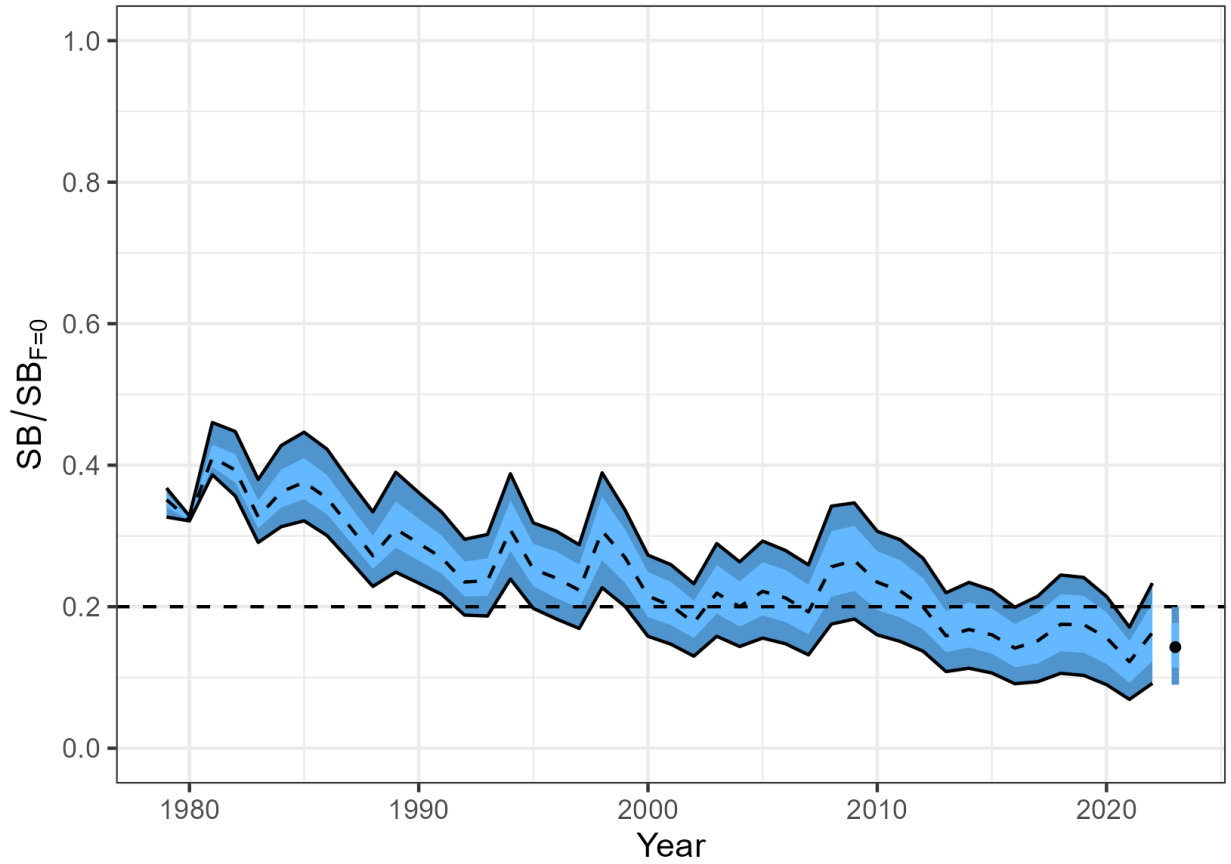
Figure 51: Annual estimated 90% (dark blue) and 75% (light blue) quantiles of $SB/SB_{F=0}$ by region from the model ensemble. The dashed line within the interval indicates the median. The bar at the right of the ribbon indicates the median (black dots) with the 10th and 90th percentiles for $SB_{\text{recent}}/SB_{F=0}$.
