SCIENTIFIC COMMITTEE
SIXTEENTH REGULAR SESSION

ELECTRONIC MEETING
12-19 August 2020

Stock assessment of yellowfin tuna in the western and central Pacific Ocean

WCPFC-SC16-2020/SA-WP-04 (Rev.3)

M. Vincent, N, Ducharme-Barth, P. Hamer, J. Hampton, P. Williams, G. Pilling

Oceanic Fisheries Programme, The Pacific Community

Details of changes made during the third revision

Corrected median values for projection depletion for years 2025, 2035, and 2048 in Table A1. Correction to x-axis label in Figure 20.

Details of changes made during the second revision

A section including the future projections from the structural uncertainty grid were added in an appendix to the end of the report.

Details of changes made during the first revision of the original report

Minor changes have been made to the stock assessment report for yellowfin tuna that had previously been uploaded to the WCPFC scientific committee website. These relate to corrections to mistakes identified in the y-axis of the maturity plot (Figure 48b), which was incorrectly labeled as "Length (cm)" was changed to "Reproductive output". The values of the size frequency weighting in the structural uncertainty grid (Table 4) was corrected for the first value of 10 to 20. A section was added to the appendix about the Hessian and an additional Likelihood profile was added for the diagnostic case.

Contents

1 Executive Summary

This paper describes the 2020 stock assessment of yellowfin tuna (*Thunnus albacares*) in the western and central Pacific Ocean. A further three years of data were available since the last full stock assessment conducted in 2017, and the model extends through to the end of 2018. New developments to the stock assessment include addressing recommendations of the 2017 stock assessment report (Tremblay-Boyer et al., 2017), new information on age and growth from otoliths and the integration of growth information from tag recaptures, as well as the implementation of the Richards growth model; updates to the definition of reproductive potential; implementation of a composite 'index' longline fishery for each model region which received that region's standardized CPUE index, and concurrently a 'pseudo catch conditioned' approach taken for the assessment. Based upon recommendations of SPC's 2020 pre-assessment workshop, only the 10° N spatial structure was considered within this assessment. The assessment is supported by the analysis of longline catch-per-unit-of-effort (CPUE) data (Ducharme-Barth et al., 2020b), background analyses of biological parameters, compilation of the tagging data and definition of the fisheries structures (Vincent and Ducharme-Barth, 2020), tagging data treatment (Peatman, 2020; Scutt Phillips et al., 2020), preparation of the length and weight composition data (Peatman et al., 2020), new otolith aging and growth work (Farley et al., 2020) and estimation of growth using a tag-integrated model (Eveson et al., 2020). Key changes made in the progression from the 2017 to 2020 diagnostic models include:

- Updating all data up to the end of 2018.
- Implementation of updated models for tag data, purse seine catch estimates and size composition data.
- Implementation of the 'index fishery' approach, which used a geo-statistically standardized CPUE index.
- Utilizing updated biological parameters for the length-weight relationship and reproductive potential, and extension of the number of quarterly age classes in the model to 40.
- Changes to gear selectivity settings.
- Implementation of growth using the conditional age-at-length otolith data.

In addition to the diagnostic model, we report the results of one-off sensitivity models to explore the relative impacts of key data and model assumptions for the diagnostic model on the stock assessment results and conclusions. We also undertook a structural uncertainty analysis (model grid) for consideration in developing management advice where all possible combinations of those key areas of uncertainty were included.

As per recent Scientific Committee practice, it is recommended that management advice is formulated from the results of the structural uncertainty grid. Across the 72 models of the structural uncertainty grid, the most important uncertainties were growth, mixing period, and steepness.

Incorporation of age-at-length information for estimating growth led to more optimistic depletion estimates and exploitation rates. Models where the tag mixing period was assumed to be equal to one quarter resulted in more pessimistic stock status compared to a mixing period of two quarters. The steepness axis displayed largely predictable results, with steepness of 0.65 and 0.95 producing more pessimistic and optimistic estimates than the 0.8 assumed in the diagnostic model, respectively. The lower the steepness the more depleted the stock, and the higher the $F_{recent}/F_{\mathrm{MSY}}$.

The general conclusions of this assessment are as follows:

- Dynamic depletion remained relatively stable until the early 1970s, after which time all models in the structural uncertainty grid showed notable declines across the remaining time series. Estimated recruitment declined from the 1960s to the mid 1990s, before a slightly increasing trend.
- Estimates of stock status from the structural uncertainty grid were generally more optimistic than from the 2017 assessment. This is strongly linked to the incorporation of the new growth. However, alternative treatment of tag data and assumptions regarding selectivity to better fit the data also resulted in a more optimistic stock status.
- All models in the structural uncertainty grid show WCPO yellowfin tuna to be above $20\%SB_{F=0}$, which is consistent with the previous assessment. Median terminal depletion ($SB_{recent}/SB_{F=0}$) was 0.58 (80 percentile range: 0.51-0.64). The influence of more positive recruitments estimated in the terminal period of the previous stock assessment led to more optimistic stock status in the recent period, but have now moved through the population.
- All model regions show declines in yellowfin spawning potential over the time series from the diagnostic model. Depletion is notably greater in the tropical regions (3, 4, 7 and 8) as well as in region 5 (the south west temperate region), influenced by stronger declining signals in the regional standardized CPUE series. Other temperate regions show relatively limited estimated fishing impacts, with trends being influenced by estimated declines in recruitment. In all cases, regional terminal depletion also remained above $20\%SB_{F=0}$.
- Average fishing mortality rates for juvenile and adult age-classes increased from the 1970s, with that for juveniles plateauing from the early 2000s.
- All models in the structural uncertainty grid showed exploitation of WCPO yellowfin tuna to be below F_{MSY}. Median $F_{recent}/F_{\mathrm{MSY}}$ was 0.36 (80 percentile range: 0.27-0.47).
- A number of key research needs have been identified in undertaking this assessment that should be investigated either internally or through directed research. These include: improving our understanding of yellowfin tuna biology, including regional and spatial growth, spawning potential and natural mortality; further improving CPUE inputs to the model; and further improvements to key model settings, inputs, testing of MULTIFAN-CL developments for future use in assessments, and efforts to reduce model complexity.

2 Introduction

This paper presents the 2020 stock assessment of yellowfin tuna (*Thunnus albacares*; YFT) in the western and central Pacific Ocean (WCPO; west of 150° W). Assessments of WCPO yellowfin tuna have been conducted regularly since the late 1990s (e.g. Tremblay-Boyer et al. 2017). As for previous assessments, the objectives of the 2020 yellowfin tuna assessment are to estimate population parameters, such as time series of recruitment, biomass, biomass depletion and fishing mortality. These estimated quantities from the assessment models indicate the stock status and impacts of fishing. We summarize the stock status and fishing impacts in terms of reference points adopted by the Western and Central Pacific Fisheries Commission (WCPFC). The methodology used for the assessment is based on the general approach of integrated modeling (Fournier et al., 1998; Maunder and Punt, 2013), which is carried out using the stock assessment framework MULTIFAN-CL[2] (MFCL; Fournier et al., 1998; Hampton and Fournier, 2001; Kleiber et al., 2019). MFCL implements a size-based, age- and spatially-structured population model. Model parameters are estimated by maximizing an objective function, consisting of both likelihood (data) and prior information components (penalties).

Each new assessment of a WCPO tuna stock involves updates to fishery input and tag/recapture data, implementation of new features in the MFCL modeling software, and often the consideration of new information on biology, population structure and other population parameters. These regular changes are an important part of efforts to improve the modeling procedures and more accurately estimate fishing impacts, biological, and population processes. However, these changes can manifest in changes to the estimated status of the stock and fishing impacts. It is important to recognize that each new assessment represents a new estimation of the historic population dynamics, impacts of fishing and recent stock status. Advice from the Scientific Committee (SC) on previous assessments, and the annual pre-assessment workshops (PAW) (Hamer and Pilling, 2020) guide this ongoing process.

A feature of the tuna stock assessments in the WCPO is the use of an 'uncertainty grid'. The uncertainty grid is a group of models that are run to explore the interactions among selected axes of uncertainty that relate to biological assumptions, data inputs and data treatment. The axes are generally selected from one-off sensitivity runs of a diagnostic (or base case) model to indicate uncertainties that have notable effects on the estimates of key model parameters and management quantities. The uncertainty grid typically involves numerous models, as it includes models for all combinations of the uncertainty axes and alternative assumptions within each axis. Importantly, the uncertainty grid captures variability in model estimates due to assumptions in model structure that are not accounted for by statistical uncertainty estimated in a single model run, or a set of one-off sensitivities. Management advice should typically be based on the consideration of the range of model estimates produced by all models in the uncertainty grid and the relative plausibility of

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

each model scenario.

The most recent assessment of yellowfin tuna in the WCPO was conducted in 2017 using data up until 2015 (Tremblay-Boyer et al., 2017). The 2017 assessment indicated that stock biomass has shown a declining trend particularly since the 1970s, but that recent depletion and fishing mortality were unlikely to have exceeded their respective WCPFC limit reference points. However, there were some models in the uncertainty grid that indicated the depletion and fishing mortality limit reference points had been, or were close to being breached.

The 2020 assessment incorporates data up until and including 2018 from yellowfin fisheries across the WCPO and incorporates new biological information on age and growth from otoliths (Eveson et al., 2020; Farley et al., 2020), as recommended by the previous assessment. Preparatory work on data for each assessment is extensive and is described in limited detail in this report. We strongly advise that this assessment report is read in conjunction with several supporting papers:

- analyses of longline catch-per-unit-of-effort (CPUE) data (Ducharme-Barth et al., 2020b)
- background analyses of biological parameters, compilation of the tagging data and definition of the fisheries structures (Vincent and Ducharme-Barth, 2020)
- tagging data treatment (Peatman, 2020; Scutt Phillips et al., 2020)
- preparation of the length and weight composition data (Peatman et al., 2020)
- new otolith aging and growth work (Farley et al., 2020)
- estimation of growth using a tag-integrated model (Eveson et al., 2020)

3 Background

3.1 Stock structure

Yellowfin tuna is distributed across the Pacific, Indian and Atlantic Oceans in a continuous band between approximately 45° north and south of the equator (Grewe et al., 2015; Moore et al., 2020). Genetic studies suggest that populations in the three major oceans are largely separate (Pecoraro et al., 2018; Moore et al., 2020), although connectivity between yellowfin spawning areas in the Indian Ocean and populations in the Atlantic Ocean near south Africa has been detected (Mullins et al., 2018). Recent studies have suggested genetic differences among regions within the Pacific ocean. For example, Aguila et al. (2015) found that yellowfin tuna within the Philippines were genetically different from the Bismarck Sea. Similarly, Grewe et al. (2015, 2016) identified genetically distinct populations of yellowfin tuna in the Coral Sea, Tokelau and Baja California, and Pecoraro et al. (2018) found genetic differences between populations in the far eastern and western Pacific ocean. Analyses of SNPs suggest the occurrence of at least three stocks broadly associated

with the WPO (Western Pacific Ocean), CPO (Central Pacific Ocean) and EPO (Eastern Pacific Ocean) (Grewe et al., 2015). Observations of the distribution of yellowfin tuna larvae support that spawning occurs broadly throughout the central and western Pacific (Nishikawa et al., 1985), and an earlier genetic study by Ward et al. (1994) proposed the existence of eastern and western Pacific sub-populations separated at around 150° W. At a finer scale using close kin proximity Anderson et al. (2019) suggested that yellowfin tuna may even school with related individuals through their first year and show overlapping regional fidelity as adults. Complimentary studies using otolith chemistry have also suggested that populations at sub-regional scales may be sourced predominantly from local spawning (Wells et al., 2012; Rooker et al., 2016; Proctor et al., 2019).

The results of genetic studies are broadly consistent with tag/recapture data in suggesting that mixing between the far western and far eastern Pacific Ocean regions is limited (Moore et al., 2020). The extensive tag/recapture data available since 1989 shows that longitudinal movements among the equatorial regions of the central and western Pacific can be extensive but latitudinal movements to and from the tropical/sub-tropical latitudes may be less so (Figure 2). The longitudinal movements and continuous distribution across the Pacific suggests an isolation by distance mechanism is responsible for the genetic differences observed between the west and east Pacific regions. Despite the significant tagging and recent genetic studies there remains considerable uncertainty on sub-regional population structure in the WCPO, in particular the spatial connectivity between spawning areas, early life stages and recruitment to size classes vulnerable to fishing in different regions. This is an important area for further research. A thorough review of the status of knowledge of yellowfin tuna population structure is available in Moore et al. (2020), and concludes that the weight of evidence from both genetic and non-genetic studies supports the presence of discrete stocks of yellowfin tuna in the EPO and WCPO, as well as the potential for finer-scale spatial structuring within each of these regions. For this assessment, the stock within the domain of the model area (essentially the WCPO, west of 150° W) has been considered as a discrete stock that exhibits the same biological traits (Langley et al., 2011; Davies et al., 2014; Tremblay-Boyer et al., 2017).

Over time, the spatial complexity of the modeling of the yellowfin stock in the WCPO has increased. In the 2011 assessment the model domain was divided into 6 regions. After a review of the bigeye assessment in 2012 (Ianelli et al., 2012), the current 9 region model (Figure 1) was implemented in 2014 for both yellowfin and bigeye (McKechnie et al., 2014a; Davies et al., 2014; Harley et al., 2014), with the northern boundary of regions 3 and 4 set at 20° N. In the 2017 yellowfin assessment (Tremblay-Boyer et al., 2017) an additional option was included that involved moving the northern boundary of regions 3 and 4 to 10° N to better reflect the purse seine fishery spatial structure and the assumption of low movement rates between the equatorial and sub-tropical northern regions (Figure 2). Based on the comparisons between the 10 and 20° N options in the 2017 assessment, and the same comparisons for the concurrent bigeye assessment (Vincent et al., 2018), the 2020 PAW (Hamer and Pilling, 2020) recommended only using the 10° N option in the 2020 yellowfin

assessment (Figure 1). The 9 region spatial structure is a compromise between the limited knowledge of sub-regional population structure, fishery/fleet spatial structures and the locations of major localized tag release events (i.e. region 8 and 9).

3.2 Biological characteristics

Yellowfin tuna can reach a maximum fork length (FL) of around 180 cm and live for up to 15 years, although most fish aged to date have been less than 10 years old (Itano, 2000; Farley et al., 2020). Growth of the juvenile stage is particularly fast and they can reach a fork length of around 20-30 cm by three months of age and approximately 50 cm by 1 year (Farley et al., 2020). Length at 50% maturity in the WCPO is at around 100-110 cm (Itano, 2000) which equates to around 2 years of age. In this assessment for the purpose of computing the spawning biomass, we assume a fixed maturity schedule consistent with the observations of Itano (2000) (see Vincent and Ducharme-Barth, 2020 for details). Yellowfin tuna are thought to spawn opportunistically throughout the Pacific in waters warmer than 26° C (Itano, 2000). Larval stages are found widely in surface waters throughout the central and western Pacific (Nishikawa et al., 1985; Servidad-Bacordo et al., 2012) and at least some spawning appears to occur year-round in the WCPO. However, understanding of spatio-temporal variation in spawning fraction is limited. Important areas for spawning are thought to occur in the Banda Sea in Indonesia, the north-western Coral Sea, the eastern and southern Philippines, northeast of Solomon Islands, and around Fiji (McPherson, 1991; Gunn et al., 2002; Servidad-Bacordo et al., 2012). Juvenile yellowfin (several months of age) are prevalent in commercial fisheries in the Philippines and eastern Indonesia (Williams and FFA, 2019), suggesting this region is important for the juvenile stages.

Growth parameters can be highly influential in the estimates of management parameters by stock assessment models (McKechnie et al., 2017a; Vincent et al., 2018), and the most recent stock assessment of yellowfin in the WCPO identified growth as an important uncertainty requiring further research. Two companion papers to this assessment discuss recent work on yellowfin tuna growth involving aging from otoliths (Farley et al., 2020) and the integration of growth data from tag/recaptures and otolith aging (Eveson et al., 2020). These new studies resulted in different estimates of growth rate parameters to those previously estimated by the MFCL model based on length composition modal progression. There were also some differences between growth patterns estimated from otoliths and tag/recaptures, and the validation of age interpretation from otoliths is subject to further work. As such growth is an important axis in the uncertainty grid for this assessment. Finally, understanding of spatio-temporal variation in growth is limited and is insufficient to consider such effects in the current assessment, although previous studies summarized in Tremblay-Boyer et al. (2017) suggest it is likely for yellowfin tuna.

Natural mortality (M) rate of yellowfin tuna varies with size/age (Hampton, 2000). Mortality is highest for the smaller juveniles and estimated to be lowest for the pre-adult stage (50-80 cm FL)

0.6-0.8 yr^{-1} (Hampton, 2000). After reaching maturity it is thought that mortality increases with age, particularly in females. Sex ratios of yellowfin tuna are commonly observed to be biased toward males at larger sizes, and it is thought that this may relate to the higher mortality rates of mature females due to the physiological stresses related to spawning or a combination of this and different growth rates between males and females at older ages (Schaefer et al., 1963; Hampton, 2000; Fonteneau, 2002; Sun et al., 2006; Zhu et al., 2008). For the purpose of computing the spawning potential and mortality schedules, a good understanding of sex ratio with length is important. For this assessment we used data on sex ratios collected by observers in the WCPO, which show a similar trend to other studies with males becoming dominant in samples from around 130 cm FL (Vincent and Ducharme-Barth, 2020).

3.3 Fisheries

Yellowfin tuna is an important component of tuna fisheries throughout the WCPO. They are harvested with a wide variety of gear types, from small-scale artisanal fisheries in Pacific Island and southeast Asian waters to large, distant-water longliners and purse seiners that operate widely in equatorial and tropical waters (Williams and Thomas Ruaia, 2020). Purse seiners catch a wide size range of yellowfin tuna, however, smaller yellowfin often dominate catches associated with FADs (fish aggregation devices), whereas the longline fishery takes mostly larger adult fish (Vidal and Hamer, 2020; Williams and Thomas Ruaia, 2020).

The annual yellowfin tuna catch in the WCPO increased from 100,000 mt in 1970 to between 600,000 – 700,000 mt in recent years, mainly due to increased catches in the purse seine fishery, (Williams and Thomas Ruaia, 2020). The 2017 catch was the highest recorded at approximately 713,000 mt, and the 2018 (last of year of this assessment) catch was slightly lower at approximately 670,000 mt (Figure 3). Purse seiners harvest the majority of the yellowfin tuna catch (54% in 2019), while the longline fleet accounted for 15-20% of the catch in recent years, primarily in the equatorial regions (Figure 4, Williams and Thomas Ruaia, 2020). The remainder of the catch is dominated by the domestic fisheries of the Philippines and Indonesia, principally catching smaller individuals using a variety of small-scale gear types (e.g. pole-and-line, ringnet, gillnet, handline and seine net). Small to medium sized purse seiners based in those countries also catch fish of sizes more typical of the purse seine fisheries elsewhere (Williams and Thomas Ruaia, 2020).

Yellowfin tuna typically represent 15–20% of the overall purse-seine catch in recent years, and may contribute higher percentages of the catch in individual sets. Yellowfin tuna are often directly targeted by purse seiners, especially within unassociated schools (free schools) that are comprised of larger yellowfin compared to those associated with FADs. Unassociated sets account for approximately 70% of purse seine sets in the last few years (Williams and Thomas Ruaia, 2020).

Since 2010, annual catches of yellowfin tuna by longline vessels in the WCPO have varied between approximately 85,000 to 105,000 mt (Williams and Thomas Ruaia, 2020). The highest longline

catch recorded was 125,000 mt in 1980 (Figure 3). Annual catches from the domestic fisheries of the Philippines and eastern Indonesia are highly uncertain, particularly prior to 1990. Recent estimates for pole and line and other gears have reached approximately 150,000 mt in recent years (Figure 3).

Figure 5 shows the spatial distribution of yellowfin tuna catch in the WCPO for the past 10 years. Most of the catch is taken in western equatorial areas, with less catch by both purse-seine and longline toward the east. The east-west distribution of catch is strongly influenced by ENSO events, with larger catches taken east of 160 ° E during El Niño episodes. Catches from outside the equatorial region are relatively minor (5%) and are dominated by longline catches south of the equator and purse-seine and pole-and-line catches in the north-western area of the WCPO (Figure 4 and Figure 5).

Improved catch statistics in recent years for the Indonesian, Philippines, and Vietnamese fleets have resulted from collaborative work between the fisheries agencies of these countries and the SPC, WCPFC, and CSIRO. In some instances data are available at the individual fisheries level (e.g., longline or large-fish handline), but often statistics are aggregated across a variety of gears that typically catch small yellowfin tuna, e.g., ring-net, handline, and troll. Data for these fisheries have been included in this assessment, as described below.

4 Data compilation

The data used in the yellowfin tuna assessment conducted in MFCL consist of catch, effort, length-frequency and weight-frequency data for the fisheries defined in the analysis (Figure 6), and tag release-recapture data. Detailed summaries of the analyses and methods of producing the necessary input data are detailed in the supporting papers referred to at the end of Section 2. Conditional age-at-length data derived from the otolith direct aging methods described in Farley et al. (2020) were incorporated into this assessment. Full details of the inputs and methods conducted are not repeated here, rather we provide a brief overview of the key features and direct readers to the relevant papers throughout this section.

4.1 Spatial stratification

The broad geographic area considered in the 2020 yellowfin tuna assessment covers the Pacific Ocean to the west of 150° W, and between 50° N and 40° S (Figure 1). The eastern boundary of the assessment excludes the WCPFC Convention area component that overlaps with the IATTC area. The region extends to 140° E off south-east Australia, 110° E in south-east Asia and 120° E in the north western Pacific. Within this area there are 9 spatial assessment regions identified based on consideration of fishery characteristics and movement information from tagging studies (Tremblay-Boyer et al., 2017). The configuration for the 2020 assessment has the northern boundaries of

regions 3 and 4 set at 10° N consistent with the diagnostic model in the previous assessment. The rationale for this is described more fully in McKechnie et al. (2017b). Some small regions are included, i.e. region 8 designed to approximate the archipelagic waters of PNG and Solomon Islands, where considerable tagging effort has occurred, and region 9 that was established in 2014 to better model the tagging data from the Coral Sea (Davies et al., 2014).

4.2 Temporal stratification

The time period covered by the assessment is 1952–2018 which includes all significant post-war tuna fishing in the WCPO. Within this period, data were compiled into quarters (1; Jan–Mar, 2; Apr–Jun, 3; Jul–Sep, 4; Oct–Dec). As agreed at SC12, the assessment does not include data from the most recent calendar year. This is because these data are only finalized very late and often subject to significant revision post-SC, in particular the longline data on which this assessment greatly depends.

4.3 Definition of fisheries

MFCL requires the definition of *fisheries* within the model to consist of relatively homogeneous fishing units. Ideally, the defined fisheries will have selectivity and catchability characteristics that do not vary greatly over time and space, although in the case of catchability some allowance can be made for time series variation. For most pelagic fisheries assessments, fisheries are defined according to gear type, fishing method and region, and sometimes by vessel flag or fleet.

There are 41 fisheries defined for this assessment (Table 1) consisting of index, longline, purse seine, pole and line and various miscellaneous small-fish fisheries in Indonesia and the Philippines. The fisheries definitions for the 2020 assessment are consistent with those used in the 2017 assessment, except for the addition of an index fishery for each region. A graphical summary of the availability of data for each fishery used in the diagnostic model is provided in Figure 6. A summary of the tag/recapture data available for this assessment is provided in Table 2.

Equatorial purse seine fishing activity was aggregated over all nationalities, but stratified by region and set type, in order to sufficiently capture the variability in fishing operations. Set types were grouped into associated (i.e. log, FAD, whale, dolphin, and unknown set types) and unassociated (free-school) sets. Further fisheries were defined for pole-and-line fisheries and miscellaneous fisheries (gillnets, ringnets, hook-and-line, handlines etc.) in the western equatorial area. At least one longline fishery was defined in each region, although in regions 3 and 7 longline fishing was separated into distant water and offshore components to account for the apparent differences in fishing practices (including captured fish sizes) for these fleets in these regions.

4.4 Index fisheries

In previous yellowfin assessments, a longline fishery in each region (one per region to avoid conflicts) received standardized effort from an external CPUE analysis. These fisheries were assumed to have a fixed catchability across regions and years. As such, their CPUEs served to index abundance across space and time. In this assessment, we used an "index" fisheries approach, first applied in the 2018 assessment of South Pacific albacore (Tremblay-Boyer et al., 2018). The methodology is described in detail in that paper. Briefly, an index fishery is defined for each of the 9 model regions as a composite fishery composed of all longline fisheries operating in the assessment area. The full longline operational data-set, described in McKechnie et al. (2014b) and Ducharme-Barth et al. (2020b), was used as the basis for the index fisheries. A trivial amount of catch (1 fish per quarter) was assigned so that index fisheries are in effect non-extractive fisheries. Effort for each time step is adjusted, as with the longline fisheries receiving standardized effort in previous assessments, such that the original standardized CPUE derived using a spatio-temporal delta-Generalized Linear Mixed Model or "geostatistical" CPUE standardization model is preserved (see Ducharme-Barth et al., 2020b). The standardized indices for each region are scaled by the regional scaling factors derived from the geostatistical CPUE standardization model. Catchability for the index fisheries is then assumed to be constant over time and shared across the 9 assessment regions in order to scale the population. The regular capture (or extractive) longline fisheries are based on the same data set, but are disaggregated into the longline fisheries defined in Table 1. The catchabilities of these extraction fisheries are defined individually as they no longer need to be shared to scale abundance between the assessment regions. The size composition data (length and weight-frequency) for the extraction fisheries represents the actual composition of the removed fish for any space-time strata, and in the data preparation process are weighted by the catch in order to represent the fisheries at the spatial (region) and temporal (quarter) resolution of the model (Tremblay-Boyer et al., 2017; Peatman, 2020). For the index fisheries, the size data are weighted by CPUE (rather than by catch) so that the data are more representative of the abundance of the underlying population in each region and time period. Further, because the size data for the index and respective extractive fisheries are effectively being used twice, the likelihood weighting is adjusted such that the original intended weight (i.e. effective sample size) in the likelihood is preserved.

Overall, the index fisheries approach to generating abundance indices for model fitting is considered an improvement over the previous approach because it:

- Provides the best possible spatial and temporal coverage for the indices of relative abundance in the assessment,
- Allows the size data to be weighted by CPUE for the index fisheries, thus better representing the composition of the population, while maintaining a catch-based weighting for the capture fisheries.

4.5 Catch and effort data

4.5.1 General characteristics

Catch and effort data were compiled according to the fisheries defined in Table 1. Catches by the longline fisheries were expressed in numbers of fish, and catches for all other fisheries expressed in weight (mt). This is consistent with the form in which the catch data are recorded for these fisheries.

It is also worth noting that this assessment is implemented using a "pseudo" catch-conditioned approach. To accommodate this, effort was removed for all fisheries except for the index fisheries. This also eliminated the need to account for fisheries effort creep through the estimation of time-varying changes in catchability. A nominal effort of one was included for the final year of the model to allow the estimation of a catchability coefficient to assist with projection analyses.

Total annual catches by major gear categories for the WCPO are shown in Figure 3 and a regional breakdown is provided in Figure 4. The spatial distribution of catches over the past 10 years is provided in Figure 5. Discarded catches are estimated to be minor and were not included in the analysis. Annual catches in the northern region have been relatively stable over much of the assessment period. Most of the catch occurs in the tropical regions (3, 4, 7, and 8).

A number of significant trends in the fisheries have occurred over the model period, specifically:

- The steady increases in total yellowfin catch over most of the assessment period in the equatorial region, with 2018 being the third highest catch on record for the WCPO.
- The relatively stable catches of yellowfin in the northern temperate region by longline vessels, and to a much lesser degree, Japanese pole and line and purse seine vessels in region 1.
- The development of the equatorial purse-seine fisheries from the mid-1970s, and corresponding catches, particularly in equatorial regions 3, 4 and 8, with the purse seine catch recently at 3-5 times higher than the longline catch.
- Large changes in the purse seine fleet composition and the increasing size and efficiency of the fleet.
- The steady increase in catch for the domestic fisheries of Indonesia and the Philippines since 1970, where mostly small juveniles are taken.

4.5.2 Purse seine

In the previous assessment, purse seine catches by species were estimated using the procedure documented in Hampton and Williams (2016) as Method 3. SC15 agreed to a number of changes to the methodology (WCPFC, 2019), including the approach used to account for grab sample bias,

based on recommendations arising from WCPFC Project 60. The revised estimation approach is documented in Peatman et al. (2019).

4.5.3 Longline fisheries

The CPUE indices for the main longline fisheries in each region are one of the most important inputs to the assessment, as they provide information on trends in abundance over time for each region, and scale relative abundance among the regions. These indices are implemented using the 'index' fishery approach described in Section 4.4 and inform the model of trends and variation in the components of the stock vulnerable to fishing over time.

The index-fishery CPUE time series for the 2020 assessment (Figure 7) were derived from the operational longline dataset for the WCPO region. This dataset is an amalgamation of operational level data from the distant-water fishing nations (DWFN) and Pacific-Island countries and territories (PICTs) longline fleets operating in the Pacific basin. It represents the most complete spatio-temporal record of longline fishing activity in the Pacific spanning from 1952 through to the present and is the result of collaborative ongoing data-sharing efforts from many countries. This data-set was first created in 2015 in support of the Pacific-wide bigeye tuna stock assessment (McKechnie et al., 2015), and was subsequently analyzed to generate indices of relative abundance for the 2017 WCPFC bigeye and yellowfin tuna stock assessments (McKechnie et al., 2017a; Tremblay-Boyer et al., 2017). The spatial distribution and decadal variation in nominal longline CPUE for yellowfin tuna in the WCPO is shown in Figure 8.

In 2017, the first application of spatio-temporal modeling approaches was used in support of WCPFC stock assessments (Tremblay-Boyer et al., 2017). For this assessment we built on these previous efforts and analyses of the operational longline data, as well as the spatio-temporal modeling done for the 2019 WCPFC skipjack tuna stock assessment (Ducharme-Barth et al., 2019; Kinoshita et al., 2019). Two sets of indices were used in the stepwise model development of the 2020 yellowfin tuna diagnostic model. The first index replicated the approach taken in the previous stock assessments for bigeye and yellowfin tuna (McKechnie et al., 2017a; Tremblay-Boyer et al., 2017) and utilized a delta-GLM fit to independent partitions of data from the operational longline data-set within each assessment region. This index was only used in the initial phases of model development in order to maintain continuity in the stepwise model progression from the 2017 diagnostic model to the 2020 diagnostic model.

The primary index used in the 2020 stock assessment was estimated using the VAST spatio-temporal modeling framework (Thorson, 2019). A detailed description of the method for generating the spatio-temporal index from the longline operational level CPUE data is provided in Ducharme-Barth et al. (2020b). Spatio-temporal models have been shown to be more accurate and less biased than equivalently structured delta-GLMs when fit to fisheries dependent data (Grüss et al., 2019; Zhou et al., 2019). Additionally, explicitly modeling the spatio-temporal structure of the data

allows these models to cope with non-stationary effort distributions like the ones exhibited in the operational longline dataset (Ducharme-Barth et al., 2019). After the data-set was cleaned, the VAST approach was used to generate the relative abundance indices. In the VAST framework, the relative abundance index is the spatial average of predicted abundance once catchability effects have been "standardized" out. The model implemented by the VAST package is a spatio-temporal delta generalized linear mixed model (GLMM), an extension of the delta-GLM (Thorson et al., 2015; Ducharme-Barth et al., 2020b).

The final model applied to generate the CPUE indices included catchability covariates for vessel nationality grouping and hooks-between-floats (HBF), which is a proxy for fishing depth and species targeting. Previously, the coefficients of variation (CVs) for region-specific standardized effort were rescaled to an average level of 0.2 over a reference period of 1980-1990. Since a single spatio-temporal model was used to generate the standardized CPUE, the CVs for the standardized effort assigned to the index fishery were rescaled to an average level of 0.2 across all regions and time periods. This was done in order to more appropriately capture the relative differences in uncertainty, due in part to varying levels of sampling intensity, across spatial and temporal strata. In this way, MFCL is able to account for the time-varying nature of the CVs such that the fit to the CPUE data is given greater influence in the likelihood in time-steps with more precise estimates of abundance.

4.5.4 Other fisheries

There has been continual improvement in the catch estimates from Indonesia and the Philippines through the West Pacific East Asia (WPEA) Project and, since the 2014 assessment, some catch data from the small-fish fisheries in Vietnam have also been included. These improved catch estimates have been incorporated into the current assessment.

4.6 Size composition data

Available length-frequency data for each of the defined fisheries were compiled into 95 x 2-cm size classes (10-12 cm to 198-200 cm). Weight data were compiled into 200 x 1kg size classes (0–1 kg to 199–200 kg). Most weight data were recorded as processed weights (usually recorded to the nearest kilogram). Processing methods varied between fleets requiring the application of fishery-specific conversion factors to convert the available weight data to whole fish equivalents. Details of the conversion to whole weight are described in Langley et al. (2006). Data were either collected on-board by fishers, through observer programs, or through port sampling. Davies et al. (2011) provides more details on the source of the size data. Each length-frequency record in the model consisted of the actual number of yellowfin tuna measured. Figure 6 provides details of the temporal availability of length and weight (for longline) frequency data. Note that a maximum sample size of 1,000 is implemented in MFCL when using the robust normal likelihood for size composition

data. In practice, this effective sample size was further down-weighted, with the sensitivity to the magnitude of the down-weighting investigated in the sensitivity and structural uncertainty analyses.

4.6.1 Purse seine

Only length-frequency samples are used in the yellowfin assessment for purse seine fisheries. Prior to 2014, assessments used only observer samples which were corrected for grab-sample bias. As observer coverage had been very low and unrepresentative in early years, there were many gaps and the time series of size data did not show evidence of modal progression. Two major changes were implemented for the 2014 assessment and are described in detail in Abascal et al. (2014): first the long time series of port sampling data from Pago Pago was included, and second all samples were weighted by the catch - both at the set and strata level, with thresholds applied to ensure that small samples from important catch strata did not get too much weight (consistent with the approach taken for the longline fishery). The pre-processing of the purse seine length composition data for this assessment is based on the approach described in Abascal et al. (2014), further detailed in Peatman et al. (2020). Length-frequency data were sparse for the "all flags" associated purse seine fishery in region 7 (Fishery 30, Figure 6). In the model, this fishery was assumed to share a selectivity with the "all flags" associated purse seine fishery in the adjacent region 3 (Fishery 13).

4.6.2 Longline

A detailed review of all available length and weight frequency data for yellowfin tuna is described in McKechnie (2014). That paper along with Peatman et al. (2020), provides details for the analytical approach used to construct the longline size composition data inputs for the current assessment (Vincent and Ducharme-Barth, 2020).

The key principle used in constructing the data inputs was not to use weight and length data at the same time, even if it was available, as it would either introduce conflict (if data were in disagreement) or over-weight the model fit (if they were in agreement). The general approach used by McKechnie (2014) and Peatman et al. (2020) for the "extraction" fisheries was that weight frequency samples should be weighted with respect to the spatial distribution of flag-specific catch within each region. This is done so that catch is extracted from the population at the appropriate size. Weight-frequency data was used based on the spatio-temporal coverage and number of samples. In the previous stock assessment, Japanese weight data were not available for regions 4, 5, and 6 toward the end of the model period and was supplemented by 'all flags' length data. Additional, Japanese weight-frequency data has since been made available and is used in the current assessment. However, the number of available weight-frequency samples declines in recent years. Switching to length-frequency data for the longline fisheries in regions 4, 5, and 6 beginning in the mid-2000's is investigated as a model sensitivity.

Size composition data were prepared similarly for the index fisheries (Peatman et al., 2020). As mentioned above, the approach differed from that described for the extraction fisheries in that the size-frequency samples were re-weighted with respect to the spatial distribution of abundance as predicted by the spatio-temporal CPUE standardization model (Ducharme-Barth et al., 2020b). This is to allow size compositions to inform temporal variation in population abundance and size. To generate the size composition data for the index fisheries, data were first subset to match the nationalities of the "all flags" longline fisheries in each region. This was done to prevent shifts in size composition as a result of a change in sampling between fisheries. Sensitivity to this assumption was tested by including weight-frequency data from the US longline fleet and the Australian longline fleet to be included in the index fishery composition data for regions 2 and 5, respectively. Additionally, sensitivity to the declining availability of weight composition data in regions 4, 5, and 6 was tested by switching to length-frequency data. A third sensitivity was also tested which used all available weight composition within the the index fishery. Given that the same data were used for both the extraction and index fisheries, the observed number of size-frequency samples input into the assessment was divided by 2 for both the extraction and index fisheries. The maximum effective sample size in the stock assessment model was also divided by two for these fisheries (i.e. 500 as opposed to the default value of 1,000 assumed for the other fisheries).

4.6.3 Other fisheries

Size composition data for the Philippines domestic fisheries, both small-fish fisheries (Fishery 17) and large-fish handline fisheries (Fishery 18), were derived from a number of sampling programs conducted in the Philippines since the 1980s. In more recent years, size-sampling data have been substantially augmented by the work of the WPEA project. Length frequency data for the domestic Indonesia fishery (fishery 23) and the domestic Vietnam fishery (Fishery 32) are now included in the assessment as a result of recent data collection efforts by these countries. As in the previous assessments the length frequency samples from the Philippines domestic small fish miscellaneous fishery (Fishery 17) were adjusted to exclude all reported fish lengths greater than 90 cm. These large fish were also excluded from the length frequency data for both the domestic Indonesia (Fishery 23) and the domestic Vietnam (Fishery 32) small-scale fisheries in region 7. This was done on the basis that it is suspected that the presence of these large fish may be due to mis-reporting of the fishing gear in some of the regional sampling programs. Large-fish handline fishing sometimes takes place on mixed gear trips with other gears such as hook-and-line targeting smaller fish. To avoid "contaminating" the length frequency data for this fishery with fish that were reported as being caught using handline, fish smaller than 70 cm were excluded. Length data from the Japanese coastal purse-seine and pole-and-line fleets were provided by the National Research Institute of Far Seas Fisheries (NRIFSF). For the equatorial pole-and-line fishery, length data were available from the Japanese distant-water fleet (sourced from NRIFSF) and from the domestic fleets (Solomon Islands and PNG). Since the late 1990s, most of the length data were collected by

observers covering the Solomon Islands pole-and-line fleet.

4.7 Tagging data

A considerable amount of tagging data were available for incorporation into the current yellowfin stock assessment (Table 2). A summary of the data characteristics and the process of constructing the MFCL tagging file are described in detail by Vincent and Ducharme-Barth (2020). The data were available from the Regional Tuna Tagging Project (RTTP) during 1989–92 (including affiliated in-country projects), more recent (1995, 1999-2001) releases and returns from the Coral Sea Tagging Programme (CSTP) by CSIRO, and the Pacific Tuna Tagging Programme (PTTP) carried out during the period 2006 until the end of 2017. Additional data have become available since the 2014 assessment for the Japanese Pacific Tagging Programme (JPTP), conducted by NRIFSF and the Ajinomoto Co. Inc, over the period 2000–2017. The tagging data from the JPTP provided information about the fishing mortality and movement rates in region 1 for this assessment.

Tags were released using standard tuna tagging equipment and techniques by trained scientists and technicians. Tags have been returned from a range of fisheries, having been recovered on-board or via processing and unloading facilities throughout the Asia-Pacific region.

In the current assessment, the numbers of tag releases input to the assessment model were adjusted for a number of sources of tag loss and unusable recaptures due to the lack of adequately resolved recapture data, estimates of tag loss (shedding and initial mortality) due to variable skill of taggers, and estimates of base levels of tag shedding/tag mortality. An additional issue is that there are tag returns that were released within the WCPO but recaptured to the east, outside the assessment region, these are problematic due to not being assigned to a fishery. This adjustment or rescaling process preserves the recovery rates of tags from individual tag groups that would otherwise be biased low given that a large proportion of recaptures cannot be attributed to a recapture category in the assessment. These procedures were first described in Berger et al. (2014) and McKechnie et al. (2016). For the current assessment, Vincent and Ducharme-Barth (2020) describe the analyses done to prepare the tagging data. Additionally, the model used to adjust tags due to variability in tagger ability or 'tagger effect' has been changed to more appropriately account for covariates influencing yellowfin tuna tag recapture rates, see Scutt Phillips et al. (2020).

After tagged fish are recaptured, there is often a delay before the tag is reported and the data are entered into the tagging databases. If this delay is significant, then reported recapture rates for very recent release events will be biased low and will impact estimates of fishing mortality in the terminal time periods of the assessment. For this reason, release events that occurred after the end of 2017 were excluded from the assessment.

The largest tag data set available for inclusion in the current assessment was that from the recent PTTP, which was mainly undertaken in the western tropical Pacific from Indonesia to the Gilbert

Islands of Kiribati over the last decade. This data set was expanded since the previous assessment to 17,002 recaptures across 53 release groups (Table 2).

For incorporation into the assessment, tag releases were stratified by release region, time period of release (quarter) and the same size classes used to stratify the length-frequency data. A total of 116,125 effective releases were classified into 145 tag release groups. The returns from each size-class of each tag release group (22,406 effective, usable tag returns in total) were then classified by recapture fishery and recapture time period (quarter). Because tag returns by purse seiners were often not accompanied by information concerning the set type, tag return data were aggregated across set types for the purse seine fisheries in each region. The population dynamics model was in turn configured to estimate tag recaptures by grouping returns from these fisheries. The likelihood penalties or "priors" used for the reporting rates of the grouped tag return fisheries have been updated relative to that used in the previous assessment based on the analysis of tag seeding experiments (Peatman, 2020).

5 Model description: Structural assumptions, parameterization, and priors

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

5.1 Population dynamics

The model partitions the population into 9 spatial regions and 40 quarterly age-classes. The last age-class comprises a 'plus group' in which mortality and other characteristics are assumed to be constant. The population is 'monitored' in the model at quarterly time steps, extending through a time window of 1952–2018. The main population dynamics processes are as follows.

5.1.1 Recruitment

Recruitment is defined as the first age class in which the fish become vulnerable to a fishing gear (i.e., age = 1 quarter). Tropical tuna spawning does not typically follow a clear seasonal pattern but occurs sporadically, with more intense spawning thought to occur when food supplies are

plentiful (Itano, 2000). In this assessment it was assumed that recruitment occurs instantaneously at the beginning of each quarter. This is a discrete approximation to continuous recruitment, but provides sufficient flexibility to allow a range of variability to be incorporated into the estimates as appropriate.

Spatially-aggregated recruitment (over all model regions) was assumed to have a weak relationship with spawning potential via a Beverton and Holt stock-recruitment relationship (SRR) with a fixed value of steepness (h). Steepness is defined as the ratio of the equilibrium recruitment produced by 20% of the equilibrium unexploited spawning potential to that produced by the equilibrium unexploited spawning potential (Francis, 1992; Harley, 2011). Typically, fisheries data are not very informative about steepness (ISSF, 2011); hence, the steepness parameter was fixed at a moderate value (0.80) and the sensitivity of the model results to the value of steepness was explored by setting it to lower (0.65) and higher (0.95) values.

In the diagnostic model, it was assumed that annual recruitment was related to annual mean spawning potential, which was recommended by the 2011 Bigeye Tuna Peer Review (Ianelli et al., 2012) and was previously assumed for the South Pacific albacore and WCPO skipjack assessments (Tremblay-Boyer et al., 2018; Vincent et al., 2019).

The SRR is incorporated in MFCL mainly so that yield analysis and population projections can be undertaken for stock assessment purposes, particularly the determination of equilibrium- and depletion-based reference points. We therefore applied a weak penalty (equivalent to a CV of 2.2) for deviation from the SRR so that it would have negligible effect on recruitment and other model estimates (Hampton and Fournier, 2001), but still allow the estimation of asymptotic recruitment. This approach was recommended (rec. 20) by the 2011 bigeye assessment review (Ianelli et al., 2012). Consistent with the 2017 assessment, the SRR was calculated over the period from 1962–mid-2017, to prevent the early recruitments (which may not be well estimated), and the terminal recruitments (which are not freely estimated), from influencing the relationship.

In recent assessments of tuna in the WCPO the terminal recruitments have often been fixed at the mean recruitment of the rest of the model period. This acknowledges that these estimates are poorly supported by data and, if unconstrained, can vary widely, with potentially large consequences for stock projections. This approach has been continued here for the six terminal recruitments (quarters). The distribution of recruitment among the model regions was estimated within the model and allowed to vary over time in a relatively unconstrained fashion.

5.1.2 Initial population

The population age structure in the initial time period in each region was assumed to be in equilibrium and determined as a function of the average total mortality during the first 20 quarters. This assumption avoids having to treat the initial age structure, which is generally poorly determined,

as independent parameters in the model.

As noted above, the population is partitioned into quarterly age-classes with an aggregate class for the maximum age (plus-group). The aggregate age class makes the accumulation of old and large fish possible, which is likely in the early years of the fishery when exploitation rates were very low.

5.1.3 Growth

The standard assumptions for WCPO assessments fitted in MFCL were made concerning age and growth: i) the lengths-at-age are normally distributed for each age-class; ii) the standard deviations of length for each age-class are a log-linear function of the mean lengths-at-age; and 3) 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 regionally and temporally invariant.

As recommended by the previous assessments, and supported by SC15, two new growth data-sets were prepared for the current assessment: i) a new otolith aging data-set (Farley et al., 2020), and ii) an integrated otolith and tag-recapture growth data-set (Eveson et al., 2020). Both the traditional von Bertalanffy (VB) growth and the more flexible Richards growth (Richards, 1959) were explored to model the growth of yellowfin tuna.

Four alternative approaches to modeling growth were investigated in the current assessment:

- Otolith: This growth curve is a fixed Richards growth curve based on high-readability otoliths and the new method to estimate decimal age discussed in Farley et al. (2020). The two length-at-age standard deviation (SD) parameters, the "generic" SD and the "age dependent" SD (Kleiber et al., 2019) were freely estimated in the assessment model, while the main growth parameters were not estimated internally.

- Tag-Oto: This is a VB growth curve based on the same high readability otolith data-set as the oto-only growth in addition to yellowfin tuna tag-recapture data filtered to comprise higher confidence returns with respect to release and recapture dates and length (Eveson et al., 2020). Additionally, the two length-at-age SD parameters were freely estimated to account for variability in the length at age, but the main growth parameters were fixed.

- CondAge: A conditional age-length data-set was constructed from the otolith dataset (Farley et al., 2020) and was input to MFCL. This allowed for modal progressions, apparent in various size composition data, in addition to the otolith data to inform the estimation of all growth parameters (L1, L2, K, two SD parameters). This conditional age-length data-set makes use of the improved method for estimating decimal age as it allows for fish to be assigned to quarterly age classes without making a birth-date assumption.

- Modal: This involved estimating the growth using size composition only (i.e. modal progres-

sion) from a model with size composition not down-weighted, and estimating deviates for the first eight quarters. This is consistent with the method used in previous yellowfin tuna assessments (Tremblay-Boyer et al., 2017). The estimated growth curve parameters were then fixed for application in the structural uncertainty grid, with the exception of the two SD parameters.

5.1.4 Movement

Movement was assumed to occur instantaneously at the beginning of each quarter via movement coefficients that connect regions sharing a common boundary. Note that fish can move between non-contiguous regions in a single time step due to the 'implicit transition' computational algorithm employed (see Hampton and Fournier (2001) and Kleiber et al. (2017) for details). Movement is parameterized by a pair of bi-directional coefficients at each region boundary in each of the four quarters. Each of these coefficients is estimated independently resulting in 104 estimated movement coefficients (2 x no.region boundaries (13) $\times$ 4 quarters). There are limited data from which to estimate long-term, annual variation in movement or age-specific movement rates. As such, the estimated seasonal pattern is assumed to be fixed across years and the movement coefficients are invariant with respect to age. The default "prior" of zero is assumed for all movement coefficients, with a low penalty is applied to deviations from the "prior".

5.2 Natural mortality

In MFCL, natural mortality (M) can either be held fixed at pre-determined, age-specific values or estimated as age-specific parameters. As in previous assessments despite the male sex bias with length, males and females are not treated separately and natural mortality is treated as a rate for males and females combined by using a three phase mortality curve to account for the higher mortality of females at older ages (Hampton and Fournier, 2001; Maunder and Watters, G.M., 2001) (Figure 9). M-at-age was calculated using an approach applied to other tunas in the WCPO and EPO (Harley and Maunder, 2003; Hoyle, 2008; Hoyle and Nichol, 2008). The generally increasing proportion of males observed in the catch with increasing size is assumed to be due to an increase in the natural mortality of females at larger sizes (Section 2). Since fixed values of M-at-age are initially calculated at-length and then back transformed to age using a growth curve, it is important to calculate a specific M-at-age for each growth curve used in the modeling. This process is described in Vincent and Ducharme-Barth (2020) for the alternative growth curves considered in the current assessment (Figure 9). For this assessment it should be noted that we use a specific M-at-age for each of the fixed growth curves used in the assessment (Vincent and Ducharme-Barth, 2020). For the models in the structural uncertainty grid where growth was estimated from the conditional age-at-length, the natural mortality calculated from the growth curve of the diagnostic model was applied.

Given the new growth data for yellowfin tuna in the WCPO (Farley et al., 2020; Eveson et al., 2020), the existing assumption around the mean level of M was revisited. A meta-analytic framework, based on life-history theory and empirical estimates (Piner and Lee, 2011), was used to generate an updated estimate for M (Vincent and Ducharme-Barth, 2020). This analysis yielded an estimated quarterly M of 0.11-0.15 which is less than that used in the last assessment and the diagnostic model, i.e. 0.23. Sensitivity analyses were conducted to explore how model estimated management quantities changed when the age-specific deviates from the diagnostic model were applied to the new estimates of average quarterly M.

5.3 Reproductive potential

The reproductive potential ogive is an important component of the assessment structure as it translates model estimates of total population biomass to the relevant management quantity, spawning potential biomass (SB). Subsequent to the 2017 stock assessment, a new feature in MFCL was developed allowing the reproductive potential ogive to be defined and input into the assessment as a function of length. This length-based ogive is then converted internally to a reproductive potential-at-age using a smooth-spline approximation (Davies et al., 2019). This allows for a more natural definition of reproductive potential as the product of three length-based processes: proportion females-at-length (sex-ratio) (sex-ratio-at-length was calculated from Regional Observer Program data in SPC's holdings through 2018), proportion of females mature-at-length, and the fecundity-at-length of mature females (Figure 11). Another added benefit is that this reproductive potential ogive specified by length is growth invariant. The previous stock assessments were required to redefine the reproductive potential-at-age ogive for each different growth curve included in the assessment.

In the previous stock assessment, female sex ratio was estimated from reproductive data collected by fisheries observers, and then converted to proportion female-at-age for input into MFCL. This length-to-age conversion resulted in a greater decline in the proportion female-at-age than that observed from the longline observer data. This exaggerated decline in proportion female-at-age used in the previous stock assessment resulted in a dome-shaped reproductive potential ogive (see Figure 11; Tremblay-Boyer et al., 2017) and implied that 3–4 year old individuals made the largest contribution to spawning potential. Using the length-based input of female sex-ratio-at-length, the reproductive potential ogive, when converted to age, results in a larger contribution from older age classes, but fish of age 3–4 years still have the highest reproductive potential.

The second update to the reproductive potential ogive is the removal of spawning fraction. Previous assessments included spawning fraction at age from EPO yellowfin tuna (Schaefer, 1998) based on the sensitivity analysis of Hoyle and Nicol (2008). Given the potential regional differences in the Pacific-basin between the WCPO and EPO, we felt it more appropriate to remove this component from the definition of the reproductive potential ogive. Model sensitivity to the removal of spawning

fraction from the reproductive potential ogive is assessed in the stepwise model development.

5.4 Fishery dynamics

The interaction of the 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 are a random effect in the fishing effort – fishing mortality relationship.

5.4.1 Selectivity

Selectivity was modeled using a cubic spline, as in the previous yellowfin tuna assessment (Tremblay-Boyer et al., 2017). This allows for greater flexibility than assuming a functional relationship with age (e.g. logistic curve to model monotonically increasing selectivity or double-normal to model fisheries that select neither the youngest nor oldest fish), and requires fewer estimated parameters than modeling selectivity with separate age-specific coefficients. This is a form of smoothing, but the number of parameters for each fishery is the number of cubic spline 'nodes' that are deemed sufficient to characterize selectivity over the age range. We use five nodes which appears sufficient to allow for reasonably complex selectivity patterns.

In all cases, selectivity is assumed to be fishery-specific and time-invariant. However, in the absence of sufficient size composition to inform the shape of the selectivity curve, some fisheries were grouped with another fishery that exhibited similar operational characteristics or operated in a similar area (Table 1. This was the case for the longline fisheries (Fishery 27 Australian longline and Fishery 29 longline all) in region 9 which were each grouped with the respective longline fisheries in region 5 (Fishery 10 Australian longline, and Fishery 11 longline all). The unassociated purse seine in region 7 (Fishery 30) was grouped with the unassociated purse seine in region 3 (Fishery 13) as were the selectivities for the associated purse seine fisheries in these regions (Fishery 31 and Fishery 14). Additionally, the assumption was made that at least one longline extraction fishery within each of the 9 spatial regions (excluding the index fisheries) would have selectivity that was penalized to be increasing as a function of length in order to prevent the accumulation of an invulnerable, cryptic biomass within the model. Selectivity for the index fisheries was assumed to be shared across regions and penalized to be increasing as a function of length.

