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

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

WCPFC-SC19-2023/SA-WP-05 (Rev. 2)

Revision 2 September 15, 2023

J. Day[1], A. Magnusson[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 2:

This revision of the paper adds:

- Corrected y-axis numbers on annual recruitment plots, calculated as the sum rather than the average across seasons, effectively multiplying by four (Figures 45, 46, 49).
- Corrected y-axis labels on Majuro and Kobe plots, adding the subscript 'recent' (Figure 68).

Revision 1:

This is a revision of the first complete version, which was labelled 2.01. This revision of the paper adds:

- Corrected proportion-by-source-region plot Figure 48
- Corrected Majuro and Kobe plots Figure 68
- New dynamic MSY plot Figure 72

Contents

1 Executive Summary

This paper describes the 2023 stock assessment of bigeye tuna *Thunnus obesus* 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 applies the same 9–region model structure that was used for management advice from the 2020 assessment. New developments to the stock assessment include:

- Conversion from a catch-errors to a catch-conditioned modelling framework, and the inclusion of a likelihood component for the CPUE from the index 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 using the Lorenzen functional form for natural mortality at age.
- Additional procedures adopted to achieve more reliable model convergence, including extensive jittering and checking the 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 used in the structural uncertainty grid.
- Use of conditional age-at-length data, and internal estimation of growth, with alternative weighting of these data included in the structural uncertainty grid.

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), improvement to data for length-weight conversion factors (Macdonald et al., 2023b) and developments to MFCL software (Davies et al., 2023).

The assessment implemented a more rigorous approach to achieve more reliable and stable model convergence than seen in previous bigeye assessments, which was beneficial in achieving a positive definite Hessian for the 2023 diagnostic model. Subject to the caveat that models in the stepwise development did not involve the extensive level of jittering analysis that was applied to the diagnostic and grid models and, as a result, stepwise models are unlikely to have converged, the most influential steps in the development of the 2023 diagnostic model appeared to be: restricting the constraint of non-decreasing selectivity to a single fishery and estimating growth internally, which increased the final estimate of $SB/SB_{F=0}$; and applying the revised tagger effects method (as recommended by an expert workshop and supported by the Pre-assessment Workshop), which decreased the final estimate of $SB/SB_{F=0}$. The 2020 bigeye stock assessment estimated the median $SB_{\text{recent}}/SB_{F=0}$ across the model grid to be 0.41, 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 0.35. Overall, the current diagnostic model estimates a more pessimistic stock status, or lower $SB/SB_{F=0}$, than the 2020 diagnostic model.

In addition to the diagnostic model, we report the results of one-off sensitivities 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 (for an uncertainty 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 full uncertainty for these management quantities. The ability to include estimation uncertainty in addition to 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 including estimation uncertainty did not influence the management advice. It is, however, recommended that management advice is formulated using the results of the structural uncertainty grid combined with the estimation uncertainty 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 in the structural uncertainty grid, the most important factors when evaluating stock status were; the assumed tag mixing period and the steepness of the stock recruitment relationship, followed by weighting of the size composition data. The move away from using the complex fixed functional form for natural mortality at age function, as used in the previous assessment, 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. The internal estimation of natural mortality allowed natural mortality to be dropped from the structural uncertainty grid.

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.
- Average fishing mortality rates for juvenile and adult age-classes have increased throughout the period of the assessment until around 2000, after which they have stabilised, but with high inter-annual variability for juveniles. Juveniles have experienced considerably higher fishing mortality than adults.
- Overall, the median depletion from the uncertainty grid for the recent period (2018-2021; $SB_{\text{recent}}/SB_{F=0}$) is estimated at 0.35 (80 percentile range including estimation and structural uncertainty 0.30–0.40, full range 0.25–0.46)
- No models from the uncertainty grid, including estimation uncertainty, estimate the stock to be below the LRP of 20% $\text{SB}_{F=0}$.
- CMM 2021-01 contains an objective to maintain the spawning biomass depletion ratio above the average for 2012-2015, $SB_{2012-2015}/SB_{F=0}$, which is a value of 0.34 calculated across the unweighted grid. Based upon the estimates of $SB_{\text{recent}}/SB_{F=0}$ of 0.35 this objective has currently been met.
- Recent (2017–2020) median fishing mortality ($F_{\text{recent}}/F_{\text{MSY}}$) was 0.59 (80 percentile range, including estimation and structural uncertainty 0.46–0.74, full range 0.37–0.99).
- Assessment results suggest that the bigeye 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:

- Increased representative biological sampling, especially of age data.
- Continued resources to develop staff technical skills, improve software platforms and data preparation and quality control.
- Greater time and resources to conduct these assessments, to explore these complex models and data sources.
- Continued collection of more representative tagging data.
- A comprehensive review of size compositional data to ensure the data being used in the assessment models is representative of the population.

Tag mixing is a key issue for a tag based model and one that deserves more time. Preliminary analysis of individual tagging events was a good start, but more could potentially be achieved by

extending this approach. Tags from region 9 are clearly influential and appear to behave differently, when looking at preliminary Piner plot (which were not finished in time to include in this report).

Data weighting is an area that is difficult given the complexity and running time for these models. Iterative schemes have been used elsewhere to weight compositional data and conditional age-at-length data. While running these iterative approaches to convergence is clearly not practical, this is an area that could be explored.

Additional biological sampling, including age samples, and potentially the use of epigenetic ageing is very important for these models. The sample size of only 1004 otoliths is a major weakness in the data input to this assessment. A well balanced statistically designed sampling program to achieve representative temporal and spatial samples over the full range of the assessment area would be beneficial to this assessment and an increase in investment will improve the reliability of the model outputs. As always, quality control over the sampling and sampling protocols is important to avoid "garbage in, garbage out".

While the tagging data is very valuable for this assessment, the current tagging programs could be better balanced spatially and temporally to inform the assessment.

Simpler regional structures are an area that deserves more attention. These assessments are large and complex, some may say unwieldy, so there is potentially much to gain though judicious simplification. While a simpler structure could have modelling benefits, there are also potential issues with the eastern boundary. Recaptures of tagged fish demonstrate that fish cross this boundary and the large catches of bigeye near this boundary warrant consideration of alternatives assessment structures.

2 Introduction

This paper presents the 2023 stock assessment of bigeye tuna (*Thunnus obesus*; BET) in the western and central Pacific Ocean (WCPO; west of 150° W). Assessment of WCPO bigeye tuna has been conducted regularly since the late 1990s (Langley et al., 2008; Harley et al., 2009, 2010; Davies et al., 2011; Harley et al., 2014; McKechnie et al., 2017a; Vincent et al., 2018; Ducharme-Barth et al., 2020a). As in previous assessments, the objectives of the 2023 bigeye 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 2.2.2.0; 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) guide 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 assessments are conducted. Changes to aspects of an assessment can result in changes to the estimated status of the stock and fishing impacts, and resultant management advice. It is important to recognize that each new assessment represents a new estimation of the historical population dynamics and recent stock status, and each new assessment team strive to provide the best possible assessment with the time and data available.

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

[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 the assessment.

The 2023 bigeye 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 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 MFCL developments (Davies et al., 2023) and the paper on conceptual models of bigeye and yellowfin population structure (Hamer et al., 2023). Finally, the planning for this assessment was informed by the discussion 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).
- 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 start 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).
- Improve 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).
- Reduce the number of fisheries with a non-decreasing selectivity constraint to a single extraction fishery.
- Ensure 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 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

Bigeye are distributed throughout the tropical and sub-tropical waters of the Pacific, Indian and Atlantic Oceans. Genetic studies indicate that Atlantic bigeye have minimal mixing with bigeye in the Indo-Pacific regions (Chow et al., 2000; Gonzalez et al., 2008; Grewe et al., 2019). However, there is currently no clear evidence for genetic population structure in the Pacific Ocean (Grewe and Hampton, 1998; Moore et al., 2020; Natasha et al., 2022). While genetic studies are largely uninformative on the rates of mixing of bigeye tuna throughout the Pacific, they are broadly consistent with the results of tagging experiments conducted by SPC and the Inter-American Tropical Tuna Commission (IATTC) (Schaefer et al., 2015; Moore et al., 2020). These studies show that while the majority of tagged bigeye have been recaptured in the general regions of their release (i.e. within 1,500 nautical miles) some tagged bigeye move large distances across the Pacific Ocean (i.e. >4,000 nautical miles). These occasional large-scale movements and the continuous distribution of bigeye tuna across the Pacific would explain the lack of broad-scale genetic structure. The cumulative information on movement from conventional and electronic tag-recapture data, particularly efforts focused in the equatorial central Pacific, have indicated more extensive longitudinal, particularly west to east, displacements with few movements between tropical tag release sites and temperate zones (Schaefer et al., 2015; Moore et al., 2020; McKechnie et al., 2017a). While there is apparent high movement throughout the equatorial Pacific and lack of genetic structure, this does not preclude spatial structure in population processes that are important in stock assessment (i.e., growth, mortality, recruitment). For example, growth rates have been shown to differ between bigeye tuna in the WCPO and EPO, with smaller lengths-at-age observed for bigeye in the WCPO, and suggestion of a clinal increase in growth rates from the west to east Pacific (McKechnie et al., 2015a; Aires-da Silva et al., 2015; Farley et al., 2017). This is also reflected in observations of increased size at maturity for bigeye in the EPO compared to the WCPO (Farley et al., 2017). Further, there is some evidence from otolith chemistry studies for structuring of recruitment sources. For example, Rooker et al. (2016) indicated that 1–2 year old fish sampled from the Marshall Islands were derived from local sources, whereas fish caught off Hawaii were suggested to have originated more broadly from the central equatorial region. In turn, otolith chemistry and genetic studies of juvenile bigeye in the Indonesian archipelagic waters suggested juvenile residency and low mixing with the broader western Pacific (Moore et al., 2019; Proctor et al., 2019).

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)), plus 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 bigeye are likely one genetic stock in the Pacific, and that there is no clear boundary between the WCPO

and EPO populations in the central Pacific region. However, there is indication of substructure of bigeye across the Pacific, and that the populations across the west and east Pacific, and the north and south Pacific are unlikely to be extensively mixed. 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 region warrants being considered a separate spatial strata. The available information did not clearly imply a finer structure, however, it was noted that finer spatial strata are likely required to accommodate tagging data and related assumptions. The review recommended alternative simplified spatial stratification that could be considered for bigeye tuna assessments in the WCPO and this is discussed further in Section 4.2. Continuing to improve our understanding of spatial population structure and processes for bigeye in the WCPO, and more broadly in the Pacific, remains an important area of research.

Despite uncertainties in population connectivity across the Pacific a sensitivity analysis by McKechnie et al. in 2015 showed that despite extensive longitudinal movements in the equatorial Pacific, the discrete WCPO stock assumption was capable of accurately capturing the dynamics and stock status indicators for WCPO bigeye. The current assessment continues the approach of a separate assessment for the WCPO region and the stock within the domain of the model (essentially the WCPO, west of 150° W) is considered to be a discrete stock unit (Langley et al., 2008; Harley et al., 2009, 2010; Davies et al., 2011; Harley et al., 2014; McKechnie et al., 2017a; Vincent et al., 2018; Ducharme-Barth et al., 2020a).

Spatial stratification of the WCPO bigeye assessment has changed over time (Hamer et al., 2023). The 9 region spatial structure (Figure 1) applied in the 2020 assessment is the basis for the current assessment and is a compromise between the limited knowledge of sub-regional population structure, fishery spatial structures and the locations of major tag release events (i.e., regions 4, 8 and 9). Ultimately we were able to succeed in improving the performance and convergence properties of 9 region structure and it has been maintained for this 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 do recommend a thorough exploration of alternative spatial structures (prior to the next production assessment) if an alternative spatial structure is to be considered as the basis for management advice in future assessments. We also note that we did conduct a preliminary model with a 6 region structure, and this looked promising, but required more time to explore thoroughly.

3.2 Biological characteristics

Bigeye tuna are moderately fast growing for tuna, and have a maximum fork length (FL) of about 200 cm (Aires-da Silva et al., 2015; Farley et al., 2017), with an estimated average maximum

length of 157 cm in the WCPO (Eveson et al., 2020; Farley et al., 2020). The growth parameters for bigeye tuna are influential biological inputs to the stock assessment (McKechnie et al., 2017a; Vincent et al., 2018). Studies to improve growth information for bigeye have occurred in recent years, notably by Eveson et al. (2020); Farley et al. (2020). There are no new studies of bigeye growth to inform this assessment since (Farley et al., 2020). However, recent bomb radiocarbon age work has validated the interpretation of growth bands in bigeye otoliths as annual growth bands (Andrews et al., 2022) and confirmed the oldest 'validated' age for a bigeye tuna in the WCPO at 13 years of age. This compares to the longest period at liberty for a recaptured bigeye tuna tagged in the WCPO of approximately 14 years, for a fish estimated at 1-2 years old at release (SPC, unpublished data). The oldest estimated fish aged from otoliths in the WCPO is approximately 15 years (Farley et al., 2020).

Bigeye tuna in the WCPO become reproductively active from about 80 cm FL, and nearly all individuals >120 cm FL are reproductively mature (Farley et al., 2017). There is some evidence for regional variation in maturity-at-length in the WCPO (Nicol et al., 2011; Farley et al., 2017), and bigeye tuna appear to be reaching maturity at larger sizes, but at similar ages, in the EPO (Schaefer et al., 2005). As with other tunas, the population sex ratio of bigeye tuna changes at around the age/size of reproductive maturity to favor males at larger size (see Figure 8 in Sun et al., 2006; McKechnie et al., 2017b). This information is used to define spawning potential (rather than spawning biomass) as a product of maturity status, female sex-ratio in the population, and fecundity. Spawning typical occurs in water temperatures above 24° C, with an approximate daily cycle, and occurs all year in the tropical Pacific, but more seasonally in the warmer months in sub-tropical regions (Sun et al., 2006; Farley et al., 2017). Areas of higher larval densities have been found in the western Pacific from Papua New Guinea to Philippines, Chinese Taipei to the Okinawa Islands, also several areas in the central Pacific, and in the eastern Pacific off Panama and Colombia (Nishikawa et al., 1985; Ijima and Jusup, 2023).

The natural mortality (M) rate of bigeye tuna is likely to vary with size, with rates of < 0.5 yr^{-1} for bigeye >40 cm FL and >2 years age and higher rates for the youngest age-classes (<2 years) (Hampton, 2000). Tag recapture and otolith data indicate that significant numbers of bigeye reach at least eight years of age (Hampton and Williams, 2005; Farley et al., 2020) and can likely reach 15 years age in the WCPO (Farley et al., 2020). The 2020 bigeye assessment conducted a life-history based meta-analysis (Piner and Lee, 2011) of natural mortality for bigeye and indicated an envelope of potential average quarterly M rates of lower 95% confidence interval (0.109), mean (0.127) and upper 95% confidence interval (0.146) (Ducharme-Barth et al., 2020a). Additionally, natural mortality of female bigeye tuna is thought to increase at around the age of reproductive maturity, due to the physiological stresses of spawning, which, as noted above, is hypothesized to drive the occurrence of a male-biased sex ratio at larger sizes.

3.3 Fisheries

Bigeye tuna are an important component of tuna fisheries throughout the Pacific Ocean. Larger bigeye are targeted by longlines, predominantly in tropical regions of the WCPO. Smaller juveniles are caught incidentally by purse seine vessels in association with Fish Aggregation Devices (FADs). They are a principal target species in tropical waters for both the large, distant-water longline fleets of Japan, Korea, China and Chinese Taipei and the smaller, fresh sashimi longline fleets based in several Pacific Island countries and Hawaii. The prices paid for both frozen and fresh product on the Japanese sashimi market are generally the highest of all the tropical tunas.

While they are not targeted by purse seine vessels, bigeye are taken almost exclusively from sets on natural and artificial floating objects (FADs), and comprise mostly small juveniles. Estimation of the bigeye tuna catch from associated sets has been the focus of considerable research over recent years (Peatman et al., 2019, 2023c). This purse seine fishery expanded rapidly from the early 1980's and the estimated annual bigeye catches for this gear have been in the vicinity of 45,000-75,000 mt over the last 10 years in the WCPO (i.e., the region of this assessment which excludes the overlap area with IATTC). This compares to the WCPO longline catch of between approximately 35,000-70,000 mt (Hare et al., 2022; Williams and Ruaia, 2023).

The highest estimated bigeye tuna total catch from the WCPO was approximately 175,000 mt in 2004 (195,000 mt in the same year for the entire WCPFC-CA). The WCPO bigeye catches have declined since around 2010, after which they have ranged between around 110,000–160,000 mt (135,000-175,000 mt for entire WCPFC-CA) (Hare et al., 2022; Williams and Ruaia, 2023). Over the recent period (2015-2022), approximately 50% of the catch by weight, but 10% by numbers was taken by longline, reflecting the selection of longline gear for larger fish. In contrast, 50% of the catch by numbers, but only 35% by weight was taken by purse seine sets associated with FADs. A small purse seine fishery also operates in the coastal waters off Japan with an annual bigeye catch of less than 500 mt in recent years. A slightly higher level of bigeye catch is taken by the coastal Japanese pole-and-line fishery relative to the coastal Japanese purse seine, however, overall these catches are negligible.

In recent years, collaborative work between SPC, WCPFC, CSIRO (primarily in Indonesia), and fisheries agencies in Indonesia, the Philippines, and Vietnam has resulted in improved catch statistics for their fleets. 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 bigeye tuna, e.g., ring-net, handline, and troll. Data for these fisheries are included in the assessment, and account for a significant component of the bigeye catch, particularly in terms of numbers.

3.4 Key changes from the last assessment

3.4.1 Catch conditioned approach

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

In an effort to reduce complexity and paramterisation this assessment makes use of a relatively new feature of MFCL first applied to the 2022 skipjack tuna assessment in which catch is assumed to have no error, i.e., the model is "catch-conditioned" (Davies et al., 2022). This makes it possible to solve the catch equation for fishing mortality exactly, using a Newton-Raphson sub-iterative procedure. The main benefit of this approach is that effort deviation coefficients and catchability-related parameters do not require estimation as model parameters. Effort data for extraction fisheries is not required at all but can be used if available to estimate catchability through regressions of fishing mortality and effort, and this is important for making stock projections based on future effort scenarios. The reduction in parameters has enabled more rapid model convergence and Hessian matrix computation. The only cost of this approach is that missing catches, which could be accommodated in the catch-errors version if there was an accompanying effort observation, are no longer straight forward to account for. However, this is not an impediment for the key WCPO tuna assessments. The catch conditioned approach allows (but does not require) the specification of index fisheries to provide indices of relative abundance, these are discussed in section Section 4.4. In the stepwise model development runs conducted for this assessment, the transition from a 'catch-errors' to a "catch-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 bigeye 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 (Teears et al., 2023), and in Figure 6.

4.2 Spatial stratification

The geographical area considered in the assessment corresponds to the WCPO (from 50°N to 20°S between 120°E and 150°W) and oceanic waters adjacent to the east Asian coast (110°E between 20°N and 20°S). The eastern boundary of the assessment excludes the WCPFC Convention area component that overlaps with the Inter American Tropical Tuna Commission (IATTC) area. The assessment area is stratified into 9 regions that are consistent with the 2020 assessment(Figure 1). The 9 region stratification was first used in Harley et al. (2014), and was slightly modified in the 2020 assessment by moving the northern boundary of regions 3 and 4 from 20°N to 10°N to better isolate the tropical region surface and longline fisheries (Ducharme-Barth et al., 2020a). Some small regions are included, i.e., region 8 designed to approximate the archipelagic waters of Papua New Guinea and Solomon Islands, where considerable tagging effort has occurred and the analyses show more persistent residence compared to the wider western equatorial region, and region 9 that was established in 2014, after the previous bigeye assessment peer review (Ianelli et al., 2012), to better model the tagging data from the historical Coral Sea tagging cruises (Harley et al., 2014).

While the 9 region spatial stratification was the basis for the development of the diagnostic model for this assessment, exploring simpler spatial stratifications was recommended by the peer review of the 2020 yellowfin assessment (which applied the same 9 region stratification) and the authors of the previous WCPO bigeye assessment. A revision of the 9 region stratification, and consideration of alternative spatial stratification, based on a conceptual model, including additional analysis of size composition data was suggested by the peer review panel. This work is described in Hamer et al. (2023), which suggested a simpler 6 region model stratification could be considered for bigeye. This was also discussed and received support at the SPC Pre-assessment Workshop (PAW) (Hamer, 2023), along with other possibilities. Under the time constraints and other difficulties posed by this assessment, we could not complete a thorough investigation of the various alternative stratifications discussed at the PAW. The review paper indicated that the current 9 region model remains suitable, but alternative model structures may have advantages and further exploration of model spatial stratifications is warranted. As we were able to improve the performance and convergence properties of the 9 region model we chose to maintain it for the current assessment. We conducted preliminary models of a 6 region spatial stratification but due to time constraints we could not include this in the current assessment. We suggest a stand alone project is required to fully explore and compare

the benefits and limitations of alternative spatial structures for WCPO bigeye assessment, with review by SC and their advice on a preferred option for future assessments.

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 the 9 region model 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 Indonesia/Philippines/Vietnam region. The fisheries definitions for the 2023 assessment are consistent with those used in the 2020 9 region assessment. 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 3 and 7 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 (Section 3.4.1) allows the specification of "index fisheries" that are used to provide standardised CPUE indices of abundance. Index fisheries are akin to "survey fisheries" as described for other software such as Stock Synthesis, and may be the same fisheries as the extraction fisheries, but when used as index fisheries they do not take any

catch, and must have effort data to allow modelling of CPUE. For this assessment one index fishery is defined for 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 data-set (described in McKechnie et al., 2015b; Ducharme-Barth et al., 2020b) 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 Section 4.5.3).

The standardised indices for each region are derived by summing outputs at the spatial scale the spatio-temporal CPUE standardization model into each of the MFCL regions. 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.

Alternative spatial stratification: For the exploration of the 6 region spatial stratifications 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 models 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 than changing both the spatial stratification and fishery definitions together. This way the effects of simplified stratification could be isolated. If a simpler model stratification is preferred in future, revision of fisheries definitions would be considered.

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 7 - 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 3, 4, 8) 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 fisheries 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 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 (3, 4, 7, and 8).

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

- The steady increases in total bigeye catch over most of the assessment period in the equatorial regions (regions 3, 4, 7, 8), and steady declines since around 2012.
- The relatively stable catches of bigeye in the northern temperate region by longline vessels (regions 1, 2)
- The development of the equatorial purse-seine fisheries from the mid-1970s and the widespread

use of FADs since the mid-1990s, allowing an expansion of the purse-seine fishery, and corresponding increases in catch of bigeye, particularly in equatorial regions 3, 4 and 8.

- Large changes in the purse seine fleet composition and increasing size and likely efficiency of the fleet.
- The steady increase in catch for the domestic fisheries of Indonesia and the Philippines (region 7) since 1970.
- The apparent stabilisation of catches of bigeye for most gears after the mid 2000's, and declines in overall catches since around 2010.

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 relative robust to the lower observer coverage (Hamer, 2023). A sensitivity exploring the effects of doubling the purse seine catch in regions 3, 4 and 8 was conducted to investigate this issue.

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 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 data-set was first created in 2015 in support of the Pacific-wide bigeye tuna stock assessment (McKechnie et al., 2015b), 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., 2017a; Vincent et al., 2018; Ducharme-Barth et al., 2020a). Since 2017 spatiotemporal approaches (Thorson et al., 2015) 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 sdmTMB (Anderson et al., 2022). The sdmTMB package is preferred over VAST due to 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 bigeye included HBF, vessel FLAG, season, 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° x 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 however, these 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, with the catch-conditioned approached effort data are not necessary 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 bigeye 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 Section 5.5.2. Summaries of the available size composition data by year and fishery are provided in Figure 13 and Figure 14.

4.6.1 Purse seine

Only length-frequency samples are used in the bigeye 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 7 (Fishery 30). In the model, it was assumed to share a selectivity with the "all flags" associated purse seine fishery in the adjacent region 3 (Fishery 13).

4.6.2 Longline

A review of all available longline length and weight-frequency data for bigeye 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 to 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 7 (Fishery 32).

Size data were missing for the Indonesian-Philippines ex-EEZ purse seine fishery in region 7 (Fishery 24). Based on an investigation of the length frequency data of the other tropical tunas (skipjack and yellowfin) available for this fishery, selectivity was assumed to be shared with the Philippines small-fish fishery in region 7 (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 7 (Fishery 23) and the domestic Vietnam small-scale fishery in region 7 (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 7 (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

Of the three main tropical tuna species, bigeye tuna has the least amount of tagging data available. Information on the bigeye 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 15, 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 bigeye assessment is that there are a considerable number of 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. This pooling of the two species also permitted

estimation of correction factors for bigeye tuna released during the RTTP, which previously had no detectable tagging effects. 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 2.

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 $\sim$ 24 cm FL) in the population. Bigeye 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 (Schaefer et al., 2005; Sun et al., 2006; Itano, 2000). The assessment model assumed that recruitment occurs instantaneously at the beginning of each quarter. This is a discrete approximation to continuous recruitment, but provides sufficient flexibility to allow a range of variability to be incorporated into the estimates as appropriate.

In recent assessments of tuna in the WCPO, the last few recruitment deviations have little information to inform them and these are usually fixed at the mean of the estimated recruitment deviations, which also reduces problems with retrospective analyses. This approach has been continued here with the last six recruitment deviations fixed at the geometric mean of the estimated recruitment deviations, 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. In the 2020 assessment, the initial total mortality at the start of the model was set to the average total mortality (Z) for the first 20 time periods. In the 2023 assessment, the initial fishing mortality was assumed to be zero ($Z = M$). This change was introduced as part of the stepwise model development. in order to implement the a 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 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 function based on high-readability otoliths (Farley et al., 2020), a fixed external Richards growth curve based on the same high-readability otolith data-set 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 (Ducharme-Barth et al., 2020a). The assessment was sensitive to approaches used for growth and both the internal CAAL and external otolith only growth curves led to unsatisfactory results, therefore the external fixed tag-otolith integrated 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 the otolith sampling protocols, with otoliths selected according to length (i.e., attempting to achieve sufficient numbers of otolith samples across the full length range, plus bias in otolith readability with age), that any external growth curves

[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)

would likely be biased. They strongly recommended that for this type of otolith sampling regime that age data should be conditional on length and the growth is estimated internally, using conditional age-at-length data within MFCL. 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.

5.2.4 Movement

Movement was assumed to occur instantaneously at the beginning of each quarter via movement parameters 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 parameters 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 (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 parameters are invariant with respect to age. A "prior" of 0 is assumed for all movement parameters, 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. We ran a preliminary model with fixed movement parameters, using outputs from the SEAPODYM model, but this model resulted in very poor fits to the CPUE data and there was insufficient time to explore this further.

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 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 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 use of this method. 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 estimation of M-at-age for tuna. Other CAPAM reviews of approaches for estimating M have also, where possible, favoured 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. This change from the previous approach of setting a fixed value of M-at-age to the Lorenzen form with internal estimation of the scalar 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 (Ducharme-Barth et al., 2020a). 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 processes: proportion females-at-length[7] (sex-ratio), proportion of females mature-at-length[8], and the fecundity-at-length of mature females[9] (Figure 17). Another added benefit is that this reproductive

[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.

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 in the 2020 assessment, we have not included spawning fraction in the reproductive potential calculation, as these data are still not adequate for bigeye in the WCPO.

5.3 Fishery dynamics

5.3.1 Selectivity

Selectivity is estimated as a function of age within MFCL. Estimated selectivity curves at age are included in Figure 36 and Figure 38. These selectivity curves can be transformed to equivalent selectivity by length and these are shown in Figure 37 and Figure 39. Selectivity is modelled using a cubic spline, as in the 2020 assessment. This allows for greater flexibility than assuming a functional relationship with age (e.g. a logistic curve to model monotonically increasing selectivity or double-normal curve to model fisheries with reduced selectivity for both the youngest and oldest fish), and requires fewer estimated parameters than modelling selectivity with separate age-specific parameters. This provides a form of smoothing, and 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 was sufficient for all fisheries in this case. Some other modifications to selectivity curves were made to improve fits to size composition data and these are noted in Section 6.1.

In all cases, selectivity is assumed to be time-invariant and fishery-specific. However, a single selectivity function cab be "grouped" or "shared" among a group of fisheries that have similar size compositions or were assumed to operate in a similar manner. This grouping of selectivity facilitates a reduction in the number of parameters estimated and allows selectivity to be set for fisheries with either limited or missing size composition data. Selectivity groupings are indicated in Table 1.

Length-based selectivity is not currently implemented in MFCL. Efforts were made to implement this feature in 2023 in response to a YFT peer review recommendation, but there was insufficient time to implement and fully test this feature and apply it in 2023. To prevent a model from producing a cryptic biomass unavailable to any fishery, which can produce good model fits and likelihoods, but with little evidence to support the existence of these "invisible" fish, it is good practice to set a "non-decreasing" penalty on selectivity for at least one fishery to encourage that selectivity to be estimated as non-decreasing, as a function of age. The penalty is actually imposed on decreasing selectivity with age, and it is important to note that this penalty can be overidden if the improvement in likelihood obtained from fitting a decreasing selectivity outweighs the cost of the penalty, in likelihood terms. In the 2020 assessment, this "non-decreasing" penalty was imposed

on all of the longline fisheries and all of the index fisheries, so all of these fisheries were "encouraged" to produce non-decreasing selectivity. For the 2023 assessment, a decision was made to only impose this non-decreasing penalty on a single fishery, to allow improved fits to size composition data of the other fisheries. The single fishery chosen for this purpose was an extraction longline fishery in region 7. This fishery (6.LL.OS.7) catches both larger fish and larger quantities of large fish than most other fisheries (it has a fat right hand tail) and it was considered the most suitable fishery for this purpose (see Section 6.1 for further details).

5.4 Dynamics of tagged fish

Tagged fish are modelled 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 helps to address computational and dimensionality constraints.

5.4.1 Tag reporting

In theory, tag-reporting rates can be estimated internally within the model. In practice, experience has shown that independent information on tag-reporting rates for at least some fisheries tends to be required for reasonable model behavior to be obtained. We provided reporting rate priors for reporting groups (combinations of 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 have similar characteristics to those with independent estimates, but increased the prior variance for these fisheries. For others where 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, tag reporting rates for these 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 Section 6.1). For this assessment, as previously noted, we did not estimate reporting rates for tag reporting groups with five or fewer tag returns. 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 newly tagged fish, i.e., an individual tag and release event is a recruitment event for the 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 due to the time required for mixing with the untagged population to take place.

Depending on the distribution of fishing effort in relation to tag release sites, the probability of capture of tagged fish soon after release may be different to that for the untagged fish. 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 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 location 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 such an approach could potentially be used in future assessments.

The diagnostic model of the previous bigeye assessments assumed that tagged fish gradually mix with the untagged population at the region-level and that this mixing process is complete by the end of the second quarter following the quarter in which the fish are released (ie., "mixing period" assumption is two quarters). A sensitivity model was included whereby the tag mixing period assumption was one quarter (Ducharme-Barth et al., 2020a). 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 bigeye as previous work on skipjack suggested that acceptable tag mixing may take longer than two quarters (Kolody and Hoyle, 2014). In response, 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 quarters referring to recaptures with the same quarter of release, 1 quarter referring to recaptures in the first quarter after release etc.). The recaptures are scaled to the purse seine catches (the purse fishery accounts for 90% of bigeye 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 regions, similar the approach of Langley and Million (2012). When recaptures are observed in 1 x 1 degree 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 full 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 the use of mixing periods of either 1 or 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 an additional sensitivity, although perhaps with less support than using 1 or 2 quarters as a mixing period. It was noted that some tag release groups show more obvious evidence for mixing than others depending on how many tags were released (i.e., the qualitative assessments were 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 the basis used 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 bigeye 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 directly model the CPUE for 'index' fisheries. While these index fisheries exist in the model and in the data inputs as defined 'fisheries', they differ from the regular 'extraction' fisheries in that no catch is extracted and their CPUE is modelled directly as a lognormal likelihood contribution. The CPUE index series for each region is assigned a likelihood weight in the form of a region-specific σ, representing a variance term, describing the magnitude of observation error. The procedure used for estimating σ is based on maximum likelihood estimated, by calculating σ as the standard deviation of log-residuals from a model part way through the stepwise development, at which stage this value is adjusted. The region-specific values for σ are then set at these values for all subsequent models in the stepwise development, for the diagnostic model, and for the structural uncertainty grid.

5.5.2 Length and weight frequency

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

extraction fisheries was divided by two as it is essentially being used twice, and then the sample sizes were further divided by 20 before use within MFCL for the diagnostic model. Alternative divisors of 10 and 40 for the input sample size were explored in sensitivity analyses. We also explored a self-scaling method for weighting the size composition data, the Dirichlet-multinomial likelihood (Thorson et al., 2017). The Dirichlet-multinomial likelihood approach suggested relatively high weights for the size composition data, which resulted in deterioration of the fits to the abundance indices. We therefore decided not to apply the Dirichlet-multinomial likelihood weighting method, and consider that this approach requires further investigation. However, based on the preliminary results of the Dirichlet-multinomial approach suggest that a divisor of 10 is more consistent with the Dirichlet-multinomial approach than the other two divisors used.

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 parameterisation 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 corresponds to a variance twice that of the Poisson distribution. 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 deal appropriately with this. While this should provide a more realistic weighting of the tag return data in the overall likelihood and allow this source of variation to affect the confidence intervals of estimated parameters, the model estimation may be sensitive to the value assumed for (τ). We conducted a sensitivity analysis with values for τ of 1 and 4. A complete derivation and description of the negative binomial likelihood function for tagging data is provided in Kleiber et al. (2019).

5.5.4 Conditional age-at-length data

A further likelihood component involves using conditional age-at-length (CAAL) data, as recommended by the yellowfin tuna assessment peer review (Punt et al., 2023). These data are included in the assessment and con contribute to estimation of the growth parameters, as they provide direct observations of the distribution of ages, generally measured with some error, within each observed length class. For each each fishery and time period (quarter) aged fish are separated into observed (measured) length classes, with the distribution of ages tabulated as a function of the known length (ideally with ageing measurement error). The measured age distribution within each length interval is assumed to be multinomially distributed, and this forms the basis of the likelihood component for these data. However, despite the otolith data being collected across a range of locations and times the conditional age-at-length data generally not independent has a degree of non-independence, as is the case for size compositional data, so these conditional age-at-length data are downweighted by

a factor of 0.75. We also ran sensitivities with weighting factors of 0.5 and 1. The latter is consistent with the weighting used for conditional age-at-length data in 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 is continued either until a pre-specified minimum gradient is achieved or until a maximum number of iterations is reached. This optimisation is performed by an efficient algorithm using exact derivatives with respect to the model parameters (auto-differentiation, Fournier et al. (2012)). Estimation was conducted in a sequence of phases, the first of which uses somewhat arbitrary starting parameter values. A bash shell script, or "doitall", implements the phased sequence for gradually estimating more and more parameters for the model. Some parameters were assigned starting values consistent with available biological information. The values of these parameters are provided in the .ini input file or within the MFCL code.

Generally, models were run with a gradient criterion of 10^{-5} and a with the maxium iteration criterion of greater than 10 000. This generally ensures that model runs complete in less than 24 hours, allowing a routine of continuous model development. During the stepwise development, some model runs were later demonstrated to achieve sub-optimal local minima. This was demonstrated when fits with improved objective functions were subsequently obtained by running several replicates where selected parameters were perturbed during the optimisation process, in a process known as jittering. Following these adjustment to parameter values, the optimisation was continued with additional iterations for each replicate jitter. In many cases, the new jittered solution achieved an improved objective function value, and a set of estimated model parameters that produced notable changes to derived management quantities. This was cause for considerable concern and demonstrated that additional work was required to find more accreptable solutions to these parameter estimates that clearly provided suboptimal solutions, and which provide reliable and stable estimates of the derived quantities of most interest for management advice. A thorough jittering process was therefore adopted for the both the diagnostic model and the structural uncertainty grid, exploring the parameter space more thoroughly and selecting the parameter estimates possible to exlpain the data, those parameter values that produced the best likelihood, or best value for the objective function. These improved solutions may sometimes have poor gradients and Hessian properties than models at sub-optimal local minima. Thus, a positive definite Hessian (PDH) solution was considered less important in this process than achieving 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 a PDH, wher the objective function was subsequently improved, in some cases substantially, through jittering.

Additional steps were incorporated in the process adopted to find estimated model parameters with the best possible objective function value, compared to those 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 (also referred to as "twerking") could sometimes be found, but these were often less than 10 units of log likelihood and typically the secondary round of jittering produced little change to the stock status. However the initial round of jittering produced changes to spawning biomass depletion of 5% or more. This iterative 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 Section 6.1). While there is no guarantee that this is the optimal solution, it was clearly a better solution than those obtained without jittering.

The 2020 assessment ran jitters on the diagnostic model but did not find the same instability in relation to changes to the 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 to a model in the structural uncertainty grid, while perhaps indicative, could not really be trusted, reagardless of the objective function value, gradient or Hessian status. While the grid models based on the diagnostic model are generally expected to have good starting parameter values, jittering each model in the grid is now considered important to increase confidence that these models achieve stable solutions, both in terms of an improved objective function value and reliable management quantities. However, under time constraints the jittering of models in the grid was restricted to 20 jitters per model, with the Hessian calculated on the best jittered solution, where "best" is judged solely on the objective function values achieved. The requirement to jitter and run Hessians on grid models again compounded the already heavy workload and computational requirements for this assessment. However, this was all required to meet recommendations of the yellowfin peer review and the SC18 requests for greater attention to 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 the best possible objective function value 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 little influence on the estimations of 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 solutions, one with and one without a PDH, that the estimation of parameter 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 non-influential number of small negative eigenvalues are found and if estimates of the parameter variances can be obtained anyway.

Estimation uncertainty A positive definite Hessian is desirable to calculate parameter or estimation uncertainty for individual models, although as noted, MFCL has advanced procedures for computing estimation uncertainty in the absence of a positive definite Hessian. with negative eigenvalues. 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 for this assessment but this request has added significant time, and computational burden, leading to delays in finishing the assessment and leaving some model explorations unfinished.

Structural uncertainty This 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, absolute estimates of total abundance can be unstable. Therefore, a likelihood profile analysis was undertaken of the marginal penalized likelihood in respect of population scaling, following the procedure outlined by McKechnie et al. (2017a) and Tremblay-Boyer et al. (2017). The results of these procedures are presented in the appendices (Appendices Section 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). Retrospective analyses for the 2023 diagnostic model are presented in the appendices (Appendices Section 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 $SB_{F=0[t]}$, incorporate recruitment variability, their ratio at each quarterly time step (t) of the analysis, $SB_t/SB_{F=0[t]}$, can be interpreted as an index of fishery depletion. The computation of unexploited biomass includes an adjustment in recruitment to acknowledge the possibility of reduction of recruitment in exploited populations through stock-recruitment effects. To achieve this the estimated recruitment deviations are multiplied by a scalar based on the difference in the SRR between the estimated fished and unfished spawning potential estimates.

A similar approach can be used to estimate depletion associated with specific fisheries (fisheries impact analysis) or groups of fisheries. Here, fishery groups of interest - 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 bigeye 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 bigeye tuna stock however CMM 2021-01 states in para 11 "Pending agreement of a target reference point the spawning biomass depletion ratio ($SB/SB_{F=0}$) is to be maintained at or above the average $SB/SB_{F=0}$ for 2012–2015". Stock status was referenced against these points by calculating the reference points; $SB_{recent}/SB_{F=0}$ and $SB_{latest}/SB_{F=0}$ where $SB_{F=0}$ is calculated over 2012–2021

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_{recent}/F_{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 Section 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 (Section 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}) are computed by repeating the yield analysis for each year in turn. This enabled temporal trends in the reference points to be estimated and a consideration of the differences in MSY levels under historical patterns of age-specific exploitation. This analysis is presented in the form of dynamic Kobe plots and "Majuro plots", which have been presented for all stock assessments in recent years.

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.

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 use the data inputs from the 2020 assessment and implementing the main changes to the modelling methods used for the 2023 assessment on this old dataset. This development of the model using old data is necessary as there is insufficient time to complete these analyses using updated data between the finalisation of WCPFC data submissions and the SC meeting. These time constraints made it impossible to rerun the stepwise model development on the updated data inputs. As a result, updating the data is incorporated into the last steps in the development of a new 2023 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 listed to describe each step. Each step builds from the previous step, and in most cases retains all the previous changes. Many of the steps listed in this stepwise model development include a number of sub-steps, which result in negligible changes in model outcomes. Each sub-step incorporates the smallest practical number of changes to the model to allow any changes in model outcomes to be fully attributed. The steps listed here include details on sub-steps, but a step is only named and listed here if it involves important

structural changes, or if there is a notable change in the model outcomes.

1. **[00BETDiag2020]** The 2020 bigeye diagnostic model, using data from 2020.
2. **[01SingleStep]** The 2020 bigeye diagnostic model featured a two stage optimisation, with an additional "FixSelBump" step added in stage two, after all other parameters were estimated in stage one. In 2020, FixSelBump constrained the selectivity for large fish to ensure better fits to length data for a group of miscellaneous fisheries (17, 21, 22, 23, 28 and 32). This additional stage was added just before the 2020 diagnostic model was finalised. At the start of the stepwise in 2023, this change to selectivity was incorporated into a single stage optimisation, with the timing of this modification to selectivity also altered in the sequence when parameters were introduced to the estimation. This selectivity change was moved from the last phase, in 2020, to the first phase of estimation in 2023. As a result, constraints on selectivity for all fisheries are now set in the same phase (Phase 1).
3. **[02NewExe]** This step updates the MULTIFAN-CL (MFCL) executable. This step incorporates a number of changes, removing some redundant MFCL flag settings and setting maturity for very young ages to a very small non-zero value (1×10^{-9}), moving these values away from the bound of zero, as required by the most recent version of MFCL, version 2.2.2.0.
4. **[03PreCatchCond]** Multiple changes are required to the model structure to enable the transition to a catch-conditioned model, and to ensure stable model behaviour. These changes include two initial sub-steps of relaxing a bound on the total mortality level in later phases of the model, and assuming zero initial fishing mortality, prior to the initial year of data input to the model, rather than setting the initial fishing mortality to the average fishing mortality estimated from the first few years of data. The third sub-step involved decreasing the penalty on non-zero regional movement and this substep resulted in notable changes. These three sub-steps were incorporated separately prior to the catch-conditioning steps, to enable the effects of catch-conditioning to be examined in isolation. These steps all involve running catch-errors models without using index fisheries and their associated CPUE likelihood.
5. **[04aCatchCondOldCPUE]** This step implements an initial catch-conditioned model, by removing many of the catch-errors specific MFCL flags, introducing grouped index fisheries with common catchability but retaining the old-style catch-errors CPUE likelihood, through estimating the relationship between fishing mortality level and effort. This step represents a major change in the model structure.
6. **[04bCatchCondNewCPUE]** This step takes the initial catch-conditioned model, with grouped catchability for the index fisheries, but no longer estimates the relationship between fishing mortality level and effort. Instead it incorporates the CPUE likelihood component from the index fisheries and again is a significant change to the model structure.

7. **[05SelChanges]** This step incorporates a number of sub-steps which have negligible impact including: adding the real effort data for the last 12 quarters for some extraction fisheries to enable effort projections for these fisheries at a later step; relaxing a constraint on estimation of recruitment in region 9; setting the period for maximum sustainable yield (MSY) calculations to the last 212 quarters; adjusting selectivity (forcing it to zero) for the youngest age classes for some longline fisheries; and adjusting selectivity (forcing it to zero) for the five youngest age classes for fishery 18.HL.PHID.7. The important change in this step involved adjusting the choice of fisheries which were constrained to have non-decreasing selectivity. In the 2020 bigeye diagnostic model, all the index fisheries and all of the longline fisheries were constrained to have non-decreasing selectivity, to prevent the model from having dome-shaped selectivity for all fisheries which could allow a large cryptic biomass to be estimated. An intermediate sub-step was initially examined where only the nine index fisheries were constrained to have non-decreasing selectivity (relaxing this constraint for all longline fisheries). After further examination of the length data, a decision was made to apply this non-decreasing selectivity only to a single longline extraction fishery, 6.LL.OS.7, a fishery which has some of the largest fish and the best representation of large fish in the weight composition data. Unsurprisingly, this relaxation allowed for considerable improvements in the objective function.

8. **[07Growth]** Once again, this step includes two sub-steps with negligible changes, both relating to tags. The first sub-step involved relaxing the upper bound on estimated tag group reporting rates from 0.9 to 0.99 and the second sub-step excludes any tag release groups with five or fewer recaptures from the set of groups for which tag reporting rates are estimated. An attempt after this sub-step to estimate the tag overdispersion parameter τ, in a separate exploration resulted in an estimate of τ on the lower bound. This suggests that it is not possible to get a plausible estimate of τ, so it was fixed at a value of 2 which represents a moderate level of overdispersion relative to the Poisson distribution. The important change in this step was the final sub-step, where all the von Bertalanffy growth parameters were estimated, in contrast to the 2020 diagnostic model which used fixed growth parameters (estimated externally) and only estimated two variances parameters for growth. This sub-step required inclusion of conditional age-at-length data (from a total of 1004 otolith readings) and estimation of growth within the model, so this is a major structural change using additional data. As a result, the total likelihood is not comparable to the previous step. The starting values for growth parameters can be important for complex models when there are sensitivities to initial conditions, and sometimes optimisation algorithms can have difficulty exploring the parameter space to find the optimal solution. This sensitivity to initial conditions when trying to estimate growth was a problem identified in the 2020 assessment (Ducharme-Barth et al., 2020a). As part of this sub-step, the starting values for the growth parameters were updated to set them closer to the estimated values, in an iterative process, to improve the

chances of the model finding a "good" solution and avoid being stuck in a section of the parameter space with a sub-optimal local minimum. Jittering analysis showed that more complex models, discussed later in this stepwise development, are particularly sensitive to these starting values, and starting close to the estimated optimal values improved model performance and stability. Once this instability was identified, and to increase the chance of finding a "good" solution at each step of the stepwise, the starting growth parameter values were reset at this step, and for all successive steps, and all subsequent steps rerun with these new starting values for the growth parameters.

9. **[08DataWeights]** Again this step involved an exploratory sub-step that was not adopted in the full stepwise. In this case, the Dirichlet multinomial approach to estimate a scalar to weight the length and weight composition data was explored. This approach suggested that the composition data should be upweighted, but this was at the expense of deterioration in the fits to the CPUE data. While it would be useful to adopt a self scaling approach to weighting this composition data, the consequences to the fits to the CPUE were considered to be unacceptable. The 2022 yellowfin peer review (Punt et al., 2023) recommended that the age data was previously overweighted in the 2020 yellowfin assessment. Due to computational and time constraints, it was not possible to consider an iterative reweighting of either the size composition data (Francis, 2011) or of the conditional age-at-length data Punt (2015) in this assessment, but further exploration of this approach could prove useful in future assessments. Alternative data weightings were included in the uncertainty grid, with diagnostic model choices made for the default divisors of 20 for the composition data (unchanged from 2020), based partly on initial results from exploring the Dirichlet multinomial approach, and 0.75 for the conditional age-at-length data (changed from the value of 1.0 which was used for the yellowfin assessment in 2020 where conditional age-at-length data was used with growth estimated internally in that assessment). The sub-step including changes to the weighting of the conditional age-at-length data produced minimal change to the final stock status and the changes to the likelihood are not comparable. The important change in this step involved allowing different values for the input CPUE variance for each index fishery, and hence for each region. Note that this feature was not available in the MFCL version used in 2020. The value of the input variance for each index fishery is set by calculating the maximum likelihood estimate of the variance of the observed CPUE data points compared to a modelled (expected) value from a CPUE fit from a previous step in the stepwise development, fitting to the log-residuals. This allows this variance to be estimated from the input data and the fit to a similar model, rather than using an estimate of variance obtained directly from CPUE analysis, which can often be unrealistically small due to the large samples sizes involved, and which may not adequately represent the process error observed. Again, the total likelihood is not comparable to the previous step.

10. **[09NatMort]** The first sub-step used the same shape for natural mortality as used in the 2020 diagnostic model, but estimated the scale for this mortality form. This sub-step resulted in minimal change. The second sub-step used the Lorenzen form of natural mortality, again with the scale estimated. This form features a much higher mortality rate for young fish than was used in the 2020 diagnostic model, and also features monotonically decreasing mortality with age. The mortality form used for the 2020 diagnostic model featured an increase in mortality at around 3 years of age that was used to model observed changes in sex-ratio (and implied changes in mortality by sex) for older bigeye. The use of Lorenzen mortality and estimating the scale, if possible, was recommended as good practice in the recent CAPAM tuna stock assessment good practices workshop held in March 2023, https://www.capamresearch.org/Tuna_Stock_Assessment_Good_Practices_Workshop.

11. **[10TaggerEffect]** 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 old (2020) tagging data. This revised model uses different assumptions to the model used in 2020, reverting to a model with assumptions similar to those used in 2017. This update to the method results in changes to the 2020 input data, so the objective functions can not be meaningfully compared.

12. **[11NewCPUEMethod]** The updated CPUE method using sdmTMB with new covariates (Teears et al., 2023) was applied to the 2020 dataset. With different data inputs, the objective function is not comparable with the previous step.

13. **[12aLengthWeight]** This step incorporates the updated length-weight conversion factors (Macdonald et al., 2023b).

14. **[12bNewIniM]** Like the growth parameters, the starting values for the mortality scale parameter can be important for complex models that have sensitivities to initial conditions. In this step, the starting value for the mortality scalar was updated, and in all following steps, with values subsequently reset closer to estimated values from a later step, again through an iterative process, to improve the chances of the model finding a "good" solution.

15. **[12cNewData]** This step incorporates all the new data and historical data revisions up to the end of 2021, applying all the previous structural changes to the model. This included revisions to data from 1952-2018, and 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, new CPUE data and applies the updated tagger effects model to the new tagging data. There is no new conditional age-at-length data available since the 2020 assessment. There are also an additional 3 years of recruitment deviations estimated as part

of this step. Ideally, all of these changes would be separated and incorporated one at a time, but this was not possible due to the limited time between receiving the finalised new data, with associated quality checks and exploration of various filtering options, and producing the assessment report. The unjittered comparable final stock status (at the start of 2019) was very similar to both the previous step and the 2020 diagnostic case, initially suggesting that further separation of this step was not critical. The objective function is not comparable between this step and the previous step because all components of the data have changed, with the single exception of the conditional age-at-length component.

16. **[12dF17_24ungp]** With the new data first included in the previous step, length composition data became available for fishery 24.PS.PHID.7 for the first time. Previously this fishery had been grouped with fishery 17.MISC.PH.7 for selectivity, as it appeared to be the most similar fishery. New composition data makes it possible to ungroup these two fisheries and estimate selectivity separately for each fishery.

