SCIENTIFIC COMMITTEE
SIXTEENTH REGULAR SESSION
Online
11–20 August 2020

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

WCPFC-SC16-2020/SA-WP-03 [REV3]

N. Ducharme-Barth[1], M. Vincent[1], J. Hampton[1], P. Hamer[1], P. Williams[1], G. Pilling[1]

[1]Oceanic Fisheries Programme, The Pacific Community

Note: At the time of paper submission, the computationally intensive calculations for the likelihood profile and Hessian had not completed for the diagnostic case model. Section 11.2 will be updated as a revision to the paper [REV1].

[REV1] Results from the likelihood profile (Section 11.2) and the Hessian diagnostic (Section 11.5) are included in the Appendix. Figures 25 and 28 have been updated to include the asymptotic 95% confidence interval as calculated using the delta-method. The first paragraph of Appendix Section 11.8 was also revised to more clearly state the assumptions made in the simplified model that was investigated. A subsequent revision to the paper will include the projection analysis [REV2].

[REV2] Results from the stochastic projections are included in Section 11.6.

[REV3] Figure 52 (c) has been revised to not show the last two years of total recruitment as these are fixed at the mean level and are thus misleading when attempting to discern retrospective bias.

Contents

1 Executive Summary

This paper describes the 2020 stock assessment of bigeye tuna (*Thunnus obesus*) in the western and central Pacific Ocean. A further three years of data were available since the last full stock assessment conducted in 2017 (that in 2018 being an update assessment), and the model extends through to the end of 2018. New developments to the stock assessment include:

- Addressing recommendations of the 2017 stock assessment report (McKechnie et al., 2017a), enhanced growth information through the addition of age-at-length information from smaller fish using daily ageing, and the integration of growth information from tag recaptures, as well as the implementation of the Richards growth model.
- Updates to the definition of reproductive potential
- Implementation of a composite "index" longline fishery for each model region which received that region's standardized CPUE index, and concurrently a "pseudo catch conditioned" approach was taken for the assessment

Based upon recommendations of SPC's 2020 pre-assessment workshop, only the 10° N spatial structure was considered within this assessment.

The assessment is supported by the analysis of longline catch-per-unit-of-effort (CPUE) data (Ducharme-Barth and Vincent, 2020), background analyses of biological parameters, compilation of the tagging data and definition of the fisheries structures (Vincent and Ducharme-Barth, 2020), tagging data treatment (Peatman, 2020; Scutt Phillips et al., 2020), preparation of the length and weight composition data (Peatman et al., 2020), new otolith ageing and growth work (Farley et al., 2020) and estimation of growth using a tag-integrated model (Eveson et al., 2020). Key changes made in the progression from the 2017 to 2020 diagnostic case models include:

- Updating all data up to the end of 2018.
- Implementation of updated models for tag data (tagger effects, tag reporting rates, sliding window for tag mixing period, tag reporting rate priors), purse seine catch estimates and size composition data.
- Implementation of the "index fishery" approach and utilization of spatiotemporal model standardized CPUE indices.
- Utilizing updated biological parameters for the length-weight relationship and reproductive potential.
- Implementation of growth as defined by the tag-integrated Richards model.

In addition to the diagnostic case model, we report the results of one-off sensitivity models to explore the relative impacts of key data and model assumptions for the diagnostic case model on the stock assessment results and conclusions. We also undertook a structural uncertainty analysis

(model grid) for consideration in developing management advice where all possible combinations of those areas of uncertainty from the key one-off models were included.

As per recent Scientific Committee practice, it is recommended that management advice is formulated from the results of the structural uncertainty grid. Broadly speaking, the results from the current assessment are consistent with those from the previous stock assessment (McKechnie et al., 2017a; Vincent et al., 2018) following the adoption of the *Updated New* growth and the *10° N* regional structure. The most important factor contributing to the uncertainty around the estimated stock status is the level of data-weighting given to the size-frequency data in the model. Greater weight placed on the size-frequency data resulted in more optimistic assessment outcomes, while increased down-weighting produced more pessimistic estimates and stock trajectories. Within levels of the different size-frequency weightings, different assumptions on the average level of natural mortality and the assumed steepness had a lesser but predictable effect. Higher assumed average natural mortality and/or steepness resulted in a stock that was estimated to be more productive and less depleted relative to the other levels of natural mortality and steepness.

The main conclusions of the current assessment are summarized as follows:

- All models in the structural uncertainty grid show WCPO bigeye tuna to be above $20\%SB_{F=0}$.
- However, there is evidence to suggest that the overall stock status estimated by the model is being "buffered" by the temperate regions. The most pessimistic predictions of overall stock status correspond to models where depletion in these temperate regions is predicted to be high.
- The equatorial regions appear to show consistent levels of regional depletion that approach estimates of $20\%SB_{F=0}$ across models.
- A substantial decline in bigeye abundance was estimated by all models in the assessment. With respect to the difference in recent levels of depletion and those from the beginning of the model period, the impacts of fishing pressure appear to be persistent and meaningful, at least on a multi-year scale.
- The effects of the relatively large recruitments seen at the end of the previous stock assessment do not appear to have continued in recent years. Most models across the grid show a downturn in stock status in recent years, which appears to match the pattern seen in the standardized CPUE.

2 Introduction

This paper presents the 2020 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). As in previous assessments, the objectives of the 2020 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[2] (MFCL; Fournier et al., 1998; Hampton and Fournier, 2001; Kleiber et al., 2019). MFCL implements a size-based, age- and spatially-structured population model. Model parameters are estimated by maximizing an objective function, consisting of both likelihood (data) and "prior"[3] information components (penalties).

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

A feature of the tuna stock assessments in the WCPO is the use of an "uncertainty grid". The uncertainty grid is a group of models that are run to explore the interactions among selected "axes" of uncertainty that relate to biological assumptions, data inputs and data treatment. The axes are generally selected from one-off sensitivity runs of a diagnostic (or base case) model to indicate uncertainties that have notable effects on the estimates of key model parameters and management quantities. The uncertainty grid approach may involve many models depending on the number of axes and the number of alternative assumptions for each axis, as models are run for all combinations of the axes and alternative assumptions within each axis. Importantly, the uncertainty grid captures variability in model estimates due to assumptions in model structure that are not accounted for by statistical uncertainty estimated in a single model run, or a set of one-off sensitivities. Management

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

[3] 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.

advice should typically be based on consideration of the range of model estimates produced by all models in the uncertainty grid and the relative plausibility of each model scenario.

The most recent assessment was conducted in 2017 (McKechnie et al., 2017a). The 2017 assessment utilized a new growth data set from otolith ageing and considered alternative spatial structures. These changes, in particular the new growth data, had an important influence on the outcome of the assessment, which suggested that the stock was less depleted and fishing pressure was not as high as indicated by the previous assessment. A range of alternative models were presented to cover what were considered the key areas of uncertainty. Overall, the previous assessment indicated that depletion and exploitation status of bigeye in the WCPO were somewhat uncertain, but that the majority of models indicated that the stock was not overfished or undergoing overfishing according to WCPFC reference points. This assessment was revisited in 2018 with a re-evaluation of the 2017 assessment to further explore the alternative growth and regional structures (Vincent et al., 2018), with SC14 electing to make management recommendations based on the "new" growth from otolith ageing (WCPFC, 2018).

The 2020 assessment includes further development of growth information by enhancing the previous otolith ageing with smaller fish using daily ageing (Farley et al., 2020), and the integration of growth information from tag-recaptures (Eveson et al., 2020). Preparatory work on data for each assessment is an extensive and time-consuming part of the stock assessment process. The preparatory work is covered in limited detail in this paper and this assessment report should be read in conjunction with several supporting papers:

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

3 Background

3.1 Stock structure and movement

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 (Alvarado Bremer et al., 1998; Chow et al., 2000; Gonzalez et al., 2008). However, there is currently no clear evidence for genetic population structure in the Pacific Ocean

(Grewe and Hampton, 1998; Moore et al., 2020). 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), but while longitudinal movements are common, latitudinal movements are thought to be less extensive (Moore et al., 2020). These occasional large-scale movements and the continuous distribution of bigeye tuna across the Pacific would explain the lack of broad-scale genetic structure. This does not necessarily preclude regional 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 (McKechnie et al., 2015a; Aires-da Silva et al., 2015; 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. Despite uncertainties in population connectivity within the Pacific, stock assessment outcomes for bigeye tuna in the WCPO have been shown to be relatively insensitive to inclusion of data from the eastern Pacific Ocean (EPO) (McKechnie et al., 2015a), supporting the continued approach of separate assessments for the WCPO and EPO regions. Improving understanding of spatial population processes for bigeye in the WCPO, and more broadly in the Pacific, remains an important area for research.

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). For the current bigeye tuna assessment, the stock within the domain of the model area (essentially the WCPO, west of 150° W) is considered to be a discrete stock unit as in previous assessments (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). A sensitivity analysis by McKechnie et al. in 2015 showed that despite extensive longitudinal movements in the equatorial Pacific, the discrete stock assumption was capable of accurately capturing the dynamics and stock status indicators for WCPO bigeye tuna.

Within the WCPO, the spatial complexity of the modeling of bigeye tuna has increased. In the 2011 assessment the model domain was divided into 6 regions. After a review of the bigeye assessment in 2012 (Ianelli et al., 2012), a 9 region structure was implemented in 2014 for both yellowfin and bigeye (McKechnie et al., 2014; Davies et al., 2014; Harley et al., 2014), with the northern boundary of regions 3 and 4 set at 20° N latitude. In the 2017 bigeye tuna stock assessment (McKechnie et al., 2017a) an additional option was included that involved moving the northern

boundary of regions 3 and 4 to 10° N latitude to better reflect the spatial structure of the purse seine fisheries (which are largely restricted to the zone between 10°S and 10°N) and the assumption of low movement rates between the equatorial and sub-tropical northern regions. The implications of these alternative regional structures were previously investigated in the 2017 assessment and 2018 update assessment (McKechnie et al., 2017a; Vincent et al., 2018). Based on these analyses, the 2020 PAW (Hamer and Pilling, 2020) made the recommendation to only consider the 10° N option in the 2020 assessment of bigeye tuna. This 9 region spatial structure (Figure 1) 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. region 8 and 9).

3.2 Biological characteristics

Bigeye tuna are relatively fast growing, 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). Based on the recommendations of SC14 (WCPFC, 2018), recent work has been conducted to improve the knowledge and confidence of bigeye tuna growth parameters in the WCPO region. This involved analyses of over 1200 otoliths, including small fish from assessment region 7 (Farley et al., 2020). Additionally, growth increment data from over 2,500 tag-recaptures was combined with the otolith data set in an integrated analysis of bigeye tuna growth in the WCPO (Eveson et al., 2020). The results of these analyses show that while lengths at age are estimated to be similar from the otolith only and tag-integrated based growth data-sets, there are slight differences in the estimated growth parameters between the two growth models.

Available data indicate that 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 sex ratio of bigeye tuna changes at around the age/size of reproductive maturity to favor males at larger size (see Figure 8 in 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.

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; Bousquet et al., 2012). 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). The longest period at liberty for a recaptured bigeye tuna tagged in the WCPO is approximately 14 years, for a fish 1-2 years old at release (SPC unpublished data) and the oldest fish aged from

otoliths in the WCPO is approximately 15 years (Farley et al., 2020).

Given the recently completed studies on bigeye tuna age and growth (Eveson et al., 2020; Farley et al., 2020), a meta-analysis using different life-history based M estimators was conducted to provide a range of plausible natural mortality values for the assessment (Piner and Lee, 2011). 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. As in previous assessments, this feature of the population dynamics has also been incorporated into the natural-mortality-at-age schedules used in the current assessment.

3.3 Fisheries

Bigeye tuna are an important component of tuna fisheries throughout the Pacific Ocean and are taken by both surface gears, mostly as juveniles, and longline gear, as more valuable adult fish. They are a principal target species in tropical waters of 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 (Allain et al., 2016). In recent years the prices paid for both frozen and fresh product on the Japanese sashimi market are the highest of all the tropical tunas. The bigeye tuna longline catch in the WCPFC area had a "delivered" value in 2018 of approximately US $660 million (Williams and FFA, 2019).

