SCIENTIFIC COMMITTEE
FIFTEENTH REGULAR SESSION
Pohnpei, Federated States of Micronesia
12–20 August 2019

Stock assessment of SW Pacific striped marlin in the WCPO

WCPFC-SC15-2019/SA-WP-07

N. Ducharme-Barth[1], G. Pilling[1], J. Hampton[1]

[1]Oceanic Fisheries Programme, The Pacific Community

Contents

1 Executive summary

This paper describes the 2019 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 six years of data were available since the previous assessment was conducted in 2012, with the current assessment extending the model period through 2017. The model structure was updated to reflect recommendations from the 2012 stock assessment report (Davies et al., 2012), the 2019 pre-assessment workshop (PAW) (Pilling and Brouwer, 2019), updates to the MULTIFAN-CL software (Davies et al., 2019), and to explore uncertainties in the model, particularly as related to the biological assumptions made. This assessment report is supported by Ducharme-Barth and Pilling (2019) which details the fisheries definitions, analyses of longline CPUE data and background analyses related to data inputs and model structures.

Key changes made in the progression from the 2012 reference case to the 2019 diagnostic case model include:

- Updating all data through to the end of 2017.
- Using standardized CPUE for the Japanese and Chinese Taipei longline fisheries calculated using a geostatistical model.
- Updating the biological information on maturity and defining this process as a function of length and not age.

In addition to the diagnostic case model, a number of other models were investigated (one-off sensitivities) to assess the relative impacts of key data and model assumptions on the estimated assessment results and conclusions. Based on these model runs, the input from PAW, and recommendations from previous assessments; a model grid was constructed across all combinations of the key sources (axes) of uncertainty to characterize the uncertainty in estimates of stock status relative to maximum sustainable yield (MSY) and depletion based reference points. It is recommended that management advice is formulated based on the results from the ensemble of models included in this structural uncertainty model grid.

Uncertainty in the stock status and key reference points was high, though a consensus of models indicated a clear, declining trend in stock status. As noted in the previous assessment, lack of knowledge on key biological processes (natural mortality and steepness) contributed to the overall level of uncertainty in the assessment. Three different, fixed levels were considered for the baseline level of average annual natural mortality (0.3, 0.4, and 0.5) and steepness (0.65, 0.8, and 0.95) in the structural uncertainty grid. Across grid runs, models assuming higher values for either of these two quantities generally estimated a more optimistic stock status. Appropriate levels for these values are informed by meta-analyses based on life-history theory, which generally rely heavily on the growth relationship. A high research priority should be placed on verifying the aging method used to derive the growth relationship in order to inform levels of biological uncertainty assumed in the grid.

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), persistent declines in biomass and spawning biomass were estimated since the start of the assessment period. Recent years show a slight improvement in stock status relative to a low point at the beginning of the current decade (2010s).

- The negative trend in recruitment identified in the previous two stock assessments remains a feature of the current model. Recruitment variability appears to have reduced in the last decade as spawning stock biomass has decreased.
- Fishing mortality has gradually increased over time. The rate of increase accelerated for both the juvenile and adult components in the early 2000s before peaking at the beginning of the current decade (2010s). Fishing mortality is estimated to have declined since then.
- With respect to maximum sustainable yield (MSY) based reference points, 69% of runs estimate recent spawning biomass to be less than than the spawning biomass that supports MSY ($SB_{recent}/SB_{\text{MSY}} < 1$).
- In terms of spawning biomass depletion, 50% of runs indicate that recent spawning biomass is at less than 20% of the unfished level of spawning biomass ($SB_{recent}/SB_{F=0} < 0.2$).
- With respect to fishing mortality, 56% of model runs estimate recent levels of fishing mortality to be less than the fishing mortality that would result in MSY ($F_{recent}/F_{\text{MSY}} < 1$), though only marginally less (median 0.91; range 0.03–3.50).
- This assessment concludes that SWPO striped marlin stock is likely overfished, and close to undergoing overfishing according to MSY-based reference points.

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).

Specifically in the SW 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). 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)).

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}). Recent assessments within the WCPO have indicated that the SWPO component of the stock is close to overfished but not undergoing overfishing (Davies et al., 2012) while the northern component of the stock is overfished and undergoing overfishing relative to maximum sustainable yield (MSY) based reference points (ISC-BWG, 2015, 2019).

This report describes a new assessment for SWPO striped marlin which builds upon the previous

stock assessment undertaken by Davies et al. (2012). The current assessment updates the catch, effort, and composition data through 2017. Additional changes include: using a geostatistical methods to standardize the CPUE data, redefining natural mortality as a function of age, updating the biological information on maturity and defining maturity as a function of length. Model assumptions for fisheries selectivity, and recruitment specification have been updated.

3 Background

The background material in this report largely repeats that of 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 are 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 of the 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 behavior (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 (10–30°N), June–November in the central-eastern Pacific and between November and December in the SWPO (10–30°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 within the broad longitudinal band described by Nakamura (1985) (Kopf et al., 2012). Length data from the Japanese distant water fleet (see Section 4.5) indicate that juvenile striped marlin are predominantly captured in the tropical regions of the SWPO within 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). 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). 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 mature at around 140 cm (males) - 180 cm (females) EFL (Kopf et al., 2012) and 27–40 kg corresponding to approximately 2 years of age (Skillman and Yong, 1976; Nakamura, 1985). Striped marlin live for at least 10–12 years (Melo-Barrera et al., 2003; Kopf et al., 2005, 2011) and can exceed more than 260 cm EFL and 240 kg.

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 southwest 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 estimate 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).

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 (northern and southern Pacific) model (Hinton and Bayliff, 2002). However, a number of studies have strengthened arguments for the occurrence of a semi-independent stock in the southwest Pacific. Conventional tag recapture data indicated no trans-basin movements by striped marlin tagged in the southwest Pacific, though a single individual tagged to the northeast of Hawaii with a pop-up satellite tag (PSAT) traveled 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 for 1954. 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 1). Much of this decline is attributable to lower catches in the Japanese longline fishery.

Catches of striped marlin were dominated by the Japanese longline fleet until the early 1990s (Figure 1). 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. 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 catches in the last 15 years have been offset by recent increases in catches in the Chinese longline fishery. 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 these domestic fleets. Catches by Australian longline fleets rapidly increased until 2003 due, at least in part, to specific targeting of striped marlin during some periods of the year. These catches have declined and stabilized 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 support recreational fisheries in the north of the country (Kopf et al., 2005).

Extensive recreational fisheries exist throughout the southwest 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 (tagged and) released by recreational fisheries in the WCPO (up to 60%, Holdsworth et al., 2019). However, studies into the survival of recreationally captured marlin have estimated that between 0–50% of marlin suffer post-release mortality due to hook damage, stress or increased susceptibility to predation (Pepperell and Davis, 1999). In the limited amount 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 what was considered in the 2006 stock assessment, the issues identified by Langley et al. (2006) persist in the recent years of data: low recovery rates, short times at liberty, and small movement distances (< 500km). 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 2 quarters. These recaptures would be excluded from the model under the common assumption made in WCPO tuna assessments (McKechnie et al., 2017; Tremblay-Boyer et al., 2018; Vincent et al., 2019) for the amount of time needed for tagged individuals to be well-mixed with the non-tagged population.

Though MULTIFAN-CL has the capability to integrate tagging data in the assessment model, the tagging data from the striped marlin fishery would have limited direct application in the current

assessment procedure given the issues described above. Nevertheless, the tag data are useful in consideration of the appropriate regional structure for the model. Individuals tagged off of the east coast of Australia appear to show higher residence times and lower levels of dispersion relative to fish tagged off of the North Island of New Zealand. However, given the issues identified in the previous paragraph, there was not sufficient confidence in the data to appropriately estimate sub-regional movement rates for use in a spatially explicit model. Additionally, available satellite tagging data summarized 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 were 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), though 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 4.1).

4 Data compilation

Data used in the striped marlin assessment for the SWPO consisted of fishery-specific catch and effort data, length-frequency data, and weight-frequency data. Again, much of the content in this section repeats that of Davies et al. (2012) since the fisheries definitions, and spatiotemporal model frame remain largely unchanged from the 2012 stock assessment. Catch, effort, length-frequency and weight-frequency data were updated through 2017 (Figure 2).

4.1 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 3). This represents the region of the SWPO where most striped marlin catches have been reported since 1952. 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.

The assessment modeled a single population of striped marlin within the region, assuming individuals were well-mixed within the region. However, four sub-regions were defined based on qualitative and quantitative assessments of the distribution of fishing effort and catch for the major fleets, the size composition of the catch (Figure 4) and the qualitative assessment of available tagging and recapture data (Langley et al., 2006; Davies et al., 2012). These sub-regions were used to define the spatial boundaries of the individual fisheries operating within the assessment region. 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.

Fleets 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 fleets. Japanese vessels have been the dominant fleet within sub-region 1. However, vessels from other distant water fishing nations (DWFNs) also operate within the sub-region. In addition, domestic fleets of PICTs in the region developed during the 1990s. Fleets in sub-region 1 historically reported moderate catches of striped marlin which were characterized by small individuals, but since 2003 these fleets accounted for the highest catches within the SWPO (Figure 4 and Figure 5).

The highest overall catches of striped marlin have been reported from sub-region 2 of the model region (Figure 5). Longline fleets 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 effort has declined in this sub-region (Figure 3) since the early 1990s. The Australian longline fleet and longline fleets 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 fleets 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 from sub-region 3 are currently of a similar magnitude to those from sub-region 2 (Figure 5).

Catches of striped marlin are relatively low from sub-region 4 (Figure 5). They are a commercially retained bycatch species for longline fleets 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 fleets 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 are relatively insignificant. Limited size data were available for this sub-region of the model.

4.2 Temporal stratification

Data used in the current analyses covered the period 1952–2017. 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). As a result, data were divided into quarters (January–March, April–June, July–September, and October–December).

4.3 Definition of fisheries

The sub-regions of the model region were applied to define the spatial boundaries of the specific fisheries in the SWPO. A total of 12 longline fisheries and 2 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 Ducharme-Barth and Pilling (2019).

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. Given this, 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. In order to account for this, 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 & 3 off of 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). Other recreational fisheries capturing striped marlin exist in the assessment region (Whitelaw, 2001). However, these other recreational fisheries are relatively small and catch and effort data were not readily available.

4.4 Catch and effort data

For all fisheries, catch data were expressed as the number of striped marlin captured (Table 1). For all longline fisheries, fishing effort was defined as the number of hooks set (in hundreds of hooks). For recreational fisheries, effort data were supplied as number of days. Catch and effort data for all fisheries were aggregated within the quarterly time intervals.

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 $5^\circ \times 5^\circ$ spatial resolution early in the model period, and more recently operational-level logsheet data were available for many of the longline fleets. 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 for were supplied by the National Research Institute for Far Seas Fisheries (NRIFSF) stratified by spatial cell ($5^\circ \times 5^\circ$), 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).

Fishing effort by the Japanese distant-water fleet (Fisheries 1–4) were standardized using a geostatistical delta-generalized linear mixed model (geostats) approach described in Ducharme-Barth and Pilling (2019). The geostats model included the following variables: year-quarter, sea surface temperature, target species cluster, and HBF. The resulting CPUE indices are presented in Fig-

ure 6. For each year-quarter, an index of standardized effort was calculated by dividing the total quarterly catch by the CPUE index derived from the geostats model. Estimates of time-variant precision for each standardized index were calculated internal to the model using the generalized delta method. The widest confidence intervals (95%) occurred in the early periods before gradually diminishing over time (Figure 6).

Chinese Taipei distant-water longline fishery (Fishery 5; sub-region 4): Catch and effort data for this fleet were available aggregated by $5^\circ \times 5^\circ$ cell and month. Data were supplied by the Fisheries Agency - Chinese Taipei. A geostats approach (Ducharme-Barth and Pilling, 2019) was applied similar to that described for the Japanese fisheries to derive a standardized CPUE series (Figure 6). Estimates of time-variant precision were calculated as described for Japan.

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 a delta-generalized linear model (delta-GLM) approach described in Campbell (2012, 2018). The delta-GLM model was applied to operational-level data to derive a standardized effort series for the fisheries in sub-regions 2 and 3. The delta-GLM models included the following variables: a specific function for the year/quarter/area interaction, HBF, time of setting, number of light sticks deployed on the longline, bait type, and a number of environmental factors (Campbell, 2012, 2018). The resulting CPUE indices are presented, along with the time-variant precision estimates, in Figure 6.

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 underestimated because operational data provided does not represent 100% of the vessels. Therefore, the operational data was raised based on the annual catch estimates provided (i.e. raised by total target tuna annual catch estimates / total target tuna operational data).

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. All data for these other fisheries were supplied as logsheet data and/or aggregated spatial data, with nominal effort and catches raised as appropriate.

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 last assessment were unavailable. However, an updated CPUE index was calculated through 2017 from tournament data using the method described by Ghosn et al. (2012) and is shown in Figure 6 along with the estimates of time-variant precision.

New Zealand recreational fishery (Fishery 10, sub-region 3): Information for this fishery was supplied by Holdsworth et al. (2019) with catch data for the period 1952–2017 and standardized CPUE indices from 1975-2017. Data were obtained from extensive fishing club records. The GLM index is presented in Figure 6 along with the estimates of time-variant precision.

4.5 Length and weight-frequency data

Length-frequency and/or weight-frequency 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 (Figure 7).

Length data were provided based on four different length measurement methods: eye orbit–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 the AFMA represented the most extensive collection of size composition data used in this analysis (approximately 3,000 length measurements and 65,000 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 (Ghosn et al., 2012). All weights were supplied 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.

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 fishery 11 beginning in 1996; 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 time strata (year/quarter). Length data were aggregated into 52 6-cm size classes (20–326 cm EFL) consistent with the 2012 stock assessment. Weight data were aggregated into 123 2-kg intervals (5–249 kg whole weight), which is a coarser stratification than that used in 2012 (246 1-kg intervals). Length or weight data were not available for all quarters for the period of data supplied for each fishery (Figure 7). 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 JP 2 LL, JP 3 LL, AU 2 LL, AU 3 LL, and the NZ 3 REC fisheries; and those with comprehensive length-frequency data available included OTHER 1 LL and OTHER 2 LL (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 (Figure 4). 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 the two previous assessments (Langley et al., 2006; Davies et al., 2012), 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. For fishery 14 in the period 1995 - 2001 samples consisted of only small fish, which was clearly atypical 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 atypical of 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 the AU 3 REC fishery 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.

5 Model description

5.1 General characteristics

As with any model, various structural assumptions were made in the SWPO striped marlin stock assessment. Such assumptions are a trade-off between the need to keep the parameterization as simple as possible, and to allow sufficient flexibility so that important characteristics of the fisheries and population are captured in the model. This section summarizes the key structural assumptions made in each component of the assessment model: the population dynamics of the fish stock, the dynamics of the fishery, observation models for the data (likelihood components), and the parameter estimation procedure. Detailed technical descriptions of these components are given in Hampton and Fournier (2001) and Kleiber et al. (2019). Ancillary analyses for the interpretation of the stock assessment results are also described.

5.2 Population dynamics

The model partitions the population into 10 annual age-classes with the last age-class consisting of a 'plus group' in which mortality and other characteristics are assumed to be constant. The main population dynamics processes which are described below are assumed to occur at an annual time step beginning in 1952 and running through 2017.

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 spatiotemporal differences in the size structure of the catch, a sub-regional structure was used to define the fisheries. This fleets-as-regions approach allowed the flexibility to estimate specific size-based selectivity functions to match the pattern of observations in each sub-region. Additionally, seasonal and spatial variations in the catch rates between fisheries was accounted for by estimating fishery-specific catchability parameters incorporating seasonal variation. A one-off sensitivity analysis investigated the effect of dividing the assessment region into 2 model regions at 165°E (Section 6.2.8).

While there is evidence to suggest that a degree of sexual dimorphism exists for striped marlin (Kopf et al., 2011, 2012), a sex-aggregated model was used in the assessment. A one-off sensitivity analysis using a sex-disaggregated model was conducted to examine the impact of sexually dimorphic growth and weight-length relationships (Section 6.2.7).

Following the research activity at the start of the current decade (Piner and Lee, 2011; Kopf et al., 2011, 2012), very little progress has been made to address the key biological uncertainties for SWPO striped marlin raised by the 2012 assessment. As a result, the biological assumptions relating to the population dynamics made in the current assessment (Figure 8) are mostly unchanged, apart from improvements to their implementation in MULTIFAN-CL and the inclusion of new biological information on maturity, from those made by Davies et al. (2012).

5.2.1 Recruitment

Recruitment in terms of the MULTIFAN-CL model is the appearance of age-class 1 fish 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 be an annual event that

occurred 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. The previous assessment assumed recruits entered the fishery at the beginning of the first quarter (January) following a one year lag. Sensitivity to the timing of when recruits entered the model was considered during the stepwise development of the diagnostic case (Section 6.1).

Recruitment was assumed to have a relationship with spawning biomass via a Beverton and Holt stock-recruitment relationship (SRR) with a fixed value of steepness (h). Steepness is defined as the ratio of the equilibrium recruitment produced by 20% of the equilibrium unexploited spawning 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 the steepness parameter was fixed at a moderate value (0.80) and the sensitivity of the model results to the value of steepness was explored by setting it to lower (0.65) and higher (0.95) values (Section 6.3). 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: 80^{th} percentile). 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 stock assessments (McKechnie et al., 2017; Tremblay-Boyer et al., 2018; Vincent et al., 2019) 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 the previous stock assessment for striped marlin (Davies et al., 2012) 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 2012 for the diagnostic case, however the sensitivity to this assumption is accounted for in the structural uncertainty grid (Section 6.3).

5.2.2 Initial population

The population age structure in the initial time period was assumed to be in unfished equilibrium and determined as a function of the average natural mortality used in the model. This avoids having to treat the initial age structure, which is generally poorly determined, as independent parameters in the model. Though the assumption of starting the age structure at the unfished equilibrium appears reasonable given that industrial fishing was largely absent from the SWPO prior to the model start (the global Great Depression and World War II limited the development of fishing fleets, and the post-war MacArthur lines limited Japanese fishing activity to the northwest Pacific prior to 1952 (Ahrens, 2010)), a one-off sensitivity analysis explored beginning the model at a fished state in 1955 (Section 6.2.1).

5.2.3 Growth

MULTIFAN-CL makes the following structural assumptions regarding age and growth: lengths-at-age are normally distributed for each age-class, mean lengths-at-age follow a von Bertalanffy (VB) growth curve, the standard deviations of length for each age-class are a log-linear function of the mean lengths-at-age, and the probability distributions of weights-at-age are a deterministic function

of the lengths-at-age and a specified weight-length relationship. These processes are assumed to be temporally invariant.

VB growth is parameterized in MULTIFAN-CL using the length-at-first-age (L1), length-at-last-age (L2), and K as leading parameters. Following the previous assessment, growth was fixed in the 2019 diagnostic case using to the values (L1 = 121.03 cm EFL; L2 = 220.53 cm EFL; K = 0.4494) from the "back-calculated 1" growth model averaged for both sexes from Kopf et al. (2011). While the mean lengths-at-age were fixed at the published values, the standard deviations of lengths-at-age were estimated when fitting the population model. Preliminary results from otolith aging of striped marlin conducted by CSIRO indicate that striped marlin may live longer than Kopf et al. (2011) estimated from dorsal fin-spines. An alternative growth hypothesis based on this new research was considered as a part of the structural uncertainty grid (Section 6.3).

Weight-at-age (defined as $W_a = \alpha L_a^{\beta}$) was fixed at the values used in the previous assessment: $\alpha = 4.4990 \times 10^{-7}$ and $\beta = 3.6165$.

5.2.4 Natural mortality

The previous stock assessment assumed a value of 0.4 for all ages based on the mean of the age-specific natural mortality for striped marlin calculated in the Piner and Lee (2011) meta-analysis. In the current diagnostic case, the age-specific natural mortality values from (Piner and Lee, 2011) are used instead of the average value of 0.4. Natural mortality (M) is considered to be a key source of uncertainty, therefore alternative scenarios for M were considered in the structural uncertainty grid (Section 6.2.4 & Section 6.3).

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 current assessment revises what was previously done in the 2012 stock assessment and updates the maturity information regarding SWPO striped marlin to be consistent with the findings of Kopf et al. (2012). The age-at-maturity ogive in the previous assessment was based on studies from both the northern WCPO and SWPO (Skillman and Yong, 1976; Hanamoto, 1977). The current assessment also takes advantage of a relatively new feature in MULTIFAN-CL allowing for the relative reproductive potential to be specified by length class, with conversion to maturity-at-age done internal to the model (Davies et al., 2019). 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 Kopf et al. (2012). Further details on the calculation of the length specific maturity ogive used in the current assessment can be found in Ducharme-Barth and Pilling (2019).

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 effort deviations which allow for random variation in the fishing effort - fishing mortality relationship. 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 case, 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 2012 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 Davies et al. (2012), 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 each sub-region (i.e. fisheries 11–14) were assumed to have equivalent selectivities to the corresponding Japanese longline fisheries. An exception to this is the OTHER 2 LL fleet (fishery 12; sub-region 2) which had enough composition data to be estimated independently. Very limited size data were available from the TW LL 4 fishery and the selectivity was assumed to be equivalent to the JP LL 4 and OTHER 4 LL fisheries, although limited size data were available from either fishery. Separate cubic spline selectivities were estimated for the two Australian longline fisheries. The selectivity of the New Zealand longline fishery was assumed the same as the Australian longline fishery in sub-region 3. With the exception of fishery 12, these groupings are consistent with those used in Davies et al. (2012). The selectivity for the Australian recreational fishery was not grouped with any of the others and was estimated using a cubic spline.

The selectivity for the New Zealand recreational fishery (fishery 10; sub-region 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.

5.3.2 Catchability

Catchability was assumed to be time invariant for the fishery with the standardized CPUE time series that the model was being fit to (the "index fishery" for the model[2]). This assumption is similar to assuming that the CPUE for this fishery indexes the exploitable abundance over time. For the diagnostic case, the index fishery was JP 2 LL, the same that was used in the 2012 stock assessment.

For all other fisheries, catchability was allowed to vary slowly over time (akin to a random walk) using a structural time-series approach. Random walk steps were taken every two years, and the deviations were constrained by a penalty with mean zero and standard deviation of 0.1. Seasonal

[2]Note that this terminological definition of an index fishery is different from Tremblay-Boyer et al. (2018).

variation in catchability was also allowed to explain the strong seasonal variability in CPUE for most of the fisheries.

Given that the indices showed slightly different trends for the 9 fisheries where standardized CPUE indices were provided, sensitivity to the choice of the index fishery was investigated in Section 6.2.3 and accounted for in the structural uncertainty grid (Section 6.3).

5.3.3 Effort deviations

Effort deviations, constrained by penalties with mean zero and a specified variance, were used to model the random variation in the effort–fishing mortality relation. There were three categories of fisheries with respect to the effort deviation penalties applied, these are outlined in Table 3 and presented in Figure 9. The "index fishery" described in Section 5.3.2 was the principal index fit to in the model and had a time varying CV (as estimated using by the geostats model described in Ducharme-Barth and Pilling, 2019) rescaled to a mean of 0.2. This was then transformed to an effort deviate penalty corresponding to each observation for the fishery input into MULTIFAN-CL. Effort deviate penalties in MULTIFAN-CL are inversely related to the variance, such that lower effort penalties are associated with indices having high variance and thus have less influence on the model fit.

Fisheries that received standardized CPUE but were not selected as the "index fishery" (Fisheries 1, 3-7, and 9-10) received effort deviate penalties with a mean CV of 0.5, and were given less influence on the model fit. Additionally, the penalties of these fisheries were rescaled with respect to the square root of the standardized effort (and thus made time varying) so that lower observed effort received lower penalties than instances of high effort. Though time varying penalty weights (as calculated from the estimated variance from their respective standardized indices) were available for all fisheries with standardized effort, they were only used for Fisheries 5 and 6, and only when these were used as the index fishery in the structural uncertainty grid.

Finally, fisheries where the nominal effort was used (Fisheries 8 and 11-14) were determined to have the least influence on the model fit and received effort deviate penalties equivalent with a mean CV of 0.71. Again, these penalties were rescaled with respect to the square root of the nominal effort.

5.4 Likelihood components

There are 3 main data components that contribute to the overall log-likelihood objective function that is being optimized by the model: the total catch likelihood, the length composition likelihood, and the weight composition likelihood. The observed total catch data are assumed to be unbiased and relatively precise, with a residual CV of 0.007.

The probability distributions for the length and weight-frequency proportions are assumed to be approximated by robust normal distributions, with the variance determined by the effective sample size (ESS) and the observed length and weight-frequency proportion. The size frequency samples receive an ESS lower than the number of fish measured in recognition that size composition samples are not truly random (fish tend to school in similar sizes so multiple samples from a single school would not represent independent samples with respect to size) and that not all process error is accounted for in the model (the observed sample variability is likely an underestimate of true variability).

Effective sample size for the weight frequency samples was assumed to be 5% of the actual sample size for all fisheries, with a maximum effective sample size of 50. Most of the WCPO tuna assess-

ments fit to either the weight or the length frequency data but not usually 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 types of composition data were included. Consistent with the previous assessment, length frequency data were assumed to have an ESS of 2.5% of the observed sample size. Alternative specifications for the effective sample size were included in the structural uncertainty grid (Section 6.3).

For fisheries 6-8, the length frequencies were further down-weighted. Individuals sampled by the Australian longline fisheries (6 & 7) showed a discrepancy in the fit to the length composition data versus the weight composition data. Given that there was substantially more weight composition data for these two fisheries (Table 2), the length composition data for these fisheries was down-weighted (ESS of 1% of the observed sample size) so as to not interfere with the model fit. A one-off sensitivity model was run to explore this assumption (Section 6.2.5). The length-frequency dataset from the New Zealand longline (fishery 8) had fewer than 50 total observations. Given that this fishery was assumed to share selectivity patterns with fishery 7, these length samples were down-weighted to 1% of the effective sample size.

5.5 Parameter estimation & Uncertainty

The parameters of the model were estimated by maximizing the log-likelihood of all data components plus the log of the probability density functions of the penalties specified in the model. The maximization to a point of model convergence was performed by an efficient optimization using exact derivatives with respect to the model parameters (Fournier et al., 2012). Estimation of model parameters was done in phases by sequentially activating groups of "flags" that initialized different components of the model. A bash shell script (referred to as the *doitall* in Kleiber et al., 2019) was used to implement and document the phases used in the model fit. Model convergence was assumed if all estimated parameters had gradients less than 10^{-3}. Quality of fit was further assessed by inspecting the likelihood profile with respect to the total average population biomass, and routine stock assessment diagnostics.

An additional diagnostic, the retrospective analysis, was used to assess the overall stability of the model and to identify any persistent biases in estimated quantities as a result of possible model mis-specification. A robust and well-specified model should produce similar estimates when data from the terminal year (2017-1998) is sequentially removed (Mohn, 1999; Cadigan and Farrell, 2005). The results of the retrospective analysis for the 2019 diagnostic case model are contained in the Appendix (Section 12.1).

Two methods were used to assess the level of uncertainty in the estimated quantities. The first deals with the within model statistical uncertainty of a single model, while the second is concerned with structural uncertainty across a range of models, as a result of the suite of assumptions made when specifying the assessment model. In this case, the within model statistical uncertainty was estimated by first calculating the Hessian matrix for the diagnostic case to obtain the model covariance matrix. This is used in combination with the delta method to compute approximate confidence intervals for parameters of interest (e.g. spawning biomass or recruitment). In the second case, a factorial grid of models was used to quantify the uncertainty stemming from different structural assumptions made during model development. Selection of axes and axis levels was based on the results of the one-off sensitivity analyses and the findings from previous assessments. The uncertainty around management reference points is derived from the estimates from all models considered in the structural uncertainty grid.

5.6 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).

5.6.1 Yield analysis

The yield analysis consists of computing equilibrium catch (or yield) and biomass, conditional on a specified basal level of age-specific fishing mortality (F_a), a series of fishing mortality multipliers ($fmult$), the natural mortality-at-age (M_a), the mean weight-at-age (W_a) and the SRR parameters. All of these parameters, apart from $fmult$, 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 $fmult$ can easily be determined using the formulas given in Kleiber et al. (2019), and is equivalent to the MSY. Similarly the spawning biomass at MSY (SB_{MSY}) can also be determined. The ratios of the current (or recent average) levels of fishing mortality and biomass to their respective levels at MSY are of management interest as reference points.

Fishing mortality-at-age (F_a) for the yield analysis was determined as the mean over a recent period of time (2013-2016). The terminal year (2017) was not included in the average as fishing mortality tends to have high uncertainty for the terminal data year of the analysis and the terminal recruitment in this year is constrained to be the geometric mean over the full time-series, which affects F for the youngest age-classes.

A time dynamic MSY was also computed using the average annual F_a from each year included in the model (1952-2017). This enabled temporal trends in MSY (and MSY based reference points) to be assessed and a consideration of the differences in MSY levels under historical patterns of age-specific exploitation.

5.6.2 Depletion and fishery impact

Fishery depletion is assessed by computing the unexploited spawning biomass (or total biomass) time series using the estimated model parameters, and assuming that fishing mortality was zero. Since both the fished estimate of spawning biomass, SB_t, and the unexploited spawning biomass $SB_{F=0,t}$, incorporate recruitment variability, their ratio at each time step $SB_t/SB_{F=0,t}$, can be interpreted as an index of fishery depletion. There are two ways of defining the level of recruitment used by MULTIFAN-CL for the $SB_{F=0,t}$ calculation: either at the absolute level of recruitment or at a level scaled according to the SRR (Berger et al., 2013). In the absolute level case, the recruitments in the unfished condition are assumed equivalent to the estimated recruits in the fished condition. In the SRR case, the recruitment levels for the unfished stock were rescaled by adding the recruitment deviates to the level of recruitment predicted by the SRR. This rescaling acknowledges the possibility that recruitment may be reduced in exploited populations. Based on the recommendations made by SC 9 (WCPFC, 2013), current WCPO assessments calculate $SB_{F=0,t}$ using the SRR to adjust the recruitments. This recommendation was incorporated into the current assessment during the stepwise development of the diagnostic case model.

The depletion associated with specific groups of fisheries was assessed in a similar manner. The groups of interest were "turned-off" in separate simulations. The resulting changes in depletion from these runs was indicative of the depletion attributed to each of the fisheries groups that were removed from the analysis. For the 2019 striped marlin stock assessment 6 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 & 3 combined, and the combined recreational fisheries for Australia and New Zealand.

5.6.3 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 formerly 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 $SB_{recent}/SB_{\text{MSY}}$, $SB_{latest}/SB_{\text{MSY}}$, and $F_{recent}/F_{\text{MSY}}$ (Table 4). SB_{recent} and SB_{latest} are the mean of estimated spawning biomass over the period 2014-2017 and the estimate in the last year of the model (2017), respectively. This is consistent with decisions made at SC13. The other key reference point, $F_{recent}/F_{\text{MSY}}$, is the estimated average fishing mortality over a recent period of time (F_{recent}: 2013-2016 for the current assessment) divided by the fishing mortality producing MSY (as described in Section 5.6.1).

Depletion based reference points, as used for the tropical tuna species, are also provided. Stock status is referenced against these points by calculating $SB_{recent}/SB_{F=0}$ and $SB_{latest}/SB_{F=0}$, where $SB_{F=0}$ is calculated as the average over the period 2007-2016 (Table 4).

5.6.4 Kobe and Majuro plots

For the standard yield analysis (Section 5.6.1), the fishing mortality-at-age, F_a, is determined as the average over some recent period of time (2013-2016 for the current assessment). In addition to this approach, the MSY-based reference points (F_t/F_{MSY}, and SB_t/SB_{MSY}) and the depletion-based reference point ($SB_t/SB_{F=0,t}$) were also computed using the average annual F_a from each year included in the model (1952-2017) by repeating the yield analysis for each year in turn. This enabled temporal trends in the reference point variables to be estimated, accounting for differences in MSY levels under varying 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 WCPO stock assessments in recent years.

6 Model runs

6.1 Model developments from the 2012 assessment

Model development from the 2012 reference case to the 2019 diagnostic case occurred incrementally via successive changes. These stepwise changes were done in order to identify the impact of each modification to the assessment outcomes. Changes made to the previous assessment model include: additional input data for the years 2012-2017, the use of geostats methods to standardize the CPUE data (Ducharme-Barth and Pilling, 2019), redefining natural mortality as a function of age, updating the biological assumptions regarding maturity, and defining maturity as a function of length. Additionally, model assumptions for fisheries selectivity, and recruitment specification have been updated. These changes occurred in the following sequence of steps:

– The 2012 reference case model

- **Step 2** [*new MFCL*]: Upgrade the MULTIFAN-CL executable to version 2.0.6.1 (Davies et al., 2019).
- **Step 3** [*2017 data*]: Update the assessment model with data through 2017, keeping the assumptions and methods used in 2012.
- **Step 4** [*2kg wt*]: Aggregate the weight composition data from 1 kg bins to 2 kg bins as inspection of the composition data appeared to indicate that data were reported at the 2kg bin size for certain fisheries.
- **Step 5** [*geostats*]: Replace the standardized effort for fisheries 1-5 with the standardized effort calculated using the geostats CPUE indices from Ducharme-Barth and Pilling (2019).
- **Step 6** [*AU LF*]: Down-weight the length-frequency data for fisheries 6-7.
- **Step 7** [*split sel. 2 & 12*]: Estimate separate selectivities for fisheries 2 and 12.
- **Step 8** [*M@age*]: Define M as a function of age based on the Piner and Lee (2011) meta-analysis by providing fixed values for the age-specific deviations from M.
- **Step 9** [*maturity@length*]: Update maturity and define as a function of length according to Davies et al. (2019) and as calculated in Ducharme-Barth and Pilling (2019).
- **Step 10** [*free sel.*]: Remove non-decreasing penalty on the selectivities for all fisheries except for fishery 10.
- **Step 11** [*rec. timing*] Re-specify the timing for when recruits enter the fishery.
- **Step 12** [*SRR def*] Fix the terminal two recruitment periods to the mean recruitment level and exclude them from the calculation of the SRR.
- **Step 13 - 2019 Diagnostic Case** [*depletion def.*]: Update the method for how unfished biomass is calculated in the fisheries impacts analysis in order to be in line with the recommendations from WCPFC (2013).

6.2 Sensitivity analyses

During the course of model development for the 2019 striped marlin stock assessment several hundred models were run to explore the effects of changing the assumptions governing the population dynamics, fisheries dynamics, parameter estimation and weighting of data components in the likelihood. The presentation of the results focuses on the subset of analyses that were most influential to the stock assessment outcomes and/or those that were identified as questions of interest either in the previous assessment or the 2019 pre-assessment workshop (PAW; Pilling and Brouwer, 2019). From this subset and based on the findings of previous WCPO stock assessments, models were selected for inclusion in the structural uncertainty analyses. This process entails running a full-factorial set of models where all combinations of key structural assumptions are explored (Section 6.3).

One-off sensitivities were conducted as single stepwise changes from the 2019 diagnostic case. The purpose of these sensitivity runs was not to provide absolute estimates of management quantities but to assess the *relative* change that resulted from the various assumptions.

6.2.1 Model start year

Reliability of catches in the early model period (particularly the high catch event by the JP 2 LL fishery in 1954) was voiced as a concern at the PAW, especially if they were strongly influential to model estimates. These data were scrutinized both internal to SPC and with the NRIFSF and were deemed acceptable for inclusion in the assessment. However, a one-off sensitivity was undertaken to assess the effect on model outputs of excluding pre-1955 data and beginning the model in 1955. Given that substantial catches were believed to have occurred, the model was assumed to begin in a fished condition in 1955. Two scenarios were explored for determining the age-structure of the initial population: age-structure was assumed to be in equilibrium at the average of the total mortality from the period 1955-1959 (*init.pop.Z*), or that the initial age structure was freely estimated as independent parameters (*init.pop.free*).

6.2.2 Recreational release mortality

When incorporating the recreational fisheries into the assessment model, only the landed component of the catch was included. However, this ignores the potential discard mortality of fish that are tagged and released from the recreational fisheries. As previously mentioned in Section 3.2, this discard mortality rate is thought to be approximately 25% (Brodziak, 2012) though it can be as high as 50% (Pepperell and Davis, 1999) in related species. Given that the number of tagged and released individuals has been substantially larger than the number of landed individuals since the early 1990s (Figure 10), total fishing mortality for these fisheries may be underestimated. In order to account for this potential extra source of fishing mortality, a suite of sensitivity runs were undertaken assuming a range of release mortalities: 10%, 20%, 30%, and so on up to and including an extreme case of 100% release mortality. This additional source of mortality was calculated in each year based on the number of released individuals and then added to the existing annual landed catch. Size composition data for released individuals were not included as these estimates were not felt to be reliable, but this is unlikely to appreciably impact the overall model results.

6.2.3 CPUE indices

The standardized CPUE for the JP 2 LL fishery was used in fitting the diagnostic case model as done in the 2012 stock assessment. However, there were 8 other candidate indices that could have been selected to be used as the index fishery. The sensitivity to the choice of candidate index was explored in a set of 8 sensitivity runs, one for each of the remaining fisheries.

6.2.4 Natural mortality

In the previous stock assessment, the fixed level of assumed natural mortality was shown to be influential to the assessment outcomes. Additionally, a substantial amount of uncertainty exists as to what the assumed level of natural mortality should be. A suite of sensitivity runs were undertaken to assess the impact of specifying the average level of M between 0.2 and 0.6. The age specific deviates from Piner and Lee (2011) were added to this baseline level of M to rescale the fixed age-specific values.