17. **[13a0TailCompress]** Tail compression is commonly applied in stock assessments to avoid problems with zero observations for either end of the size composition data (i.e. for very large or very small fish). Tail compression is often used to aggregate a small fraction of the total distribution into one size bin, at either end of the distribution, to minimise the number of zero observations and to avoid problems with likelihood calculations and was recommended by the 2012 bigeye peer review (Ianelli et al., 2012), and subsequently introduced to MFCL in 2015 (Davies et al., 2015). Tail compression was not applied in the 2020 bigeye assessment, so all zero "observations" were included in the likelihood calculation. Using 2 cm length bins in the range 10 cm to 200 cm and 1 kg weight bins from 0 kg to 200 kg resulted in many quarterly size composition records having long zero tails, especially on the upper end of the distributions. Without compression of these tails, these zero "observations" are compared to the expected values with each size bin contributing to the likelihood. While it is preferable to also include a small non-zero fraction, such as 0.001, in the tail compression, as a first step to dealing with tail compression, zero tail compression was applied to both the observations and to the corresponding expected values matching each zero tail compressed observation, to prevent these strings of zeros from contributing to the likelihood. This step had a large effect on the likelihood values for the composition data, but due to the effective "change in data" these likelihood values are not comparable.

18. **[13bMinSamp50]** This step utilises an MFCL filter on input sample sizes for length and weight composition data, excluding all samples with an input sample size of 50 or fewer. As with tail compression, it is common practice to exclude small samples from size composition data, as they are often unrepresentative of the catches of population and this was also introduced to MFCL in 2015 (Davies et al., 2015). As with tail compression, the likelihood

components from the composition data are not comparable, so the objective function cannot be compared with the previous step.

19. [**13cNewInitVal**] An early version of the previous step was jittered, to see if a better solution could be found. As a result of these jitters and subsequent analysis, initial values for growth and mortality were updated to try to improve the solution found.

20. [**14Diagnostic2023**] This final step involves jittering the previous model, with updated initial values for mortality and growth, followed by a further round of jittering.

6.2 Sensitivity analyses and structural uncertainty

6.2.1 Sensitivities

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. Due to the time constraints and additional requirements to run hundreds of models to achieve a well fit and reliable diagnostic model, plus the need run jitter analyses on all grid models, there was limited time to explore sensitivities. We note that additional models were explored with alternative size data filtering ((Peatman et al., 2023b)), SEAPODYM movement coefficients, and simpler spatial structure but we did not have sufficient time to fully explore these models with additional steps required to achieve a reliable and stable convergence as described in Section 5.6.

One additional sensitivity was conducted regarding purse seine catches. Estimates of purse seine catches rely on statistical analysis of species composition data provided by the regional observer programme. As a result of COVID restrictions, purse seine observer coverage was very low in 2020 and 2021 compared to previous years (Peatman et al., 2023b). This affected the precision of estimates of species composition, and therefore there is greater uncertainty in purse seine catches of bigeye in these years than previously. To test the impact of any substantial under-estimation of bigeye catch in 2020 and 2021, we undertook a one-off sensitivity in which the purse seine catches for both associated and unassociated fisheries in regions 3, 4 and 8 were doubled from their estimated values.

We primarily focused on the key sensitivities that form the structural uncertainty grid.

1. Steepness: 0.65, 0.95
2. Tag mixing: 1 quarter
3. Size data weighting: 10, 40
4. Conditional age-at-length data weighting: 0.50, 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 bigeye 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 quarter).

The values for the diagnostic model are in bold and the alternative values used in the grid are in italics.

1. Steepness [*0.65*, **0.8**, *0.95*]
2. Tag mixing [*1 quarter*, **2 quarter**]
3. Size data weighting [*10*, **20**, *40*]
4. Conditional age-at-length data weighting [*0.50*, **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 the 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 10, 25, 75 and 90 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 small number of 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 small number (maximum of 1 in the case of bigeye) of 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 Section 6.1 and the results are displayed in Figure 19 and Figure 20. Traditionally, the stepwise analysis is presented simply by running MFCL at each step, plotting the results and attributing the change in results to changes in the modelling, the data or structural changes involved. Through the process of building this stepwise development, it became apparent that model convergence is a significant issue for these complex models for which thousands of parameters are estimated, and jittering is an important process to refine and improve the best solution found from any of these models. As these issues with model convergence were not well understood at the start of this process, and due to constraints in both time and computational resources, we were unable to conduct a jitter analyses for each step in the stepwise development. Such analysis was completed for the diagnostic model, the final step in the stepwise development, and for all of the models in the uncertainty grid, but there was not sufficient time to return to the stepwise and jitter each step. As a result, the successive estimations of spawning potential SB and dynamic spawning potential depletion $SB/SB_{F=0}$ in the stepwise development should only be considered indicative of potential changes. Ideally, all steps would be jittered to be confident that any changes are a result of the model changes and not a result of variable levels of convergence. We also note the stepwise development of the 2023 diagnostic model involved running a large number of models, with running times typically ranging between 12 and 48 hours. In addition, we explored a range of additional models which were deemed unsuitable for a variety of reasons. It is impractical to detail the results

of all of these investigations. A total of 19 significant steps are described here, with steps which involve either notable changes in the model structure or changes in the model outcomes. These steps incorporate an additional 16 potentially important sub-steps that were ultimately included in the 19 major steps (Section 6.1). A brief description of the consequences of this progression through the major steps is provided below, focusing on the key management quantities of spawning potential SB and dynamic spawning potential depletion $SB/SB_{F=0}$. As noted, care needs to be taken in drawing definitive conclusions from these stepwise changes in the absence of jittering.

1. [**00BETDiag2020**] The 2020 bigeye diagnostic model, using data from 2020.
2. [**01SingleStep**] This step produces a small decrease in both $SB/SB_{F=0}$ and SB and a minor improvement to the objective function.
3. [**02NewExe**] This step results in no change in $SB/SB_{F=0}$ or SB, and a minor deterioration in the objective function.
4. [**03PreCatchCond**] This step results in a small decrease in $SB/SB_{F=0}$ and SB, and negligible change to the objective function.
5. [**04aCatchCondOldCPUE**] This step represents a change in the model structure, using index fisheries, and is an intermediate step towards a fully catch-conditioned model, resulting in a small decrease in $SB/SB_{F=0}$ and SB, with a change in shape for the $SB/SB_{F=0}$ time series and in the unfished biomass time series. The objective function for this intermediate step is not comparable to the objective functions from either the previous or subsequent steps.
6. [**04bCatchCondNewCPUE**] This step represents a major change to the model structure incorporating CPUE likelihood from the index fisheries and a reduction from 11,421 to 2,950 parameters estimated, compared to the 03PreCatchCond step. There is another small decrease in $SB/SB_{F=0}$ and SB, contributing to a moderate change from the previous catch-errors model (03PreCatchgCond). Once again, changes to the objective function are not meaningful given the structural changes to the model.
7. [**05SelChanges**] This step resulted in considerable improvements in the objective function, mostly to the weight component of the likelihood but also to the penalties, and also a considerable increase to both $SB/SB_{F=0}$ and SB, largely reversing the incremental decreases in $SB/SB_{F=0}$ from earlier steps.
8. [**07Growth**] The changes in this step include the use of additional conditional age-at-length data, so comparison of the objective function with the previous step is not meaningful. As a result, the total likelihood is not comparable to the previous step, but estimation of growth parameters saw improvements to the CPUE, length composition data and tag components of the likelihood and a deterioration to the weight component of the likelihood. There is

also another increase in both $SB/SB_{F=0}$ and SB at this step, similar in magnitude to the increases seen in the previous step. This step features the second largest values for $SB/SB_{F=0}$ in Figure 20.

9. [**08DataWeights**] Once again, the total likelihood is not comparable to the previous step, but unsurprisingly there were improvements in the CPUE likelihood component. There is also a further increase in both $SB/SB_{F=0}$ and SB although smaller than the increases in the previous two steps. The increases in this step and the previous two steps are shown in successively lighter shades of green in Figure 19 and Figure 20, with each of these steps reflecting a shift upwards in these plots. This step features the largest values for $SB/SB_{F=0}$ in Figure 20, and the second largest values for SB in Figure 19.

10. [**09NatMort**] This step resulted in a deterioration in the objective function, but with improvements in the tag likelihood component. There is also a decrease in both $SB/SB_{F=0}$ and SB, shown as the palest shade of green in Figure 19 and Figure 20.

11. [**10TaggerEffect**] This update to the tagger effects model results in changes to the 2020 input data, so the objective function cannot be meaningfully compared. This step results in a deterioration in the objective function and also produces a decrease in $SB/SB_{F=0}$ but with a corresponding increase in SB, resulting in a similar final $SB/SB_{F=0}$ outcome to that produced by the 2020 diagnostic model. Changes in the tagger effects model resulted in a reduction in the effective number of tag releases seen by the model due to the estimation of larger tagging effects.

12. [**11NewCPUEMethod**] With different data inputs, once again the objective function is not comparable with the previous step. This step produces a small increase in $SB/SB_{F=0}$ and a larger increase in SB. This step features the largest values for SB in Figure 19.

13. [**12aLengthWeight**] The objective function deteriorated in this step, largely due to the weight composition component, but likelihood values cannot be used to reject these updated conversion factors. There was also a decrease in both $SB/SB_{F=0}$ and SB.

14. [**12bNewIniM**] Jittering analysis showed later stepwise models to be particularly sensitive to these starting values, and starting close to the estimated optimal values improved model performance. This step had minimal impact on $SB/SB_{F=0}$ and SB. However the objective function improved, largely through improvements to the weight component of the likelihood, although in contrast, the tag and length composition components of the likelihood deteriorated slightly. This indicates some instability in the model and was an early indication that jittering would be an important part of this stepwise development.

15. [**12cNewData**] The unjittered comparable final values of $SB/SB_{F=0}$ (at the start of 2019) was very similar to both the previous step and the 2020 diagnostic case, initially suggesting

that further separation of this step was not critical. The objective function is not comparable between this step and the previous step because all components of the data have changed, with the single exception of the conditional age-at-length component. Considering only the steps which incorporate new data (i.e. those with data 2019–2021), this step has the largest values for $SB/SB_{F=0}$ in Figure 20 and likewise the largest values for SB in Figure 19.

16. **[12dF17_24ungp]** This step had a large effect on the likelihood values for the composition data, but due to the effective "change in data" these likelihood values are not comparable. The fits to both the age and CPUE components of the likelihood were improved considerably as well as the penalties, with a smaller deterioration in the fits to the tag component of the likelihood. The estimation of separate selectivity for fisheries 24.PS.PHID.7 and 17.MISC.PH.7 is desirable due to the clear differences in the distribution of the length composition data for these two fisheries, and led to improvements to the objective function and in particular to the length composition component of the likelihood. This improvement is also clear when simply looking at aggregated fits to the length composition data for these two fisheries. This step produces a considerable apparent decrease in $SB/SB_{F=0}$ and SB. Further investigation of this somewhat surprising difference, by jittering the previous step, 12cNewData, indicated convergence issues with that model, which suggest that the apparent changes resulting from minor changes to selectivity actually occurred in previous steps. The jittered version of 12cNewData produced very similar results in terms of $SB/SB_{F=0}$ and SB to the results from this step, albeit without also jittering this step. Considering only the steps which incorporate new data (i.e. those with data from 2019–2021), this step has the smallest values for $SB/SB_{F=0}$ in Figure 20 and likewise the smallest values for SB in Figure 19.

17. **[13a0TailCompress]** With zero tail compression, the zero size frequency observations in each tail are excluded, so that the observed size frequencies in each tail are compressed to the smallest, and largest, interval having a non-zero observation. Accordingly, the corresponding predicted sample composition is also compressed according to the corresponding minimum and maximum size class intervals of the tails of the compressed observed distribution. This approach excludes a large number zeros in the tails. This step had a large effect on the likelihood values for the composition data, but due to the effective "change in data" these likelihood values are not comparable. The fits to both the age and CPUE components of the likelihood were improved considerably, as well as the penalties, with a smaller deterioration in the fits to the tag component of the likelihood. This resulted in an increase in both $SB/SB_{F=0}$ and SB.

18. **[13bMinSamp50]** As with tail compression, the likelihood components from the composition data are not comparable, so the overall objective function cannot be compared with the previous step. There is a small improvement in the tag data component of the likelihood,

with a slightly larger deterioration in the fit to the age component at this step. This resulted in a decrease in both $SB/SB_{F=0}$ and SB.

19. [**13cNewInitVal**] An early version of the previous step was jittered, to see if a better solution could be found by modifying key parameter values in an additional phase of estimation. Out of 60 jitters conducted, 48 models produced a better objective function, with improvements in the objective function of up to 141 likelihood units. These improved fits also showed a considerable decrease in the estimated final $SB/SB_{F=0}$, indicating that the unjittered model had not converged to a reliable stable solution. As a result, initial parameter values for growth were reset to values close to the estimated values at this stage, with changes made to the starting values, and the stepwise development repeated for all steps from 07Growth onwards. As already described, the initial value of the scalar for natural mortality had already been updated from step 12bNewIniM. With these new starting values, the requirement to jitter this step to ensure better convergence was reduced. To ensure continued stability the starting values for natural mortality and growth parameters were again reset at 13cNewInitVal, using updated starting values close to the values estimated from 13bMinSamp50. The objective function obtained by updating these starting values shows an improvement of 4.8 likelihood units over the best solution obtained from jittering 13bMinSamp50 with alternative starting values for the growth and natural mortality parameters (even with 60 separate jitters), highlighting the general sensitivity to initial conditions for this model, and the highly complex likelihood surface. Unlike the large changes seen in $SB/SB_{F=0}$ and SB at 13cNewInitVal in a preliminary version of this stepwise development, resetting some initial parameter values in earlier steps, resulted in small changes at 13cNewInitVal in this final stepwise development and resulted in a considerable improvement to the overall objective function, largely through improvements to the weight composition component of the likelihood, with a small improvement in the CPUE component as well. There was a very small decrease in the estimated $SB/SB_{F=0}$ and SB at this step.

20. [**14Diagnostic2023**] The 2023 diagnostic model was obtained by running a final round of 60 jitters on the previous step. Out of 60 new jittered model runs, 10 runs produced a better solution than the previous step, with the best jittered solution having a maximum improvement of 8.87 likelihood units. Notably, there were no changes to $SB/SB_{F=0}$ and SB, for any of these 10 jittered solutions with improved objective functions, or for the model that these jittered runs were derived from in the previous step (13cNewInitVal).

The 2020 and 2023 diagnostic models estimate similar values for $SB/SB_{F=0}$ for the first half of the estimated time series, with the trajectories starting to diverge around 1985, with the 2023 stock assessment estimating a value for $SB/SB_{F=0}$ between 5-10 percentage points lower than the 2020 diagnostic model from 1990 onwards.

Overall, the 2023 diagnostic model estimated the recent $SB/SB_{F=0}$ at 34.2% to 2021 compared to 42.3% for the 2020 diagnostic model to 2018 (Figure 20). Note that the trajectory of $SB/SB_{F=0}$ is fairly flat over the period 2018–2021, so the estimated value of $SB/SB_{F=0}$ in 2021 for the 2023 diagnostic model is similar to the value in 2018.

From 1960 onwards, the 2023 diagnostic model estimated a slightly increased spawning potential SB compared to the 2020 diagnostic model, with a very similar estimated spawning potential prior to that (Figure 20). Both the 2020 and 2023 diagnostic models indicated a declining trend in spawning potential since around 1960, with a relatively steep decline until the mid 1970s and then a gradual decline through to 2010 with a relatively stable trend since 2010.

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 restricting the constraint of non-decreasing selectivity to a single fishery and estimating growth internally, in terms of increasing the estimate of final value of $SB/SB_{F=0}$ and applying the revised tagger effects method, in terms of decreasing the estimate of the final value of $SB/SB_{F=0}$ (Figure 19, Figure 20). However, these indicative influential steps should be treated with considerable caution, and to be definitive about the most influential steps, the full stepwise development would need to be appropriately jittered.

7.2 Fit of the diagnostic model to data sources

7.2.1 Standardized CPUE: index fisheries

The model fits to the index fishery standardised CPUE data were generally very good for all model regions (Figure 21, Figure 22). The model was able to predict the longer term trends and short-term variation at the sub-decadal scale. The residual plots (Figure 23, Figure 24) do not show any clear periods of bias or trends in the patterns of the residuals, although there tends to be a greater spread in the residuals earlier in the time series associated with the higher CPUE levels. There are a number of positive residuals of larger magnitude around the late 1990s and early 2000s for regions 8 and 9. However, region 8 is relatively small and region 9 is tiny, in relation to the spatiotemporal standardisation method and there is imprecision around that period (Figure 12).

7.2.2 Size composition data

This discussion is focused on aggregated fits to the weight and length composition, aggregated over all time periods, rather than the individual quarterly based size composition fits that contribute directly to the likelihood. Supplementary information to show length and weight composition fits for each quarter will hopefully be provided soon via a link on the WCPFC website, including bubble plots indicating the residuals for each length and weight category for all time periods for

each fishery in a single plot, in addition to individual plots for each quarter showing fits to the size composition data for every fishery.

Longline fisheries: The aggregate model fits to the weight composition data for the longline extraction fisheries (fisheries 1-29, Figure 26), were good for all fisheries, and most importantly for those fisheries accounting for the majority of catches, e.g., fisheries 9, 2, 1, 4 and 6, listed in order of the size of the catch from each fishery, (Figure 7), and also noting the variable y-axis scale in these figures, where the absolute size can be used to indicate the relative size of composition samples for each fishery in these figures. Fisheries 1, 7, 8, and 10 show some multi-model structure in the aggregated weight composition, particularly with some peaks at smaller weights that the model could not closely fit, although the relative sample size in fisheries 7, 8 and 10 is comparatively low. Fisheries 1, 7 and 8 occur in the western model region, that encompass the Philippines, Indonesia, Vietnam, Papua New Guinea and Solomon Islands where smaller bigeye are expected to be more abundant.

The index fisheries all have a common shared selectivity, and while the fits are still relatively good, they are not quite as good as the fits for the extraction fisheries, with a slight over-prediction of larger fish for fisheries 33 (region 1), 34 (region 2) and 40 (region 8), and an under prediction of small fish in fisheries 33 (region 1) and 39 (region 7).

Other fisheries: The aggregated length composition fits are also quite good for most of the non-longline fisheries, with the exception of fisheries 19 and 20 (Figure 25). 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 bigeye, so the inability to fit those compositions is not so important compared to other fisheries with larger catches. There is a considerable improvement in the length fits to fishery 17 (a miscellaneous fishery in the Philippines), compared to similar fits from the 2020 assessment, and similarly for fishery 22 (pole and line in region 8).

7.2.3 Tagging data

When aggregated, the model tag attrition estimates fit the observed tagging data relatively well (Figure 29), albeit overestimating the number of tag returns after 2 quarters at liberty, and underestimating the returns in periods 3 and 4. When compared at the tag program scale, there are some differences in the quality of fit (Figure 30). 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 bigeye, and this series is difficult to fit well, with a pattern that appears to show some seasonal cycles. The fit to the JPTP program is relatively good, especially given the increase in the number of tag returns between periods 2 and 4 seen in the data. At the regional scale, the fits are good for those regions with the most recaptures,

respectively regions 4, 3 and 8, with adequate fits for region 1 (Figure 31). The observed return data for region 9 seem to show a strong seasonal pattern, making these data difficult for the model to fit well, given there is no seasonal component in the modelled tag recaptures.

For the tag returns by quarter of recapture, aggregated over all programs, 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 32). The number of tag returns is low for most longline fisheries, so the fits to this patchy data are of minor consequence to the likelihood (Figure 33). The one exception to the small number of tag returns in the longline fisheries is fishery 27, the Australian longline fishery in the Coral Sea, whch has a noticeably higher number of tag recaptures than any other longline fishery. Here, tag recaptures seem to be influenced by fish returning regularly and seasonally to a potential spawning area, producing a noisy data series with some notable spikes, which is difficult to fit well.

The fits to the fisheries other than longline are good for those fisheries with the most observed tag returns, namely fisheries 13, 14, 15 and 16 (purse seine in regions 3 and 4). The fits are acceptable, albeit with some spikes missed, for fisheries 25 and 26 (purse seine in region 8) and reasonable for the other fisheries with small numbers of tag returns (Figure 34).

7.2.4 Conditional age-at-length

The available conditional age-at-length data is restricted to only 1004 otoliths in the 2023 assessment and hence has limited spatial and temporal coverage. There is a critical need to age existing otoliths and to collect additional ongoing conditional age-at-length data as a part of a structured spatial and temporal sampling plan in order address data gaps (e.g., lack of samples from the north central Pacific ocean) and help resolve issues with estimating growth in future assessments.

Aggregated fits to the conditional age-at-length data are shown in Figure 16. For most length bins, the variance around the fitted growth curve covers the mean and the distribution of ages adequately, noting that the conditional age-at-length data are not the only data source contributing to the estimation of growth in this integrated assessment.

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. The age-specific selectivity is shown in Figure 36 and weight-specific selectivity is shown in Figure 37.

Fishery 6 (OS in region 7) is the only fishery with a penalty imposed on non-decreasing selectivity and as a result this is the only fishery with an asymptote at full selectivity for the oldest fish.

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, with the value of this asymptote varying from fishery to fishery. The selectivity by weight shows similar patterns, but expressed in weight rather than age, with fishery 6 reaching full selectivity at around 60 kg and the other fisheries typically achieving maximum selectivity at around 30–40 kg.

Of the non-longline fisheries, only fishery 18 (handline in the Philippines and Indonesia) is estimated to have non-decreasing selectivity, with the asymptotic selectivity approaching full selectivity for the oldest age and length classes, with zero selectivity for the first 5 quarters, or up to around 70 cm, and relatively low selectivity between quarters 5 and 15, or around 75–125 cm length. All other non-longline fisheries have estimated selectivity which is high for young fish, with full selectivity achieved before 5 quarters or about 60 cm in length and then declining to zero, mostly somewhere between 5 and 10 quarters of age, or equivalently around 100 cm in length. Some older and large fish are caught in purse seine fisheries, especially unassociated, in regions 3, 4 and 8 and by Japanese pole and line fisheries in region 1. Fisheries in region 7 often catch younger and smaller fish, with the exception of fishery 18 as noted.

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 40 and in Figure 42. The tag return data shows generally low movement between most regions, with notable movement, as a proportion of the number of tags returned from each region of release, recorded from region 8 to 3, from region 5 to region 8, especially in quarter 1, and some movement from region 4 to region 3. Movement between regions estimated by the model most closely matches the observed movement from region 8 to region 3, especially in quarters 1 and 2. This is also the most significant proportional movement between any two regions (Figure 43). Small amounts of movement are estimated from regions 3 to 8, from 3 to 4, and from 7 to 3, albeit with the latter estimated from very low numbers of tags released in region 7. There are very small estimates of movement from region 8 to 7, with otherwise very limited movement estimated between regions. While region 9 represents only a very small area and a small proportion of the total population, the proportion of movement estimated from region 9 to region 5 is similar to that from region 3 to region 8, but is much less than the estimated proportion of movement from region 8 to region 3.

Figure 48 shows the proportional region of estimated recruitment of the equilibrium bigeye tuna biomass for each region. This suggests that the majority of estimated recruitment is local for two of the temperate regions (regions 1 and 6) and these same temperate regions (regions 1 and 6) provide the largest proportion of estimated recruits for adjacent temperate regions (regions 2

and 5 respectively), albeit with significant local contributions within these regions as well. The recruitment in region 9 is estimated to be entirely local. There is more movement for the tropical regions (3, 4, 7 and 8). However, the majority of estimated recruitment is local in region 7, and this region also provides the largest proportion of estimated recruits to all the other tropical regions, regions 3, 4 and 8. 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 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 natural mortality with the scale fixed. The comparison between the estimated natural mortality at age in the 2023 diagnostic model and the fixed natural mortality at age is shown in Figure 18. Compared to the fixed mortality used in the 2020 diagnostic model, the Lorenzen form of natural mortality features considerably higher levels of natural mortality at age for very young fish, and a slightly lower asymptote for mortality at age for the oldest age classes.

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 17) 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 44. 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. This removed 21 out of 41 tag reporting rate groups, and an additional three 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 17 tag reporting rate groups where the reporting rates were estimated. Three of these had estimated tag reporting group rates on the upper bound (0.99), RTTP tags captured by the Australian longline fishery, PTTP tags captured by domestic fisheries in Indonesia and the Philippines and JPTP tags captured by the Japanese pole-and-line fishery in region 1. There is a case that tag reporting is

genuinely high in the Australian longline fishery and for JPTP tags recaptured in the pole-and-line fishery close to Japan because of good industry cooperation. In the case of the Indonesian and Philippines domestic fisheries, the estimated reporting rate is likely to be positively biased, but this involves relatively low bigeye tuna catches by these fisheries and just 130 tag recaptures, therefore the over-estimated reporting rate is not likely to be consequential.

7.3.6 Growth

Growth was estimated in the 2023 diagnostic model using a von Bertalannfy growth form, estimating the three standard parameters and two variance parameters. The difference in the internally estimated 2023 diagnostic model growth curve and the fixed (externally estimated) 2020 diagnostic model growth curve are shown in Figure 16.

7.4 Stock assessment results

7.4.1 Recruitment: diagnostic model

While the estimated total annual recruitment across all regions (Figure 45) shows considerable inter-annual variation, the trend in recruitment is stable since 1960. The high recruitment estimated in the first few years (1952-1958) may be a result of the assumptions made on initial fishing mortality prior to 1952. To implement a catch conditioned model, the 2023 diagnostic model assumes that there is no fishing mortality prior to 1952. This estimated high early recruitment may simply allow the model to settle into a more stable age and size structure. The estimates of high early recruitment on quantities of management interest will have no effect on the model after 1980, once these early cohorts have passed through the population. As in the 2020 diagnostic model assessment, the final 6 quarterly estimates of recruitment at the aggregated WCPO scale are constrained to equal the geometric mean of recruitment over the entire assessment period (Rice et al., 2014). This has little impact on the spawning potential or other reference points as recruits from these most recent quarters have not yet entered the spawning biomass. Overall, the 2023 model estimates a smaller average annual recruitment than the 2020 diagnostic model, at around 120 million fish per year compared to a value closer 200 million, although with changes in mortality, comparison of absolute recruitment numbers is not that meaningful.

Figure 47 shows the estimated recruitment by region and quarter, averaged over all time periods, indicating some seasonal variation in recruitment in the temperate regions 1 and 2 and also shows the relative contribution to recruitment from each region. At the regional scale, the patterns in recruitment in the 2023 diagnostic model (Figure 46) are largely similar to those seen in the 2020 diagnostic model, albeit at a reduced scale, although some patterns have changed a little. As an example, the 2020 diagnostic model features a spike in recruitment very close to the start for many of the regional recruitment time series, and this good early recruitment seems to be spread over

more years in the 2023 diagnostic model. Recruitment in region 4 featured an early spike around 1960, which is no longer present at all in the 2023 diagnostic model. The 2023 diagnostic model features a recruitment outlier, a late spike in estimated recruitment in 2020, which is more than twice the value of any other estimated annual recruitment from this time series in region 4. Further, this occurs in a year for which recruitment is only estimated in two of the four quarters. In contrast the 2020 diagnostic model featured a series of good recruitment events for each of the last five years in which recruitment was estimated in region 4. The relative changes in regional recruitment are also shown in Figure 49, which highlights the good overall recruitment at the start of the time series, a single year in the late 1970s with notably poor recruitment, and the spike in recruitment in 2021 for regions 4 and 7. Figure 49 also highlights changes in the overall recruitment over time and shows that in some years, there can be considerable change to the relative distribution of recruitment among regions, without changing the overall recruitment level. Recruitment in region 3 was estimated to be strong from 2004-2013, but appears to have dropped considerably since then (Figure 46, Figure 49). Recruitment in region 2 is estimated to be strong prior to 1960, to fit the strongly declining CPUE index in that region when fishing mortality is not sufficient to explain this decline. This contributes to estimates showing region 2 initially comprising a large relative proportion of the spawning potential (Figure 49), with this proportion moderating over time as the recruitment in region 2 also reduces, relative to the early years.

The fit to the stock recruitment relationship for the 2023 diagnostic model is shown in Figure 50.

7.4.2 Biomass: diagnostic model

The 2023 diagnostic model estimates an initial decline in both the total biomass and spawning potential, from the late 1950s through until the mid 1970s, followed by a more gradual decline through to the present (Figure 49, Figure 51).

The pattern in the decline in spawning potential over time, which is consistent with the CPUE indices, is generally similar in all regions, with the early decline through to the mid 1970s perhaps a little more noticeable in regions 1 and 2 than in the other regions (Figure 51).

7.4.3 Depletion: diagnostic model

The 2023 diagnostic model shows an initial gradual decline in spawning potential depletion, $SB/SB_{F=0}$ through to the mid 1970s, followed by a faster decline through to about 2010, at a level of around 0.34, followed by a 10 year period of stability to 2021 (Figure 52). This pattern varies regionally, with the lowest values for $SB/SB_{F=0}$ in the equatorial regions (3, 4, 7 and 8), with $SB/SB_{F=0}$ approaching values close to 0.2 for all four of these equatorial regions from around 2010 onwards. In contrast, in the temperate regions (1, 2, 5, 6 and 9) $SB/SB_{F=0}$ is higher than 0.6 for the entire time series

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

To fully interpret the trends in spawning depletion it is important to compare individual trends in spawning potential, SB_t, with the predicted spawning potential in the absence of fishing (unfished) $SB_{F=0}$, at both the spatially aggregated level and for each region separately (Figure 53). At the spatially aggregated level $SB_{F=0}$ declines from the late 1950s to the mid 1970s and then gradually increases, before stabilising from about 2005 onward, at a level a little lower than in the late 1950s. This pattern is not repeated at the regional level, where typically the equatorial regions feature the same general overall shape as the spatially aggregated trajectory, but they finish/stabilise with a value of $SB_{F=0}$ higher than the level in the late 1950s, with the divergence between the fished and unfished trajectories beginning around 1970 and increasing progressively until the end of each time series. In contrast, in the temperate regions, there is typically a general decline in $SB_{F=0}$ over the full time series and generally a small difference between the fished and unfished trajectories. This suggests that the equatorial regions have become more productive from 1970 onwards, with the temperate regions either simply maintaining their productivity or seeing a slight decline. 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

Differences over time in the estimated adult and juvenile fishing mortality are shown in Figure 54, with a gradual increase in adult fishing mortality up until around 2000, with relatively stable adult mortality since then. Fishing mortality in juveniles starts to increase from around 1970 with the expansion of the pole-and-line fisheries, then increases at a faster rate with the expansion of the purse seine fisheries from 1980–2000. This is followed by a period of relative stability in estimated annual juvenile mortality, albeit with considerable variability from year to year, from 2000 onwards, with annual juvenile mortality in this period around 0.35, compared to values of around 0.1 for adults in the same period.

The relative instantaneous fishing mortality by age varies by region as shown in Figure 55, with higher fishing mortality for older fish in the temperate regions generally, particularly regions 2, 5, 6 and 9, and higher instantaneous fishing mortality for juveniles in the equatorial regions 3, 4, 7 and 8. The trends in fishing mortality-at-age by decade (Figure 56) show a general increase in fishing mortality for all age-classes over the assessment period, although the increase for the youngest age classes is notably larger. There has been a temporal shift in the age distribution towards younger age classes as the stock has been fished down (Figure 56).

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 Figure 57, Figure 58, Figure 59 and Figure 60. These comparisons show that estimates of both spawning depletion and spawning potential, were somewhat sensitive to the choice of tag mixing period (Figure 57), while spawning depletion was also somewhat sensitive to the assumed steepness value (Figure 58), and spawning potential to the assumed size data weighting (Figure 59).

Under the alternative assumptions for tag mixing periods, the depletion trajectories were comparable up until the late 1990s, when the trajectories diverge, with the results from the 1 quarter mixing scenario indicating a slightly higher value for $SB/SB_{F=0}$. The 2 quarter mixing period scenario scales the spawning potential higher over the whole the time series, compared to the 1 quarter mixing period (Figure 57). The different steepness scenarios had no impact on the estimated spawning potential, while the estimated value for $SB/SB_{F=0}$ increased with increased values for steepness (Figure 58). The scalar for weighting the size data had little effect on the estimated spawning depletion trajectory, but had some effect on the estimated spawning potential, with increasing the weight on the size composition data (or a lower scalar value) resulting in lower values for the estimated spawning potential (Figure 59). The age data weighting also had little effect on the estimated spawning depletion trajectory. While the estimated spawning potential for the age data weightings of 1.0 and 0.75 were comparable, the spawning potential for the weighting of 0.5 was scaled slightly lower (Figure 60).

A sensitivity to the tag overdispersion parameter (τ), with alternative values of τ set to 1 and 4, was compared to a value of 2 for the diagnostic model. This showed very little effect either on spawning depletion or spawning potential trajectories for the diagnostic model, and the results are not shown here.

The sensitivity on doubling purse seine catches in regions 3, 4 and 8 in 2020 and 2021 resulted in a model with a large spike in estimated recruitment, mainly in region 4, with that spike occurring just before these catches were doubled, and just prior to the point where recruitment is no longer estimated (the third quarter of 2020) and fixed to the historical mean value. As a result, the fit to the CPUE series in region 4 deteriorates further towards the end of time series, with an increased spike at the end of this series, which is not supported by increases in the input CPUE data. The status of spawning potential and spawning potential depletion is not substantially changed by this spike in recruitment as neither these additional recruits predicted by the model, nor the purse-seine-sized bigeye caught in the last two years of the time series in this sensitivity, have begun to contribute to the spawning potential by the end of the model time period. Therefore purse

seine catch uncertainty in 2020 and 2021 has not impacted the main stock status indicators in this assessment, but may become an issue in future assessments.

Some preliminary sensitivities to regional structure were also investigated. A 5 region model using the same regional structure as proposed in the 2023 yellowfin stock assessment was explored for bigeye, combining regions 1 and 2 into a single region, combining regions 3 and 4 into another single region and combining three regions, region 7, 8 and 9 into further single region. This resulted in a model with problems fitting the tag data from releases in the Coral Sea area (region 9 in this model). Similar problems were noted by the 2012 bigeye peer review (Ianelli et al., 2012), which resulted in the introduction of a new small region in the Coral Sea (region 9) in later bigeye assessments, to specifically address the problem of poor fits to tagging data from that region when it was a small component of a larger region. It was not considered appropriate to reintroduce these problems, especially given the information these tags provide through some observations of long periods between release and recapture. An alternative simplified structure with 6 regions, maintaining the current region 9 separate to the combined southern region (including regions 5 and 6) was examined and this model gave qualitatively similar results to the 9 region model and resulted in a model with a positive definite Hessian. While it would be useful to explore this model in greater detail, given a positive definite Hessian had already been obtained for a potential 9 region diagnostic model, there was not the same incentive, or at that point, the time, to adopt a simpler regional structure for bigeye as there was for yellowfin, and this 6 region model was not developed further in this assessment.

We ran a one off sensitivity using fixed movement parameters derived from the SEAPODYM model. This model resulted in very poor fits to the CPUE data and there was insufficient time to explore this in greater detail.

7.5.2 Structural uncertainty grid

We chose to maintain the original proposed structural uncertainty grid, with axes for tag mixing, data weighting (separated for age and size composition) and steepness. While steepness may not be that influential on spawning potential reference points it can be quite influential on spawning depletion through unfished biomass and MSY based reference points. As such it is standard practice to include steepness as an uncertainty axis in tuna assessments. While growth and mortality are often included in an uncertainty grid, the growth parameters were able to be estimated internally, which is an approach which is supported by Punt et al. (2023). Using the Lorenzen functional form for mortality with the scale parameter estimated is also considered good practice and this approach is implemented for this stock assessment. Therefore we considered it unnecessary to include axes for growth or mortality in the uncertainty grid.

While in previous assessments, constructing the uncertainty grid has been computationally expen-

sive, but relatively routine. The process of establishing grid models in 2023 involved first running each model, followed by running 20 jitters on each model and for some models running extra evaluations, or making small manual jitter steps to allow a model to move to a better solution, usually based on a known better likelihood achieved for a similar grid model with difference only to the steepness settings. Changes to steepness typically result in changes to the spawning depletion, but typically with minimal change to the objective function.

The grid can be separated into two tag mixing halves. Each half needs a base model used to construct the grid. For the mix 2 half of the grid, the base model used was the model found in the penultimate step before settling on the 2023 diagnostic model. This is a model which runs from (modified) initial starting values and gives a likelihood around 10 likelihood units worse than the diagnostic model. Given every model in the grid was to be jittered, this "almost diagnostic" model was used as the basis for the mix 2 half of the grid.

For the mix 1 half of the grid, a new base model was required, as a different tag input file and model structure is used with a different tag mixing period. This mix 1 base model was derived from the mix 2 base model. After modifying the tag input file, this models was jittered 60 times, initial growth and mortality parameters were adjusted and the model was "twerked", or jittered again. For this half of the grid, it was not possible to find a base model that was close to the best jittered solution. In contrast to the mix 2 half of the grid, the base grid entries were based on a jittered solution, which required starting each model from a partially converged solution, albeit from a different model. We took advantage of the similarities between "steepness triplets" and proceeded to jitter one third of the grid, with steepness set to 0.8, and then used the jittered solutions to derive the grid members for steepness set to 0.65 and 0.95, for each "triplet", with extra iterations to try to achieve a model with a suitable gradient. On occasions, some of the grid members still had poor gradients, or objective functions worse than other members of the triplet, and extra iterations were run on these until a suitable likelihood value and gradient had been obtained. This process was time consuming, and required manual intervention at several steps.

Once the structural uncertainty grid was finalised, which was a journey in itself, the Hessian was calculated for every element of the grid and the results for each element were tabulated, showing results for the initial objective function, $SB/SB_{F=0}$, the initial gradient achieved, the number of negative eigenvalues, the value of the minimum eigenvalue if it was negative, as an indication of how close the model was to having a positive definite Hessian, and then all of these same parameters calculated for the jittered and improved result. This allows investigation of the gains achieved from jittering, in terms of changes to the objective function and the effects on $SB/SB_{F=0}$. The changes to $SB/SB_{F=0}$ were both increases and decreases and were all less than 4 percentage points change, remembering that these base models were already highly jittered and twerked. Larger changes to $SB/SB_{F=0}$ were found for models that were only jittered in a single cycle (no twerking), but the

starting point for the grid models (the base models) were already highly refined models. These results are listed in great detail in Table 6.

For the unjittered uncertainty grid, we started with 46 grid models which had a positive definite Hessian. Given that the objective function was improved for every one of these grid models through the process described above, none of the grid models was at a global minimum, even those 46 models which started with positive definite Hessians.

After jittering 48 of the 54 grid elements had positive definite Hessians, including all 8 of the base models that started without a positive definite Hessian. Of the 6 grid elements which started with a positive definite Hessian and ended without one after jittering, the improvement in objective function ranged from 29 to 52 likelihood units, so in each case there was a significant improvement in the fit obtained form jittering, even with the loss of PDH status. In each case, there was only one negative eigenvalue, and the absolute size was less than 10^{-6} in each case, so it was a very small in absolute size, so an excellent candidate for positivisation though MFCL, to get a workable estimate of estimation uncertainty.

These hard fought 54 grid models formed the structural uncertainty grid, representing an uncertainty grid that has probably never been worked so hard in the history of WCPO stock assessments, but a grid where it was possible to calculate estimation uncertainty. Phew!

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. Tag mixing and steepness have the largest influence in the uncertainty grid, with the size and age data weighting relatively less important (Figure 61).

The distribution of recruitment across model regions and quarters for all models in the structural uncertainty grid is summarised in Figure 62. Proportionally, recruitment is predicted to be highest in temperate Regions 1, 2 with notable seasonal variation, with relatively high recruitment also in Regions 6 and 7, and proportionally low recruitment in region 9, the smallest region. Quarters 1 and 4 tend to exhibit a greater proportion of recruitment in key regions (Figure 62).

The time series for spawning depletion ($SB_{\text{recent}}/SB_{F=0}$) and spawning potential SB_t across grid models are shown in Figure 63 and Figure 64. Majuro and Kobe plots showing the estimates of $F_{\text{recent}}/F_{\text{MSY}}$, $SB_{\text{recent}}/SB_{F=0}$, $\text{SB}/\text{SB}_{MSY}$, along with $SB_{\text{latest}}/SB_{F=0}$ and $\text{SB}_{latest}/\text{SB}_{MSY}$ across all models in the grid (Figure 68, Figure 69). The averages and quantiles across the 54 models in the grid for all of the reference points and other quantities of interest are shown in (Table 5). For key management quantities ($SB_{\text{recent}}/SB_{F=0}$, $SB_{\text{recent}}/SB_{\text{MSY}}$ and $F_{\text{recent}}/F_{\text{MSY}}$) we also include the additional estimate of estimation uncertainty for management advice (Table 5).

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

- The grid contains 54 models with a moderate range of estimates of stock status and reference points, and suggest that, overall, the stock is slightly more depleted (lower $SB_{\text{recent}}/SB_{F=0}$) and with higher spawning potential than the estimates from the 2020 assessment (Table 5).
- The most influential axis is the tag mixing period with an assumed mixing period of 2 quarters resulting in slightly greater levels of depletion (lower $SB_{\text{recent}}/SB_{F=0}$) and higher $F_{\text{recent}}/F_{\text{MSY}}$ than those in the 1 quarter mixing period (Figure 61).
- The second most influential axis of uncertainty in the grid is 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 61).
- Estimates of $SB_{\text{recent}}/SB_{F=0}$ and fishing mortality were relatively insensitive to the assumed size data weighting. Assuming a scalar of 40 led to marginally lower estimates of $SB_{\text{recent}}/SB_{F=0}$ than those for scalars of 20 and 10 (Figure 61).
- The conditional age-at-length data weighting axis also had limited impact on management quantities, with the tail of the 0.50 age data weighting scenario producing slightly lower estimates of $SB_{\text{recent}}/SB_{F=0}$ (Figure 61).
- The decline in $SB_{\text{recent}}/SB_{F=0}$ has been continuous across most of the time series, with $SB_{\text{recent}}/SB_{F=0}$ stabilising around 2010.
- Spawning depletion estimates in the tropical regions show notable declines across the time period through to 2010, approaching 20% $SB_{F=0}$ in the most recent model years for some of the grid models (Figure 63). In the most recent period, depletion levels have stabilised, with some regions showing a slight increase in $SB/SB_{F=0}$ at the end of the time series. Temperate regions have tended to show slow but steady declines in $SB/SB_{F=0}$ across the time period (Figure 63).
- Estimated spawning potential shows earlier declines in the time series by region than seen for $SB/SB_{F=0}$, with the estimated spawning potential subsequently stabilising later in the time series for many regions. Historically, region 2 is estimated to have the greatest spawning potential (Figure 64).
- None of the models in the structural uncertainty grid have an overall spawning potential depletion below the LRP (20% $SB_{F=0}$) with the median $SB_{\text{recent}}/SB_{F=0}$ at 0.35 and varying from 0.31–0.40 between the 10$^{\text{th}}$ to the 90$^{\text{th}}$ percentiles (Table 5).
- All models in the structural uncertainty grid show exploitation to be below F_{MSY}. Median $F_{\text{recent}}/F_{\text{MSY}}$ is 0.68 and varying from 0.53–0.86 over the 10$^{\text{th}}$ to the 90$^{\text{th}}$ percentile range

(Table 5).

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 65 for $SB_{\text{recent}}/SB_{F=0}$, Figure 66 for $F_{\text{recent}}/F_{\text{MSY}}$ and Figure 67 for $SB_{\text{recent}}/SB_{\text{MSY}}$.

Presenting the estimates arising from these two approaches to incorporating uncertainty separately, both the results from the uncertainty grid and from 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 that incorporating estimation uncertainty are identical, with slight differences seen when considering the 80$^{\text{th}}$ percentile range values. As expected, the tails of the distribution are slightly longer when incorporating estimation uncertainty.

For MSY-related quantities, incorporation of estimation uncertainty had 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 lower and median $SB_{\text{recent}}/SB_{\text{MSY}}$ is slightly lower when incorporating estimation uncertainty. While the 80$^{\text{th}}$ percentile range values for $SB_{\text{recent}}/SB_{\text{MSY}}$ are comparable, both bounds for the $F_{\text{recent}}/F_{\text{MSY}}$ 80$^{\text{th}}$ percentile range are lower when incorporating estimation uncertainty. The overall distribution of $F_{\text{recent}}/F_{\text{MSY}}$ and $SB_{\text{recent}}/SB_{\text{MSY}}$ values are also wider, with some values 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 66) and lower estimates of $SB_{\text{recent}}/SB_{\text{MSY}}$ (Figure 67).

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 including dynamic Majuro and Kobe analyses and fisheries impacts analyses.

We do not present the results of all analysis for all models in this report. In this section, we rely largely on the tabular results of the structural uncertainty grid (Table 5) and the dynamic spawning depletion and spawning potential plots for the models in the structural uncertainty grid (Figure 63 and Figure 64). We also refer to the fished and unfished spawning potential trajectories for the diagnostic model discussed previously (Figure 53) and the dynamic Majuro and Kobe plots (Figure 69).

Dynamic Majuro and Kobe plots and comparisons with Limit and Target Reference Points: The section summarising the structural uncertainty grid (Section 7.5.2) presents terminal estimates of stock status in the form of Majuro plots. Further analyses can estimate the time-series of stock status in the form of Majuro and Kobe plots, with methods are discussed in Section 5.7.4. The dynamic Majuro and Kobe plots for the diagnostic model models are shown in Figure 69.

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 late 1960s. The dynamic Majuro plot indicates that while $SB/SB_{F=0}$ stabilises toward the end of the assessment time period, fishing mortality has varied notably. 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 69).

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), in order to estimate which types of fishing activity have the most impact on the spawning potential (Figure 70). Fishing impacts were estimated to be very minor in all regions before about 1970, resulting primarily due to the longline fishery. The impact of this and pole and line gears has increased very slightly over the time series. In the early 1970s, catch information from the miscellaneous fisheries (Region 7) leads to an increase in impact, particularly in that region. Subsequently, the onset of notable impacts due to purse seine associated fishing occurs from the 1980s in all tropical regions. Examining the overall impact, the purse seine associated fishery has the most impact, with that of the miscellaneous and longline fisheries also notable.

The greatest fishing impacts are in the tropical regions, by the purse seine associated fishery and miscellaneous fishery, dependent on the region. Impacts in other regions are primarily due to longline fishing, with that of the other fishery components also present.

Yield analysis: The yield analyses conducted in this assessment incorporates the spawner recruitment relationship (Figure 50) 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 assume 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 71 for select models representing different axes of the structural uncertainty grid (specifically, the two levels of tag mixing). For the diagnostic model, it is estimated that MSY would be achieved by increasing fishing mortality by a factor of 1.3, although the resulting increase in yield would be relatively small (3%). The different example yield curves under the alternative mixing assumption 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.

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 (Figure 72). Prior to 1970, the WCPO bigeye fishery was almost exclusively conducted using longline gear, with a low exploitation of small bigeye. The associated age-specific selectivity pattern resulted in a much higher MSY in the early period compared to the recent estimates. This pronounced decline occurred after the expansion of the small-fish fisheries in region 7 and, soon after, the rapid expansion of the purse seine fishery which shifted the age composition of the catch towards much younger fish. This lower MSY is due to a combination of fish being removed from the system at smaller sizes and also before they have the chance to reproduce.

8 Discussion and conclusions

8.1 Stock Status from the uncertainty grid

The 2023 WCPO bigeye tuna stock assessment estimates that the median recent spawning depletion ($SB_{\text{recent}}/SB_{F=0}$) is well above the limit reference point, at the stock-wide scale, for all models in the in the uncertainty grid (Figure 63) and $F_{\text{recent}}/F_{\text{MSY}}$ is less than one (Table 5).

Considering reference points calculated from the uncertainty grid results, which also incorporate estimation uncertainty, the median value for $SB_{\text{recent}}/SB_{F=0}$ is 0.35 (Table 5). $F_{\text{recent}}/F_{\text{MSY}}$ is less than one, with a median value of 0.59.

The estimated depletion across the whole model region shows a long-term decline to around 2010 and then stabilising, while the spawning potential shows a steeper initial decline before also stabilising in the recent period. The most notable declines are in the tropical regions, with slightly smaller relative declines in the temperate regions.

Overall, the outcomes of this assessment suggest that the bigeye stock in the WCPO is not overfished and is not 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 34% calculated across the unweighted grid). Based upon the estimates of $SB_{\text{recent}}/SB_{F=0}$ (35% of $SB_{F=0}$) this objective has currently been met.

8.2 Changes to the previous assessment

The major changes from the 2020 assessment assessment include the following.

- Conversion from a catch-errors to a catch-conditioned modelling framework, 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 functional form of natural mortality at age.
- Additional procedures adopted to achieve more reliable model convergence, including extensive jittering and checking the 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 used in the structural uncertainty grid.
- Use of conditional age-at-length data, and internal estimation of growth, with alternative weighting of these data included in the structural uncertainty grid.

An additional three years of data (including new catch, effort, size composition and tagging data) was incorporated in this assessment. The conversion from a catch-errors model, with 11,421 estimated parameters, to a catch-conditioned model, with 3,067 estimated parameters, significantly reduced the complexity of the model and resulted in many intermediate models in the stepwise development having positive definite Hessians. Estimating mortality and growth parameters allowed a change in the axes selected for the uncertainty grid, removing growth but including tag mixing period, and data weighting of the age data.

8.3 Stock status incorporating estimation uncertainty

The assessment approach in 2023 has allowed the inclusion of estimation uncertainty, in addition to the standard approach of applying a structural uncertainty grid. This allows a more complete representation of the uncertainty in the estimation of key quantities of interest for management advice (Table 5) and is a significant advance in the history of bigeye stock assessments in the WCPO.

8.3.1 Model diagnostics

Of the 54 models in the structural uncertainty grid, 48 of them had positive definite Hessians. Retrospective analyses (Section 12.2) indicate that there are no retrospective patterns of concern. Likelihood profiles and Piner plots indicate that there is some conflict in the data, both between major data source, with weight composition data in conflict with the length composition and the tag data, but also within fisheries for some of these likelihood components.

Fits to the CPUE data were generally very good, and fits to the composition data were acceptable, with some notable improvements to the time aggregated fits, compared to those achieved in the 2020 assessment. Fits to the tagging data were acceptable, given the limitations of this dataset and the peculiarities in tagging data from some regions.

While we would like to have had time to examine more model diagnostics, there was not sufficient time to expand this list of diagnostics any further.

8.3.2 Further information to consider

The sensitivity using fixed movement parameters derived from the SEAPODYM model seemed like a promising approach, given the apparent differences in the estimated movement between regions, and regional movement present in the input data. We hoped this may provide an alternative perspective to possible movement in this model. The model has considerable freedom to adjust movement and recruitment parameters, which can sometimes be used to provide better fits to components of the data other than tag data, and indeed to trade off regional recruitment with movement between regions. It is not clear that the movement estimates produced by this model best describe movement of the actual population as a result. Alternative approaches looking at fine scale modelling of tagged fish would be worth exploring in future.

Sensitivities were explored to alternative options for refining the fits to size composition data (Peatman et al., 2023b), with stronger filtering of size composition data in the hope of improving fits to size composition data and exploring the use of sets as a measure of the raw input sample sizes, rather than number of fish, for the purse seine fisheries. This work was initiated in response to recommendations from the 2022 yellowfin peer review (Punt et al., 2023). While this approach also looks promising, it became apparent that this required more time than was available for this assessment, although considerable progress has been made in setting up data structures required to explore these options. This initial work provides a solid basis for a possible future project to explore these options and to find an objective method to select from the multiple options available. Models were explored with stronger data filtering, but it is clear that more dedicated time and research is required to examine this thoroughly.