5.4.2 Catchability

Constant (time-invariant) catchability was assumed for the index fisheries that received standardized indices of relative abundance. This assumption is similar to assuming that the CPUE for these fisheries indexes the exploitable abundance over time. These index fisheries were grouped for the

purpose of initial catchability, and to maintain the relativity of catch rates among regions. This provides the model with information on the relative population sizes among regions, which previous experience suggests is difficult to estimate without this assumption.

As a result of the removal of effort data from all of the extraction fisheries, there were no penalties applied to the catchability deviations or effort deviations. Thus a "pseudo catch conditioned" approach was taken whereby the model estimated effort deviations from an assumed value. A single catchability coefficient was estimated for each fishery based on the nominal values in the final quarters to be used for projections.

5.4.3 Effort deviations

Effort deviations were used to model the random variation in the effort - fishing mortality relationship, and are constrained by pre-specified 'prior' distributions (on the log-scale). The region specific CPUE indices implemented through the index fisheries represent the principal indices of stock abundance, and the extent to which the model can deviate from the indices is moderated by the penalty weights assigned to the standardized effort series. For these fisheries the 'prior' was set to have a mean of zero and the CV was allowed to be time-variant and based on the variance estimates (using the canonical variance method of Francis (1999)) from the spatio-temporal standardization model Ducharme-Barth et al. (2020b). As explained previously, the regional differences in the estimated CVs from the spatio-temporal standardization model were preserved since they were generated from a unified analysis across regions. The CV across all index fisheries and time periods was rescaled to a value of 0.2. The resulting scaled CVs were transformed to an effort deviate penalty for each CPUE observation in MFCL. Penalties are inversely related to variance, such that lower effort penalties are associated with indices having high variance, consequently these indices are less influential in fitting the model. As seen in Figure 12, more influence is given to the CPUE from regions well sampled by the collective longline fisheries (regions 1–6 particularly later in the time series), and comparatively little weight is given to regions less well sampled by longline vessels (regions 7 and 8). Region 9 is quite small and sporadically sampled, resulting in the lowest relative penalty weight to the fit to the CV.

Finally, for all the extraction fisheries with nominal effort in the final year, the effort deviates were given effort deviation penalties equivalent to a CV of 0.41 to prevent the CPUE of these fisheries from influencing population dynamics in the final year.

5.5 Dynamics of tagged fish

5.5.1 Tag reporting

In principle, tag-reporting rates can be estimated internally within the model. In practice, experience has shown that independent information on tag-reporting rates for at least some fisheries tends to be required for reasonable model behavior to be obtained. In addition to varying by fishery, we allowed reporting rates to also vary among tagging 'programs' implemented at different times in the history of the fishery, or conducted by different agencies. We provided reporting rate 'priors' for all fishery/tagging program groups that reflect independent estimates of the reporting rates and their variances. These were derived from analyses of tag seeding experiments (Peatman, 2020). For the RTTP and PTTP, relatively informative 'priors' were formulated for the equatorial purse seine fisheries given that tag seeding experiments were focused on purse seiners. All reporting rates were assumed to be time-invariant, and have an upper bound of 0.9. Tag recapture and reporting rate groupings are provided in Table 2.

5.5.2 Tag mixing

The population dynamics of the fully recruited tagged and untagged populations are governed by the same model structures and parameters. The populations differ in respect of the recruitment process, which for the tagged population is the release of tagged fish, i.e. an individual tag and release event is the "recruitment" for that tagged population. Implicitly, we assume that the probability of recapturing a given tagged fish is the same as the probability of catching any given untagged fish in the same region and time period. For this assumption to be valid either the distribution of fishing effort must be random with respect to tagged and untagged fish and/or the tagged fish must be randomly mixed with the untagged fish. The former condition is unlikely to be met because fishing effort is almost never randomly distributed in space. The second condition is also unlikely to be met soon after release because of insufficient time for mixing to take place.

Depending on the distribution of fishing effort in relation to tag release sites, the probability of capture of tagged fish soon after release may be different to that for the untagged fish in that model region. It is therefore desirable to designate one or more time periods (quarters) after release as "pre-mixed" and compute fishing mortality for the tagged fish based on the actual recaptures, corrected for tag reporting, rather than use fishing mortalities based on the general population parameters. This in effect de-sensitizes the likelihood function to tag recaptures in the pre-mixed periods while correctly removing fish from the tagged population for the recaptures that occurred. We assume that tagged yellowfin gradually mix with the untagged population at the region-level and that this mixing process is complete by the end of the second quarter after release. We investigate the robustness to this assumption in sensitivity analyses.

Tagged fish are modeled as discrete cohorts based on the region, year, quarter and age at release for

the first 30 quarters after release. Subsequently, the tagged fish are pooled into a common group. This is to limit memory and computational requirements.

5.6 Likelihood components

There are four data components that contribute to the log-likelihood function for this assessment - the total catch data, the length-frequency data, the weight-frequency data and the tagging data. For the model in which the conditional age-at-length data were included, this constituted a fifth data component to the log-likelihood. Fit to the CPUE data does not influence the fit as an explicit likelihood component but rather as a penalty on the effort deviates.

As mentioned previously, the observed total catch data are assumed to be unbiased and relatively precise, with the SD of residuals on the log scale being 0.002.

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-frequency proportion. Size frequency samples are assigned ESS lower than the number of fish measured. Lower ESS recognize that (i) length- and weight-frequency samples are not truly random (because of non-independence in the population with respect to size) and would have higher variance as a result; and (ii) the model does not include all possible process error, resulting in further under-estimation of variances. We applied a maximum observed sample size of 1000, and then applied a divisor to obtain the ESS for each length and weight sample for a fishery. Alternative divisors for specifying ESS were explored in sensitivity analyses. We also examined in a sensitivity analysis the application of a new size-likelihood based on the Self-scaling Multinomial plus Random Effects (SSMULT-RE). This method improves upon the traditional Multinomial likelihood because it; i) does not assume that the variance is proportional to the sample size, ii) accounts for the positive correlation in residuals through using auto-correlated random effects (Davies et al., 2019), and iii) is self-scaling. This allows for the estimation of ESS from the data, and removes the need for arbitrary assumptions about ESS. As this is a new feature of MFCL (Davies et al., 2020), it was not used as the baseline likelihood structure in the diagnostic model or the structural uncertainty grid. Rather, it was investigated as a one-off sensitivity to the diagnostic model. We expect further investigation into this promising new method will occur.

A log-likelihood component for the tag data was computed using a negative binomial distribution. The negative binomial is preferred over the more commonly used Poisson distribution because tagging data often exhibit more variability than can be attributed by the Poisson. We employed a parameterization of the overdispersion parameter such that as it approaches 0, the negative binomial approaches the Poisson. Fish pars(4) in MFCL are all set to zero, which gives a tau=2, hence we assume twice the variance of the Poisson. Therefore, if the tag return data show high variability (for example, due to contagion or non-independence of tags), then the negative binomial is able to recognize this. This should then provide a more realistic weighting of the tag return

data in the overall log-likelihood. A complete derivation and description of the negative binomial likelihood function for tagging data is provided in Kleiber et al. (2019).

A further log-likelihood component is introduced for models that include the conditional age-at-length dataset (Vincent and Ducharme-Barth, 2020). These data can be included in the assessment to assist in estimating growth parameters because they provide direct observations of the distribution of fish ages within length classes. Each otolith sample was assigned to a corresponding length and age class in addition to a fishing incident based on the collection date of the sample and the gear by which it was captured. The model fits to the observed age-at-length data along with information from size mode progression to influence the estimation of the growth curve. The observed age composition within each length interval is assumed to be multinomially distributed, and this forms the basis of the likelihood component for this data source.

5.7 Parameter estimation and 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 'priors' and 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 (auto-differentiation) (Fournier et al., 2012). Estimation was conducted in a series of phases, the first of which used relatively arbitrary starting values for most parameters. A bash shell script, "doitall", implements the phased procedure for fitting the model. Some parameters were assigned specified starting values consistent with available biological information.

In this assessment two approaches were used to describe the uncertainty in key model outputs. The first estimates the statistical uncertainty within a given assessment model, while the second focuses on the structural uncertainty in the assessment by considering the variation among a suite of models. For the first approach, the Hessian was calculated for the diagnostic model run to obtain estimates of the covariance matrix, which is used in combination with the delta method to compute approximate confidence intervals for parameters of interest (for example, the biomass and recruitment trajectories). For the second approach, a factorial grid of model runs (described in the introduction) was undertaken which incorporated many of the options of uncertainty explored in one-off sensitivity analyses. This procedure attempts to describe the main sources of structural and data uncertainty in the assessment.

For highly complex population models fitted to large amounts of often conflicting data, it is common for there to be difficulties in estimating absolute abundance. As in the previous assessment, a likelihood profile analysis was conducted for the marginal posterior likelihood in respect of the total average population biomass as a measure of population scaling (Lee et al., 2014), with the definition of this parameter detailed in Kleiber et al. (2019). Rationale for profiling the likelihood with respect to total average population biomass instead of the total population scaling parameter,

along with a description of how this profile is generated can be found in Section 5.6 of the previous assessment report (Tremblay-Boyer et al., 2017). Reasonable contrast in the profile obtained using this method is taken to indicate that sufficient information existed in the data for estimating absolute abundance, and also offered confirmation that the maximum likelihood estimate obtained represented a global solution, at least with respect to total population scaling. This procedure is presented in the Appendix (11.1), including examination of the profiles for the individual data components.

Retrospective analyses are also undertaken as a general test of the stability of the model, as a robust model should produce similar output when rerun with data for the terminal year/s sequentially excluded (Cadigan and Farrell, 2005). The retrospective analyses for the 2020 diagnostic model are presented in Appendix (11.2).

Jitter analyses were performed on the diagnostic model as a diagnostic for model stability and to ensure convergence to a global solution (Cass-Calay et al., 2014). Jitter analysis entails randomly perturbing parameter estimated from the diagnostic model and refitting the model to the data. This 'jittering' can be conducted multiple times with differing levels of perturbations to determine how well behaved the model is. This is the first time these analyses have been conducted on MULTIFAN-CL models and requires further development and investigation. The results of the jitter analyses are presented in Section A.3.

5.8 Stock assessment interpretation methods

Several ancillary analyses using the fitted model/suite of 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).

5.8.1 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) for the entire model domain, 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 formulae given in Kleiber et al. (2019), and is equivalent to the MSY. Similarly the spawning potential 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 determined for all models of interest, including those in the structural uncertainty grid, and so alternative values of steepness were assumed for the SRR in many of them.

Fishing mortality-at-age (F_a) for the yield analysis was determined as the mean over a recent period of time (2013–2017). We do not include 2018 in the average as fishing mortality tends to have high uncertainty for the terminal data year of the analysis and the terminal recruitments in this year are constrained to be the average over the full time-series, which affects F for the youngest age-classes. MSY was also computed using the average annual F_a from each year included in the model. This enabled temporal trends in MSY to be assessed and a consideration of the differences in MSY levels under historical patterns of age-specific exploitation.

5.8.2 Depletion and fishery impact

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

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

A similar approach can be used to estimate depletion associated with specific fisheries or groups of fisheries. Here, fishery groups of interest - LL, PS associated sets, PS unassociated sets, miscellaneous PS fisheries and Other fisheries, are removed in turn in separate simulations. The changes in depletion observed in these runs are then indicative of the depletion caused by the removed fisheries.

5.8.3 Reference points

The unfished spawning potential ($SB_{F=0}$) in each time period was calculated given the estimated recruitments and the Beverton-Holt SRR. This offers a basis for comparing the exploited population relative to the population subject to natural mortality only. The WCPFC adopted $20\% SB_{F=0}$ as a limit reference point (LRP) for the yellowfin stock where $SB_{F=0}$ is calculated here as the average over the period 2008–2017. Stock status was referenced against these points by calculating $SB_{recent}/SB_{F=0}$ and $SB_{latest}/SB_{F=0}$, where SB_{latest} and SB_{recent} are the estimated spawning

potential in 2018, and the mean over 2015–2018, respectively (Table 3).

The other key reference point, $F_{recent}/F_{\text{MSY}}$ (Table 3), is the estimated average fishing mortality over the full assessment area for a recent period of time (F_{recent}; 2014–2017 for this stock assessment) divided by the fishing mortality producing MSY which is produced by the yield analysis and has been detailed in Section 5.8.1.

5.8.4 Majuro and Kobe plots

For the standard yield analysis (Section 5.8.1), the fishing mortality-at-age, F_a, is determined as the average over some recent period of time (2014–2017 herein). 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, with no value calculated for the terminal year) by repeating the yield analysis for each year in turn. This enabled temporal trends in the reference point variables to be estimated taking account of the 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 Developments from the last assessment

Model development from the 2017 diagnostic model to the 2020 diagnostic model occurred incrementally via successive steps. 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 2016-2018, use of a spatio-temporal standardization model to prepare the CPUE data (Ducharme-Barth et al., 2020b), modified methods in producing the input files (Peatman, 2020; Peatman et al., 2020; Scutt Phillips et al., 2020; Vincent and Ducharme-Barth, 2020), updated biological information (Eveson et al., 2020; Farley et al., 2020; Vincent and Ducharme-Barth, 2020), and incorporation of new features to MFCL developed since the last assessment (Davies et al., 2020). These changes occurred in the following sequence of steps:

1. The 2017 diagnostic model [Diag17].
2. The 2017 diagnostic model run with the updated MFCL executable, v2.0.7.0 [NewExe].
3. Change of model misspecification of tag return grouping where tags from PL in different regions had been modeled together, and update of the reporting rate [TagGrp].

4. A complete update of the 2017 diagnostic model - all inputs extended from 2015 to 2018 using identical methodology for CPUE, size frequencies etc, and the same MFCL model settings, but tagging data that utilized the new tagger effects correction [Update].

5. The previous model with the purse seine catch estimates based on the methodology documented in Peatman et al. (2019) [PeatmanPS].

6. The previous model with the CPUE coming from the spatio-temporal model described in Ducharme-Barth et al. (2020b) [GeoStats].

7. The previous model with the size composition data prepared according to Peatman et al. (2020) [SizeComp].

8. The previous model with the switch to the index fisheries approach [Index].

9. The index fishery approach with removal of effort information from all other fisheries [Idx-NoEff]

10. Ungrouping of selectivity parameters between regions to better fit the data [SelUngroup]

11. The previous model where tag mixing was enforced to be 182 days for each release event based on the recommendations of the 2020 pre-assessment workshop, incorporation of tags with incomplete recapture location within the mixing period and addition of the Japanese Tagging Program. These steps were combined because they collectively had nearly identical trajectories [JPTP].

12. The previous model but extending the number of age classes in the model to 40 (10 years old) and updating the length-weight relationship; the latter had minimal impact on management quantities [Age 10-LW].

13. Incorporation of otolith data into the estimation of growth through the conditional age-at-length feature in MFCL [CondAge].

14. Use of maturity at length information to estimate the spawning potential given the growth curve [MatLength].

15. The previous model with the spawning fraction removed from the calculation of the reproductive potential ogive [NoSpnFrac].

16. Increasing the scalar on size composition to a value of 60 attempting to replicate the likelihood of the other data components as estimated by a model using SSMULT [Size60].

17. Fixing selectivity for age class 1 to 0 for the purse seine and pole-and-line fisheries in an attempt to prevent recruitment only coming from the temperate regions. This step became the diagnostic model for the 2020 assessment [Diag20].

6.2 Sensitivity analyses

During the course of model development for the 2020 yellowfin tuna 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 2020 pre-assessment workshop (Hamer and Pilling, 2020). 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 entailed 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 2020 diagnostic model. 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. For this reason, we only provide tables of management quantities for the structural uncertainty grid.

6.2.1 Steepness [*H0.65, H0.95*]

Steepness is a particularly difficult parameter to estimate in stock assessment models. Yet if it is fixed in the model, the choice of value may have significant influence on most reference points used in management. As was the case in other tropical tuna and albacore tuna assessments, we assume a value of 0.8 for the diagnostic model, but examined values of 0.65 (*H0.65*) and 0.95 (*H0.95*) in sensitivity runs. This choice of values is consistent with the results of the meta-analysis conducted on tuna stock-recruitment data and has been well established in previous Scientific Committees.

6.2.2 Tag mixing period [*Mix1*]

The tag mixing period is imposed to allow tagged fish to distribute themselves throughout the region of tagging, although it is somewhat difficult to ascertain how long this period should be. In the diagnostic model the mixing period was set at two quarters and an alternative model was run assuming a mixing period of one quarter (*Mix1*).

6.2.3 Relative weighting of length and weight frequency data [*Size20, Size60, Size200, Size500*]

The difficulties in assigning weighting to the size-frequency data were discussed previously. To assess the sensitivity of model results to the weighting of these data, four alternative models were considered; a model where size frequency data were up-weighted (corresponding to a maximum effective sample size of 50 fish) relative to the diagnostic model (*Size20*); and several where the

size-frequency data were increasingly down-weighted to maximum effective sample sizes of 16.5 (*Size60*), 5 (*Size200*), and 2 (*Size500*). Note that in each case these maximum effective sample sizes are further reduced by a factor of two for the longline all and index fisheries. This was done because size composition samples in these fisheries were from the same database and thus would be double counted otherwise.

6.2.4 Self-scaling Multinomial plus Random Effects likelihood for the size-frequency data [SSMULT-RE]

The Self-scaling Multinomial plus Random Effects (SSMULT-RE) likelihood is a new development in MFCL (Davies et al., 2020), and offers the potential to estimate the effective sample size of frequency samples used in the model, thus obviating the need for largely arbitrary assumptions that determine the overall weighting of the size data in the model likelihood. We applied the SSMULT-RE in a sensitivity analysis in this assessment.

6.2.5 Alternative growth functions [*Otolith, Tag-Oto, CondAge, Modal*]

The importance of the growth function and the details of the alternative methods used in the assessment are detailed in Section 5.1.3, with three methods additional to the diagnostic model (conditional age-at-length = *CondAge*) being implemented as sensitivity runs. The diagnostic model estimated growth by including the conditional age-length data, constructed from the otolith age data, within the model-fitting procedure of MFCL. The second sensitivity considered the Richards Otolith (otolith only) growth model described by Farley et al. (2020) (*Otolith*). The third sensitivity considered the *Tag-Oto* growth model (otolith and tag data) described by Eveson et al. (2020) and the fourth sensitivity considered growth estimated internal by MFCL from the size composition modal progressions (*Modal*). This growth curve was estimated from a model size composition data were not down-weighted and then later fixed in each relevant model.

6.2.6 Model start year [***Start1962***]

One of the features of the model seen in the previous assessment was a high spike in recruitment paired with a rapid decline in biomass within the first 10 years of the model despite relatively low levels of catch. This indicates that, at least in the early years, the CPUE may be hyper-depleted given the level of catch that was removed from the system. An alternative models was investigated to test the robustness of the model to including this early period by beginning the model period in 1962 (*Start1962*).

6.2.7 Natural mortality [*M 0.11, M 0.13, M 0.15*]

As mentioned before in Section 5.1.3, the new growth information available for this assessment provided an opportunity to revisit the assumption made on the mean level of M using a meta-analytic approach. Model sensitivity to changing the assumed mean level of M was explored by applying the age-specific deviates to the mean level of M from the meta-analysis (M-mid = 0.13), and the levels corresponding to the lower (M-low = 0.11) and upper (M-hi = 0.15) 95% confidence intervals.

6.2.8 Size-frequency data [*WtAll, WtJP, WtUSAU, WtLen*]

The composition data for the index fisheries were restricted to weight-frequency data coming from fishing nations defined for the longline ALL fisheries. This assumption ignores the large amount of additional weight-frequency data available from other fleets. Therefore, a sensitivity model was conducted to determine if the inclusion of the additional weight composition from the off-shore fisheries impacted stock status estimates (*WtAll*). Similarly, weight composition available from the US and Australian longlines was incorporated into the size composition for the region 2 index fishery and the region 5 index fishery, respectively (*WtUSAU*). The weight frequency data from these two fisheries (Fishery 3 and 10) were correspondingly down-weighted to account for their inclusion as a part of the index fisheries' weight-frequency data. There was also concern that different flags may have different selectivity patterns and the ad hoc collection of samples for these fleets may overly influence trends in mean size in the weight composition. Therefore, we conducted a sensitivity where only the Japanese weight data were incorporated into the index fishery (*WtJP*). The availability of weight-frequency data has declined for the longline ALL and index fisheries in regions 4-6 beginning in the mid-2000s. An additional one-off sensitivity (*WtLen*) was conducted to investigate the effects of switching to length-frequency data for these fisheries in the mid-2000s.

6.2.9 Sensitivity to alternative CPUE series [*PSCPUE, HLPSPHPS8*]

Yellowfin tuna captured in purse seine sets associated with drifting FADs are predominately smaller juveniles and sub-adults. A recent geo-spatial analysis was conducted in an attempt to create a standardized CPUE index for yellowfin tuna from purse seine sets on drifting FADs that could index the juvenile component of the stock (Vidal and Hamer, 2020). The analysis was focused on regions 3, 4 and 8, where smaller yellowfin are particularly abundant. This new index, which is being further developed, was applied as a sensitivity in this assessment (*PSCPUE*). Similar to the 2017 diagnostic model, standardized CPUE indices for the Philippines-Indonesia handline fishery, Philippines-Indonesia purse seine fishery, and the region 8 purse seine fishery were incorporated into a sensitivity model (*HLPSPHPS8*). These indices were previously used to account for the lack

of an abundance index from the longline fishery in these regions, but the use of the spatial-temporal CPUE index allows estimation for these areas.

6.3 Structural uncertainty

Stock assessments of pelagic species in the WCPO use an approach to assess the structural uncertainty in the assessment model by running a "grid" of models to explore the interactions among selected "axes" of uncertainty. The grid contains all combinations of two or more parameter settings or assumptions for each uncertainty axis. The axes are generally selected from those key factors explored in the one-off sensitivities with the aim of providing an approximate understanding of variability in model estimates due to assumptions in model structure not accounted for by statistical uncertainty estimated in a single model run, or over a set of one-off sensitivities.

The structural uncertainty grid for the 2020 assessment was constructed from 4 axes steepness (3 levels), growth (3), size data weighting (4), and tag mixing period (2) with the settings used directly comparable to those presented in Section 6.2 through identical notation. The final grid consisted of 72 models (Table 4). The Tag-Oto growth model was not included in the uncertainty as stepwise model tests suggested that it performed similar to the *Otolith* model. However, it is subject to further investigation, and was included as a one-off sensitivity.

7 Results

7.1 Consequences of model development steps

In order to examine the impacts of stepwise developments from the 2017 yellowfin assessment (Tremblay-Boyer et al., 2017) to the diagnostic model for the current assessment, for each of the stepwise developments described in Section 6.1 the resulting temporal trajectories of spawning potential and spawning potential depletion ($SB_{[t]}/SB_{F=0[t]}$) are compared (Figure 13a and Figure 14a). A summary of the consequences of this progression through the model development steps is as follows:

2017 diagnostic model with new executable
Diag17 $\rightarrow$ *NewExe.*

Applying the latest version of the MULTIFAN-CL software (version 2.0.7.0)(NewExe) reduced spawning potential compared to the 2017 diagnostic model in the early period of the time series, but by the recent period this effect was minor and terminal spawning potential was similar (Figure 13a). For depletion, the NewExe was similar to the 2017 diagnostic at the start of the time series, but by the recent period under the NewExe the stock was estimated to be more depleted than for the 2017 diagnostic (Figure 14a). These differences were a result of a better fit to the data based on a

larger likelihood value. However, when the new executable was applied to the previous structural uncertainty grid the management advice was unchanged.

Correction of model misspecification of tag groupings and update reporting rate
Diag17–NewExe → TagGrp.

The correction of model misspecification of tag return grouping where tags from PL in regions 3 and 8 had been modeled together, and the update of the reporting rate had the effect of increasing the spawning potential slightly and similarly across the time series, and reducing the depletion slightly, particularly for the last 10 years (Figures 13a and 14a).

Complete update of 2017 diagnostic, all inputs extended from 2015 to 2018, same methods and model settings, new tagger effects correction
Diag17–NewExe–TagGrp → Update.

Updating of the 2017 diagnostic model with data up until 2018 and applying the CPUE standardization approach, and MFCL model settings, but with the new tagger effects correction, resulted in a notable increase in spawning potential in the early years of the time series, but by the recent years this effect was minimal (Figure 13a). Depletion was slightly lower in terminal years (Figure 14a) though the overall trend was shifted marginally upward.

Purse seine catch estimates using Peatman et al. (2019) method
Diag17–NewExe–TagGrp–Update → PeatmanPS.

The change to the purse seine catch estimation method led to a slight reduction in spawning potential across the time series, and had a slightly more depleted status in recent years compared to the previous step (Figure 13a and Figure 14a).

CPUE from the spatio-temporal model described in Ducharme-Barth et al. (2020b)
Diag17–NewExe–TagGrp–Update–PeatmanPS → Geostats.

Using the CPUE from the geostats method resulted in a minor change in spawning potential and spawning depletion in the last year of the model (Figure 13a and Figure 14a).

Size composition data prepared using methods of Peatman et al. (2020)
Diag17–NewExe–TagGrp–Update–PeatmanPS–Geostats → SizeComp.

Applying the size composition data preparation method outlined in Peatman et al. (2020) resulted in an increase in spawning potential across the entire time series (Figure 13a). For depletion this step resulted in a more optimistic state across all model years, with the largest difference in recent years (Figure 14a).

Switch to index fisheries
Diag17–NewExe–TagGrp–Update–PeatmanPS–Geostats–SizeComp → Index.

Applying the index fisheries resulted in a reduction of spawning potential for most of the time

series but similar spawning potential as the previous step in the recent years (Figure 13a). Similarly, spawning depletion was nearly identical to the previous step in the last years of the model (Figure 14a).

Index fisheries with effort removed from other fisheries

Diag17–NewExe–TagGrp–Update–PeatmanPS–Geostats–SizeComp–Index → IdxNoEff.

Applying the index fisheries with effort removed for the other fisheries resulted in a decrease in the spawning potential, particularly in the early years of the model (Figures 13a and 13b). This change thus resulted in a slightly more depleted status in the terminal year (Figures 14a and 14b).

Separation of selectivity patterns among regions, assigning non-decreasing penalty to one fishery per region

Diag17–NewExe–TagGrp–Update–PeatmanPS–Geostats–SizeComp–Index–IdxNoEff → SelUngroup.

Estimating selectivity parameters for fisheries separately resulted in a larger spawning potential across the entire time series (Figure 13b). Depletion estimates were more optimistic, particularly in the final years (Figure 14b).

Change to mixing period, inclusion of tags with incomplete recapture data, include JPTP

Diag17–NewExe–TagGrp–Update–PeatmanPS–Geostats–SizeComp–Index–IdxNoEff–SelUngroup → JPTP.

Forcing the mixing period of the tag recaptures to be 182 days for each tag release had the largest effect on this combined step. The inclusion of tags in the mixing period without recapture locations in the purse seine fisheries and the addition of the Japanese tagging program data had a net effect the same as the mixing period enforcement (Figure 13b). This step resulted in a more optimistic depletion state compared to the previous step (Figure 14b).

Extend age classes to 40

Diag17–NewExe–TagGrp–Update–PeatmanPS–Geostats–SizeComp–Index–IdxNoEff–SelUngroup–JPTP → Age10-LW.

Extending the model to 40 age classes (from 27) resulted in an increase in the spawning potential (Figure 13b). However, the depletion in the last year for this model was nearly identical to the JPTP step (Figure 14b).

Addition of otolith data through conditional age-at-length

Diag17–NewExe–TagGrp–Update–PeatmanPS–Geostats–SizeComp–Index–IdxNoEff–SelUngroup–JPTP–Age10-LW → CondAge.

The addition of the otolith data and the removal of the deviates from the growth curve in the first 8 age classes resulted in a small decrease in the spawning potential (Figure 13b). Conversely, this model resulted in a more optimistic estimation of depletion (Figure 14b).

Maturity-at-age calculated internally from maturity-at-length

Diag17–NewExe–TagGrp–Update–PeatmanPS–Geostats–SizeComp–Index–IdxNoEff–SelUngroup–JPTP–Age10-LW–CondAge → MatLength.

The calculation of maturity at age internally within the assessment model from the maturity-at-length ogive resulted in a higher spawning potential (Figure 13b). However, the estimate of depletion was relatively unaffected by this change (Figure 14b).

Maturity at length removing the spawning fraction from the ogive calculation

Diag17–NewExe–TagGrp–Update–PeatmanPS–Geostats–SizeComp–Index–IdxNoEff–SelUngroup–JPTP–Age10-LW–CondAge–MatLength → NoSpnFrac.

The removal of the spawning fraction from the maturity ogive had limited impact on the spawning potential (Figure 13b) and the depletion compared to the previous step (Figure 14b).

Down-weighting the size composition data with a scalar of 60

Diag17–NewExe–TagGrp–Update–PeatmanPS–Geostats–SizeComp–Index–IdxNoEff–SelUngroup–JPTP–Age10-LW–CondAge–MatLength–NoSpnFrac → Size60.

The size composition down-weighting with a divisor of 60, in an attempt to match the fits to the other data sources by a model using SSMULT, resulted in a marginally smaller spawning potential for most years except the last few compared to the previous model (Figure 13b). This model has a similar depletion to the previous step (Figure 14b).

Constraint to fix selectivity in age class 1 to 0 for the purse seine and pole-and-line fisheries

Diag17–NewExe–TagGrp–Update–PeatmanPS–Geostats–SizeComp–Index–IdxNoEff–SelUngroup–JPTP–Age10-LW–CondAge–MatLength–NoSpnFrac–Size60 → Diag20.

In order to fix an overestimation of the fish caught by the purse seine and pole-and-line fisheries less than 30 cm we fixed the selectivity for the first age class for these fisheries to equal 0. This step resulted in a decrease in the estimated spawning potential (Figure 13b) and a slightly more depleted status (Figure 14b).

Overall, the effects of the stepwise developments from the 2017 diagnostic model to the 2020 diagnostic model were a notable increase in historical spawning potential of around 65% in the 1950s, but an even greater increase in the recent spawning potential of 80–100%. While the trends were somewhat similar, particularly in the last 20 years, the development of the new diagnostic model has involved a rescaling of the spawning potential. This effect was not related to the updating of 2017 diagnostic case model with the new data, or applying the the new MFCL executable. While the new executable resulted in a reduction in estimated spawning potential and higher depletion, consecutive steps involving changes to data treatment (i.e., purse seine composition, geostats CPUE, size comp data treatment, index fisheries, removing effort for the extraction fisheries) each had minor effects, mostly involving slight increases in spawning potential, so that by the terminal years

these steps resulted in spawning potential and depletion levels similar to those estimated by the 2017 diagnostic case (Figure 13a and Figure 14a). Changes to the selectivity assumptions led to the more optimistic stock status of the 2020 diagnostic model. However, it was the enforcing of a mixing period of 182 days for the tagging data and the incorporation of the conditional age-at-length data that resulted in the largest changes in depletion status. The changes to the tagging data led to reduced information available to inform the model on fishing mortality. This is due to changing the attrition curve during the first period at liberty after the mixing period, which is arguably the most influential in estimating fishing mortality (Figure 24).

7.2 Model fit for the diagnostic model

7.2.1 Catch data

As very high penalties were applied to the catch data for all fisheries, model-predicted catches closely matched the observed data (Figure 15), with a SD of residuals on the log scale of 0.002.

7.2.2 Standardized CPUE - Geostats

An important change from the 2017 assessment was the development of the goeostats (VAST) abundance indices from the CPUE data. In theory the VAST approach should provide an improved representation of spatio-temporal abundance trends compared to the traditional delta-GLM approach. The companion paper on the geostats CPUE modeling makes some important observations (Ducharme-Barth et al., 2020b). Firstly, the geostats approach while smoothing out some of the short-term spikes in nominal CPUE to a greater extent than the delta-GLM, overall was very similar to the nominal and delta-GLM indices (Figure 7). It was suggested that the lack of standardization effect on CPUE over the 60+ year time series, during which significant improvements in fishing efficiency are highly likely, indicates that the standardization model is lacking covariates to appropriately measure changes in catchability of the longline fleets over time (Ducharme-Barth et al., 2020b). The implications of this are discussed further in the discussion section. The model-predicted CPUE fit the overall trend from the VAST CPUE indices; however seasonal variability in these indices were not fit very well, particularly early in the time seres (Figure 16). In general, the model captures the longer-term declines seen in the longline CPUE trends, as well as the more recent dynamics, with the exception of region 7 where the model overestimates CPUE for a period around 2010 (Figure 16). Increased variability early in the time series leads to some degradation of fit, but the overall trend is still captured. The early variability cannot be explained by seasonal effects on their own, and may reflect lower sample sizes (and generally greater uncertainty) early in the operational dataset together with a small proportion of the assessment region actually being fished. Consequently, these values have very high CVs (lower penalties) (Figure 12), which allows

the model to deviate from them to an extent, likewise for region 7 where there is lower confidence in the CPUE in recent years due to small sample sizes.

These overall patterns of CPUE fit are reflected in the time-series of effort deviations (Figure 17) where there is an even spread of positive and negative deviations across the time series. The exceptions are region 7 for a period around 2010 (negative effort deviates) and region 8 where a trend from negative to positive effort deviations occurs over the last decade.

7.2.3 Size frequency data

Time-aggregated length frequency data were adequately fit by the diagnostic model where sample sizes were sufficiently high (Figure 18). Similar to the previous assessment, high frequencies of lower sizes (40-50cm) were under-fit by the model. This is allowed to occur due to an over-estimation of very small (less than 30 cm) and larger size classes. This is likely related to model misspecification of the selectivity function. Temporal trends in median lengths are well fit for most fisheries (Figure 19). However, the model underestimates median length for PS.UNASS.4 toward the end of the time series, but overestimates the median length for the associated purse seine fishery in that region (PS.Assoc.4). The model also generally overestimates the median length in region 8 for both the associated and unassociated purse seine. The drop seen in the observed median length for the PS.JP.1 in the late 1980s is not well fit by the model.

Time-aggregated model fits to the weight-frequency data for the longline, index fisheries and the HL.IDPH7 were generally good (Figure 20). There were slight underestimations of the modal weight frequencies for the Australian, US and region 8 longline fisheries (LL.AU.6, LL.AU.9, LL.ALL.8, LL.US.2). The model overestimated the presence of fish greater than 70 kg in the extraction and index fisheries in regions 3 and 7, while underestimating fish of this size caught by the Index fisheries in regions 2, 6, and 9.

The model fits the median weights across time for most fisheries reasonably well (Figure 21). However, there was a clear overestimation the recent decade for several fisheries (i.e., LL.ALL.1, LL.ALL.2, LL.ALL.7, Index 1, 2, 7). Similarly, the median weight for region 8 (Index 8 and LL.All.8) is overestimated for the entire time series (Figure 21). The decline observed in the time series for the LL.OS.3 is not fit well by the model.

7.2.4 Tagging data

The model fit to the overall tagging data aggregated across programs and release events, but excluding those recaptures within the mixing period, was good (Figure 22), with some underestimation of recaptures noted for the early 1990s and around 2008-2010 and 2014-2015. The observed and model-predicted recapture numbers show large increases coinciding with the two tagging programs that released the most fish (RTTP; early 1990's, PTTP; 2006-present) (Figure 22).

For the individual recapture groups in the purse seine and miscellaneous fisheries, the model fit to the observed recapture rates were good for fisheries with high numbers of recaptures (Figure 23). For the longline and pole and line fisheries there were much lower recaptures and the model fits were consequently poor (Figure 23).

The model predicted recaptures with respect to time at liberty, fitted the observed tag attrition reasonably well (Figure 24a). With the exception of three quarters at release, the model generally underestimates the number of tag returns in the first 6 quarters and then slightly overestimates the returns after that (Figure 24b). The same trend is observed on the log-scale. At the program scale there are generally the same trends as overall, except that the tag returns in the third quarter are underestimated for the JPTP (Figure 25). At the regional scale, large differences in the scales between regions make interpretation slightly difficult. However, residuals in region 5 suggest poor fits to the data but this is likely due to the small number of recaptures.

7.3 Model parameter estimates (diagnostic model)

7.3.1 Selectivity

Selectivity curves are shown for the different fishery groupings in Figure 26. There are a variety of patterns in selectivity with logistic patterns for the longline index fisheries that asymptote at older ages, while the selectivity of LL.ALL.2 and LL.ALL.3 fisheries peaked around age-class 10 (i.e., 2.5 years). The large fish handline fishery in region 7 also showed asymptotic selectivity, with very few fish younger than about 10 quarters of age being estimated to be caught by this fishery. The PS, PL and miscellaneous small fish fisheries in region 7 displayed a variety of relatively complex selectivity functions (Figure 26). This is due to the same number of nodes restricted to a smaller range of ages due to a lack of fish greater than 90 caught by these fisheries. The selectivities of the domestic fisheries (ID, PH, and VN) were most selected for younger and smaller fish. The pole-and-line and purse seine fisheries generally show strongly dome shaped selectivities that peak at around 5 quarters, though the unassociated purse seine fisheries also catch older fish. The trend of associated purse-seines catching a greater proportion of smaller fish than the corresponding unassociated fisheries is generally maintained across regions.

7.3.2 Movement

Movement between the 9 regions is difficult to show in a clear manner. Therefore, we display the movement rates in two forms. The first plot shows the contributions of each region to total biomass within a given region, which is based on the average regional recruitment and the movement rates. This plot suggests that region 2 contributes a large portion of biomass to region 1. Region 4 contributes significant biomass to regions 2 and 6. Region 5 appears to be a source of biomass for regions 3, 8, and 9. The biomass in region 8 appears to come primarily from other regions, with

only about a quarter coming from within this region. Region 7 contributes nominal amounts of biomass to all regions. The alternative method to display the movement among regions is to show the quarterly movement rates between each region (Figure 28). Generally, most regions show a high rate of residency (the red diagonal through each grid). However, there is relatively large movement in quarters 1, 2, and 3 from region 8 into region 5 and slightly less movement from region 5 into region 8. About 20% of the fish in region 6 move into region 4 in the third quarter and a similar proportion move from region 2 into region 4. There is also a large amount of movement out of region 3 in the first quarter that goes into regions 4, 5, and 8. Almost 30% movement occurs from region 9 into region 5 in the fourth quarter.

7.3.3 Tag Reporting Rates

Tag reporting rates estimated by the diagnostic model for the various fishery groups are displayed in Figure 29. There were seven fishery groups/tag program combinations where the diagnostic model estimated tag reporting rates at the upper bounds: PS_1_RTTP, PL3_PL8_PTTP, LL_AU_RTTP_PTTP, PL3_PL8_RTTP, ID_PH_Dom_JPTP, PL7_RTTP, PL1_JPTP. Reporting rate parameters at the upper bounds are often indicative that the mortality rates are being underestimated by the model relative to what is suggested by the tagging data. While high reporting rates could be plausible in some fisheries where the incentives to report tags (i.e. monetary rewards) are meaningful and fishers are involved in the programs (i.e. perhaps PL8), reporting rate parameters at the upper bounds should be treated with caution.

7.3.4 Growth

The model fit and residuals to the conditional age-at-length used for the diagnostic model is shown in Figure 30 and the estimated curve and associated standard deviation at age are shown in Figure 31. The diagnostic model growth function estimated an L2 of 153 cm, L1 of 20 cm and k = 0.1.

7.4 Stock assessment results

7.4.1 Recruitment

As MFCL has the ability to use both movement and regional recruitment to distribute the population among regions to maximize the total objective function, cautious interpretation of the results of the estimated regional recruitment distribution is required.

The diagnostic model recruitment estimates for each region and the overall assessment region are shown in Figures 32 and 33. The overall trends in recruitment for the WCPO (Figure 32) are challenging to interpret due to the greater uncertainty in the historical estimates. Compared to the previous assessment, (see Fig. 30 Tremblay-Boyer et al., 2017) the recruitments estimated by

the 2020 diagnostic model are substantially higher (approximately double) across all regions, and region 4 is afforded much higher recruitment relative to the other regions compared to the 2017 diagnostic model. Quarterly recruitment as a proportion of the total annual recruitment across regions is highest within region 7, consistent with the previous assessment, followed by regions 3, 6, 4 and 5 (Figure 34). The model predicts there to be practically no recruitment occurring in region 8 (Figure 34). The regional estimates display notable inter-annual variability and variation on longer time scales (Figure 32). Recruitment is estimated to be very high in most regions up to 1965. However, the variability in recruitment estimates in these early periods is very high and is cut off in the plot shown. Uncertainty in the early recruitments underpinned the decision to assume the period 1968-2017 as the basis for the long-term average recruitment and estimation of the BH-SRR. Over this period, recruitment on average displays very little trend, and the uncertainty decreases substantially since the mid 1960's.

The estimate of the SRR is presented in Figure 35. Recruitment levels have declined over time as the stock size has declined, but have stabilized over the last 20 years and inter-annual variability has decreased.

7.4.2 Biomass

The pattern in spawning potential by region is presented in Figure 36 and Figure 33. Spawning potential is estimated to have declined steadily over the model period, with notable declines seen in all regions except regions 1, 2 and to a lesser extent region 9 (Figure 36). However, these regions account for smaller proportions of the overall WCPO biomass. There have been several short term increases in spawning potential at various times during the assessment period, but these were never sustained more than a few years. In the most recent years, spawning potential has stabilized, and shown a small increase in the mid-late 2010s. Consistent with the recruitment patterns, spawning potential at the start of the model is estimated to be substantially higher than that estimated for the period 1960 onward. This feature was also present in previous assessments and is seen in the bigeye assessment, which uses the same source of data to standardize CPUE (Ducharme-Barth et al., 2020b). This substantial decline occurs despite the low and relatively stable catch levels in the periods up to 1975 (Figure 3). Therefore the high early biomass and subsequent decline around 1965 is influenced by the uncertain and high recruitment estimates preceding 1965.

The distribution over all years of proportion of the overall number at age within a year are shown for each region and quarter as estimated by the diagnostic model in Figure 37. The largest proportion of annual recruitment over the years occurs in regions 3, 6, and 7. Age 1 fish are almost completely absent from region 8. At older ages the proportion of the population in region 8 increases, whereas the proportion in regions 6, 7, and 9 decrease with older ages. This matches with estimates of movement and recruitment distributions (Sections 7.3.2 and 7.4.1).

7.4.3 Fishing mortality

Average fishing mortality rates on both juvenile and adult age-classes have increased markedly starting in 1970. Fishing mortality for juveniles is estimated to have stabilized since around 2000, but there is notable variability since that time. Adult fishing mortality is estimated to have continued to increase and is at the highest level in the most recent years (Figure 38). Estimates are comparable for juvenile and adult age-classes throughout the model period, although the juvenile rates are generally slightly higher, except for the most recent years.

Changes in fishing mortality-at-age and population age structure are shown for decadal time intervals in Figure 39. Since the 1970's exploitation rates have increased on the youngest age classes through to the most recent decade. This is associated with increased catches by the purse-seine fishery in tropical regions and the Philippines, Indonesian and Vietnamese fisheries in region 7 (Figure 40). There is also a high exploitation rate on the older age classes (6 and older), which coincides with the peak in the selectivity of the unassociated purse-seine fisheries and longline selectivity. Overall, there has been a decline in the proportion of old (greater than age class 15) fish in the population since the 1970s (Figure 39). Fishing mortality-at-age estimates by region (Figure 40) show expected results, with temperate regions dominated by LL fishing showing highest fishing mortality for the older age classes. In the equatorial regions, where the PS and miscellaneous small fish fisheries occur, a higher fishing mortality occurs on the youngest age-classes.

7.5 Multi-model inference – sensitivity analyses and structural uncertainty

7.5.1 One-off changes from the structural uncertainty analysis

Comparisons of the spawning potential and depletion trajectories for the diagnostic model case and one-off sensitivity runs are provided in graphical form in Figures 41 to 44 and 47. The key reference points for the structural uncertainty grid are provided in Table 5.

Natural mortality [*M 0.11, M 0.13, M 0.15*]

Sensitivities to the assumed mean natural mortality resulted in large differences in the estimated spawning potential and depletions. As a result of the lower values of M, the spawning potential increased and the stock was estimated to be less depleted (Figure 41). As natural mortality declined the estimate of fishing mortality also declined (thus an overall decrease in total mortality). The lower values of natural mortality resulted in slightly less recruitment, but due to much greater survival in all age classes the spawning potential increased substantially. Though the decline in biomass over the time-series was greater in magnitude compared to the diagnostic model, the stock was estimated to be less depleted with lower values of mean natural mortality. The estimated spawning potential with the low values of natural mortality were higher than was considered to be biologically reasonable for yellowfin tuna. Additionally, these models fit the data worse than the

diagnostic model by between 500 and 2000 points different in the likelihood. While these results are being investigated further, we therefore only used the value assumed from previous assessments in the structural uncertainty grid (base level of 0.2 per quarter). Further work to estimate natural mortality for available tagging data is recommended.

Alternative Size Compositions [*WtAll, WtJP, WtUSAU, WtLen*]

Use of alternative size compositions data in the index fishery and the LL extraction fisheries had little impact on spawning potential or biomass relative to the diagnostic model (Figure 42). This impact was lower for the diagnostic model compared to initial model testing of these different data sources (not shown). The initial tests of these data sources occurred before the implementation of selectivity changes and changes to the tagging data. Compared to the diagnostic model the use of only the Japanese data in the index fishery was nearly identical in terms of depletion and spawning potential to the diagnostic model. The addition of weight composition from the offshore fisheries in the index fishery resulted in the most increase in spawning potential compared to the diagnostic model in this set of sensitivities. In general, these different data sources made little impact on the overall depletion or spawning potential.

Alternative standardized CPUE indices [*HLPSPHPS8, PSCPUE*]

The inclusion of indices for the Indonesia & Philippines handline, Indonesia & Philippines purse seine, and the associated purse seine in region 8 resulted in similar estimates of spawning potential compared to the diagnostic model. There was a slightly higher biomass estimated in this model (HLPSPHPS8) compared to the diagnostic model. The estimated depletion for the majority of the time series resulted in a marginally more depleted stock, but during the last few years had identical estimates compared to the diagnostic model.

The model that incorporated a standardized index for the associated purse seine fisheries in region 3, 4, and 8 estimated spawning potential that was very similar to the diagnostic model, with only a few deviations in the 2000s (PSCPUE, Figure 43). The estimated depletion for this model was slightly lower compared to the diagnostic for the entire time series. Both of these sensitivity models retained the index fisheries and CPUE from the geostats model used in the diagnostic model, but incorporated additional indices for the respective fisheries. The impact of these additional CPUEs had was limited due to the regional scaling imposed by the index fishery framework.

Model beginning in 1962 [*Start1962*]

A model that started in 1962 to test the impact of inclusion of the initial high recruitments resulted in a slightly larger estimated spawning potential compared to the diagnostic model during the 1970s and 1980s (Figure 43). However, by the end of the time series the spawning potential estimates were nearly identical. The depletion status of the Start1962 model generally tracked the same trends as the diagnostic model, but by the end of the time series was in a slightly more depleted state than the diagnostic model.

Alternative growth [*Tag-Oto, Modal, Otolith*]

The growth curve used in MULTIFAN-CL assessments is a very important factor in the estimated stock status. During initial model development the tag-otolith integrated model had very similar estimated depletion compared to the otolith only Richards growth curve. Therefore, it was decided to use three growth curves obtained in the following manner:

1. estimated using the conditional age-at-length,
2. estimated from modal progressions in a model that did not down-weight the size composition data, but then fixed in subsequent models, and
3. the external estimate from the otolith-only Richards growth curve.

Subsequent to that decision, additional changes to the model were made, including the discovery of the model misspecification of the tag returns by the pole-and-line fisheries in multiple regions, fixing the selectivity for the purse seine and pole-and-line fisheries in the first age class to zero and down-weighting the size composition data. These changes had unexpected consequences on the *Tag-Oto* model where the estimated spawning potential was much lower than the diagnostic model and the depletion was similar to the model that used growth estimated from the modal progression (Figure 44). The otolith only growth curve estimated marginally higher spawning potential in the last few years of the assessment compared to the diagnostic model and a slightly less depleted state in these years. The model with growth estimated from modal progression resulted in a similar spawning potential but a more depleted state compared to the diagnostic model.

A comparison of the mean lengths from the three growth curves used in the structural uncertainty grid was made with selected size composition from the longline US fishery in region 2 for the period 2004 to 2005, and the associated purse seine between 1998 and 1999 (Figures 45 and 46). The vertical lines in the plots show the mean length at age for the selected growth curve. Each panel shows the size composition for sequential quarters in the selected time period. Therefore, it is expected that a mode in the size composition that matched with one of the lines (mean length at age) would have a mode in the panel below centered at the line to the right of it. From the weight composition data, it is apparent that there are larger growth increments for age classes 6-10 in the *Modal* progression growth curve compared to the growth curves estimated by the *Otolith* and *CondAge* models (Figure 46). Conversely, the length composition plots show that the *Modal* progression growth estimates a near cessation in growth between the third and fourth quarters (Figure 45). This cessation in growth makes the fish vulnerable to the purse seine fishery for one additional quarter for the *Modal* growth compared to the *Otolith* growth curve and the *CondAge* growth curve.

Model with tag mixing period of 1 Quarter [*Mix1*]

The sensitivity of the model to the assumed mixing period of the tagging data was assessed by setting the mixing period to 1. The resulting model estimated less spawning potential compared

to the diagnostic model (Figure 44) and depletion for this model was estimated to be lower.

Steepness [*H0.65, H0.95*]

The low penalties on the SRR relationship resulted in the assumed value of steepness having no impact on model fit and time-series estimates of spawning potential (Figure 47). The lines in the plot for the diagnostic model are obscured by the lines for *H0.95* and *H0.65*. The estimates of depletion changed as a result of the different steepness values. The stock was estimated to be less depleted with a steepness of 0.95 and more depleted with a steepness of 0.65 (Figure 47).

Weighting of size frequency data [*Size20, Size200, Size500, SSMULT*]

The scalar assumed for the size composition data resulted in various changes to the estimated trends in spawning potential (Figure 47). A smaller scalar relative to the diagnostic model (20) gave more influence to the size composition data resulted in a larger spawning potential for the entire time series. The trajectory of the depletion for this model was similar to the diagnostic model early in the time series, was less depleted during the late 2000s, but had nearly identical depletion in the terminal year. The model with a scalar of 200 estimated lower spawning potential for the time series and depletion that was marginally lower compared to the diagnostic model. The model with a scalar of 500 resulted in spawning potential that was similar for the majority of the time series. However, the estimated depletion of this model was more optimistic compared to the diagnostic model. A model that used the self-scaling multinomial distribution (*SSMULT*) to model the size composition data resulted in the largest change in spawning potential and depletion. The model predicted a lower spawning potential and a much more pessimistic depletion by approximately 15 percentage points compared to the diagnostic model.

7.5.2 Structural uncertainty analysis

Axes for the structural uncertainty grid were selected partly by considering the of axes of uncertainty from the previous yellowfin tuna stock assessment (Tremblay-Boyer et al., 2017), but primarily from the recommendations made at the PAW (Hamer and Pilling, 2020). Consequently, the values for the axes of uncertainty for size composition scalars, steepness, tag mixing, and growth were chosen to encompass the greatest source of plausible uncertainty. The results of the structural uncertainty analysis are summarized in several forms – estimates of specific parameters across models in the grid (Figures 48 to 50), violin-plots of $F_{recent}/F_{\text{MSY}}$ and $SB_{recent}/SB_{F=0}$ for the different levels of each of the four axes of uncertainty (Figures 51 and 52), time-series plots of fisheries depletion for all models in the grid (Figures 53 and 54), Majuro plots showing the estimates of $F_{recent}/F_{\text{MSY}}$ and $SB_{recent}/SB_{F=0}$ across all models in the grid (Figures 55 and 56), and averages and quantiles across the full grid of 72 models for all the reference points and other quantities of interest (Table 5).

Many of the results of the structural uncertainty analysis show consistent trends with previous assessments of tuna stocks in the WCPO for similar axes of uncertainty. However, the estimated

status from the current structural uncertainty grid is more optimistic than previous assessments of yellowfin tuna.

Parameter estimates from across the grid

Recent assessments have progressed toward the use of a structural uncertainty grid to provide management advice. However, estimates from the grid are only shown in terms of management quantities of interest. At SC15 there was a request to provide additional diagnostics from across the uncertainty grid. In a step toward this, we provide the estimated growth curves for the conditional age-at-length models in the structural uncertainty grid (Figure 48a). Estimates of growth in the structural uncertainty grid are generally consistent with the external otolith growth curve but show variability in the L2 parameter. Estimates of L1 from the conditional age-at-length were generally on the lower bound of 20 cm. As a result of the use of the maturity-at-length feature in MULTIFAN-CL and the estimated growth curves, a different maturity-at-age for each of the conditional age-at-length models is estimated (Figure 48b). The maturity-at-age ogives for the conditional age-at-length models are generally between those of the modal progression and the otolith models, but have a slightly larger decline at older ages.

Reporting rate parameter estimates from models within the structural uncertainty grid are presented in Figure 49. Reporting rates that were on the upper bound in the diagnostic model were estimated on the upper bound for all models in the structural uncertainty grid. Variability in reporting rate estimates from the structural uncertainty grid varied notably between groups.

Movement rates from different models across the structural uncertainty grid are presented in Figure 50. The movement estimates across the structural uncertainty grid were affected by changes in the assumed growth and the weighting of the size composition. Movement rates were generally the largest for models that assumed the modal growth curve and/or a size composition scaler of 20. The lowest movement rates were estimated by a model where the size composition were down-weighted by a scalar of 500.

