SCIENTIFIC COMMITTEE
NINETEENTH REGULAR SESSION
Koror, Palau
16–24 August 2023

Stock assessment of yellowfin tuna in the western and central Pacific Ocean: 2023

WCPFC-SC19-2023/SA-WP-04 (Rev. 3)

15 September 2023
minor revision 19 May 2026

A. Magnusson[1], **J. Day**[1], **T. Teears**[1], **J. Hampton**[1], **N. Davies**[2], **C. Castillo Jordán**[1], **T. Peatman**[3], **R. Scott**[1], **J. Scutt Phillips**[1], **S. McKechnie**[1], **F. Scott**[1], **N. Yao**[1], **R. Natadra**[1], **G. Pilling**[1], **P. Williams**[1], **P. Hamer**[1]

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

Revision 3, from 19 May 2026

A minor revision, three years after the stock assessment:

- Corrected Figure 22 to show composite length frequency plots for yellowfin tuna rather than bigeye tuna.

Revision 2, from 15 September 2023

This revision of the report adds:

- Corrected y-axis numbers on annual recruitment plots, calculated as the sum rather than the average across seasons, effectively multiplying by four (Figures 41, 42, 45).
- Corrected y-axis labels on Majuro and Kobe plots, adding the subscript 'recent' (Figure 64).

Revision 1, from 23 August 2023

This is a revision of the first complete version which was labelled 1.03. This revision of the report adds:

- Corrected proportion-by-source-region plot (Figure 44).
- Corrected Majuro and Kobe plots (Figure 64).
- New dynamic MSY plot (Figure 68).

Contents

1 Executive Summary

This paper describes the 2023 stock assessment of yellowfin tuna (*Thunnus albacares*) in the western and central Pacific Ocean. An additional three years of data were available since the previous assessment in 2020, and the model extends through to the end of 2021. The assessment moved to a new 5 region spatial structure with improved convergence properties and which achieved a positive definite Hessian solution, which was a requirement for future assessments from SC18. This change was made during the stepwise model development process. The 5 region model, given its superior convergence properties to the original 9 region model, is used as the basis for the structural uncertainty grid and stock status conclusions. Additional new developments to the stock assessment, many of which have emanated from the independent peer review of the 2020 yellowfin assessment, include:

- Conversion from a catch-errors to a catch-conditioned approach, and the inclusion of a likelihood component for the CPUE from the index fisheries.
- Change from using VAST to sdmTMB to standardise the input CPUE series and the inclusion of additional covariates in the CPUE model.
- Different CPUE variances used for the CPUE associated with each index fishery, applying a new approach to estimate these variances.
- Internal estimation of natural mortality and application of the Lorenzen form of natural mortality at age.
- Additional procedures implemented for achieving more reliable model convergence, including jittering and checking positive definite Hessian status for all grid models.
- Integration of parameter estimation uncertainty with model-based uncertainty across the model grid for the key management reference points.
- Additional size composition filtering.
- Modifications to selectivity estimation settings, changes to fisheries with non-decreasing selectivity.
- Adoption of revised tagger effect modelling framework, reverting to assumptions similar to those used in 2017.
- Changes to size data weighting and downweighting the conditional age-at-length data for internal growth estimation.

This assessment is supported by the analysis of catch and effort data for longline fisheries to provide regional abundance indices (Teears et al., 2023), revised analysis of tagger effects and tag reporting

rates (Peatman et al., 2023a; Peatman, 2023), size composition data analyses and preparation (Peatman et al., 2023b), improvements to data for length-weight conversion factors (Macdonald et al., 2023b), review and analyses to inform considerations of alternative spatial structures, and developments to the MFCL software (Davies et al., 2023).

This assessment implemented a more rigorous approach to achieve more reliable and stable model convergence, which was beneficial for achieving a positive definite Hessian for the 2023 diagnostic model. The cumulative effects of the stepwise changes between the 2020 diagnostic model and the current diagnostic model is a reduction of $SB/SB_{F=0}$. The difference between the 2020 diagnostic model and the current diagnostic model is considerable in terms of $SB/SB_{F=0}$ but much smaller in terms of SB. Therefore, it appears that the main difference between the two diagnostic models is the estimate of higher ($SB_{F=0}$) by the 2023 diagnostic model. This difference appeared to occur mainly in the step that introduced the estimation of Lorenzen natural mortality, although the change in spatial structure and application of the revised approach to modelling tagger effects further contributed to the reduction of $SB/SB_{F=0}$. The move away from the more complicated externally calculated natural mortality at age function used previously to estimating natural mortality internally with a Lorenzen functional form followed recommendations from various reviews on stock assessment methods, and was supported by a recent tuna stock assessment good practices workshop. Subject to the caveat that none of the steps in the stepwise development involved jittering, in terms of decreasing the final value of $SB/SB_{F=0}$ the most influential steps in the development of the 2023 diagnostic model were; the estimation of natural mortality using the Lorenzen curve, applying the revised tagger effects method (which was also a recommendation for an expert workshop and supported by the Pre-assessment Workshop), and updating the CPUE spatio-temporal analysis. Finally, the 2020 yellowfin stock assessment estimated the median $SB_{\text{recent}}/SB_{F=0}$ across the model grid to be 0.58, where 'recent' was the period 2015-2018. Calculating the equivalent median depletion from the 2023 stock assessment grid, $SB_{2015-2018}/SB_{F=0}$ is 0.47. Overall the changes in assessment methods and the updated data produced a less optimistic estimation of stock status than the 2020 assessment.

In addition to the diagnostic model, we report the results of one-off sensitivity models to explore the impact 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 with 54 models) for consideration in developing management advice that includes combinations of those areas of uncertainty considered important. Finally, we have also estimated the parameter (estimation) uncertainty for the key management reference points $SB_{\text{recent}}/SB_{F=0}$ and $F_{\text{recent}}/F_{\text{MSY}}$ which is combined with the structural uncertainty to provide the final uncertainty for these management quantities. The ability to include estimation uncertainty on top of structural uncertainty for the key management quantities, $SB_{\text{recent}}/SB_{F=0}$ and $F_{\text{recent}}/F_{\text{MSY}}$, is an improvement from previous assessments, however, in this case its inclusion did not influence the management advice. It is,

however, recommended that management advice is formulated from the results of the structural uncertainty grid with the estimation uncertainty included for $SB_{\text{recent}}/SB_{F=0}$ and $F_{\text{recent}}/F_{\text{MSY}}$. The results below are based on equal weighting of all grid models.

Across the 54 models of the structural uncertainty grid run in this assessment, the most important factors when evaluating stock status were the steepness of the stock recruitment relationship, weighting of the size composition data, and tag mixing period. Unlike the previous assessment, growth was not included as an uncertainty axis, which was partly due to the recommendation of the peer review of the 2020 yellowfin assessment that external growth curves would likely be biased due to the way in which otoliths were selected for developing the growth curves.

The general conclusions of this assessment are as follows:

- The spawning potential of the stock has become more depleted across all model regions until around 2010, after which it has become more stable, or shown a slight increase.
- Average fishing mortality rates for juvenile and adult age-classes have increased throughout the period of the assessment, although more so for juveniles which have experienced considerably higher fishing mortality than adults. In the recent period a sharp increase in juvenile fishing mortality is estimated, while adult fishing mortality has stabilised.
- Overall, median depletion from the model grid for the recent period (2018–2021; $SB_{\text{recent}}/SB_{F=0}$) is estimated at 0.47 (80 percentile range including estimation and structural uncertainty 0.42–0.52, full range 0.33–0.60)
- No models from the uncertainty grid, including estimation uncertainty, estimate the stock to be below the LRP of $20\%SB_{F=0}$.
- CMM 2021-01 contains an objective to maintain the spawning biomass depletion ratio above the average of 2012–2015, $SB_{2012-2015}/SB_{F=0}$, which is a value of 0.44 calculated across the unweighted grid. Based upon the estimates of $SB_{\text{recent}}/SB_{F=0}$ of 0.47, this objective has currently been met.
- Recent (2017–2020) median fishing mortality ($F_{\text{recent}}/F_{\text{MSY}}$) was 0.50 (80 percentile range, including estimation and structural uncertainty 0.41–0.62, full range 0.26–0.78).
- Assessment results suggest that the yellowfin stock in the WCPO is not overfished, nor undergoing overfishing.

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:

1. Continued work examining appropriate approaches for modeling natural mortality for the WCPO yellowfin assessment.

2. Further simplifying the assessment by combining fisheries within regions.
3. Evaluation of growth parameter settings.
4. Improved sampling of biological data across the WCPO region for yellowfin.
5. Succession planning for MFCL.
6. Tropical focused model investigation.

2 Introduction

This paper presents the 2023 stock assessment of yellowfin tuna (*Thunnus albacares*; YFT) in the western and central Pacific Ocean (WCPO; west of 150° W). Assessment of WCPO yellowfin tuna has been conducted regularly since the late 1990s (Hampton and Kleiber, 2003; Hampton et al., 2005; Langley et al., 2011; Davies et al., 2014; Tremblay-Boyer et al., 2017; Vincent et al., 2020). As in previous assessments, the objectives of the 2023 yellowfin tuna assessment are to estimate population level parameters which indicate the stock status and impacts of fishing, such as time series of recruitment, biomass, biomass depletion and fishing mortality. We summarize the stock status 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 and Archibald, 1982), which is carried out using the stock assessment framework MULTIFAN-CL[4] (MFCL version number 2.2.x.0; Fournier et al., 1998; Hampton and Fournier, 2001; Kleiber et al., 2019). MFCL implements a size-based, age- and spatially-structured population model. Model parameters are estimated by maximizing an objective function, consisting of both likelihood (data) and "prior"[5] information components (penalties).

Each new assessment of a WCPO tuna stock typically involves updates to fishery catch, effort, and size composition data, updates to tag-recapture data when tagging data is used, implementation of new features in the MFCL modeling software, changes to preparatory data analysis, such as CPUE standardisations, and consideration of new information on biology, population structure and potentially other population parameters. These changes are an important part of efforts to continually improve the modeling procedures and more accurately estimate fishing impacts, biological and population processes and quantities used for management advice. Advice from the Scientific Committee (SC) on previous assessments, and the annual SPC pre-assessment workshops (PAW) (Hamer, 2023) guides this ongoing process. Furthermore, due to changes in assessment staff, new assessments often involves staff that did not participate in the previous versions and this may also influence differences in how assessment are conducted. Changes to aspects of an assessment can result in changes to the estimated status of the stock and fishing impacts, and resultant management advice. It is important to recognize that each new assessment represents a new estimation of the historical population dynamics and recent stock status, and each new assessment team strive to provide the best possible assessment with the time and data available.

The assessment uses an 'uncertainty grid' of models as the basis for management advice. The uncertainty grid is a group of models that are run to explore the interactions among selected

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

[5] Note that any mention of a "prior" in this report does not refer to a prior in the Bayesian sense, though the effect on the parameter estimate is similar, but rather a penalty placed on the likelihood such that the estimated parameter does not deviate too much from the specified "prior" value. The magnitude of the deviation from the "prior" is dependent on the information content of the data and the strength of the likelihood penalty applied.

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

The 2023 yellowfin tuna assessment occurs after the 2022 peer review of the WCPO 2020 yellowfin tuna assessment (Punt et al., 2023). The peer review outcomes have implications for the current assessment and these are noted where relevant. Notable new features of the 2023 assessment are summarised below and this assessment report should be read in conjunction with several supporting papers, specifically the paper on CPUE analyses and other data inputs (Teears et al., 2023), the paper on size composition data preparations and weighting (Peatman et al., 2023b), the papers on tag reporting rates and tagger effects estimations (Peatman, 2023; Peatman et al., 2023a), the paper on improved conversion factors and data on fish weights and lengths (Macdonald et al., 2023a), the paper on developments in the MFCL software (Davies et al., 2023) and the paper on conceptual models of yellowfin and bigeye population structure (Hamer et al., 2023). Finally, the planning for this assessment was informed by the discussions at the 2023 SPC PAW (Hamer, 2023).

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

- Conversion from a catch-errors to a catch-conditioned approach, and the inclusion of a likelihood component for the CPUE from the index fisheries (peer review supported this).
- Adoption of a simpler spatial structure (5 model regions). Detailed review of information and development of conceptual models for spatial structure including both size composition analysis (regression trees) and CPUE time series analysis (peer review recommendation, supported by PAW).
- Change from using VAST to sdmTMB to standardise the input CPUE series and increased the spatial resolution of the mesh configuration. Various alternative CPUE model structures and analyses explored resulting in the inclusion of additional covariates in the CPUE model (peer review recommendation).
- Different CPUE variances were used for the CPUE associated with each index fishery, using new approaches to estimate these variances. Modifications were required to MFCL to enable the index fisheries to have separate CPUE variances while maintaining shared selectivity (peer

review recommendation).

- Internal estimation of natural mortality and application of the Lorenzen form of natural mortality (recommendation of 2023 CAPAM Tuna Good Practices Workshop), also an MFCL modification to allow input of Lorenzen starting parameter values with improved parameter scaling.
- Additional procedures implemented for achieving more reliable model convergence, including jittering and checking positive definite Hessian status for all grid models (improvements to convergence criteria requested by SC18, recommendation by peer review to provide Hessian diagnostics).
- Integration of estimation uncertainty with model-based uncertainty across the grid (SC18 request for inclusion of estimation uncertainty). An MFCL development to enable calculation of variances only for the key derived quantities required for the uncertainty grid was implemented, reducing the computational load to estimate uncertainty for management quantities.
- Use of MFCL tail compression feature, applied only to zero values (and could apply a further tail compression proportion of 0.001 in future).
- Improved size composition data filtering approaches to reduce influence of low/unrepresentative sampling. Also explored alternatives for specifying input samples sizes, such as numbers of sets (but ran out of time to fully explore a range of filtering options). Applied a minimum input sample size of 50 for size composition data in MFCL (peer review recommendation to reduce unrepresentative size composition data).
- Continued use of conditional age-at-length data and internal estimation of growth, but down-weighting this data in the likelihood as an initial step in data weighting (peer review recommendations).
- Reduced the number of fisheries with a non-decreasing selectivity constraint to just the index fisheries.
- Ensured that tag reporting rate groups are not estimated for groups with zero tag recoveries, and extending this to tag reporting rate groups with fewer than 6 tag recoveries.
- Adoption of revised tagger effect modelling framework (recommended by expert workshop) with separate treatment of PTTP Central Pacific tag releases; use of multi-species models; model selection based on predictive accuracy; and, reverted to assumptions similar to those used in 2017.
- Initial explorations of the use of Dirichlet multinomial for self-scaling size composition data weighting (peer review supported this), and modification (reduction) to size composition data

weighting divisors in the grid as a result.

- Qualitative analysis of tag recapture data to inform tag mixing assumptions (PAW recommendation, and peer review).

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 45° 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). Grewe et al. (2015) showed strong genetic differences between samples from Baja California in the eastern Pacific, Tokelau in the central Pacific and the Coral Sea in the western Pacific. Similarly, Pecoraro et al. (2018) found genetic differences between populations in the far eastern and western Pacific Ocean. Evidence of finer scale genetic structure of yellowfin in the western and central Pacific Ocean is less clear and varies between different studies (Appleyard et al., 2001; Aguila et al., 2015; Pecoraro et al., 2018; Anderson et al., 2019; Evans et al., 2019). An earlier genetic study by Ward et al. (1994) proposed the existence of eastern and western Pacific subpopulations separated at around 150° W. Observations of the distribution of yellowfin tuna larvae indicate that spawning occurs broadly throughout the central and western tropical Pacific, with spawning all year in the equatorial region, and seasonally in warmer months to the north and south (Nishikawa et al., 1985; Ijima and Jusup, 2023). Studies using otolith chemistry have 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 most recent otolith chemistry study provided good evidence that majority of small (30-40 cm) juvenile yellowfin captured around Japan originated from very small juveniles ($<$ 10 cm) sampled further south in the western Pacific tropical region (Satoh et al., 2023). This suggests dispersal/movement from the equatorial western Pacific spawning areas via the western boundary current (Kuroshio) is important for juvenile recruitment around Japan.

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; Hamer et al., 2023). 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 populations. 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 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), a 9 region model (Figure 1) was implemented in 2014 for both yellowfin and bigeye (Davies et al., 2014; Harley et al., 2014; McKechnie 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 recommended only using the 10° N option in the 2020 yellowfin assessment. The 9 region spatial structure was viewed as 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. regions 4, 8 and 9).

The paper by Hamer et al. (2023) considers published information on genetic and non-genetics indicators of population structure (mostly covered in the review by Moore et al. (2020)), larval distribution patterns, and also includes analyses of spatial heterogeneities in size composition data and CPUE time series for the Pacific longline fisheries. That paper suggests that while yellowfin tuna are likely one genetic stock in the Western and Central Pacific Ocean (WCPO), there is indication of substructure with the WCPO assessment region. The paper notes that the tropical region has several features that warrant it being considered as a separate spatial strata from the northern and southern sub-tropical/temperate regions. Likewise the paper suggests that the Indonesia/Philippines/Vietnam/South China Sea region warrants being considered a separate spatial strata. There was also support for a separate region/fisheries around Hawaii based on size composition data, localised spawning and otolith chemistry studies. The review did not conclude that the

previous 9 region model structure was inappropriate but did suggest alternative simplified spatial stratifications that could be considered for yellowfin tuna assessment in the WCPO and this is discussed further in subsection 4.2. Continuing to improve understanding of spatial population structure and processes for yellowfin in the WCPO, and more broadly in the Pacific, remains an important area of research.

The 9 region spatial structure (Figure 1) applied in the 2020 assessment was the basis for the current assessment. However, based on the review and analyses in Hamer et al. (2023) and the recommendation from the peer review to explore plausible simpler spatial structures (Punt et al., 2023), we considered an alternative simplified 5 region structure. Ultimately, the 5 region structure was adopted for this assessment due to better model performance and convergence properties (see subsection 6.1), including the requirement to achieve a positive definite Hessian. A thorough exploration of alternative spatial structures besides the 5 region structure was not possible in the time available, but is still recommended as future work.

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) and the maximum validated age is 13 years (Andrews et al., 2022). 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 24° C (Itano, 2000; Reglero et al., 2014). Larval stages are found widely in surface waters throughout the central and western Pacific (Nishikawa et al., 1985; Servidad-Bacordo et al., 2012; Ijima and Jusup, 2023) 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; Ijima and Jusup, 2023). Juvenile yellowfin (several months of age) are prevalent in commercial fisheries in the Philippines and eastern Indonesia (Hare et al., 2023; Williams and Ruaia, 2023), suggesting this region is important for the juvenile stages, but they are also found more widely in the equatorial Pacific.

Growth parameters can be highly influential in the estimates of management parameters by stock assessment models, and the most recent stock assessments of yellowfin in the WCPO identified

growth as an important uncertainty requiring further research (McKechnie et al., 2017; Vincent et al., 2018, 2020). Considerable work was carried our on developing external growth curves from otoliths and otolith-tag integrated growth models for the previous assessment (Eveson et al., 2020; Farley et al., 2020). Those studies resulted in different estimates of growth rate parameters to those previously estimated by the MFCL model based on length composition modal progression. Annual aging of yellowfin using otoliths has recently been validated using the bomb radiocarbon method (Andrews et al., 2022), but there has been no further otolith aging of yellowfin since the 2020 assessment. The previous assessment explored alternative growth options including external fixed growth curves from both otolith aging an internal modal length analyses (applied as external fixed curves), as well as a fully internal growth estimation by inputting the otolith data (with associated length data) as conditional-age-at-length data. The latter was used as the approach for the diagnostic model, and has since been recommended as the most suitable approach for growth estimation by the yellowfin assessment peer review (Punt et al., 2023). Understanding of spatio-temporal variation in growth is limited and is insufficient to consider such effects in the current assessment, nor is it feasible to implement spatially varying growth using the MFCL model framework.

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, data on sex ratio at length is important. For the previous assessment data on sex ratios collected by observers in the WCPO was incorporated into the estimation of spawning potential and M at age (Vincent and Ducharme-Barth, 2020). We utilise the same data source in this assessment. The 2020 yellowfin assessment conducted a life-history based meta-analysis of natural mortality for yellowfin (Vincent and Ducharme-Barth, 2020) indicating an envelope of potential quarterly M rates of lower 95% confidence interval (0.1100), mean (0.1298) and upper 95% confidence interval (0.1495). More recently natural mortality of yellowfin has been reviewed by Hoyle et al. (2023), and this review is taken into account when considering options for natural mortality in the current assessment.

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 Ruaia, 2023). 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 Ruaia, 2023).

The annual yellowfin tuna catch in the WCPO increased from 100,000 mt in 1970 to between 700,000 and 750,000 mt in recent years, mainly due to increased catches in the purse seine fishery, (Hare et al., 2023; Williams and Ruaia, 2023). The 2022 catch was approximately 720,000 mt, and the 2021 (last of year of this assessment) catch was slightly higher at approximately 730,000 mt (Figure 3). Purse seiners harvest the majority of the yellowfin tuna catch (around 50–55% since 2018), while the longline fleet accounted for around 10-15% of the catch in recent years, primarily in the equatorial regions (Figure 4, Williams and Ruaia, 2023). 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.

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 (associated sets). Unassociated sets account for the majority of purse seine sets, however, many of these sets fail (skunks) and when considering successful sets only, the numbers of associated and unassociated sets are similar (Hare et al., 2022).

Since 2010, annual catches of yellowfin tuna by longline vessels in the WCPO have varied between approximately 74,000 to 105,000 mt (Williams and Ruaia, 2023). The highest longline catch recorded was around 125,000 mt in 1980 (Figure 3). Annual catches from the domestic fisheries of the Philippines and eastern Indonesia area are highly uncertain, particularly prior to 1990. Recent estimates for pole and line and other gears have reached approximately 220,000–260,000 mt in the last 5 years years, and for purse seine 350,000–400,000 mt (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.

3.4 Key changes from the last assessment

3.4.1 Catch conditioned approach

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

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

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

4 Data compilation

4.1 General notes

Data used in the yellowfin tuna stock assessment using MFCL consist of catch, effort, length & weight-frequency data for the fisheries defined in the analysis, and tag-recapture data. Conditional age-at-length data are also used directly as data in the assessment model, as was recommended by the peer review of the 2020 yellowfin tuna assessment (Punt et al., 2023). Improvements in these data inputs are ongoing and readers should refer to the companion papers highlighted at the end of section 2 for detailed descriptions of how the data and biological inputs were formulated as only brief overviews are provided below. A summary of the data available for the assessment is provided in Figure 6.

4.2 Spatial stratification

The geographical area considered in the assessment corresponds to the WCPO (from 50°N to 40°S between 120°E and 150°W) and oceanic waters adjacent to the east Asian coast (110°E between 20°N and 10°S). The eastern boundary of the assessment excludes the WCPFC Convention area component that overlaps with the Inter American Tropical Tuna Commission (IATTC) area. We began the stepwise model development with the previous 9 region model structure (Vincent et al., 2020)(Figure 1), but as we progressed through the stepwise model development, aspects of model convergence began to deteriorate and a decision was made to implement the 5 region structure which had better convergence properties, including a positive definite Hessian, which was indicated as being essential for diagnostic models by SC18. The 5 region stratification was supported by the review paper (Hamer et al., 2023), and we maintained the fisheries definitions for the extraction fisheries as applied in the 9 region model. That is, gear/flag specific fisheries that were defined by separate regions in the 9 region stratification, remained defined as separate fisheries within the larger regions of the simplified stratification, noting that the simplified stratification involved merging regions of the 9 region structure rather than altering boundaries and creating entirely new regions. This is akin to a fleets-as-areas approach within the larger simplified regions. Maintaining the extraction fisheries definitions was partly for efficiency and partly to maintain the fishery definitions rather that changing both the spatial stratification and fishery definitions together. In this way the effects of simplified stratification could be isolated. This also accounts for heterogeneities in fisheries composition data such as for the region around Hawaii. The 5 region structure maintains the region around the Papua New Guinea and Solomon Islands area to accommodate the longer residency of yellowfin in these archipelagic waters and the tagging data and associated mixing period assumptions.

While the 5 region structure was preferable over the 9 region structure in terms of model performance, and provides what we feel is a suitable spatial structure, with more time we may have been able to improve the performance or the 9 region model. We suggest a stand alone project is required to fully explore and compare the benefits and limitations of alternative spatial structures, with review by SC and their advice on a preferred option for future assessments.

Readers should be aware of the differences in the region numbering between the 5 and 9 region structures, for example region 8 (Papua New Guinea/Solomon Islands) in the 9 region structure is region 3 in the 5 region structure. When model region numbers are referred to they relate to the 5 region structure unless otherwise specified.

4.3 Temporal stratification

The time period covered by the assessment is 1952–2021 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 as this is considered incomplete at the time of formulating the assessment inputs. Recent year data are also often subject to significant revision post-SC, in particular the longline data on which this assessment greatly depends.

4.4 Definition of fisheries

MFCL requires "fisheries" to be defined that 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. For most pelagic fisheries assessments, fisheries are typically defined according to combinations of gear type, fishing method and region, and for some, also flag or fleet. There are 41 fisheries defined for both the 5 and 9 region models used in this assessment (Table 1) consisting of two fishery types: "index fisheries", that are used for generating indices of abundance (see further below), and "extraction fisheries" that account for the catches removed from the stock. Extraction fisheries include longline, purse seine, pole and line and various miscellaneous fisheries in the Indonesia/Philippines/Vietnam region. The fisheries definitions for the 2023 assessment are consistent with those used in the 2020 9 region assessment, but region numbers change to account for the 5 region structure. A graphical summary of the availability of data for each fishery used in the assessment model is provided in Figure 6.

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 and selectivity of different purse seine set types. Set types were grouped into associated (i.e. log, FAD, whale, dolphin, and unknown set types) and unassociated (free-school) sets. Additional 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 index fishery was defined in each region, although in regions 2 and 4 extraction longline fishing was separated into distant water and offshore components to account for the apparent differences in fishing practices and selectivity for these fleets in these regions.

Index fisheries: The catch-conditioned approach (subsubsection 3.4.1) allows the specification of "index fisheries" that are used to provide standardised CPUE indices of abundance for each model region. Index fisheries are akin to "survey fisheries" as described for other software such as Stock Synthesis, and may be the same fisheries as the extraction fisheries, but when used as index fisheries they do not take any catch, and must have effort data to allow modelling of CPUE. For this assessment one index fishery is defined for each model region as a composite fishery composed of all longline fisheries operating in each assessment region (Teears et al., 2023). Index fisheries may be grouped if it is felt that the CPUE reflects differences in average abundance among regions. For this assessment, index fisheries are grouped which allows the standardised CPUE to provide information on regional as well as temporal relative abundance. The full longline operational dataset, described in McKechnie et al. (2015), Ducharme-Barth et al. (2020b), and Teears et al. (2023), was used as the basis for developing the index fisheries CPUE. The CPUE standardisation approach for the index fisheries is described in detail in Teears et al. (2023), see further subsubsection 4.5.3.

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 assessment regions in order to scale the population. This means that the assessment model estimates relative abundance among spatial strata that is generally similar to the scaled CPUE relative abundance. The regular longline extraction fisheries are based on the same data set, but are disaggregated into the longline fisheries defined in Table 1.

The size composition data (length and weight-frequency) for the extraction fisheries is assumed to represent 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 extractions at the spatial (region) and temporal (quarter) resolution of the model (Peatman et al., 2023b). However, for the index fisheries, while the same aggregation process is conducted, the size data are weighted by standardised CPUE (rather than by catch) so that the size 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 extraction fisheries are effectively being used twice (but weighted differently), the likelihood weighting for the size composition is adjusted such that the original intended weight (effective sample size) in the likelihood is preserved.

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 and reported for these fisheries. The catches are aggregated at 5° x 5° and quarterly resolution, with the aggregation process either conducted by SPC, where operational data is available to inform this, or by the particular countries following statistical procedures that are reported to the Commission. For some fisheries, notably those in region 2 - Indonesian/Philippines/Vietnam - operational information on quarterly or spatial patterns in catches is poor so the annual catches are aggregated evenly across quarters and spatial cells. This is done by SPC.

In the catch-conditioned model, effort is not essential but is required (at least for a recent period of time) for projection analyses involving fisheries managed under effort rather than catch controls. The effort data are necessary to derive recent estimates of catchability for running the effort based projections. In this case the main industrial purse seine fisheries operating in the tropical region (i.e., regions 2, 3, 4) are managed under effort control. Effort data for these purse seine fisheries are defined as number of sets specified by set type (associated or unassociated), and are included for the last 12 quarters to facilitate projections. The period of 12 quarters is consistent with previous projections using catch errors models. For this assessment several other fisheries also have effort included to allow effort based projection for management purposes, these are the longline extraction fisheries, with effort measured as numbers of hooks per set, and the Japanese pole and line fishery with effort measured in vessel fishing days.

Total annual catches by major gear categories for the WCPO are shown in Figure 3 and a regional breakdown is provided in Figure 4. Catches by fishery groups are provided in Figure 7, Figure 8 and Figure 9. The spatial distribution of catches over the past ten years is provided in Figure 5. Discarded catches are estimated to be minor and were not included in the analysis. Catches in the northern region are low and highly seasonal and the annual catch has been relatively stable over much of the assessment period. Most of the catch occurs in the tropical regions (2, 3, and 4).

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

- The steady increase in total yellowfin catch over most of the assessment period, with the highest overall catches reported in the most recent years.
- The steady increase in catch for the domestic fisheries of Indonesia and the Philippines (region 2) since 1970, where mostly small juveniles are taken, and more significant increase in the catches over the last 15 years. Some of this trend can be related to improved information to estimate catches.

- The relatively stable and low catches of yellowfin in the northern and southern temperate regions by longline vessels (regions 1, and 5).
- The development of the equatorial purse-seine fisheries from the mid-1970s, and corresponding increased catches, particularly in equatorial regions, 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 likely efficiency of the fleet.

4.5.2 Purse seine

For the industrial purse seine fisheries predominantly operating in tropical regions 3, 4 and 8, catch by species within each set type (associated or unassociated) is determined by applying estimates of species composition from observer-collected samples to total catches estimated from raised logsheet data (Hampton and Williams, 2016; Peatman et al., 2021, 2023c). For the Japanese (JP) fleet for which there is greater confidence in species-based reporting, reported catch by species is used. Purse seine catch for Philippines (PH) and Indonesian (ID) domestic purse seine fisheries, predominantly operating in Region 7, was derived from raised port sampling data provided by these countries. We note that the COVID-19 pandemic resulted in low observer coverage of the purse seine fleets for the last two years of the assessment period. The implications of the low observer coverage on the purse seine catch composition estimates could not be fully explored under the time constraints, but preliminary analysis suggest the estimates have been relatively robust to the lower observer coverage (Hamer, 2023).

4.5.3 Longline

For the longline fisheries catches in number of fish by species are derived from raised logbook data or annual catch estimates provided by specific countries. Effort is in terms of hooks per set.

The longline CPUE indices are one of the most important inputs to the assessment as they provide indices of abundance over time for each region. The CPUE indices are implemented as "index" fisheries where they are assumed to have the same catchability and are grouped to provide information on biomass scaling among the model regions.

The index fishery CPUE time series for the 2023 assessment were derived from the operational longline dataset for the Pacific region. This dataset is an amalgamation of operational level data from the distant-water fishing nations (DWFN), United States, Australian, New Zealand and Pacific Island countries and territories (PICTs) longline fleets operating in the Pacific basin. It represents the most complete spatiotemporal 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 dataset 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 and 2020 WCPFC bigeye and yellowfin tuna stock assessments (McKechnie et al., 2017; Ducharme-Barth et al., 2020b). Since 2017 spatiotemporal approaches have been used for CPUE modeling in WCPFC stock assessments (Tremblay-Boyer et al., 2017; Ducharme-Barth et al., 2020b). For this assessment we build on these previous efforts and have transitioned from using the VAST software (Thorson, 2019; Thorson et al., 2015) for these analyses to using the sdmTMB package (Anderson et al., 2022). sdmTMB was preferred over VAST due to it's greater computational efficiency, ease of use, and the ready availability of online support from a larger user community than VAST.

A detailed description of the methods for generating the spatiotemporal abundance indices is provided in Teears et al. (2023). Briefly, it was first confirmed that the sdmTMB package could closely replicate the previous VAST indices using the data from the 2020 assessment. After this step a model was run with an increased density of mesh knots (371 versus 154) and the same spatiotemporal subsampling design as the previous assessment. Following this, further exploration of alternative models was conducted considering additional covariates in addition to those applied in the 2020 assessment. These included density covariates of SST, depth of the 15° C isotherm, and difference between the depth of the 12° C isotherm and the 18° C isotherm. As per the previous assessment, catchability covariates of hooks between floats (HBF) and vessel FLAG were included. A vessel ID covariate was considered, but there were over 6,000 unique vessel IDs and this was not considered computationally feasible. El Niño Southern Oscillation data were also included as a potential covariate but caused model instability and therefore, this was not included in the analyses. A model selection process described in Teears et al. (2023) was followed and the final model for yellowfin included HBF, vessel FLAG, depth of 15° C isotherm, and the difference between the depth of the 12° C isotherm and the 18° C isotherm.

In response to the yellowfin peer review, two additional analyses were conducted. One analysis involved running separate models for northern, equatorial, and southern regions with 'non-viable' (poorly sampled) 5° × 5° grid cells removed and comparing the predictions to the results of the same aggregated northern, equatorial, and southern regions from the Pacific-wide indices. Results indicated differences in spatial characterization, although the differences were in areas with comparatively low abundance and had limited implications. An analysis was also conducted comparing a principal-fleet model (Japanese fleet only) to the multi-fleet results to assess the effects of combining fleets. The indices derived from multiple fleets were very similar to the principal-fleet results. It was decided that the outcomes of these additional analyses did not warrant changing the initial approach (see Teears et al. (2023)).

4.5.4 Other fisheries

Effort data for the ID, PH, and VN surface fisheries and Japanese research longline fisheries are unavailable. However, as these fisheries are not part of the index fisheries, the catch-conditioned approach does not require effort data for these extraction fisheries. Catch estimates for the ID/PH/VN fisheries are derived from various port sampling programmes dating back to the 1960s for ID and the PH, and early 2000s for VN (Williams and Ruaia, 2023).

4.6 Size data

Available length-frequency data for each of the fisheries were compiled into 95 x 2cm size classes from 10–12 cm to 198–200 cm. Weight data were compiled into 200 x 1 kg size classes from 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 Macdonald et al. (2023b). Data were either collected onboard by fishers, through observer programs, or through port sampling. Each size-frequency record in the model consisted of the actual number of yellowfin tuna measured and Figure 6 provides details of the temporal availability of length and weight-frequency data (also see Teears et al., 2023). Note that a maximum effective sample size of 1,000 is implemented in MFCL when using the robust normal likelihood for size composition data. The effective sample size was further down-weighted as explained in subsubsection 5.5.2. Summaries of the available size composition data by year and fishery are provided in Figure 12 and Figure 13.

4.6.1 Purse seine

Only length-frequency samples are used in the yellowfin assessment for purse seine fisheries. Prior to 2014, the assessments used only observer samples which had been 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 the current assessment is described in Peatman et al. (2023b). Length-frequency data were unavailable for the "all flags" associated purse seine fishery in region 2 (Fishery 30). In the model, it was assumed to share a selectivity with the "all flags" associated purse seine fishery in the adjacent region 4 (Fishery 13).

4.6.2 Longline

A review of all available longline length and weight-frequency data for yellowfin was undertaken by McKechnie (2014). Details on the data and analytical approach used to construct the size data inputs for the current assessment are in Peatman et al. (2023b) and Teears et al. (2023). The key principle used in constructing the size composition 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 in previous assessments 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 and is not biased by issues such as small catches with lots of weight frequency samples. Weight-frequency data were used over length frequency based on the spatiotemporal coverage and number of samples. However, despite additional weight frequency data being provided by Japan for the 2020 assessment, the number of available weight-frequency samples has declined in recent years. The 2020 assessment conducted a sensitivity analysis involving switching from weight to length-frequency data for the longline fisheries in regions 4, 5, and 6 of the 9 region structure beginning in 2000. The results were relatively insensitive to this change. We suggest that the next assessment could develop a longline size composition data set that optimises the use of both length and weight frequency data with respect of maximising spatial and temporal coverage, and transitioning to length composition data for the recent years.

Size composition data were prepared similarly for the index fisheries (Peatman et al., 2023b). The approach for the index fisheries differed from the one briefly described above for the extraction fisheries in that the size-frequency samples were weighted with respect to the spatial distribution of abundance as predicted by the spatiotemporal CPUE standardization model (Teears et al., 2023). 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.

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 port 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 West Pacific East Asia (WPEA) data improvement project. Additionally, recent data collection efforts in both Indonesia and Vietnam have provided new length-frequency data for inclusion in the recent assessments for both the domestic Indonesia small-scale fishery (Fishery 23) and the domestic Vietnam small-scale fishery in region 2 (Fishery 32).

Size data were missing for the Indonesian-Philippines ex-EEZ purse seine fishery in region 2 (Fishery 24). Based on an investigation of the length frequency data of the other tropical tunas available for this fishery, selectivity was assumed to be shared with the Philippines small-fish fishery in region 2 (Fishery 17) as this fishery had the most similar size composition for the other tropical tuna species.

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 from the current assessment. These large fish were also excluded from the new length-frequency data for both the domestic Indonesia small-scale fishery in region 2 (Fishery 23) and the domestic Vietnam small-scale fishery in region 2 (Fishery 32). 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.

The Indonesia–Philippines domestic handline fishery in region 2 (Fishery 18) consistently catches the largest individuals in the WCPO. Handline fishing often 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 mis-reported as being caught using a 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 reasonable amount of tagging data is available for yellowfin tuna, although it is mostly constrained to the tropical region. Information on the yellowfin tag data characteristics and the process of

constructing the MFCL tagging file are available in Peatman et al. (2023a); Teears et al. (2023). A summary of the tagging data is in Figure 14, and maps displaying tag displacements are in Figure 2. Data were available from the Regional Tuna Tagging Project (RTTP) during 1989–92 (including affiliated in-country projects in the Solomon Islands, Kiribati, Fiji and the Philippines), historical (1995, 1999-2001) data from the Coral Sea tagging cruises by CSIRO (Evans et al., 2008), and the ongoing Pacific Tuna Tagging Programme (PTTP) that began in 2006. Data for the PTTP is included up until the end of 2021, with tag releases included until end of 2020 and recaptures until end of 2021. The 2020 assessment added data from the Japanese Tagging Programme (JPTP) conducted by NRIFSF and the Ajinomoto Co. Inc, over the period 2000–2020, and these data are included in the 2023 assessment. The new tagging data for the 2023 assessment comes primarily from PTTP.

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 onboard or via processing and unloading facilities throughout the Asia-Pacific region.

In this assessment, the numbers of tag releases input to the assessment model were adjusted for a number of sources of tag loss, unusable recaptures due to lack of adequately resolved recapture data, estimates of tag loss (shedding and initial mortality) due to variable skill of taggers (i.e., tagger effects), and estimates of base levels of tag shedding and tag mortality. These adjustments are described in more detail in Peatman et al. (2023a). An additional issue for the yellowfin assessment is that there are tag returns that were released within the WCPO but recaptured to the east of longitude 150°W, outside the WCPO assessment region. The adjustment or rescaling of releases for recaptures in the EPO preserves the recovery rates of tags from individual tag groups that would otherwise be biased low given that a considerable 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, Peatman et al. (2023a) and Teears et al. (2023) describe the approaches to prepare the tagging data. Additionally, the model used to adjust tag releases due to variability in tagger ability or "tagger effects" has changed from that used in the 2020 assessment. This change was the outcome of an expert workshop to review and recommend the approach for modelling tagger effects and providing the correction factors to adjust the tag release numbers (Peatman et al., 2022). The approach recommend from that workshop was applied to the 2022 WCPO skipjack assessment (Castillo Jordán et al., 2022) and is applied to this assessment. The new approach differs from that applied in 2020, in that it reintroduces individual tagging events as a term in the model selection process whilst also keeping cruise leg covariates, whose inclusion were supported for PTTP bigeye releases. It also estimates separate models for central Pacific and western Pacific cruises, given their difference in tagging platforms and associated station and tagger effects, but fits models pooling both yellowfin and bigeye tuna releases, allowing species-specific differences in tagging effects to be accounted for where supported by the data. These

changes result in stronger tagger effects being predicted and therefore generally larger adjustment (reductions) to tag releases. This has an important effect of increasing the recapture rates, which has implications for model estimation of fishing mortality and population scale.

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, any release events occurring after the end of 2020 were excluded from the assessment, as noted.

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.

The likelihood penalties or "priors" used for the reporting rates of the grouped tag return fisheries has been updated relative to those used in the previous assessment based on the analysis of tag seeding experiments (Peatman, 2023). Tag reporting was assumed to be similar between the RTTP and CSTP (which were actually targeted cruises of the RTTP) so reporting rates estimates were shared across these two programs to reduce model dimensionality. For this assessment we have also excluded tag release groups with 5 or less recaptures from the estimation of reporting rates, as we felt there was insufficient information to inform model estimation of the reporting rates. Tag reporting rate groupings are included in Table 1.

5 Model description

5.1 General characteristics

The model can be considered to consist of several components, (i) the dynamics of the fish population; (ii) the fishery dynamics; (iii) the dynamics of tagged fish populations (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), and summarised below. In addition, we describe the procedures followed for estimating the parameters of the model and the way in which stock status conclusions are drawn relative to a series of reference points.

5.2 Population dynamics

The model partitions the population into spatial regions, depending on the specified spatial stratification, 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 1952–2021. The main population

dynamics processes are as follows.

5.2.1 Recruitment

Recruitment is defined as the appearance of age-class 1 quarter fish (i.e. fish averaging $\sim$ 20 cm FL) in the population. Yellowfin tuna spawning does not typically follow a clear seasonal pattern, especially in the tropical waters where most of the spawning occurs. Rather it occurs sporadically all year and is thought to be influenced by food availability (Itano, 2000). The assessment model assumed that recruitment occurs instantaneously at the beginning of each quarter. This is a discrete approximation of continuous recruitment, but provides sufficient flexibility to allow a range of variability to be incorporated into the estimates as appropriate.

The proportion of total recruitment occurring in each region was initially set relative to the variation in average regional catch and then estimated during the later phases of the fitting procedure. Time-series variability in this proportion was estimated within the model and allowed to vary in a relatively unconstrained fashion.

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 to reduce the instability that has been detected by retrospective analyses. This approach has been continued here with the terminal six recruitments fixed at the geometric mean, which is appropriate for a log-normally distributed random variable.

Spatially-aggregated (over all model regions) recruitment was assumed to have a weak relationship with annual mean 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 the steepness parameter of the SRR parameters (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. The high CV (2.2) of the log-recruitment deviates, computed annually, ensured that the SRR had negligible impact on the estimation of recruitment and other model parameters, as recommended by Ianelli et al. (2012). The SRR was estimated over the period 1962–2020 to prevent the earlier recruitments (which may not be well estimated), and the terminal recruitments (which are not freely estimated), from influencing the relationship.

The SRR was incorporated mainly so that yield analysis and population projections could 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 (recommendation 20) by the 2011 bigeye assessment review (Ianelli et al., 2012).

5.2.2 Initial population

The population age structure in the initial time period in each region was assumed to be in equilibrium. For this assessment the calculation of the equilibrium initial population was changed to no longer use the average total mortality (Z) for the first 20 time periods for the equilibrium initial population. Instead, for the initial equilibrium condition we applied an initial Z value equal to the natural mortality (M). The change to the settings for the initial population was introduced as part of the stepwise model development to implement the fully catch conditioned 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 it possible for accumulation of old and large fish, which is likely in the early years of the fishery when exploitation rates were very low.

5.2.3 Growth

The standard assumptions for WCPO assessments fitted in MFCL were made concerning age and growth: i) the lengths-at-age are normally distributed for each age-class; ii) the standard deviations of length for each age-class are a log-linear function of the mean lengths-at-age; and 3) the probability distributions of weights-at-age are a deterministic function of the lengths-at-age and a specified weight-length relationship[6]. These processes are assumed to be regionally and temporally invariant.

In the previous assessment several approaches to growth estimation were explored, including an external otolith based growth curve with a fixed Richards growth curve based on high-readability otoliths (Farley et al., 2020), a fixed external Richards growth curve based on the same high-readability otolith dataset but with the addition of tag-recapture growth increment data (Eveson et al., 2020), an internal MFCL growth estimation using a conditional-age-at-length (CAAL) dataset from the otolith dataset including daily and annual ages (Farley et al., 2020), which also made use of the modal progressions apparent in various size composition data (Vincent et al., 2020).

The assessment was sensitive to approaches used for growth and the internal CAAL estimate of the von Bertalanffy growth curve was the approach used for the diagnostic model.

The peer review of the yellowfin assessment (Punt et al., 2023) carefully considered the approaches to estimating growth. They concluded that based on how the otolith samples were selected according to length bins (i.e., attempting to achieve sufficient numbers of otolith samples for length

[6]The length-weight relationship has been updated for the current assessment based on an analysis of current and historical port sampling data under WCPFC Project 90 (Macdonald et al., 2023b)

bins across the length range, plus bias in otolith readability with age), that any external growth curves would likely be biased. They strongly recommended that for this type of otolith sampling regime the data should be used as CAAL data and the growth be estimated by the MFCL model if this is possible. We also note that in exploring the internal growth estimations we uncovered some features with the MFCL code for the internal Richards growth estimation that require attention, however, this could not be dealt with under the time constraints. For this assessment we use the CAAL and size composition data to estimate the von Bertalanffy growth curve internal to the MFCL model similar to the 2020 diagnostic model. When estimating the growth curve, similar to the previous assessment, we found the L_1 parameter tended to hit the lower bounds of 20 cm FL, so it was necessary to fix L_1 at 19.8 cm, consistent with the average size at the age of one quarter from the otoliths (Farley et al., 2020).

5.2.4 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., 2019 for details). Movement is parameterized by a pair of bi-directional coefficients at each region boundary. Movement is therefore possible in both directions across each regional boundary in each of the four quarters. Each of these coefficients is estimated independently resulting in 104 estimated movement parameters for the 9 region spatial structure and 76 for the 5 region spatial structure (adopted as the 2023 diagnostic structure) (2×no.region boundaries (13)×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. A "prior" of 0 is assumed for all movement coefficients, and a low penalty is applied to deviations from the "prior". We had hoped to explore different settings for modelling movement, as this is an area of uncertainty, however, this was not possible to do properly under time constraints.

5.2.5 Natural mortality

The instantaneous rate of natural mortality (M) consists of an average over age classes and age-specific deviations from the average M. Average M can be estimated internally by the model or specified as a fixed external value. Internal estimation can be constrained by providing a prior value and a penalty weight for deviations from the prior. Age specific deviations from average M can also be estimated or input as a fixed M-at-age function. The later approach was taken in the 2020 assessment for both bigeye and yellowfin (Ducharme-Barth et al., 2020a; Vincent et al., 2020) where 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 Nicol, 2008). The peer review had some difficulty with tracing the sources of data for some of the inputs to the external M-at-age function applied in 2020 that requires fitting a model that depends on empirical data on the sex-ratio at length, a growth curve, a base M assumption for males, assumptions on critical lengths for inflections and a multiplier that determines the linear decline in M for young ages to the base M, plus a length at which female mortality is assumed to begin to increase. There are uncertainties in all these empirical data and assumptions and the data for sex ratio is limited especially for larger and older fish. While the peer review generally supported the method, the construction of the external M-at-age function, notably the limitations in some of the available data required to estimate it, the complexity of the calculations and the various assumption required are problematic. Furthermore, the external M-at-age function requires re-estimation if different growth curves are applied, further complicating its use. After the yellowfin peer review, a Tuna Stock Assessment Good Practices workshop run by the "Center for the Advancement of Population Assessment Methodology" (CAPAM) was held (March 2023, https://www.capamresearch.org/Tuna_Stock_Assessment_Good_Practices_Workshop) and provided a strong endorsement for applying a simpler Lorenzen functional form for M-at-age estimation for tuna. Other CAPAM reviews of approaches for estimating M have also favoured where possible estimating M within the model, "let the data speak for themselves", while being careful to review model diagnostics and plausibility of model estimated M against external estimates (Hamel, 2023; Punt, 2023).

In this assessment we chose to use the Lorenzen functional form for M-at-age (with M-at-age being inversely proportional to the mean length at age) and to estimate the asymptotic value of M, i.e. the M for the oldest fish. The change from the previous external M-at-age to the Lorenzen form with internal M estimation was made as part of the stepwise development of the 2023 diagnostic model.

5.2.6 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). Assumptions about maturity do not affect the process of fitting the model, only the reference point values. The approach for calculating the reproductive potential at length ogive in this assessment is the same as the 2020 assessment (Vincent et al., 2020). MFCL estimates the reproductive potential at age ogive internally from the externally calculated reproductive potential ogive at length. This length-based ogive is converted internally to the 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 pro-

cesses: proportion females-at-length[7] (sex-ratio), proportion of females mature-at-length[8], and the fecundity-at-length of mature females[9] (Figure 16). Another added benefit is that this reproductive potential ogive is growth invariant. The previous stock assessments had to redefine the reproductive potential-at-age ogive for each different growth curve included in the assessment. As for the 2020 assessment we have not included spawning fraction in the reproductive potential calculation as the data are still not adequate for yellowfin in the WCPO.

5.3 Fishery dynamics

5.3.1 Selectivity

Fishery selectivity coefficients at length and weight (for longline) are included in Figure 34 and Figure 36 and at age in Figure 33 and Figure 35. Selectivity was modeled using a cubic spline, as in the previous assessment. 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. The number of nodes may vary among fisheries to allow for reasonably complex selectivity patterns, however we found that 5 nodes seemed sufficient for all fisheries in this case. Some other modifications to selectivity curves were made to better fit size composition data and these are noted in subsection 6.1.

In all cases, selectivity was assumed to be time-invariant and fishery-specific. However, a single selectivity function could be "shared" among a group of fisheries that have similar length compositions or were assumed to operate in a similar manner. This grouping facilitates a reduction in the number of parameters estimated and can provide insight into the regional abundance of fish of specific sizes. Selectivity groupings are indicated in Table 1.

While length-based selectivity is not currently permitted in MFCL (Note: efforts to implement this feature in 2023 were made in response to a peer review recommendation but there was insufficient time to complete and test this feature), the age-based selectivity functions are penalized such that selectivity of age-classes that are similar in size will have similar selectivities for a given fishery or group of fisheries. Additionally, the assumption was made that at least one fishery within each of the model regions 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. In the previous assessment this was typically one of the extraction longline fisheries, though in region 2

[7]For the current assessment, female sex-ratio-at-length was calculated from Regional Observer Program data in SPC's holdings through to 2018, consistent with the previous assessment as few data have been collected since 2018.
[8]Taken from Farley et al. (2017) as in the previous assessment.
[9]Taken from Sun et al. (2006) and standardized per kg of body weight at length as in the previous assessment.

(region 7 of the 9 region structure) it was assumed to be the Indonesia-Philippines handline fishery (Fishery 18), plus the index fisheries were also assumed to have selectivity penalised to be non-decreasing with length. However, in this assessment we chose not to penalize any of the extraction fisheries to have non-decreasing selectivity with increasing length and only apply this to the index fisheries which are assumed to share selectivity across regions (see further subsection 6.1).

5.4 Dynamics of tagged fish

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.