Bigeye caught by purse seine vessels are taken almost exclusively from sets on natural and artificial floating objects (FADs). Estimation of the bigeye (and to a lesser extent yellowfin) tuna catch from associated sets has been the focus of considerable research over several years (Peatman et al., 2019). The purse seine fishery mostly targets skipjack, and to a lesser extent yellowfin, though significant incidental catches of small bigeye occur. This fishery expanded rapidly from the early 1980's and the estimated annual bigeye catch for this gear has recently been in the vicinity of 40,000-70,000 mt. The highest estimated bigeye tuna total catch from the WCPO was approximately 192,000 metric tonnes in 2004, but in recent years has been around 140,000-150,000 metric tonnes (Williams et al., 2020). Over the recent period (2015-2019), 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, due to this fishing method selecting for smaller fish. 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.

In recent years, collaborative work between SPC, WCPFC, CSIRO (primarily in Indonesia), and fisheries agencies in Indonesia, 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 have been included in the assessment, as described below.

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 may also be used directly as data in the assessment model. Several companion papers detail the work on input data to support this assessment. Readers should refer to these papers highlighted at the end of Section 2 for detailed descriptions of how the data and biological inputs were formulated, as only brief descriptions will be provided in this report.

4.2 Spatial stratification

The broad geographic area considered in the 2020 assessment covers the Pacific Ocean to the west of 150° W, and between 50° N and 40° S. The eastern boundary of the assessment excludes the WCPFC Convention area component that overlaps with the IATTC area. The region extends to 140° E off south-east Australia, 110° E in south-east Asia and 120° E in the north western Pacific (Figure 1). Within this area there are 9 spatial assessment regions identified based on consideration of fishery characteristics and movement information from tagging studies (Harley et al., 2014). Consistent with the 2017 assessment, the current configuration has the northern boundaries of regions 3 and 4 being set at 10° N. The rationale for this is described more fully in McKechnie et al. (2017b). Some small regions are included, i.e. region 8 designed to approximate the archipelagic waters of PNG and Solomon Islands, where considerable tagging effort has occurred and the analyses show more persistent residence compared to the wider western equatorial region, and region 9 that was established in 2014 to better model the tagging data from the Coral Sea (Harley et al., 2014).

4.3 Temporal stratification

The time period covered by the assessment is 1952–2018 which includes all significant post-war tuna fishing in the WCPO. Within this period, data were compiled into quarters (1; Jan–Mar, 2; Apr–Jun, 3; Jul–Sep, 4; Oct–Dec). As agreed at SC12, the assessment does not include data from the most recent calendar year 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 defined according to gear type, fishing method and region, and also by vessel flag or fleet. There are 41 fisheries defined for this assessment (Table 1 & Table 2) consisting of index, longline, purse seine, pole and line and various miscellaneous small-fish fisheries in Indonesia and the Philippines. The fisheries definitions for the 2020 assessment are consistent with those used in the 2017 assessment, except for the addition of index fisheries for each region. A graphical summary of the availability of data for each fishery used in the diagnostic case model is provided in Figure 2.

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

4.4.1 Index fisheries

In previous WCPO bigeye assessments, a longline fishery in each region received standardized effort from an external CPUE analysis. These fisheries were assumed to have a fixed catchability across regions and years. As such, their CPUEs served to index abundance across space and time. In this assessment, we used an "index" fisheries approach, first applied in the 2018 assessment of South Pacific albacore (Tremblay-Boyer et al., 2018). Briefly, an index fishery is defined for each of the 9 model regions as a composite fishery composed of all longline fisheries operating in the assessment area. The full longline operational data-set (described in McKechnie et al., 2015b) was used as the basis for the index fisheries. A trivial amount of catch (1 fish per quarter) was assigned so that index fisheries are in effect non-extractive fisheries. Effort for each time step is adjusted, as with the longline fisheries receiving standardized effort in previous assessments, such that the original standardized CPUE derived using a spatiotemporal delta-Generalized Linear Mixed Model or "geostatistical" CPUE standardization model is preserved (see Ducharme-Barth and Vincent, 2020 and Section 4.5.3). The standardized indices for each region are scaled by the regional scaling factors derived from the geostatistical CPUE standardization model. Catchability for the index fisheries is then assumed to be constant over time and shared across the 9 assessment regions in order to scale the population. The regular capture (or extractive) longline fisheries are based on the same data set, but are disaggregated into the longline fisheries defined in Table 1. The size composition data (length & 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 at the spatial (region) and temporal (quarter) resolution of the model (McKechnie, 2014; Peatman et al., 2020). However, for the index fisheries, while the same aggregation process is conducted, the size data is weighted by CPUE (rather than by catch) so that the size data is 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 is adjusted such that the original intended weight (effective sample size) in the likelihood is preserved.

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

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

4.5 Catch and effort data

4.5.1 General characteristics

Catch and effort data were compiled according to the fisheries defined in Table 1 & Table 2. Catches by the longline fisheries were expressed in numbers of fish, and catches for all other fisheries expressed in weight (mt). This is consistent with the form in which the catch data are recorded for these fisheries.

It is also worth noting that this assessment is implemented using a "pseudo" *catch-conditioned* approach (see Section 5.3.2). To accommodate this, effort was removed for all fisheries except for the index fisheries. This also eliminated the need to account for fisheries effort creep through the estimation of time-varying changes in catchability. A nominal effort of one was included for the final year of the model to allow the estimation of a catchability coefficient to assist with projection analyses.

Total annual catches by major gear categories for the WCPO are shown in Figure 3 and a regional breakdown is provided in Figure 4. The spatial distribution of catches over the past 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:

- Bigeye catch by longline vessels steadily increased through the mid-2000s in the equatorial and southern regions. However, these catches have since shown a decline.
- The relatively stable catches of bigeye in the northern temperate region by longline vessels, and to a lesser degree, Japanese pole and line and purse seine vessels in region 1.
- The development of the equatorial purse-seine fisheries from the mid-1970s and 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 efficiency of the fleet.
- The steady increase in catch for the domestic fisheries of Indonesia and the Philippines since 1970.
- The apparent stabilisation of catches of bigeye for most gears after the mid 2000's.

4.5.2 Purse seine

In the previous assessment, purse seine catches by species were estimated using the procedure documented in Hampton and Williams (2017) as Method 3. SC15 agreed to a number of changes to the methodology (WCPFC, 2019), including the approach used to account for grab sample bias, based on recommendations arising from WCPFC Project 60. The revised estimation approach is documented in Peatman et al. (2019).

4.5.3 Longline fisheries

The longline CPUE indices are one of the most important inputs to the assessment as they provide indices of abundance over time for each region, and help scale abundance across the regions. These indices are implemented using the "index" fishery approach described in Section 4.4.1 and inform the model of trends and variation in the components of the stock vulnerable to fishing.

The index-fishery CPUE time series for the 2020 assessment were derived from the operational longline dataset for the WCPO region. This dataset is an amalgamation of operational level data from the distant-water fishing nations (DWFN) and Pacific-Island countries and territories (PICTs) longline fleets operating in the Pacific basin. It represents the most complete 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 WCPFC bigeye tuna stock assessment (McKechnie et al., 2017a; Vincent et al., 2018). In 2017, the first application of spatiotemporal modeling approaches was used for a WCPFC stock assessment (Tremblay-Boyer et al., 2017). For this assessment we built on these previous efforts including also the spatiotemporal

modeling done in support of the 2019 WCPFC skipjack tuna stock assessment (Ducharme-Barth et al., 2019; Kinoshita et al., 2019).

Two sets of indices were used in the stepwise model development of the 2020 bigeye tuna diagnostic case. The first index replicated the approach taken in the previous assessment (McKechnie et al., 2017a,b) and utilized a delta-GLM fit to independent partitions of data from the operational longline data-set within each assessment region. This model structure was not altered as this index was only used in the initial phases of model development in order to maintain continuity in the stepwise model progression from the 2017 diagnostic case to the 2020 diagnostic case. The primary index used in the 2020 stock assessment (Figure 6) was estimated using the VAST spatiotemporal modeling framework (Thorson, 2019).

A detailed description of the process and workflow for generating the spatiotemporal index is provided in Ducharme-Barth and Vincent (2020). Briefly, after the data-set was cleaned, the VAST approach was used to generate the relative abundance indices. In the VAST framework, the relative abundance index is the spatial average of predicted abundance once catchability effects have been "standardized" out. The model implemented by the VAST package is a spatiotemporal delta generalized linear mixed model (GLMM), an extension of the delta-GLM (Thorson et al., 2015). The final model applied to generate the CPUE indices included catchability covariates of a vessel nationality grouping variable and hooks-between-floats (HBF), which is a proxy for fishing depth and species targeting.

Previously, the coefficients of variation (CVs) for region-specific standardized effort were rescaled to an average level of 0.2 over a reference period of 1980-1990. Since a single spatiotemporal model was used to generate the standardized CPUE, the CVs for the standardized effort assigned to the index fishery were rescaled to an average level of 0.2 across all regions and time periods. This was done in order to more appropriately capture the relative differences in uncertainty, due in part to varying levels of sampling intensity, across spatial and temporal strata. In this way, MFCL is able to account for the time-varying nature of the CVs such that the fit to the CPUE data is given greater influence in the likelihood in time-steps with more precise estimates of abundance.

4.5.4 Other fisheries

There has been continual improvement in the catch estimates from Indonesia and the Philippines through the GEF-WPEA project and since the 2014 assessment catch data from fisheries in Vietnam have also been included. These improved catch estimates have been incorporated into the current assessment.

4.6 Size data

Available length-frequency data for each of the fisheries were compiled into 95 2cm size classes (10–12 cm to 198–200 cm). Weight data were compiled into 200 1kg size classes (0–1 kg to 199–200 kg).

Most weight data were recorded as processed weights (usually recorded to the nearest kilogram). Processing methods varied between fleets requiring the application of fishery-specific conversion factors to convert the available weight data to whole fish equivalents. Details of the conversion to whole weight are described in Langley et al. (2006). Data were either collected 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 2 provides details of the temporal availability of length and weight-frequency data. Note that a maximum effective sample size of 1,000 is implemented in MFCL when using the robust normal likelihood for size composition data. In practice, this effective sample size was further down-weighted, with the sensitivity to the magnitude of the down-weighting investigated in the sensitivity and structural uncertainty analyses.

4.6.1 Purse seine

Only length-frequency samples are used in the bigeye assessment for purse seine fisheries. Prior to 2014, 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 based on the Abascal et al. (2014) approach is described in further detail in Peatman et al. (2020).

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 detailed review of all available longline length and weight-frequency data for bigeye has previously been undertaken by McKechnie (2014). That paper, along with Peatman et al. (2020), provides details for the analytical approach used to construct the data inputs for the current assessment. The key principle used in constructing the data inputs was not to use weight and length data at the same time, even if it was available, as it would either introduce conflict (if data were in disagreement) or over-weight the model fit (if they were in agreement). The general approach used by McKechnie (2014) and Peatman et al. (2020) for the "extraction" fisheries was that weight-frequency samples should be weighted with respect to the spatial distribution of flag-specific catch within each region. This is done so that catch is extracted from the population at the appropriate size. Weight-frequency data were used based on the spatiotemporal coverage and number of samples. In the previous stock

assessment Japanese weight data were not available for regions 4, 5, and 6 towards the end of the model period and were supplemented by "all flags" length data. Additional Japanese weight-frequency data has since been made available and is used in the current assessment. However, the number of available weight-frequency samples declines in recent years. Switching to length-frequency data for the longline fisheries in regions 4, 5, and 6 beginning in the mid-2000's is investigated as a model sensitivity.