6.2.5 AU size composition data

Both length and weight composition data were available for the Australian longline fisheries (6 & 7). In developing the diagnostic case model it became readily apparent that the model was fitting the weight composition better and was underestimating the length composition relative to

the observations. The length composition were subsequently down-weighted to remedy this data-conflict. The sensitivity to this assumption was tested by fitting a single model which did not down-weight the length composition data for these fisheries and but down-weighted the weight composition data for these fisheries instead.

6.2.6 Recruitment penalty CV

As noted in Section 5.2.1 the CV on the recruitment penalty used in the diagnostic case is more restrictive than what is typically assumed for the tropical tuna assessments conducted in the WCPO. Sensitivity to the choice of CV for this penalty was assessed through two additional models using alternative CV values of 2.2 (assumption in WCPO tropical tuna stock assessments), and an intermediate value of 0.5.

6.2.7 Sex disaggregated

Previous research found some evidence of sexual dimorphism in striped marlin. Kopf et al. (2011) showed sex-specific differences in the length-at-age and weight-at-length relationships for SWPO striped marlin, with females growing to be larger than males. An accompanying study on reproductive behavior and characteristics noted that the sex-ratio at larger lengths skewed towards females (Kopf et al., 2012). One of the recommendations made in the 2012 stock assessment was to consider an alternative sex-disaggregated model structure in order to more appropriately account for these sex-specific processes. A sex-disaggregated model, building off the work of Takeuchi et al. (2018) for SWPO swordfish, was attempted for striped marlin in order to account for these apparent sex-specific differences. A more detailed description of the background analyses done in preparation for this sensitivity model is presented in Ducharme-Barth and Pilling (2019).

6.2.8 Spatially disaggregated

A spatially disaggregated one-off sensitivity was suggested by the PAW as an additional avenue for exploration. In this model the outer boundaries of the assessment region in SWPO were kept the same, but the assessment region was divided into two regions at 165°E. Previous tagging research conducted on striped marlin, using both conventional and electronic tags, appears to indicate different patterns of movement on either side of the 165°E boundary (Ortiz et al., 2003; Domeier, 2006; Sippel et al., 2011). Fish tagged off of the east coast of Australia (west of the boundary) appear to show higher residence times and lower levels of dispersion relative to fish tagged off of the North Island of New Zealand (east of the boundary). These different signals appear to challenge the assumption of a well-mixed population that is made by the single region structure assumed in the current stock assessment. A series of two-region models, making different assumptions about movement rates between the two regions, were run to evaluate the model sensitivity to the assumption of a single assessment region. Further detail on the background analyses done in support of this analysis are contained in Ducharme-Barth and Pilling (2019).

Development of this special uncertainty scenario was done in stepwise fashion. Prior to any of the multi-region models being run, an additional single region model was run as a bridge between the 2019 diagnostic case and the models with the newly defined regional structure. This model assumed a single spatial region but used the new fisheries definitions (described in Ducharme-Barth and Pilling (2019)) that were defined for the two-region models. Given the lack of a clear hypothesis on movement patterns or rates between the two regions (Section 3.3), a "mini-grid" of possible movement scenarios was investigated. To reflect the possible movement hypotheses

suggested by the tagging data the following 4 quarterly movement scenarios were considered as plausible:

- No movement west-to-east & moderate movement (0.1 longitudinal degrees traveled per day) east-to-west
- Moderate movement west-to-east & moderate movement east-to-west
- No movement west-to-east & high movement (0.4 longitudinal degrees traveled per day) east-to-west
- Moderate movement west-to-east & high movement east-to-west

These movement rates, suggested by the PAW, were converted to bulk transfer rates using the method of Kolody and Davies (2008) and were described in Ducharme-Barth and Pilling (2019).

6.3 Structural uncertainty grid

Stock assessments of pelagic species in the WCPO in recent years have run a "grid" of models to assess the uncertainty resulting from structural assumptions in the assessment model and to explore the interactions among the selected "axes of uncertainty". This grid contains all combinations of two or more parameter settings or assumptions for each uncertainty axis. These axes were selected based on the models explored in developing the 2019 diagnostic case, those in the one-off sensitivity analyses, from the findings of previous WCPO stock assessments, and from input at the PAW. The aim of the structural uncertainty grid is to provide an understanding of the variability arising from different structural assumptions that is not accounted for in the statistical uncertainty from a single model run or from a set of one-off sensitivities. The structural uncertainty grid for the 2019 striped marlin stock assessment contained 6 different axes (Table 5).

6.3.1 Steepness

As outlined in Section 5.2.1, the steepness parameter of the SRR is often poorly estimated and is therefore assumed to be fixed in WCPO stock assessments. However, the choice of value is known to be highly influential to the estimated management reference points. Consistent with previous WCPO stock assessments, steepness values of 0.6, 0.8, and 0.95 were considered as values in the uncertainty grid.

6.3.2 Growth

Growth is a major source of uncertainty within the assessment model, and was considered as an axis in the previous stock assessment. Ongoing research from annual otolith aging of striped marlin conducted by CSIRO showed that striped marlin may live longer than was found by Kopf et al. (2011) using dorsal fin-spines, and indicated inconsistencies between the age determined from dorsal spines and the annual age determined from otolith aging. An alternative growth hypothesis to the Kopf et al. (2011) growth considered in the diagnostic case, based on this new research, was considered as an option in the uncertainty grid (Figure 11). This alternative growth was estimated using a subset of the individuals used by Kopf et al. (2011) except that for the larger individuals annual otolith ages were used instead of ages determined from dorsal fin-spines. Smaller individuals used the same daily increment ages as determined by Kopf et al. (2011). The parameterization of the alternate growth that was used is given by: L1 = 140.97 cm EFL, L2 = 215.11 cm EFL, and K = 0.6733.

6.3.3 Natural mortality

In the one-off sensitivity analysis, natural mortality was shown to be highly influential to assessment outcomes. Additionally, this is a parameter for which a considerable amount of uncertainty exists. Based on the recommendations made in the previous striped marlin assessment, the uncertainty grid considered 3 options for the average level of natural mortality: 0.3, 0.4, and 0.5 (Figure 12). Note that the age-specific deviates from Piner and Lee (2011) were applied to these values.

6.3.4 CPUE

The one-off sensitivities showed that assessment results were sensitive to the choice of standardized CPUE used as the index fishery. Standardized indices for the 3 major longline fisheries operating in the assessment region (JP 2 LL, TW 4 LL, and AU 2 LL) were selected as options in the uncertainty grid.

6.3.5 Size frequency weighting

The difficulties in assigning weighting to the size frequency data were discussed in Section 5.4. To assess the sensitivity of model results to the weighting of these data, two alternative "scalars" were considered as sensitivities beyond what was used in the diagnostic case model. One scalar (ESS divisors of 10 and 20 for weight and length composition data, respectively) up-weighted the data relative to the diagnostic case, while the other down-weighted the data (divisor of 50 and 100 for weight and length composition data, respectively). This approach is consistent with other WCPO stock assessments.

6.3.6 Recruitment penalty CV

As described in Section 6.2.6, the assumption of fixing this penalty at different levels appeared to be influential in the diagnostic case. In addition to the level assumed in the diagnostic cases of both the 2012 and 2019 stock assessments, two additional models using alternative CV values of 2.2 (assumption in WCPO tropical tuna stock assessments), and an intermediate value of 0.5 were included in the grid.

7 Results

7.1 Consequences of stepwise model development

The progression of model development from the 2012 reference case model to the 2019 diagnostic case model is outlined in Section 6.1 and results are displayed in Figure 13 with respect to both the change in spawning biomass (SB) and the level of depletion of SB relative to the unfished condition. A summary of the consequences of this progression through the models is as follows:

- The 2012 reference case model (black dotted line)
- **Step 2** [*new MFCL*]: Updating the MULTIFAN-CL executable to the most recent version (**cyan**), produced a virtually identical fit to the 2012 reference case.
- **Step 3** [*2017 data*]: Adding the 6 additional years of data (**light blue**) shifted SB slightly, but ultimately ended up at roughly the same level of depletion.

- **Step 4** [*2kg wt*]: Changing the aggregation of the weight composition data to 2 kg bins (**teal**), resulted in a substantial downwards revision in both starting *SB* and total depletion. Likelihood profiling of the model with 1 kg bins indicated an optimum solution at a lower biomass than what was estimated by the model. Aggregating the data to the 2 kg bins resulted in the model solution converging to this lower optimum biomass estimate.
- **Step 5** [*geostats*]: Replacing the standardized CPUE indices for fisheries 1-5 with the geostats indices (**light green**) revised the starting *SB* upwards, and provided a slightly more optimistic fit than the previous step.
- **Step 6** [*AU LF*]: Down-weighting the Australian longline's length composition data (**forest green**) did not appreciably alter the magnitude or trend of either *SB* or depletion.
- **Step 7** [*split sel. 2 & 12*]: Allowing for the independent estimation of the previously shared selectivities for fisheries 2 and 12 (**lime green**) resulted in a better fit to the composition data (weight for fishery 2 and length for fishery 12) of both of these fisheries. This did not alter the trend but resulted in an upwards shift in *SB* and a marginally more optimistic level of depletion.
- **Step 8** [*M@age*]: Changing the definition of M from an age-invariant 0.4 across all ages to age-specific deviates from 0.4 (**yellow**) did not result in a noticeable change from the previous step.
- **Step 9** [*maturity@length*]: Updating the maturity ogive and specifying maturity as a function of length (**orange**) resulted in a substantial downwards shift of the *SB* and also resulted in a more pessimistic level of overall depletion. The magnitude of this change can be attributed to updating the ogive to be consistent with the current biological information and also to incorporating the sex-ratio at length into the calculation.
- **Step 10 - 12**[*free sel., rec. timing and SRR def*]: These three steps had negligible impacts on the model and estimated quantities: freeing up the selectivity estimation (**coral**), changing the timing of when recruits enter the fishery (**maroon**), and fixing the recruits in the final two years of the model (**violet**).
- **Step 13 - 2019 Diagnostic Case** [*depletion def.*]: Updating the definition for how depletion was calculated internally to the model (**indigo**) did not change *SB* or any of the equilibrium MSY based reference points, but it did result in a lower level of depletion relative to $SB_{F=0}$.

7.2 Model fit - diagnostic case model

This section discusses the model results for the diagnostic case model, defined by the final step in the stepwise model development (13 - depletion def; Figure 13). The presentation focuses on the fit to the various sources of data, and the biological and fisheries-related parameters.

7.2.1 Catch data

A relatively high penalty was imposed on the fit to the catch, resulting in a strong fit to the catch and tight pattern in the relative log-residuals (Figure 14). It is worth noting that the fit to the catch was weakest for the index fishery (fishery 2) relative to the other fisheries with weaker effort deviation penalties (Figure 9), indicating the conflict in the likelihood between the fit to the catch and to the CPUE index for this fishery.

7.2.2 CPUE index

The general trend in CPUE is fairly consistent across the different standardized indices. CPUE declined from a peak at the start of the model period, rebounded slightly in the early 1970s and then continued to decline for the duration of the model, albeit with another slight increase in the early 2000s (Figure 15). Strong seasonality in the observed catch rate was also evident. Model fit to the observed standardized CPUE for the index fishery (fishery 2) and to the other longline fisheries was generally good, though it consistently underestimated the highest observed catch rates throughout the model period. This can be partially attributed to the model's inability to replicate the significant variation between time steps. Fit to the lower observations of CPUE was good but the model was unable to match the prediction of the adjacent high CPUE observations. The model predicted moderate seasonal variation in CPUE, though the magnitude was less than observed in the standardized indices.

The fit to the two recreational indices was quite poor in comparison to the fit to the major longline fisheries. For the Australian recreational fishery, the standardized CPUE increased rapidly before leveling off. This is contrary to what the model predicts for this fishery which is a sharp decline followed by a relatively flat trend. The fit to the standardized CPUE for New Zealand is better in comparison through the first half of the period of record, however in the last decade the observed and predicted trends diverge.

The plots of effort deviates for each fishery over time indicate an adequate fit to the CPUE indices (Figure 16), with no systematic departures from a distribution around zero for the fisheries which received standardized effort. For the New Zealand longline fishery (fishery 8), some trend in the effort deviations is visible. There are negligible catches reported for this fishery despite a considerable amount of effort. Since 1987, it has been illegal to commercially land striped marlin in New Zealand, therefore providing little incentive to accurately report interactions with the species.

7.2.3 Size frequency data

Overall the model fit to the aggregated size frequency data was good, for both weight (Figure 17) and length (Figure 18), when there was a sufficient number of observations. The major longline fisheries (fisheries 2,3,6,7,11, and 12) all appeared to interact with a single mode of individuals of between 3-5 years of age. The recreational fisheries reported the largest individuals, with the New Zealand component being the largest.

Fisheries 1 and 4 had few observations and it was not felt that selectivity curves could be estimated independently for these fisheries. As a result they were grouped with the other longline fisheries operating in those model sub-regions. The poor aggregate estimates for these fisheries can be attributed to the model estimating a selectivity curve that better fit the composition data of the fishery it was paired with.

For fisheries 6 & 7 the model consistently underestimated the length composition data given the observed and predicted weight frequency data. This data conflict could arise due to model misspecification if these samples are coming from a component of the stock that are "skinnier" than the weight-length relationship used in the model would expect. Another possibility is that the conversion factors (Williams and Smith, 2018) used to convert weight measurements to whole weight and length measurements to eye-orbital fork-length need to be revisited for this fishery.

Investigating the temporal trends in predicted and observed median size composition, the fit to the weight frequency observations for the major longline fisheries (fisheries 2,3,6 and 7) appears

reasonably good (Figure 19). The best fit is to the long record of catch-at-weight information from the New Zealand recreational fishery (fishery 10). This fishery showed a persistent decrease in median weight over time from the beginning of the assessment period, which was well captured by the model fit.

In general there were much fewer length frequency data available to be fit in the model, likely due to the low level of observer coverage in the region. The data that were available predominantly came from the PICT longline fleets grouped together in the mixed longline fisheries 11-14. Fit to the length frequency for these fisheries appeared reasonable (Figure 20) other than the aforementioned issues with fisheries 6 and 7. Fishery 11 showed a shift in the median observed lengths in the mid-2000s. This is not unsurprising given that the length frequency observations come from an amalgamation of different sampling programs. In this instance, the shift coincides with a large influx of composition samples from Chinese Taipei in 2006 that persist for the next 8 years (Figure 12; Ducharme-Barth and Pilling, 2019). In the future, it may be advantageous to dis-aggregate these mixed longline fisheries if there is sufficient size frequency data or to standardize the composition data to account for the source of the samples.

7.2.4 Likelihood

The change in negative log-likelihood of average total biomass relative to the minimum is shown for the total likelihood (black line) and the individual data components (colored lines) in Figure 21. As evident from the figure, the likelihood profile for this model is quite asymmetrical. The biomass estimate is defined by a very hard lower bound as all components' relative likelihoods increase quite rapidly at lower total biomass values. The likelihood increases substantially at higher total biomass values but this increase is less extreme and indicative of a longer uncertainty 'tail' in this direction. Additionally, the model fit seems to be driven almost exclusively by the fit to the weight composition data which suggests a lower biomass relative to the model estimate. In contrast the likelihood profile for CPUE, other than on the low end of biomass values, is very flat and comparatively uninformative across a wide range of biomass values. The minimum of this component occurs at a biomass estimate roughly 2.5 times higher than the current model fit. The conflict between these two components results in the fit that is achieved since all of the other components are even less informative than CPUE.

The fishery specific likelihood profiles by data component are shown in Figure 22. From these fishery specific profiles it can be seen that the overall model fit is being driven by the fit to the long record of weight composition data from the New Zealand recreational fishery (Section 7.2.3), and to a lesser extent the weight composition observations from the two Australian longline fisheries. All three of these fisheries have well defined weight composition likelihoods indicating a lower total biomass estimate. The CPUE fit to the index fishery (JP 2 LL) is also shown, with this component favoring an average biomass estimate roughly twice that of the overall model estimate.

7.3 Parameter estimates - diagnostic case model

Estimates from the diagnostic case are presented here in a detailed form, by fishery, so that model behavior can be assessed.

7.3.1 Catchability

In fitting the 2019 diagnostic case, the catchability for the index fishery (JP 2 LL) was held constant over time (though seasonal variability was permitted). For all other longline fisheries,

catchability was allowed to vary biennially (long term trend) as well as seasonally (Figure 23). The two recreational fisheries were only assumed to have biennial changes in catchability as they only had catches in the 1^{st} quarter of the year. Time varying trends in catchability could be indicative of a number of different factors: change in spatial distribution of the fleet, change in gear configuration or target species not accounted for in the CPUE standardization, and/or change in the composition of the fleet.

The catchability patterns observed for the 2019 diagnostic case are largely consistent with those seen in the 2012 reference case. Catchability is highest for the fisheries operating in model sub-region 2 which is consistent with the high removals seen in this sub-region. High estimates of catchability were also seen for fishery 11 (the mixed-longline fishery in sub-region 1), and this coincides with a period where catches from Korean vessels made up a large component of the catch (Figure 12; Ducharme-Barth and Pilling, 2019) relative to the other DWFN/PICT vessels it was grouped with. Additionally, strong seasonal trends in catchability are evident from most fisheries. Catchability for the longline fisheries was highest in the 4^{th} quarter of the calendar year.

The temporal trends in catchability also illustrated conflict between the index fleet standardized CPUE and other fisheries where standardized effort was provided. Despite being standardized, the catchability trends for both the recreational fisheries increase over time (1990s for the AU 3 REC and 1970s for the NZ 3 rec). This conflicts with the declining trend of the index fishery (JP 2 LL) over the same period since catchability for that fishery is constant over time.

In contrast to the two recreational fisheries, the two Australian longline fisheries (AU 2 LL & AU 3 LL) showed a fairly flat catchability trend over the model period. This suggests consistency with the trend shown by the index fishery (JP 2 LL) for that same time period. A difference from the 2012 reference case is that the catchability trends for the JP 1 LL & JP 3 LL fisheries do not show persistent increases as they did in the previous assessment. This is likely a result of the geostats modeling approach that simultaneously estimated the standardized CPUE for the Japanese longline fleets in all four sub-regions.

As mentioned earlier, changing catchability can be a result of changing fleet composition. This is illustrated in the catchability trends for the mixed longline fisheries. In OTHER 1 LL the increase in catchability corresponds with the increase in catches from Korean vessels. In the OTHER 2 LL & OTHER 4 LL fisheries, the development in the early 1990s of PICT fisheries in these regions (particularly Fiji in sub-region 2, and French Polynesia in sub-region 4) coincide with the decline in catchability that is seen (Figure 13 & 15; Ducharme-Barth and Pilling, 2019). Additionally, apparent changes in catchability could be related to reporting issues including species identification.

7.3.2 Selectivity

Longline fisheries selectivity functions are typically thought to be asymptotic, meaning the gear has equal probability of catching all individuals greater than a certain size. However, given that this is a single region model and that a fleets-as-areas approach was used to capture the spatiotemporal patterns seen in the composition data, the estimated selectivity functions are a product of both the gear selectivity and the "availability" of fish to that gear. Modeling this "total selectivity" of each fishery with a cubic spline allowed for the estimation of non-asymptotic shapes for many of the longline fisheries (Figure 24).

Overall, striped marlin tended to recruit to the gear beginning at age 2 (except in sub-region 1 where very small striped marlin were observed in the catch) and become fully recruited to the major

longline fisheries operating in sub-regions 2 & 3 by age 3-4. Although a spline function, without a non-decreasing penalty, was assumed for fisheries in sub-regions 1 and 4, asymptotic selectivity was estimated most likely due to the broad size range in catches. Relative to sub-region 2, the estimated selectivities for the longline fisheries in sub-region 3 showed higher selectivity at older ages reflecting the larger size composition in the catch. The two longline selectivity patterns in sub-region 3 (groups 3 & 6) showed remarkably similar forms despite being estimated independently.

A difference from the 2012 stock assessment was that the selectivity for the AU 3 REC fishery was allowed to be non-asymptotic to account for the differences in size composition as compared to striped marlin caught by the NZ 3 REC. In both cases, fish are not estimated to recruit to the gear before age 3. Fish are fully recruited to AU 3 REC by age 4 and closer to age 6 for NZ 3 REC.

7.4 Stock assessment results - diagnostic case model

7.4.1 Recruitment

There is considerable temporal variation in recruitment over the model period (Figure 25), and the recruitments show a similar pattern to that described in the 2006 and 2012 SWPO striped marlin stock assessments (Langley et al., 2006; Davies et al., 2012). Annual recruitment is estimated to have been high prior to 1971 (43% greater than the mean recruitment) and fluctuated about a lower level (82% of the mean) for the subsequent period. Since the mid-2000s the relative variability in recruitment has been considerably less than that observed in earlier time periods. The last two years of recruitment were not estimated and were fixed at the mean level. The recruitment estimates have broad confidence intervals (the statistical uncertainty derived from the Hessian) particularly during the early period (1950s and 1960s), prior to the availability of size data.

The estimated SRR for the diagnostic case model is presented in Figure 26.

7.4.2 Biomass

The annual trends in total (Figure 27) and adult biomass (Figure 28) are similar to the temporal trend in recruitment described in the previous section. Biomass was estimated to be high during the 1950s and declined sharply until the late 1960s. Following a brief increase in the early 1970s, biomass is predicted to have declined steadily through to the late 1990s. Since that time, the decline has leveled off and biomass has remained fairly constant. The most recent years of the model appear to show a slight increase in trend. Again similar to recruitment, the model specific statistical uncertainty as calculated from the Hessian is greatest in the 1950s and 1960s.

These patterns are mirrored in the estimates of vulnerable (exploitable) biomass for each of the fisheries in the model (Figure 29).

7.4.3 Fishing mortality

Fishing mortality (exploitation) rates for adult striped marlin are estimated to have increased sharply in the mid 1950s (Figure 30 & Figure 31) following the development of the Japanese longline fishery in sub-region 2 and the initial period of high catches (Figure 5). Since that time, fishing mortality rates for adult striped marlin have steadily increased to the early 2000s where the adult mortality rate jumped up sharply to a peak of close to 0.8. An equally rapid decline in fishing mortality rate to around 0.35 a year has occurred in the last two years of the model. This is likely a combination of the decline in total catch over the last 3 years, and the increase in recruitments in the last two years due to being fixed at the mean level.

Fishing mortality rates for juvenile striped marlin are estimated at a lower level throughout the model period because these fish (< 3 years-old) are generally not fully recruited to the main longline fisheries (Figure 24), although they have slowly increased through time, particularly for age 2 individuals following the mid-1990s (Figure 32). This is perhaps a result of the development of DWFN and PICT longline fleets in the equatorial sub-region of the model over that time period.

7.4.4 Fishery impact

An indicator of the impact of fishing on the stock is to compare the biomass trajectories with fishing and the predicted biomass trajectory in the absence of fishing. The impact can be expressed as a proportional reduction in biomass ($1 - B_t/B_{t,F=0}$) and calculated for either the total stock or just individuals contributing to the spawning biomass. It is then possible to attribute the fishery impact to specific fishery components in order to see which types of fishing activity have the largest impact on the total biomass and spawning biomass (Figure 33).

The diagnostic case model indicates that the entire fishery has had a substantial impact on the levels of total and spawning biomass (Figure 33), with current levels being 36% of the total biomass and 17% of the spawning biomass that would have occurred in the absence of fishing. The impact of the entire fishery is estimated to have increased sharply in the 1950s with a more gradual increase since that time. Longline fisheries in sub-region 2 (excluding Australia and New Zealand) have had the largest impact, however as Japanese longline fishing effort and catch has declined in this sub-region, so has the impact of these grouped longline vessels. The decline in sub-region 2 impact has almost been completely offset by a series of progressive fishery expansions in sub-region 1. The increase in Korean longline effort in sub-region 1 during the late 1970s, followed by the increase in Chinese Taipei longline catches in that sub-region in the early 2000s and lastly the increase in Chinese longline catches in the late 2000s (Figure 12; Ducharme-Barth and Pilling, 2019) all coincide with "stepwise" increases to the fishery impact for sub-region 1 longlines. These increases, combined with the rapid development of the Australian longline fisheries in the late 1990s (Figures 7 & 8; Ducharme-Barth and Pilling, 2019) contribute to the overall increase in fisheries impact, and fishing mortality estimated by the model at the turn of the century (Figure 32).

The estimated fishery impacts are low for the recreational fisheries.

7.5 Multi-model inference

7.5.1 Sensitivity analyses

Comparisons of the spawning biomass, and depletion trajectories for the diagnostic case and key sensitivity analyses are presented here. In addition to discussions at the PAW and recommendations made in the previous assessment, the results from these one-off sensitivities informed the selection of the axes and axes-levels used in the structural uncertainty grid.

- **Model start year**: The model appears robust to the exclusion of data prior to 1955 under both scenarios considered (Figure 34): beginning at the equilibrium fished condition (*init.pop.Z*) or estimating the age structure in the first year as free parameters (*init.pop.free*). Though starting points for the spawning biomass and depletion estimates were different, both models converged to a solution very similar to the 2019 diagnostic case, though marginally more optimistic.
- **Recreational release mortality**: Including an additional source of mortality for the recreational fisheries, in the form of tag and release mortality, did not appear to make much of a

difference to either the estimated spawning biomass or the level of depletion relative to the 2019 diagnostic case (Figure 35). Starting biomass crept upwards with increases in release mortality, but all sensitivity models ultimately resulted in nearly identical levels of spawning biomass and depletion, even in the extreme case where all released striped marlin suffered mortality.

– **CPUE indices**: There were 9 fisheries with standardized CPUE indices available in the 2019 SWPO striped marlin stock assessment. In this sensitivity analysis, each model run used a different one of these fisheries as the index fishery (constant catchability assumed and time varying effort deviate penalty weights calculated from the CVs of the provided standardized index). The choice in index fishery resulted in significant variability in the estimated levels of spawning biomass and adult depletion (Figure 36). In terms of the spawning biomass, starting levels and terminal levels partitioned into three distinct groups. For the starting spawning biomass estimate, the assumed index fishery in the diagnostic case (fishery 2) had the highest estimate at just over 17,000 mt, estimates from fisheries 1 & 5 were closer to 15,000 mt, and the remaining fisheries (3, 4, 6, 7, 9 & 10) estimated starting spawning biomass around 13,500 mt. These groupings were not consistent in terms of the terminal spawning biomass estimate. The highest estimate was from Fishery 9 ($\sim 4,000$ mt), followed by estimates from fisheries 5-7 & 10 ($\sim 3,000$ mt), and fisheries 1-4 estimated terminal spawning biomass at the lowest level ($\sim 1,750$ mt). The groupings for terminal biomass were consistent with the groupings seen for level of depletion. Given this variability, the choice of fishery used as the index fishery was included as an axis in the uncertainty grid and indices from the 3 major longline fisheries operating in the model region (JP 2 LL, TW 4 LL, and AU 2 LL) were selected as levels for this axis.

– **Natural mortality**: Natural mortality was identified in the previous assessment as a parameter of influence on stock assessment outputs. Sensitivity runs were undertaken using levels of average natural mortality-at-age from 0.2 (lower bound in 2012 uncertainty grid) to 0.6 (upper bound in 2012 uncertainty grid) by increments of 0.025 (Figure 37). Starting spawning biomass was highly variable with respect to the assumed level of M, though the terminal estimates were very consistent. This translated to a large range in the estimated level of adult depletion. Given that the model begins at the unfished equilibrium, it is intuitive that lower values of M would result in higher initial biomass estimates as there would be a greater proportion of older, larger individuals in the population.

  From these sensitivity runs, a likelihood profile was calculated for M (Figure 38) which indicated an optimal value of 0.3. There was no clear data component driving this estimate, and the fit was achieved by finding a balance between the CPUE and the weight composition data. M was estimated using models fit to the 9 different possible index fisheries to account for the effect of choice of CPUE index on the optimal value for M. Most models consistently estimated the mean level of M around 0.3, however the model fitting the index from Australian recreational fishery estimated a mean value closer to 0.5 (Figure 39). All values were within the range considered in the uncertainty analysis.

  Three different fixed levels of M were ultimately included in the structural uncertainty analysis: 0.3 (close to model best estimate, and recommendation from 2012 stock assessment), 0.4 (diagnostic case), and 0.5 (recommendation from 2012 stock assessment).

– **AU size composition data**: Given the conflict between the fits to the weight and length frequency data from the Australian longline fishery, the impact of down-weighting either the

weight (*fit2wf*) or the length (*fit2lf*) for these fisheries was assessed in a sensitivity analysis (Figure 40). Choice of data component made little difference overall. Starting spawning biomass was slightly larger when fitting to the length data which resulted in a negligible improvement in terms of depletion.

- **Recruitment penalty CV**: Estimated model outputs were sensitive to the choice of recruitment penalty CV, given the other assumptions made when parameterizing the 2019 diagnostic case model (Figure 41). Reducing the penalty led to progressively larger estimates of starting spawning biomass (via an increase in the level of average recruitment (bottom panel; Figure 42)) and reduced levels of adult depletion relative to the level estimated in the 2019 diagnostic case, though the trend in recruitments also became more extreme (top panel; Figure 42). Terminal spawning biomass estimates were similar. To account for these differences, the CV for the recruitment penalty was selected as an axis in the uncertainty grid with 3 levels: 0.2, 0.5, and 2.2.
- **Sex disaggregated model**: Attempts at running a sex-disaggregated model for 2019 SWPO striped marlin were not to be successful. Though estimates of spawning biomass appeared to be reasonably similar to those estimated in the 2019 diagnostic case, the effort deviates and fishing mortalities estimated did not seem plausible. Despite reasonable amounts of composition data in the recent period, there was very little sex-specific composition data early in the model to inform the population sex-ratio. As a result, the model struggled to account for the large removals early in the model using the estimated shared selectivity at length between the two sexes given the similarity in growth between the two.
- **Spatially disaggregated model**: A request was made at the 2019 PAW to consider a spatially disaggregated model structure by dividing the existing model region in two at 165°E. This request was investigated in a side analysis, the results of which are presented in the Appendix (Section 12.2). Given the large uncertainty in key assumptions (e.g. movement and proportion of recruitment assigned to each region), these models should not be considered for management purposes as the assumptions made for movement and recruitment partitioning between regions are strongly influential to the model results.

  Additionally, from a stock assessment modeling standpoint, one of the reasons for moving to a regionally explicit model structure is to avoid using "fleets-as-areas". The arbitrary placement of a boundary at 165°E does not accomplish this, so all of the longline fisheries are still estimated with flexible cubic splines (allowing for a descending limb) since selectivity is still a product of availability and gear selectivity. As a result, the level of depletion is driven by the shape of the estimated selectivity curves (non-decreasing leading to more depleted and dome-shaped being less depleted). If a regional model is to be seriously considered for striped marlin, the regional structure should be selected in a way as to remove "fleets-as-areas" so that selectivities are based on the gear and with the availability component driven by movement. This would remove the uncertainty caused by the changing shape of the selectivity curves since a non-decreasing form could be used for the longline fisheries. However, this would require more confidence on the movement rates between regions otherwise running the model across a grid of movements as presented in the Appendix (Section 12.2) will need to be continued.

7.5.2 Structural uncertainty grid

The results of the structural uncertainty analysis are summarized in several forms: time-series plots of spawning biomass (SB_t), fisheries depletion ($SB_t/SB_{t,F=0}$), $F_t/F_{MSY,t}$, and $SB_t/SB_{MSY,t}$ for all

models in the grid (Figure 43); violin-plots of $SB_{recent}/SB_{\text{MSY}}$, $F_{recent}/F_{\text{MSY}}$ and $SB_{recent}/SB_{F=0}$ for the different levels of each of the six uncertainty axes (Figure 44); Kobe (Figure 45) and Majuro (Figure 46) plots showing the estimates of $F_{recent}/F_{\text{MSY}}$, $SB_{recent}/SB_{\text{MSY}}$, $SB_{recent}/SB_{F=0}$ for the different uncertainty axes in the grid; and averages and quantiles across the full grid of 300 models for all the reference points and other quantities of interest (Table 6). Though the full-factorial combination of grid axis levels resulted in 486 models being run, a number of these models (186) did not meet the retention criteria and were thus excluded in a manner similar to the 2012 stock assessment. Models were excluded from the analysis if they had a maximum gradient component greater than 10^{-3} (1% of excluded runs) or if they estimated implausible starting total biomass values >500,000 mt (99% of excluded runs). For reference, in the diagnostic case, the starting total biomass estimated was 30,629 mt and a maximum gradient component of 1.27×10^{-6}.

Many of the results of the structural uncertainty analysis are consistent with the results of previous WCPO stock assessments that used the same uncertainty axes, notably for natural mortality and steepness. However, additional axes have been included in this assessment and these have consequences for summaries of stock status and resulting management implications. The general features of the analysis are as follows:

- There appears to be a higher than normal level of uncertainty associated with this assessment, however this simplistic characterization is a bit misleading. The uncertainty distributions for key parameter estimates and time dynamic quantities are highly asymmetrical (positive/right skew) with long tail giving the impression that uncertainty is larger than it actually is. The core of the distributions (50^{th} percentile) are fairly tight and compare favorably to the uncertainty presented in WCPO tropical tuna stock assessments. Translating this uncertainty to the MSY based reference points: 69% of model runs estimated $SB_{recent}/SB_{\text{MSY}}$ to be less than 1, and 44% of model runs estimated $F_{recent}/F_{\text{MSY}}$ greater than 1. In terms of depletion, 92% of runs estimated SB_{recent} to be less than 50% of $SB_{F=0}$ and 50% of runs estimated SB_{recent} to be less than 20% of $SB_{F=0}$.
- Given the long tails, there was no clear separation among axis levels (at the 80^{th} percentile band), though trends across levels did exist. Natural mortality and steepness both showed the strongest trend across levels for $SB_{recent}/SB_{\text{MSY}}$, $F_{recent}/F_{\text{MSY}}$ and $SB_{recent}/SB_{F=0}$. Across the different levels of the natural mortality axis: median $SB_{recent}/SB_{\text{MSY}}$ increased from 0.50 (M=0.3) to 0.78 (M=0.4) to 1.13 (M=0.5); median $SB_{recent}/SB_{F=0}$ increased from 0.12 (M=0.3) to 0.19 (M=0.4) to 0.29 (M=0.5) ; and $F_{recent}/F_{\text{MSY}}$ decreased from 1.33 (M=0.3) to 0.88 (M=0.4) to 0.54 (M=0.5). For the 3 different levels of steepness: median $SB_{recent}/SB_{\text{MSY}}$ increased from 0.53 (h=0.65) to 0.73 (h=0.8) to 1.32 (h=0.95); median $SB_{recent}/SB_{F=0}$ increased from 0.18 (h=0.65) to 0.21 (h=0.8) to 0.26 (h=0.95); and $F_{recent}/F_{\text{MSY}}$ decreased from 1.28 (h=0.65) to 0.92 (h=0.8) to 0.50 (h=0.95).
- A weaker trend existed for both the CPUE and the composition likelihood weighting axes. For CPUE, the model runs using the Australian longline index estimated higher median $F_{recent}/F_{\text{MSY}}$ (1.28) and lower $SB_{recent}/SB_{\text{MSY}}$ (0.58) & $SB_{recent}/SB_{F=0}$ (0.16) than runs fitting to either the Japanese (0.77 $F_{recent}/F_{\text{MSY}}$; 0.78 $SB_{recent}/SB_{\text{MSY}}$; 0.23 $SB_{recent}/SB_{F=0}$) or Chinese Taipei (0.90 $F_{recent}/F_{\text{MSY}}$; 0.78 $SB_{recent}/SB_{\text{MSY}}$; 0.20 $SB_{recent}/SB_{F=0}$) index. In terms of the composition likelihood weighting, when the composition data were given greater importance in the likelihood median model estimates were higher for $F_{recent}/F_{\text{MSY}}$ (1.14) and lower for $SB_{recent}/SB_{\text{MSY}}$ (0.63) & $SB_{recent}/SB_{F=0}$ (0.17) than runs with either moderate (0.86 $F_{recent}/F_{\text{MSY}}$; 0.81 $SB_{recent}/SB_{\text{MSY}}$; 0.23 $SB_{recent}/SB_{F=0}$) or low likelihood importance (0.62 $F_{recent}/F_{\text{MSY}}$; 0.99 $SB_{recent}/SB_{\text{MSY}}$; 0.29 $SB_{recent}/SB_{F=0}$) importance.

The composition likelihood weighting axis saw the greatest disparity between axis levels in the proportion of runs that were excluded from the grid, 94% of runs with the high likelihood importance were retained compared to 62% retention for the moderate level of importance and only 29% retention for the low level of likelihood importance. Given the nature of the likelihood structure in this model, this patterns make sense. The weight composition data was the most well defined component of the likelihood. Reducing the weighting on this component causes the likelihood surface to be less well defined and thus increases the probability that the MULTIFAN-CL optimization routine "wanders-off" into an implausible parameter space.

– Neither the growth axis or the CV on the recruitment penalty were major drivers for trends in the uncertainty grid as there was little separation between the different levels. For the growth axis, median $F_{recent}/F_{\text{MSY}}$, $SB_{recent}/SB_{\text{MSY}}$, and $SB_{recent}/SB_{F=0}$ were marginally more pessimistic for the Kopf growth (0.99, 0.66, and 0.18; respectively) than for the CSIRO otolith aging growth (0.80, 0.78, and 0.25). A CV of 0.2 had the most pessimistic outcomes in terms of median $F_{recent}/F_{\text{MSY}}$, $SB_{recent}/SB_{\text{MSY}}$, and $SB_{recent}/SB_{F=0}$ (1.02, 0.65, and 0.18) compared to the medians for a CV of 0.5 (0.82, 0.81, and 0.24) and 2.2 (0.90, 0.76, and 0.20).