5.4.1 Tag reporting

In theory, tag-reporting rates can be estimated internally within the stock assessment 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. We provided reporting rate priors for reporting groups (combinations for tagging programme and fishery) that reflect independent estimates of the reporting rates and their variance (Peatman, 2023). We also make some assumptions regarding fisheries that were similar to those with independent estimates, but increased the prior variance. For others where we felt there was very little information to inform priors and variance, uninformative priors were allocated, or in cases where there were 5 or less tag returns, those groups were not estimated. In these cases, the small numbers of tag returns were removed from the tagging data input file and the reporting rate for these groups set to a fixed value of zero. The prior reporting rates and penalty terms were informed from analyses of tag seeding experiments reported in Peatman (2023). 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 0.99 (increased from 0.9 in previous assessment, see subsection 6.1). For this assessment, as previously noted, we did not estimate reporting rates for tag reporting groups with five or few tag return. Tag recapture and reporting rate groupings are provided in Table 2. Previous assessments have assumed fishery-specific reporting rates are constant over time. This assumption was reasonable when most of the tag data were associated with a single tagging program. However, tag reporting rates may vary considerably between tagging programs due to changes in the composition and operation of individual fisheries, and different levels of awareness and follow-up. Consequently, fishery-specific tag reporting rates that are also specific to individual tagging programs were estimated.

5.4.2 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 with the untagged population to have taken 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. It is therefore desirable to designate one or more time periods after release as "pre-mixed" and compute fishing mortality for the tagged fish during this period based on the actual recaptures, corrected for tag reporting and tagging effects, rather than use fishing mortalities based on the general population parameters. This, in effect, desensitizes the likelihood function to tag recaptures in the specified "pre-mixed" periods while correctly removing fish from the tagged population that is present after the "pre-mixed" period. Herein we refer to the "pre-mixed" period as the "mixing period".

The allocation of appropriate "mixing periods" in stock assessments that use tag-recapture data is problematic as model estimations are sensitive to this assumption and misspecification can bias estimation of management quantities (Kolody and Hoyle, 2014). Mixing rates may vary depending on release locations and times depending on various factors, including; oceanographic dynamics, contexts of releases (e.g., FADs versus free schools, archipelagic waters versus oceanic), fishing effort distribution and environmental/food conditions that influence movements. The yellowfin peer review discussed mixing period assumptions and supported an approach applied to the 2022 skipjack assessment (Castillo Jordán et al., 2022; Punt et al., 2023). In this approach an external individual based model was used to estimate mixing periods 'specifically' for each release group, taking into account the locational and temporal varying conditions of each release event constituting the group that may result in different rates of mixing of the released fish (Scutt Phillips et al., 2022). It applied the individual based Lagrangian model (Ikamoana) (Scutt Phillips et al., 2018) to simulate movement of individual fish (particles) and quantify the fishing pressure (observed) that individuals experienced across their dispersal trajectories in comparison to a population of simulated untagged particles. While this approach is promising, it requires substantial work to develop and it was not practical or possible to develop this approach for the current assessment,

but could potentially be explored in future.

In the previous yellowfin assessments the diagnostic model assumed 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 the quarter in which the fish were releases (ie., "mixing period" assumption is two quarters). A sensitivities was include whereby the tag mixing period assumption was one quarter (Vincent et al., 2020). Discussion at the PAW (Hamer, 2023) recommended that external analysis of tag-recapture patterns would be useful to inform consideration of tag mixing assumptions for yellowfin as previous work on skipjack suggested that acceptable tag mixing may take longer than two quarters (Kolody and Hoyle, 2014). In response to this the background analysis paper for this assessment includes a series of maps (see appendices, Teears et al. (2023)) that show tag release distributions and their related recapture distributions for individual tagging cruise and model regions at 0, 1, 2, 3 and 4 quarters after release (0 quarter meaning recapture with the same quarter of release, 1 quarter meaning recapture in the first quarter after release etc.). The recaptures are scaled to the purse seine catches (the purse fishery accounts for 94% of yellowfin tag returns) and plotted as numbers of tags recaptured per 100 mt of catch in 1 x 1 degree grids cells. The maps provide spatial distributions of relative tag recapture rates within a model region, similar to the approach of Langley and Million (2012). When recaptures are observed in 1 x 1 cells spread throughout the model region in relation to the catch distribution and with roughly similar rates of recapture it could be considered a qualitative diagnostic that tag mixing was achieved. These plots were used to assign indicative mixing periods for fish released from the individual tag cruises and these were summarised to provide an indication of 'reasonable' mixing period assumptions to apply in sensitivity analyses (Teears et al., 2023). As further information to support this qualitative assessment, the distributions of displacement distances of tag recaptures at 1, 2, 3, 4, 5, and 6 quarters after release, for releases from model regions 3, 4, 7 and 8 of the 9 region model (see Teears et al. (2023)) were considered.

Overall, for tag releases with good numbers of recaptures these qualitative analyses supported mixing period sensitivities of 1, and 2 quarters (Teears et al., 2023), consistent with the previous assessment. Some tag releases were likely not mixed until at least 3 quarters, which could be considered as a third sensitivity, although perhaps less supported than 1 and 2 quarters. It was noted that some tag release groups showed more obvious evidence for mixing than others depending on how many tags were released (i.e., the qualitative assessments were likely more reliable for larger tag releases). While these qualitative assessments of mixing period assumption could be improved with more time, they provide a stronger basis for mixing assumptions than in the previous assessment.

As per the 2020 assessment the tag return files were created using a sliding window to calculate the mixing period for each release group as per the recommendation of the 2020 PAW, which was

further endorsed by the yellowfin assessment peer review (Punt et al., 2023). This approach ensures that the mixing periods are applied faithfully in respect of actual times at liberty. It means that tags had to have a time at liberty of at least 182 days for a mixing period of 2 quarters as is assumed in the diagnostic case, or 91 days for a mixing period of 1 quarter.

5.5 Likelihood components

There are four data components that contribute to the log-likelihood function for the yellowfin stock assessment: the index fishery CPUE data, the length and weight frequency data, the tagging data and the conditional age at length data.

5.5.1 Index fishery CPUE likelihood

In previous catch-errors models, abundance indices were constructed for extraction fisheries by assuming that catchability remained constant over time. In catch-conditioned models, a new approach has been implemented to model directly the CPUE for 'index' fisheries. While such index fisheries exist in the model and in the data inputs as defined 'fisheries', they differ from the regular 'extraction' fisheries in that no catch is extracted and their CPUE is modelled directly as a lognormal likelihood contribution. The CPUE index series for each region is assigned a likelihood weight in the form of a region-specific σ, describing the magnitude of observation noise. The computational procedure for estimating σ is based on maximum likelihood estimation, by calculating σ as the standard deviation of log-residuals. The residuals are from a fitted model in the stepwise development where data weighting is adjusted. These maximum likelihood estimates of region-specific σ are then kept the same throughout the stepwise development, diagnostic model, and structural uncertainty grid.

5.5.2 Length and weight frequency

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 or weight-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. The observed sample sizes (OSS) are capped at 1,000 internal to MFCL, the sample size for the composition data used in the common index and extraction fisheries was divided by two to account if being used twice, and then the sample sizes were further divided by 20 for the diagnostic model, resulting in a 'maximum' ESS of 50 for each length or weight sample for a fishery. Alternative divisors for specifying ESS were

explored in sensitivity analyses, these were 10 and 40. We further explored a self-scaling method for weighting the size composition date, the Dirichlet-multinomial (DM) likelihood (Thorson et al., 2017). The DM likelihood approach tended to put high weight on the size composition data leading to deterioration of the fits to the abundance indices. We therefore decided not to apply the DM likelihood weighting method, which requires further exploration. However, based on the testing of the DM weighting, the divisor of 10 would be more consistent with the DM weighting than the other two sensitivities.

5.5.3 Tag data

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 (τ) such that as it approaches 1, the negative binomial approaches the Poisson. In the current assessment we assume $\tau = 2$, which is a variance twice that 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 provide a more realistic weighting of the tag return data in the overall log-likelihood and allow the variability to impact the confidence intervals of estimated parameters.

5.5.4 Conditional age-at-length data

A further log-likelihood component involves the conditional age-at-length (CAAL) dataset, the inclusion of which was recommended by the yellowfin tuna assessment peer review (Punt et al., 2023). These data are 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 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. However, despite the otolith data being collected across a range of locations and times the CAAL data may have a degree of non-independence so we have downweighted the otolith samples for the diagnostic model by multiplying the input samples size by 0.75. We also ran sensitivities with downweighting of 0.5 and no downweighting (i.e,. 1). That latter is consistent with the 2020 yellowfin assessment.

5.6 Parameter estimation and uncertainty

The parameters of the model were estimated by minimising the sum of the negative log-likelihoods associated with each of the data components plus the negative log of the probability density functions of the priors and penalties specified in the model. The optimisation to a specified gradient level or to a maximum number of function evaluations, if the specified gradient level is not met, 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. The values of these parameters are provided in the .ini input file.

As a rule, models were run with a gradient criterion of 10^{-5} and a maximum of 10 000 function evaluations. This ensured that model runs would complete in less than 24 hours, allowing a daily routine of continuous model development. During the stepwise development, some model fits were later demonstrated to be at sub-optimal local minima when fits with improved objective functions were subsequently obtained by running several replicates with selected parameters perturbed, in a process referred to as jittering, and continuing the optimisation process with additional function evaluations for each replicate. In many cases, the new jittered solution that had the best objective function value was a considerably different model fit in terms of derived management quantities. This was clearly cause for concern and required additional work to achieve parameter estimates that we could be confident had achieved a fit in terms of obtaining a good objective function value and reliable and stable estimates of the derived quantities of most interest for management advice. A thorough jittering process was therefore adopted for the diagnostic model and the structural uncertainty grid, selecting the model fit that has the best objective function value. These best fits may sometimes have worse gradients and Hessian properties than models at sub-optimal local minima. Thus, a positive definite Hessian (PDH) solution was less important in this process than considering the best value for the objective function and stable estimates of management quantities. We reached this conclusion after obtaining several apparently well-converged model solutions with PDH's that were subsequently improved, in some cases substantially, by jittering.

Additional steps were incorporated in a process adopted to find estimated model parameters with an improved objective function value, compared to solutions obtained without jittering. Due to the sensitivity of these models to initial conditions, an iterative process was adopted where key parameters for growth and mortality were reset close to values obtained from a previous run of the model. These values could either be reset at the initial phase or in an intermediate phase of the optimisation. This process was combined with repeated cycles of jittering, with up to 60 replicate jitters, and on occasions a secondary round of jittering on the best jittered solution found from the

primary jitter. Typically, improvements in likelihood were routinely found in the primary round of jitters, sometimes with improvements to the objective function of tens to hundreds of units of log likelihood. Improvements in the secondary round of jittering (aka 'twerking') could sometimes be found, but these were usually less than 10 units of log likelihood. Typically the secondary round of jittering produced little change to the stock status, but the primary jittering could produce changes of 5% or more to the spawning biomass depletion stock status indicator. This process was used to select the 2023 diagnostic model and resulted in improvements to the objective function and overall likelihood. However, this added significant time to the assessment work. Once a stable 'good fit' model was obtained, we calculated the Hessian. This process resulted in a well-converged, jittered diagnostic model with a PDH (see subsection 6.1).

The 2020 assessment ran a jitter analysis on the diagnostic model but did not find the same instability of solutions in relation to changes to the estimated stock status. This greater stability noted in 2020 may relate to that model using fixed (external) M-at-age and a fixed (externally estimated) growth curve, or perhaps there was insufficient time to investigate jittering more thoroughly.

This exhaustive process adopted to achieve a suitable diagnostic model in 2023, also meant that a single fit of a model in the structural uncertainty grid, while perhaps indicative, could not really be trusted, irrespective of the objective function value, gradient or Hessian status. While the grid models based off the diagnostic model are generally expected to have good starting parameter values, a jittering step for each model in the grid is now considered important to increase confidence that these models had also achieved stable solutions, both in terms of an improved objective function value and reliable management quantities. However, under time constraints the jittering of grid models was reduced to 20 jitters per model, with the Hessian calculated on the best jittered solution, where 'best' is judged solely on the objective function values. The requirement to jitter and run Hessians on grid models again compounded the workload and computation time required for the assessment. However, this was all required to meet recommendations of the yellowfin peer review and the SC18 regarding model convergence.

Finally, we considered the SC18 concerns over a PDH being a mandatory diagnostic that needs to be achieved for a model to be included for management advice. We argue that while a PDH is desirable it is not as important as a stable 'good fit' objective function and stable management quantities for these complex assessment models. Often the lack of PDH in these models with several thousand parameters estimated, is due to one or a few very small negative eigenvalues that have no influence on the estimations of the key management quantities. MFCL has advanced methods for computing estimation uncertainty for Hessians with negative eigenvalues, allowing the parameter estimation uncertainty to be calculated for Hessians with small numbers of small negative eigenvalues. We have shown by comparing pairs of similar models with the same structure, but with slightly different fits, one with and one without a PDH, the parameter estimation uncertainty is very similar (e.g.,

appendix 15.3 Castillo Jordán et al. (2022)). While we expect that a PDH solution is an important diagnostic to meet for the diagnostic model, this criterion may not be essential (or practicable) for all grid models, especially if only a small number of small non-influential negative eigenvalues are found.

Estimation uncertainty

A positive definite Hessian is desirable to calculate parameter or estimation uncertainty for individual models, although as noted MFCL has advanced procedures for computing estimation uncertainty in the absence of a positive definite Hessian. Hessian calculations can take considerable computer time for these models, however, SC18 requested the Hessian status be reported for all models used for management advice (which includes all grid models). This has been completed in this assessment but has added significant time, and computational burden, leading to delays in finishing the assessment.

Structural uncertainty

The structural uncertainty grid attempts to describe the main sources of structural and data uncertainty in the assessment. Previous experience has shown that overall uncertainty is dominated by the structural uncertainty grid. For this assessment we have continued with a factorial grid of model runs which incorporates selected uncertainties explored in one-off sensitivity analyses.

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

Likelihood profiles

For highly complex population models fitted to large amounts of often conflicting data, it is common for there to be difficulties in estimating total abundance. Therefore, a likelihood profile analysis was undertaken of the marginal posterior likelihood in respect of population scaling, following the procedure outlined by McKechnie et al. (2017) and Tremblay-Boyer et al. (2017). The results of these procedures are presented in the appendices (Appendices subsection 12.1). Likelihood profiles are only presented for the diagnostic model.

Retrospective analyses were conducted 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 quarters sequentially excluded (Cadigan and Farrell, 2005). The retrospective analyses for the 2023 diagnostic model are presented in the appendices (Appendices subsection 12.2).

5.7 Stock assessment interpretation methods

5.7.1 Depletion and fishery impact

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

This problem is better approached by computing the spawning potential time series (at the model region level) using the estimated model parameters, but assuming that fishing mortality was zero. Because both the estimated spawning potential SB_t (with fishing), and the unexploited spawning potential $SB_{t,F=0}$, incorporate recruitment variability, their ratio at each quarterly time step (t) of the analysis, $SB_t/SB_{t,F=0}$, 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 - longline, purse seine associated sets, purse seine unassociated sets, pole and line 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.

5.7.2 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 $SB_{F=0}$ for this assessment is calculated as the average over the period 2012–2021. There is no agreed WCPFC target reference point for the yellowfin tuna stock however CMM 2021-01 states in para 11 "Pending agreement of a target reference point the spawning biomass depletion ratio ($SB_{F=0}$) is to be maintained at or above the average $SB_{F=0}$ for 2012–2015". Stock status was referenced against these points by calculating the reference points; $SB_{\text{recent}}/SB_{F=0}$ and $SB_{\text{latest}}/SB_{F=0}$ where $SB_{F=0}$ is calculated over 2011–2020 and SB_{recent} and SB_{latest} are the mean of the estimated spawning potential over 2018–2021, and 2021 respectively (Table 4).

The other key reference point, $F_{\text{recent}}/F_{\text{MSY}}$, is the estimated average fishing mortality at the full

assessment area scale over a recent period of time (F_{recent}; 2017–2020 for this stock assessment) divided by the fishing mortality producing MSY which is a product of the yield analysis and is detailed in subsubsection 5.7.3.

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

5.7.3 Yield analysis

The yield analysis consists of computing equilibrium catch (or yield) and spawning potential, conditional on a specified basal level of age-specific fishing mortality (F_a) for the entire model domain, a series of fishing mortality multipliers (*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 be determined using the formulae given in Kleiber et al. (2019), and is equivalent to the MSY. Similarly, the spawning potential at MSY SB_{MSY} can be determined from this analysis. The ratios of the current (or recent average) levels of fishing mortality and spawning potential to their respective levels at MSY are determined for all models of interest. This analysis was conducted for all models in the structural uncertainty grid and thus includes alternative values of steepness assumed for the SRR.

Fishing mortality-at-age (F_a) for the yield analysis was determined as the mean over a recent period of time (2017–2020). We do not include 2021 in the average as fishing mortality tends to have high uncertainty for the terminal data year of the analysis and the catch and effort data for this terminal year are potentially incomplete. Additionally, recruitments for the terminal year of the model are constrained to be the geometric mean across the entire time series, which affects the F for the youngest age classes.

MSY was also computed using the average annual F_a from each year included in the model (1952–2021). This enabled temporal trends in MSY to be assessed and a consideration of the differences in MSY levels under historical patterns of age-specific exploitation.

5.7.4 Kobe analysis and Majuro plots

For the standard yield analysis (subsubsection 5.7.3), the fishing mortality-at-age, F_a, is determined as the average over some recent period of time (2017–2020). In addition to this approach the MSY-based reference points (F_t/F_{MSY}), and SB_t/SB_{MSY}) were also computed by repeating the yield analysis for each year in turn. This enabled temporal trends in the reference points to be estimated and a consideration of the differences in MSY levels under historical patterns of age-specific exploitation. This analysis is presented in the form of dynamic Kobe plots and "Majuro

plots", which have been presented for all stock assessments in recent years.

5.7.5 Stock projections from the structural uncertainty grid

Projections of stock assessment models can be conducted within MFCL to ensure consistency between the fitted models and the simulated future dynamics, and the framework for performing this exercise is detailed in Pilling et al. (2016). Typically, stochastic 30 year projections of recent catch and effort (2019-2021) are conducted from each assessment model within the uncertainty grid developed. For each model, 100 stochastic projections, which incorporate future recruitments randomly sampled from historical deviates, are performed. The results of stock projections are included in the appendices (subsection 12.3).

6 Model runs

6.1 Developments from the last assessment

The progression of the model development (referred to as the 'stepwise') from the 2020 diagnostic model to the 2023 diagnostic model is described here. Most steps in the stepwise model development process were implemented using the data inputs from the 2020 assessment, before the new 2023 data had been fully processed and made available. The final input files were ready a few weeks before the SC meeting, with just enough time to adapt the latest development model to the new data and make decisions on the final structural changes to the diagnostic model.

6.1.1 Stepwise model development

The major changes incorporated at each step in the diagnostic model development are summarised below, including the model names used in the figures describing the steps. Each model builds from the previous step, retaining all previous changes. Some steps combine a major change with one or more minor changes, especially those that result in negligible changes in model outcomes. The model names describe an important feature or structural change introduced in that step.

1. **Diag2020**. The 2020 yellowfin diagnostic model, using data from 2020.
2. **NewExe**. The MULTIFAN-CL (MFCL) executable was updated from version 2.0.7.0 (from 17 January 2020) to version 2.2.0.0 (30 June 2023), accumulating three years of software development, including features that were used in this assessment. Later patches of MFCL 2.2.x.0 were released during the stepwise model development period, introducing new features while retaining compatibility with 2.2.0.0. The full backward compatibility meant that it was not necessary to rerun the entire stepwise sequence of models each time a new MFCL patch was released.

3. **PreCatchCond**. A set of four changes were applied to the model structure to aid the transition to a catch-conditioned model, and to ensure stable model behaviour:

   (a) A penalty on regional differences in recruitment was reduced from 1.0 to the MFCL default value of 0.1.

   (b) The maximum mortality rate in the model was changed from the MFCL default value of 5.0 to become a gradual process during the model estimation, where the upper limit starts at 0.7 and is gradually increased to 3.0.

   (c) The assumed fishing mortality before the first year of the model was set to zero instead of the average of the first five years.

   (d) A penalty on movement coefficients was reduced from the MFCL default value of 5.0, applied to differences from zero movement, to a penalty of 0.1 applied to movement coefficients deviating from the prior mean of 0.1.

4. **CatchCond**. Moving from a catch-errors model to catch-conditioned model was an important structural change in this year's assessment. By assuming that the total catches are observed with negligible error, instead of estimating effort deviates as was done in previous models, the number of model parameters decreased from 11,668 down to 2,963. This reduced the run time of each model by approximately one third and the expectation was that the smaller number of estimated parameters would lead to more reliable model convergence, which had been identified as a key challenge in previous assessments. This stepwise change was implemented in two parts, in an attempt to isolate the effect of the structural transformation from the accompanying change in CPUE likelihood:

   (a) The intermediate 'Old CPUE' model implements an initial catch-conditioned model, introducing grouped index fisheries with common catchability but retaining the catch-errors CPUE likelihood, through estimating the relationship between fishing mortality level and effort.

   (b) The fully converted 'New CPUE' model includes a likelihood component that measures directly the goodness of fit to the observed CPUE indices, instead of estimating the relationship between fishing mortality and effort.

5. **SelChanges**. In the 2020 assessment, all index fisheries and some of the extraction longline fisheries were modelled with a penalty constraining their selectivity to be non-decreasing, to avoid cryptic biomass in the model. After experimental analyses and evaluations for this year's model development, however, it was deemed enough to constrain only the index fisheries to avoid cryptic biomass in the model. This step estimated selectivity for all extraction longline fisheries without the non-decreasing constraint.

6. **TagStructure**. Two changes were made to the modelling of tagging data in the assessment model:

   (a) The upper limit of the estimated tag reporting rates was increased from 0.90 to 0.99, in an attempt to relieve the problem of estimated parameters running into bounds.

   (b) Tag release groups were excluded from tag reporting rate estimation if they had 5 or fewer recaptures.

7. **Growth**. The von Bertalanffy growth parameters L_1, L_A, and K, were all estimated in the 2020 assessment. However, the L_1 parameter was estimated at a lower bound that was set at 20.0 cm. When this bound was lowered, the estimated value of L_1 went down to 7.9 cm, well below any observations found in the data. At this stage in the stepwise development, options were considered to alleviate this problem and a decision made to fix the L_1 parameter at the average length of age 1 quarter fish, based on observed length measurements from the otolith data, where $\overline{L_1} = 19.8$ cm ($n = 66$). The growth parameters L_A and K, along with the growth variability parameters σ_1 and σ_A, were estimated in this model and subsequent steps without running into bounds.

8. **DataWeights**. Three adjustments were made to the data weighting in the model:

   (a) CPUE data weights are expressed as σ_I, the magnitude of observation noise around the CPUE indices. In the previous assessment, σ_I was set to 0.20 in all regions. For this assessment, a statistical approach was used to estimate maximum likelihood region-specific σ_I values as the standard deviation of log-residuals. The residuals are from a fitted model at this point in the stepwise development. These maximum likelihood σ_I values are then kept the same throughout the stepwise development, diagnostic model, and structural uncertainty grid. The statistically derived σ_I values are 0.25 on average, varying slightly between regions.

   (b) Age data weights for the otoliths are expressed on a scale between 0 and 1, where 1 would mean that the otoliths sampled were a perfect representation of the yellowfin tuna population across areas, years, quarters, age, etc. To acknowledge that this is not the case, a value of 0.75 was chosen as a reasonable value, which is lower than the value of 1 used in the last assessment.

   (c) Size composition data weights are expressed as a denominator to divide the number of fish measured to approximate an effective sample size for a given fishery in a year-quarter. A denominator of 20 was used in this assessment, lower than the value of 40 from the previous assessment. One reason to change this value was that the size composition data were to be filtered at a later step (see 'FilterSizeComps' below), resulting in a smaller number of measured fish in the final input data.

9. **NatMort**. The natural mortality at age used in the last assessment posed practical problems related to its estimation outside of the assessment model and its dependency on sex-specific catch data, as well as the requirement of updating the analyses whenever growth parameters estimates were updated. Furthermore, the last assessment report commented that the natural mortality rates seemed too high when compared to a meta analysis based on life-history theory and empirical estimates. To overcome these challenges, and to incorporate recommendations from the 2023 CAPAM Tuna Stock Assessment Good Practices workshop, this year's assessment used an internally estimated Lorenzen curve for natural mortality. This was a major structural change in the assessment model, as the shape of the Lorenzen curve differs substantially from the previous curve and the overall scale of the natural mortality rate was estimated rather than fixed.

10. **TaggerEffect**. The recommended tagger effects model, based on the 2022 workshop and simulation study (Peatman et al., 2022, 2023a) with the recommended separate tagger effects models for the western Pacific and central Pacific tagging cruises, was applied to the tagging data. The processing of the tagging data uses a model that is different from the 2020 assessment, reverting to assumptions similar to those used in 2017.

11. **NewCPUEMethod**. This year, the spatio-temporal analysis to produce CPUE indices was updated, adding new covariates and using the sdmTMB modelling platform instead of VAST (Teears et al., 2023).

12. **NewData**. This step included revisions to data from 1952–2018 and added three years of new data from 2019–2021, including new tagging data, minor revisions to data filtering protocols for composition data, new length and weight composition data, and new CPUE data. No new conditional age-at-length data were available since the 2020 assessment. The length-weight conversion factors were updated in a separate stepwise model run before the main data update. Due to the limited time between receiving the finalised new data and the SC meeting, all other data components were updated together rather than one at a time.

13. **FilterSizeComps**. Two changes were made in the treatment of size composition data in this year's assessment, both to length and weight compositions:

    (a) Tail compression was applied to remove the lower and upper tails of all size frequency distributions that only contain zero-frequency observations. This treatment results in size distributions that have the same range as the observed data. When MFCL calculates a predicted size frequency distribution, it accumulates the very smallest and largest sizes and adds them to the first and last observed size bin.

    (b) An MFCL setting was enabled to exclude length and weight frequency samples of less than 50 fish.

The purpose of both these changes was to facilitate the estimation of selectivities so that each fishery removed fish of the right sizes in the stock assessment model. The estimation of selectivity shape parameters has been identified as a challenge in previous assessments and the expectation was that reducing noise from the observed data could result in improved convergence and more reliable parameter estimates.

14. **FiveRegions**. Following recommendations from the 2022 stock assessment review and the 2023 pre-assessment workshop, alternative regional structures were considered for this year's assessment. The conclusion was to adopt a 5 region structure, merging selected regions from the 9 region structure that was used in the last assessment. The new structure substantially reduced model complexity, especially in terms of estimated region-year-quarter recruitment deviations and movement between regions, decreasing the number of model parameters from 3,069 down to 1,901. This change resulted in reducing the run time of each model by approximately one third and the expectation was that the smaller number of estimated parameters would lead to more reliable model convergence, which had been identified as a key challenge in previous assessments.

15. **Diagnostic2023**. The final refinements for the 2023 diagnostic model involved estimating natural mortality and growth parameters in the last estimation phase and updating the starting values of these parameters, based on a jittering analysis. This change resulted in a considerable improvement in the objective function value and a positive definite Hessian.

6.2 Sensitivity analyses and structural uncertainty

6.2.1 Sensitivities

Various one-off sensitivity models were explored to understand the sensitivity of the diagnostic model estimations to structural and data uncertainties. Each one-off sensitivity model was created by making a single change to the 2023 diagnostic model. These sensitivities are described below:

1. Steepness: 0.65, 0.95
2. Tag mixing: 1 quarter
3. Size data weighting: 10, 40
4. Conditional age-at-length data weighting: 0.5, 1

6.2.2 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 that explore the interactions among

selected 'axes' of uncertainty. The grid contains all combinations of levels of several model quantities, or assumptions, and allows the sensitivity of stock status and management quantities to this uncertainty to be determined and factored into management advice. The axes are generally selected from 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 2023 yellowfin stock assessment was constructed from 4 axes of uncertainty with 1–3 levels for each (below), resulting in a total of 54 models (Table 3). The previous assessment included axes for steepness (same values as current assessment), growth (internal estimate based on length modes, external otolith growth curve, conditional age-at-length internal), size data weighting (20, 60, 200, 500), and tag mixing (1 quarter, 2 quarters).

The values for the diagnostic model are in bold and the levels used in the grid are directly comparable to those presented in subsubsection 7.5.2 through identical notation. The levels of the grid are:

1. Steepness: 0.65, **0.8**, 0.95
2. Tag mixing: 1 quarter, **2** quarters
3. Size data weighting: 10, **20**, 40
4. Conditional age-at-length data weighting: 0.5, **0.75**, 1

6.2.3 Integrated model and estimation uncertainty for key management quantities

For a full picture of uncertainty for the key management quantities ($SB_{\text{recent}}/SB_{F=0}$, $SB_{\text{recent}}/SB_{\text{MSY}}$ and $F_{\text{recent}}/F_{\text{MSY}}$) we attempted for the first time to integrate estimation uncertainty for individual grid models with the variability in best estimates of these quantities across the grid. The procedure that we adopted involved the following:

1. Obtain the best estimates of the key management quantities for each of the 54 grid models;
2. Obtain Hessian-based estimates of the standard deviations for these quantities using the variance-covariance matrix of the model parameters and the Delta Method:
3. Generate 1,000 random draws from normal distributions with mean and standard deviation specified as per steps 1 and 2, above, for each of the 54 grid models; in the case of $SB_{\text{recent}}/SB_{F=0}$, which was estimated on the log scale, transform the random deviates to normal space by taking their exponent;
4. Compute the mean, median, and 10th, 25th, 75th and 90th percentiles of the 54,000 values of each management quantity.

Note that the above procedure implicitly gives equal weight to each of the 54 grid models, which we felt was appropriate for this assessment. However, different relative weights could easily be given by varying the number of random draws of the management quantities from each grid model.

With respect to step 2, we were able to derive SDs even in cases where the Hessian matrix was not positive definite, but with a few very small negative eigenvalues, using a Hessian 'positivisation' process that has been coded in MFCL. By comparison of similar models from the grid that did and did not have zero negative eigenvalues, we were able to establish that the estimates of standard deviations of the key management quantities were completely unaffected by the Hessian not being positive definite, but with a few (maximum of 1 in the case of bigeye) very small negative eigenvalues. Therefore, we opted to include the estimates of estimation error from the few models that did not have positive definite Hessians.

7 Results

7.1 Consequences of key model developments

The progression of model development from the 2020 diagnostic model to the 2023 diagnostic model is described in subsection 6.1 and the results are displayed in Figure 18 and Figure 19. In previous assessments, the stepwise analysis has been presented simply by running MFCL at each step, plotting the results and attributing the change in results to changes in the model or the data. Through the process of building this stepwise development, it became apparent that model convergence is a significant issue for these complex models, and jittering is an important process to refine and improve the best solution found for any model.

Due to constraints in both time and computational resources, it was not feasible to conduct a jitter analysis during the full breadth of model exploration involved in the stepwise development. Such analysis was completed for the diagnostic model and for every model in the uncertainty grid, but not for the stepwise development. As a result, the successive estimations of two key management quantities, dynamic spawning potential depletion $SB/SB_{F=0}$ and spawning potential SB, in the stepwise development should only be considered indicative of potential changes.

1. **Diag2020**. The 2020 yellowfin diagnostic model, using data from 2020.
2. **NewExe**. Updating the executable resulted in no change in $SB/SB_{F=0}$ and a small decrease in SB.
3. **PreCatchCond**. The set of four changes applied before catch-conditioning resulted in a small increase in $SB/SB_{F=0}$ and SB.
4. **CatchCond**. The intermediate 'Old CPUE' model change resulted in a decrease in $SB/SB_{F=0}$ and SB, but the fully converted 'New CPUE' model change resulted in a substantial increase

in $SB/SB_{F=0}$ and SB, compared to the intermediate model. The overall effect of the combined 'CatchCond' step was a net increase in both quantities.

5. **SelChanges**. Removing constraints on longline selectivities had negligible effects on $SB/SB_{F=0}$ and SB.

6. **TagStructure**. Changes to the modelling of tagging data resulted in a decrease in $SB/SB_{F=0}$ and a small decrease in SB.

7. **Growth**. Fixing the L_1 growth parameter resulted in an increase in both quantities.

8. **DataWeights** Adjustments to data weighting in the model resulted in a small decrease in $SB/SB_{F=0}$ and a decrease in SB.

9. **NatMort**. Estimating Lorenzen natural mortality resulted in a substantial decrease in $SB/SB_{F=0}$, despite a substantial increase in SB, suggesting the biggest influence of this step was to increase the estimate of $SB_{F=0}$.

10. **TaggerEffect**. Updating the tagger effects model resulted in a decrease in both quantities.

11. **NewCPUEMethod**. Updating the CPUE spatio-temporal analysis resulted in a decrease in both quantities.

12. **NewData**. Adding the new data resulted in a decrease in both quantities.

13. **FilterSizeComps**. Changing the treatment of size composition data resulted in an increase in $SB/SB_{F=0}$ and no change in SB.

14. **FiveRegions**. Adopting the 5 region structure resulted in a decrease in $SB/SB_{F=0}$, despite a substantial increase in SB, suggesting the biggest influence of this step was to increase the estimate of $SB_{F=0}$.

15. **Diag2023**. The final refinements to improve the objective function value of the diagnostic model resulted in a decrease in $SB/SB_{F=0}$ and substantial decrease in SB.

Overall, the changes in $SB/SB_{F=0}$ are small until it decreases in steps 9 through 12, and again in step 15. The cumulative effect of all changes between the 2020 diagnostic model and the current diagnostic model is a reduction of $SB/SB_{F=0}$ from 55% to 43% for the year 2018, the last year where they can be compared. The current diagnostic model estimates the final $SB/SB_{F=0}$ in 2021 also at 43%.

The difference between the 2020 diagnostic model and the current diagnostic model is considerable in terms of $SB/SB_{F=0}$ but much smaller in terms of SB. In other words, it is not the numerator (SB) that makes $SB/SB_{F=0}$ different between the two diagnostic models but the denominator ($SB_{F=0}$). The increase in $SB_{F=0}$ occurs mainly in step 9, estimating Lorenzen natural mortality.

Subject to the caveat that none of the steps in the stepwise development involved jittering, the most influential steps in the development of the 2023 diagnostic model appear to be the estimation of natural mortality using the Lorenzen curve, applying the revised tagger effects method, and updating the CPUE spatio-temporal analysis, in terms of decreasing the final value of $SB/SB_{F=0}$.

7.2 Fit of the diagnostic model to data sources

7.2.1 Standardised CPUE from index fisheries

The observed CPUE indices show a general and steady decline from the 1950s until around 2000 and show a relatively stable trend after that. Comparing the range of recent indices to that of historical ones, the catch rates in region 1 have declined less than in other stock regions.

The values of the observed CPUE indices in each region act as regional scalers, since the catchability coefficient is the same across all regions. The observed CPUE index in the final year 2021, averaged over quarters, is higher in region 4 than in other regions, implying that this region contains the highest abundance of yellowfin tuna. Expressed as proportions, the implied abundance from the observed 2021 CPUE index suggests that 43% of the abundance is in the higher latitude regions 1 and 5, while 57% of the abundance is in the equatorial regions 2, 3, and 4.

The model fits to the index fishery standardised CPUE data were generally very good for all model regions (Figure 20). The model was able to predict the longer term trends and short-term variation at the sub-decadal scale. The residual plots (Figure 21) generally show agreeable residual distributions, although there tends to be a greater spread in the residuals earlier in the time series associated with the higher CPUE levels. There are sequences of negative residuals in regions 2 and 4 between 2009 and 2014, but the model fits the main CPUE trends after that.

7.2.2 Size composition data

Longline fisheries: The aggregate model fits to the weight composition data for the longline extraction fisheries (fisheries 1–29, Figure 23) were relatively good for all fisheries, and most importantly for the fisheries accounting for the largest catches, such as 4, 6, and 9 (Figure 7) in regions 2 and 4. One example of a fishery where the model does not fit the weight composition well is fishery 27, whose selectivity is linked with that of fishery 10, where the weight frequency distribution is bimodal. Both of those fisheries have a considerable number of weight samples but the annual catches are very small. Fishery 3 has a bimodal distribution that the model manages to fit quite well.

The index fisheries all have a common shared selectivity, where the model fit to fisheries 35 and 36 (regions 3 and 4) underestimates the proportion of small fish and overestimates the proportion of large fish. For the index fisheries in regions 1, 2, and 5, the model fits the weight compositions

more closely.

Other fisheries: The aggregated length composition fits are also mostly good for the non-longline fisheries, with the exception of fisheries 19 and 20 (Figure 22). These two fisheries in region 1 are the Japanese purse seine and pole and line fisheries, respectively. Both fisheries feature multi-modal distributions in the input data, and both fisheries catch relatively small amounts of yellowfin, so the inability to fit those compositions is not so important compared to other fisheries with larger catches. The non-longline fisheries with the largest annual catches are fisheries 13, 14, 17, 23, and 26 (Figure 8, Figure 9) consisting of miscellaneous gears in region 2 and purse seine in regions 3 and 4.

7.2.3 Tagging data

When aggregated, the tag attrition estimates fit the tagging data relatively well (Figure 26), albeit overestimating the number of tag returns after 2 periods at liberty, and underestimating the returns in periods 3–5. When compared at the tag program scale, there are some differences in the quality of fit (Figure 27). The fit to the PTTP generally reflects the fit to the aggregated scale, as this represents a large majority of the tags used in the assessment. The RTTP program is the second largest in terms of tag input data for yellowfin, and the data and fit also resemble the aggregated scale. The fit to the JPTP program is not as good, underestimating the tag returns after 2–6 periods at liberty and overestimating the tag returns for longer periods at liberty. At the regional scale, region 1 looks similar to the JPTP program and regions 2–4 look similar to the PTTP and RTTP programs (Figure 28). Region 5 has very few tag returns, with irregular patterns in the observed tag return data that the model does not fit well. Most of the tag returns occur in regions 3 and 4.

For the tag returns by year/quarter of recapture, the model-predicted tag returns show relatively good agreement with the observed data, albeit with some spikes missed and some over- and under-prediction at various periods (Figure 29). The number of tag returns is low for most longline fisheries, so the fits to these patchy data are of minor consequence to the likelihood (Figure 30). The fits to fisheries other than longline are good for those fisheries with the most observed tag returns, namely fisheries 25 and 26 (purse seine in region 3), fisheries 13, 14, 15, and 16 (purse seine in region 4), and fisheries 17, 18, 23, and 24 (miscellaneous gears in region 2). The model fits are reasonable for the other fisheries with small numbers of tag returns (Figure 31).

7.2.4 Conditional age-at-length

The available conditional age-at-length data consists of 1471 otoliths sampled between 1990 to 2018. The geographic distribution of samples (Figure 32) corresponds reasonably well to the range where yellowfin tuna are caught in great quantities. The model fit to the conditional age-at-length

data are shown in Figure 15 with 95% prediction intervals. This year's diagnostic model fits the data substantially better than the 2020 diagnostic model, with over 200 units of log-likelihood gain for this data component, when similar data weights are used.

7.3 Population dynamics estimates

7.3.1 Selectivity

A range of selectivity curves are estimated for the different fisheries in the model and can be largely classified by gear type. For longline fisheries, the age-specific selectivity is shown in Figure 33 and weight-specific selectivity is shown in Figure 34. For other fisheries, the age-specific selectivity is shown in Figure 35 and length-specific selectivity is shown in Figure 36.

The five index fisheries have a shared selectivity with a penalty constraining their shape to be non-decreasing, to avoid cryptic biomass in the model. Some of the longline extraction fisheries are also asymptotic or estimate the oldest age classes to be nearly fully selected, specifically fisheries 1, 2, 3, 10, 12, and 27, all in the higher latitude regions 1 and 5. All other longline fisheries have estimated selectivity with a peak at around 10 to 15 quarters, with selectivity then declining to some asymptote at an intermediate value between zero and full selectivity for the oldest fish. The selectivity by weight shows similar patterns, but expressed in weight rather than age, where the aforementioned higher latitude longline fisheries tend to reach full selectivity around 40 kg and the other longline fisheries typically achieving maximum selectivity around 30 kg.

The selectivity for the non-longline fisheries generally targets younger fish than longline. The main fisheries in region 2, such as fisheries 17, 23, and 24, have an estimated selectivity that peaks at age 2–3 quarters and then declines, reaching zero selectivity by age 10 quarters. The exception to this is fishery 18 that has a selectivity that is asymptotic and increases gradually with age, reaching half selectivity around age 18 quarters. The major purse seine fisheries 13, 14, and 26, in regions 3 and 4, have estimated selectivity curves that fully select ages 4–5 quarters and then decline gradually for older ages. Expressed in length, fisheries 17, 23, and 24 in region 2 reach full selectivity between 35 and 50 cm, while fisheries 13, 14, and 26 in regions 3 and 4 achieve maximum selectivity between 60 and 70 cm.

7.3.2 Movement

Observed patterns of tag releases and returns among regions are compared to the estimated movement coefficients between regions for each quarter from the diagnostic model in Figure 37 and in Figure 38.

Summing the observed tagging data over all quarters, the implied retention rate varies between regions, where over 90% of the tags released in regions 1, 2, 3 are recaptured within the same

region, 80% for region 4, but only 12% of tags released in region 5 are recaptured within that region. Most of the tags released in region 5 are recaptured in region 3, whereas almost no tags released in region 3 are recaptured in region 5. Thus, the data suggest a substantial northward movement from region 5 to 3. Another movement direction suggested by the data comes from the observation that 20% of the tags released in region 4 are recaptured in region 3.

The movement estimated in the model, averaging over quarters, shows somewhat different trends, with movement coefficients around 0.90 staying within regions 1, 2, 4, and 5 but a lower rate of around 0.60 staying in region 3. The estimated movement from region 3 is mainly into region 5 but also into region 4. The main movement within the model, southward from region 3 to region 5, is in the opposite direction from what the tagging data indicate. This reflects that the tagging data are not the only data source contributing to the estimation of movement in this integrated assessment. MFCL has considerable freedom to trade off recruitment and regional movement, so care should be taken in drawing conclusions on one of these estimates without considering the other.

7.3.3 Natural mortality

The Lorenzen form of natural mortality M is used in the 2023 assessment, with the scale of this curve estimated. As noted in the methods, this is considered good practice. This is in contrast to the 2020 diagnostic model which used a different form for M with the scale fixed. The comparison between the estimated M in 2023 diagnostic model and the fixed M is shown in Figure 17. Compared to the fixed M used in the 2020 diagnostic mode, the Lorenzen form of M features considerably higher levels of M for the youngest fish, and a lower asymptote for the older age classes.

The scale of M in the 2020 diagnostic model was determined by a geometric mean parameter which was set at 0.232 per quarter. The geometric mean of the Lorenzen M curve in the 2023 diagnostic model can be calculated as 0.153 per quarter, but these quantities are not directly comparable as the shape of the curve is quite different. The Lorenzen scale parameter that is estimated in the current diagnostic model determines M at the highest age in the model. This parameter is estimated by the diagnostic model as 0.119 per quarter, with 95% confidence limits at 0.113 and 0.126, based on the Hessian.

7.3.4 Maturity

Maturity-at-age is derived from the fixed maturity-at-length (fixed at the same values used in the 2020 diagnostic model) and applying the estimated growth curve to this to get maturity-at-age for the 2023 diagnostic model. This maturity-at-age curve differs slightly from the 2020 diagnostic model (Figure 16) due to the differences in growth curves in these two models.

7.3.5 Tag reporting rates

The estimated tag reporting rates by fishery recapture groups (see groupings in Table 1) are shown in Figure 40. As expected, the reporting rate estimates differ among fisheries groups and across tagging programs. In most cases, the reporting rate estimates for those groupings that received higher penalties were relatively close to the prior mean. Any fishery recapture groups for which there are no reported tag recaptures have tag reporting rates fixed at zero. In addition, groups with low numbers of tag recaptures (less than six) were also fixed at zero and those recaptures removed from the input file. This left 20 tag reporting rate groups where the reporting rates were estimated, of which 7 tag reporting group rates estimated on the upper bound at 0.99. The change in the assessment model this year, raising the bound from 0.90 to 0.99, did not prevent the tag reporting rate from running into bounds during the parameter estimation.

7.3.6 Growth

Growth was estimated in the 2023 diagnostic model using a von Bertalanffy growth form, where L_1 is the mean length at age 1 quarter, L_A is the mean length at oldest age in the model, K is a shape parameter, σ_1 is the length variability at age 1 quarter, and σ_A is the length variability at the oldest age. As described in subsection 6.1, the L_1 parameter is fixed, while all other growth parameters are estimated with the assessment model. The estimated growth curves from the 2020 diagnostic model and the 2023 diagnostic model are shown in Figure 15. Compared to the growth curve from the 2020 assessment, the current growth curve predicts slightly larger body size for the younger ages and smaller body size for ages above 10 quarters.

7.4 Stock assessment results

7.4.1 Recruitment: diagnostic model

The estimated recruitment aggregated across all regions (Figure 41) shows interannual variation, especially in the earlier decades of the assessment period that does not have size data to inform recruitment estimation. The variation in recruitment estimates also fluctuates in the last few years, as these fish are still quite young and have in many cases not yet been observed in the fisheries data. The total recruitment for the final 6 quarters is set to the arithmetic mean recruitment, a model setting that affects only the last two points in Figure 41. The long-term trend is that the estimated recruitment is somewhat higher in the first decade and last five years, but these are also the periods where recruitment estimates are least reliable.

Overall, region 2 is estimated to have contributed around 40% of the recruitment to the stock, while regions 1, 4, and 5 are close to 20% each (Figure 43). Region 3 is estimated to have practically no recruitment, but regions with essentially zero recruitment have been an unwanted and recurring feature in previous yellowfin assessments (Davies et al., 2014; Tremblay-Boyer et al., 2017; Vincent

et al., 2020) with the geographic location of missing recruitment varying between the assessments (Hamer, 2023). The substantial recruitment estimated in region 2 feeds into regions 1, 3, and 5, where region 2 is the largest source of unfished total biomass (Figure 44). Region 4, on the other hand, has more equally distributed source regions.

The temporal recruitment trends within individual regions (Figure 42) vary more than the sum of all regions (Figure 45). The estimated recruitment in region 2 has increased steadily since 1990 and currently contributes around 50% of recruits to the stock. Region 1 has relatively high estimated recruitment between 1980 and 2000, while recruitment in region 4 is estimated very low from 1990 to 2010. Region 5 has a more stable long-term trend in recruitment, and region 3 has effectively zero estimated recruitment.

The estimated relationship between spawning potential and recruitment is shown in Figure 46 with the assumed steepness of 0.8. The estimated recruitment variability is considerably higher in the earlier decades of the stock assessment, when size data are limited. The model excludes the early periods prior to 1968 from the estimation of the stock-recruitment curve.

7.4.2 Biomass: diagnostic model

The estimated total biomass and spawning potential declined steadily from 1960 to 2000, followed by a relatively stable population size since then (Figure 45). In 1960, the total biomass and spawning potential, averaged across quarters, are estimated to have been 10.6 and 7.5 million tonnes, respectively, and by 2000 they had declined to 4.3 and 2.4 million tonnes. In 2021, the final year of the assessment, the total biomass and spawning potential are estimated at 4.9 and 2.6 million tonnes.

The long-term trends vary between regions (Figure 47), with most regions having a current spawning potential that is close to one third of the 1960 levels. The exception is region 1, which is around two thirds of the the 1960 levels. These trends match the observed long-term CPUE trends in each region (Figure 20).

The current spawning potential of 2.6 million tonnes is partioned between region 1 = 460 thousand tonnes, region 2 = 370 thousand tonnes, region 3 = 210 thousand tonnes, region 4 = 980 thousand tonnes, and region 5 = 600 thousand tonnes. The combined higher latitude regions 1 and 5 have a current spawning potential of 1,070 thousand tonnes, while the combined equatorial regions 2, 3, and 4 have 1,550 thousand tonnes. Thus, the proportion of the spawning potential in the higher latitude regions is estimated at 41%, with 59% in the equatorial regions. These proportions are similar to the proportions of observed CPUE indices in subsubsection 7.2.1.

Analyses conducted at the 2022 peer review of the 2020 yellowfin assessment showed that the biomass partitioning between regions is effectively determined by the regional scaling of the observed

CPUE indices (Figure 6 in Punt et al., 2023).

7.4.3 Depletion: diagnostic model

The estimated spawning potential depletion $SB/SB_{F=0}$ aggregated across all regions shows an initial gradual decline until 1970, followed by a faster decline to about 2005, and is relatively stable after that (Figure 48). The estimated $SB/SB_{F=0}$ for all regions combined in the final year of 2021 is 0.43. This pattern varies regionally, as the current level of $SB/SB_{F=0}$ is around 0.50 in regions 1, 3, 4, and 5 but a substantially lower level of 0.25 in region 2.

The steepest decline in $SB/SB_{F=0}$ occurred from 1990 to 2000, especially in regions 2 and 4, corresponding to higher levels of annual catches in that decade compared to previous decades (Figure 4). The current level of $SB/SB_{F=0}$ is similar to the year 2010 in regions 1, 3, 5, while region 4 shows a gradual increase in this recent period. Region 2 shows a substantial recent decline from 0.39 in 2010 down to 0.25 in 2021.

7.4.4 Fished (SB) versus unfished ($SB_{F=0}$) spawning potential: diagnostic model

To interpret the trends in spawning depletion it is useful to compare the trends in spawning potential, SB, with the predicted spawning potential that would have occurred in the absence of fishing, $SB_{F=0}$, also called the unfished biomass (Figure 49). Unfished biomass is the denominator in the depletion ratio.

The total unfished biomass follows the same decline as the spawning potential during the first two decades of the stock assessment period, indicating that the steep decline in the estimated spawning potential before 1970 can be primarily explained by the estimated recruitment trends rather than fishing. From 1970 onwards, the unfished biomass is estimated to have been relatively stable for the next decades, until a recent increase in the last few years near the end of the assessment period. This matches the estimated recent recruitment increase seen in subsubsection 7.4.1.

The individual regions have comparable long-term trends in the estimated unfished biomass, with a noticable hump around 1990 that can be traced to an estimated recruitment pulse a few years earlier. Region 4 has the highest estimated unfished biomass among the regions, while region 1 stands out for having a current unfished biomass level near its historical maximum. Regional differences in the behaviour of the unfished trajectory are an important component in subsequent differences in $SB/SB_{F=0}$.

7.4.5 Fishing mortality: diagnostic model

The temporal trend in the adult fishing mortality has been a gradual increase until around 2010 and a slight decline since then (Figure 50), averaging 0.13 in the last ten years. Juveniles, as defined

by the maturity ogive (Figure 16), have fishing mortality rates that are generally around two times higher than that of adults, with annual fluctuations. The large difference between juvenile and adult fishing mortality is a change from the previous yellowfin assessment, where juvenile and adult fishing mortality were both around 0.15 for recent years. During the stepwise development of the current diagnostic model, this change occurs when natural mortality is estimated, leading to a different level and shape of the natural mortality curve.

The juvenile fishing mortality is estimated to have increased rapidly in the last few years, from 0.22 in 2015 to 0.46 in 2021. This increase matches the rapidly increasing catches in fishery 23 (Figure 9), consisting of miscellaneous gears in Indonesian waters targeting juvenile fish at age 2 and 3 quarters (Figure 35). The annual catches in fishery 23 have increased from 58 thousand tonnes in 2015 to 169 thousand tonnes in 2021, currently more than double that of any other fishery (Figure 9).

Regional comparison shows that fishing mortality rates are highest for the youngest ages in region 2 but higher for the older ages in the temperate regions 1 and 5, while regions 3 and 4 have fishing mortality rates that apply both to the younger and older fish in the population (Figure 51). Decadal comparison shows a recent increase in juvenile fishing mortality rates and that the age distribution in the population has remained relatively stable over time (Figure 52).