Size composition data were prepared similarly for the index fisheries (Peatman et al., 2020). As mentioned in Section 4.4.1, the approach differed from the one briefly described for the extraction fisheries in that the size-frequency samples were reweighted with respect to the spatial distribution of abundance as predicted by the spatiotemporal CPUE standardization model (Ducharme-Barth and Vincent, 2020). This is to allow size compositions to inform temporal variation in population abundance and size. To generate the size composition data for the index fisheries, data were first subset to match the nationalities of the "all flags" longline fisheries in each region. This was done to prevent shifts in size composition as a result of a change in sampling between fisheries. Sensitivity to this assumption was tested by including weight-frequency data from the US longline fleet and the Australian longline fleet in the index fishery composition data for regions 2 and 5, respectively. Additionally, sensitivity to the declining availability of weight composition data in regions 4, 5, and 6 was tested by switching to length-frequency data.

Given that the same data were used for both the extraction and index fisheries, the observed number of size-frequency samples input into the assessment was divided by 2 for both the extraction and index fisheries. The maximum effective sample size in the stock assessment model was also divided by two for these fisheries (i.e. 500 as opposed to the default value of 1,000 assumed for the other fisheries).

4.6.3 Other fisheries

Size composition data for the Philippines domestic fisheries, both small-fish fisheries (Fishery 17) and large-fish handline fisheries (Fishery 18), were derived from a number of sampling programs conducted in the Philippines since the 1980s. In more recent years, size-sampling data have been substantially augmented by the work of the GEF-WPEA project. Additionally, recent data collection efforts in both Indonesia and Vietnam have made available new length-frequency data for inclusion in the assessment 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 tuna and yellowfin tuna) 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. A summary of its characteristics and the process of constructing the MFCL tagging file are presented in detail by Vincent and Ducharme-Barth (2020) and in Table 3. 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), more recent (1995, 1999-2001) data from the Coral Sea Tagging Programme (CSTP) by CSIRO (Evans et al., 2008), and the Pacific Tuna Tagging Programme (PTTP) carried out during the period 2006 until the end of 2017. Additional data have been incorporated since the 2017 assessment for the Japanese Tagging Programme (JPTP), conducted by NRIFSF and the Ajinomoto Co. Inc, over the period 2000–2020. The tagging data from the JPTP provide information about the fishing mortality and movement rates in region 1 for these assessments. Additional tagging on a longline vessel near Tonga (*SM* cruise) that took place in the fourth quarter of 2000 was included because it had information from bigeye larger than 100 cm. There were 95 tags released with 4 usable recoveries. Though the previous assessment (McKechnie et al., 2017a) restricted the inclusion of the JPTP to a sensitivity analysis due to the slower model run times (MFCL has to track an extra "population" of individuals for each release event), these data were included in the 2020 diagnostic case as they provided tag/recapture information outside of the main equatorial regions. Sensitivity to the inclusion of these data is shown in the stepwise model progression to the diagnostic case.

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, and estimates of base levels of tag shedding/tag mortality. An additional problem 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, outside the assessment region. This adjustment or rescaling preserves the recovery rates of tags from individual tag groups that would otherwise be biased low given that a large proportion of recaptures cannot be attributed to a recapture category in the assessment. These procedures were first described in Berger et al. (2014) and McKechnie et al. (2016). For the current assessment, Vincent and Ducharme-Barth (2020) describes the analyses taken to prepare the tagging data. Additionally, the model used to adjust tags due to variability in tagger ability or "tagger effect" has been substantially changed to more appropriately account for covariates influencing bigeye tuna tag recapture rates. The report from Scutt Phillips et al. (2020) describes the changes to the "tagger effects" model in greater detail.

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 2017 were excluded from the assessment.

For incorporation into the assessment, tag releases were stratified by release region, time period of release (quarter) and the same size classes used to stratify the length-frequency data. A total of 36,847 effective releases were classified into 112 tag release groups (Vincent and Ducharme-Barth, 2020). The returns from each size-class of each tag release group (9,256 effective, usable tag returns in total) were then classified by recapture fishery and recapture time period (quarter). Because tag returns by purse seiners were often not accompanied by information concerning the set type, tag return data were aggregated across set types for the purse seine fisheries in each region. The population dynamics model was in turn configured to predict equivalent estimated tag recaptures by these grouped fisheries.

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, 2020). Tag reporting was assumed to be similar between the RTTP and CSTP so reporting rates estimates were shared across these two programs to reduce model dimensionality.

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; (iv) the observation models for the data; (v) the parameter estimation procedure; and (vi) stock assessment interpretations. Detailed technical descriptions of components (i)–(iv) are given in Hampton and Fournier (2001) and Kleiber et al. (2019). In addition, we describe the procedures followed for estimating the parameters of the model and the way in which stock assessment conclusions are drawn using a series of reference points.

5.2 Population dynamics

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

5.2.1 Recruitment

Recruitment is defined as the appearance of age-class 1 quarter fish (i.e. fish averaging $\sim$ 20–30 cm given current growth curves) in the population. Tropical tuna spawning does not always follow a clear seasonal pattern but is thought to occur sporadically when food supplies are plentiful (Itano, 2000). In the assessment model, it was assumed that recruitment occurs instantaneously at the beginning of each quarter. This is a discrete approximation to continuous recruitment, but provides sufficient flexibility to allow a range of variability to be incorporated into the estimates as appropriate.

Spatially-aggregated (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 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). The SRR was calculated over the period from 1962–mid-2017 to prevent the early recruitments (which appear to be part of a different "regime" to subsequent estimates and may not be well estimated in any case), and the terminal recruitments (which are not freely estimated), from influencing the relationship, which is consistent with the approach of the 2017 assessment.

In recent assessments of tuna in the WCPO the terminal recruitments have often been fixed at the mean recruitment of the rest of the model period. This acknowledges that these estimates are poorly supported by data and if unconstrained can vary widely, with potentially large consequences for stock projections. This approach has been continued here by fixing the six terminal recruitments at the mean of the recruitments over the rest of the assessment period.

The distribution of recruitment among the model regions was estimated within the model and allowed to vary over time in a relatively unconstrained fashion.

5.2.2 Initial population

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

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

5.2.3 Growth

The standard assumptions for WCPO assessments fitted in MFCL were made concerning age and growth: i) the lengths-at-age are normally distributed for each age-class; ii) the standard deviations of length for each age-class are a log-linear function of the mean lengths-at-age; and 3) the probability distributions of weights-at-age are a deterministic function of the lengths-at-age and a specified weight-length relationship[4]. These processes are assumed to be regionally and temporally invariant. The growth curves were defined using L1 (length at first age-class) and L2 (length at terminal age-class) calculated from the midpoint of each age bin (i.e. length at age 0.5 quarters for L1 and length at 39.5 quarters for L2) in order to more appropriately model the variability in growth for each age class.

Given the mandate for further investigation and research on growth by SC14, two new data-sets

[4]The length-weight relationship which has been updated for the current assessment based on an analysis of current and historical port sampling data (Vincent and Ducharme-Barth, 2020)

were prepared for the current bigeye tuna assessment: i) an improved otolith ageing data-set (Farley et al., 2020), and ii) an integrated otolith and tag-recapture data-set (Eveson et al., 2020). New for this assessment is the assumption of the more flexible Richards growth (Richards, 1959) as both the otolith only (*oto-only*) and the tag-integrated (*tag-int*) models showed a better fit to the data than for the traditional von Bertalanffy (VB) growth curve (Eveson et al., 2020; Farley et al., 2020). Three alternative approaches to modeling growth were investigated in the current assessment. All assumed the Richards growth function, and all met recommendations made by SC14 and the 2020 pre-assessment workshop:

- *Oto-Only:* This growth curve is a fixed Richards growth curve based on high-readability otoliths (Farley et al., 2020). This curve is derived from an expanded version of the combined daily and annual aged otolith data-set described in Farley et al. (2017) to define the "new growth" used in the previous stock assessment (McKechnie et al., 2017a; Vincent et al., 2018) and used to inform management (WCPFC, 2018). Changes from the previous analysis include an improved method for estimating decimal age using otolith size and the inclusion of very small bigeye that were captured by the Philippines domestic fisheries in region 7. The two length-at-age standard deviation (SD) parameters, the "generic" SD and the "age-dependent" SD (Kleiber et al., 2019) were freely estimated in the assessment model. Variation in asymptotic length for the otolith dataset is likely to be underestimated. The estimated age-dependent SD allows the model to fit the size data adequately, especially for large fish caught in the longline and handline fisheries.

- *Tag-Int:* This growth curve is a fixed Richards growth curve based on the same high-readability otolith data-set as the *oto-only* growth in addition to bigeye tuna tag-recapture data (Eveson et al., 2020). Additionally, the two length-at-age SD parameters were freely estimated to account for the large fish caught by the longline and handline fisheries.

- *Est-Richards:* A conditional age-length data-set was constructed from the combined daily and annual otolith dataset (Farley et al., 2020) and was input to MFCL. This allowed for modal progressions, apparent in the temperate region size composition data, in addition to the otoliths to inform the estimation of all growth parameters (L1, L2, K, two SD parameters, and the Richards parameter). This conditional age-length data-set makes use of the improved method for estimating decimal age as it allows for fish to be assigned to quarterly age classes without making a birth-date assumption. Additionally, though the small fish from region 7 were not collected until the first quarter of 2020, including these small fish is important for the estimation of the L1 parameter. As a result these fish were included in the assessment model as being caught in the first quarter 2018. The implicit assumption being made is that had the Philippines fishery been sampled for otoliths in 2018, these fish would have been present in the population and sampled at that time.

5.2.4 Movement

Movement was assumed to occur instantaneously at the beginning of each quarter via movement coefficients that connect regions sharing a common boundary. Note that fish can move between non-contiguous regions in a single time step due to the "implicit transition" computational algorithm employed (see Hampton and Fournier, 2001 and Kleiber et al., 2019 for details). Movement is parameterized by a pair of bi-directional coefficients at each region boundary. Movement is possible in both directions across each each regional boundary in each of the four quarters. Each of these coefficients is estimated independently resulting in 104 estimated movement parameters (2×no.region boundaries (13)×4 quarters). There are limited data from which to estimate long-term, annual variation in movement or age-specific movement rates. As such, the estimated seasonal pattern is assumed to be fixed across years and the movement coefficients are invariant with respect to age. A "prior" of 0 is assumed for all movement coefficients, and a low penalty is applied to deviations from the "prior".

5.2.5 Natural mortality

In MFCL, natural mortality (M) can either be held fixed at pre-determined, age-specific values or estimated as age-specific parameters. As in previous assessments, 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 externally-estimated M-at-age function was input to MFCL as fixed values and is shown in Figure 7 for the diagnostic case model. The generally increasing proportion of males observed in the catch with increasing size is assumed to be due to an increase in the natural mortality of females, associated with sexual maturity and the onset of reproduction. Since fixed values of M-at-age are initially calculated at-length and then back transformed to age using a growth curve, it is important to calculate a specific M-at-age for each growth curve used in the modeling. This process is described in McKechnie et al. (2017b) and is repeated for the alternate growth curves considered in this assessment. For this assessment we use a specific M-at-age for each of the fixed growth curve used. When growth was estimated internally to MFCL using the conditional age-length data-set, the natural mortality was assumed to follow that of the growth curve corresponding to the parameter starting values, either *Tag-Int* or *Oto-Only*.

Given the updates in bigeye tuna growth over the past several years (Eveson et al., 2020; Farley et al., 2017, 2020), the existing assumption around the mean level of M was revisited. A meta-analytic framework, based on life-history theory and empirical estimates, was used to generate an updated estimate for M (Vincent and Ducharme-Barth, 2020). This analysis yielded an estimated quarterly M of 0.1275 (0.1090 – 0.1459; 95% confidence interval) which is consistent with the assumed mean level of quarterly M used in the diagnostic case, 0.1124. Sensitivity analyses were conducted to explore how model estimated management quantities changed when the age-specific deviates from the diagnostic case were applied to the new estimates of average quarterly M.

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). Subsequent to the 2017 stock assessment, a new feature in MFCL was developed allowing the reproductive potential ogive to be defined and input into the assessment as a function of length. This length-based ogive is then converted internally to a reproductive potential-at-age using a smooth-spline approximation (Davies et al., 2019). This allows for a more natural definition of reproductive potential as the product of three length-based processes: proportion females-at-length[5] (sex-ratio), proportion of females mature-at-length[6], and the fecundity-at-length of mature females[7] (Figure 8). Another added benefit is that this reproductive potential ogive is growth invariant. The previous stock assessments had to redefine the reproductive potential-at-age ogive for each different growth curve included in the assessment.