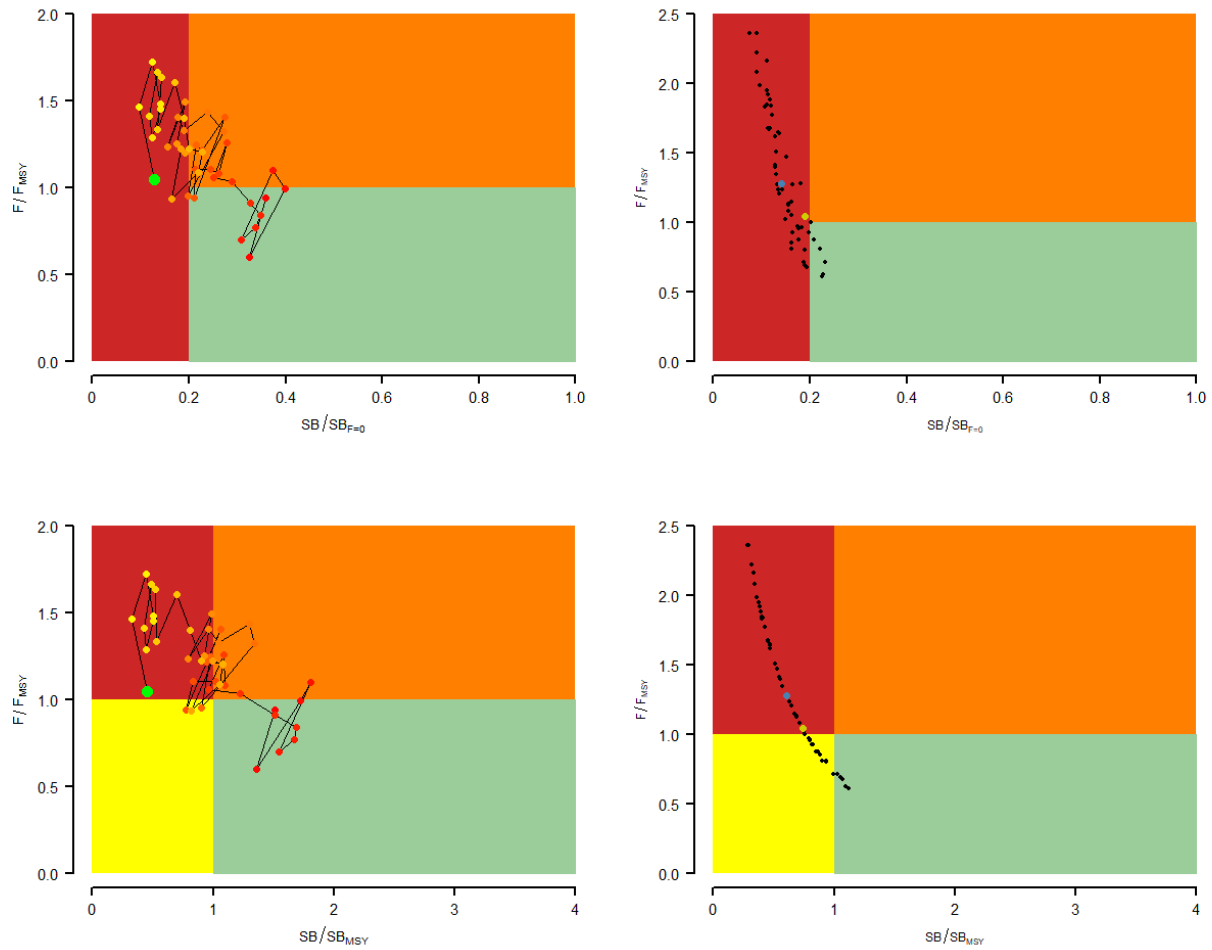
Figure 52: Majuro plots (top) and Kobe plots (bottom) summarising the results for the dynamic MSY and depletion analysis (left) and each of the models in the model ensemble for the recent period (2019–2022; right). Colours for dynamic MSY and depletion analysis transition from red to green over time. The blue point in model ensemble represents the median and the yellow point is the diagnostic model.