7.5 Multi-model inference: sensitivity analyses and structural uncertainty

7.5.1 One-off sensitivity analyses

Comparisons of the spawning depletion and spawning potential trajectories for the diagnostic model and the related one-off sensitivity models are provided in Figures 53, 54, 55, and 56.

These comparisons show that estimates of both spawning depletion and spawning potential were somewhat sensitive to the choice of tag mixing period, while spawning depletion was also somewhat sensitive to the assumed steepness value, and spawning potential to the assumed size data weighting.

Under the alternative assumed mixing periods, the depletion trajectories followed similar trajectories, diverging from the late 1970s. Results from the 1 quarter mixing scenario indicating a slightly less depleted state than the assumption of a 2 quarter mixing scenario. In terms of spawning potential, that from the 1 quarter mixing scenario was scaled higher across the time series.

With regards steepness, as expected a lower steepness assumption implied a more depleted stock. There was no impact on the estimated spawning potential.

Assumed size data weighting had small impacts on the estimated spawning depletion trajectory, with the divisor of 10 implying a more depleted stock at the end of the time series, and a slightly different trajectory. The assumed size data weighting also scaled the estimated spawning potential.

The age data weighting also had little impact on the estimated spawning depletion trajectory or the estimated spawning potential (Figure 56).

7.5.2 Structural uncertainty grid

Results of the structural uncertainty analysis are summarized in box and violin plots of $F_{\text{recent}}/F_{\text{MSY}}$ and $SB_{\text{recent}}/SB_{F=0}$ for the different levels of each of the four axes of uncertainty (Figure 57).

The distribution of recruitment across model regions and quarters for all models in the structural uncertainty grid is summarised in Figure 58. Time series of spawning depletion ($SB_{\text{recent}}/SB_{F=0}$) and spawning potential SB across grid models are shown in Figure 59 and Figure 60. Majuro and Kobe plots showing the estimates of $F_{\text{recent}}/F_{\text{MSY}}$, $SB_{\text{recent}}/SB_{F=0}$ and SB/SB_{MSY} across all models in the grid are presented in Figure 64. The averages and quantiles across the 54 models in the grid for all the reference points and other quantities of interest are presented in Table 5. For key management quantities ($SB_{\text{recent}}/SB_{F=0}$; $F_{\text{recent}}/F_{\text{MSY}}$; $SB_{\text{recent}}/SB_{\text{MSY}}$) the table alsos includes the additional estimate of estimation uncertainty for management advice.

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

- The grid contains 54 models that display a moderate range of estimates of stock status relative to reference points, and suggest that, overall, the stock is moderately more depleted than estimates from the 2020 assessment (Table 5).
- The results of the jittering process to improve the fit of all models in the grid is shown in Table 6. The jittering always achieved to improve the objective function value, but the gradient could often get worse after jittering and the Hessian for a given grid model could become non-positive definite.
- The most influential axis was steepness, which 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} (Figure 57). The assumed steepness level results in a clear clustering of stock status levels on the Kobe plot (Figure 64), particularly in relation to estimates of SB_{MSY}.
- Across the tag mixing period axis, results of models with the 2 quarter mixing period implied a slightly greater level of depletion and higher F/F_{MSY} than those with the 1 quarter mixing period.
- The estimates of depletion and fishing mortality for a size composition divisor of 10 were more pessimistic than those for 20 and 40, which were relatively consistent with one another.

With an assumed size composition divisor of 10 there was a slight skew in the distribution toward more optimistic stock status.

- The conditional age-at-length data weighting axis had limited impact on management quantities, with all levels showing similar ranges of depletion and F/F_{MSY}.
- Spawning depletion was generally low in the initial time period and started to increase in the early 1970s. Depletion stabilised in the mid-2000s through to the most recent period.
- Spawning depletion estimates in region 2 are approaching the limit reference point $20\%SB_{F=0}$ in the most recent model years for some of the grid models (Figure 59). While not to the same extent, depletion in the remainder of the tropical region (regions 3 and 4) have shown notable declines, but some recovery in the most recent period. Declining trends in the temperate regions are less severe and have also levelled out (Figure 59).
- Similar patterns are seen in spawning potential, with declines most notable in the tropical region, as well as in region 5 (Figure 60).
- Recruitment is predicted to be highest in region 2, with moderate recruitment in regions 1, 4 and 5, and lower recruitment in region 3. Across the scenarios, region 3 recruitment estimates are influenced by the tag mixing assumption, with zero recruitment in this region under all 2 quarter mixing scenarios but recruitments present where 1 quarter tag mixing is assumed. There is no consistent pattern in quarterly recruitment between regions (Figure 58).
- None of the models in the structural uncertainty grid had an overall spawning potential depletion below the LRP ($20\%SB_{F=0}$); median $SB_{\text{recent}}/SB_{F=0}$ was 0.47 (80 percentile range: 0.42 to 0.52) (Table 5).
- All models in the structural uncertainty grid showed exploitation to be below F_{MSY}. Median $F_{\text{recent}}/F_{\text{MSY}}$ was 0.50 (80 percentile range 0.42 to 0.61).

7.5.3 Integration of estimation and model uncertainty for key management quantities

Estimation uncertainty across the grid of 54 models was calculated for the key management quantities $SB_{\text{recent}}/SB_{F=0}$, $F_{\text{recent}}/F_{\text{MSY}}$ and $SB_{\text{recent}}/SB_{\text{MSY}}$ (Table 5). Distributions of the resulting quantities broken down by element for each of the four grid axes are presented in Figure 61 for $SB_{\text{recent}}/SB_{F=0}$, Figure 62 for $F_{\text{recent}}/F_{\text{MSY}}$ and Figure 63 for $SB_{\text{recent}}/SB_{\text{MSY}}$.

Presenting the estimates arising from these two approaches to incorporating uncertainty, results from the uncertainty grid of models and incorporating estimation uncertainty, allows the impact of the additional estimation uncertainty to be examined.

The median values for $SB_{\text{recent}}/SB_{F=0}$ from the grid and after incorporating estimation uncertainty

are identical. The tails of the distribution are also the same, after rounding to two decimal places.

For MSY-related quantities, incorporation of estimation uncertainty had slightly larger impacts than the changes seen when incorporating estimation uncertainty into $SB_{\text{recent}}/SB_{F=0}$. Median $F_{\text{recent}}/F_{\text{MSY}}$ is the same but median $SB_{\text{recent}}/SB_{\text{MSY}}$ is slightly lower when incorporating estimation uncertainty. d the 80 percentile ranges of $F_{\text{recent}}/F_{\text{MSY}}$ and $SB_{\text{recent}}/SB_{\text{MSY}}$ become somewhat wider, with some values below the 10th percentile of $SB_{\text{recent}}/SB_{\text{MSY}}$ falling below 1. These values are influenced by the levels of mixing and assumed steepness with mix 2 and lower steepness assumptions leading to higher estimates of $F_{\text{recent}}/F_{\text{MSY}}$ (Figure 62) and lower estimates of $SB_{\text{recent}}/SB_{\text{MSY}}$ (Figure 63).

It is recommended that management advice is based on the estimated management quantities including both the uncertainty grid and estimation uncertainty. The values of $SB_{\text{recent}}/SB_{F=0}$ are all above the LRP (20% $SB_{F=0}$ SBF=0), and the values of $F_{\text{recent}}/F_{\text{MSY}}$ are all below 1.

7.5.4 Analyses of stock status

There are several ancillary analyses related to stock status that are typically undertaken on the diagnostic model (e.g., dynamic Majuro and Kobe analyses, fisheries impact analyses etc.).

We do not present the results of all analyses for all models in the stock assessment paper. In this section, we rely largely on the dynamic spawning depletions and spawning potential plots for the models in the structural uncertainty grid (Figures 59 and 60). We also refer to the fished and unfished spawning potential trajectories for the diagnostic model discussed previously (Figure 49) and the dynamic Majuro and Kobe plots (Figure 65).

Dynamic Majuro and Kobe plots and comparisons with Limit Reference Points: The section summarizing the structural uncertainty grid (subsubsection 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 and Kobe plots, the methods of which are presented in subsubsection 5.7.4. The dynamic Majuro and Kobe plots for the diagnostic model are presented in Figure 65.

Both the dynamic Majuro and Kobe plots show the steady increase in depletion of the stock since the 1950s, with an increase in fishing mortality from the 1960s. The dynamic Majuro plot indicates that while depletion stabilised toward the end of the assessment time period, fishing mortality tended to increase. However, the terminal spawning potential is well above both SB_{MSY} and $20\%SB_{F=0}$, and the fishing mortality is well below F_{MSY} (Figure 65).

Fishing impact: In addition to the above analysis, it is possible to attribute the fishery impact with respect to depletion levels to specific fishery components (i.e., grouped by gear type), to estimate which types of fishing activity have the most impact on the spawning potential (Figure 66).

Fishing impacts were estimated to be very minor in all regions before about 1970, resulting primarily due to longline and pole and line fisheries. The impact of these gears has increased slightly over the time series. In the early 1970s, catch information from the miscellaneous fisheries leads to an increase in impact, with the onset of notable impacts due to purse seine fishing from the 1980s. Examining the overall impact, the miscellaneous and purse seine fisheries (associated and unassociated sets) have the major impact, with that of the miscellaneous fisheries increasing in the most recent period, and the impact of purse seine being equally split between associated and unassociated sets towards the end of the time series.

The greatest fishing impact is in region 2, where the miscellaneous fishery in this region has the largest proportional impact. The impact of the miscellaneous gears is also seen to a lesser extent in other regions, due to movement. Impact in the other tropical regions (regions 3 and 4) are primarily due to purse seine fishing, with a similar pattern seen in region 5.

Yield analysis: The yield analysis conducted in this assessment incorporates the spawner recruitment relationship (Figure 46) into the equilibrium biomass and yield computations. Importantly, in the diagnostic 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.

The yield distributions under different values of fishing effort relative to the current effort are shown in Figure 67 for select models representing different axes of the structural uncertainty grid (specifically, different levels of steepness). For the diagnostic model, it is estimated that MSY would be achieved by increasing fishing mortality by a factor of 1.65, although the resulting increase in yield would be relatively small (10%). The different example yield curves under the alternative steepness assumptions display a similar pattern over the scale of fishing mortality although the absolute value of the yield curve and behaviour of the descending limb differs significantly.

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 68. 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 juvenile fisheries in region 2 and, soon after, the rapid expansion of the PS fishery which shifted the age composition of the catch toward younger fish.

8 Discussion and conclusions

8.1 Stock status

The 2023 WCPO yellowfin tuna stock assessment estimated that the median recent spawning depletion ($SB_{\text{recent}}/SB_{F=0}$) at the stock-wide scale for all models in the grid are well above the limit reference point (Figure 59) and F/F_{MSY} is less than one.

The reference points calculated from the uncertainty grid results, incorporating estimation uncertainty, suggest that the median $SB_{\text{recent}}/SB_{F=0}$ is 0.47 (Table 5, Figure 59). F/F_{MSY} is less than one, with a median value of 0.50.

Estimated depletion and the spawning potential across the whole model region showed a long term decline to around the mid-2000s and remained stable after that, while the spawning potential was estimated to have increased slightly in the most recent period. The most notable declines are in the tropical region, with slightly lesser declines in the temperate regions.

Overall, the outcomes of this assessment suggest that the yellowfin stock in the WCPO is not overfished or undergoing overfishing.

CMM 2021-01 contains an objective to maintain the spawning biomass depletion ratio above the average $SB/SB_{F=0}$ for 2012-2015 (which is a value of 44% calculated across the unweighted grid). Based upon the estimates of $SB_{\text{recent}}/SB_{F=0}$ ($47\%SB_{F=0}$) this objective has currently been met.

8.2 Changes to the previous assessment

The addition of three more years of data (tagging, catch, effort, size compositions) and several other model changes were introduced to the 2023 assessment. These included:

- Conversion from a catch-errors to a catch-conditioned approach, and the inclusion of a likelihood component for the CPUE from the index fisheries.
- Change from using VAST to sdmTMB to standardise the input CPUE series and the inclusion of additional covariates in the CPUE model.
- Different CPUE variances used for the CPUE associated with each index fishery, applying a new approach to estimate these variances.
- Internal estimation of natural mortality and application of the Lorenzen form of natural mortality at age.
- Additional procedures implemented for achieving more reliable model convergence, including jittering and checking positive definite Hessian status for all grid models.

- Integration of parameter estimation uncertainty with model-based uncertainty across the model grid for the key management reference points.
- Additional size composition filtering.
- Modifications to selectivity estimation settings, changes to fisheries with non-decreasing selectivity.
- Adoption of revised tagger effect modelling framework, reverting to assumptions similar to those used in 2017.
- Changes to size data weighting and downweighting the conditional age-at-length data for internal growth estimation.

8.3 Model diagnostics and commentary

8.3.1 Mode of operation in early stepwise development

During the stepwise model development, diverse modelling options were explored and evaluated. This involved checking the model fit to data, both visually and by tabulating the likelihood components and maximum gradient to compare alternative models. When models achieved a low gradient, Hessian calculations were sometimes run to see if it was positive definite, which would take another day of computing time. By investing a little bit of human time, managing scripts and files, the Hessian can be run in parallel on multiple cores to get the results on the same day. The workflow to conduct Hessian calculations on dozens of models simultaneously has been streamlined by the newly developed R package called hessian, tailored for MFCL and SPC-specific servers.

More often than not, though, the decision to adopt a certain stepwise modelling option was not based on gradients or Hessians, or the total objective function, as changes in the model or data treatment would cause the likelihood components to be not comparable. The reasoning behind the adoption of each stepwise change is outlined in subsection 6.1.

It is worth noting that during the entire stepwise pathway from the 2020 diagnostic model to the 2023 diagnostic model, the yellowfin model showed consistently worse gradients than the bigeye model, even though they are very similar models with only a few differences. The main difference lies in the data, where the yellowfin data seem to pose a more irregular likelihood surface for the model. This was confirmed in the likelihood profile analysis, where the same computational approach for profiling resulted in more uneven likelihood surfaces (Figure 12.1) for the yellowfin assessment than for bigeye.

8.3.2 Mode of operation in late stepwise development

At the stepwise change when the estimation of M was introduced, it became apparent that this change was resulting in noticeably worse gradients than from the earlier model with a fixed M. The uncertainty about M can either be addressed by estimating M within the model or as a structural uncertainty with a number of fixed M options. After explorations and evaluations, it was decided that the best approach was to estimate M within the model.

At this point, it became clear that jittering was necessary to assist the model to find the best fit. A set of 10, 20, or 50 jitters were used and the finding was that the very best likelihood is found in a rough and uneven area of the likelihood surface, where gradients are high and the Hessians not positive definite. A nearby model could often be found at a local minimum, with quite similar parameter values, a good gradient and a positive definite Hessian, but with a worse objective function value and unfortunately often with substantially different estimates of management quantities.

One important lesson from this is that at least for the yellowfin assessment, a positive definite Hessian is a highly unreliable indicator of having found the best model fit to the data for a specific model configuration. Jittering is the tool to find the best fit, and in the case of yellowfin, the best fit is unlikely to have a positive definite Hessian.

8.3.3 Refining the diagnostic model

In the best case, a model run that starts from a standard MFCL `.ini` file results in a fit whose objective function value is not quite as good as a jittered version of the same model, but the parent model before jittering has similar estimated parameter values and management quantities. The 2023 diagnostic model is such a model.

The 2023 diagnostic model has both an ancestor model and offspring models. All three 'generations' share the same model parametrization and data handling. The ancestor model, named `14_Five_Regions` in the stepwise development, gave a very different estimate of depletion than the main `15_Diag_2023` diagnostic model (Figure 19). The refinement from step 14 to 15 involved an extensive exploration of the likelihood surface using jittering, revising initial parameter values, and changing the estimation phases.

The main diagnostic model runs from a standard `.ini` file, converges to a good objective function value and has a positive definite Hessian. However, when the diagnostic model is jittered from its final `.par` file, slightly better fits can be found with a lower objective function value that have similar estimates of management quantities. Such jitter offspring models can only be run from a final `.par` file and not from a standard `.ini` file, where users generally specify initial values and other model settings. Furthermore, these jitter offspring models have higher gradients and do not have a positive definite Hessian. Overall, it does not seem that a positive definite Hessian should

be seen as a strong requirement or guarantee of a well converged model.

8.3.4 Model reliability and challenges

The retrospective analysis (subsection 12.2) shows a pattern of a slight but consistent underestimation of biomass and depletion, which is 'corrected' each time a year of data is added. These results are similar to the retrospective analysis presented in the 2020 assessment.

During the stepwise model development, as well as jittering, two alternative models could often have a comparable objective function value, but one had a better likelihood for the length compositions and the other model had a better likelihood for the weight compositions. The likelihood profile in Figure 12.1 shows the conflict between these data components very clearly.

Overall, it's fair to say that the yellowfin assessment is subject to model convergence problems that pose both statistical and operational challenges for the assessment. Simplifying the regional structure has not solved the model convergence issues in terms of large final gradients and the need for jittering, as well as estimating zero recruitment in region 3. One important benefit of the simplified regional structure was to reduce the run time of a single model from around 17 hours to 12 hours, which allowed a substantially more rapid cycle of model development and testing.

8.4 Recommendations for further work

Changing the spatial structure from 9 regions to 5 regions involved some extra work that resulted in benefits such as reducing both the number of parameters and the computational time each run takes, as well as slight improvements in model convergence in terms of Hessians and gradients. To simplify the model further, it is likely that the number of fisheries could be reduced, merging similar fisheries that are now in the same region. Simplifying the fisheries could help with model convergence and parameter estimability, as selectivities are a challenging aspect for the model fitting, as spline node parameters frequently run into bounds. When deciding which fisheries could be merged, one would compare the size frequencies of those fisheries, tagging data, as well as fisheries management aspects. This would be a good topic to examine in the next assessment.

The large proportion of the stock biomass estimated in the higher latitude regions 1 and 5 are a concern, given that there are almost no catches observed in those regions. Specifically, 41% of the current biomass is estimated to be in regions 1 and 5, while 6% of the catches in tonnes are observed in those regions. A research effort was made in this year's assessment to examine whether there is evidence in the data that such a large proportion of the biomass might be in regions 1 and 5. Alternative approaches were explored for the CPUE data preparation, but did not lead to important findings regarding this issue. The 2022 yellowfin assessment review recommended implementing an equatorial-only model to examine this issue, which would exclude the higher latitude regions but

would still contain 94% of the fishery catches. This regional study should be conducted outside of the main assessment period.

This year's stepwise model development included a stage where alternative options to model natural mortality (M) were explored and evaluated. The options included: (1) same M-at-age as used in the 2020 assessment; (2) an M-at-age curve that has the same shape as used in the 2020 assessment but estimating the overall scale; (3) Lorenzen M curve estimating the overall scale; (4) a two-stage M-at-age curve with the original scale tabulated in Hoyle et al. (2023); and (5) the same two-stage M-at-age curve but estimating the overall scale. All options had some strengths and weaknesses with respect to likelihoods, residual patterns, model convergence, and other model selection criteria. A decision was made that the Lorenzen curve was the best option, noting that it would have been good to have more time to explore the M options. This was at a stepwise development stage where the assessment model had 9 regions and a few other model changes were also yet to be made. It would be a worthwhile research topic to revisit the question of how best to model M in the yellowfin assessment.

Another decision during the stepwise model development was to fix the L_1 parameter, as a workaround to prevent unreasonably low estimates of 9 cm that were occurring in the 9 region model at that stage in the model development. As a very last check in the assessment analysis, at the time of this writing, an unofficial model run was conducted where L_1 was estimated in the 5 region model and the result was that L_1 was estimated as 18 cm. It is possible that with the simplified model structure that L_1 is now estimable and does not need to be fixed. In general, it is better to estimate such parameters than to fix them, so this would be a good potential improvement to evaluate in the next assessment's stepwise development.

The biological data from the yellowfin stock assessment suffer from some geographic gaps in the data. Ideally, all data components should be sampled so that the observations reflect the entire stock. There are ongoing and planned research projects that relate to this objective, which will directly improve the reliability of the yellowfin stock assessment results, including the estimation of the relative stock status.

In the near future, there is a need to explore and develop alternative stock assessment software platforms to succeed MFCL that are capable of appropriately utilising the data relevant to WCPFC tuna stocks.

8.5 Main assessment conclusions

The general conclusions of this assessment are as follows:

- The spawning potential of the stock has become more depleted across all model regions until around 2010, after which it has become more stable, or shown a slight increase.

- Average fishing mortality rates for juvenile and adult age classes have increased throughout the period of the assessment, although more so for juveniles which have experienced considerably higher fishing mortality than adults. In the recent period a sharp increase in juvenile fishing mortality is estimated, while adult fishing mortality has stabilised.
- Overall, median depletion from the model grid for the recent period (2018–2021; $SB_{\text{recent}}/SB_{F=0}$) is estimated at 0.47 (80 percentile range including estimation and structural uncertainty 0.42–0.52, full range 0.36–0.59).
- No models from the uncertainty grid, including estimation uncertainty, estimate the stock to be below the LRP of $20\%SB_{F=0}$.
- CMM 2021-01 contains an objective to maintain the spawning biomass depletion ratio above the average of 2012-2015, $SB_{2012-2015}/SB_{F=0}$, which is a value of 0.44 calculated across the unweighted grid. Based upon the estimates of $SB_{\text{recent}}/SB_{F=0}$ of 0.47, this objective has currently been met.
- Recent (2017–2020) median fishing mortality ($F_{\text{recent}}/F_{\text{MSY}}$) was 0.50 (80 percentile range, including estimation and structural uncertainty 0.41–0.62, full range 0.26–0.78).
- Assessment results suggest that the yellowfin stock in the WCPO is not overfished, nor undergoing overfishing.

9 Acknowledgements

We thank the various fisheries agencies and regional fisheries observers for their support with data collection, provision and preparatory analysis. We thank participants at the preparatory stock assessment workshop for their contributions to the assessment. We especially thank the SPC data management team for their hard work and support to provide the data fuel for the assessment. We particularly thank the three independent peer review experts for the insightful comments that helped advance and improve this assessment immeasurably. We are grateful to Fabrice Bouyé for ensuring smooth operation of the computing resources required to do the assessment. Funding from Pacific-European Union (EU) Marine Partnership (PEUMP) supported the work on CPUE analysis. Finally, we highlight the lifetime of work that Dave Fournier put into MULTIFAN-CL, without which we could not have performed this assessment.

10 Tables

Table 1: Definition of fisheries for the 2023 yellowfin stock assessment in the WCPO.

Fishery	Fishery label	Flag	Region	Sel group	% Catch last 10 yrs	% Catch all yrs
F1	1.LL.ALL.1	ALL	1	1	0.9	1.7
F2	2.LL.ALL.1	ALL	1	2	0.5	0.7
F3	3.LL.US.1	US	1	3	0.2	0.1
F4	4.LL.ALL.4	ALL	4	4	1.0	3.5
F5	5.LL.OS.4	OS	4	8	1.4	0.8
F6	6.LL.OS.2	OS	2	9	7.5	8.5
F7	7.LL.ALL.2	ALL	2	10	0.1	1.3
F8	8.LL.ALL.3	ALL	3	11	0.7	1.0
F9	9.LL.ALL.4	AU	4	5	2.6	4.7
F10	10.LL.AU.5	ALL	5	12	0.2	0.2
F11	11.LL.ALL.5	ALL	5	7	1.3	1.8
F12	12.LL.ALL.5	ALL	5	6	1.7	1.3
F13	13.PS.ASS.4a	ALL	4	13	6.9	10.4
F14	14.PS.UNA.4	ALL	4	16	10.0	8.9
F15	15.PS.ASS.4b	ALL	4	14	6.8	4.6
F16	16.PS.UNA.4	ALL	4	17	7.4	4.2
F17	17.MISC.PH.2	PH	2	19	5.6	10.6
F18	18.HL.PHID.2	ID.PH	2	20	6.2	4.0
F19	19.PS.JP.1	JP	1	21	0.7	1.3
F20	20.PL.JP.1	JP	1	22	0.5	1.3
F21	21.PL.ALL.4	ALL	4	23	0.0	0.1
F22	22.PL.ALL.3	ALL	3	24	0.0	0.3
F23	23.MISC.ID.2	ID	2	25	15.9	8.9
F24	24.PS.PHID.2	ID.PH	2	28	1.9	2.6
F25	25.PS.ASS.3	ALL	3	15	4.1	5.9
F26	26.PS.UNA.3	ALL	3	18	11.3	6.5
F27	27.LL.AU.5	AU	5	12	0.0	0.0
F28	28.PL.ALL.2	ALL	2	26	3.3	3.0
F29	29.LL.ALL.5	ALL	5	7	0.0	0.0
F30	30.PS.ASS.2	ALL	2	13	0.0	0.7
F31	31.PS.UNA.2	ALL	2	16	0.3	0.4
F32	32.MISC.VN.2	VN	2	27	0.8	0.5
Index fisheries						
F33	LL-ALL-1	ALL	1	29		
F34	LL-ALL-2	ALL	2	29		
F35	LL-ALL-3	ALL	3	29		
F36	LL-ALL-4	ALL	4	29		
F37	LL-ALL-5	ALL	5	29		

Table 2: Definition of fisheries and associated tag recapture and reporting rate groupings (three columns at right) for the 2023 MULTIFAN-CL yellowfin tuna stock assessment in the WCPO. RTTP=Regional Tuna Tagging Program, PTTP=Pacific Tuna Tagging Program, JPTP – Japanese Tagging Program. An asterisk (*) indicates groups with 5 or fewer tag returns for which reporting rates were not estimated. The Recap group column indicates the fishery groupings for tag recapture data that were necessary due to tag returns by purse seine fisheries rarely including information on the set type (associated or unassociated) in which recaptures occurred.

Fishery	Fishery label	Flag	Model region	Recap group	RTTP	PTTP	JPTP
F1	1.LL.ALL.1	ALL	1	1	RTTP_L	RTTP_L	JPTP_L
F2	2.LL.ALL.1	ALL	1	2	RTTP_L	RTTP_L	JPTP_L
F3	3.LL.US.1	US	1	3	RTTP_L(US)*	RTTP_L(US)*	JPTP_L(US)*
F4	4.LL.ALL.4	ALL	4	4	RTTP_L	RTTP_L	JPTP_L
F5	5.LL.OS.4	OS	4	5	RTTP_L	RTTP_L	JPTP_L
F6	6.LL.OS.2	OS	2	6	RTTP_L	RTTP_L	JPTP_L
F7	7.LL.ALL.2	ALL	2	7	RTTP_L	RTTP_L	JPTP_L
F8	8.LL.ALL.3	ALL	3	8	RTTP_L	RTTP_L	JPTP_L
F9	9.LL.ALL.4	AU	4	9	RTTP_L	RTTP_L	JPTP_L
F10	10.LL.AU.5	ALL	5	10	RTTP_L(AU)	RTTP_L(AU)	JPTP_L(AU)*
F11	11.LL.ALL.5	ALL	5	11	RTTP_L	RTTP_L	JPTP_L
F12	12.LL.ALL.5	ALL	5	12	RTTP_L	RTTP_L	JPTP_L
F13	13.PS.ASS.4a	ALL	4	13	RTTP_S(a)-4	PTTP_S(a)-4	JPTP_S(a)-4*
F14	14.PS.UNA.4	ALL	4	13	RTTP_S(a)-4	PTTP_S(a)-4	JPTP_S(a)-4*
F15	15.PS.ASS.4b	ALL	4	14	RTTP_S(b)-4	PTTP_S(b)-4	JPTP_S(b)-4
F16	16.PS.UNA.4	ALL	4	14	RTTP_S(b)-4	PTTP_S(b)-4	JPTP_S(b)-4
F17	17.MISC.PH.2	PH	2	15	RTTP_S(PH,ID)-2	PTTP_S(PH,ID)-2	JPTP_S(PH,ID)-2
F18	18.HL.PHID.2	ID.PH	2	15	RTTP_S(PH,ID)-2	PTTP_S(PH,ID)-2	JPTP_S(PH,ID)-2
F19	19.PS.JP.1	JP	1	16	RTTP_S(JP)-1*	PTTP_S(JP)-1*	JPTP_S(JP)-1
F20	20.PL.JP.1	JP	1	17	RTTP_P(JP)-1*	PTTP_P(JP)-1*	JPTP_P(JP)-1
F21	21.PL.ALL.4	ALL	4	18	RTTP_P-3-4	PTTP_P-3-4	JPTP_P-3-4*
F22	22.PL.ALL.3	ALL	3	19	RTTP_P-3-4	PTTP_P-3-4	JPTP_S-3*
F23	23.MISC.ID.2	ID	2	15	RTTP_S(PH,ID)-2	PTTP_S(PH,ID)-2	JPTP_S(PH,ID)-2
F24	24.PS.PHID.2	ID.PH	2	15	RTTP_S(PH,ID)-2	PTTP_S(PH,ID)-2	JPTP_S(PH,ID)-2
F25	25.PS.ASS.3	ALL	3	20	RTTP_S-3	PTTP_S-3	JPTP_P-3*
F26	26.PS.UNA.3	ALL	3	20	RTTP_S-3	PTTP_S-3	JPTP_P-3*
F27	27.LL.AU.5	AU	5	21	RTTP_L(AU)	RTTP_L(AU)	JPTP_L(AU)*
F28	28.PL.ALL.2	ALL	2	22	RTTP_P-2	PTTP_P-2	JPTP_P-2*
F29	29.LL.ALL.5	ALL	5	23	RTTP_L	RTTP_L	JPTP_L
F30	30.PS.ASS.2	ALL	2	24	RTTP_S-2	PTTP_S-2*	JPTP_S-2*
F31	31.PS.UNA.2	ALL	2	24	RTTP_S-2	PTTP_S-2*	JPTP_S-2*
F32	32.MISC.VN.2	VN	2	25	RTTP_Z(VN)-2*	PTTP_Z(VN)-2*	JPTP_Z(VN)-2*

Table 3: Structural uncertainty grid for the 2023 WCPO yellowfin tuna stock assessment. Bold values indicate settings for the diagnostic model.

Axis	Levels	Option 1	Option 2	Option 3
Steepness	3	0.65	**0.8**	0.95
Tag mixing (# quarters)	2	1	**2**	
Size data weighting divisor	3	10	**20**	40
Age data weighting	3	0.5	**0.75**	1

Table 4: Description of symbols used in the yield and stock status analyses.

Symbol	Description
C_{latest}	Catch in the last year of the assessment (2021)
F_{recent}	Average fishing mortality-at-age for a recent period (2017–2020)
$Y_{F_{\text{recent}}}$	Equilibrium yield at average fishing mortality for a recent period (2017–2020)
f_{mult}	Fishing mortality multiplier at maximum sustainable yield (MSY)
F_{MSY}	Fishing mortality-at-age producing the maximum sustainable yield (MSY)
MSY	Equilibrium yield at F_{MSY}
$F_{\text{recent}}/F_{\text{MSY}}$	Average fishing mortality-at-age for a recent period (2017–2020) relative to F_{MSY}
SB_{latest}	Spawning biomass in the latest time period (2021)
SB_{recent}	Spawning biomass for a recent period (2018–2021)
$SB_{F=0}$	Average spawning biomass predicted in the absence of fishing for the period 2011–2020
SB_{MSY}	Spawning biomass that will produce the maximum sustainable yield (MSY)
$SB_{\text{MSY}}/SB_{F=0}$	Spawning biomass that produces maximum sustainable yield (MSY) relative to the average spawning biomass predicted to occur in the absence of fishing for the period 2011–2020
$SB_{\text{latest}}/SB_{F=0}$	Spawning biomass in the latest time period (2021) relative to the average spawning biomass predicted to occur in the absence of fishing for the period 2011–2020
$SB_{\text{latest}}/SB_{\text{MSY}}$	Spawning biomass in the latest time period (2021) relative to that which will produce the maximum sustainable yield (MSY)
$SB_{\text{recent}}/SB_{F=0}$	Spawning biomass for a recent period (2018–2021) relative to the average spawning biomass predicted to occur in the absence of fishing for the period 2011–2020
$SB_{\text{recent}}/SB_{\text{MSY}}$	Spawning biomass for a recent period (2018–2021) 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 $t-1$ (2011–2020)

Table 5: Summary of reference points over the 54 individual models in the structural uncertainty grid, along with results incorporating estimation uncertainty.

	Mean	Median	Min	10%ile	90%ile	Max	Diagnostic model
C_{latest}	751657	751856	750785	750860	752268	752337	751908
F_{MSY}	0.07	0.07	0.06	0.06	0.09	0.09	0.07
f_{mult}	1.96	2.00	1.47	1.64	2.38	2.50	1.89
$F_{\text{recent}}/F_{\text{MSY}}$	0.51	0.50	0.40	0.42	0.61	0.68	0.53
MSY	697874	700400	616800	644320	739560	771600	671600
SB_0	5761796	5729000	4455000	4817200	6640900	7279000	5216000
$SB_{F=0}$	5633743	5603267	4624645	4907798	6280841	6825888	5173954
SB_{latest}/SB_0	0.49	0.50	0.41	0.44	0.54	0.56	0.49
$SB_{\text{latest}}/SB_{F=0}$	0.50	0.50	0.41	0.45	0.55	0.58	0.49
$SB_{\text{latest}}/SB_{\text{MSY}}$	2.49	2.48	1.78	1.91	3.11	3.16	2.44
SB_{MSY}	1177733	1160500	740400	838260	1538200	1707000	1044000
SB_{MSY}/SB_0	0.20	0.20	0.17	0.17	0.23	0.24	0.20
$SB_{\text{MSY}}/SB_{F=0}$	0.21	0.21	0.16	0.17	0.24	0.25	0.20
$SB_{\text{recent}}/SB_{F=0}$	0.47	0.47	0.38	0.42	0.52	0.54	0.46
$SB_{\text{recent}}/SB_{\text{MSY}}$	2.31	2.30	1.68	1.77	2.89	2.94	2.27
$Y_{F_{\text{recent}}}$	157188	155300	141400	145150	172270	173300	152500
Including estimation uncertainty:							
	Mean	Median	Min	10%ile	90%ile	Max	
$SB_{\text{recent}}/SB_{F=0}$	0.47	0.47	0.36	0.42	0.52	0.59	
$F_{\text{recent}}/F_{\text{MSY}}$	0.51	0.50	0.26	0.41	0.62	0.78	
$SB_{\text{recent}}/SB_{\text{MSY}}$	2.31	2.28	0.93	1.73	2.95	3.59	

Table 6: Overview of grid models. Each grid model was jittered 20 times and the model run that had the best objective function value was selected. The columns show the final gradient, final dynamic depletion, number of negative eigenvalues, and whether the Hessian was positive definite.

	Grid model	Njitters	ObjFun	Gradient	Depletion	Neigen	PDH
1	m1_s10_a050_h65	20	-838059.98	0.00229	0.431	3	
2	m1_s10_a050_h80	20	-838060.01	0.00018	0.468	3	
3	m1_s10_a050_h95	20	-838060.02	0.00414	0.492	2	
4	m1_s10_a075_h65	20	-837227.54	0.01139	0.430	4	
5	m1_s10_a075_h80	20	-837227.56	0.01899	0.467	4	
6	m1_s10_a075_h95	20	-837227.58	0.00139	0.491	4	
7	m1_s10_a100_h65	20	-836390.85	0.00095	0.440	0	Yes
8	m1_s10_a100_h80	20	-836390.87	0.00016	0.476	0	Yes
9	m1_s10_a100_h95	20	-836390.89	0.00637	0.500	0	Yes
10	m1_s20_a050_h65	20	-747997.99	0.00398	0.455	3	
11	m1_s20_a050_h80	20	-747998.01	0.00051	0.495	3	
12	m1_s20_a050_h95	20	-747998.02	0.00034	0.521	3	
13	m1_s20_a075_h65	20	-747182.59	0.00722	0.462	4	
14	m1_s20_a075_h80	20	-747182.61	0.00077	0.501	4	
15	m1_s20_a075_h95	20	-747182.62	0.00242	0.528	5	
16	m1_s20_a100_h65	20	-746365.46	0.00184	0.464	3	
17	m1_s20_a100_h80	20	-746365.49	0.00291	0.504	2	
18	m1_s20_a100_h95	20	-746365.50	0.00804	0.530	2	
19	m1_s40_a050_h65	20	-650118.01	0.00015	0.435	0	Yes
20	m1_s40_a050_h80	20	-650118.04	0.00540	0.478	0	Yes
21	m1_s40_a050_h95	20	-650118.05	0.00010	0.507	0	Yes
22	m1_s40_a075_h65	20	-649312.61	0.00044	0.470	3	
23	m1_s40_a075_h80	20	-649312.63	0.00329	0.511	1	
24	m1_s40_a075_h95	20	-649312.65	0.00181	0.539	3	
25	m1_s40_a100_h65	20	-648510.68	0.00398	0.447	3	
26	m1_s40_a100_h80	20	-648510.69	0.00235	0.490	1	
27	m1_s40_a100_h95	20	-648510.71	0.00018	0.518	2	
28	m2_s10_a050_h65	20	-839609.44	0.00024	0.396	3	
29	m2_s10_a050_h80	20	-839609.46	0.00691	0.434	4	
30	m2_s10_a050_h95	20	-839609.49	0.00025	0.460	3	
31	m2_s10_a075_h65	20	-838777.79	0.00058	0.383	3	
32	m2_s10_a075_h80	20	-838777.83	0.00534	0.422	3	
33	m2_s10_a075_h95	20	-838777.85	0.00033	0.448	3	
34	m2_s10_a100_h65	20	-837950.60	0.00546	0.383	4	
35	m2_s10_a100_h80	20	-837950.64	0.00170	0.422	4	
36	m2_s10_a100_h95	20	-837950.66	0.00231	0.448	3	
37	m2_s20_a050_h65	20	-749472.93	0.00455	0.412	1	
38	m2_s20_a050_h80	20	-749472.97	0.00109	0.454	1	
39	m2_s20_a050_h95	20	-749472.98	0.00356	0.482	1	
40	m2_s20_a075_h65	20	-748647.13	0.00236	0.417	1	
41	m2_s20_a075_h80	20	-748647.17	0.00034	0.459	1	
42	m2_s20_a075_h95	20	-748647.18	0.00147	0.487	1	
43	m2_s20_a100_h65	20	-747819.87	0.00204	0.423	1	
44	m2_s20_a100_h80	20	-747819.90	0.01627	0.464	1	
45	m2_s20_a100_h95	20	-747819.92	0.00417	0.492	1	
46	m2_s40_a050_h65	20	-651492.37	0.00078	0.426	0	Yes
47	m2_s40_a050_h80	20	-651492.41	0.00089	0.471	0	Yes
48	m2_s40_a050_h95	20	-651492.42	0.00748	0.501	0	Yes
49	m2_s40_a075_h65	20	-650675.52	0.00025	0.440	3	
50	m2_s40_a075_h80	20	-650675.55	0.00049	0.484	3	
51	m2_s40_a075_h95	20	-650675.56	0.00081	0.513	2	
52	m2_s40_a100_h65	20	-649867.26	0.00464	0.443	3	
53	m2_s40_a100_h80	20	-649867.29	0.00061	0.487	3	
54	m2_s40_a100_h95	20	-649867.30	0.00279	0.516	3	

11 Figures

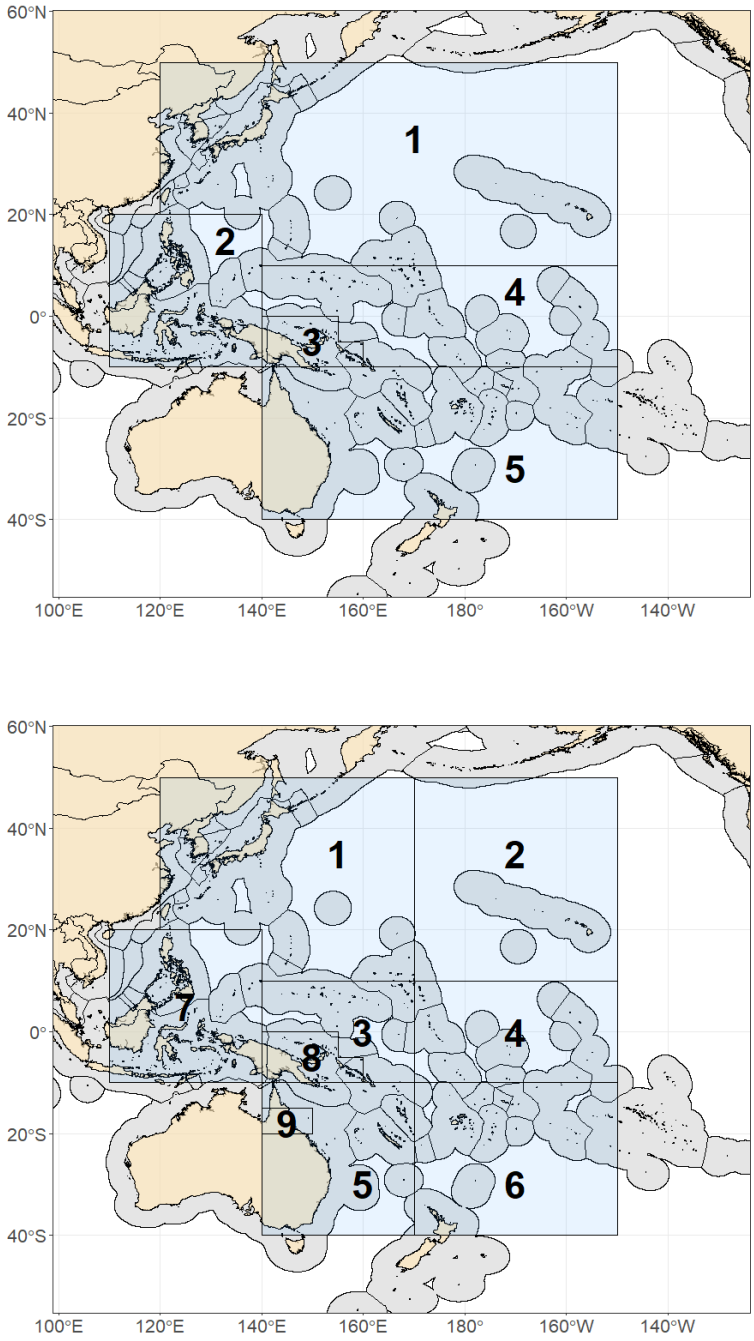

Figure 1: The geographical area covered by the stock assessment and the boundaries of the model regions for the 5 region structure that was used for 2023 WCPO yellowfin tuna assessment (top), and (bottom) the previous 9 region model structure that was used as the base structure for the stepwise model development.

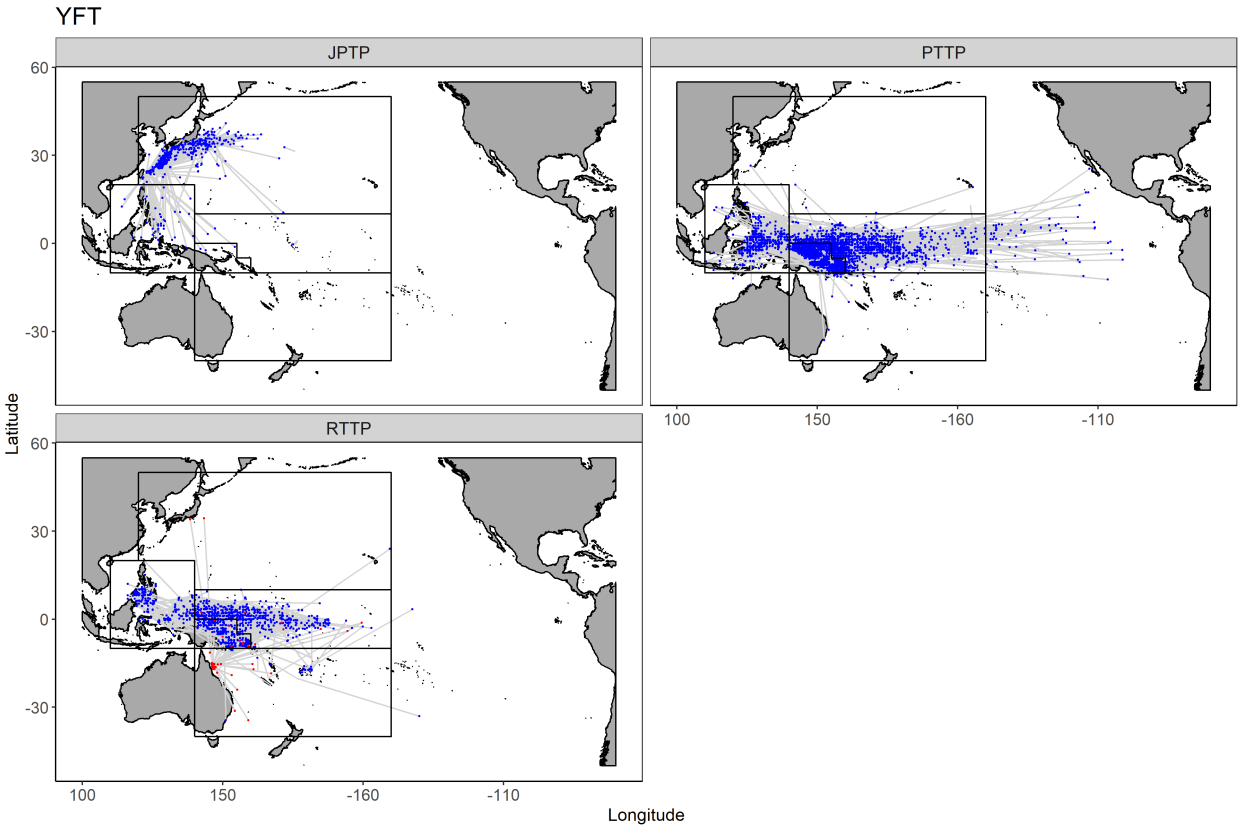

Figure 2: Map of tag recaptures. The panels show the distributions of release and recapture displacements for the different tagging programs: Pacific Tuna Tagging Program (PTTP), Regional Tuna Tagging Program (RTTP) and the Japanese Tagging Program (JPTP). Dots indicate recapture locations. Red dots in RTTP plot are the targeted Coral Sea tagging cruises.

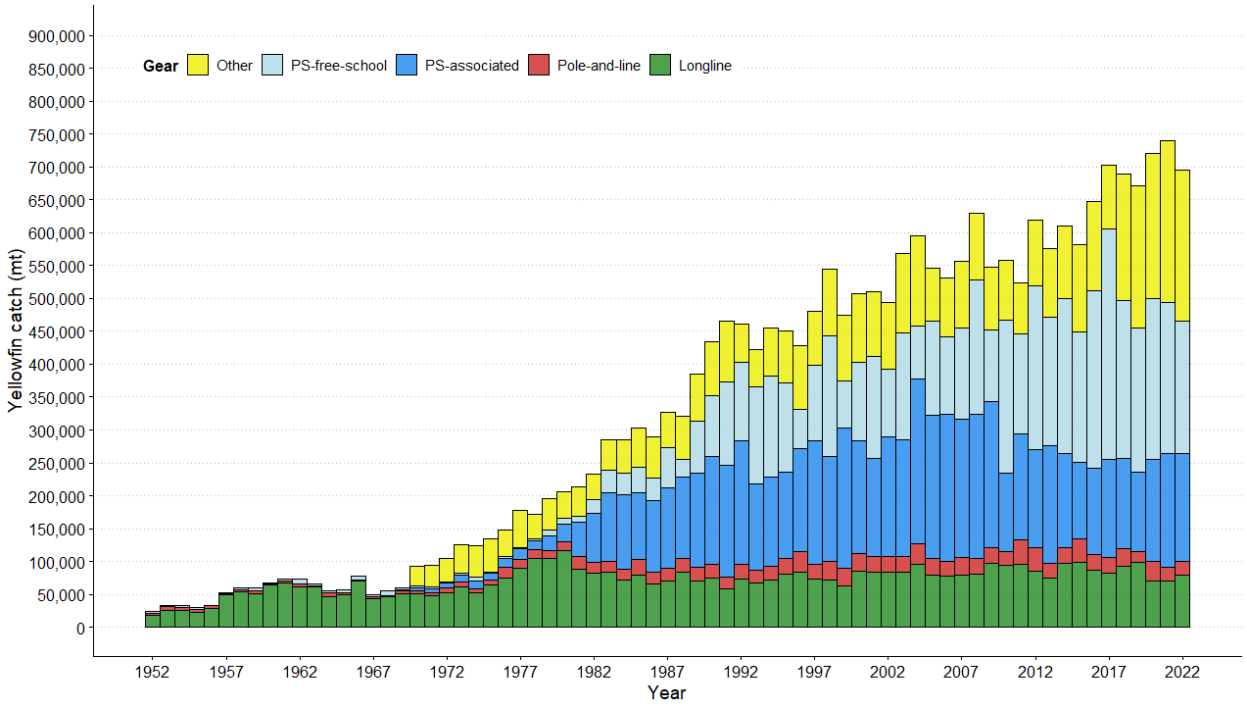

Figure 3: Annual catches of yellowfin by gear type in the WCPO area covered by the assessment.

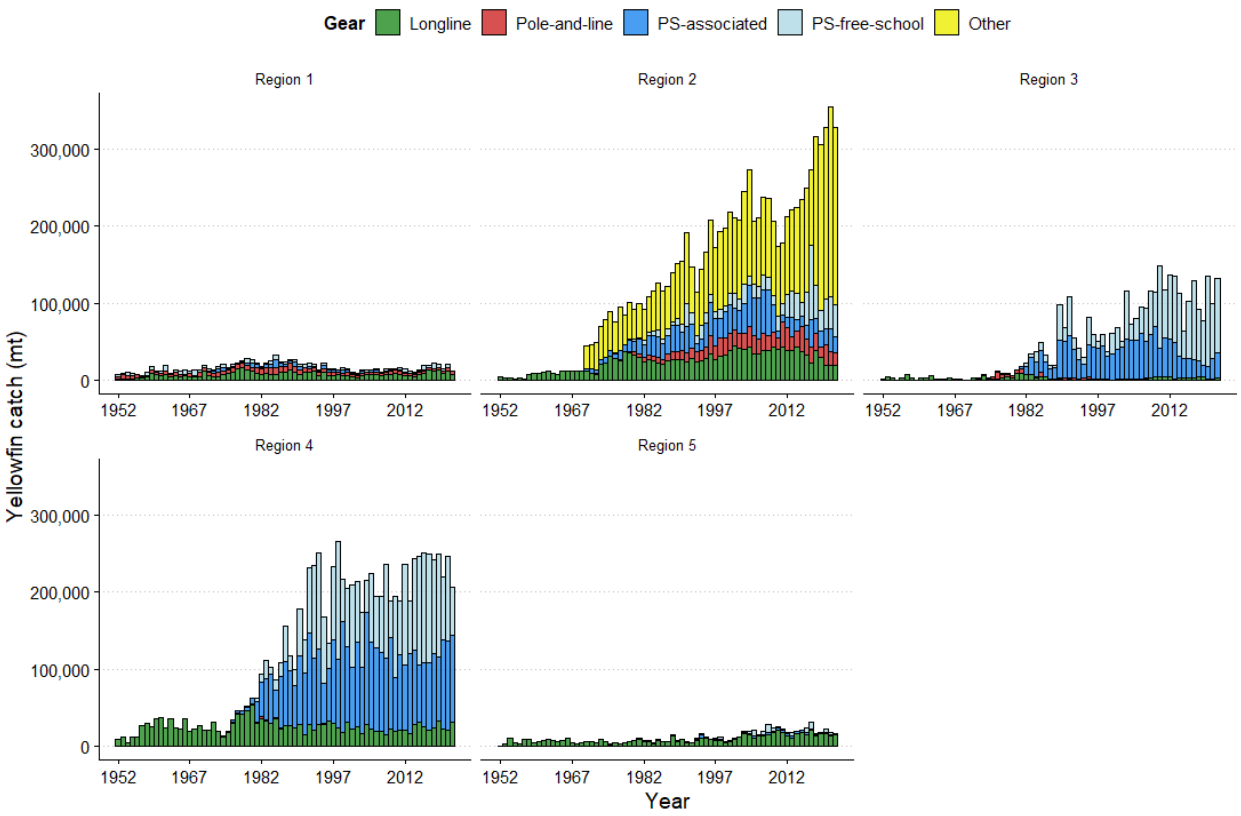

Figure 4: Annual catches of yellowfin by gear type for each of the five model regions.