In addition to the structural change in how reproductive potential is input and modeled, there are two updates to how the reproductive potential ogive was defined relative to the previous assessments. As mentioned in the previous paragraph, female sex-ratio at length is one of the three components that make up the reproductive potential ogive. In the past this was estimated from sex-at-length data collected by fisheries observers, and then converted to proportion female-at-age for input into MFCL. This length-to-age conversion resulted in a much greater decline in the proportion female-at-age than was indicated by the observer data. This exaggerated decline in proportion female-at-age used in the previous stock assessments resulted in a dome-shaped reproductive potential ogive (see Figure 11; McKechnie et al., 2017a) and implied that 5–7 year old individuals made the largest contribution to spawning potential. Using the more appropriate length-based input of female sex-ratio-at-length, the reproductive potential ogive is shifted to older individuals (Figure 9).

The second update to the reproductive potential ogive is the removal of spawning fraction. Previous assessments included spawning fraction at age from EPO yellowfin tuna (Schaefer, 1998) based on the sensitivity analysis of Hoyle and Nicol (2008). Given the differences between species and regional differences in the Pacific-basin between the WCPO and EPO, we felt it more appropriate to remove this component from the definition of the reproductive potential ogive. Although Farley et al. (2017) collected some information on WCPO bigeye tuna spawning fraction, we did not feel that sample sizes were sufficiently large (n = 168) to estimate spawning fraction across all length bins of spawning capable individuals. Model sensitivity to the removal of spawning fraction from the reproductive potential ogive is assessed in the stepwise model development.

[5]For the current assessment, female sex-ratio-at-length was calculated from Regional Observer Program data in SPC's holdings through 2018.

[6]Taken from Farley et al. (2017) as in the previous assessment.

[7]Taken from Sun et al. (2006) and standardized per kg of body weight at length as in the previous assessment.

5.3 Fishery dynamics

The interaction of the fisheries with the population occurs through fishing mortality. Fishing mortality is assumed to be a composite of several separable processes - selectivity, which describes the age-specific pattern of fishing mortality; catchability, which scales fishing effort to fishing mortality; and effort deviations, which are a random effect in the fishing effort - fishing mortality relationship.

5.3.1 Selectivity

Selectivity was modeled using a cubic spline (Harley et al., 2014; McKechnie et al., 2017a). This allows for greater flexibility than assuming a functional relationship with age (e.g. logistic curve to model monotonically increasing selectivity or double-normal to model fisheries that select neither the youngest nor oldest fish), and requires fewer estimated parameters than modeling selectivity with separate age-specific coefficients. This is a form of smoothing, but the number of parameters for each fishery is the number of cubic spline "nodes" that are deemed sufficient to characterize selectivity over the age range. We use five nodes, which appears sufficient to allow for reasonably complex selectivity patterns. Model sensitivity to increasing the number of nodes to 10 was investigated but not pursued further as it did not result in perceptible change to the estimated management quantities. Additionally, for the fisheries showing strong domed shaped selectivity patterns (purse seine, pole-and-line, and miscellaneous fisheries), selectivity was assumed to be fixed at zero beyond the length of the last observation in order to prevent spurious selectivity estimation by the cubic spline beyond the range of the observations.

In all cases, selectivity is assumed to be fishery-specific and time-invariant. However, where size composition data were insufficient to inform the proper shape of the selectivity curve, some fisheries were grouped with another fishery that exhibited similar operational characteristics or operated in a similar area (Table 1). This was the case for the longline fisheries (Fishery 27 Australian longline and Fishery 29 longline all) in region 9 which were each grouped with the respective longline fisheries in region 5 (Fishery 10 Australian longline and Fishery 11 longline all). Additionally, the unassociated purse seine in region 7 (Fishery 30) was grouped with the unassociated purse seine in region 3 (Fishery 13) and as mentioned previously, the Indonesia-Philippines domestic purse seine fishery in region 7 (Fishery 24) was grouped with the Philippines small-fish fishery in region 7 (Fishery 17).

While full length-based selectivity is not currently permitted in MFCL, the age-based selectivity functions are penalized such that selectivity of age-classes that are similar in size will have similar selectivities for a given fishery or group of fisheries. Additionally, the assumption was made that at least one fishery within each of the 9 spatial regions (excluding the index fisheries) would have selectivity that was penalized to be increasing as a function of length in order to prevent the accumulation of an invulnerable, cryptic biomass within the model. This was typically one of the longline fisheries, though in region 7 it was assumed to be the Indonesia-Philippines handline fishery (Fishery 18) as it caught the largest fish in the region (Table 1).

Lastly, selectivity for the index fisheries was assumed to be shared across regions and penalized to be increasing as a function of length.

5.3.2 Catchability

Constant (time-invariant) catchability was assumed for the index fisheries that received standardized indices of relative abundance. This assumption is similar to assuming that the CPUE for these fisheries indexes the exploitable abundance over time. These index fisheries were grouped for the purpose of initial catchability, and to maintain the relativity of catch rates among regions. This provides the model with information on the relative population sizes among regions, which previous experience suggests is difficult to estimate without this assumption.

Previous assessments assumed a very high penalty on the fit to the catch (SD of residuals on the log scale of 0.002) which is essentially equivalent to treating the catch as "error-free" as done in "catch-conditioned" modeling approaches. As a result, there was no need to estimate both effort and catchability deviates for the extraction fisheries. Pending the completed development of the fully catch-conditioned option in MFCL (Davies et al., 2020), a pseudo-catch conditioned approach was taken in this assessment for all non-index fisheries. All effort was removed (except for the nominal effort needed in the final 4 quarters to serve as a basis for projections), and partial fishing mortalities were estimated consistent with the observed catches by the function minimizer. Therefore, catchability deviations are not estimated for these fisheries. Only a single catchability value was estimated for each fishery to scale the nominal effort in the final year.

5.3.3 Effort deviations

Effort deviations were used to model the random variation in the effort - fishing mortality relationship, and are constrained by pre-specified "prior" distributions (on the log-scale). The region-specific CPUE indices implemented through the index fisheries represent the principal indices of stock abundance, and the extent to which the model can deviate from the indices is moderated by the penalty weights assigned to the standardized effort series. For these fisheries the "prior" was set to have a mean of zero and the CV was allowed to be time-variant and based on the variance estimates (using the canonical variance method of Francis, 1999) from the spatiotemporal standardization model (Ducharme-Barth and Vincent, 2020). As explained in Section 4.5.3, the regional differences in the estimated CVs from the spatiotemporal standardization model were preserved since they were generated from a unified analysis across regions. The CV across all index fisheries and time periods was rescaled to a value of 0.2. The resulting scaled CVs were transformed to an effort deviate penalty for each CPUE observation in MFCL. Penalties are inversely related to variance, such that lower effort penalties are associated with indices having high variance, consequently these indices are less influential in fitting the model. As seen in Figure 10 more influence is given to the CPUE from regions well sampled by the collective longline fisheries (regions 1 – 6 particularly later in the time series), and comparatively little weight is given to regions less well

sampled by longlines (regions 7 & 8). Region 9 is quite small and sporadically sampled, resulting in the lowest relative penalty weight to the fit to the CV.

Finally, for all the extraction fisheries (Table 1) with nominal effort in the final year, the effort deviates were given penalties equivalent to a CV of approximately 0.7 to prevent the CPUE of these fisheries from influencing population dynamics in the final year.

5.4 Dynamics of tagged fish

5.4.1 Tag reporting

In principle, tag-reporting rates can be estimated internally within the model. In practice, experience has shown that independent information on tag-reporting rates for at least some fisheries tends to be required for reasonable model behavior to be obtained. In addition to varying by fishery, we allowed reporting rates to also vary among tagging "programs" implemented at different times in the history of the fishery, or conducted by different agencies. We provided reporting rate "priors" for all fishery/tagging program groups that reflect independent estimates of the reporting rates and their variances. These were derived from analyses of tag seeding experiments (Peatman, 2020). For the RTTP and PTTP, relatively informative "priors" were formulated for the equatorial purse seine fisheries given that tag seeding experiments were focused on purse seiners. All reporting rates were assumed to be time-invariant, and have an upper bound of 0.9. Tag recapture and reporting rate groupings are provided in Table 1 & Table 2.

5.4.2 Tag mixing

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

Depending on the distribution of fishing effort in relation to tag release sites, the probability of capture of tagged fish soon after release may be different to that for the untagged fish in that model region. It is therefore desirable to designate one or more time periods (quarters) after release as "pre-mixed" and compute fishing mortality for the tagged fish based on the actual recaptures, corrected for tag reporting, rather than use fishing mortalities based on the general population parameters. This in effect desensitizes the likelihood function to tag recaptures in the pre-mixed

periods while correctly removing fish from the tagged population for the recaptures that occurred. We assume that tagged bigeye gradually mix with the untagged population at the region-level and that this mixing process is complete by the end of the second quarter after release. We investigate the robustness to this assumption in sensitivity analyses.

The tag return files were created using a sliding window to calculate the mixing period for each release group per the recommendation of the 2020 ePAW (Hamer and Pilling, 2020). This was done to ensure that tags had 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. Tags that were recaptured within these time frames were assigned to the quarter of release if recaptured prior to 91 days at liberty and to the quarter after release for time at liberty between 92 and 182 days. This is an added step taken to ensure that tags that had not fully mixed with the entire population did not influence the estimation of fishing mortality.

Tagged fish are modeled as discrete cohorts based on the region, year, quarter and age at release for the first 30 quarters after release. Subsequently, the tagged fish are pooled into a common group. This is to limit memory and computational requirements.

5.5 Likelihood components

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

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

The probability distributions for the length- and weight-frequency proportions are assumed to be approximated by robust normal distributions, with the variance determined by the effective sample size (ESS) and the observed length-frequency proportion. Size frequency samples are assigned ESS lower than the number of fish measured. Lower ESS recognize that (i) length- and weight-frequency samples are not truly random (because of non-independence in the population with respect to size) and would have higher variance as a result; and (ii) the model does not include all possible process error, resulting in further under-estimation of variances. The observed sample sizes are capped at 1,000 internal to MFCL, and these were further divided by 20, resulting in a maximum ESS of 50 for each length and weight sample for a fishery. Alternative divisors for specifying ESS were explored in sensitivity analyses.

We also examined in a sensitivity analysis the application of a new size-likelihood based on the Self-scaling Multinomial plus Random Effects (SSMULT-RE) This method improves upon the traditional Multinomial likelihood because it i) does not assume that the variance is proportional

to the sample size ii) accounts for the positive correlation in residuals through using auto-correlated random effects (Davies et al., 2019), and iii) is self-scaling. This allows for the estimation of ESS from the data, and removes the need for arbitrary assumptions about ESS. As this is a new feature of MFCL (Davies et al., 2020), it was not used as the baseline likelihood structure in the diagnostic case or structural uncertainty grid. Rather, it was investigated as one-sensitivity to the diagnostics. Further investigation into this promising new method should be conducted.

A log-likelihood component for the tag data was computed using a negative binomial distribution. The negative binomial is preferred over the more commonly used Poisson distribution because tagging data often exhibit more variability than can be attributed by the Poisson. We employed a parameterization of the overdispersion parameter (τ) such that as it approaches 1, the negative binomial approaches the Poisson. In the current assessment we assume $\tau = 2$, which is a variance twice that of the Poisson. Therefore, if the tag return data show high variability (for example, due to contagion or non-independence of tags), then the negative binomial is able to recognize this. This should then provide a more realistic weighting of the tag return data in the overall log-likelihood and allow the variability to impact the confidence intervals of estimated parameters. A complete derivation and description of the negative binomial likelihood function for tagging data is provided in Kleiber et al. (2019).