Structural uncertainty results

The general features of the structural uncertainty grid are as follows:

- Estimates of stock status, reference points, and trends in abundance were more optimistic compared to the previous assessment. The terminal depletion estimated for all models was above the $20\% SB_{F=0}$, with the range of the grid of $SB_{recent}/SB_{F=0}$ between 0.51 and 0.64.
- The most influential axes of uncertainty with respect to estimated stock status was growth. The most pessimistic model estimates occurred with models that assumed growth estimated from the modal progression information in the size composition data. The most optimistic stock status estimates were obtained from models that used the growth curve estimated externally from otolith data. Models where growth was estimated by the conditional age-at-

length resulted in estimates that were in between the other two, but were more consistent with the otolith growth curve models.

- The next most influential axis in the grid was the assumption of the tag mixing period. Models with a tag mixing period of one quarter resulted in more pessimistic stock status compared to a mixing period of two quarters.
- The steepness axis displayed results consistent with previous structural uncertainty grids. Models with steepness of 0.95 were the more optimistic compared to the steepness of 0.8 assumed in the diagnostic model, while a steepness of 0.65 was the most pessimistic. The lower the steepness the more depleted the stock and the higher the fishing mortality with respect to F_{MSY}.
- The influence of the different levels of the scalar for the size composition were more complicated. The estimates of depletion and fishing mortality for a scalar of 20 or 60 were relatively consistent with one another. With a size composition scalar of 200 there was a slight skew in the distribution toward more pessimistic status. The distribution of models with a scalar of 500 had both the most and least depleted status in the structural uncertainty grid.
- Dynamic depletion across the grid was estimated to be more optimistic in most time periods since the last assessment, but showed a more pessimistic trend in the last quarters of 2018.
- Equatorial regions (3, 4, 7, and 8) are the most depleted but are above a regional 20%$SB_{F=0}$ across all models.
- There is evidence to suggest that the overall stock status is "buffered" or kept optimistic by low exploitation in the temperate regions (1, 2, 6 and 9).

7.5.3 Further analyses of stock status

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 use of multi-model inference, defining management reference points based on the ensemble of grid runs, complicates the presentation of these results for each individual model. These analyses are presented with respect to the diagnostic case.

Fishery impacts by region

Fishery impact is measured at each time step as the ratio of the estimated spawning potential relative to the spawning potential that is estimated to have occurred in the historical absence of fishing. This is a useful quantity to monitor, as it can be computed both at the region level, and for the WCPO as a whole. This information is plotted in two ways, firstly as the fished and

unfished spawning potential trajectories (Figure 57), and secondly as one minus the depletion ratios (Figure 58). The latter is relevant for the agreed reference points and example plots of these values are displayed for the diagnostic case model.

Significant declines in spawning potential were estimated for each region (Figure 57), with a general pattern of steady declines over the majority of the time series and moderate increases over the most recent 5-10 years, in most regions. The overall pattern of declines was not only estimated to occur for the fished stock, but also the unfished stock in all regions. This indicates that much of the decline in spawning potential can be attributed to other factors affecting the stock, such as changes in recruitment levels.

Fishery impact by gear and across regions

The comparison of fished and unfished stock trajectories can be extended to examine the impacts of individual gear types, within regions, and the results are displayed in Figure 58. Fishing impacts were estimated to be very minor in all regions before about 1970, when catches started to increase significantly. Subsequent to this, the estimated fisheries impacts differ considerably among regions. Fisheries impacts in several temperate regions (2, 6 and 9) were estimated to be very limited at well under 20%. In the remaining temperate regions (1 and 5) moderate (20–40%) fishing impacts were estimated, and the most influential gears in both cases were the associated and free-school purse seine fisheries. Although catches by these gears in these regions are low to moderate, the catches in adjoining regions influence biomass among regions via movement across region borders.

The equatorial regions (3, 4, 7 and 8) are more impacted by fishing than the temperate regions. In regions 3, 4 and 8 similar results were obtained with fishing impacts of over 40% estimated. The most influential fisheries in these declines were the purse seine associated and free-school fisheries, with both displaying relatively similar impacts, and to a much lesser extent the longline and 'other' fisheries. The western equatorial region (7) showed slightly different dynamics with a substantial impact of fishing (>50%), although much of this could be attributed to the 'other' fisheries occurring in region 7 rather than the purse seine fisheries which are more influential to the east.

When fishing impacts are considered at the regional level (bottom right panel of Figure 58), it is evident that the whole range of major gears being fished in the WCPO are having some impact on the abundance of yellowfin. Purse seine fishing is estimated to be having the largest impact although the relative influence between associate and free school fishing has shifted from associated sets having the most impact during the period ∼1990–2010 to the present situation where free school sets are impacting spawning biomass more. Since about 2000, the 'other' fisheries in the Philippines and Indonesia are estimated to be having a significant influence on abundance, almost at the level of the free school purse seine fishery. The impact of the longline and pole and line fisheries has been less severe and remained relatively stable for the last two decades.

Yield analysis and equilibrium estimates across the grid

The yield analyses conducted in this assessment incorporates the spawner recruitment relationship (Figure 35) into the equilibrium biomass and yield computations. Importantly in the diagnostic case model the steepness of the SRR was fixed at 0.8 so only the scaling parameter was estimated. Other models in the one-off sensitivity analyses and structural uncertainty analyses assumed steepness values of 0.65 and 0.95. Across the structural uncertainty grid the equilibrium virgin spawning potential in the absence of fishing ($SB_{F=0}$) was estimated to be between 3,231,353 and 4,050,429 mt (Median = 3,603,980 mt; Table 5), and the spawning potential that would support the MSY (SB_{MSY}) was estimated to be between 590,090 and 1,114,400 mt. The ratio of SB_{MSY} to $SB_{F=0}$ was estimated to be between 0.18 and 0.28 (median = 0.24). Levels relative to the $20\%SB_{F=0}$ and F_{MSY} have been discussed above (see Section 7.5.2).

A plot of the yield distribution under different values of fishing effort relative to the current effort are shown in Figure 59 for one-off sensitivity models. The current effort levels are estimated to be well below the levels supporting MSY for all models. The fishing mortality multiplier necessary to achieve the maximum yield was relatively consistent across each of the sensitivity models at between about 2 and 3.5. The MSY values for this suite of models varied between about 950,000 and 1,300,000mt.

The yield analysis also enables an assessment of the MSY level that would be theoretically achievable under the different patterns of age-specific fishing mortality observed through the history of the fishery. We present a plot for the diagnostic case model in Figure 60. Prior to 1970, the WCPO yellowfin fishery was almost exclusively conducted using LL gear, with a low exploitation of small yellowfin. Fisheries other than longline were known to operate in the region before 1970, but no catch estimates are available. The associated age-specific selectivity pattern resulted in a much higher MSY in the early period compared to the recent estimates. A pronounced decline occurred after the expansion of the small-fish fisheries in region 7 and, soon after, the rapid expansion of the PS fishery which shifted the age composition of the catch dramatically toward much younger fish.

Dynamic Majuro plots and comparisons with Limit Reference Points:

The section summarizing the structural uncertainty grid (Section 7.5.2) presents terminal estimates of stock status in the form of Majuro plots. Further analyses can estimate the time-series of stock status in the form of Majuro plots, the methods of which are presented in Section 5.8.4. The large number of model runs in the structural uncertainty grid precludes undertaking and presenting this process for all runs, however an example for the diagnostic case model is shown in Figure 61. These produce intuitive results with respect to the terminal results already presented in previous sections. The stock at the start of the assessment period was estimated to be close to an $SB/SB_{F=0}$ of one and an F/F_{MSY} approaching zero, but it progressively tracked toward the overfishing and overfished definitions over the remaining period. The diagnostic case model never reaches a point close to $20\%SB_{F=0}$ or an F/F_{MSY} of 1, and the status of the stock improves slightly in recent years.

8 Discussion and conclusions

8.1 Changes to previous assessment

There were numerous changes made to the assessment model in 2020 compared to the 2017 assessment, including the incorporation of three years of additional tagging, catch, effort and size composition data. Key changes made to this assessment related to the implementation of an index fishery approach that received the standardized CPUE indices from the spatial-temporal (geostats) model, and changes to the preparation of the size composition and tagging data. These changes had moderate but varied effects on the estimated status of the stock.

The largest changes in estimated stock status occurred due to enforcement of a 182 day mixing period for the tagging data and model modifications related to the size composition, e.g. selectivity assumptions, data weighting and growth inputs. Changes made to the tagging data resulted in less information available to estimate the fishing mortality rate as a result of the change in the attrition curve. This change was hidden in initial tests of these data inputs due to the model misspecification where the tag recaptures from the pole-and-line fisheries in regions 3 and 8 were grouped.

Following the outputs of Project 82, age-at-length data were available to the yellowfin stock assessment for the first time. The move from growth estimated through the modal progression in the size composition with deviates from the growth curve in the first 8 quarters (as used in the 2017 assessment), to the incorporation of conditional age-at length data led to more optimistic stock status. As a result of less information on fishing mortality due to the changes made to the tagging data, the model estimated a larger biomass in order to fit the length composition data better. When the growth curve is changed, fish are vulnerable to the purse seine fishery for one less quarter. Assuming that fishing mortality remains relatively consistent despite a change in the growth curve, the model estimates an increase in the population size to fit the catch data from the purse seine fisheries.

The model interacts with the size composition data through the estimation of selectivity and changes to assumptions regarding these parameters have a large impact on the estimated stock status. The model was found to estimate high recruitment (20 cm fish) in the temperate regions (e.g., 5) and to move fish into the equatorial regions (e.g., 8). This is thought to be unlikely given that most tagging studies, otolith chemistry, and parasite assemblages from yellowfin tuna indicate that movement is relatively limited for small fish, though our understanding of movement dynamics of large fish is limited (Moore et al., 2020). The model also predicted selectivities for purse seine and pole-and-line gears that implied they were able to catch a large amount of very small fish (less than 30cm) that were not observed in the size composition data for these fisheries. In an attempt to remedy these model issues, we fixed selectivity for these fisheries in the first age class to be zero. This resulted in changes to the estimated movements and a slight negative change in overall biomass and depletion estimates in the stepwise model progression. However, there was a marked increase in spawning

potential and depletion when this change was implemented for models where conditional age-at-length data were not included. Therefore, there are often complex and unexpected interactions between changes made to the assessment models, which are difficult to explore and understand due to the long run time and complexity of the current assessment models.

8.2 Sources of uncertainty

8.2.1 Growth

Growth is an important model component in an integrated assessment model. For example, it defines how many quarters a fish remains within a size range that is vulnerable to a dome shaped selectivity pattern (Figure 45). It is also important in determining the number of age classes of natural mortality fish experience before they become vulnerable to a gear that catches primarily older fish, e.g. longline (Figure 46). Therefore, correctly estimating the growth of a species is key for a reliable length-based stock assessment. Previous assessments of yellowfin tuna in the WCPO have relied on modal progressions in the size composition data to estimate growth parameters. However, those estimates can be confounded with the selectivity parameters in the model and thus are difficult to estimate, as indicated by the jitter analysis (Section A.3).

The current assessment incorporates the newly available otolith increment data (Project 82), through conditional age-at-length and otolith only growth curves. The resulting differences in estimated stock status appear to be driven by the large difference in the estimates of the growth parameter k and the exclusion of deviates for the first 8 age classes. Comparing the growth estimates from the modal progression data indicate that the mean lengths of fish aged 3 and 4 quarters are approximately 47 cm and 51 cm, respectively. In contrast, the otolith only growth curve puts the size for these age classes at 48 cm and 59 cm (Eveson et al., 2020). The models that used the growth estimated from the modal progression data thus have an additional quarter during which they are within the most selected size range for the associated purse seine fisheries. Models that use the otolith growth curve or conditional age at length data account for differences in the growth curve by increasing the recruitment and thus ultimately the biomass of the model to fit the catch. It is currently unclear whether the "two phase" growth seen in the modal progression data is a confounding of the estimation of the highly dome shaped selectivity in the purse seine fisheries and the deviations from the growth curve or a biological phenomenon. However, none of the models in the structural uncertainty grid are able to fit the high peak seen in the size composition data around 45–50 cm for the purse seine fisheries. Therefore, these models could be underestimating the fishing mortality on these small fish, which would have large implications on the depletion estimate. The overestimation of older larger fish in these fisheries is additionally concerning because these fish weigh more and could allow the model to fit the catch by spreading out the fishing mortality over a broader range of age classes than may actually occur. Further investigation of alternative selectivity assumptions within the stock assessment model on a fishery specific basis is warranted.

Changes in growth resulted in differing movement patterns estimated by the model (Figure 50). Changes to estimated movement patterns also occurred under differing assumptions of the size composition scalar values. Why changes to the size composition, growth, and selectivity have such a large impact on the movement rates and estimated biomass is currently not fully understood in this assessment model and requires additional investigation.

The diagnostic model incorporated the conditional age-at-length data with the aim of balancing the information in the modal progressions of the size composition data and the otolith length-at-age information. However, the likelihood profile of total biomass indicates that the conditional age-at-length data is strongly influencing the estimated scale of the population. We also note that within the uncertainty grid, the growth curves from conditional age-at-length models are very similar to models that use the growth curve estimated externally from the otoliths only, and hence the structural uncertainty grid may be skewed by double counting the influence of the otolith data. Despite the addition of the otolith data to the assessment, uncertainty surrounding the growth remains due to the conflict between the otolith data, the tag increment data, and the size composition data. Therefore, there is a need for validation studies of growth estimates from otolith aging; for example, bomb radiocarbon approaches or strontium chloride tagging, but preferably both, are recommended to validate the growth curves.

There is some indication that regional patterns in yellowfin growth may exist. For example, studies in the WCPO indicate that the length at 50% maturity may differ spatially, which could coincide with differences in growth (Moore et al., 2020). The size composition data used within the assessment indicates that the maximum weight or length reached by yellowfin tuna may differ among regions, as did analysis of otolith data to some extent (Farley et al., 2020). Within the weight composition data, the maximum size of fish generally seems to be largest in the temperate regions and smallest in regions 3 and 7, a pattern seen across the time series of available data. The model thus compensates for this by indicating that the fishing mortality in these regions must not be high to allow these large fish to survive at the frequency seen in comparison to the equatorial regions. Biomass in the temperate regions (particularly regions 5 and 6) is therefore estimated to be high and relatively undepleted. Growth studies should be conducted in a manner that would allow for verification of the presence (or absence) of regional differences in growth (see below Section 8.4).

8.2.2 Natural Mortality

Natural mortality within a stock assessment model scales the population relative to the catch within an assessment and determines the overall biomass of the population. The current assessment used the same base level of natural mortality as in the 2017 assessment but the shape of the curve varied depending on the estimated growth curve. We originally intended to include natural mortality as an axis of uncertainty in the grid, but the model estimates of biomass were unreasonably high from the candidate levels. When natural mortality was lowered the stock assessment model estimated

a lower fishing mortality as well. This may indicate that the total fishing mortality rate is poorly estimated within this model and thus there may be unaccounted for uncertainty in the estimates of depletion. Additional investigation of natural mortality estimation using the tag data external to the MULTIFAN-CL stock assessment model is recommended (see below Section 8.4).

8.2.3 CPUE

Within a stock assessment model it is generally believed that the CPUE should be the primary driving force behind the trends in biomass and depletion. Trends of high recruitment early in the time series of this assessment appear to be due to a large drop in the CPUE, but not a noticeable increase in the catch. Therefore, the model is able to fit this trend by having recruitment be very high during the 1950s and 1960s without a change in depletion. However, starting the model in 1962, after this initial phase, resulted in a similar estimate of depletion compared to inclusion of the entire time series (Figure 43).

The regional scaling from the spatial-temporal model (geostats) can be very important in estimating the regional populations in multi-region models (see McKechnie et al., 2017a; Tremblay-Boyer et al., 2017). The difference between the two regional structures in the previous assessments appear to be primarily due to the difference in the regional scaling. The temperate regions are generally estimated to have a low exploitation rate due to a smaller catch compared to the equatorial regions. However, the biomass in these regions is often estimated to be at the same scale as the equatorial regions due to the regional scaling of the CPUE indices. This high biomass in temperate regions can thus lead to an optimistic estimate of depletion. Determination of the biomass in these regions and appropriate regional scaling from a geostats model is difficult due to a lack of contrast in the CPUE data and fewer fishing observations compared to the equatorial regions. Uncertainty surrounding the regional scaling estimates from various geostatistical models is needed in future assessments.

Investigation of alternative CPUE indices in addition to the spatial-temporal index had relatively little effect on the estimated depletion and biomass. A standardized index for the associated purse seine was intended to act as an indicator of recruitment, but this fishery also catches large sizes and does not catch the smallest fish observed in the assessment (20 cm). Thus this index did not influence recruitment estimates as was expected (Vidal and Hamer, 2020). This index will require further refinement to explore the possibilities of hyper-stability in the index and to correct for this if possible. The lack of influence of these additional indices on the overall estimated depletion was likely a result of maintaining the regional scaling estimated by the geostats index fishery framework.

All indices used in this assessment are inherently fishery-dependent and thus need to be corrected for changes in efficiency and fishing approach (day or night, hook depth, bait type, etc.). If these are done incorrectly or the model is unable to adequately correct the changes in CPUE due to a lack of information, then the resulting estimated trend in CPUE will be biased, leading to biases in the estimated biomass and thus management advice. Additional challenges with CPUE can

occur if there is no contrast in the data due to a continually increasing fishing effort (the one-way trip). Stock assessments for many fisheries have avoided potential issues in standardizing fishery dependent catch rates by conducting a fishery independent survey that uses a standardized method, which is kept constant over time. This allows changes in the observed catch rate to be attributed to changes in the biomass, but only if the method of catching fish is kept constant over time. In order for a fishery independent survey to be representative of the population it must cover the entire geographic extent of the population and have numerous samples to reduce variability in estimates. A fishery independent survey would also be useful in the collection of biological information that is relevant to all species caught in the WCPO. If designed appropriately, a single fishery independent survey could be used as an index of abundance for all four main tuna species in the WCPO (yellowfin, bigeye, skipjack, and albacore). Design of such an inherently challenging program would require coordination between stock assessment scientists, statisticians, fishermen, boat captains, and biologists.

8.2.4 Conflict among data sources

This assessment was fraught with strife due to conflict among data inputs to this assessment. As noted, a key area of uncertainty, and hence axis of the uncertainty grid, was growth of yellowfin tuna. The tag increment growth data and the otolith data showed different trends, which prevented a Richards growth model from being fit (Eveson et al., 2020). The externally estimated growth curve was inconsistent with modal progressions in the size composition data and resulted in very different estimated status from these models (Figures 45 and 46). Conflict between data sources also was evident in the likelihood profile (Figure A2). The conditional age-at-length data and the length data had the best fit to the data when estimating a higher biomass, whereas the weight composition, CPUE, and tagging data were better fit to a lower total biomass. However, the latter data components had a relatively flat likelihood and the conditional age-at-length data tended to drive the estimation of total biomass. The 'jumps' seen in the likelihood profile may also be indicative of some other conflict within the assessment model that has not been identified, potentially a conflict between the estimation of growth and the selectivity parameters that may be highly correlated with one another (Cass-Calay et al., 2014). The effective sample size of the multinomial distribution for the conditional age-at-length data can be adjusted to allow a difference in the weight given to this likelihood component of an assessment model. For this assessment we assumed an effective sample size equal to the number of otolith samples. Further investigation of alternative weightings for this data source are recommended for models that use conditional age-at-length data. A model that used the *SSMULT* likelihood resulted in very different trends in spawning potential and depletion compared to any of the robustified normal distribution scalars chosen. This is likely do to the flexibility of this likelihood to estimate differing influences between the various assumed groupings of length and weight composition data. This new feature in MULTIFAN-CL is promising, but requires additional investigation before inclusion in stock assessments.

An additional conflict evident in the previous assessment (Tremblay-Boyer et al., 2017) that we were unable to resolve was the conflict between the CPUE data and the tagging data. The reporting rate on the upper bound for the Australian LL fishery effectively lowers the fishing mortality estimate for these fisheries and allows for a large biomass in region 5. Though the reporting rate estimate for the purse seine fishery in region 8 are not estimated on the upper bound in the diagnostic model, the much higher reporting rate compared to the prior could indicate that the fishing mortality informed by the CPUE data is less than the fishing mortality that would best fit the tagging data in this region. Similarly, the reporting rate of the pole-and-line fishery in regions 8 and 7 were estimated at the upper bound for the RTTP and PTTP, which may limit the fishing mortality in these regions, if to a lesser extent because the reporting rate for these fisheries may truly be high.

8.3 Further information for consideration

Although the structural uncertainty grid presents a relatively positive indication of stock status, there is reason for caution. The stock status appears influenced by model estimates of high biomass with low exploitation, and relatively high recruitment levels in the temperate regions (in the diagnostic case model). The model is able to estimate high recruitment in these temperate regions because there is no information provided to the model for juvenile fish (i.e., only longline fisheries in these regions are modeled that catch big fish). The recruits are then redistributed into other more highly exploited tropical regions as they grow older, allowing the model to fit the size composition the best way possible (Figure 37). The high estimated movement rates particularly between regions 5 and 8 may not be consistent with what is known about the limited movement of juvenile yellowfin tuna (Moore et al., 2020). There are two hypotheses for the current model behavior. The first hypothesis is that either a lack of contrast in the CPUE or low sample sizes of fishing incidences for the temperate regions estimates the region scaling to have a high biomass in these regions. The alternative hypothesis is that spatial differences in growth is currently mis-specified and the models are attempting to compensate through the resulting movement and abundance patterns.

We note that there has been a marked decline in the CPUE especially in the equatorial regions, which is considered to be the core area of yellowfin tuna (Figure 8). Similarly, the size composition data from the purse seine fisheries Figure 62 shows that in the late 1990s and early 2000s there was a wide range of sizes, and recruitment events could be followed through time (modal progressions). However, since 2010 the prevalence of fish greater than 75 cm observed in the size composition data for the associated purse seine has reduced. In turn, there is a greater proportion of fish at sizes less than 50 cm observed in the size composition data for this fishery. This may indicate the impact of fishing pressure on these smaller fish, with few fish seen in large proportions greater than 75 cm due to their absence from the population. Alternatively, this could indicate a change in selectivity that is currently unaccounted for by the assessment. The estimated stock status appears to be at odds with these observed trends, despite both these data sources being included in the assessment. These discrepancies could be due to estimates of regional scaling from the geostatistical

CPUE standardization model being poorly estimated or the under-estimation of the contribution of smaller fish to the catch in the PS size compositions, and overestimating the contribution of larger fish (Figures 18 and 19). Both of these factors could cause an underestimation of the fishing mortality and thus a more optimistic stock status.

There is some evidence to suggest that the current structure of the yellowfin tuna stock assessment is overly-complex and over-parameterized. A number of MULTIFAN-CL developments are in active development to help address the issue of complexity. The bigeye report (Ducharme-Barth et al., 2020a) provides more detail on this issue and we will not repeat that here, apart from noting that similar considerations apply to yellowfin tuna.

8.4 Recommendations for further work

A key issue identified in this, and other WCPO tuna stock assessments, is the need to better understand the biology of the species (Minte-Vera et al., 2017). This includes growth, reproductive biology, and natural mortality.

- **Growth** Given the influence of growth on the current assessment result, and the uncertainty that remains, further collection and analysis of biological samples to enhance the growth curve is needed as a priority, and should include validation studies of growth estimates from otolith aging. Focus on collecting samples from the largest fish and from very small fish (less than 30 cm) in all regions is needed to accurately estimate growth at the assessment region level. In turn, investigation of the conflict between the otolith data, the tag increment data, and the size composition data is warranted. Potential strategies to collect biological information from the purse seine fisheries in a spatially structured manner is outlined in Macdonald et al. (2020). A similar approach would also need to be applied to longline vessels in order to provide valuable information on the oldest and largest fish. However, there would still remain a lack information for the very small fish (less than 20 cm) for each region that is required for estimation of a growth curve. This would require directed fishery independent sampling to collect small fish in all regions (not just region 7).

- **Spawning Potential** The spawning potential ogive used in stock assessments is an important, but often overlooked component of determining stock status for management advice. However, the components of the spawning potential ogive used in this assessment have not been updated since 2008 due to a lack of new data. The spatial representativeness of this relationship is also uncertain. Additional collection of biological information for yellowfin tuna is therefore needed, with particular improvement in our understanding of the spawning fraction and location of spawning within the WCPO needed (see Vincent and Ducharme-Barth, 2020 for further discussion).

- **Natural Mortality** As noted, the value of natural mortality assumed had significant and

unexpected impacts on estimated stock status. In order to accurately estimate natural mortality, a good understanding of the reporting rate and tag mortality are needed. Tag seeding experiments show that there is a potential for a decrease in the tag reporting rates in recent years (Peatman, 2020). Unfortunately, the number of tag seeding experiments have also decreased in recent years and thus estimates of these reporting rates are highly uncertain. A greater focus on tag seeding experiments across the entire fleet is needed in order provide accurate information to the assessments.

- **Conversion factors** The sample sizes of assumed biological relationships such as gilled gutted to whole weight conversion are currently low. Further efforts to improve these biological relationships, including through Project 90, are encouraged.
- **Spatial patterns in biology and population structuring** The MULTIFAN-CL model must currently assume a common biological function across the spatial domain of the assessment, although future developments will relax this assumption. The design of any study to improve our biological understanding of yellowfin tuna should include investigation of potential spatial differences. For example, sufficient data to discern regional growth differences are currently lacking, and sample sizes from the current otolith study are insufficient to accurately estimate growth for all regions. Continued collection and aging of yellowfin tuna otoliths needs to be conducted across the entire WCPO. This may required direct collection of samples that do not come from the fisheries in some regions. This will be important in determining whether there are sub-populations within the WCPO stock, which is currently poorly understood (Moore et al., 2020). If regional differences in growth or sub-populations are found to exist, these dynamics could possibly be incorporated into future assessment models, or separate assessment models could be conducted.

Further work is also recommended in improving the CPUE information available to the assessments:

- **Regional scaling** The potential impact of regional scaling assumptions on model estimates has been noted within this report. Further analysis of the impact of alternative regional scaling inputs estimated from alternative CPUE models is recommended.
- **Investigation of covariates** Noting that the availability of covariates varies within the available time series of CPUE data, the investigation of alternative separations of that CPUE time series to incorporate available covariates for standardization is warranted. Investigation of the observed increase in CPUE for regions 1 and 2 in the past decade is needed to determine if this is a signal from the population or due to an increase in efficiency.
- **Environmental covariates** Given that regional oceanography and spatial patterns of oceanography within specific model regions can strongly influence achieved CPUE, the identification and testing of the inclusion of environmental covariates within the CPUE standardization for yellowfin tuna is recommended. Further investigation of these factors with attention to the effect on the regional scaling is needed.

Finally, further investigation of key model settings and inputs, and MULTIFAN-CL developments is recommended:

- **Tag mixing period** The sliding window approach to correcting the tag data for mixing period was implemented within the current assessment. Based upon this experience, further enhancements to the approach can be examined, including alternative corrections for mixing period that maintain the attrition rate for data not adjusted.
- **SSMULT** The SSMULT was trailed as a one-off sensitivity model within the current assessment, and showed considerable promise as an approach to independently weighting size composition data. Indeed, the size composition scalar of 60 chosen for the diagnostic model was based on attempting to replicate the likelihood fits of the other data sources from a model that used the SSMULT feature. However, the SSMULT was not ultimately used in these assessments due to a lack of complete understanding of the methodology, insufficient testing of the feature and computation limitations - the SSMULT models take a very long time to run and requires large amounts of memory, and thus requires additional computational resources. In order to adequately explore models (e.g., jittering analyses) and run the structural uncertainty grid with these computationally intensive models, significant additional computational resources will be required. Additional investigation and optimization of the SSMULT feature is therefore warranted.
- **Spatial functionality** The potential for spatial patterns in biological stock characteristics has been suggested within this report. Approaches to model regional differences in growth within the stock assessment model are encouraged. This may require additional development of MULTIFAN-CL or use of the multi-species feature.
- **Selectivity** The impact of selectivity assumptions on estimated stock status was demonstrated within the stepwise model development. A thorough investigation of the appropriate number of nodes in the selectivity curves for each fishery and their interaction with the growth curve (both assumed and estimated from the modal progression) is warranted. The development of length based selectivity curves in MULTIFAN-CL should also be investigated. Finally, the consideration of time-block selectivities for some fisheries, based on changes in fishing behavior, may be warranted, though care must be taken when increasing the complexity of these models to not extend beyond the information content of the data.
- **Reduce model complexity** Investigation of approaches to simplify future assessments is recommended. As part of this, a review of the current assessment structure is warranted to determine whether the current spatial structure is tenable with the available data and uncertainties surrounding basic biological processes such as growth.

8.5 Main assessment conclusions

The current assessment derived results from a total uncertainty grid of 72 model runs. The Scientific Committee will need to decide whether all axes of uncertainty are included when deriving final management advice, and/or any relative weightings applied to different axis elements. Based upon the results from the diagnostic case model and full uncertainty grid, the general conclusions of this assessment are as follows:

1. Spawning potential remained relatively stable until the early 1970s, after which time all models in the structural uncertainty grid showed notable declines across the remaining time series. Estimated recruitment showed declines from the 1960s to the mid 1990s, before showing a slightly increasing trend.

2. Estimates of stock status from the structural uncertainty grid from the 2020 assessment were generally more optimistic than from the 2017 assessment. This is strongly linked to the incorporation of the new growth (similar to the 2017 bigeye assessment McKechnie et al., 2017a). However, alternative treatment of tag data, assumptions regarding selectivity to better fit the data, and the use of maturity at length also resulted in a more optimistic stock status.

3. All models in the structural uncertainty grid show WCPO yellowfin tuna to be above $20\%SB_{F=0}$, which is consistent with the previous assessment. Median terminal depletion ($SB_{recent}/SB_{F=0}$) was 0.58 (80 percentile range: 0.51-0.64). The influence of more positive recruitments estimated in the terminal period of the previous stock assessment led to more optimistic stock status in the recent period, but have now moved through the population.

4. All model regions show declines in yellowfin spawning potential over the time series from the diagnostic case model. Depletion is notably greater in the tropical regions (3, 4, 7 and 8) as well as in region 5 (the south west temperate region), influenced by stronger declining signals in the regional standardized CPUE series. Other temperate regions show relatively limited estimated fishing impacts, with trends being influenced by estimated declines in recruitment. In all cases, regional terminal depletion also remained above $20\%SB_{F=0}$.

5. Average fishing mortality rates for juvenile and adult age-classes increased equivalently from the 1970s, with that for juveniles plateauing from the early 2000s.

6. All models in the structural uncertainty grid showed exploitation of WCPO yellowfin tuna to be below F_{MSY}. Median $F_{recent}/F_{\text{MSY}}$ was 0.36 (80 percentile range: 0.27-0.47).

7. The three most influential uncertainty grid axes were growth, tag mixing, and steepness. Incorporation of age-at-length information within growth estimates led to more optimistic depletion estimates and exploitation rates. The assumption of a tag mixing period of one quarter resulted in a more pessimistic exploitation rates and depletion status. The steepness axis displayed largely predictable results, with steepness of 0.65 and 0.95 producing more

pessimistic and optimistic estimates, respectively, than the 0.8 assumed in the diagnostic case model. The lower the steepness the more depleted the stock, and the higher the fishing mortality with respect to F_{MSY}.

8.6 Acknowledgments

We thank Keisuke Satoh and Takayuki Matsumoto for the provision of the JPTP tagging data and the Japanese government, NRIFSF and Ajinomoto Co. Inc. for supporting the JPTP tagging program. We also thank participants at the pre-assessment workshop (Noumea, April 2017) for their contributions to these analyses. We would also like to thank Sam McKechnie's contribution to this report through analysis of the jittering, additions to the text, alteration of plots, citation/reference management, and humorous relief. We acknowledge his desire to not be included in the list of authors.

References

Abascal, F., Lawson, T., and Williams, P. (2014). Analysis of purse seine size data for skipjack, bigeye and yellowfin tunas. Technical Report WCPFC-SC10-2014/SA-IP-05, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

Aguila, R. D., Perez, S. K. L., Catacutan, B. J. N., Lopez, G. V., Barut, N. C., and Santos, M. D. (2015). Distinct yellowfin tuna (Thunnus albacares) stocks detected in the western and central Pacific Ocean (WCPO) using DNA microsatellites. *PLOS ONE*, 10(9):e0138292.

Anderson, G., Lal, M., Hampton, J., Smith, N., and Rico, C. (2019). Close Kin Proximity in Yellowfin Tuna (Thunnus albacares) as a Driver of Population Genetic Structure in the Tropical Western and Central Pacific Ocean. *Front. Mar. Sci.*, 6:341.

Berger, A. M., McKechnie, S., Abascal, F., Kumasi, B., Usu, T., and Nichol, S. J. (2014). Analysis of tagging data for the 2014 tropical tuna assessments: data quality rules, tagger effects, and reporting rates. Technical Report WCPFC-SC10-2014/SA-IP-06, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

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.

Cass-Calay, S., Tetzlaff, J., Cummings, N., and Isley, J. (2014). Model diagnostics for StockSynthesis 3: Examples from the 2012 assessment of cobia in the U.S. Gulf of Mexico. Technical Report 70(5), ICCAT.

Davies, N., Fournier, D., Bouye, F., and Hampton, J. (2020). Developments in the MULTIFAN-CL Software 2019-20. Technical Report WCPFC-SC16-SA-IP-01.

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., Harley, S., Hampton, J., and McKechnie, S. (2014). Stock assessment of yellowfin tuna in the Western and Central Pacific Ocean. Technical Report WCPFC-SC10-2014/SA-WP-04, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

Davies, N., Hoyle, S., Harley, S., Langley, A., Kleiber, P., and Hampton, J. (2011). Stock assessment of bigeye tuna in the western and central Pacific Ocean. Technical Report WCPFC-SC7-2011/SA-WP-02, Pohnpei, Federated States of Micronesia.

Ducharme-Barth, N., Vincent, M., Hampton, J., Hamer, P., and Pilling, G. (2020a). Stock assessment of bigeye tuna in the western and central Pacific Ocean. Technical Report SC16-SA-IP-07.

Ducharme-Barth, N., Vincent, M., Pilling, G., and Hampton, J. (2019). Simulation analysis of pole and line CPUE standardization approaches for skipjack tuna in the WCPO. Technical Report WCPFC-SC15-2019/SA-WP-04, Pohnpei, Federated States of Micronesia 12-20 August 2019.

Ducharme-Barth, N., Vincent, M., Vidal, T., and Hamer, P. (2020b). Analysis of Pacific-wide operational longline dataset for bigeye and yellowfin tuna catch-per-unit-effort (CPUE). Technical Report SC16-SA-IP-07.

Eveson, P., Vincent, M., Farley, J., Krusic-Golub, K., and Hampton, J. (2020). Integrated growth models from otolith and tagging data for yellowfin and bigeye tuna in the western and central Pacific Ocean. Technical Report SC16-SA-IP-03.

Farley, J., Krusic-Golub, K., Eveson, P., Clear, N., Rouspard, F., Sanchez, C., Nicol, S., and Hampton, J. (2020). Age and growth of yellowfin and bigeye tuna in the western and central Pacific Ocean from otoliths. Technical Report SC16-SA-WP-02.

Fonteneau, A. (2002). Estimated sex ratio of large yellowfin taken by purse seiners in the Indian Ocean: comparison with other oceans. Technical Report IOTC Proc., 5: 279–281.

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, \textbackslashtextbackslashtextbackslashemphThunnus 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. (1999). The impact of correlations in standardised CPUE indices. NZ Fisheries Assessment Research Document 99/42, National Institute of Water and Atmospheric Research.

Grewe, P. M., Feutry, P., Hill, P. L., Gunasekera, R. M., Schaefer, K. M., Itano, D. G., Fuller, D. W., Foster, S. D., and Davies, C. R. (2015). Evidence of discrete yellowfin tuna (Thunnus albacares) populations demands rethink of management for this glglobal important resource. *Nature Science Reports*, 5(16915).

Grewe, P. M., Irianto, H. E., Proctor, C. H., Adam, M. S., Jauhary, A. R., Schaefer, K., Itano, D., Killian, A., and Davies, C. R. (2016). Population structure and provenance of tropical tunas: recent results from high throughput genotyping and potential implications for monitoring and assessment. Technical Report WCPFC-2016-SC12/SA-WP-01, CSIRO, Stones Hotel, Kuta, Bali, Indonesia.

Grüss, A., Walter, J. F., Babcock, E. A., Forrestal, F. C., Thorson, J. T., Lauretta, M. V., and Schirripa, M. J. (2019). Evaluation of the impacts of different treatments of spatio-temporal variation in catch-per-unit-effort standardization models. *Fisheries Research*, 213:75–93.

Gunn, J. S., Clear, N. P., Ward, R., Appleyard, S., Grewe, P. M., and Innes, B. (2002). The origin of recruits to the east coast yellowfin tuna fishery and the delineation of the structure of yellowfin stocks in the western Pacific. Final Report to the Fisheries Research and Development Corporation, CSIRO Marine Research, Hobart, TAS, Australia.

Hamer, P. and Pilling, G. (2020). Report from the SPC pre-assessment E-workshop, Noumea, April 2020. Technical Report WCPFC-SC16-2020/SA-IP-02, Pacific Community.

Hampton, J. (2000). Natural mortality rates in tropical tunas: size really does matter. *Canadian Journal of Fisheries and Aquatic Science*, 57:1002–1010.

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

Hampton, J. and Williams, P. (2016). Annual estimates of purse seine catches by species based on alternative data sources. Technical Report WCPFC-SC12-2016/ST-IP-03, Bali, Indonesia, 3–11 August 2016.

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.

Harley, S. J., Davies, N., Hampton, J., and McKechnie, S. (2014). Stock assessment of bigeye tuna in the Western and Central Pacific Ocean. Technical Report WCPFC-SC10-2014/SA-WP-01, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

Harley, S. J. and Maunder, M. N. (2003). A simple model for age-structured natural mortality based on changes in sex ratios. Technical Report SAR-4-01, Inter-American Tropical Tuna Commission, La Jolla, California, USA, 19–21 May 2003.

Hoyle, S. and Nichol, S. (2008). Sensitivity of bigeye stock assessment to alternative biological and reporductive assumptions. Technical Report WCPFC-SC4-2008/ME-WP-01, Port Moresby, Papua New Guinea, 11–22 August 2008.

Hoyle, S. D. (2008). Adjusted biological parameters and spawning biomass calculations for south Pacific albacore tuna, and their implications for stock assessments. Technical Report WCPFC-SC4-2008/ME-WP-02, Port Moresby, Papua New Guinea, 11–22 August 2008.

Ianelli, J., Maunder, M. N., and Punt, A. E. (2012). Independent review of the 2011 WCPO bigeye tuna assessment. Technical Report WCPFC-SC8-2012/SA-WP-01, Busan, Republic of Korea, 7–15 August 2012.

ISSF (2011). Report of the 2011 ISSF stock assessment workshop. Technical Report ISSF Technical Report 2011-02, Rome, Italy, March 14–17.

Itano, D. (2000). The reproductive biology of yellowfin tuna (ıt Thunnus albacares) in Hawaiian waters and the western tropical Pacific Ocean: Project summary. JIMAR Contribution 00-328 SOEST 00-01.

Kinoshita, J., Aoki, Y., Ducharme-Barth, N., and Kiyofuji, H. (2019). Standardized catch per unit effort (CPUE) of skipjack tuna of the Japanese pole-and-line fisheries in the WCPO from 1972 to 2018. Technical Report WCPFC-SC15-2019/SA-WP-14, Pohnpei, Federated States of Micronesia.

Kleiber, P., Fournier, D., Hampton, J., Davies, N., Bouye, F., and Hoyle, S. (2017). *MULTIFAN-CL User's Guide*. http://www.multifan-cl.org/.

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

Langley, A., Hoyle, S. D., and Hampton, J. (2011). Stock assessment of yellowfin tuna in the Western and Central Pacific ocean. Technical Report WCPFC-SC7-2011/SA-WP-03, Pohnpei, Federated States of Micronesia, 9–17 August 2011.

Langley, A., Okamoto, H., Williams, P., Miyabe, N., and Bigelow, K. (2006). A summary of the data available for the estimation of conversion factors (processed to whole fish weights) for yellowfin and bigeye tuna. Technical Report WCPFC-SC2-2006/ME-IP-03, Manila, Philippines, 7–18 August 2006.

Lee, H. H., Piner, K. R., Methot, R. D., and Maunder, M. N. (2014). Use of likelihood profiling over a global scaling parameter to structure the population dynamics model: An example using blue marlin in the Pacific Ocean. *Fisheries Research*, 158:138–146.

Macdonald, J., Roupsard, F., Leroy, B., Sanchez, C., Hosken, M., Hampton, J., Anderson, G., and Nicol, S. (2020). Appraisal of new biological sampling approaches for tropical tunas on purse seiners. Technical Report WCPFC-SC16-2020/RP-35b-02, Online.

Maunder, M. and Watters, G.M. (2001). Status of yellowfin tuna in the eastern Pacific Ocean 2001. Technical Report InterAmer. Trop. Tuna Comm., Stock Assessment Report 1: 5-86.

Maunder, M. N. and Punt, A. E. (2013). A Review of integrated analysis in fisheries stock assessment. *Fisheries Research*, 142:61–74.

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

McKechnie, S., Harley, S. J., Davies, N., Rice, J., Hampton, J., and Berger, A. (2014a). Basis for regional structures used in the 2014 tropical tuna assessments, including regional weights. Technical Report WCPFC-SC10-2014/SA-IP-02, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

McKechnie, S., Harley, S. J. Chang, S.-K., Liu, H.-I., and Yuan, T.-L. (2014b). Analysis of longline catch per unit effort data for bigeye and yellowfin tunas. Technical Report WCPFC-SC10-2014/SA-IP-03, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

McKechnie, S., Ochi, D., Kiyofuji, H., Peatman, T., and Caillot, S. (2016). Construction of tagging data input files for the 2016 skipjack tuna stock assessment in the western and central Pacific Ocean. Technical Report WCPFC-SC12-2016/SA-IP-05, Bali, Indonesia, 3–11 August 2016.

McKechnie, S., Pilling, G., and Hampton, J. (2017a). 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.

McKechnie, S., Tremblay-Boyer, L., and Harley, S. J. (2015). Analysis of Pacific-wide operational longline CPUE data for bigeye tuna. Technical Report WCPFC-SC11-2015/SA-WP-03, Pohnpei, Federated States of Micronesia, 5–13 August 2015.

McKechnie, S., Tremblay-Boyer, L., and Pilling, G. (2017b). Background analyses for the 2017 stock assessments of bigeye and yellowfin tuna in the western and central Pacific Ocean. Technical Report WCPFC-SC13-2017/SA-IP-06, Rarotonga, Cook Islands.

McPherson, G. (1991). Reproductive biology of yellowfin tuna in the eastern Australian Fishing Zone, with special reference to the north-western Coral Sea. *Mar. Freshwater Res.*, 42(5):465.

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

Moore, B. R., Bell, J. D., Evans, K., Farley, J., Grewe, P. M., Hampton, J., Marie, A. D., Minte-Vera, C., Nicol, S., Pilling, G. M., Scutt Phillips, J., Tremblay-Boyer, L., Williams, A. J., and Smith, N. (2020). Defining the stock structures of key commercial tunas in the Pacific Ocean I: Current knowledge and main uncertainties. *Fisheries Research*, 230:105525.

Mullins, R. B., McKeown, N. J., Sauer, W. H. H., and Shaw, P. W. (2018). Genomic analysis reveals multiple mismatches between biological and management units in yellowfin tuna (Thunnus albacares). *ICES Journal of Marine Science*, 75(6):2145–2152.

Nishikawa, Y., Honma, M., Ueyanagi, S., and Kikawa, S. (1985). Average distribution of larvae of oceanic species of scombroid fishes, 1956–1981. *Far Seas Fish.Res.Lab., 99 p.*

Peatman, T. (2020). Analysis of tag seeding data and reporting rates. Technical Report SC16-SA-IP-04.

Peatman, T., Ducharme-Barth, N., and Vincent, M. (2020). Analysis of purse-seine and longline size frequency data for bigeye and yellowfin tuna in the WCPO. Technical Report SC16-SA-IP-18.

Peatman, T., Fukofuka, S., Park, T., Williams, P., Hampton, J., and Smith, N. (2019). Better purse seine catch composition estimates: progress on the Project 60 work plan. Technical Report.

Pecoraro, C., Babbucci, M., Franch, R., Rico, C., Papetti, C., Chassot, E., Bodin, N., Cariani, A., Bargelloni, L., and Tinti, F. (2018). The population genomics of yellowfin tuna (Thunnus albacares) at global geographic scale challenges current stock delineation. *Sci Rep*, 8(1):13890.

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

Proctor, C., Lester, R., Clear, N. P., Grewe, P. M., Moore, B. R., Eveson, J. P., Lestari, P., Wudji, A., Taufik, M., Wudianto, Lansdell, M., Hill, P., Dietz, C., Thompson, J., Cutmore, S., Foster, S., Gosselin, T., and Davies, C. R. (2019). Population structure of yellowfin tuna (Thunnus albacares) and bigeye tuna (T. obesus) in the Indonesian region. Technical Report Final Report as output of ACIAR Project FIS/2009/059, Australian Centre for International Agricultural Research, Canberra.

Richards, F. K. (1959). A Flexible Growth Function for Empirical Use. *Journal of Experimental Botany*, 10(2):290–301.

Rooker, J. R., David Wells, R. J., Itano, D. G., Thorrold, S. R., and Lee, J. M. (2016). Natal origin and population connectivity of bigeye and yellowfin tuna in the Pacific Ocean. *Fish. Oceanogr.*, 25(3):277–291.

Schaefer, K. M. (1998). Reproductive biology of yellowfin tuna (ıt Thunnus albacares) in the eastern Pacific Ocean. Technical Report 21: 205–272., IATTC Bulletin.

Schaefer, M., Broadhead, G., and Orange, C. (1963). Synopsis on the Biology of Yellowfin Tuna Thunnus (neothumus) Albacares (bonnaterre) 1788 (Pacific Ocean). Technical report.

Scutt Phillips, J., Peatman, T., Vincent, M., and Nicol, S. (2020). Analysis of tagging data for the 2020 tropical tuna assessments: tagger and condition effects. Technical Report SC16-SA-IP-05.

Servidad-Bacordo, R., Dickson, A., Nepomuceno, L., and Ramiscal, R. (2012). Composition, distribution and abundance of fish eggs and larvae in the Philippine Pacific Seaboard and Celebes Sea with focus on tuna larvae (Family: Scombridae). Information Paper WCPFC-SC8-2012/SA-IP-03, Eighth Regular Session of the Scientific Committee of the Western and Central Pacific Fisheries Commission. Busan, Republic of Korea.

Sun, C.-L., Chu, S.-L., and Yeh, S.-Z. (2006). Reproductive biology of bigeye tuna in the Western and Central Pacific Ocean. Technical Report WCPFC-SC2-2006/BI-WP-01, Manila, Philippines, 7–18 August 2006.

Thorson, J. T. (2019). Measuring the impact of oceanographic indices on species distribution shifts: The spatially varying effect of cold-pool extent in the eastern Bering Sea. *Limnology and Oceanography*, 64(6):2632–2645.

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

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

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

Vidal, T. and Hamer, P. (2020). Developing yellowfin tuna recruitment indices from drifting FAD purse seine catch and effort data. Technical Report WCPFC-SC16-SA-IP-08.

Vincent, M. and Ducharme-Barth, N. (2020). Background analyses for the 2020 stock assessments of bigeye and yellowfin tuna in the western and central Pacific Ocean. Technical Report SC16-SA-IP-06.

Vincent, M., Pilling, G., and Hampton, J. (2018). Incorporation of updated growth information within the 2017 WCPO bigeye stock assessment grid, and examination of the sensitivity of estimates to alternative model spatial structures. Technical Report WCPFC-SC14-2018/ SA-WP-03, Oceanic Fisheries Programme, The Pacific Community.

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, R., Elliott, N., and Grewe, P. (1994). Allozyme and mitochondrial DNA variation in yellowfin tuna (Thunnus albacares) from the Pacific Ocean. *Marine Biology*, 118:531–539.

WCPFC (2019). WCPFC-Scientific Committee Fifteenth Regular Session Summary Report. Technical report. Technical report.

Wells, R., Rooker, J., and Itano, D. (2012). Nursery origin of yellowfin tuna in the Hawaiian Islands. *Mar. Ecol. Prog. Ser.*, 461:187–196.

Williams, P. and FFA (2019). Overview of tuna fisheries in the WCPO, including economic conditions – 2018. Technical Report WCPFC-SC15-2019/GN-IP-01.

Williams, P. and Thomas Ruaia (2020). Overview of tuna fisheries in the western and central Pacific Ocean, incuding economic conditions-2019. Technical Report WCPFC-SC16-2020/GN IP-1.

Zhou, S., Campbell, R. A., and Hoyle, S. D. (2019). Catch per unit effort standardization using spatio-temporal models for Australia's Eastern Tuna and Billfish Fishery. *ICES Journal of Marine Science*.

Zhu, G., Xu, L., Zhou, Y., and Song, L. (2008). Reproductive biology of yellowfin tuna T. albacares in the west-central Indian Ocean. *J. Ocean Univ. China*, 7(3):327–332.

9 Tables

Table 1: Summary of the groupings of fisheries within the assessment for estimation of selectivity, catchability (used for the implementation of regional weights), tag recaptures (Recaps), and tag reporting rates (Tag RR). Note that effort is missing for all non-index fisheries so effort deviation penalties only apply to the last four quarters (seeSection 5.4.3).]

Fishery	Region	Selectivity	Catchability	EffPen	EffPenCV	Recaps	Tag RR
F1 L-ALL-1	1	1	1	constant	0.41	1	1
F2 L-ALL-2	2	2	2	constant	0.41	2	1
F3 L-US-2	2	3	3	constant	0.41	3	2
F4 L-ALL-3	3	4	4	constant	0.41	4	1
F5 L-OS-3	3	8	5	constant	0.41	5	1
F6 L-OS-7	7	9	6	constant	0.41	6	1
F7 L-ALL-7	7	10	7	constant	0.41	7	1
F8 L-ALL-8	8	11	8	constant	0.41	8	1
F9 L-ALL-4	4	5	9	constant	0.41	9	1
F10 L-AU-5	5	12	10	constant	0.41	10	3
F11 L-ALL-5	5	7	11	constant	0.41	11	1
F12 L-ALL-6	6	6	12	constant	0.41	12	1
F13 S-ALL-3	3	13	13	constant	0.41	13	4
F14 S-ALL-3	3	16	14	constant	0.41	13	4
F15 S-ALL-4	4	14	15	constant	0.41	14	5
F16 S-ALL-4	4	17	16	constant	0.41	14	5
F17 Z-PH-7	7	19	17	constant	0.41	15	6
F18 H-ID.PH-7	7	20	18	constant	0.41	15	6
F19 S-JP-1	1	21	19	constant	0.41	16	7
F20 P-JP-1	1	22	20	constant	0.41	17	8
F21 P-ALL-3	3	23	21	constant	0.41	18	9
F22 P-ALL-8	8	24	22	constant	0.41	19	9
F23 Z-ID-7	7	25	23	constant	0.41	15	6
F24 S-ID.PH-7	7	28	24	constant	0.41	15	6
F25 S-ALL-8	8	15	25	constant	0.41	20	10
F26 S-ALL-8	8	18	26	constant	0.41	20	10
F27 L-AU-9	9	12	27	constant	0.41	21	3
F28 P-ALL-7	7	26	28	constant	0.41	22	11
F29 L-ALL-9	9	7	29	constant	0.41	23	1
F30 S-ALL-7	7	13	30	constant	0.41	24	12
F31 S-ALL-7	7	16	31	constant	0.41	24	12
F32 Z-VN-7	7	27	32	constant	0.41	25	13
F33 I-I-1	1	29	33	time-variant	0.20	25	13
F34 I-I-2	2	29	33	time-variant	0.20	25	13
F35 I-I-3	3	29	33	time-variant	0.20	25	13
F36 I-I-4	4	29	33	time-variant	0.20	25	13
F37 I-I-5	5	29	33	time-variant	0.20	25	13
F38 I-I-6	6	29	33	time-variant	0.20	25	13
F39 I-I-7	7	29	33	time-variant	0.20	25	13
F40 I-I-8	8	29	33	time-variant	0.20	25	13
F41 I-I-9	9	29	33	time-variant	0.20	25	13

Table 2: Summary of the number of release events, tag releases and recoveries by region and program

Prog		JPTP			PTTP			RTTP	
Years		1999–2017			2006–2017			1989–1995	
Category	Grps	Rel	Rec	Grps	Rel	Rec	Grps	Rel	Rec
1	52	10295	1013	0	0	0	0	0	0
2	1	14	0	0	0	0	0	0	0
3	1	11	0	15	7800	1343	6	1837	235
4	3	211	11	13	3236	568	4	1951	188
5	0	0	0	5	3870	479	2	1112	209
6	0	0	0	0	0	0	3	830	41
7	1	20	0	3	9563	1558	7	9653	2361
8	0	0	0	17	54871	13054	9	8631	1269
9	0	0	0	0	0	0	3	2221	77
Total	58	10551	1024	53	79339	17002	34	26235	4380

Table 3: 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 2014–2017 and for SB is the average over the period 2015–2018 and "latest" is 2018

Symbol	Description
C_{latest}	Catch in the last year of the assessment (2018)
F_{recent}	Average fishing mortality-at-age for a recent period (2014–2017)
YF_{recent}	Equilibrium yield at average fishing mortality for a recent period (2014–2017)
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_{latest}	Spawning biomass in the latest time period (2018)
SB_{recent}	Spawning biomass for a recent period (2015–2018)
$SB_{F=0}$	Average spawning biomass predicted in the absence of fishing for the period 2008–2017
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 2008–2017
$SB_{latest}/SB_{F=0}$	Spawning biomass in the latest time period (2018) relative to the average spawning biomass predicted to occur in the absence of fishing for the period 2008–2017
$SB_{latest}/SB_{\mathrm{MSY}}$	Spawning biomass in the latest time period (2018) relative to that which will produce the maximum sustainable yield (MSY)
$SB_{recent}/SB_{F=0}$	Spawning biomass for a recent period (2015–2018) relative to the average spawning biomass predicted to occur in the absence of fishing for the period 2008–2017
$SB_{recent}/SB_{\mathrm{MSY}}$	Spawning biomass for a recent period (2015–2018) relative to the spawning biomass that produces maximum sustainable yield (MSY)
$20\%SB_{F=0}$	WCPFC adopted limit reference point – 20% of spawning biomass in the absence of fishing average over years $t-10$ to $t-1$ (2008–2017)

Table 4: 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, 0.80*, or 0.95
Growth	3	Modal estimate, External otolith, Cond age-at-length*
Size frequency weighting	4	sample sizes divided by 20, 60*, 200, or 500
Tag mixing	2	1 quarter, 2 quarters*

Table 5: Summary of reference points over the 72 models in the structural uncertainty grid.