7.5.3 Further analyses of stock status

There are several ancillary analyses related to stock status that are typically undertaken on the diagnostic case model (dynamic Majuro analyses, yield analyses, etc.). The shift towards multi-model inference, defining management reference points based on the ensemble of grid runs, makes it impractical to present results from each individual model. This section summaries these analyses across all retained models within the structural uncertainty grid.

– **Yield analysis**: The yield analyses conducted in this assessment incorporate the SRR (Figure 26) into the equilibrium biomass and yield computations. The diagnostic case used a steepness value of 0.8, and a recruitment penalty CV of 0.2 to estimate the recruits coming from this relationship. In the structural uncertainty grid, values of 0.65 and 0.95 were also considered for steepness and values of 0.5 and 2.2 were considered for the recruitment penalty CV. Both of these axes along with the choice of CPUE the model fit to influenced the shape and magnitude of model estimated yield curves (Figure 47). Across grid runs, higher levels of steepness and recruit CV generally contributed to higher levels of equilibrium MSY and could sustain a broader range of fishing mortality values. The median MSY increased across steepness levels from 1,989 mt, to 1,998 mt, and 2,189 mt, and generally increased with increasing recruit CV levels from 1,945 mt, to 2,181 mt, and 2,091 mt. Additionally, grid models fitting to the JP 2 LL index showed higher median MSY levels (2,250 mt) relative to TW 4 LL (1,975 mt) and AU 2 (LL 1,855 mt). For reference, the median catch rate since 1980 has been approximately 1,650 mt per year and 1,300 mt per year over the last 10 years (Figure 1).

  The remaining three axes also showed trend across their respective levels, though they did not alter the shape of the yield curve as much as the previous three. Median estimated MSY across grid runs was positively correlated with natural mortality (1,956, 2,005, 2,216 mt) and negatively correlated with the level of composition data weighting (2,004, 2,035, and 2,216 mt). Lastly, the Kopf growth showed lower median levels of MSY (2,000 mt) than the growth derived from otolith aging (2,108 mt).

  In general, equilibrium MSY was achieved at higher fishing mortality values than recently seen (i.e. the estimated fishing mortality multiplier at maximum of the yield curve is greater than 1). However, a handful of grid axis levels indicated that MSY could be achieved at lower

than recent levels of fishing mortality: high levels of size frequency weighting in the likelihood (0.87; fishing mortality multiplier), natural mortality of 0.3 (0.75), steepness of 0.65 (0.78), AU 2 LL CPUE index (0.78), and a recruitment CV of 0.2 (0.98).

– **Dynamic MSY**: Given the assumptions made in the grid with regards to the SRR, the ratio of SB_{MSY} to SB_0 was estimated to be between 0.121 and 0.305 with a median of 0.228. This metric has gradually increased over time (from $\sim$ 0.20 in 1952 to $\sim$ 0.25 in 2017), as calculated using the time-dynamic MSY described in Section 5.6.1 (Figure 48). Additionally, the trimodal distribution that is seen corresponds to the 3 different levels of steepness used in the grid with the left peak corresponding to a steepness of 0.95, the middle a steepness of 0.8, and the right peak a steepness of 0.65. The higher the productivity of the stock according to the SRR, the greater the level of depletion (relative to SB_0) at MSY. In terms of SB_{MSY} relative to $SB_{F=0}$ the median across the grid was estimated to be higher at 0.271 (range: 0.159–0.621).

  Examining the other time-dynamic MSY based reference points, the modal estimate of time-dynamic MSY (MSY level that would be theoretically achievable under the different patterns of age-specific fishing mortality observed through the history of the fishery) is stable across the model period, although the long tail and positive skew of the distribution is evident (Figure 48). The time-dynamic distribution of fishing mortality is also fairly stable with perhaps a slight decline in recent years. The two modes of this distribution again correspond to the different values of steepness assumed in the grid with the right mode characterized by models with a steepness of 0.95 and the left mode characterized with models having steepnesses of 0.65 and 0.8.

– **Dynamic Kobe & Majuro**: The section summarizing the structural uncertainty grid (Section 7.5.2) presents terminal estimates of stock status in the form of Kobe and Majuro plots. Further analyses can estimate the time-series progression of these metrics (Section 5.6.4). The large number of models in the grid makes it unwieldy to present the time-dynamic Kobe and Majuro plots for each, however the median time-series progression of the Kobe management reference points across all grid runs is presented in Figure 49. The equivalent dynamic Majuro plot is shown in Figure 50.

  The stock begins at the unfished equilibrium in 1952 and rapidly declines through the 1950s. The stock continues to decline, though more gradually, in the 1970s, and in the early-1990s crosses into an overfished status ($SB_t/SB_{MSY,t} < 1$) though overfishing is not occurring ($F_t/F_{MSY,t} < 1$) at that point according to MSY based definitions. With regards to depletion, spawning biomass is at approximately 30% of $SB_{F=0,t}$ at this point. In the late-1990s the stock status recovers to greater than $SB_{MSY,t}$ levels but quickly returns to an overfished state in the early-2000s. At this point overfishing, as defined by MSY, is also determined to be occurring and continues until the early-2010s. At the beginning of the current decade, the SWPO striped marlin stock was overfished and undergoing overfishing according to MSY reference points (based on the median uncertainty grid estimates), and was close to 18% of $SB_{F=0,t}$ levels. Since then, the stock has marginally recovered. According to the median estimate of the uncertainty grid, overfishing is no longer occurring in the latest year of the model, though the stock remains in an overfished state and at approximately 25% of $SB_{F=0,t}$ levels.

8 Discussion

8.1 General remarks

There are a number of factors that contribute to the challenges associated with assessing SWPO striped marlin, and to the high levels of uncertainty presented in this assessment. Striped marlin are almost exclusively caught by longline fisheries which select for the largest, oldest individuals in the population. As a result there is very little information on the smallest individuals, which from an assessment standpoint provide a wealth of information. Modal progression in the size frequency data of juvenile fish entering and moving through the population can provide information on age-specific processes such as growth, mortality, and recruitment variability. However, limited amounts of length and weight composition information exist for SWPO striped marlin, relative to the tropical tunas. Even fewer data exist for small individuals which make these age-specific quantities difficult to estimate. The inability to estimate these quantities internal to MULTIFAN-CL places extra importance on having a good biological understanding of the species to inform model assumptions, and reduce uncertainty.

Unfortunately, a sound understanding of key biological processes is lacking for striped marlin. Part of the reason is the highly-migratory nature of the species which makes it challenging to study and appropriately sample. Additionally, striped marlin is largely a bycatch species. The limited research funding and pelagic research programs that exist are usually directed, understandably, towards the more economically valuable tropical tuna species. Furthermore, recent genetic research further suggests that SWPO striped marlin form a genetically distinct stock (Mamoozadeh et al., 2018). As a result, future studies to resolve the biological uncertainties should focus on sampling individuals in the SWPO as data collected from individuals in the northern WCPO and EPO may not be representative.

Despite these biological uncertainties, the model is still able to estimate a decline in stock status. One of the principles of population dynamics is that sustained levels of exploitation lead to the erosion of age structure in the population. This erosion is often manifested as a decrease in the median length or weight of the population over time (Beverton and Holt, 1957). As shown in the likelihood profile, the most influential data component in the assessment is the continuous record from the beginning of the assessment period of landed catch-at-weight data from the New Zealand recreational fishery. These data show a decline from the beginning of the assessment period (Figure 51), which the model attributed to fishing mortality. It is possible that since 1988, further decline in median weight has been masked by a voluntary policy by this fishery of releasing striped marlin weighing less than 90 kg (Holdsworth et al., 2019).

From a behavioral and reproductive standpoint, striped marlin show different patterns to the tropical tunas. Typically solitary individuals, striped marlin are believed to spawn on an annual cycle in spatially confined aggregations. These aggregations are believed to occur across a wide band around 20° (particularly in the western part of sub-region 2) during the 4^{th} quarter of the year (Kopf et al., 2012). This aggregating behavior during the spawning season has implications from a fisheries perspective. Bringing together widely dispersed individuals could increase their catchability to the gear, and lead to smaller amounts of effort having disproportionately large impacts on the stock. This provides context to some of the stock assessment findings. The seasonal catchability patterns are highest in the 4^{th} quarter of the calendar year indicating a greater rate of interaction with the gear during the time period when spawning aggregations are believed to occur. Unintentional or directed targeting of the spawning aggregation during this time period could result in high numbers of removals and an associated decline in spawning biomass.

An investigation of the historical pattern of fisheries removals appears to support this. The large catches in the 1950s and 1960s by DWFNs, coinciding with increasing levels of depletion, occur primarily in the 4^{th} quarter of the year in sub-region 2 making it likely that increased catchability during the spawning season played a role. Later, the period of high fishing mortalities in the late 1990s through the 2000s coincides with the development of the Australian longline fishery in this same area. This same seasonal pattern in catch is strongly evident in the sub-region 2 component of the fishery but not as evident in the sub-region 3 component, which is south of where the major spawning activity is believed to occur. Additionally, the weight composition data for the Australian longline fishery in sub-region 2 shows a seasonal increase in the median weight of individuals caught during the 4^{th} quarter of the year which is consistent with an influx of larger, more mature individuals coming together to form spawning aggregations. Though there are a number of other factors that also likely contributed (such as longline expansion and development in sub-region 1), the pattern in the fishery described here provides some context to the declines estimated in the current assessment.

8.2 Improvements to the assessment

Numerous developments to MULTIFAN-CL, improvements to stock assessment best practices, and data analytical advances occurred since the previous SWPO striped marlin stock assessment in 2012. Two key improvements have been the modeling of maturity as a length specific process, and the use of a geostats model to standardize the CPUE for the major DWFN longline fisheries. Though growth was not estimated within this assessment, modeling maturity as a length-based process allows for the maturity-at-age ogive used within MULTIFAN-CL to be correctly specified if growth is estimated internal to the model. Compared to the previous assessment, there was substantially less conflict between the standardized CPUE indices estimated in the different model sub-regions and the assessment outputs among models fitting to these different CPUE indices. This diminished the amount of uncertainty propagated by this particular axis.

Smaller improvements included re-binning the weight composition data, defining natural mortality as an age-specific process and updating the method for calculating depletion in the assessment model. Not only did re-binning the weights allow the model to converge to a more optimal solution, but it had the added benefit of improving the computational efficiency of the MULTIFAN-CL model. Updating the method of depletion calculation to use the equilibrium unexploited recruitment deviates as the basis for computing dynamic unfished spawning biomass, was one of a number of small changes made to bring the SWPO striped marlin stock assessment more in line with current WCPO stock assessments. These changes included redefining the period over which the SRR was calculated, fixing the terminal recruitments at the mean value, and scaling the effort deviates to be proportional to the square root of effort for fisheries with nominal effort.

8.3 Uncertainty

Uncertainty in model estimates was accounted for using a structural grid approach with different axes representing the major sources of uncertainty. As was found in the previous assessment, lack of knowledge on key biological processes contributed to the overall level of uncertainty in the estimates of management quantities and stock status. The three life-history traits (growth, natural mortality, and steepness) included in the grid are all linked. Previous meta-analysis for striped marlin in the northern WCPO which estimated M (Piner and Lee, 2011) or steepness (Brodziak et al., 2015) rely on life-history and/or reproductive ecology theory which rely in turn, on growth. It is therefore critical to reduce uncertainty in growth estimates and in particular to provide estimates

representative of SWPO striped marlin.

Recent growth work on swordfish in the WCPO discovered that relative to otolith aging, determining age using hard-parts such as fin spines could underestimate the age of the fish (Farley et al., 2016). Additionally, research on tropical tunas has shown that estimating age using daily increment counts can also underestimate the age of older, larger individuals as the increments become closer together and more difficult to differentiate (Williams et al., 2013). In light of this, growth for SWPO striped marlin should be revisited and alternative aging methods including otolith aging explored. The current best scientific estimate of growth for SWPO striped marlin utilizes daily increment counts of otoliths to age small individuals (100-200 cm, EFL), and fin spine aging to age large striped marlin. If either or both of these components systematically underestimate age, it would overestimate the growth rate and the implied productivity of the stock. This could then change the assumptions regarding appropriate levels for natural mortality or steepness.

One of the additional sources of uncertainty in the grid was the relative weight of the composition data in the likelihood. Some tension did exist between the two main likelihood components, weight composition and CPUE, and different weightings were given to the composition data to try and allow for a better fit to the CPUE. Changing the weightings did result in slight differences in the estimated quantities which contributed to overall model uncertainty. In order to reduce the uncertainty from this axis, future assessments should explore the application of "self-scaling" in order to assign a more statistically appropriate weight to the composition data (Francis, 2017).

Though not explicitly modeled using MULTIFAN-CL, the implicit spatial structure of the model using "fleets-as-areas" contributed to model uncertainty via the estimated shape of the fisheries selectivity curves. Though all fisheries in the model either used gear that selected for (longline) or specifically target (recreational anglers) larger individuals, all fisheries except the NZ 3 REC were defined with cubic spline selectivities in order to account for the spatiotemporal patterns observed in the composition data. Within the uncertainty grid, the more optimistic estimates typically estimated selectivity curves that were more dome-shaped than asymptotic. Dome-shaped selectivity curves resulted in large estimates of "cryptic biomass", as individuals became invulnerable to fishing mortality beyond a certain size. With the appropriate spatial structure, and either better estimates of movement or better tagging data to estimate movement internal to MULTIFAN-CL; the selectivity functions for the longline fisheries could be redefined in terms of gear selectivity rather than the combination of gear selectivity and availability. Redefining these curves as asymptotic would reduce the uncertainty caused by selectivity-driven cryptic biomass and would allow for the spatiotemporal patterns seen in the composition data to be more realistically modeled as a combination of growth, movement, and fishing mortality.

8.4 Stock assessment conclusions

The main conclusions of the current assessment are based upon the 300 plausible models that made up the structural uncertainty grid. The general conclusions of this assessment are summarized as follows:

- Despite the high levels of uncertainty shown in the grid, a consensus of models indicated a clear, persistent decline in spawning biomass and increase in fishing mortality from the beginning of the model period, which is consistent with what was found in the previous assessment. Fishing mortality increased in the 2000s to a peak at the beginning of the current decade (2010s) which also coincided with the lowest estimates of spawning biomass. Recent years show a slight improvement in stock status relative to the early 2010s.

- The negative trend in recruitment identified in the previous two stock assessments remains a feature of the current model. Recruitment variability appears to have reduced in the last decade. Limited composition data for very small fish exists within the model to inform the recruitment deviations, and they are likely poorly estimated.
- The WCPFC has yet to formally agree to a limit reference point (LRP) for SWPO striped marlin, therefore the main stock assessment results are presented in terms of both MSY-based and spawning biomass depletion based reference points. Looking at the recent period, 69% of runs indicate that $SB_{recent}/SB_{\text{MSY}}$ is less than 1 (median = 0.737; range 0.152 – 3.312), and 50% of runs indicate that $SB_{recent}/SB_{F=0}$ is less than 20% of the unfished condition (median = 0.198; range 0.038 – 0.977). These results are marginally more optimistic in the terminal (2017) year of the assessment, 61% of runs indicate that $SB_{latest}/SB_{\text{MSY}}$ is less than 1 (median = 0.898; range 0.174 – 3.924), and 39% of runs indicate that $SB_{latest}/SB_{F=0}$ is at less than 20% of the unfished condition (median = 0.238; range 0.044 – 1.158).
- With respect to fishing mortality, median $F_{recent}/F_{\text{MSY}}$ was estimated across grid models to be 0.911 (range 0.030 – 3.5) with 44% of runs greater than 1. However, given the potential effects of the fixed recruitment on the terminal estimates of F, the window over which $F_{recent}/F_{\text{MSY}}$ is defined may need to be reconsidered.
- Based on these findings, this assessment concludes that the SWPO striped marlin stock is likely overfished, and close to undergoing overfishing according to MSY-based reference points.
- Though recent research indicates that this population is a genetically distinct stock, some connectivity with the northern WCPO, the EPO and the eastern Indian Ocean stocks has been shown (Lam et al., 2019; Mamoozadeh et al., 2018). The results of the current SWPO assessment could have implications for other components of the striped marlin metapopulation structure.

8.5 Research recommendations

In order to progress the assessment of SWPO striped marlin, it is recommended that future research investigates the following topics:

- As detailed in Section 8.3, future research should focus on resolving the biological uncertainty surrounding natural mortality, and growth. This can involve: dedicated tagging studies to estimate natural mortality, collection of catch-at-age data, or focused (either via onboard observers or a fisheries independent survey) sampling in the tropical region of the model to improve collection of size frequency information on small individuals. Collection of otoliths and hard-parts for aging should also pe prioritized as future work is also needed to validate the existing method of aging striped marlin and to further develop the estimation of growth using otolith ages. Finally, meta-analyses for natural mortality and steepness should be updated to include data inputs specific to the SWPO striped marlin stock.
- Continued development of length-based selectivity functions in MULTIFAN-CL is needed. Modeling selectivity as a function of length using a cubic spline could lead to better estimates of age related processes such as fishing mortality.
- As the amount of composition data available to the assessment model increases, it may be possible to revisit the fisheries definitions particularly for the mixed-longline fisheries 11-14. Doing so may better explain the patterns seen in the catchability estimates and fits to the

size composition data.

- In order to properly develop a spatial stock assessment model for SWPO striped marlin, better estimates of movement are required. An emphasis should be placed on characterizing long term movement patterns ($>$ 180 days) to identify mixing rates across parts of the model region. Additionally, any attempts to assess striped marlin in a meta-population framework will need a better understanding for the level of mixing that occurs between the different sub-stocks.
- Given the evidence for sexual dimorphism with respect to weight-at-length and maturity, a sex-disaggregated model remains an area of focus. Improving the data collection with respect to sex-specific size composition data, natural mortality, and movement could improve the performance of the sex-disaggregated model.
- Particularly for a bycatch species like striped marlin, data collection of demographic information (length, weight, sex, and location of capture) from the recreational component of the fishery provides a critical input to the assessment. Maintaining and improving the existing data collection process, particularly information on landed and released individuals is important. Collection of otoliths and histological samples from landed individuals would be a useful addition. Additionally, total catch estimates are not available for the Australian component of the fishery beyond 2011, attempts should be made to remedy this prior to the next assessment.
- Though it was not addressed in this report, changing oceanographic conditions in the SWPO as a result of climate change could alter the distributional assumptions made regarding this species. Efforts should be made to increase sampling (or data availability for the assessment from the southern bluefin tuna longline fisheries) in the southern portion of the species range (south of 40°S) to identify any changes in the distribution.

9 Acknowledgments

The 2019 SWPO striped marlin stock assessment relied on input from a number of organizations and individuals in order to come together in its final form. We thank the various fisheries agencies for the provision of the catch, effort and size frequency data used in this analysis, and are grateful for the work done by fisheries observers throughout the region to collect observations and measurements at sea. We thank R. Campbell for providing standardized CPUE for the Australian longline fisheries and J. Farley for the work done analyzing otolith aging and growth for striped marlin. We thank J. Holdsworth for providing a standardized CPUE index, catch information, weight composition, and tagging information for the New Zealand recreational fishery. We thank D. Ghosn for providing an updated standardized CPUE index for the Australian recreational fishery, and we thank J. Pepperell for collating and providing sex-specific demographic information for individuals captured by the Australian recreational fishery. Additionally, we thank P. Williams and the data management team at SPC for their work compiling the data inputs for the assessment. S. McKechnie provided valuable feedback throughout the assessment process.

References

Ahrens, R. (2010). *A global analysis of apparent trends in abundance and recruitment of large tunas and billfishes inferred from Japanese longline catch and effort data*. PhD thesis, University of British Columbia, Vancouver.

Berger, A., Pilling, G., Kirchner, C., and Harley, S. (2013). Determination of appropriate time-windows for calculation of depletion-based limit reference points. Technical Report WCPFC-SC9-2013/MI-WP-02, Pohnpei, Federated States of Micronesia.

Beverton, R. J. H. and Holt, S. J. (1957). On the dynamics of exploited fish populations. *Fish and Fisheries Series*, 11:1–537.

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. (2012). Meta-analysis of post-release mortality in striped (Kajikia audax) and blue marlin (Makaira nigricans) using pop-up satellite archival tags. Technical Report ISC/12/BILLWG-1/07.

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

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: Journal du Conseil*, 62(2):256–265.

Campbell, R. (2012). Abundance indices for striped marlin and broadbill swordfish in the south-west Pacific based on standardised CPUE from the Australian longline fleet. Technical Report WCPFC-SC8-SA-IP-13.

Campbell, R. (2018). Aggregate and size-based standardised CPUE indices for longline target species caught within the ETBF – 2018 Update. Working paper presented to the 21st meeting of the Tropical Tuna RAG, Mooloolaba,16-17 July 2018.

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.

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., and Hampton, J. (2019). Developments in the MULTIFAN-CL software 2018-2019. Technical Report WCPFC-SC15-2019/SA-IP-02, Pohnpei, Federated States of Micronesia.

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

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

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. and Pilling, G. (2019). Background analyses for the 2019 stock assessment of SW Pacific striped marlin. Technical Report WCPFC-SC15-2019/SA-IP-07, Pohnpei, Federated States of Micronesia.

Farley, J., Clear, N., Kolody, D., Krusic-Golub, K., Eveson, P., and Young, J. (2016). Determination of swordfish growth and maturity relevant to the southwest Pacific stock. Technical Report WCPFC-SC12-2016/ SAWP-11, Bali, Indonesia, 3–11 August 2016.

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, \emphThunnus 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., Nielson, 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 (ıt Hoplostethus atlanticus) on the Chatham Rise, New Zealand. *Canadian Journal of Fisheries and Aquatic Science*, 49:922–930.

Francis, R. I. C. C. (2017). Revisiting data weighting in fisheries stock assessment models. *Fisheries Research*, page 11.

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.

Hampton, J. and Fournier, D. (2001). A spatially-disaggregated, length-based, age-structured population model of yellowfin tuna (\emphThunnus 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., 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.

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.

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

Kolody, D. and Davies, N. (2008). Spatial structure in South Pacific swordfish stocks and assessment models. Technical Report WCPFC-SC4-2008/SA-IP-2, Port Moresby, Papua New Guinea, 11–22 August 2008.

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.

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.

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.

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

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

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).

Ortiz, M., Prince, E. D., Serafy, J. E., Holts, D. B., Davy, K. B., Pepperell, J. G., Lowry, M. B., and Holdsworth, J. C. (2003). Global overview of the major constituent-based billfish tagging programs and their results since 1954. *Marine and Freshwater Research*, 54(4):489.

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.

Pilling, G. and Brouwer, S. (2019). Report from the SPC Pre-assessment Workshop, Noumea, April 2019. Technical Report WCPFC-SC15-2019/SA-IP-01, Pohnpei, Federated States of Micronesia.

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

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.

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.

Takeuchi, Y., Davies, N., Fournier, D., Pilling, G., and Hampton, J. (2018). Testing MULTIFAN-CL developments for multispecies/multi-sex assessments, using SW Pacific swordfish. Technical Report WCPFC-SC14-2018/SA-IP-10, Busan, South Korea, 8-16 August 2018.

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.

Vincent, M., Pilling, G., and Hampton, J. (2019). Stock assessment of skipjack tuna in the WCPO. Technical Report WCPFC-SC15-2019/SA-WP-05, Pohnpei, Federated States of Micronesia.

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.

WCPFC (2013). Report of the Scientific Committee Ninth Regular Session. Summary Report, Commission for the Conservation and Management of Highly Migratory Fish Stocks in the Western and Central Pacific Ocean, Pohnpei, Federated States of Micronesia, 6–14 August 2013.

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

Williams, A., Leroy, B., Nicol, S., Farley, J., Clear, N., Krusic-Golub, K., and Davies, C. (2013). Comparison of daily- and annual- increment counts in otoliths of bigeye (Thunnus obesus), yellowfin (T. albacares), southern bluefin (T. maccoyii) and albacore (T. alalunga) tuna. *ICES Journal of Marine Science*, 70(7):1439–1450.

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.

10 Tables

Table 1: Definition of fisheries for the 2019 MULTIFAN-CL SWPO striped marlin stock assessment.

Fishery	Nationality	Gear	Sub.region	Catch	Effort	Std.CPUE	Years
1: JP 1 LL	JP	Longline	1	Number	Hooks	Y	1952-2017
2: JP 2 LL	JP	Longline	2	Number	Hooks	Y	1952-2017
3: JP 3 LL	JP	Longline	3	Number	Hooks	Y	1952-2017
4: JP 4 LL	JP	Longline	4	Number	Hooks	Y	1954-2017
5: TW 4 LL	TW	Longline	4	Number	Hooks	Y	1967-2017
6: AU 2 LL	AU	Longline	2	Number	Hooks	Y	1987-2017
7: AU 3 LL	AU	Longline	3	Number	Hooks	Y	1987-2017
8: NZ 3 LL	NZ	Longline	3	Number	Hooks	N	1989-2017
9: AU 3 REC	AU	Recreational	3	Number	Days	Y	1952-2017
10: NZ 3 REC	NZ	Recreational	3	Number	Days	Y	1952-2017
11: OTHER 1 LL	DWFN/PICT	Longline	1	Number	Hooks	N	1962-2017
12: OTHER 2 LL	DWFN/PICT	Longline	2	Number	Hooks	N	1964-2017
13: OTHER 3 LL	DWFN/PICT	Longline	3	Number	Hooks	N	1968-2017
14: OTHER 4 LL	DWFN/PICT	Longline	4	Number	Hooks	N	1964-2017

Table 2: Number of total striped marlin length and weight-frequency samples for each of the defined fisheries in the 2019 diagnostic case.

Fishery	Length.Frequency	Weight.frequency
1: JP 1 LL	311	1,701
2: JP 2 LL	853	27,923
3: JP 3 LL	586	12,879
4: JP 4 LL	0	148
5: TW 4 LL	92	0
6: AU 2 LL	1,300	51,791
7: AU 3 LL	1,272	13,103
8: NZ 3 LL	43	0
9: AU 3 REC	0	1,805
10: NZ 3 REC	1,102	20,779
11: OTHER 1 LL	9,377	0
12: OTHER 2 LL	17,539	0
13: OTHER 3 LL	0	0
14: OTHER 4 LL	1,438	0

Table 3: Summary of the groupings of fisheries used within the diagnostic case for estimation of selectivity, catchability, and the effort deviates. See Table 1 for further details on each fishery.

Fishery	Sub.region	Sel.	Sel.Term.Age	Non.Decrease.Pen	q	Seas.q	Temp.q	Temp.q.CV	Eff.Pen	Eff.Pen.CV
1: JP 1 LL	1	1	6	N	1	Y	Y	0.1	scaled	0.50
2: JP 2 LL	2	2	7	N	2	Y	N	0.1	time-variant	0.20
3: JP 3 LL	3	3	7	N	3	Y	Y	0.1	scaled	0.50
4: JP 4 LL	4	4	6	N	4	Y	Y	0.1	scaled	0.50
5: TW 4 LL	4	4	6	N	5	Y	Y	0.1	scaled	0.50
6: AU 2 LL	2	5	7	N	6	Y	Y	0.1	scaled	0.50
7: AU 3 LL	3	6	7	N	7	Y	Y	0.1	scaled	0.50
8: NZ 3 LL	3	6	7	N	8	Y	Y	0.1	scaled	0.71
9: AU 3 REC	3	7	7	N	9	N	Y	0.1	scaled	0.50
10: NZ 3 REC	3	8	9	Y	10	N	Y	0.1	scaled	0.50
11: OTHER 1 LL	1	1	6	N	11	N	Y	0.1	scaled	0.71
12: OTHER 2 LL	2	9	7	N	12	N	Y	0.1	scaled	0.71
13: OTHER 3 LL	3	3	7	N	13	N	Y	0.1	scaled	0.71
14: OTHER 4 LL	4	4	6	N	14	N	Y	0.1	scaled	0.71

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 2013–2016 and for SB is the average over the period 2014–2017 and "latest" is 2017.

Symbol	Description
C_{latest}	Catch in the last year of the assessment (2017)
F_{recent}	Average fishing mortality-at-age for a recent period (2013–2016)
YF_{recent}	Equilibrium yield at average fishing mortality for a recent period (2013–2016)
fmult	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}
$F_{recent}/F_{\mathrm{MSY}}$	Average fishing mortality-at-age for a recent period (2014–2017) relative to F_{MSY}
SB_0	Equilibrium unexploited spawning potential
SB_{latest}	Spawning biomass in the latest time period (2017)
SB_{recent}	Spawning biomass for a recent period (2014–2017)
$SB_{F=0}$	Average spawning biomass predicted in the absence of fishing for the period 2007–2016
SB_{MSY}	Spawning biomass that will produce the maximum sustainable yield (MSY)
$SB_{MSY}/SB_{F=0}$	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 2007–2016
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 (2017) relative to the equilibrium unexploited spawning potential
$SB_{latest}/SB_{F=0}$	Spawning biomass in the latest time period (2017) relative to the average spawning biomass predicted to occur in the absence of fishing for the period 2007–2016
$SB_{latest}/SB_{\mathrm{MSY}}$	Spawning biomass in the latest time period (2017) relative to that which will produce the maximum sustainable yield (MSY)
$SB_{recent}/SB_{F=0}$	Spawning biomass for a recent period (2014–2017) relative to the average spawning biomass predicted to occur in the absence of fishing for the period 2007–2016
$SB_{recent}/SB_{\mathrm{MSY}}$	Spawning biomass for a recent period (2014–2017) relative to the spawning biomass that produces maximum sustainable yield (MSY)

Table 5: Description of the structural sensitivity grid used to characterise uncertainty in the assessment. Levels used under the diagnostic case are starred.

Axis	Levels	Option
Steepness	3	0.65, 0.8* or 0.95
Growth	2	Kopf et al. 2011* or otolith age
Natural mortality	3	0.3, 0.4* or 0.5
CPUE	3	JP 2 LL*, TW 5 LL or AU 6 LL
Size frequency weighting	3	Weight/length samples divided by 10/20, 20/40* or 50/100
Recruitment penalty CV	3	0.2*, 0.5 or 2.2

Table 6: Summary of reference points over the 300 models retained in the 2019 structural uncertainty grid.

	Mean	Median	Min	10%	90%	Max
C_{latest}	1124	1130	1065	1077	1165	1197
YF_{recent}	1966	1920	235	1488	2655	3044
fmult	1.895	1.098	0.286	0.529	3.191	33.180
F_{MSY}	0.259	0.241	0.152	0.172	0.357	0.466
MSY	2672	2039	1742	1845	3535	23710
$F_{recent}/F_{\text{MSY}}$	1.029	0.911	0.030	0.313	1.891	3.500
SB_0	16142	13195	7038	8944	22790	101400
$SB_{F=0}$	12205	10759	5450	7039	19060	44940
SB_{MSY}	3620	3032	960	1396	6109	20890
SB_{MSY}/SB_0	0.221	0.228	0.121	0.140	0.291	0.304
$SB_{MSY}/SB_{F=0}$	0.281	0.271	0.159	0.181	0.368	0.621
SB_{latest}/SB_0	0.209	0.196	0.051	0.100	0.342	0.499
$SB_{latest}/SB_{F=0}$	0.294	0.238	0.044	0.106	0.533	1.158
$SB_{latest}/SB_{\text{MSY}}$	1.062	0.898	0.174	0.383	1.979	3.924
$SB_{recent}/SB_{F=0}$	0.247	0.198	0.038	0.093	0.464	0.977
$SB_{recent}/SB_{\text{MSY}}$	0.895	0.737	0.152	0.334	1.635	3.312

11 Figures

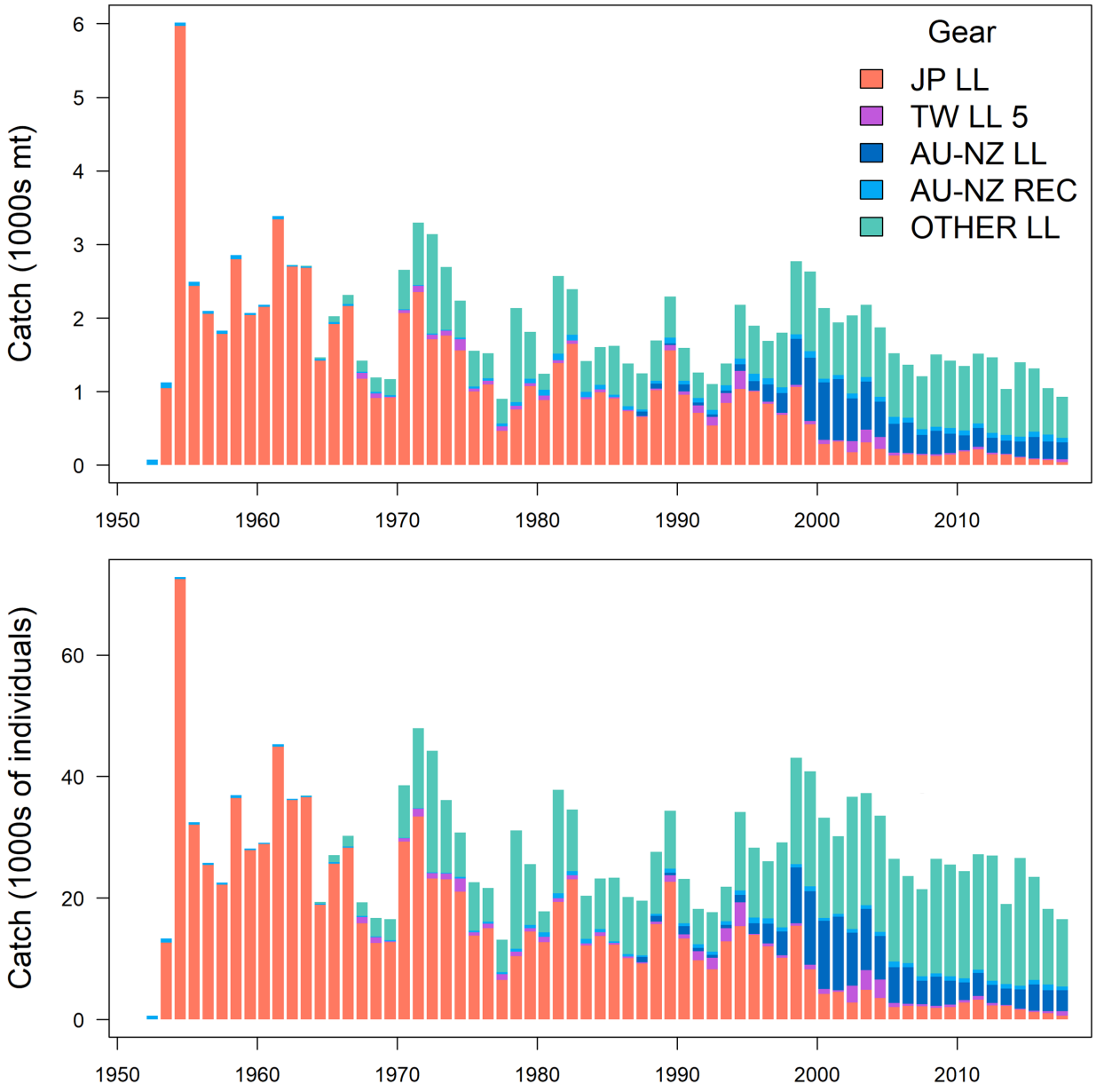

Figure 1: Total striped marlin catches (biomass – top; biomass – bottom) grouped by major fisheries in the model region, 1952–2017: JP LL - Japanese longline (red), TW LL 5 – Chinese Taipei longline sub-region 4 (purple), AU-NZ LL – Australia and New Zealand longline sub-regions 2 & 3 (dark blue), AU-NZ REC - Australia and New Zealand recreational sub-region 3 (light blue), OTHER LL – distant water fishing nation and Pacific Island country and territory longline (teal).