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

5.6 Parameter estimation and uncertainty

The parameters of the model were estimated by maximizing the log-likelihood of all data components plus the log of the probability density functions of the "priors" and penalties specified in the model. The maximization to a point of model convergence was performed by an efficient optimization using exact derivatives with respect to the model parameters (auto-differentiation, Fournier et al., 2012). Estimation was conducted in a series of phases, the first of which used relatively arbitrary starting values for most parameters. A bash shell script, "doitall", implements the phased procedure for fitting the model. Some parameters were assigned specified starting values consistent with available biological information.

In this assessment two approaches were used to describe the uncertainty in key model outputs. The first estimates the statistical uncertainty within a given assessment model, while the second

focuses on the structural uncertainty in the assessment by considering the variation among a suite of models.

For the first approach, the Hessian was calculated for the diagnostic case model run to obtain estimates of the covariance matrix, which is used in combination with the delta method to compute approximate confidence intervals for parameters of interest (for example, the biomass and recruitment trajectories). For the second approach, a factorial grid of model runs was undertaken which incorporated many of the options of uncertainty explored in one-off sensitivity analyses. This procedure attempts to describe the main sources of structural and data uncertainty in the assessment.

For highly complex population models fitted to large amounts of often conflicting data, it is common for there to be difficulties in estimating absolute abundance. As in the previous assessment, a likelihood profile analysis was conducted for the marginal posterior likelihood in respect of the total average population biomass as a measure of population scaling ((Lee et al., 2014), with the definition of this parameter detailed in (Kleiber et al., 2019)). Rationale for profiling the likelihood with respect to total average population biomass instead of the total population scaling parameter, along with a description of how this profile is generated can be found in Section 5.6 of the previous assessment report (McKechnie et al., 2017a). Reasonable contrast in the profile obtained using this method is taken to indicate that sufficient information existed in the data for estimating absolute abundance, and also offered confirmation that the maximum likelihood estimate obtained represented a global solution, at least with respect to total population scaling. This procedure is presented in the Appendix (Section 11.2), including examination of the profiles for the individual data components.

Retrospective analyses are also undertaken as a general test of the stability of the model, as a robust model should produce similar output when rerun with data for the terminal year/s sequentially excluded (Cadigan and Farrell, 2005). The retrospective analyses for the 2020 diagnostic case model are presented in the Appendix (Section 11.3).

Jitter analyses were performed on the diagnostic model as a check for model stability and to ensure convergence to a global solution (Cass-Calay et al., 2014). Jitter analysis entails randomly perturbing parameter estimated from the diagnostic model and refitting the model to the data. This "jittering" can be conducted multiple times with differing levels of perturbations to determine how well behaved the model is. This is the first time these analyses have been conducted on MULTIFAN-CL models and requires further development and investigation. The results of the jitter analyses are presented in the Appendix (Section 11.4).

5.7 Stock assessment interpretation methods

Several ancillary analyses using the fitted model/suite of models were conducted in order to interpret the results for stock assessment purposes. The methods involved are summarized below and further

details can be found in Kleiber et al. (2019).

5.7.1 Yield analysis

The yield analysis consists of computing equilibrium catch (or yield) and biomass, conditional on a specified basal level of age-specific fishing mortality (F_a) for the entire model domain, a series of fishing mortality multipliers (*fmult*), the natural mortality-at-age (M_a), the mean weight-at-age (w_a) and the SRR parameters. All of these parameters, apart from *fmult*, which is arbitrarily specified over a range of 0–50 (in increments of 0.1), are available from the parameter estimates of the model. The maximum yield with respect to *fmult* can easily be determined using the formulae given in Kleiber et al. (2019), and is equivalent to the MSY. Similarly the spawning potential at MSY (SB_{MSY}) can also be determined. The ratios of the current (or recent average) levels of fishing mortality and biomass to their respective levels at MSY are determined for all models of interest, including those in the structural uncertainty grid, and so alternative values of steepness were assumed for the SRR in many of them.

Fishing mortality-at-age (F_a) for the yield analysis was determined as the mean over a recent period of time (2013–2017). We do not include 2018 in the average as fishing mortality tends to have high uncertainty for the terminal data year of the analysis and the terminal recruitments in this year are constrained to be the average over the full time-series, which affects F for the youngest age-classes.

MSY was also computed using the average annual F_a from each year included in the model (1952–2018). 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. More details of this approach are provided in Section 5.7.4.

5.7.2 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.

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

SRR between the estimated fished and unfished spawning potential estimates.

A similar approach can be used to estimate depletion associated with specific fisheries or groups of fisheries. Here, fishery groups of interest - 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.

5.7.3 Reference points

The unfished spawning potential ($SB_{F=0}$) in each time period was calculated given the estimated recruitments and the Beverton-Holt SRR as outlined in Section 5.7.2. 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 $SB_{F=0}$ is calculated as the average over the period 2008–2017. Stock status was referenced against these points by calculating $SB_{recent}/SB_{F=0}$ and $SB_{latest}/SB_{F=0}$, where SB_{latest} and SB_{recent} are the estimated spawning potential in 2018, and the mean over 2015–2018, respectively (Table 4).

The other key reference point, $F_{recent}/F_{\mathrm{MSY}}$ (Table 4), is the estimated average fishing mortality over the full assessment area over a recent period of time (F_{recent}; 2014–2017 for this stock assessment) divided by the fishing mortality producing MSY which is produced by the yield analysis and has been detailed in Section 5.7.1.

5.7.4 Majuro and Kobe plots

For the standard yield analysis (Section 5.7.1), the fishing mortality-at-age, F_a, is determined as the average over some recent period of time (2014–2017 herein). In addition to this approach the MSY-based reference points (F_t/F_{MSY}, and SB_t/SB_{MSY}) and the depletion-based reference point ($SB_t/SB_{F=0[t]}$) were also computed using the average annual F_a from each year included in the model (1952–2017, with no value calculated for the terminal year) by repeating the yield analysis for each year in turn. This enabled temporal trends in the reference point variables to be estimated taking account of the differences in MSY levels under varying historical patterns of age-specific exploitation. This analysis is presented in the form of dynamic Kobe plots and "Majuro plots", which have been presented for all WCPO stock assessments in recent years.

6 Model runs

6.1 Developments from the last assessment

Model development from the 2017 diagnostic case to the 2020 diagnostic case occurred incrementally via successive changes. These stepwise changes were done in order to identify the impact of each modification to the assessment outcomes. Changes made to the previous assessment model include:

additional input data for the years 2016-2018, use of a spatiotemporal standardization model to prepare the CPUE data (Ducharme-Barth and Vincent, 2020), modified methods in producing the input files (Peatman, 2020; Peatman et al., 2020; Scutt Phillips et al., 2020; Vincent and Ducharme-Barth, 2020), updated biological information (Eveson et al., 2020; Farley et al., 2020; Vincent and Ducharme-Barth, 2020), and incorporation of new features to MFCL developed since the last assessment (Davies et al., 2020). These changes occurred in the following sequence of steps:

1. The 2017 diagnostic case model [*Diag17*]
2. The 2017 diagnostic case model run with the updated MFCL executable, v2.0.7.0 [*Diag17: New MFCL*]
3. A complete update of the 2017 diagnostic case model - all inputs extended from 2015 to 2018 using identical methodology for CPUE, tagging, size frequencies etc, and the same MFCL model settings [*Diag17: Data Update*].
4. The previous model with the tagging data adjusted using the new tagger effects model described in Scutt Phillips et al. (2020) [*New tagger effects*].
5. The previous model with the purse seine catch estimates based on the methodology documented in Peatman et al. (2019) [*Peatman catch*].
6. The previous model with the CPUE coming from the spatiotemporal model described in Ducharme-Barth and Vincent (2020) [*Geostats CPUE*].
7. The previous model with the size composition data prepared according to Peatman et al. (2020) [*Peatman comp*].
8. The previous model with the switch to the index fisheries approach. This also included removing the effort for the extraction fisheries [*Index fisheries*].
9. The previous model with the tag reporting rate priors updated following the analysis of Peatman (2020) [*New RR*].
10. The previous model with the mixing period defined using a sliding window for each release group based on the recommendations of the 2020 ePAW [*Force Mix*].
11. The previous model but including tags captured within the mixing period but were not included previously due to incomplete recapture location. Including these tags does not directly impact the estimation of fishing mortality since they were captured within the mixing period, however they maximize the information provided by the tagging data by reducing the tag usability correction factor and providing additional information for estimating the tag reporting rates [*Mix2*].
12. The previous model with the inclusion of tags released as a part of the Japanese Tagging Program, and a small number of large bigeye tuna released as a part of the central Pacific

SM longline tagging cruise [*JPTPSM*].

13. The previous model with and updated length-weight relationship estimated from port sampling data [*New L-W*].
14. The previous model with the reproductive potential ogive defined as a function of length [*Repro@Length*].
15. The previous model with the spawning fraction removed from the reproductive potential at length ogive [*Remove SF*].
16. The previous model with the growth defined according to the *Tag-Int* model described in (Eveson et al., 2020) [*Tag-Int*].
17. The previous model with the selectivity groupings defined in Table 1 and the assumption of increasing selectivity as a function of age applied to one fishery in each region [*Diagnostic*].

6.2 Sensitivity analyses

During the course of model development for the 2020 bigeye tuna stock assessment, several hundred models were run to explore the effects of changing the assumptions governing the population dynamics, fisheries dynamics, parameter estimation and weighting of data components in the likelihood. The presentation of the results focuses on the subset of analyses that were most influential to the stock assessment outcomes and/or those that were identified as questions of interest either in the previous assessment or the 2020 ePAW (Hamer and Pilling, 2020). From this subset and based on the findings of previous WCPO stock assessments, models were selected for inclusion in the structural uncertainty analyses. This process entailed running a full-factorial set of models where all combinations of key structural assumptions are explored (Section 6.3).

One-off sensitivities were conducted as single stepwise changes from the 2020 diagnostic case. The purpose of these sensitivity runs was not to provide absolute estimates of management quantities but to assess the *relative* change that resulted from the various assumptions, and for that reason the reference points are presented in the Appendix (Section 11.7).

6.2.1 Relative weighting of length and weight frequency data [*Size10, Size60, Size200, Size500*]

The difficulties in assigning weighting to the size-frequency data were discussed in Section 5.5. To assess the sensitivity of model results to the weighting of these data, four alternative models were considered; a model where frequency data were up-weighted (corresponding to a maximum effective sample size of 100 fish) relative to the diagnostic case model (*Size10*); and several where the size-frequency data were increasingly down-weighted to maximum effective sample sizes of 16.5 (*Size60*), 5 (*Size200*), and 2 (*Size500*). Note that in each case these maximum effective sample sizes are further reduced by a factor of two for the longline all, and index fisheries.

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

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

6.2.3 Alternative growth functions [*EstRichardsT, EstRichardsO, Oto-Only*]

Two methods additional to the diagnostic case being implemented as sensitivity models. The first set of sensitivities estimated growth by including the conditional age-length data, constructed from daily and annual otolith readings, within the model-fitting procedure of MFCL. This data set therefore contributes a likelihood component to the model. The previous bigeye tuna assessment (Vincent et al., 2018) noted starting point dependency issues with the estimation of growth parameters. To account for this, two models were run, with growth parameters starting equal to those from the *Tag-Integrated* (*EstRichardsT*) or *Oto-Only* (*EstRichardsO*) Richards growth curves. The second sensitivity considered the *Oto-Only* growth model described by (Farley et al., 2020). This growth curve is similar to the one assumed in the diagnostic case as they are estimated from the same set of otolith data.

6.2.4 Steepness [*h0.65, h0.95*]

Steepness is a particularly difficult parameter to estimate in stock assessment models, but if it is fixed in the model, the choice of value may have significant influence on most reference points used in management. As was the case in other tropical tuna and albacore tuna assessments, we assumed a value of 0.8 for the diagnostic case, but examined values of 0.65 (*h0.65*) and 0.95 (*h0.95*) in sensitivity runs. This choice of values is consistent with the results of the meta-analysis conducted on tuna stock-recruitment data and has been well established in previous Scientific Committees.