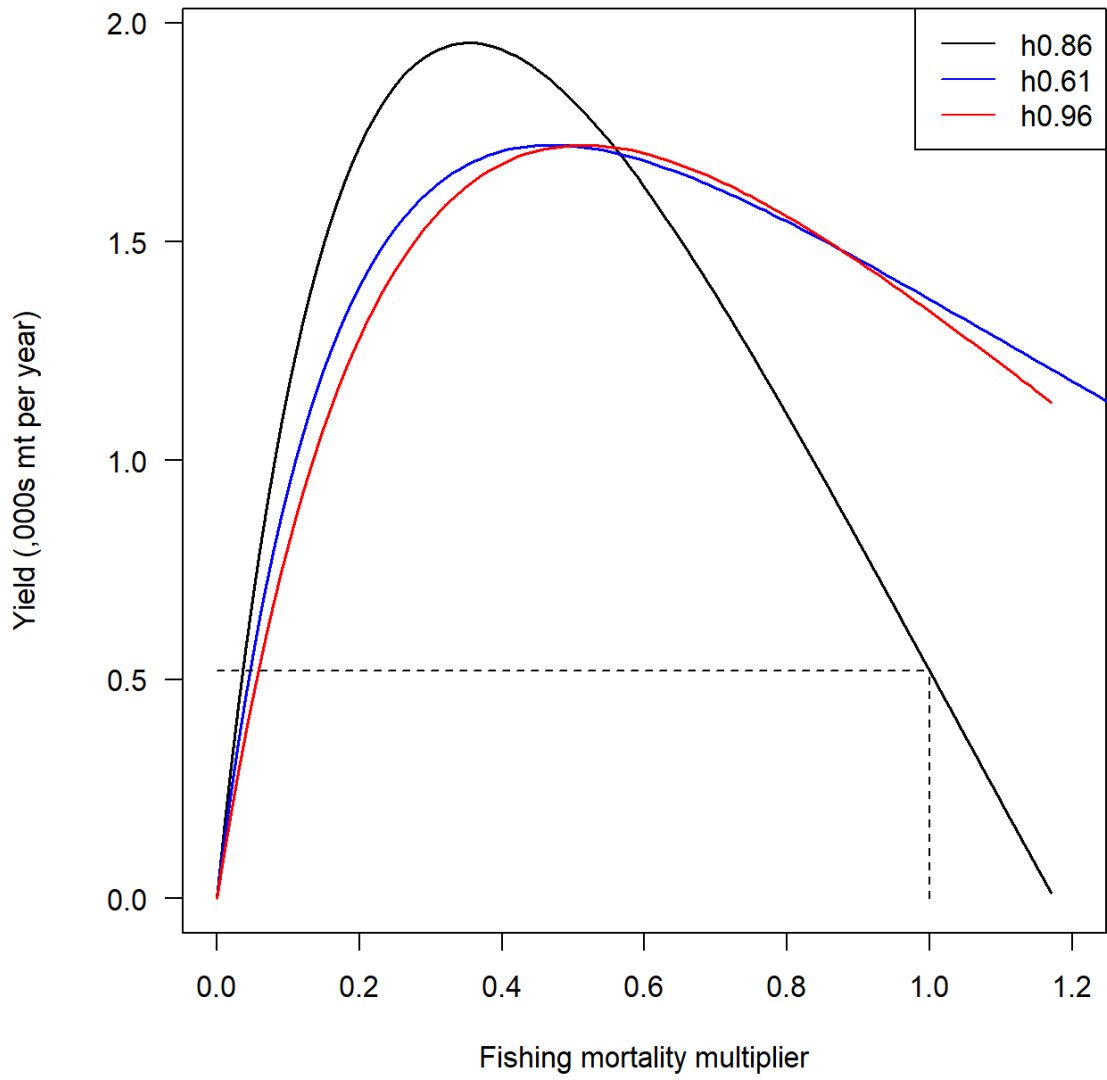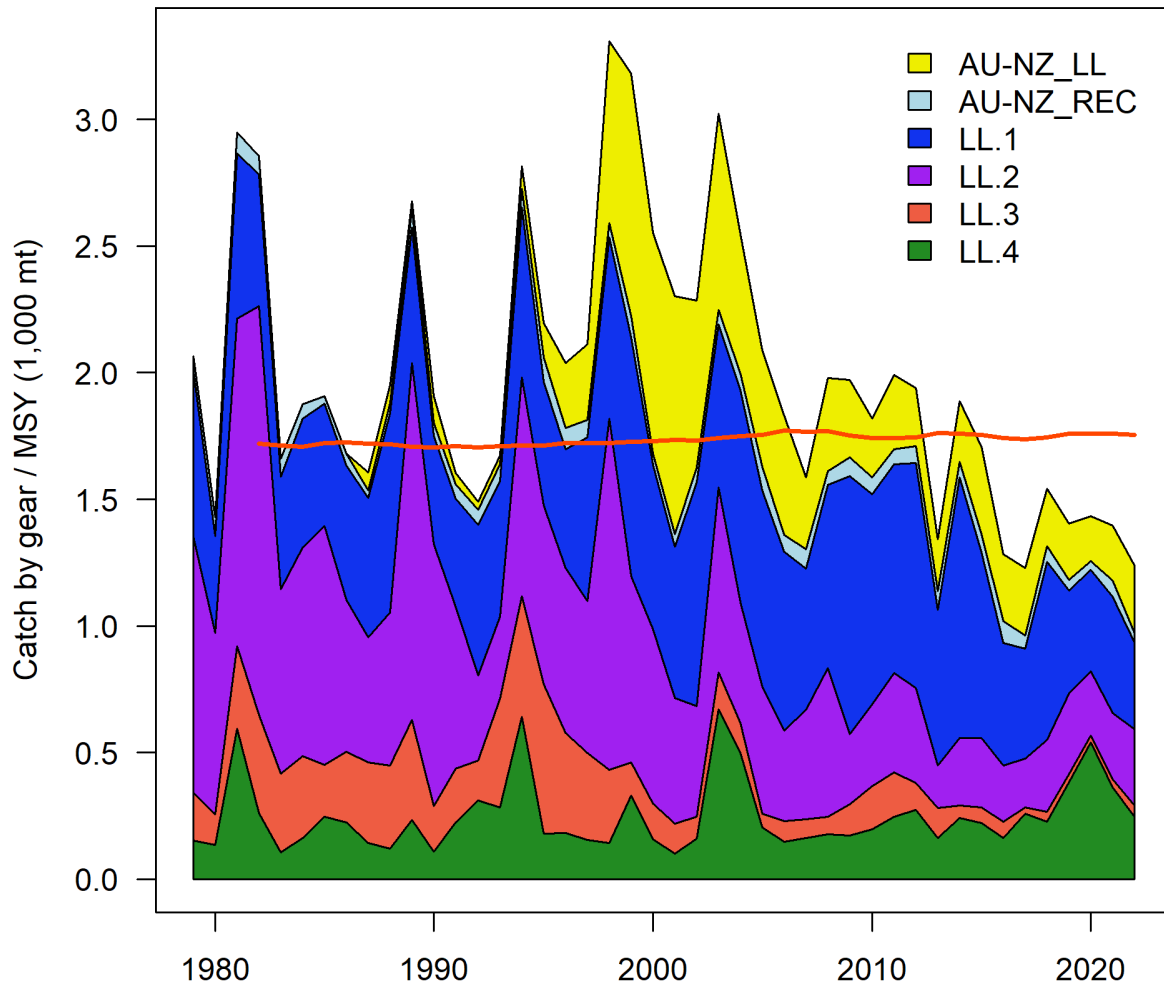
Figure 53: Yield analysis.

Figure 54: History of the annual estimates of MSY (red line) for the diagnostic model compared with annual catch by the main gear types. Note that this is a "dynamic" MSY.