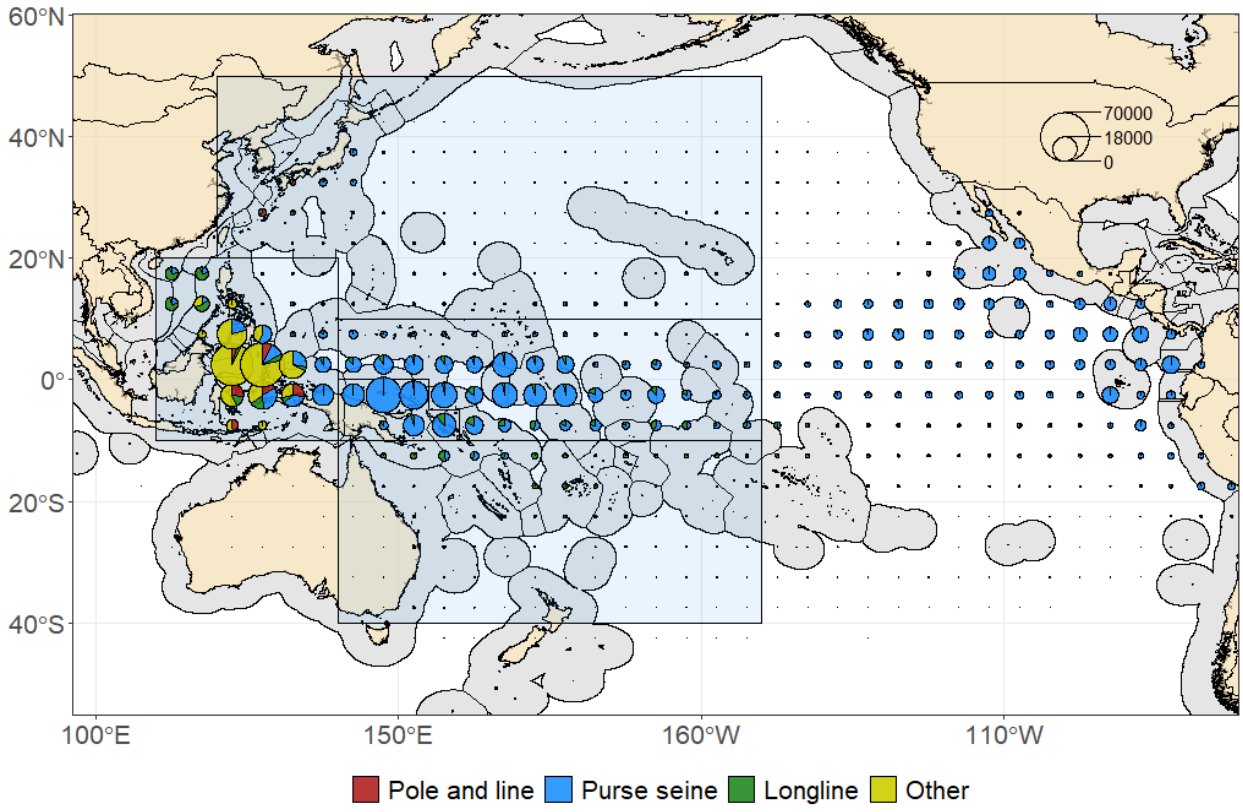

Figure 5: Distribution and magnitude of yellowfin catches (mt) by gear type summed over the last 10 years (2012–2021) for 5 × 5 degree cells.

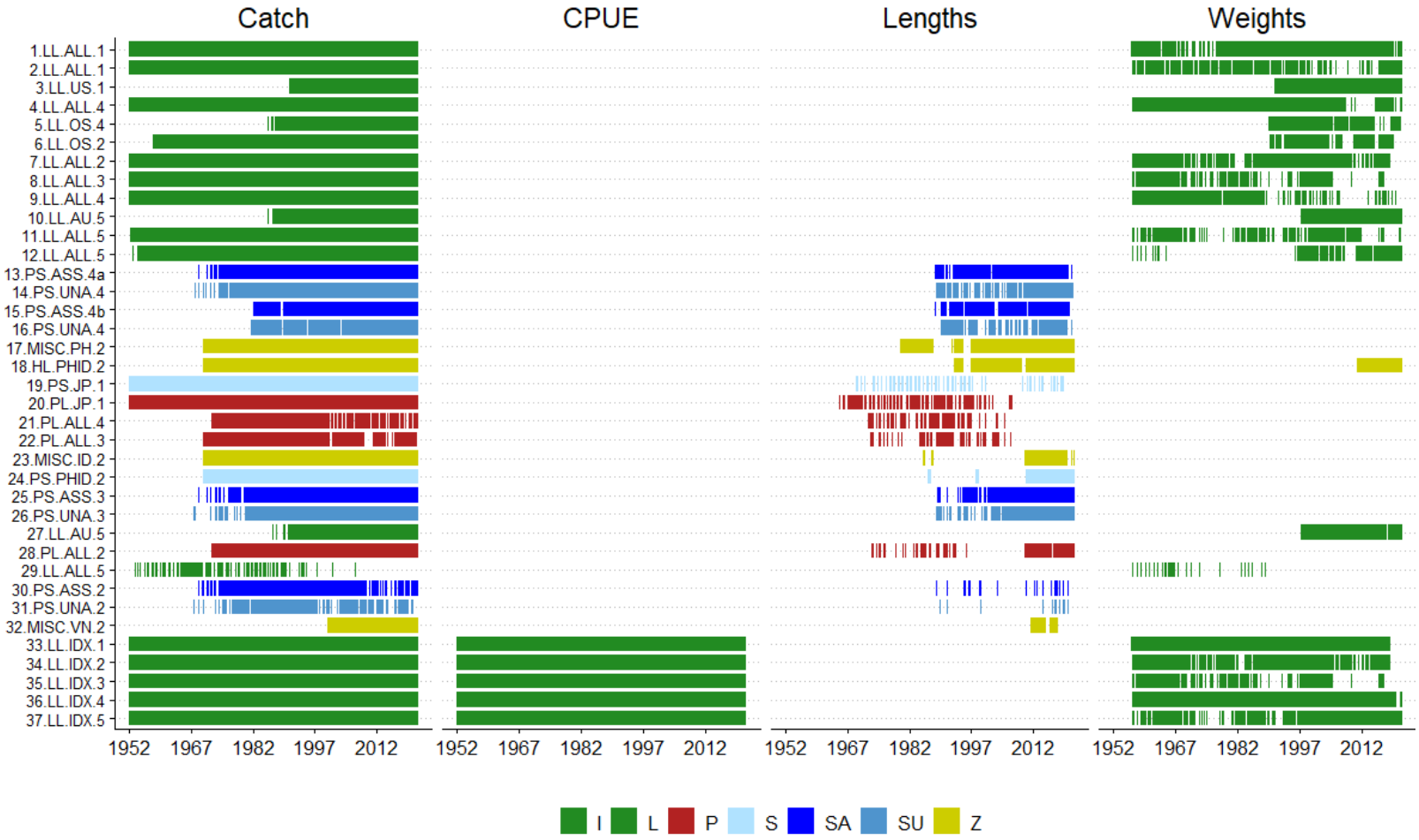

Figure 6: Summary of data coverage by fishery for the WCPO 2023 yellowfin assessment. I=index fisheries, L=longline, P=pole and line, S=purse seine (unspecified), SA=purse seine associated, SU=purse seine unassociated, Z=miscellaneous gears.

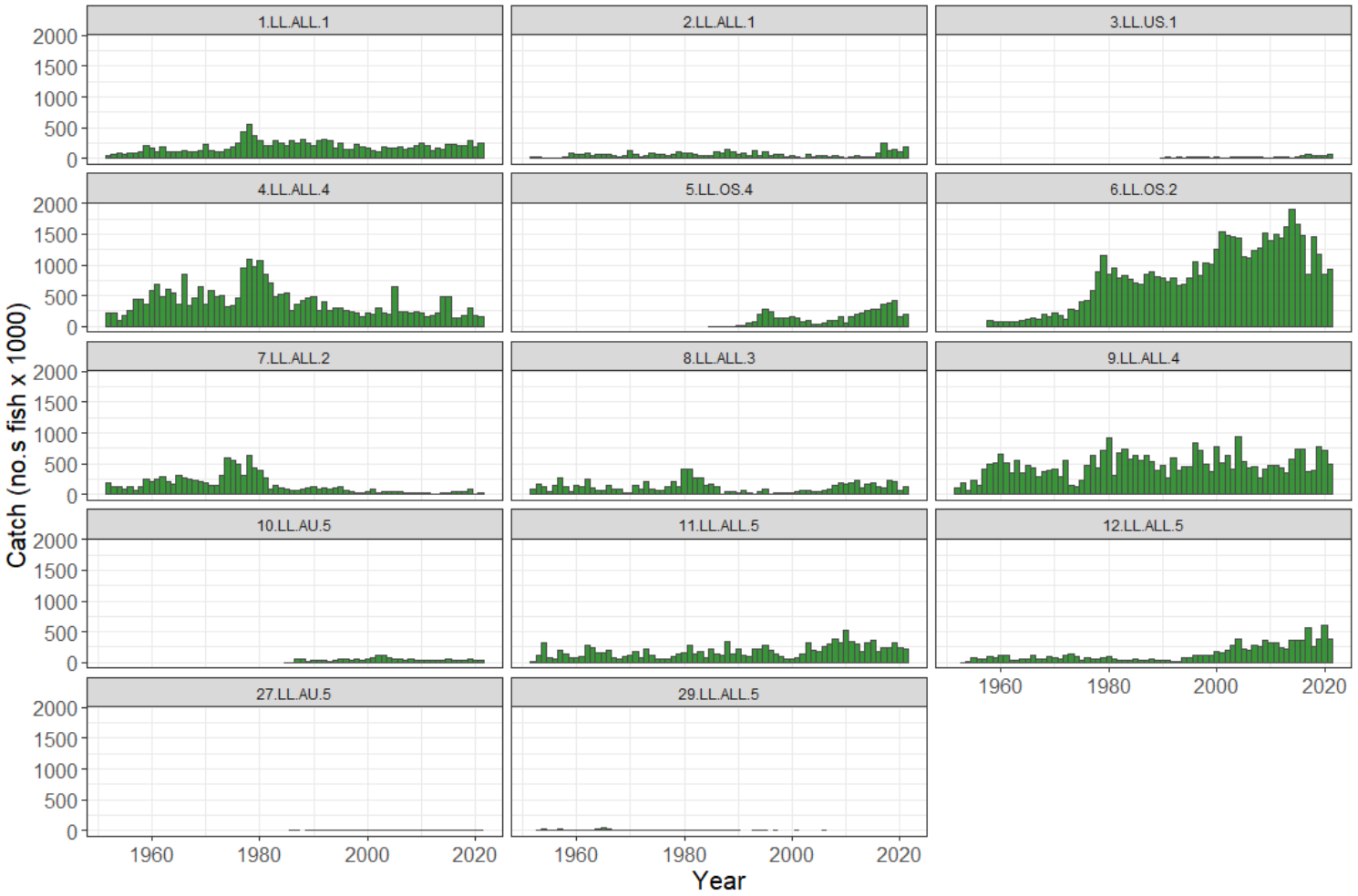

Figure 7: Time series of annual catches (numbers of fish) by fishery: longline.

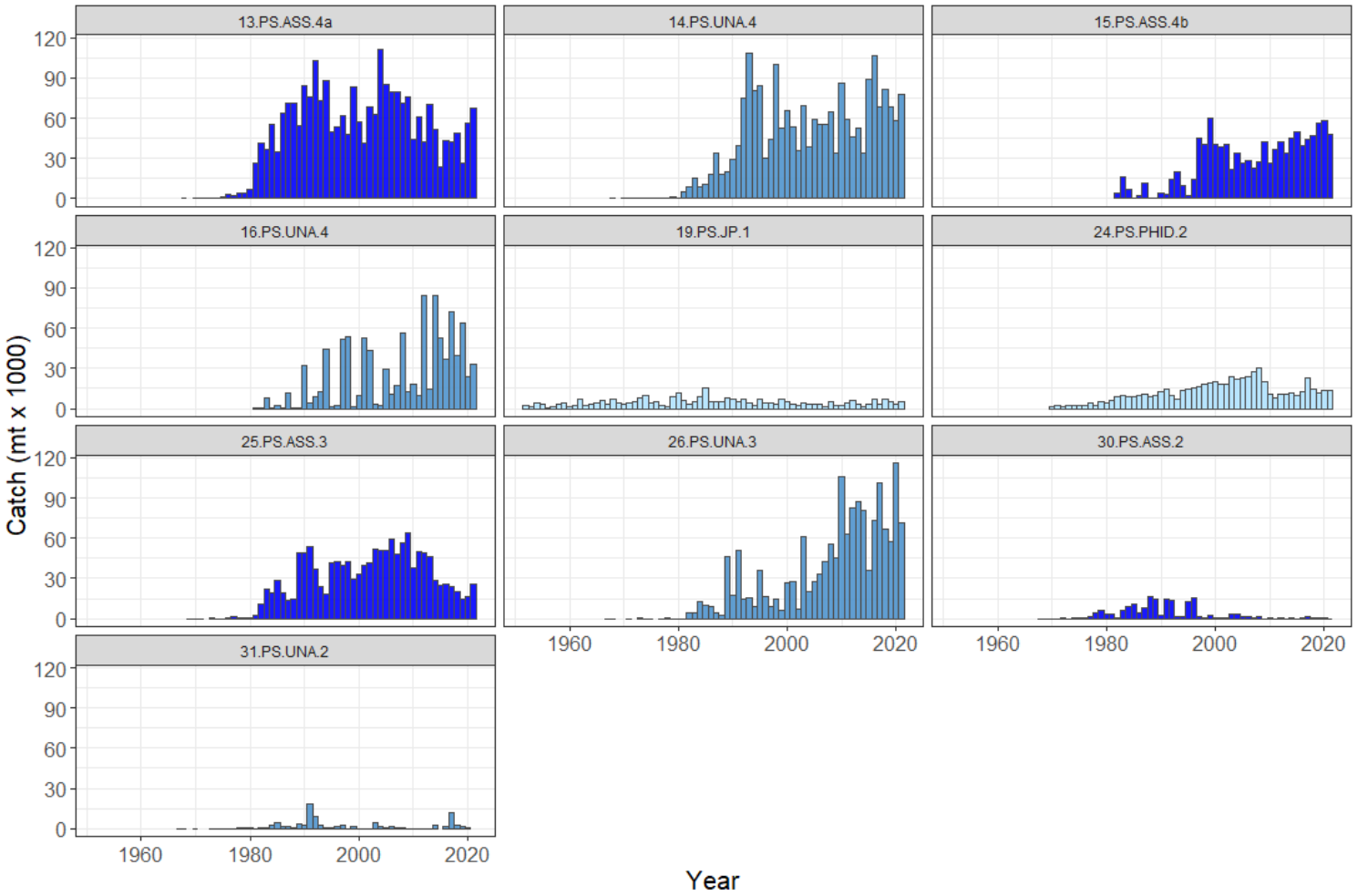

Figure 8: Time series of annual catches (mt) by fishery: purse seine.

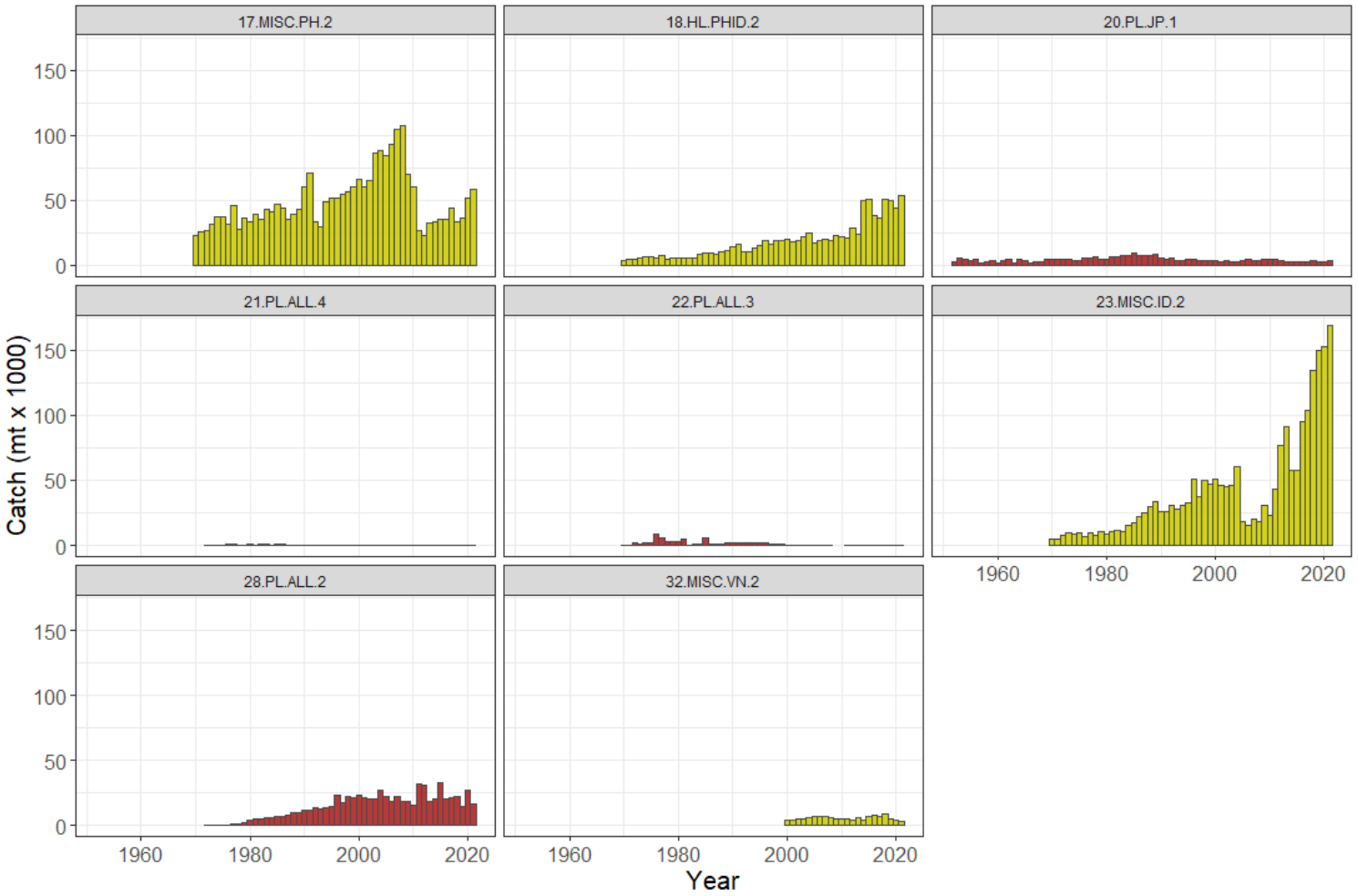

Figure 9: Time series of annual catches (mt) by fishery: other.

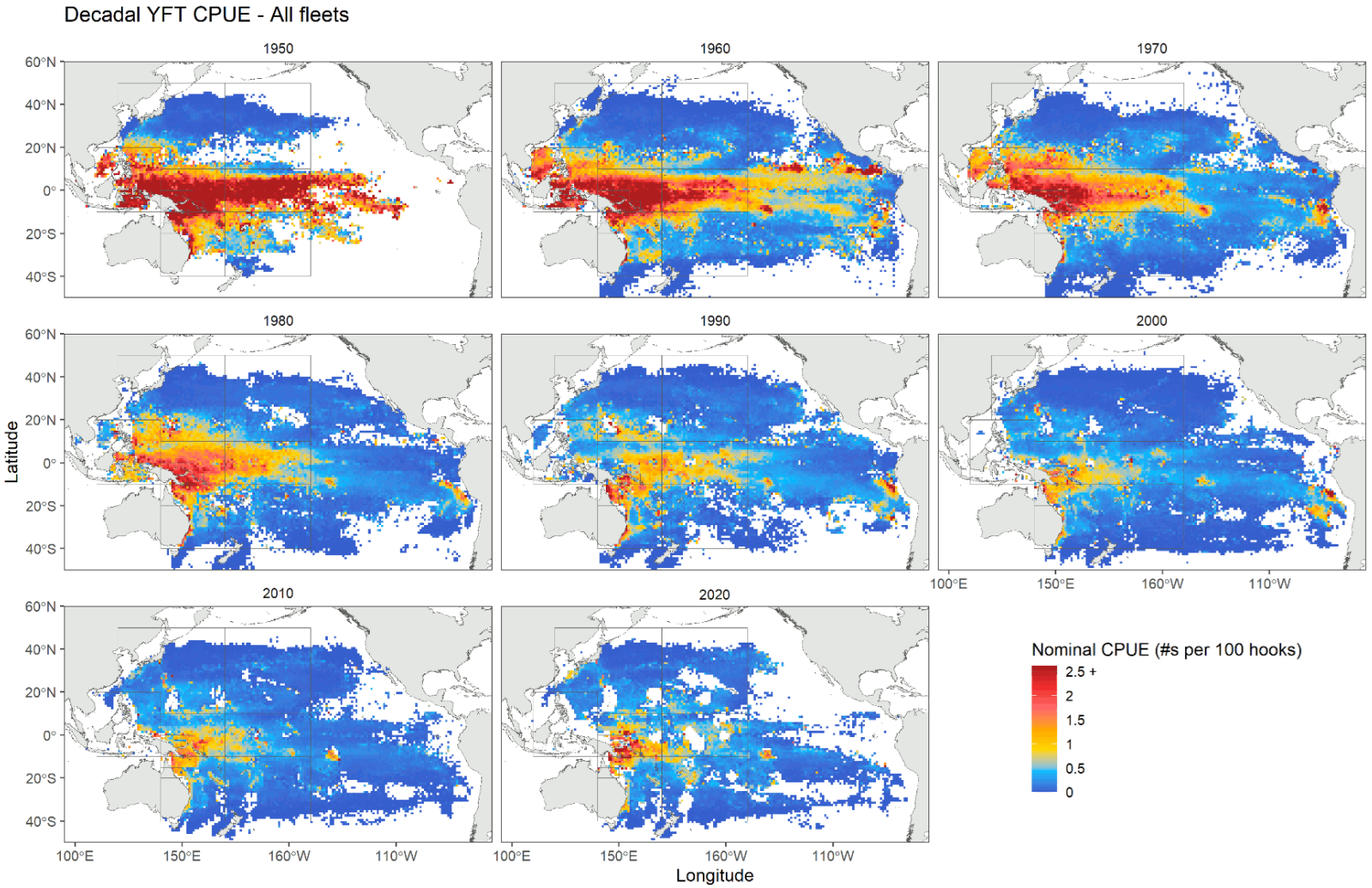

Figure 10: Spatial distribution of nominal longline CPUE (all fleets) for yellowfin in the Pacific.

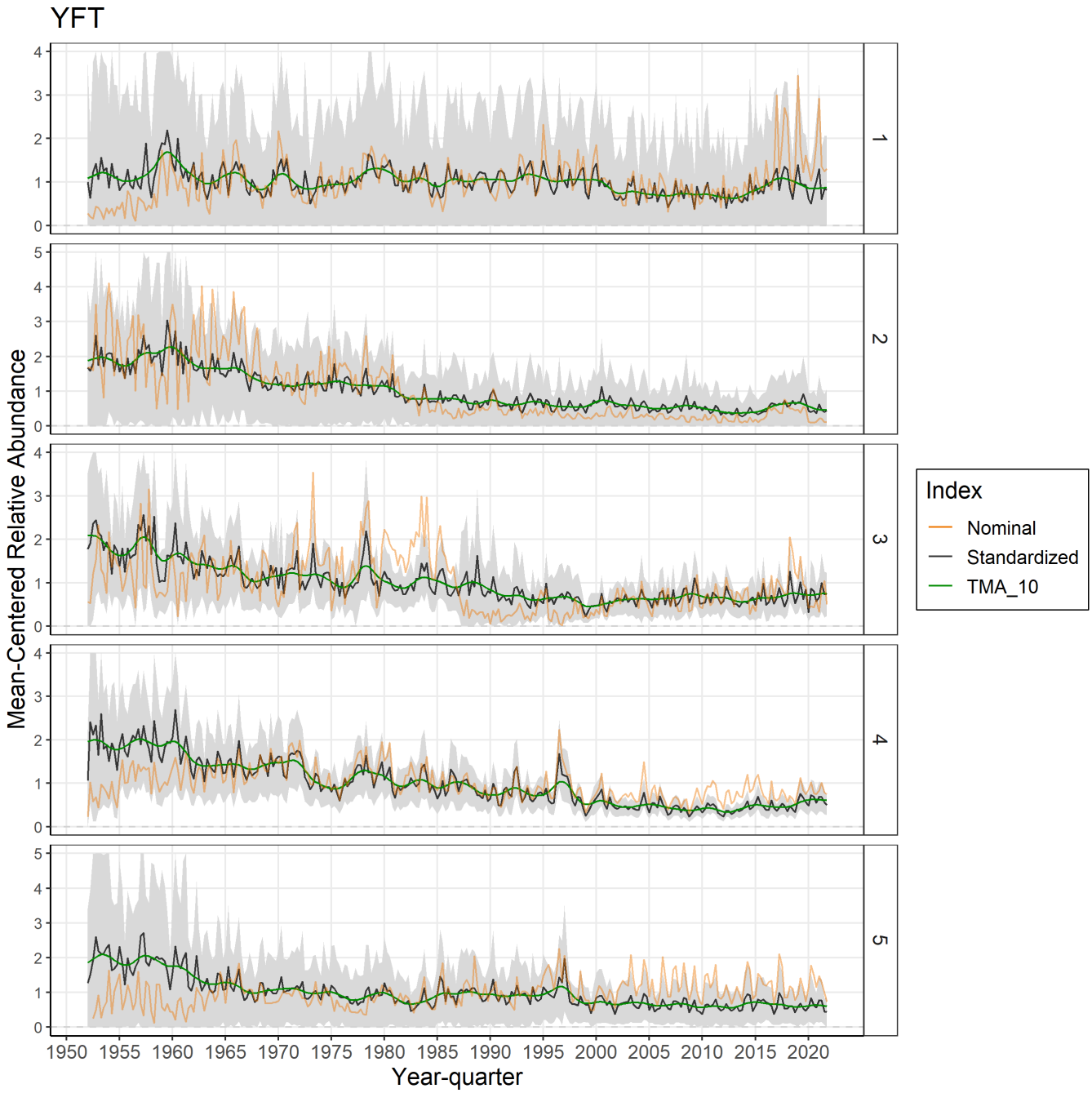

Figure 11: Standardised (black line) and nominal (orange) CPUE for the longline index fisheries in each model. Gray band is 95% CI. Triangular moving average smoothing function applied to demonstrate overall trend (green line; smoothing window = 10).

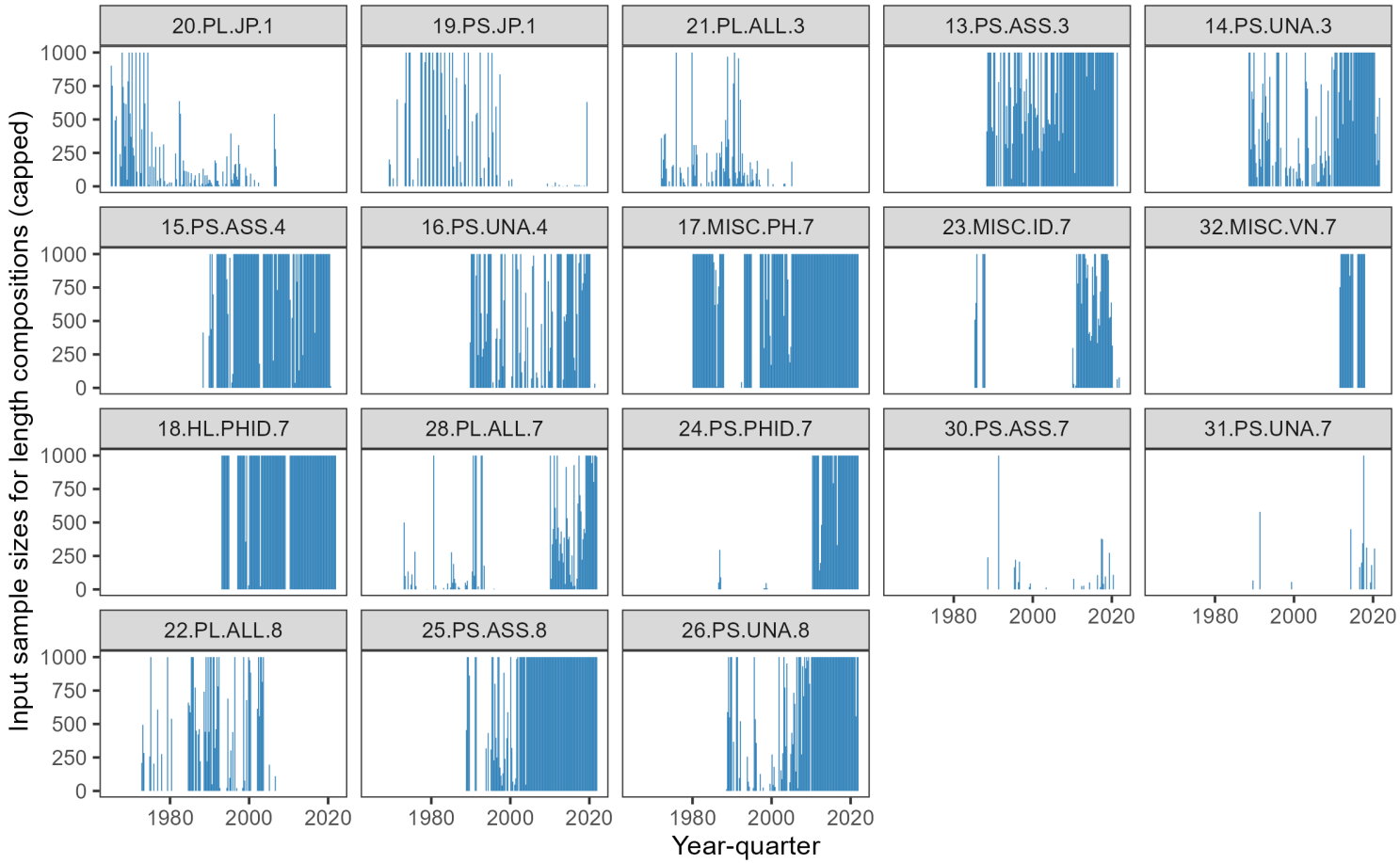

Figure 12: Plots of samples sizes (capped at 1,000) for length composition for each fishery in the model across the model time period.

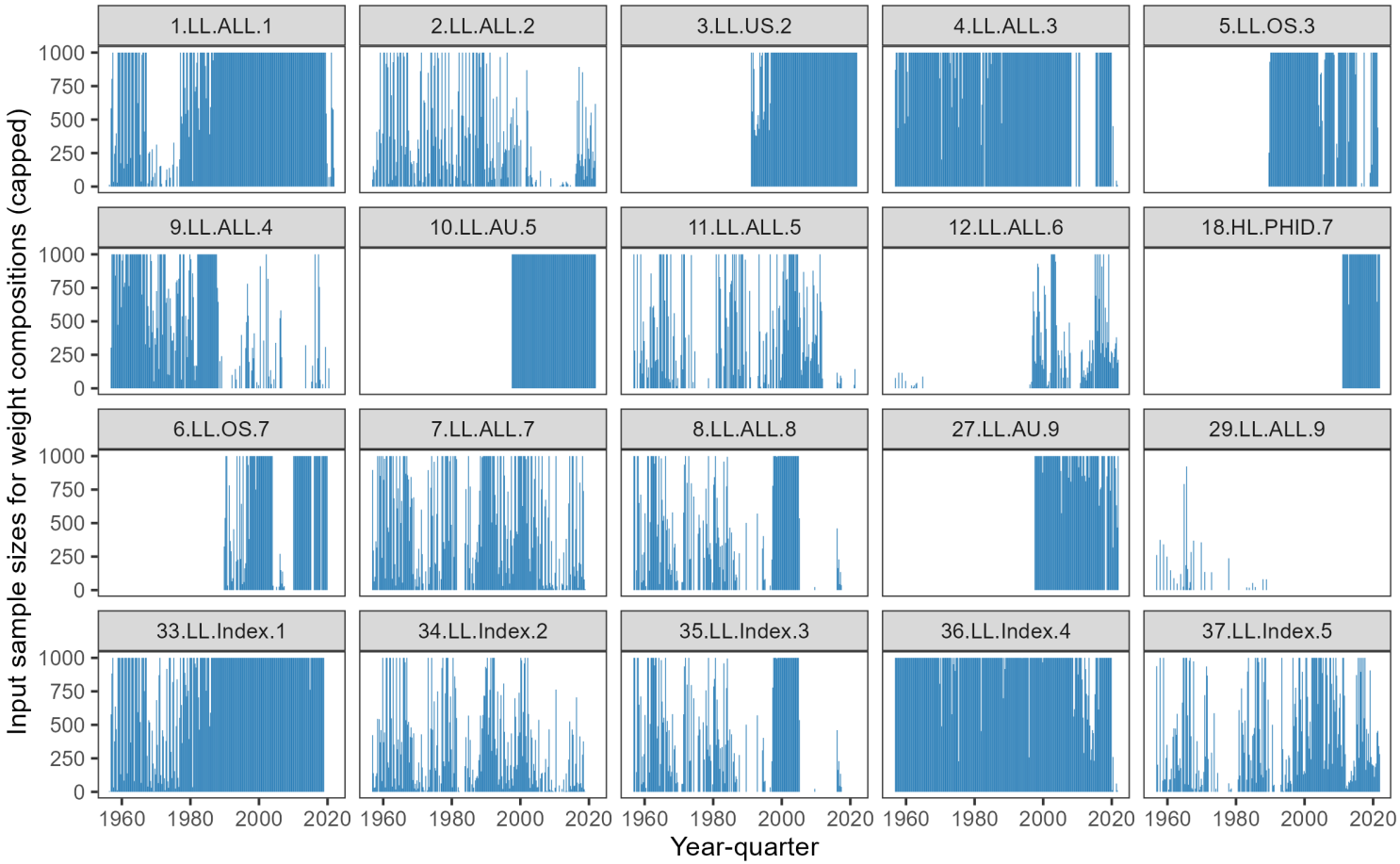

Figure 13: Plots of samples sizes (capped at 1,000) for weight composition for each fishery in the model across the model time period.

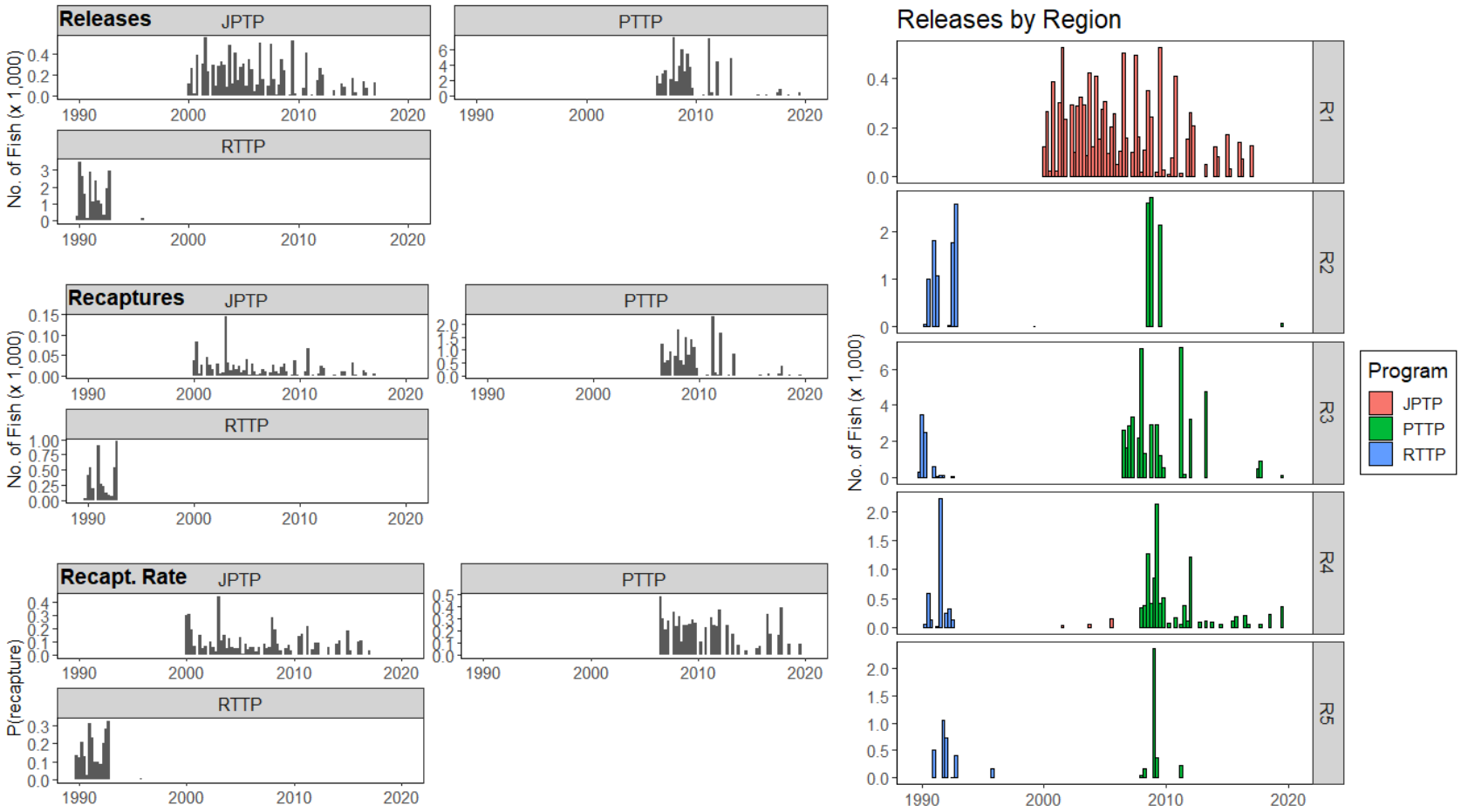

Figure 14: Summary plots of the number of releases, recaptures, and recapture rate of tags, by tagging program and region.

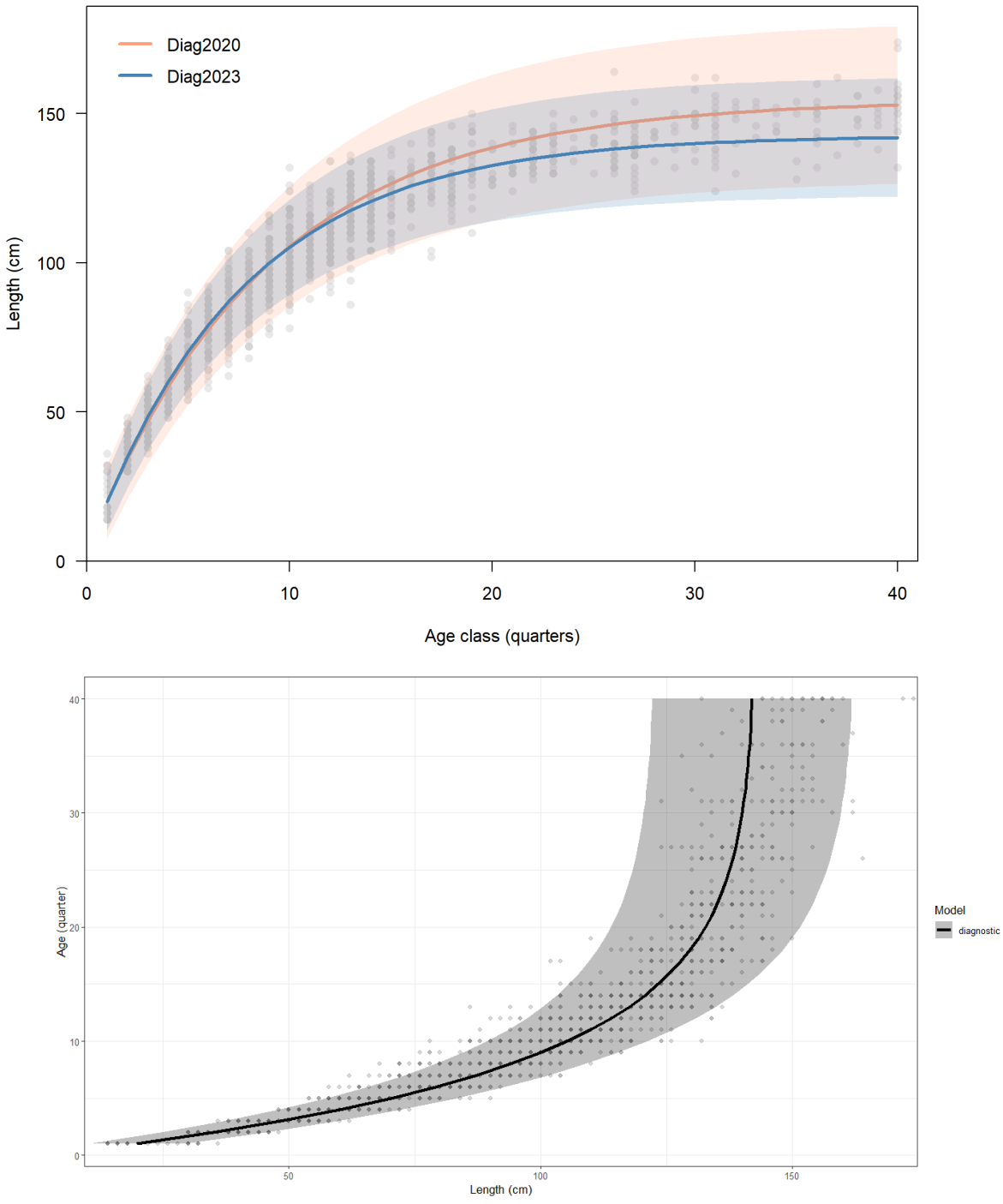

Figure 15: Growth curves compared between the 2020 diagnostic model and 2023 diagnostic model (top). Transposed version 2023 diagnostic model for conditional-age-at-length, vertical distribution of the points indicates the distributions of observed ages for each length (bottom). Points are age-length samples for the 2023 assessment.

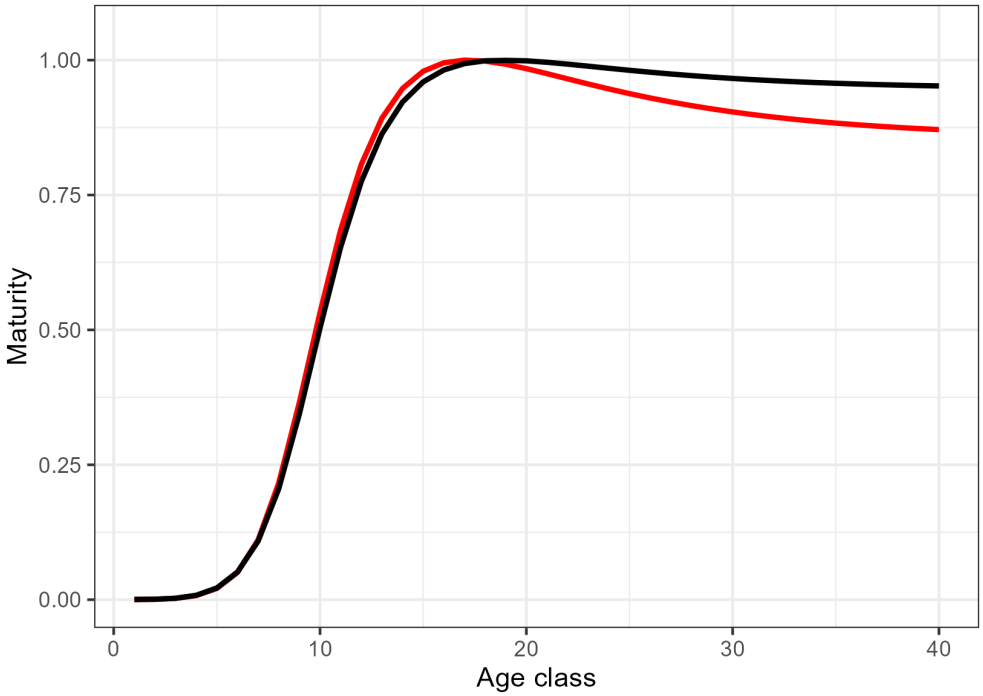

Figure 16: Maturity-at-age ogive compared between 2020 diagnostic model (red) and 2023 diagnostic model (black).

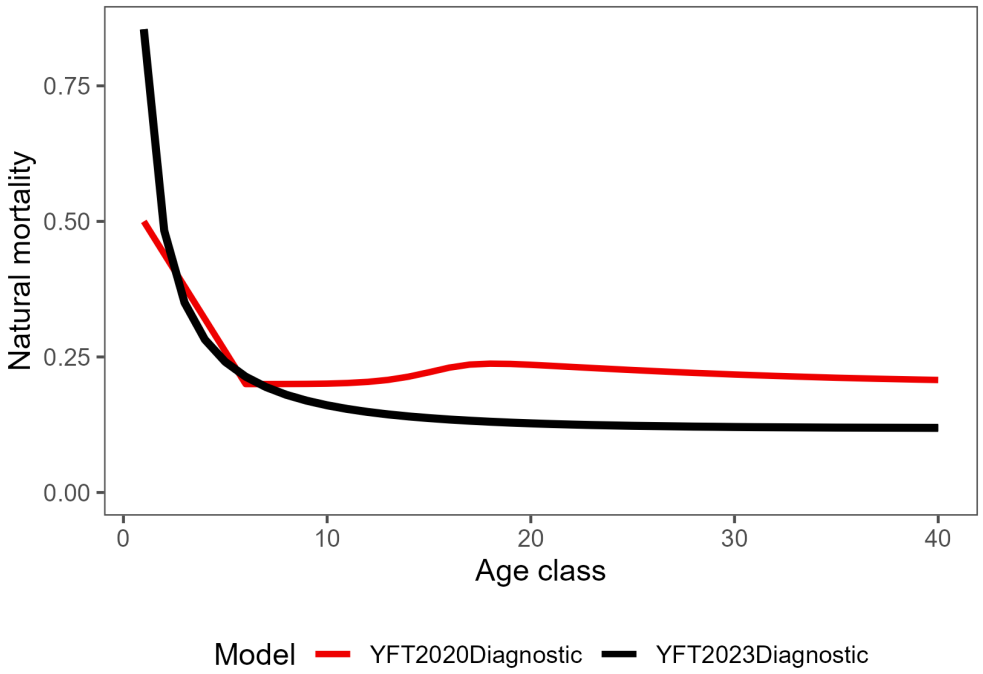

Figure 17: Natural mortality-at-age (quarters) for the 2020 diagnostic model (red) and the 2023 diagnostic model (black).

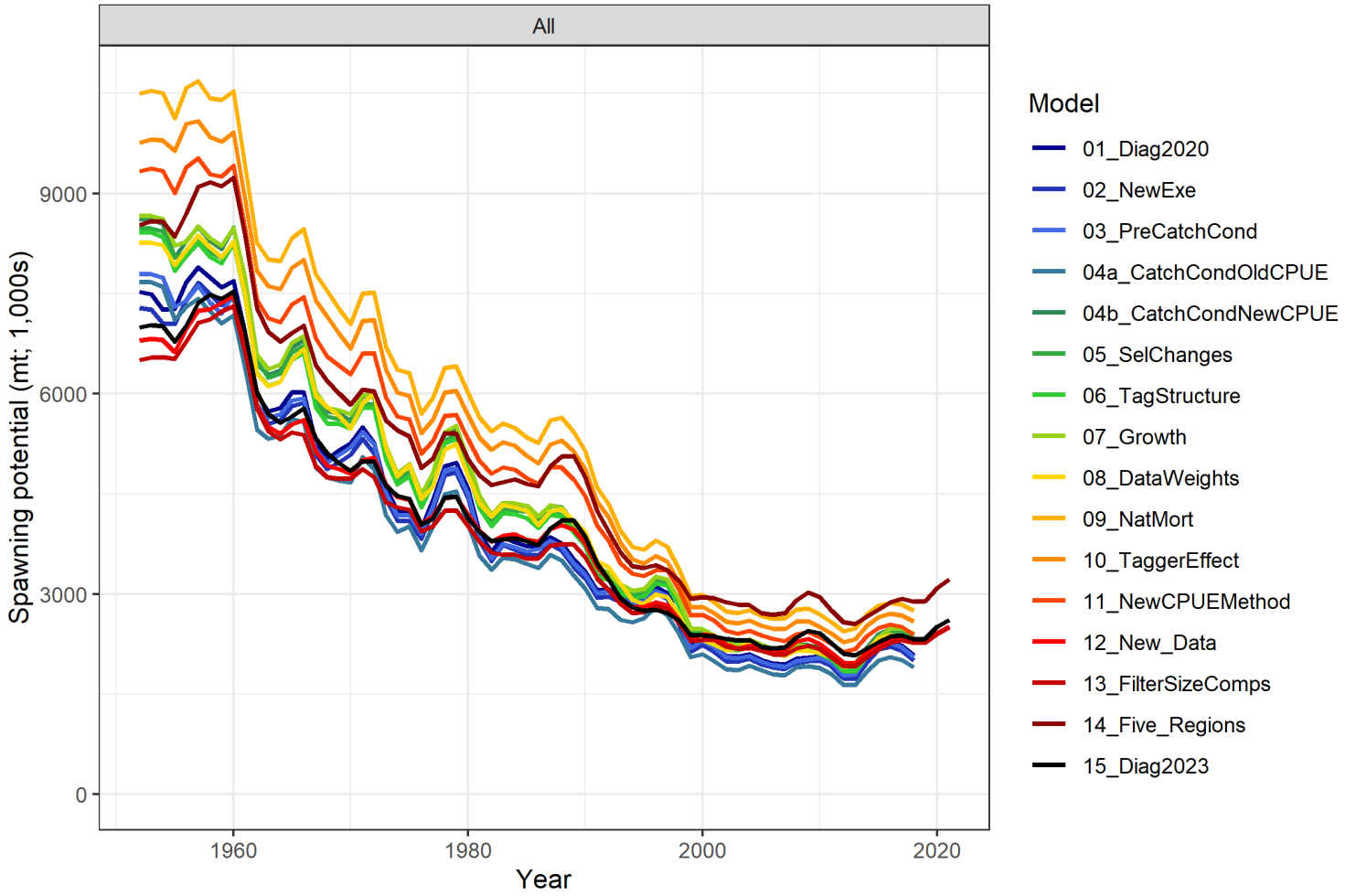

Figure 18: Estimated spawning potential, SB_t, trajectories for each of the main steps in the stepwise model runs, final diagnostic model is black.