6.2.5 Tag data [*Mix1, TagFree*]

The tag mixing period is imposed to allow tagged fish to distribute themselves throughout the region of tagging, although it is somewhat difficult to ascertain how long this period should be. In the diagnostic case model the mixing period was set at two quarters and an alternative model was run assuming a mixing period of one quarter (*Mix1*). Additionally, in the stepwise model development it was noted that the model was sensitive to the assumptions made in preparing the tagging data. An alternative model (*TagFree*) was developed to test the effect of removing all tagging data from the model.

6.2.6 Natural mortality [*M-low, M-mid, M-hi*]

The new growth information provided an opportunity to revisit the assumption made on the mean level of M using a meta-analytic approach. Model sensitivity to changing the assumed mean level of M was explored by applying the age-specific deviates of to the mean level of M from the meta-analysis (*M-mid*), and the levels corresponding to the lower (*M-low*) and upper (*M-hi*) 95% confidence intervals.

6.2.7 Selectivity groupings and structure [*Alt, FreeSel*]

There are a number of different assumptions that can be made when defining the fisheries selectivity curves. We investigated the robustness of the model to different fisheries groupings (*Alt*), and removing the assumption of having a selectivity be non-decreasing as a function of age for at least one extraction fishery in each region (*FreeSel*).

6.2.8 Model start year [*Model1962, CPUE1962*]

One of the features of the model seen in the previous assessment was a high spike in recruitment paired with a rapid decline in biomass within the first 10 years of the model despite relatively low levels of catch. This indicates that, at least in the early years, the CPUE may be hyper-depleted given the level of catch that was removed from the system. Two alternative models were investigated to test the robustness of the model to including this early period of CPUE: beginning the model in 1962 (*Model1962*) or removing the first 10 years of CPUE data from the model (*CPUE1962*).

6.2.9 Size-frequency data [*idxAU-US, len456*]

The composition data for the index fisheries was restricted to weight-frequency data coming from fishing nations defined for the longline ALL fisheries. This assumption ignores the large amount of additional weight-frequency data available for the region 2 index fishery and the region 5 index fishery available from the US and Australian longlines. Model sensitivity to this assumption was explored in an additional model (*idxAU-US*) which included the US longline weight-frequency data for the region 2 index fishery and the Australian weight-frequency data for the region 5 index fishery. The weight frequency data from these two fisheries (Fishery 3 & 10) was correspondingly down-weighted to account for its inclusion as a part of the index fisheries weight-frequency data. Additionally, it was also noted in Section 4.6.2 that the availability of weight-frequency data available for the longline ALL and index fisheries in regions 4–6 declined beginning in the mid-2000s. An additional one-off sensitivity (*len456*) was conducted to investigate the effects of switching to length-frequency data for these fisheries in the mid-2000s.

6.3 Structural uncertainty

The structural uncertainty grid for the 2020 assessment was constructed from 3 axes – steepness (3 levels), natural mortality M (2), and size data weighting (4), with the settings used directly

comparable to those presented in Section 6.2 through identical notation. The final grid consisted of 24 models (Table 5). The rationale for the selection of these axes and levels is discussed in the following Sections 7.6 & 8.3.

7 Results

7.1 Consequences of key model developments

The progression of model development from the 2017 diagnostic case model to the 2020 diagnostic case model is outlined in Section 6.1 and results are displayed in Figure 11 with respect to both the change in spawning potential (SB) and the level of depletion of spawning potential relative to the unfished condition ($SB_{F=0}$). A summary of the consequences of this progression through the models is as follows:

1. *Diag17: New MFCL* (**Step 2**; Figure 11) – Running the 2017 diagnostic case with the new MFCL executable produced virtually identical results in terminal estimates of spawning potential and depletion.

2. *Diag17: Data Update* (**Step 3**; Figure 11) – The model with fully updated datasets (catch, effort, size frequencies, tagging) increased initial estimates of spawning potential, although temporal changes and terminal estimates of both model outputs spawning potential and depletion remained similar to the previous model.

3. *New tagger effects* (**Step 4**; Figure 11) – Updating the tagger effects model resulted in a large upwards shift in spawning potential, resulting in a more optimistic level of depletion. The new tagger effects model described in Scutt Phillips et al. (2020) increased the number of effective releases relative to the previous model. Given that the number of recaptures was unchanged, this resulted in the model estimating a lower fishing mortality and higher level of biomass.

4. *Peatman catch* (**Step 5**; Figure 11) – Updating the purse seine catch estimates according to the method described in Peatman et al. (2019) resulted in minimal changes to estimated quantities relative to the previous step.

5. *Geostats CPUE* (**Step 6**; Figure 11) – Updating the CPUE using the spatiotemporal or "geostats" model described in Ducharme-Barth and Vincent (2020) resulted in an increase in the initial estimate of spawning potential relative to the previous model. However, the overall temporal trend of spawning potential and estimates of depletion up to the final two years were similar. Switching to the spatiotemporal CPUE resulted in a marginally more pessimistic depletion in the terminal years of the model.

6. *Peatman comp* (**Step 7**; Figure 11) – Moving to size composition data prepared according to Peatman et al. (2020) resulted in progressively lower estimates of spawning potential relative

to the previous model beginning in 1960. This translated to more pessimistic estimates of depletion.

7. *Index fisheries* (**Step 8**; Figure 11) – Making the switch to the index fisheries structure resulted in both spawning potential and depletion estimates that reverted back to similar levels as those estimated in the *Geostats CPUE* step of model development.

8. *New RR* (**Step 9**; Figure 11) – Updating the reporting rates from Peatman (2020) resulted in a marginally lower but more informative "priors" for the purse seine fisheries. This step also assigned uninformative priors to the domestic fisheries in region 7. These changes resulted in fewer reporting rates being estimated at the upper bound of 90%, which has been a persistent problem in past bigeye tuna stock assessments. When reporting rates are lower and not on the upper bound, the same number of recaptures translates to a higher estimate of fishing mortality resulting in lower spawning potential and more pessimistic depletion estimates. This is seen to occur as the estimated quantities return to the level seen in the *Peatman comp* step.

9. *Force Mix* (**Step 10**; Figure 11) – Changing the mixing period definition to a sliding window for each tag release group resulted in marginally higher estimates of spawning potential and a marginally less depleted stock status relative to the previous step.

10. *Mix2* (**Step 11**; Figure 11) – Increasing the number of tags included in the model had the ultimate effect of reducing the usability correction factor, and thus increased the effective number of releases. As seen in the *New tagger effects* step, this results in an upwards rescaling of the biomass and more optimistic stock status given that the level of removals is unchanged relative to the previous step.

11. *JPTPSM* (**Step 12**; Figure 11) – Including tags from the Japanese Tagging Program and the *SM* central Pacific cruise resulted in slightly lower levels of spawning potential, and slightly more pessimistic depletion relative to the previous step.

12. *New L-W* (**Step 13**; Figure 11) – Updating the length-weight relationship from the parameters used in the previous assessment (originally from Morita, 1973), did not change estimates of depletion though it served to slightly scale down the estimate of spawning potential relative to the previous step.

13. *Repro@Length* (**Step 14**; Figure 11) – Switching to a length based parameterization of the reproductive potential ogive, with a greater contribution from larger, older fish, expectedly resulted in a lower estimate of spawning potential and more pessimistic estimate of depletion relative to the previous step.

14. *Remove SF* (**Step 15**; Figure 11) – Removing spawning fraction from the definition of the reproductive potential ogive had a negligible effect on both estimates of spawning potential and depletion.

15. *Tag-Int* (**Step 16**; Figure 11) – Updating the growth to the fixed tag-integrated curve described in Eveson et al. (2020) resulted in a substantial downwards shift in spawning potential and a much more pessimistic estimate of depletion. Relative to the previous growth curve, length-at-age is higher at all age-classes (for the youngest ages ∼9 cm and ∼3 cm for the oldest). Given the higher L_∞ it is expected that outcomes are more pessimistic (Vincent et al., 2018). However, there are two additional factors that contribute to this change. With the larger size at age from the new growth, younger ages are estimated to be more vulnerable to the different longline and index fisheries. In the case of the extraction fisheries this translates to a higher fishing mortality experienced by younger age-classes. For the index fisheries, the decline in CPUE is attributed to a broader range of age-classes rather than just the very oldest individuals.

16. *Diagnostic* (**Step 17**; Figure 11) – In the process of updating the growth curve in the previous step, it was observed that all of the longline extraction fisheries (except for Fishery 6 which the previous assessment penalized to be non-decreasing) were estimated to have a descending right-hand limb in their selectivity curve. The flexibility of the cubic spline explains the lack of observations of the largest individuals as a function of them transitioning into a less vulnerable state which is unexpected given our understanding of longline selectivity patterns. Though it could be possible that there is a behavioral component affecting selectivity at age (i.e. largest fish showing a preference for deeper depths where they are less vulnerable to longline gears), McKechnie et al. (2017a) noted that this was unlikely to be the case across all longline fisheries, especially those operating in regions where the thermocline is relatively shallow. To account for this, the assumption was made that the longline extraction fishery catching the largest individuals in each region would have a non-decreasing selectivity penalty applied to it. When this assumption is made, it expectedly results in lower estimates of spawning potential and a more depleted stock status relative to the previous step.

Relative to the diagnostic case of the previous assessment in 2017, the diagnostic case for the current assessment estimates the WCPO bigeye tuna stock to have a lower level of spawning potential and a slightly more pessimistic stock status relative to depletion from the unfished condition.

7.2 Model fit for the diagnostic case model

This section discusses the diagnostics for the diagnostic case model, defined by the final step in the stepwise model development (model *Diagnostic*; Figure 11), the final step described in Section 7.1.

7.2.1 Catch data

High penalties were applied to the catch data for all fisheries and so the catch is fit very closely, with a SD of residuals on the log scale of 0.002 (Figure 12).

7.2.2 Standardized CPUE

There was substantial temporal variability in the standardized CPUE indices used in the assessment, but despite this, the model-predicted CPUE fitted the indices very well (Figure 13). In general, the model captures most of the seasonal variation in the more temperate regions and the more stable dynamics in the equatorial regions. The pulse of high CPUE that appeared in the final quarters of the previous assessment has since moved through the data. The model is able to capture this slight downturn in CPUE since 2015 in most of the spatial regions.

There are some minor lack of fit issues in several of the spatial regions, however this can be attributed either to low penalty (high CV; Figure 10) being applied to the fit to these data and/or extreme variability in the standardized index due to limited spatial sampling. In the case of regions 2 & 9, it is a combination of the two effects. The central tendency of the trend in region 9 appears to be captured through to the late-1990s, though the seasonal variability and a portion of the early-2000s is not fit well. This is not unexpected given the very low weight placed on these data. However, given the small amount of biomass estimated in this spatial region relative to the others, this is not believed to strongly influence the model outcomes. Similarly, the model captures the central tendency of the CPUE in region 2 throughout the time series but fails to capture the high seasonality seen prior to 1980. Though region 2 has a moderately high penalty weight, the period prior to 1980 had a comparatively lower penalty weight which allowed the model to deviate from these observations. Additionally, despite the higher penalty weights placed on the CPUE in region 4, the model is unable to account for the pulse in CPUE just prior to the year 2000. Since this coincides with an instance of relatively lower penalty weights, the model is able to "smooth"-through this more uncertain data. Lastly, there appears to be a persistent overestimation of the CPUE in region 7 beginning in the 1990s. This pattern is most apparent in the time series of effort deviates (Figure 14) which are consistently estimated to be negative for this region and this instance, and reflects the comparatively more uncertain estimates coming from this region due to limited spatial sampling. However, this could also be an indication of model mis-specification as the lower observed CPUE are being fit by lower catchability through the effort deviates rather than by estimating a lower level of biomass.