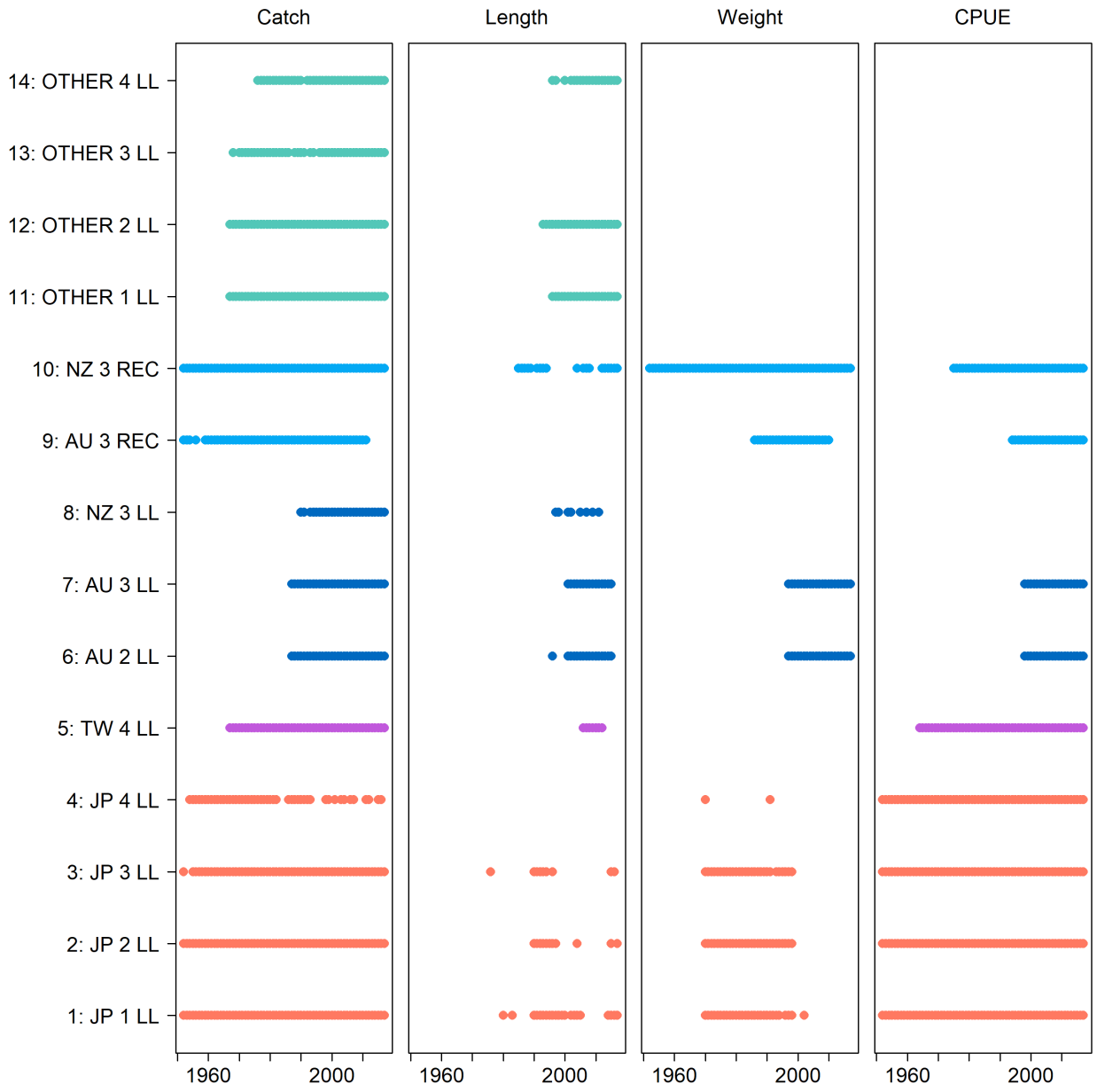

Figure 2: Presence of catch, length frequency, weight frequency, and standardized CPUE by year and fishery for the 2019 diagnostic case model. The different colors correspond to the fleet groups defined in (Figure 1).

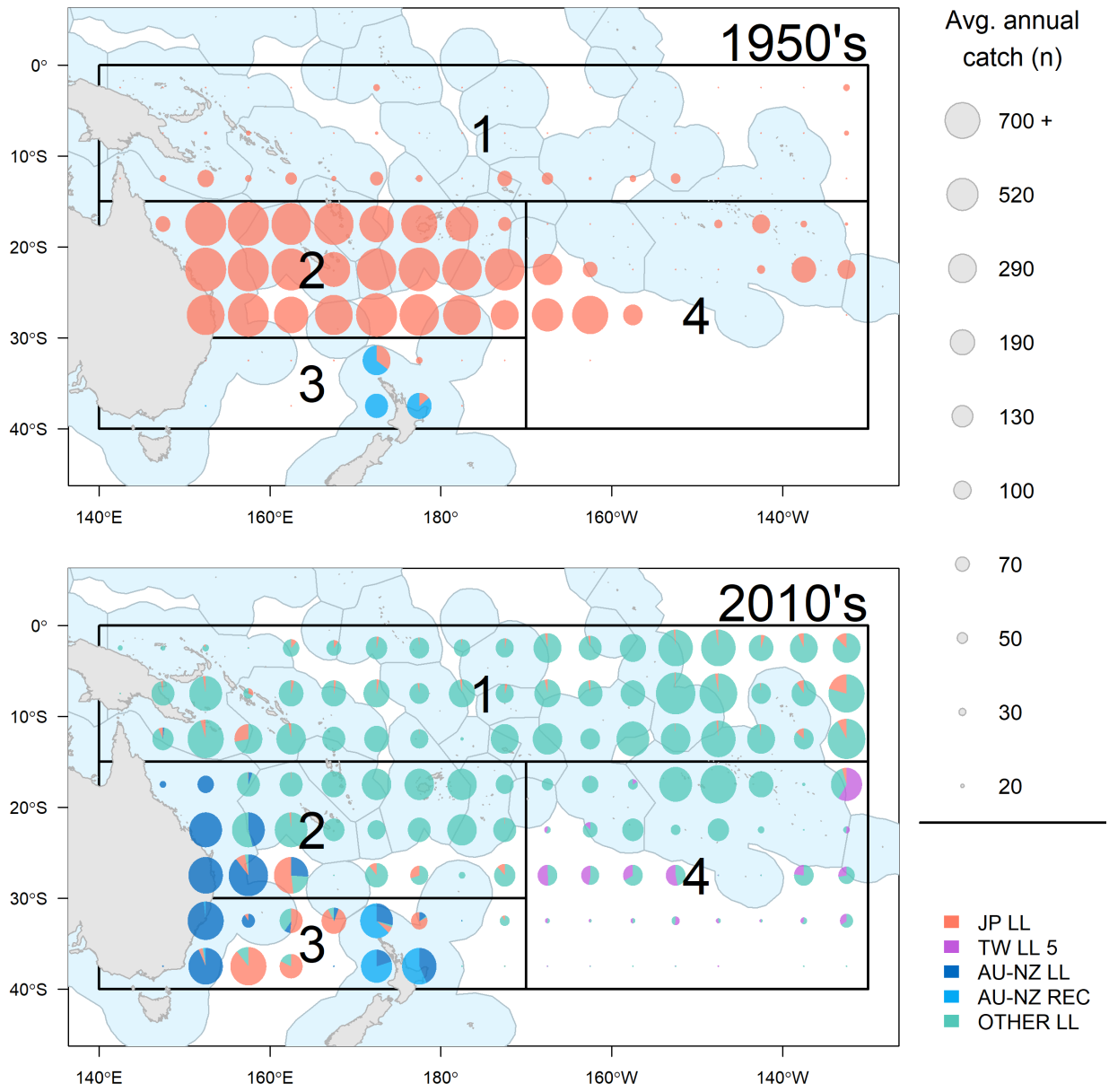

Figure 3: Average annual catches of striped marlin in the SWPO by 5°×5° cell, during the 1950s (top panel) and the 2010s (bottom panel) indicating the large shift in fisheries composition over time. The black lines represent the boundaries of the assessment region (outer lines) for striped marlin in the southwest Pacific Ocean and the four sub-regions used to define the fisheries.

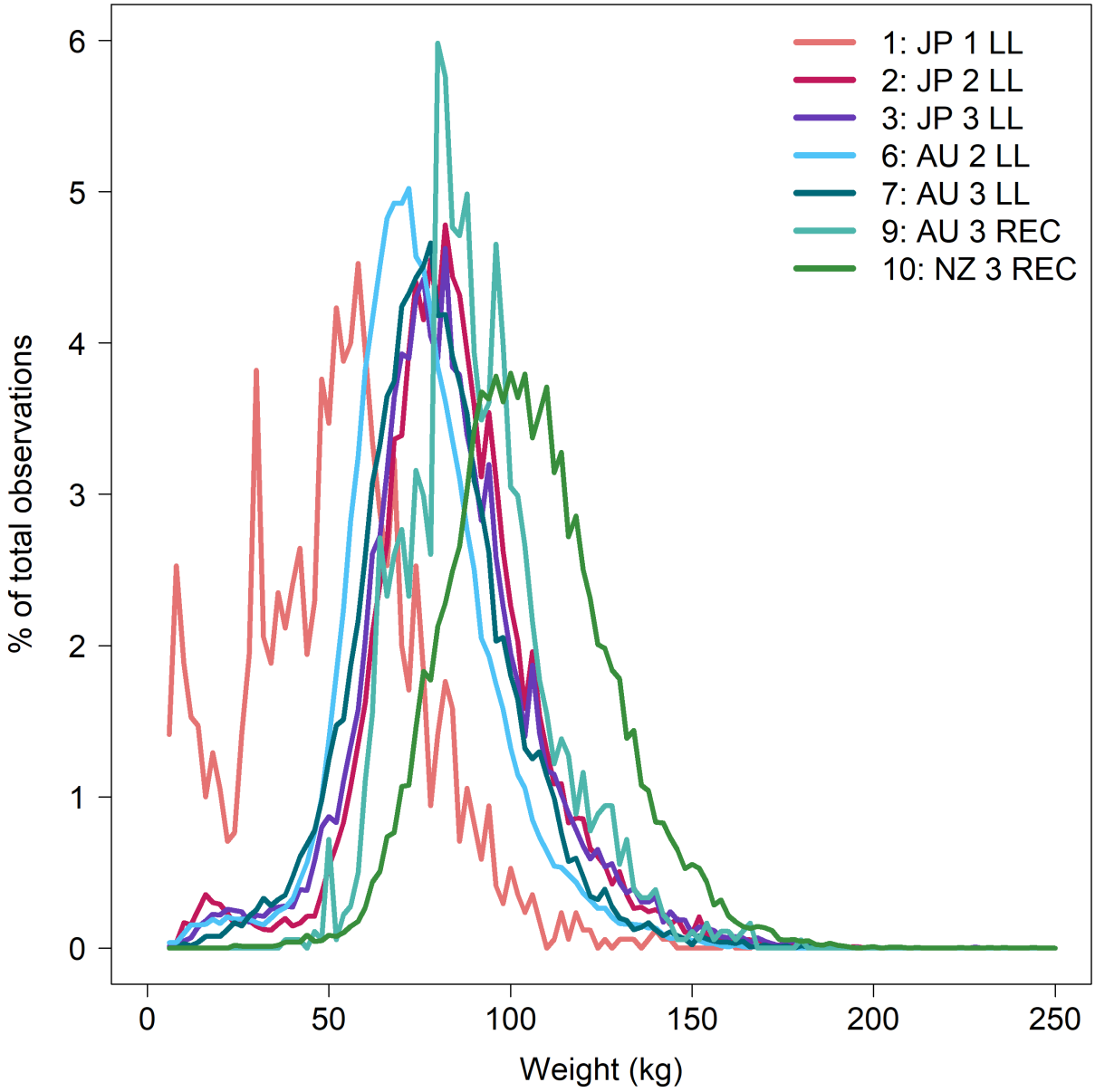

Figure 4: A comparison of the weight (whole weight, kilograms) frequency distributions of the sampled catches from the main fisheries, all years combined.

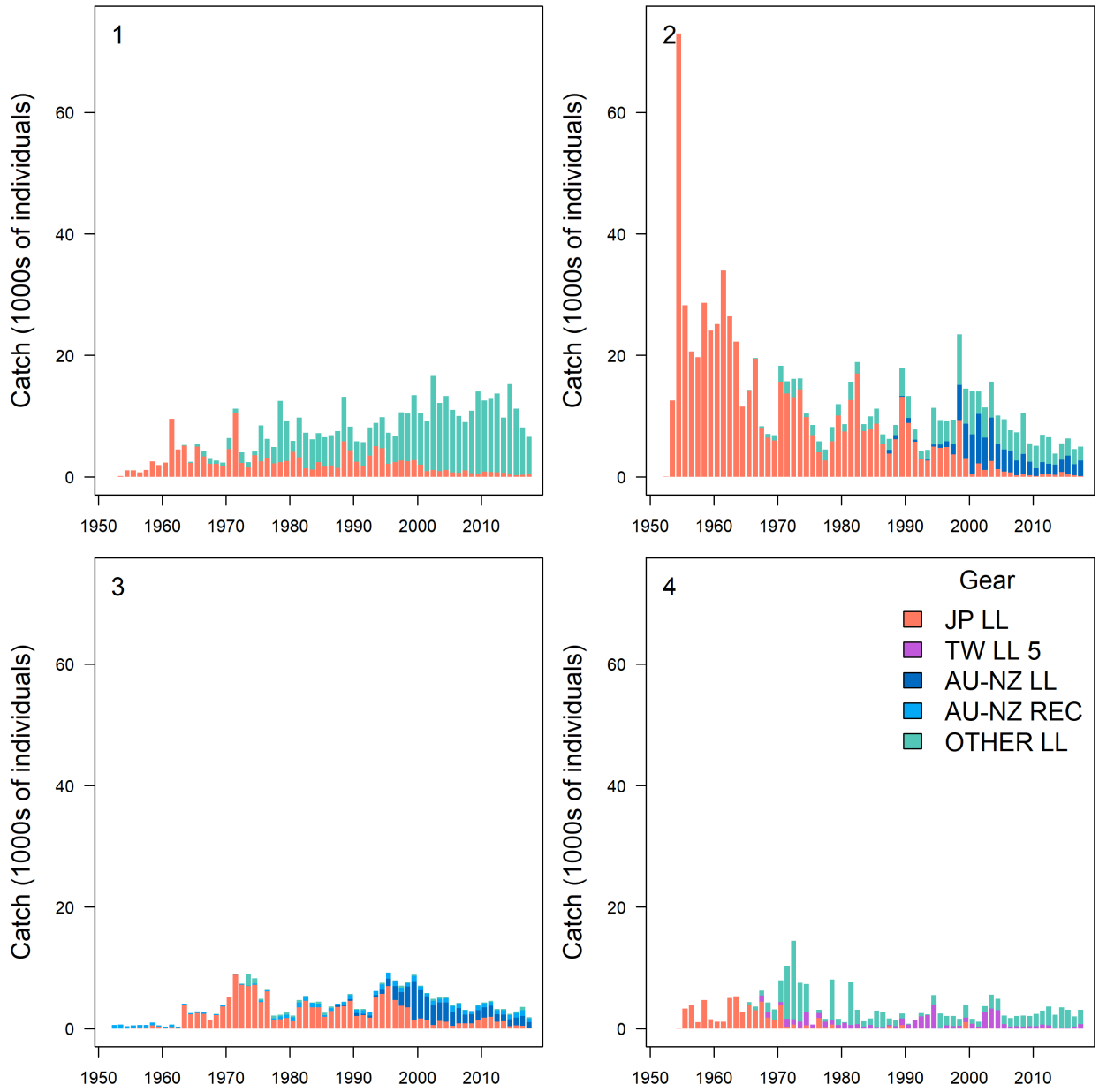

Figure 5: Total striped marlin catches (numbers) grouped by major fisheries in the model sub-regions, 1952–2017: JP LL - Japanese longline (red), TW LL 5 – Chinese Taipei longline sub-region 4 (purple), AU-NZ LL – Australia and New Zealand longline sub-regions 2 & 3 (dark blue), AU-NZ REC - Australia and New Zealand recreational sub-region 3 (light blue), OTHER LL – distant water fishing nation and Pacific Island country and territory longline (teal).

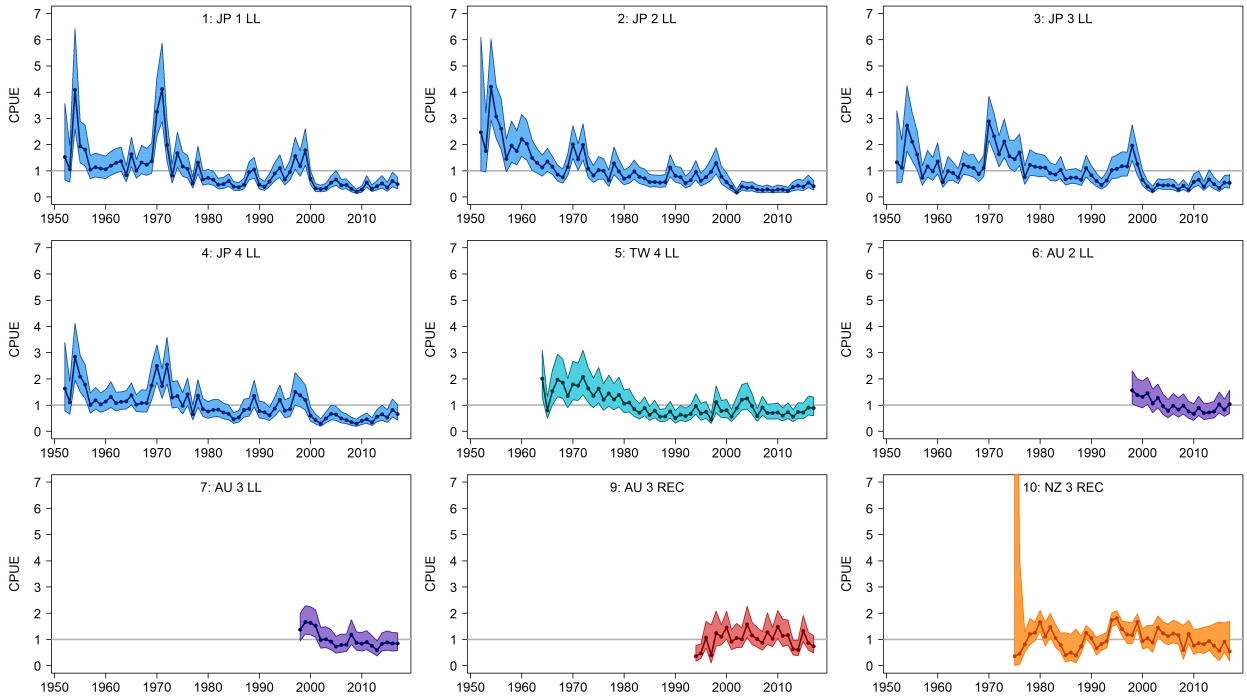

Figure 6: Standardized catch-per-unit-effort (CPUE) indices for the 9 fisheries with available standardized CPUE. The shaded band represents the 95% confidence intervals around the standardized index derived from the time variant precision used to define the effort deviation penalties for the diagnostic case model.

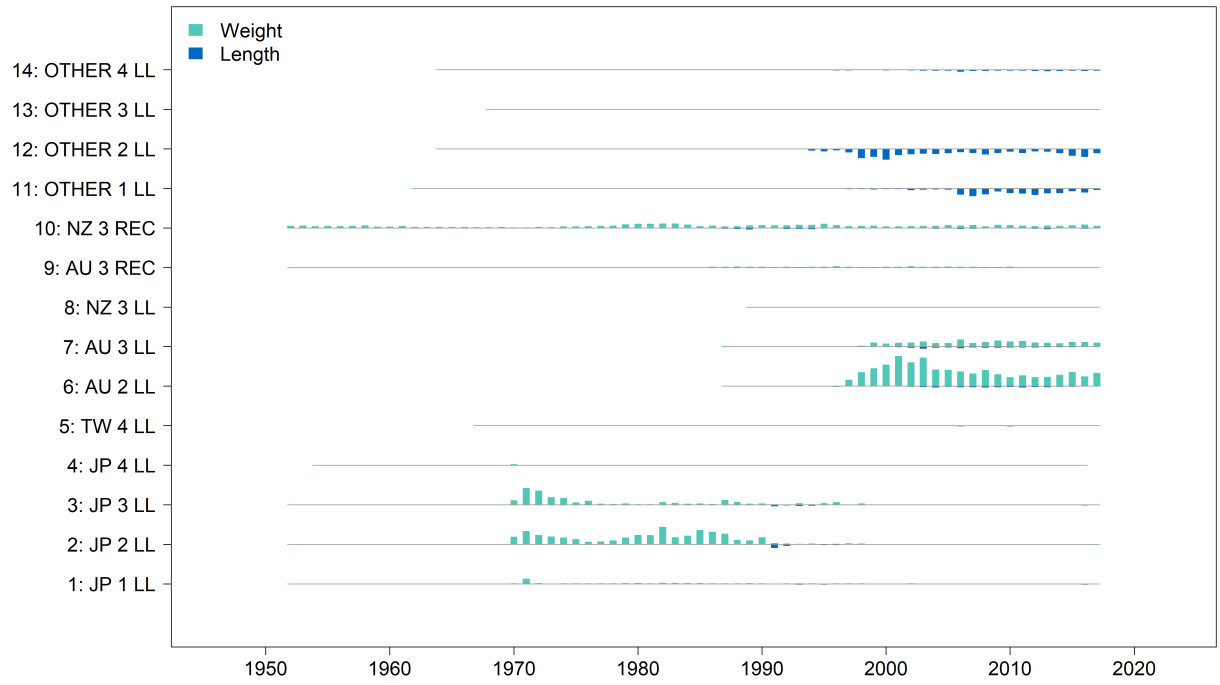

Figure 7: Number of fish size measurements by year for each fishery. The blue bars represent length measurements and the green bars represent weight measurements. The maximum bar length is 4961 fish. The extent of the horizontal lines indicates the period over which each fishery occurred.

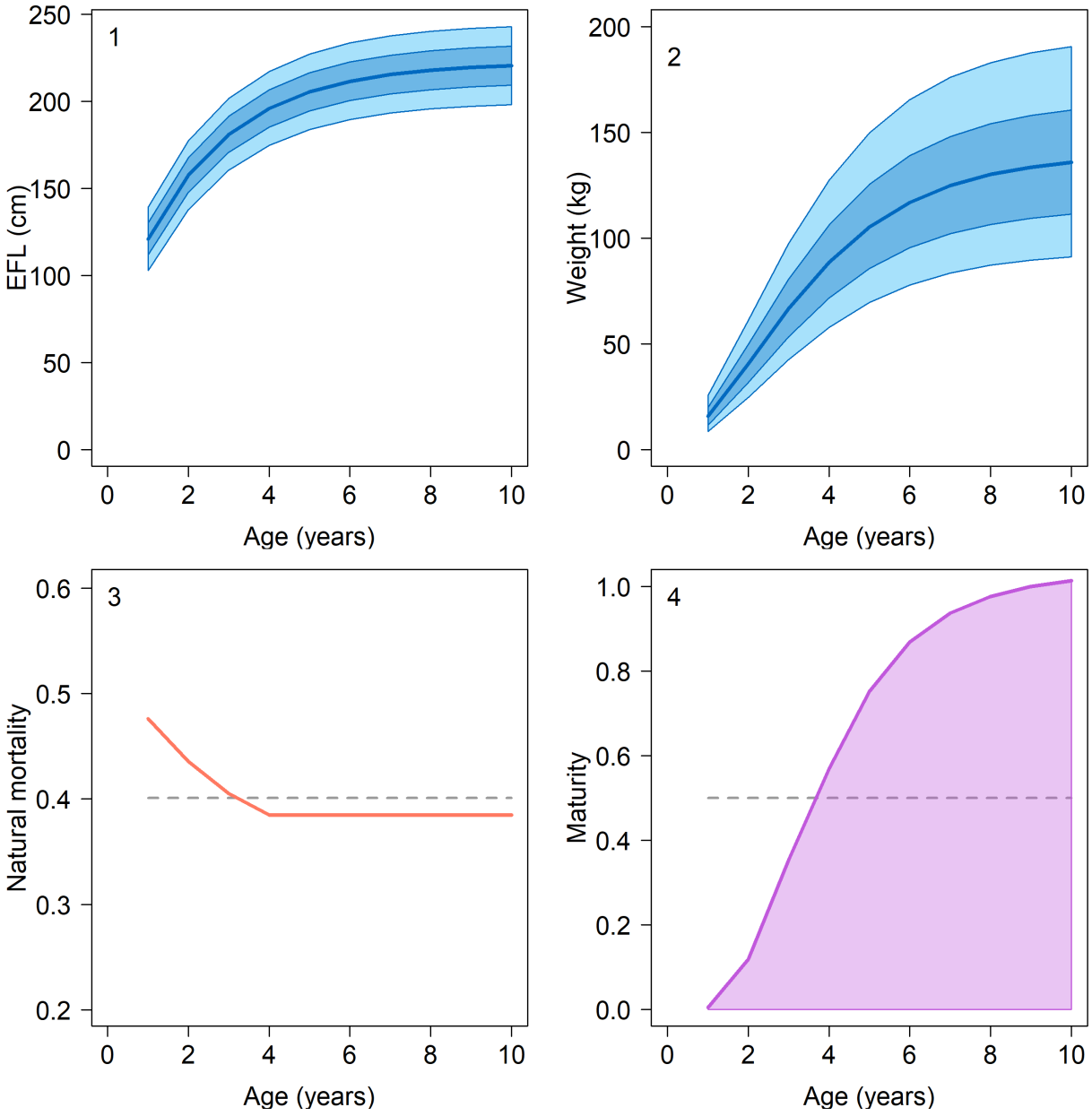

Figure 8: Biological assumptions made in the 2019 diagnostic case: length-at-age (panel 1), weight-at-age (panel 2), natural mortality-at-age (panel 3), and maturity-at-age (panel 4). For length-at-age and weight-at-age the model estimated uncertainty in growth is shown as ± 1 and ± 2 standard deviations from the mean. For natural mortality-at-age the dashed line indicates the mean mortality-at-age and for maturity-at-age the dashed line indicates the 50^{th} percentile of the ogive.

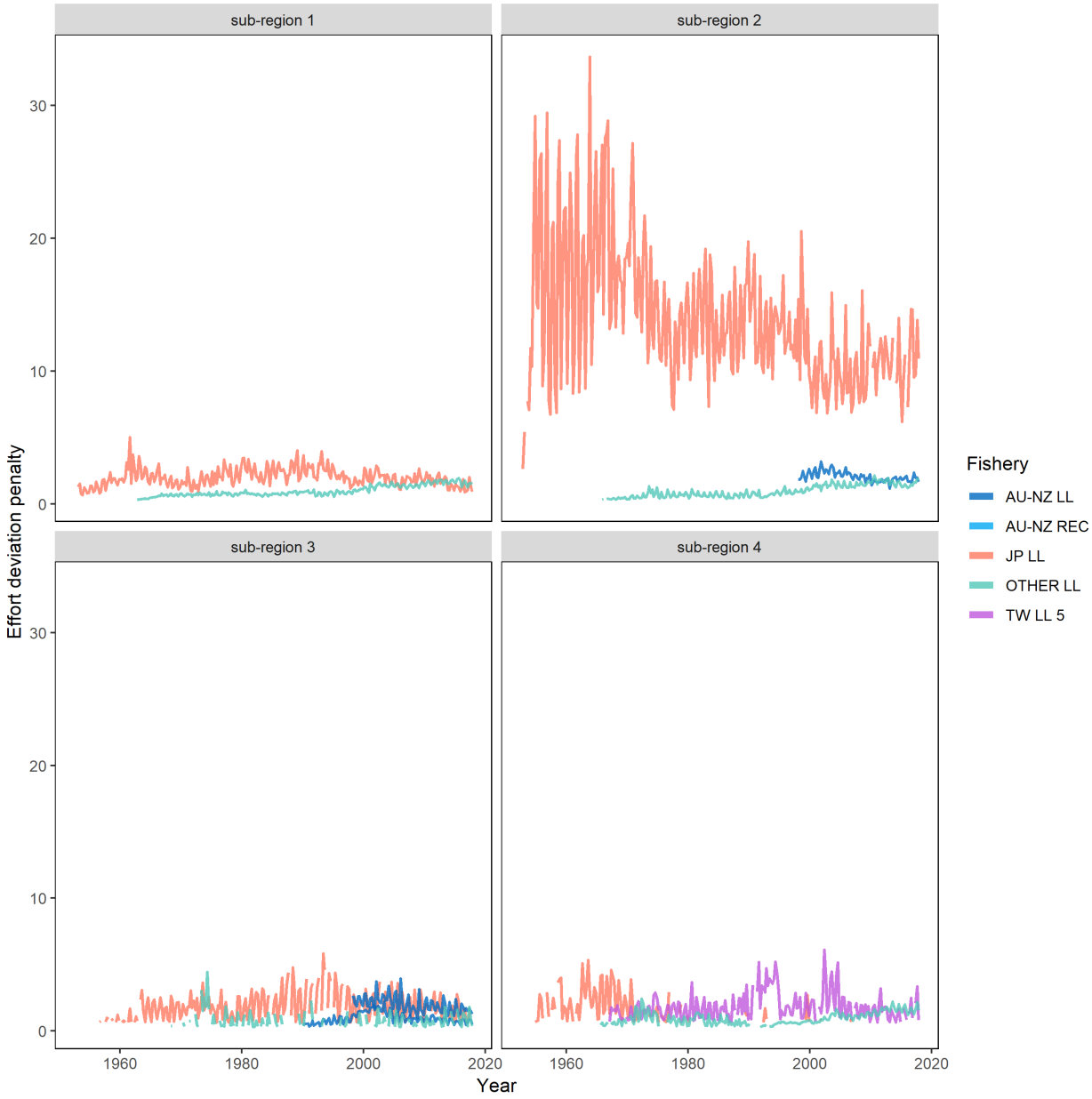

Figure 9: Plot of the effort deviation penalties applied to each fishery, by sub-region, with the colors of the lines representing the gear of the fishery. A higher penalty gives more weight to the CPUE of that fishery. The high weighting applied to the standardized CPUE of the index fishery used in the diagnostic case (JP 2 LL) is apparent.

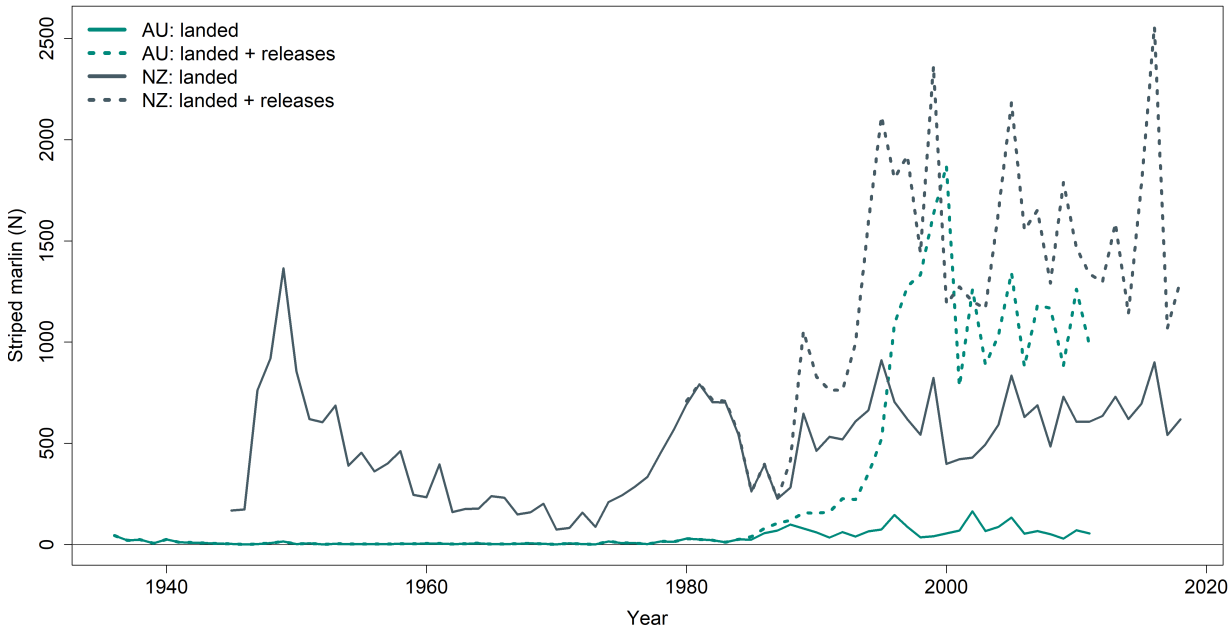

Figure 10: Annual numbers of striped marlin that are landed and or subsequently tagged and released by the two major recreational fisheries operating in the assessment region: Australia (teal) and New Zealand (gray).

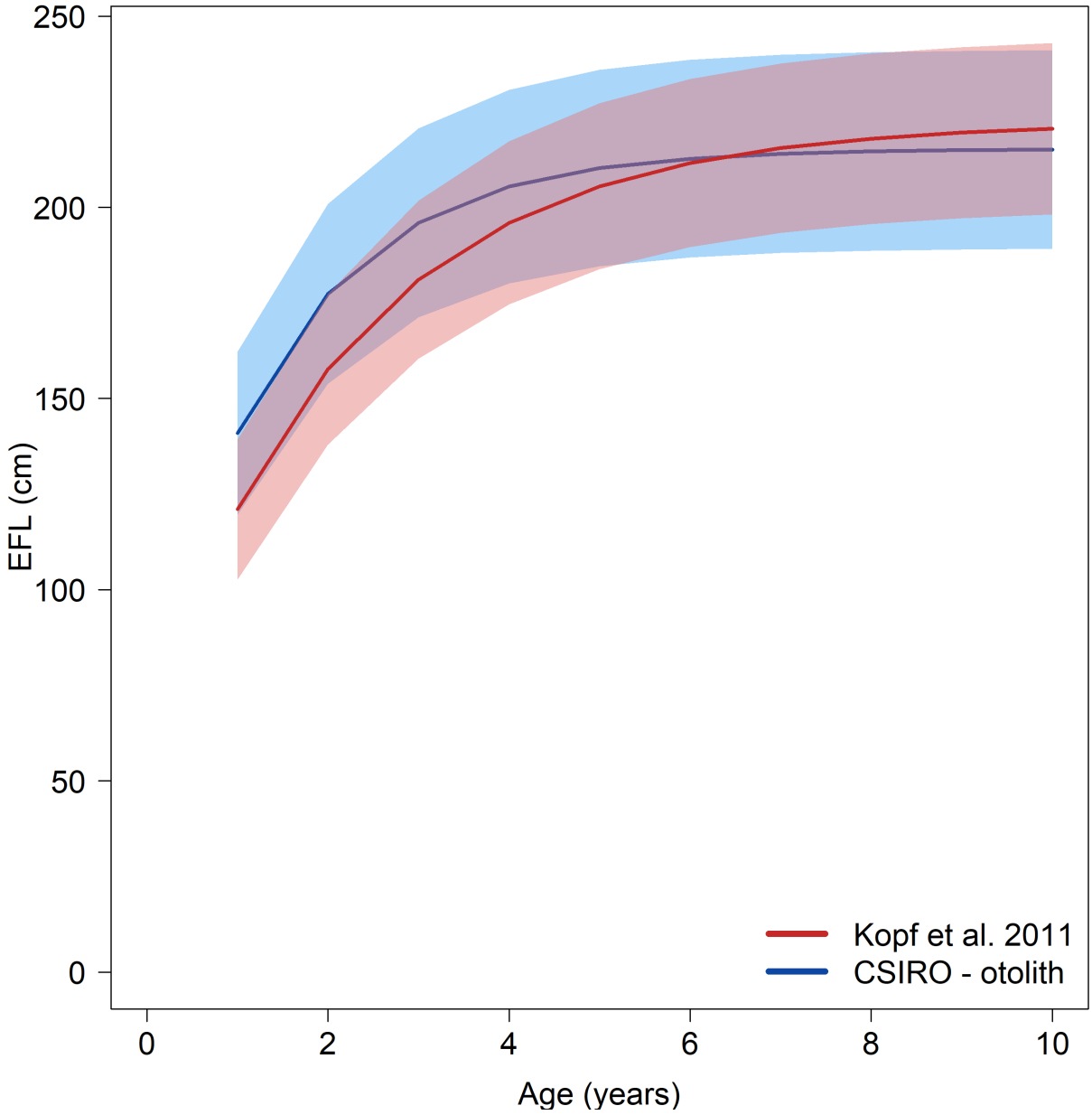

Figure 11: Growth options considered in the structural uncertainty grid along with the 95% confidence interval. The growth used in the diagnostic case (Kopf et al., 2011) is shown in red and the growth function developed by CSIRO using otolith aging is shown in blue.

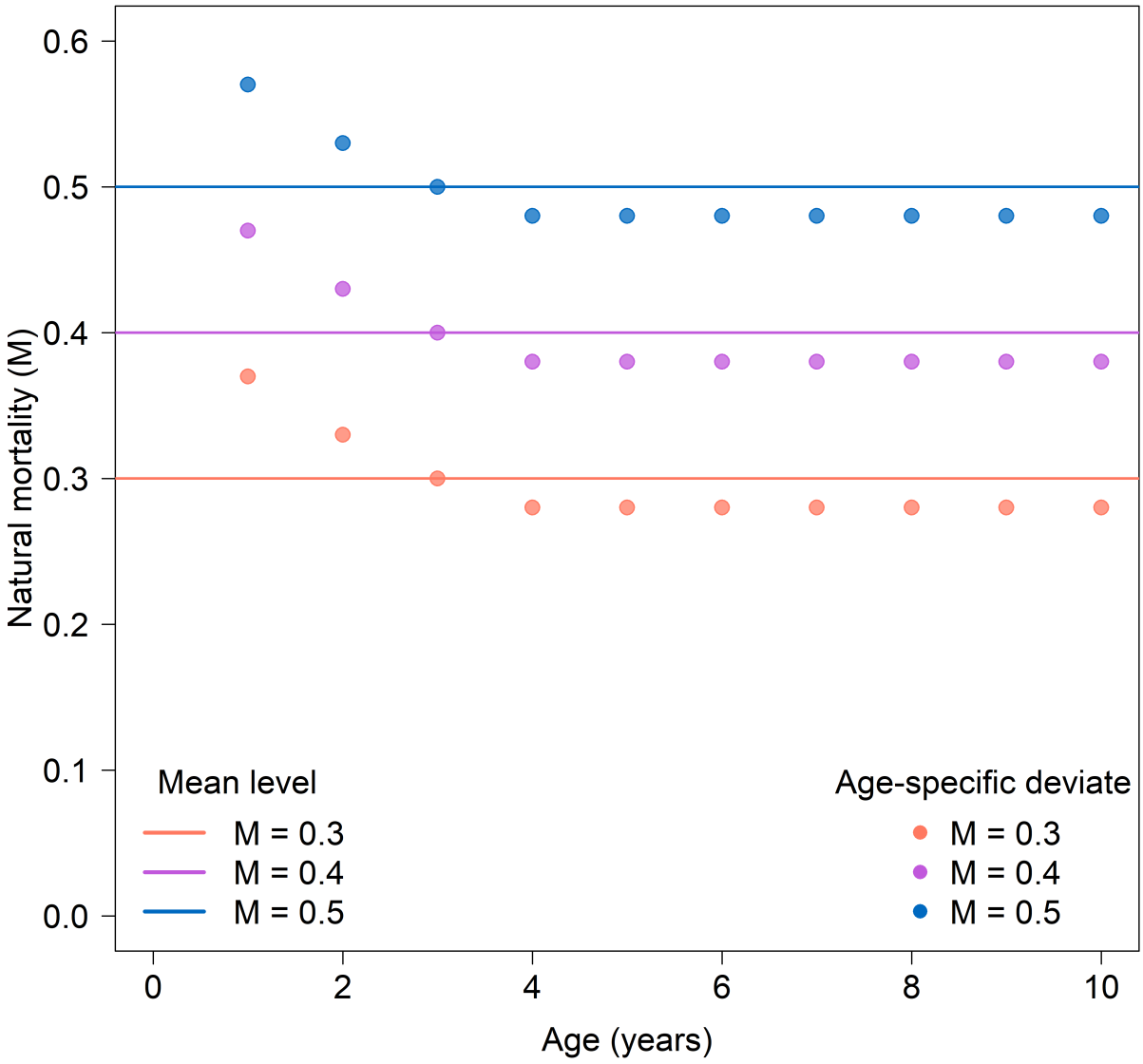

Figure 12: Age-specific natural mortality scenarios considered in the structural uncertainty grid. The age-specific deviates from Piner and Lee (2011) were applied to the average natural mortality levels of 0.3, 0.4, and 0.5.