	Mean	Median	Min	10	90	Max
C_{latest}	709389	711072	700358	702279	712761	714073
YF_{recent}	779872	784200	661600	707720	877040	908000
fmult	2.868	2.798	1.702	2.115	3.716	4.289
F_{MSY}	0.105	0.104	0.085	0.091	0.120	0.147
MSY	1090706	1091200	791600	874200	1283920	1344400
$F_{recent}/F_{\text{MSY}}$	0.366	0.357	0.233	0.269	0.473	0.588
$SB_{F=0}$	3641228	3603980	2893274	3231353	4050429	4394277
SB_{MSY}	860326	858700	349100	590090	1114400	1322000
$SB_{MSY}/SB_{F=0}$	0.233	0.236	0.121	0.175	0.278	0.302
$SB_{latest}/SB_{F=0}$	0.543	0.542	0.404	0.471	0.601	0.664
$SB_{latest}/SB_{\text{MSY}}$	2.430	2.282	1.466	1.665	3.293	4.889
$SB_{recent}/SB_{F=0}$	0.578	0.583	0.424	0.507	0.641	0.677
$SB_{recent}/SB_{\text{MSY}}$	2.591	2.432	1.538	1.773	3.571	5.267

10 Figures

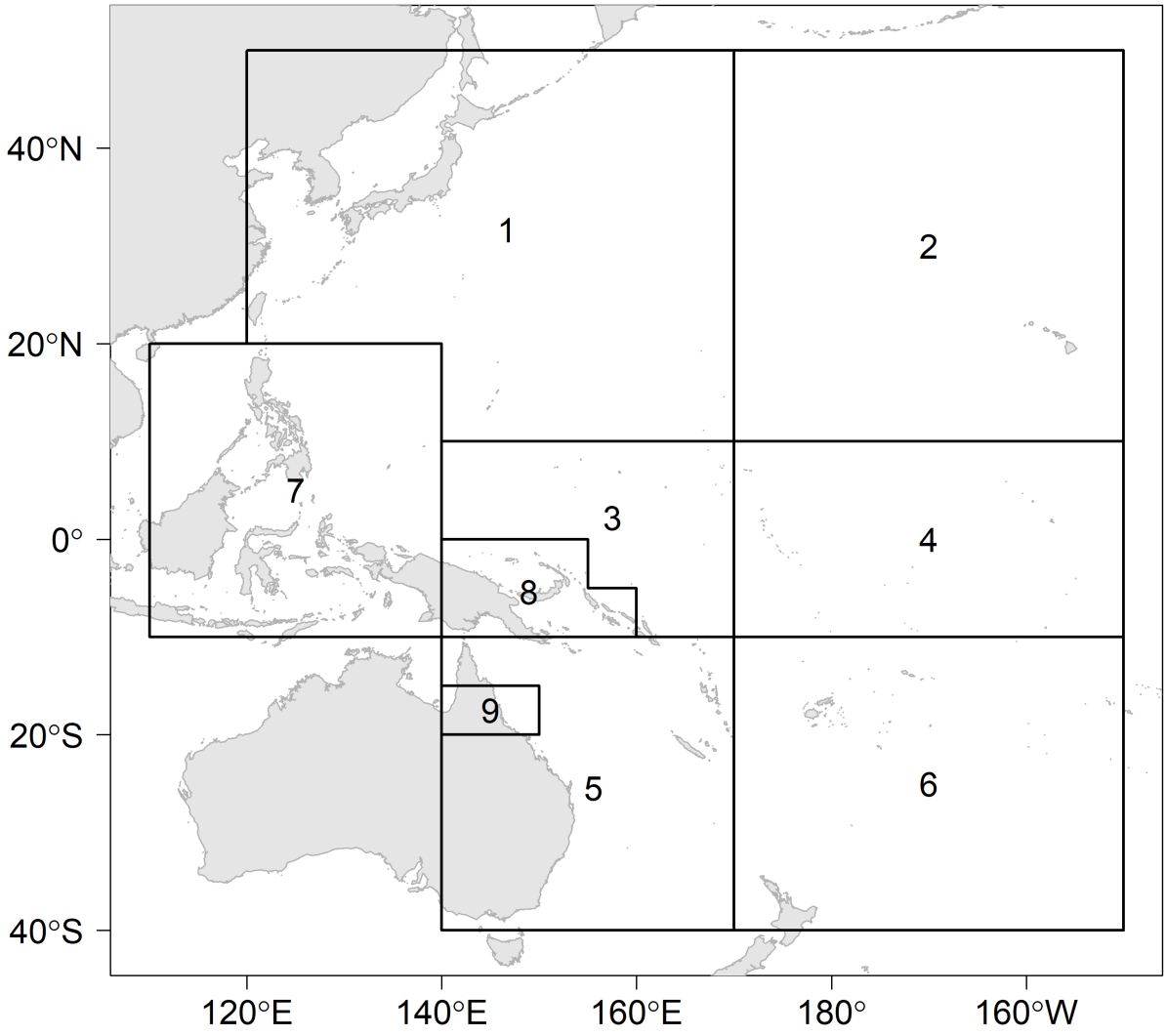


Figure 1: The geographical area covered by the stock assessment and the boundaries for the 9 regions when using the "10N regional structure".

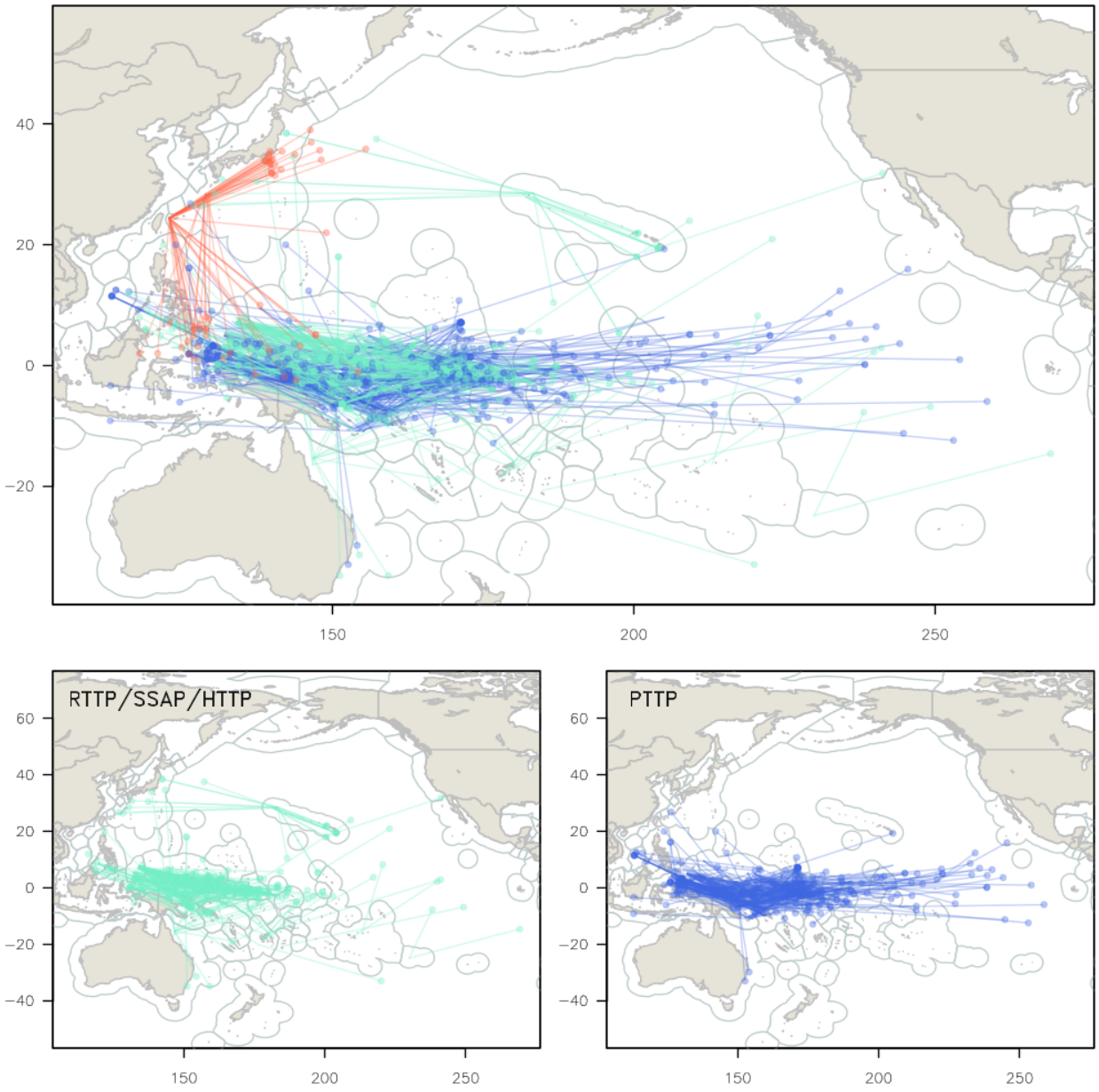


Figure 2: Map of the movements of tagged yellowfin released in the Pacific Ocean and subsequently recaptured more than 1,000 nautical miles from their release site.

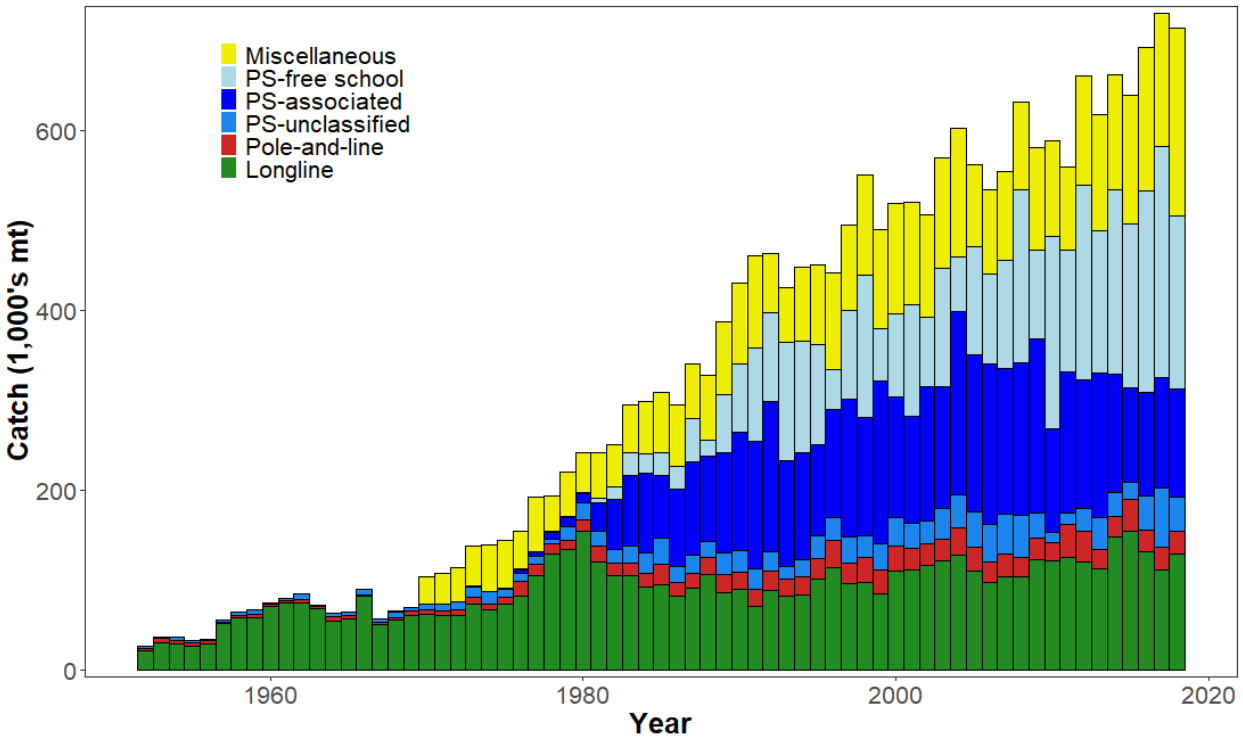


Figure 3: Time series of total annual catch (1000's mt) by fishing gear over the full assessment region and time period. The different colours denote longline (L) (green), pole-and-line (P) (red), purse seine (S) (blue), purse seine-associated (S) (dark blue), purse seine-unassociated (S) (lightblue), miscellaneous (yellow).

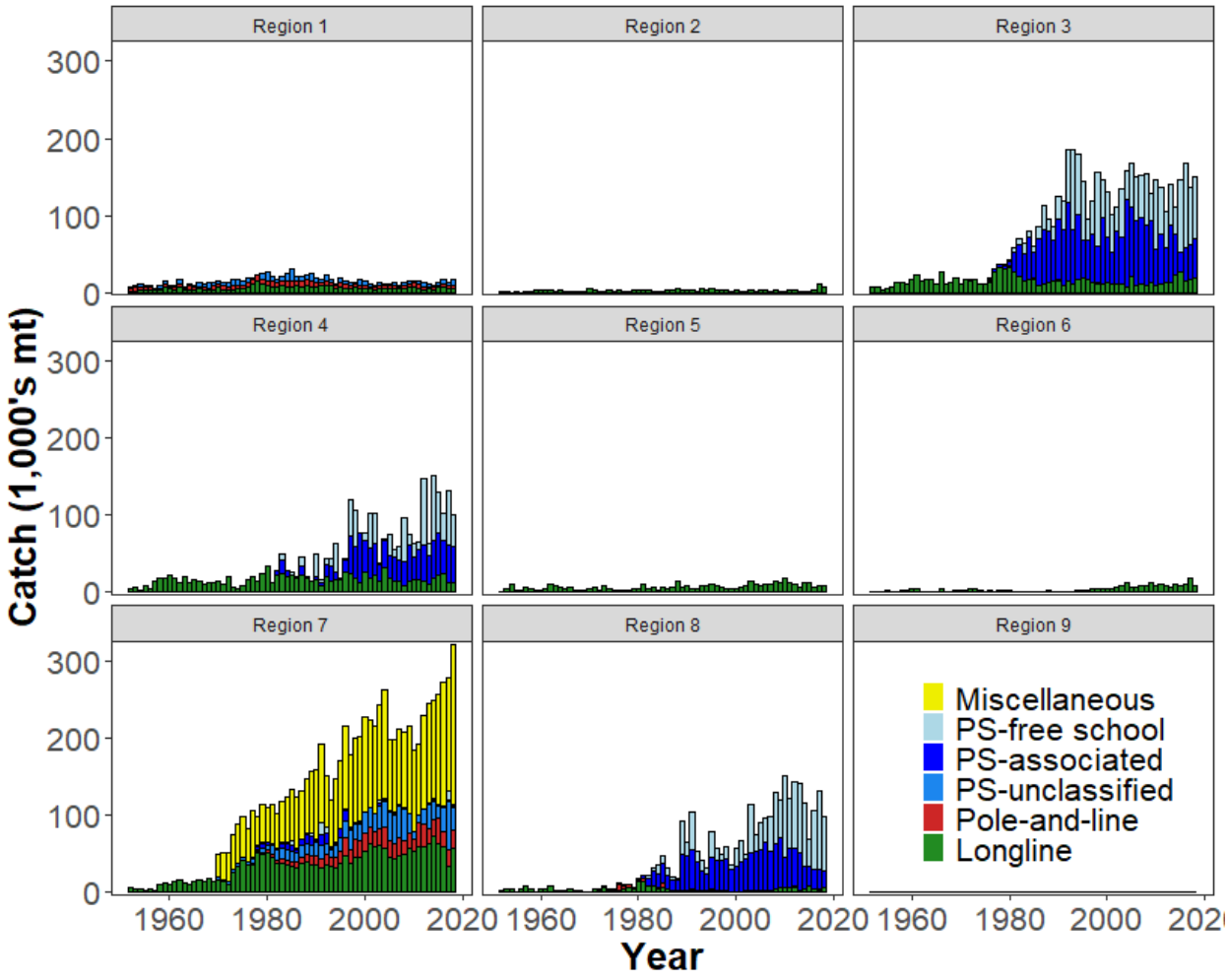


Figure 4: Time series of total annual catch (1000's mt) by fishing gear and assessment region over the full assessment period. The different colours denote longline (L) (green), pole-and-line (P) (red), purse seine (S) (blue), purse seine-associated (S) (dark blue), purse seine-unassociated (S) (lightblue), miscellaneous (yellow).

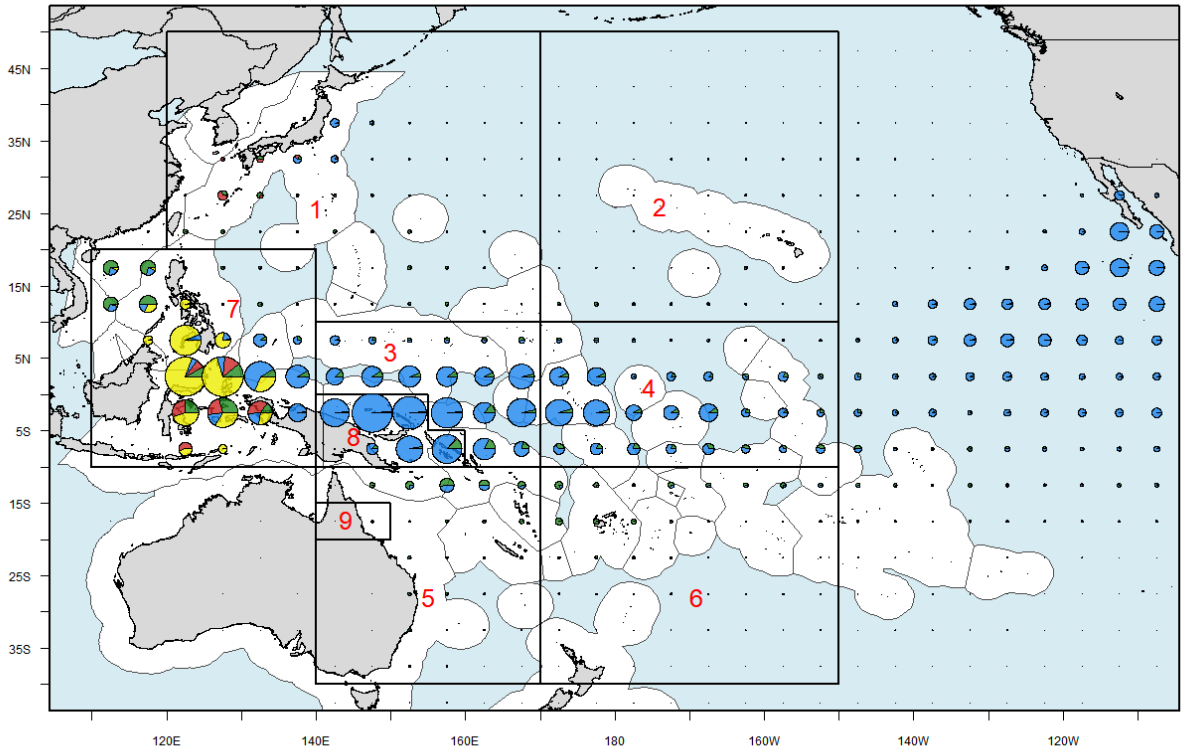


Figure 5: Distribution and magnitude of yellowfin tuna catches for the most recent decade of the stock assessment (2009-2018) by 5° square and fishing gear: longline (green), pole-and-line (red), purse seine (blue) and miscellaneous (yellow), for the WCPO and part of the EPO. Overlayed are the regional boundaries for the stock assessment.

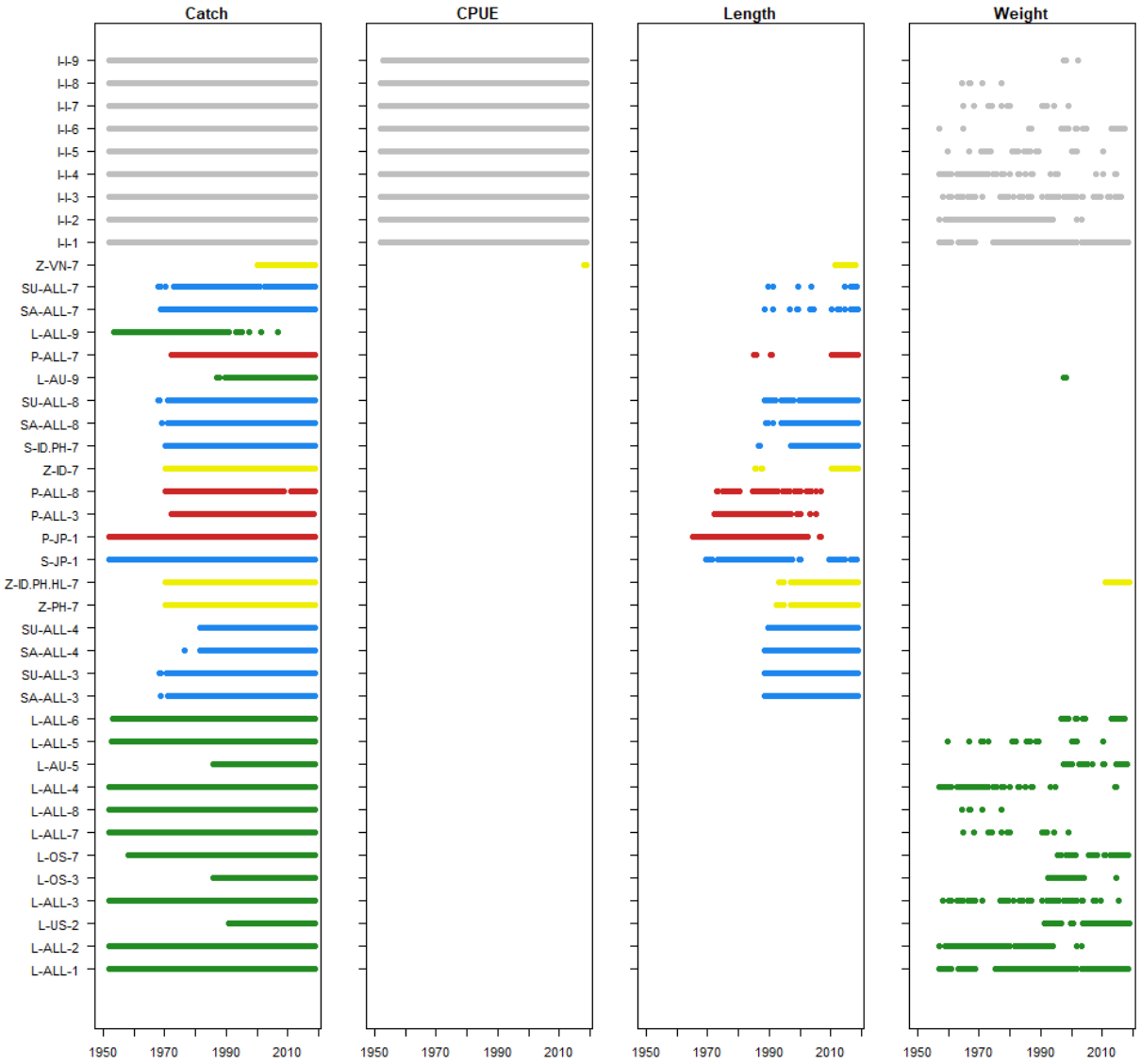


Figure 6: Presence of catch, standardised CPUE, length frequency and weight frequency data by year and fishery for the diagnostic model. The different colours denote gear-type of the fishery: index fisheries (grey); longline (green); pole-and-line (red); purse seine (blue); and miscellaneous (yellow).

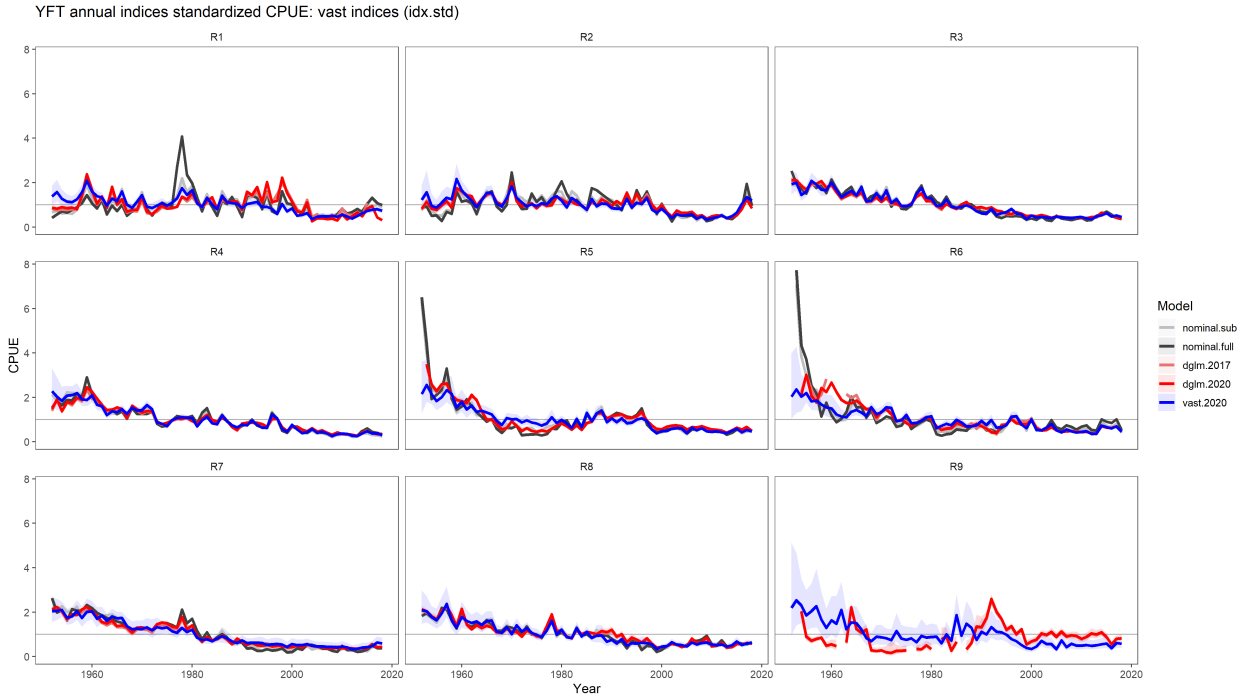


Figure 7: Standardised catch-per-unit-effort (CPUE) indices (geostats approach) for the longline fisheries in regions 1–9 used in the diagnostic model. See Ducharme-Barth et al. (2020b) for further details of the estimation of these CPUE indices. The light grey lines represent the 95% confidence intervals derived from the effort deviation penalties used in the diagnostic model.

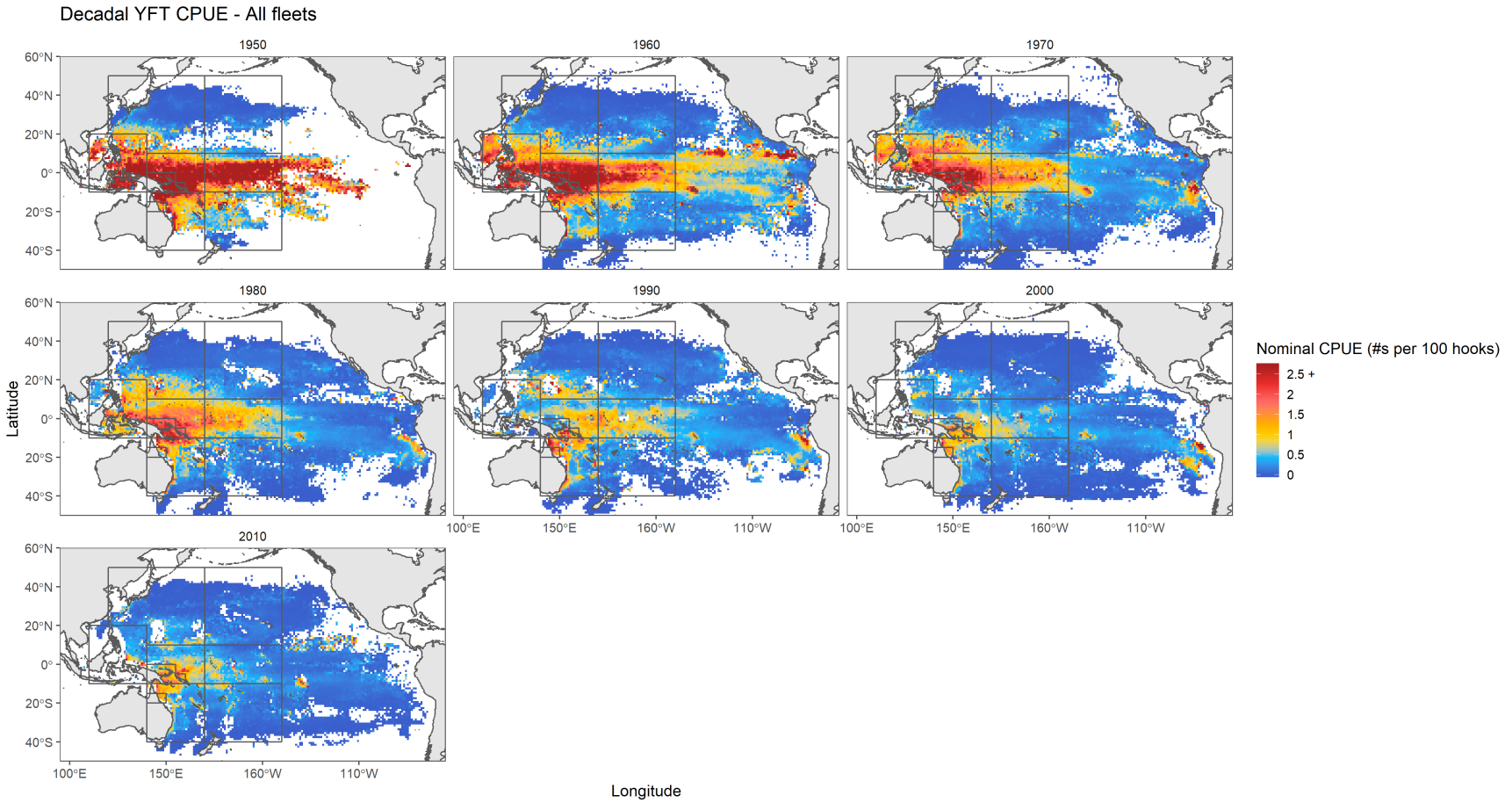


Figure 8: Spatial distribution and decadal patterns of nominal catch-per-unit-effort (CPUE) of yellowfin tuna by longline in the WCPO (all fleets)

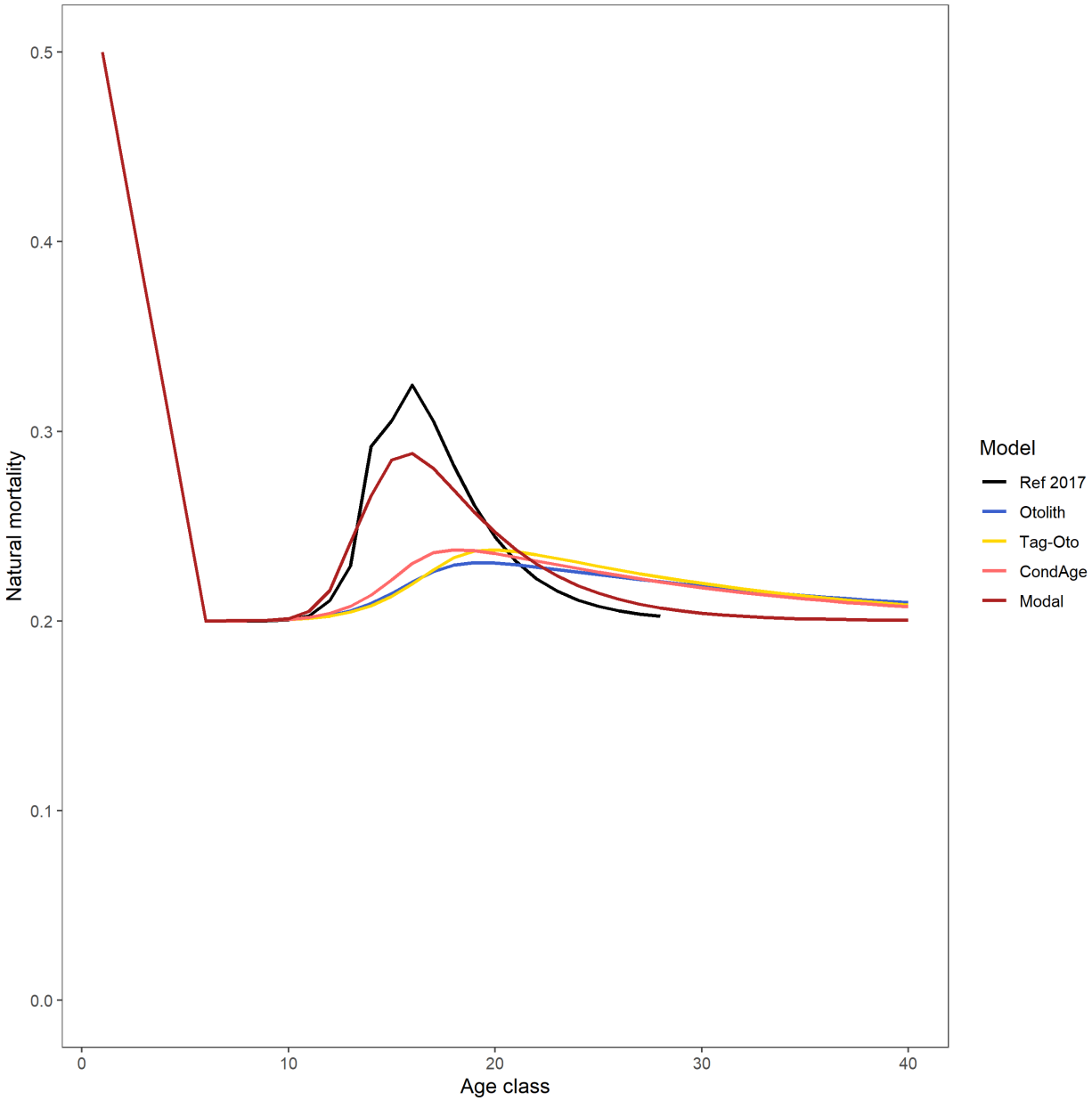


Figure 9: Quarterly natural mortality-at-age. The natural mortality are calculate from the different growth curves where Ref 2017 is M from the 2017 assessment, Otolith is from the otolith only Richards growth curve, Tag-Oto is that otolith tag integrated VB, CondAge is the growth estimated from the diagnostic model, and Modal is growth estimated from the size composition (presented in Section 5.2).

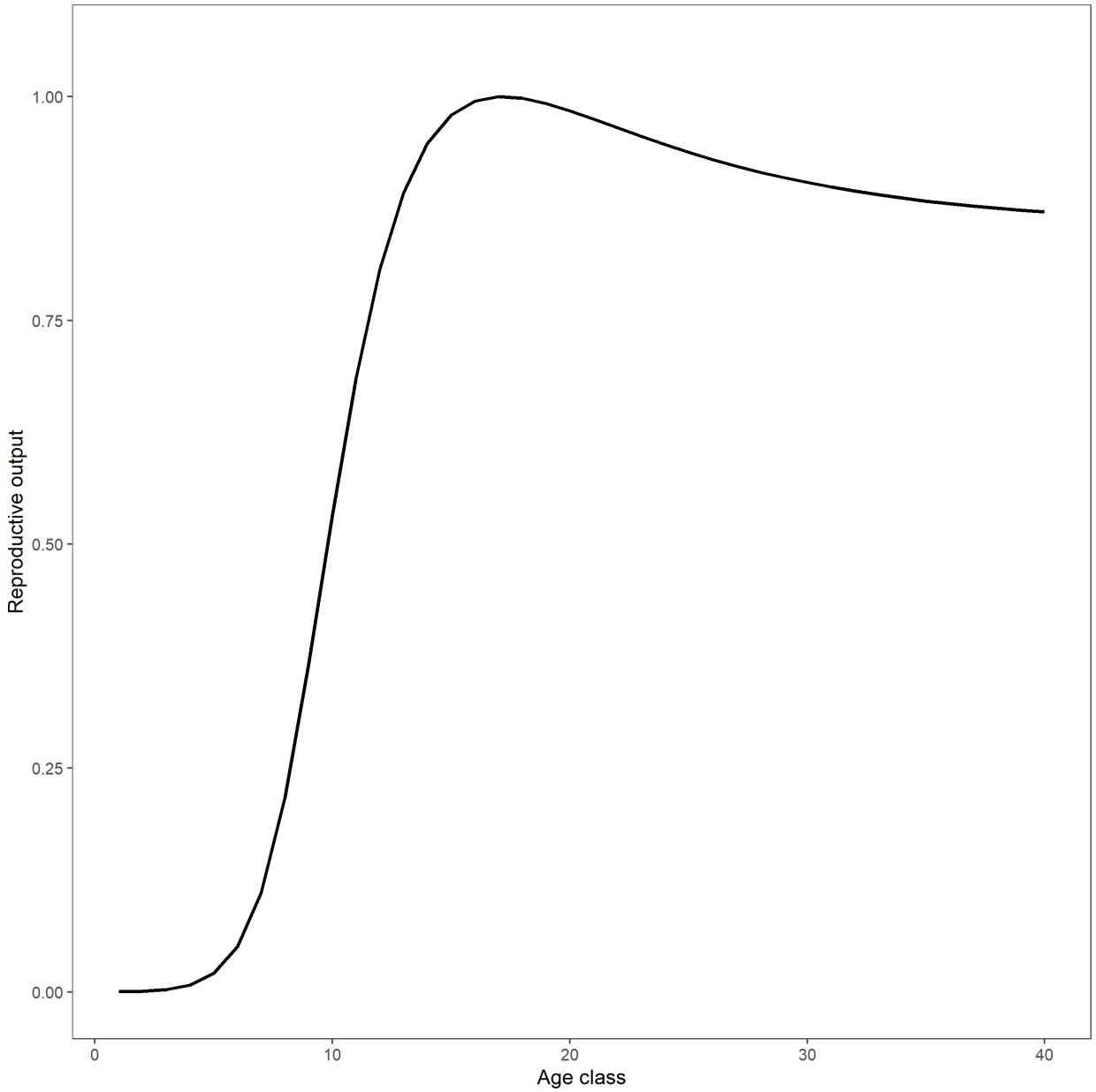


Figure 10: Maturity-at-age as estimated by MFCL using the estimated growth curve and input maturity at length for the diagnostic model.

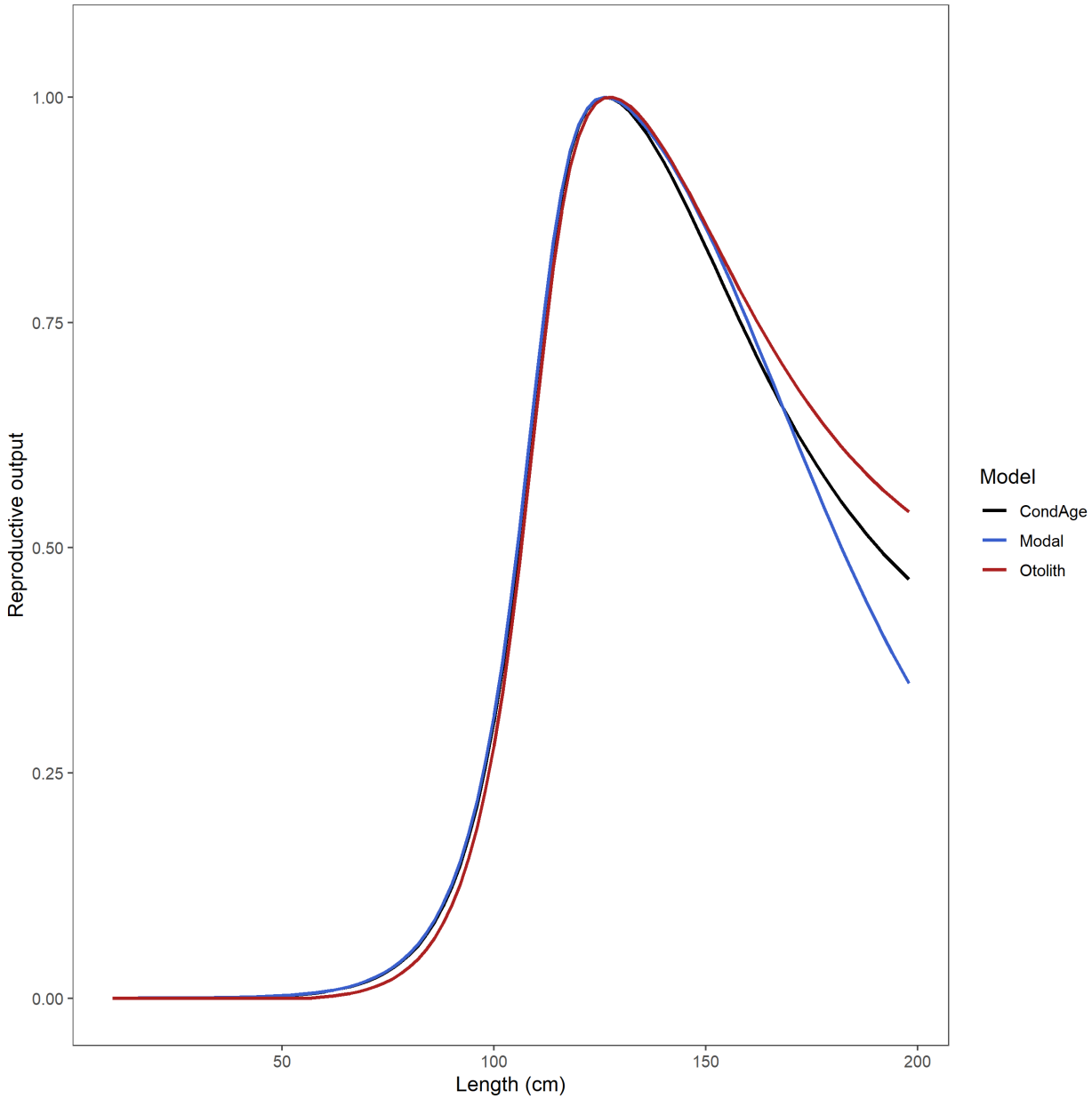


Figure 11: Maturity-at-length input for the different growths in the structural uncertainty grid, where the diagnostic model is CondAge, Modal is from the growth curve estimated by the size composition, and Otolith is from the Richards otolith only growth curve. Slight differences in the maturity-at-length are due to the fit to the proportion male from the natural mortality analysis.

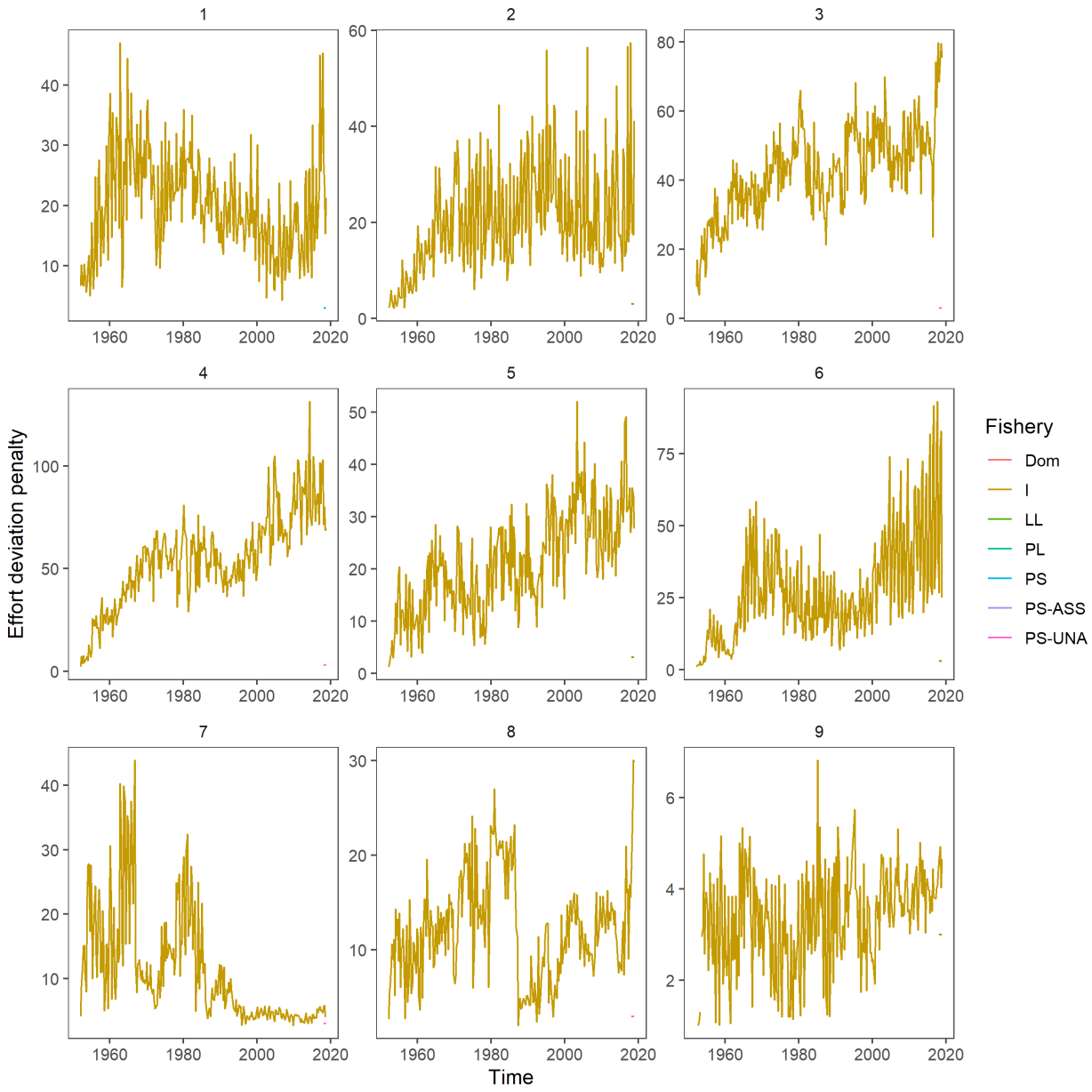
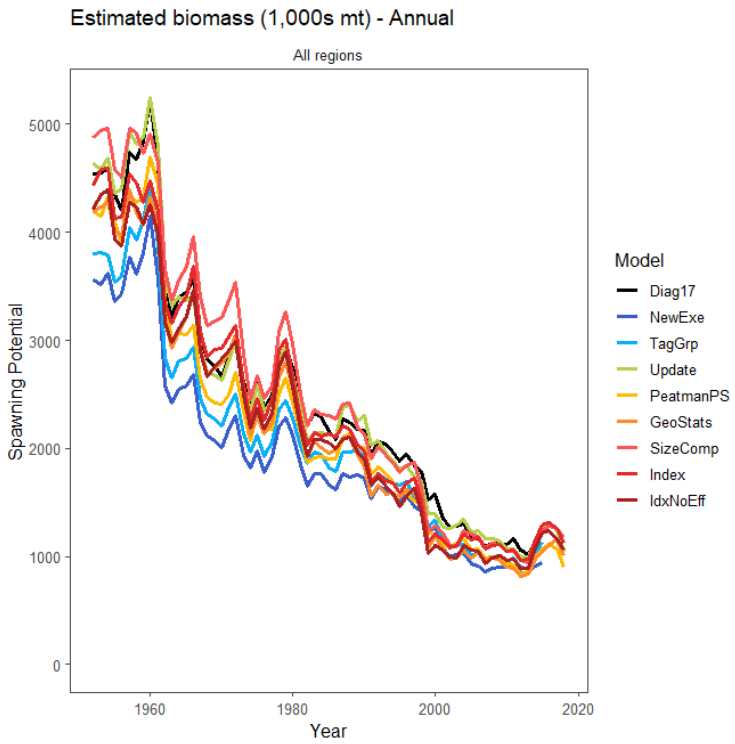


Figure 12: Plot of the effort deviation penalties applied to each fishery, by region, with the colours of the lines representing the gear type, where only the index fishery is given penalties. A higher penalty gives more weight to the CPUE of that fishery. Note the different y-axis scales

(a) Stepwise set 1



(b) Stepwise set 2

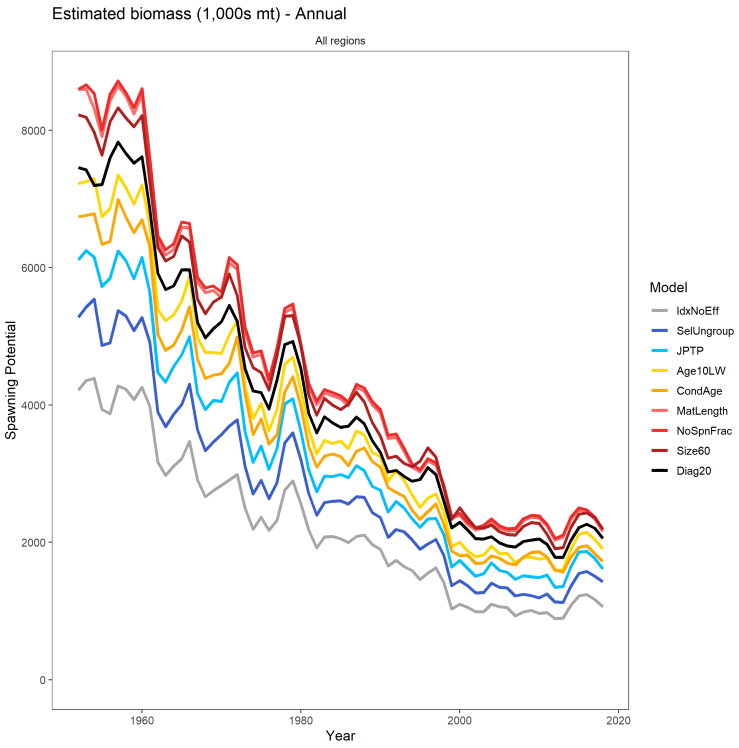
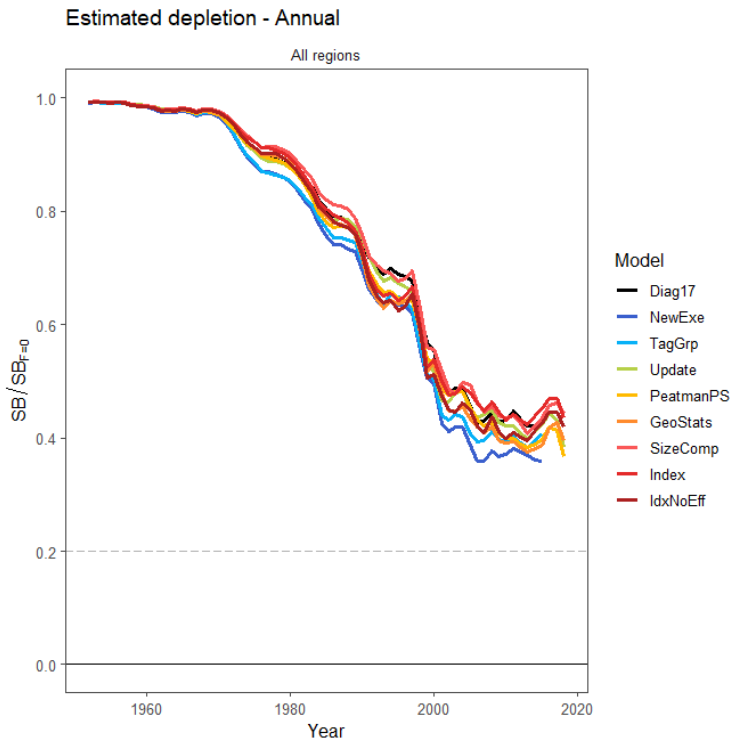


Figure 13: Stepwise changes in spawning potential from the 2017 reference case model through to the 2020 diagnostic model.