7.2.3 Size frequency data

There was a reasonable fit to the length frequency data for fisheries with adequate sample sizes (Figure 15), particularly the important PS fisheries and the domestic fisheries for the Philippines and Vietnam in region 7 (Dom.PH.7 & Dom.VN.7 in Figure 15). The exception to this is the unassociated purse seine fishery in region 4 (SU-ALL-4) which shows a lack of fit relative to the other purse seines and the previous assessment, despite being estimated independently. This is unlikely to be strongly influential due to the small amount of catch by this fishery. There is also some modality in the observed frequencies for some of these fisheries which could not be completely fit by the model and presumably relates to the diverse sources of some of these data. Several fisheries

display some lack of fit but in most cases this reflects a lack of data over most of the time-series (e.g. PL ALL 8). These are also often low volume fisheries and not likely to be influential on the assessment outcomes. Additionally, it is important to note that with a length at first quarter (L1) of 31 cm assumed by the current growth curve and structural limitations on the length-dependent component of variability around the growth curve, the model is unable to fit the smallest individuals captured in the domestic Philippines fishery in region 7.

The weight frequency data for the longline fisheries were generally well fitted by the model (Figure 16), although there is a slight overestimation of large fish in some fisheries (e.g. the LL fisheries in regions 1 and 8). As was the case in the previous assessments, not all modality in the observed data was predicted by the model (e.g. L.ALL.1) and this may be due to the difficulty of attributing size data to different fishing operations (distant water vs offshore).

These patterns were largely reflected in the comparison of temporal variation in observed and model-predicted median fish lengths (Figure 17) and weights (Figure 18). For fisheries with a high sample size, and good temporal coverage the model is able to track the patterns seen in the observed median weight. In general the model appears to show better temporal fit than the previous assessment, and the temporal variation in the weight frequencies was again better fitted than the lengths. However, some temporal lack of fit can be seen in several fisheries, particularly for the Indonesia-Philippines handline fishery in region 7 (Fishery 18), where there is apparent conflict between the lengths and weights during the middle of the past decade. The weights appear consistent with the lengths collected for the rest of the time series, so it would appear that the lower median lengths observed arise due to mis-identification of the capture gear since this is part of a mixed-gear fishery. Continued improvement in identifying the data sources in these mixed gear fisheries would be beneficial for the assessment.

7.2.4 Tagging data

In the aggregate sense the model appears to fit the tagging data fairly well (Figure 19), though there is an indication that at longer times at liberty the model is under-estimating these "rarer" recovery events (Figure 19; c). Recent developments in MFCL to develop a censored-gamma tagging likelihood function could aid in the fitting to these observations (Davies et al., 2020). Breaking down model fit by both tag program and release region (Figure 20) we see that the overall model performance is buoyed by the reasonable fits to the regions (1, 3, 4, 7, & 8) and programs (particularly the PTTP) with higher numbers of tag returns. Consistent with previous assessments, the model appears to struggle with the high temporal variability in region 9 tag returns in the Coral Sea. However, model sensitivities in the previous assessment (McKechnie et al., 2017a) indicated that this is unlikely to have an impact on model outputs.

7.3 Model parameter estimates (diagnostic case)

7.3.1 Selectivity

A very diverse set of fisheries are included in the assessment model and this is reflected in the form of the estimated selectivity functions (Figure 21). Most longline fisheries had a non-decreasing penalty applied (Fisheries 1, 3, 5, 6, 8, 9, 10, 12, 27, and 33 – 41) and were estimated to have selectivity curves that were close to asymptotic. However, even those without the penalty also estimated selectivities where the oldest fish were fully selected. The exception was the longline ALL fishery in region 3 (Fishery 4) which caught smaller fish than the offshore longline in that region (Fishery 6), and estimated a descending-limb to accommodate this difference. The large fish handline fishery in region 7 also showed longline-like selectivity with very few fish younger than about 15 quarters of age being estimated to be caught by this fishery.

The purse seine, pole-and-line and miscellaneous small fish fisheries in region 7 displayed a variety of relatively complex selectivity functions (Figure 21), although in general these fisheries tend to catch very few fish over about 20 quarters of age, and some (particularly the Philippines and Indonesia miscellaneous fisheries) mainly catch fish of the youngest few age classes.

7.3.2 Movement

Figure 22 portrays the origin (which region the fish were originally recruited into) of the equilibrium bigeye tuna biomass in each region. From this figure it is evident that the model estimates significant movements of fish between the majority of regions. In general, there is a tendency for the temperate regions (1, 2, 5, & 6) to receive most of their fish as local recruits, however in the tropical regions (3, 4, 7, & 8) a significant proportion of fish ending up in a region were actually derived from recruitment to nearby tropical regions first (4 or 7) and then moved to their final region via immigration. Due to the flexibility of MFCL with regards to assigning recruits to regions and then moving them to other regions to fit data components such as size frequencies and CPUE indices, some caution must be exercised when interpreting the movement rates.

7.3.3 Tag Reporting Rates

The estimated tag reporting rates by fishery recapture groups (see groupings in Table 1) are displayed in Figure 23. As expected, the reporting rate estimates differed 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. Additionally, there were a number of fisheries with no recaptures which, had reporting rates fixed to 0.

Reporting rates hitting the upper bound are potentially symptoms of a model that is underestimating the fishing mortality and overestimating biomass relative to what the tagging data might indicate. Compared to the previous bigeye tuna assessments which had several reporting rates on the upper bound, the current diagnostic case represents an improvement from the previous mod-

els with only the reporting rates for the RTTP/PTTP program tags recovered by the Australian longline in region 5, and the JPTP program tags recovered by the Japanese pole-and-line fishery in region 1. High estimated reporting rates are consistent with our understanding of the level of tag-reporting within these fisheries, given their proximity to and awareness of tag releases (CSTP and JPTP) conducted in their operating area.

7.3.4 Growth

The Richards growth function of the diagnostic case model specifies most rapid growth for the youngest age-classes, starting from a mean length of about 30 cm for the youngest age-class, before slowing down over older ages with a mean length of 153 cm for the oldest fish in the model (age-class 40). The estimated standard deviation of length-at-age increases significantly with age and estimates substantial variation in length-at-age (Figure 24). This is necessary to fit the size frequency data observed for the longline fisheries and is supported by the high apparent variation in growth observed in the otolith dataset investigated external to MFCL.

7.4 Stock assessment results

7.4.1 Recruitment

The estimated distribution of recruitment across regions must be interpreted with some caution, as MFCL has a high degree of flexibility to use a combination of movement and regional recruitment to distribute the population in a way that maximizes the total objective function. The diagnostic case model recruitment estimates for each region and the entire assessment domain are shown in Figures 25 and 26. The overall pattern of recruitments is similar to previous assessments: high recruitments in the first few years of the assessment period, particularly in the northern regions. Recruitment is largely constant around a level of 50 million recruits beginning in 1960 (Figure 26). The increase in recruitments over the assessment period that has been a feature of the bigeye assessment for a number of years is substantially reduced in this assessment.

The estimate of the stock-recruit relationship (SRR) is presented in Figure 27. This relationship is defined over the period 1962–mid-2017, and excludes the influence of the high initial recruitments as these are often poorly estimated. The SRR indicates that recruitments have been maintained around average levels, even as stock size has declined.

It is important to note that, although aggregate levels of recruitment have remained relatively constant throughout the assessment period this is not the case regionally. The regional dynamics can have an influence on the overall model outcomes. In this case, the overall decline in biomass appears to be driven by a decline in recruitment in region 2. This indicates a potential mismatch between the level of removals relative to the estimated biomass and the decline seen in the CPUE since the model is relying on a recruitment trend to force the biomass to match the CPUE. Furthermore, while recruitment has declined considerably in region 2 and to a lesser extent region

5, this has been offset by increases in recruitment in the equatorial regions 3, 4, 7 & 8. Given the flexibility that MFCL has with the regional recruitments and seasonal movement, it is possible that increases in fishing mortality are being "buffered" by influxes of biomass from other regions.

7.4.2 Biomass

The relative pattern in spawning potential at the regional scale is consistent with the results presented in the previous stock assessments (Figure 26 and 28), and with the abundance distribution predicted from the spatiotemporal CPUE standardization model (Figure 29). There was a period of high abundance in the first decade of the assessment period before a period of sustained decline over most of the remaining time-series. As mentioned in Section 7.4.1, this appears to be in-part driven by the persistent decline in recruitment in region 2. Within the decline, there were several short term increases in abundance throughout the assessment period but these were never sustained for more than a few years. The decline was estimated to occur in all model regions (Figure 28). Similar to previous assessments, the regional distribution of individuals is concentrated in regions 2 & 4, though the proportion of biomass concentrated in region 2 has increased somewhat relative to previous assessments.

7.4.3 Fishing mortality

A steady increase in fishing mortality of adult age-classes is estimated to have occurred over most of the assessment period (Figure 30). Juvenile fishing mortality was initially low, before increasing after 1970 with the expansion of the pole-and-line fisheries. This increase in juvenile fishing mortality becomes particularly rapid after the expansion of purse seine fishing in the 1980's. In recent years, the juvenile fishing mortality has been high, but also variable. There appears to be some stabilization in the fishing mortality rates for both juvenile and adults over the last decade.

The fishing mortality-at-age estimates, by region (Figure 31), display intuitive results: the temperate regions where longline fishing dominates show highest fishing mortality for the vulnerable older age classes while in the equatorial regions where the purse seine and miscellaneous small fish fisheries occur show a high mortality for the youngest age-classes. The trends in fishing mortality-at-age by decade are presented in Figure 32 and show the general increase in fishing mortality of all age-classes over the assessment period, although the increase for the youngest age-classes is more pronounced and leads to an almost bimodal distribution with age-classes around 10 quarters having the lowest fishing mortality. The ages younger than this are most vulnerable to purse seine and miscellaneous fishing and the ages older are more vulnerable to longline fishing. Over time there has been a shift in the age distribution of the stock towards younger age-classes as the stock has been fished down (Figure 32).

7.5 Multimodel inference: One-off sensitivities

Comparisons of the spawning potential, and depletion trajectories for the diagnostic case and one-off sensitivity runs from the structural uncertainty analysis are provided in Figures 33–34. The key reference points for the one-off sensitivity models are compared in Appendix Tables 9–11 and the likelihood components are provided in Appendix Tables 12–14.

7.5.1 Relative weighting of length and weight frequency data [*Size10, Size60, Size200, Size500*]

Four alternative models were considered, one that up-weighted the size frequency data (*Size10*) and three that increasingly down-weighted the size frequency data (*Size60, Size200, Size500*). Model estimates for the *Size10* and *Size60* models did not show meaningful change from the diagnostic case for either spawning potential or depletion (Figures 33 & 34), however the more extreme down-weightings showed progressively lower estimates of spawning potential and more pessimistic levels of depletion. As the influence of the size-frequency data to the likelihood is reduced, it allows the model to fit the other components of the model better, namely CPUE. This indicates a likely conflict in the signals in these two data sources, with the abundance trend from the CPUE likely indicating a more pessimistic outcome.

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

Similarly to the previous set of sensitivities which deal with the likelihood contribution of the size-frequency data, the SSMULT-RE also adjusts the influence of the size frequency data. However, unlike the previous approach, the SSMULT-RE is able to adjust the weighting in a much more fine-scaled manner by up-weighting & down-weighting the size-frequency (through the estimation of ESS) for each pre-specified group of fisheries. There are a number of ways to group fisheries (similarity of gear fished, comparable size compositions, or fisheries where the data sampling program is similar), sensitivity to this grouping assumption is an area of further investigation. In general the model appears to up-weight weight frequency data from the longlines in regions 5, 6, & 8 and down-weighting the other fisheries, especially the region 7 Philippines and Indonesia handline (Fishery 18). In terms of length-composition, the model on average appears to down-weight the length observations from the purse seine (more so for the associated than the unassociated fisheries), severely down-weight the observations from region 7 fisheries, and increasing the weight of the pole-and-line fisheries observations.

This more tailored re-weighting of the size-frequency data which takes into account the initial sample size and auto-correlation of the samples results in a different estimate of stock status (Figures 33 & 34). Spawning potential is estimated to be scaled down to ∼75% of that estimated in the diagnostic case, resulting in overall depletion estimates approaching 30% and depletion estimates for the equatorial regions (3, 4, 7, & 8) approaching 20%$SB_{F=0}$.