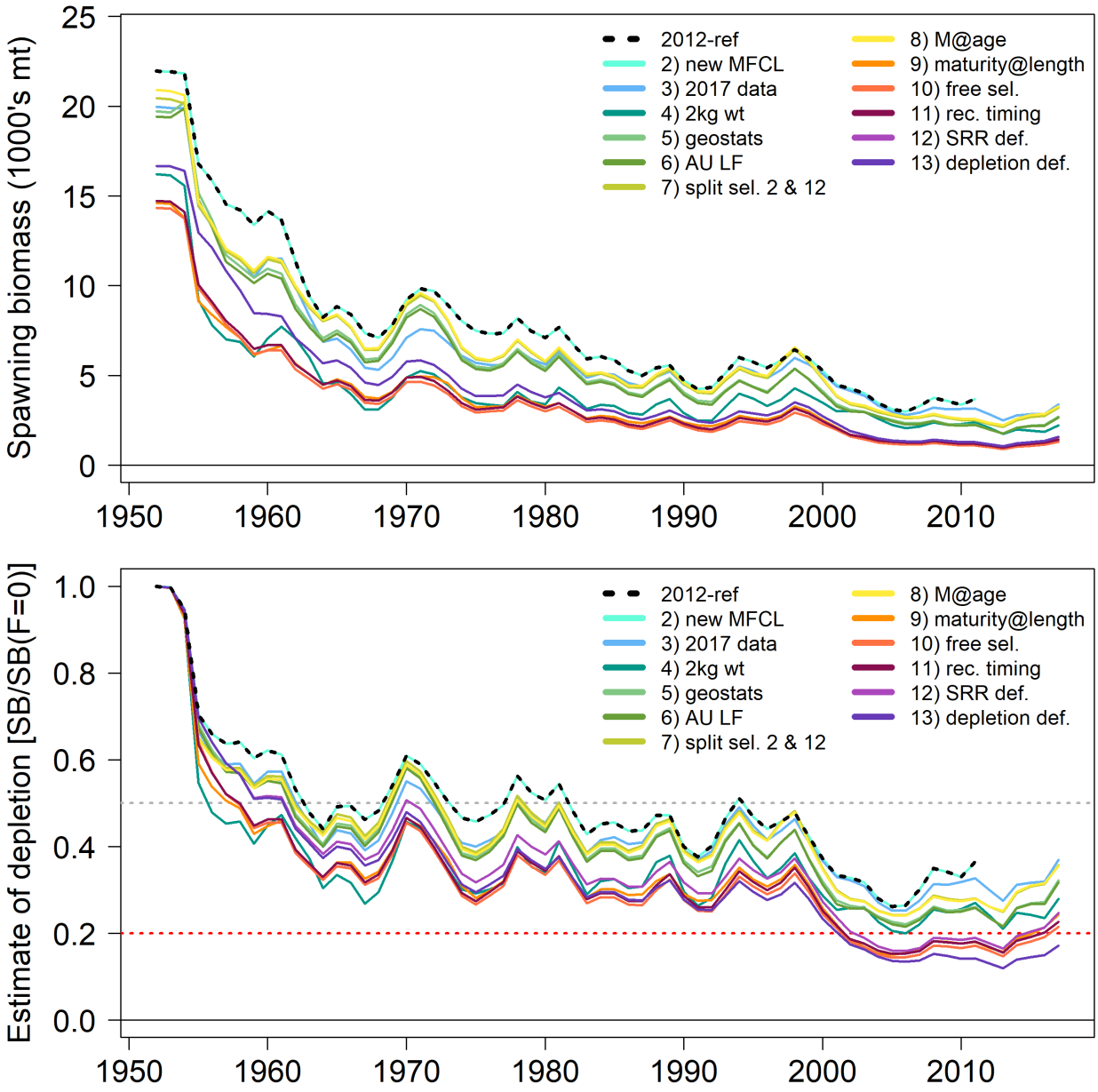

Figure 13: Stepwise progression from the 2012 reference case model to the 2019 diagnostic case. Top panel shows the spawning biomass and the bottom panel shows the relative depletion of the spawning biomass. For orientation, in the lower panel, the gray dotted line indicates the level of 50% depletion and the red dotted line indicates the level of 20% depletion.

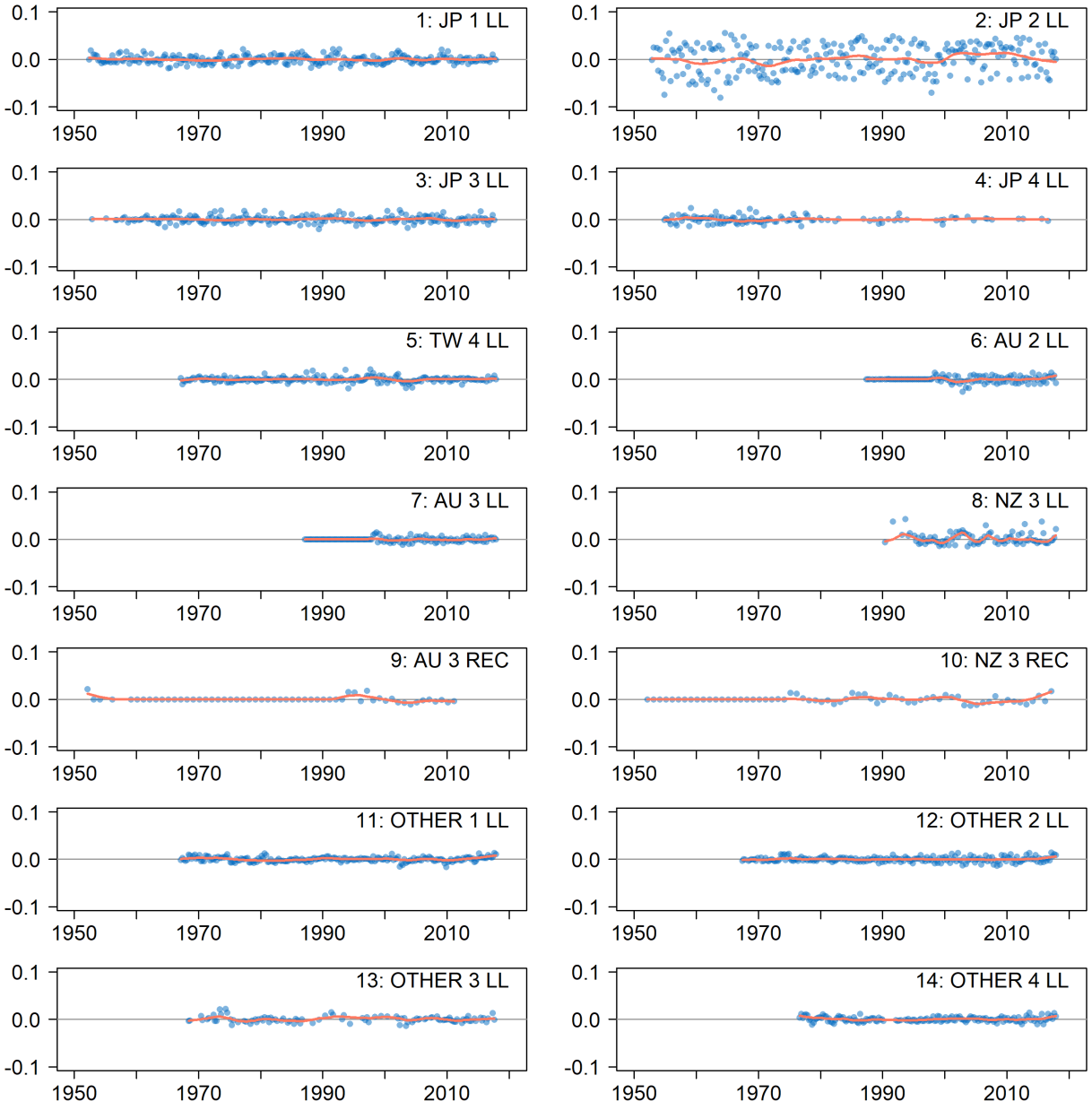

Figure 14: Relative scaled log-residuals for the 2019 diagnostic cased between the observed and predicted catch for each fishery instance in the model. The red line is a lowess smooth through the data.

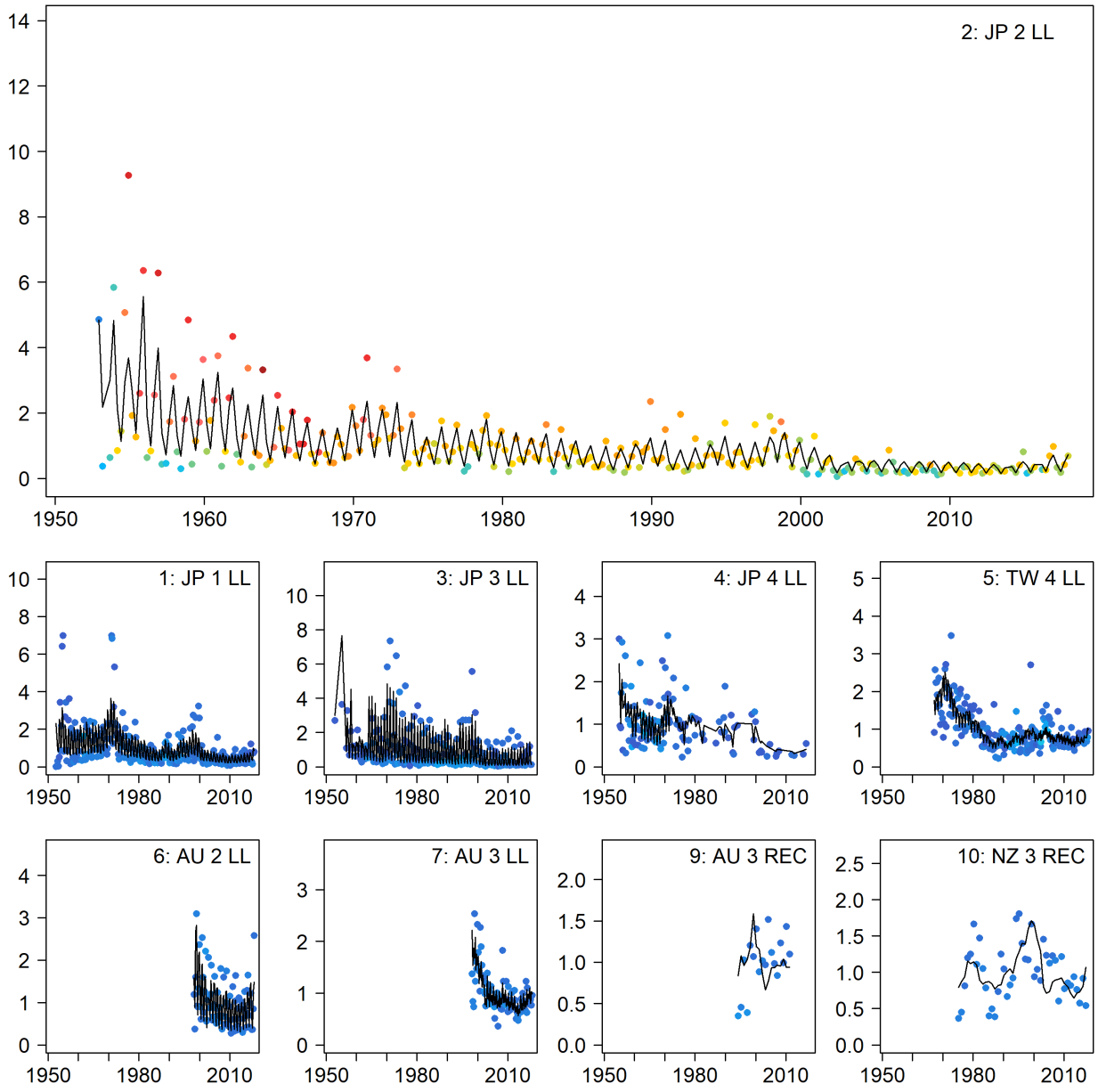

Figure 15: Observed (colored points) and model-predicted (black lines) CPUE for the fisheries which received standardized CPUE index in the diagnostic case model. The JP 2 LL fishery was the index fishery used in the diagnostic case model, and the fit to this index is shown in the main panel (top). Observed points are colored as a function of the effort deviate penalty weight applied on them, going from cooler to warmer colors as the penalty increases. Note that the y-axis is not consistent across panels, though all CPUE has been rescaled to a mean of 1.

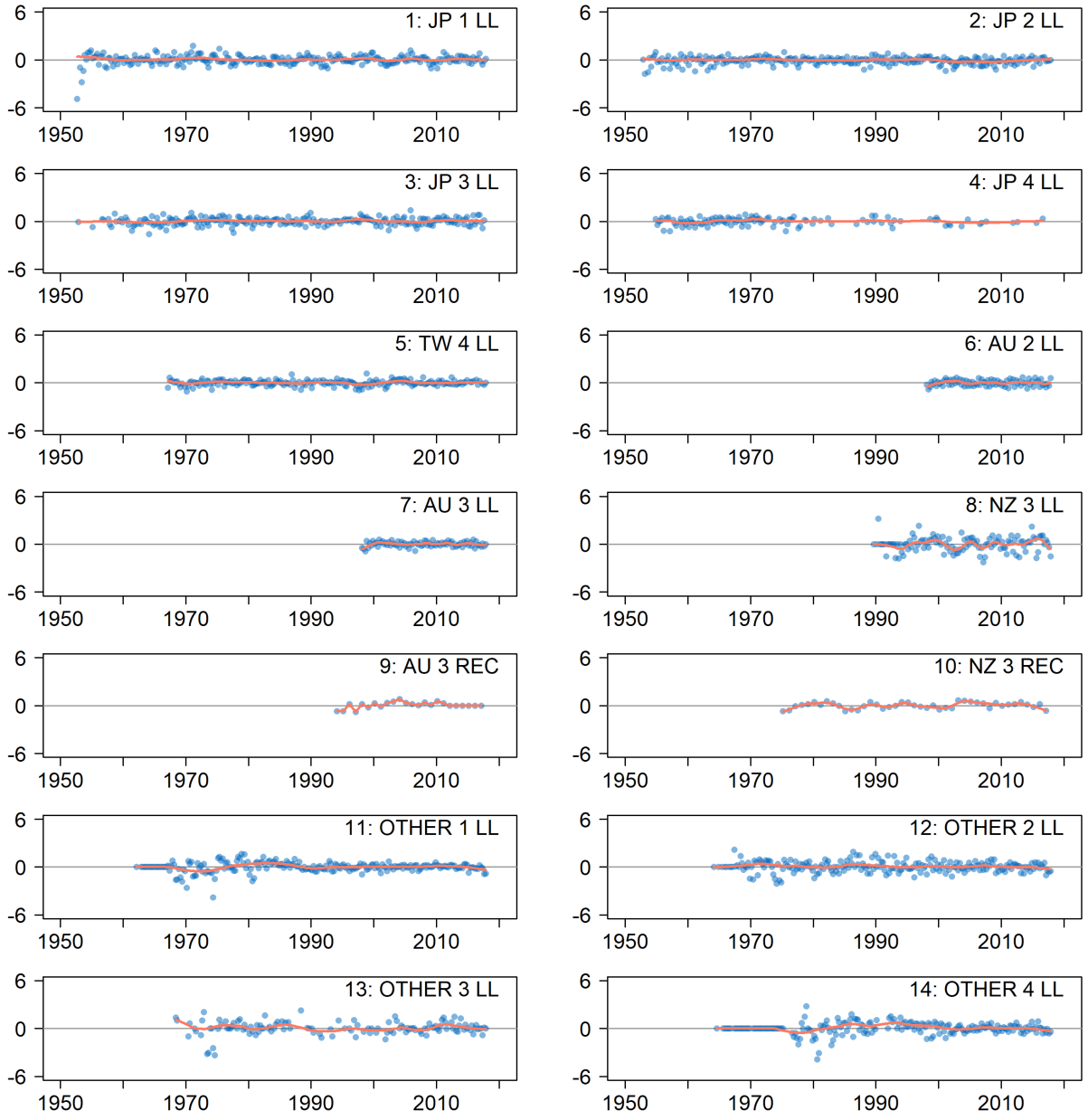

Figure 16: Effort deviations by time period for each of the fisheries receiving standardized CPUE indices in the diagnostic case model. The red line represents a lowess smoothed fit to the effort deviations.

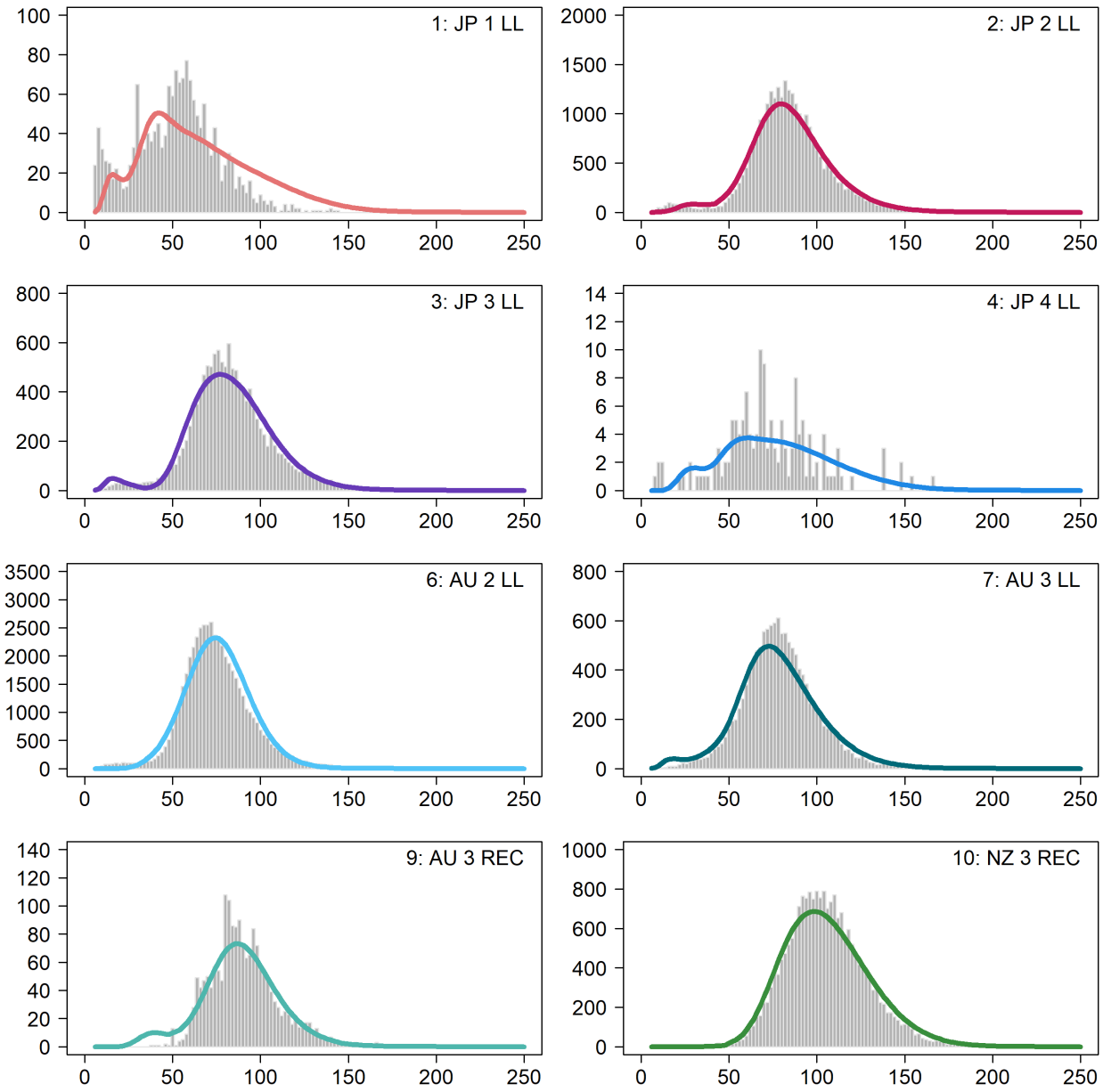

Figure 17: Composite (all time periods combined) observed (grey histograms) and predicted (colored lines) catch-at-weight for all fisheries with samples for the diagnostic case model. The colors indicate the groupings of the fisheries with respect to selectivity, fisheries with the same color in the plot share the same selectivity in the model. The x-axis is the whole weight in kg and the y-axis is numbers of observations.

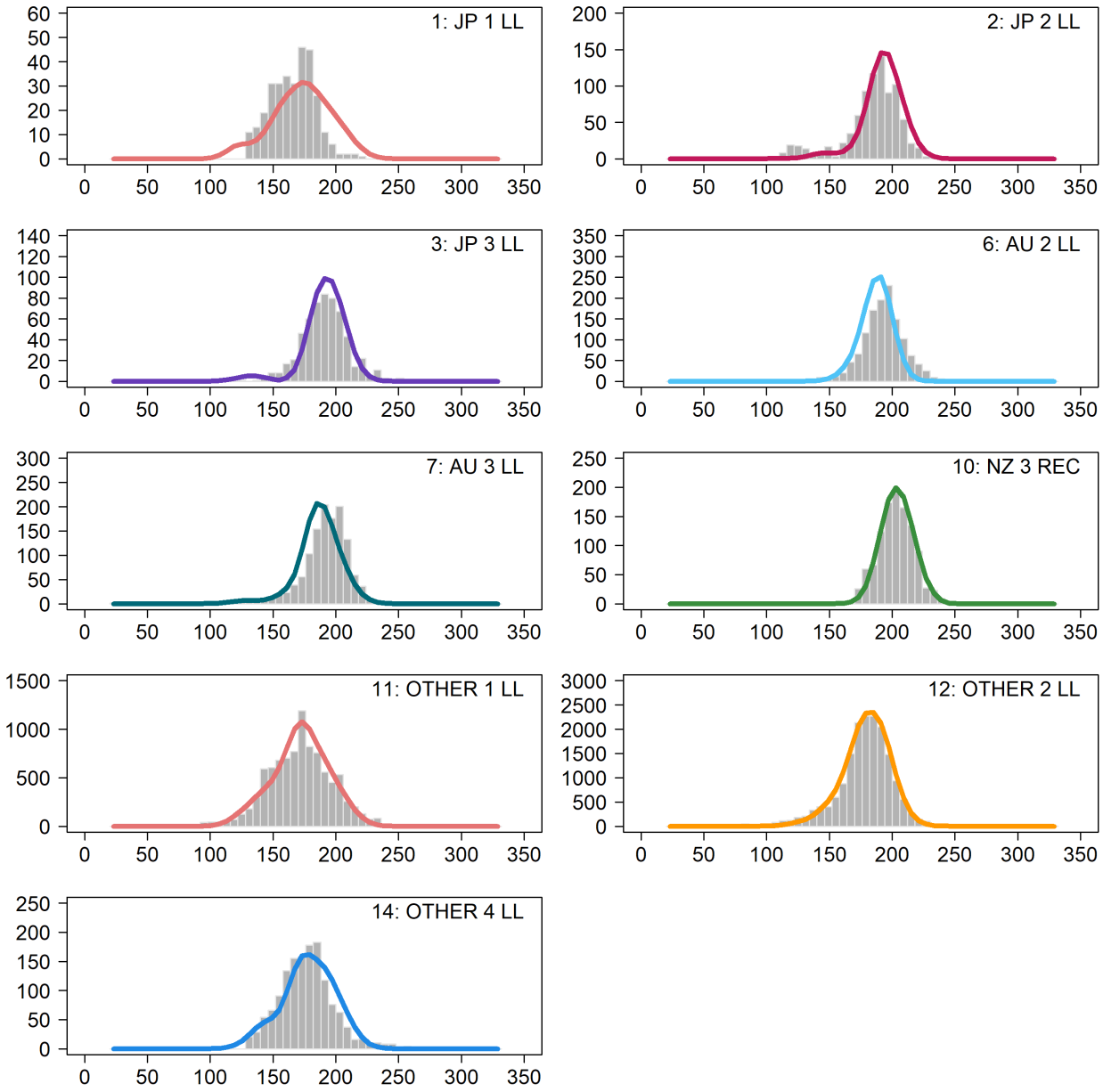

Figure 18: Composite (all time periods combined) observed (grey histograms) and predicted (colored lines) catch-at-length for all fisheries with samples for the diagnostic case model. The colors indicate the groupings of the fisheries with respect to selectivity, fisheries with the same color in the plot share the same selectivity in the model. The x-axis is the EFL in cm and the y-axis is numbers of observations.

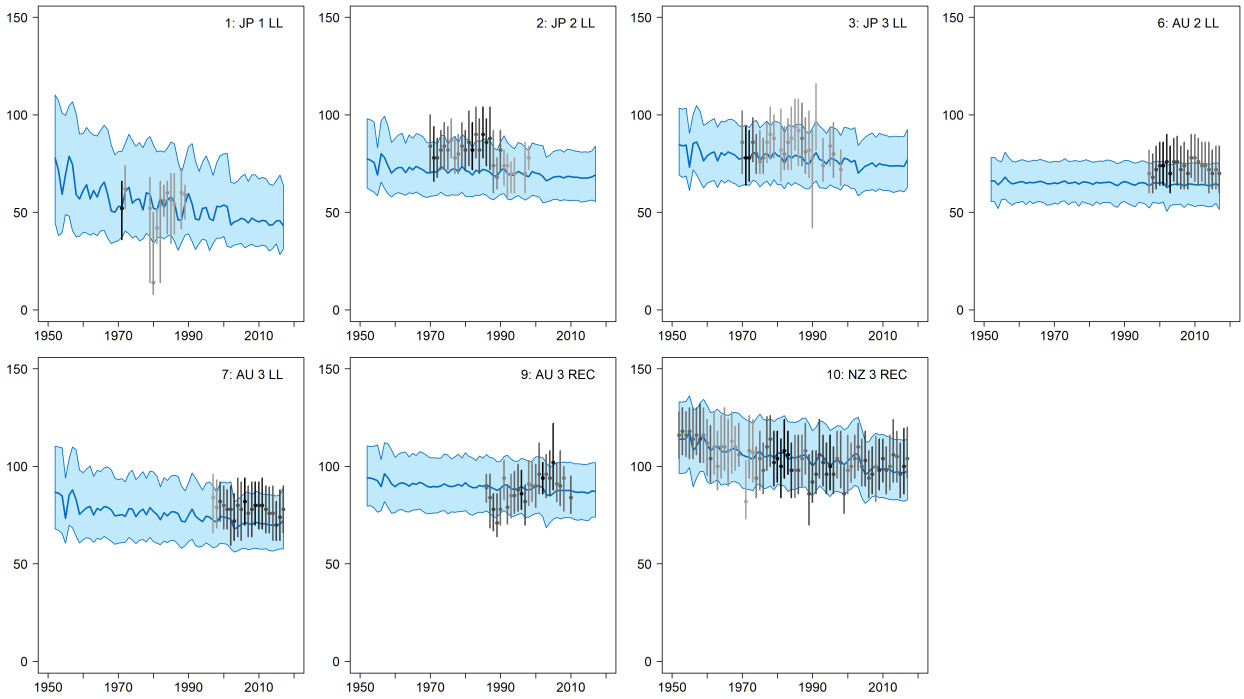

Figure 19: A comparison of the observed (gray scale points) and predicted (blue line) median weight (WW, kg) for the major fisheries with size frequency observations in the diagnostic case. The uncertainty intervals (blue shading) represent the values encompassed by the 25% and 75% quantiles. Sampling data were aggregated by year and were plotted only if there were a minimum of 30 samples per year. For each sample the bar indicates the 50^{th} percentile. The color of the observed sample gives the relative number of observations across all observations for that fishery. Darker grays indicate a larger sample size.

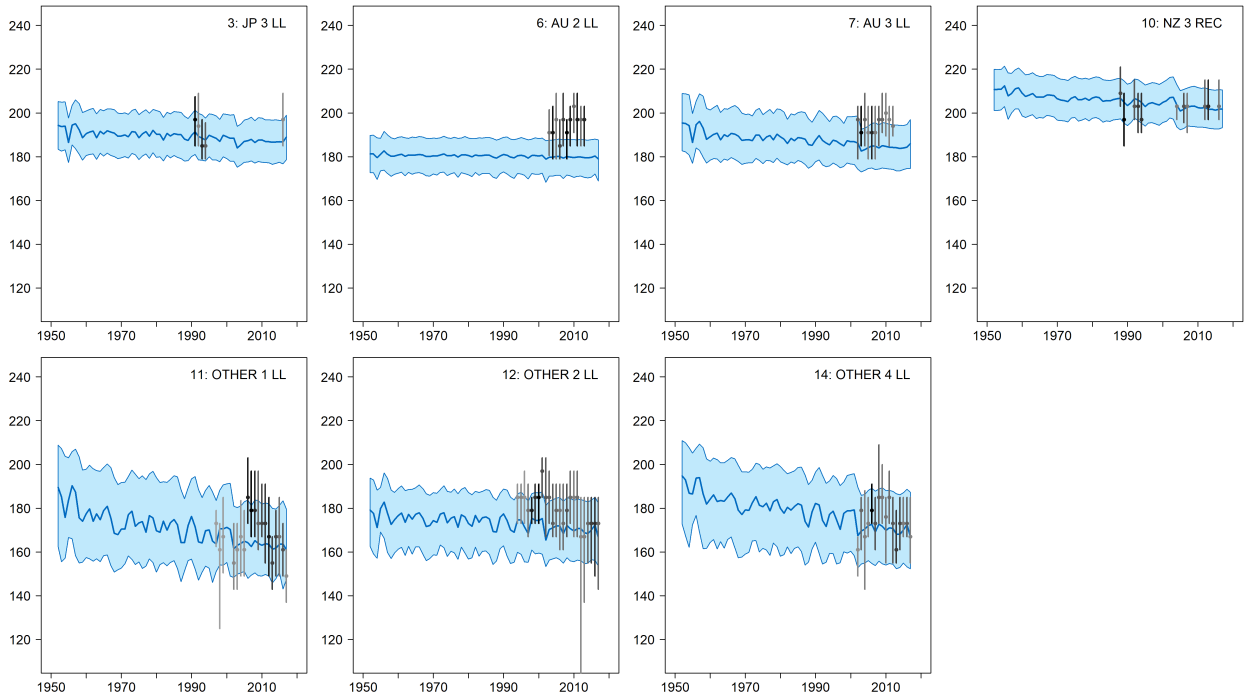

Figure 20: A comparison of the observed (gray scale points) and predicted (blue line) median length (EFL, cm) for the major fisheries with size frequency observations in the diagnostic case. The uncertainty intervals (blue shading) represent the values encompassed by the 25% and 75% quantiles. Sampling data were aggregated by year and were plotted only if there were a minimum of 30 samples per year. For each sample the bar indicates the 50^{th} percentile. The color of the observed sample gives the relative number of observations across all observations for that fishery. Darker grays indicate a larger sample size.

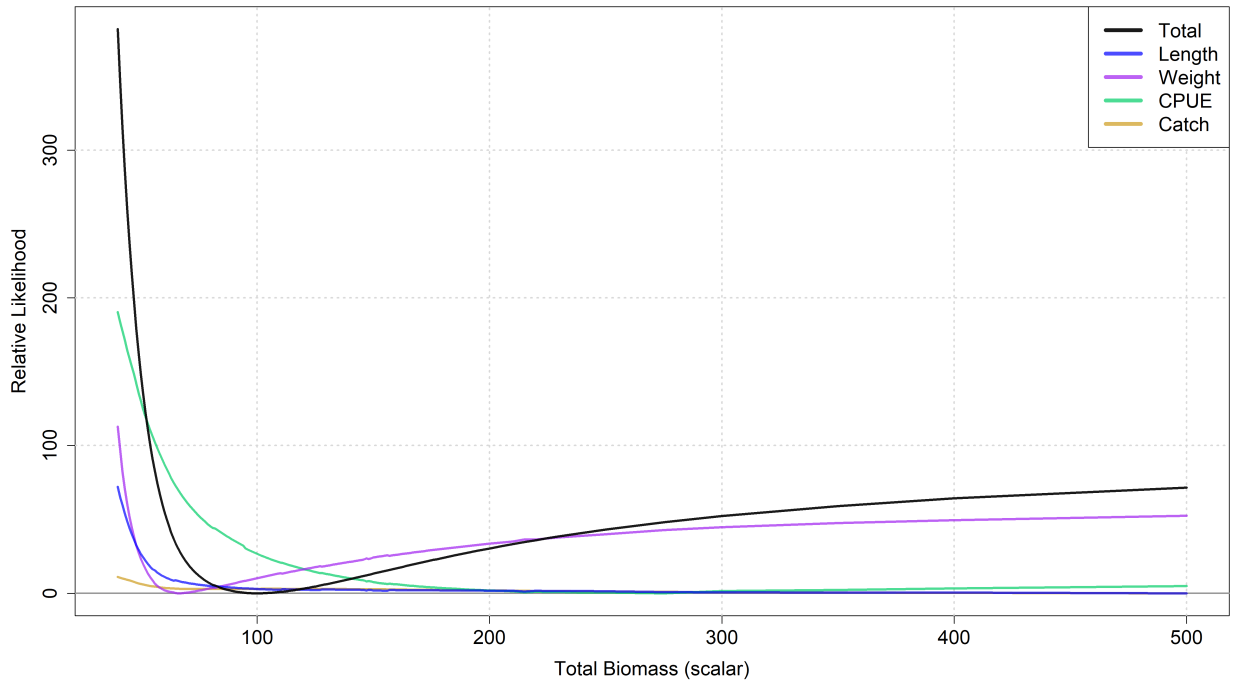

Figure 21: Change in the total, and individual data component log-likelihoods with respect to the derived parameter, mean total biomass over the assessment period, across a range of scalar values at which this parameter was penalised to fit, for the diagnostic case model. For reference, a scalar value of 100 denotes the estimate from the diagnostic case model.