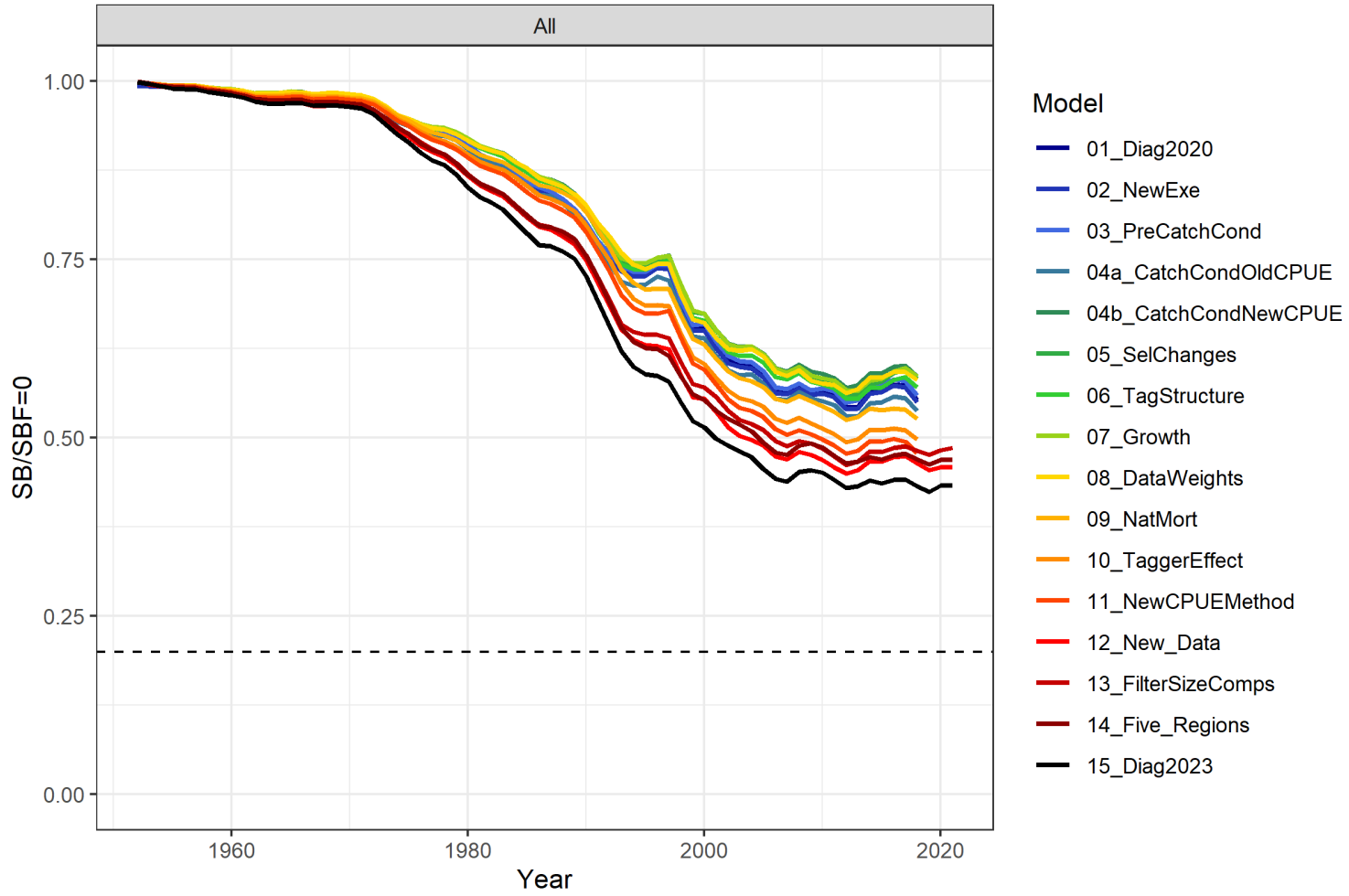

Figure 19: Estimated dynamic spawning depletion, $SB_t/SB_{t,F=0}$, trajectories for each of the main steps in the stepwise model runs, final diagnostic model is black.

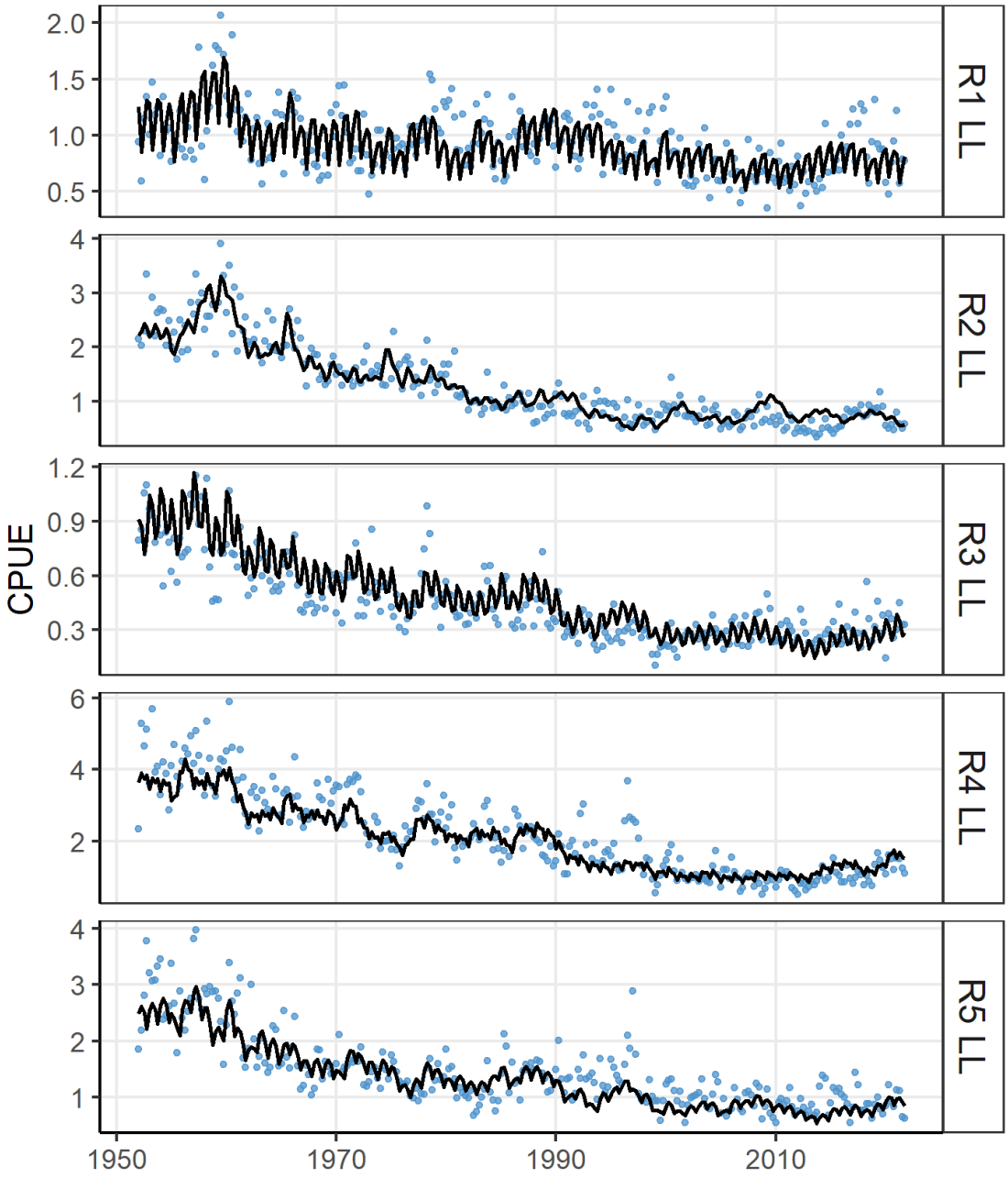

Figure 20: Comparison of model estimated (black line) and observed standardised CPUE (blue dots) for the longline index fisheries in each region.

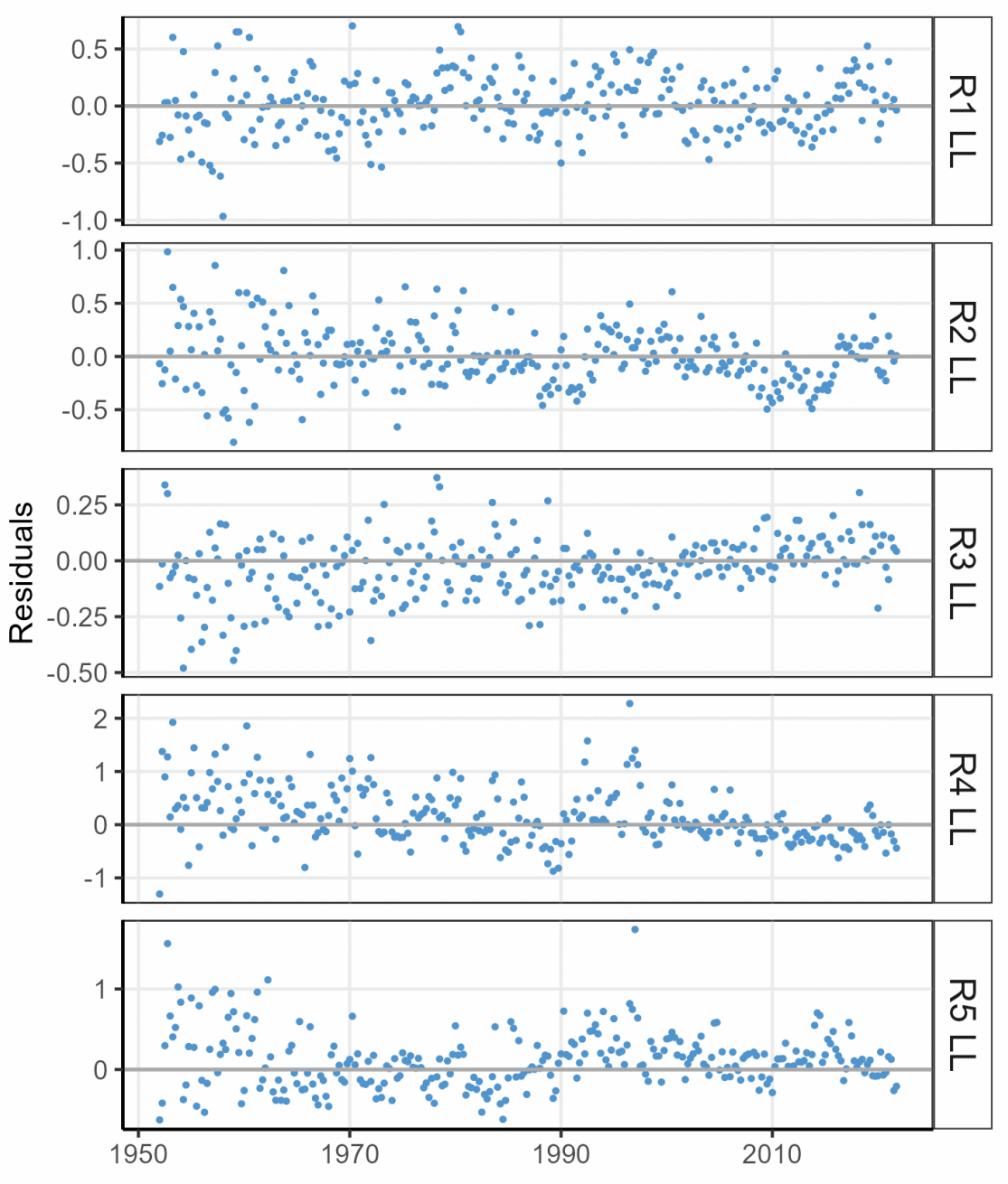

Figure 21: Plots of residuals between observed and predicted standardised CPUE for the longline index fisheries in each region.

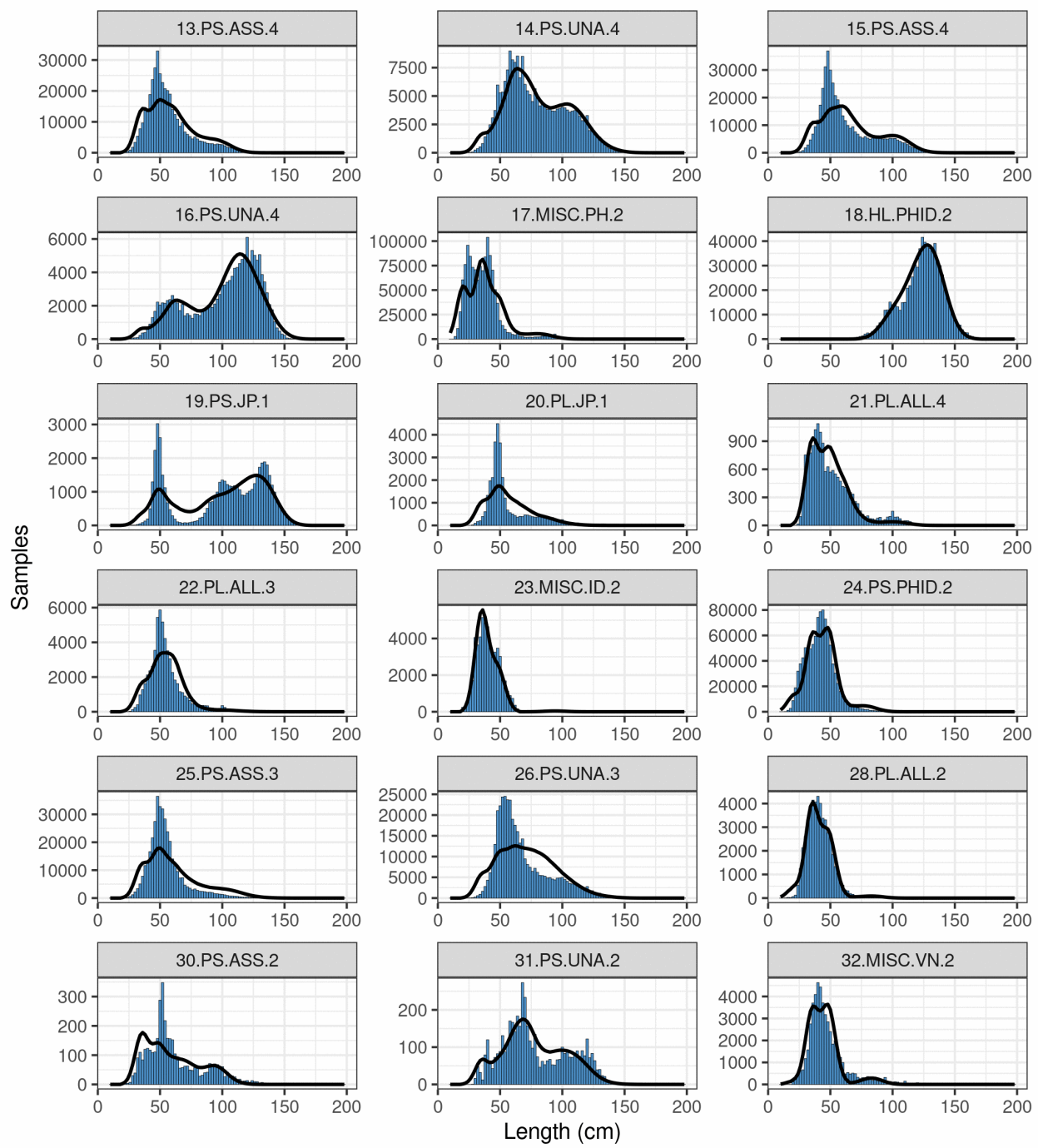

Figure 22: Composite (all time periods combined), observed (blue histograms), and predicted (black line) length frequency for fisheries with length frequency data for the diagnostic model.

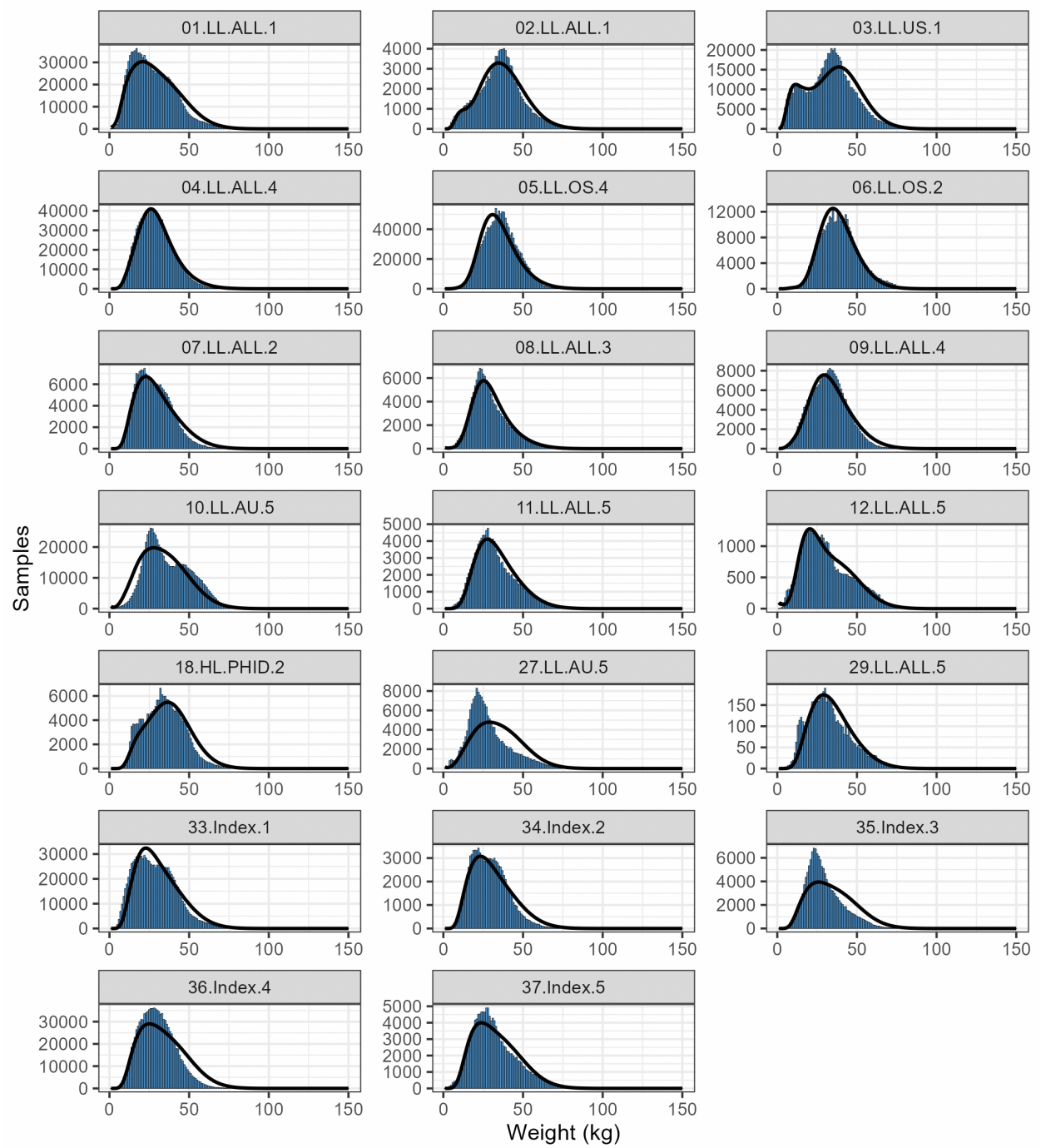

Figure 23: Composite (all time periods combined), observed (blue histograms), and predicted (black line) weight frequency for longline fisheries for the diagnostic model.

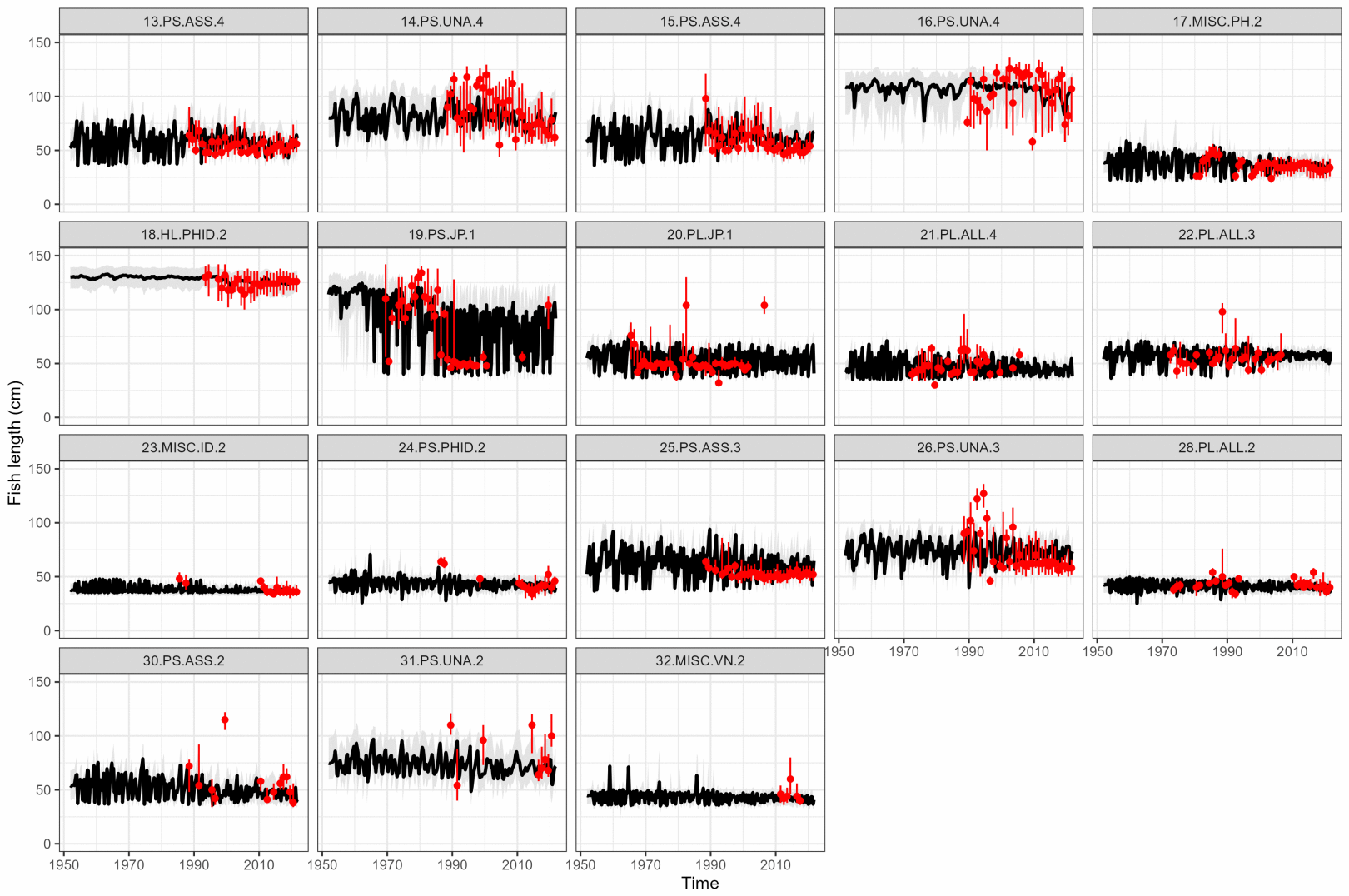

Figure 24: A comparison of the observed (red points) and predicted (black line) median fish length (FL, cm) for the fisheries with length data in the diagnostic model. The uncertainty intervals (gray shading) represent the values encompassed by the 25% and 75% quantiles. Sampling data are by quarter and only length samples more than 30 fish per quarter are plotted.

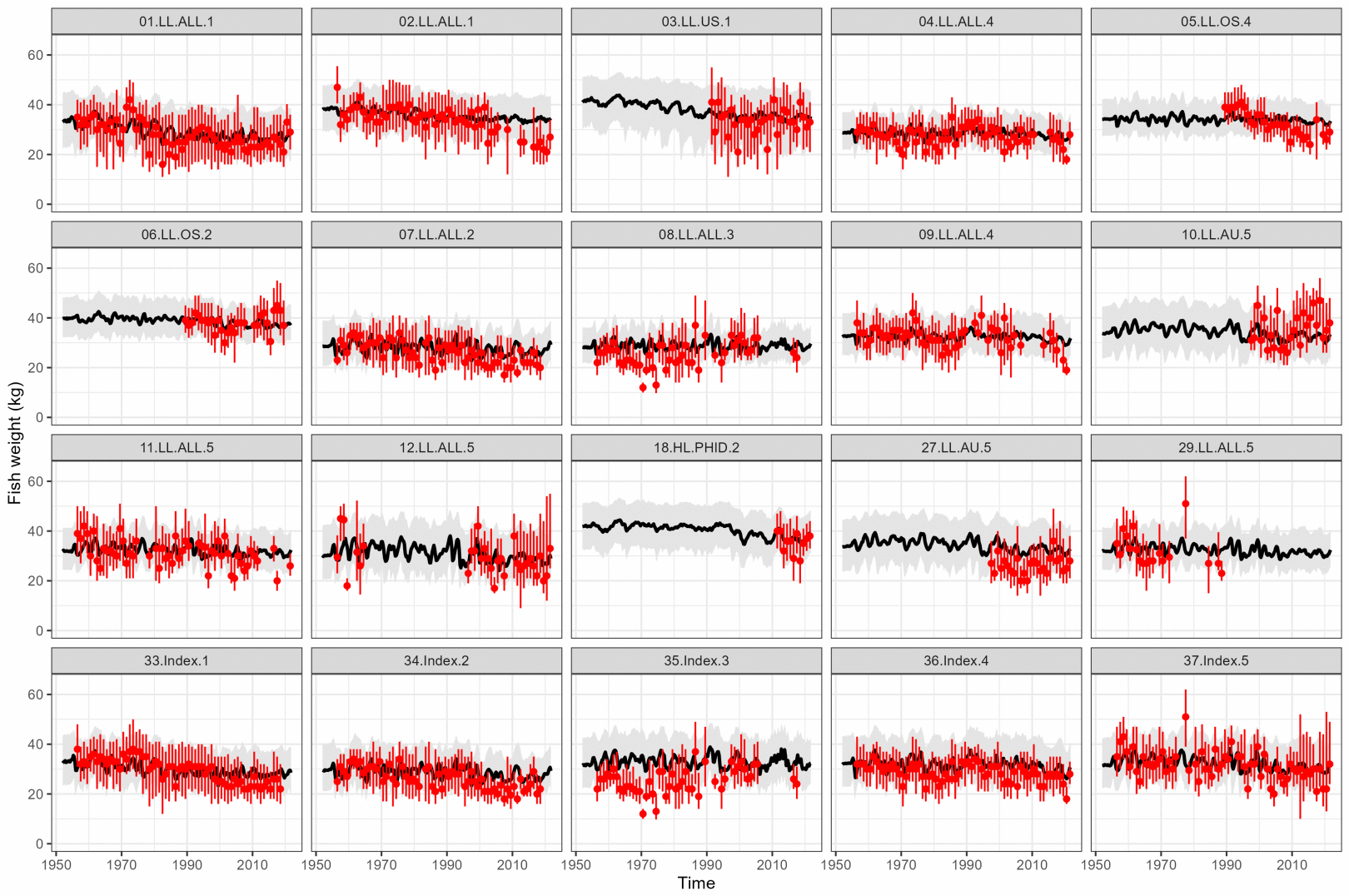

Figure 25: A comparison of the observed (red points) and predicted (black line) median fish weights (kg) for the longline fisheries in diagnostic model. The uncertainty intervals (gray shading) represent the values encompassed by the 25% and 75% quantiles. Sampling data are by quarter and only length samples more than 30 fish per quarter are plotted.

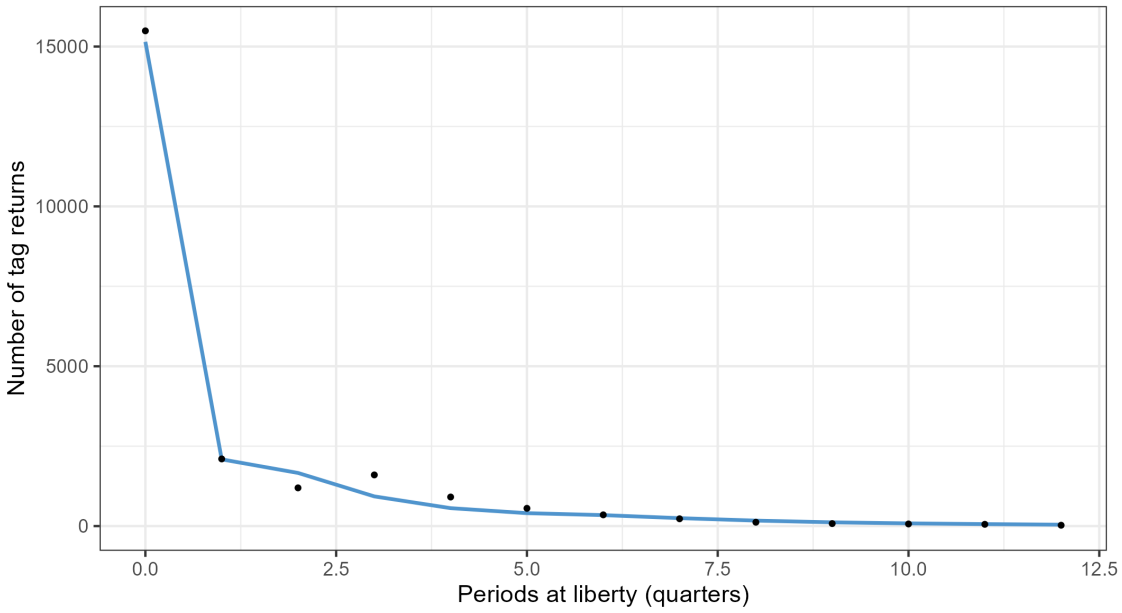

Figure 26: Observed (black points) and model-predicted (blue line) tag attrition across all tag release events for the diagnostic model.

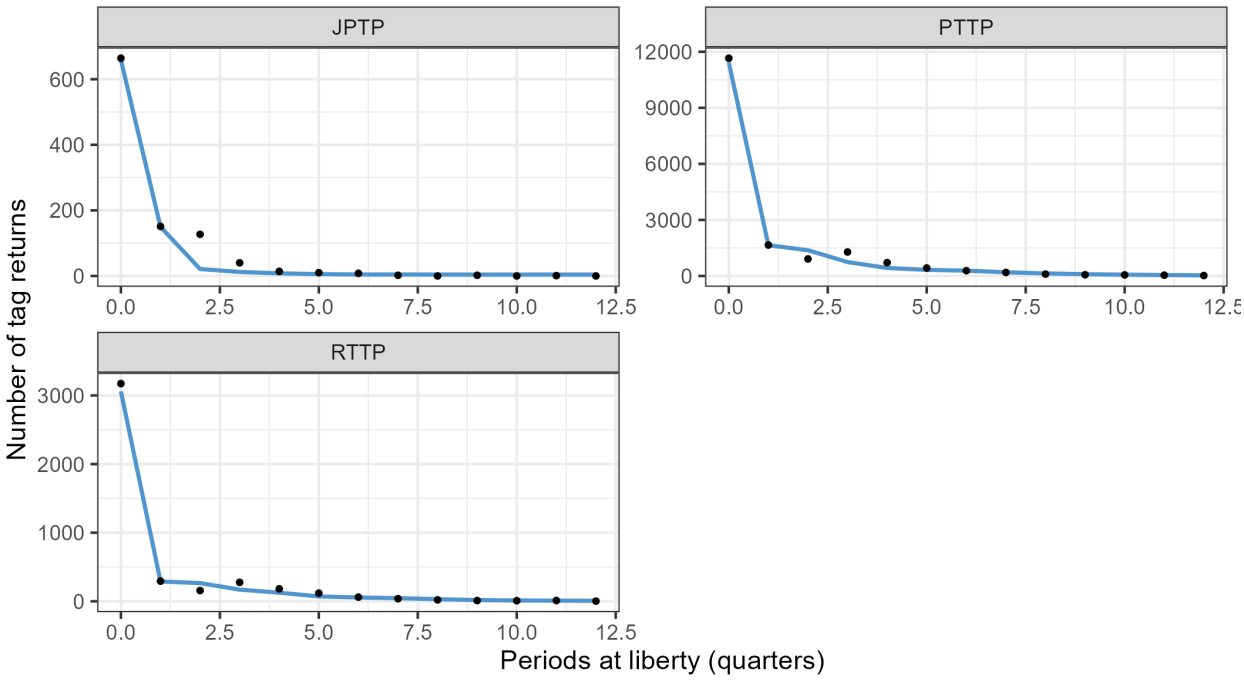

Figure 27: Observed (black points) and model-predicted (blue line) tag attrition by tagging programme for the diagnostic model.

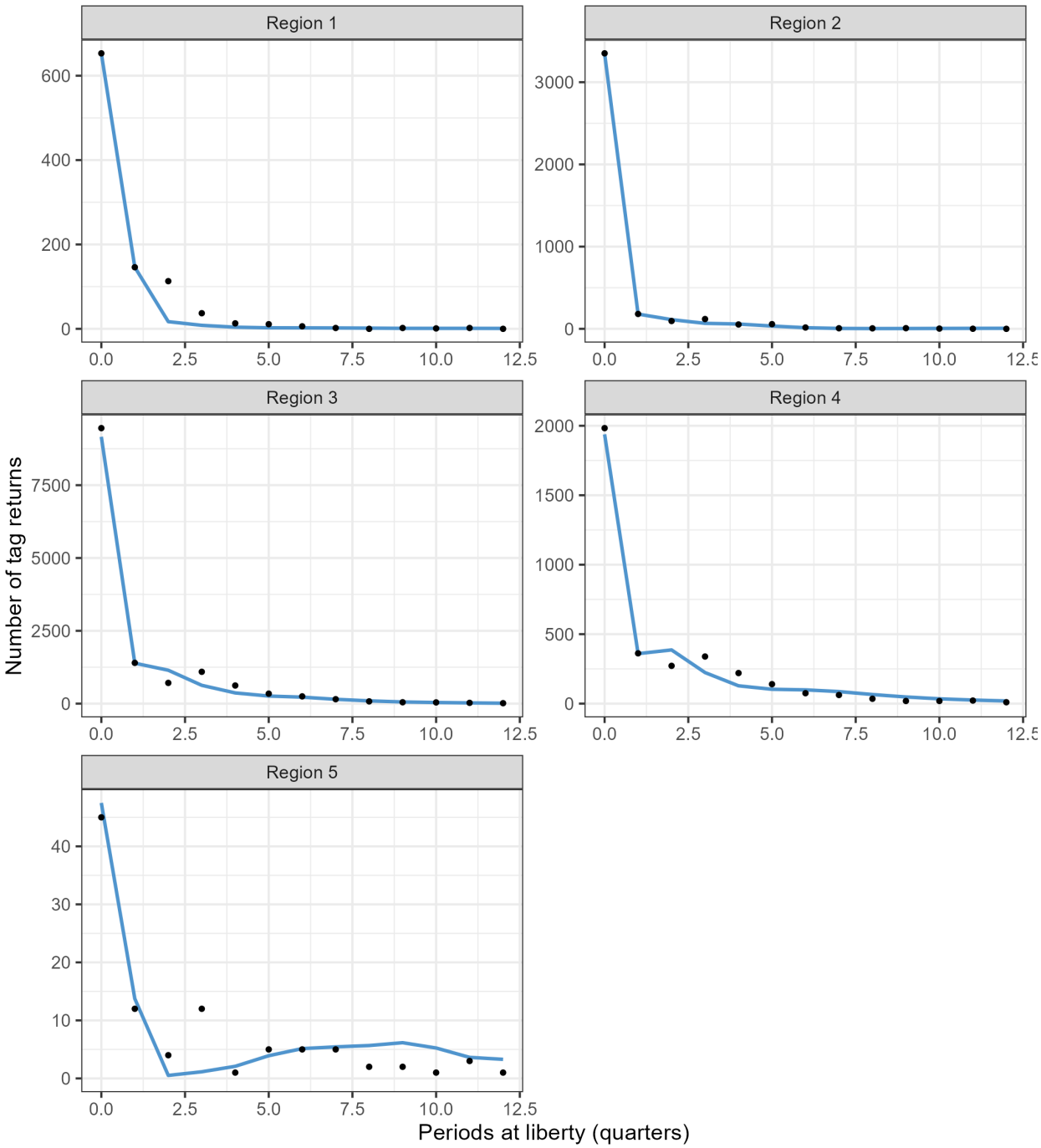

Figure 28: Observed (black points) and model-predicted (blue line) tag attrition by region for the diagnostic model.

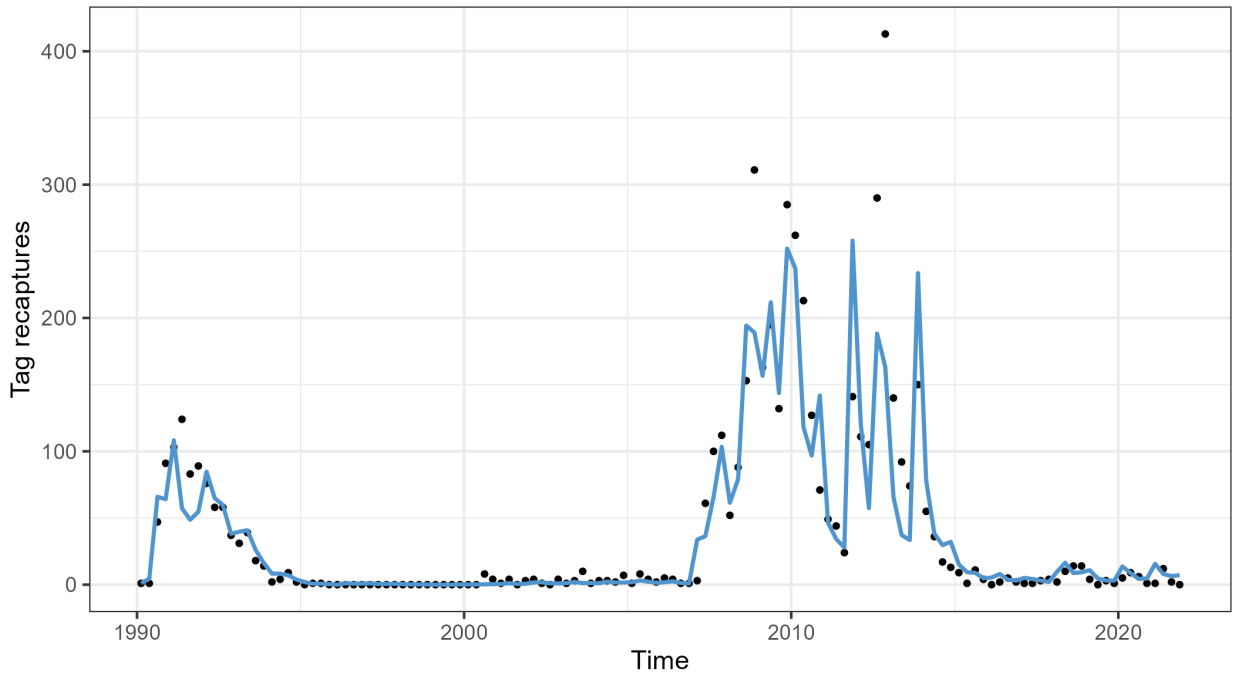

Figure 29: Observed (black points) and model-predicted (blue line) tag returns over time, with returns in the mixing period removed, for the diagnostic model across all tag release events with all tag recapture groupings aggregated.

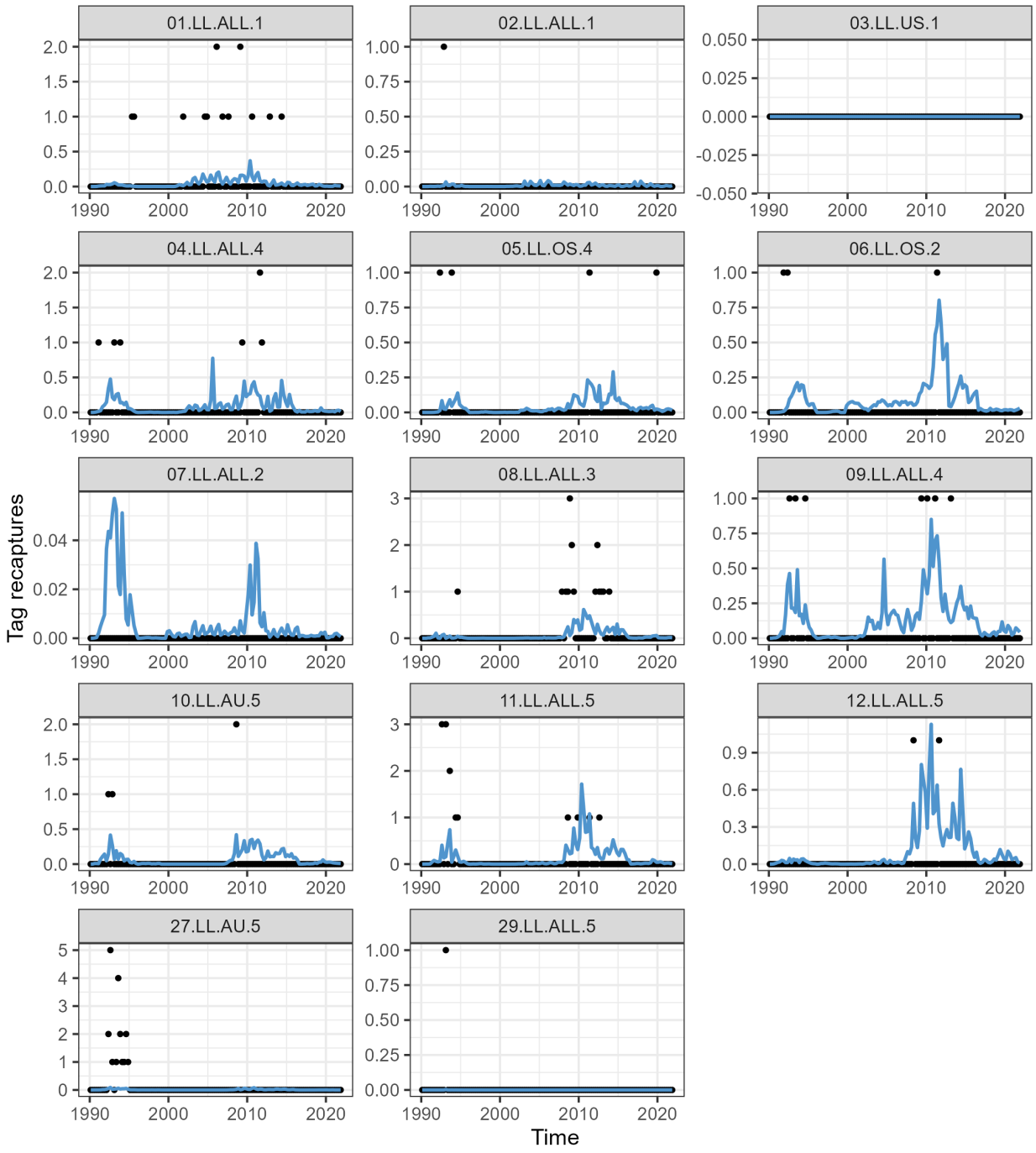

Figure 30: Observed (black points) and model-predicted (blue line) tag returns over time, with returns in the mixing period removed, for the diagnostic model for longline fisheries.

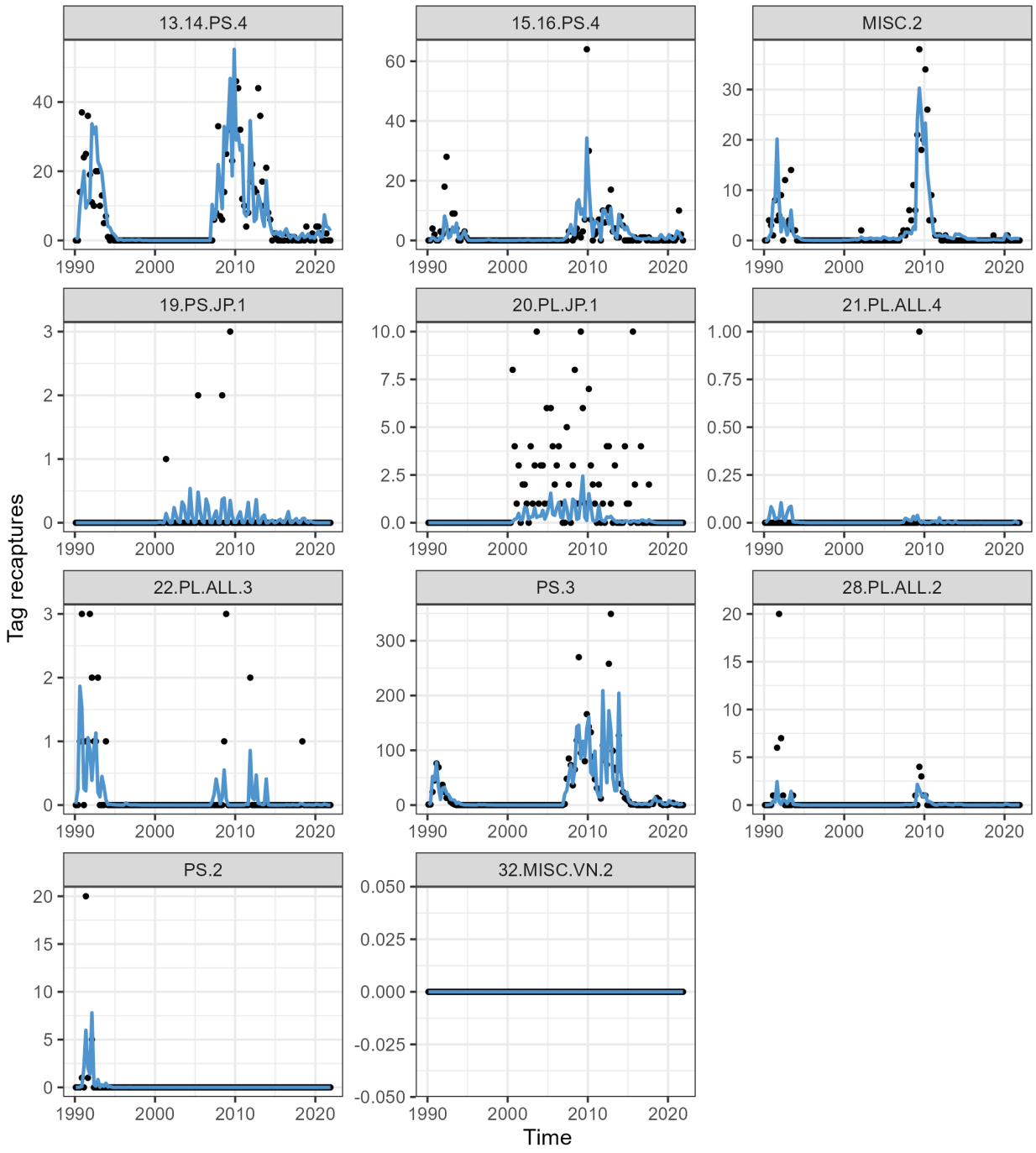

Figure 31: Observed (black points) and model-predicted (blue line) tag returns over time, with returns in the mixing period removed, for the diagnostic model for other (non-longline) fisheries.

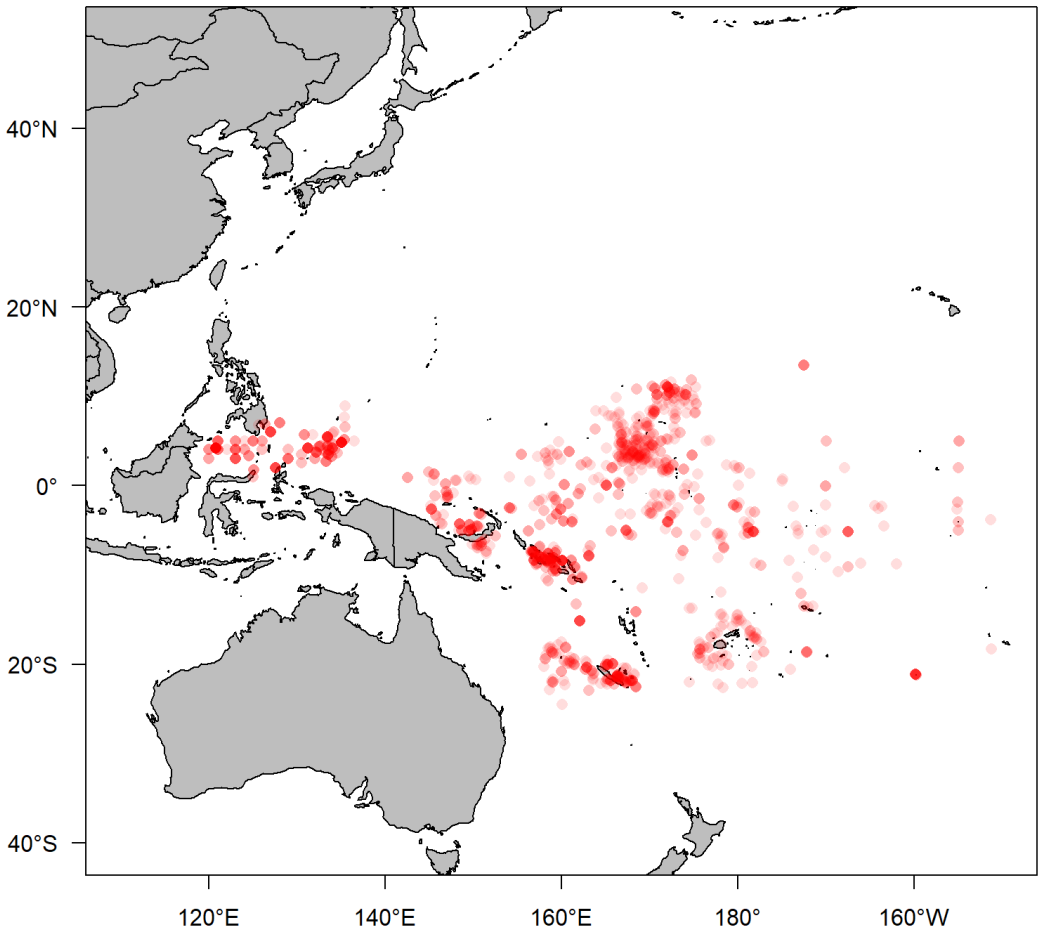

Figure 32: Sample locations of otoliths ($n = 1471$) used in the assessment model to inform internal growth estimation. Single otoliths are shown as pink circles and overlapping otoliths as progressively more saturated red circles.