Tag mixing is another area in which we hoped to explore more sensitivities, either looking at

a period of three quarters for tag mixing, or assigning different mixing periods for different tag release groups, based on analysis of spatial and temporal recapture patterns from these tag release events. Once again, there was insufficient time to explore these options in more detail, but it is clear that there are some issues that require further research in converting the definitions of quarters at liberty from the difference between quarters of release and recapture, to considering the number of days at liberty for tagged fish recaptured, both within the mixing period, but also beyond the tag mixing period tag

8.4 Recommendations for further work

Consideration could be given in the next assessment to modifying the number of years at the end of the series for which recruitment is no longer estimated, to avoid high recruitment estimates at the end of the time series which may be an artifact of trying to fit other components of the model and with insufficient number of cohorts observed in the size composition data to confirm a high recruitment, with little support in the size composition data. The sensitivity on potential COVID affected purse seine catch records highlighted the potential need to examine the variance of the last few estimated recruitment deviations, and to consider revising the number of quarters for which recruitment is no longer estimated (set to the mean) at the end of the time series. The combination of high final recruitment estimates with poor fits to the CPUE at the end of the time series in this sensitivity, was also observed in region 4 for other models. Examining the variance of the last few estimated recruitment event is good practice to justify the decision of when to stop estimating recruitment deviations, and it may be worthwhile to look more closely at this in future assessments

Additional one-off sensitivities were explored looking at alternatives to the tag mixing assumptions, using either the number of days at liberty for the mixing period only, or the number of quarter boundaries crossed as a quarterly demarcation, as used prior to 2020. This appears to be an influential decision, and alternative assumptions may be worth testing, as there appear to be some issues with both methods. Dealing with tag recaptures that need to be moved across recapture boundaries is problematic when that moves recaptures into periods which may not have any fishing by the fishery that recaptured that tag. There are also potential issues with only applying this correction to tags within the mixing period, which requires careful consideration. Preliminary sensitivities were run looking at this and the initial results suggest that this modelling decision is influential.

This assessment includes large quantities of size data which is difficult to explore and understand. It would be useful to have a comprehensive review of size compositional data to ensure the data being used in the assessment models is representative of the population and to ensure the appropriate filtering and catch-weighting of this data is carried out. Collecting representative data is very important given the scale of the assessment area, and would allow greater investigation into issues

such as potential regional variation in growth.

There were a number of other sensitivities that were "aspirational" which we would have liked to have had the time to explore. These include a single area model, an equatorial only model, a no tag model, input compositional sample sizes using sets rather than fish numbers, datasets with various degrees of filtering to try to improve fits to compositional data.

Tag mixing is a key issue for a tag based model and one that deserves more time. Preliminary analysis of individual tagging events was a good start, but more could potentially be achieved by extending this approach. Tags from region 9 are clearly influential and appear to behave differently, when looking at preliminary Piner plot (which were not finished in time to include in this report).

Data weighting is an area that is difficult given the complexity and running time for these models. Iterative schemes have been used elsewhere to weight compositional data and conditionals age-at-length data. While running these iterative approaches to convergence is clearly not practical, this is an area that could be explored.

Additional biological sampling, including age samples, and potentially the use of epigenetic ageing is very important for these models. The sample size of only 1004 otoliths is a major weakness in the data input to this assessment. A well balanced statistically designed sampling program to achieve representative temporal and spatial samples over the full range of the assessment area would be beneficial to this assessment and an increase in investment will improve the reliability of the model outputs. As always, quality control over the sampling and sampling protocols is important to avoid "garbage in, garbage out".

While the tagging data is very valuable for this assessment, the current tagging programs could be better balanced spatially and temporally to inform the assessment.

Simpler regional structures are an area that deserves more attention. These assessments are large and complex, some may say unwieldy, so there is potentially much to gain though judicious simplification. While a simpler structure could have modelling benefits, there are also potential issues with the eastern boundary. Recaptures of tagged fish demonstrate that fish cross his boundary and the large catches of bigeye near this boundary warrant consideration of alternatives assessment structures.

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.
- Average fishing mortality rates for juvenile and adult age-classes have increased throughout

the period of the assessment until around 2000, after which they have stabilised, but with high inter-annual variability for juveniles. Juveniles have experienced considerably higher fishing mortality than adults.

- Overall, the median depletion from the uncertainty grid for the recent period (2018-2021; $SB_{\text{recent}}/SB_{F=0}$) is estimated at 0.35 (80 percentile range including estimation and structural uncertainty 0.30–0.40, full range 0.25–0.46)
- No models from the uncertainty grid, including estimation uncertainty, estimate the stock to be below the LRP of 20% $\text{SB}_{F=0}$.
- CMM 2021-01 contains an objective to maintain the spawning biomass depletion ratio above the average for 2012-2015, $SB_{2012-2015}/SB_{F=0}$, which is a value of 0.34 calculated across the unweighted grid. Based upon the estimates of $SB_{\text{recent}}/SB_{F=0}$ of 0.35 this objective has currently been met.
- Recent (2017–2020) median fishing mortality ($F_{\text{recent}}/F_{\text{MSY}}$) was 0.59 (80 percentile range, including estimation and structural uncertainty 0.46–0.74, full range 0.37–0.99).
- Assessment results suggest that the bigeye 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 (PAW) 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. we highlight the lifetime of work that Dave Fournier put into MULTIFAN-CL, without which we could not have performed this assessment.

Thanks are due to so many people who contributed to this assessment, formally and informally. For some, this was there first time working on bigeye, for others there was still energy and enthusiasm for working on this stock assessment after a range of previous experience with this species and in this region. A collaborative and co-operative hard working team is a wonderful thing, but there are numerous support staff who make this possible and have tolerated way too much discussion about positive definite Hessians, jittering and twerking. Many friends and colleagues have helped

support this work by just being there, going for a run, hanging out, shooting the breeze, throwing a frisbee, camping - you probably know who you are. Others have contributed in the background, either through data collection and processing or administration. To the team who are formally authors, it has been a great journey and your input and enthusiasm has been enlightening and contagious. I hope we are all richer as a result. As lead author, I can confidently say that this is the best (bigeye) stock assessment I have ever worked on. I hope you think so too. It is also my first stock assessment - well for bigeye anyway. However, as a team, I think we have broken new ground and hopefully set some new standards, that teams to come can build on, just as we have started from a solid basis built by those who preceded us. While I am sure there are many aspects of this assessment that we could have improved, the one aspect that don't think I can fault any of my ao-authors on is their good humour, patience, resilience, tenaciousness and plain hard work. It has been a pretty long journey at times – but with a very satisfying end product. I hope others find it at least half as good as we like to think it is!

10 Tables

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

Fishery No	Fishery	Flag	Region	Sel Group	% catch last 10 yrs	% catch all yrs
F1	LL-ALL-1	ALL	1	1	4.41	7.92
F2	LL-ALL-2	ALL	2	2	4.67	9.17
F3	LL-US-2	US	2	3	3.55	1.75
F4	LL-ALL-3	ALL	3	4	3.00	6.62
F5	LL-OS-3	OS	3	8	3.83	2.34
F6	LL-OS-7	OS	7	9	7.40	6.57
F7	LL-ALL-7	ALL	7	10	1.13	2.26
F8	LL-ALL-8	ALL	8	11	0.42	0.56
F9	LL-ALL-4	AU	4	5	14.88	19.08
F10	LL-AU-5	ALL	5	12	0.30	0.24
F11	LL-ALL-5	ALL	5	7	1.12	1.51
F12	LL-ALL-6	ALL	6	6	2.69	2.07
F13	SA-ALL-3	ALL	3	13	11.59	12.05
F14	SU-ALL-3	ALL	3	16	2.86	1.77
F15	SA-ALL-4	ALL	4	14	17.98	8.08
F16	SU-ALL-4	ALL	4	17	1.08	0.47
F17	Z-PH-7	PH	7	19	1.20	3.33
F18	Z-ID.PH-7	ID.PH	7	20	1.29	0.51
F19	S-JP-1	JP	1	21	0.26	0.47
F20	P-JP-1	JP	1	22	0.96	1.92
F21	PL-ALL-3	ALL	3	23	0.01	0.05
F22	PL-ALL-8	ALL	8	24	0.00	0.01
F23	Z-ID-7	ID	7	25	8.10	3.34
F24	S-ID.PH-7	ID.PH	7	26	0.68	1.06
F25	SA-ALL-8	ALL	8	15	2.13	3.34
F26	SU-ALL-8	ALL	8	18	1.90	0.84
F27	L-AU-9	AU	9	12	0.02	0.04
F28	P-ALL-7	ALL	7	27	1.29	1.83
F29	L-ALL-9	ALL	9	7	0.00	0.02
F30	SA-ALL-7	ALL	7	13	0.03	0.31
F31	SU-ALL-7	ALL	7	16	0.05	0.05
F32	Z-VN-7	VN	7	28	1.20	0.41
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		
F38	LL-ALL-6	ALL	6	29		
F39	LL-ALL-7	ALL	7	29		
F40	LL-ALL-8	ALL	8	29		
F41	LL-ALL-9	ALL	9	29		

Table 2: Definition of fisheries and associated tag recapture and reporting rate groupings (three columns at right) for the 2023 MULTIFAN-CL bigeye tuna stock assessment in the WCPO. RTTP–Regional Tuna Tagging Program, PTTP–Pacific Tuna Tagging Program, JPTP - Japanese Tagging Program. 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. * Indicates groups with 5 or fewer tag returns for which reporting rates were not estimated.

Fishery No.	Fleet Codes	Flag	Model Region	Recap Group	RTTP	PTTP	JPTP
F1	LL-ALL-1	ALL	1	1	RTTP_L	RTTP_L	JPTP_L
F2	LL-ALL-2	ALL	2	2	RTTP_L	RTTP_L	JPTP_L
F3	LL-US-2	US	2	3	RTTP_L(US)*	RTTP_L(US)*	JPTP_L(US)*
F4	LL-ALL-3	ALL	3	4	RTTP_L	RTTP_L	JPTP_L
F5	LL-OS-3	OS	3	5	RTTP_L	RTTP_L	JPTP_L
F6	LL-OS-7	OS	7	6	RTTP_L	RTTP_L	JPTP_L
F7	LL-ALL-7	ALL	7	7	RTTP_L	RTTP_L	JPTP_L
F8	LL-ALL-8	ALL	8	8	RTTP_L	RTTP_L	JPTP_L
F9	LL-ALL-4	AU	4	9	RTTP_L	RTTP_L	JPTP_L
F10	LL-AU-5	ALL	5	10	RTTP_L(AU)	RTTP_L(AU)	JPTP_L(AU)*
F11	LL-ALL-5	ALL	5	11	RTTP_L	RTTP_L	JPTP_L
F12	LL-ALL-6	ALL	6	12	RTTP_L	RTTP_L	JPTP_L
F13	SA-ALL-3	ALL	3	13	RTTP_S-3	PTTP_S-3	JPTP_S-3*
F14	SU-ALL-3	ALL	3	13	RTTP_S-3	PTTP_S-3	JPTP_S-3*
F15	SA-ALL-4	ALL	4	14	RTTP_S-4	PTTP_S-4	JPTP_S-4
F16	SU-ALL-4	ALL	4	14	RTTP_S-4	PTTP_S-4	JPTP_S-4
F17	Z-PH-7	PH	7	15	RTTP_S(PH,ID)-7	PTTP_S(PH,ID)-7	JPTP_S(PH,ID)-2*
F18	Z-ID.PH-7	ID.PH	7	15	RTTP_S(PH,ID)-7	PTTP_S(PH,ID)-7	JPTP_S(PH,ID)-2*
F19	S-JP-1	JP	1	16	RTTP_S(JP)-1*	PTTP_S(JP)-1*	JPTP_S(JP)-1
F20	P-JP-1	JP	1	17	RTTP_P(JP)-1*	PTTP_P(JP)-1*	JPTP_P(JP)-1
F21	PL-ALL-3	ALL	3	18	RTTP_P-3*	PTTP_P-3*	JPTP_P-3
F22	PL-ALL-8	ALL	8	19	RTTP_P-8*	PTTP_P-8*	JPTP_P-8
F23	Z-ID-7	ID	7	15	RTTP_S(PH,ID)-7	PTTP_S(PH,ID)-7	JPTP_S(PH,ID)-2*
F24	S-ID.PH-7	ID.PH	7	15	RTTP_S(PH,ID)-7	PTTP_S(PH,ID)-7	JPTP_S(PH,ID)-2*
F25	SA-ALL-8	ALL	8	20	RTTP_SA-8	PTTP_S-8	JPTP_S-8
F26	SU-ALL-8	ALL	8	20	RTTP_SU-8	PTTP_S-8	JPTP_S-8
F27	L-AU-9	AU	9	21	RTTP_L(AU)	RTTP_L(AU)	JPTP_L(AU)*
F28	P-ALL-7	ALL	7	22	RTTP_P-7	PTTP_P-7	JPTP_S-8
F29	L-ALL-9	ALL	9	23	RTTP_L	RTTP_L	JPTP_L
F30	SA-ALL-7	ALL	7	24	RTTP_S-7*	PTTP_S-7*	JPTP_S-8
F31	SU-ALL-7	ALL	7	24	RTTP_S-7*	PTTP_S-7*	JPTP_S-8
F32	Z-VN-7	VN	7	25	RTTP_Z(VN)-7*	PTTP_Z(VN)-7*	JPTP_S-8

Table 3: Structural uncertainty grid for the 2023 WCPO bigeye 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)
$\text{Y}_{Frecent}$	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. Note that these values do not include estimation uncertainty, unless otherwise indicated.

	mean	median	min	10%ile	90%ile	max	diagnostic model
C_{latest}	139314	139199	138527	138947	139939	140347	139341
F_{MSY}	0.06	0.06	0.04	0.04	0.07	0.08	0.05
f_{mult}	1.69	1.67	2.27	2.17	1.35	1.22	1.54
$\text{F}_{recent}/\text{F}_{MSY}$	0.59	0.60	0.44	0.46	0.74	0.82	0.65
MSY	162248	164640	137920	143112	180820	184440	147160
SB_0	1867222	1832500	1384000	1552100	2247300	2457000	2147000
$\text{SB}_{F=0}$	1952050	1921715	1460378	1612630	2356598	2561690	2286052
$\text{SB}_{latest}/\text{SB}_0$	0.36	0.36	0.29	0.31	0.40	0.41	0.35
$SB_{\text{latest}}/SB_{F=0}$	0.34	0.34	0.27	0.30	0.38	0.40	0.33
$\text{SB}_{latest}/\text{SB}_{MSY}$	1.76	1.77	1.16	1.28	2.31	2.46	1.64
SB_{MSY}	393037	376300	225100	277230	534330	595900	464800
$SB_{\text{MSY}}/\text{SB}_0$	0.21	0.21	0.16	0.17	0.24	0.25	0.22
$SB_{\text{MSY}}/\text{SB}_{F=0}$	0.20	0.20	0.15	0.17	0.23	0.24	0.20
$SB_{\text{recent}}/SB_{F=0}$	0.35	0.35	0.28	0.31	0.40	0.41	0.34
$\text{SB}_{recent}/\text{SB}_{MSY}$	1.82	1.83	1.20	1.32	2.38	2.54	1.68
$\text{Y}_{Frecent}$	37982	37805	33400	34365	42369	42980	35820
Including estimation uncertainty							
	mean	median	min	10%ile	90%ile	max	
$SB_{\text{recent}}/SB_{F=0}$	0.35	0.35	0.25	0.30	0.40	0.46	
$\text{F}_{recent}/\text{F}_{MSY}$	0.59	0.59	0.37	0.46	0.74	0.99	
$\text{SB}_{recent}/\text{SB}_{MSY}$	1.82	1.79	0.94	1.32	2.41	2.96	

Table 6: Summary of objective functions (obj value), gradients (grad), numbers of improved jitters (imp), Hessian status (a positive definitite Hessian indicated by a 0 in the neg ev columns), minimum eigenvalues (min e-value) and $SB_{\text{recent}}/SB_{F=0}$ before jittering (recent 1), after jittering (recent 2) and the absolute difference in $SB_{\text{recent}}/SB_{F=0}$ (diff) for 54 individual grid models in the structural uncertainty grid.

model	seed	obj value 1	obj value 2	diff	grad 1	grad 2	imp	neg ev 1	min e-value 1	neg ev 2	min e-value 2	recent 1	recent 2	diff
m1_s10_a050_h65	16	-1159504.68	-1159555.03	-50	0.00006	0.00126	2	0		0		0.336	0.332	-0.004
m1_s10_a050_h80	16	-1159534.34	-1159555.06	-21	0.00011	0.00046	11	1	-0.00000320631	0		0.371	0.368	-0.003
m1_s10_a050_h95	16	-1159543.43	-1159555.08	-12	0.00014	0.00069	2	1	-0.00000290561	0		0.395	0.392	-0.003
m1_s10_a075_h65	18	-1158906.33	-1158966.78	-60	0.00031	0.00007	2	1	-0.00000000740	0		0.334	0.335	0.001
m1_s10_a075_h80	18	-1158934.91	-1158966.40	-31	0.00006	0.00011	14	1	-0.00000002280	0		0.369	0.371	0.002
m1_s10_a075_h95	18	-1158917.40	-1158966.83	-49	0.00005	0.00028	2	0		0		0.391	0.394	0.004
m1_s10_a100_h65	07	-1158346.97	-1158390.01	-43	0.00003	0.00020	2	0		0		0.336	0.336	0.000
m1_s10_a100_h80	07	-1158359.46	-1158388.52	-29	0.00022	0.00097	7	1	-0.00001327350	0		0.369	0.372	0.002
m1_s10_a100_h95	07	-1158359.47	-1158390.07	-31	0.00021	0.00002	2	1	-0.00000000780	0		0.393	0.395	0.001
m1_s20_a050_h65	16	-1027257.50	-1027271.52	-14	0.00005	0.00001	2	0		0		0.340	0.338	-0.002
m1_s20_a050_h80	16	-1027257.54	-1027271.55	-14	0.00001	0.00001	8	0		0		0.377	0.375	-0.002
m1_s20_a050_h95	16	-1027257.55	-1027271.57	-14	0.00001	0.00005	2	0		0		0.402	0.400	-0.002
m1_s20_a075_h65	12	-1026718.65	-1026723.20	-5	0.00004	0.00026	2	0		0		0.341	0.339	-0.002
m1_s20_a075_h80	12	-1026718.68	-1026723.23	-5	0.00001	0.00001	2	0		0		0.378	0.376	-0.002
m1_s20_a075_h95	12	-1026718.70	-1026723.23	-5	0.00002	0.00001	2	0		0		0.403	0.376	-0.027
m1_s20_a100_h65	20	-1026185.50	-1026207.25	-22	0.00001	0.00002	2	0		0		0.343	0.346	0.003
m1_s20_a100_h80	20	-1026185.54	-1026207.28	-22	0.00001	0.00001	1	0		0		0.380	0.390	0.011
m1_s20_a100_h95	20	-1026185.56	-1026207.30	-22	0.00001	0.00002	2	0		0		0.404	0.407	0.003
m1_s40_a050_h65	20	-890629.35	-890647.24	-18	0.00005	0.00005	2	0		0		0.329	0.332	0.003
m1_s40_a050_h80	20	-890629.39	-890647.28	-18	0.00001	0.00008	9	0		0		0.367	0.370	0.002
m1_s40_a050_h95	20	-890629.41	-890647.30	-18	0.00001	0.00002	2	0		0		0.393	0.396	0.002
m1_s40_a075_h65	07	-890098.72	-890128.85	-30	0.00001	0.00001	2	0		0		0.332	0.338	0.006
m1_s40_a075_h80	07	-890098.75	-890128.89	-30	0.00005	0.00027	10	0		0		0.370	0.376	0.006
m1_s40_a075_h95	07	-890098.77	-890128.91	-30	0.00002	0.00004	2	0		0		0.396	0.402	0.006
m1_s40_a100_h65	15	-889574.45	-889647.07	-73	0.00001	0.00001	2	0		0		0.334	0.340	0.006
m1_s40_a100_h80	15	-889574.49	-889647.09	-73	0.00001	0.00172	11	0		0		0.372	0.378	0.006
m1_s40_a100_h95	15	-889574.51	-889647.13	-73	0.00001	0.00001	2	0		0		0.398	0.403	0.006
m2_s10_a050_h065	03	-1160340.31	-1160477.78	-137	0.00035	0.00001	17	1	-0.00000000567	0		0.338	0.311	-0.027
m2_s10_a050_h080	07	-1160416.42	-1160477.82	-61	0.00063	0.00007	15	0		0		0.356	0.347	-0.008
m2_s10_a050_h095	06	-1160345.70	-1160477.84	-132	0.00024	0.00018	17	0		0		0.394	0.372	-0.022
m2_s10_a075_h065	17	-1159770.10	-1159888.69	-119	0.00003	0.00043	14	0		0		0.331	0.297	-0.034
m2_s10_a075_h080	20	-1159807.49	-1159888.73	-81	0.00018	0.00303	14	0		0		0.370	0.333	-0.037
m2_s10_a075_h095	16	-1159764.83	-1159888.76	-124	0.00014	0.00032	15	0		0		0.390	0.357	-0.032
m2_s10_a100_h065	07	-1159233.40	-1159300.07	-67	0.00046	0.00023	17	1	-0.00000000174	0		0.334	0.319	-0.016
m2_s10_a100_h080	13	-1159241.24	-1159300.11	-59	0.00053	0.00059	14	0		0		0.366	0.354	-0.012
m2_s10_a100_h095	07	-1159241.66	-1159300.13	-58	0.00001	0.00001	11	0		0		0.389	0.378	-0.011
m2_s20_a050_h065	15	-1027990.84	-1028083.13	-92	0.00024	0.00129	13	0		0		0.314	0.292	-0.022
m2_s20_a050_h080	20	-1028063.94	-1028083.18	-19	0.00001	0.00001	8	0		0		0.341	0.330	-0.011
m2_s20_a050_h095	08	-1027990.91	-1028083.20	-92	0.00001	0.00013	14	0		0		0.375	0.355	-0.020
m2_s20_a075_h065	07	-1027530.16	-1027540.39	-10	0.00001	0.00002	4	0		0		0.305	0.305	0.000
m2_s20_a075_h080	07	-1027530.20	-1027540.43	-10	0.00001	0.00025	3	0		0		0.342	0.342	0.000
m2_s20_a075_h095	17	-1027458.71	-1027540.45	-82	0.00001	0.00013	11	0		0		0.373	0.366	-0.007
m2_s20_a100_h065	03	-1026964.03	-1027015.23	-51	0.00001	0.00001	10	0		1	-0.00000017359	0.311	0.306	-0.005
m2_s20_a100_h080	04	-1026985.92	-1027015.28	-29	0.00001	0.00003	11	0		1	-0.00000017470	0.357	0.342	-0.015
m2_s20_a100_h095	13	-1026985.94	-1027015.30	-29	0.00027	0.00045	8	0		1	-0.00000018184	0.382	0.367	-0.015
m2_s40_a050_h065	04	-891356.15	-891409.65	-53	0.00005	0.00005	11	0		0		0.297	0.279	-0.018
m2_s40_a050_h080	01	-891355.15	-891409.70	-55	0.00018	0.00001	15	0		0		0.336	0.318	-0.018
m2_s40_a050_h095	01	-891322.22	-891409.73	-88	0.00028	0.00004	13	0		0		0.359	0.344	-0.016
m2_s40_a075_h065	14	-890834.85	-890888.67	-54	0.00001	0.00006	13	0		0		0.298	0.294	-0.004
m2_s40_a075_h080	03	-890834.79	-890888.72	-54	0.00001	0.00020	14	0		0		0.336	0.332	-0.004
m2_s40_a075_h095	05	-890834.82	-890888.75	-54	0.00005	0.00002	13	0		0		0.361	0.358	-0.003
m2_s40_a100_h065	10	-890324.31	-890375.89	-52	0.00001	0.00019	13	0		1	-0.00000000586	0.301	0.297	-0.003
m2_s40_a100_h080	15	-890324.08	-890375.94	-52	0.00005	0.00001	13	0		1	-0.00000001859	0.338	0.335	-0.003
m2_s40_a100_h095	11	-890324.39	-890375.97	-52	0.00001	0.00001	13	0		1	-0.00000001859	0.364	0.360	-0.003

11 Figures

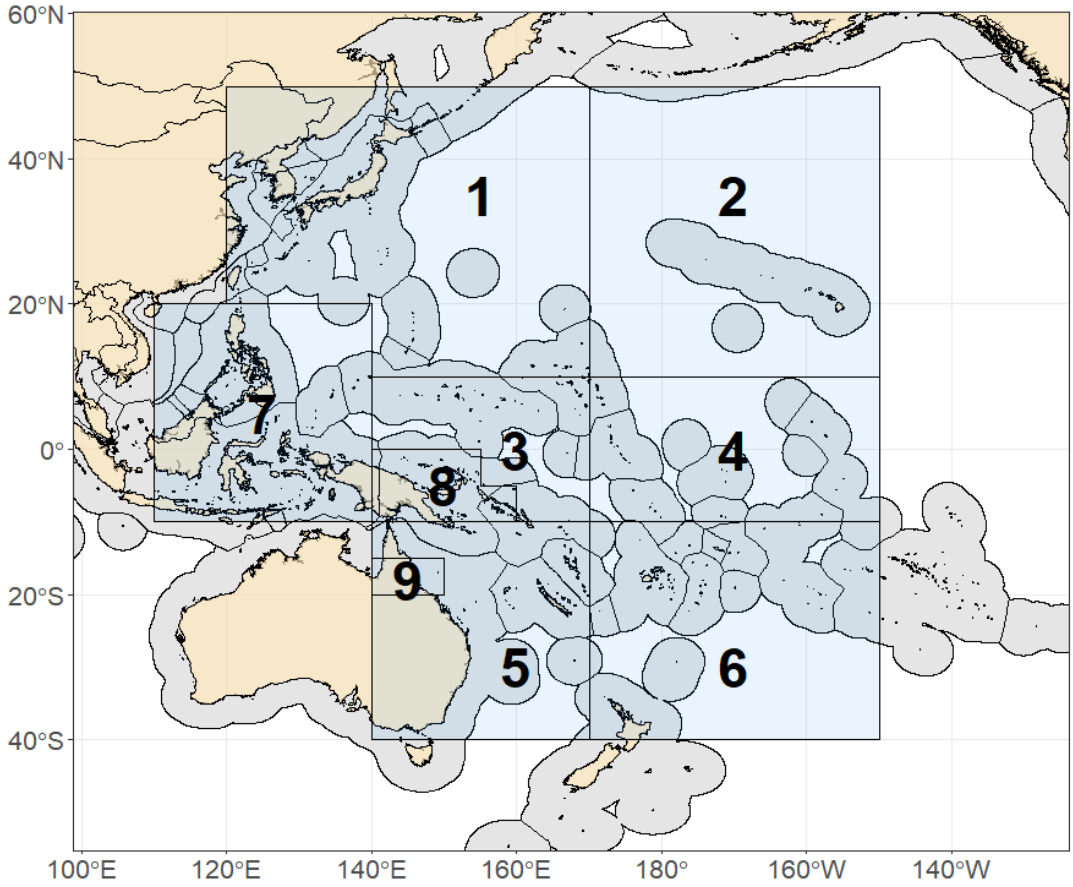

Figure 1: The geographical area covered by the stock assessment and the boundaries of the nine model regions used for 2023 WCPO bigeye tuna assessment.

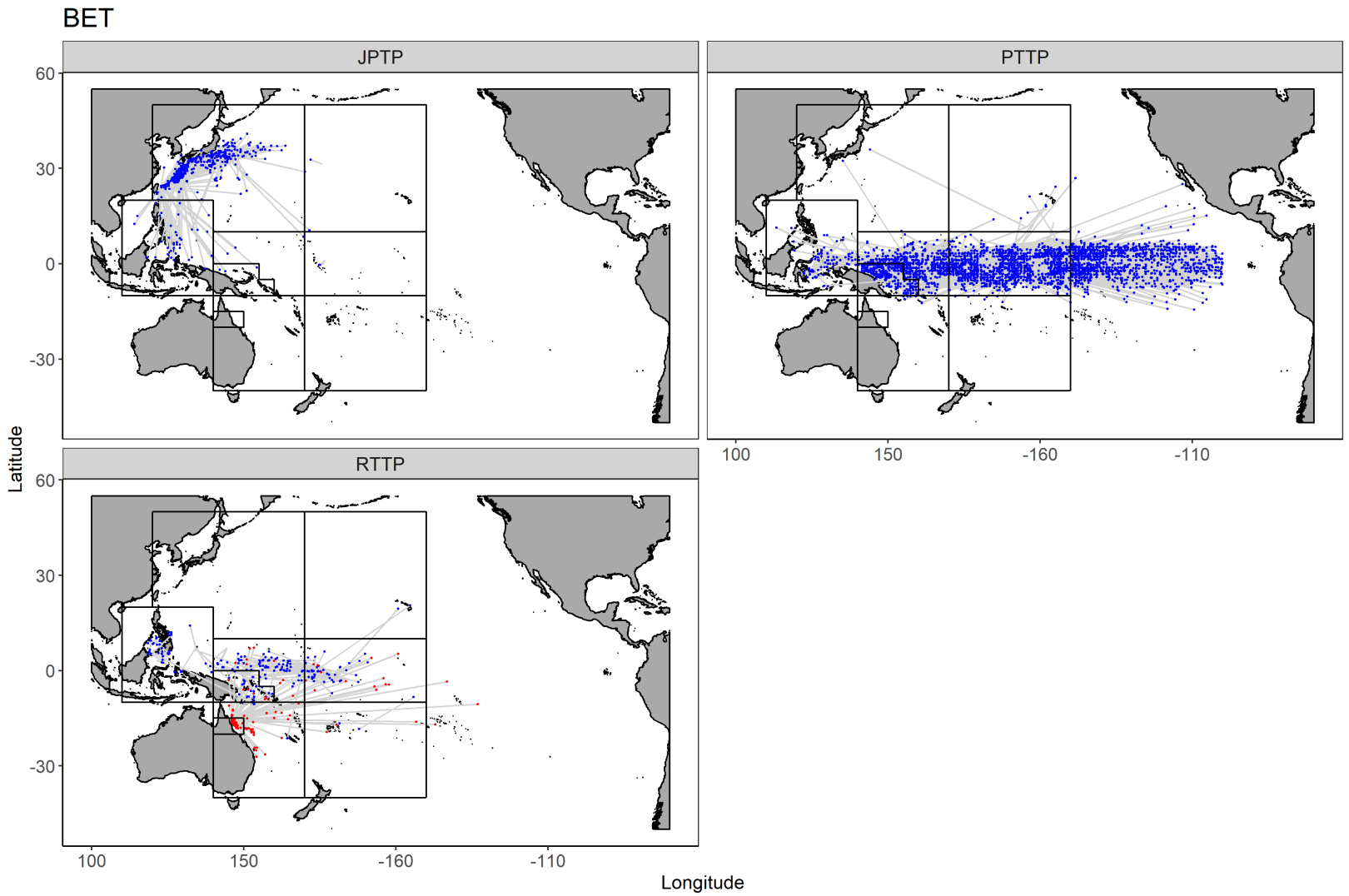

Figure 2: Map of tag recaptures. The panel shows the distributions of release and recapture displacements for the different tagging programs. Pacific Tuna Tagging Program (PTTP), Regional Tuna Tagging Program (RTTP) and the Japanese Tagging Program (JPTP). Dots indicate recapture locations. Red dots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

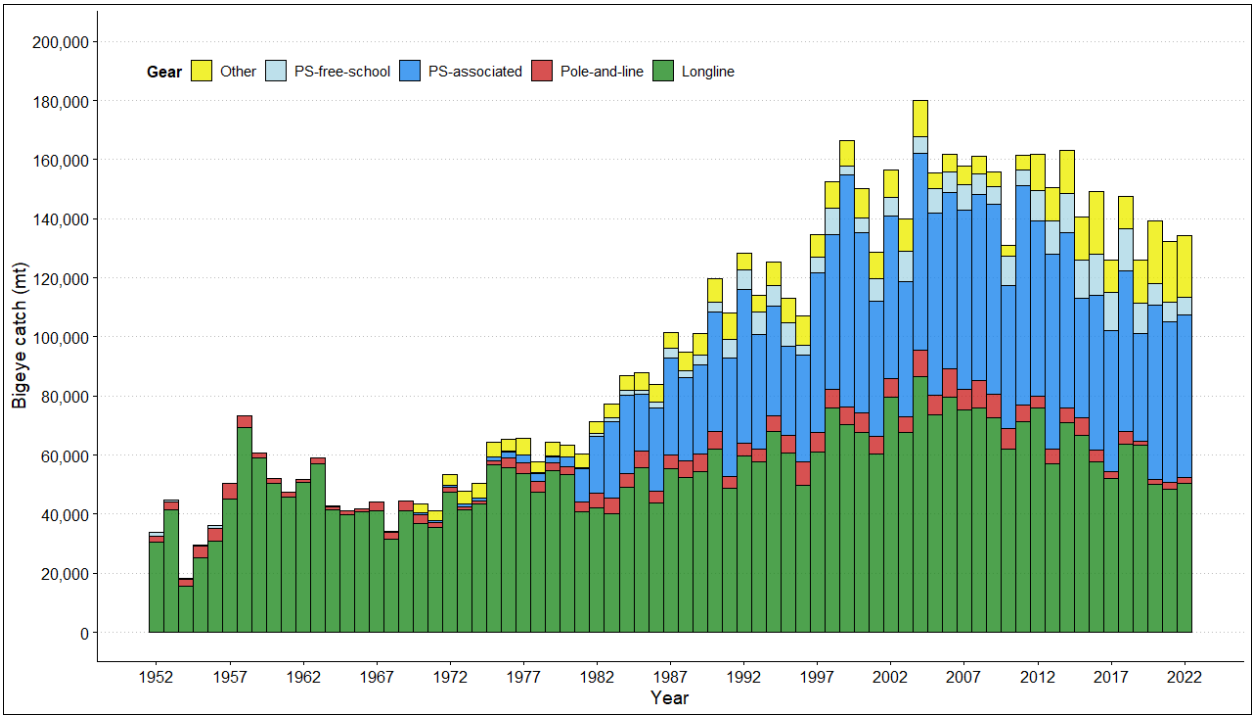

Figure 3: Annual catches of bigeye by gear type in the WCPO area covered by the assessment.

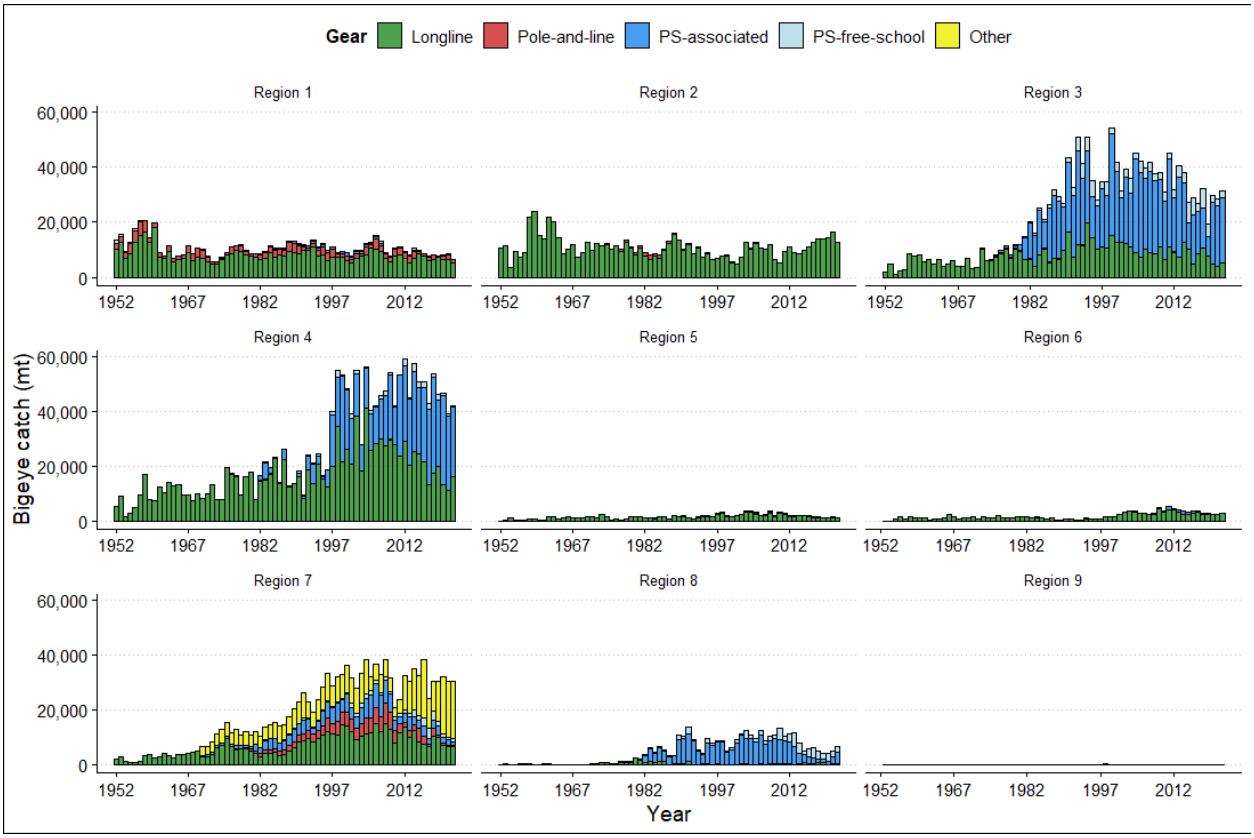

Figure 4: Annual catches of bigeye by gear type for each of the nine model regions.

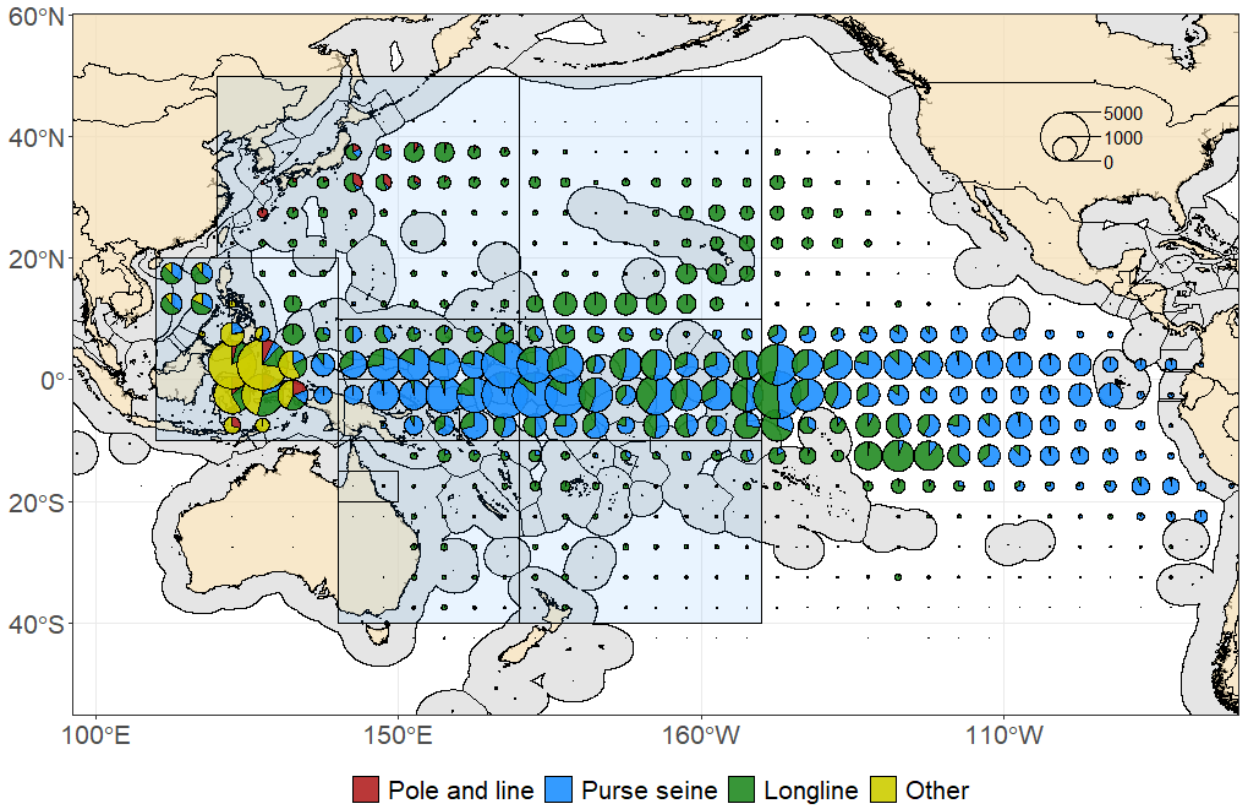

Figure 5: Distribution and magnitude of bigeye catches (mt) by gear type summed over the last 10 years (2012-2021) for 5 x 5 degree cells.

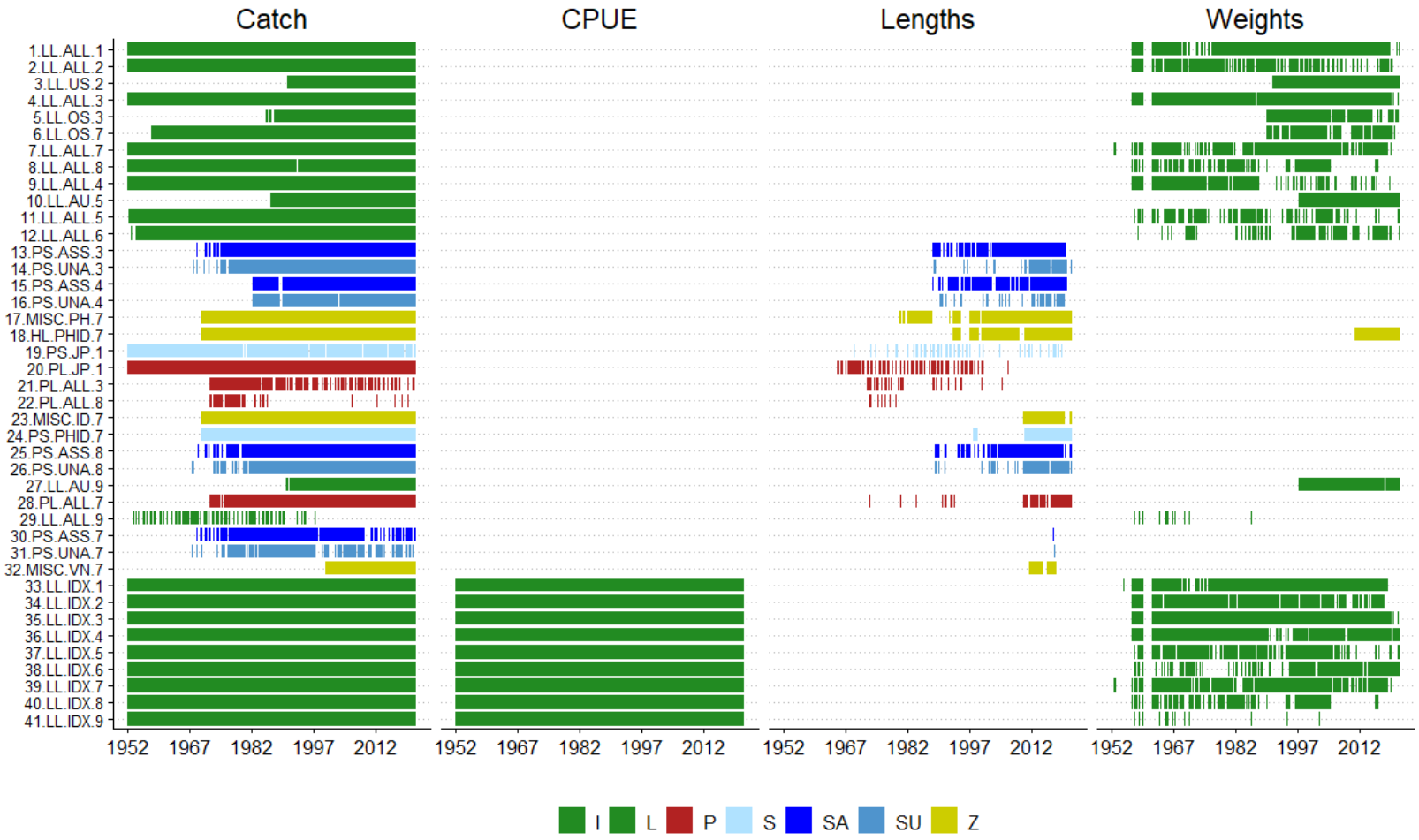

Figure 6: Summary of data coverage by fishery for the WCPO 2023 bigeye assessment. I=index fisheries, L=longline, P=pole and line, S=purse seine (unspecified), SA=purse seine associated, SU=purse seine unassociated, Z=miscellaneous gears.

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

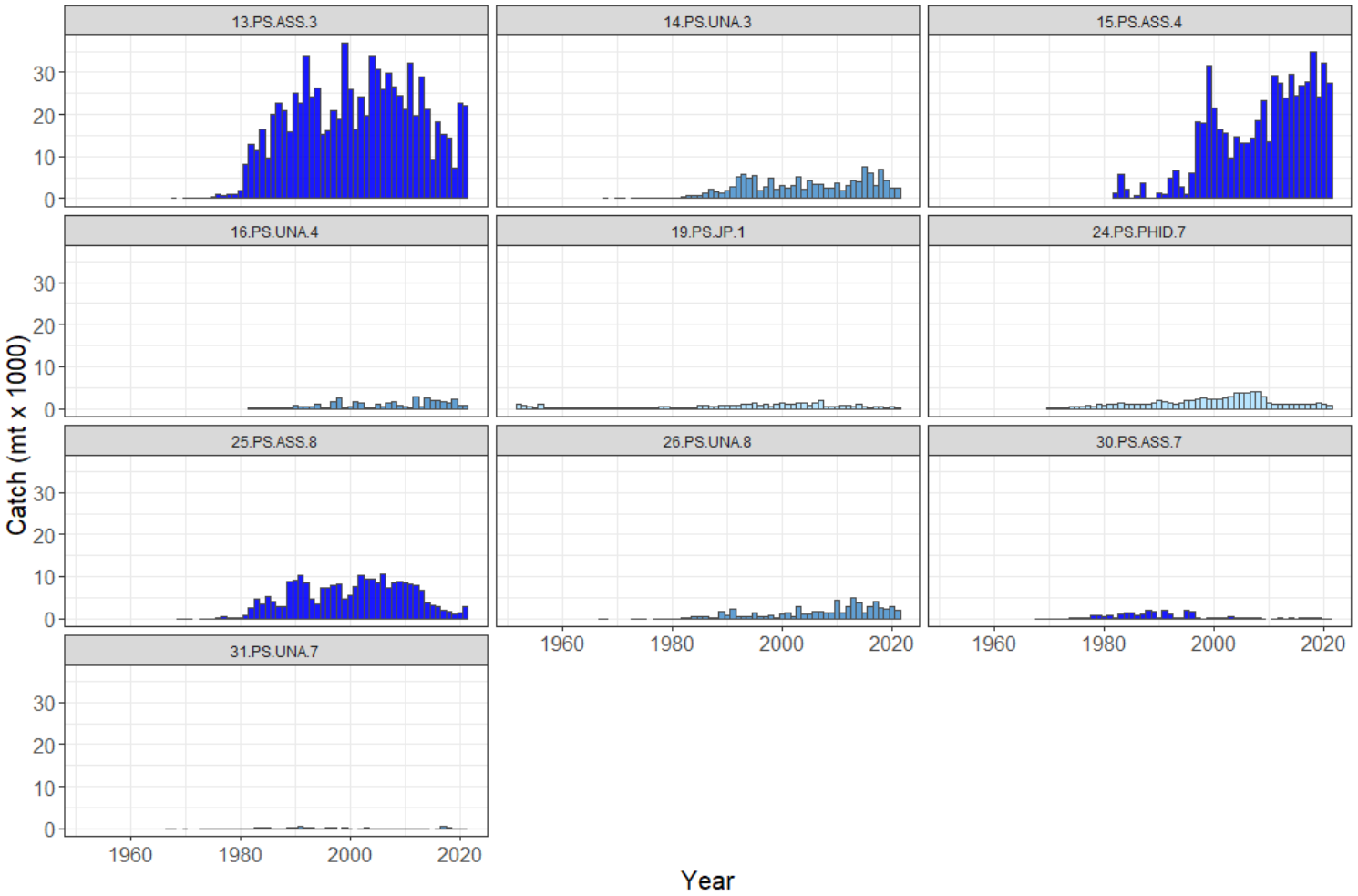

Figure 8: Time series of catches (mt) by fishery: purse seine.

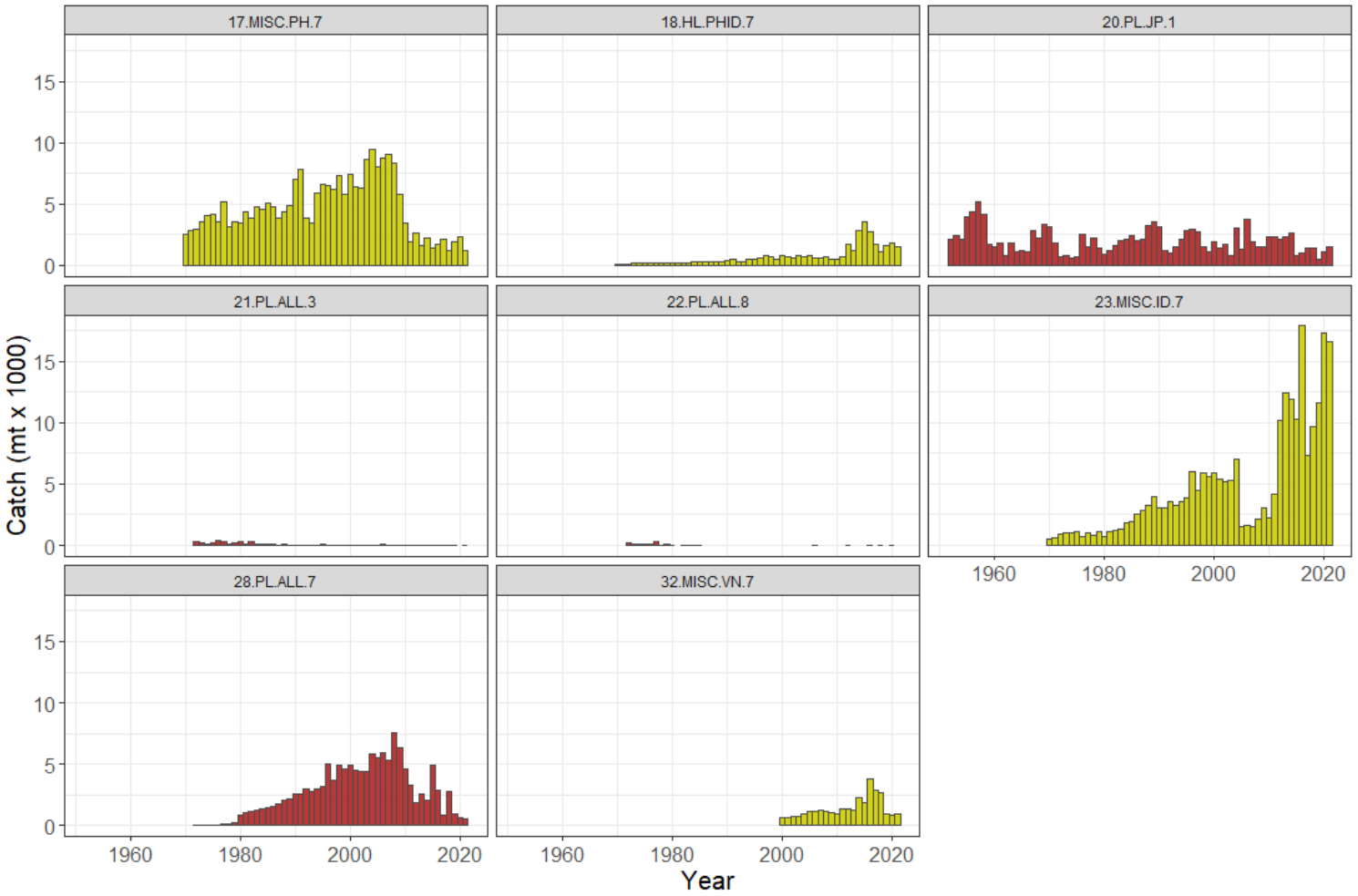

Figure 9: Time series of catches (mt) by fishery: other.