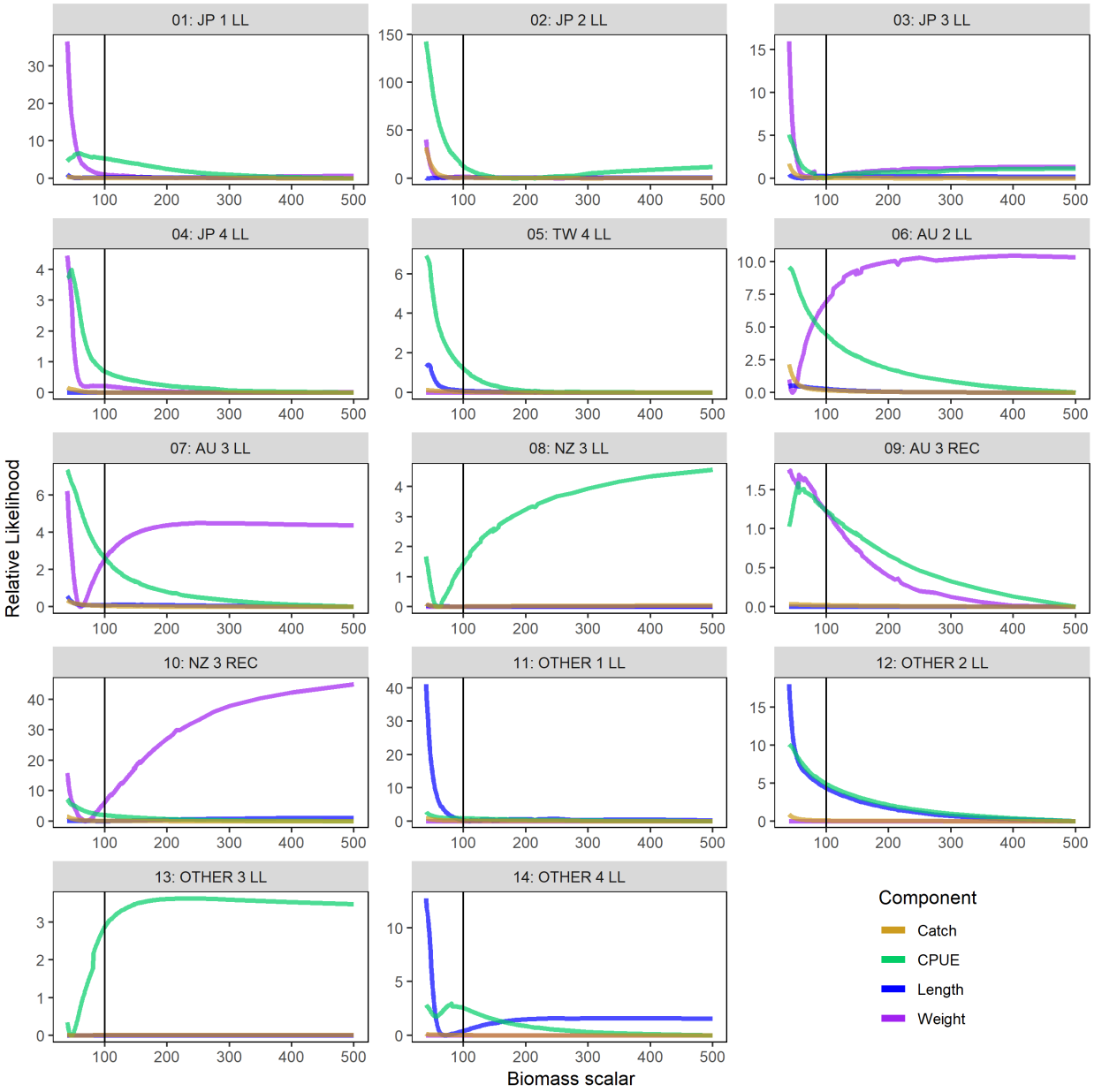

Figure 22: Change in the total, and individual data component log-likelihoods by fishery with respect to the derived parameter, mean total biomass over the assessment period, across a range of values at which this parameter was penalised to fit, for the diagnostic case model. The vertical bar at 100 denotes the total biomass estimated from the 2019 diagnostic case.

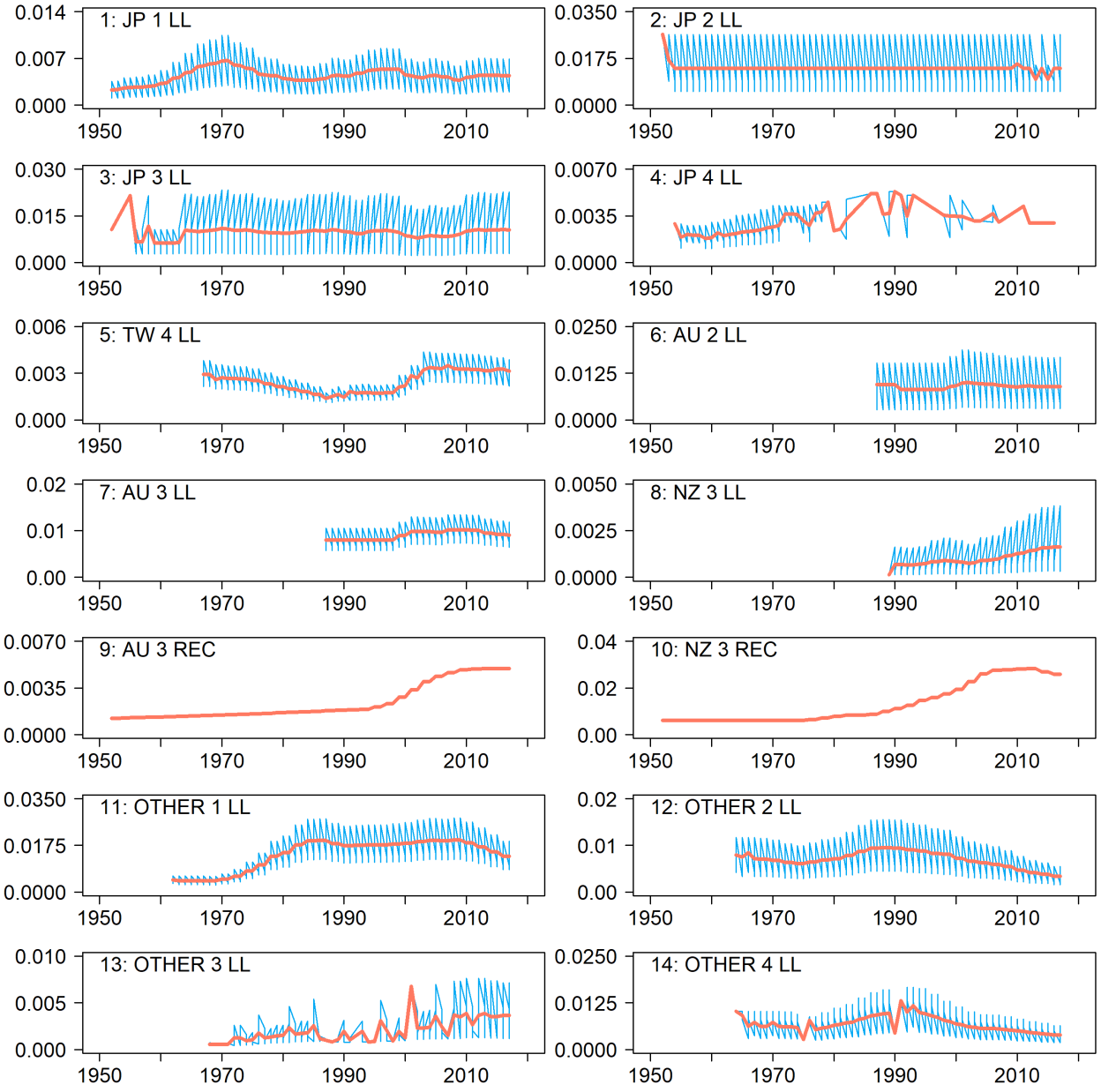

Figure 23: Estimated time series of catchability for all fisheries in the 2019 diagnostic case. The red line shows the long term biennial trend, and the blue line shows the seasonal variability associated with each fishery. The two recreational fisheries will only have a biennial trend. Note that the y-axis values are different.

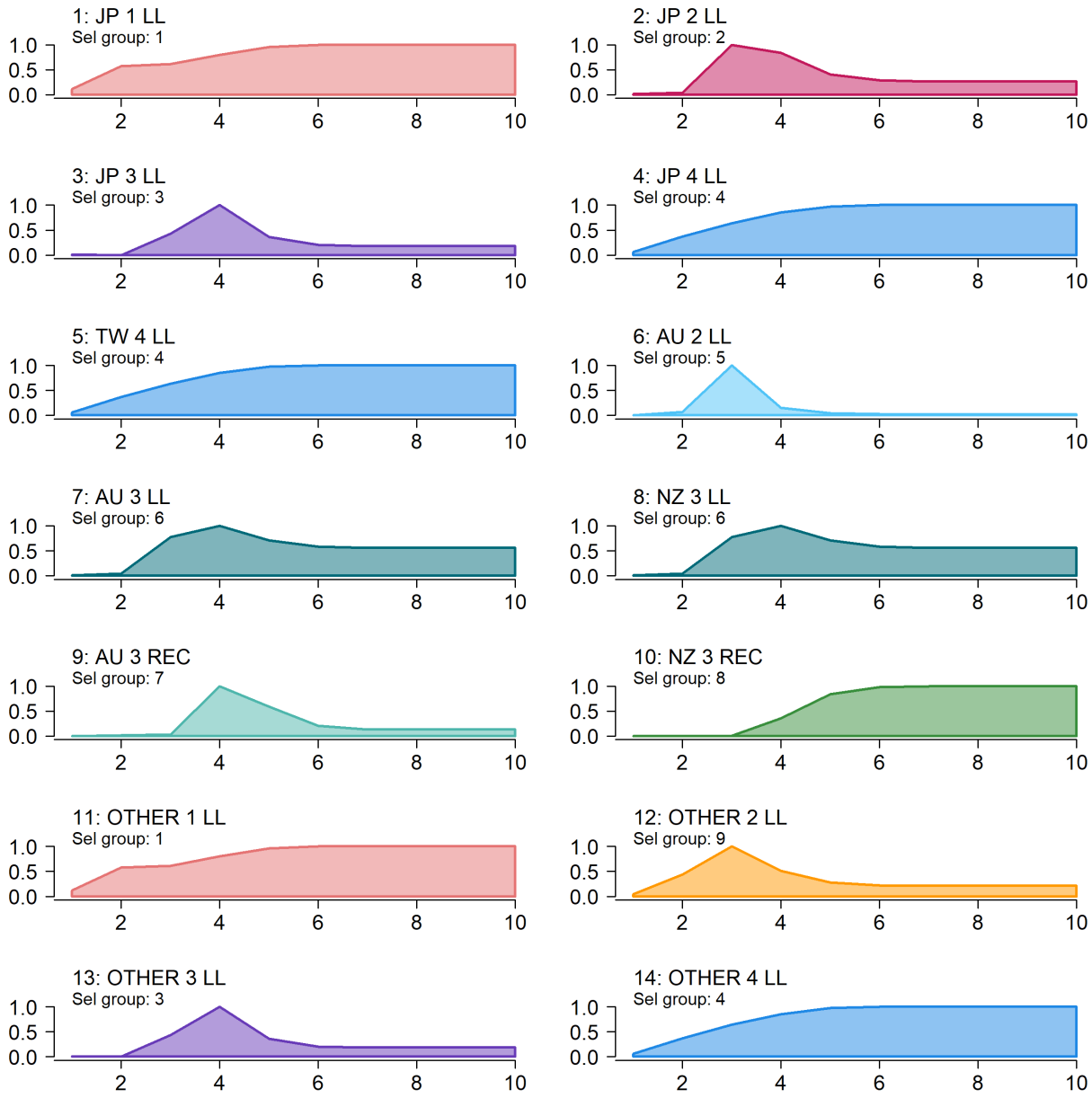

Figure 24: Estimated age-specific selectivity curves by fishery for the diagnostic case model. The colors indicate the groupings of the fisheries with respect to selectivity, fisheries with the same color were grouped together for selectivity estimation in the model.

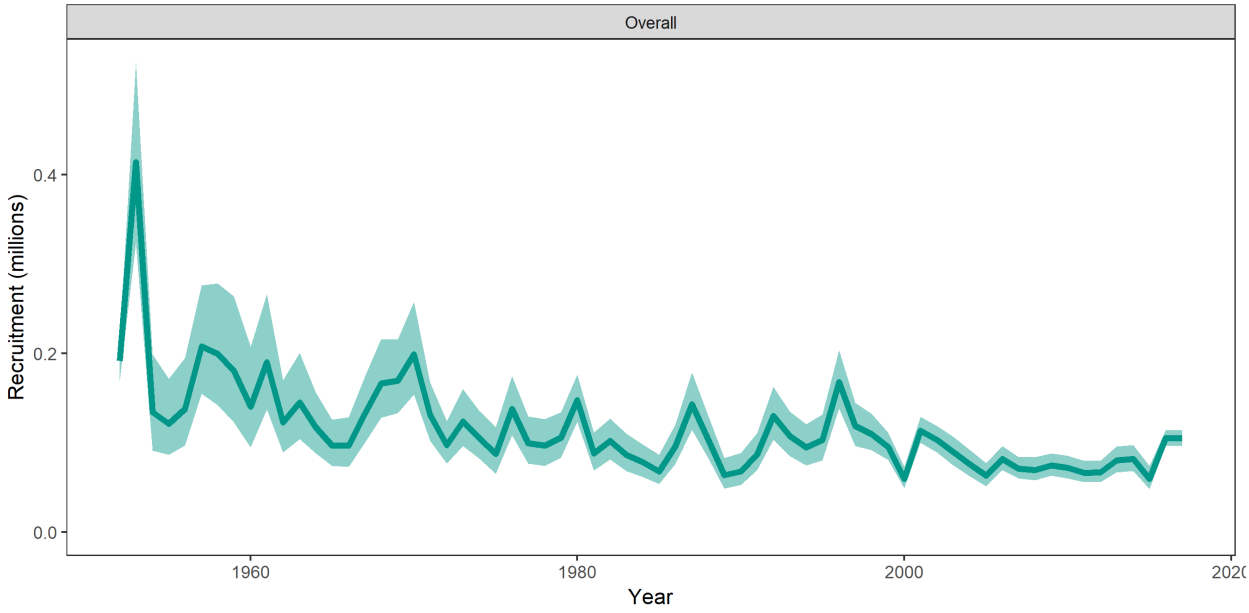

Figure 25: The estimated recruitment from the 2019 diagnostic case with the associated model specific statistical uncertainty (95% confidence interval) as calculated from the Hessian.

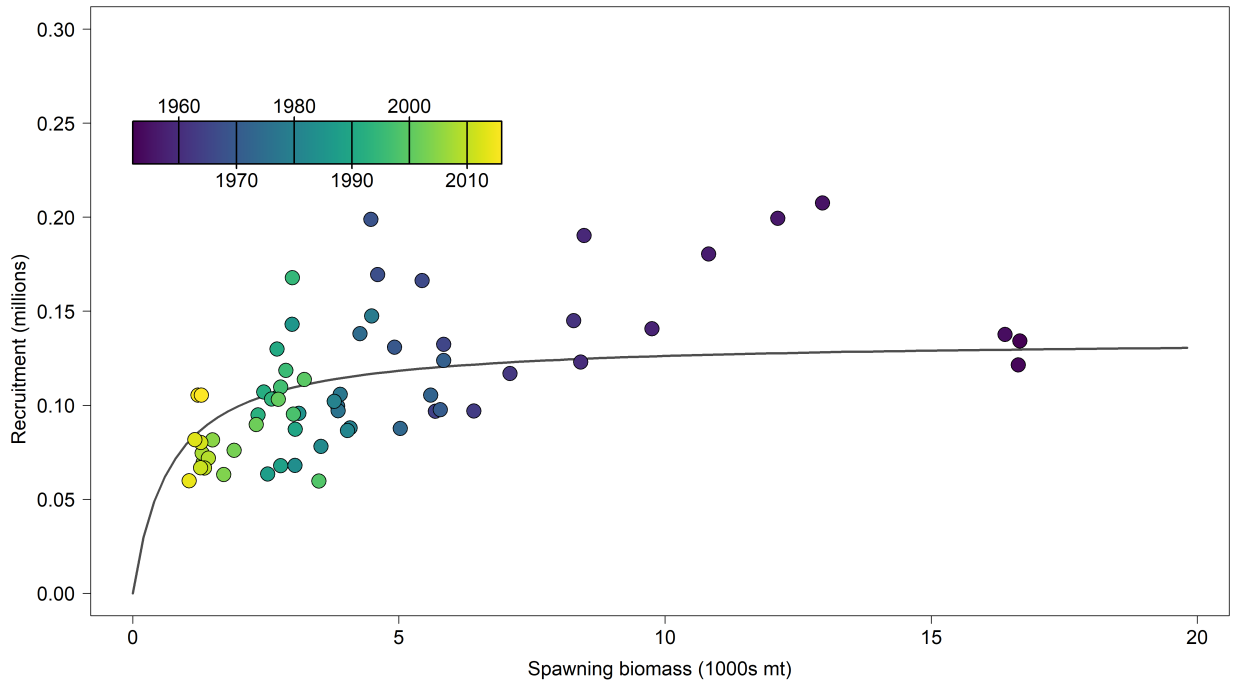

Figure 26: Estimated relationship between recruitment and spawning biomass based on annual values for the diagnostic case model. The color of the circle corresponds to the year: darker-cooler colors are early in the model period, and lighter-warmer colors are more recent years.

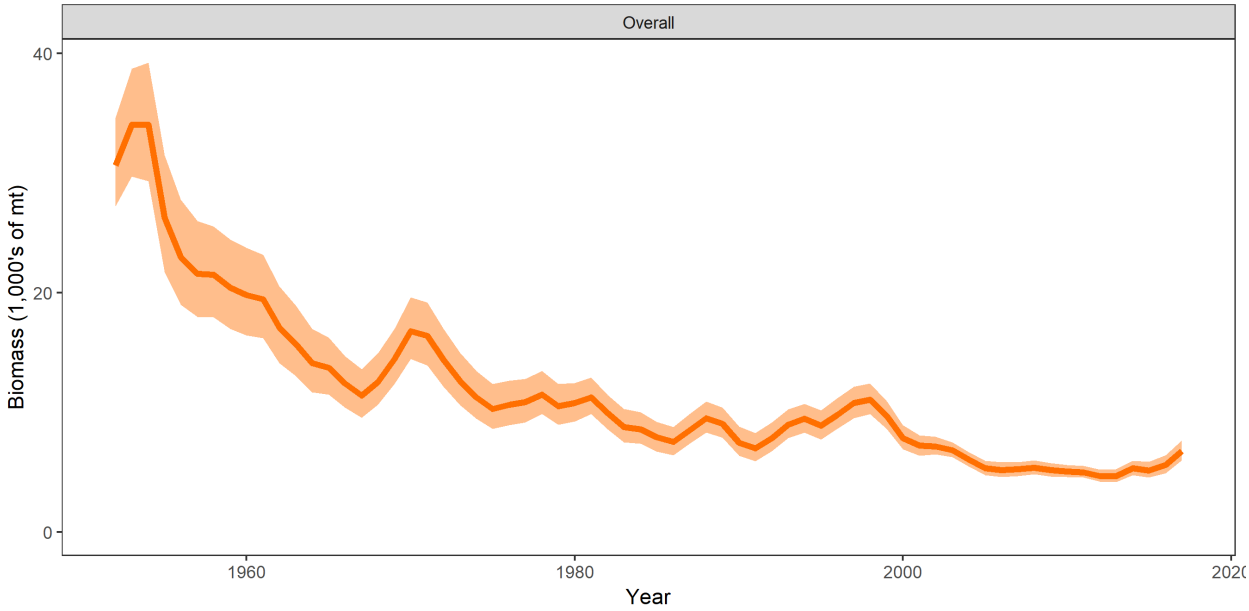

Figure 27: The estimated total biomass from the 2019 diagnostic case with the associated model specific statistical uncertainty (95% confidence interval) as calculated from the Hessian.

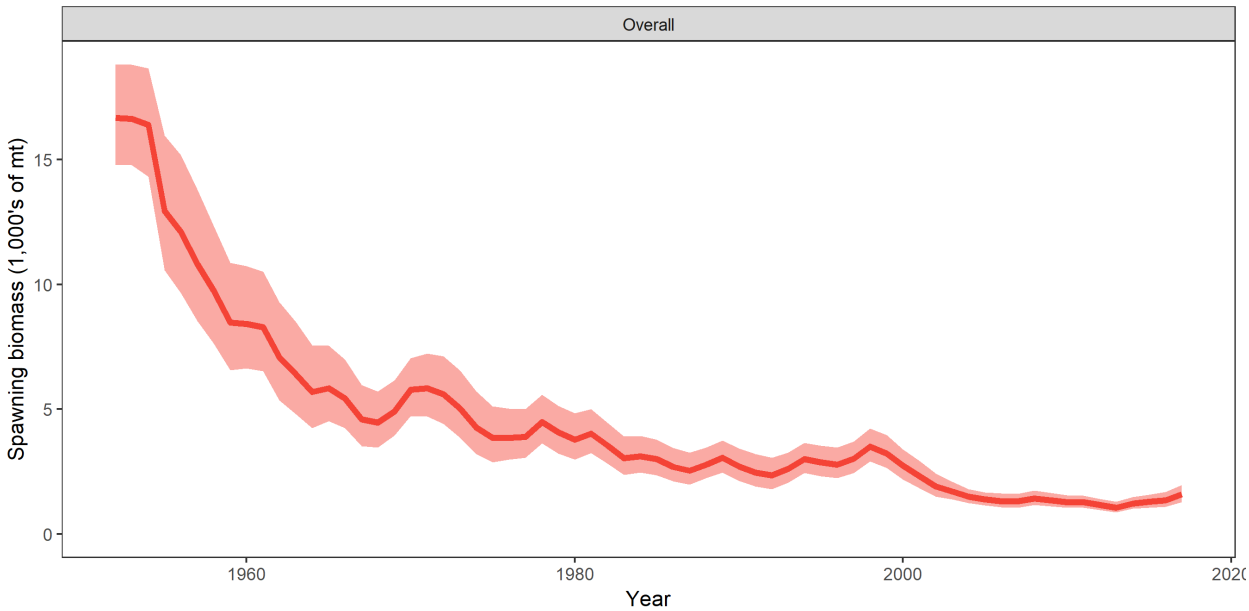

Figure 28: The estimated spawning biomass from the 2019 diagnostic case with the associated model specific statistical uncertainty (95% confidence interval) as calculated from the Hessian.

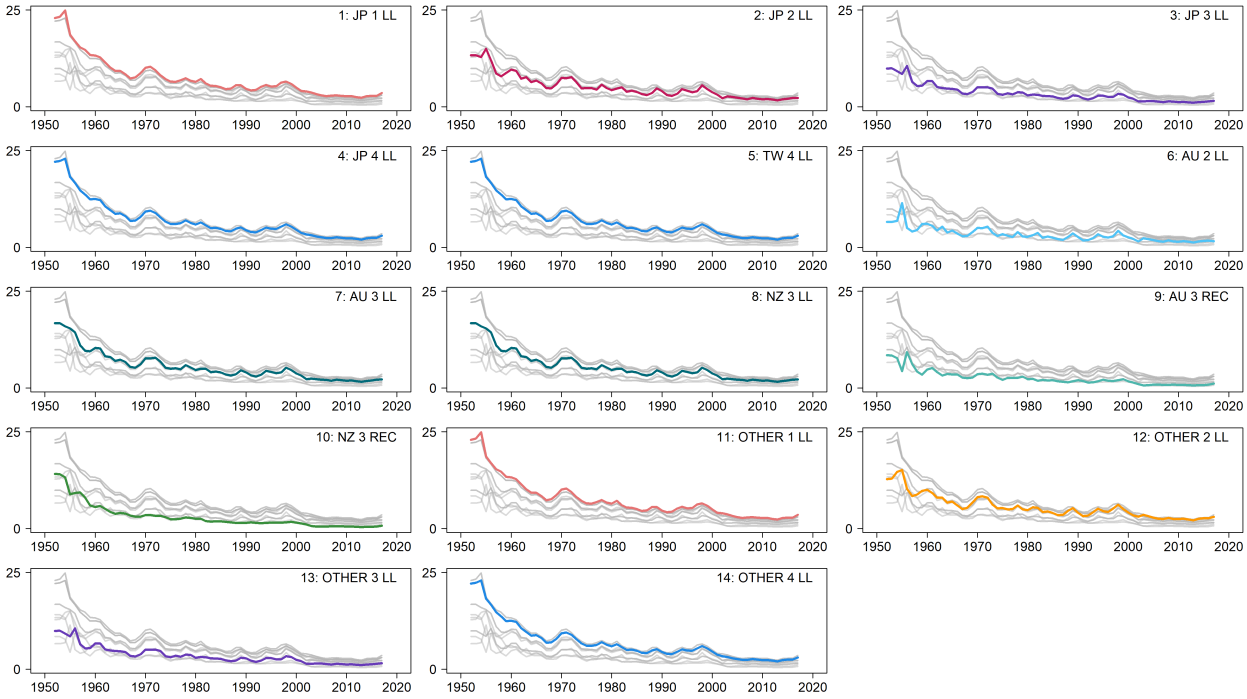

Figure 29: Vulnerable (exploitable) biomass for each fishery defined in the model, where vulnerable biomass is the product of catch in numbers by weight-at-age and selectivity for a given fishery. The colors indicate the shared selectivity groups defined in Figure 24 for that fishery. The gray lines in each panel are the other fisheries in the model. These are shown for comparison.

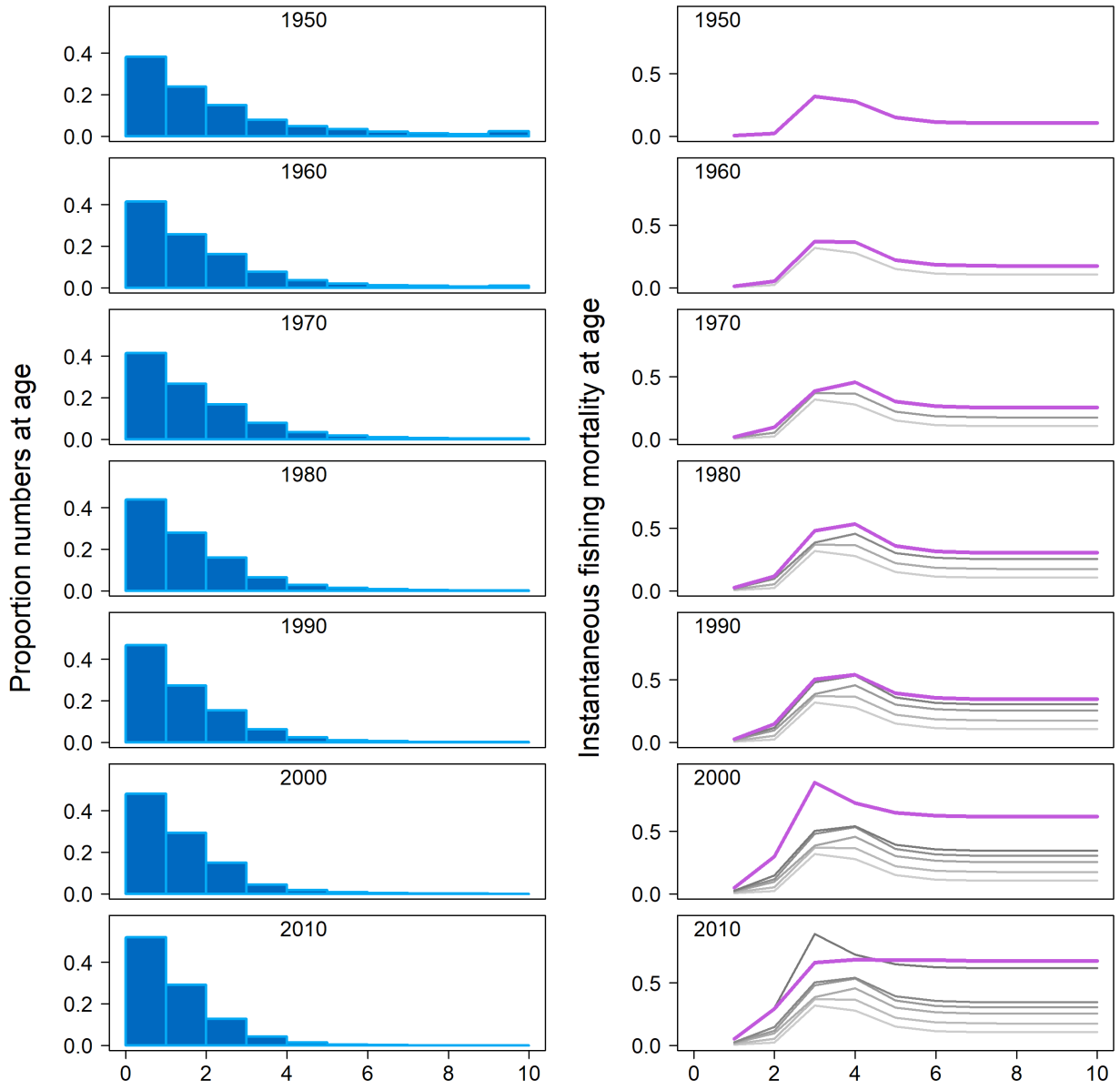

Figure 30: Estimated proportion of the average population at-age (left panels) and fishing mortality-at-age (right panels), at decadal intervals, for the diagnostic case model. In the right panels the gray lines indicate the average fishing mortality-at-age levels for preceding decades.

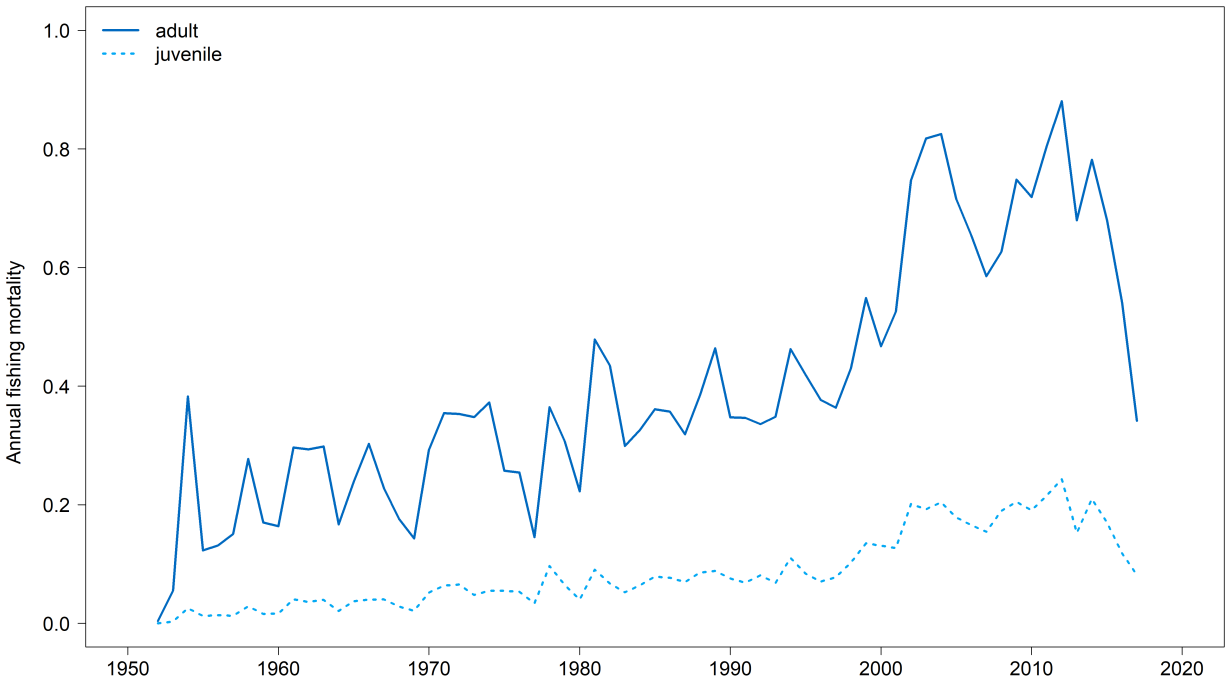

Figure 31: Estimated average annual juvenile (1-2 years) and adult (3-10 years) fishing mortality for the diagnostic case model.

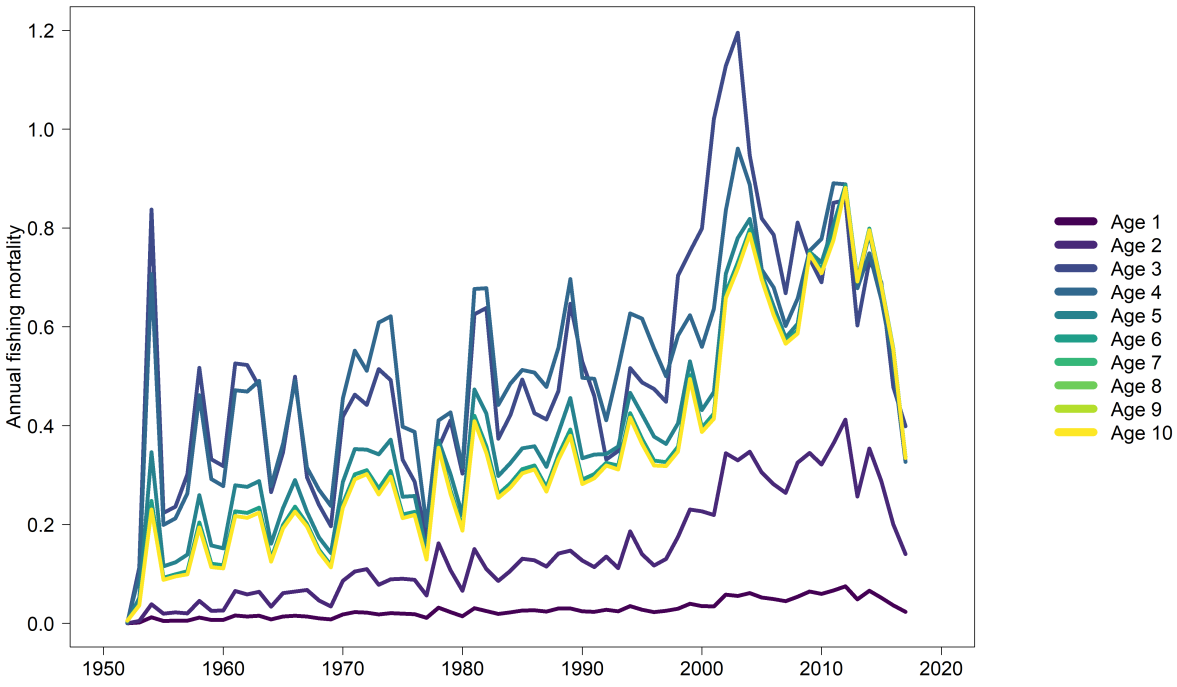

Figure 32: Estimated age-specific fishing mortality for the diagnostic case model.

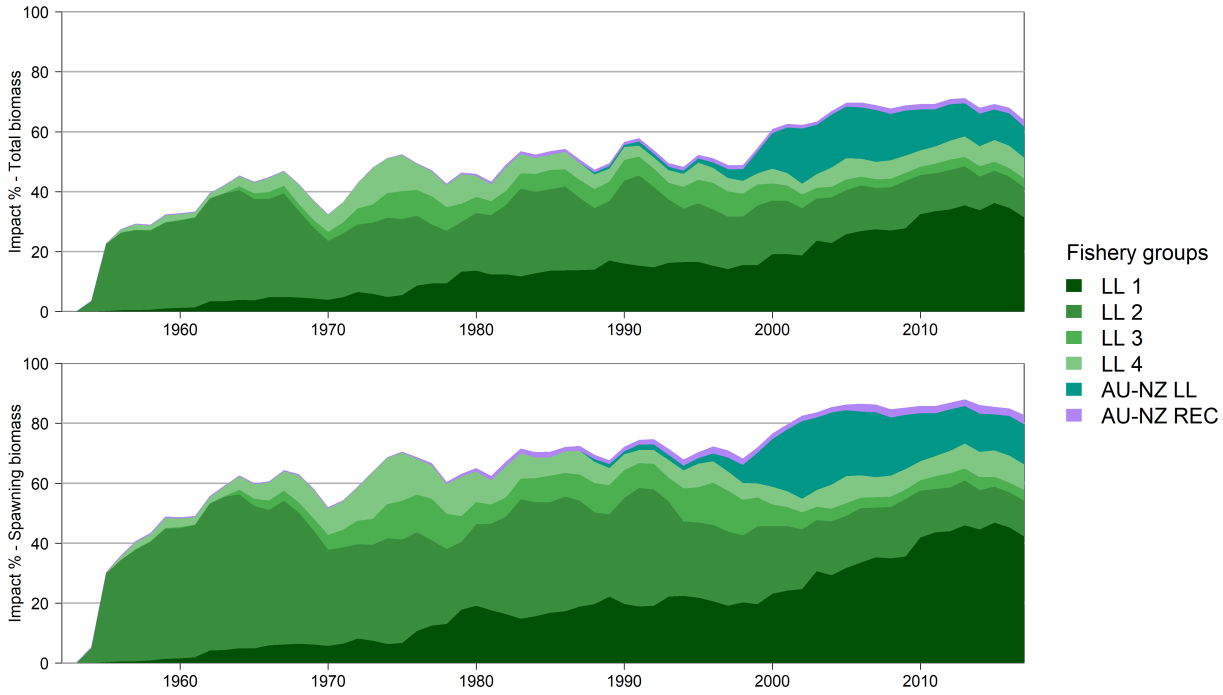

Figure 33: Estimates of reduction in total biomass (top) and spawning biomass (bottom) due to fishing (fishery impact $= 1 - B_t/B_{t,F=0}$) attributed to various fishery groups for the diagnostic case model.