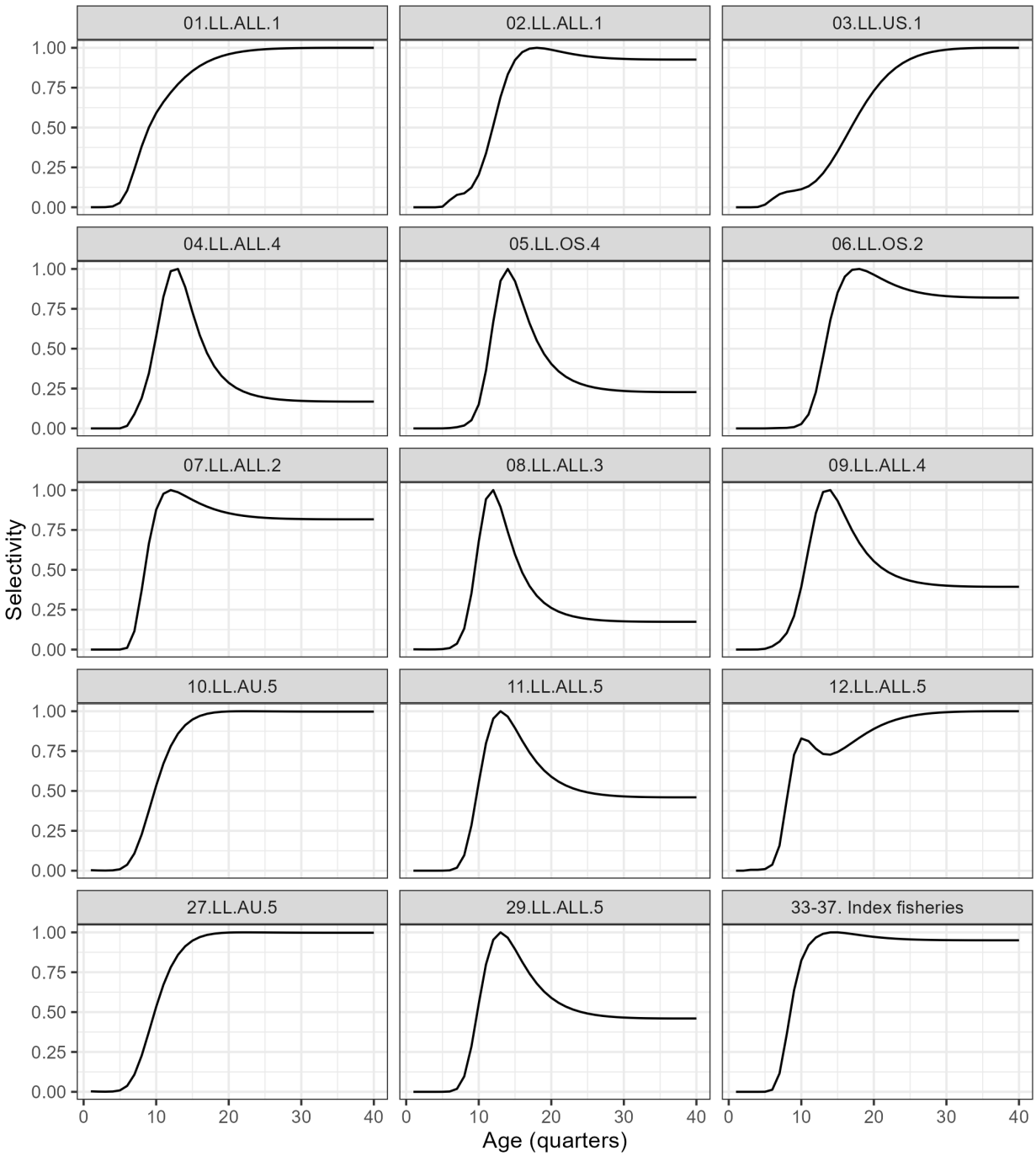

Figure 33: Age-specific selectivity coefficients for longline fisheries with shared selectivities, with one panel for index fisheries which share selectivity across model regions.

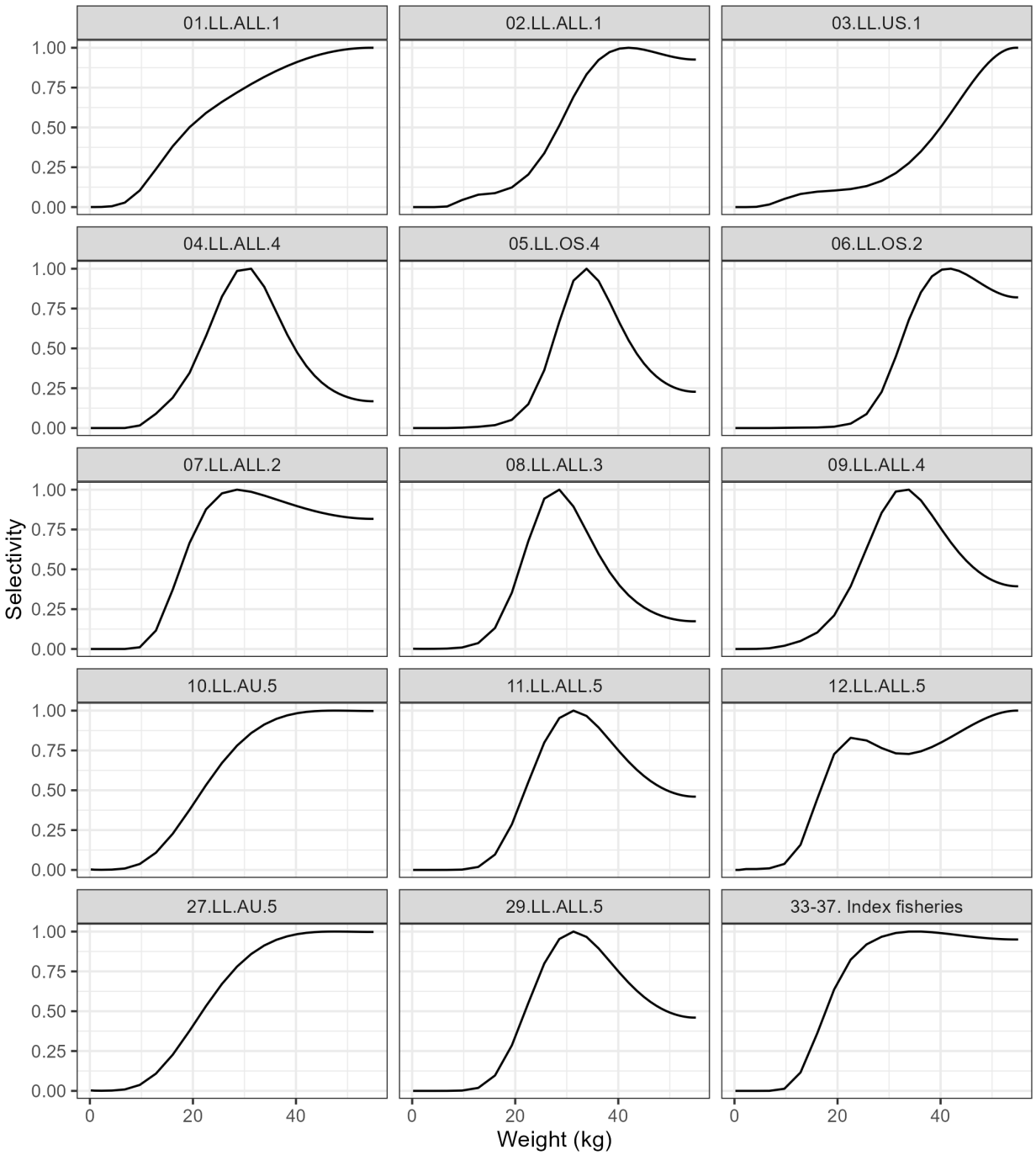

Figure 34: Weight-specific selectivity coefficients for longline fisheries with shared selectivities, with one panel for index fisheries which share selectivity across model regions.

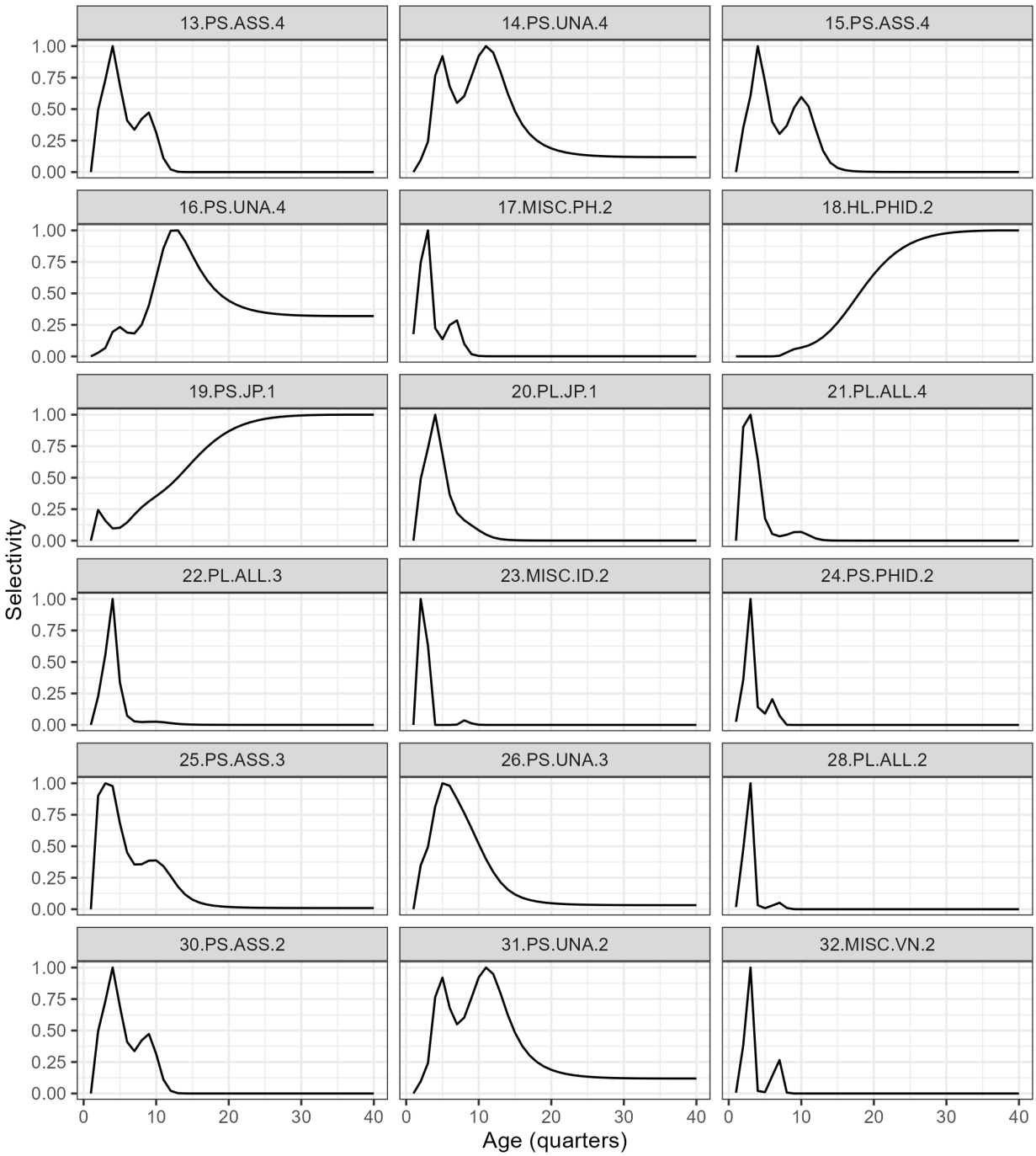

Figure 35: Age-specific selectivity coefficients by groups of fisheries with shared selectivities for miscellaneous and pole and line fisheries.

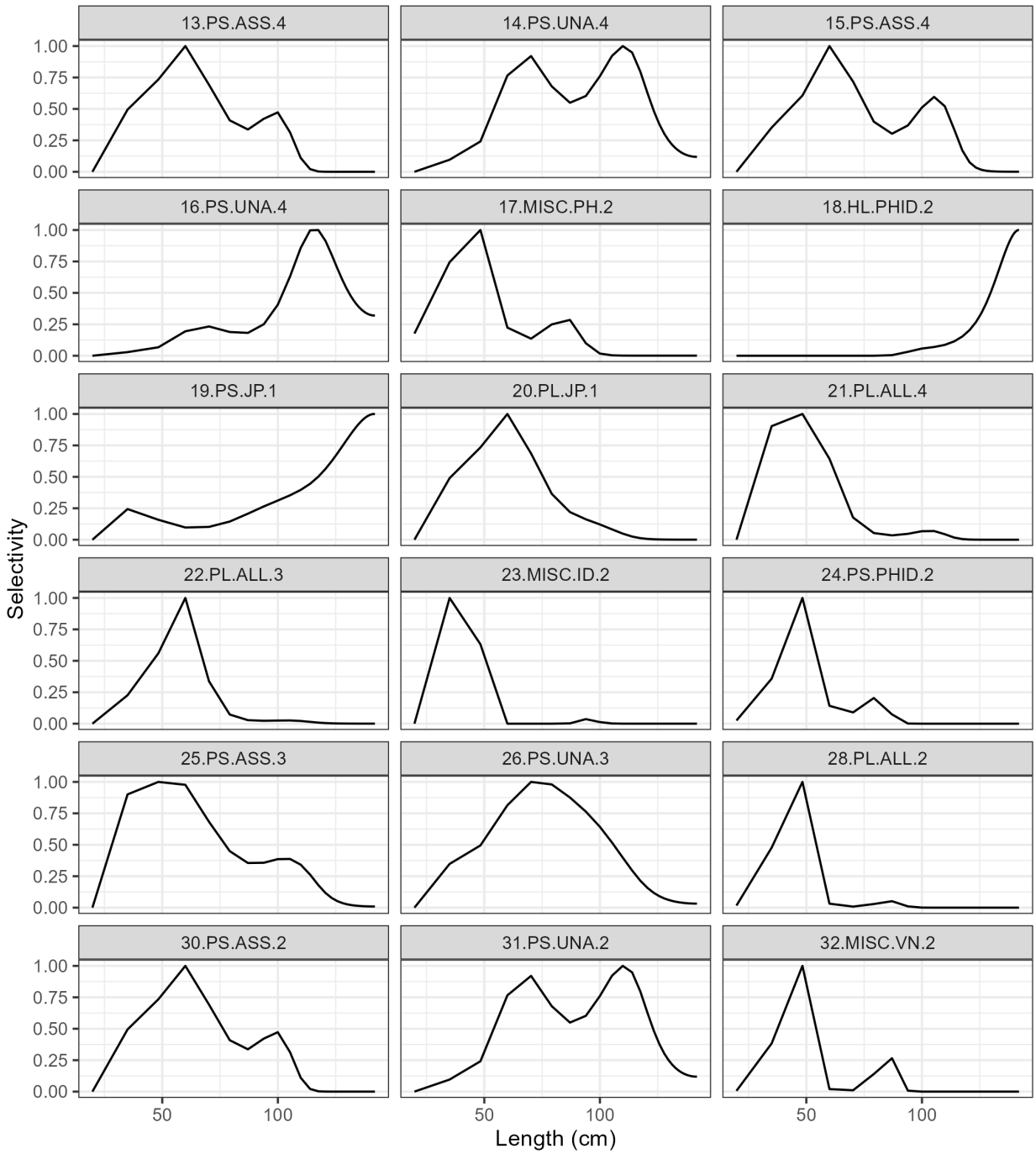

Figure 36: Length-specific selectivity coefficients by groups of fisheries with shared selectivities for miscellaneous and pole and line fisheries.

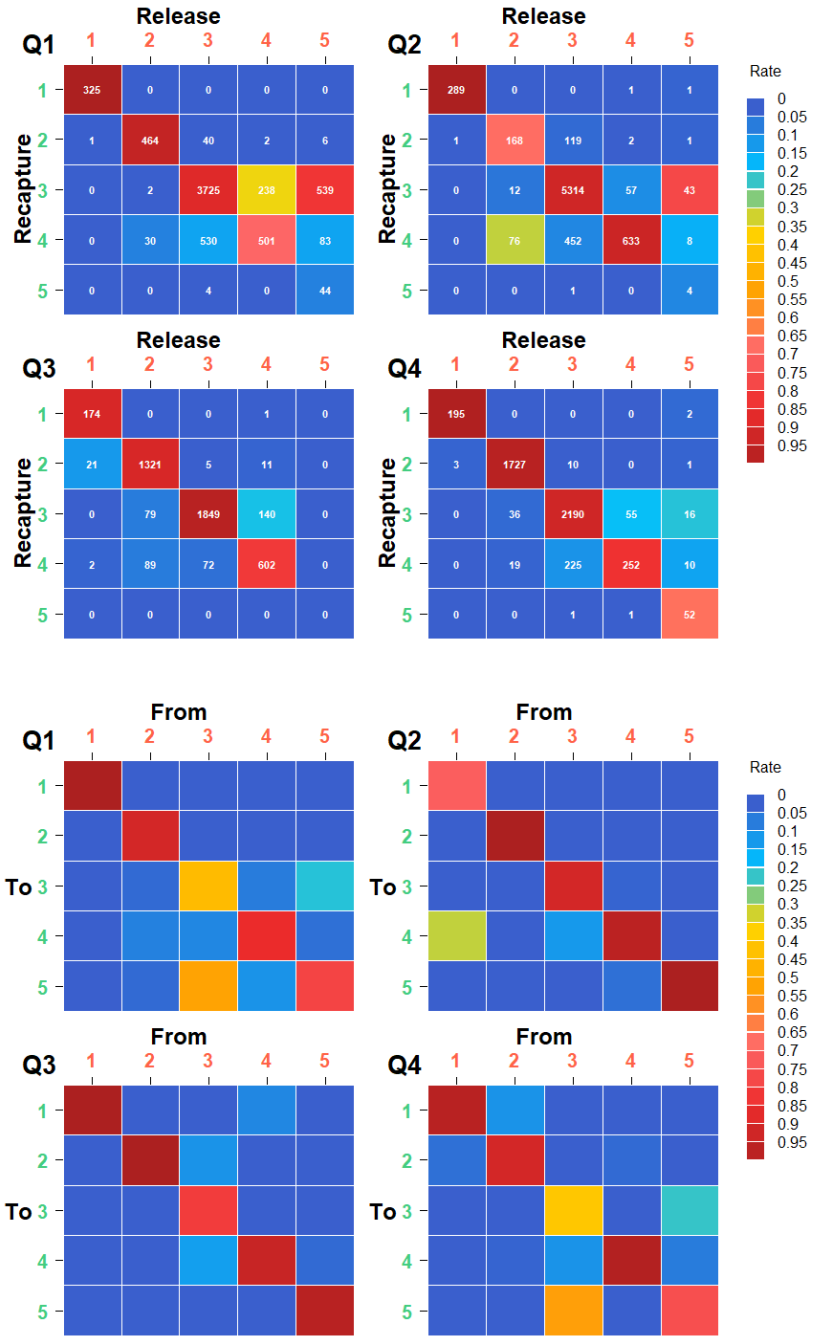

Figure 37: (Top) Observed proportion of tags returned by region of release (columns), region of recapture (rows), and quarter of recapture (panel) where the color of the tile indicates the proportion of tags returned from the region of releases, numbers in boxes indicate the actual numbers. (Bottom) Estimated movement probabilities by quarter for the diagnostic case model. The red numbers (horizontal axis) indicate the source model region (From) and the green numbers (vertical axis) indicate the receiving (To) regions. The color of the tile shows the magnitude of the movement rate (proportion of individuals moving from region x to region y in that quarter), with each column adding up to 1.

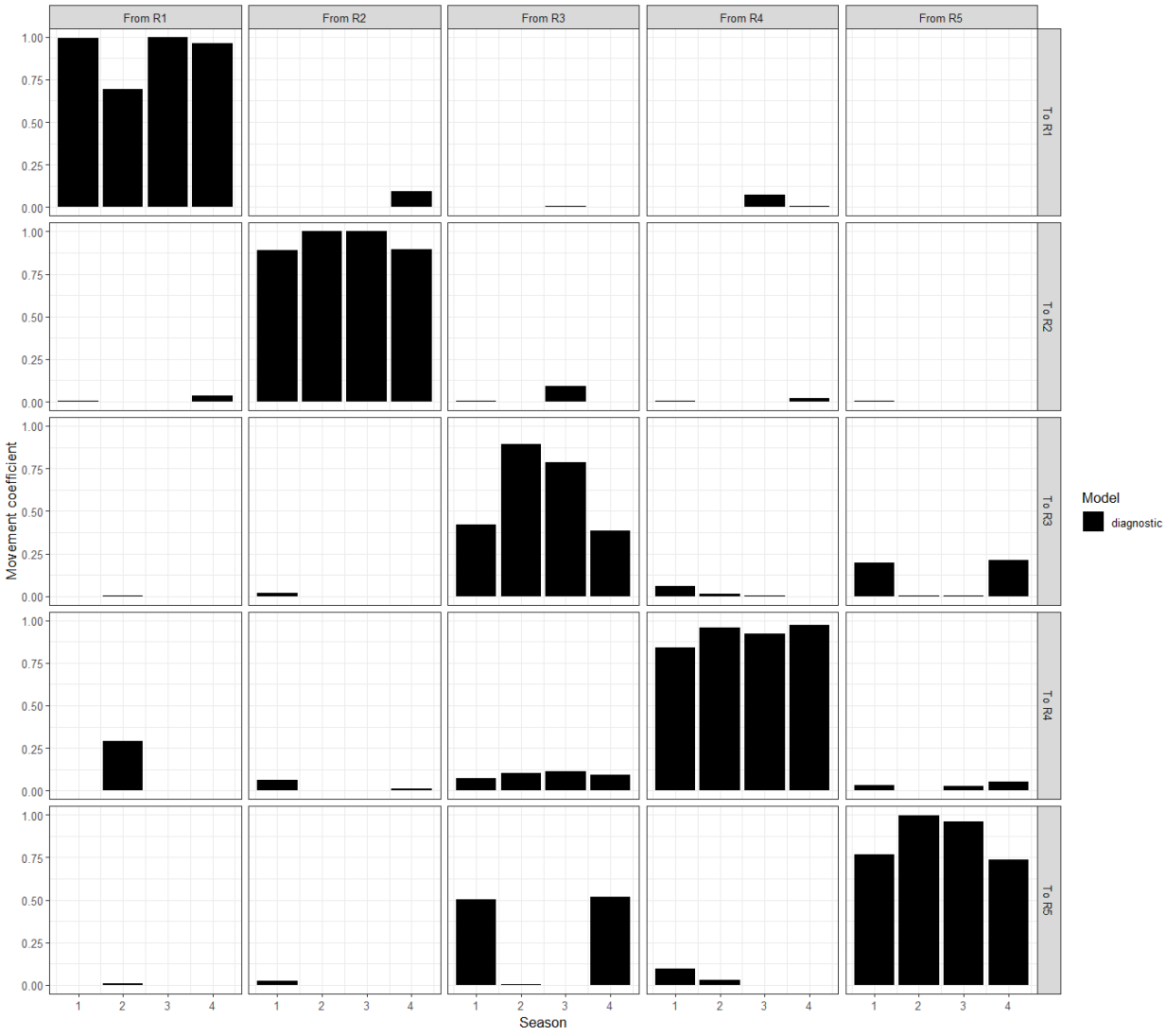

Figure 38: Season-specific movement probabilities estimated by the diagnostic model.

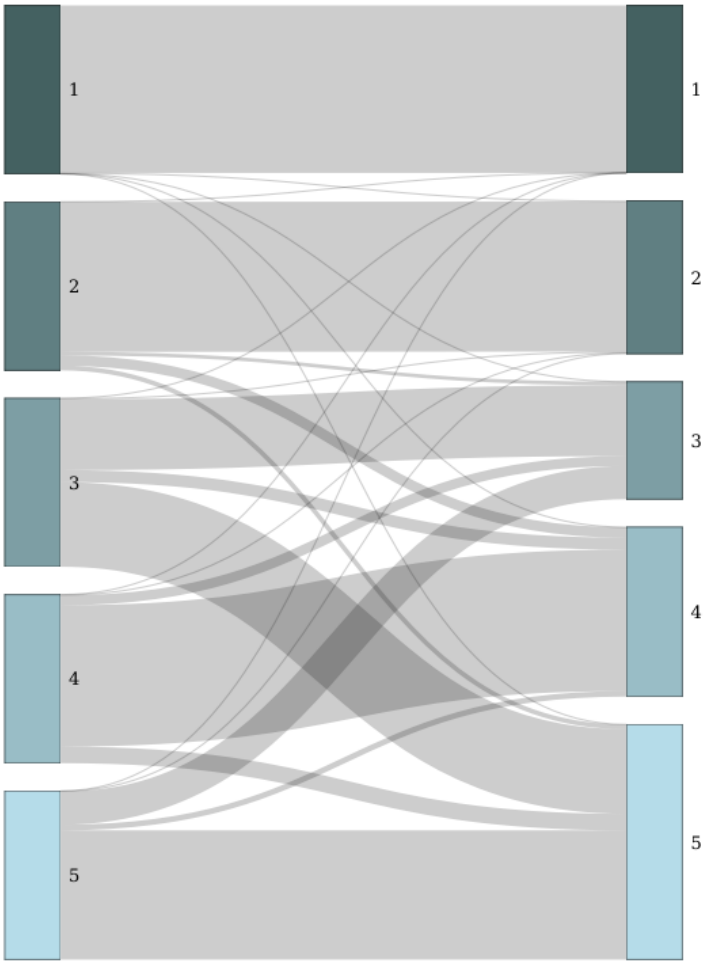

Figure 39: Stylised estimated movement rates between stock assessment regions (all ages and seasons) for the diagnostic case. Estimated movement is shown FROM the model regions on the left TO the model regions on the right.

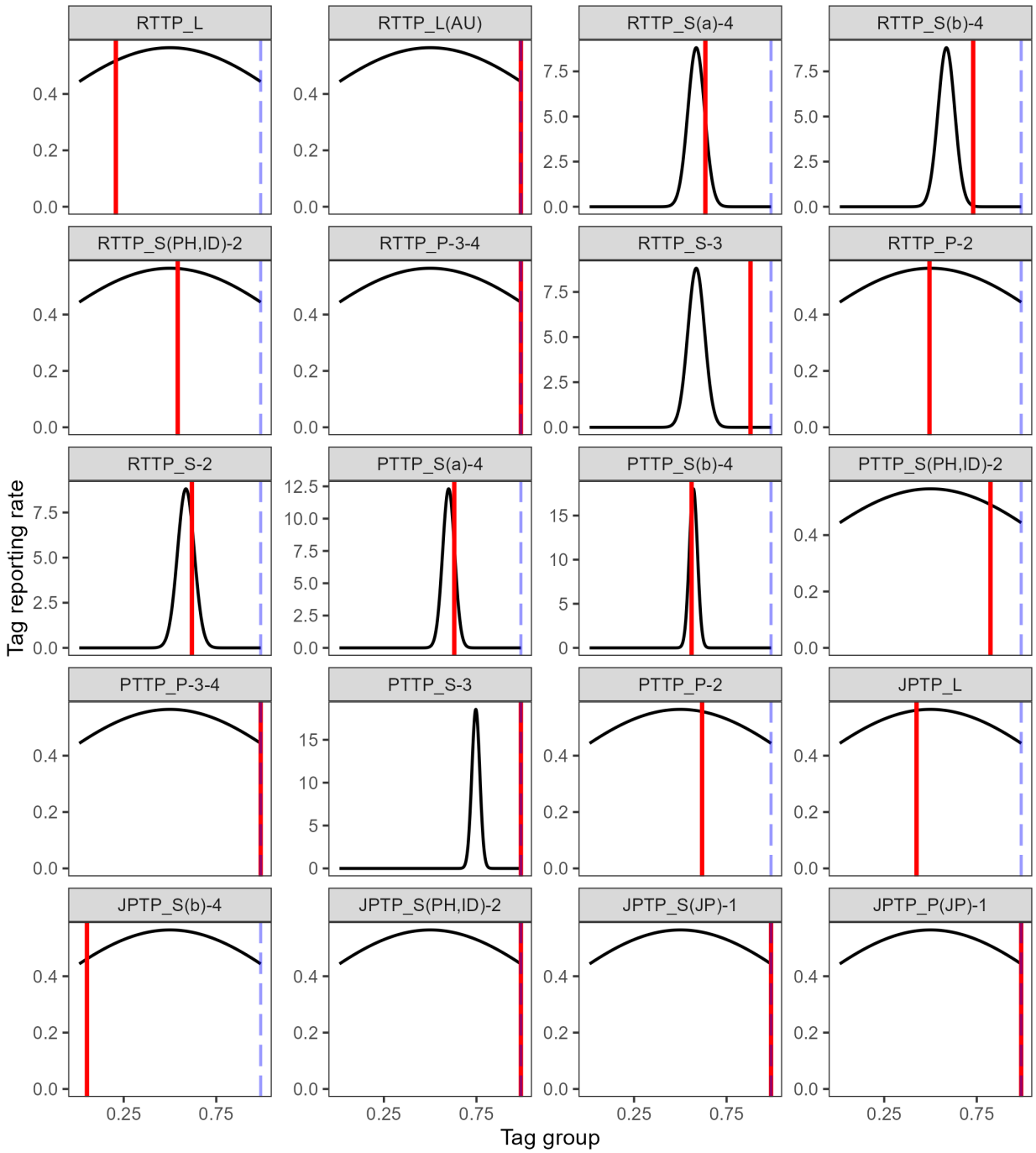

Figure 40: Estimated reporting rates for the diagnostic model (red lines) and the prior distribution for each reporting rate group (black lines). The imposed upper bound (0.99) on the reporting rate parameters is shown as a blue dashed line. Reporting rates can be estimated separately for each release program and recapture fishery group but in practice are aggregated over some recapture groups to reduce dimensionality.

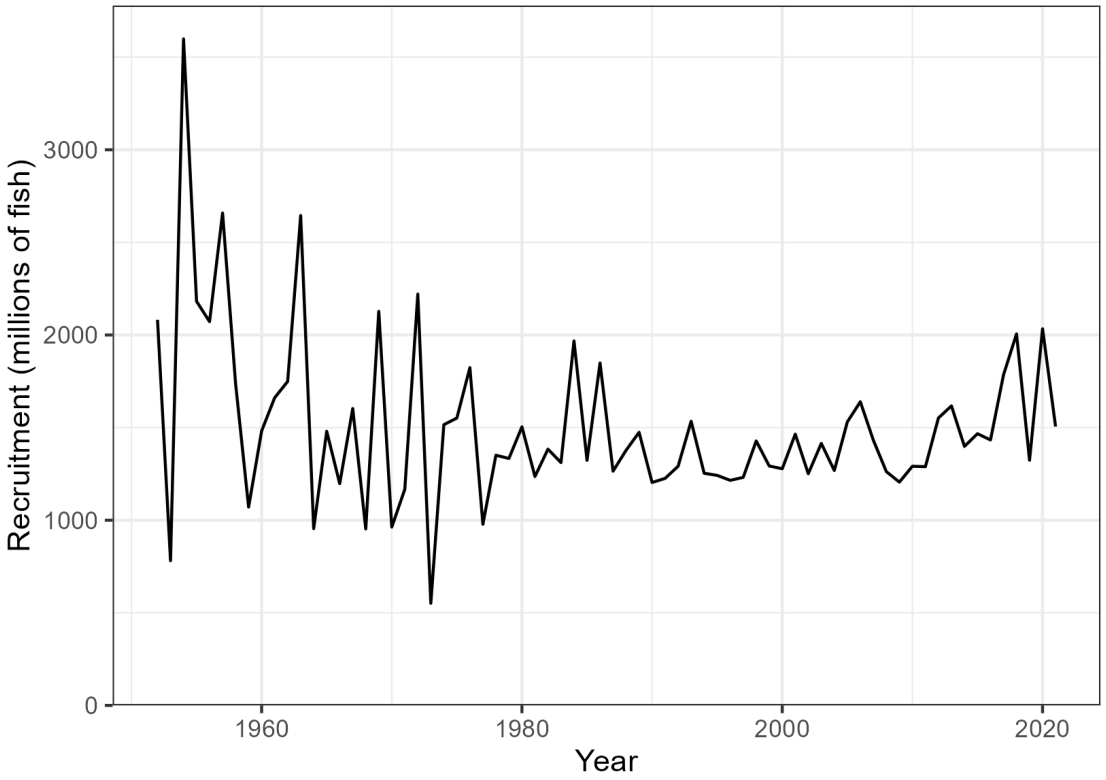

Figure 41: Time series of estimated annual recruitment summed across regions for the diagnostic model.

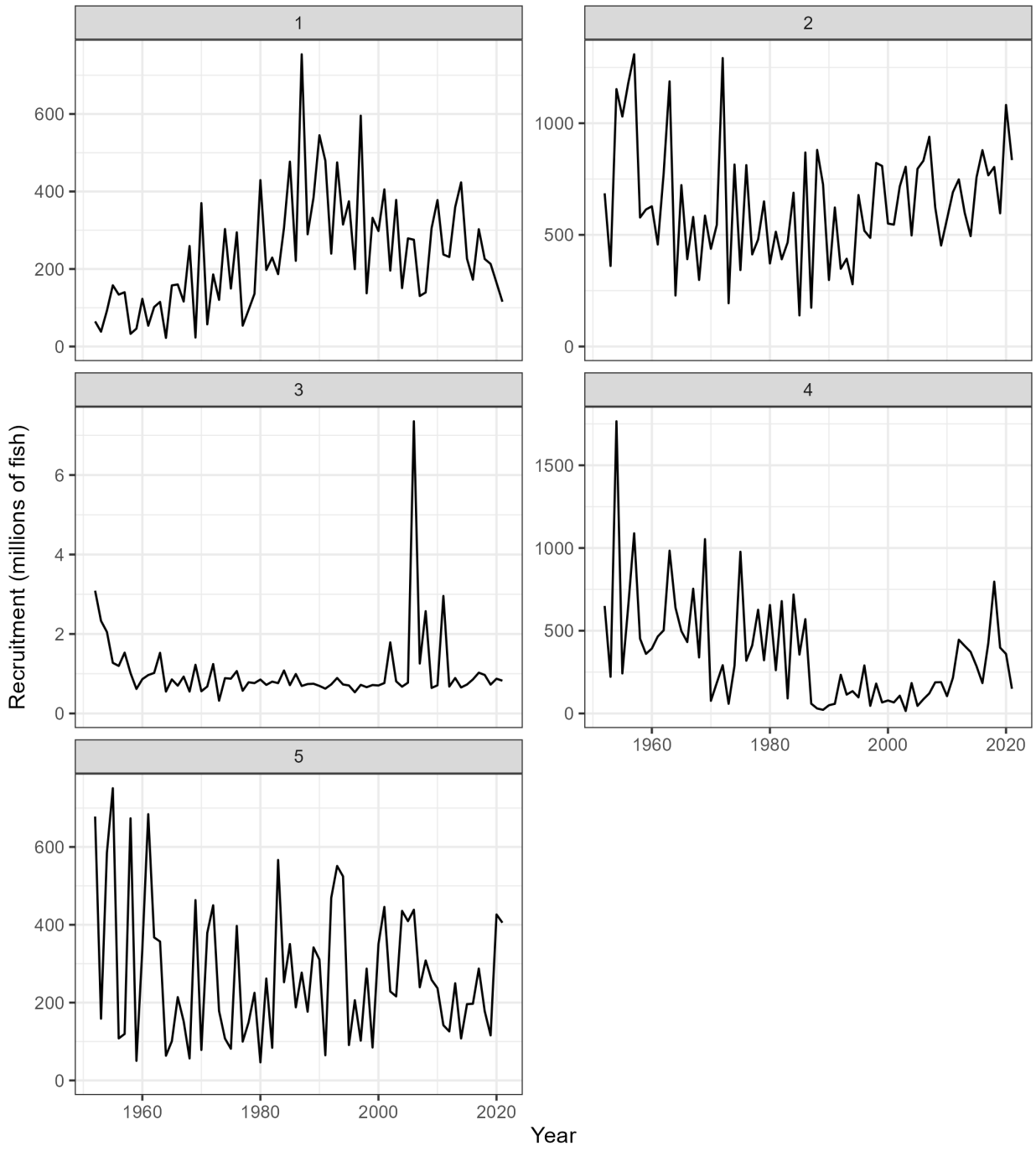

Figure 42: Time series of estimated annual recruitment by model region for the diagnostic model. Note that the scale of the y-axis is not constant across regions.

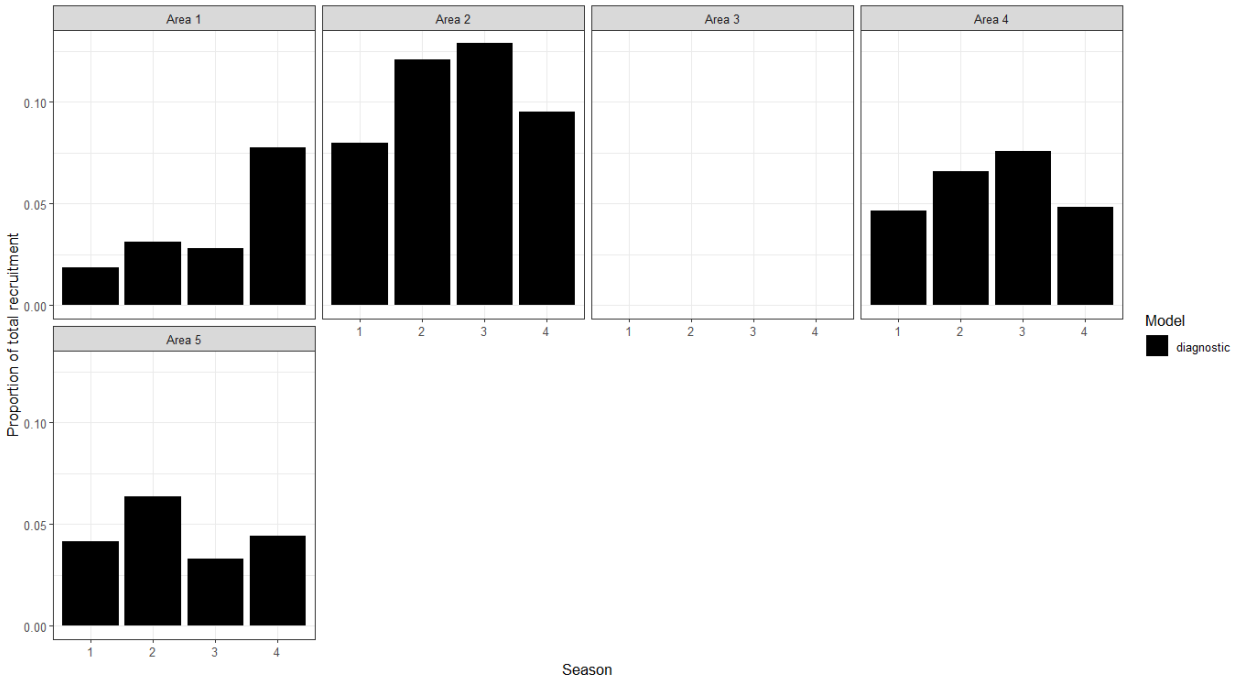

Figure 43: Estimated recruitment distribution by region and quarter.

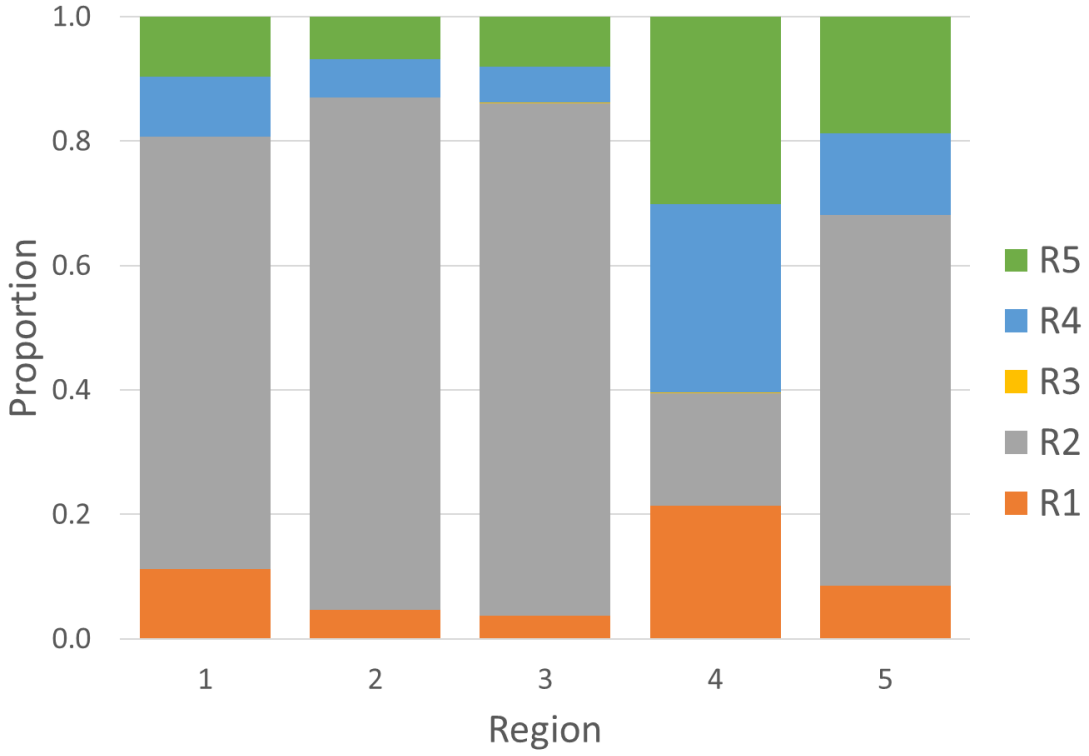

Figure 44: Proportional distribution of unfished total biomass in each region apportioned by the source region of the fish, for the diagnostic model. The colour of the originating region is presented in the legend. The biomass distributions are calculated based on the long-term average distribution of recruitment between regions, estimated movement parameters, and natural mortality.

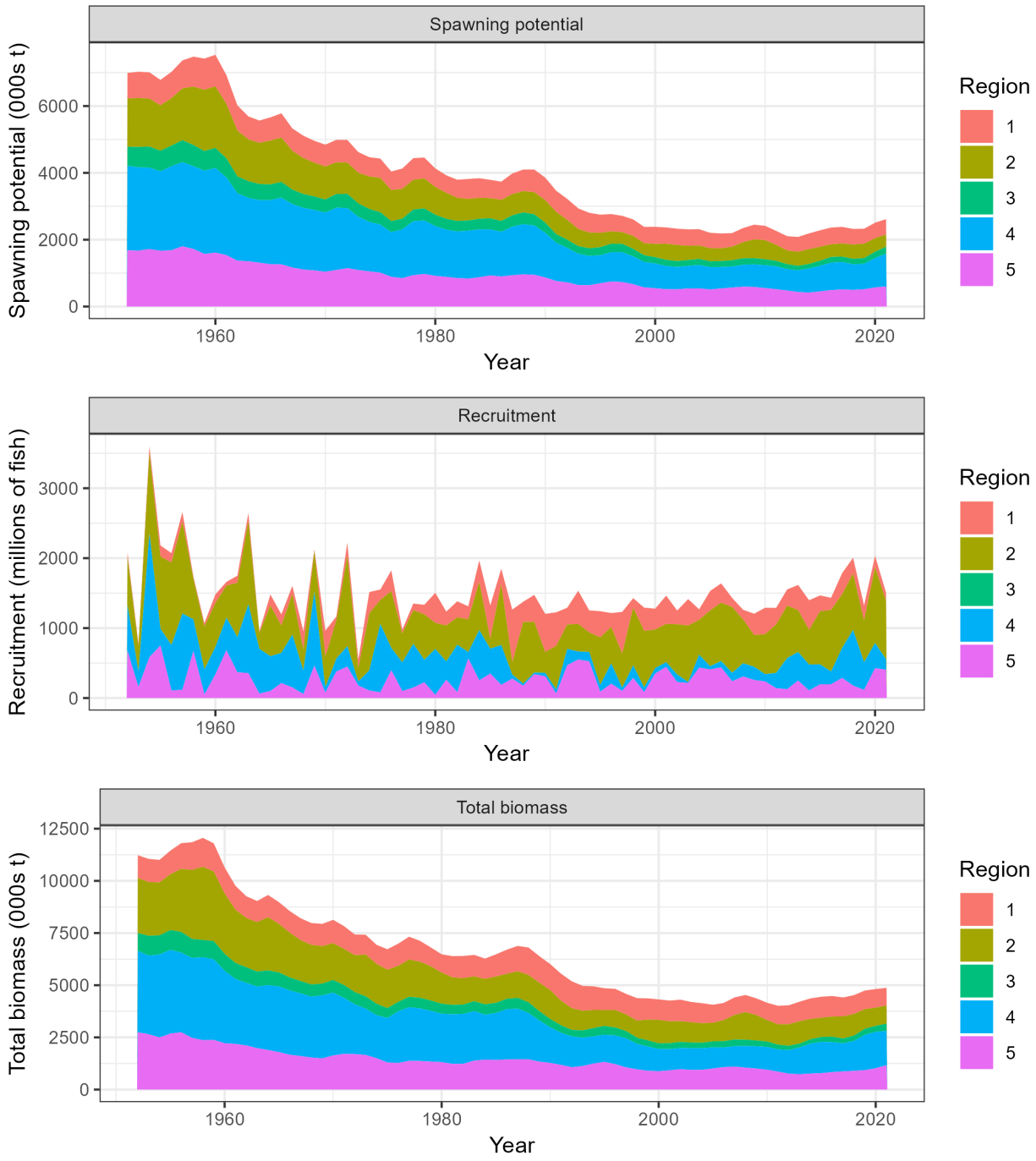

Figure 45: Time series of estimated annual spawning potential, recruitment and total biomass by model region for the diagnostic model, showing the relative proportions among regions. Note the data represent the averages of the quarterly model time steps for each year for spawning potential and total biomass and the sum of the quarterly recruitment estimates for annual recruitment.

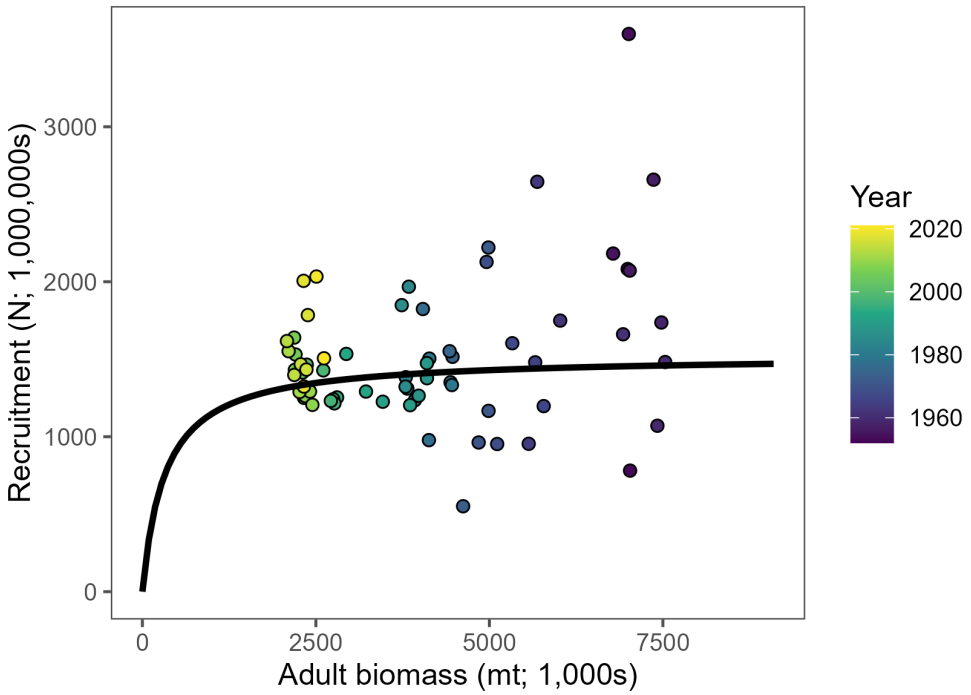

Figure 46: Estimated relationship between recruitment and spawning potential based on annual values for the diagnostic model. The darkness of the circles changes from light (more recent) to darker (earlier) through time.

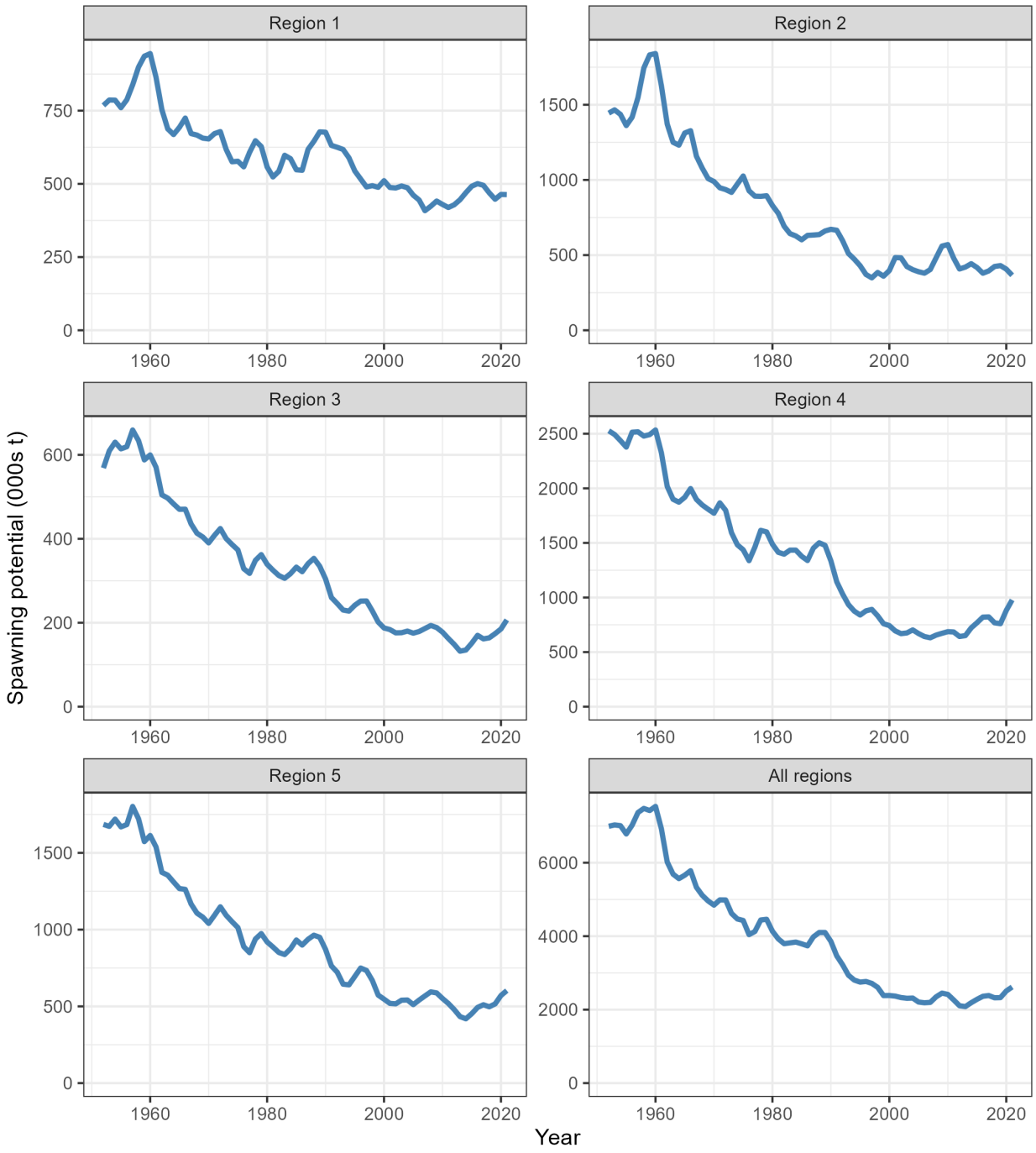

Figure 47: Estimated temporal spawning potential, SB_t, by model region, and for all model regions summed for the diagnostic model. Note that the scale of the y-axis is not constant across regions.