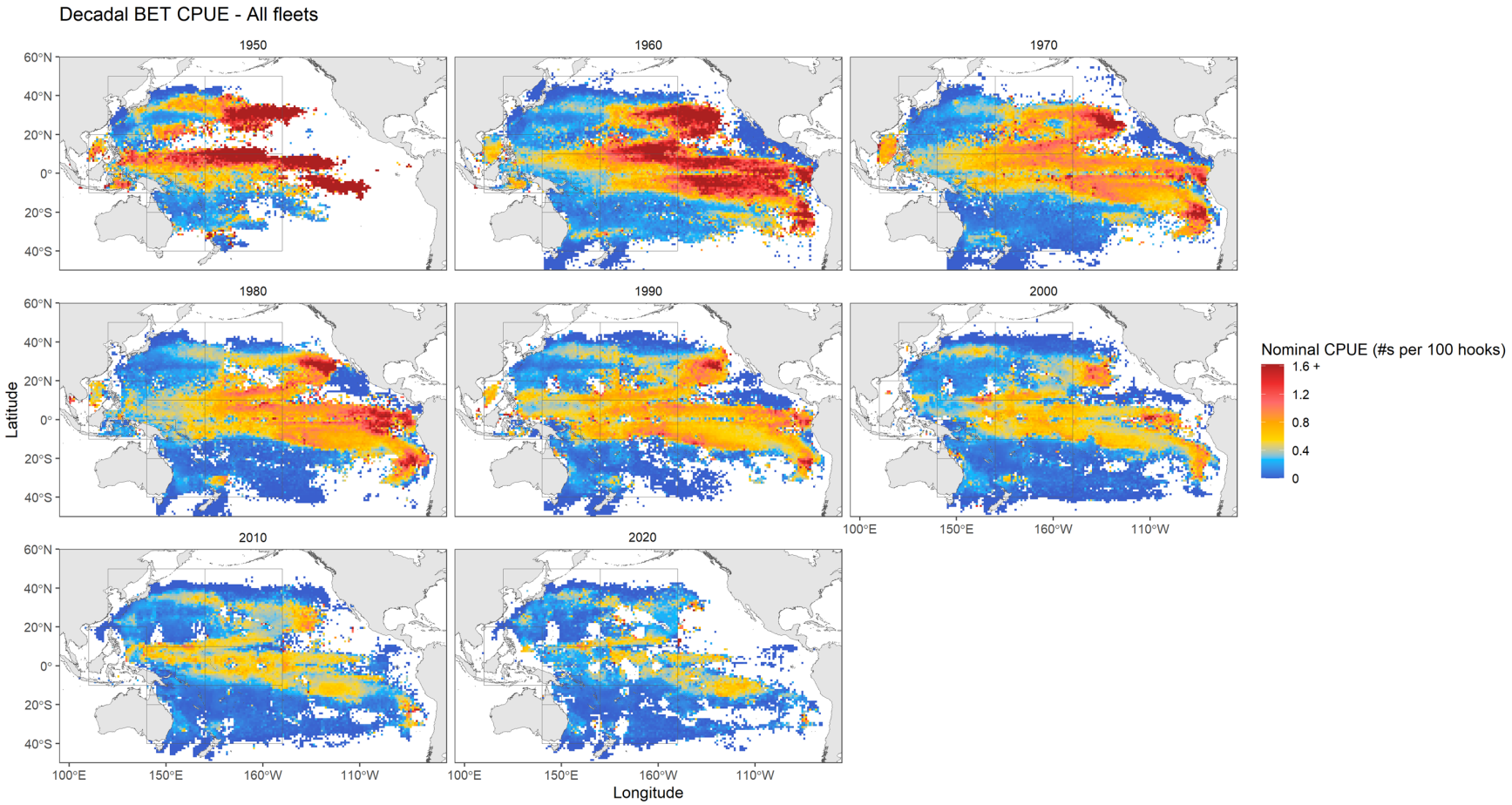

Figure 10: Spatial distribution of nominal longline CPUE (all fleets) for bigeye in the Pacific.

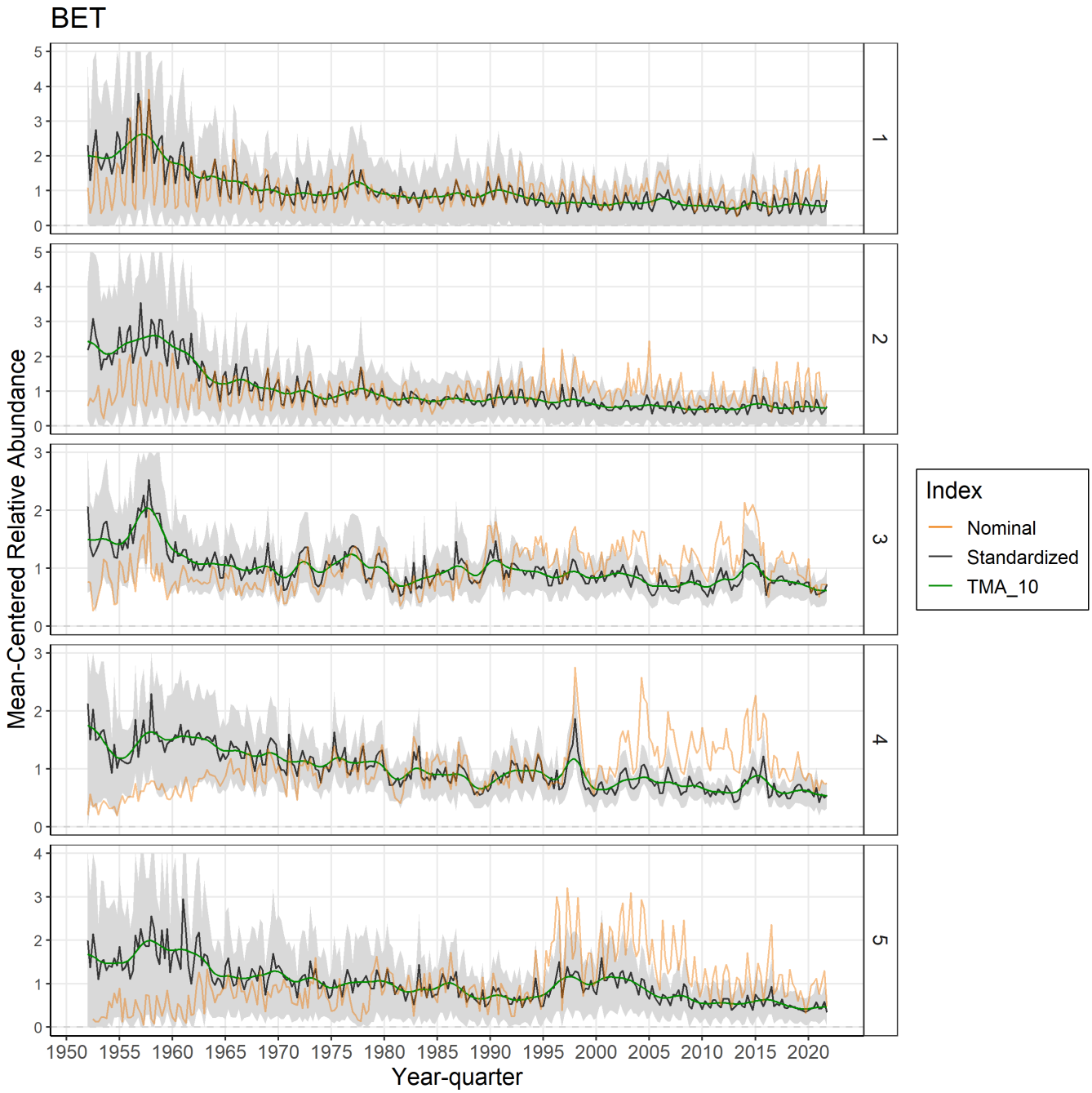

Figure 11: Standardised (black line) and nominal (orange) CPUE for the longline index fisheries in model regions 1-5. Grey band is 95% CI. Triangular moving average smoothing function applied to demonstrate overall trend (green line; smoothing window = 10).

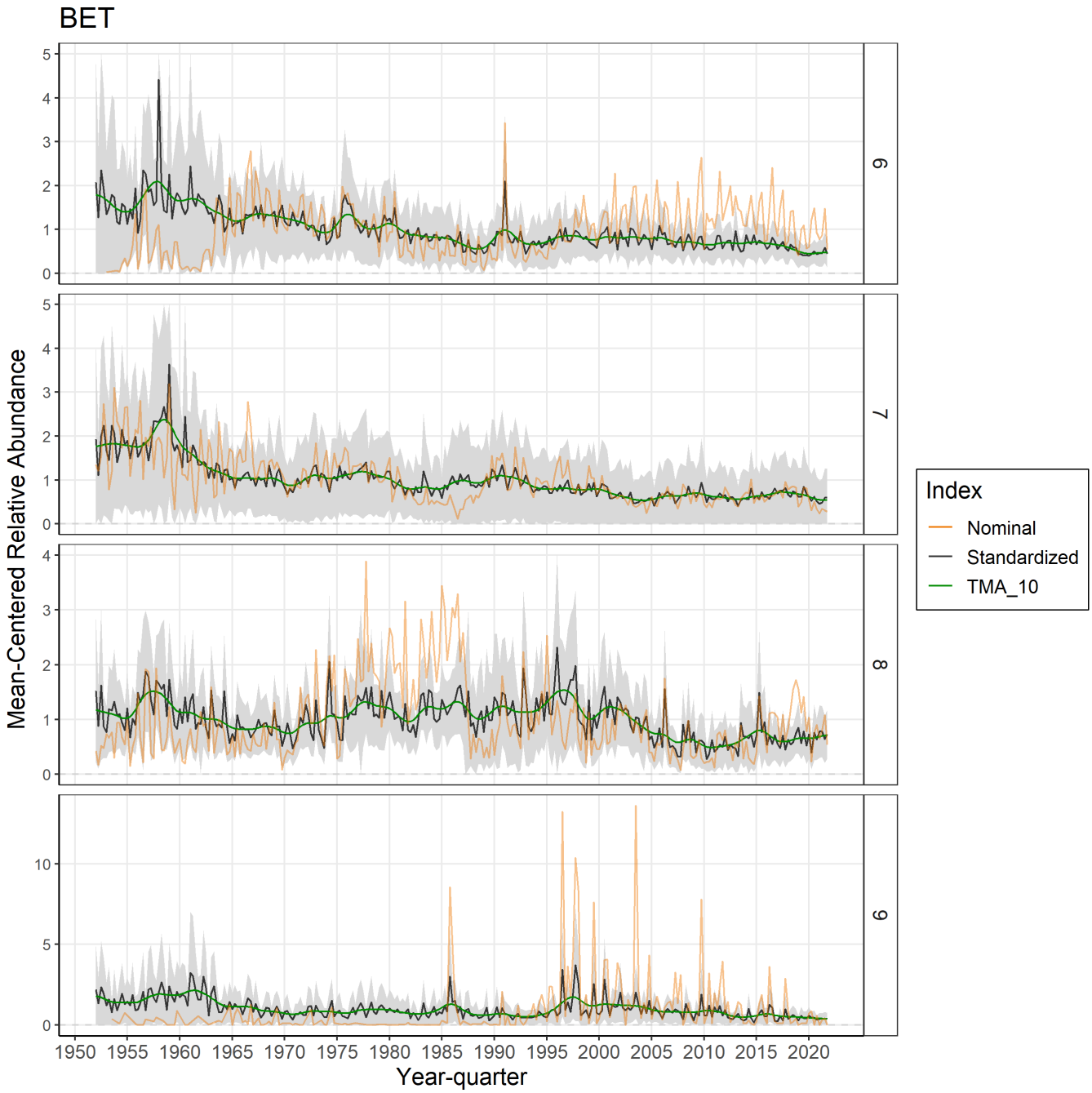

Figure 12: Standardised (black line) and nominal (orange) CPUE for the longline index fisheries in model regions 6-9. Grey band is 95% CI. Triangular moving average smoothing function applied to demonstrate overall trend (green line; smoothing window = 10).

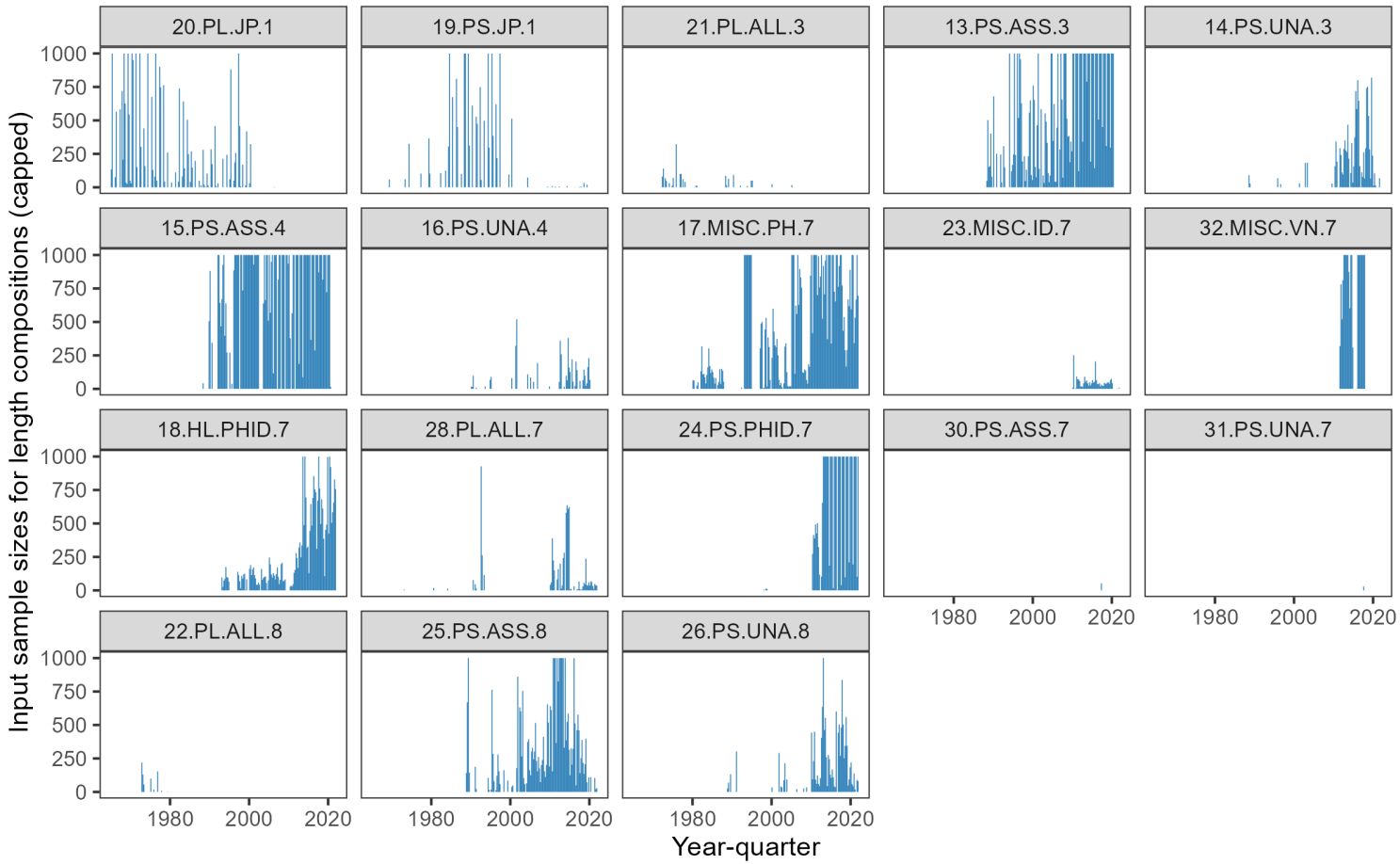

Figure 13: Plots of samples sizes (capped at 1,000) for length composition data for each fishery in the model across the model time period.

Figure 14: Plots of samples sizes (capped at 1,000) for weight composition data for each fishery in the model across the model time period.

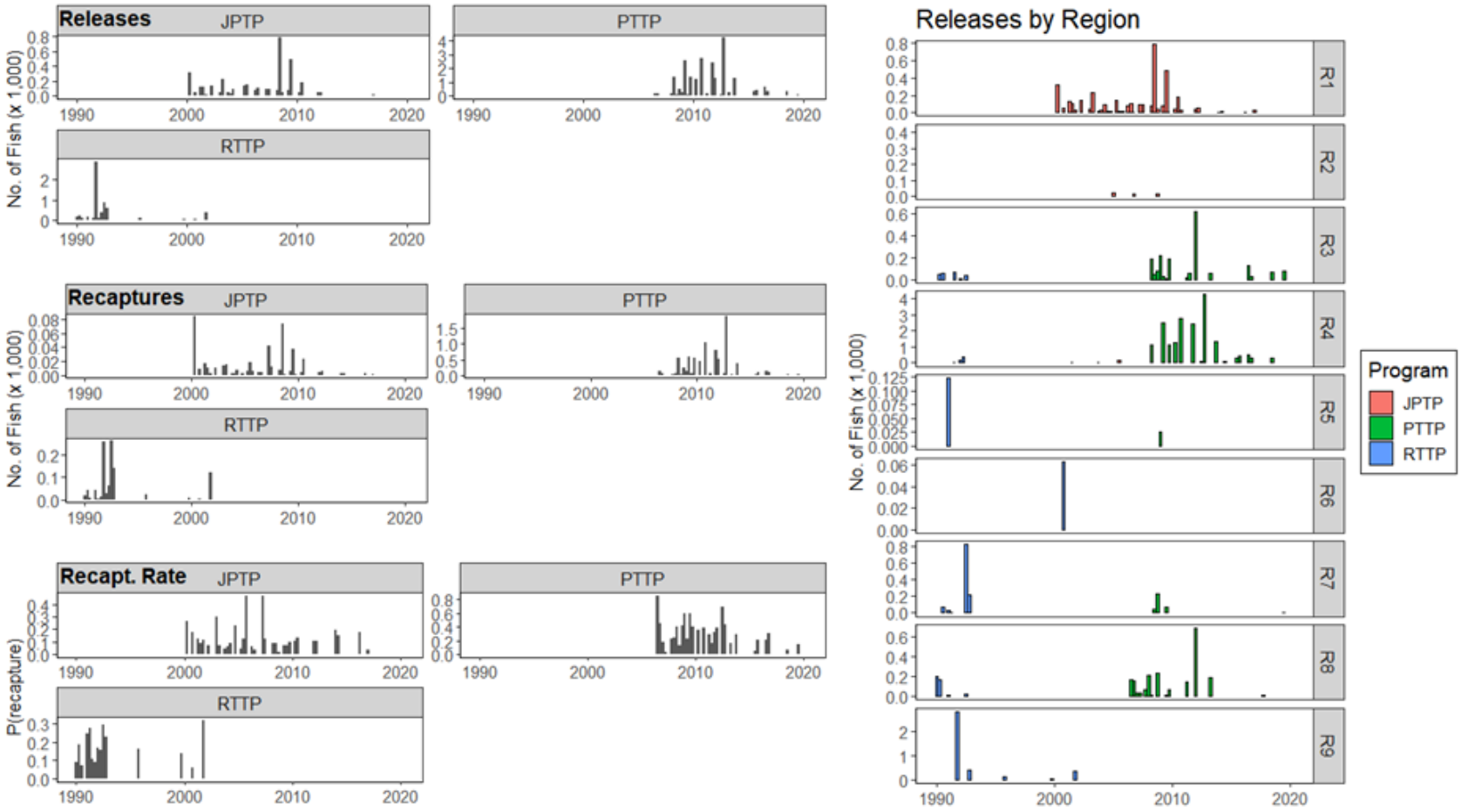

Figure 15: Summary plots of the number of releases, recaptures, and recapture rate of tags, by tagging program and region.

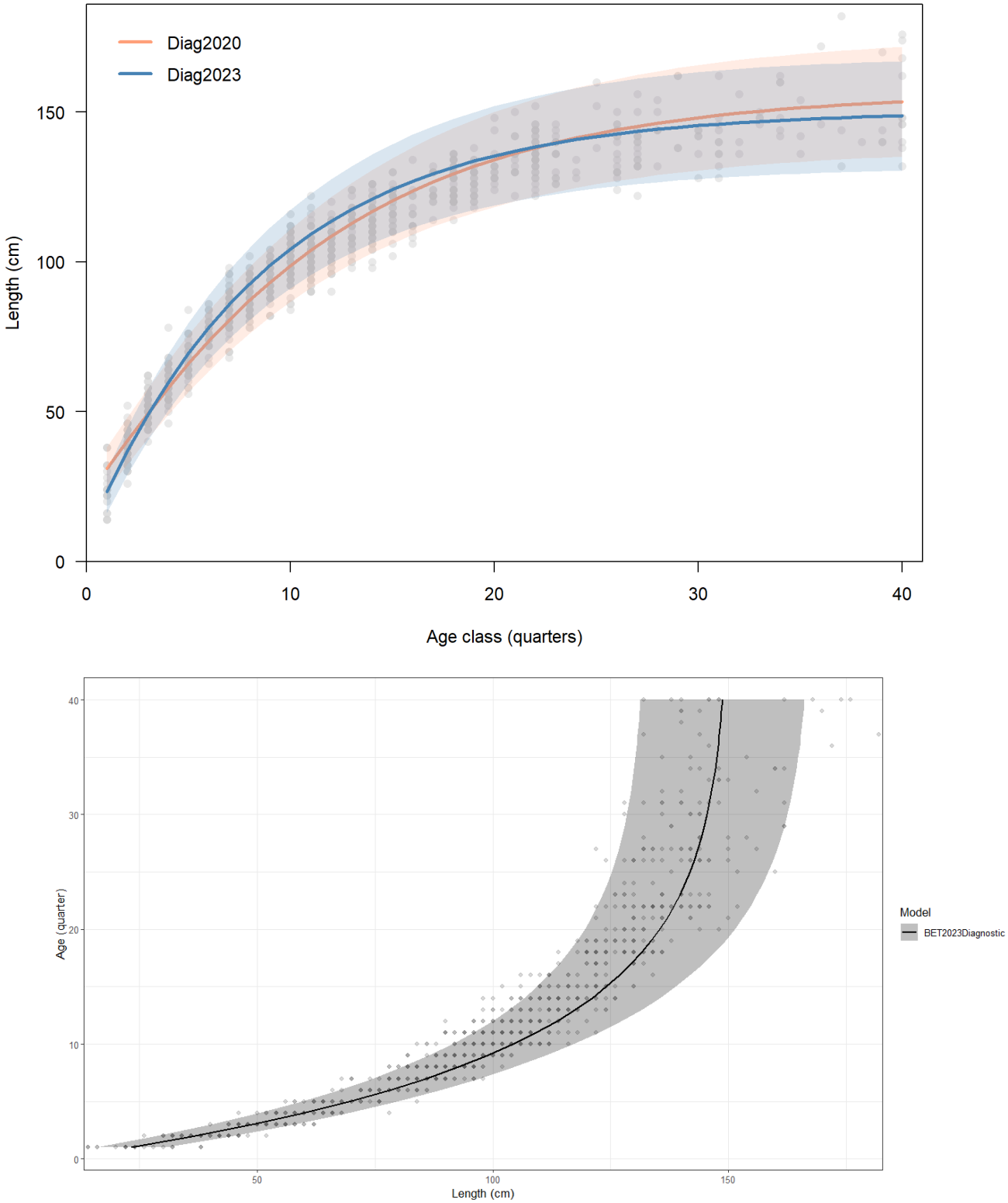

Figure 16: Growth curves plotted for the 2020 diagnostic model and 2023 diagnostic model (top). Transposed 2023 growth curve. The vertical distributions shows the range of observed ages for each length class (bottom). Plotted points are the conditional age-at-length observations used in the 2023 assessment, which is the same dataset used in 2020.

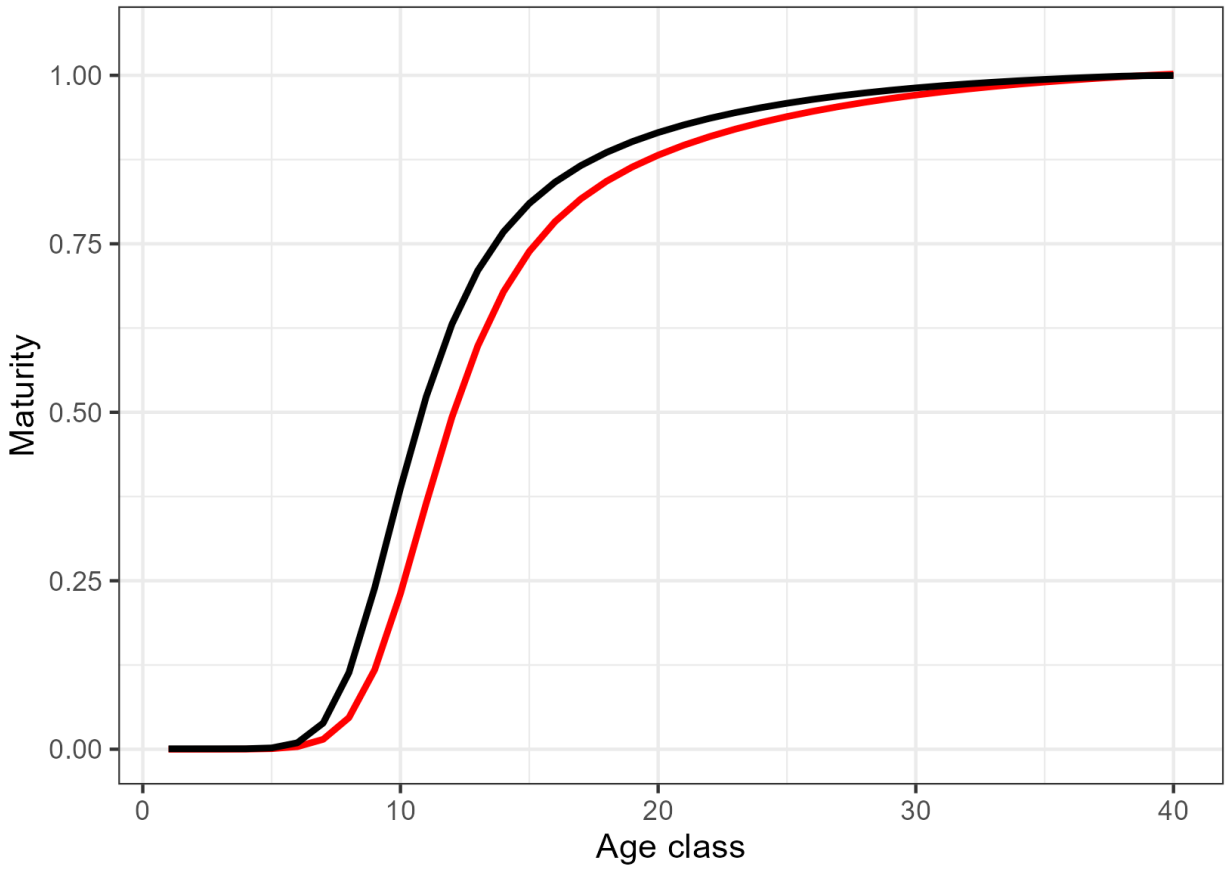

Figure 17: Maturity-at-age ogives for the 2020 diagnostic model (red) and the 2023 diagnostic model (black).

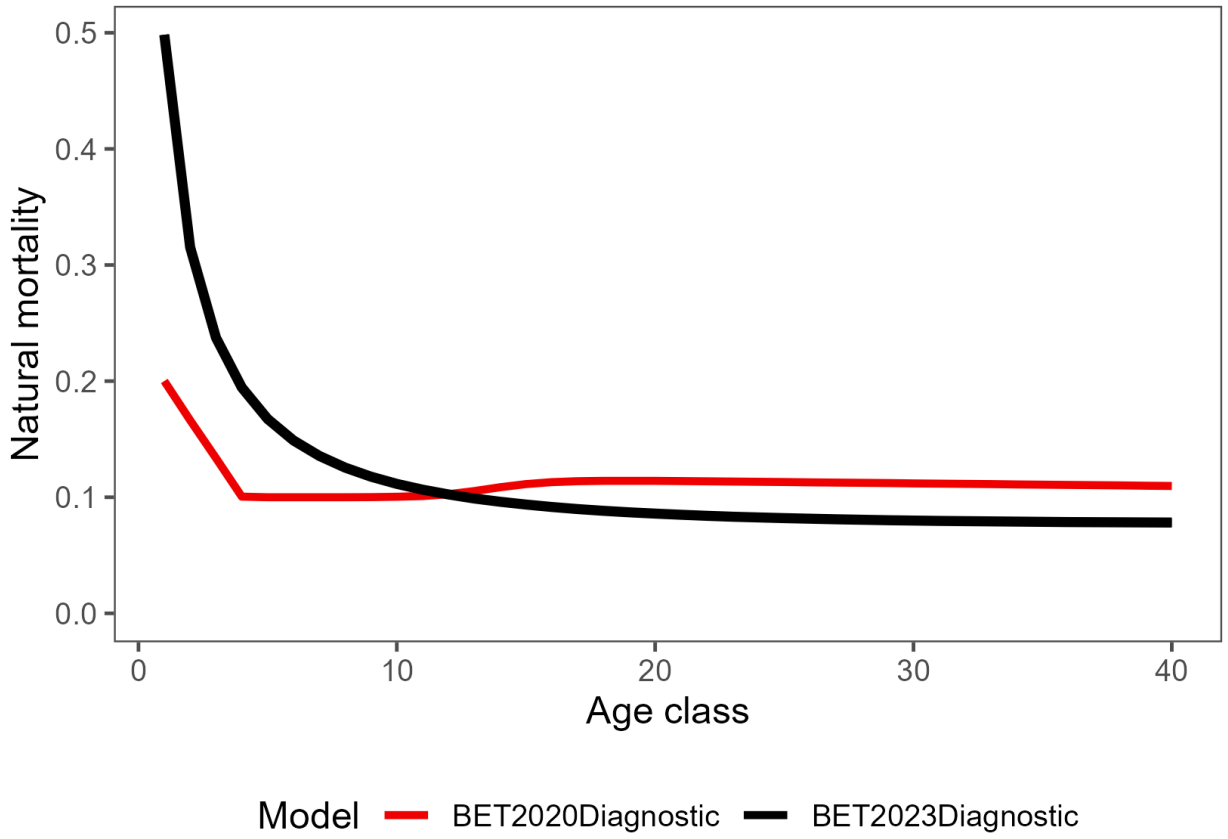

Figure 18: Natural mortality-at-age (quarters) for the 2020 diagnostic model (red) and the 2023 diagnostic model (black).

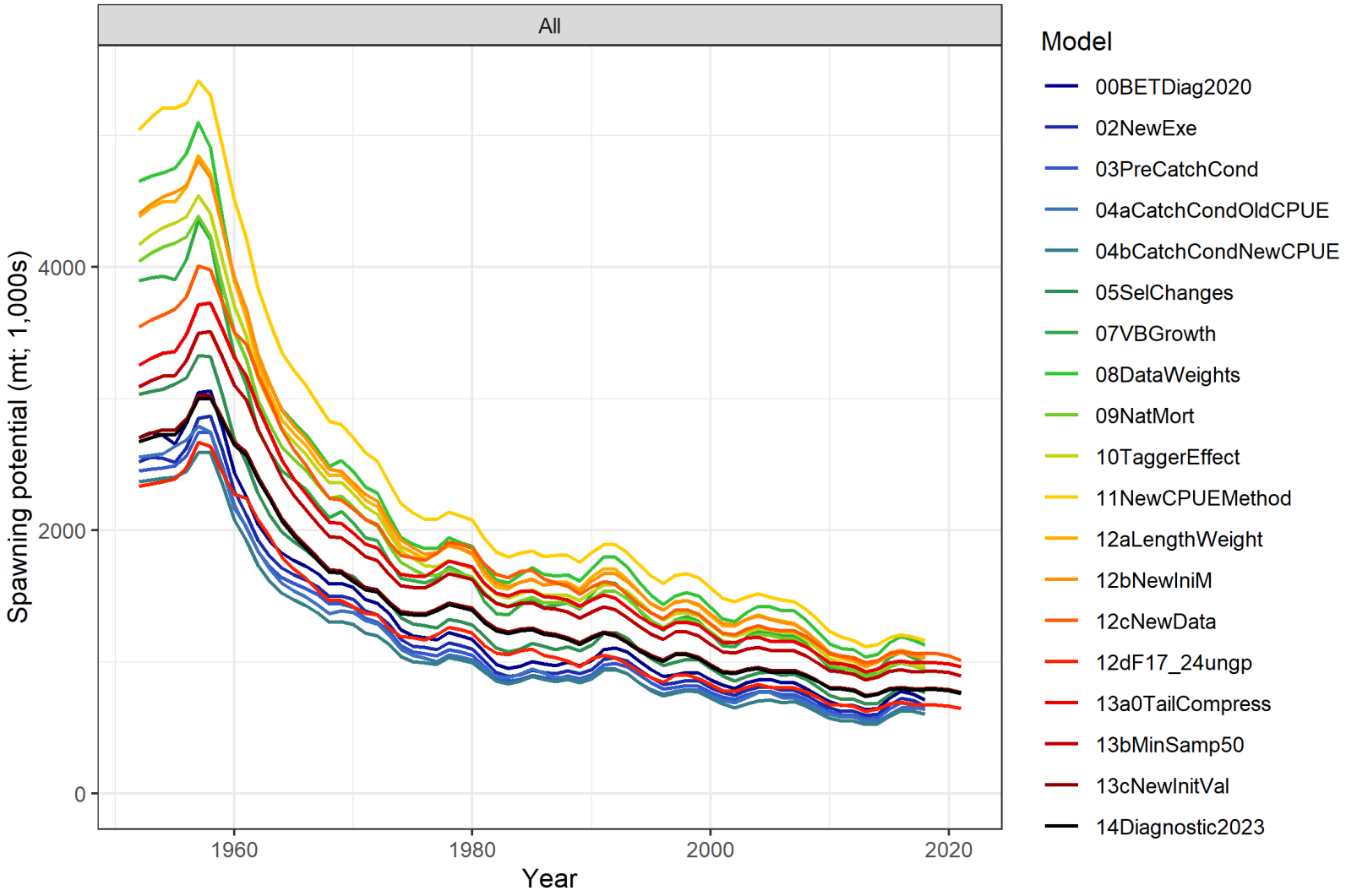

Figure 19: Estimated spawning potential, SB_t, trajectories for each of the main steps in the stepwise model development, with the 2023 diagnostic model shown in black.

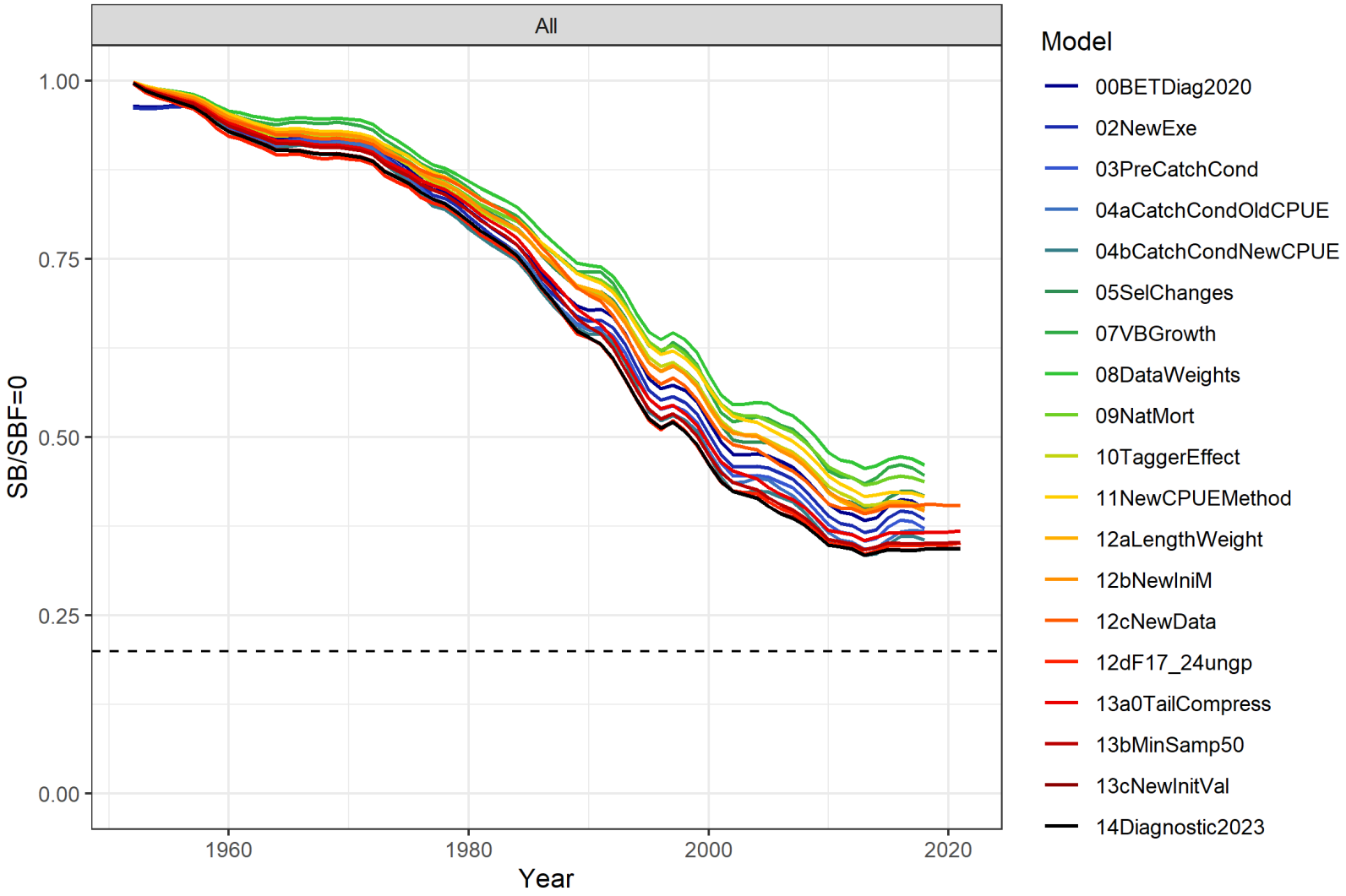

Figure 20: Estimated dynamic spawning depletion, $SB_t/SB_{t,F=0}$, trajectories for each of the main steps in the stepwise model development, with the 2023 diagnostic model shown black.

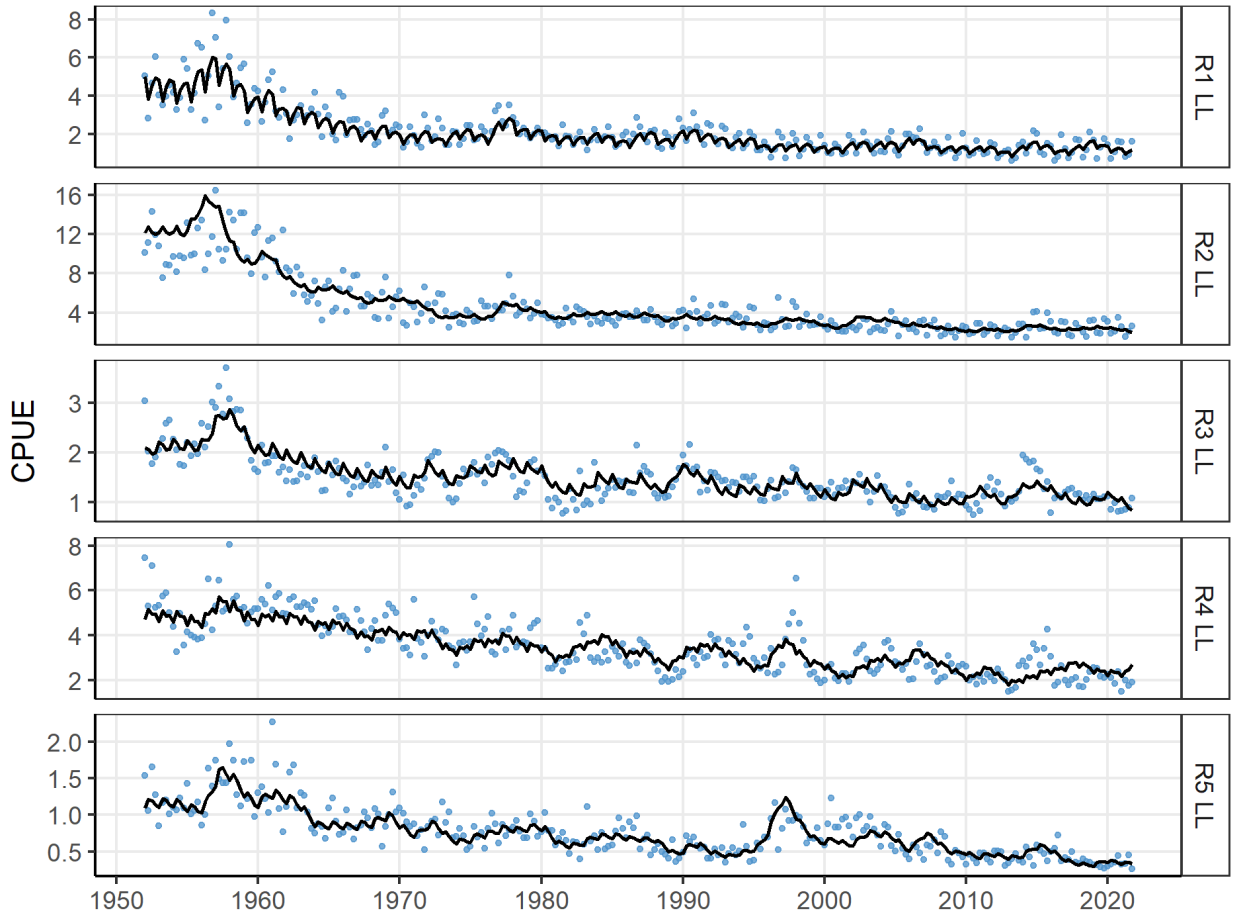

Figure 21: Fits (black line) to the standardised CPUE (blue dots) for the longline index fisheries in region 1-5.

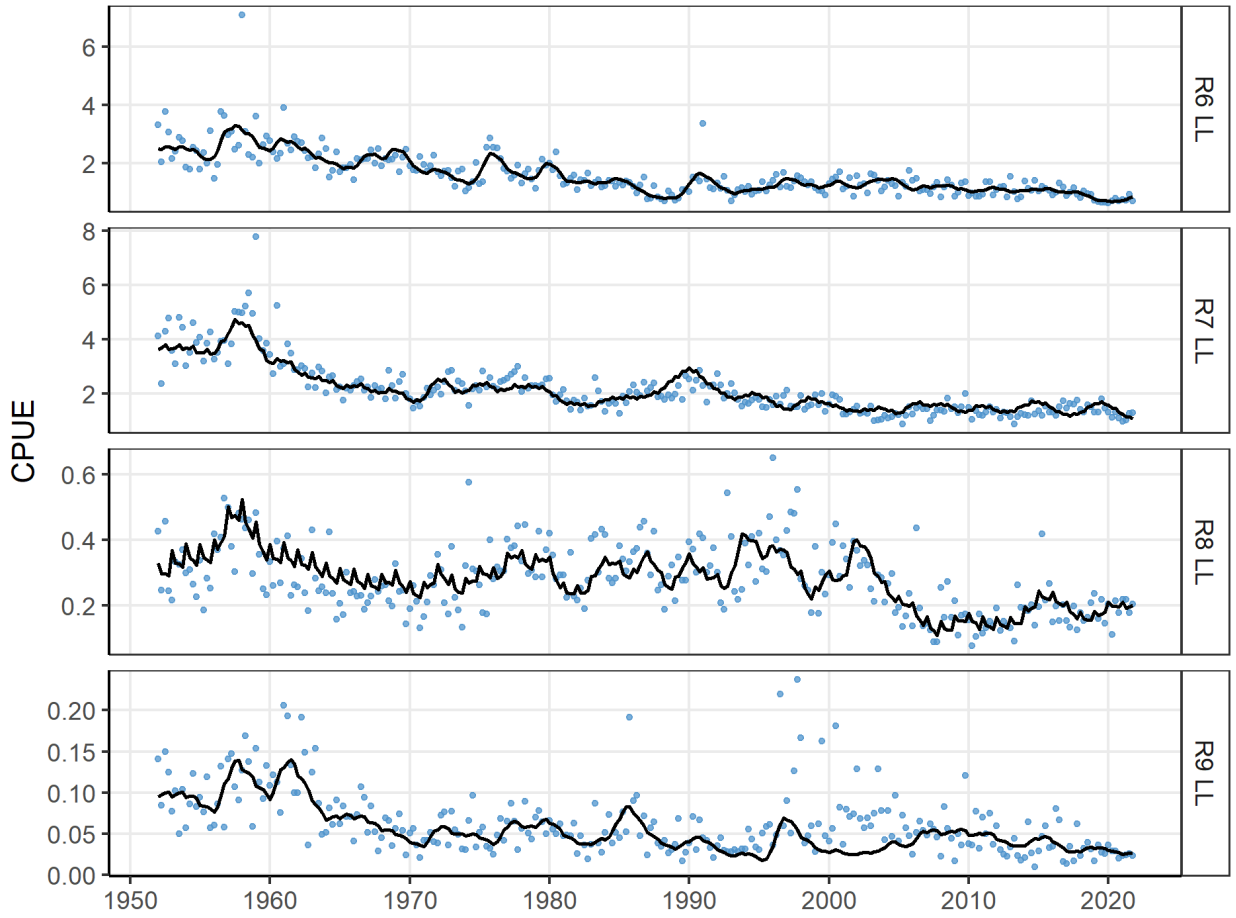

Figure 22: Fits (black line) to the standardised CPUE (blue dots) for the longline index fisheries in region 6-9.

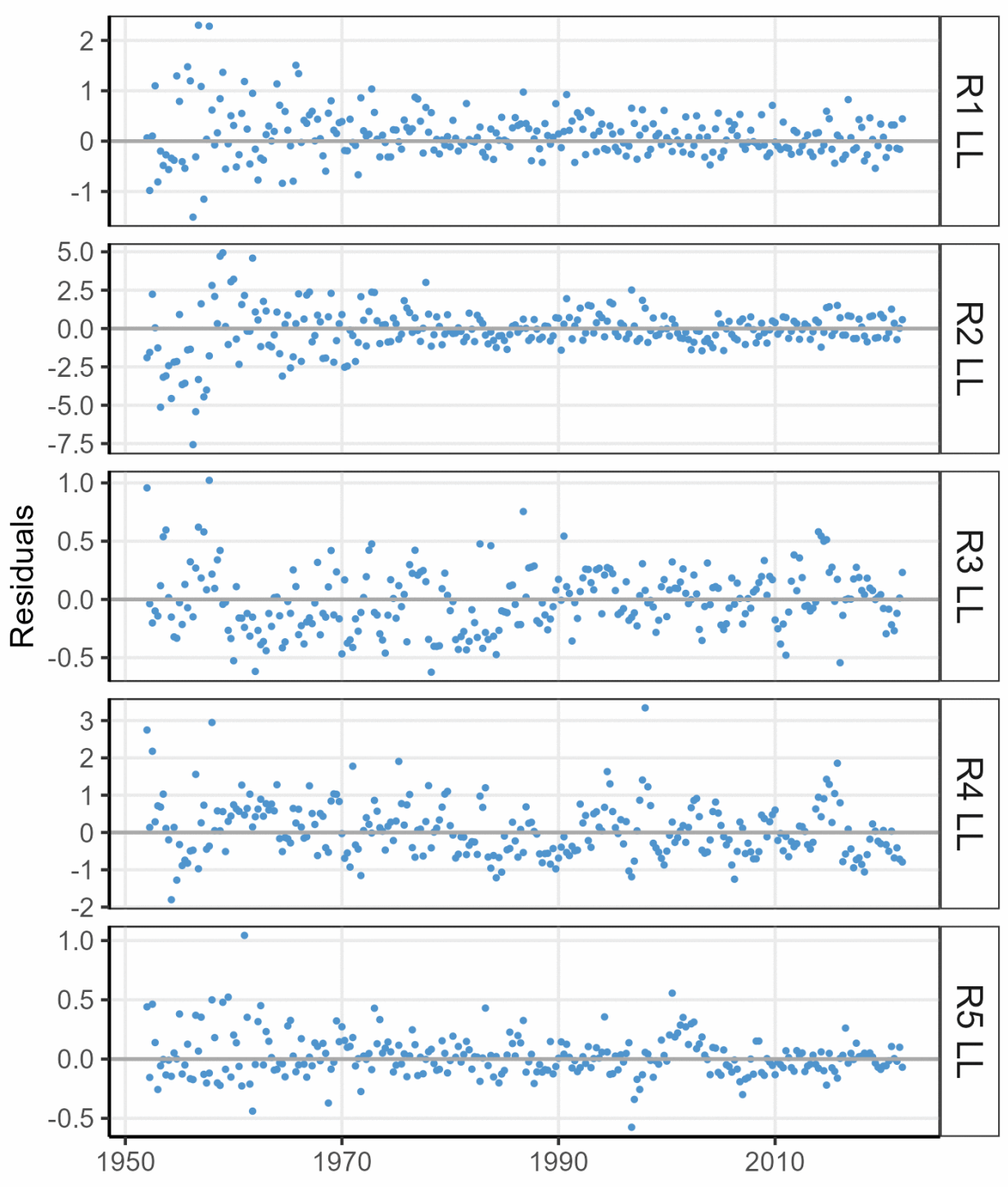

Figure 23: Residuals for the standardised CPUE for the longline index fisheries in region 1-5.