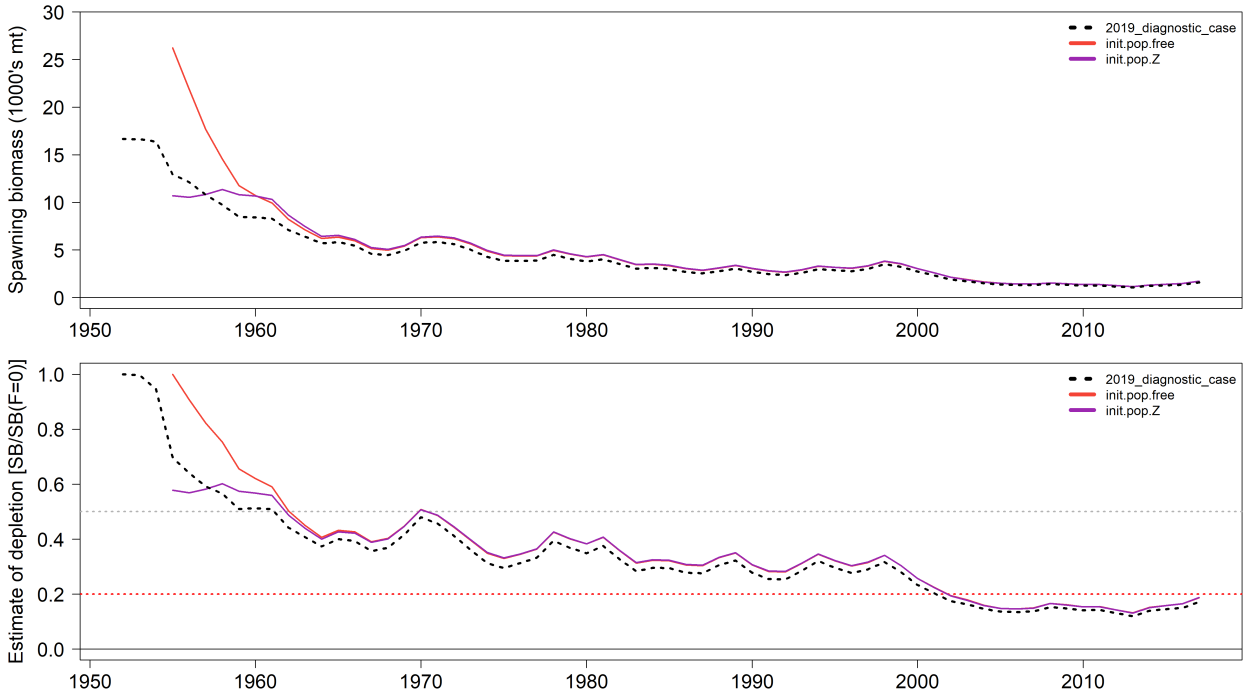

Figure 34: Comparison of sensitivity runs of starting the model in 1955 under two different starting age structure scenarios relative to the 2019 diagnostic case (black dotted line). Top panel shows the spawning biomass and the bottom panel shows the relative depletion of the spawning biomass. For orientation, in the lower panel, the gray dotted line indicates the level of 50% depletion and the red dotted line indicates the level of 20% depletion.

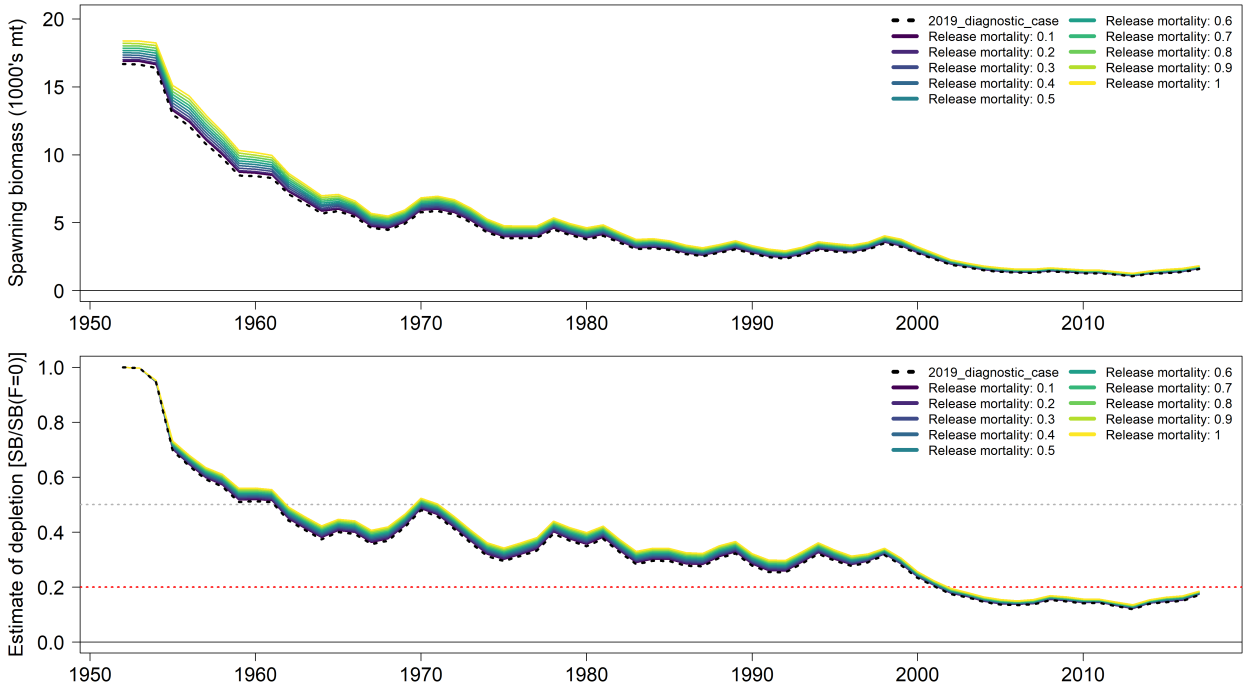

Figure 35: Comparison of sensitivity runs to including release mortality for the recreational fisheries relative to the 2019 diagnostic case (black dotted line). Ten different levels of release mortality were considered. Top panel shows the spawning biomass and the bottom panel shows the relative depletion of the spawning biomass. For orientation, in the lower panel, the gray dotted line indicates the level of 50% depletion and the red dotted line indicates the level of 20% depletion.

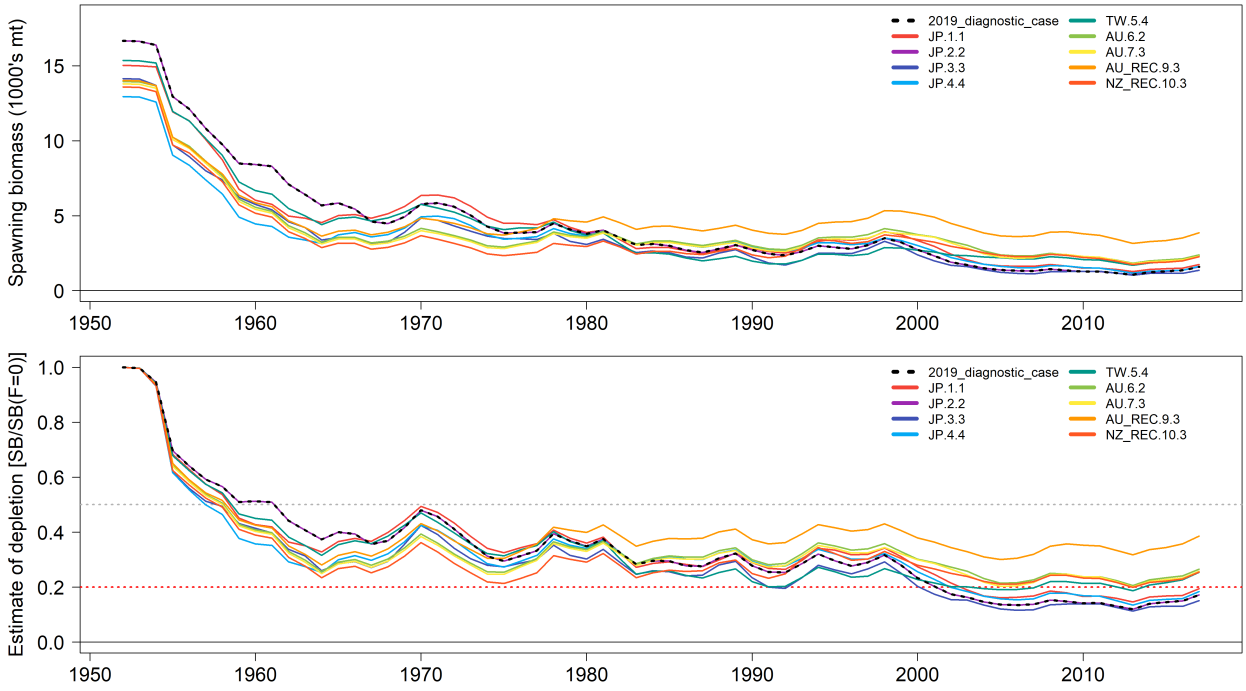

Figure 36: Comparison of sensitivity runs to the choice of index fishery relative to the 2019 diagnostic case (black dotted line). Top panel shows the spawning biomass and the bottom panel shows the relative depletion of the spawning biomass. For orientation, in the lower panel, the gray dotted line indicates the level of 50% depletion and the red dotted line indicates the level of 20% depletion.

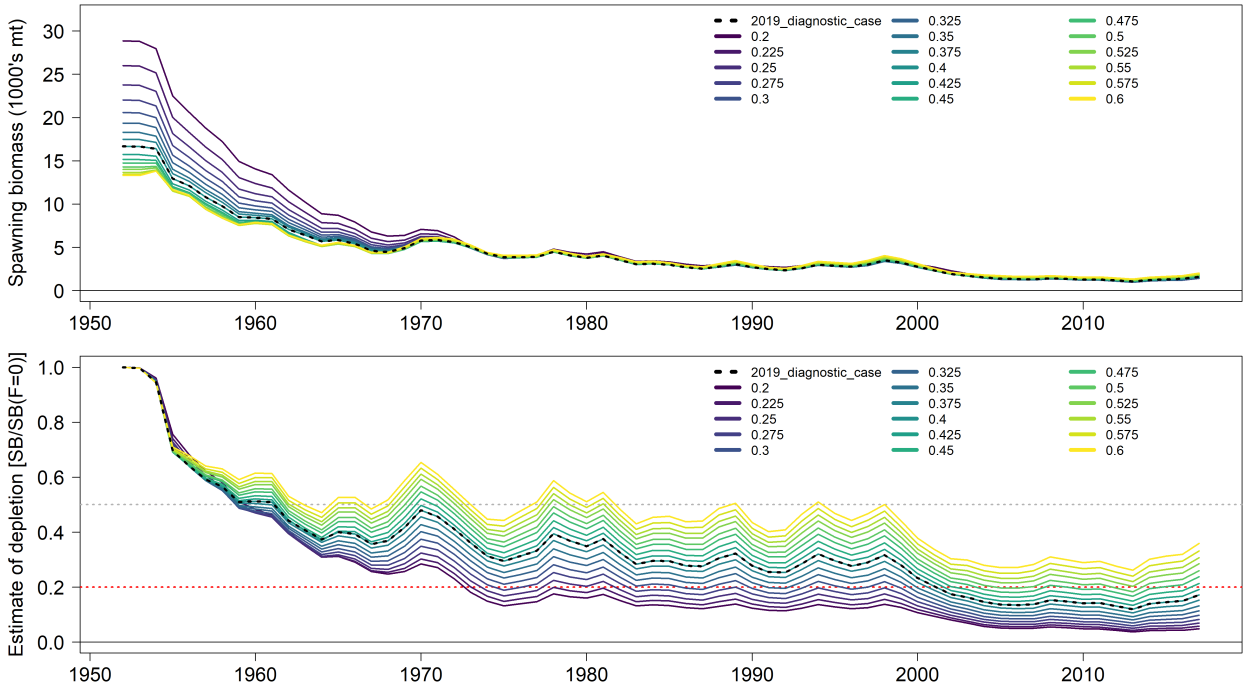

Figure 37: Comparison of sensitivity runs to the choice of average natural mortality-at-age relative to the 2019 diagnostic case (black dotted line). Top panel shows the spawning biomass and the bottom panel shows the relative depletion of the spawning biomass. For orientation, in the lower panel, the gray dotted line indicates the level of 50% depletion and the red dotted line indicates the level of 20% depletion.

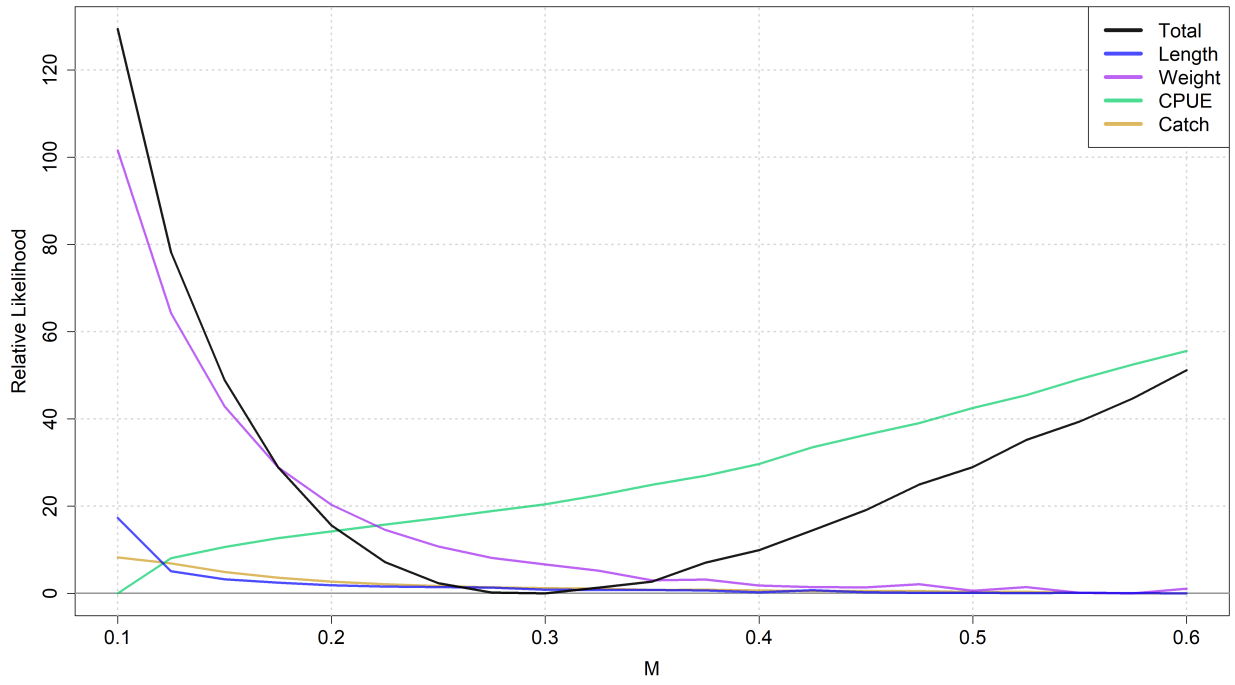

Figure 38: Change in the total, and individual data component log-likelihoods with respect to the average natural mortality-at-age (M), across a range of values at which this parameter was fixed, for the diagnostic case model.

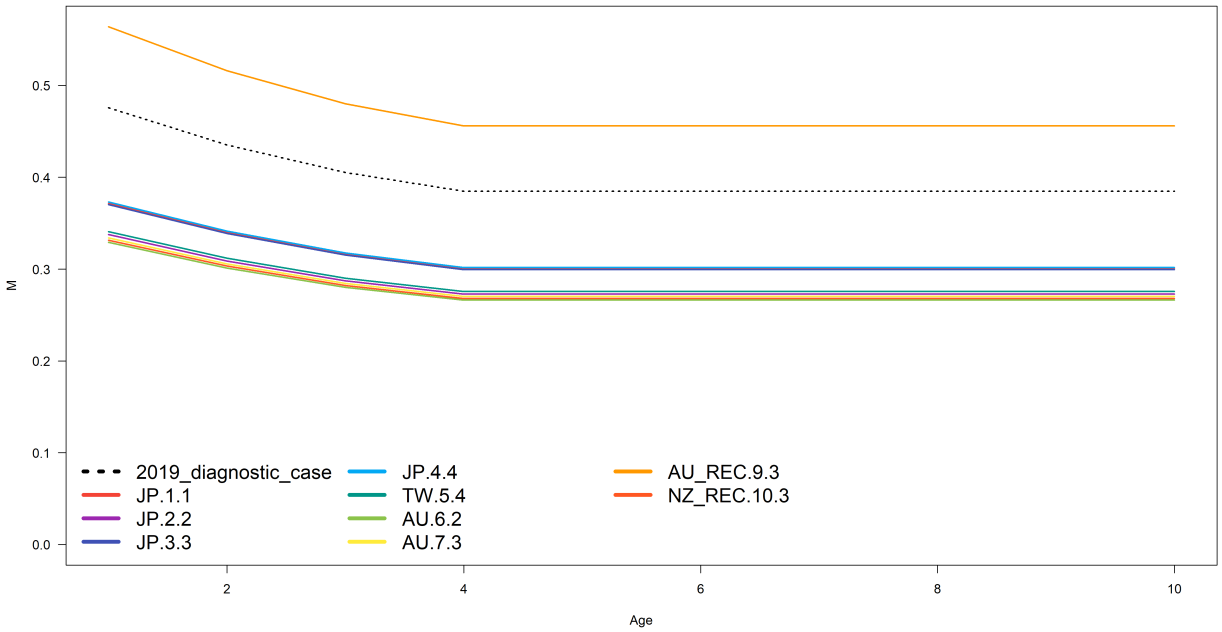

Figure 39: Estimated values of average natural mortality-at-age (M) with the addition of the fixed age-specific deviates from models using the different standardized CPUEs as the index.

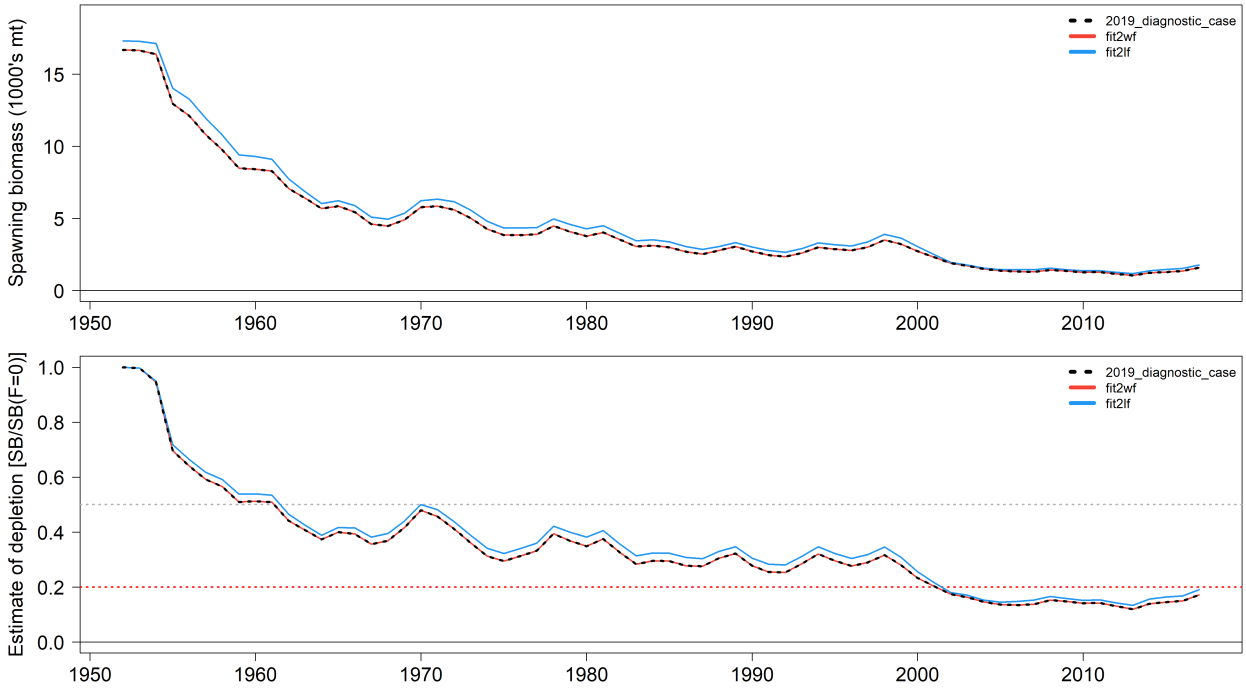

Figure 40: Comparison of sensitivity runs to either down-weighting the weight (*fit2wf*) or the length (*fit2lf*) composition data for the Australian longline relative to the 2019 diagnostic case (black dotted line). Top panel shows the spawning biomass and the bottom panel shows the relative depletion of the spawning biomass. For orientation, in the lower panel, the gray dotted line indicates the level of 50% depletion and the red dotted line indicates the level of 20% depletion.

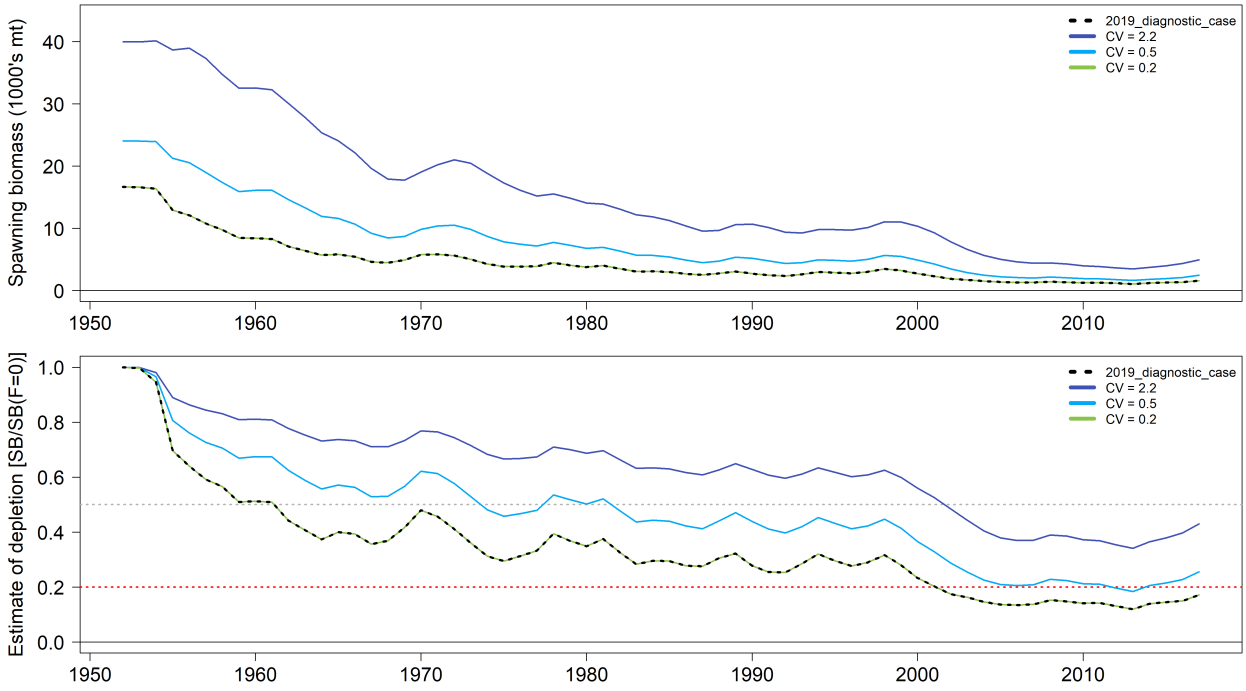

Figure 41: Comparison of sensitivity runs to changing the CV on the recruitment penalty relative to the 2019 diagnostic case (black dotted line). Top panel shows the spawning biomass and the bottom panel shows the relative depletion of the spawning biomass. For orientation, in the lower panel, the gray dotted line indicates the level of 50% depletion and the red dotted line indicates the level of 20% depletion.

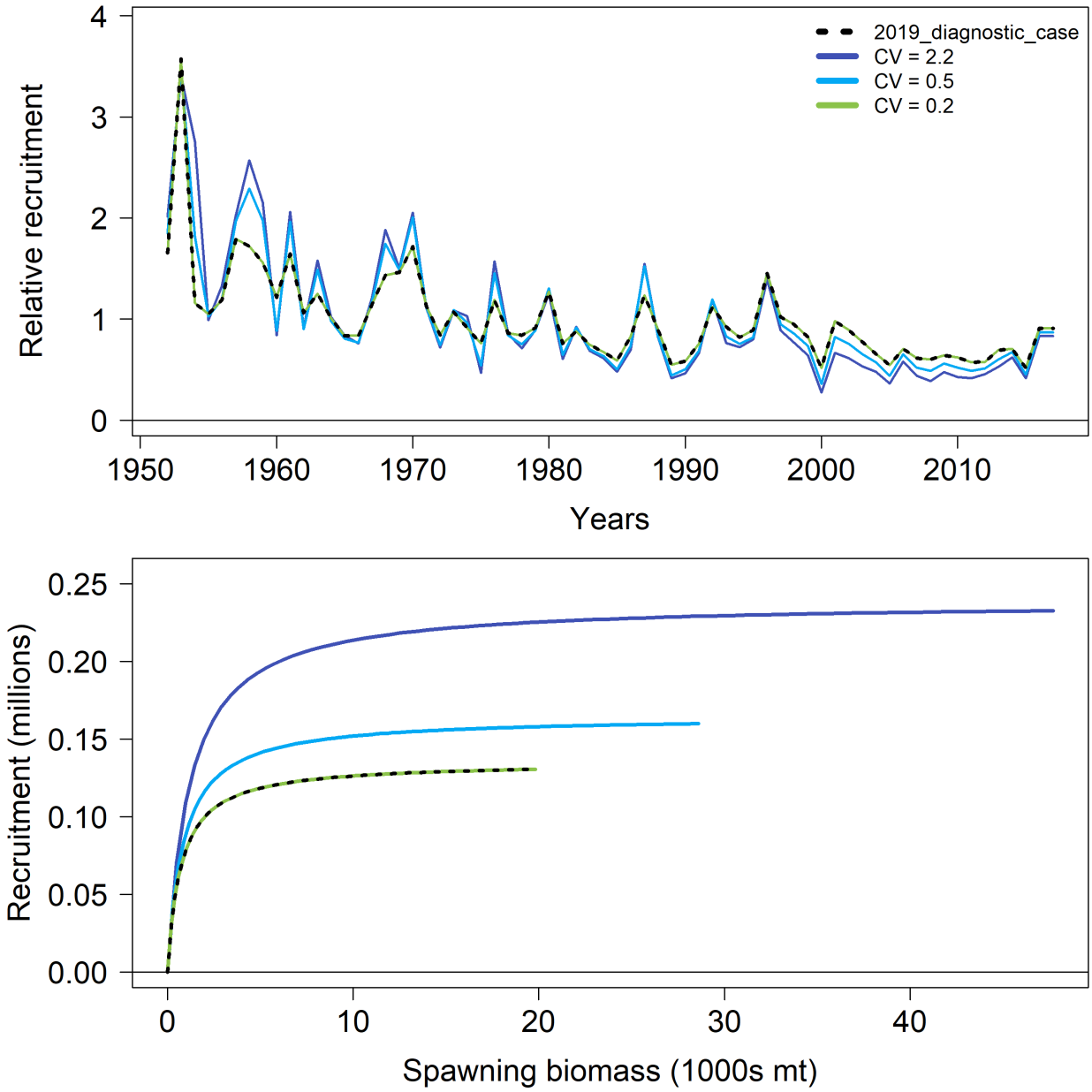

Figure 42: Relative trends in estimated recruitment (top panel) and estimated stock-recruitment relationship (bottom panel), steepness fixed at 0.8, for the 3 levels of recruitment penalty CV (0.2, 0.5, and 2.2) relative to the 2019 diagnostic case.

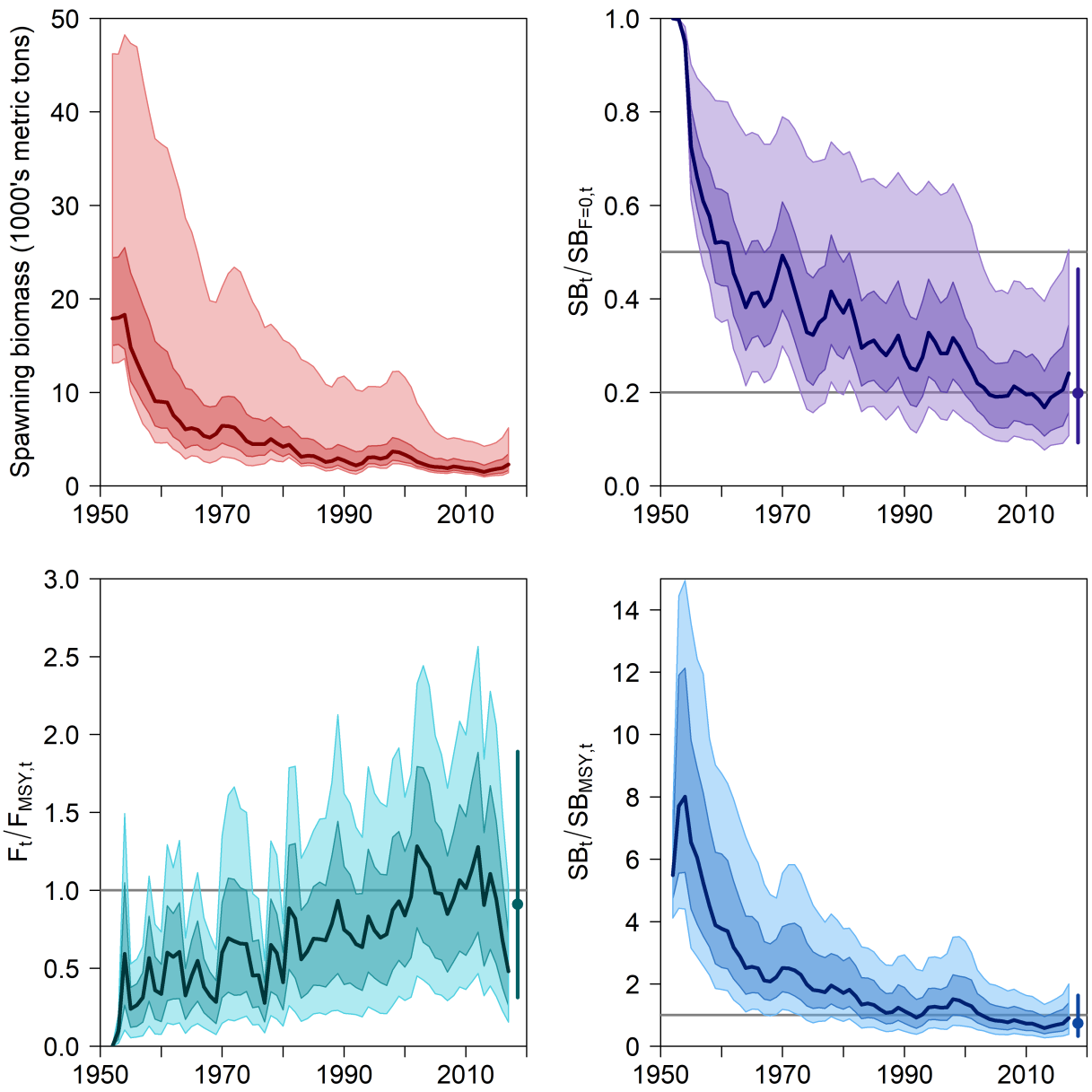

Figure 43: Distribution of time-dynamic estimates of spawning biomass (top left), adult depletion (top right), $F_t/F_{MSY,t}$ (bottom left), and $SB_t/SB_{MSY,t}$ (bottom right) across all models within the structural uncertainty grid. The median of the estimates is shown as the dark line, while the shaded regions show the 50^{th} and 80^{th} percentiles. On the right of the depletion, F_{MSY}, and SB_{MSY} panels is the median point estimate of the "recent" level reference point with the bar indicating the 80^{th} percentile.

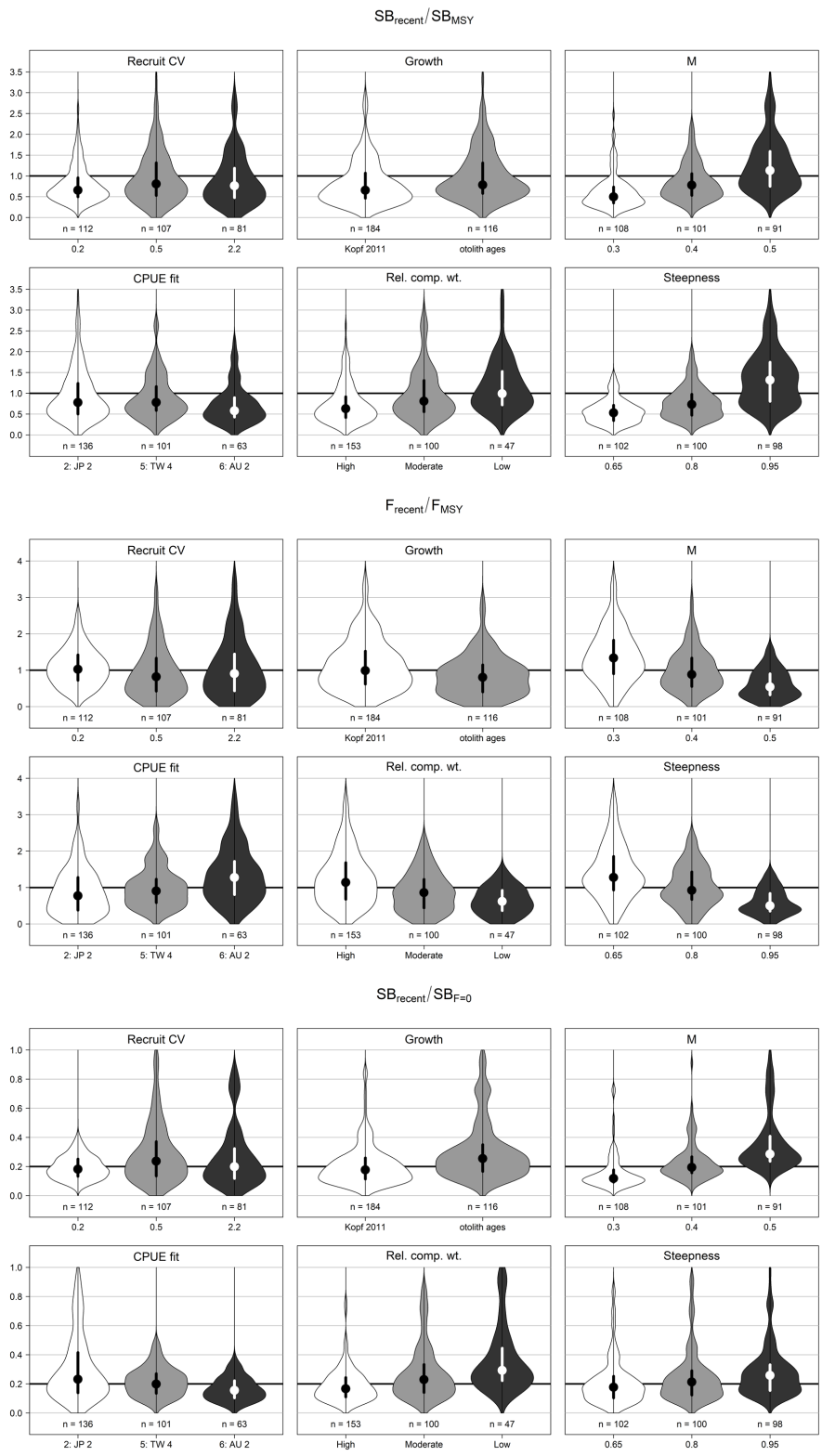

Figure 44: Violin plots showing the distribution of estimates across all axis levels in the uncertainty grid. The median of each distribution is shown by the contrasting dot, and the contrasting line gives the 50^{th} percentile range. The number below each violin gives the number of grid models with that axis level, with each panel summing to 300.

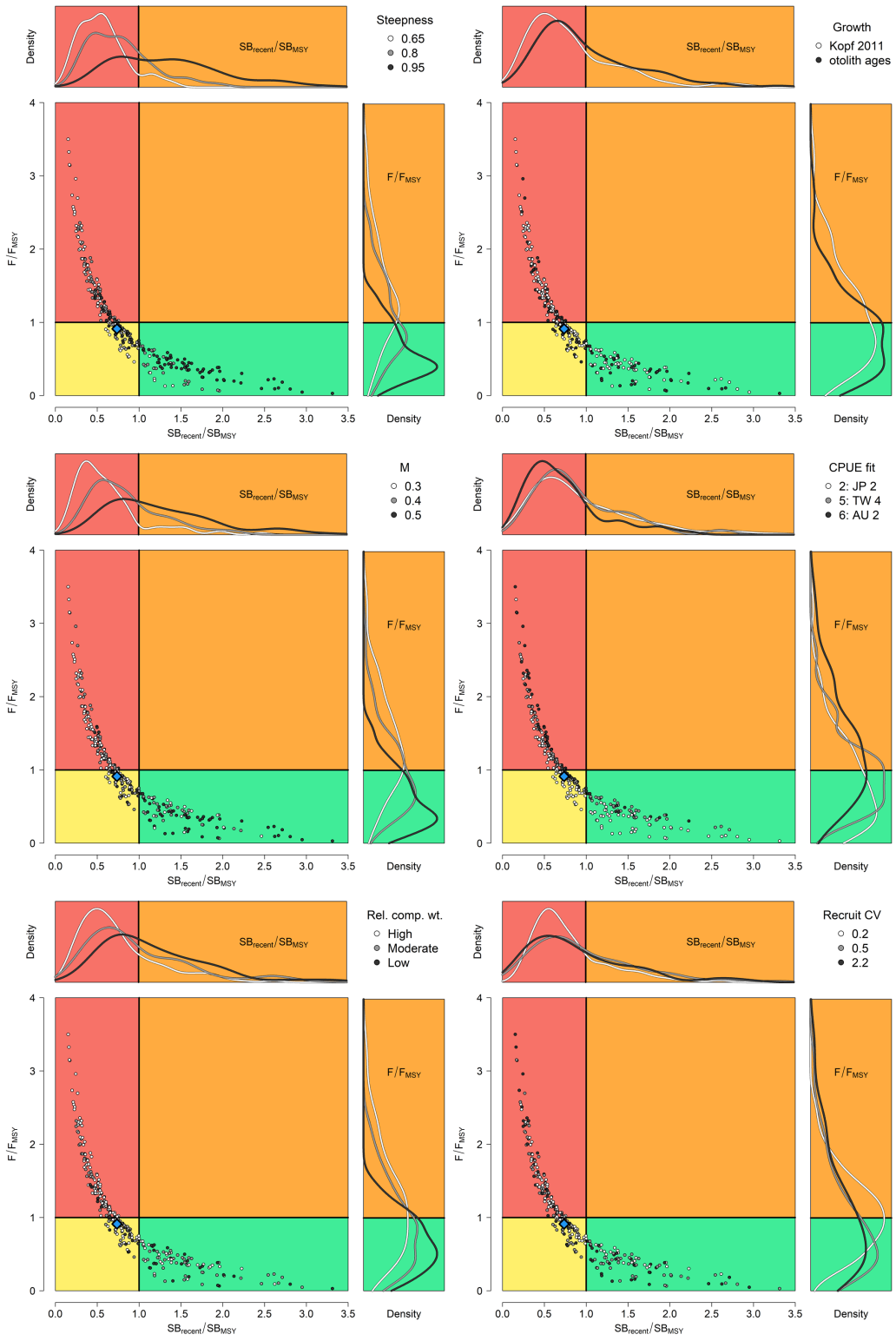

Figure 45: Kobe plots summarising the results for each of the models in the structural uncertainty grid. The plots represent estimates of stock status in terms of spawning biomass and fishing mortality. The points are colored according to the uncertainty axis level for that particular panel. The blue diamond denotes the median. On each axis of the Kobe plot is the density distribution of the management reference point with respect to the different axis levels.