(a) Stepwise set 1



(b) Stepwise set 2

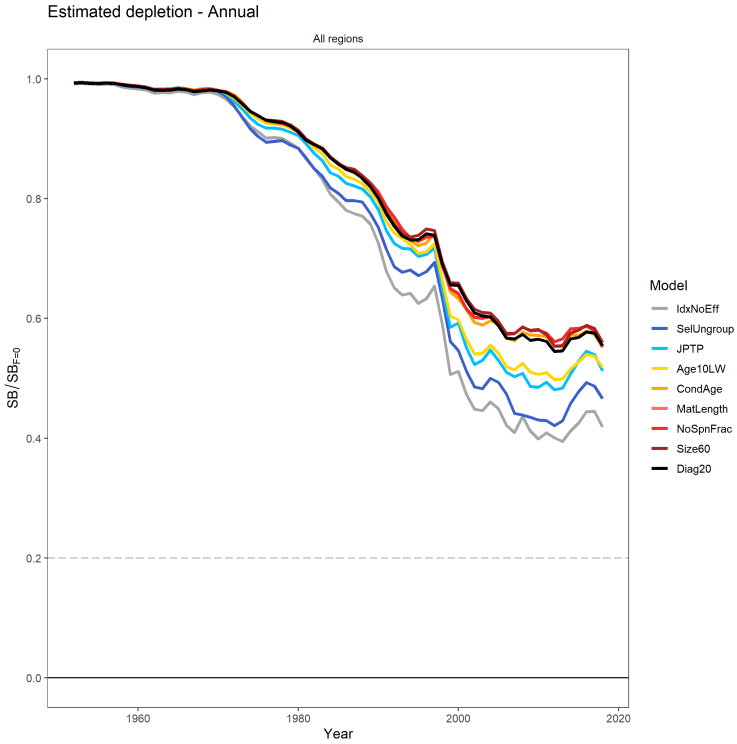


Figure 14: Stepwise changes in adult depletion from the 2017 reference case model through to the 2020 diagnostic model.

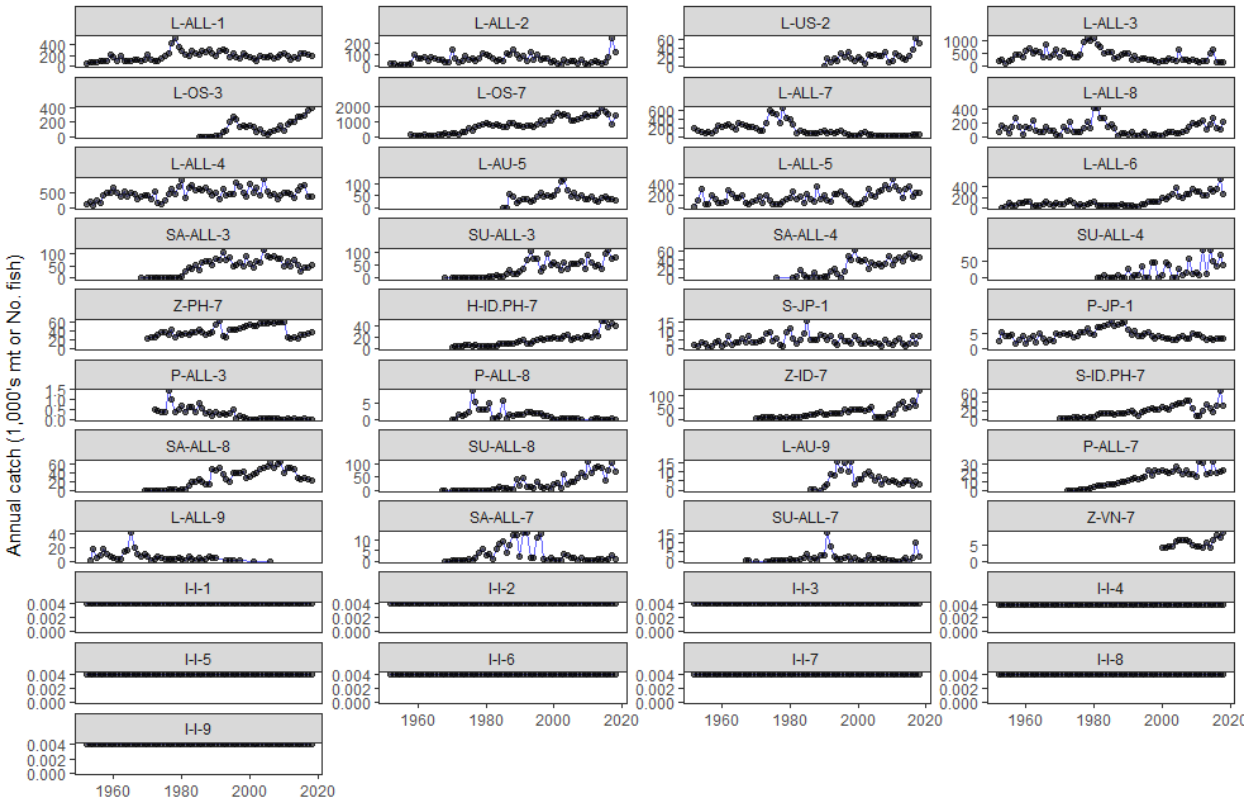


Figure 15: Observed (black points) and model-predicted (blue lines) catch for the 41 fisheries in the diagnostic model. The y-axis is in catch-in-numbers for the longline fisheries and catch-in-weight for the other fisheries, both divided by 1,000.

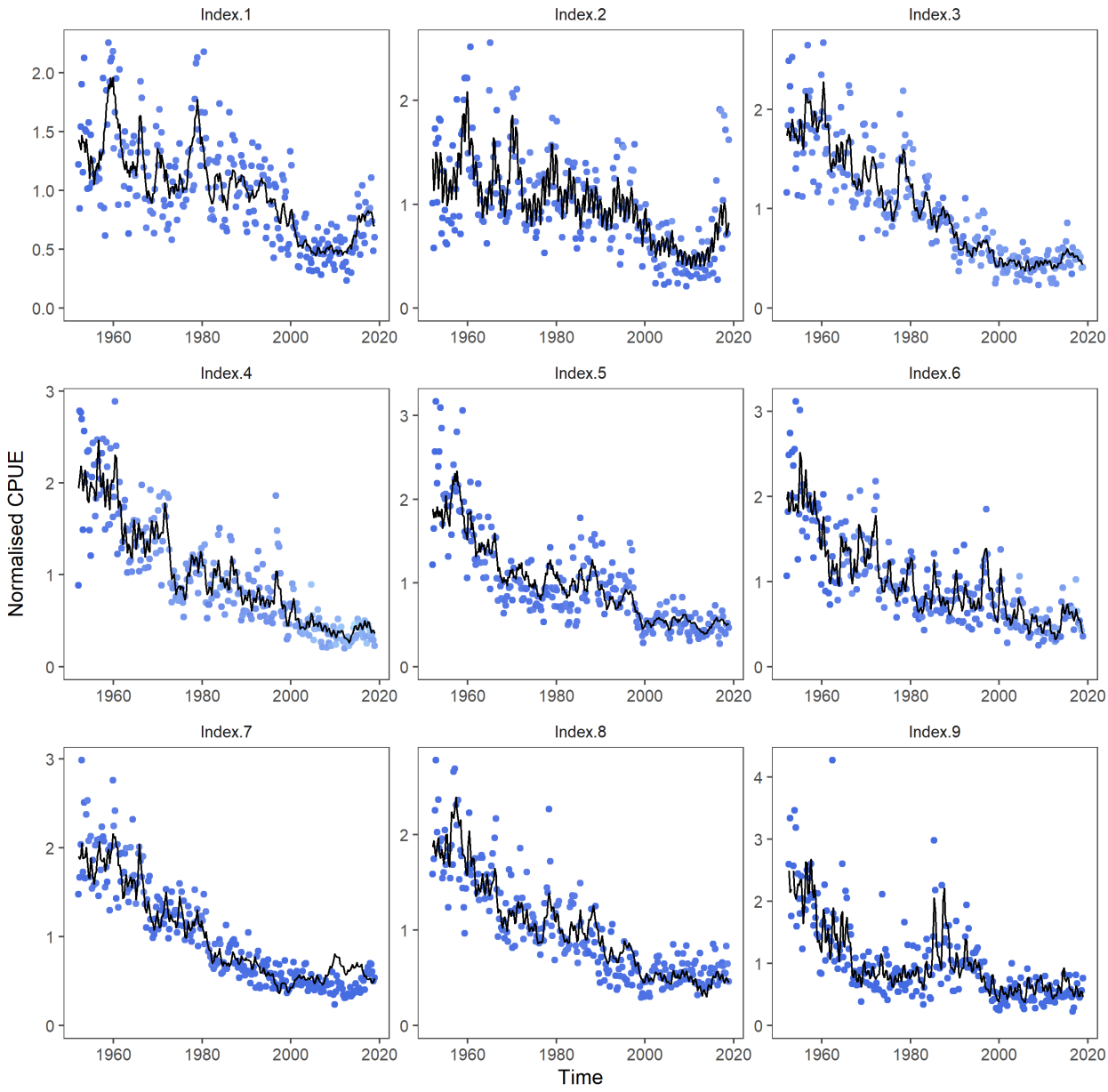


Figure 16: Observed (blue points) and model-predicted (black lines) CPUE for the 9 index fisheries which received the VAST CPUE indices in the diagnostic model.

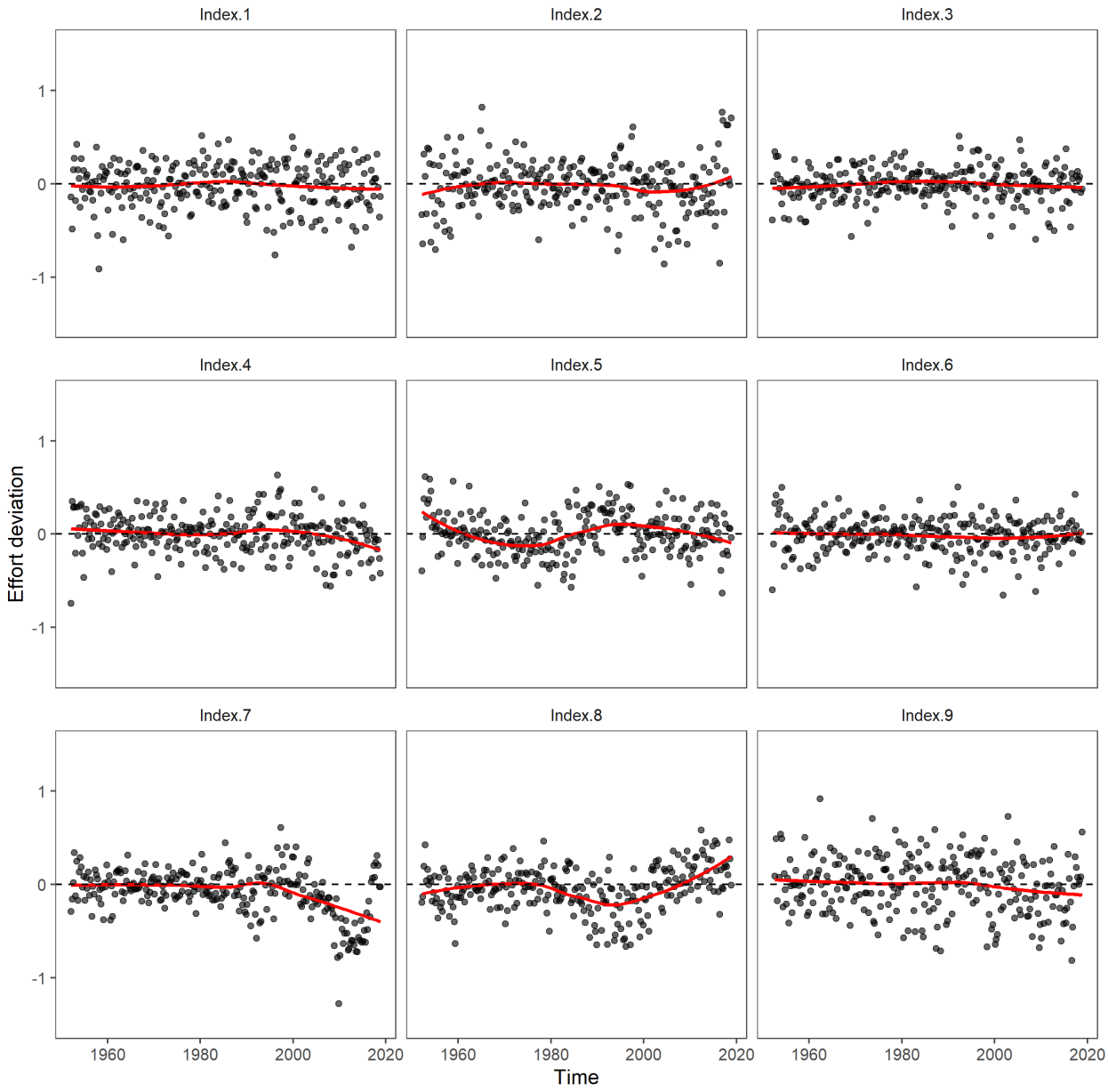


Figure 17: Effort deviations by time period for each of the fisheries receiving standardised CPUE indices in the diagnostic model. The red line represents a lowess smoothed fit to the effort deviations.

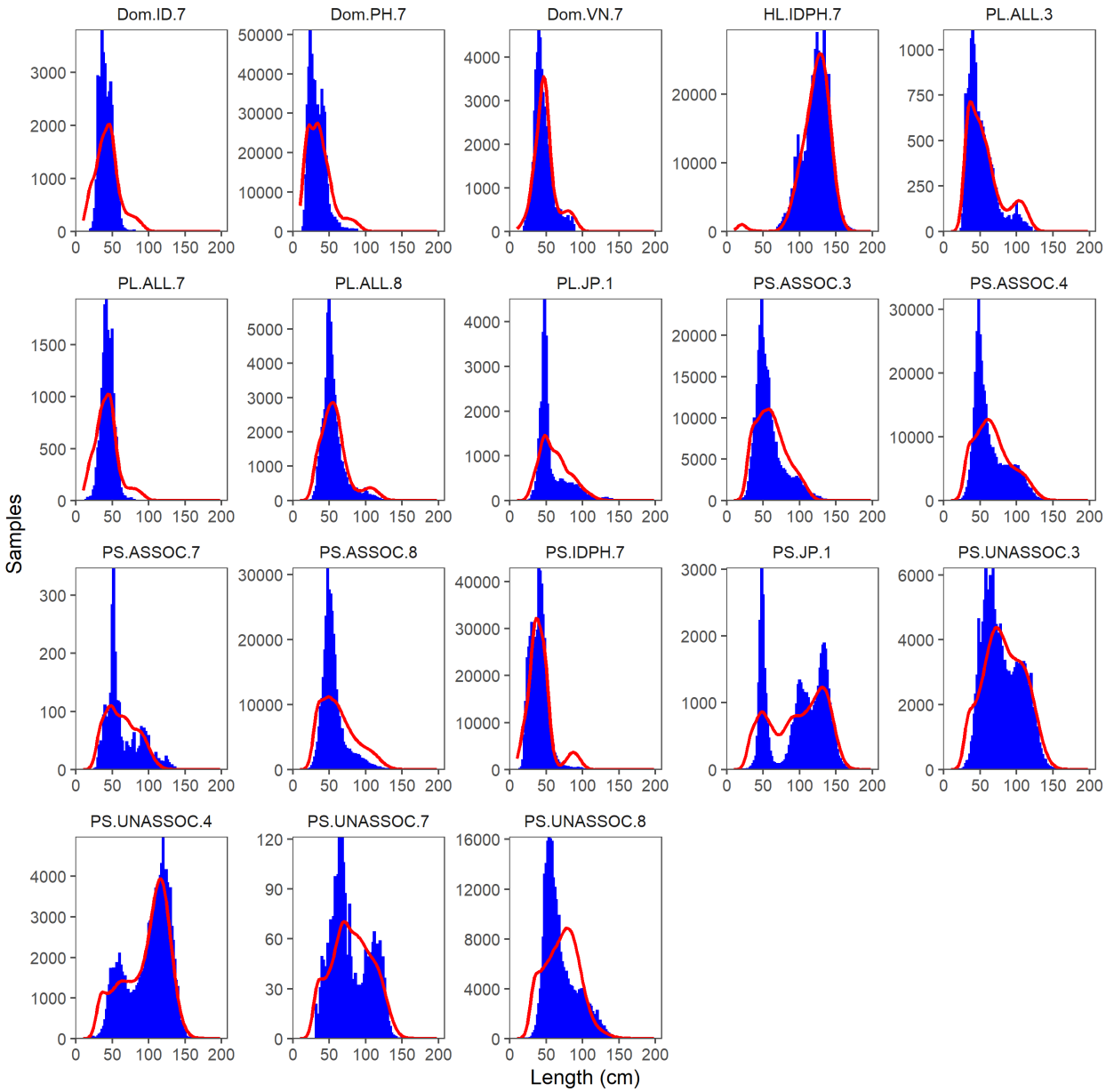


Figure 18: Composite (all time periods combined) observed (blue histograms) and predicted (red lines) catch-at-length for all fisheries with length samples for the diagnostic model.

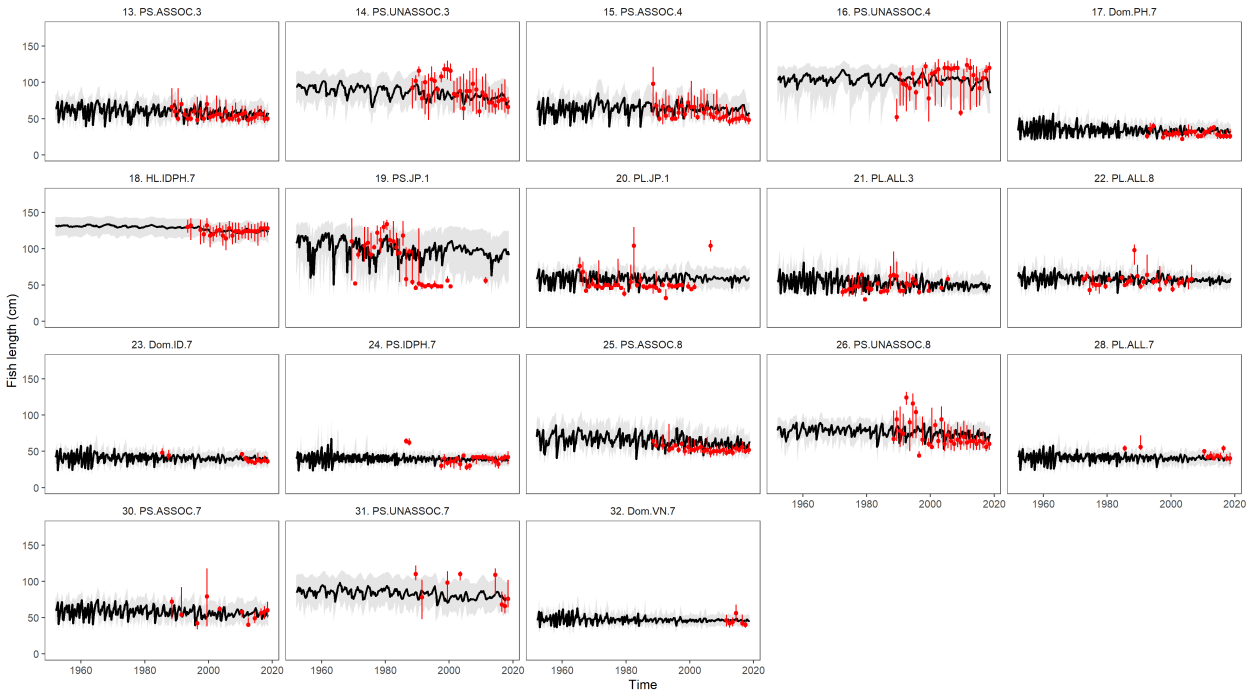


Figure 19: Comparison of the observed (red points) and predicted (grey line) median fish length (FL, cm) for all fisheries with length samples for the diagnostic model. The uncertainty intervals (grey shading) represent the values encompassed by the 25% and 75% quantiles. Sampling data are aggregated by year and only length samples with a minimum of 30 fish per year are plotted.

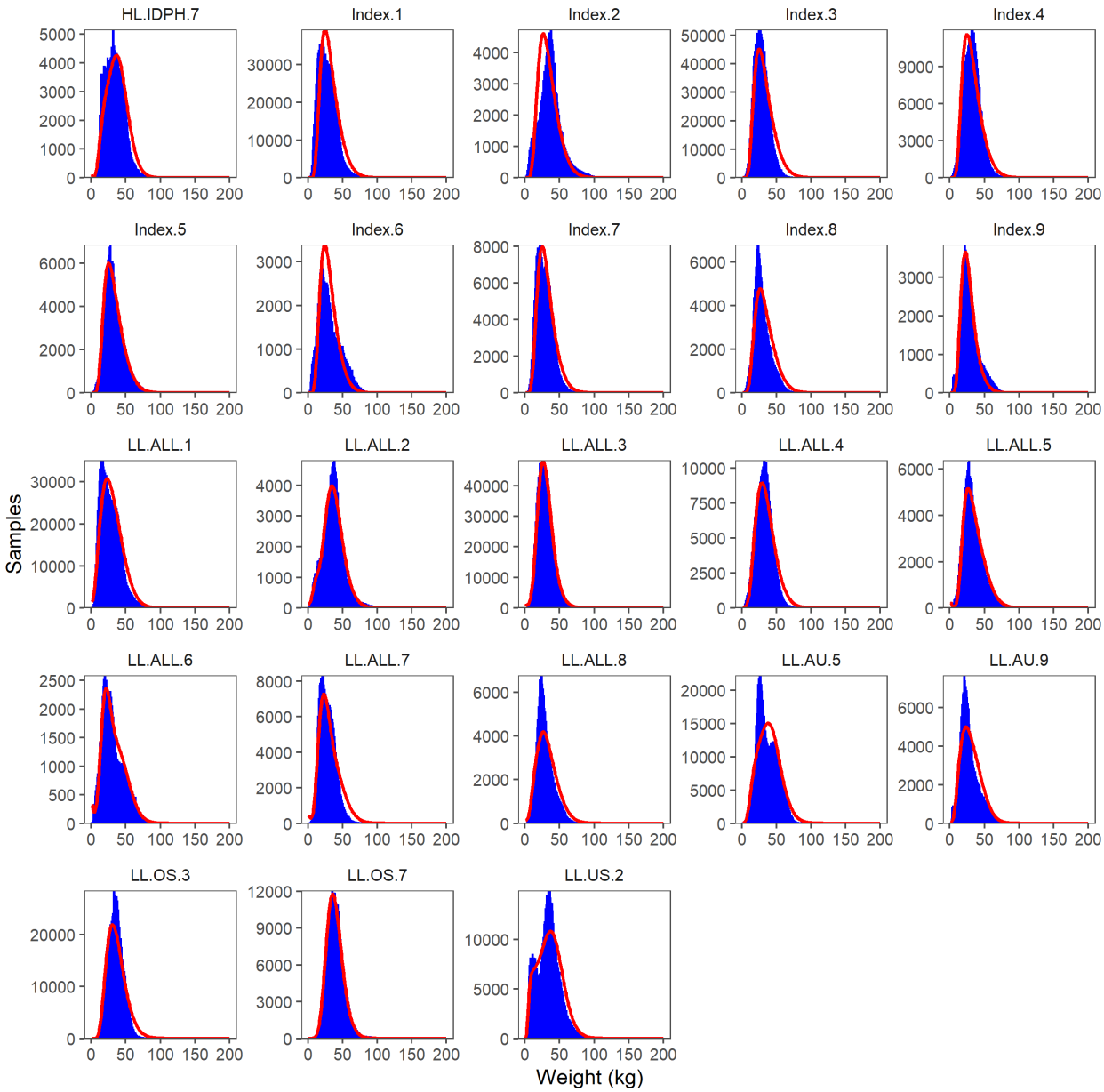


Figure 20: Composite (all time periods combined) observed (blue histograms) and predicted (red lines) catch-at-weight for all fisheries with weight samples for the diagnostic model.

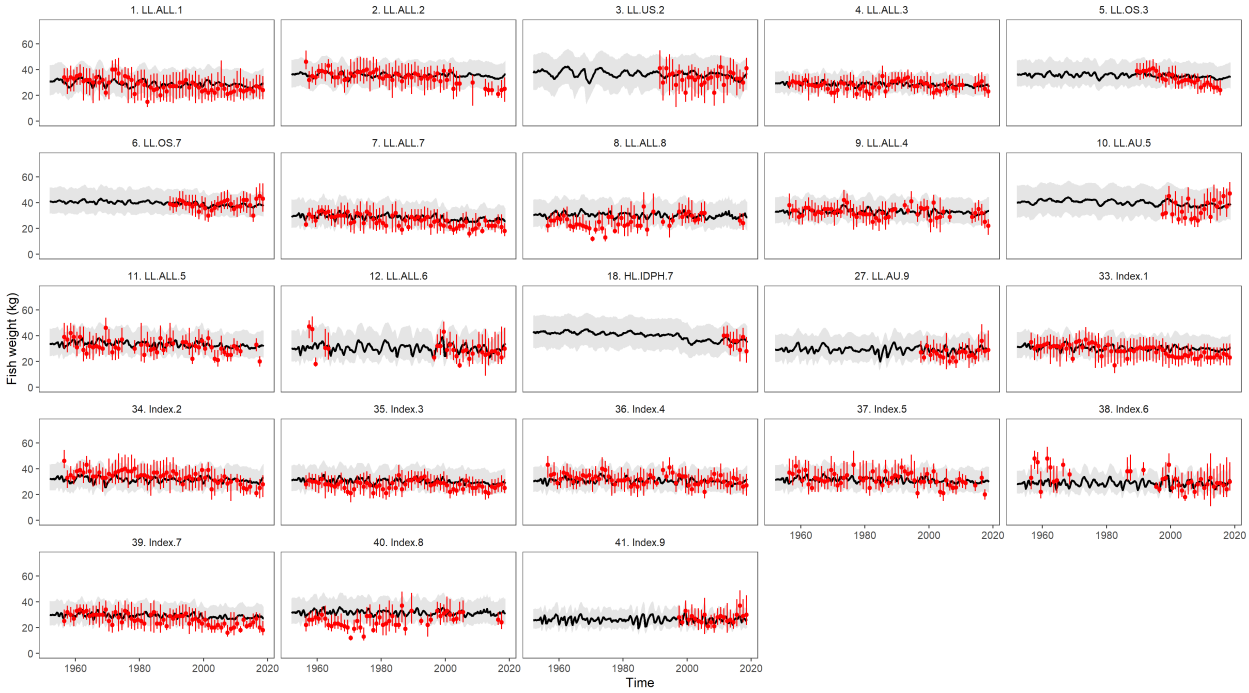


Figure 21: Comparison of the observed (red points) and predicted (grey line) median fish weight (kg) for all fisheries with weight samples for the diagnostic model. The uncertainty intervals (grey shading) represent the values encompassed by the 25% and 75% quantiles. Sampling data are aggregated by year and only weight samples with a minimum of 30 fish per year are plotted

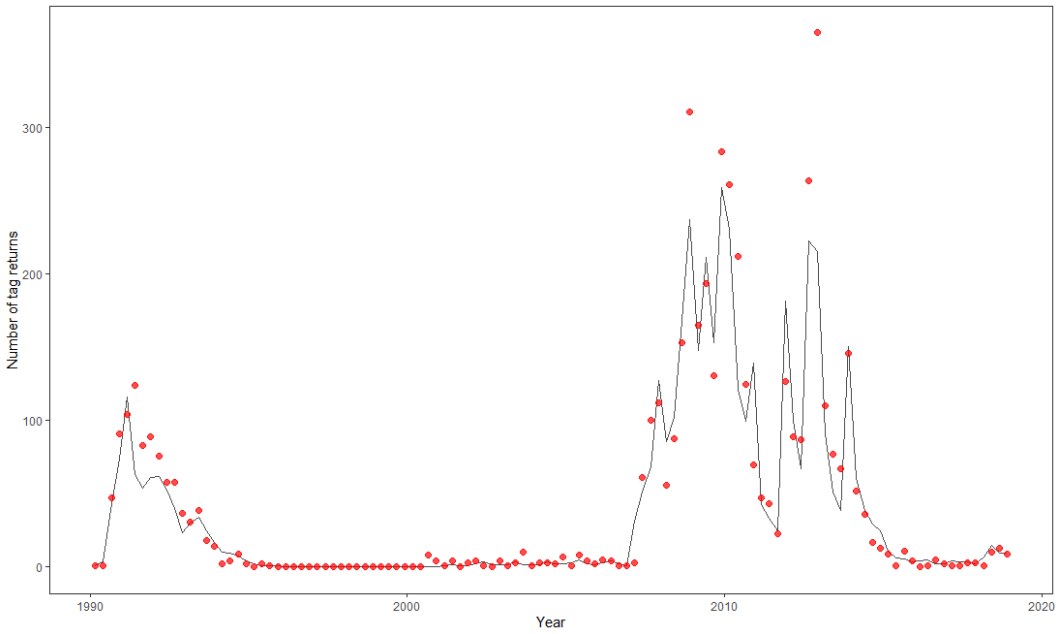


Figure 22: Observed (red dots) and model-predicted (black line) tag returns over time for the diagnostic model across all tag release events with all tag recapture groupings aggregated, but with recaptures within the mixing period excluded.

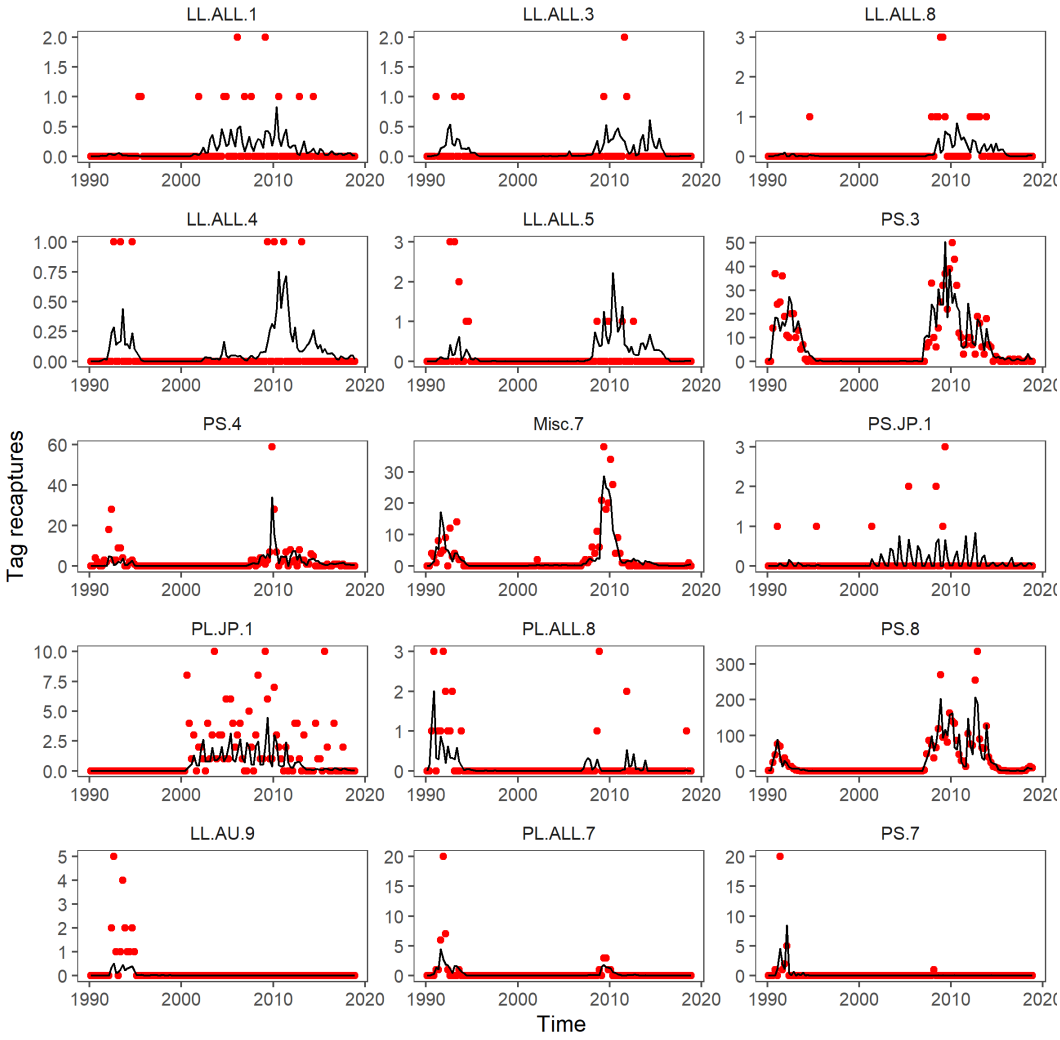
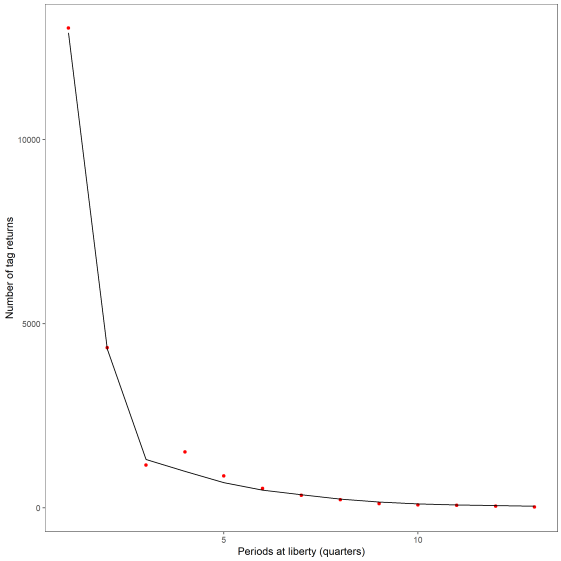
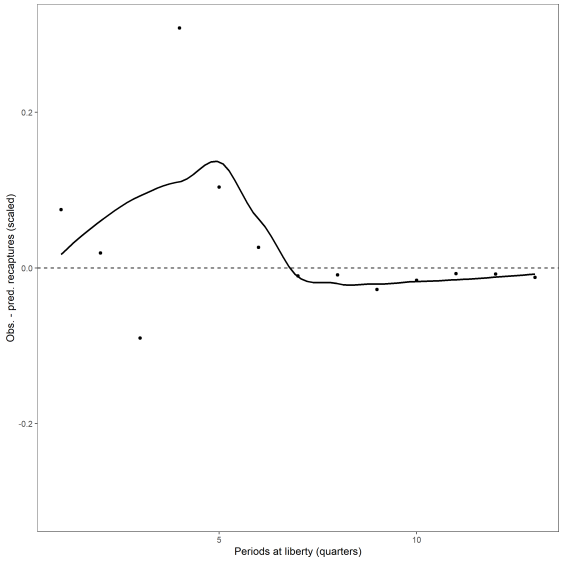


Figure 23: Observed (red dots) and model-predicted (black line) tag returns over time for the diagnostic model by recapture groups for fisheries with more than 10 tag returns after the mixing period.

(a) Observed and model-predicted tag attrition across release events.



(b) Scaled residuals.



(c) Log-scale observed and model-predicted tag attrition across release events.

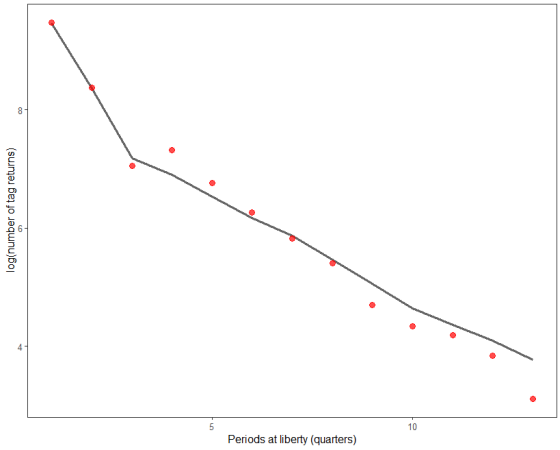
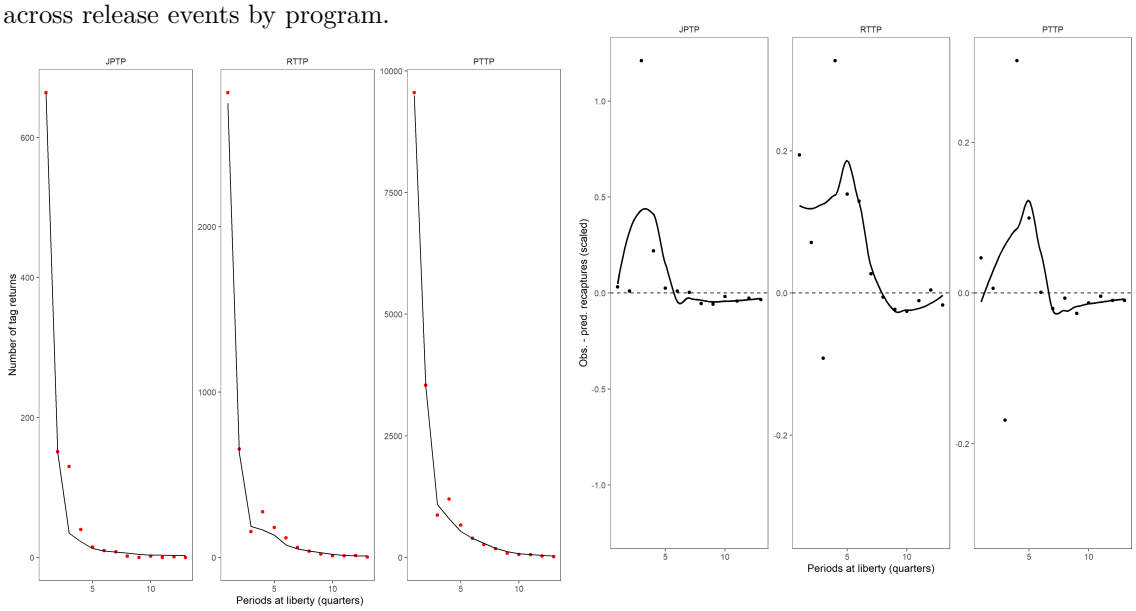


Figure 24: Observed and model-predicted (a) tag attrition, (b) scaled residuals, and (c) log-scale across all tag release events for the diagnostic model.

(a) Observed and model-predicted tag attrition across release events by program.

(b) Scaled residuals by program.



(c) Observed and model-predicted tag attrition across release events by region

(d) Scaled residuals by region.

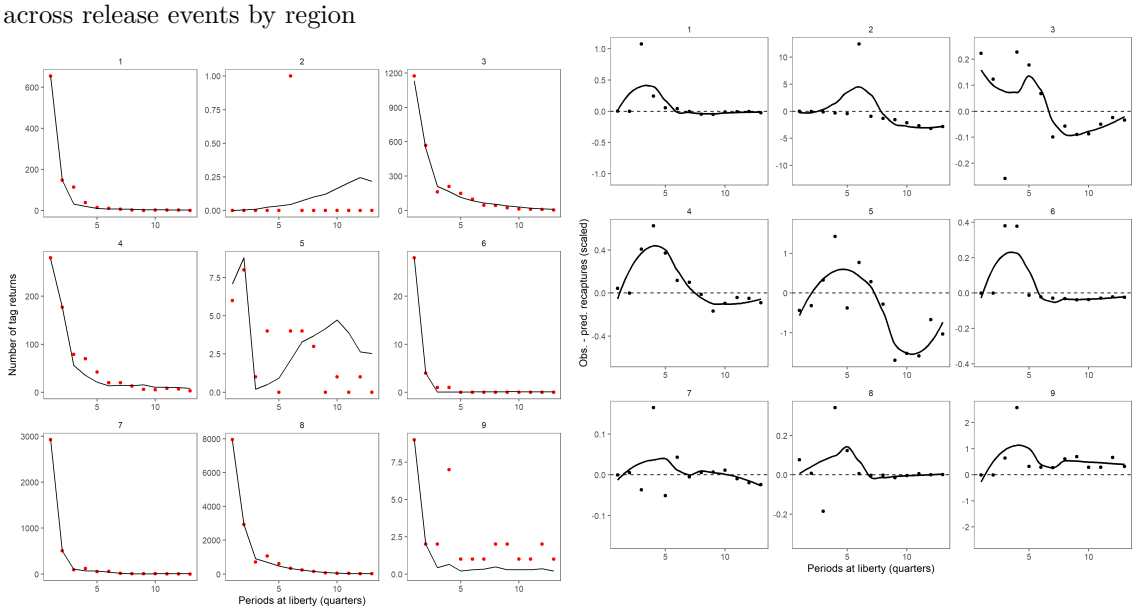


Figure 25: Observed and model-predicted tag attrition (a) program, (b) program scaled residuals, and (c) by region and (d) region residuals across all tag release events for the diagnostic model. Not the difference in the y-axis scale by panel.

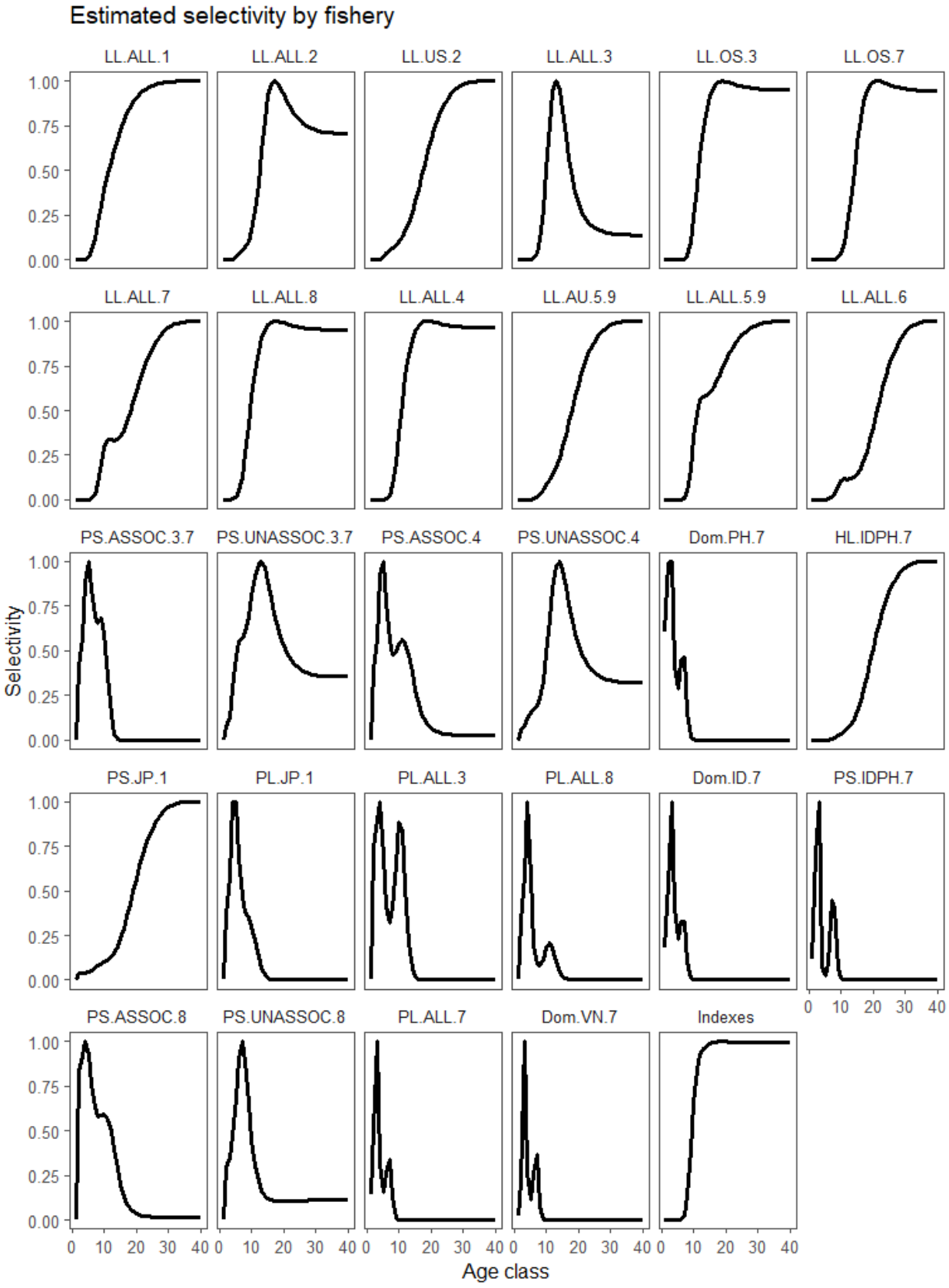


Figure 26: Estimated age-specific selectivity coefficients by fishery groupings for the diagnostic model.

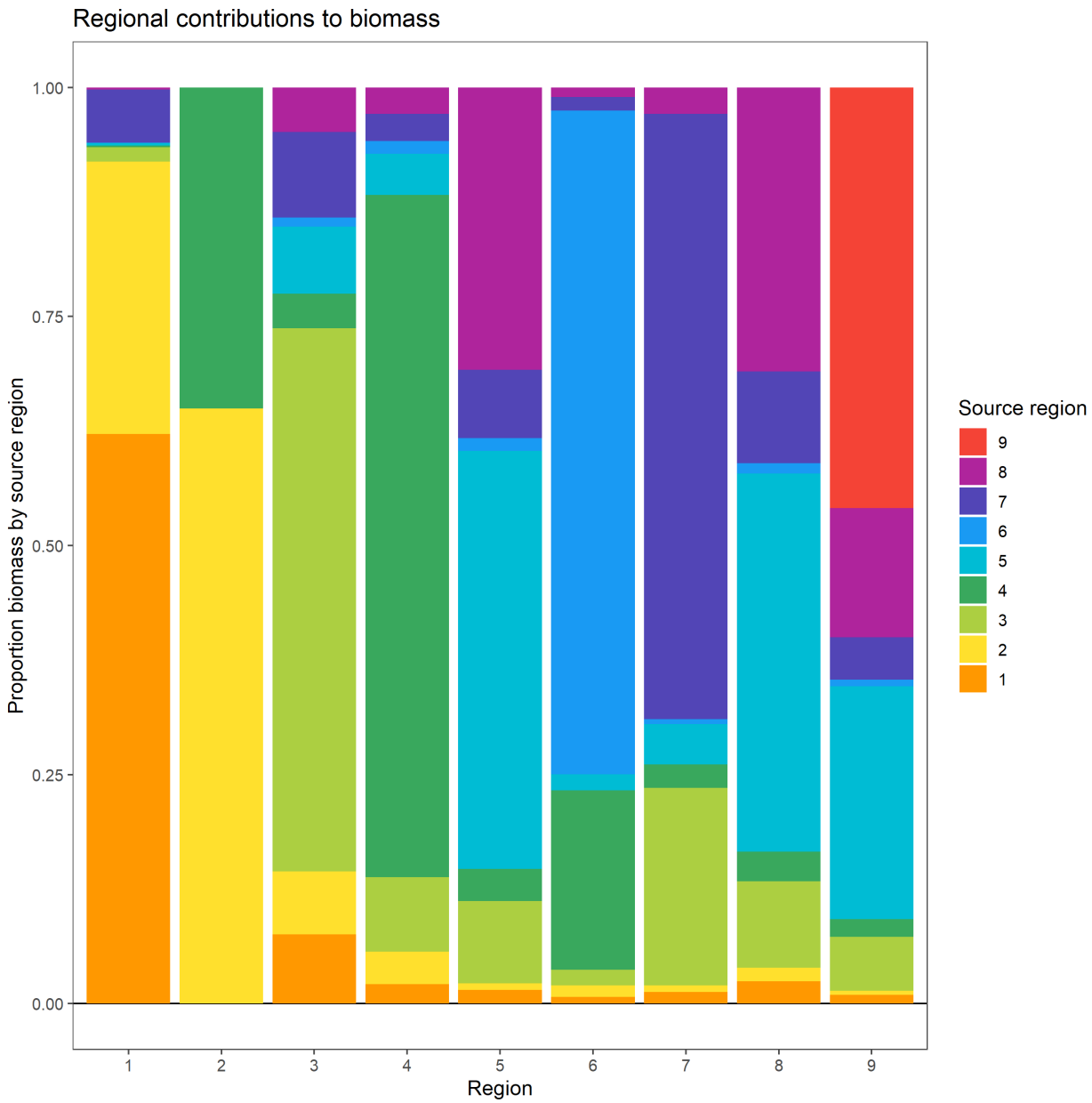


Figure 27: Proportional distribution of total biomass (by weight) in each region apportioned by the source region of the fish, for the diagnostic model. The colour of the originating region is presented in the legend to the right. The biomass distributions are calculated based on the long-term average distribution of recruitment between regions, estimated movement parameters, and natural mortality.

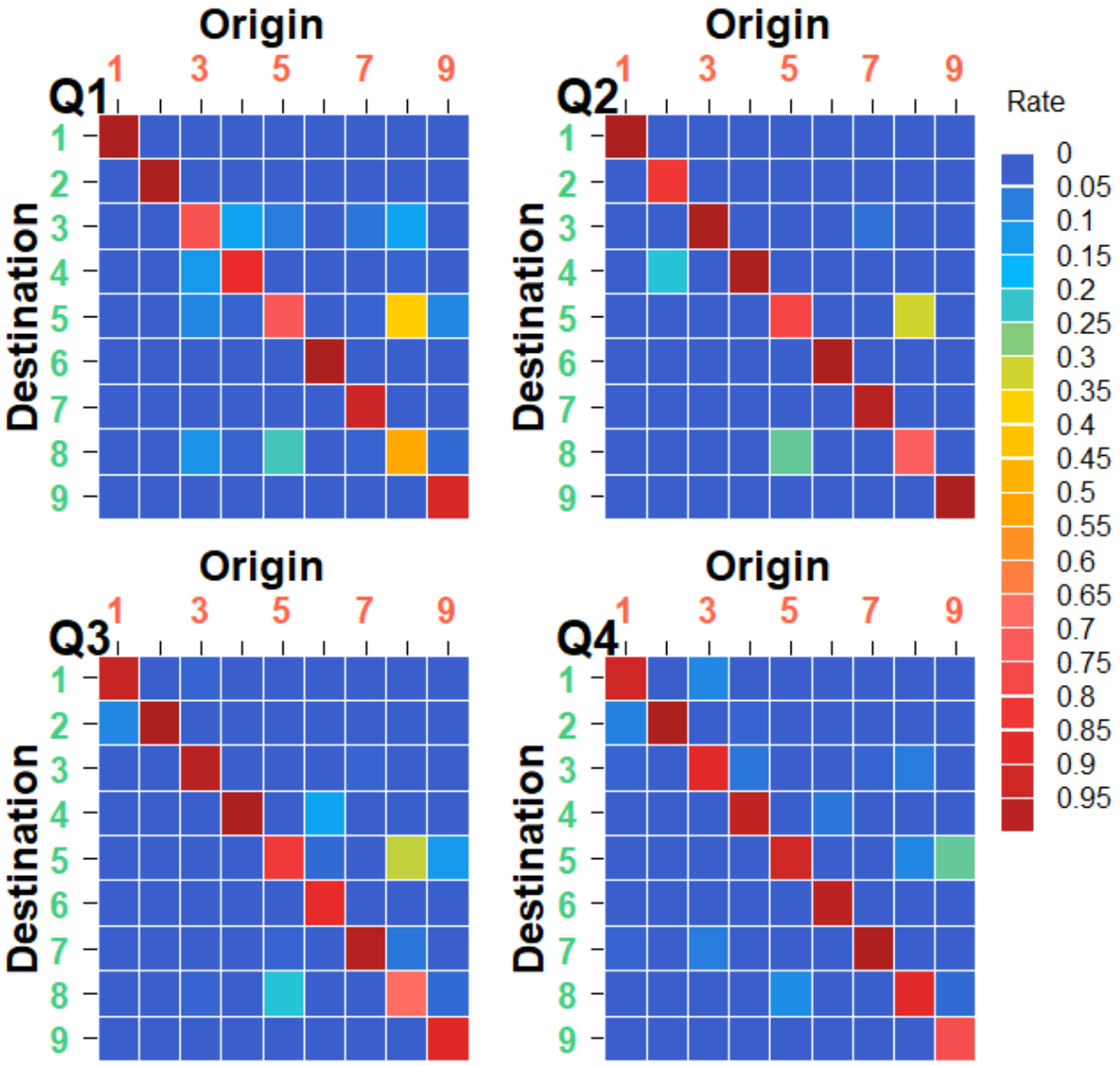


Figure 28: Estimated movement coefficients by quarter for the diagnostic case model. The red numbers (horizontal axis) indicate the source model region; the green numbers (vertical axis) indicate the receiving regions. The color of the tile shows the magnitude of the movement rate (proportion of individuals from region x moving to region y in that quarter), with each column adding up to 1 and higher proportions are more red.