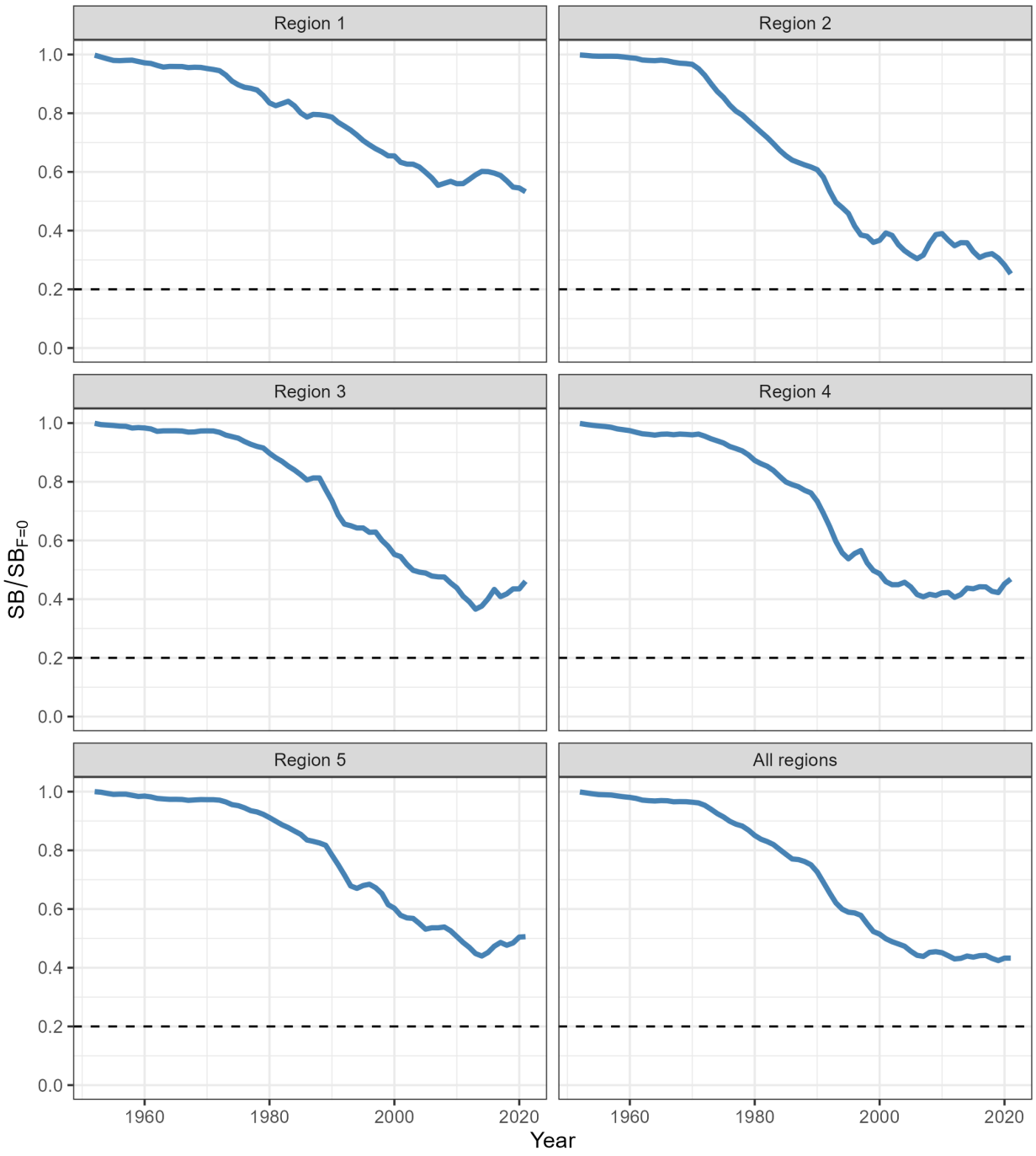

Figure 48: Estimated temporal spawning potential depletion, $SB_t/SB_{t,F=0}$, by model region, and for all model regions summed for the diagnostic model.

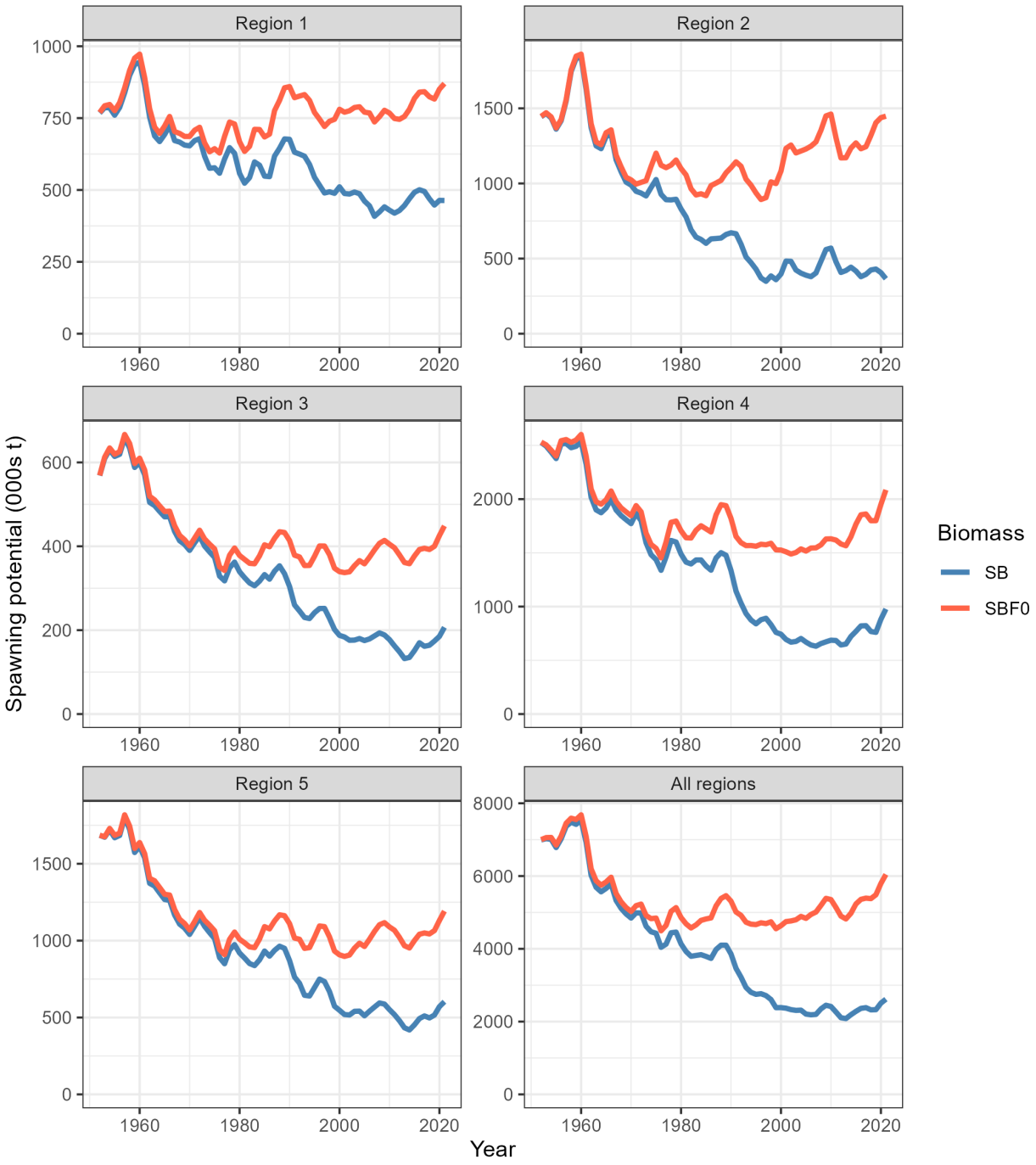

Figure 49: Comparison of the estimated annual spawning potential trajectories (lower blue lines) with the spawning potential trajectories predicted to have occurred in the absence of fishing (upper red lines) for each region and overall, for the diagnostic model. Note the scales of the Y-axes vary.

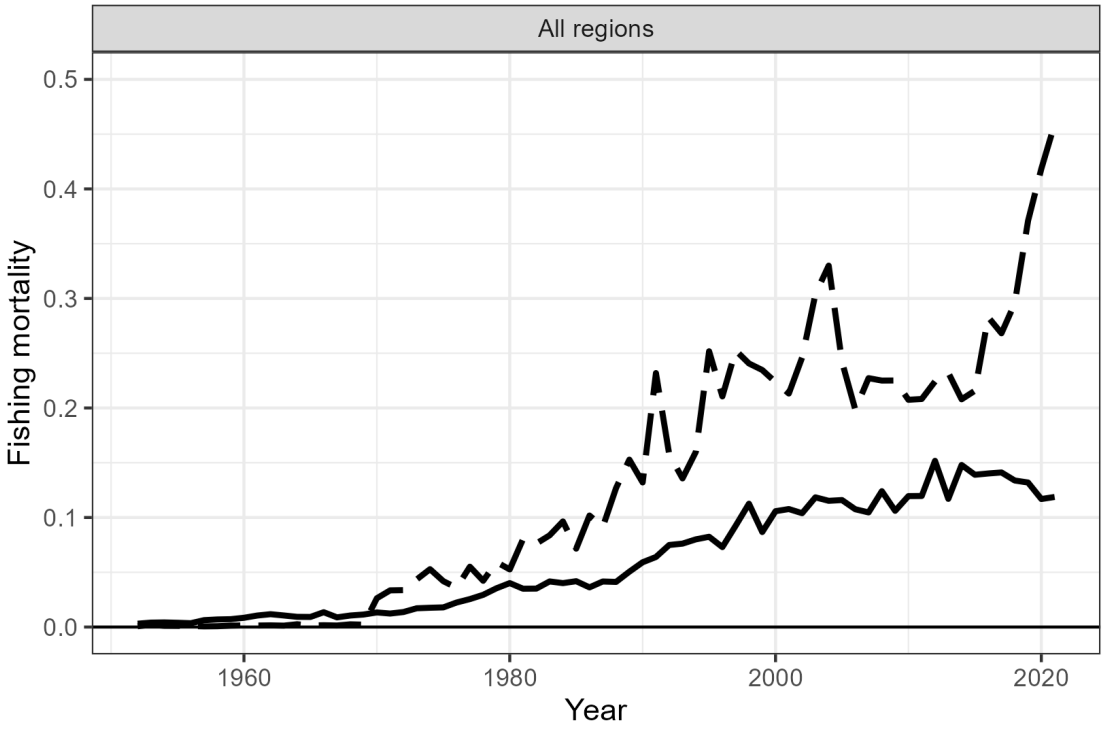

Figure 50: Estimated annual average adult (solid line) and juvenile (dashed line) fishing mortality for the diagnostic model.

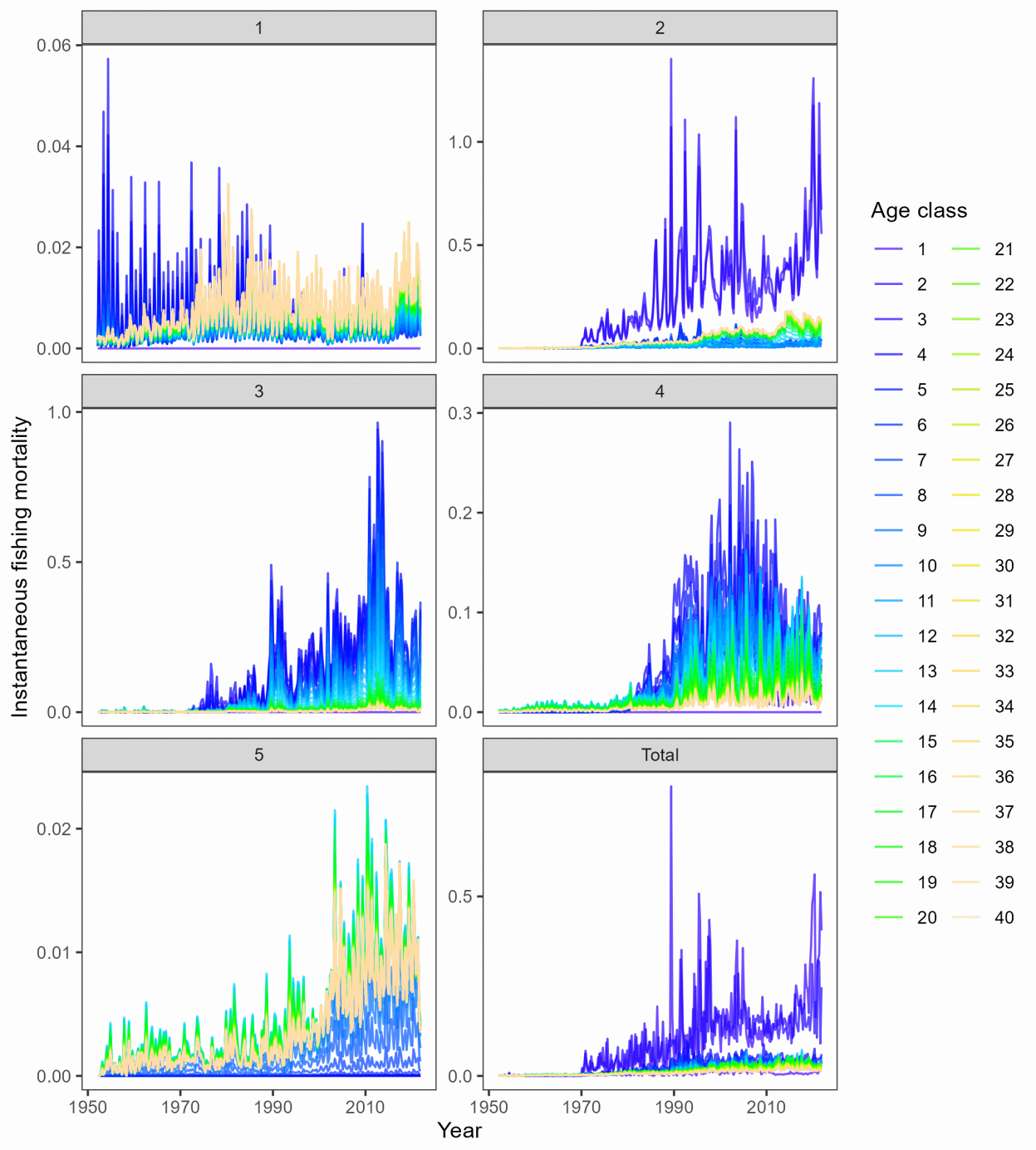

Figure 51: Estimated age-specific fishing mortality for the diagnostic model, by region and overall.

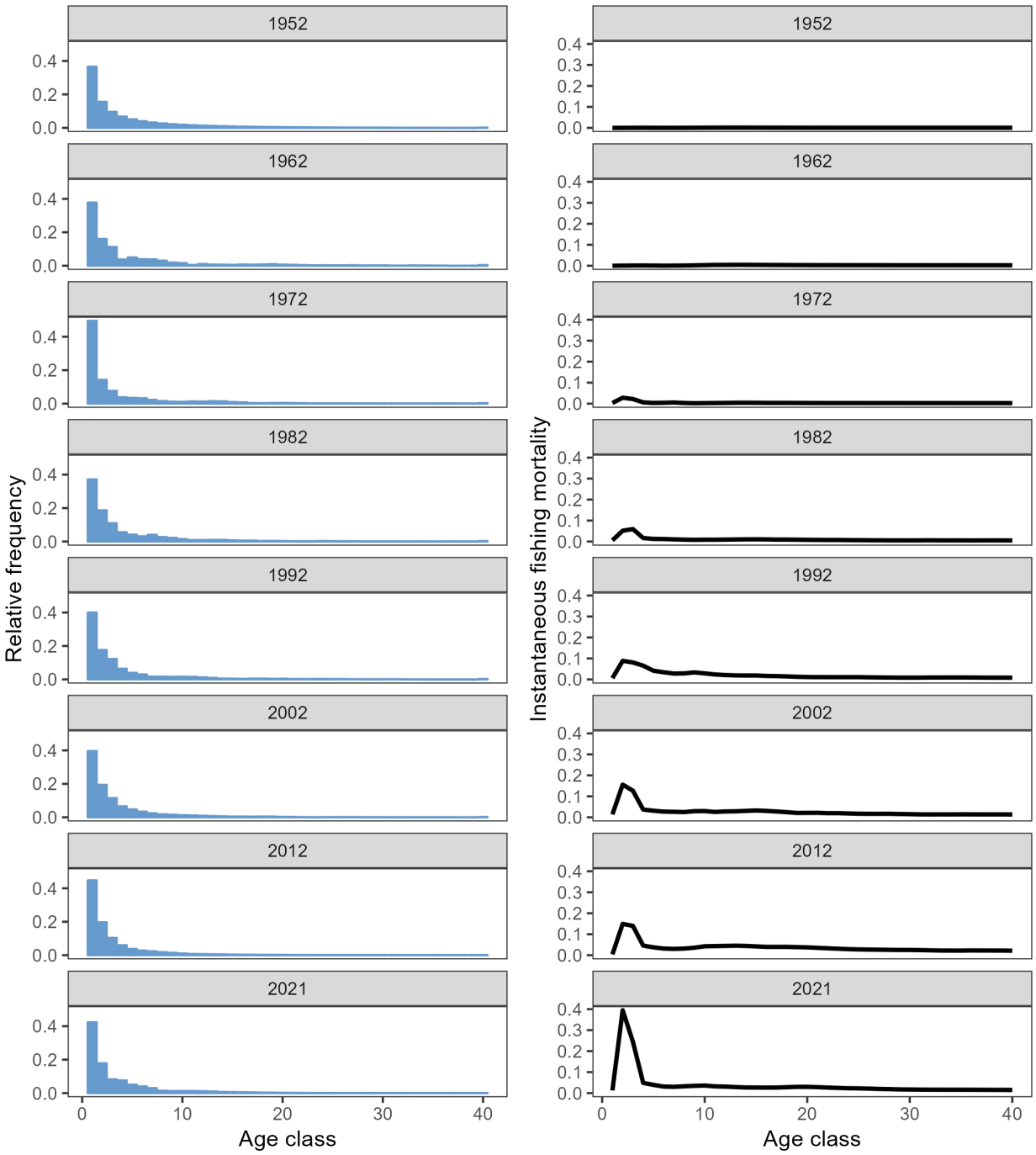

Figure 52: Estimated proportion at age (quarters) and fishing mortality at age (right), by year, at decade intervals, for the diagnostic model.

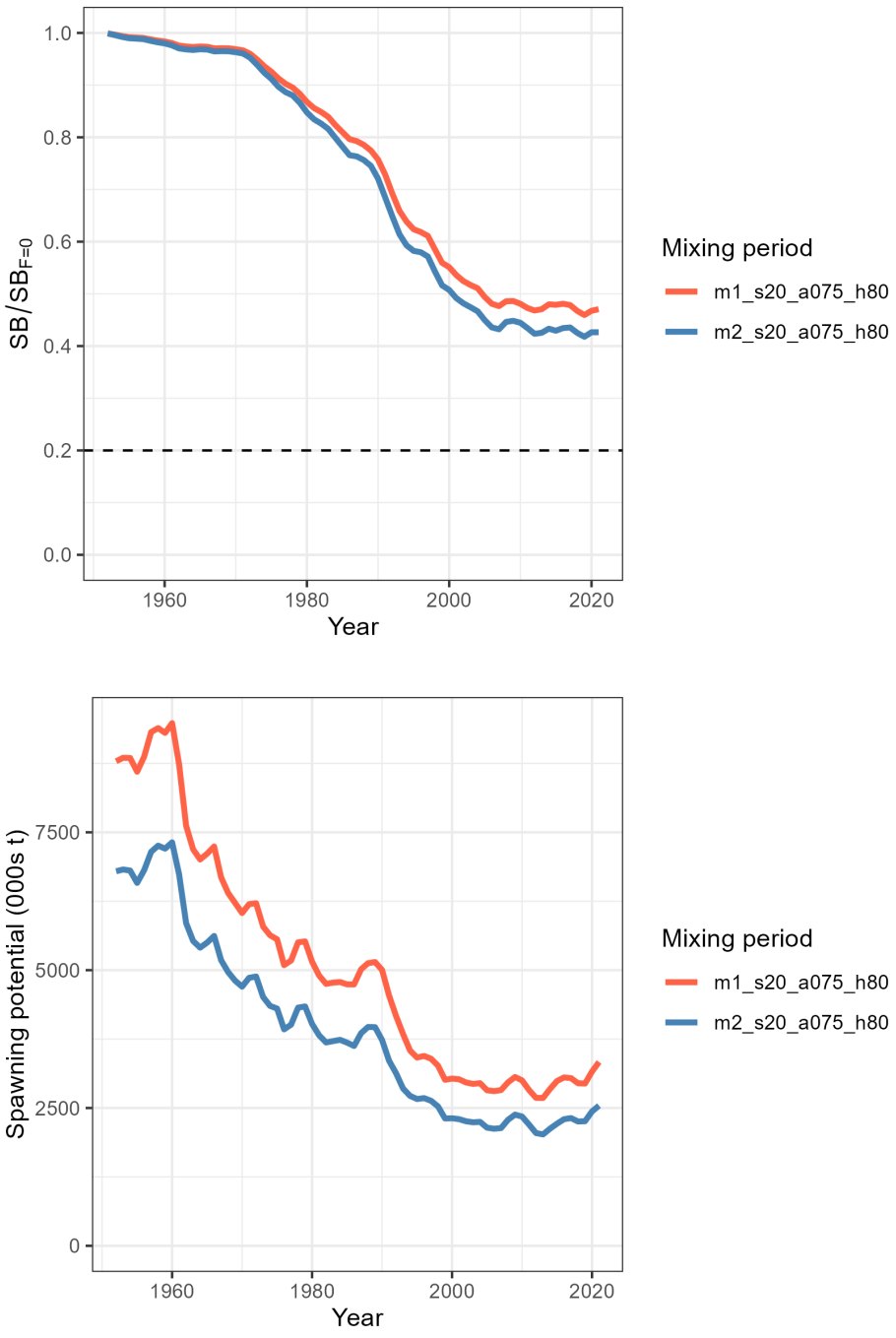

Figure 53: Estimated dynamic spawning depletion (top) and spawning potential (bottom) for the one-off sensitivities from the 2023 diagnostic model for tag mixing period. m=tag mixing, h=steepness, s=size data, a=age data.

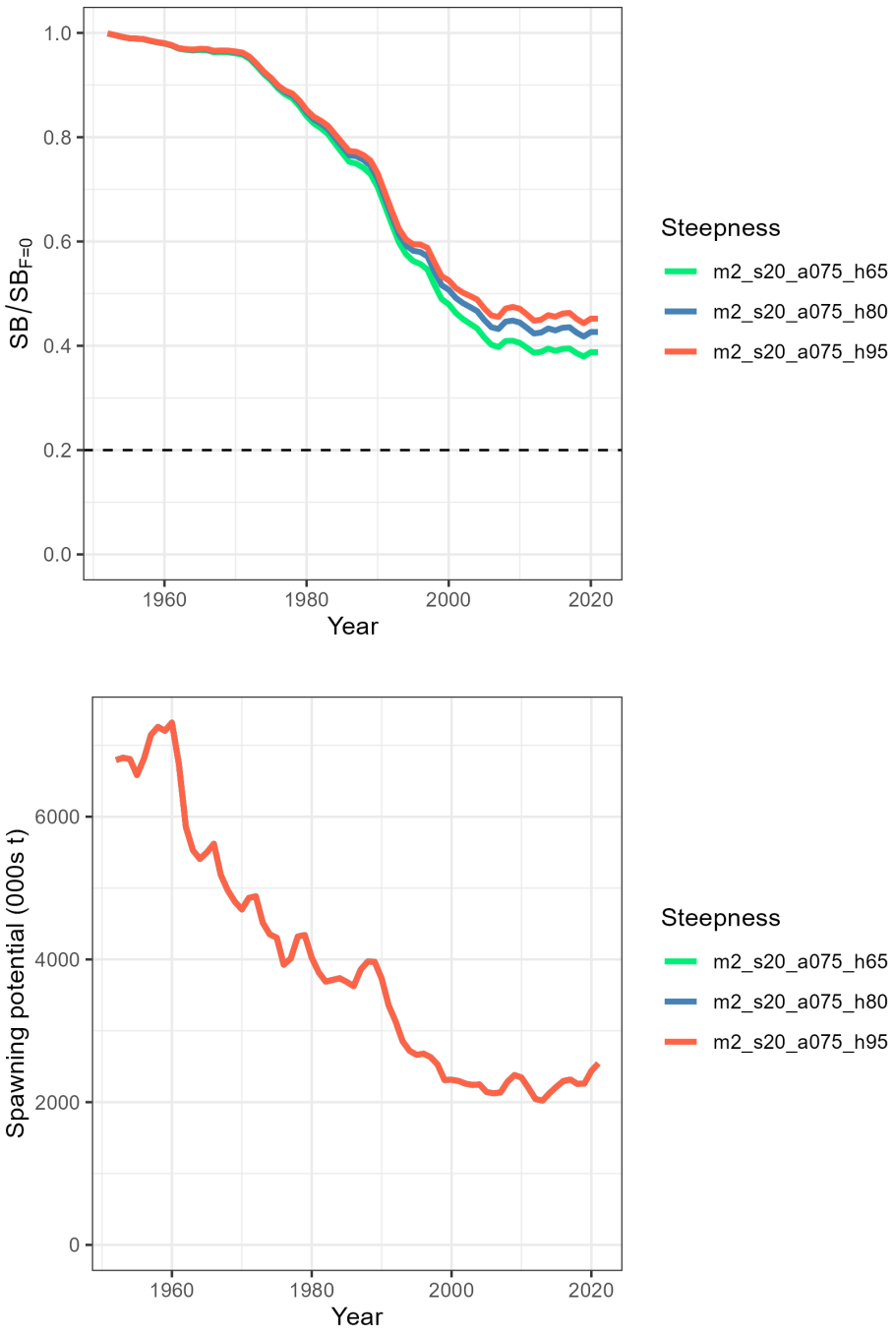

Figure 54: Estimated dynamic spawning depletion (top) and spawning potential (bottom) for the one-off sensitivities from the 2023 diagnostic model for steepness of the stock recruitment relationship. m=tag mixing, h=steepness, s=size data, a=age data.

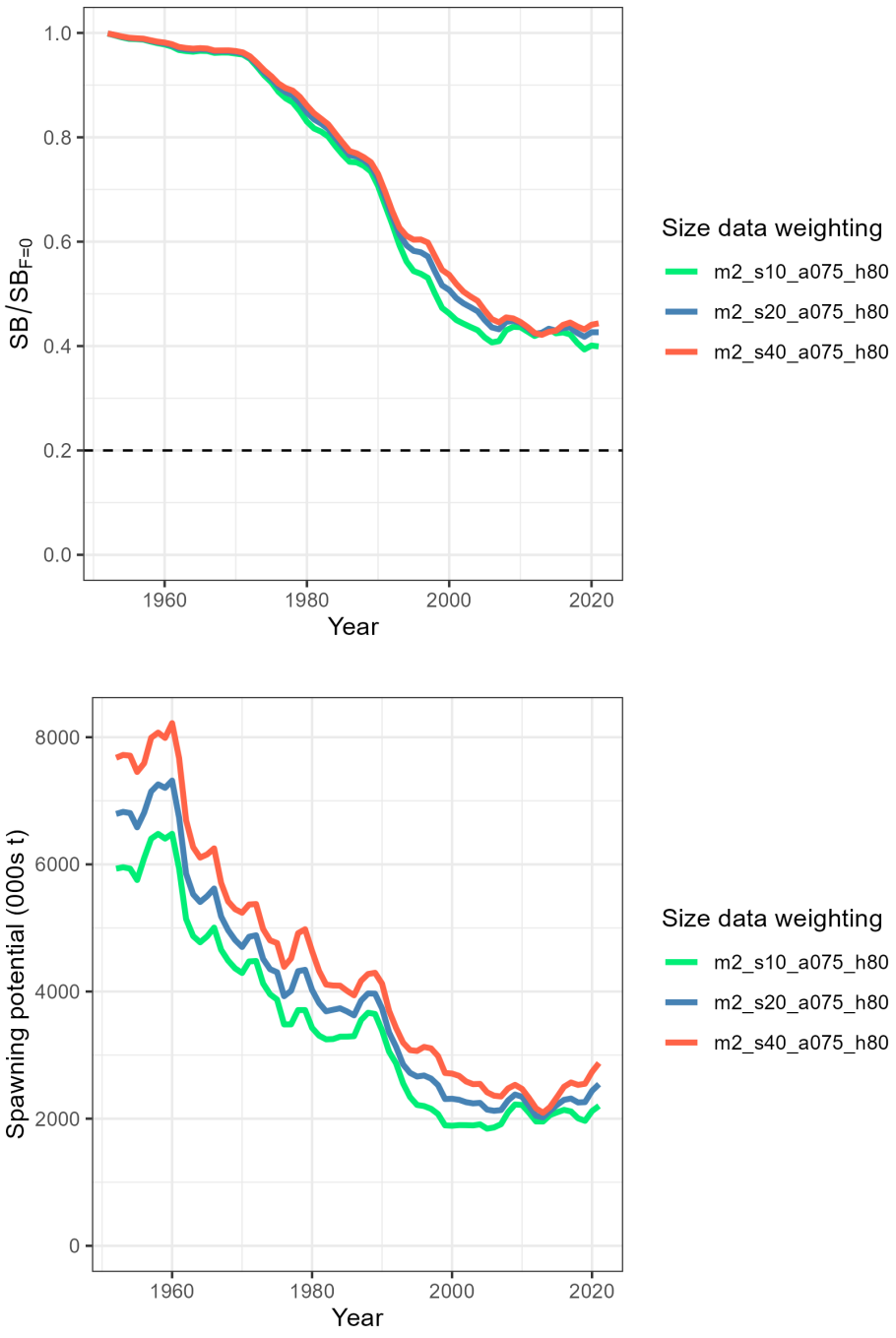

Figure 55: Estimated dynamic spawning depletion (top) and spawning potential (bottom) for the one-off sensitivities from the 2023 diagnostic model for size data weighting. m=tag mixing, h=steepness, s=size data, a=age data.

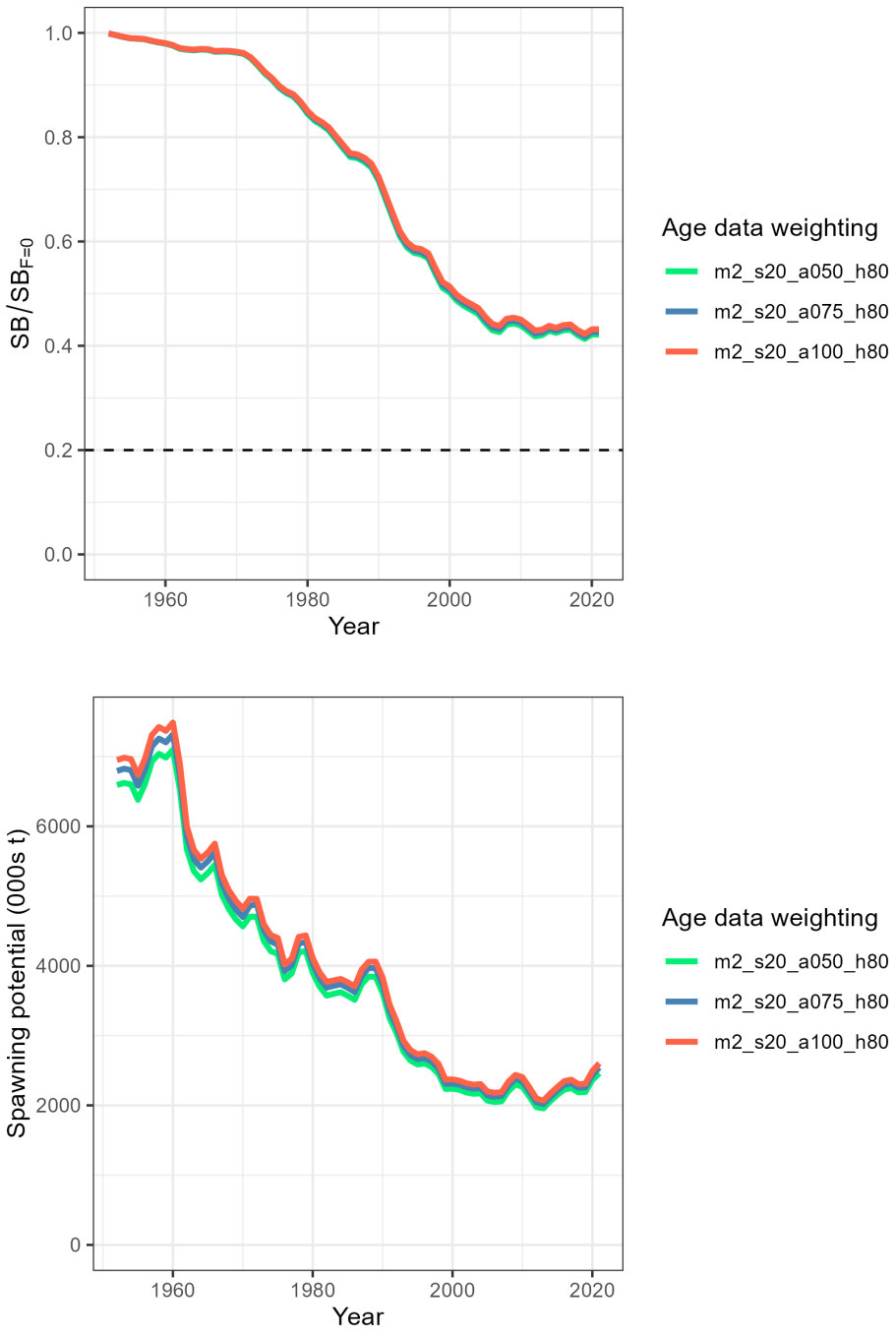

Figure 56: Estimated dynamic spawning depletion (top) and spawning potential (bottom) for the one-off sensitivities from the 2023 diagnostic model for age data weighting. m=tag mixing, h=steepness, s=size data, a=age data.

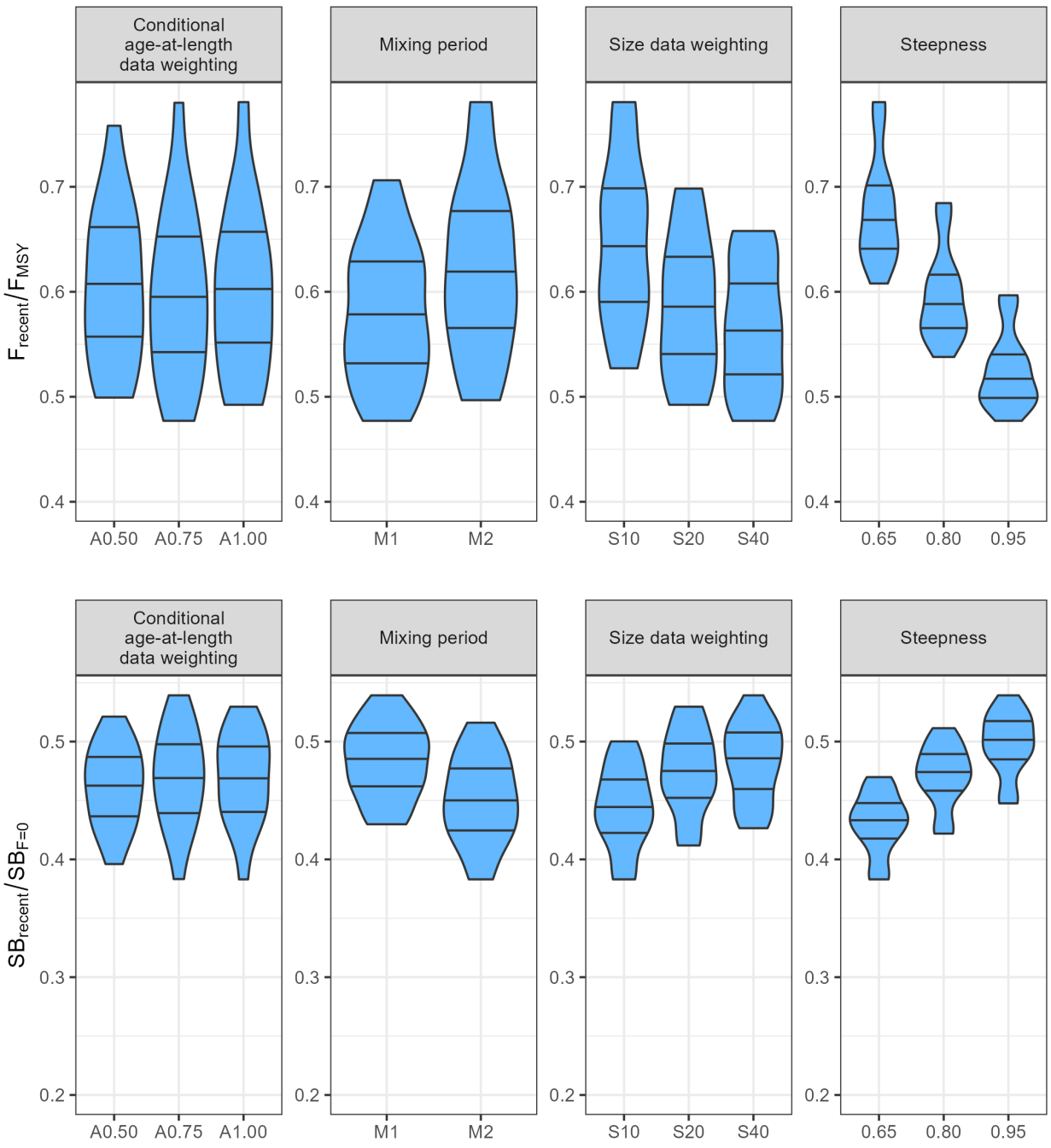

Figure 57: Box and violin plots summarizing the estimated $F_{\text{recent}}/F_{\text{MSY}}$ (top) and $SB_{\text{recent}}/SB_{F=0}$ (bottom) for each of the models in the structural uncertainty grid grouped by uncertainty axes (steepness, tag mixing, size data weighting, age data weighting). The horizontal lines are the 25th, 50th, and 75th percentiles. The shaded area shows the probability distribution (or density) of the estimates for all models in the structural uncertainty grid. These plots only include the structural uncertainty, not the estimation uncertainty.

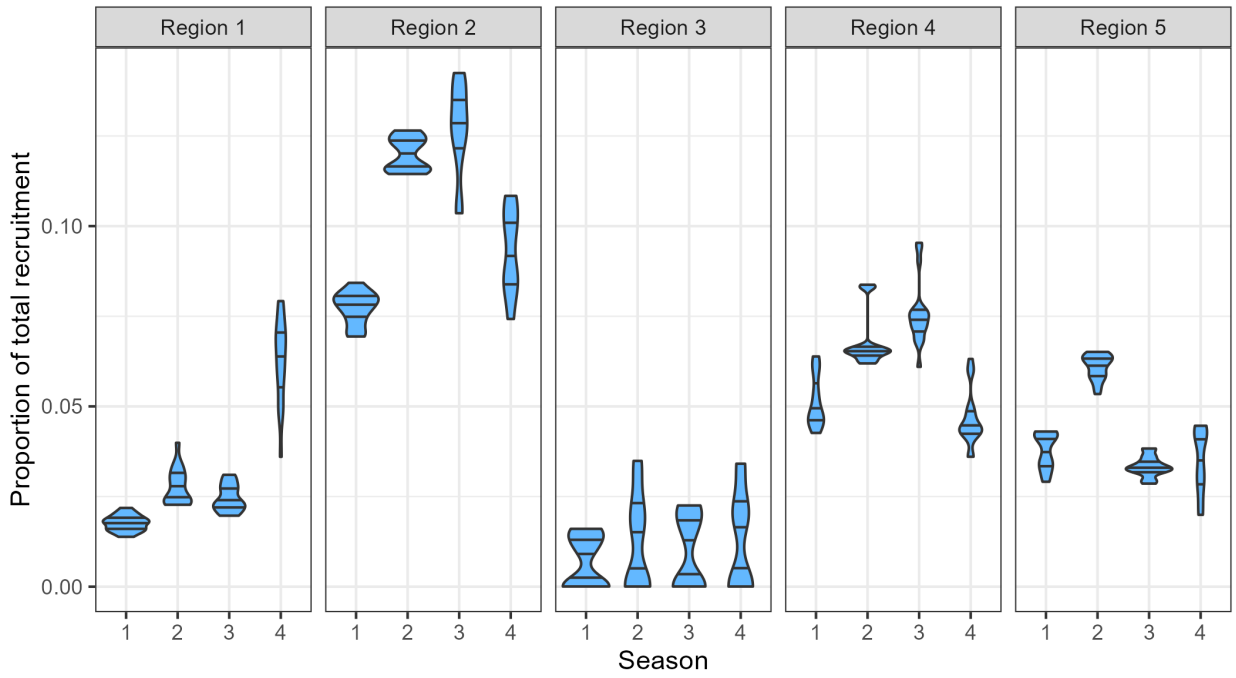

Figure 58: Box and violin plots showing the distribution of recruitment across model regions and quarters for all models in the uncertainty grid. The horizontal lines are the 25th, 50th, and 75th percentiles. The shaded area shows the probability distribution (or density) of the estimates of all models of the structural uncertainty grid. These plots only include the structural uncertainty, not the estimation uncertainty.

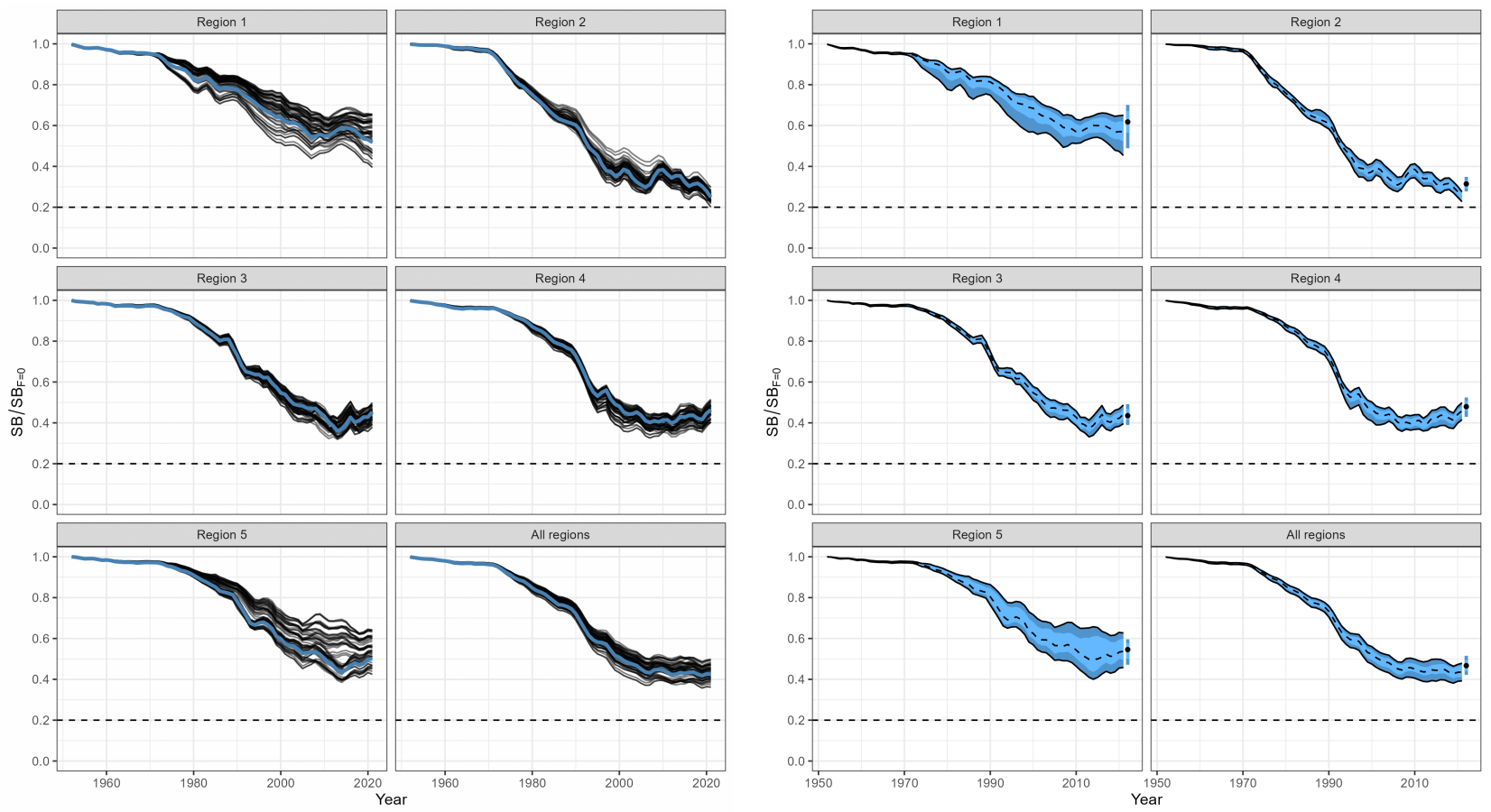

Figure 59: (Left) Trajectories of spawning potential depletion for the individual model runs included in the structural uncertainty grid over the period 1952-2021.(Right) Estimated spawning depletion across all models in the structural uncertainty grid over the period 1952-2021. The dashed line represents the median, the lighter band shows the 25th and 75th percentiles, and the dark band shows the 10th and 90th percentiles of the model estimates. The bar at the right of each ribbon indicates the median (black dots) with the 10th and 90th percentiles for $SB_{\text{recent}}/SB_{F=0}$.

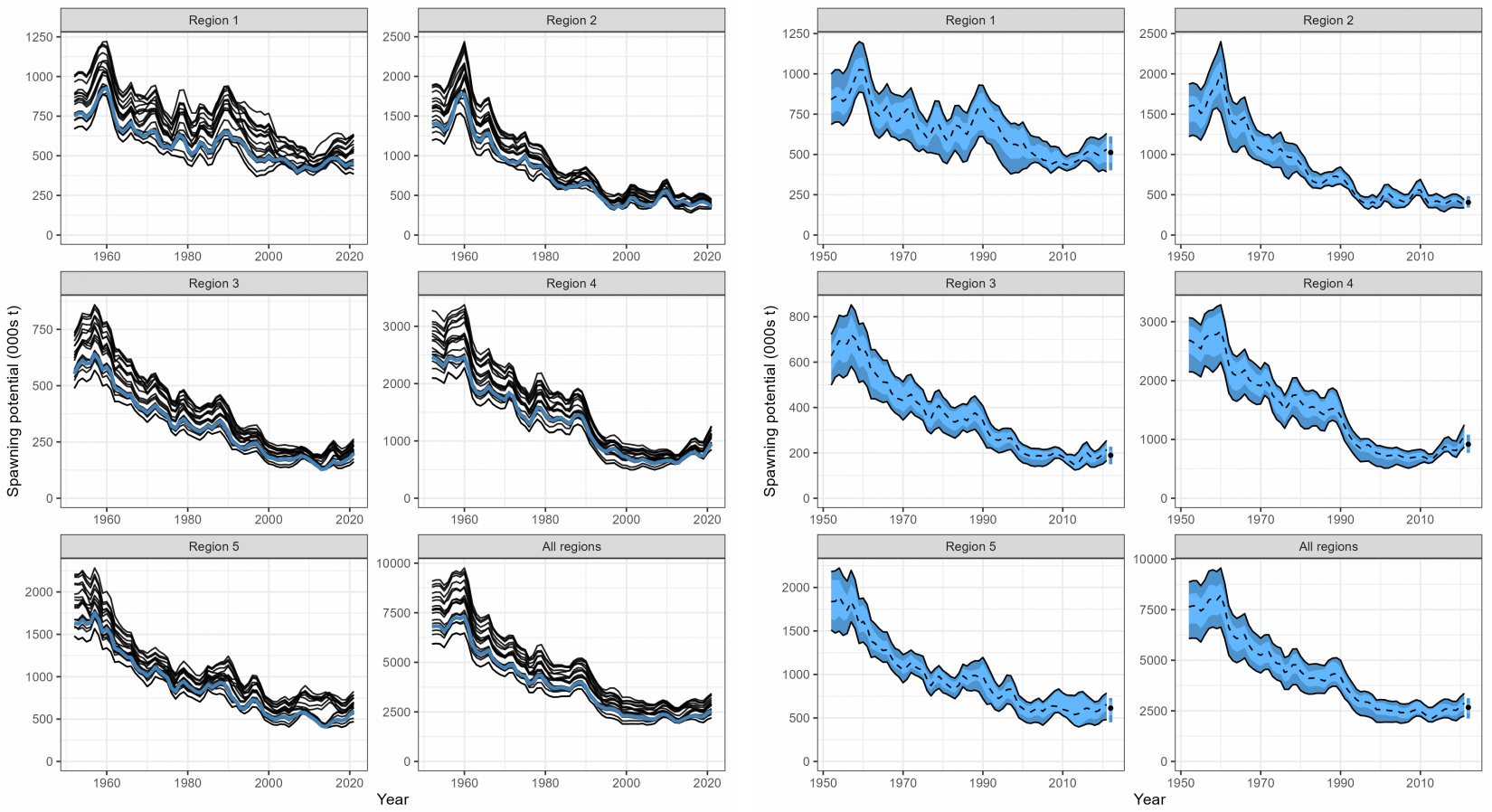

Figure 60: (Left) Trajectories of spawning potential for the individual model runs included in the structural uncertainty grid over the period 1952-2021. (Right) Estimated spawning potential across all models in the structural uncertainty grid over the period 1952-2021. The dashed line represents the median, the lighter band shows the 25th and 75th percentiles, and the dark band shows the 10th and 90th percentiles of the model estimates. The bar at the right of each ribbon indicates the median (black dots) with the 10th and 90th percentiles for SB_{recent}.

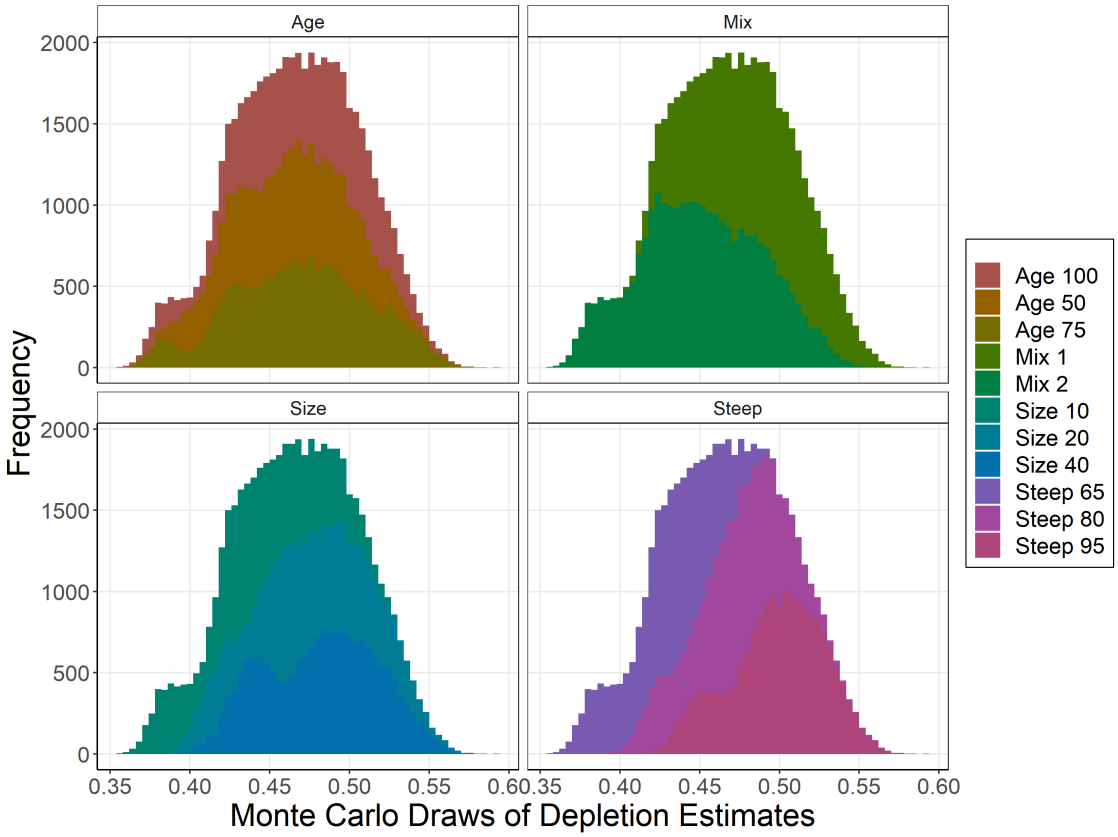

Figure 61: Distribution of $SB_{\text{recent}}/SB_{F=0}$ integrating model and estimation uncertainty, presented by uncertainty axis (panels) and axis element values (colours).