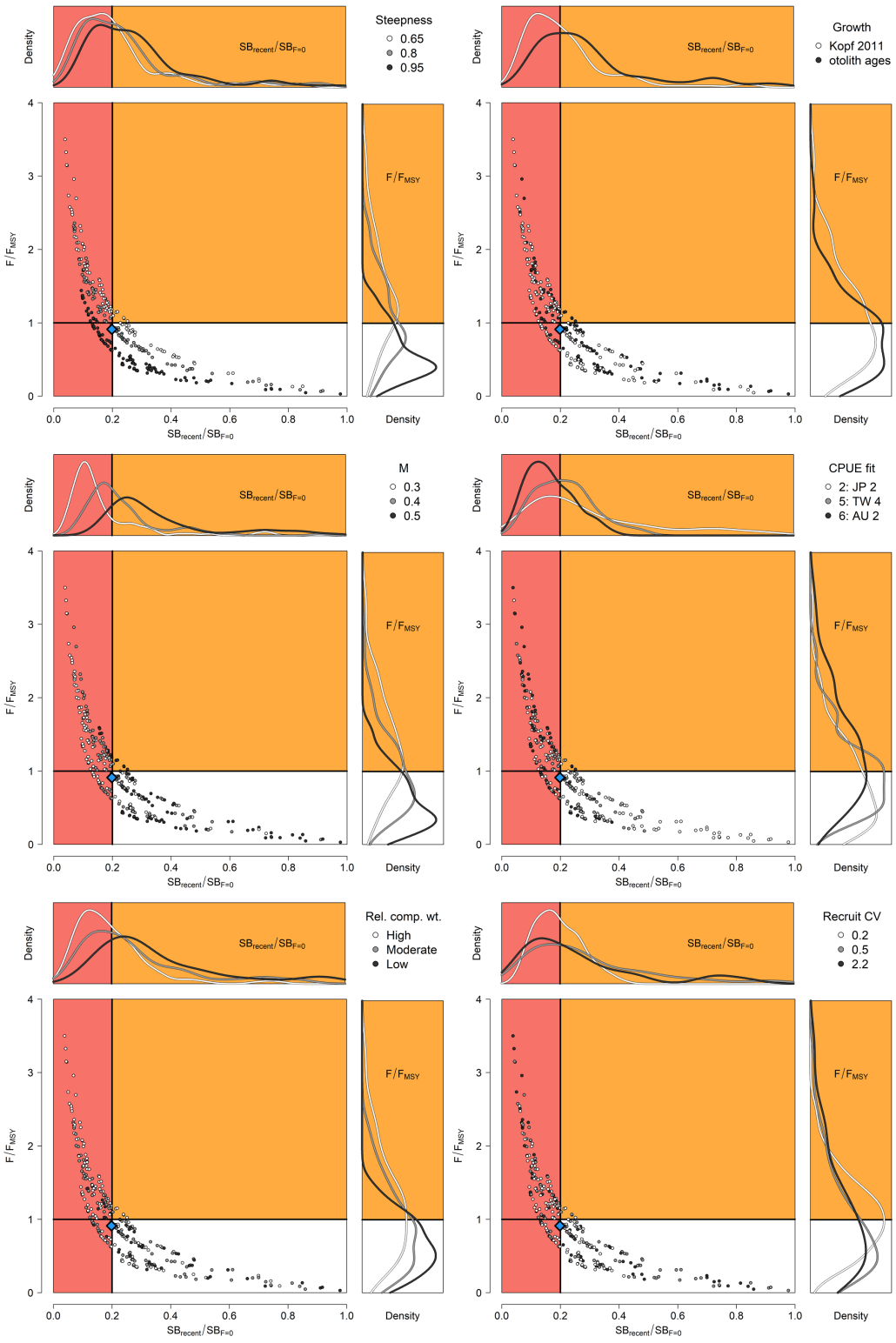

Figure 46: Majuro plots summarising the results for each of the models in the structural uncertainty grid. The plots represent estimates of stock status in terms of spawning biomass depletion and fishing mortality. The points are colored according to the uncertainty axis level for that particular panel. The blue diamond denotes the median. On each axis of the Majuro plot is the density distribution of the management reference point with respect to the different axis levels.

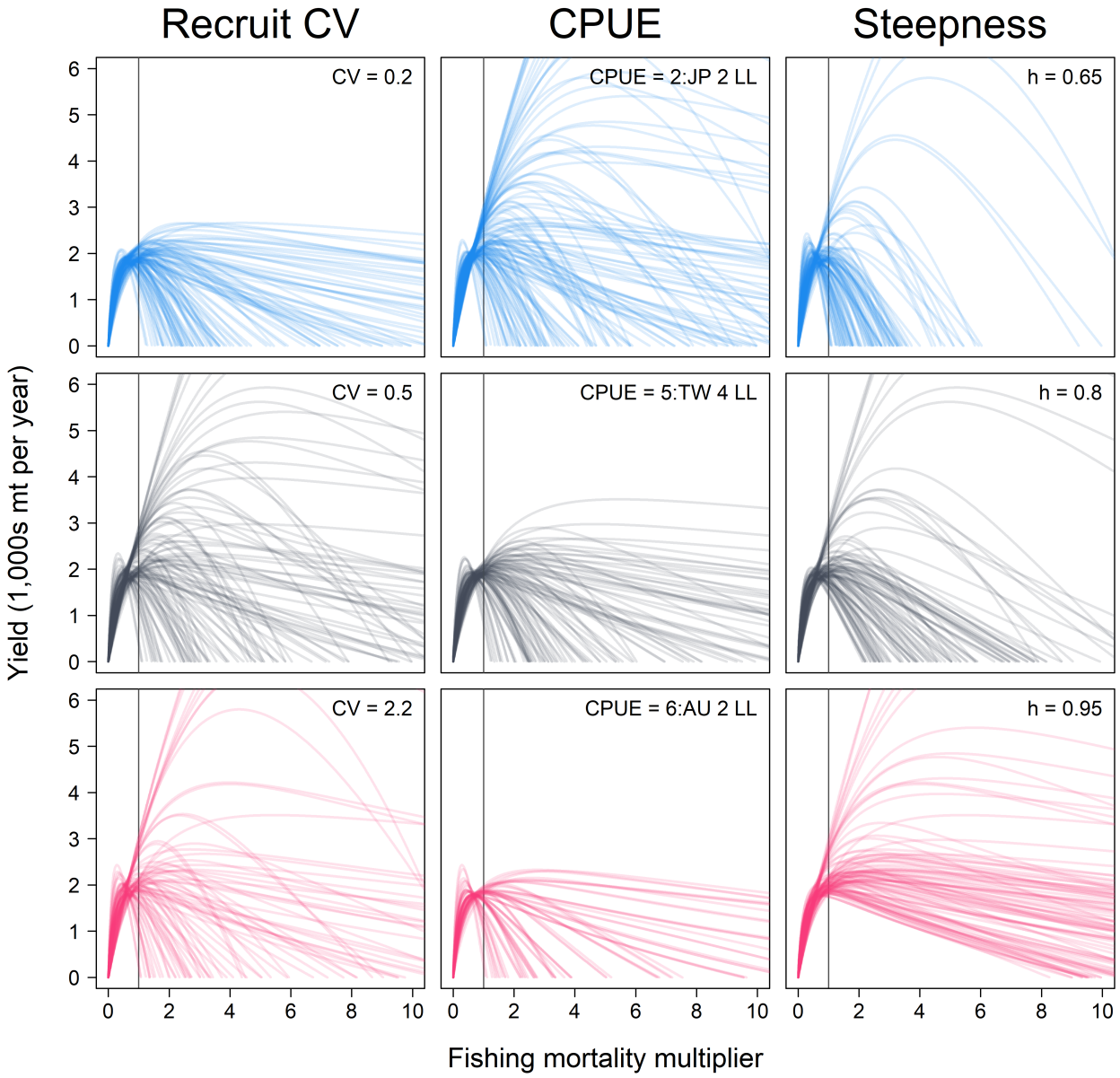

Figure 47: Estimated yield as a function of fishing mortality multiplier across all models in the statistical uncertainty grid. Three grid axes that influenced the shape and magnitude of the yield curve are highlighted (columns) where the colors correspond to the different levels of the grid axis. In each panel, the vertical bar at 1 give the yield at current levels of fishing mortality.

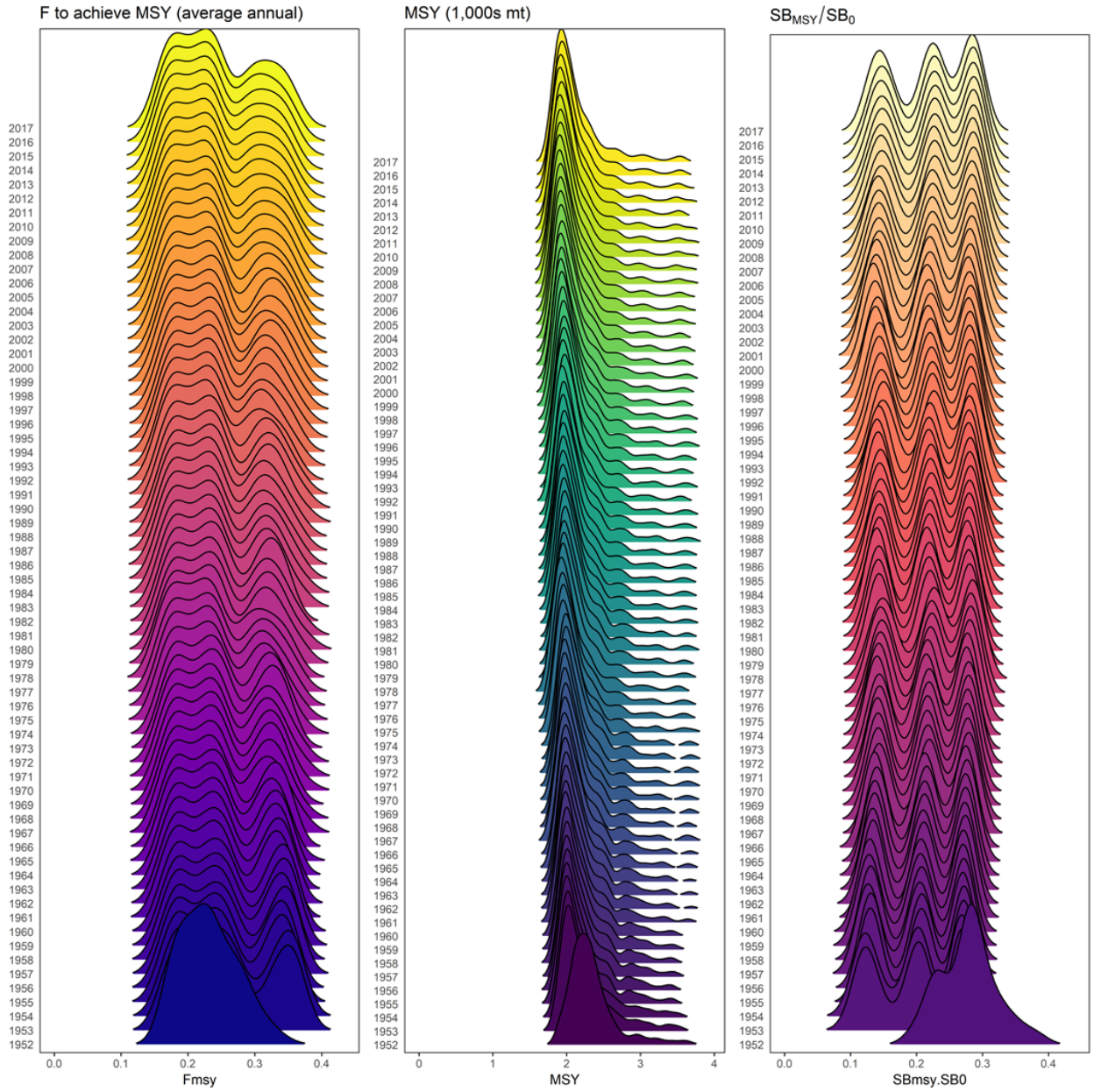

Figure 48: Time dynamic distributions of the estimates of F_{MSY}, MSY, and SB_{MSY}/SB_0 across all models in the structural uncertainty grid. The estimate across models within the first year of the assessment (1952) is shown at the bottom in dark colors. Moving progressively up the figure, the estimate across models in the final year (2017) of the assessment is shown in light colors.

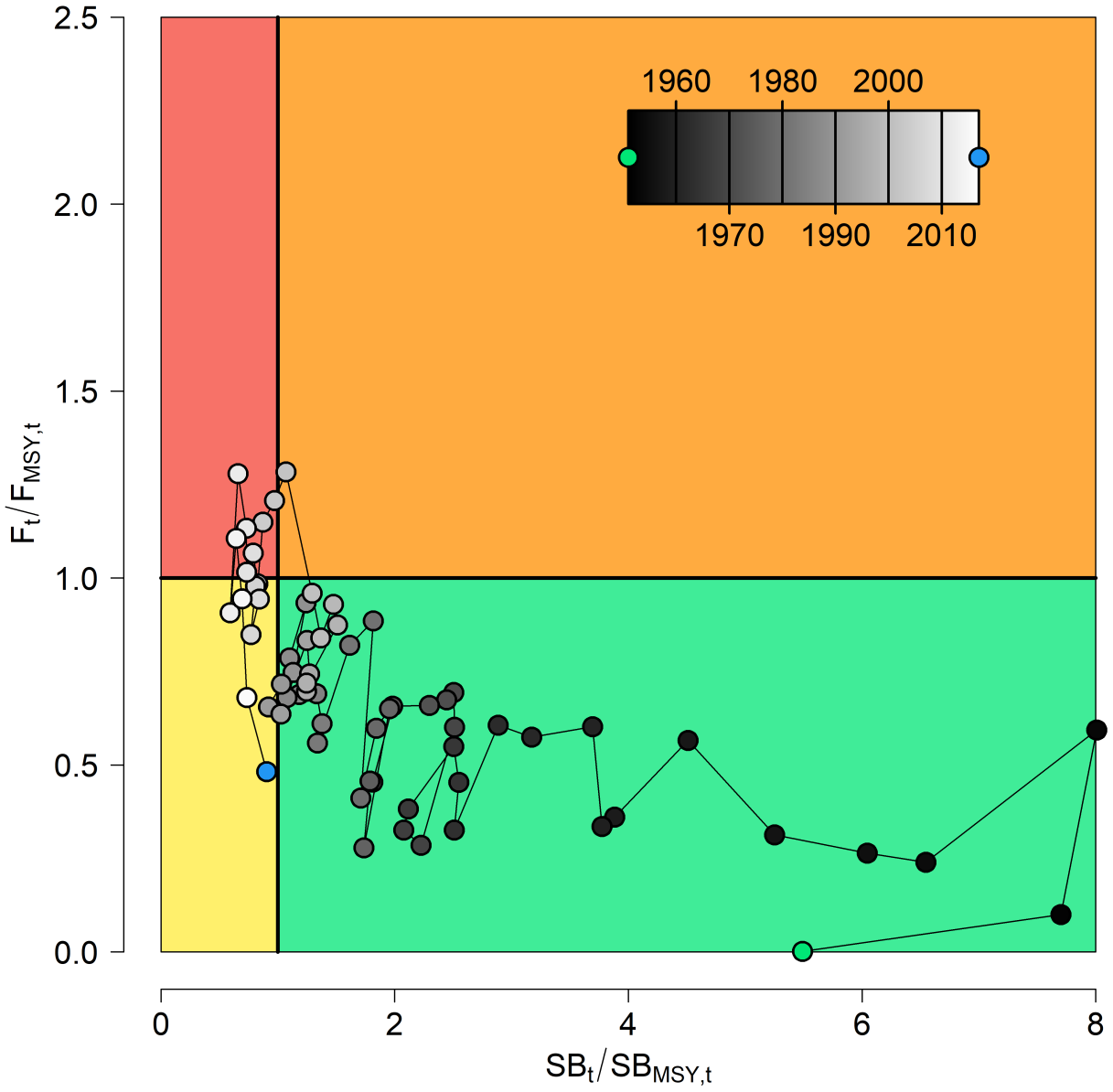

Figure 49: Time-dynamic Kobe plot based on the median estimated stock status for each model year across all uncertainty grid runs. Darker points represent estimates at the start of the assessment period (1952) with the first year shown in green. The points get progressively lighter closer to the terminal year of the assessment (2017) which is shown in blue.

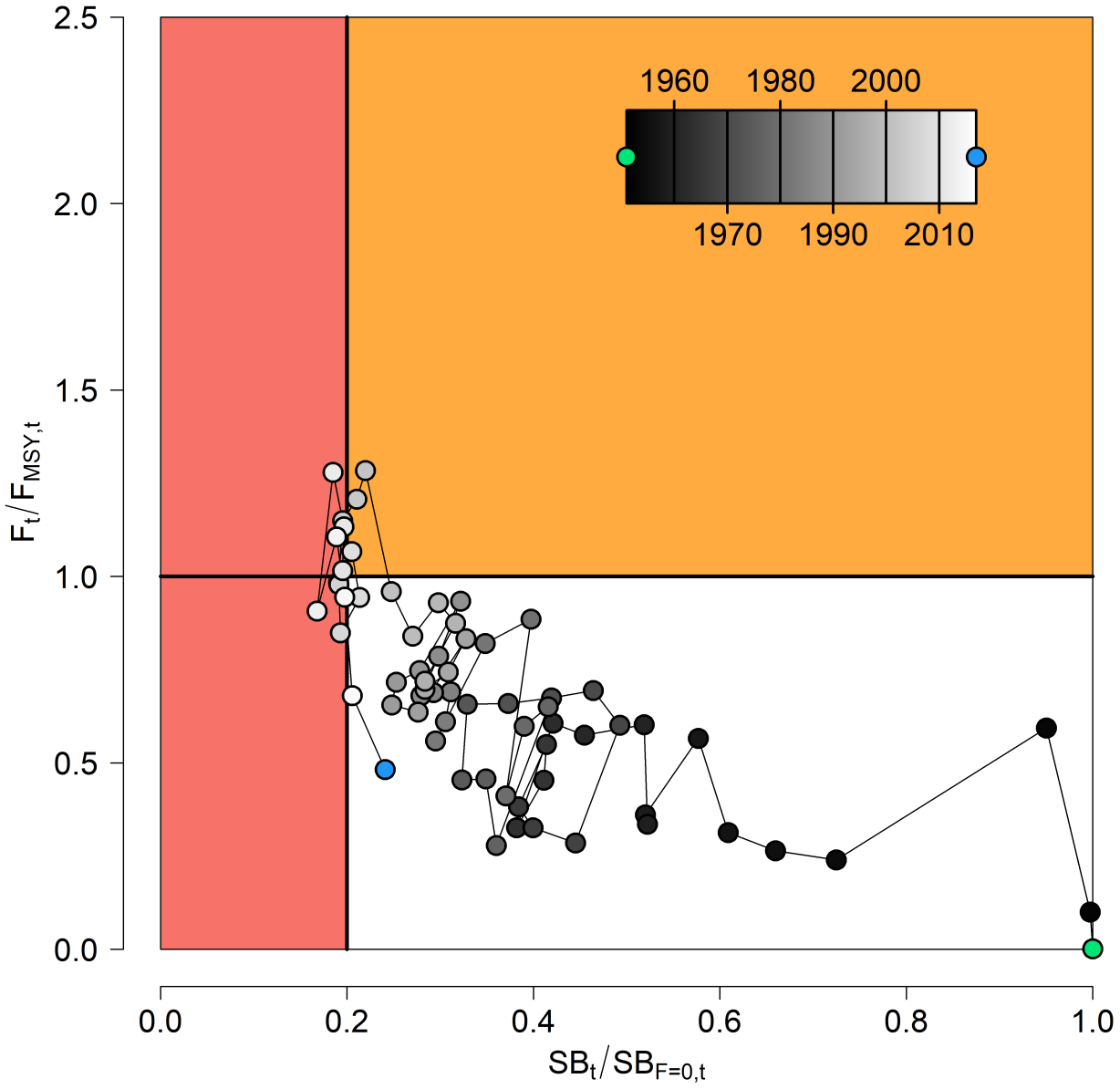

Figure 50: Time-dynamic Majuro plot based on the median estimated stock status for each model year across all uncertainty grid runs. Darker points represent estimates at the start of the assessment period (1952) with the first year shown in green. The points get progressively lighter closer to the terminal year of the assessment (2017) which is shown in blue.

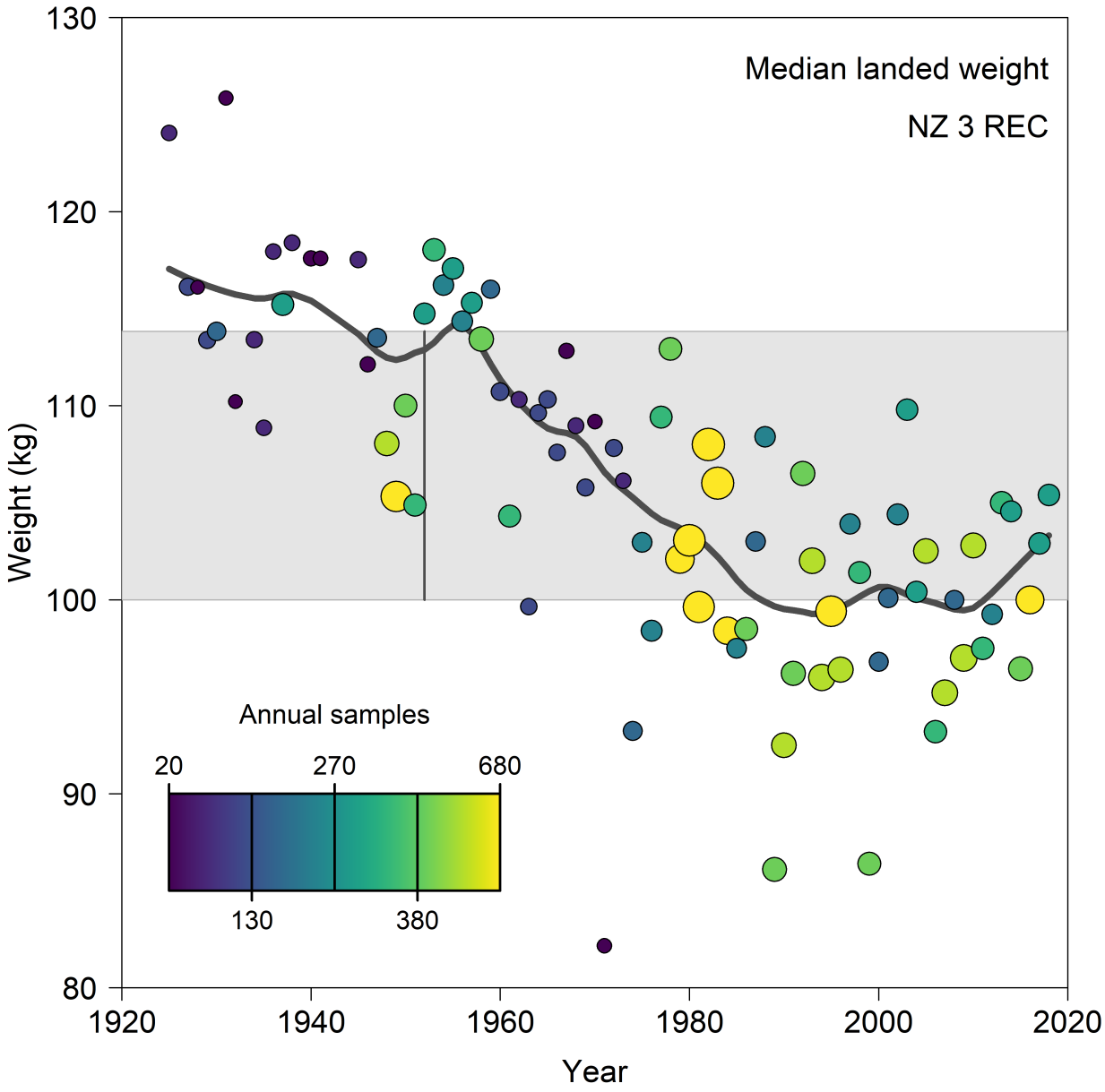

Figure 51: Median annual landed weight of striped marlin in the New Zealand recreational fishery since 1925. Color and size of dots correspond to the number of samples in a given year, and a lowess smooth in the annual trend is shown. The gray shaded region shows the change in median landed weight from the average of the pre-assessment period (< 1951) and the last 10 years of data in the assessment (2008 – 2017). The vertical bar in the shaded region denotes the start period of the model in 1952.

12 Appendix

12.1 Retrospective analyses

An additional diagnostic, the retrospective analysis, was used to assess the overall stability of the 2019 diagnostic case model and to identify any persistent biases in estimated quantities as a result of possible model mis-specification. A series of 20 additional models were fitted starting with the full data-set (through 2017), followed by models with that sequentially "peeled" away all input data for the years 2017-1998. The models are named below by the final year of data included (e.g. 1998-2017). A comparison of the spawning biomass, depletion, and recruitment trajectories are shown in Figure 52. Additionally, Mohn's ρ was calculated to ascertain the degree of retrospective bias (Hurtado-Ferro et al., 2015). All values of ρ were close to 0, and less than ± 0.2 so retrospective bias did not appear to be present. Note, that for recruitment the final two years were fixed so Mohn's ρ was calculated based off of the year of last estimated recruitment.

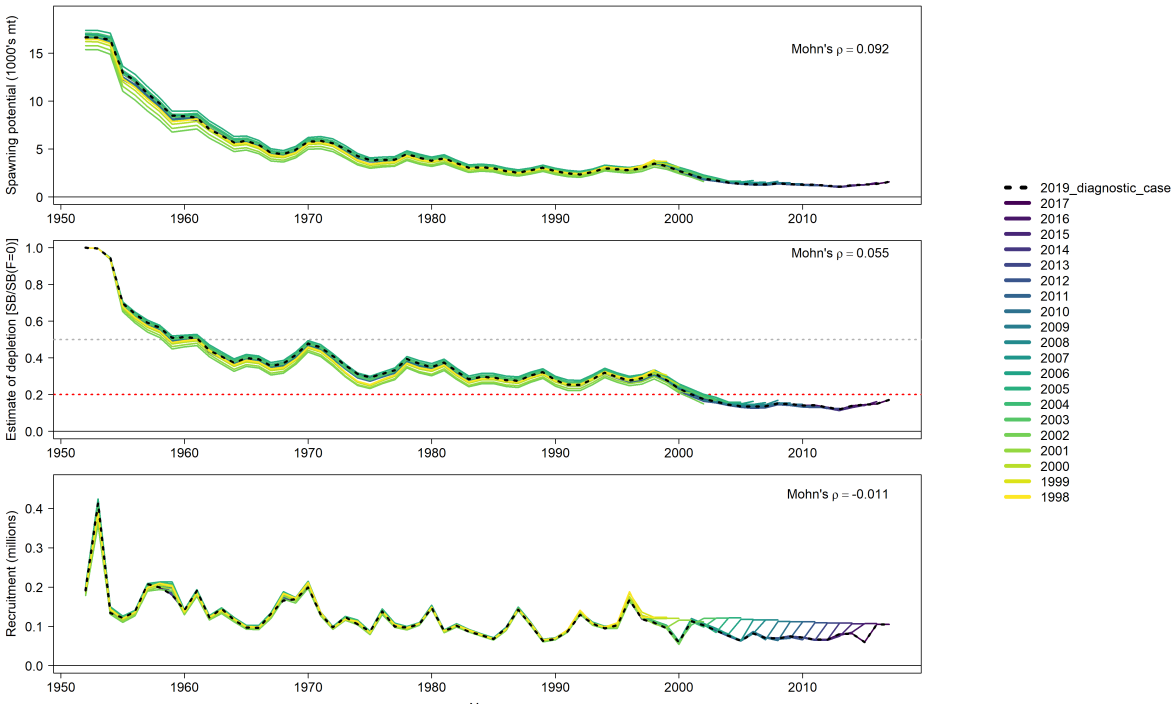

Figure 52: Estimated spawning biomass, fishery depletion ($SB_t/SB_{t,F=0}$), and recruitment for each of the retrospective models.

12.2 Sensitivity - Spatially disaggregated model

Model estimates were highly sensitive to the assumptions made for movement and recruitment partitioning between regions when using the spatially disaggregated model structure (described in Ducharme-Barth and Pilling (2019)). As shown in Figure 53 when recruitment partitioning between the two regions was estimated internally to the model biomass estimates were strongly dependent on the movement rate from region 2 into region 1 (east to west across the model region). When movement from region 2 to region 1 was moderate (runs *2area_recpartest_N.M* & *2area_recpartest_M.M*; Figure 53) estimated spawning biomass was substantially higher. Removals, particularly early in the assessment period were much higher in region 1, so limiting the flow of adults into this region of higher fishing mortality likely resulted in the higher estimated biomass. However, the estimated recruitment partitioning between the two regions seemed very low ($\sim$ 10% in region 1 & $\sim$ 90% in region 2) given that a major annual spawning aggregation of striped marlin is believed to form between 20°S and 30° within region 1 (Kopf et al., 2012). Given the limited number of small individuals in the size composition data, it is very likely that recruitment partitioning is poorly estimated to begin with. Using the *2area_recpartest_M.M* (run with the highest estimated spawning biomass) model as a baseline, recruitment partitioning was fixed at 3 levels (25%-75% split, 50%-50%, and 75%-25% split between regions 1 and 2) to explore the sensitivity of model outputs to this parameter setting (Figure 54). Doing so resulted in model estimates that were much more similar to the 2019 diagnostic case.

When movement was high from region 2 to region 1 (runs *2area_recpartest_N.H* & *2area_recpartest_M.H*; Figure 53), model results were much closer to those estimated in the 2019 diagnostic case.

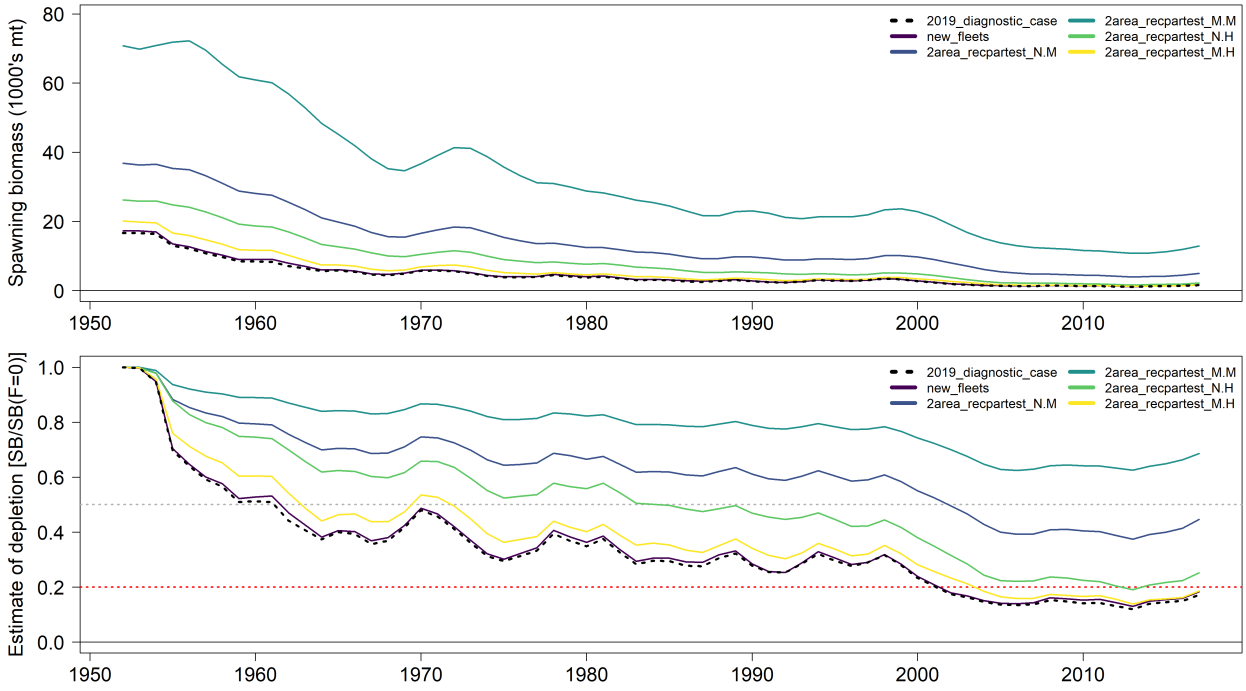

Figure 53: Comparison of sensitivity runs to the choice of movement rates between the two model regions (N = no movement; M = moderate movement; and H = high movement) relative to the 2019 diagnostic case (black dotted line). Additionally, the results from a single region model with the same fisheries structure as used in the two-region model is shown. Top panel shows this in terms of spawning biomass and the bottom panel shows it in terms of the relative depletion of the spawning biomass. For orientation, in the lower panel, the gray dotted line indicates the level of 50% depletion and the red dotted line indicates the level of 20% depletion.

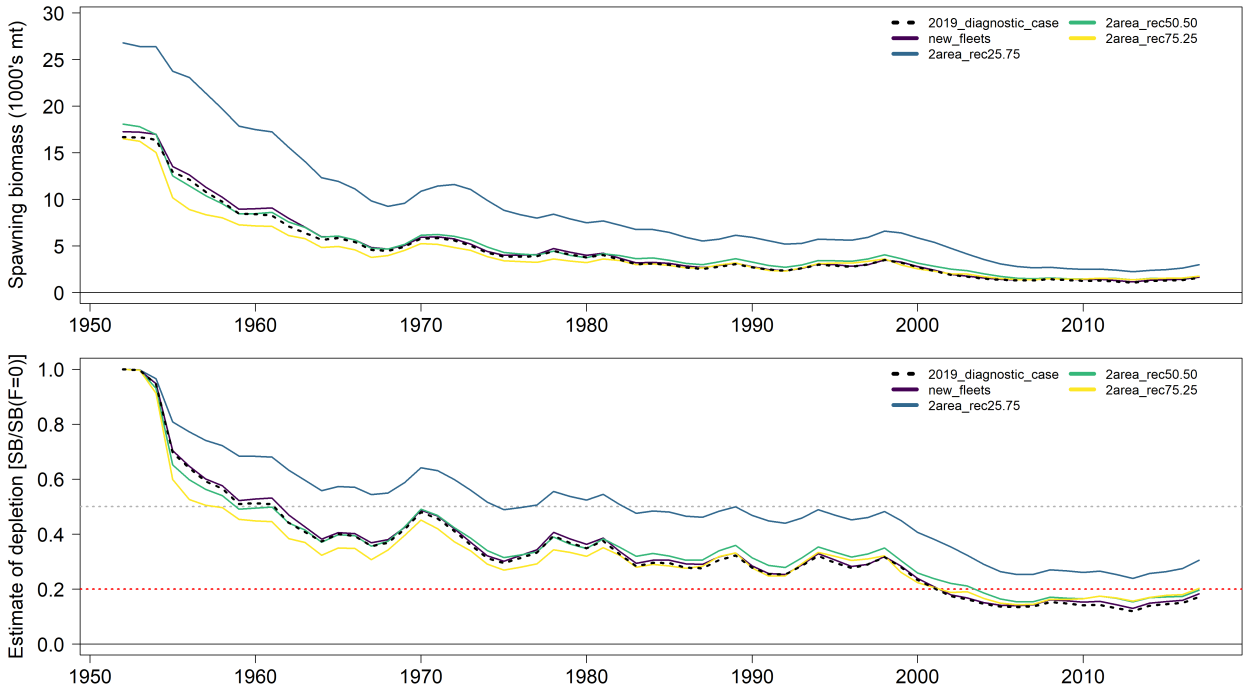

Figure 54: Comparison of sensitivity runs to the choice of recruitment paritioning between the two model regions (when movement in both directions was moderate) relative to the 2019 diagnostic case (black dotted line). Additionally, the results from a single region model with the same fisheries structure as used in the two-region model is shown. Top panel shows this in terms of spawning biomass and the bottom panel shows it in terms of the relative depletion of the spawning biomass. For orientation, in the lower panel, the gray dotted line indicates the level of 50% depletion and the red dotted line indicates the level of 20% depletion.