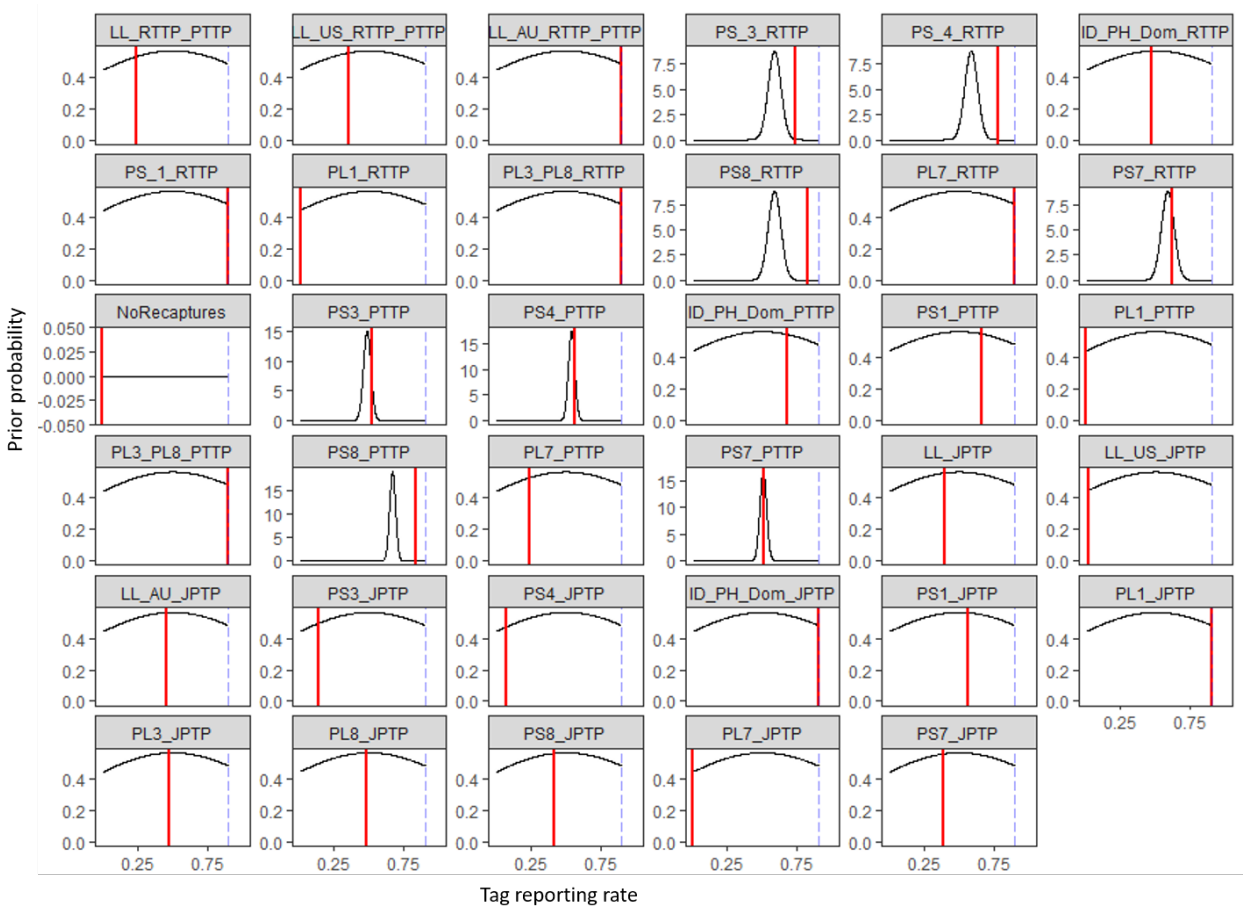
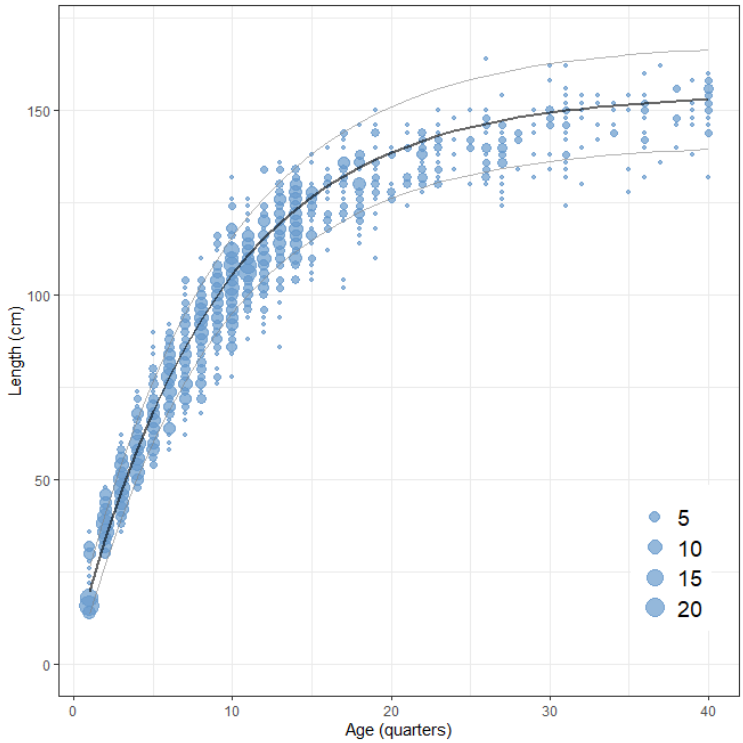


Figure 29: Tag reporting rate parameters estimated in the diagnostic model (vertical red lines) in relation to prior penalty distributions (black lines) for each fishery group. Blue dotted lines indicate the upper bound for the reporting rate parameter of 0.9.

(a) Conditional age-at-length fit



(b) Residuals to conditional age-at-length

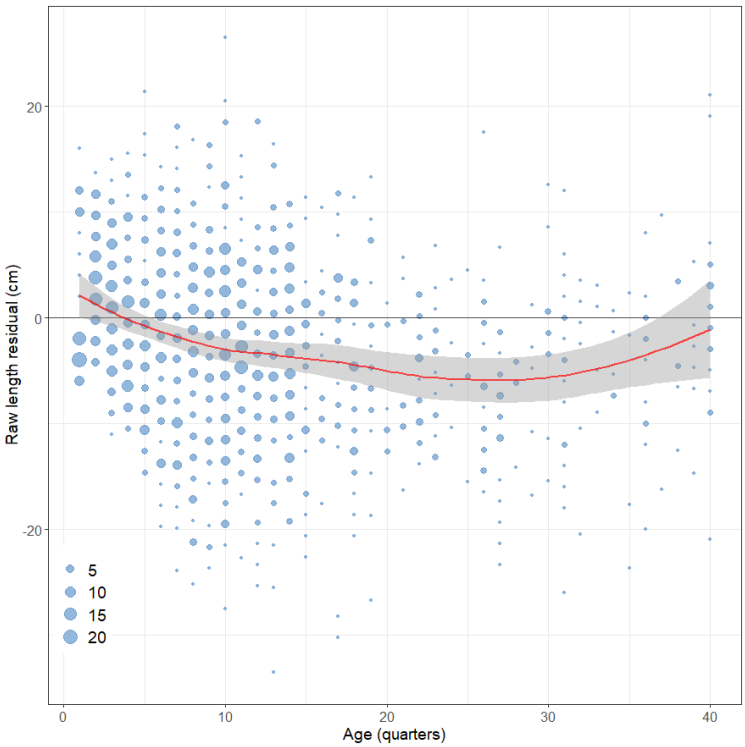


Figure 30: (a) Fit to conditional age-at-length (circle size indicate number of samples) and estimated growth functions for the diagnostic model. (b) Plot showing the raw residuals (observed lengths – mean length estimated by the MFCL growth function) across the range of age-classes included in the assessment model. The size of the points indicates the number of fish in that age and length bin and the red line indicates a lowess smoother fit to the data weighted by the sample sizes.

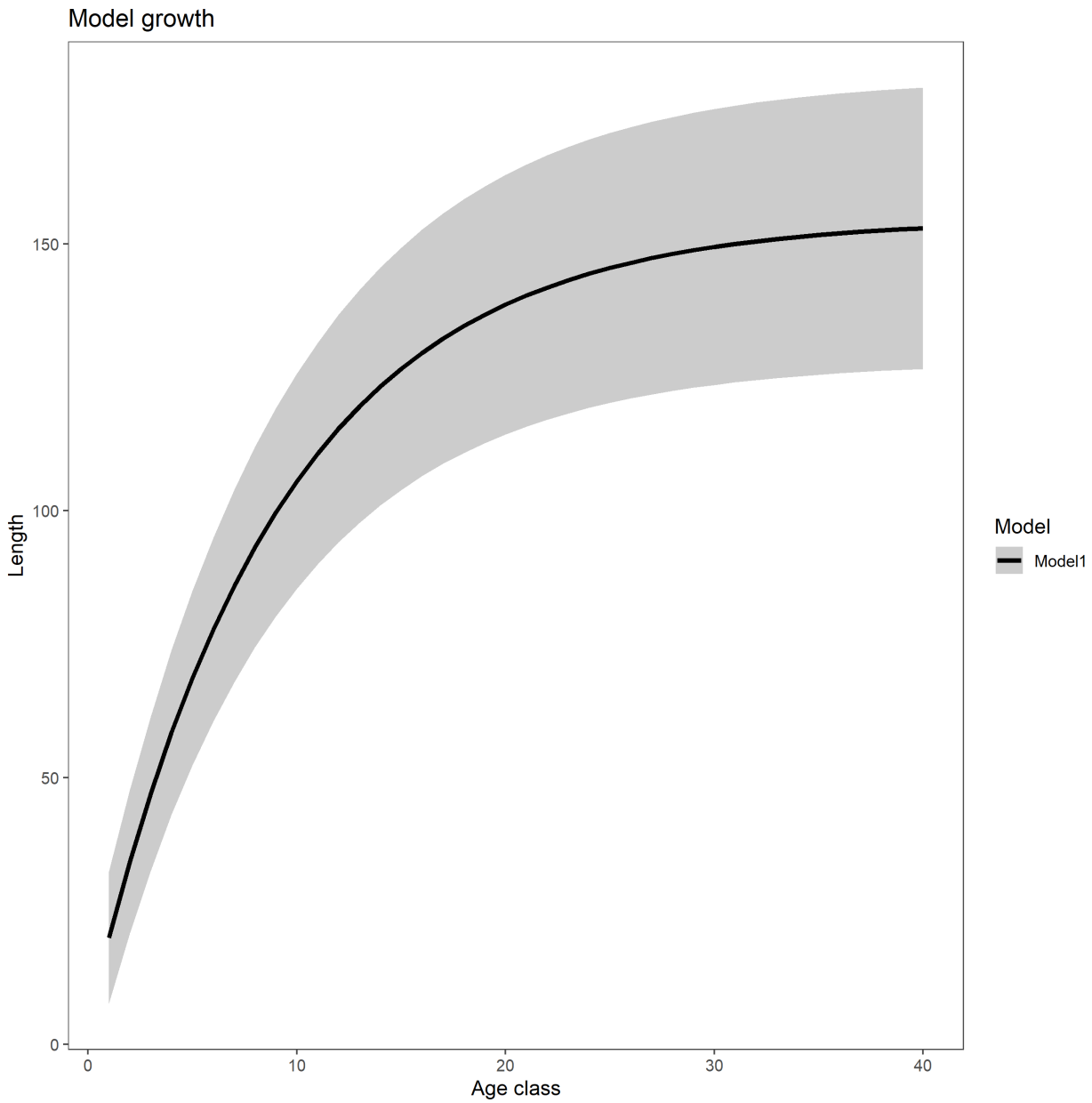


Figure 31: Plot of estimated growth curve from the diagnostic model where the black line represents the estimated mean fork length (cm) at-age and the grey region represents the length-at-age within one standard deviation of the mean.

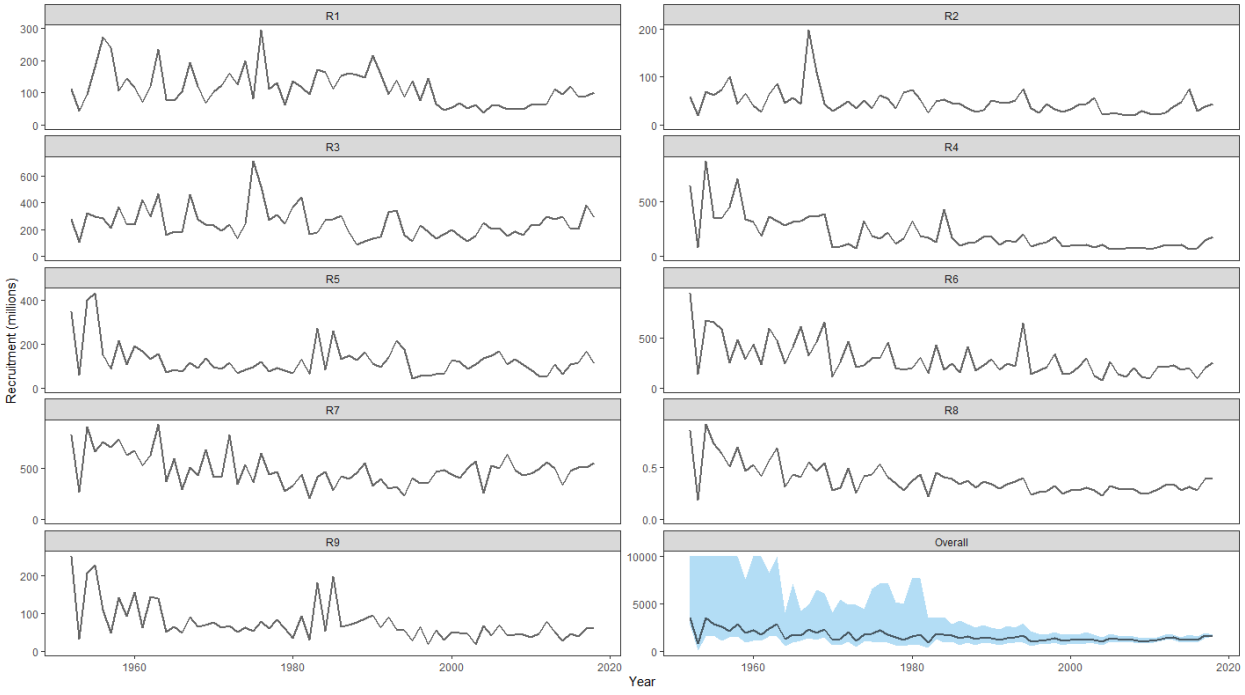
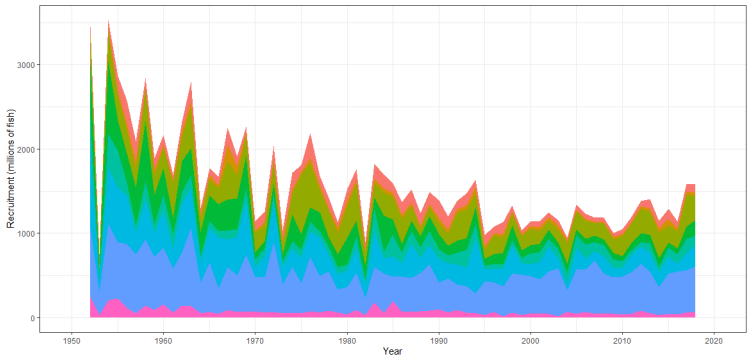
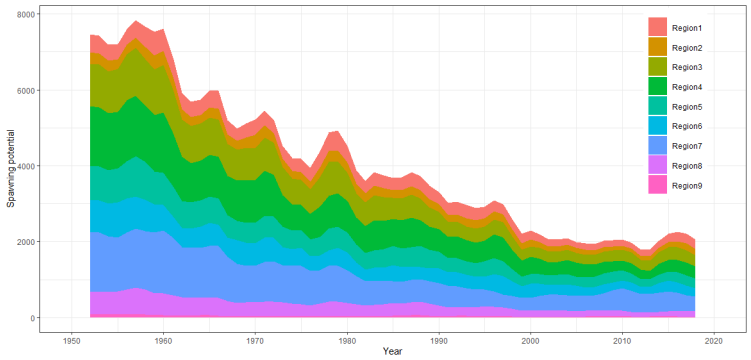


Figure 32: Estimated annual recruitment by model region and across all regions for the diagnostic model, where the blue shaded region for the overall recruitment shows the estimated 95% confidence interval based on statistical uncertainty estimated for the diagnostic model. Note that the scale of the y-axis is not constant across regions and that the very uncertain estimates from the first few years are truncated.

(a) Recruitment



(b) Spawning potential



(c) Total biomass

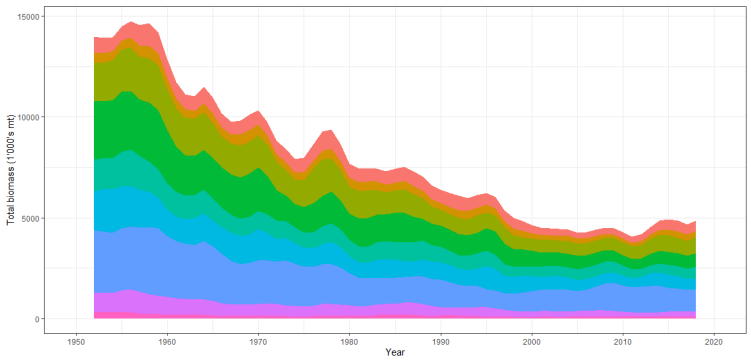


Figure 33: Estimated annual average, (a) recruitment (b) spawning potential (c) total biomass by model region for the diagnostic model, showing the relative sizes among regions.

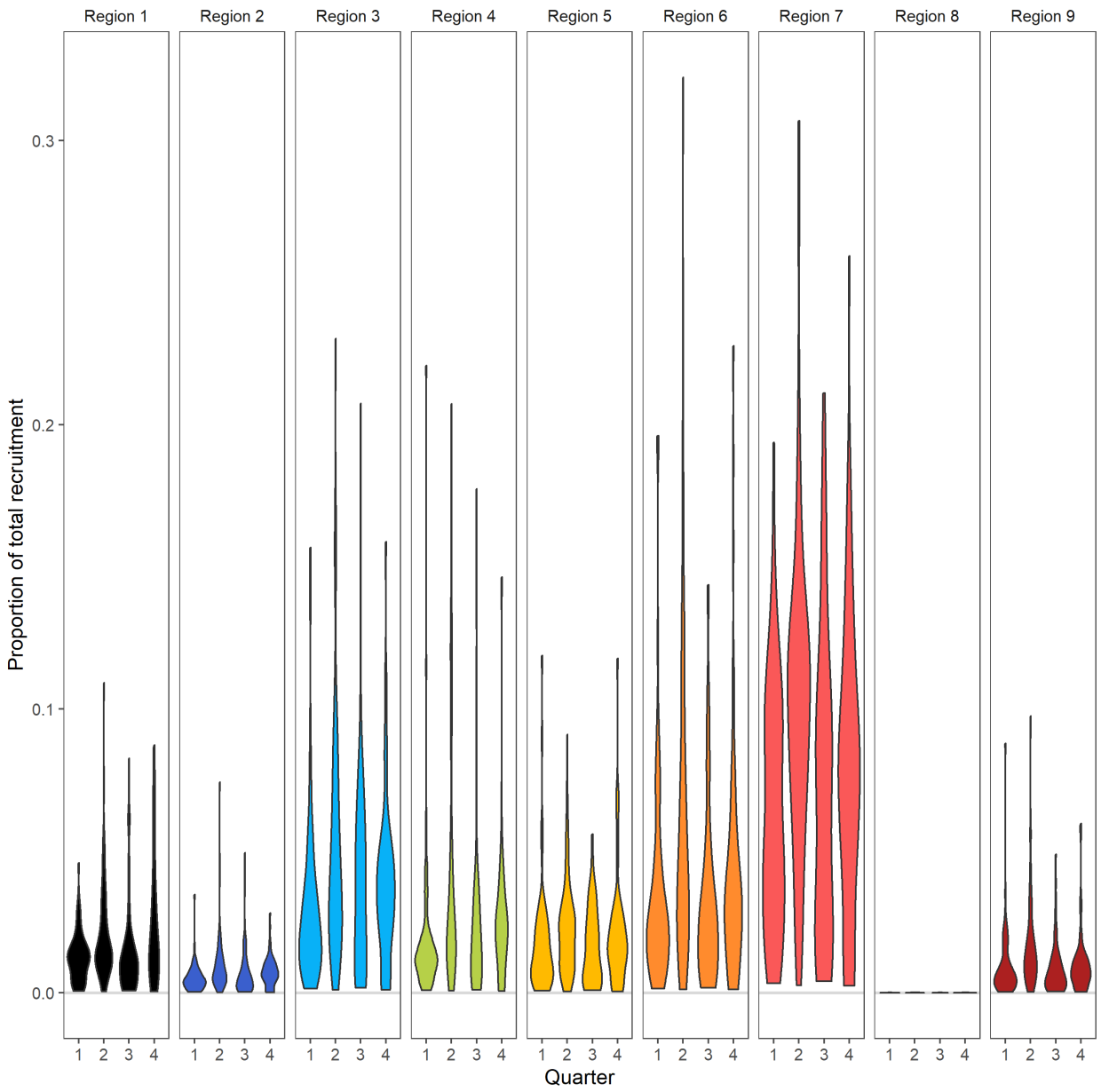


Figure 34: Distribution of annual proportions of recruitment by region (panels) and quarter (within panel group) for all years in the diagnostic model.

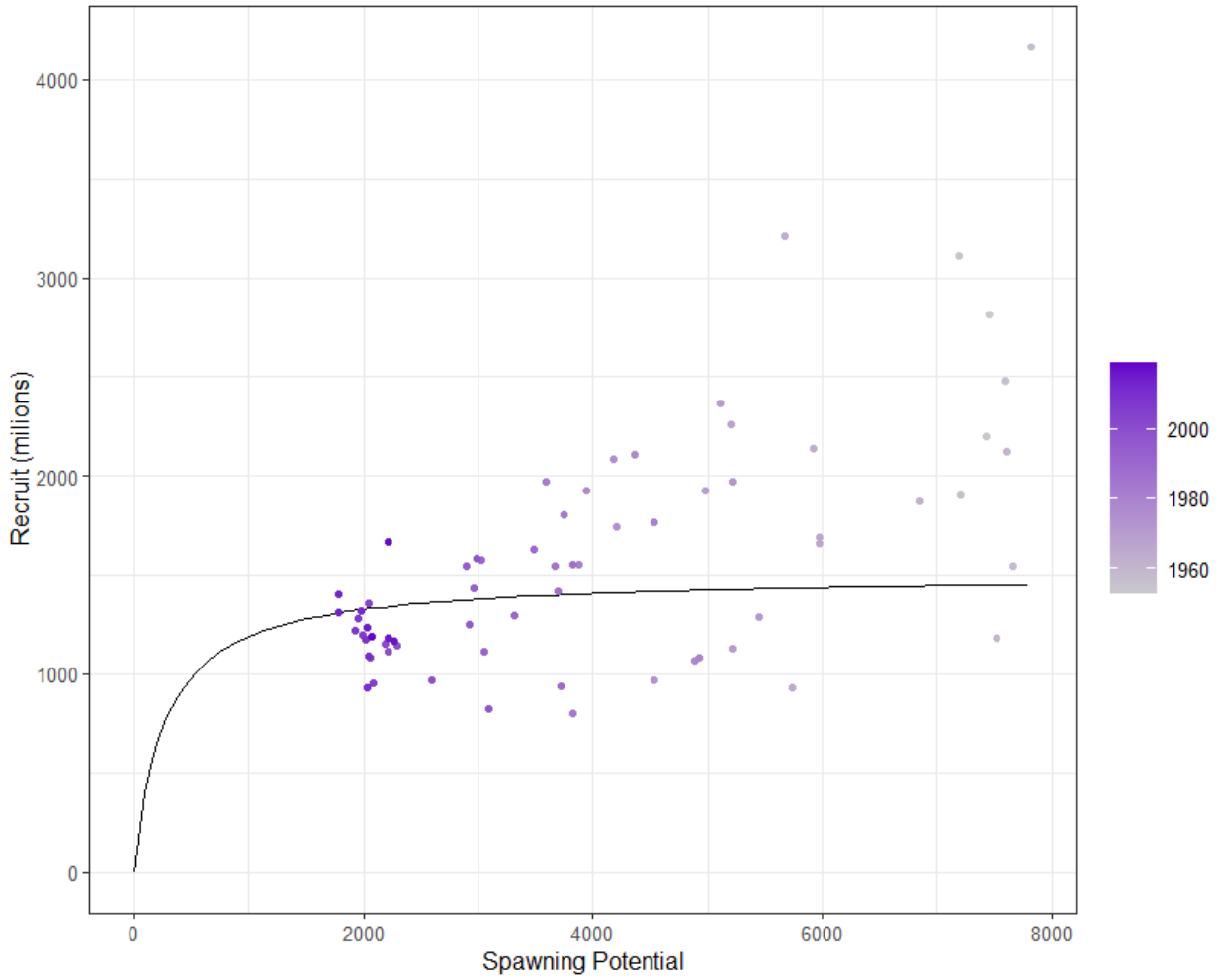


Figure 35: Estimated relationship between recruitment and spawning potential based on annual values for the diagnostic model. The darkness of the circles changes from light to dark through time.

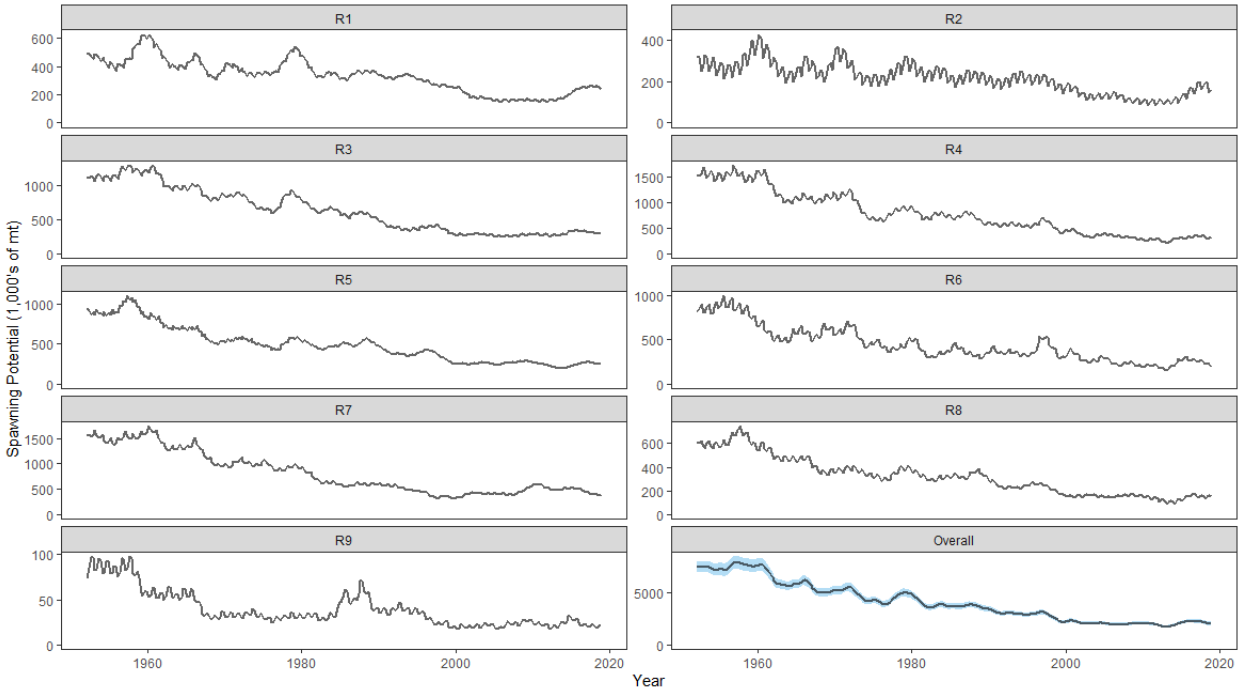


Figure 36: The temporal trend in estimated spawning potential by model region for the diagnostic model, where the blue shaded region for the overall spawning potential shows the estimated 95% confidence interval based on statistical uncertainty estimated for the diagnostic model. Note that the y-axis scale among panels are not consistent.

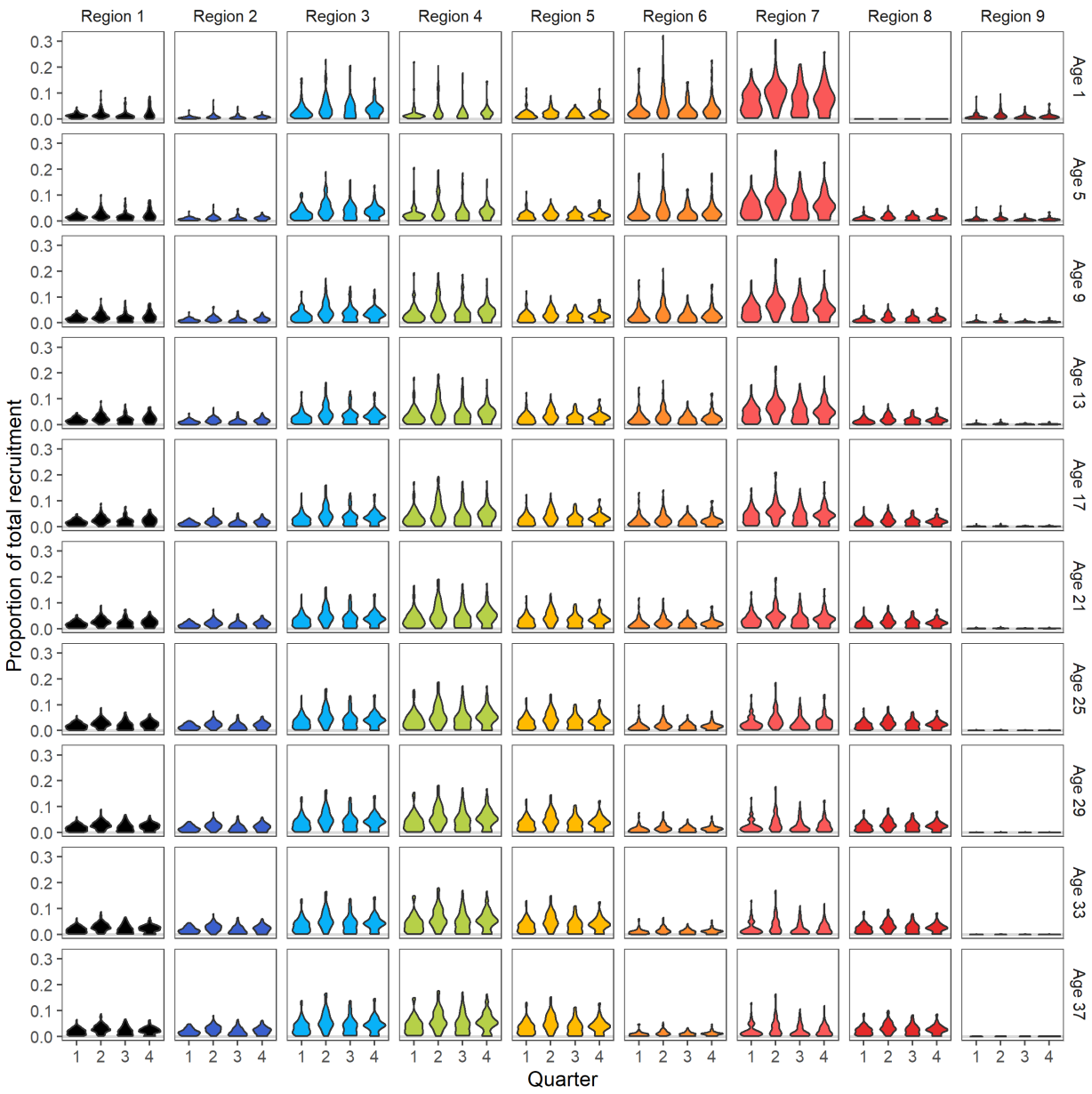


Figure 37: Distribution of annual proportion of selected ages (rows of panels) by region (columns of panel) and quarter (within panel groups) for all years estimated by the diagnostic model.

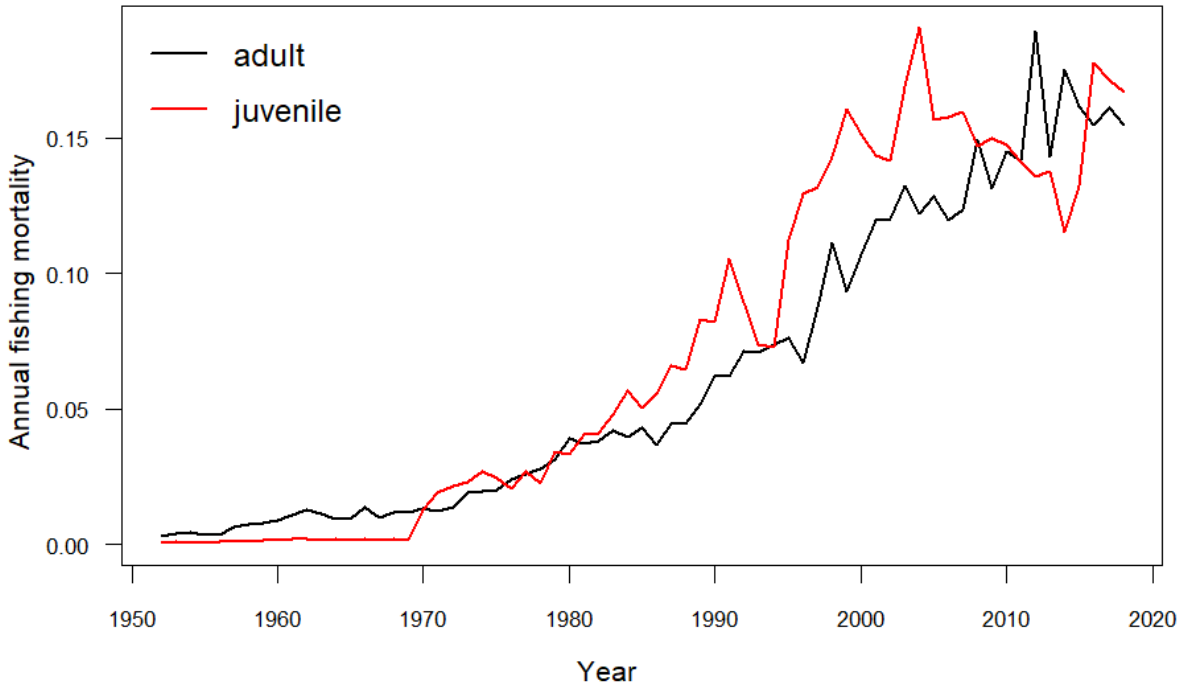


Figure 38: Estimated annual average juvenile and adult fishing mortality for the diagnostic model.

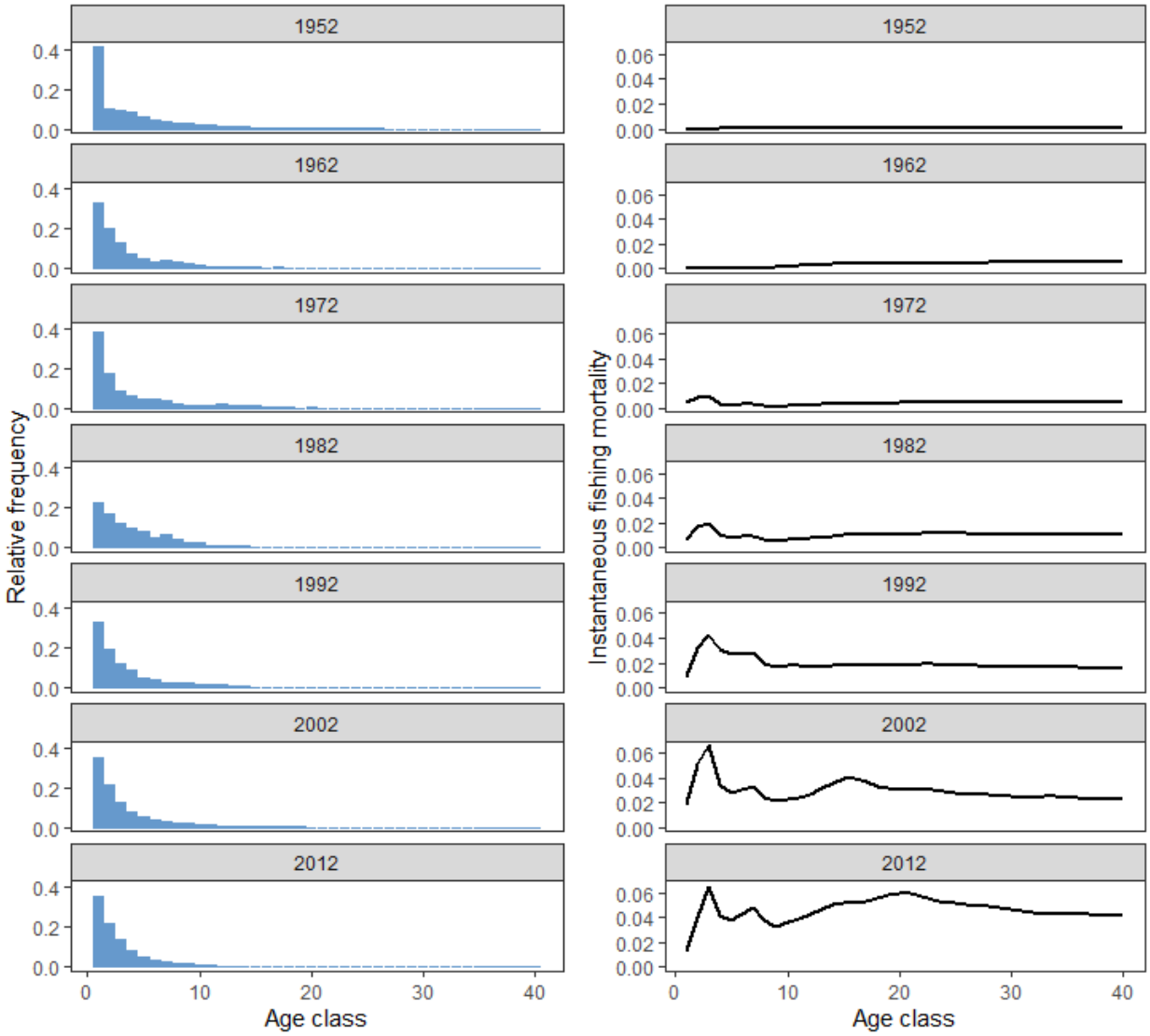


Figure 39: Estimated proportion of the population at-age (quarters; left panels) and fishing mortality-at-age (right panels), at decadal intervals, for the diagnostic model.

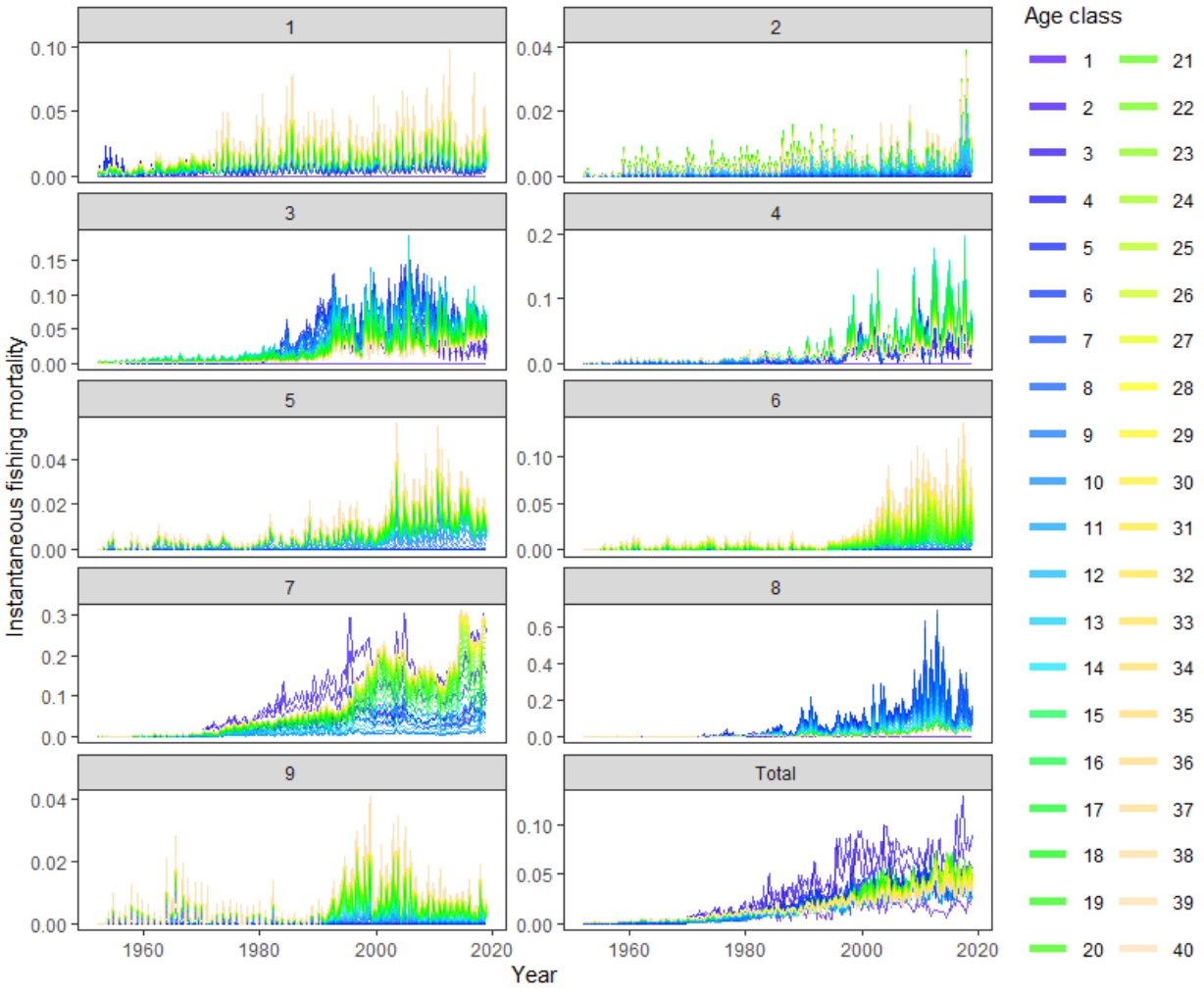
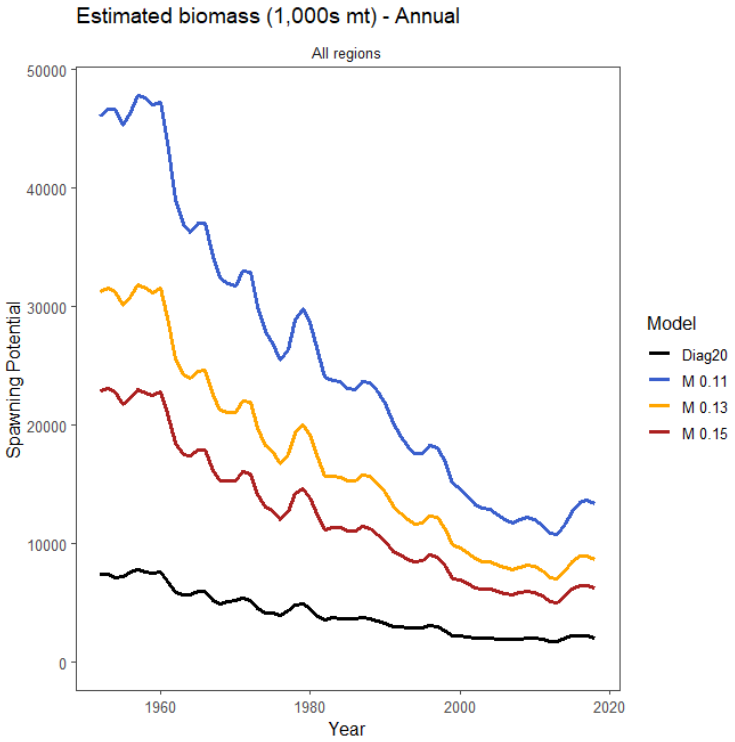


Figure 40: Estimated age-specific instantaneous fishing mortality for the diagnostic model, by region and overall.

(a) Spawning potential of sensitivities to M



(b) Depletion of sensitivities to M

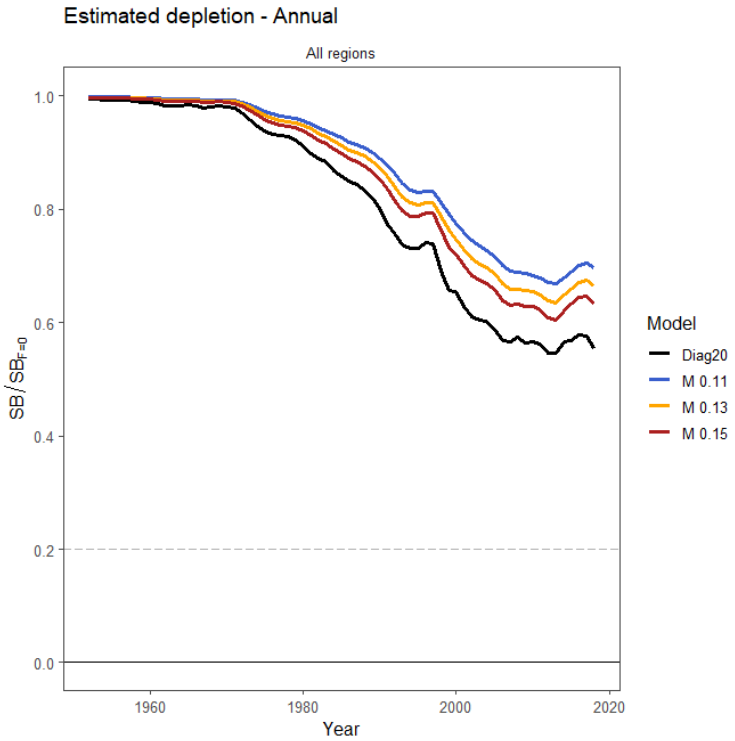
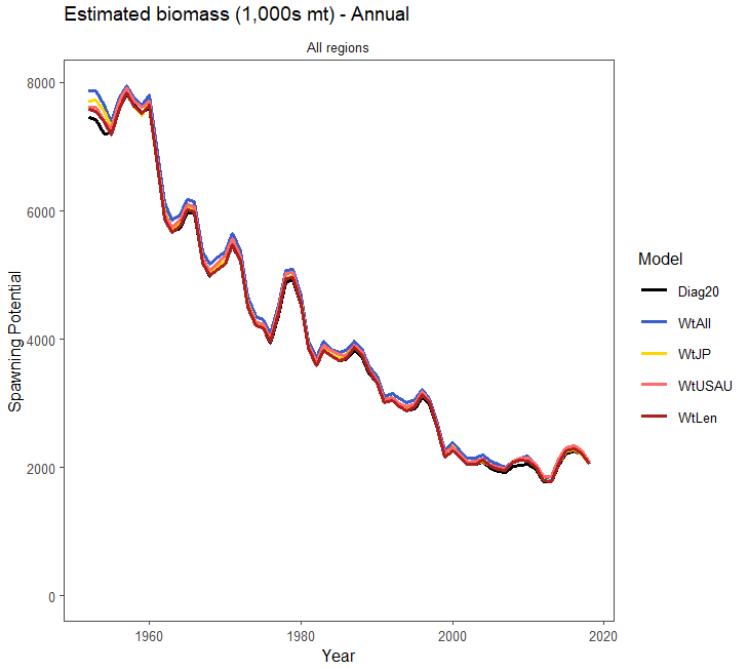


Figure 41: Spawning potential and adult depletion for sensitivities to the diagnostic model with different mean natural mortality values.

(a) Spawning potential of sensitivities to size composition data used



(b) Depletion of sensitivities to size composition data used

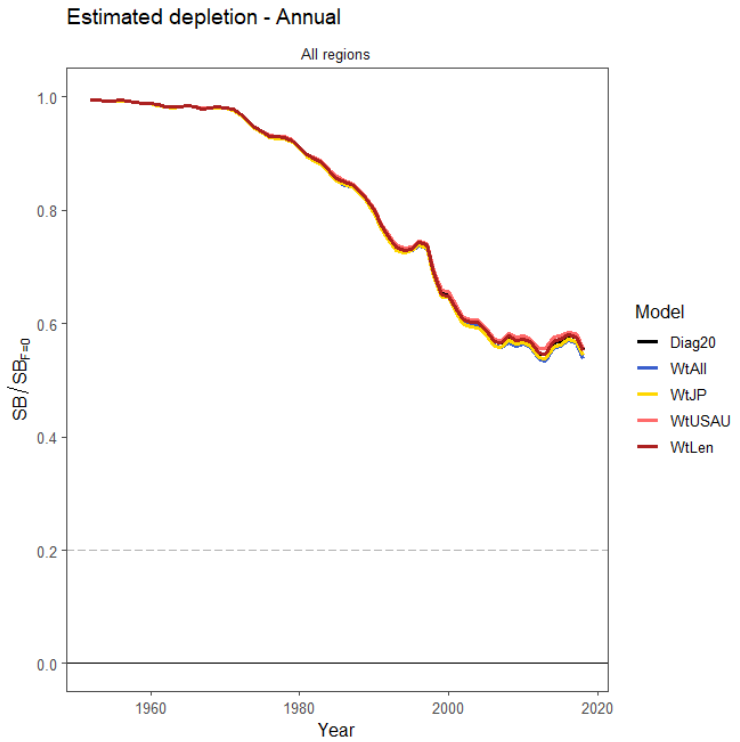
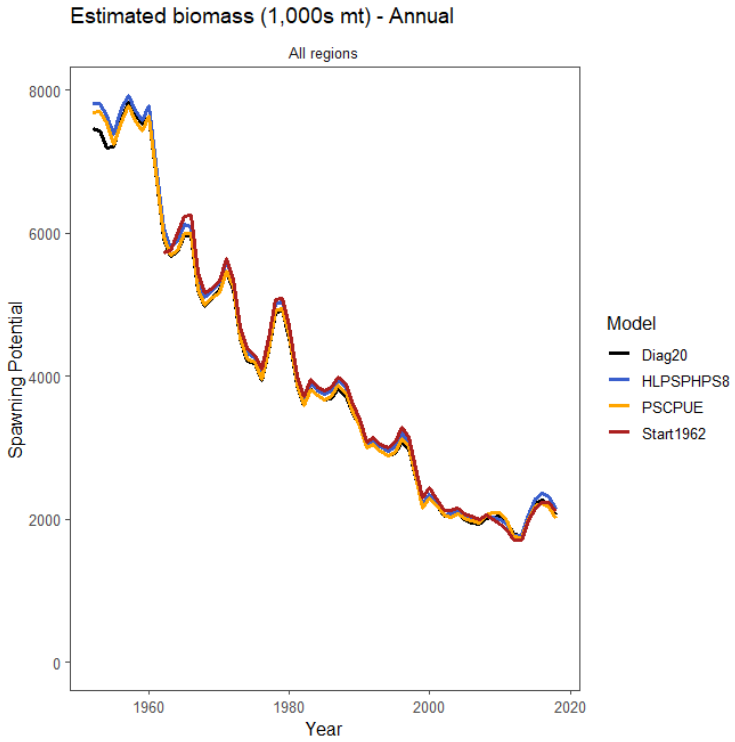


Figure 42: Spawning potential and adult depletion for sensitivities to the diagnostic model where Diag20 is the 2020 diagnostic model, WtAll is using all available weight samples in the index fishery, WTJP is using only Japanese weight data in the index fishery, WTUSAU is adding in weight samples from the US and Australia, and WtLen is using weight and length composition in recent years for regions 4 -6..

(a) Spawning potential of sensitivities to additional CPUE indices



(b) Depletion of sensitivities to additional CPUE indices

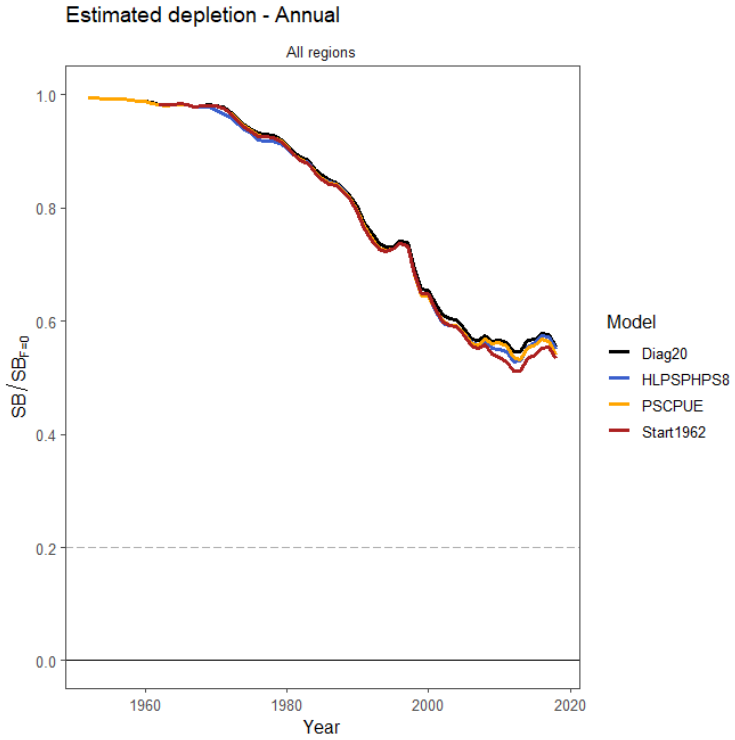
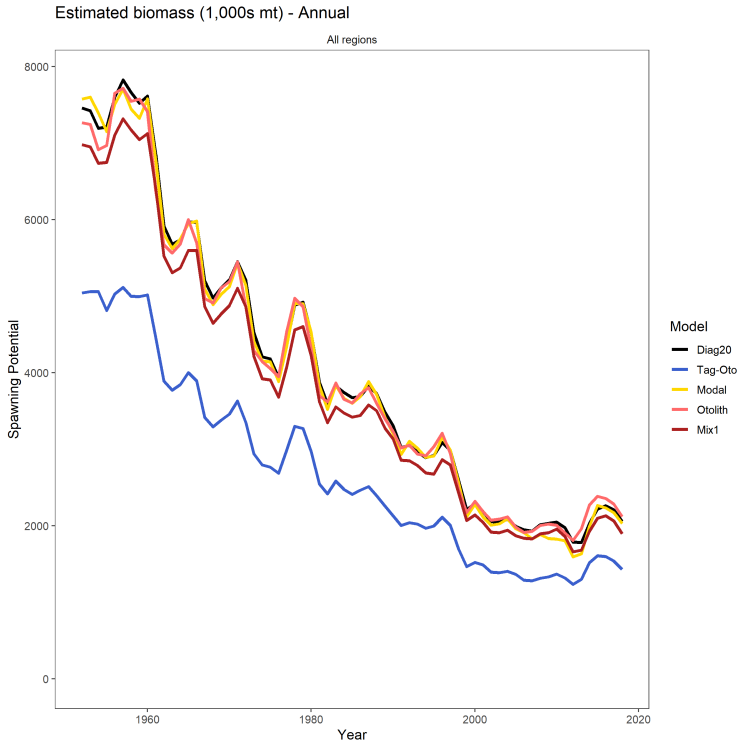


Figure 43: Spawning potential and adult depletion for sensitivities to the diagnostic model with different CPUE indices included where HLPSPHPS8 is using the indices for the PHID handline, the PHID purse seine and the region 8 associated PS, PSCPUE is using the purse seine CPUE from (Vidal and Hamer, 2020), and Start1962 is starting the assessment model in 1962.