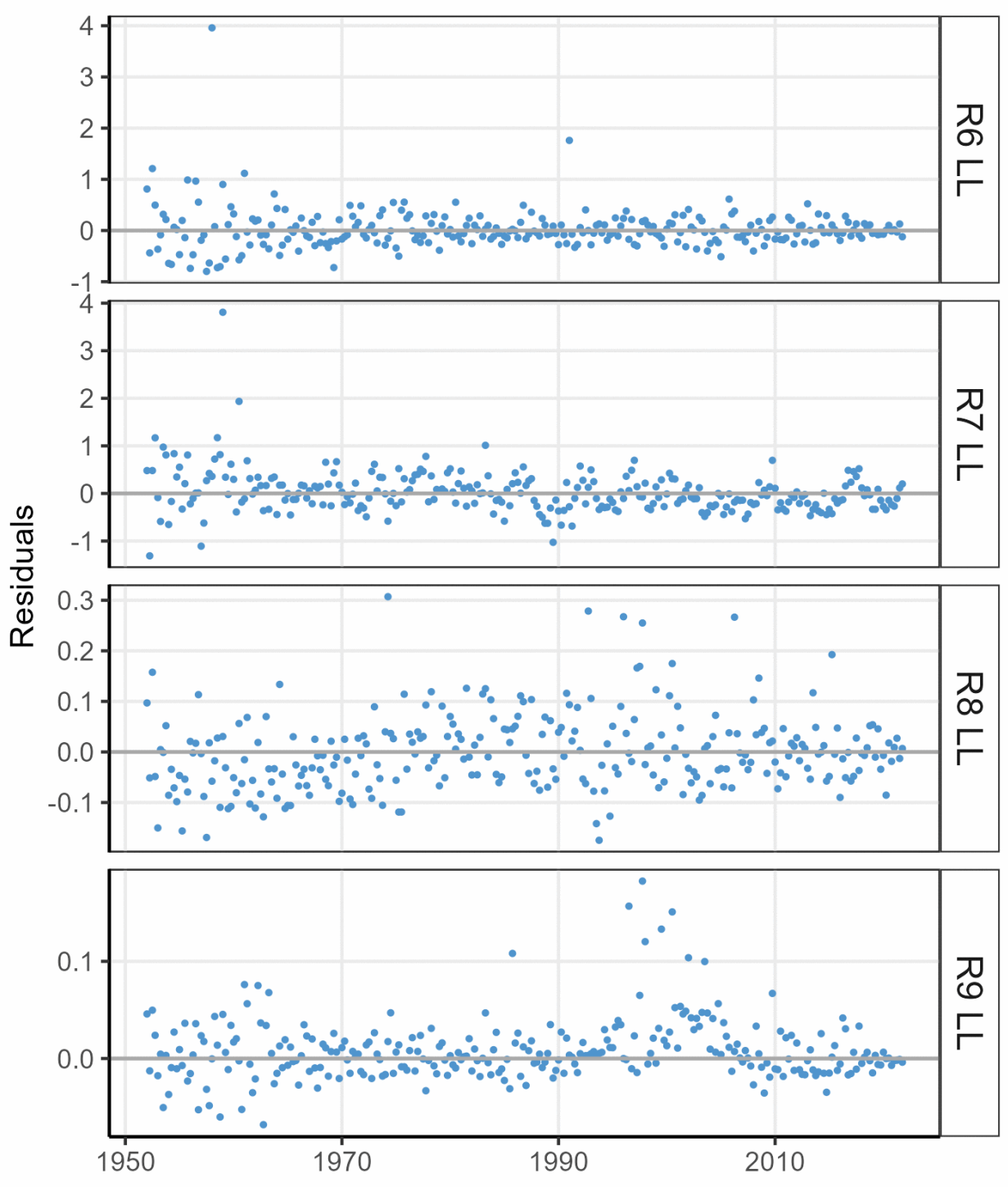

Figure 24: Residuals for the standardised CPUE for the longline index fisheries in region 6-9.

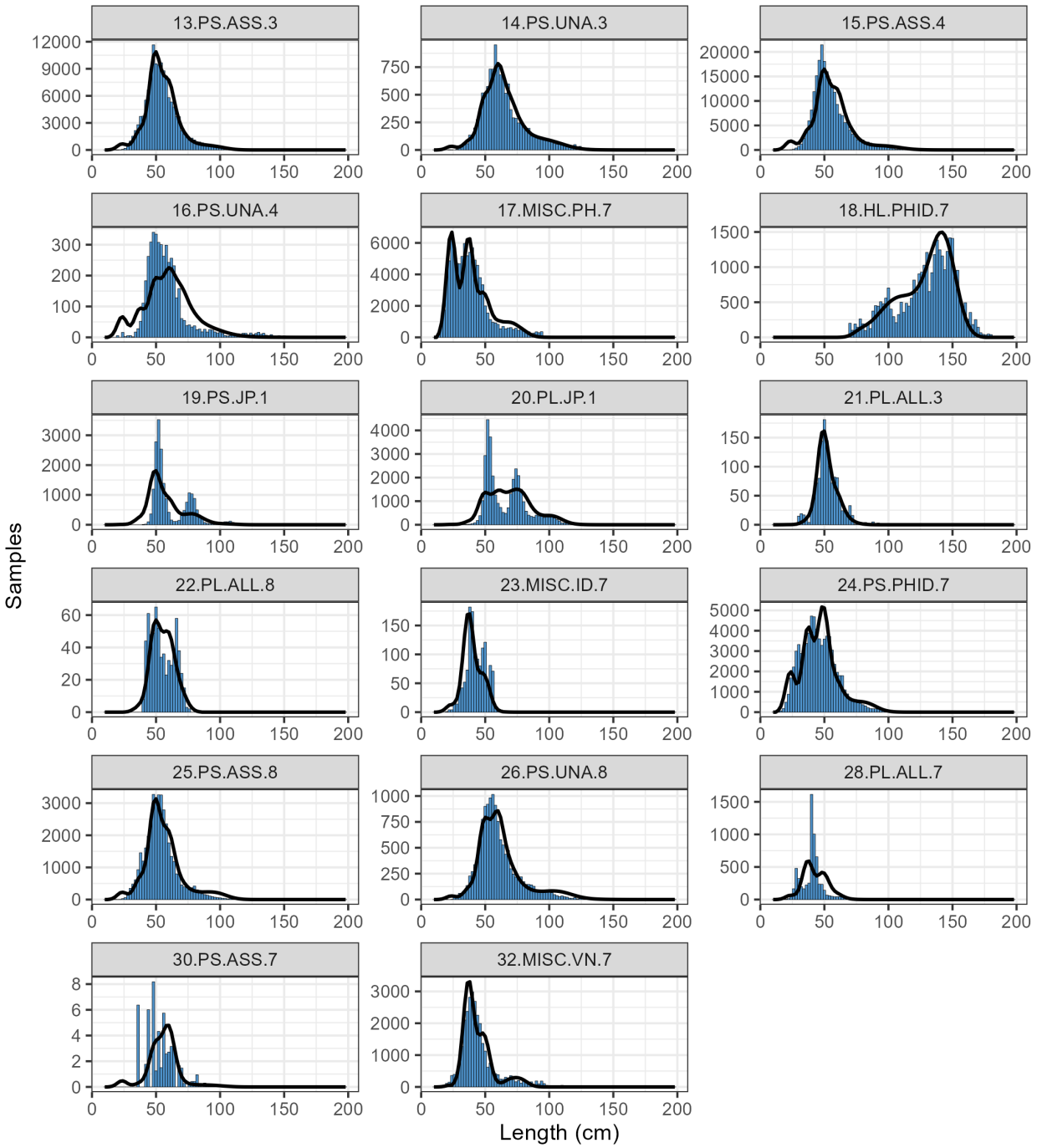

Figure 25: Composite (all time periods combined) observed (blue histograms) and predicted (black line) length frequency for fisheries with length frequency data for the 2023 diagnostic model.

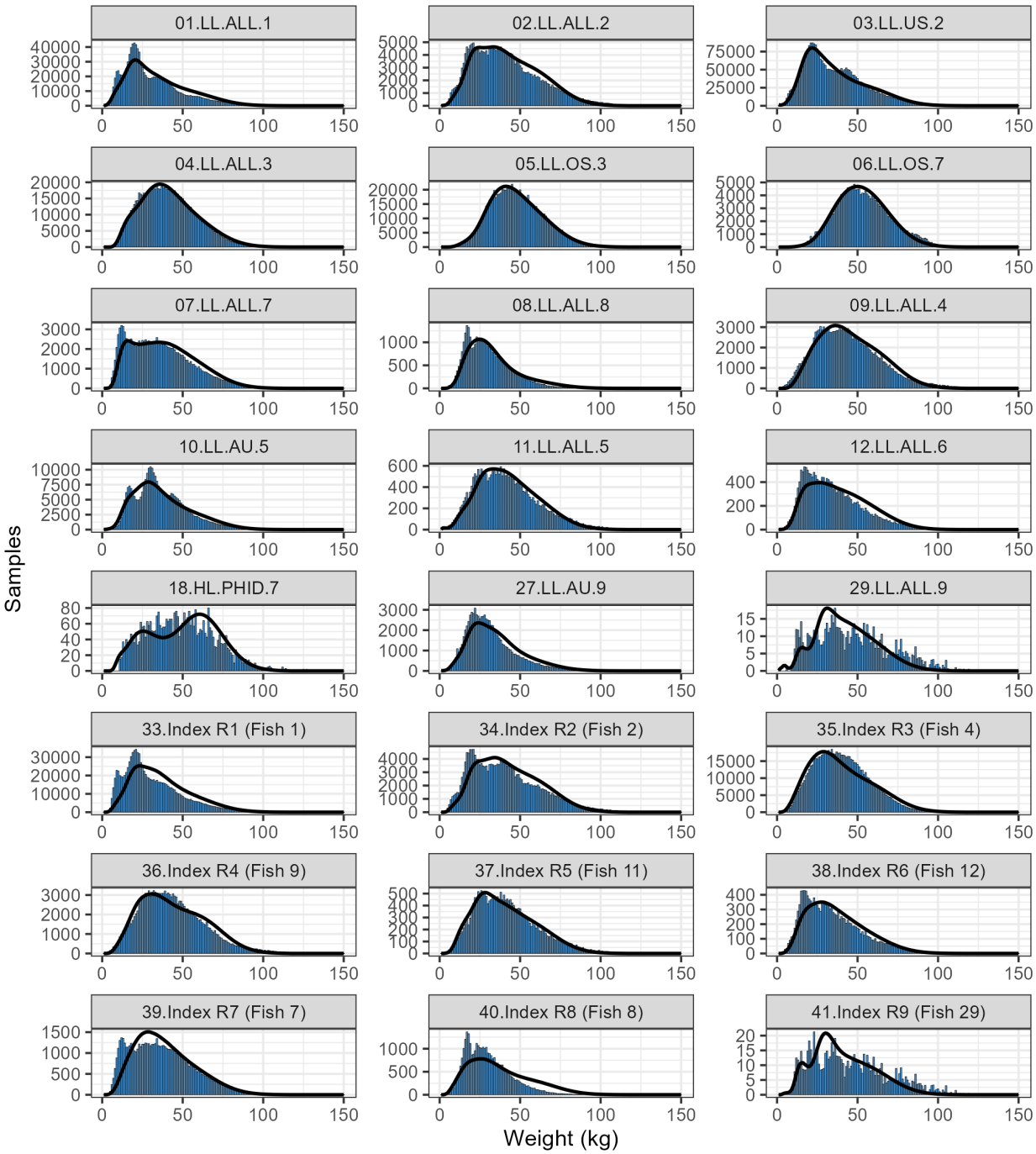

Figure 26: Composite (all time periods combined) observed (blue histograms) and predicted (black line) weight frequency for fisheries with weight frequency data for the 2023 diagnostic model.

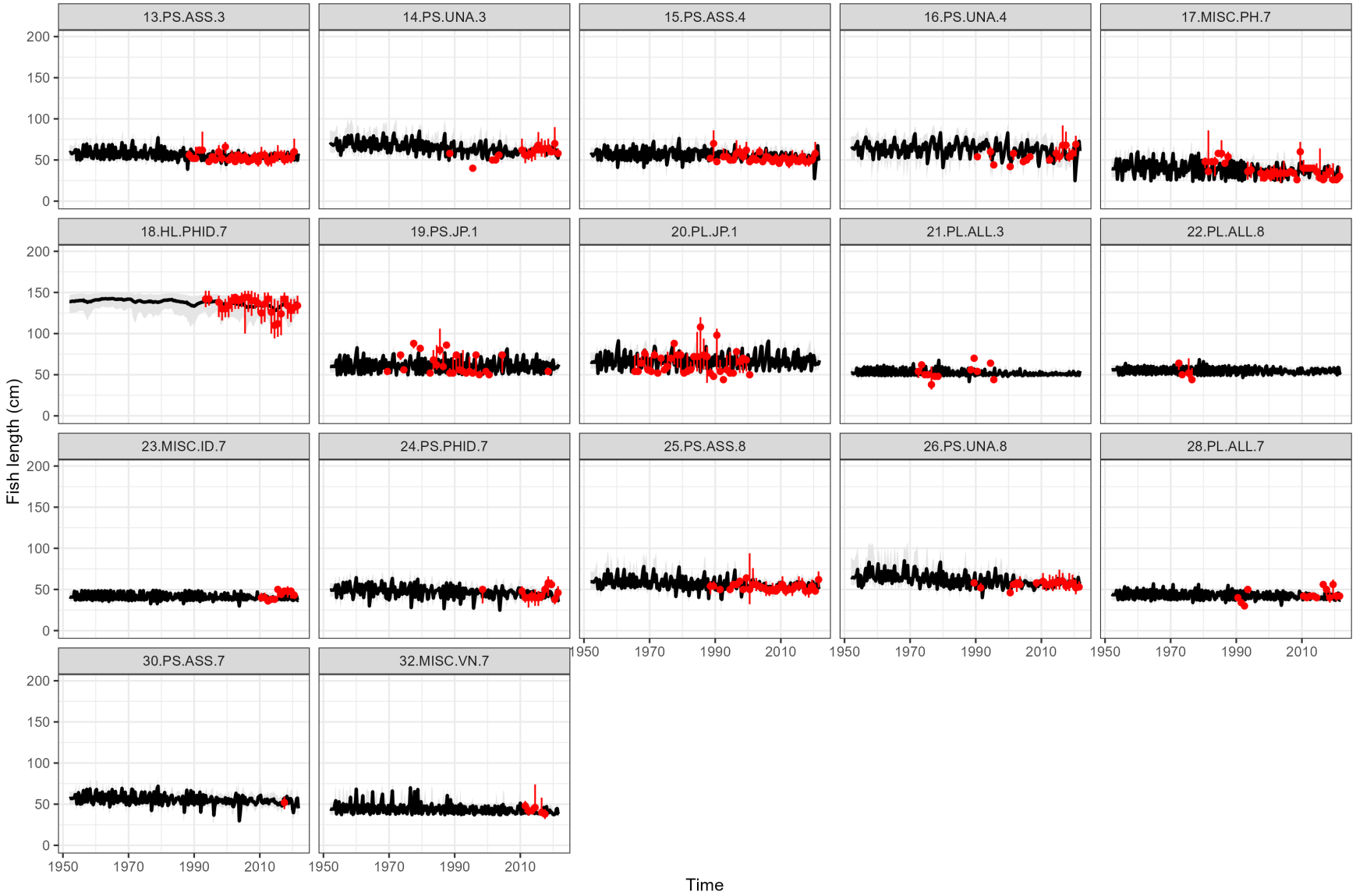

Figure 27: Observed (red points) and predicted (black line) median fish lengths (FL, cm) for the fisheries with length data for the 2023 diagnostic model. The uncertainty intervals (grey shading) represent the values encompassed by the 25% and 75% quantiles. Sampling data are by quarter and only length samples with more than 30 fish per quarter are plotted.

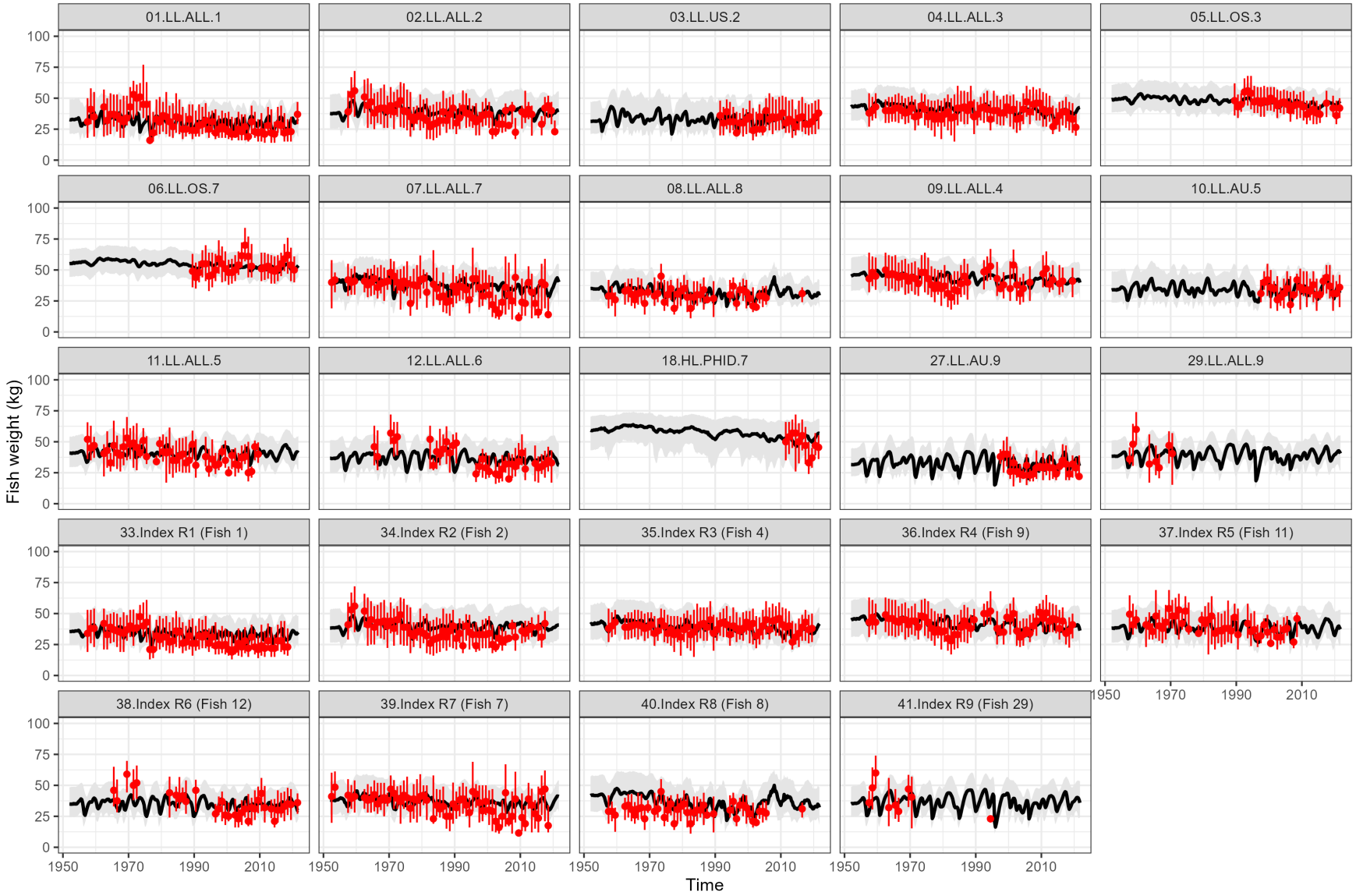

Figure 28: Observed (red points) and predicted (black line) median fish weights (kg) for the fisheries with weight data for the 2023 diagnostic model. The uncertainty intervals (grey shading) represent the values encompassed by the 25% and 75% quantiles. Sampling data are by quarter and only length samples with more than 30 fish per quarter are plotted.

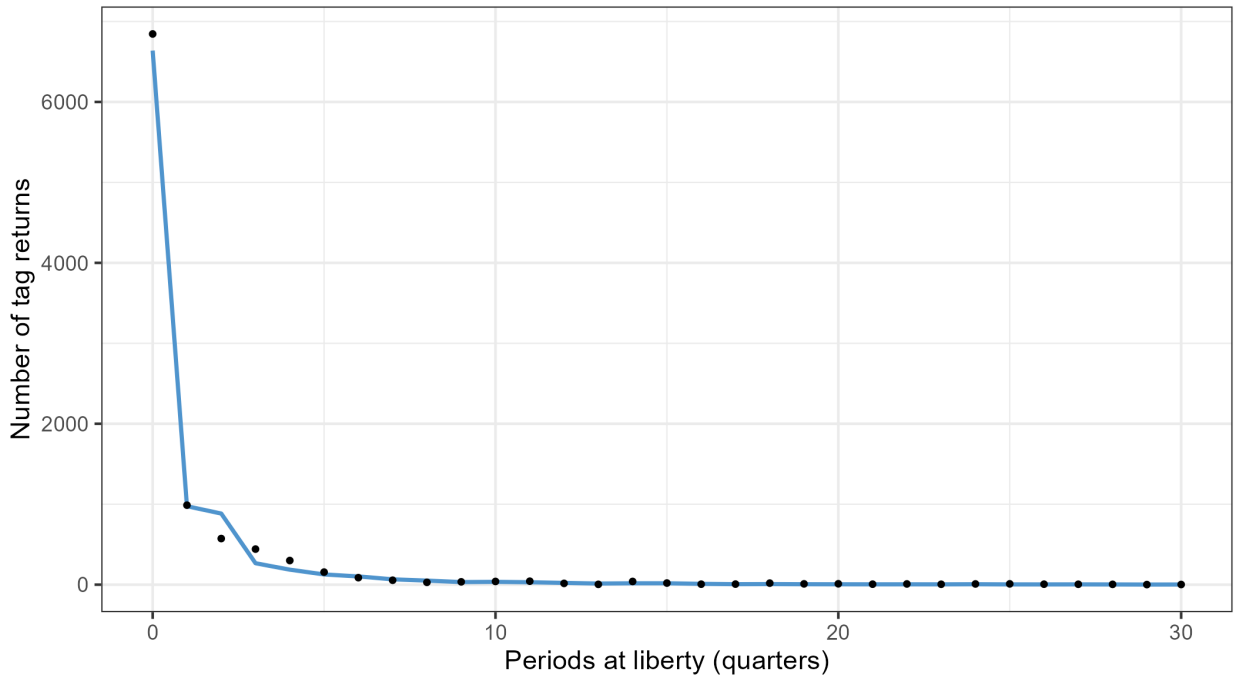

Figure 29: Observed (black points) and model-predicted (blue line) tag attrition across all tag release events for the 2023 diagnostic model.

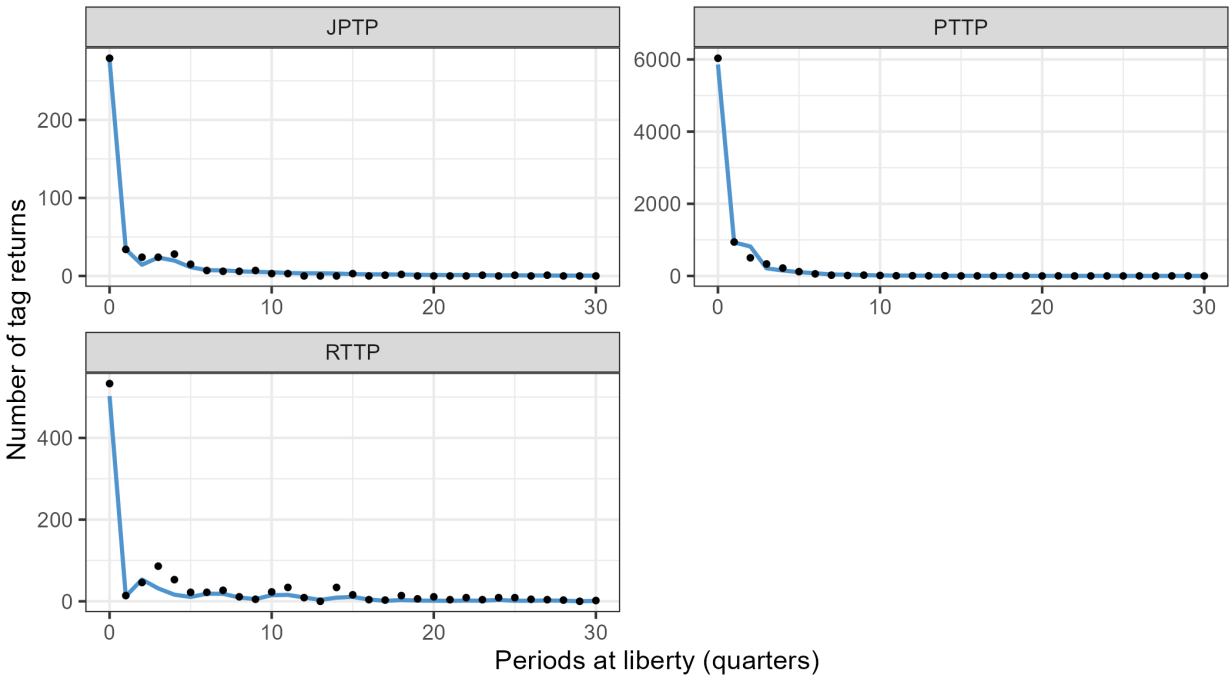

Figure 30: Observed (black points) and model-predicted (blue lines) tag attrition by tagging programme for the 2023 diagnostic model.

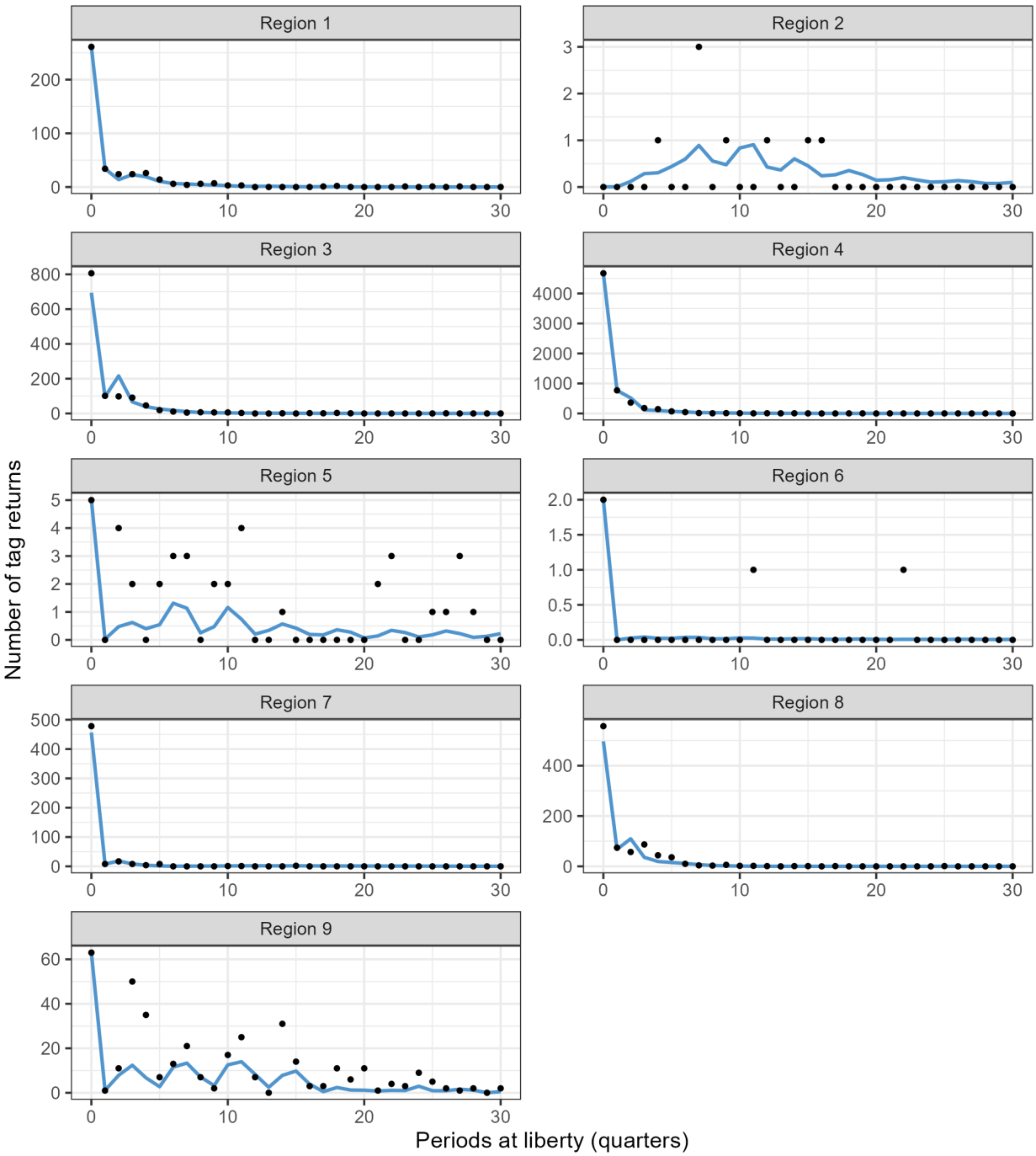

Figure 31: Observed (black points) and model-predicted (blue line) tag attrition by region for the 2023 diagnostic model.

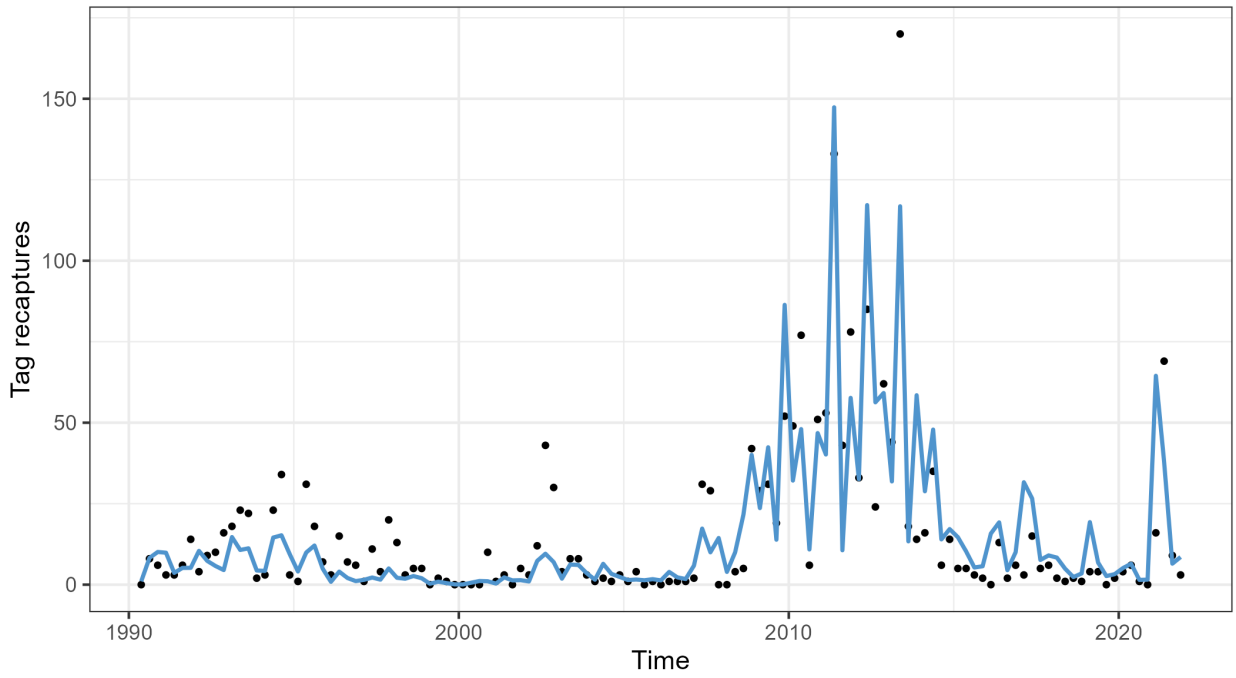

Figure 32: Observed (black points) and model-predicted (blue line) tag returns over time, with returns in the mixing period removed, for the 2023 diagnostic model across all tag release events with all tag recapture groupings aggregated.

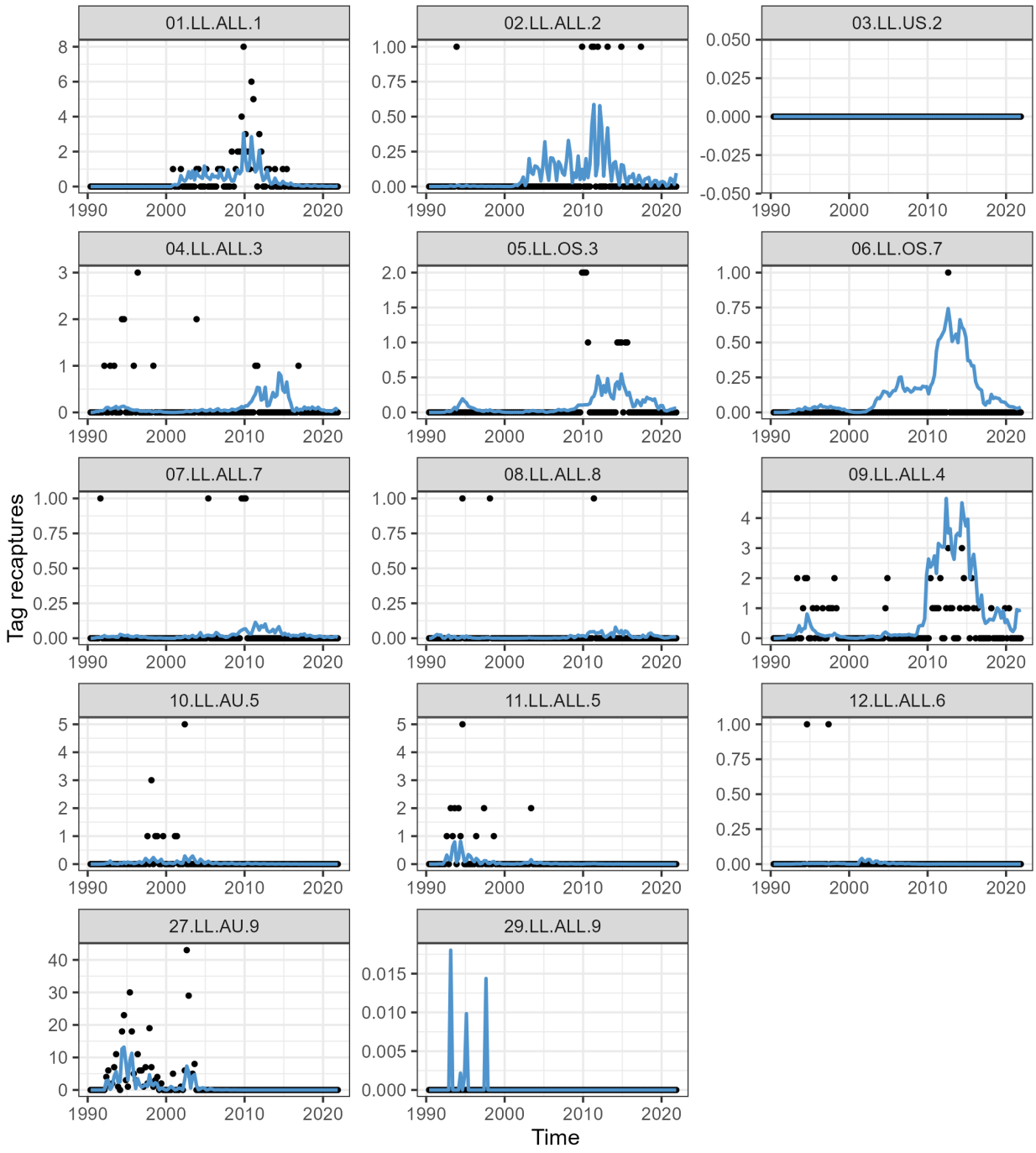

Figure 33: Observed black points) and model-predicted (blue line) tag returns over time, with returns in the mixing period removed, for the 2023 diagnostic model for longline fisheries.

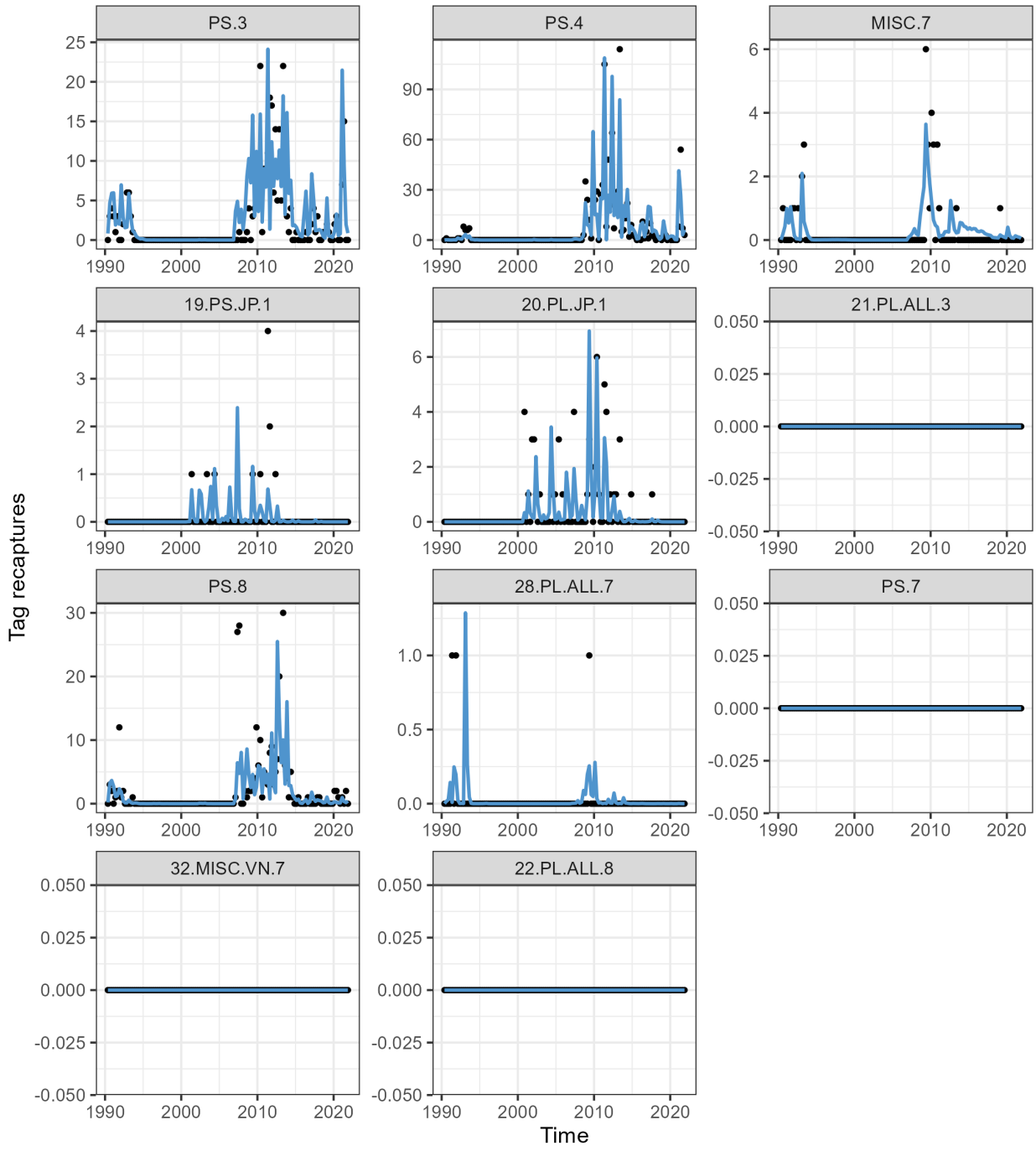

Figure 34: Observed black points) and model-predicted (blue line) tag returns over time, with returns in the mixing period removed, for the 2023 diagnostic model for other (non-longline) fisheries.

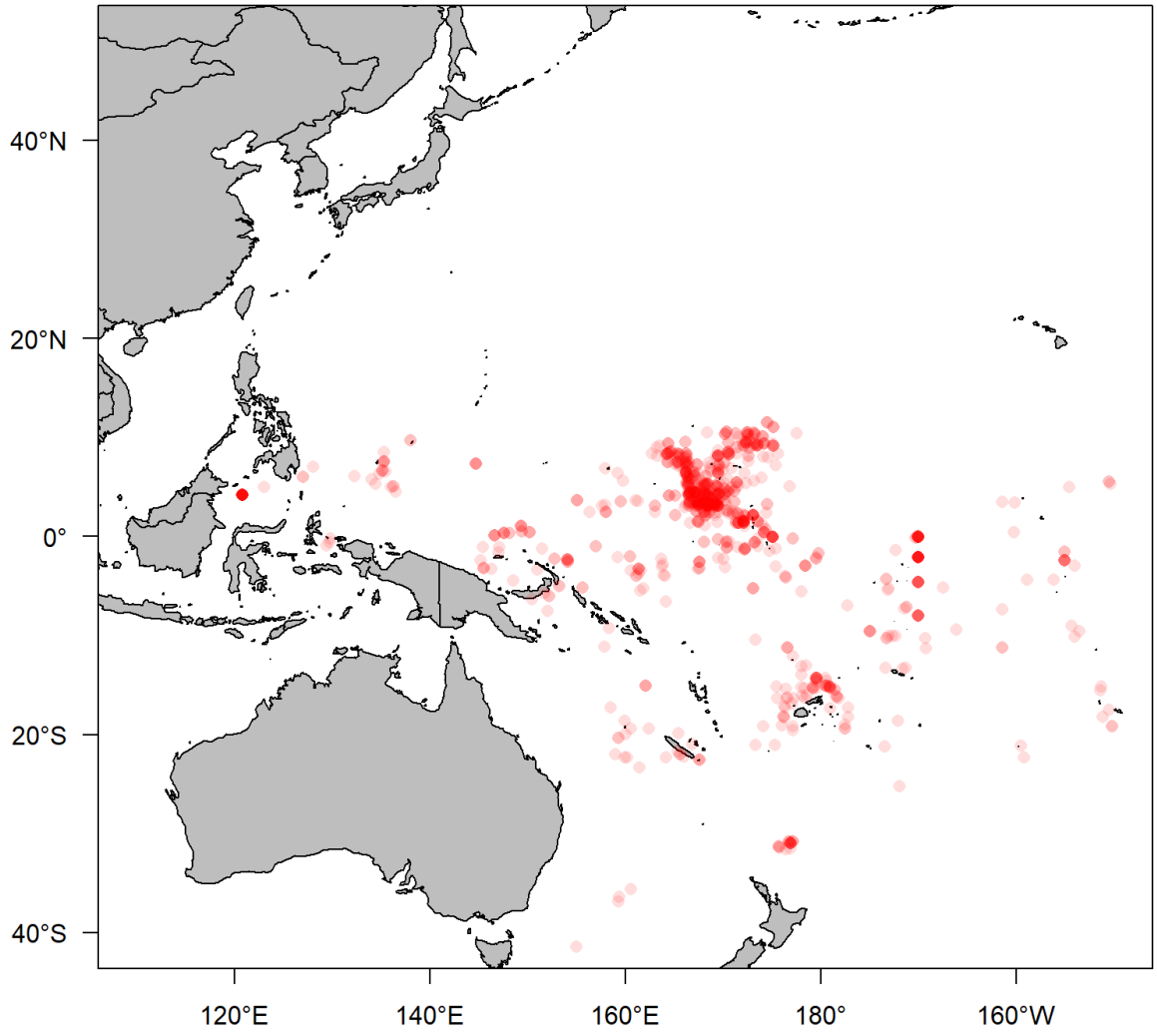

Figure 35: Sample locations of otoliths ($n = 1004$) used in the assessment model to inform internal growth estimation. Single otoliths are shown as pink circles and overlapping otoliths as progressively more saturated red circles.

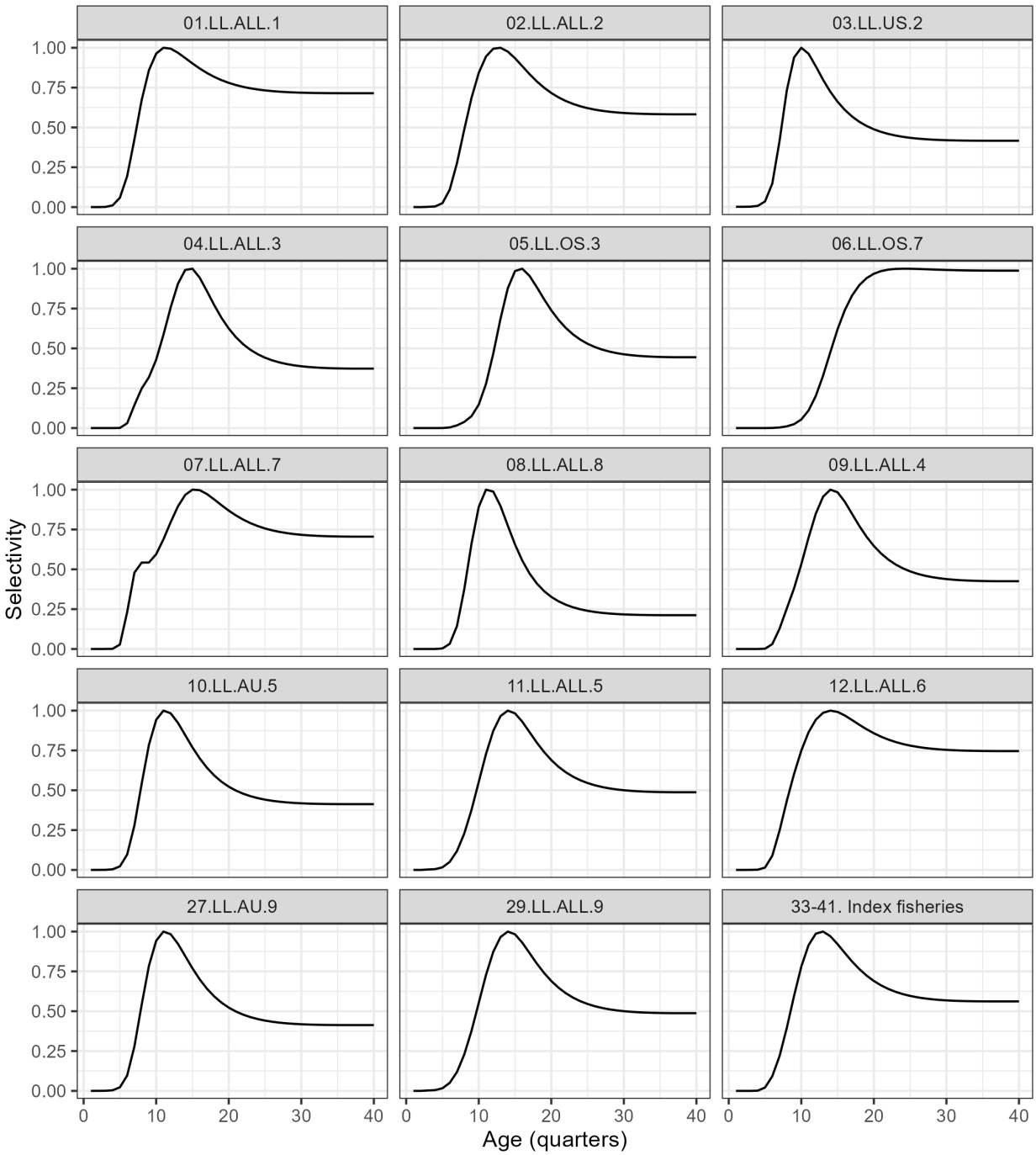

Figure 36: Age-specific (quarters) selectivity curves for longline fisheries, with one panel (bottom right) for the common (grouped) selectivity across all index fisheries. Selectivity is grouped between fisheries 10 and 27, and also between fisheries 11 and 29. The electivity for fishery 6.LL.OS.7 was encouraged (through a penalty on the objective function) to be non-decreasing with age.

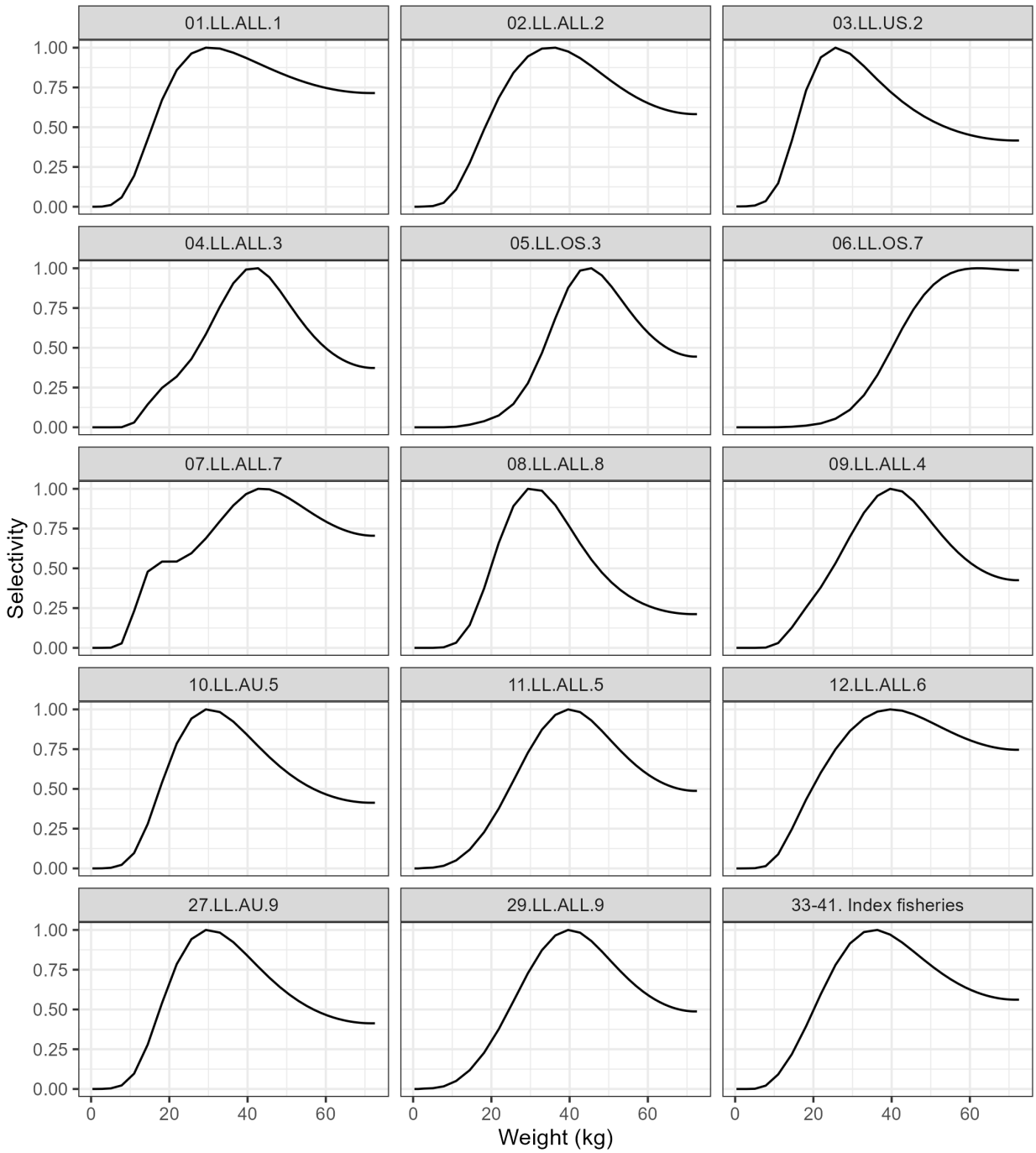

Figure 37: Weight-specific (kg) selectivity curves for longline fisheries, with one panel (bottom right) the common (grouped) selectivity across all index fisheries. Selectivity is grouped between fisheries 10 and 27, and also between fisheries 11 and 29. The selectivity for fishery 6.LL.OS.7 was encouraged (through a penalty on the objective function) to be non-decreasing with age.