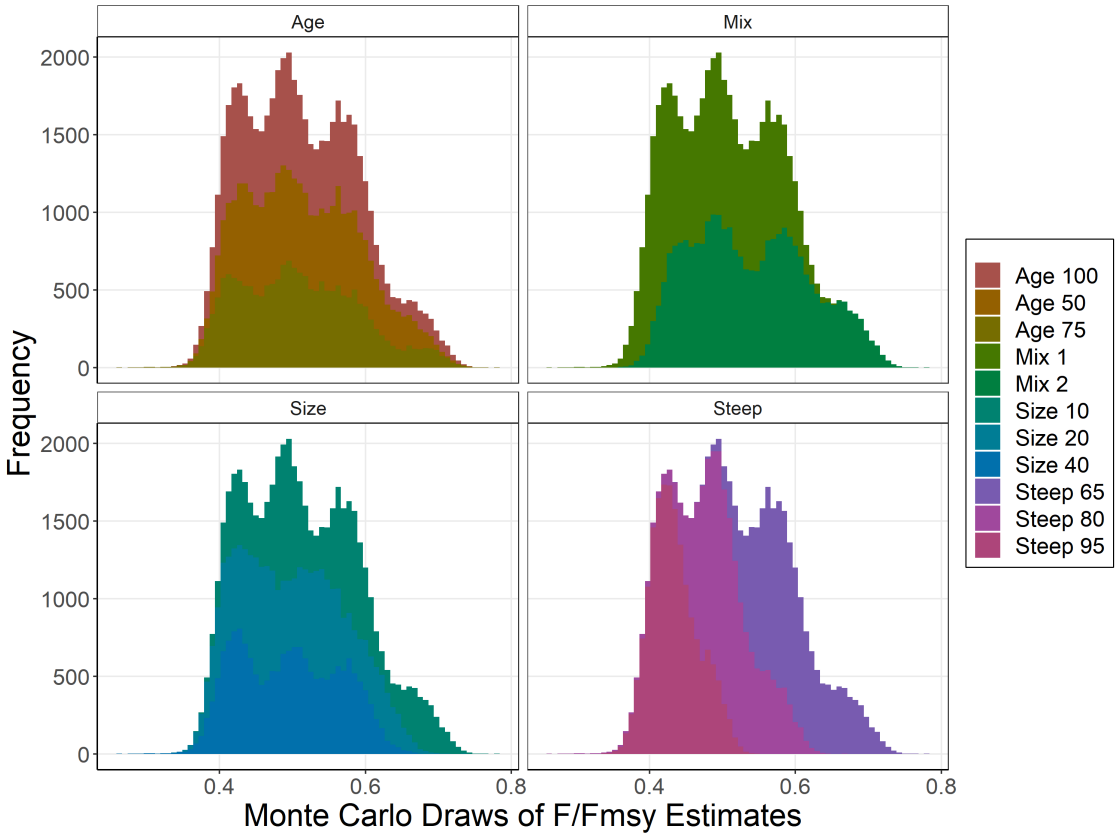

Figure 62: Distribution of $F_{\text{recent}}/F_{\text{MSY}}$ integrating model and estimation uncertainty, presented by uncertainty axis (panels) and axis element values (colours).

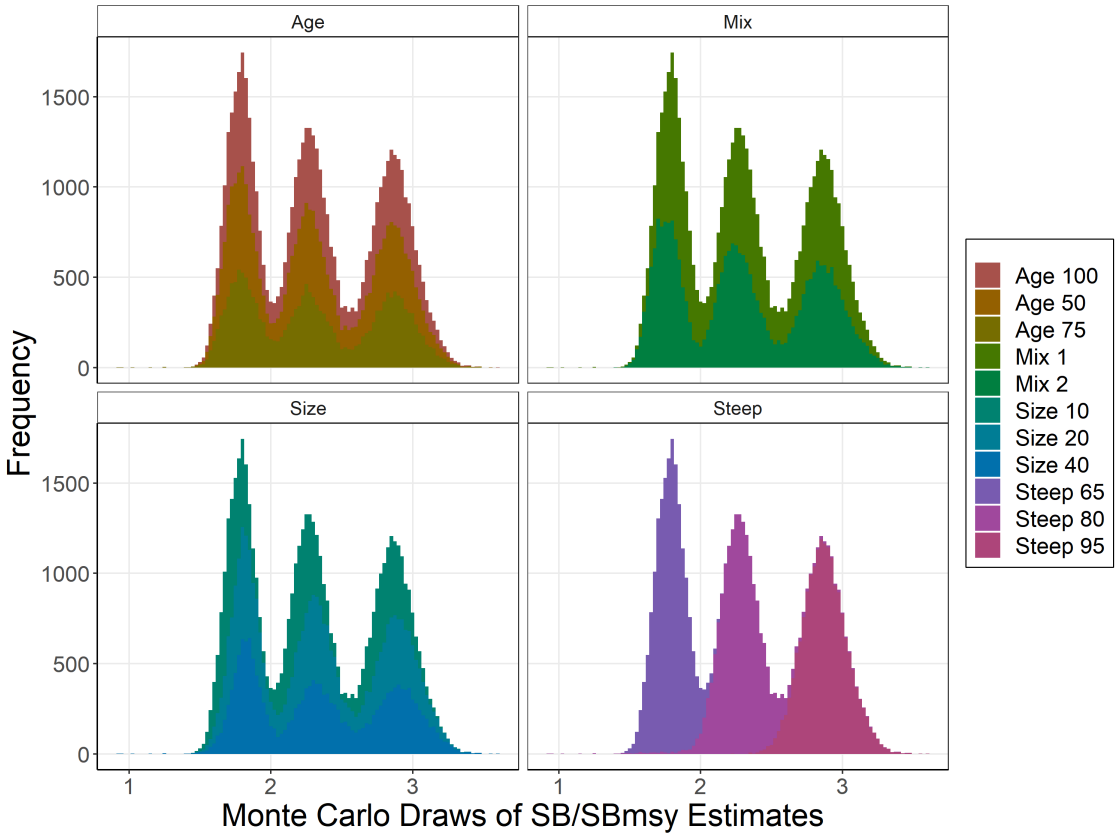

Figure 63: Distribution of $SB_{\text{recent}}/SB_{\text{MSY}}$ integrating model and estimation uncertainty, presented by uncertainty axis (panels) and axis element values (colours).

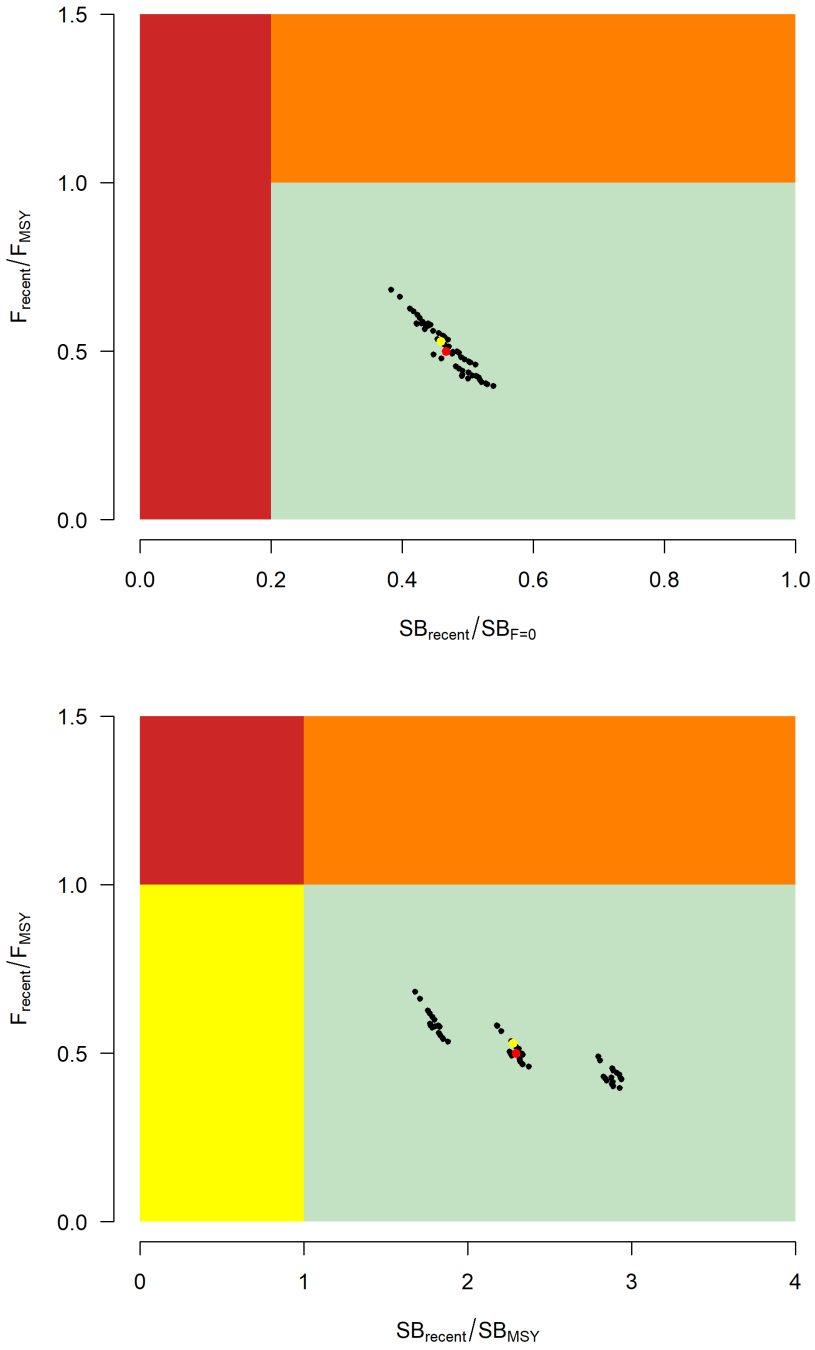

Figure 64: Majuro plot (top) and Kobe plot (bottom) summarising the results for each of the models in the structural uncertainty grid for the recent period (2018-2021). The yellow point is the 2023 diagnostic model and the red point is the median.

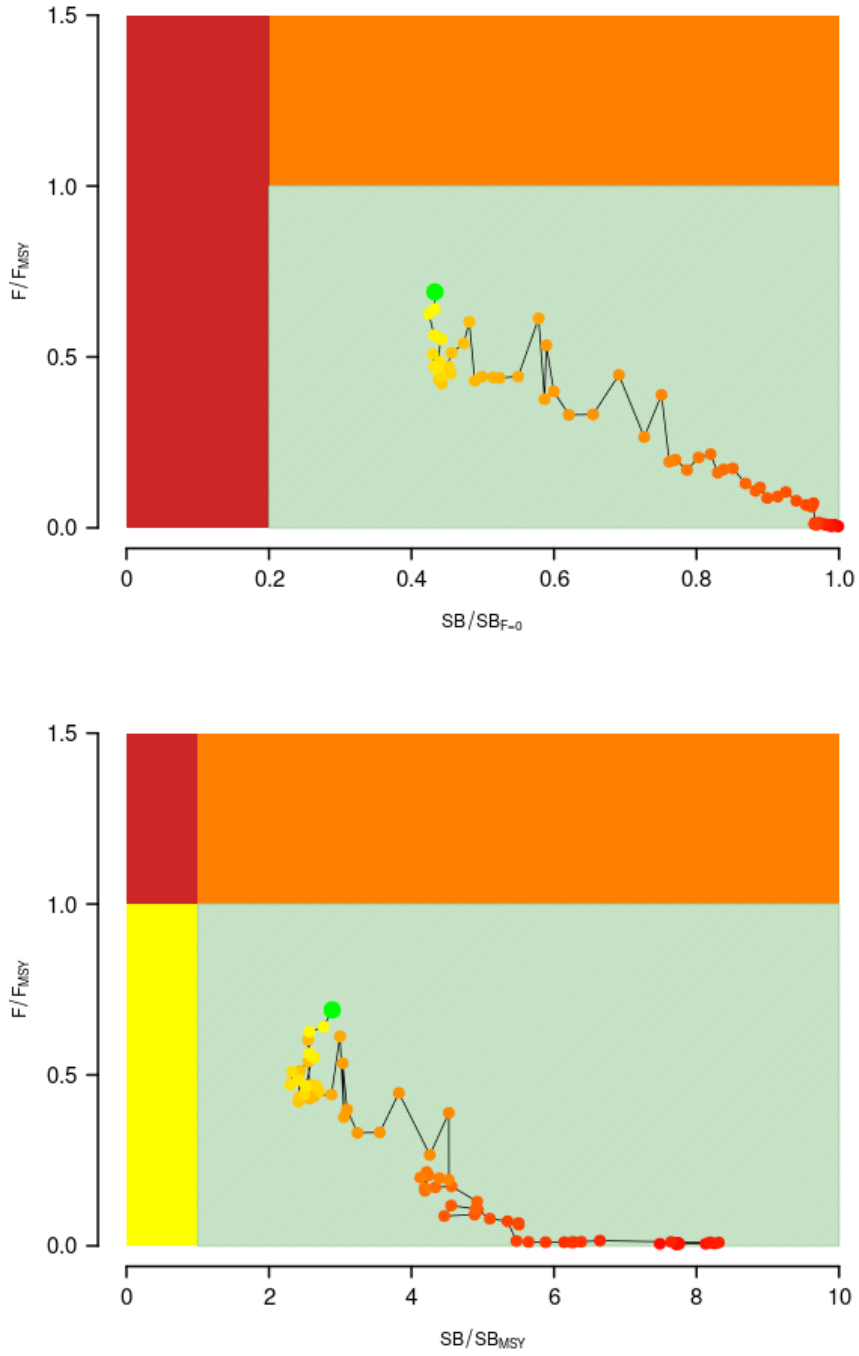

Figure 65: Time dynamic Majuro (top) and Kobe (bottom) plots summarising the results for the diagnostic case model over the model period. The green point is the estimated 2021 status, the redder the point the further back in time.

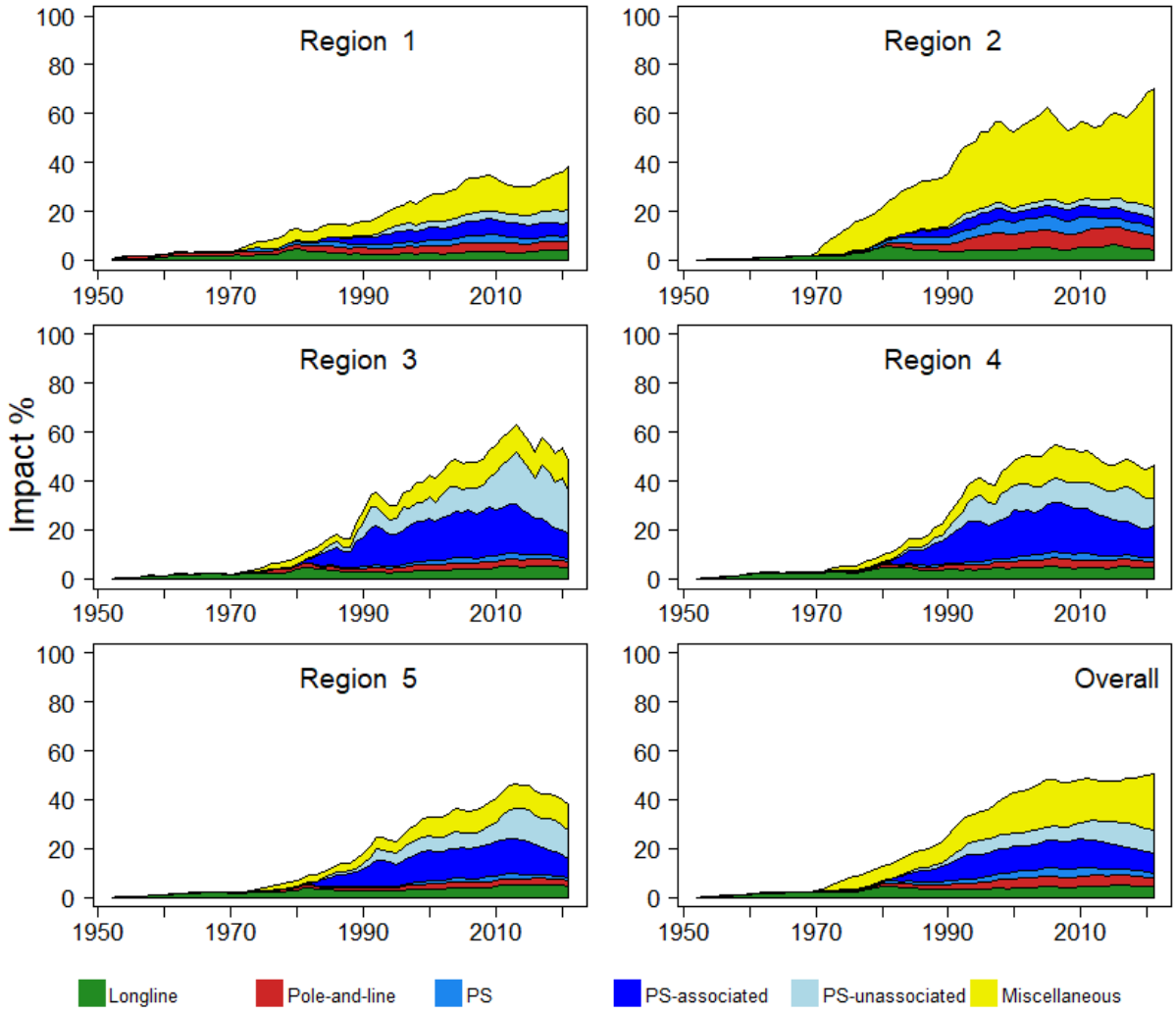

Figure 66: Estimates of reduction in spawning potential due to fishing (Fishery Impact = $1-SB_t/SB_{t,F=0}$) by region, and over all regions (lower right panel), attributed to various fishery groups for the diagnostic model.

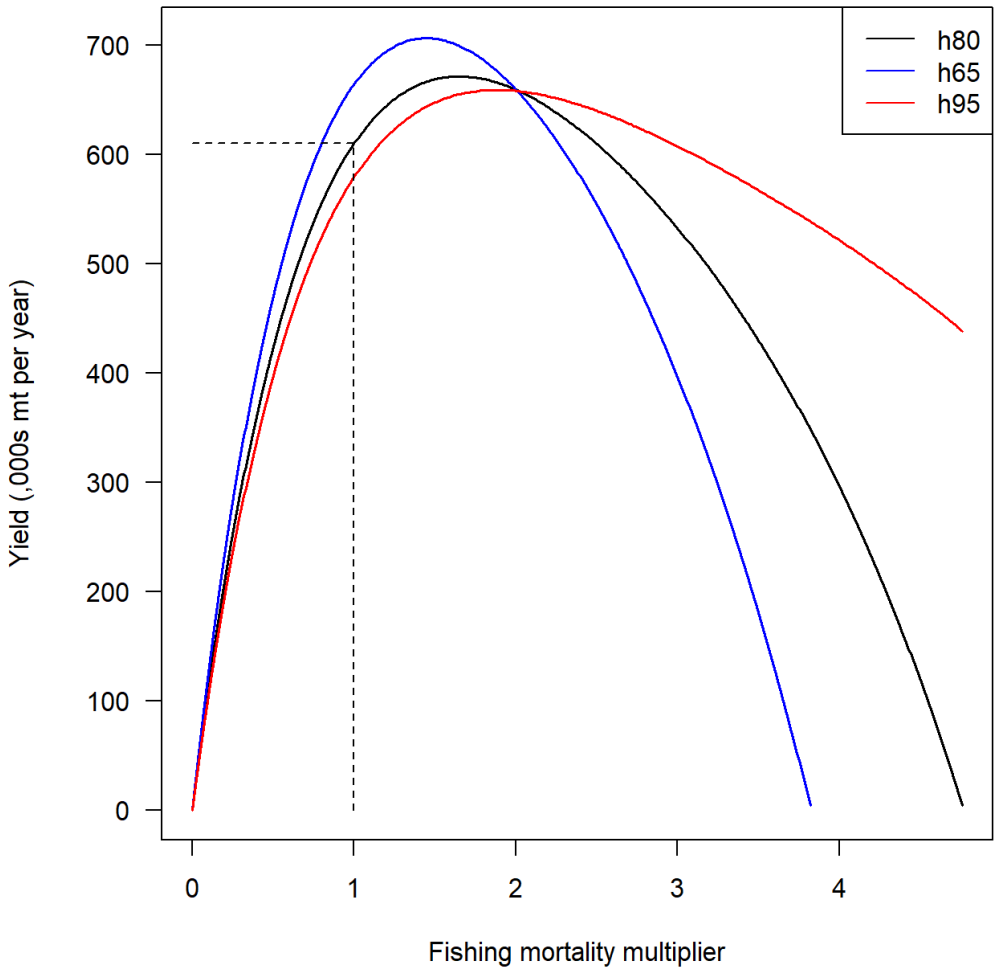

Figure 67: Estimated yield as a function of fishing mortality multiplier for the diagnostic model and a few of the one-off sensitivity models. The red dashed line indicates the equilibrium yield at current fishing mortality.

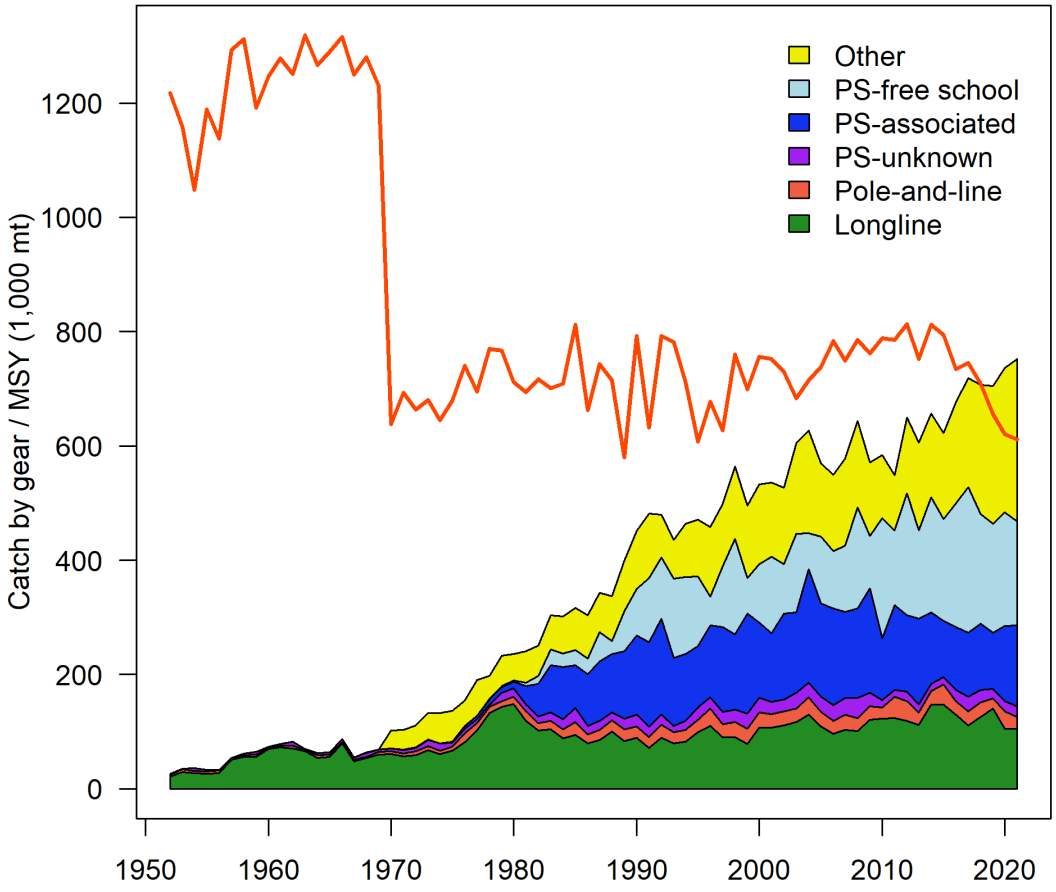

Figure 68: History of the annual estimates of MSY (red line) for the diagnostic model compared with annual catch by the main gear types. Note that this is a 'dynamic' MSY which is explained further in subsubsection 7.5.4.

12 Appendices

12.1 Likelihood profiles

The approach for calculating a likelihood profile of the total population scaling parameter is outlined in subsection 5.6. The profile reflects the loss of fit over all the data, i.e., the overall objective function value, caused by changing the population scale from that of the maximum likelihood estimated value. A range of fixed values were used until the best fit for each data source was found. The likelihood profile for the diagnostic model is shown in Figure 12.1 for narrow and wide widows around the maximum likelihood estimated value.

Likelihood profiles on the total population scaling parameter are also included for individual fisheries length and weight compositions to explore sources of conflict in these data, Figure 12.2 and Figure 12.3.

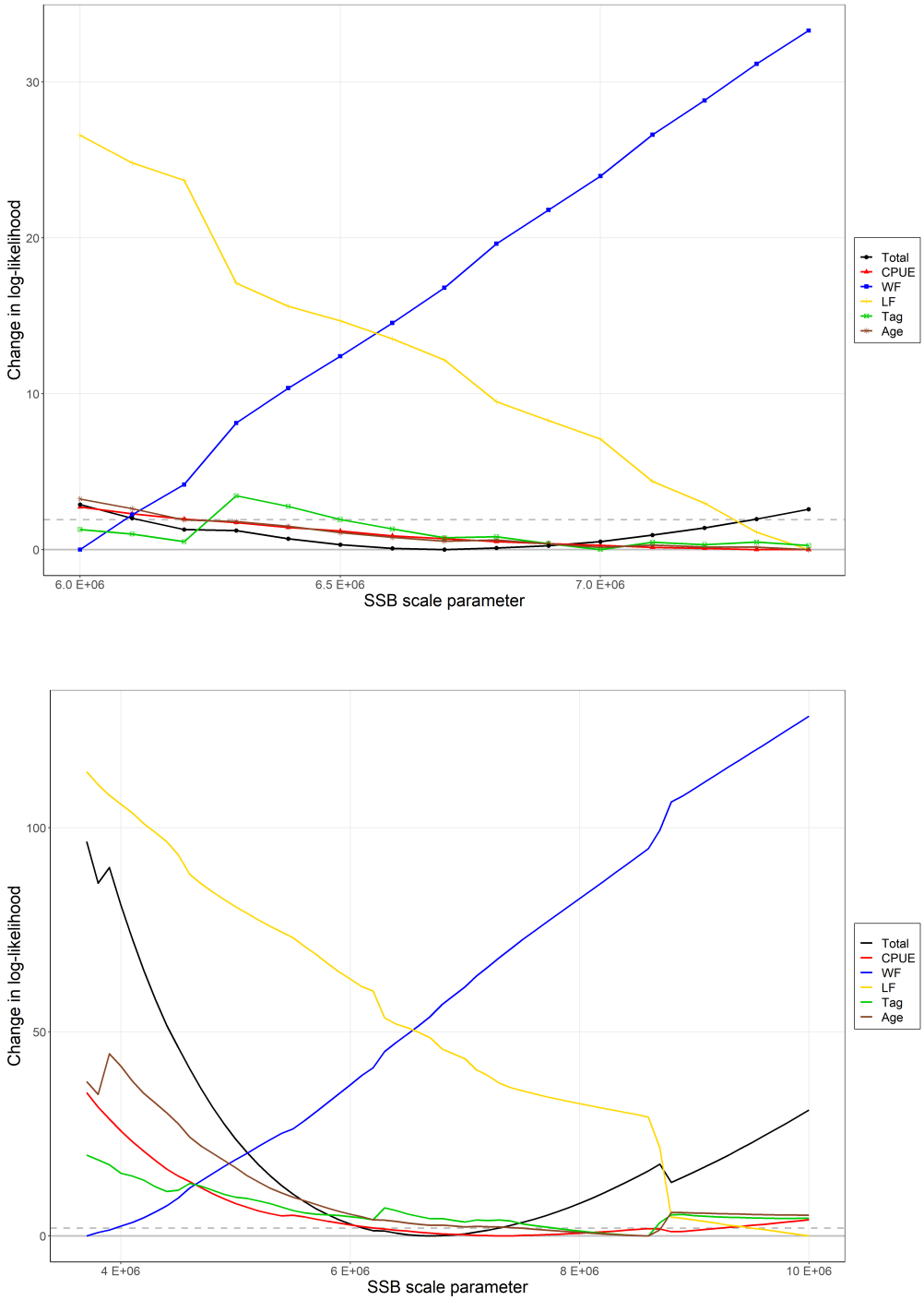

Figure 12.1: Profiles of the total log-likelihood with respect to average total biomass in million mt across a range of fixed values for the model data sources, the black line indicates the total likelihood. The top plot shows a narrow window around the maximum likelihood estimated value and the bottom plot shows a wider window.

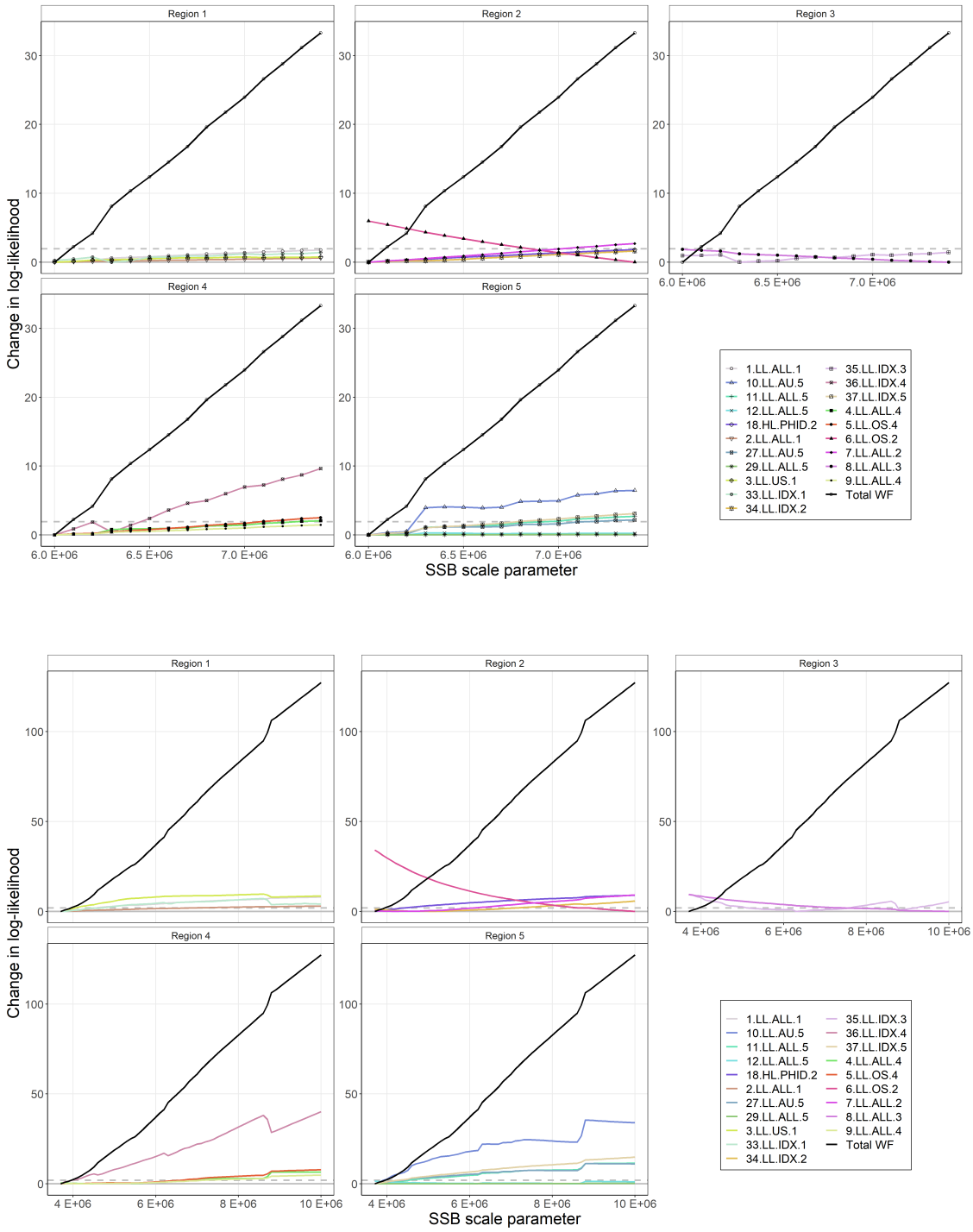

Figure 12.2: Profiles of the total log-likelihood with respect to average total biomass in million mt across a range of fixed values for the weight composition data by fishery and regions, the black line indicates the total weight composition likelihood. The top plot shows a narrow window around the total maximum likelihood estimated value and the bottom plot shows a wider window.

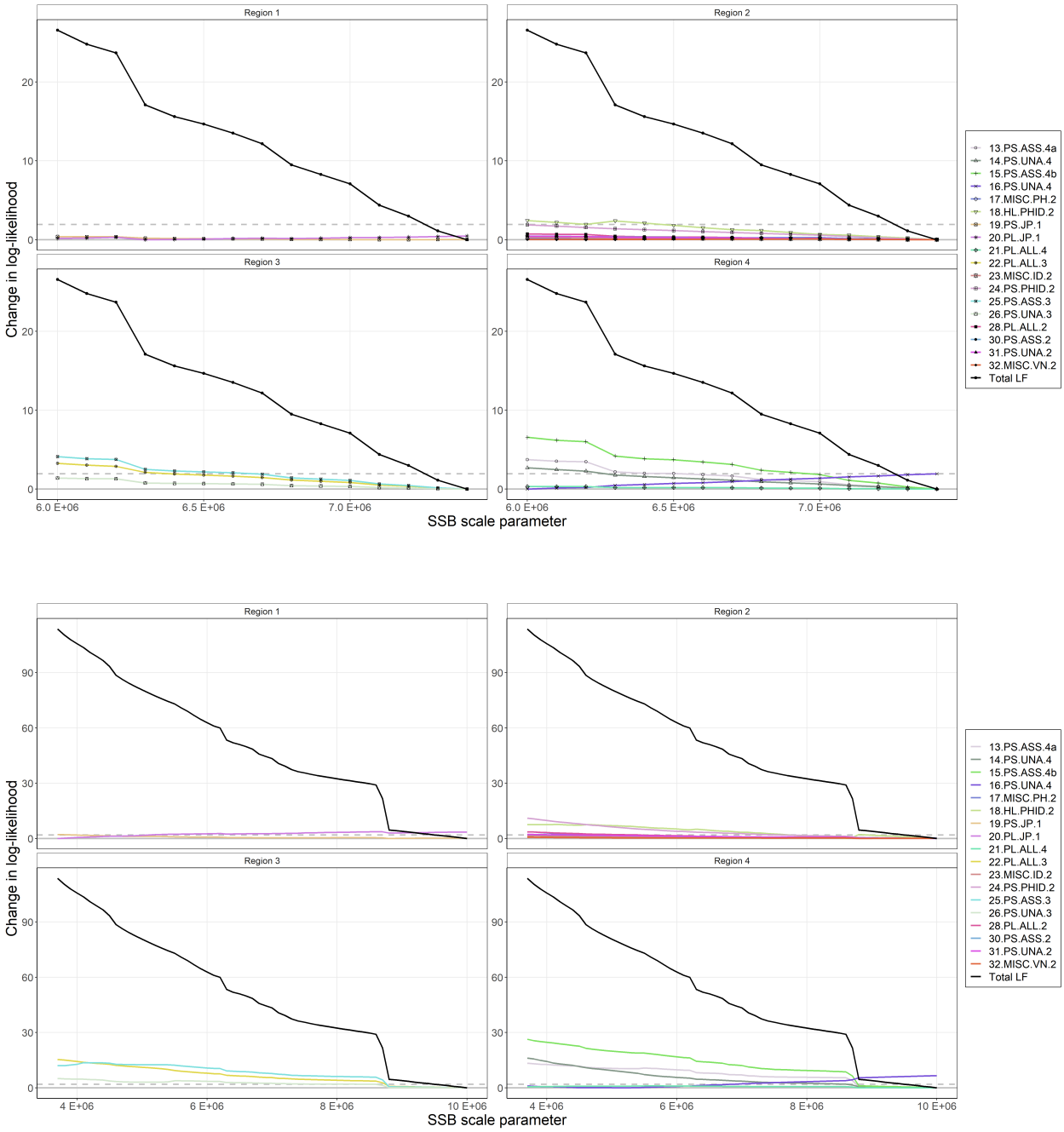

Figure 12.3: Profiles of the total log-likelihood with respect to average total biomass in million mt across a range of fixed values for the length composition data by fishery and regions, the black line indicates the total length composition likelihood. The top plot shows a narrow window around the total maximum likelihood estimated value and the bottom plot shows a wider window.

12.2 Retrospective analyses

Retrospective analysis involves rerunning the 2023 diagnostic model by consecutively removing successive years of data to estimate model bias. A series of 7 retrospective models were fitted starting with the full dataset (through 2021), followed by models with the retrospective removal of all input data for each year peel. Spawning potential and spawning depletion trajectories are shown in Figure 12.4. Mohns rho is calculated for the retrospective models of spawning depletion.

Spawning potential and spawning depletion trajectories for each of these retrospective peels are shown in Figure 12.4. Each peel produces estimates of spawning potential and spawning depletion with very similar dynamics to the diagnostic model. The value of Mohn's rho is 0.084 for the spawning depletion retrospectives, and -0.101 for spawning potential (Figure 12.4). As a general rule of thumb, values of Mohn's rho higher than 0.20 or lower than -0.15 are cause for concern in an assessment (Hurtado-Ferro et al., 2015). The values obtained for the 2023 yellowfin diagnostic model indicate that there is no strong concern for retrospective bias with the 2023 diagnostic model.

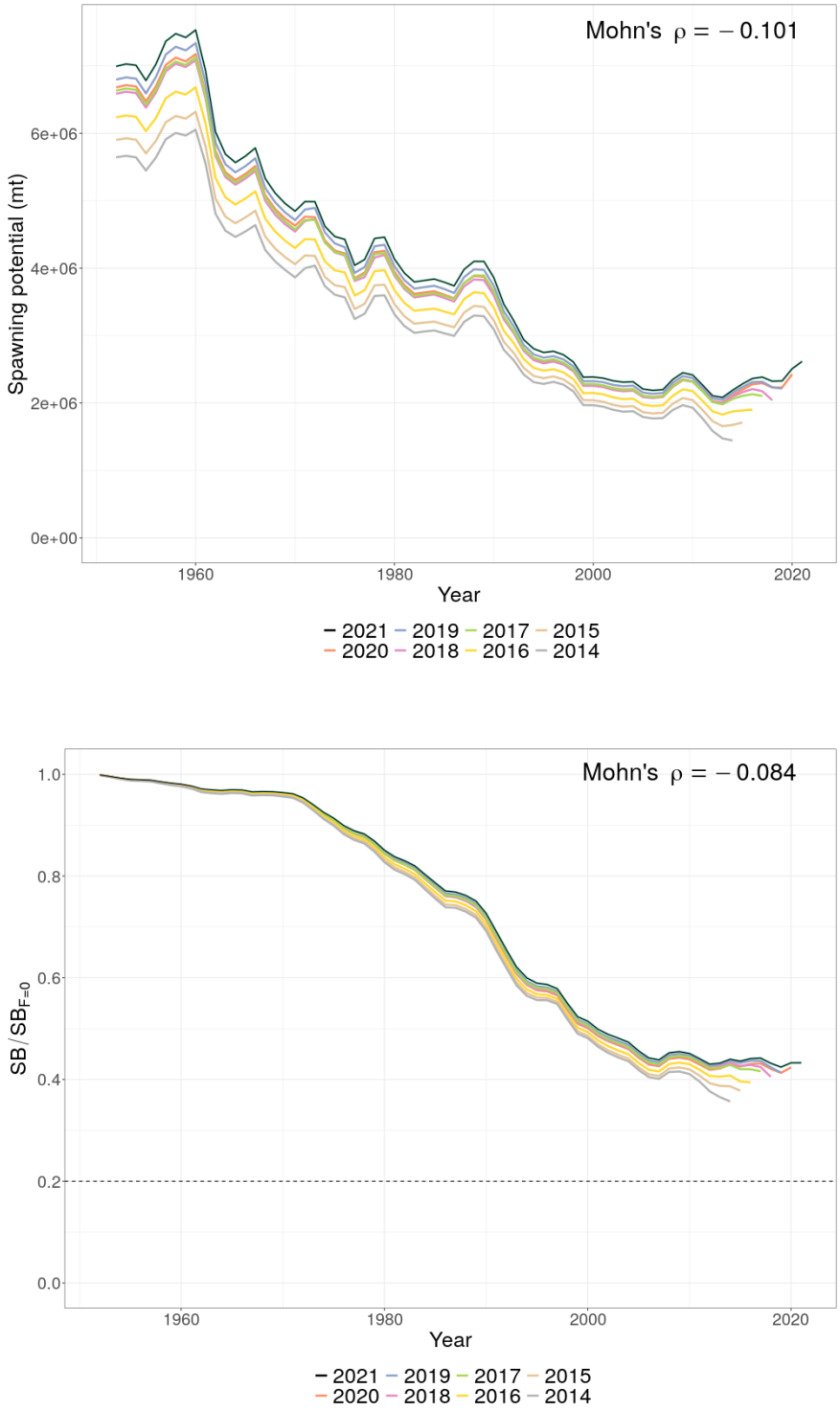

Figure 12.4: Estimated spawning potential (top) and spawning depletion (bottom) for the retrospective models.

12.3 'Status quo' stochastic stock projections for WCPO yellowfin tuna

These will be completed for the Tropical Tuna Measure meeting.

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.

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

Andrews, A., Okamoto, K., Satoh, K., Welte, K., Eveson, P., Roupsard, F., Macdonald, J., Lockheed, B., and Farely, J. (2022). Final report on bomb radiocarbon age validation for yellowfin and bigeye tunas in the WCPO (Project 105) - 2022. WCPFC Scienitific Committee paper SC18-SA-IP-14a.

Appleyard, S., Grewe, P. M., Innes, B., and Ward, R. (2001). Population structure of yellowfin tuna (*Thunnus albacares*) in the western Pacific Ocean, inferred from microsatellite loci. *Marine Biology*, 139(2):383–393.

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*, 62(2):256–265.

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

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

Davies, N., Fournier, D., Bouyé, F., Hampton, J., and Magnusson, A. (2023). Developments in the MULTIFAN-CL software 2022–2023. WCPFC-SC19-2023/SA-IP-02, Koror, Palau.

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.

Ducharme-Barth, N., Vincent, M., Hampton, J., Hamer, P., Williams, P., and Pilling, G. (2020a). Stock assessment of bigeye tuna in the western and central Pacific Ocean. Technical Report WCPFC-SC16-2020/SA-WP-03.

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.

Evans, K., Grewe, P. M., Foster, S. D., Gosselin, T., Gunasekera, R. M., and Lansdell, M. (2019). Connectivity of tuna and billfish species targeted by the Australian Eastern Tuna and Billfish Fishery with the broader Western Pacific Ocean. Technical Report WCPFC-SC15-2019/SA-IP13.

Evans, K., Langley, A., Clear, N., Williams, P., Patterson, T., Sibert, J., Hampton, J., and Gunn, J. (2008). Behaviour and habitat preferences of bigeye tuna (*Thunnus obesus*) and their influence on longline fishery catches in the western Coral Sea. *Canadian Journal of Fisheries and Aquatic Sciences*, 65:2427–2443.

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., Eveson, P., Krusic-Golub, K., Sanchez, C., Rouspard, F., McKechnie, S., Nichol, S., Leroy, B., Smith, N., and Chang, S.-K. (2017). Age, growth and maturity of bigeye tuna in the Pacific. Technical Report WCPFC-SC13- 2017/SA-WP-01, Rarotonga, Cook Islands, 9–17 August 2017.

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

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

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

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

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 globally important resource. *Nature Science Reports*, 5(16915).

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.

Hamel, O. S. (2023). Natural Mortality: Theory, Estimation, and Application in Fishery Stock Assessment Models. Publisher: Northwest Fisheries Science Center (U.S.). Fishery Resource Analysis and Monitoring Division.

Hamer, P. (2023). Report from the SPC Pre-assessment Workshop – April 2023. Technical Report SC19-SA-IP-01.

Hamer, P., Potts, J., Macdonald, J., and Senina, I. (2023). Review and analyses to inform conceptual models of population structure and spatial stratification of bigeye and yellowfin tuna assessments in the Western and Central Pacific Ocean. Technical Report SC19-SA-WP-02, Palau.

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 Kleiber, P. (2003). Stock assessment of yellowfin tuna in the western and central Pacific Ocean. Technical Report WP YFT-1, SCTB 16, Mooloolaba, Australia, 9-16 July 2003.

Hampton, J., Kleiber, P., Langley, A., Takeuchi, Y., and Ichinokawa, M. (2005). Stock assessment of yellowfin tuna in the western and central Pacific Ocean. Technical Report WCPFC SC1 SA WP-1, Noumea, New Caledonia, 8-19 August 2005.

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.

Hare, S., Pilling, G., and Williams, P. (2023). A compendium of fisheries indicators for target tuna stocks in the WCPFC Convention Area. WCPFC-SC19-2023/SA-WP-06.

Hare, S., Williams, P., Castillo Jordan, C., Hamer, P., Hampton, J., Lehodey, P., Macdonald, J., Scutt Phillips, J., Scott, R., Senina, I., and Pilling, G., editors (2022). *The western and central Pacific tuna fishery: 2021 overview and status of stocks oceanic fisheries programme*. Number no. 22 in Tuna fisheries assessment report. Secretariat of the Pacific Community, Nouméa.

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 Nicol, 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.

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

Hurtado-Ferro, F., Szuwalski, C. S., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., Licandeo, R., McGilliard, C. R., Monnahan, C. C., Muradian, M. L., Ono, K., Vert-Pre, K. A., Whitten, A. R., and Punt, A. E. (2015). Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. *ICES Journal of Marine Science*, 72(1):99–110.

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.

Ijima, H. and Jusup, M. (2023). Tuna and billfish larval distributions in a warming ocean. arXiv:2304.09442 [physics, q-bio].

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.

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

Kolody, D. and Hoyle, S. (2014). Evaluation of tag mixing assumptions in western Pacific Ocean skipjack tuna stock assessment models. *Fisheries Research*, http://dx.doi.org/101016/j.fishres.2014.05.008.

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. and Million, J. (2012). Determining an appropriate tag mixing period for the Indian Ocean yellowfin tuna stock assessment. IOTC–2012–WPTT14–31.

Macdonald, J., Williams, P., Roupsard, F., Sanchez, C., Bell, L., Chang, S.-K., Bernadette Contreras, R., Ghergariu, M., Hosken, M., Hoyle, S., Nguyen Cuu, S., Park, T., Potts, J., Schneiter, E., and Nicol, S. (2023a). Project 90 update: Better data on fish weights and lengths for scientific analyses. Technical report, Koror, Palau.

Macdonald, J., Williams, P., Sanchez, C., Schneiter, E., Prasad, S., Ghergariu, M., Hosken, M., Panizza, A., Park, T., and Nicol, S. (2023b). Project 90: Better data on fish weights and lengths for scientific analysis. WCPFC Scienitific Committee paper SC19-ST-IP-04, Palau.

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. (2014). 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., 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. (2017). Stock assessment of bigeye tuna in the western and central Pacific Ocean. Technical Report WCPFC-SC13-2017/SA-WP-05, Rarotonga, Cook Islands, 9–17 August 2017.

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.

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.

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. (2023). Analysis of tag seeding data and reporting rates for purse seine fleets. WCPFC Scienitific Committee paper SC19-SA-IP-08, Palau.

Peatman, T., Scutt Phillips, J., and Nicol, S. (2023a). Analysis of tagging data for the 2023 yellowfin and bigeye tuna assessments: corrections to tag releases for tagging conditions. WCPFC Scienitific Committee paper SC19-SA-IP-09, Palau.

Peatman, T., Scutt Phillips, J., Potts, J., and Nicol, S. (2022). Analysis of tagging data for the 2022 skipjack tuna assessment: corrections for tagging conditions. Technical Report WCPFC-SC18-2022/SA-IP-20.

Peatman, T., Williams, P., Magnusson, A., Day, J., and Teears, T. (2023b). Analysis of purse-seine and longline size frequency data for the 2023 bigeye and yellowfin tuna assessments. WCPFC Scienitific Committee paper SC19-SA-IP-03, Palau.

Peatman, T., Williams, P., and Nicol, S. (2021). Project 60: Progress towards acheiving SC16 recomendations. Technical Report WCPFC-SC17-2021/ST-IP-04.

Peatman, T., Williams, P., and Nicol, S. (2023c). Project 60: Progress report - improving purse seine species composition. WCPFC Scienitific Committee paper SC19-ST-IP-03, Palau.

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.

Pilling, G., Scott, R., Davies, N., and Hampton, J. (2016). Approaches used to undertake management projections of WCPO tuna stocks based upon MULTIFAN-CL stock assessments. Technical Report WCPFC-SC12-2016/MI-IP-04, Bali, Indonesia, 3–11 August 2016.

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.

Punt, A. E. (2023). Those who fail to learn from history are condemned to repeat it: A perspective on current stock assessment good practices and the consequences of not following them. *Fisheries Research*, 261:106642.

Punt, A. E., Maunder, M. N., and Ianelli, J. (2023). Independent review of recent WCPO yellowfin tuna assessment. Technical Report SC19-SA-WP-01.

Reglero, P., Tittensor, D., Álvarez Berastegui, D., Aparicio-González, A., and Worm, B. (2014). Worldwide distributions of tuna larvae: revisiting hypotheses on environmental requirements for spawning habitats. *Mar. Ecol. Prog. Ser.*, 501:207–224.

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.

Satoh, K., Katayama, S., Tanabe, T., and Okamoto, K. (2023). Connectivity in juvenile yellowfin tuna *Thunnus albacares* between temperate and tropical regions of the western Pacific Ocean. *Mar. Ecol. Prog. Ser.*, 713:151–172.

Schaefer, M., Broadhead, G., and Orange, C. (1963). Synopsis on the biology of yellowfin tuna *Thunnus (neothunnus) albacares* (bonnaterre) 1788 (Pacific Ocean). Technical report.

Scutt Phillips, J., Lehodey, J., Hampton, J., Senina, I., and Nicol, S. (2022). Quantifying rates of mixing in tagged, WCPO skipjack tuna. Technical Report WCPFC-SC18-2022/SA-WP-04.

Scutt Phillips, J., Sen Gupta, A., Senina, I., Sebille, E., Lange, M., Lehodey, P., J, H., and Nichol, S. (2018). An individual-based model of skipjack tuna *Katsuwonus pelamis* movement in the tropical Pacific Ocean. *Progress in Oceanography*, 164:63–74.

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.

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

Thorson, J. T. (2019). Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. *Fisheries Research*, 210:143–161.

Thorson, J. T., Johnson, K. F., Methot, R. D., and Taylor, I. G. (2017). Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. *Fisheries Research*, 192:84–93.

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

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., Ducharme-Barth, N., Hamer, P., Hampton, J., Williams, P., and Pilling, G. (2020). Stock assessment of yellowfin tuna in the western and central Pacific Ocean. WCPFC-SC16-2020/SA-WP-04.

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.

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.

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 Ruaia, T. (2023). Overview of tuna fisheries in the Western and Central Pacific Ocean, including economic conditions - 2022. Technical report.

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.