(a) Spawning potential of sensitivities to growth curves and mixing period



(b) Depletion of sensitivities to growth curves and mixing period

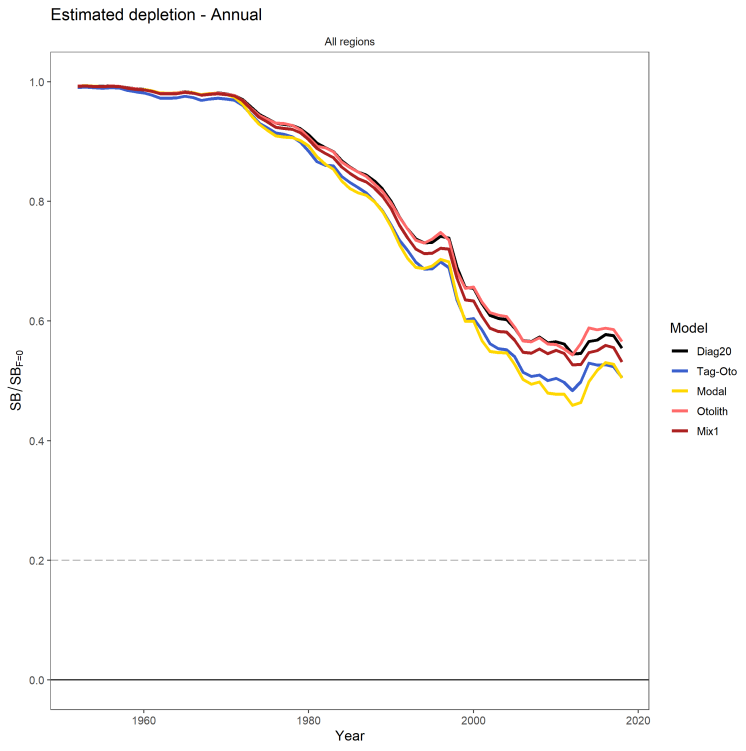


Figure 44: Spawning potential and adult depletion for sensitivities to the diagnostic model with different growth curves where Diag20 is the 2020 diagnostic model, Tag-Oto is the tag otolith integrated growth curve, Model is the growth curve estimated from the size composition, Otolith is the Richards otolith only growth curve and Mix is assuming a mixing period of 1 quarter.

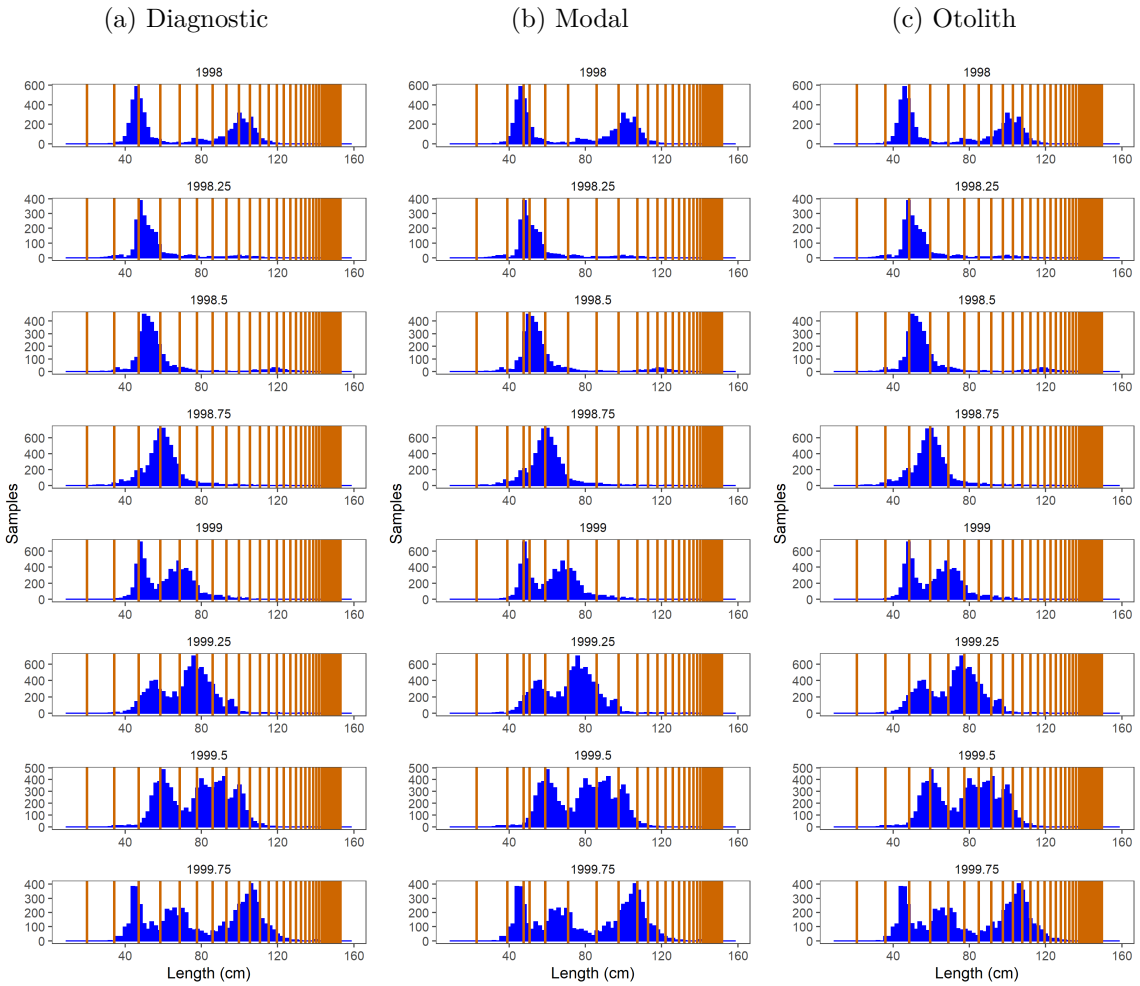


Figure 45: Observed quarterly length composition data from the associated purse seine in region 4 between 1998-1999 (blue bars) and predicted mean length at age (orange lines) from the different growth curves, where Diagnostic is estimated from the conditional age-at-length data, Modal was estimated from the modal progressions in the size composition data, and Otolith is the otolith only Richards growth curve.

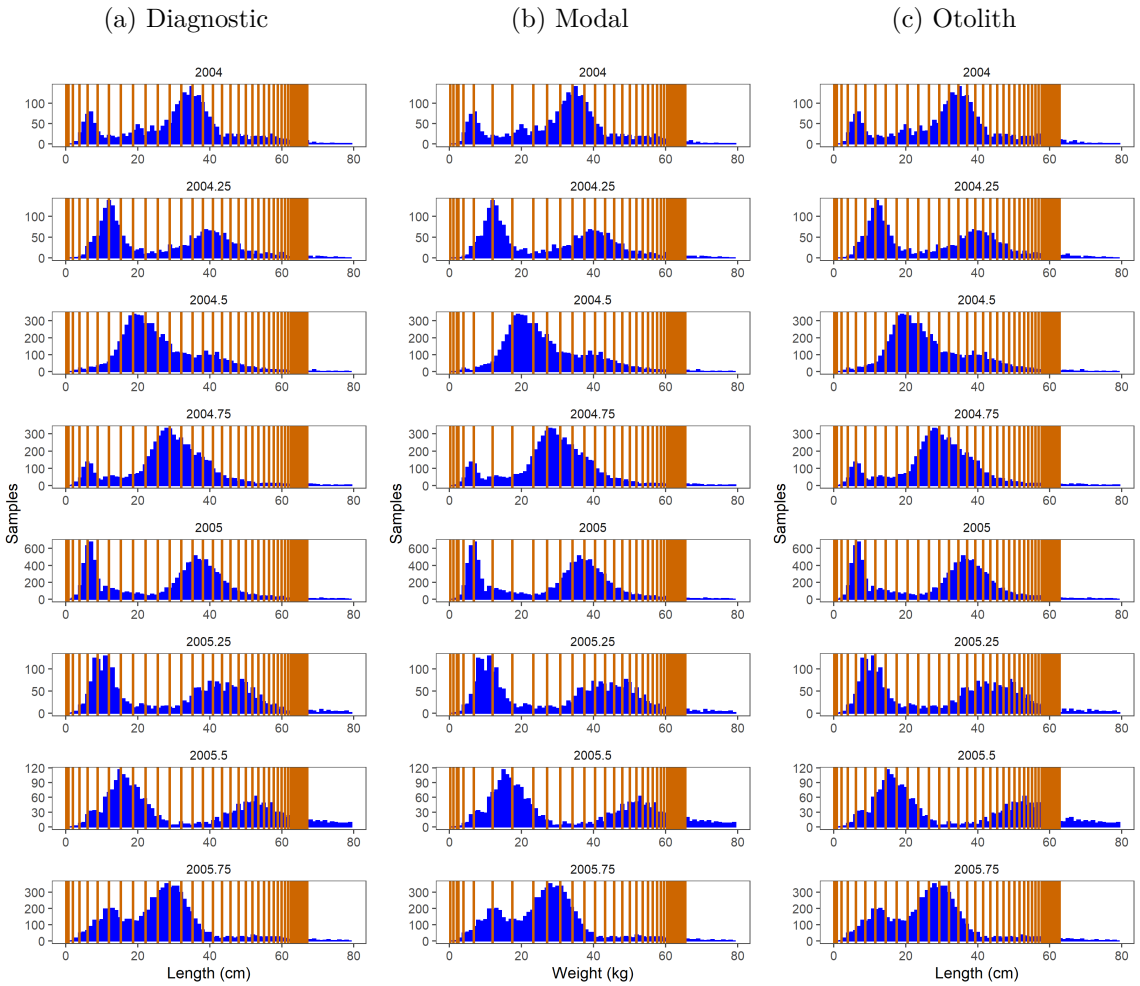
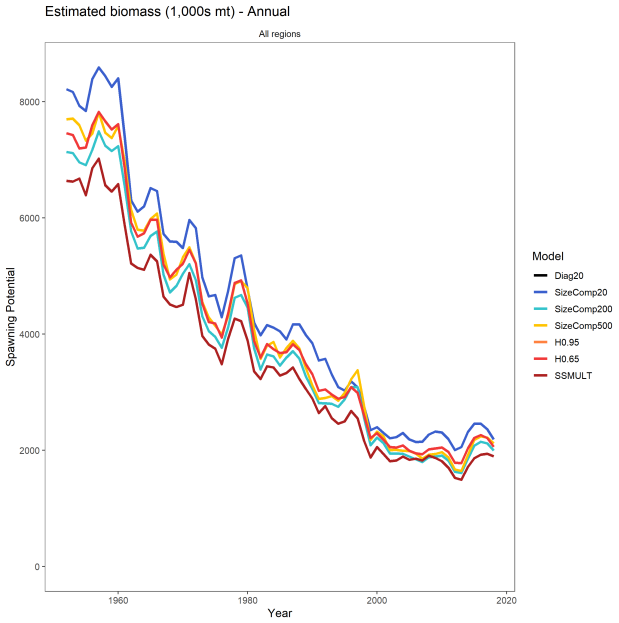


Figure 46: Observed quarterly distributions of weight composition data from the US.LL.3 between 2004-2005 (blue bars) and predicted mean length at age (orange lines) from the different growth curves, where Diagnostic is estimated from the conditional age-at-length data, Modal was estimated from the modal progressions in the size composition data, and Otolith is the otolith only Richards growth curve.

(a) Spawning potential of sensitivities to size composition weightings and steepness



(b) Depletion of sensitivities to size composition weightings and steepness

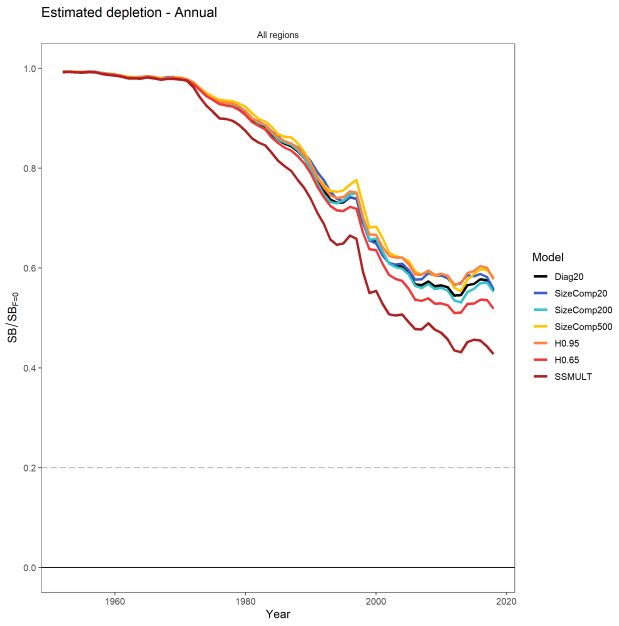
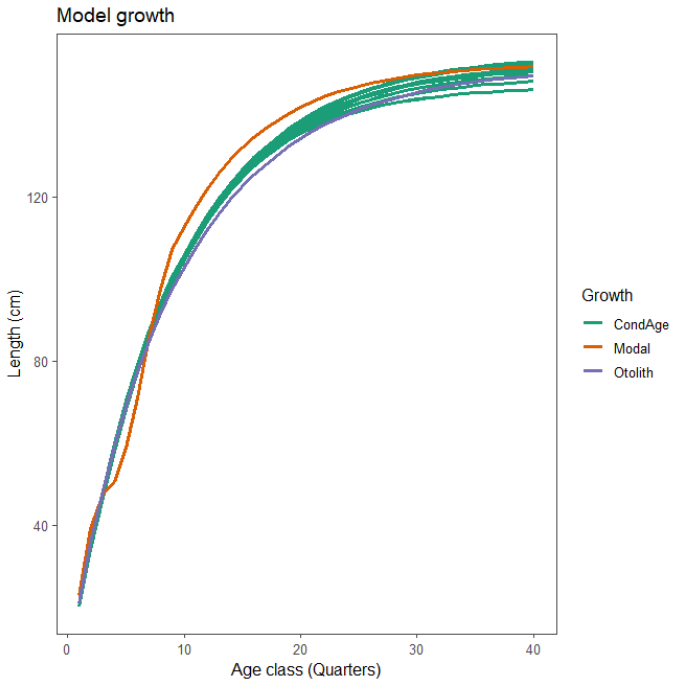


Figure 47: Spawning potential and adult depletion for sensitivities to the diagnostic model where Diag 20 is the diagnostic model, SizeComp20 is assuming a size composition scalar of 20, SizeComp200 is assuming a size composition scalar of 200, H0.95 is a steepness of 0.95, H0.65 is a steepness of 0.65, and SSMULT is a model using the self-scaling multinomial distribution (SSMULT) for the size composition.

(a) Growth



(b) Maturity

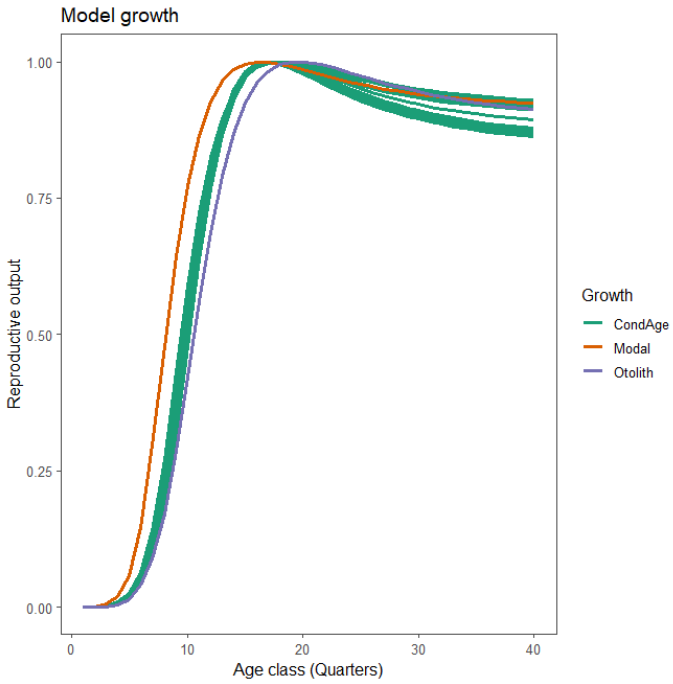


Figure 48: Growth curves and Maturity at age used in the structural uncertainty grid. Note that growth is fixed for the Otolith and Modal axis and thus maturity is also constant across models, but growth is estimated in CondAge models.

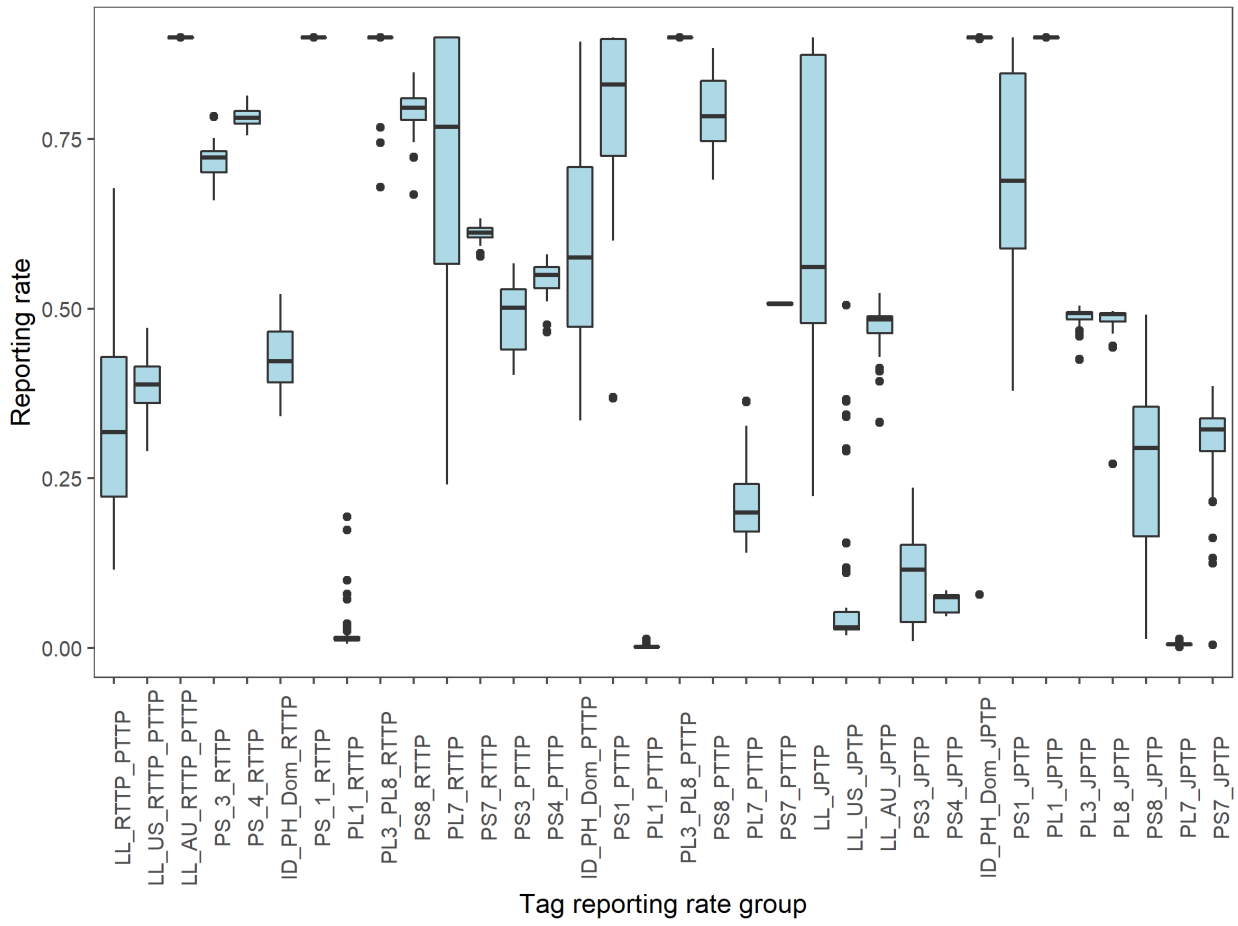
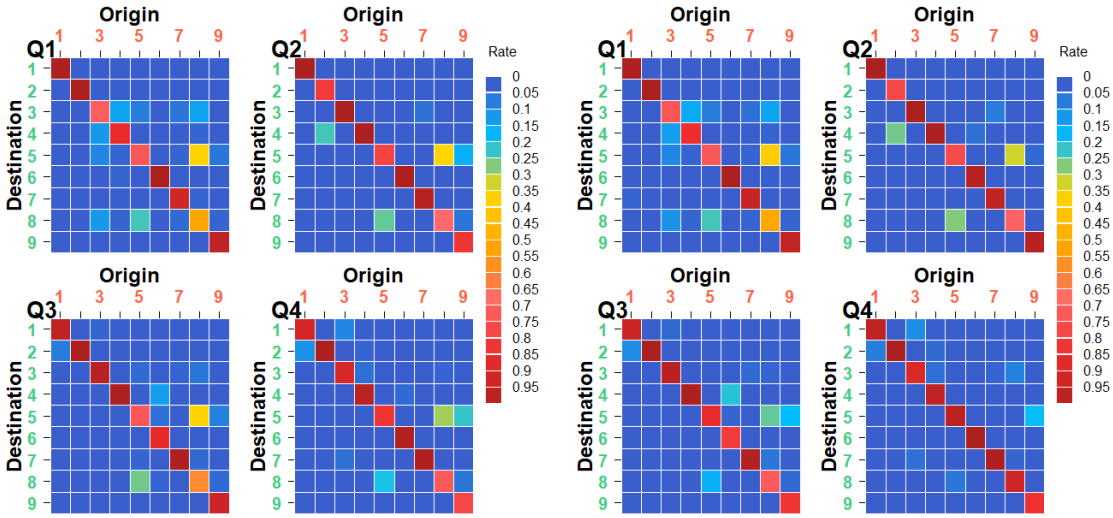


Figure 49: Boxplots of reporting rate parameters estimated across the structural uncertainty grid.

(a) Modal Growth

(b) Otolith Growth



(c) Size Composition 20

(d) Size Composition 500

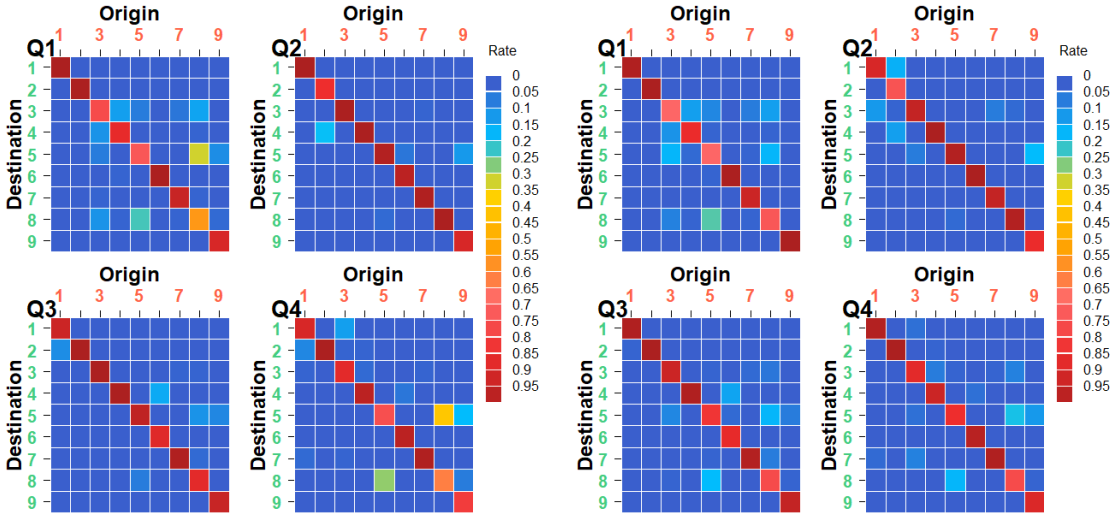


Figure 50: Movement rates estimated by models with different assumed growth functions or size composition scalars, where (a) is growth estimated from the modal progressions of size composition data and (b) is the external otolith growth curve (c) is a model with the size composition scalar of 20 and (d) is a model with the size composition scalar of 500.

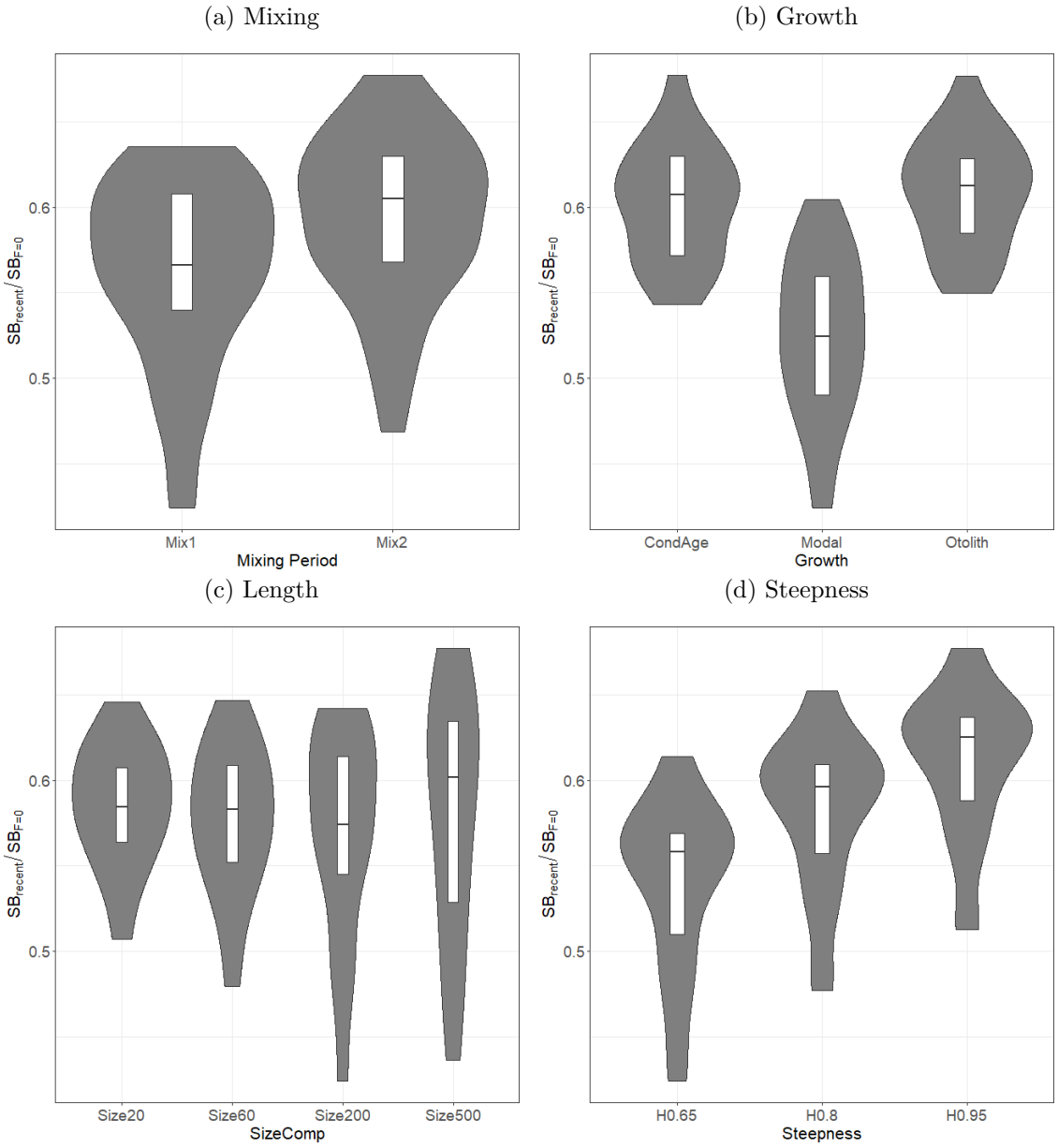


Figure 51: Box and violin plots summarizing the estimated $SB_{recent}/SB_{F=0}$ for each of the models in the structural uncertainty grid. The line in the box is the median of the estimates, while the box shows the 50^{th} percentile. The shaded area shows the probability distribution (or density) of the estimates of all models across the axes of the structural uncertainty grid.

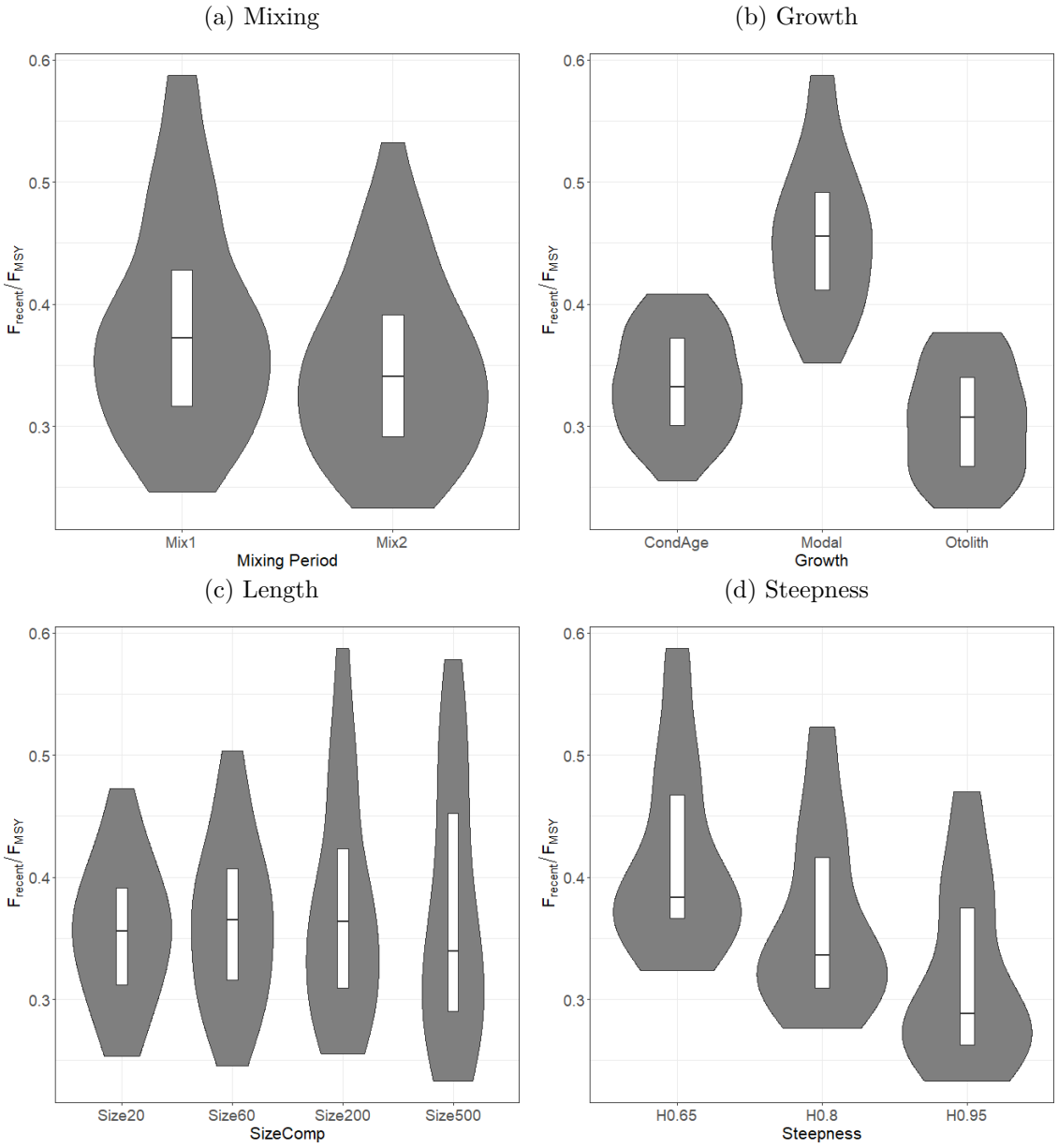
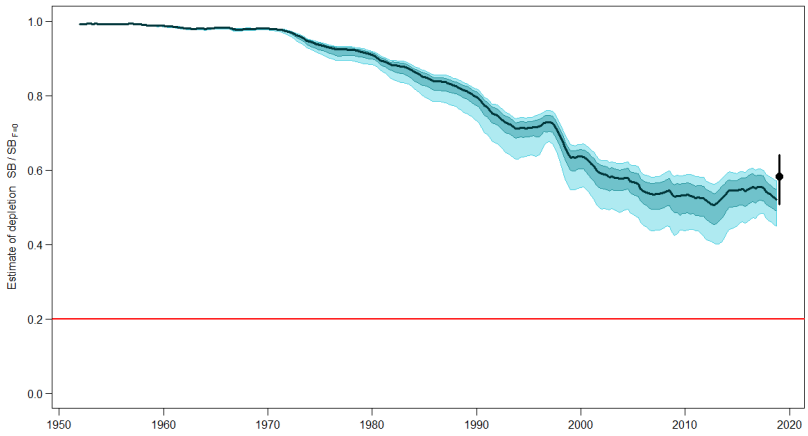


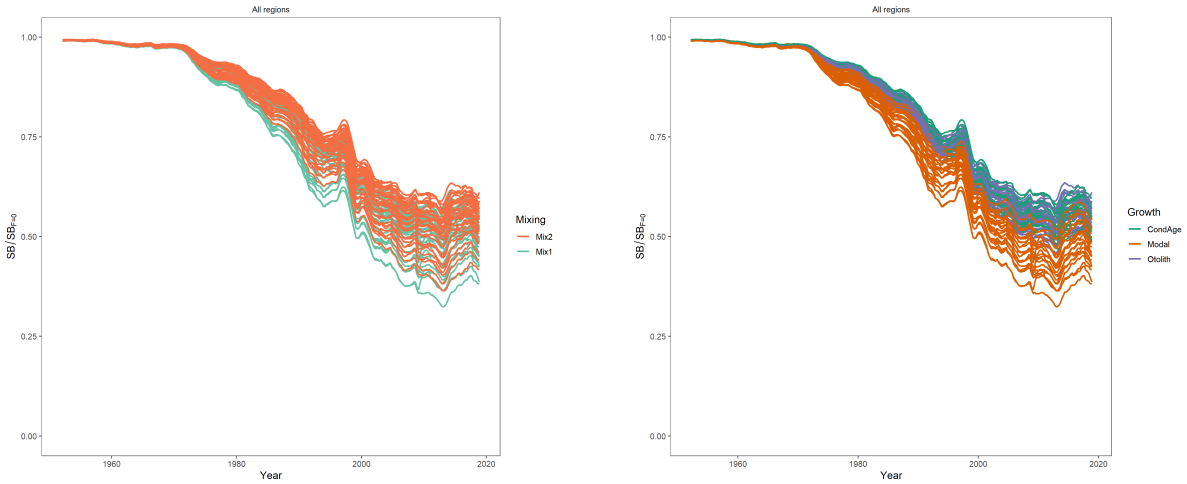
Figure 52: Box and violin plots summarizing the estimated $F_{recent}/F_{\text{MSY}}$ for each of the models in the structural uncertainty grid. The line in the box is the median of the estimates, while the box shows the 50^{th} percentile. The shaded area shows the probability distribution (or density) of the estimates of all models across the axes of the structural uncertainty grid.

(a) Depletion Quantiles



(b) Mixing

(c) Growth



(d) Size Composition

(e) Steepness

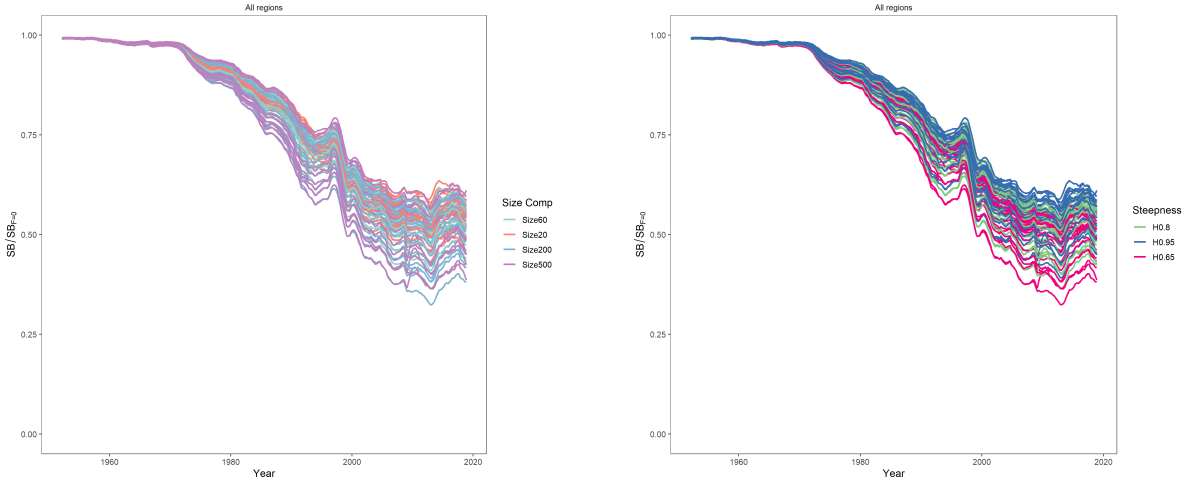


Figure 53: Plots showing the trajectories of fishing depletion of spawning potential for the models in the structural uncertainty grid (see Section 6.3 for details of the structure of the grid models). Panel 53a shows the median, 50% quantile, and 80% quantile of instantaneous depletion across the structural uncertainty grid and the point and error bars is the median and 80% quantile of estimates of $SB_{recent}/SB_{F=0}$. The remaining 4 panels show the axes used in the grid, with the color denoting the level within the axes for each model.

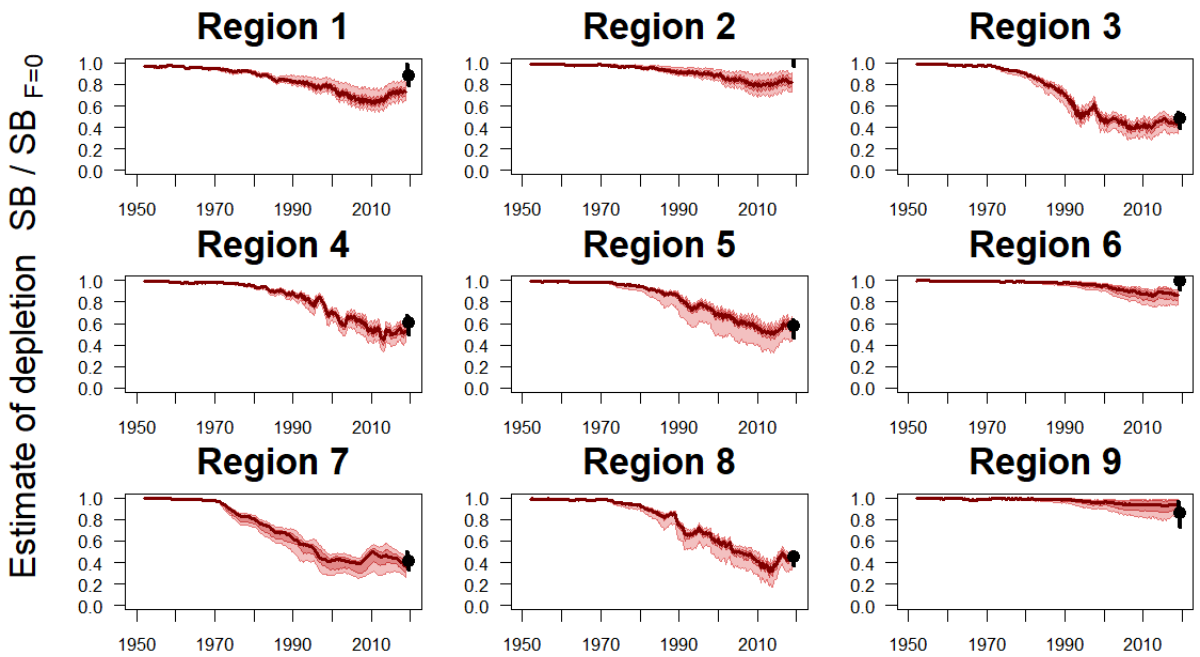
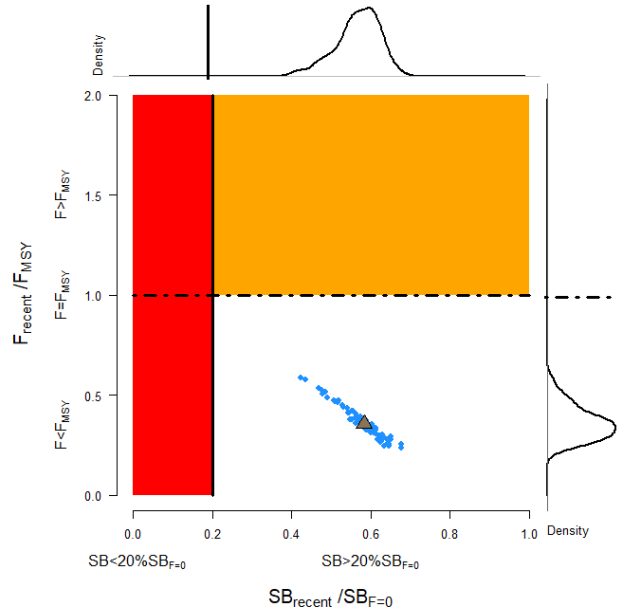


Figure 54: Distribution of time series depletion estimated for in each region in the structural uncertainty grid. The dark red line is the median, the red region represents the 50^{th} percentile range, the light red is the 80^{th} percentile range, the black point and error bars are the median and 80^{th} percentile of $SB_{recent}/SB_{F=0}$ for that region.

(a) SB Recent



(b) SB Latest

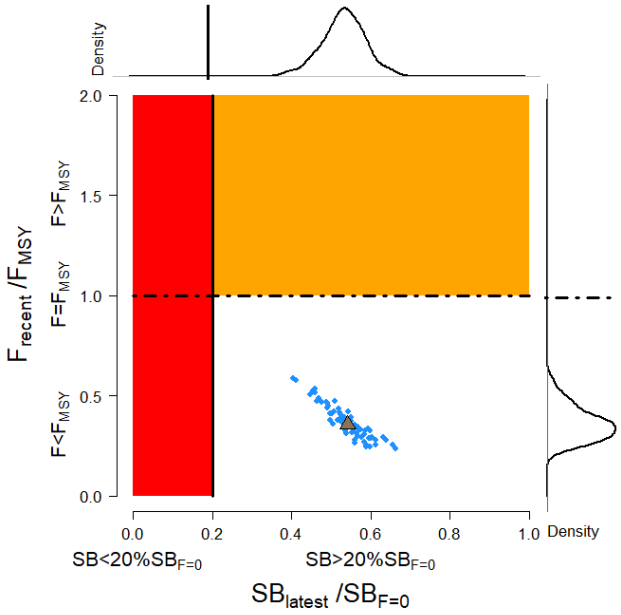


Figure 55: Majuro plots representing stock status in terms of spawning potential depletion and fishing mortality. The red zone represents spawning potential levels lower than the agreed limit reference point which is marked with the solid black line. The orange region is for fishing mortality greater than F_{MSY} (marked with the black dashed line). The plots summarize the results for each of the models in the structural uncertainty grid. The top panel displays $SB_{recent}/SB_{F=0}$ and the bottom panel displays $SB_{latest}/SB_{F=0}$.

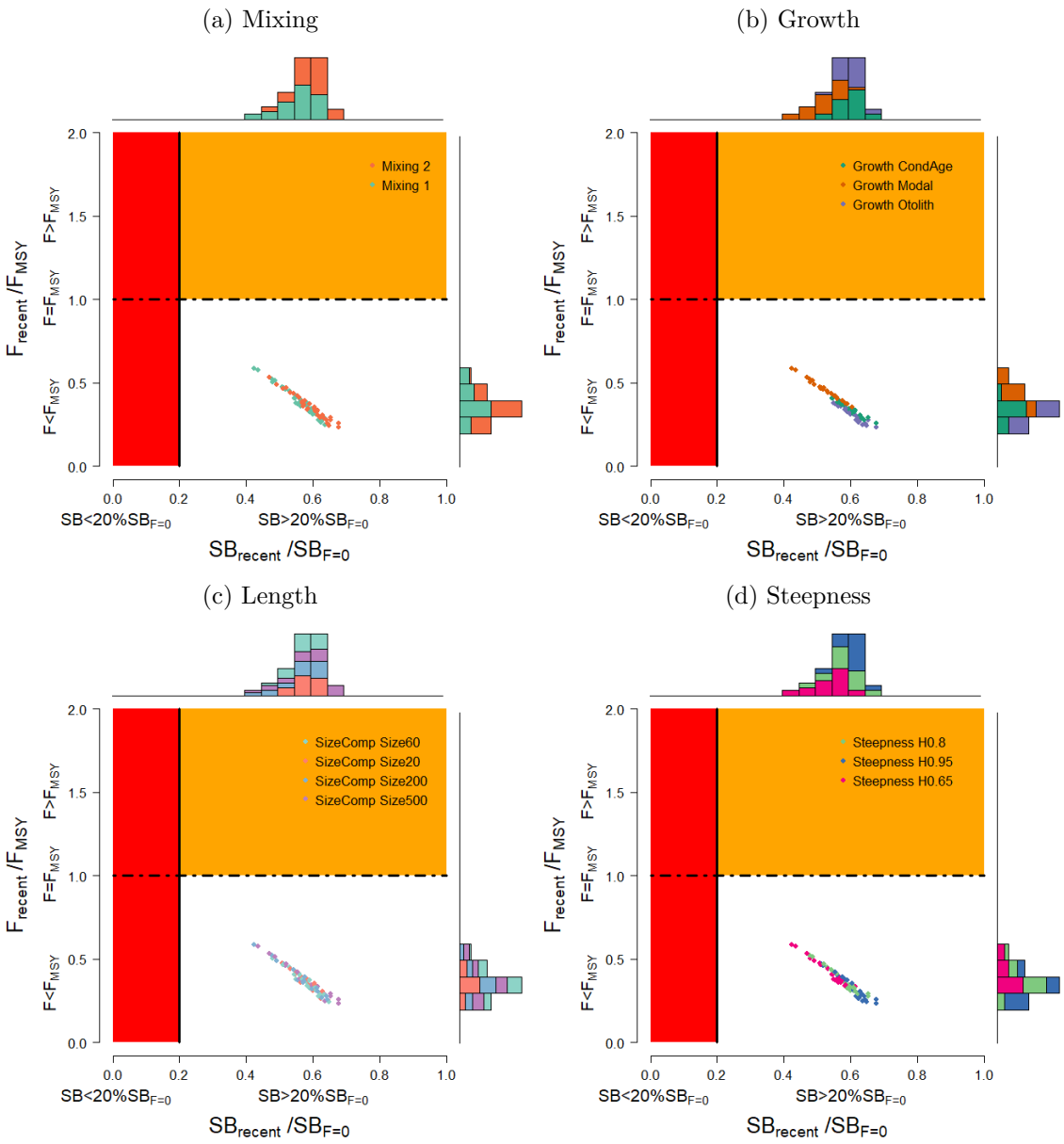


Figure 56: Majuro plots summarizing the results for each of the models in the structural uncertainty grid (see Figure 55 for explanation of the interpretation of Majuro plots). All panels display $SB_{recent}/SB_{F=0}$ for different axes of the structural uncertainty.

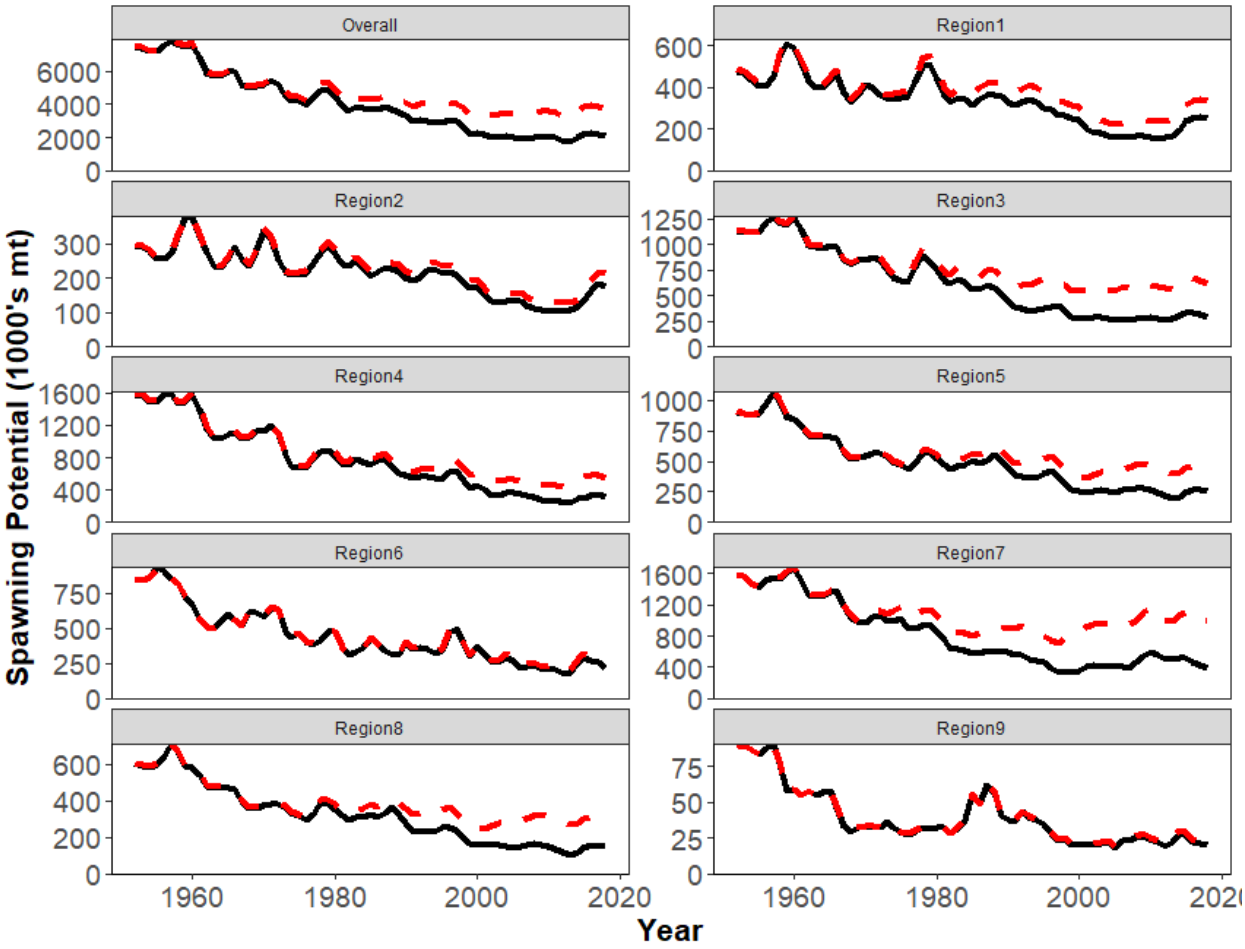


Figure 57: Comparison of the estimated annual spawning potential trajectories (lower solid black lines) with those trajectories that would have occurred in the absence of fishing (upper dashed red lines) for each region, and overall, for the diagnostic model.