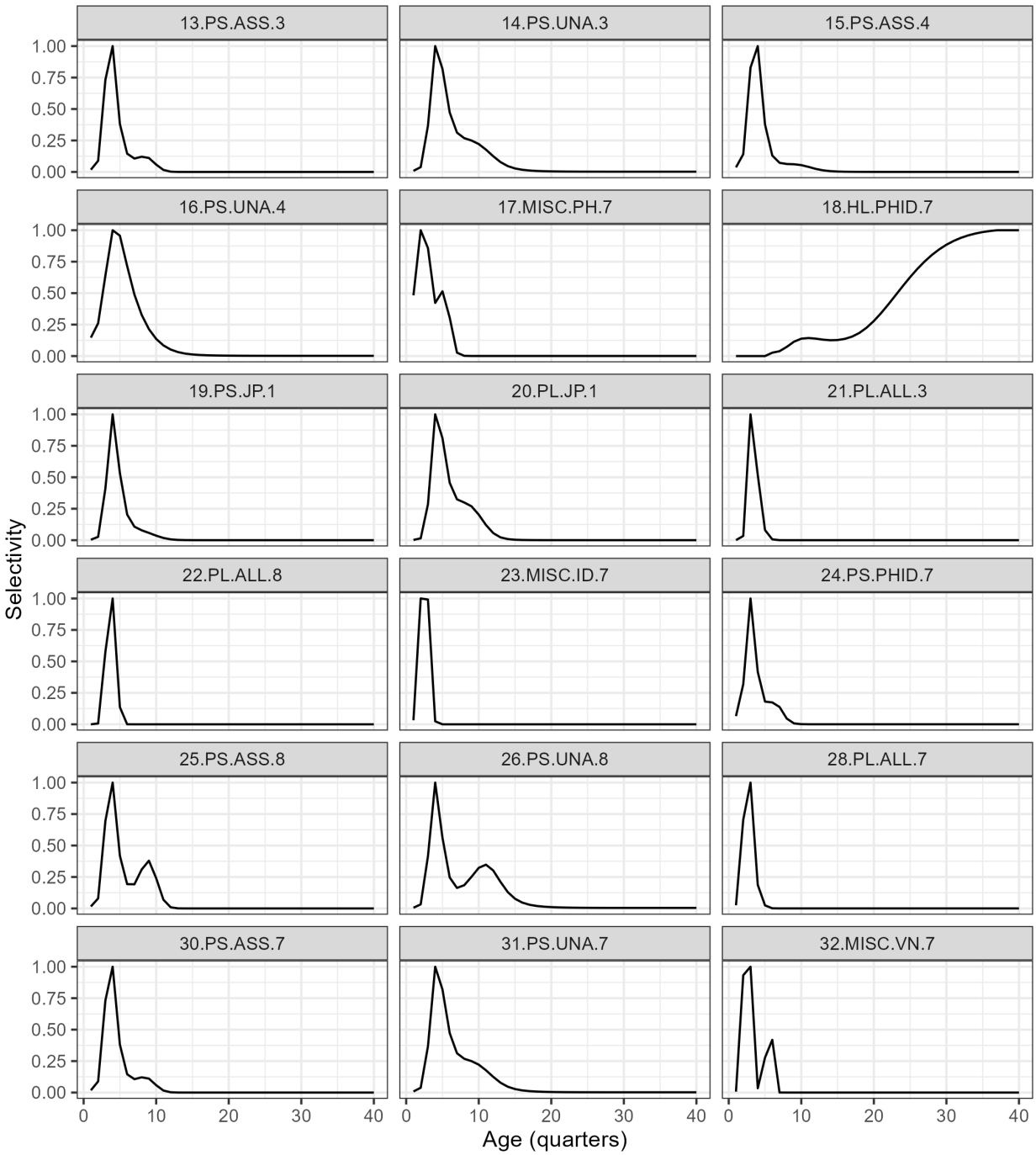

Figure 38: Age-specific (quarters) selectivity curves for non-longline fisheries. Selectivity is grouped between fisheries 13 and 30, and also between fisheries 14 and 31.

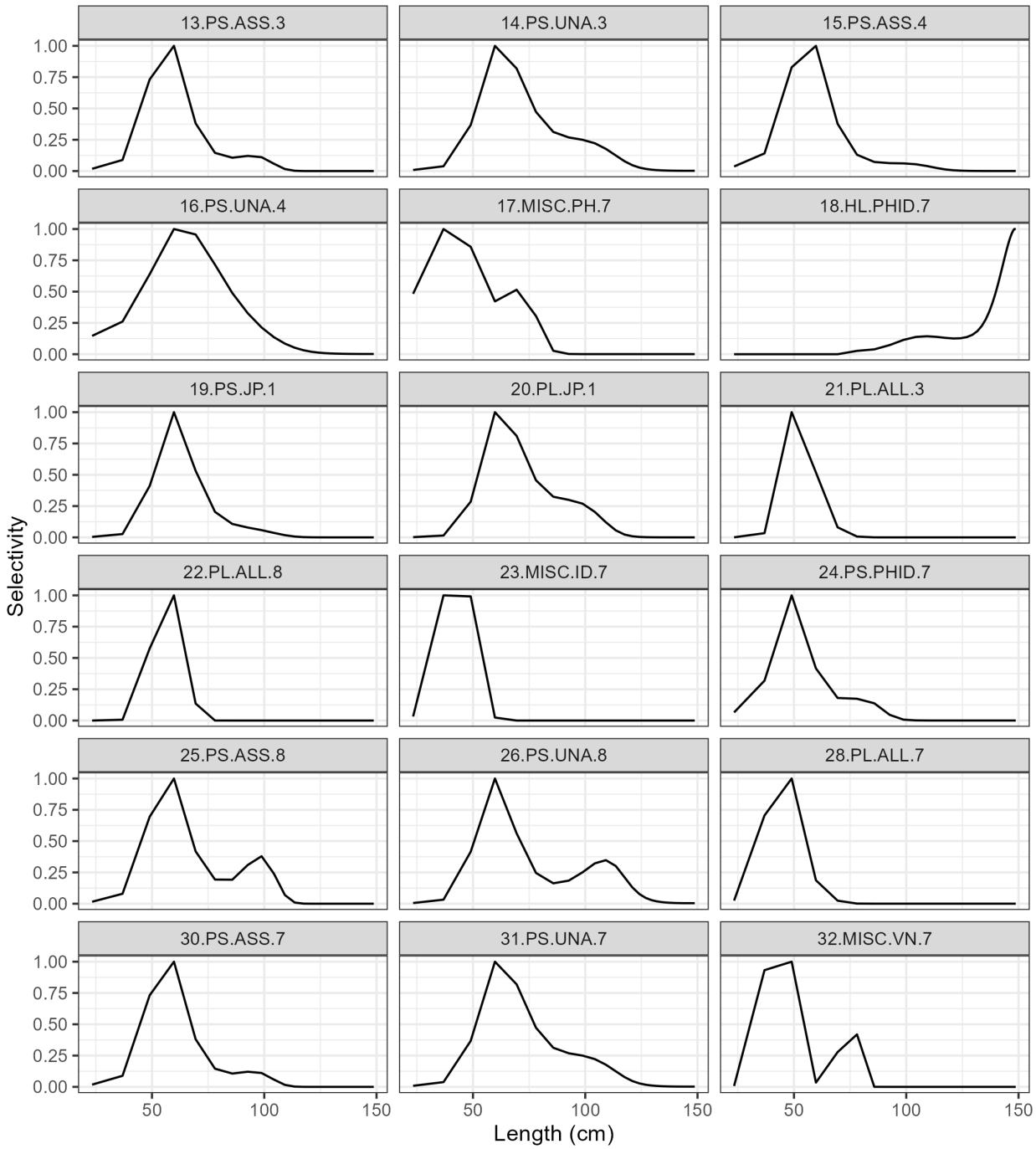

Figure 39: Length-specific (fork length, cm) curves for non-longline fisheries. Selectivity is grouped between fisheries 13 and 30, and also between fisheries 14 and 31.

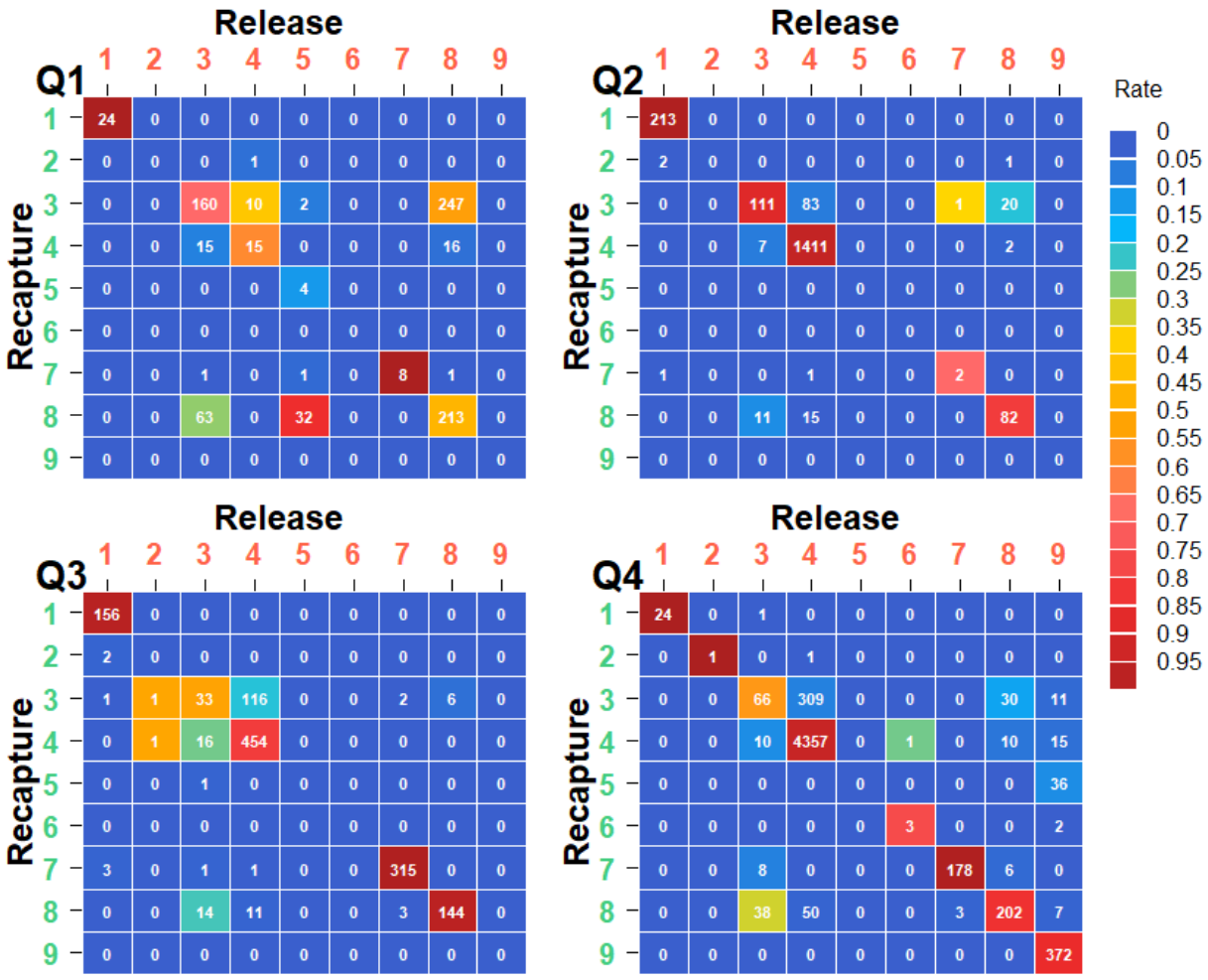

Figure 40: Observed proportion of tags returned by region of release (columns), region of recapture (rows), and quarter of recapture (panel). The colour of the tile indicates the proportion of tags returned from the region of release and the numbers in the boxes indicate the actual numbers of tags returned.

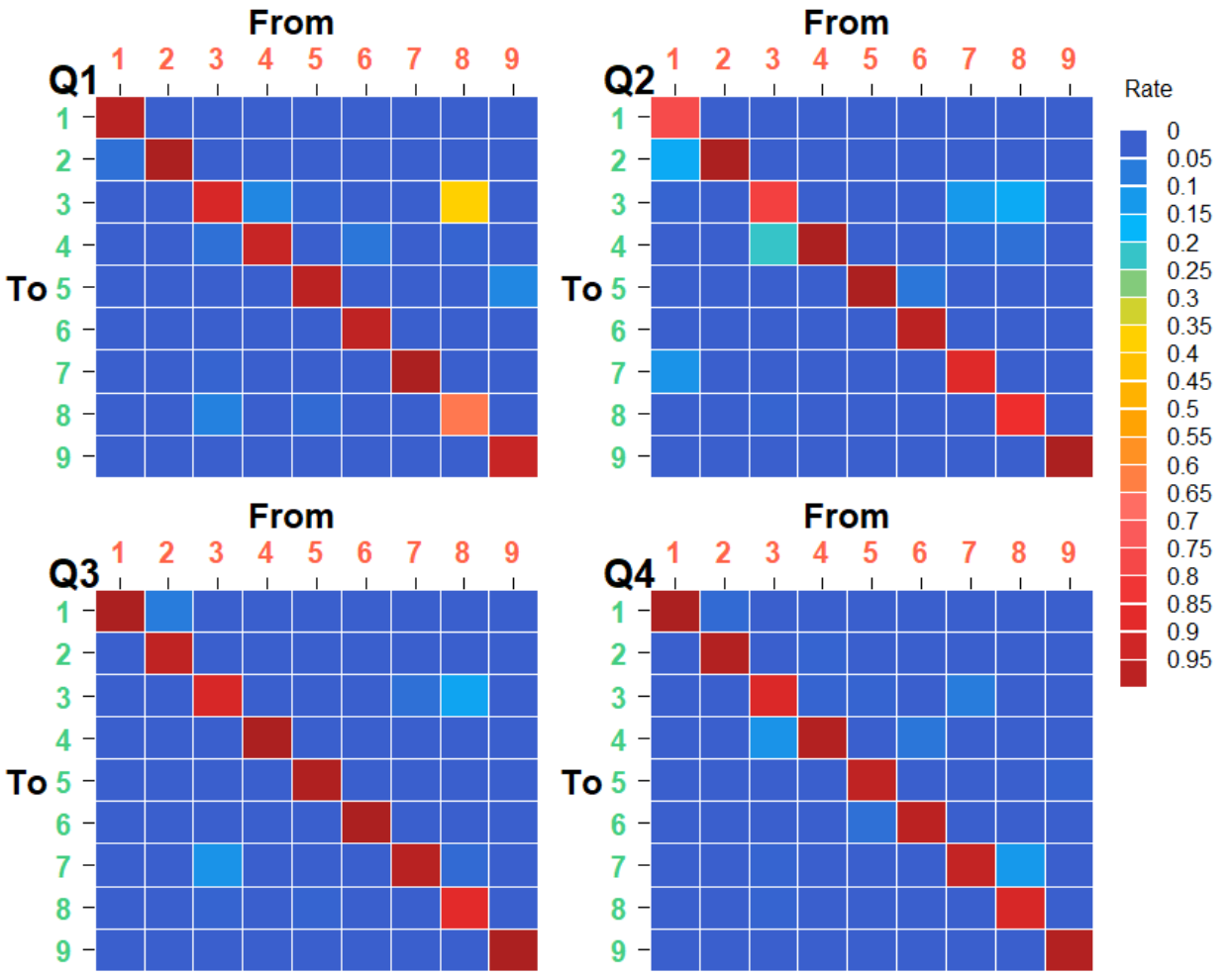

Figure 41: Estimated movement probabilities by quarter for the 2023 diagnostic model. The red numbers (horizontal axis) indicate the source model region (From); the green numbers (vertical axis) indicate the receiving (To) regions. The colour of the tile shows the magnitude of the movement rate (proportion of individuals moving from region x to region y in each quarter), with each column summing to 1.

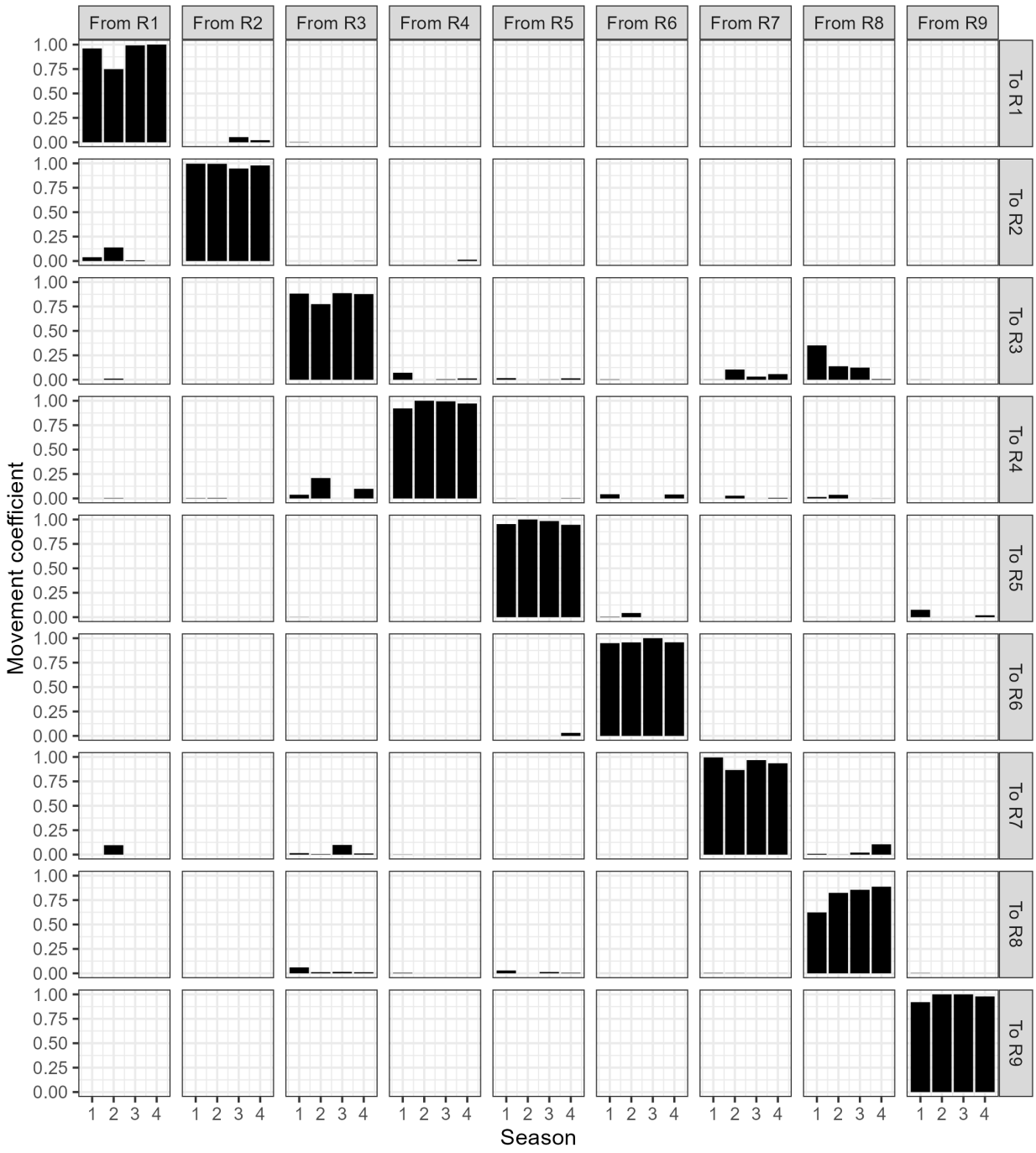

Figure 42: Season-specific movement probabilities estimated by the 2023 diagnostic model.

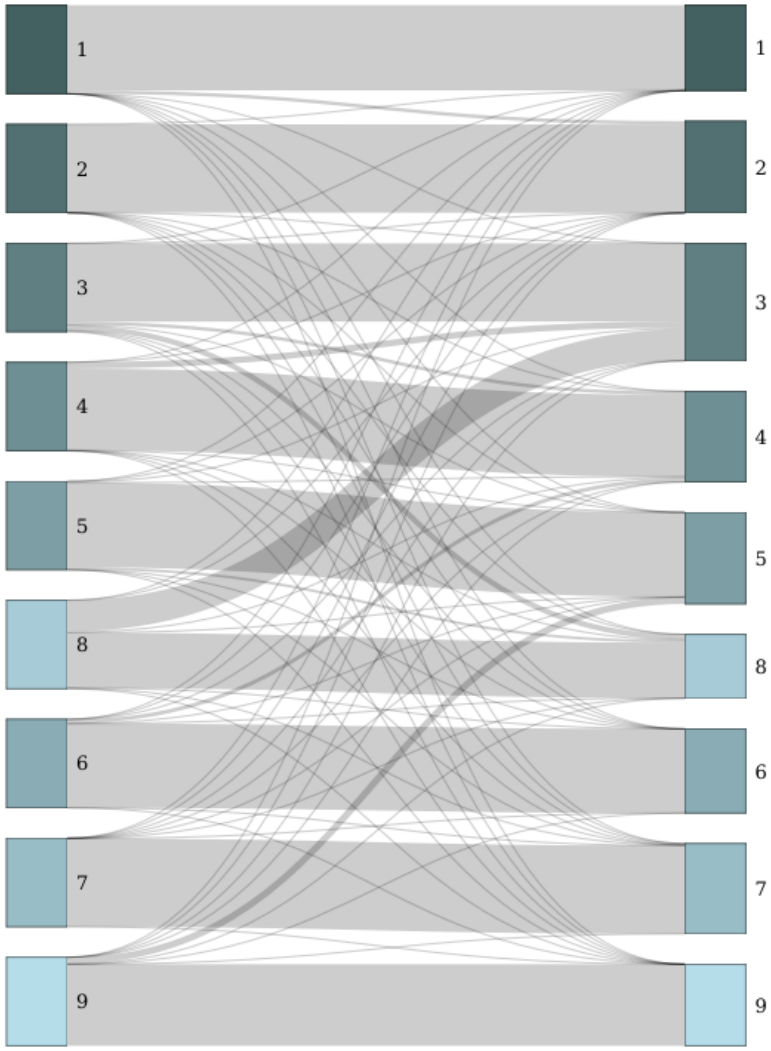

Figure 43: Stylised estimated movement rates between stock assessment regions (all ages and seasons) for the diagnostic case. Estimated movement is shown originating from the regions on the left and moving to the regions on the right.

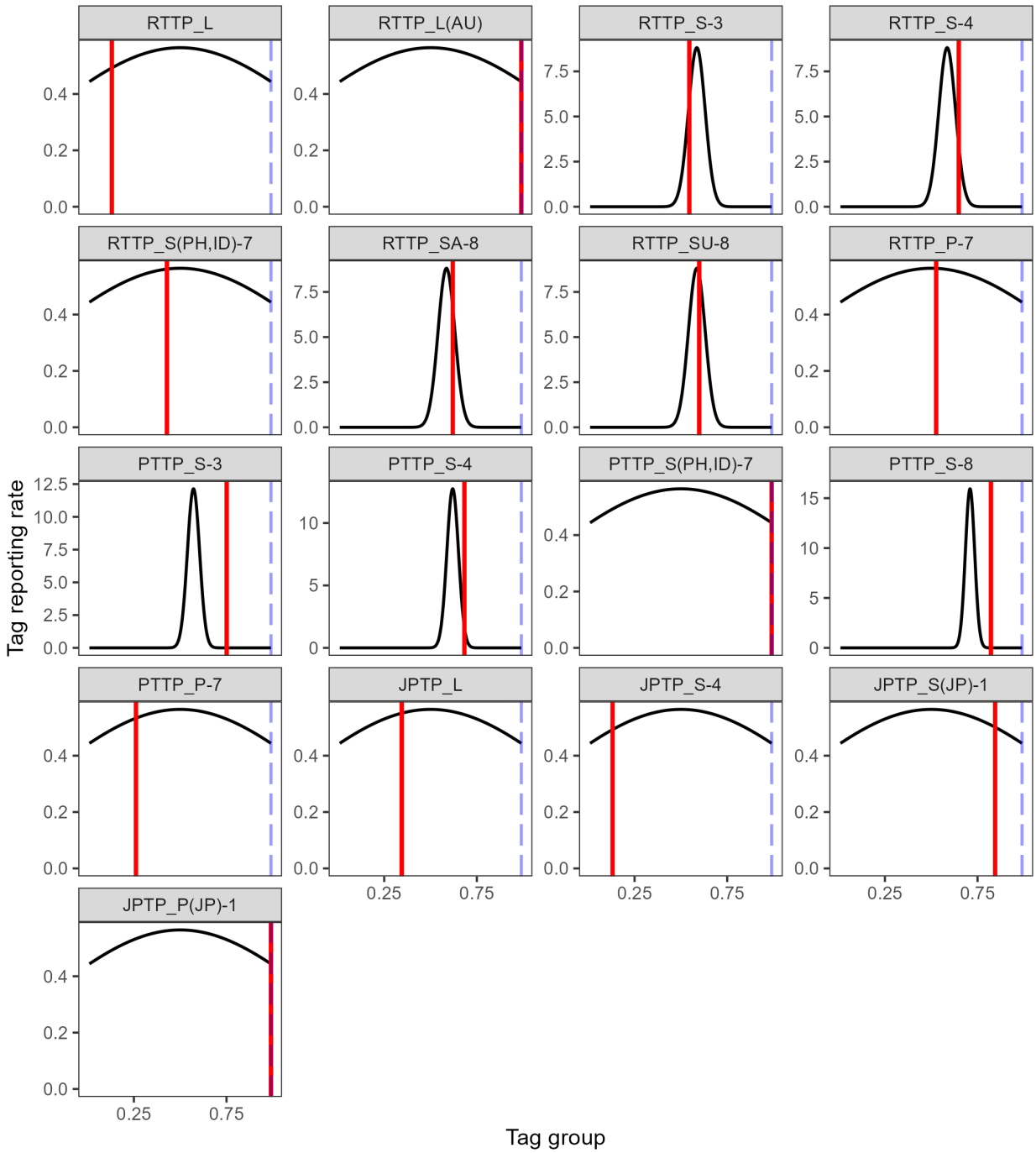

Figure 44: Estimated reporting rates for the diagnostic model (red lines) and the prior distribution for each tag reporting rate group (black lines). The upper bound (0.99) on the reporting rate parameters is shown as a blue dashed line. Reporting rates can be estimated separately for each release program and recapture fishery group but in practice are aggregated over some recapture groups to reduce dimensionality.

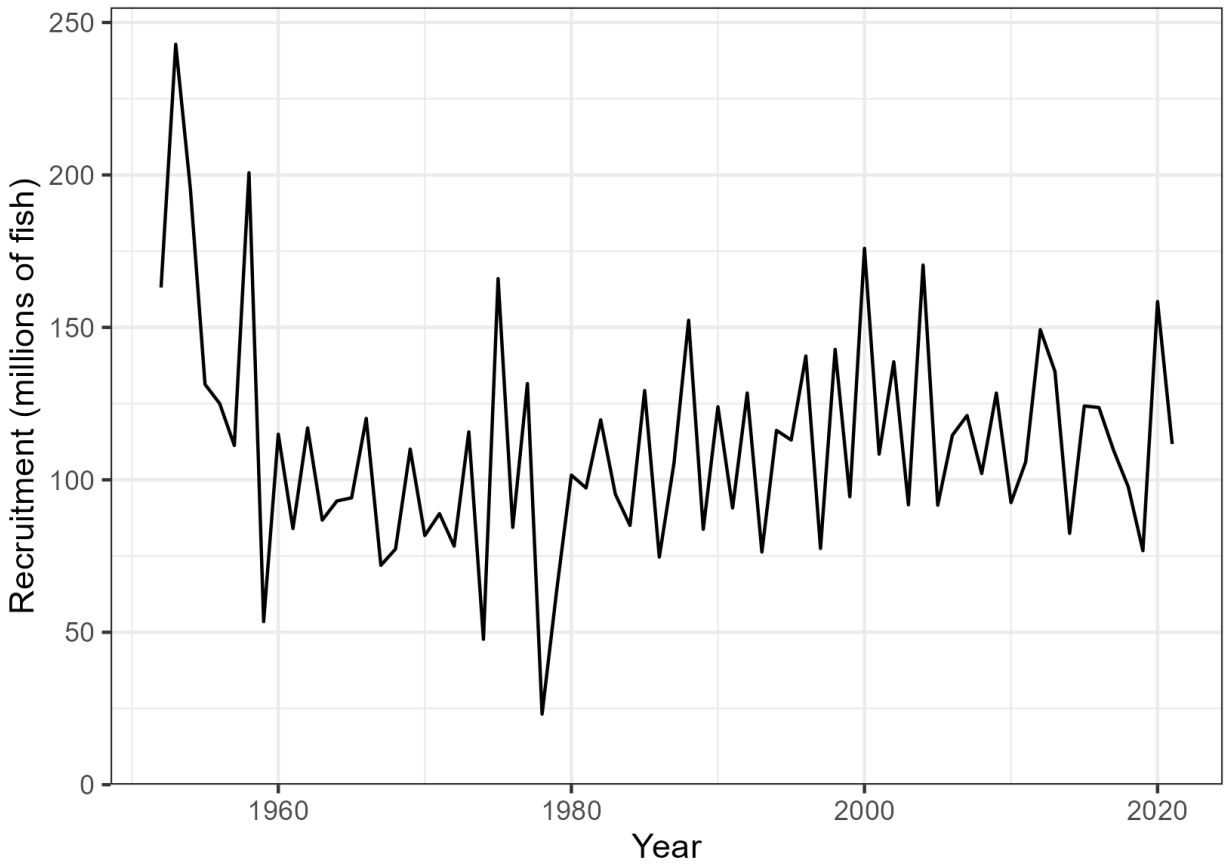

Figure 45: Time series of estimated annual recruitment summed across regions for the 2023 diagnostic model.

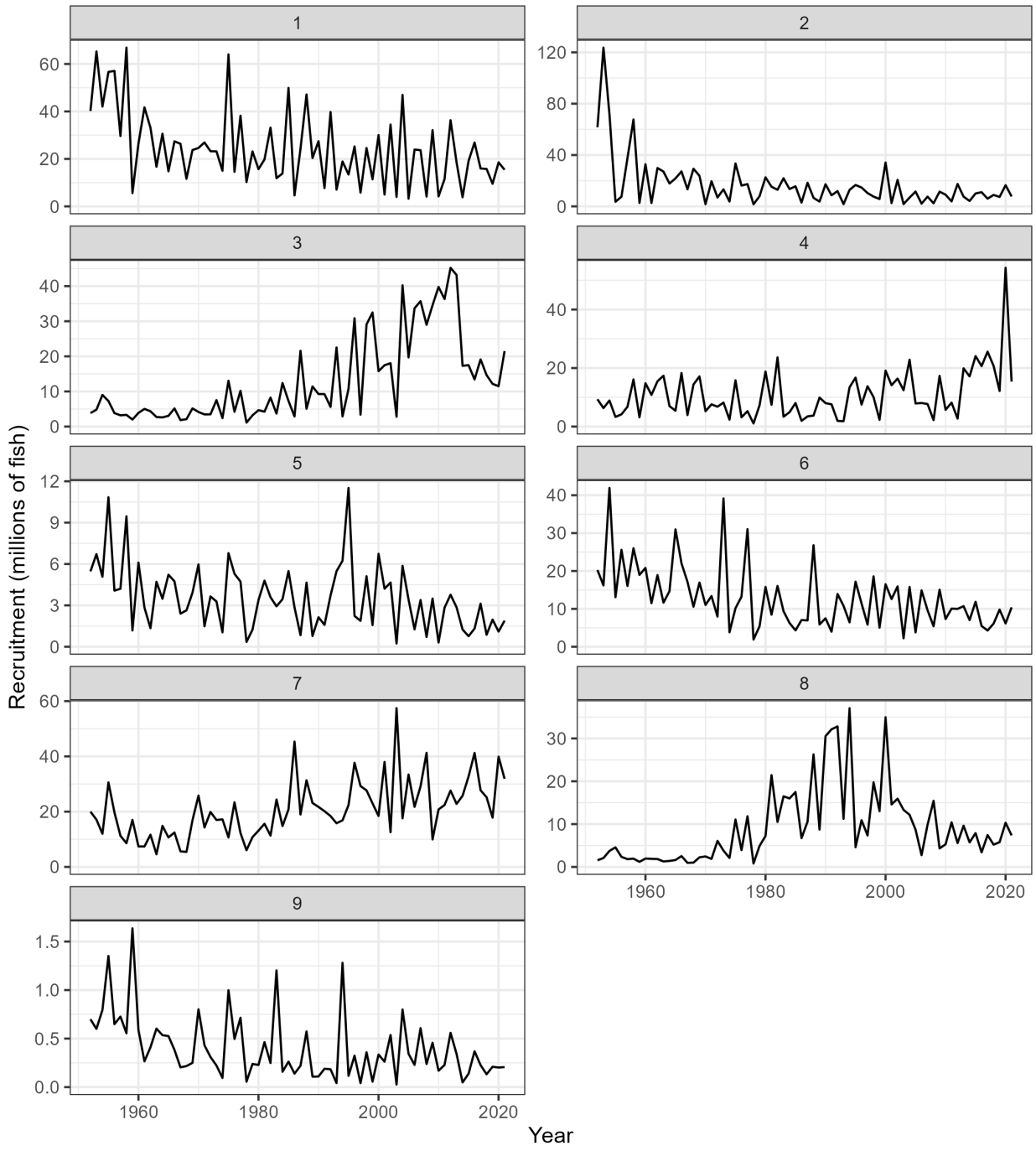

Figure 46: Time series of estimated annual recruitment by region for the 2023 diagnostic model. Note that the scale of the y-axis varies by region.

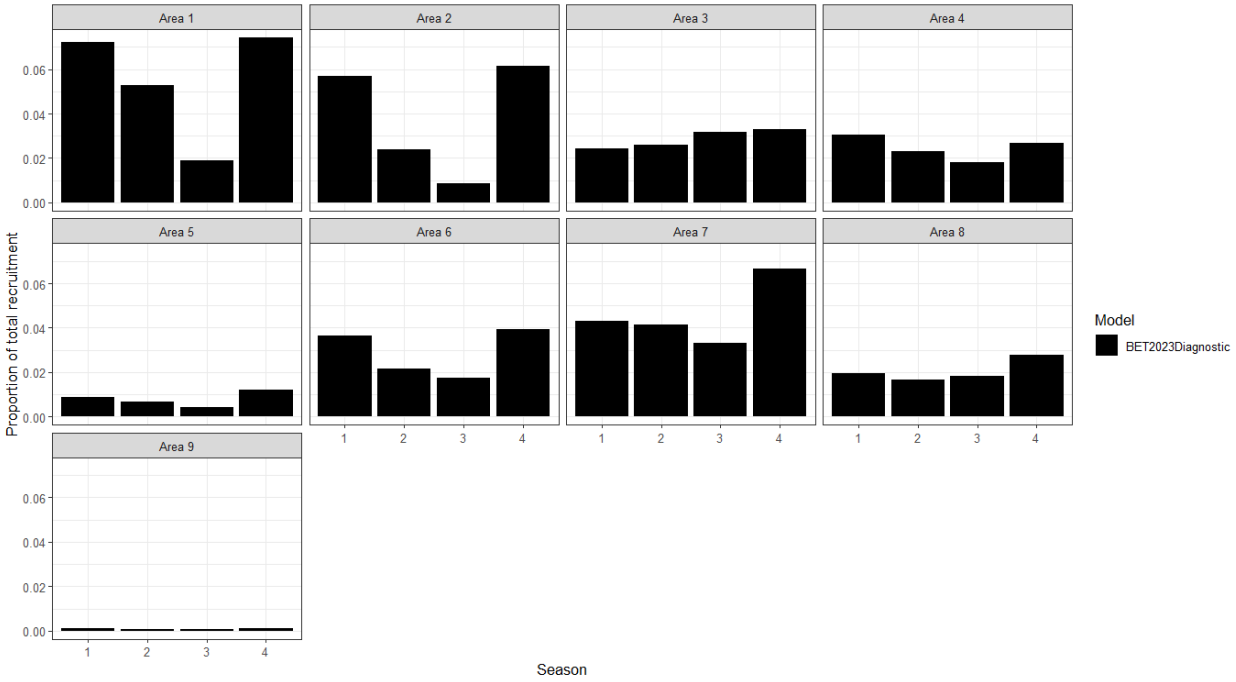

Figure 47: Estimated recruitment distribution by region and quarter.

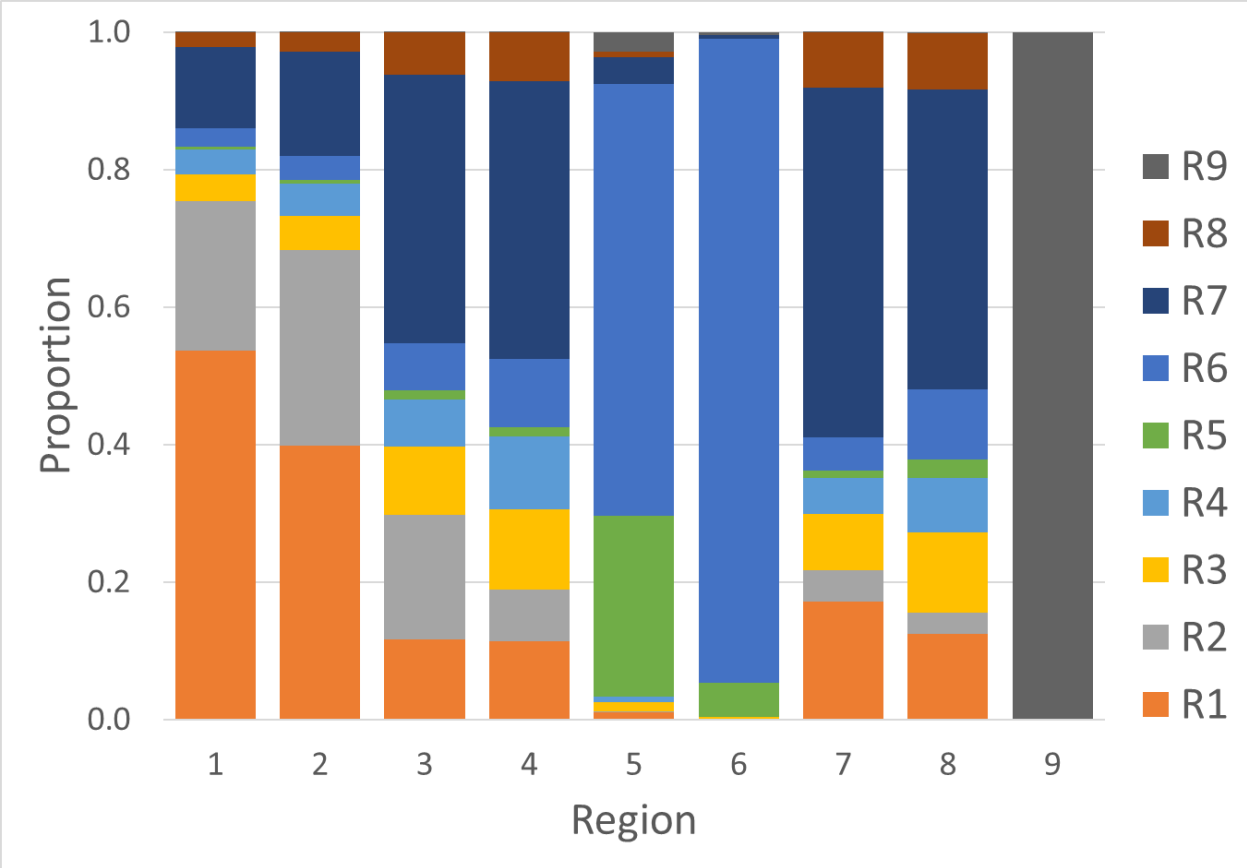

Figure 48: Proportional distribution of unfished total biomass (by weight) in each region apportioned by the region of recruitment, for the 2023 diagnostic model. The colour of the originating region is presented in the legend below. The biomass distributions are calculated based on the long term average distribution of recruitment between regions, estimated movement parameters, and natural mortality.

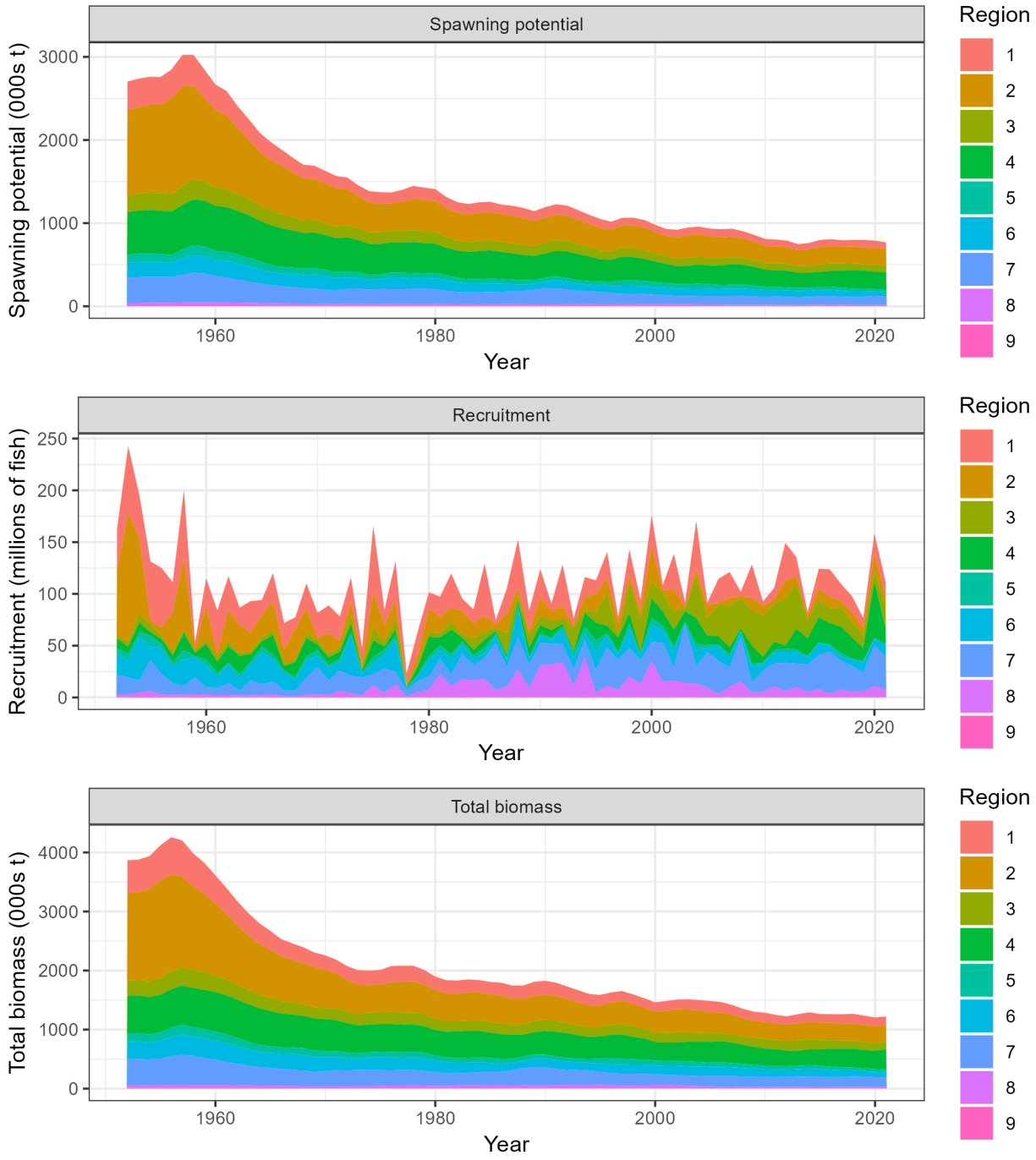

Figure 49: Time series of estimated annual spawning potential, recruitment and total biomass by model region for the diagnostic model, showing the relative proportions among regions. Note the data represent the averages of the quarterly model time steps for each year for spawning potential and total biomass and the sum of the quarterly recruitment estimates for annual recruitment.

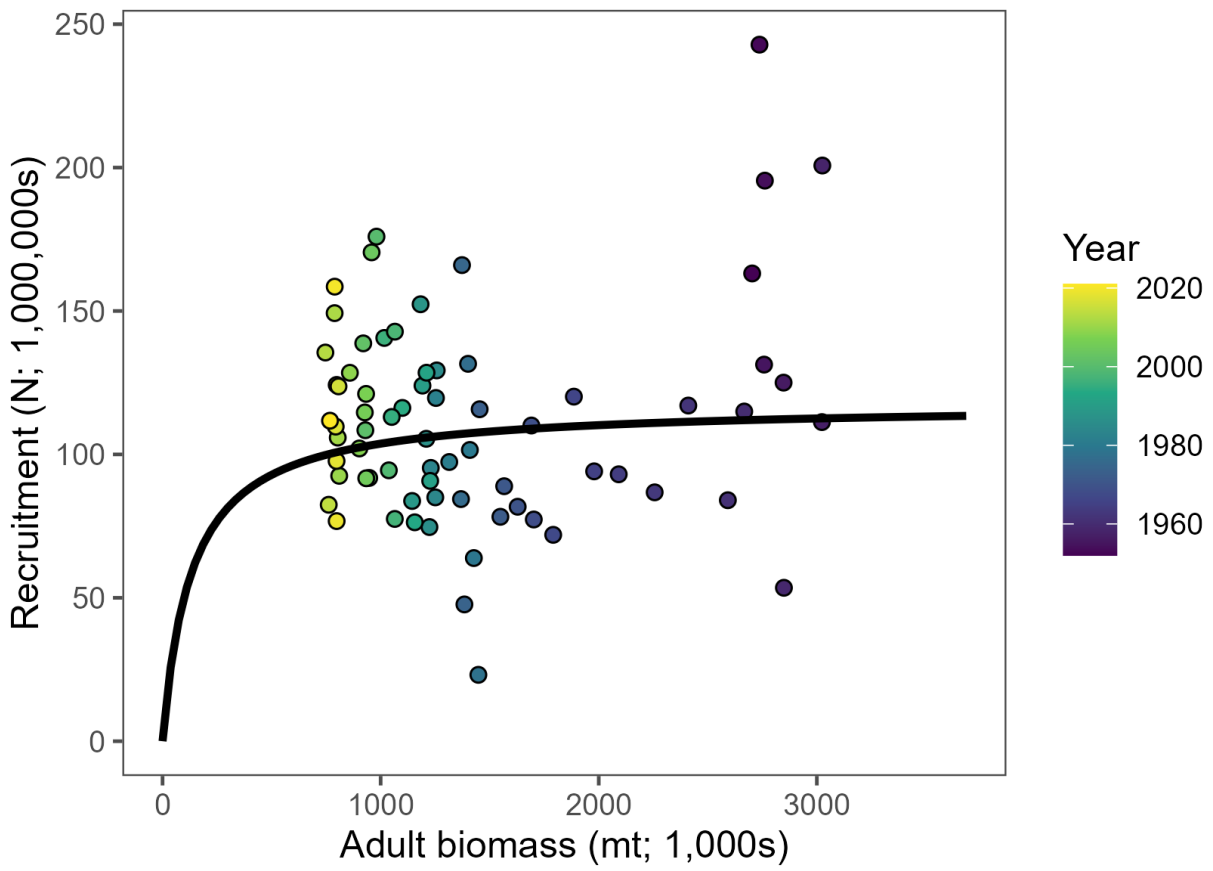

Figure 50: Estimated stock-recruitment relationship between recruitment and spawning potential based on annual estimates of recruitment for the 2023 diagnostic model. The colour of the circles transitions from purple (early in the time series) to yellow (more recent) through time. Points shown prior to 1968 are not used in the fit to the stock recruitment relationship.

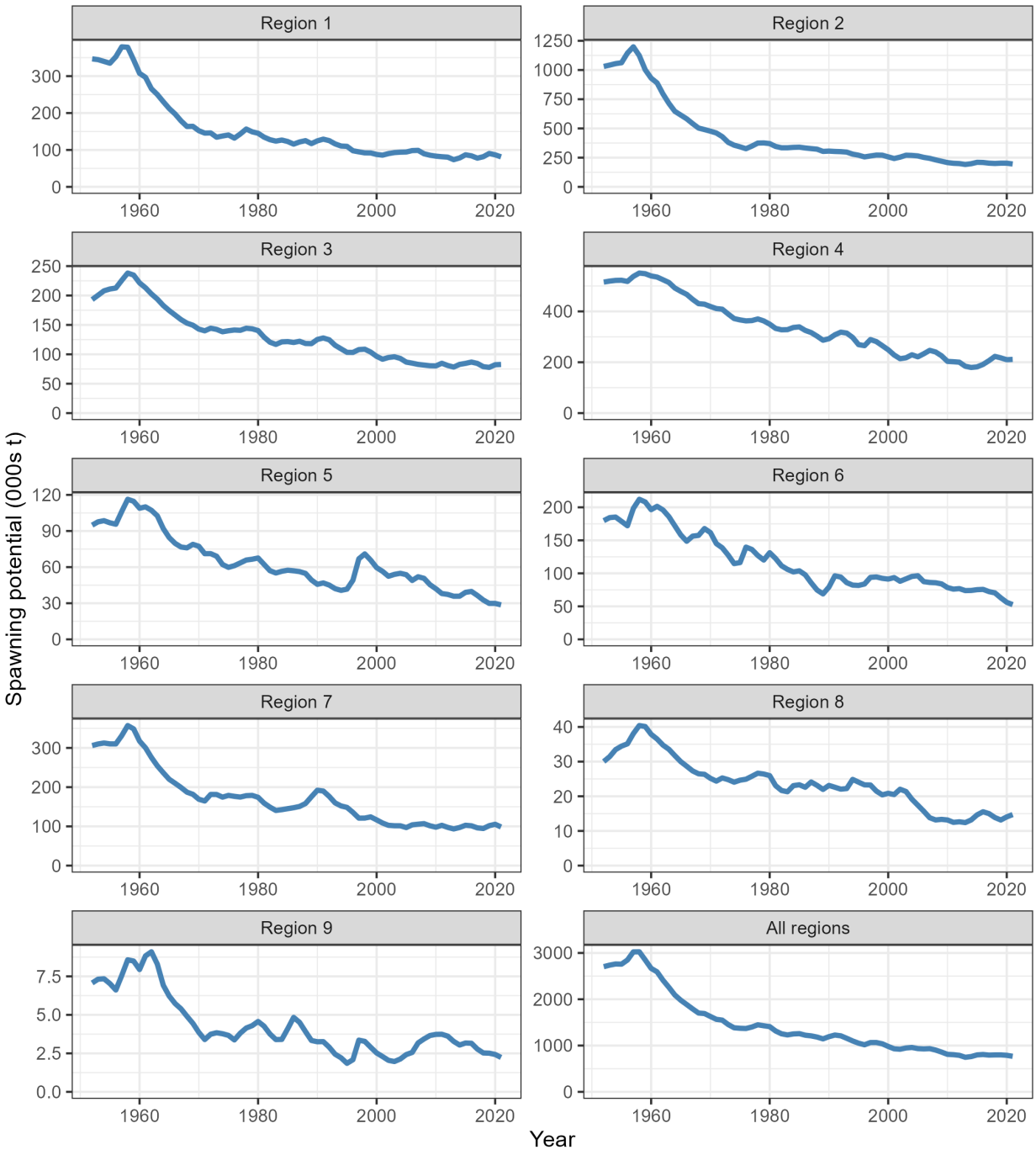

Figure 51: Estimated temporal spawning potential, SB_t, by region, and combined for all regions (bottom right) for the 2023 diagnostic model. Note that the scale of the y-axis varies for each region.