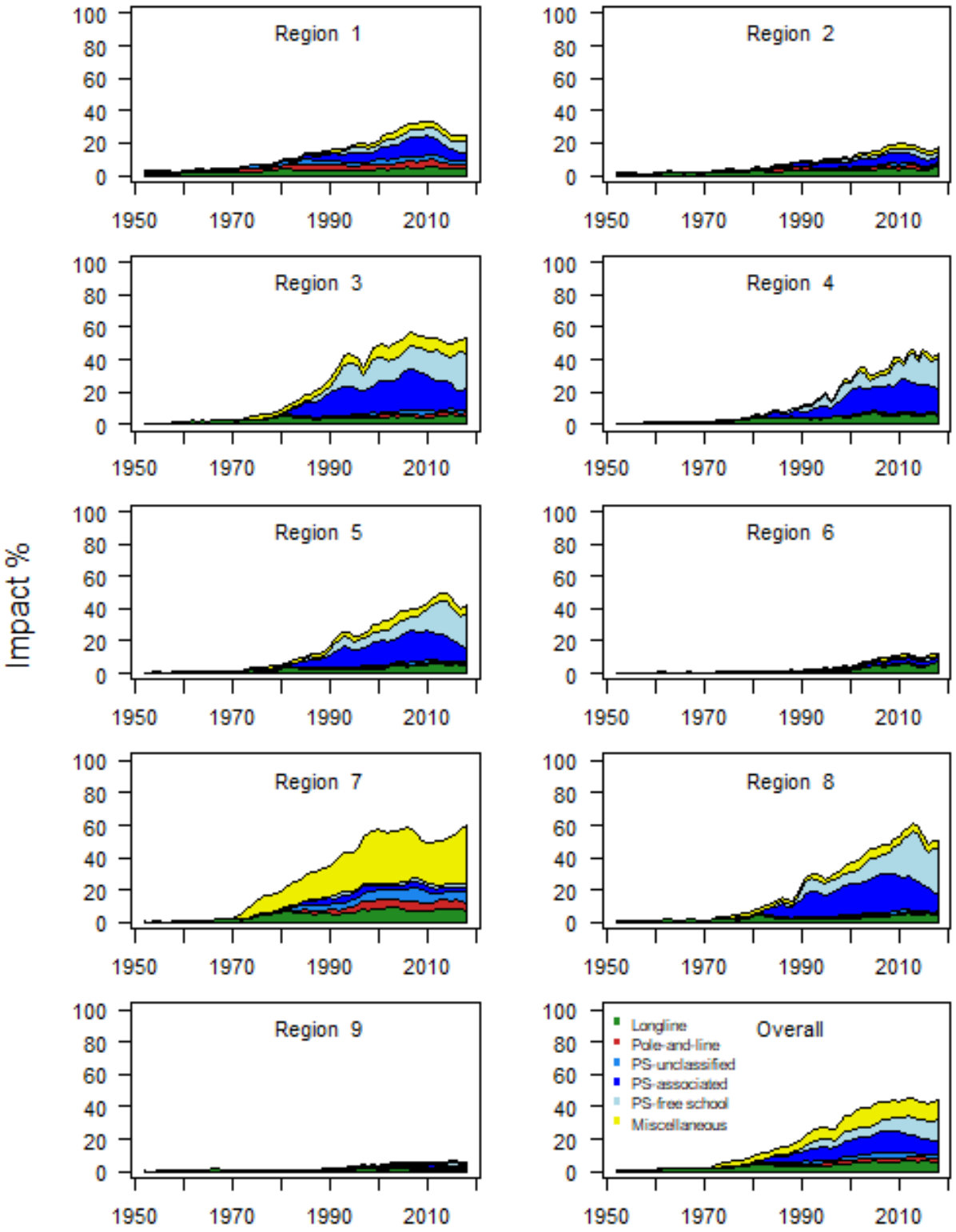


Figure 58: Estimates of reduction in spawning potential due to fishing (fishery impact = $1 - SB/SB_{F=0}$) by region, and over all regions (lower right panel), attributed to various fishery groups for the diagnostic model.

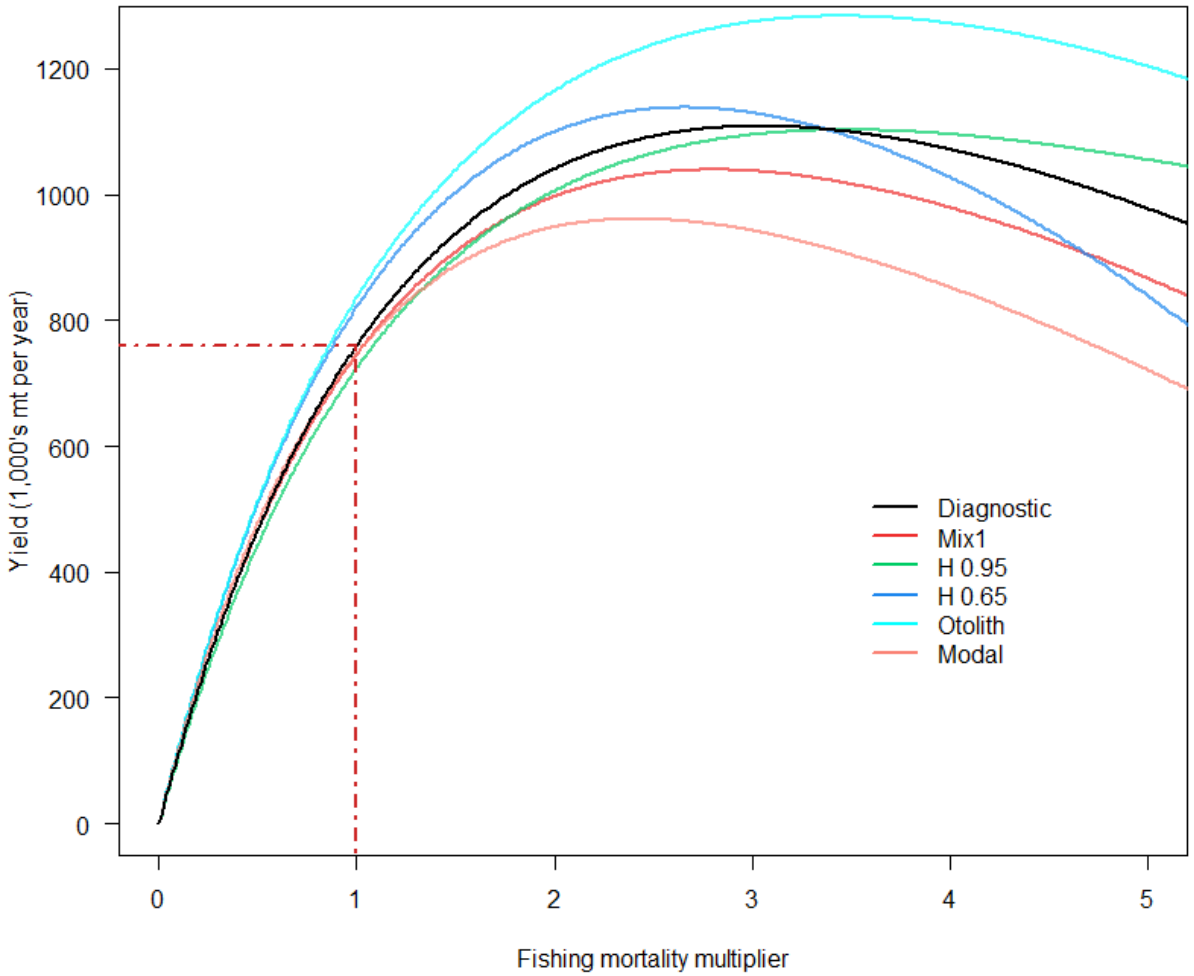


Figure 59: Estimated yield as a function of fishing mortality multiplier for the diagnostic model and a few of the one-off sensitivity models. The red dashed line indicates the equilibrium yield at current fishing mortality.

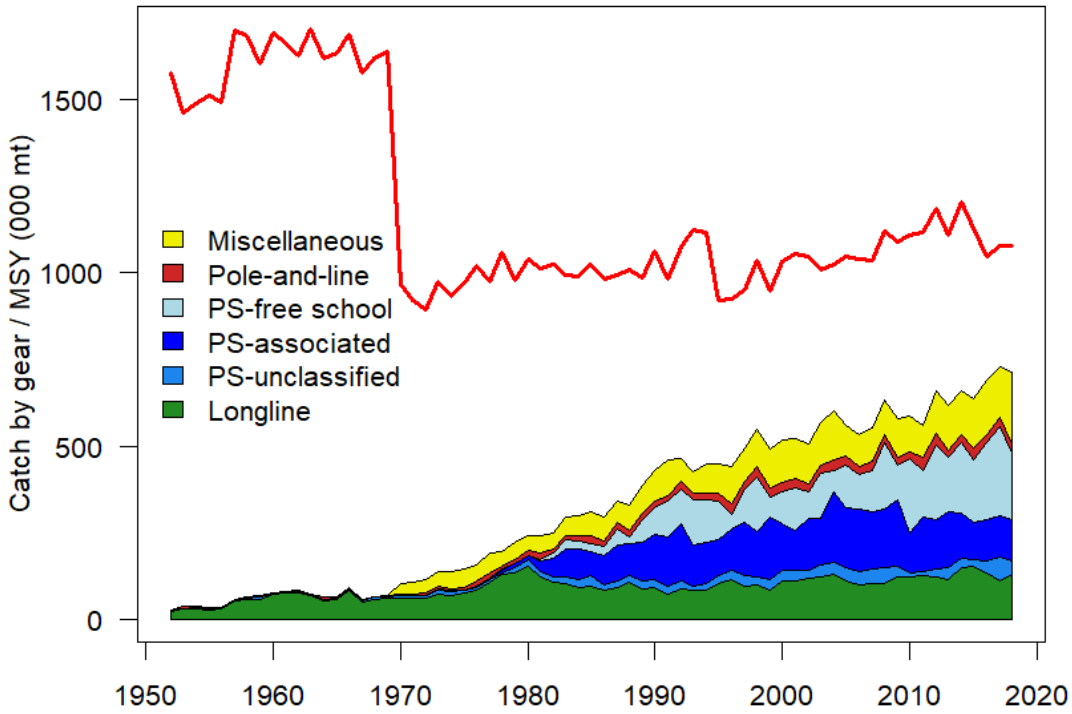
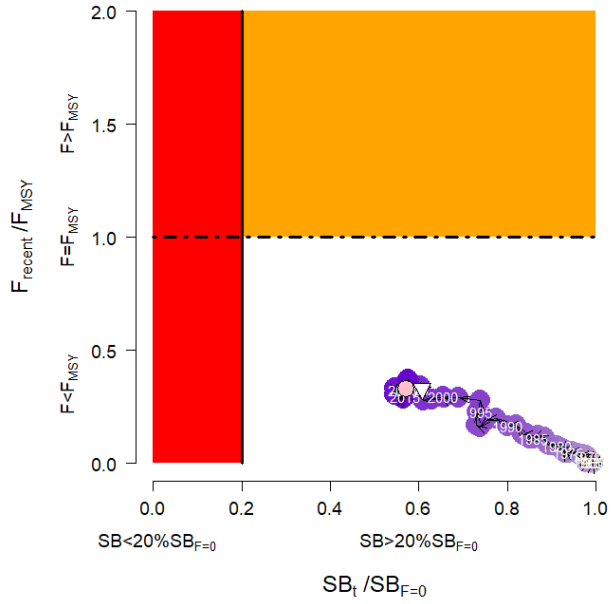


Figure 60: History of the annual estimates of MSY (red line) for the diagnostic model compared with annual catch by the main gear types. Note that this is a "dynamic" MSY which is explained further in Section 5.8.4.

(a) Dynamic Majuro



(b) Dynamic Kobe

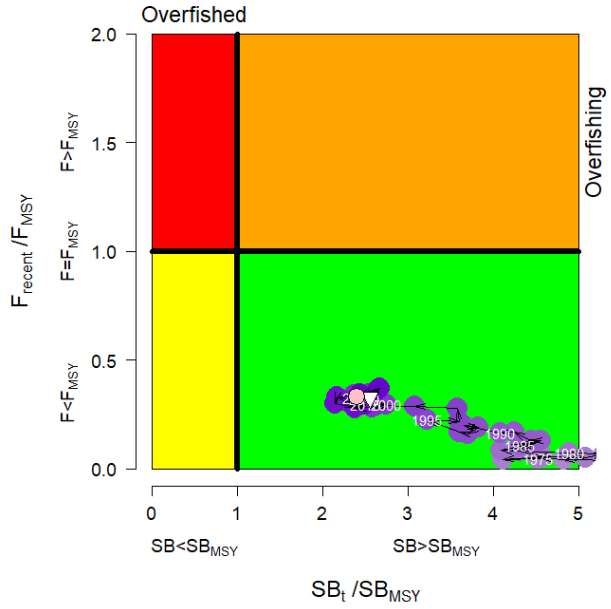


Figure 61: Majuro plot (a) for the diagnostic model (see Figure 55 for explanation of the interpretation of Majuro plots). The pink circle is $SB_{latest}/SB_{F=0}$ and the white triangle is $SB_{recent}/SB_{F=0}$, which are both detailed in Section 5.8.4. The equivalent Kobe plot for the diagnostic model (b) is provided for comparison.

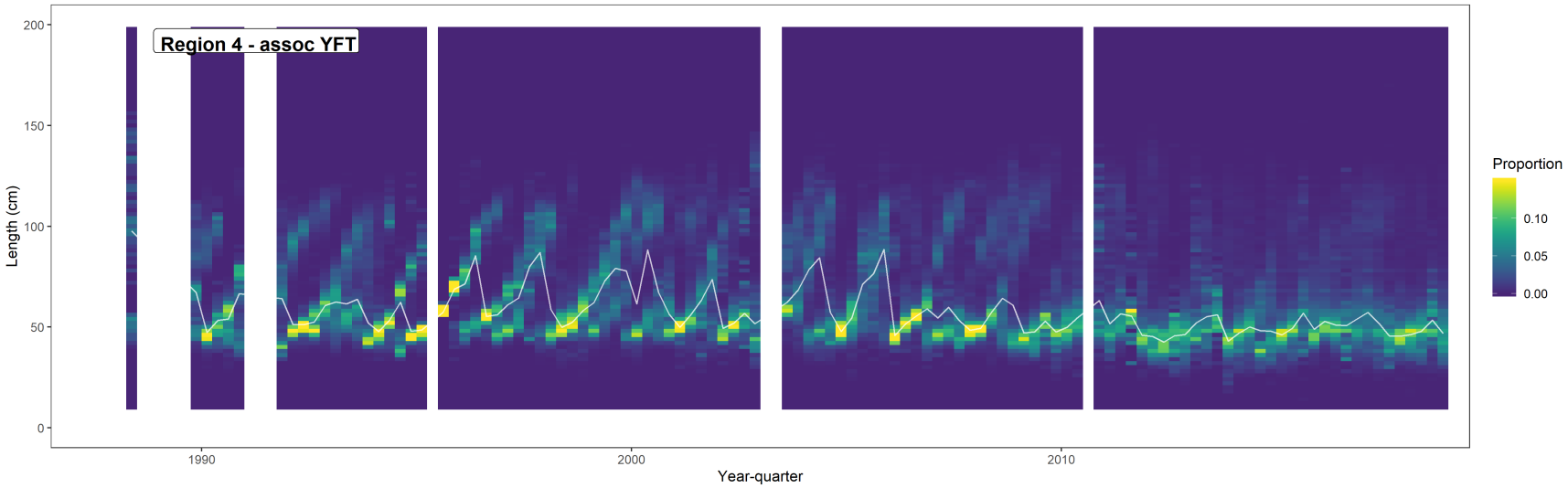


Figure 62: Length composition from the associated purse seine in region 4 that shows cohort progression early in the time series.

Appendix A

A.1 Likelihood profile

The approach for calculating a likelihood profile of the derived parameter, mean total biomass over the assessment period (to represent scale of the stock size) is outlined in Section 5.7. The profile was constructed by sequentially moving from the maximum likelihood estimate in either direction while progressively penalizing the mean total biomass at increasingly high and low values until it was determined that the minimum value had been reached for all data components. The profile reflects the loss of fit over all the data, i.e. the overall objective function value, and the individual data components, caused by changing the population size from that of the maximum likelihood estimated value. The change in likelihood relative to the maximum likelihood estimate is shown for the total likelihood (black line) and the individual data components (colored lines) in Figure A2 and displays significant declines as the parameter moves further away from the maximum value of the diagnostic model, although the curves for the individual components display different values of support for the mean total biomass. The general patterns in the influence of individual components are:

- The catch data has a flat likelihood profile indicating the negligible influence that this component has on the biomass scale in the model.
- The weight data and tagging data components display lower likelihood values at lower levels of total biomass.
- The CPUE appears to have a profile relatively similar to the total likelihood, although considerably flatter.
- The minimum of the length data likelihood appears to be at a very high total biomass and its relative likelihood increases moderately as total biomass decreases.
- The conditional age at length data also has a minimum at very high total biomass and the likelihood increases very quickly as biomass decreases, such that this component has a strong influence on the worse total fit of the model as total biomass decreases past about 7.5 million tonnes.

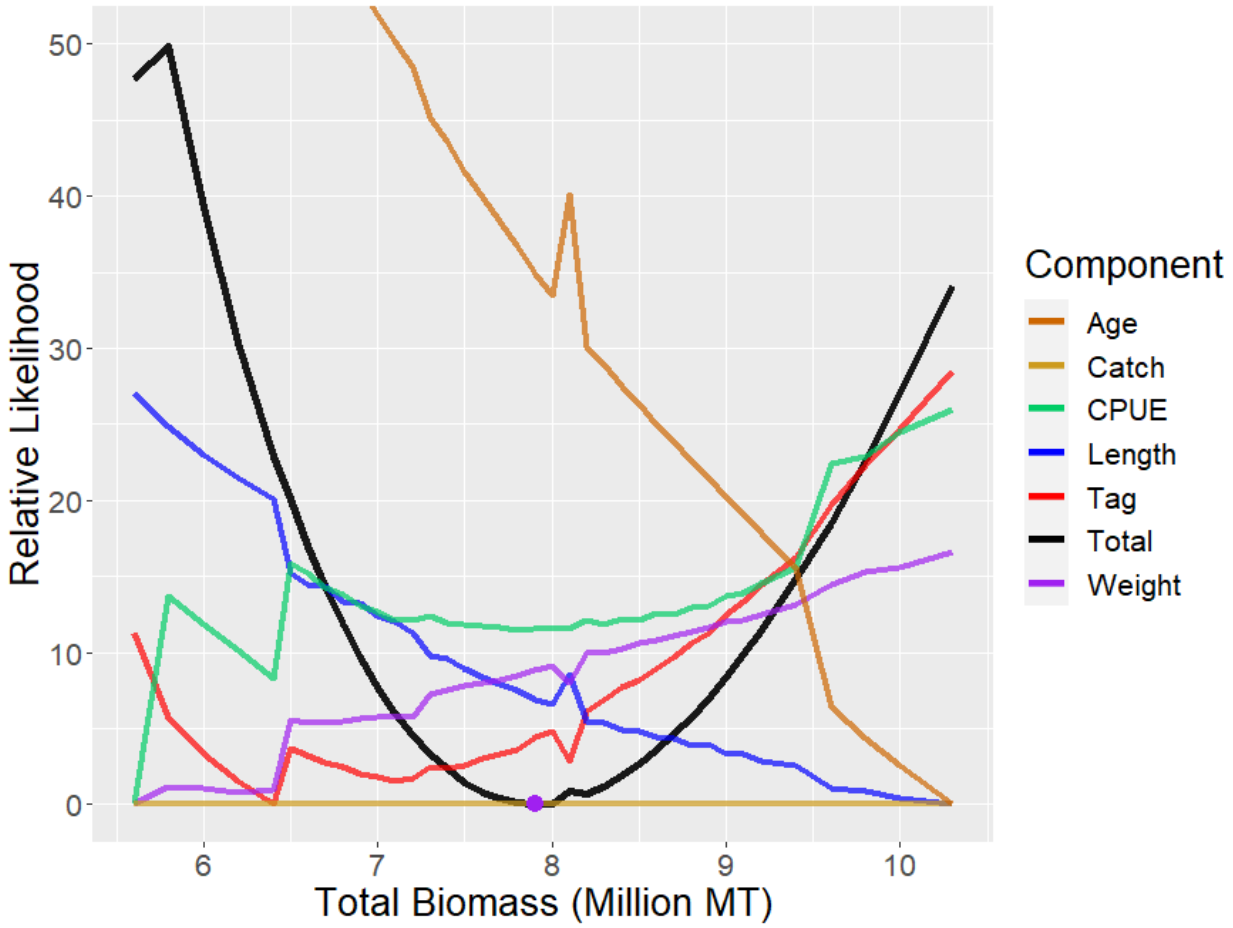


Figure A1: Change in the total (black line), and individual data component log-likelihoods with respect to the derived parameter, mean total biomass over the assessment period, across a range of values at which this parameter was penalized to fit, for the diagnostic model. The purple dot indicates the maximum likelihood estimate of the diagnostic model.

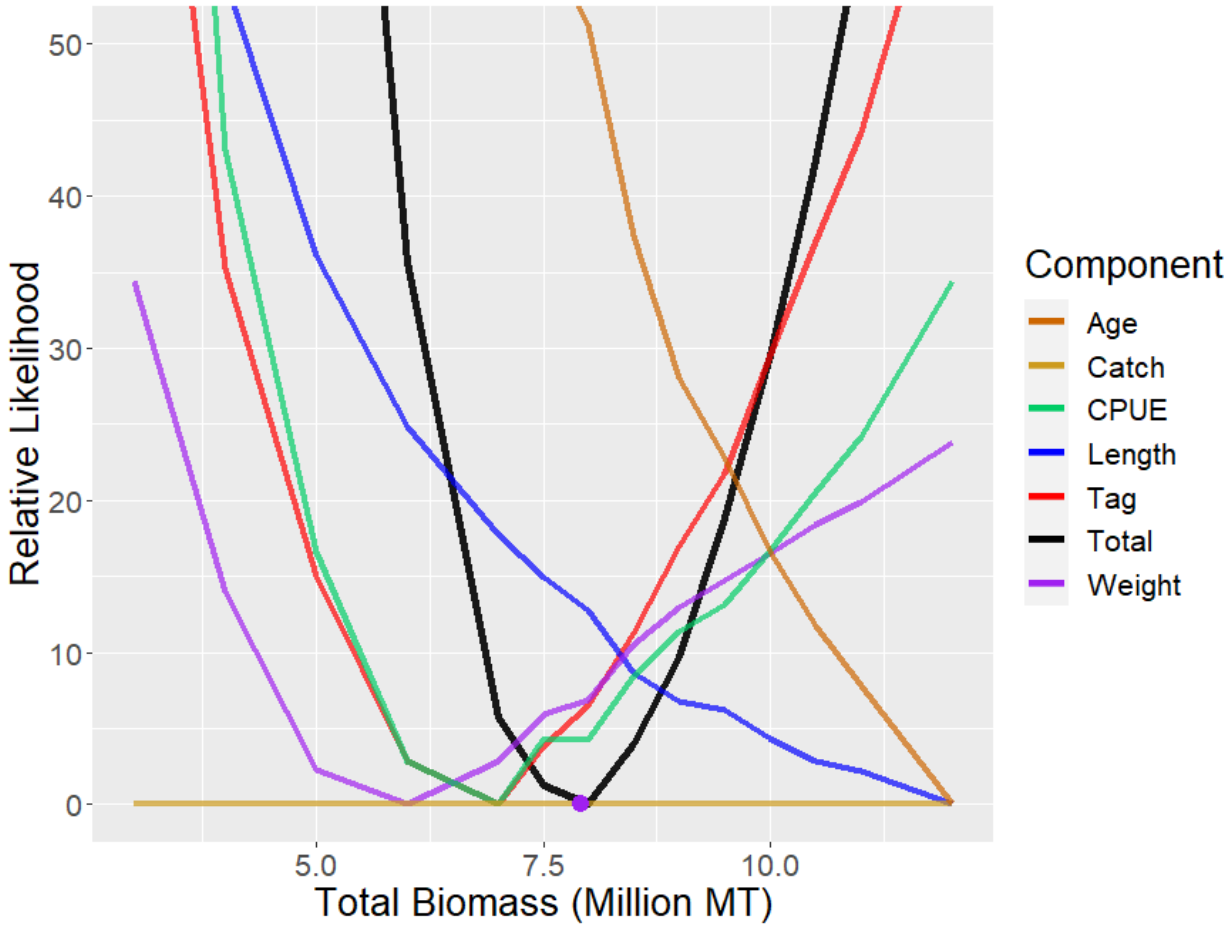
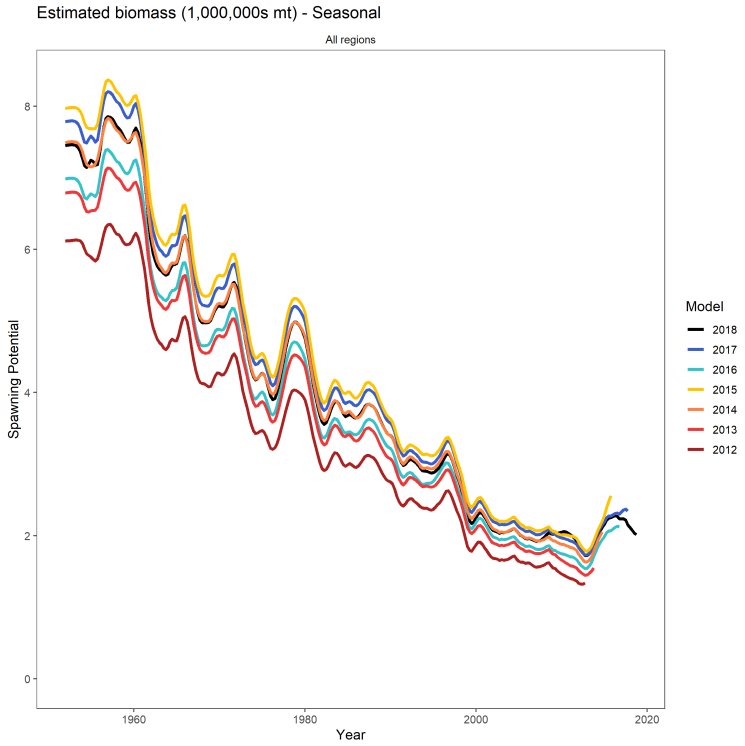


Figure A2: Change in the total (black line), and individual data component log-likelihoods with respect to the derived parameter, mean total biomass over the assessment period, across a range of values at which this parameter was penalized to fit, for the diagnostic model. The purple dot indicates the maximum likelihood estimate of the diagnostic model.

A.2 Retrospective analyses

Retrospective analyses involve rerunning the selected model by consecutively removing successive years of data to estimate model bias. A series of 6 additional models were fitted starting with the removing one year of data from the model (through 2017), followed by models with that sequentially peeled away all input data for the years 2017-2012. The models are named below by the final year of data included (e.g. 2009-2018). A comparison of the spawning potential and depletion trajectories are shown in Figures A3a and A3b. Similarly, a comparison of the estimated spawning potential and depletion from the past few stock assessments compared to the diagnostic model is presented in Figure A4.

(a) Spawning Potential



(b) Depletion

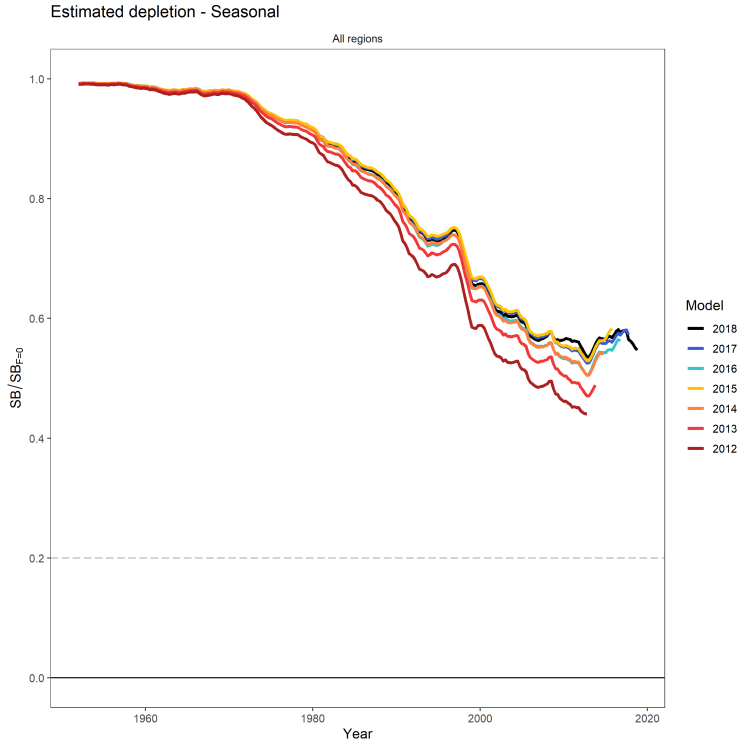
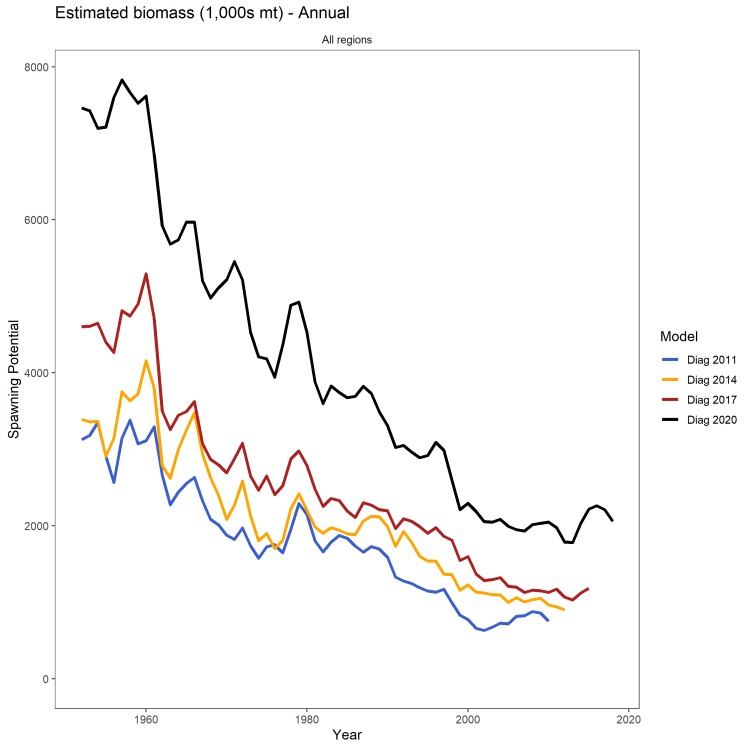


Figure A3: Estimated spawning potential(a) and fishery depletion ($SB/SB_{F=0}$) (b) for each of the retrospective models.

(a) Spawning Potential



(b) Depletion

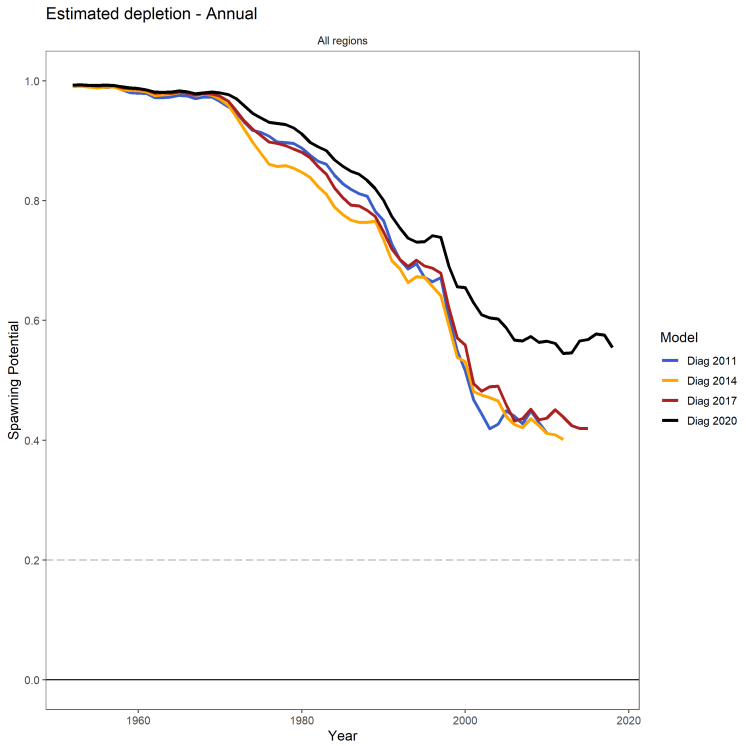


Figure A4: Estimated spawning potential(a) and fishery depletion ($SB/SB_{F=0}$) (b) for the previous three assessments.

A.3 Jittering analyses

Using the final parameter estimates from the diagnostic case model as a starting point, 47 additional runs were launched. For each of these runs, a zero-centered random deviate was added to the original parameter (or a one-centered random deviate was multiplied against the original parameter depending on the link function used) to perturb the model from its initial solution. All major estimated parameter groups (movement, selectivity, recruitment, effort deviates, tag reporting rates, etc) were perturbed. The results of this analysis are shown in Figures A5 to A7.

The jittering showed that the growth parameters were particularly difficult to estimate with a range of growth curves produced by the jittering runs (Figure A5). As a consequence, the depletion trajectories of these models were sometimes considerably different (Figure A6) with the depletion estimates being directly correlated with the growth curves. The model with the slowest growth resulted in very optimistic depletion levels, while the group of models (3 models) with much faster growth than the diagnostic case estimated much higher depletion. Most models tend to estimate similar growth to the diagnostic case; consequently, the depletion trajectories for these models formed a tight envelope. The depletion estimate of the model with the largest improvement in likelihood relative to the diagnostic model (55 points) is shown in Figure A7 and is almost identical to the depletion trajectory of the diagnostic model. Changes in the estimated depletion and difference in likelihood compared to the diagnostic model. Of the 47 converged jittered models, 33 had a better likelihood compared to the diagnostic model. Further investigation of this newly developed method for MULTIFAN-CL is encouraged to better understand the appropriate level of variability to apply to parameter estimates.

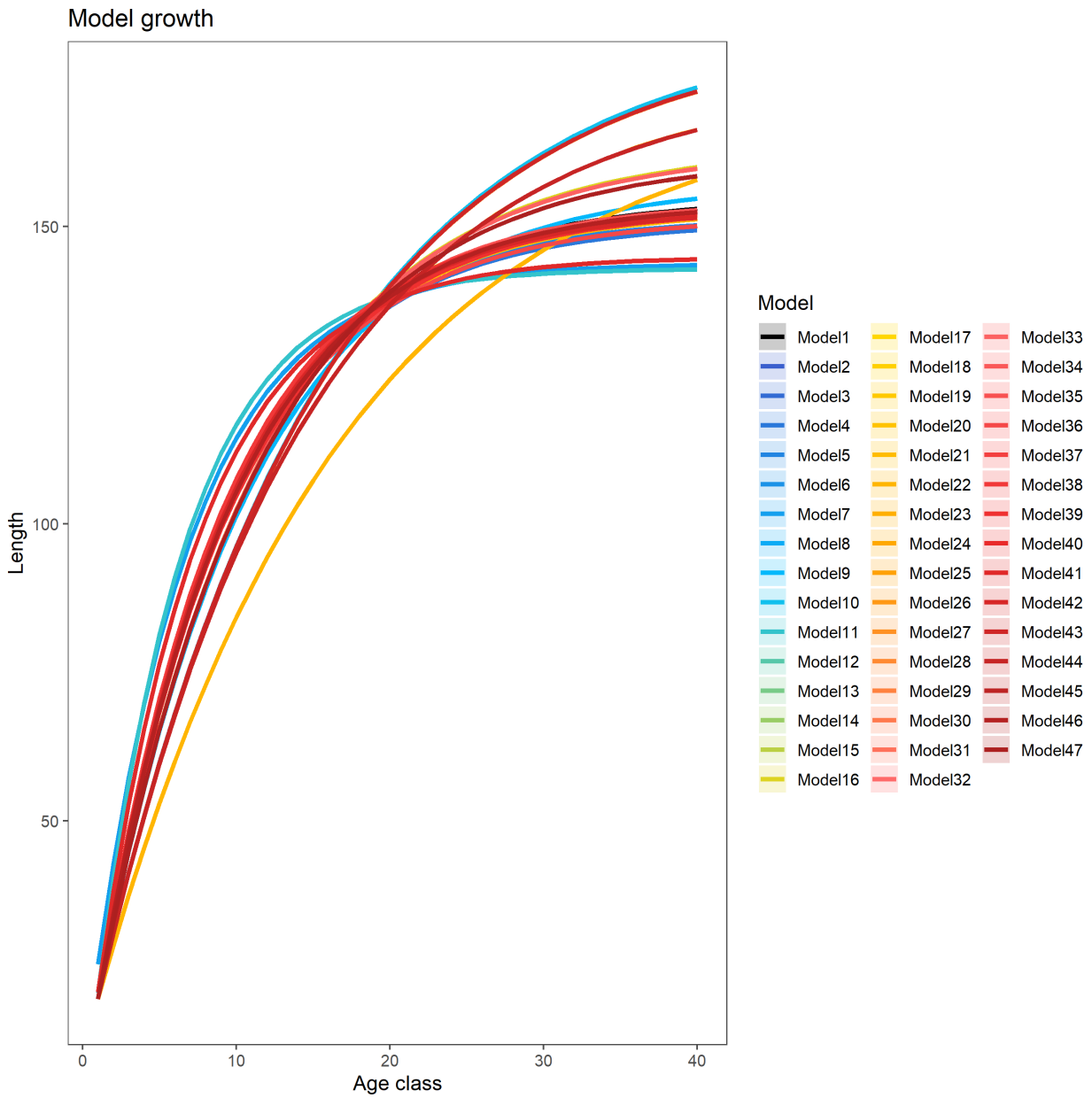


Figure A5: Growth estimates from the jitter analysis indicating that growth is difficult to estimate and starting parameters are important.

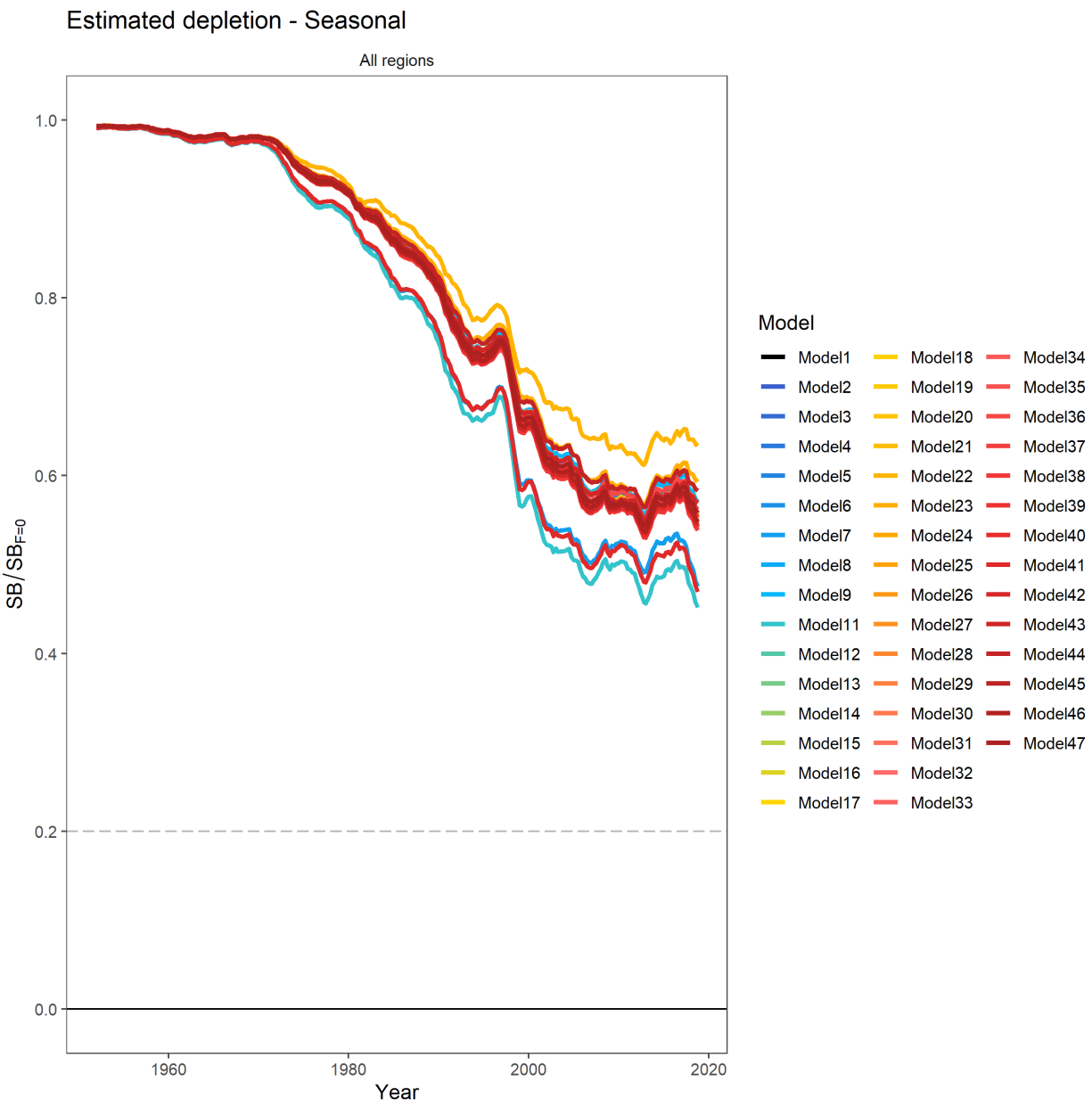


Figure A6: Estimated depletion from models jittered from the diagnostic model and ran until convergence.

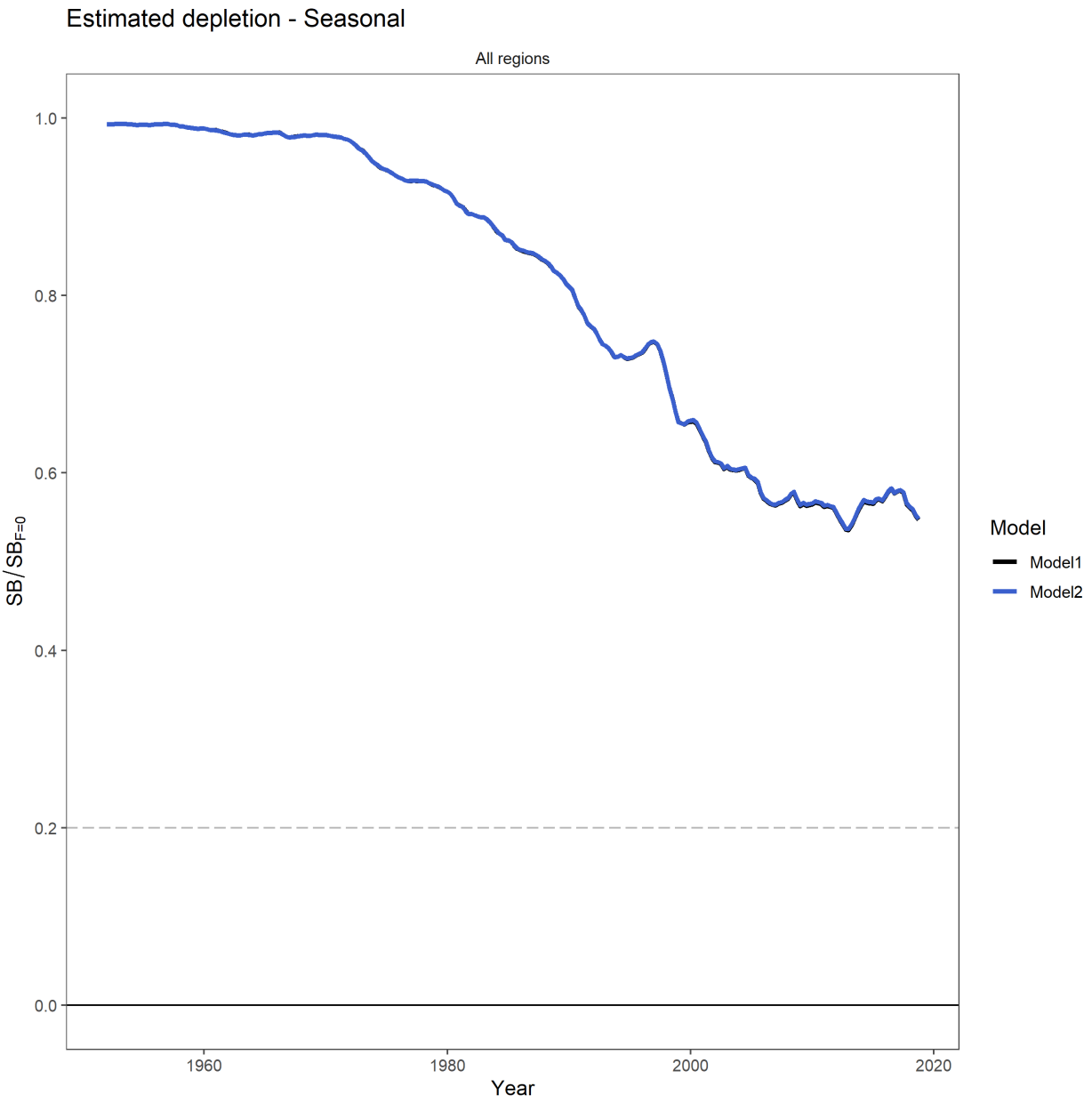


Figure A7: Comparison of depletion trajectories for the diagnostic model and the model from the jitter analysis with the best likelihood, where the lines overlay each other.

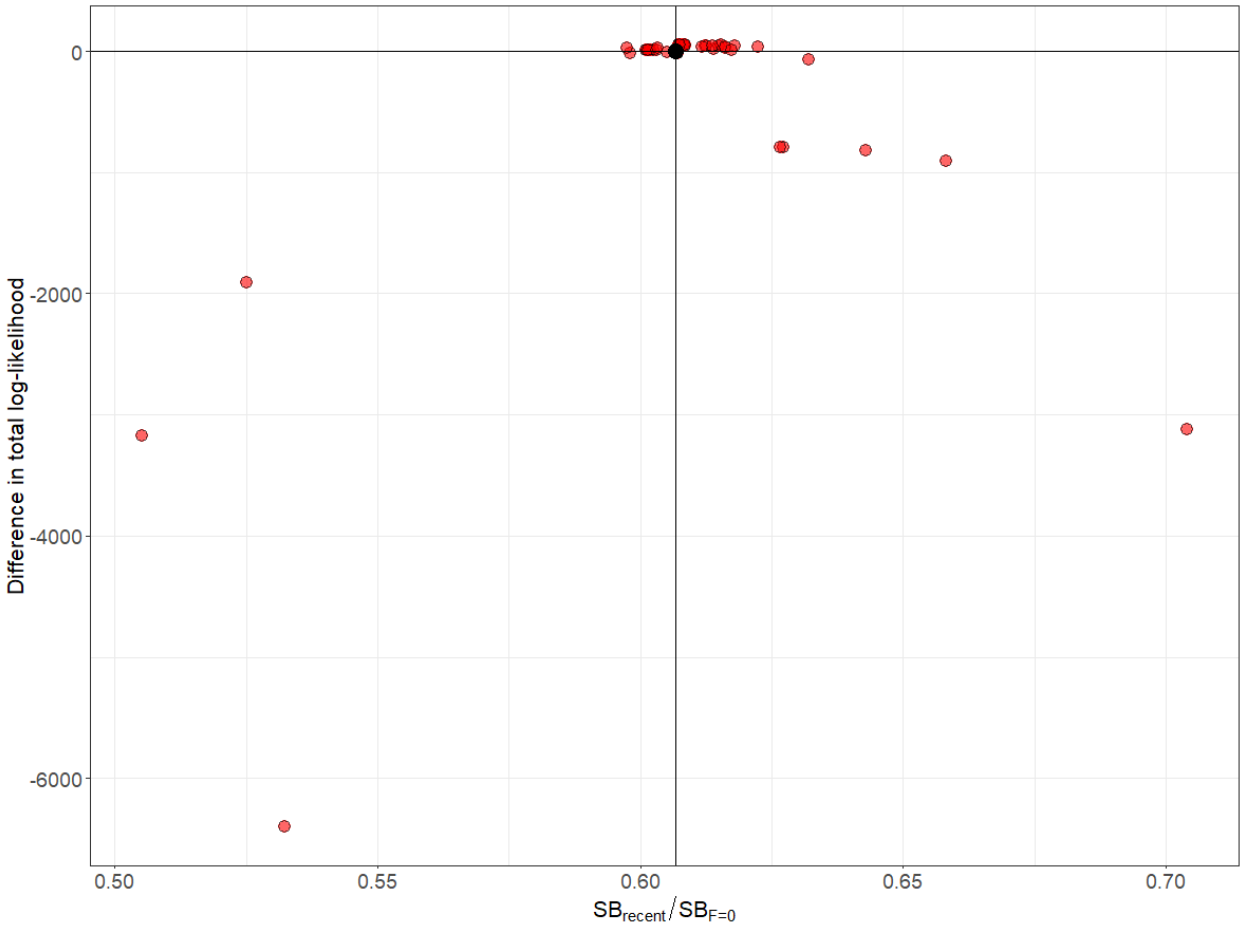


Figure A8: Difference in likelihood values relative to the diagnostic model against the estimated stock status. The points above the horizontal lines indicate a better fit to the data. The vertical line indicates the estimate of $SB_{recent}/SB_{F=0}$ from the diagnostic model.

A.4 Hessian

The Hessian matrix for the diagnostic model was found to be non-positive definite. This indicates that the model may not have converged sufficiently to the global minimum. The Choleski decomposition of the Hessian matrix indicated 2 out of 11,671 eigenvalues were negative (-2.197e-07, -3.434e-09). This model diagnostic also can indicate that there is multicollinearity among parameters in the model. A non-positive definite Hessian also supports the conclusion that the model is overly complex and there is insufficient information in the data to estimate all of the model parameters.

A.5 Stochastic projections

The potential stock consequences of fishing at 'status quo' conditions (i.e. at recent average fishing levels) were evaluated through stochastic projections, using the uncertainty framework approach previously endorsed by SC:

- Stochastic 30 year projections were conducted from each assessment model within the uncertainty grid developed for the 2020 yellowfin assessment.
- For each model, 14 stochastic projections were conducted, with future recruitments randomly sampled from historical deviates, resulting in 1008 projections across the model grid – Future recruitment in the projection period was based upon the long-term recruitment patterns (sampled from the period 1962 to 2017).
- Purse seine fisheries were projected forwards assuming 2018 effort levels. All other fisheries were projected on catch, using the recent average value (2016 to 2018).
- The outputs of the projections (median $SB_{2048}/SB_{F=0}$ and $F_{recent}/F_{\text{MSY}}$, and risk $SB_{2045}/SB_{F=0} <$ LRP) were calculated across the 72 model grid.
- Catchability was assumed to remain constant in the projection period at the level estimated in the terminal year of the assessment model.

Results of the projections are summarized in Figure A9 and Table A1.

Table A1: Summary of yellowfin stock outcomes under 2016–2018 average fishing scenario.

$SB_{2025}/SB_{F=0}$	$SB_{2035}/SB_{F=0}$	$SB_{2048}/SB_{F=0}$	Risk $SB_{2048}/SBF = 0 <$ LRP	$F_{recent}/F_{\text{MSY}}$
0.58	0.59	0.58	0%	0.349

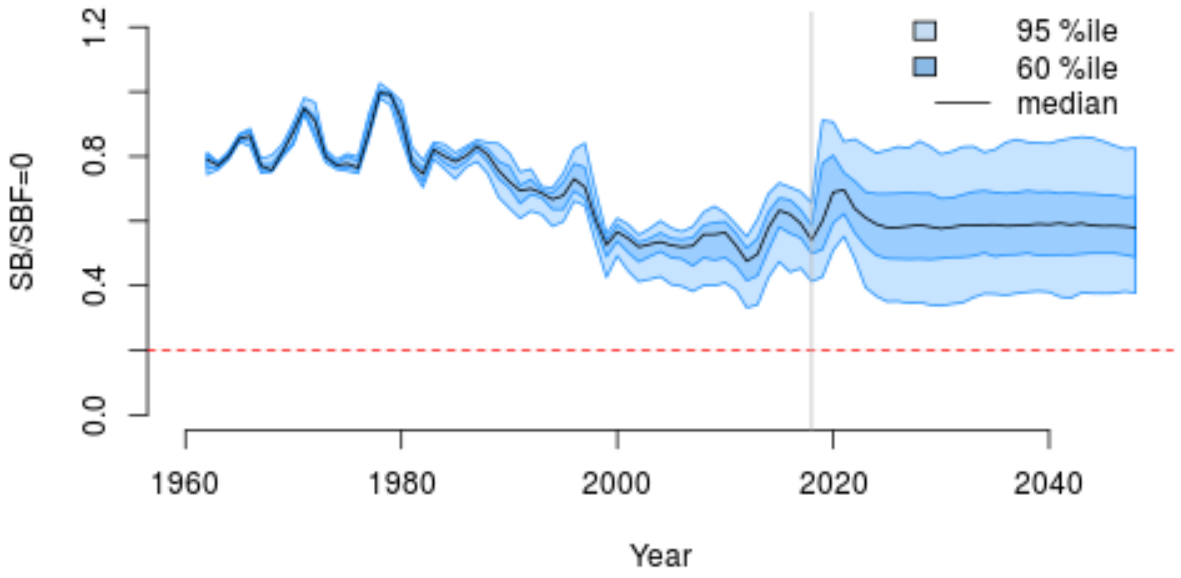


Figure A9: Time series of yellowfin tuna spawning biomass ($SB_t/SB_{t,F=0}$, where $SB_{t,F=0}$ is the average SB from t-10 to t-1) from the uncertainty grid of assessment models for the period 2000 to 2018, and stochastic projection results for the period 2019 to 2048 assuming 2016-2018/2018 fishing levels continue. Vertical gray line at 2018 represents the last year of the assessment. During the projection period (2019-2048) levels of recruitment variability are assumed to match those over the time period used to estimate the stock-recruitment relationship (1962-2017). The red horizontal dashed line represents the agreed limit reference point.