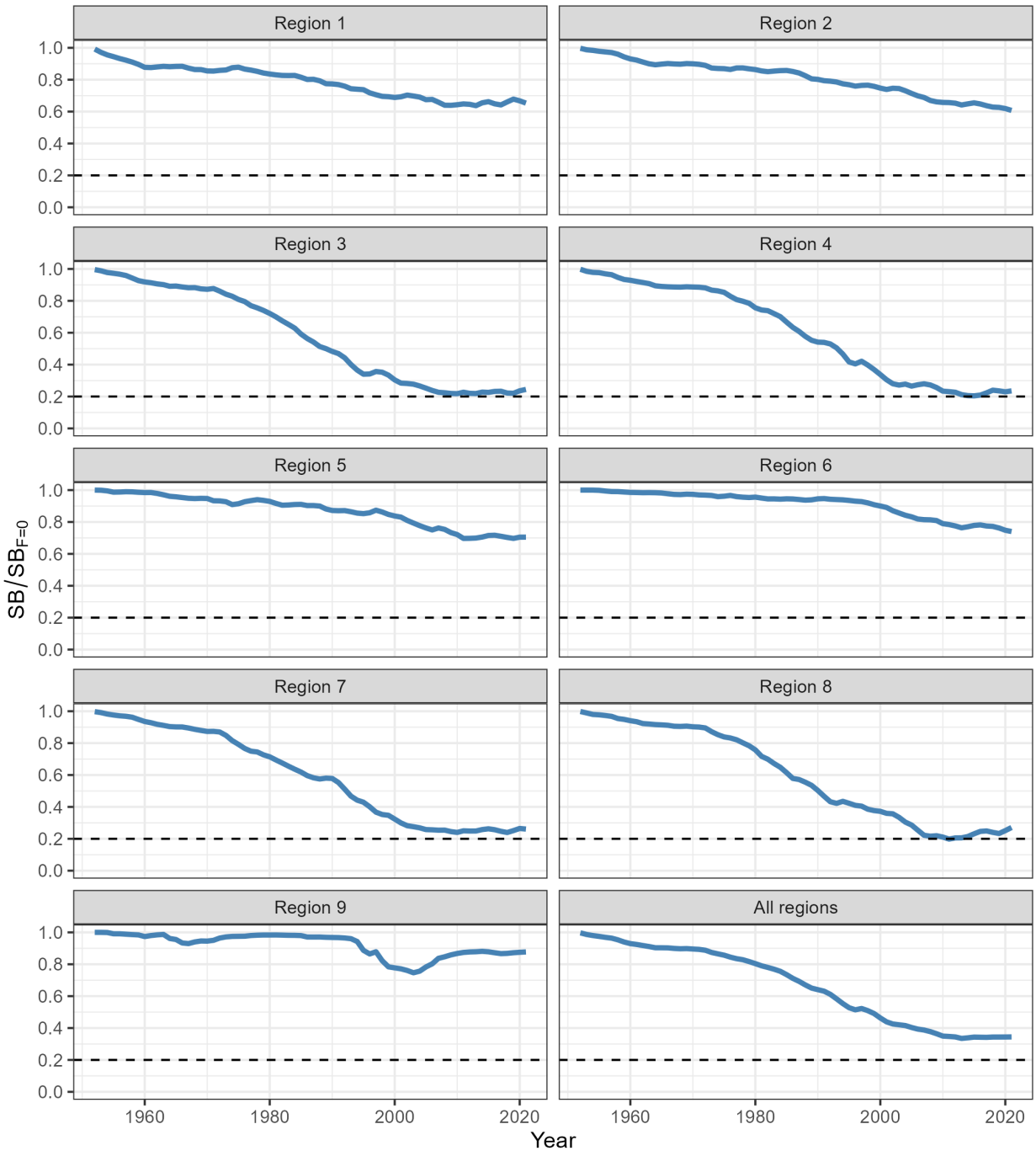

Figure 52: Estimated temporal spawning potential depletion, $SB_t/SB_{t,F=0}$, by region, and combined for all regions (bottom right) for the 2023 diagnostic model.

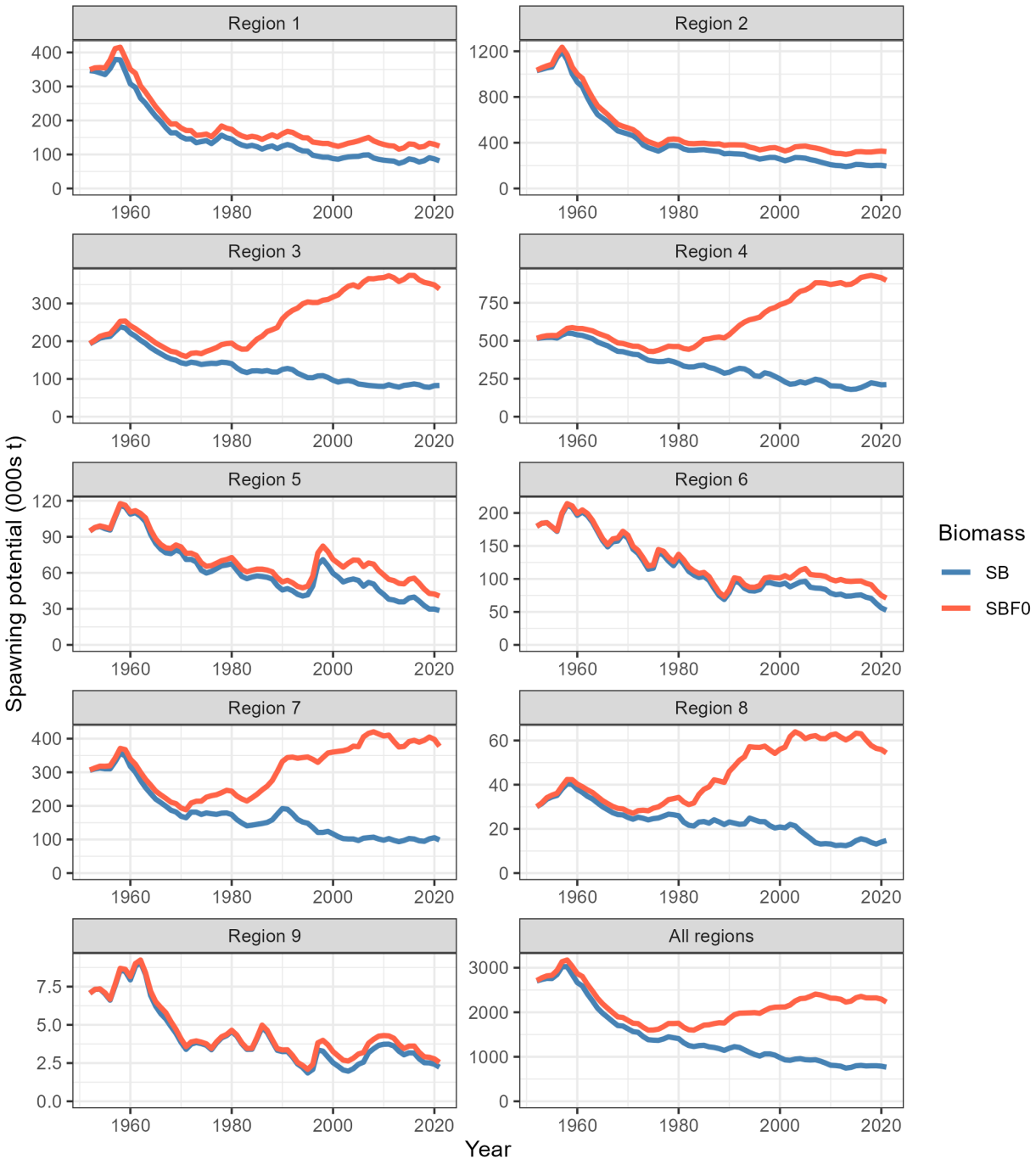

Figure 53: Comparison of the estimated annual spawning potential trajectories (lower, blue lines) with the spawning potential trajectories predicted in the absence of fishing (upper, red lines) for each region and overall, for the 2023 diagnostic model. Note the scales of the y-axis varies for each region.

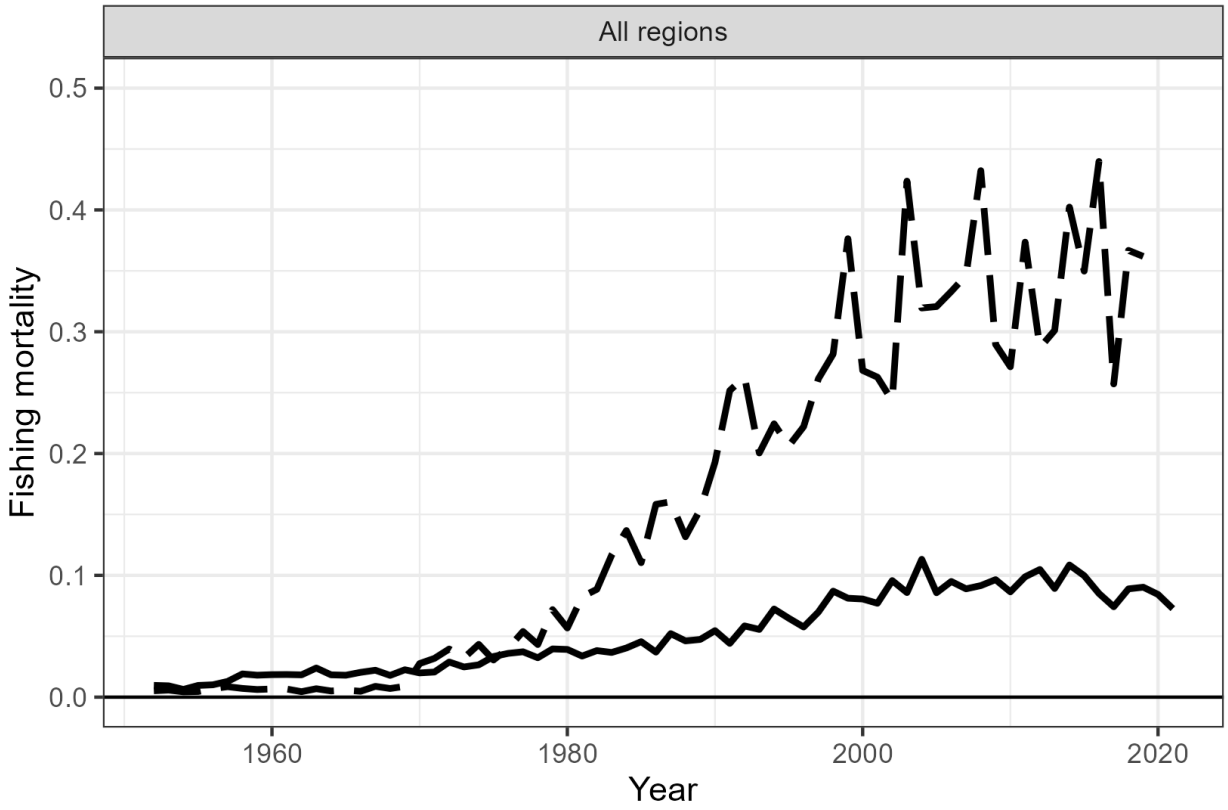

Figure 54: Estimated annual average adult (solid line) and juvenile (dashed line) fishing mortality for the 2023 diagnostic model.

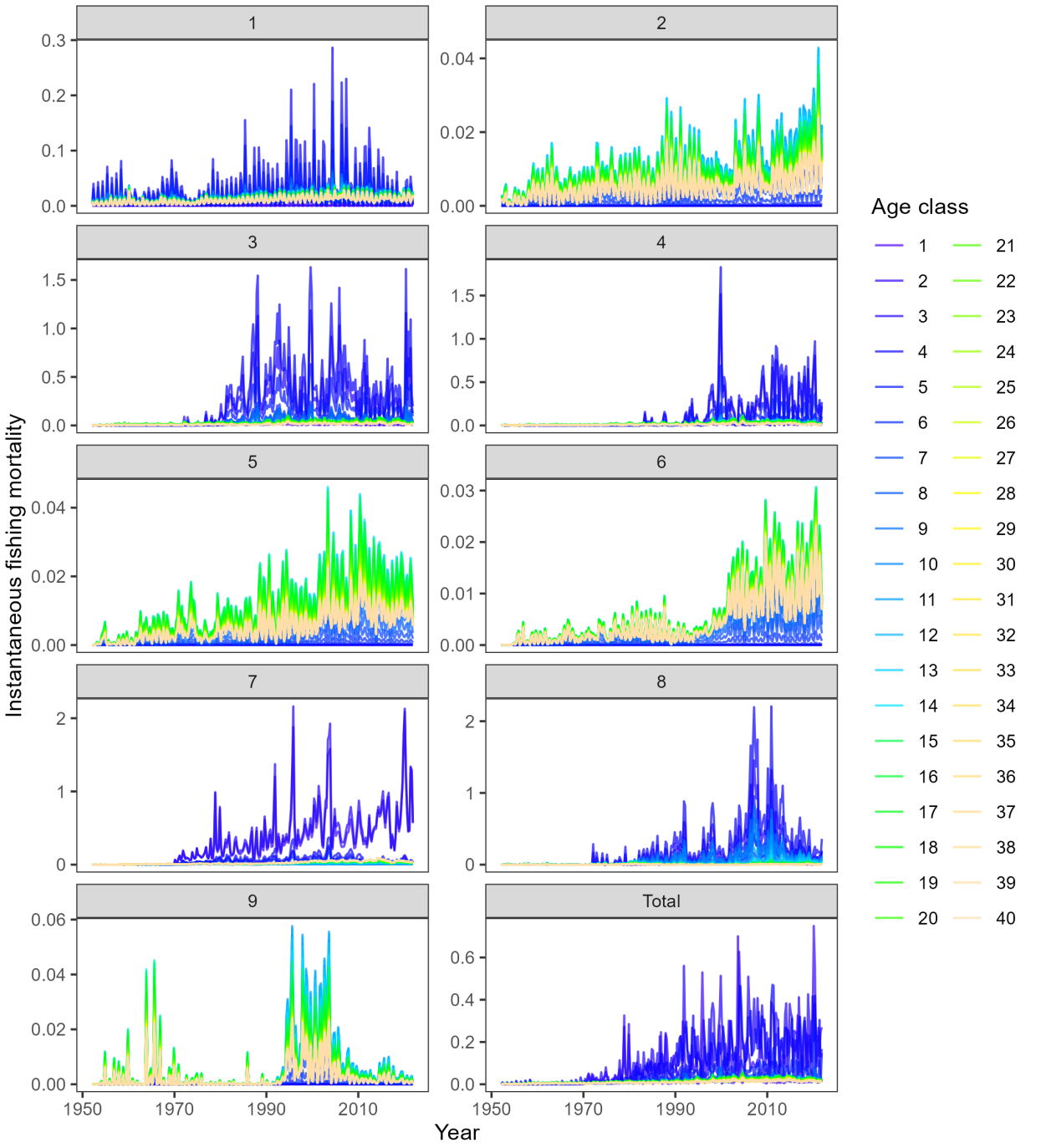

Figure 55: Estimated age-specific fishing mortality for the 2023 diagnostic model, by region and overall.

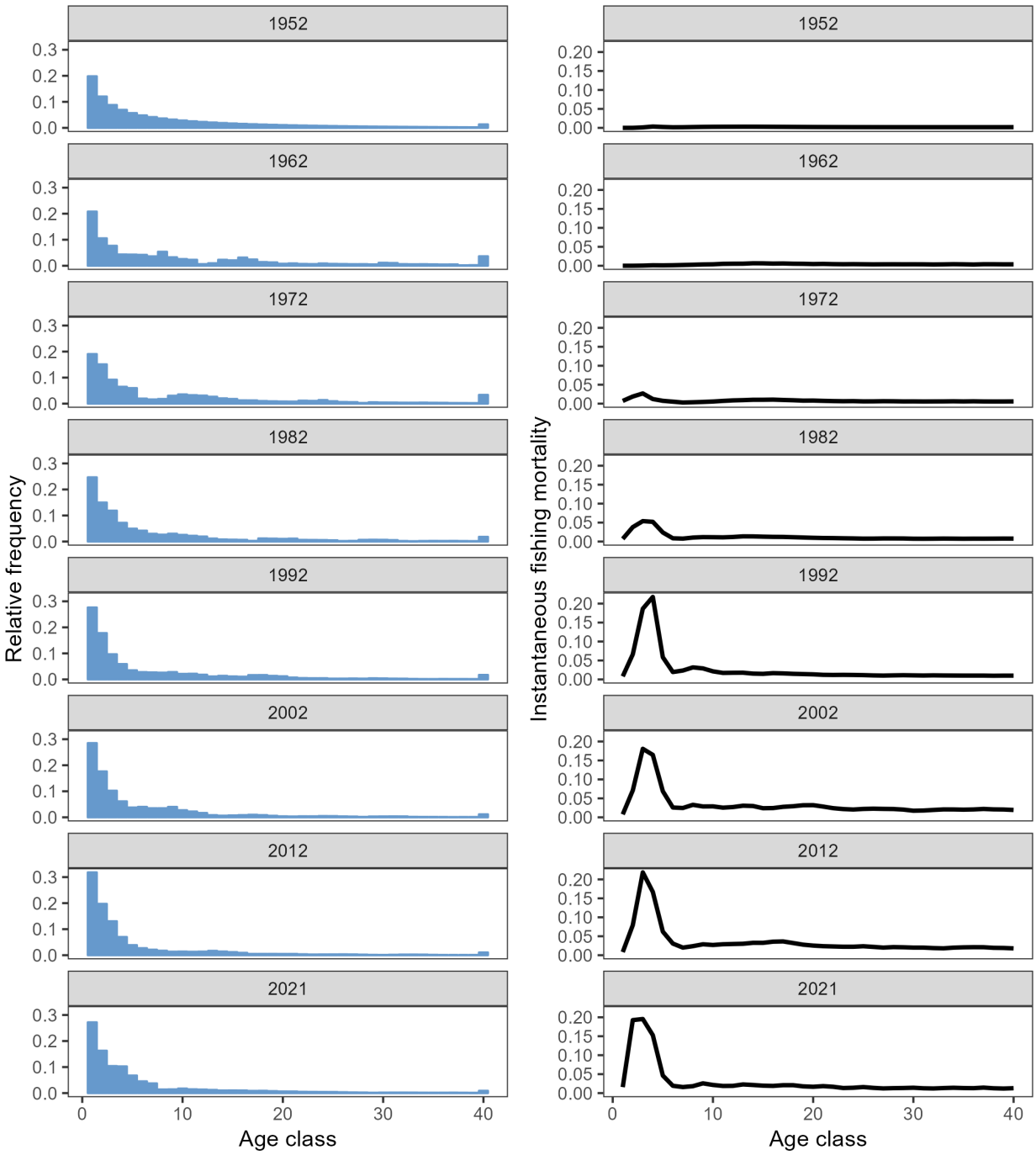

Figure 56: Estimated proportion at age (quarters) and fishing mortality at age (right), by year, for decadal intervals, for the 2023 diagnostic model.

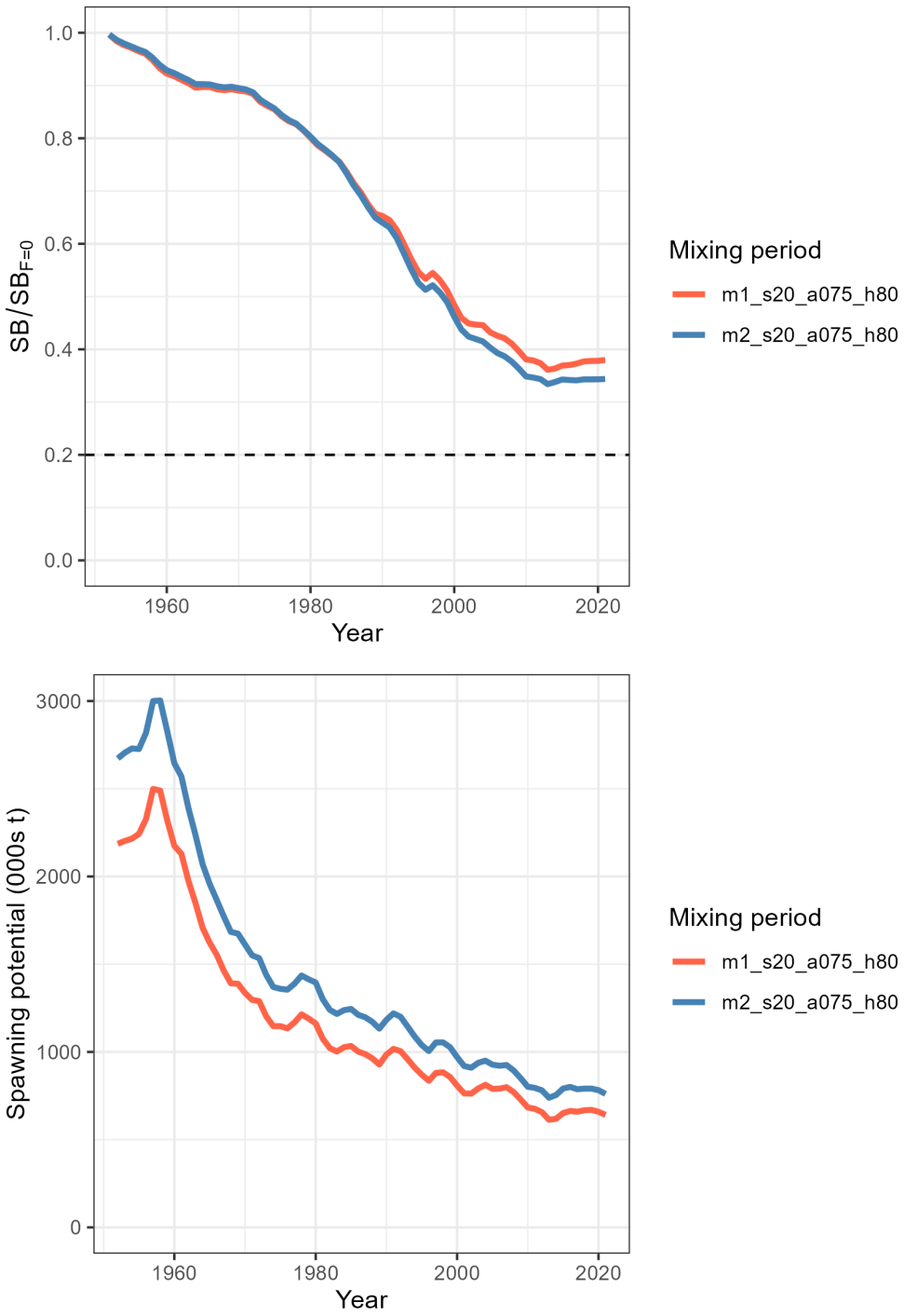

Figure 57: Estimated dynamic spawning depletion (Top) and spawning potential (Bottom) for the one-off sensitivities from the 2023 diagnostic model for tag mixing period. m = tag mixing, h = steepness, s = size data, a = age data.

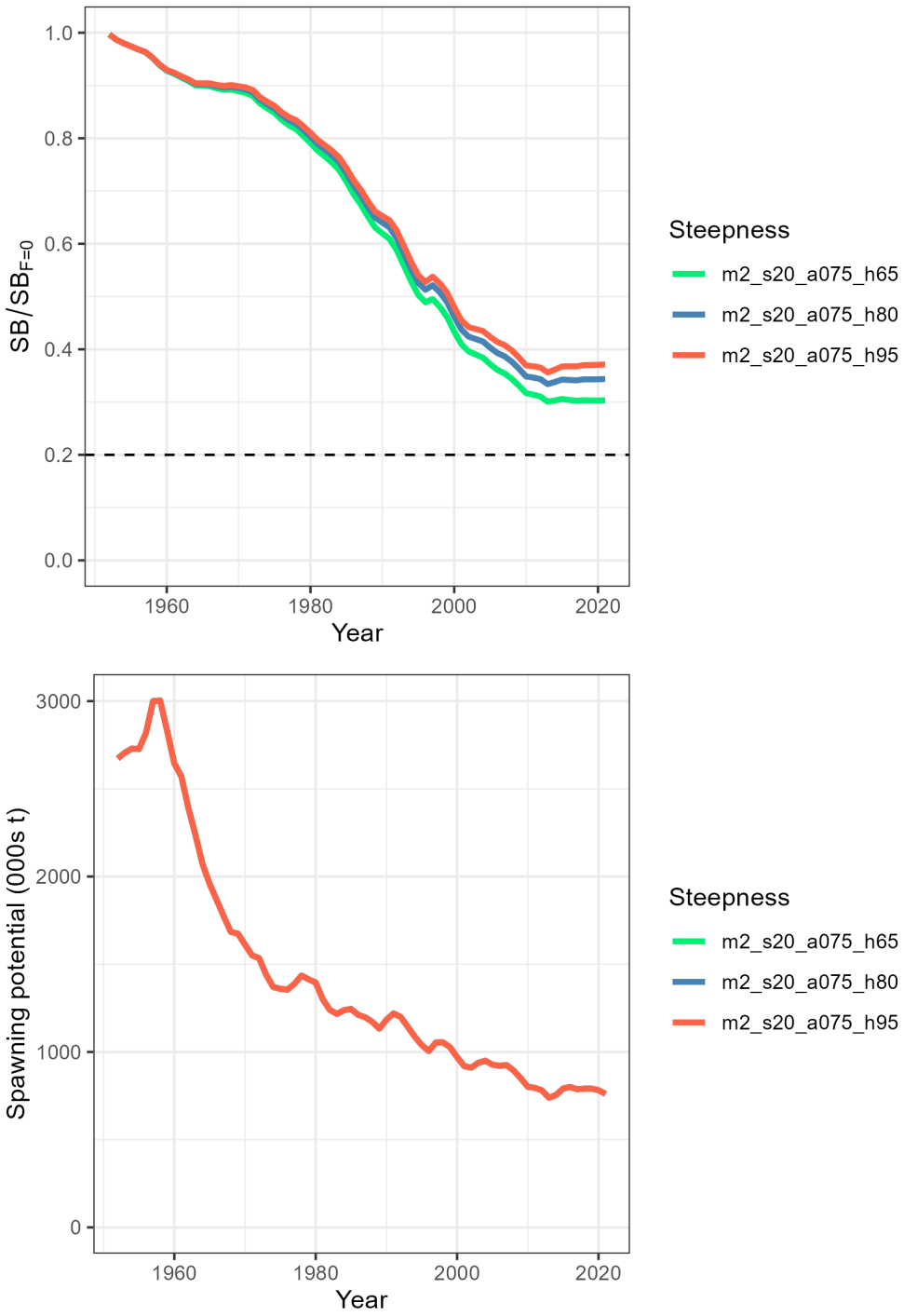

Figure 58: Estimated dynamic spawning depletion (Top) and spawning potential (Bottom) for the one-off sensitivities from the 2023 diagnostic model for steepness of the stock recruitment relationship. m = tag mixing, h = steepness, s = size data, a = age data.

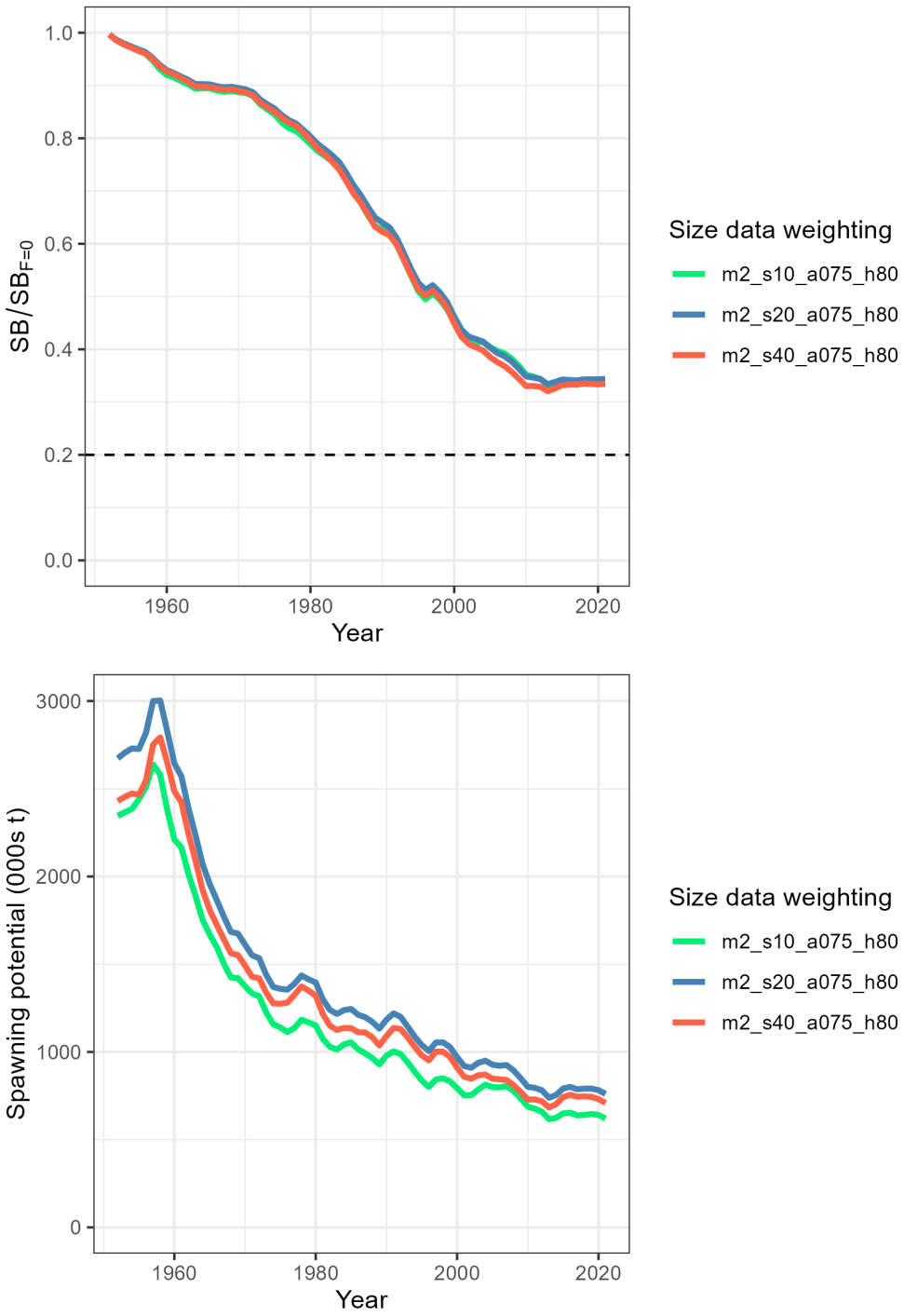

Figure 59: Estimated dynamic spawning depletion (Top) and spawning potential (Bottom) for the one-off sensitivities from the 2023 diagnostic model for size data weighting. m = tag mixing, h = steepness, s = size data, a = age data.

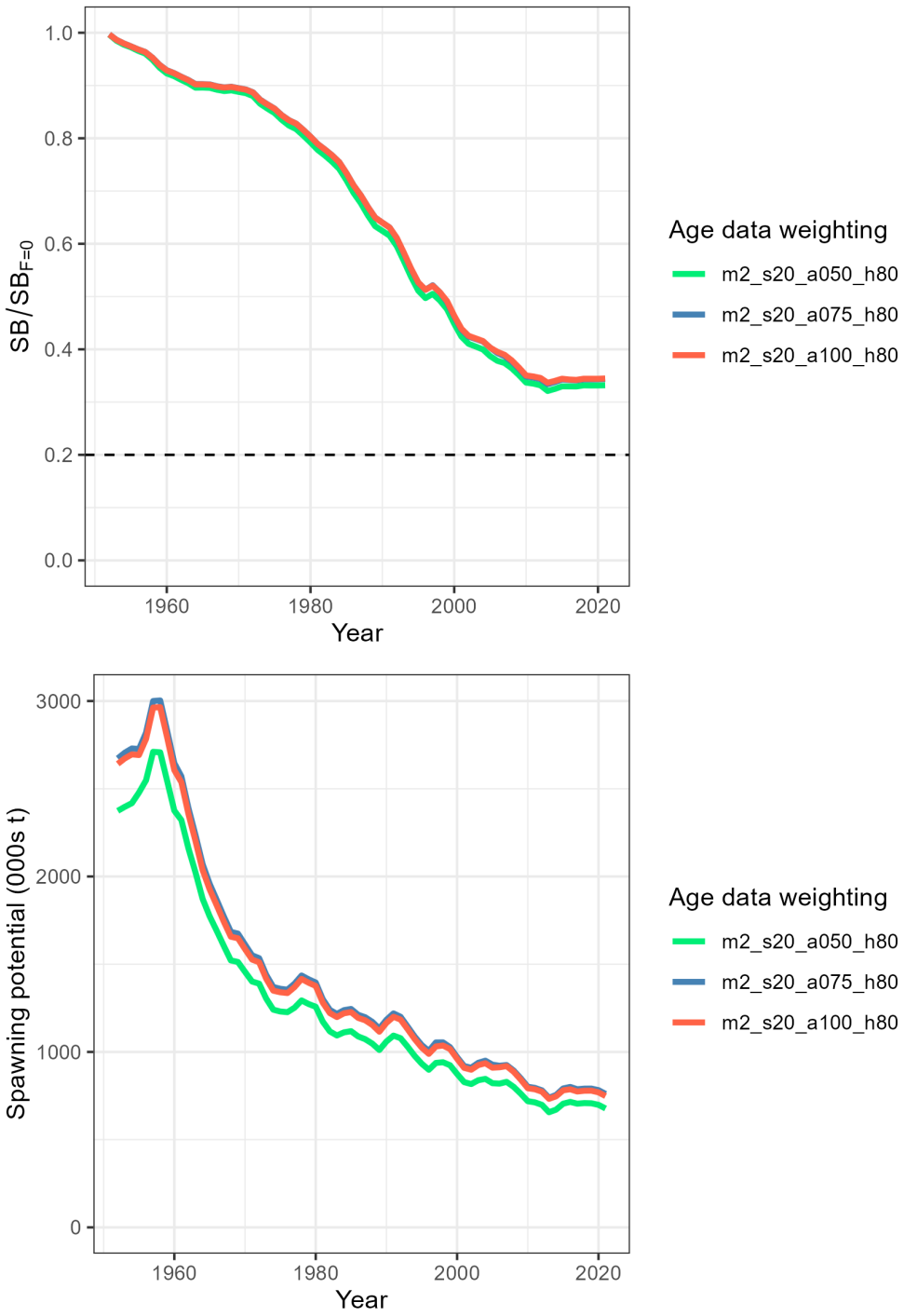

Figure 60: Estimated dynamic spawning depletion (Top) and spawning potential (Bottom) for the one-off sensitivities from the 2023 diagnostic model for age data weighting. m = tag mixing, h = steepness, s = size data, a = age data.

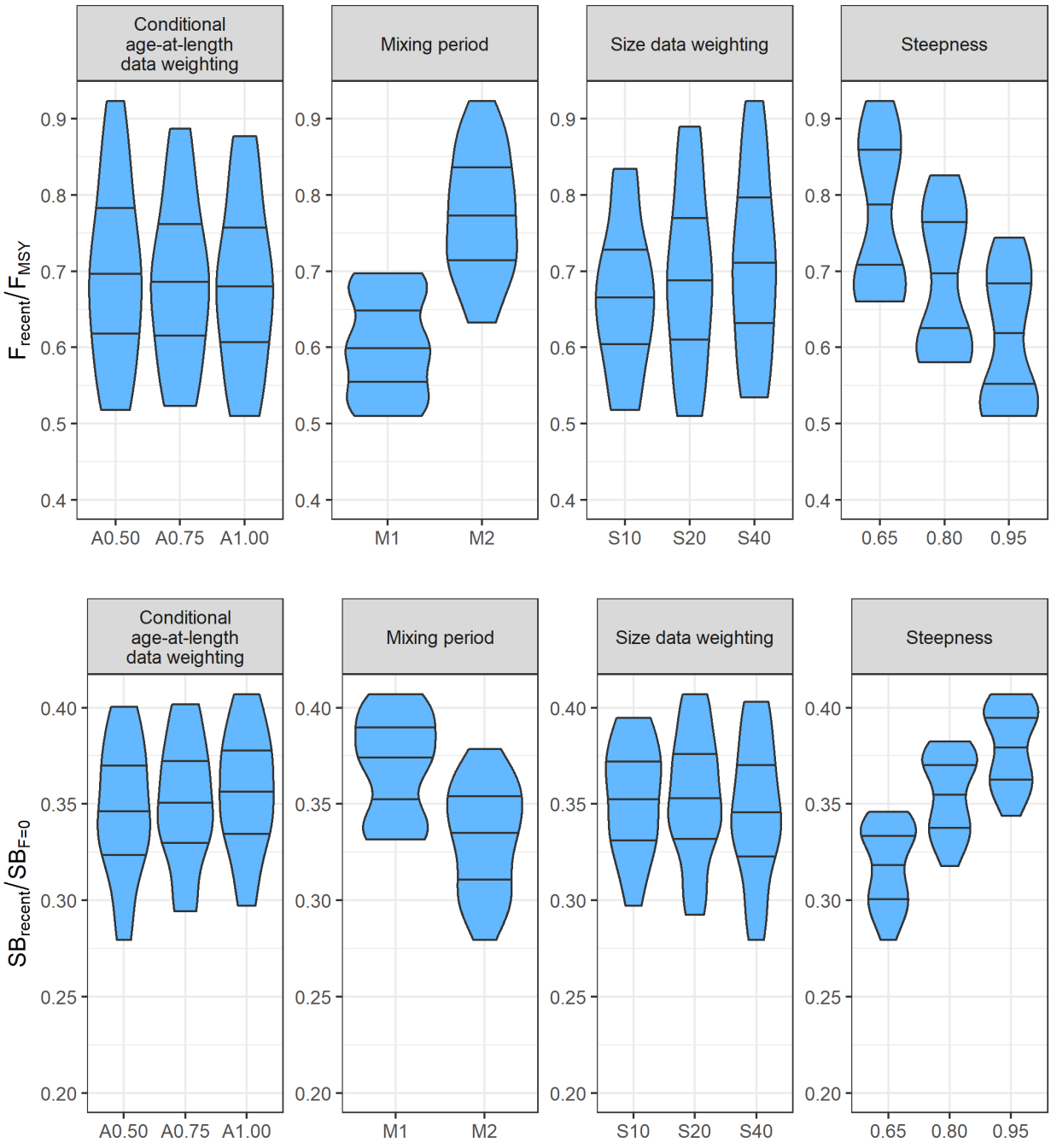

Figure 61: Box and violin plots summarizing the estimated $F_{\text{recent}}/F_{\text{MSY}}$ (top) and $SB_{\text{recent}}/SB_{F=0}$ (bottom) for each of the models in the structural uncertainty grid, excluding estimation uncertainty, grouped by uncertainty axes (steepness, tag mixing, size data weighting, 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.

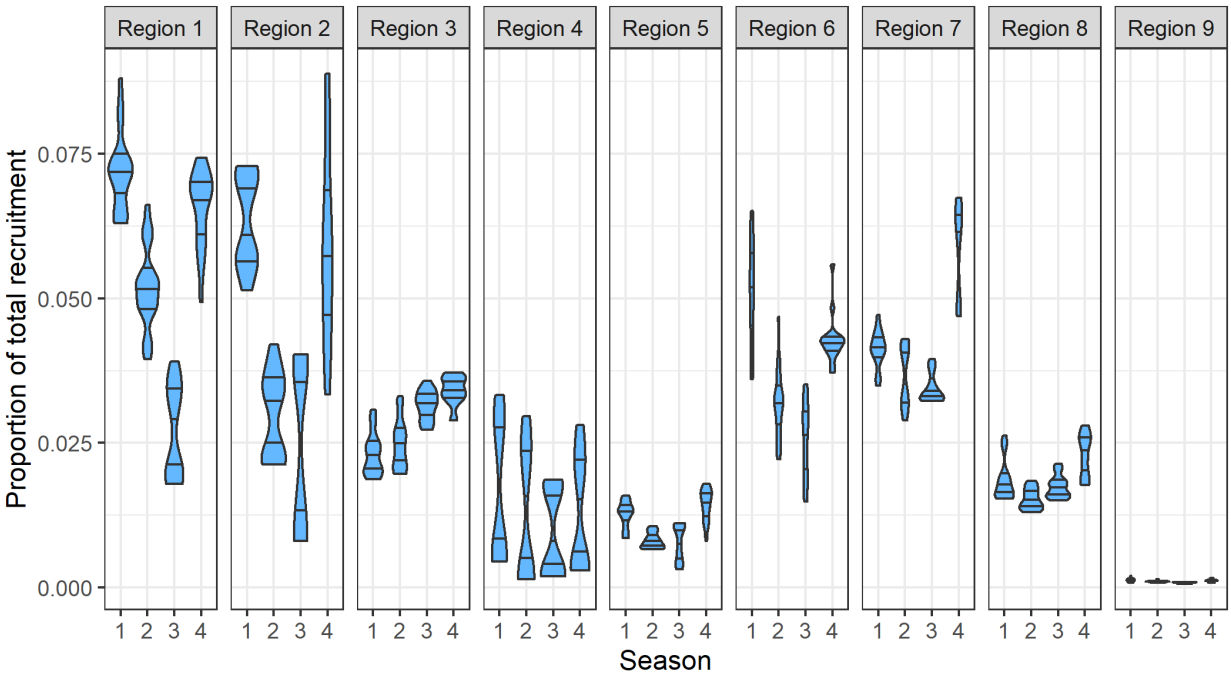

Figure 62: Box and violin plots showing the distribution or recruitment across model regions and quarters for all models in the uncertainty grid, excluding estimation uncertainty. The horizontal lines are the 25th, 50th, and 75th percentiles. The shaded area shows the probability distribution (or density) of the estimates of all models of the structural uncertainty grid.

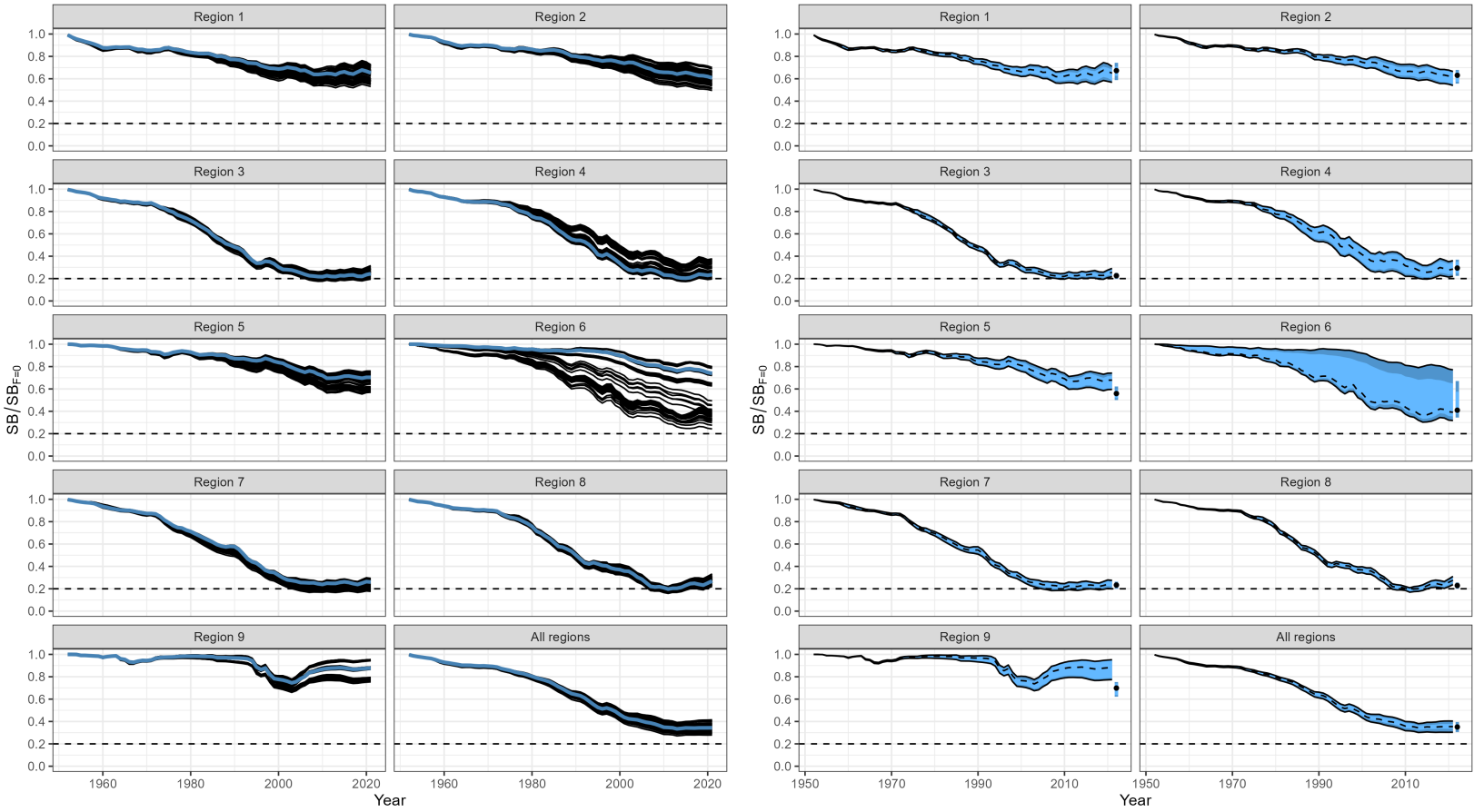

Figure 63: (Left) Trajectories of spawning potential depletion for the individual model runs included in the structural uncertainty grid over the period 1952-2021. (Right) Estimated spawning depletion across all models in the structural uncertainty grid over the period 1952-2021. The dashed line represents 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 $SB_{\text{recent}}/SB_{F=0}$.

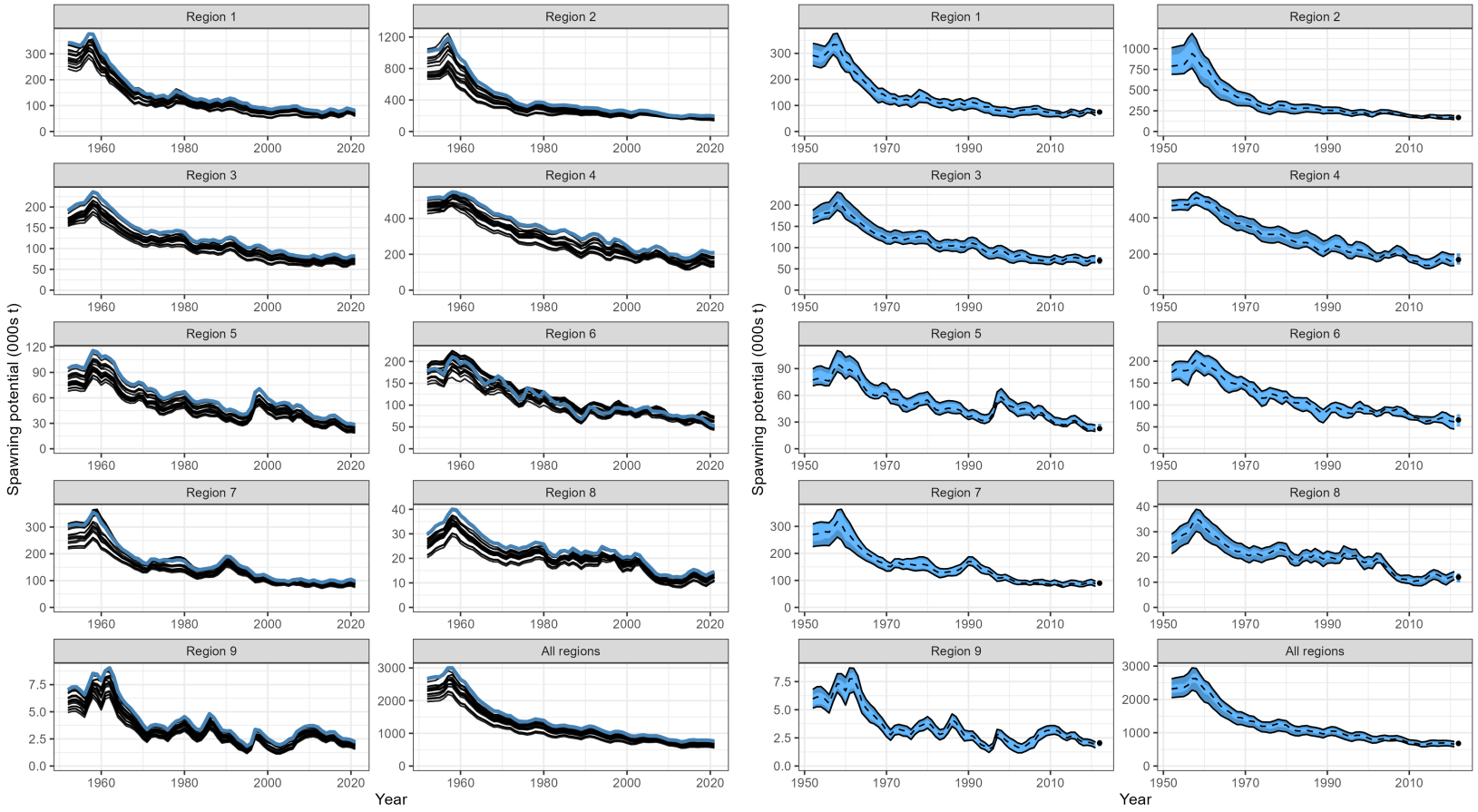

Figure 64: (Left) Trajectories of spawning potential for the individual model runs included in the structural uncertainty grid over the period 1952-2021. (Right) Estimated spawning potential across all models in the structural uncertainty grid over the period 1952-2021. The dashed line represents the median. The lighter band shows the 50th percentile, and the dark band shows the 80th percentile of the model estimates. The bar at the right of each ribbon indicate the median (black dots) and 80th percentile range for SB_{recent}.

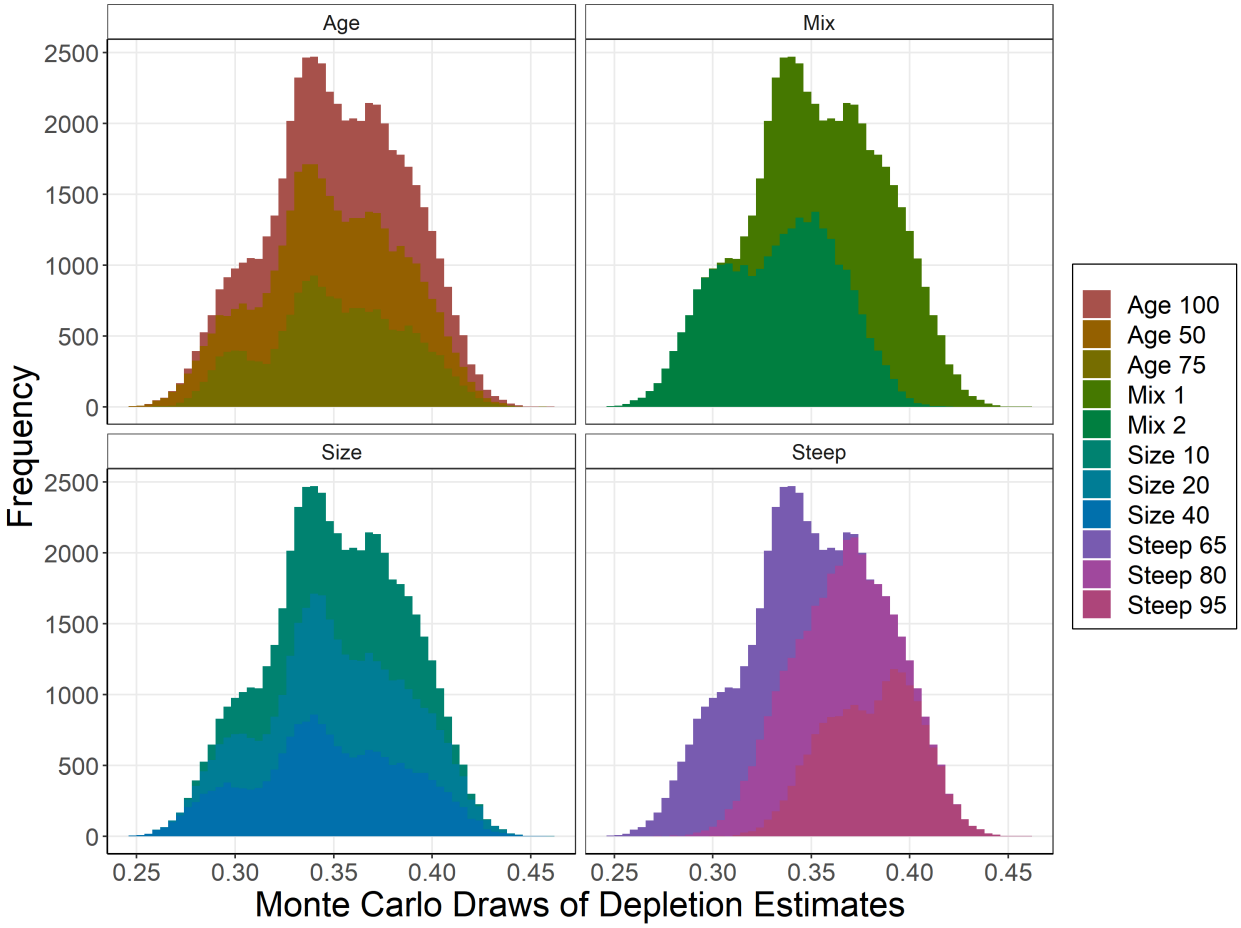

Figure 65: Distribution of $SB_{\text{recent}}/SB_{F=0}$ integrating model and estimation uncertainty, presented by uncertainty axis (panels) and axis element values (colours).

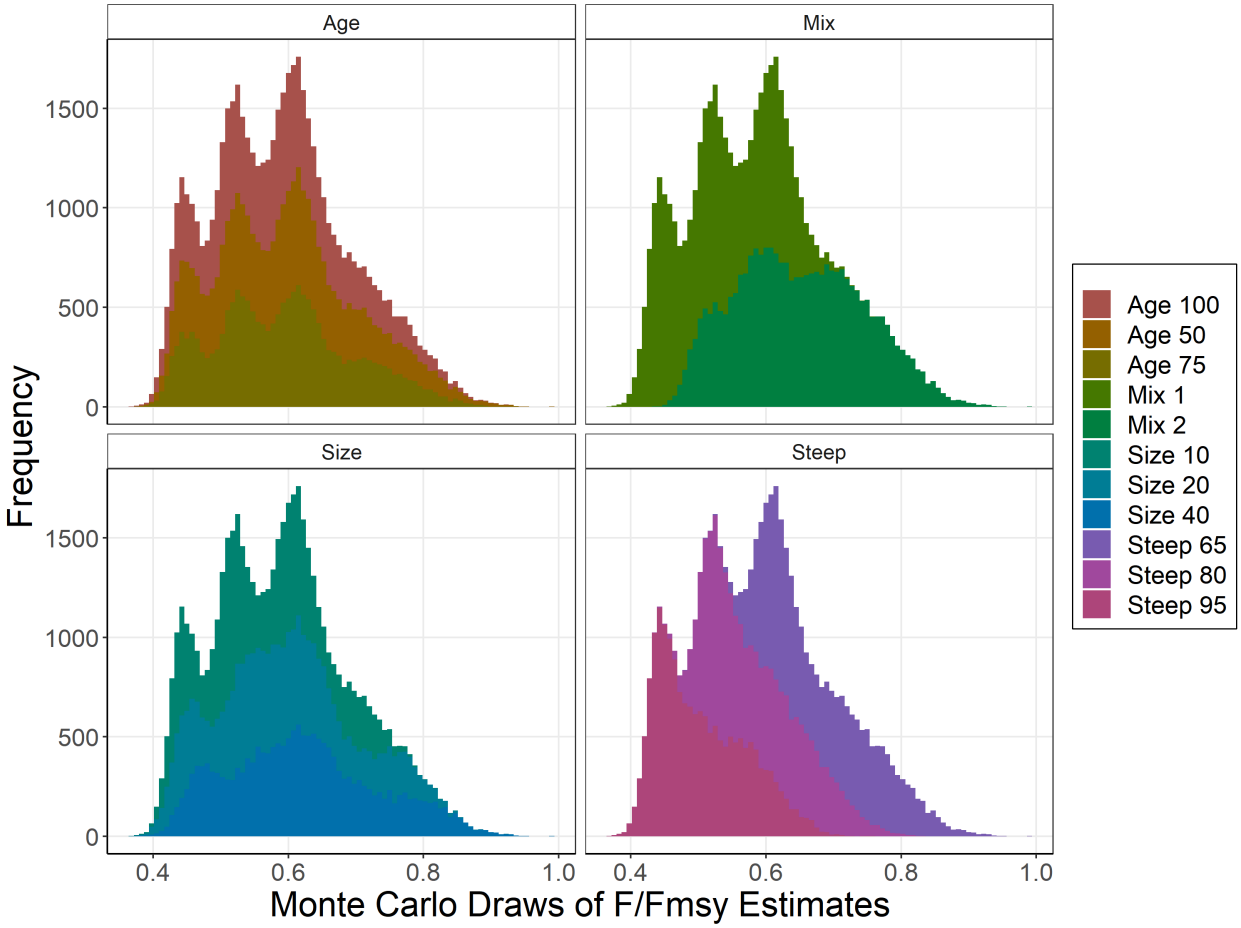

Figure 66: Distribution of $F_{\text{recent}}/F_{\text{MSY}}$ integrating model and estimation uncertainty, presented by uncertainty axis (panels) and axis element values (colours).

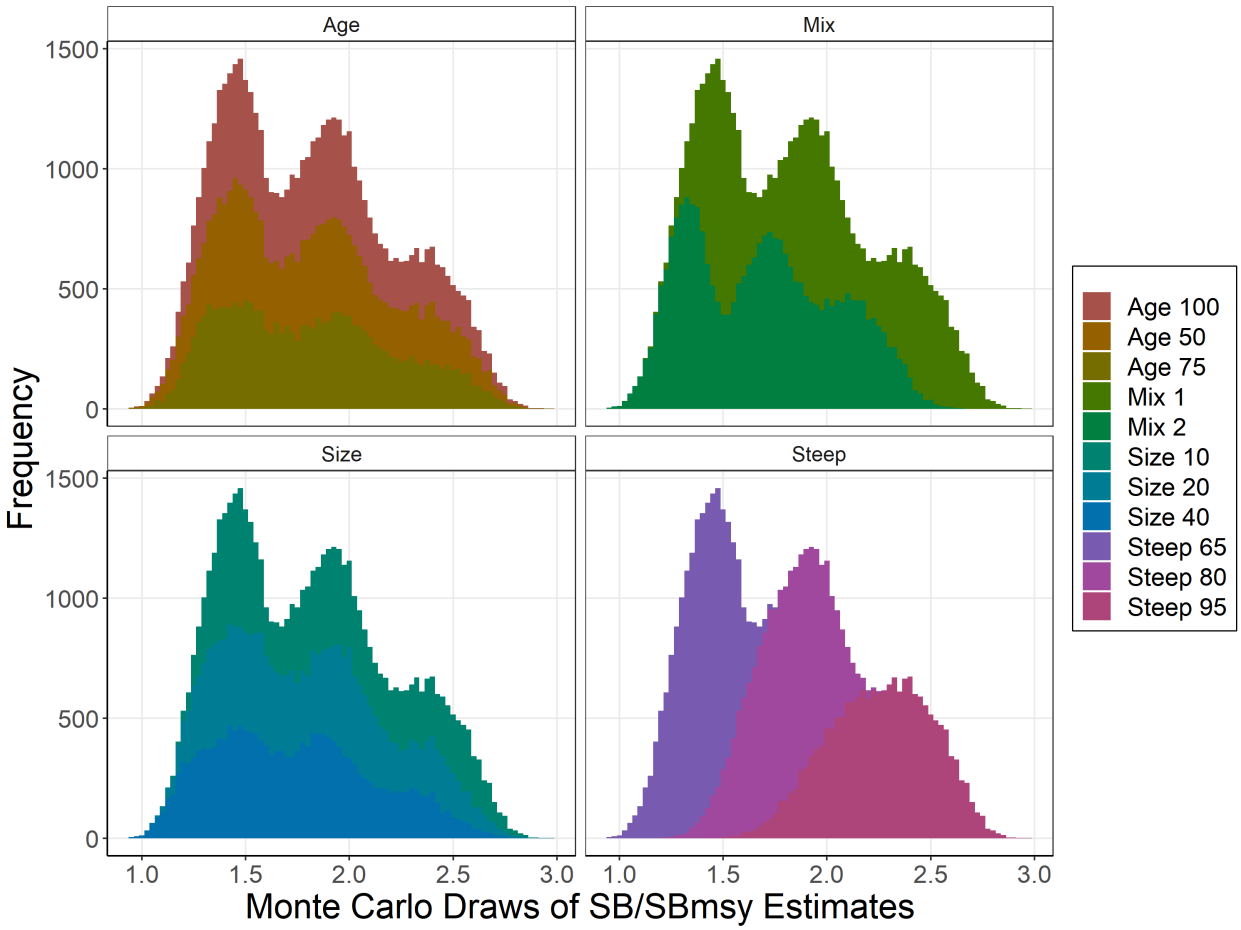

Figure 67: Distribution of $SB_{\text{recent}}/SB_{\text{MSY}}$ integrating model and estimation uncertainty, presented by uncertainty axis (panels) and axis element values (colours).

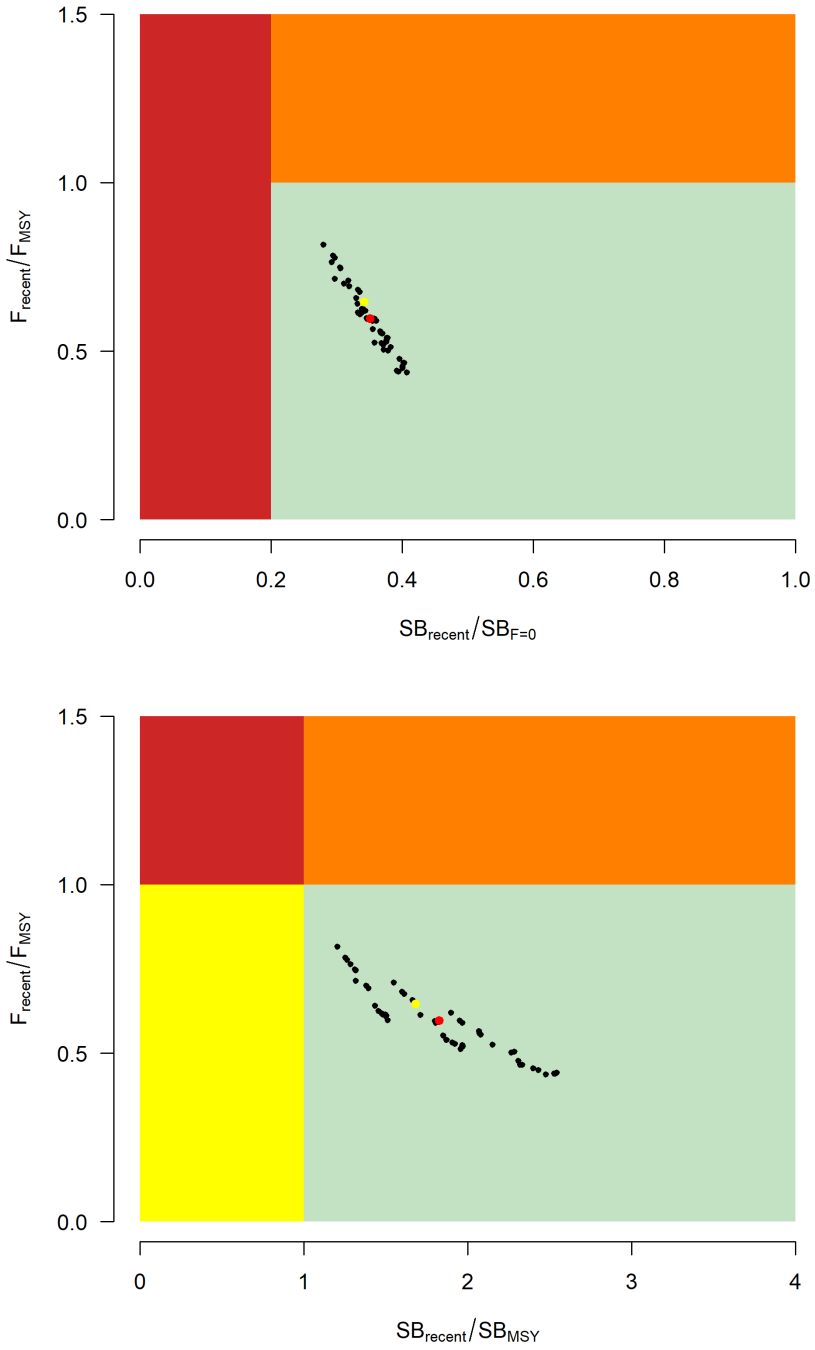

Figure 68: Majuro plot (top) and Kobe plot (bottom) summarising the results for each of the models in the structural uncertainty grid for the recent period (2018-2021). The yellow point is the 2023 diagnostic model and the red point is the median.

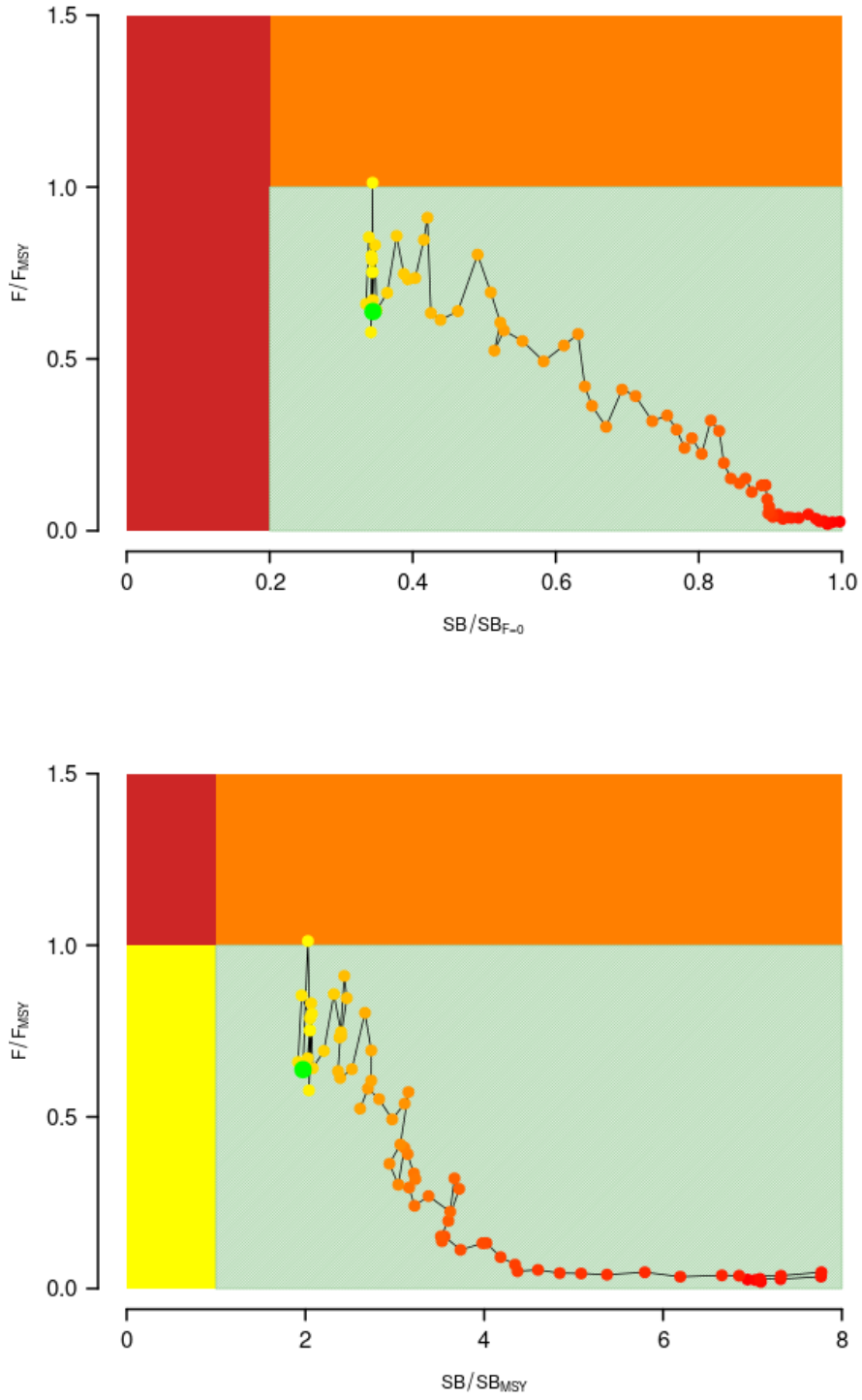

Figure 69: Time dynamic Majuro (top) and Kobe (bottom) plots summarising the results for the 2023 diagnostic model over the model period. The larger green point is the estimated 2021 status.

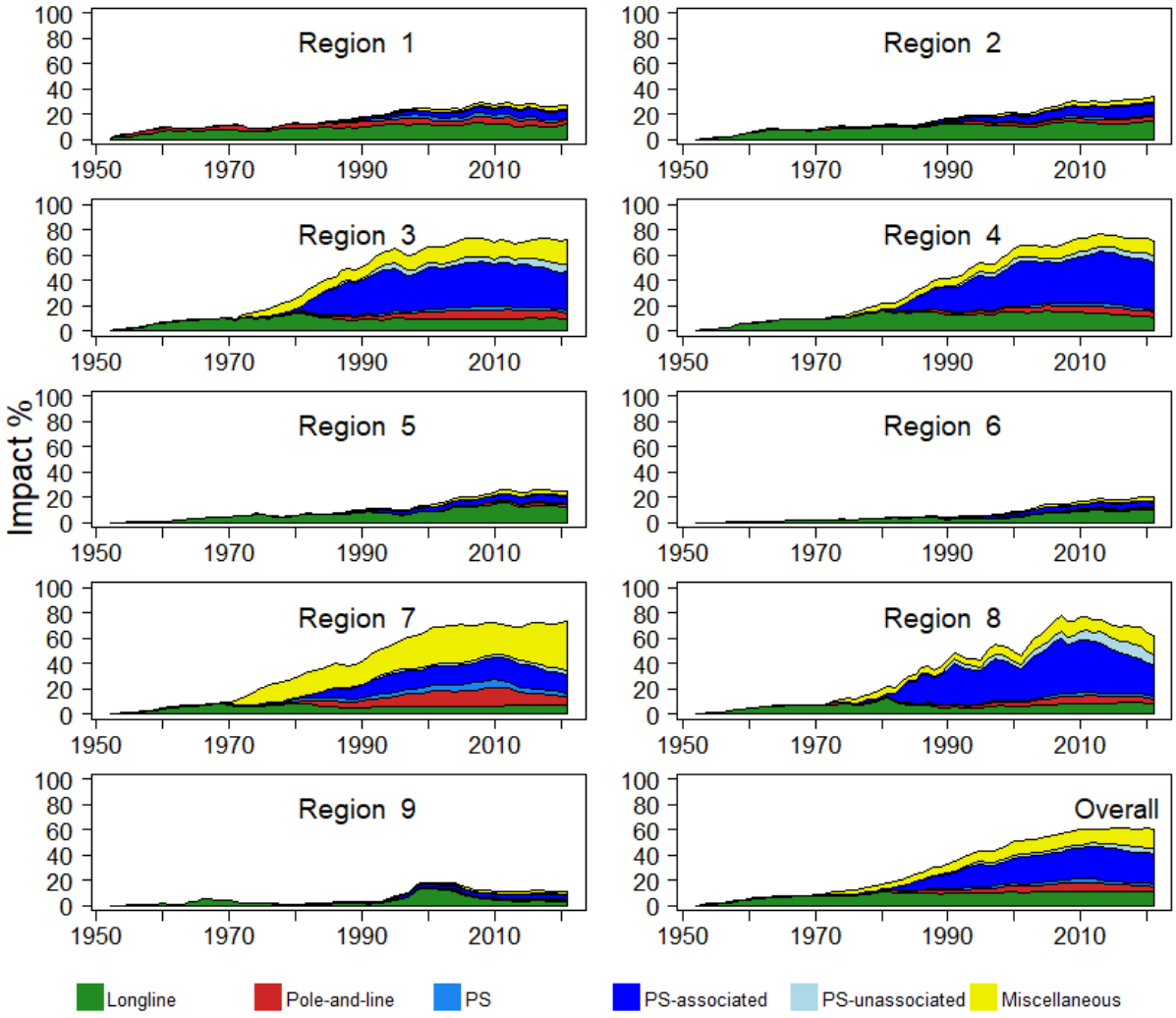

Figure 70: Estimates of fishery impact, or reduction in spawning potential due to fishing (Fishery Impact $= 1 - SB_t/SB_{t,F=0}$) by region, and over all regions (bottom right), attributed to various fishery groups for the 2023 diagnostic model.

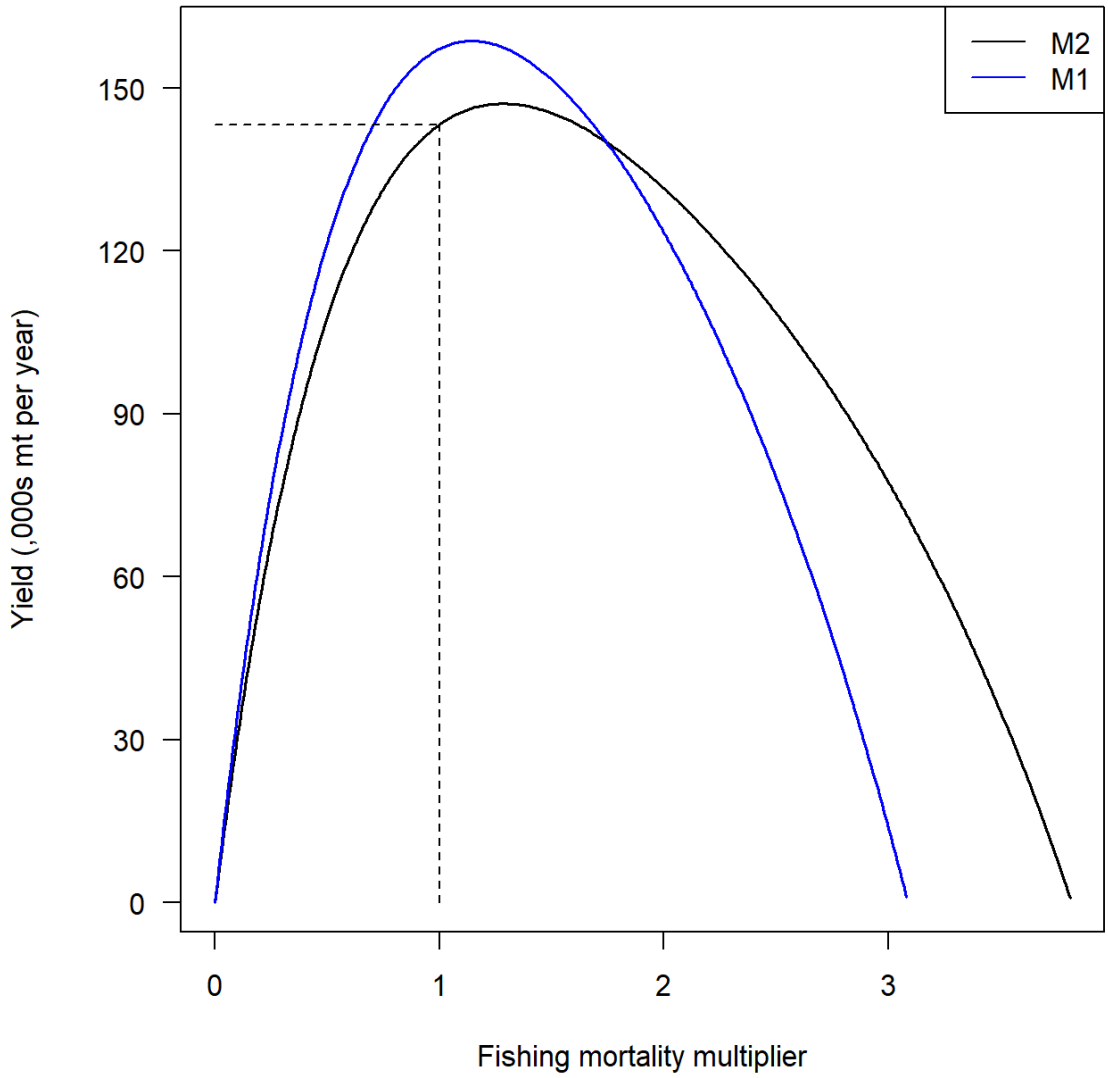

Figure 71: Estimated yield as a function of fishing mortality multiplier for the diagnostic model and the alternative mixing scenario model. The black dashed line indicates the equilibrium yield at current fishing mortality.

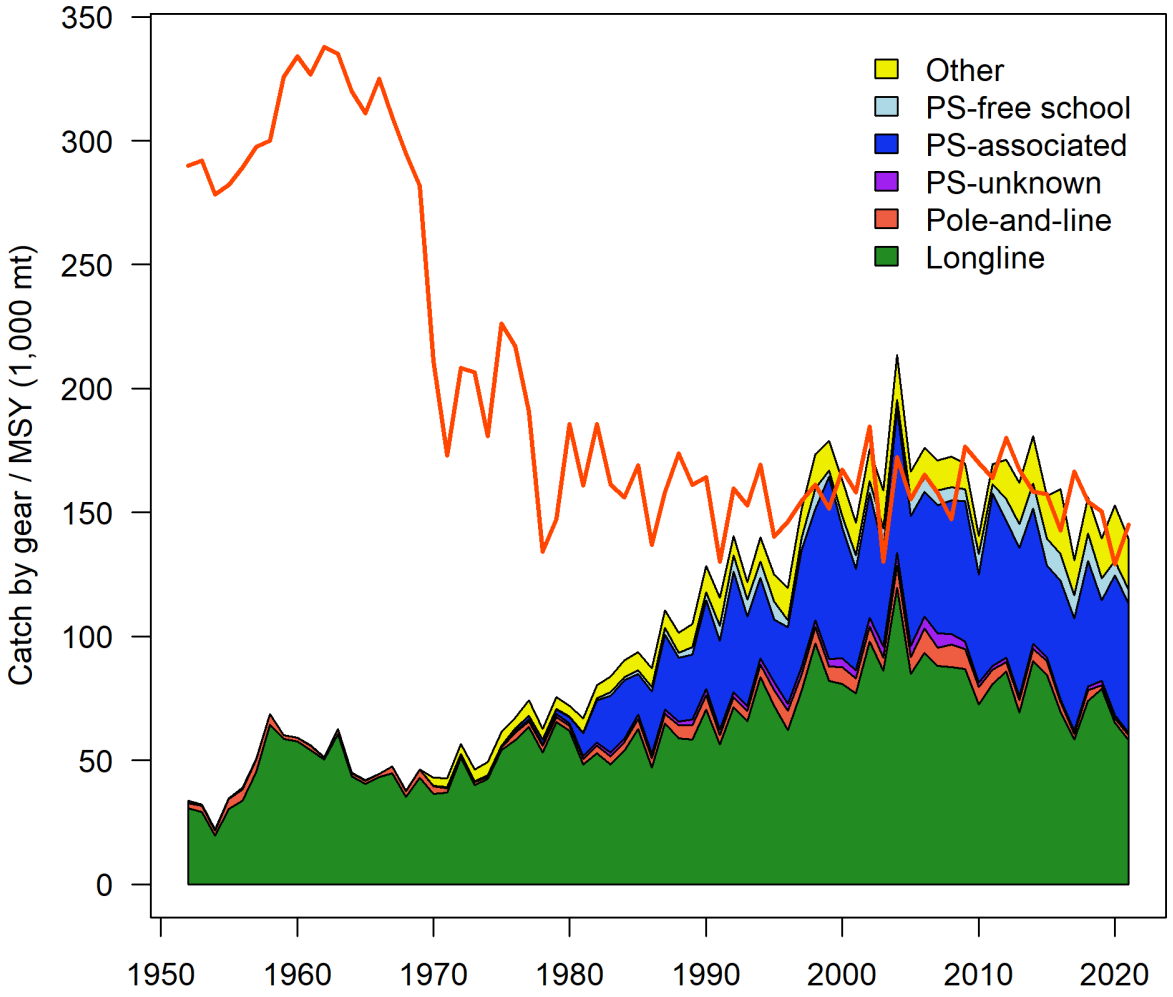

Figure 72: History of the annual estimates of MSY (red line) for the diagnostic model compared with annual catch by the main gear types. Note that this is a 'dynamic' MSY which is explained further in Section 7.5.4.

12 Appendices

12.1 Likelihood profiles

The method for calculating a likelihood profile of the total population scaling parameter is outlined in Section 5.6.

This likelihood profile (Section 5.6) of the total population scaling parameter (Figure 12.1) shows the contribution to the total likelihood from each major data source with the total likelihood shown in black and components of the total likelihood from different data sources shown in a range of colours.

The lower figure is plotted at a scale showing the population scaling parameter value where the minimum likelihood occurs for each major likelihood component. The upper figure shows the same profile but with the plot restricted between population scaling parameters, where the change in total likelihood is not statistically significant, less than 1.92 likelihood units of change, indicated by the horizontal dashed line. Despite the relatively flat curve for the total likelihood, this plot highlights considerable conflict in the data, with the preference in the weight composition data for a higher value of the estimated total population counterbalanced by a preference for lower values for the tag data, the CPUE and, to a lesser degree, the length frequency data. The age data is not influential in estimating the population scaling parameter.

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

The Piner plot for weight (Figure 12.2) shows that the weight data from most fisheries support a larger biomass, with the largest contribution from the longline offshore fishery in region 7 (6.LL.OS.7), followed by the index fishery in region 4 (36.LL.IDX.4), a longline fishery in region 3 (4.LL.ALL.3) and the index fishery in region 1 (233.LL.IDX.1). In contrast to the majority of these fisheries, a longline fishery in region 3 (5.LL.OS.3) suggests a lower biomass, with the index fishery in region 5 (37.LL.IDX.5) the only other fishery suggesting a lower biomass, albeit with a very minor influence.

The Piner plot for length (Figure 12.3) shows that the length data are largely uninformative on the population scaling parameter value for most fisheries, and a single fishery is making a major contribution to this component of the total likelihood, the miscellaneous or domestic fishery operating out of the Philippines in region 7 (17.MISC.PH.7) which supports a smaller value for biomass.

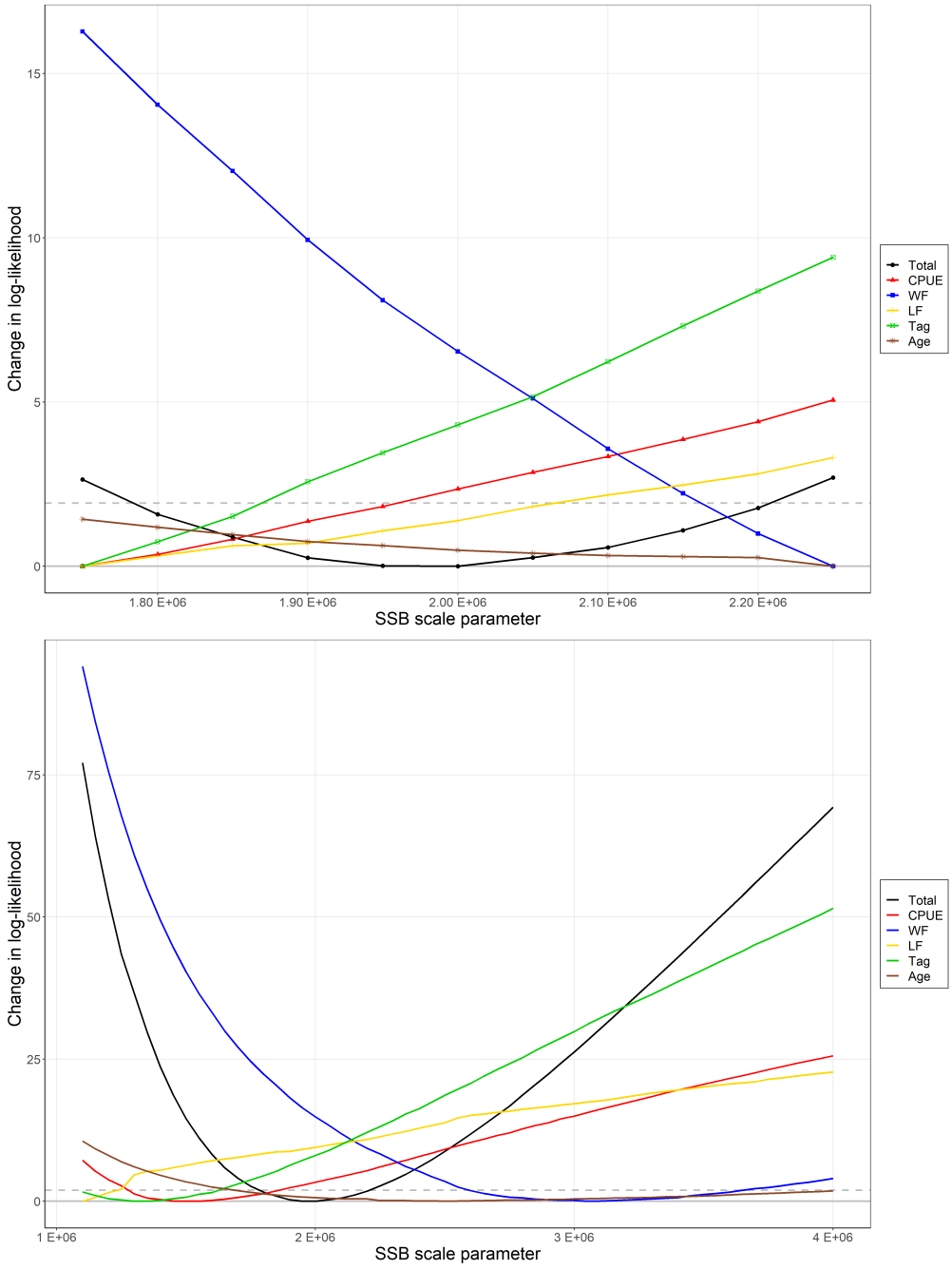

Figure 12.1: Likelihood profiles of total biomass in million mt. The black line indicates the total likelihood with the colours representing component of the total likelihood coming from different data sources. The upper plot focuses on the range of total biomass values where the change in total likelihood is not statistically significant, with less than 1.92 likelihood units of change, indicated by the horizontal dashed line. The lower plot show a broader range indicating the minima for most of the individual data sources.

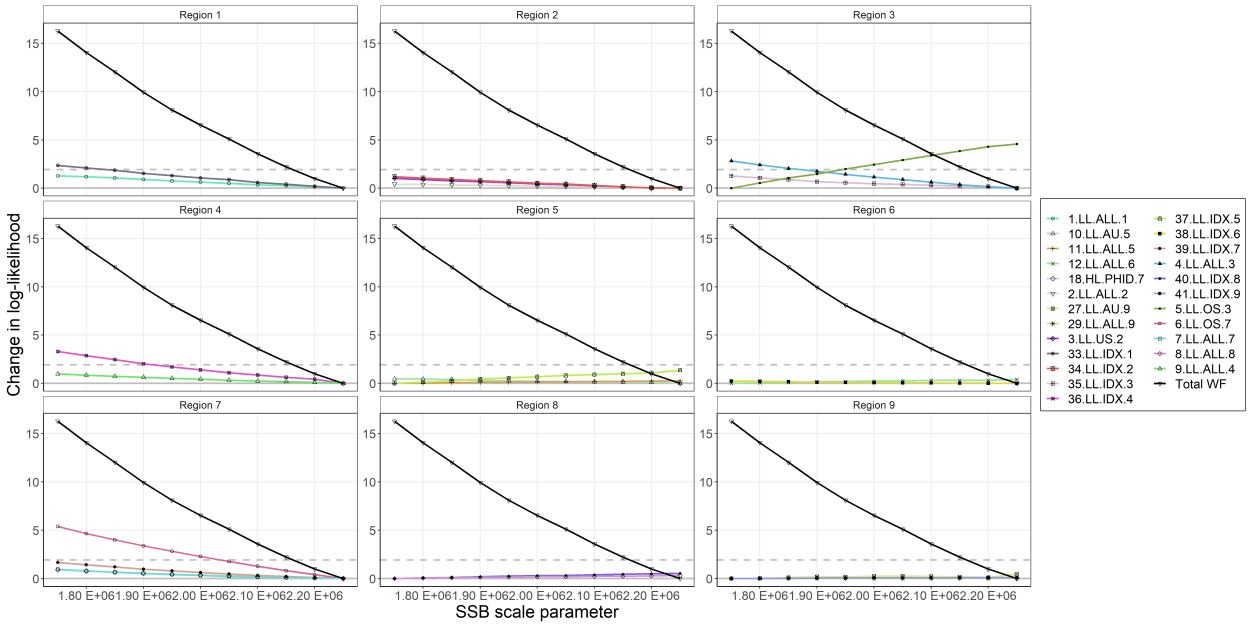

Figure 12.2: Profiles of the weight component of the total log-likelihood by individual fishery and region, plotted over the range where the change in total likelihood is not statistically significant, with less than 1.92 likelihood units of change, indicated by the horizontal dashed line. The black line indicates the weight component of the likelihood, with the colours and panels representing components of the weight likelihood separated into regions and fisheries.

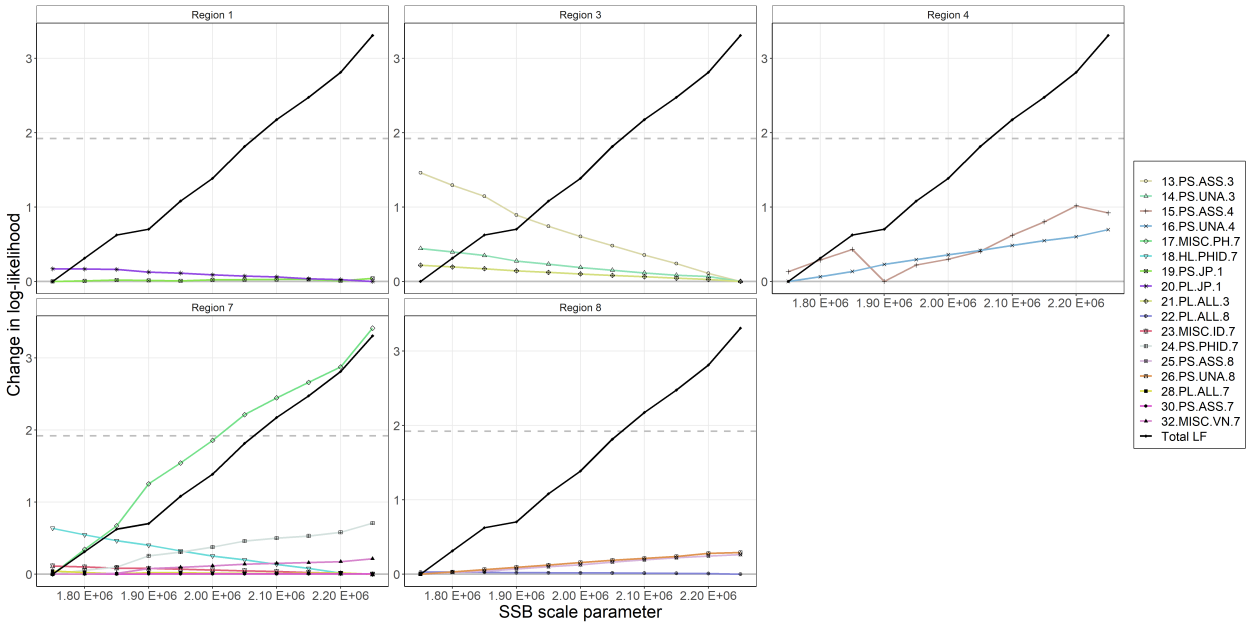

Figure 12.3: Profiles of the length component of the total log-likelihood by individual fishery and region, plotted over the range where the change in total likelihood is not statistically significant, with less than 1.92 likelihood units of change, indicated by the horizontal dashed line. The black line indicates the length component of the likelihood, with the colours and panels representing components of the length likelihood separated into regions and fisheries.

12.2 Retrospective analyses

A retrospective analysis involves running a series of models, each with less year of data than the previous model. This enables exploration of the effects of truncating the input data time series, by removing the most recent years of data to the model, one at a time. A series of seven retrospective models were fitted, starting from the diagnostic model, which has data through to 2021. These successive models (or peels), each feature one less year of data than the previous peel, by removing the last year of data from the inputs.

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 0.085 for the spawning depletion retrospectives, and -0.046 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 bigeye diagnostic model indicate that there is no concern for retrospective bias with the 2023 diagnostic model.

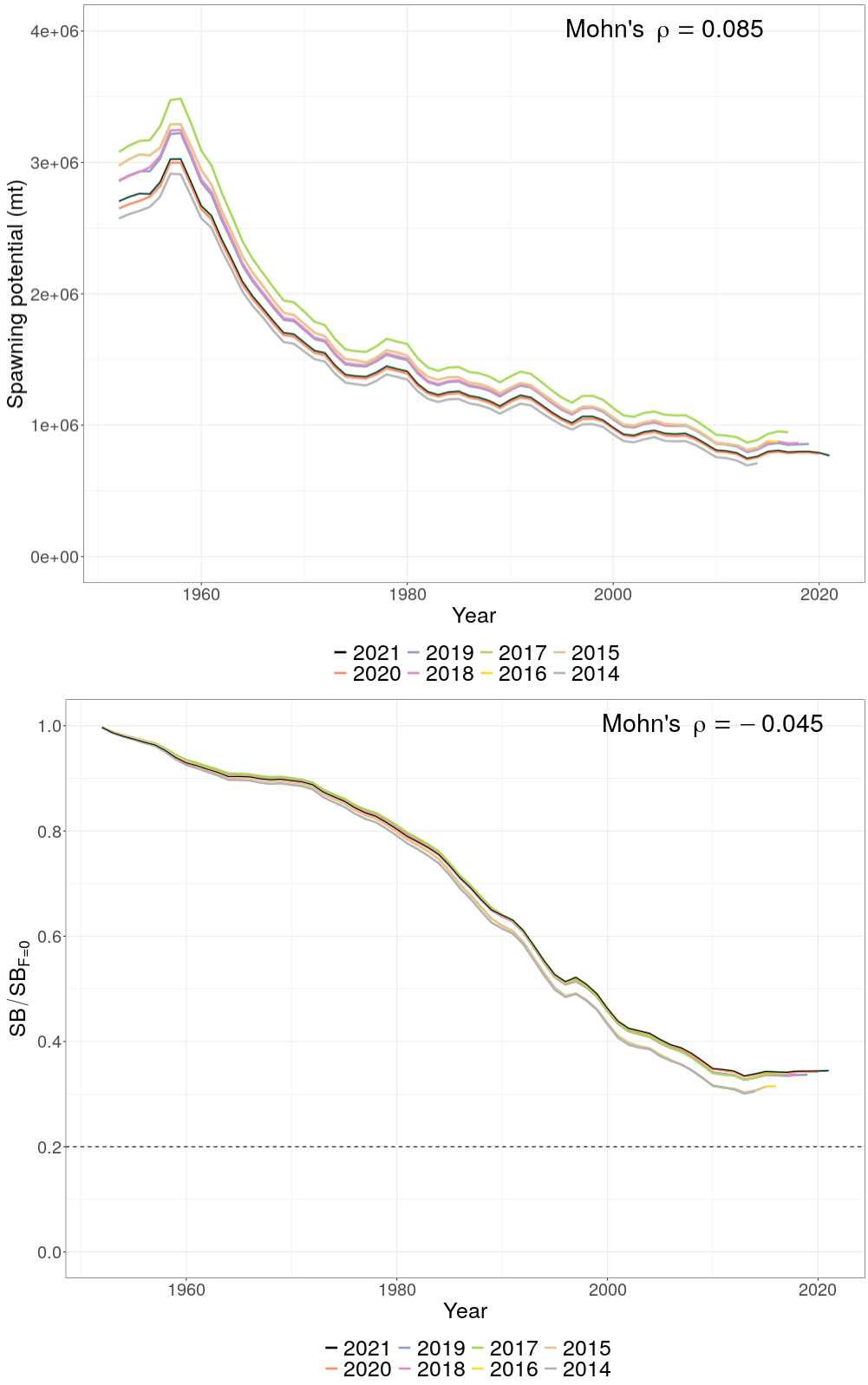

Figure 12.4: Estimated spawning potential (top), and spawning depletion (bottom) for the retrospective models.

12.3 "Status quo" stochastic stock projections for WCPO bigeye 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.

Aires-da Silva, A. M., Maunder, M. N., Schaefer, K. M., and Fuller, D. W. (2015). Improved growth estimates from integrated analysis of direct aging and tag-recapture data: An illustration with bigeye tuna (Thunnus obesus) of the eastern Pacific Ocean with implications for management. *Fisheries Research*, 163:119–126.

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.

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.

Chow, S., Okamoto, H., Miyabe, N., Hiramatsu, K., and Barut, N. (2000). Genetic divergence between Atlantic and Indo-Pacific stocks of bigeye tuna (Thunnus obesus) and admixture around South Africa. *Molecular Ecology*, 9(2):221–227.

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., Fournier, D. A., Hampton, J., and Bouyé, F. (2015). Recent developments and future plans for MULTIFAN-CL. Technical Report WCPFC-SC11-2015/SA-IP-01, Pohnpei, Federated States of Micronesia, 5–13 August 2015.

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

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

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

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

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

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

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

Gonzalez, E. G., Beerli, P., and Zardoya, R. (2008). Genetic structuring and migration patterns of Atlantic bigeye tuna, Thunnus obesus (Lowe, 1839). *BMC Evolutionary Biology*, 8(1):252.

Grewe, P. and Hampton, J. (1998). An assessment of bigeye (*Thunnus obesus*) population structure in the Pacific Ocean based on mitochondrial DNA and DNA microsatellite analysis. Technical report, JIMAR Contribution 98-330.

Grewe, P. M., Wudianto, C., Proctor, C. H., Adam, M. S., Jauhary, A. R., Schaefer, K., Itano, D. G., Evans, K., Killian, A., Foster, S. D., Gosselin, T., Feutry, P., Aulich, J., Gunasekera, R. M., Lansdell, M., and Davies, C. R. (2019). Population Structure and Connectivity of Tropical Tuna Species across the Indo Pacific Ocean Region. Technical Report WCPFC-SC15-2019/SA-IP-15.

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 Williams, P. (2005). A description of tag-recapture data for bigeye tuna in the western and central Pacific Ocean. *Col. Vol. Sci. Pap. ICCAT*, 57(2):85–93.

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., 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., Hoyle, S., and Langley, A. (2009). Stock assessment of bigeye tuna in the Western and Central Pacific Ocean. Technical Report WCPFC-SC5-2009/SA-WP-04, Port Vila, Vanuatu, 10–21 August 2009.

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., Hoyle, S., Hampton, J., and Kleiber, P. (2010). Stock assessment of bigeye tuna in the western and central Pacific Ocean. Technical Report WCPFC-SC6-2010/SA-WP-01, Nuku'alofa, Tonga, 10–19 August 2010.

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.

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., Hampton, J., Kleiber, P., and Hoyle, S. (2008). Stock assessment of bigeye tuna in the western and central Pacific Ocean, including an analysis of management options. Technical Report WCPFC-SC4/SA-WP-1.

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., Hampton, J., Abascal, F., Davies, N., and Harley, S. J. (2015a). Sensitivity of WCPO stock assessment results to the inclusion of EPO dynamics within a Pacific-wide analysis. Technical Report WCPFC-SC11-2015/SA-WP-04, Pohnpei, Federated States of Micronesia, 5–13 August 2015.

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

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

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

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

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.

Moore, B. R., Lestari, P., Cutmore, S. C., Proctor, C., and Lester, R. J. G. (2019). Movement of juvenile tuna deduced from parasite data. *ICES Journal of Marine Science*, 76(6):1678–1689.

Natasha, J., Stockwell, B. L., Marie, A. D., Hampton, J., Smith, N., Nicol, S., and Rico, C. (2022). No Population Structure of Bigeye Tunas (Thunnus obesus) in the Western and Central Pacific Ocean Indicated by Single Nucleotide Polymorphisms. *Front. Mar. Sci.*, 9:799684.

Nicol, S., Hoyle, S., Farley, J., Muller, B., Retalmai, S., Sisior, K., and Williams, A. (2011). Bigeye tuna age, growth, and reproductive biology (project 35). Technical Report WCPFC-SC7-2011/SA-WP-01, Pohnpei, Federated States of Micronesia, 9–17 August 2011.

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

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.

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.

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

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

Punt, A. E. (2015). Some insights into data weighting in integrated stock assessments. *Fisheries Research*, page 14.

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.

Rice, J., Harley, S., Davies, N., and Hampton, J. (2014). Stock assessment of skipjack tuna in the Western and Central Pacific Ocean. Technical Report WCPFC-SC10-2014/SA-WP-05, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

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

Schaefer, K., Fuller, D., Hampton, J., Caillot, S., Leroy, B., and Itano, D. (2015). Movements, dispersion, and mixing of bigeye tuna (ıt Thunnus obesus) tagged and released in the equatorial Central Pacific Ocean, with conventional and archival tags. *Fisheries Research*, 161:336–355.

Schaefer, K. M., Fuller, D. W., and Miyabe, N. (2005). Reproductive biology of bigeye tuna (ıt Thunnus obesus) in the eastern and central Pacific Ocean. Technical Report 23(1). 35p., IATTC Bulletin.

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.

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.

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-Rev2.

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.

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