Though additional evaluation needs to be completed on the set-up and configuration of this approach as mentioned above, this result gives an idea of what stock status is estimated to be under a more statistically appropriate weighting of the size composition data. This result is also only relative to the diagnostic case model; interactive effects with other sources of uncertainty such as growth and steepness in a structural uncertainty grid could result in a more optimistic or pessimistic estimate of stock status.

7.5.3 Alternative growth functions [*EstRichardsT, EstRichardsO, Oto-Only*]

On first impression it was not expected that the growth axis would have much influence as the 4 growth curves considered, appear similar in terms of the L2 parameter and show minor differences in the L1 and growth-rate (Figure 35). This is not unexpected as they are all informed, in part, by a common set of otolith data. However, information from both the tagging and length composition data appear to suggest a higher L1 parameter than the otolith data alone. Additionally, though estimated from the same data the two growth curves estimated using the conditional age-at-length data did not converge to the same estimate. This indicates that these curves are not well estimated internal to MFCL and the starting point dependency issues identified in Vincent et al. (2018) remain. Though Richards type growth curves were estimated using the conditional age-at-length data for comparison with the fixed curves, future assessments should test if using the simpler, von Bertalanffy growth curve can improve the robustness of the estimation given the starting point sensitivity.

In terms of model estimated quantities, the *Oto-only* model estimated much higher levels of biomass which corresponds to a more optimistic estimate of depletion given the same level of removals. Despite slight differences in the estimated curves the two conditional age-at-length models do not appear to converge to similar estimates of biomass or depletion. The *EstRichardsO* model estimates higher levels relative to the diagnostic case and the model *EstRichardsT* estimates levels to be intermediate to the previous model and the diagnostic case (Figures 33 & 34). Each conditional age-at-length curve produces estimates that appear to retain the influence of the initial growth parameters.

The reasons leading to the notably different estimate of biomass for the *Oto-Only* model are not yet fully understood, despite considerable efforts to identify the mechanism. Regardless of the reason, this alternate state of bigeye tuna estimated by the *Oto-Only* model shows a considerably worse fit to the size composition data overall, despite an improved fit for the smallest individuals caught in region 7. This degraded fit, in part, appears to manifest itself through selectivities that are shifted to larger, older individuals (most dramatically for the longlines). Given that for this model young fish are estimated to being largely invulnerable, the model has no option but to estimate higher levels of biomass in order to generate the appropriate numbers of removals from those selectivity patterns. A number of investigative models were run using the *Oto-only* growth with different assumptions of selectivity (10 nodes defining the cubic spline, both with and without the

non-decreasing assumption, and running the *Oto-Only* growth model with selectivity parameters fixed at the values estimated in the diagnostic case), however all yielded solutions similar, or more anomalous, than the original *Oto-Only* fit. Lastly, to test if model estimation was being trapped in a local minima, the *Oto-only* growth curve was inserted into the solution from the diagnostic case. Again, this hybrid model returned to a solution set similar to the original *Oto-only*.

Regardless of the mechanism for the divergent estimates, the high biomass estimates associated with the *Oto-Only* growth model may not be plausible when taken in the context of other WCPO tuna stocks. Over the last decade, the ratio of bigeye tuna to skipjack tuna catch in the western equatorial associated purse seine (Fishery 13 for bigeye tuna; Fishery 25 Vincent et al., 2019) appears to be 0.09. Assuming that catchability should be similar for bigeye tuna and skipjack tuna caught in the associated purse seines the model predicted vulnerable biomass for these fisheries should also preserve a similar ratio since they operate in a similar region with a similar selectivity shape. The assumption of equivalent catchability appears reasonable since they form mixed-schools of like-sized individuals and there is currently no contra-indication of behavioral differences at these sizes that would lead to differential catchabilities (Scutt Phillips et al., 2019). Comparing the model predicted ratios of vulnerable biomass of bigeye to skipjack for these fisheries from models assuming either one of the two fixed growth curves, the ratio from the *Tag-Int* model (0.06) appears to be more aligned with what is seen in the catch than the one from the *Oto-Only* model (0.43). Though this is an informal comparison, it does indicate that unless bigeye tuna have an associated purse seine catchability 5 times higher than equivalent sized skipjack the *Oto-Only* model biomass estimates are not biologically reasonable. Future research into the mixed-fisheries assessment approach is needed to more formally define how information on population scale can be shared between the tropical tunas.

7.5.4 Steepness [*h0.65, h0.95*]

The low penalties on the SRR relationship resulted in the assumed value of steepness having a negligible effect on model fit and time-series estimates of spawning potential (Figure 33). However, slightly different estimates of fisheries depletion were estimated, with the low (*h0.65*) and high (*h0.95*) steepness models suggesting a stock that is more and less depleted, respectively (Figure 34). This is consistent with previous assessments of tuna in the WCPO - namely that low and high steepness values lead to more pessimistic and optimistic estimates of stock status, respectively. This is particularly the case for MSY and MSY-based reference points (The $F_{recent}/F_{\text{MSY}}$ values for the diagnostic case level of 0.8, *h0.65* and *h0.95* are 0.63, 0.71 and 0.58, respectively) while the depletion-based reference points tend to be less sensitive to assumed steepness.

7.5.5 Tag data [*Mix1, TagFree*]

The model with the mixing period reduced to one quarter (*Mix1*) estimated negligible differences in spawning potential, and depletion (Figures 33 & 34). This indicates that the model is robust to

this change, relative to the configuration assumed for the diagnostic case.

Given the large shifts in biomass scale seen during model development due to changes in tag file preparation, a sensitivity was run without the tagging data included in the model (*TagFree*) to test for this potential source of uncertainty. Model estimates of spawning potential and depletion were surprisingly consistent albeit slightly higher than when the tagging data was included (Figures 33 & 34). This indicates that the dynamics governing movement and recruitment are being strongly influenced by another source of data (most likely the size-frequency). Additionally, although the tagging data do not appear to be very influential on the spatiotemporal dynamics of the model, they can influence the estimation of population scale particularly if interaction with the other data components results in the reporting rates being stuck on the upper bound limiting the appropriate estimation of fishing mortality and biomass.

7.5.6 Natural mortality [*M-low, M-mid, M-hi*]

From the new growth curves, a life history based meta-analysis yielded an envelope of potential average quarterly M rates: lower 95% confidence interval (0.109; *M-low*), mean (0.127; *M-mid*) and upper 95% confidence interval (0.146; *M-hi*). These values were applied to the age-specific deviates used in the diagnostic case to generate the three runs shown in Figures 33 & 34.

As expected for the *M-hi* scenario, increasing the assumed level of natural mortality reduced the level of estimated spawning potential since the higher rate prevented the accumulation of as many older individuals in the population. Even though the *M-hi* model predicted lower spawning potential, it produced a more optimistic level of stock status because these individuals were predicted to be removed from the population due to the higher level of natural mortality and not fishing mortality. Neither the *M-mid* or the *M-low* scenarios of spawning potential were very different from the estimate in the diagnostic case though this is not unexpected since M assumed in the diagnostic case (0.112) is fairly close to both. In terms of depletion the *M-low* model predicted the most pessimistic status, at just under the level seen in the diagnostic case, while the *M-mid* scenario produced estimates that were bounded by the other two models.

7.5.7 Selectivity groupings and structure [*Alt, FreeSel*]

As seen in the stepwise model development (Section 7.1), the assumptions made with regards to selectivity shape (non-decreasing for select longline fisheries; *FreeSel*) can have an effect on the estimated quantities. Here we also consider the effect of changing the way fisheries are grouped together (*Alt*). The diagnostic case selectivity groupings allowed all fisheries to be estimated independently (see Table 1; A few exceptions were made for certain fisheries which lacked sufficient samples of size data, the longline fisheries in region 9 and the purse seine fisheries in region 7). If sufficient size frequency data are available to support such independent estimation, this is a preferable approach for fisheries in which catchability is also estimated independently. In the *Alt* model, selectivities were shared between fisheries that operated similarly and/or caught like sized

individuals (e.g. all associated purse seine fisheries grouped together).

Compared to the other factors explored, assumptions on the grouping, and shape of the selectivity curve had a relatively low impact on the stock assessment outcomes (Figures 33 & 34). The *Alt* model estimated spawning potential to be scaled higher and with virtually identical levels of depletion as the diagnostic case. Removing the constraint on non-decreasing selectivity placed on some of the extraction longline fisheries (*FreeSel*) estimated a spawning potential, and a more optimistic level of depletion.

The fit to the weight composition data between the diagnostic case and the *FreeSel* model appeared relatively unchanged for most of the longline fisheries that had the non-decreasing penalty applied to them in the diagnostic case. The exceptions were the offshore longline fishery in region 3 (Fishery 5) and the longline ALL fishery in region 8 which indicated comparatively worse fits. In the future, an alternative selectivity structure relaxing the non-decreasing assumption for some of these longline fisheries could be explored. The statistical criteria for selecting an appropriate scheme will need to be explored, acknowledging that there are known issues related to model selection with complex likelihoods.

In addition to the results from the *Alt* selectivity grouping, a number of other selectivity groupings were considered spanning the spectrum between increased grouping and the independent by fishery approach of the diagnostic case. All groupings examined converged to a similar solution set.

7.5.8 Model start year [*Model1962, CPUE1962*]

Two sensitivities were run in order to investigate the impact and try to identify the source of the two high recruitment events in region 2 within the first 2 years of the assessment period. One-hypothesis was that if the standardized CPUE was mis-specified and hyper-depleted (declining at a faster rate than abundance) for this region early in the model period, then the model could compensate by using recruitment deviates to match the declining biomass trend from the CPUE. Removing the first 10 years of CPUE data and starting in 1962 (*CPUE1962*) after the largest declines in the index did not remove the estimation of these two large recruitment events indicating that the CPUE index is not responsible and that there must be information in the weight-frequency and catch data to inform these large recruitment events. Beginning in 1958 there is a 5 year period of high observed catches in this fishery and the weight-frequency from 1959 shows an increased proportion 65+ kg individuals. According to the assumed growth curve and length-weight relationship, recruits entering the model 6-7 years prior would be at the appropriate size to account for the high number of removals at these larger sizes around 1959. Starting the model in 1962 (*Model1962*) did not show similar high estimates of recruitment at the start of this new model period indicating that the high recruitments seen in the diagnostic case are likely informed by the pattern in catch and weight-frequency data. Regardless, these recruitments do not seem to be impacting the terminal estimates of spawning potential and depletion (Figures 33 & 34) likely as a result of the stock-recruit relationship being defined from 1962–mid-2017.

7.5.9 Size-frequency data [*idxAU-US, len456*]

In the preparation of the size-frequency data two sets of assumptions were made: the weight-frequency data for the index fisheries would come from flags belonging to the longline ALL fisheries, and that weight-frequency data would be used exclusively for the longlines rather than switching to length-frequency samples for regions 4, 5, & 6 post-2000 when that form of sampling became more prevalent. Sensitivity of model outputs to both of these assumptions are discussed here. Making the decision to include the additional weight-frequency data from the Australian and US longlines (*idxAU-US*), in regions 5 and 2 respectively did not measurably impact the estimates of either spawning potential or depletion (Figures 33 & 34). Switching from weight-frequency to length frequency for the longline fisheries in region 4, 5, & 6 post-2000 (*len456*) had a small impact, with spawning potential scaled up and depletion slightly more optimistic. Investigation of the size-frequency data showed that this change appears to be driven by a reduction in the median weight for these fisheries that is not matched in the length-frequency data.

7.6 Multimodel inference: structural uncertainty analysis

Axes for the structural uncertainty grid were selected after considering the results from the one-off sensitivities, previous bigeye tuna stock assessments (McKechnie et al., 2017a; Vincent et al., 2018), and feedback from the PAW (Hamer and Pilling, 2020). On this basis, axes for composition data-weighting, natural mortality, and steepness were selected as these encompassed the greatest sources of plausible variability. Growth was intended as an axis of uncertainty, however this was discarded, for the reasons mentioned in Section 7.5.3, due to a lack of sensible alternatives to the *Tag-Int* model.

The results of the structural uncertainty analysis are summarized in several forms – time-series plots of fisheries depletion for all models in the grid (Overall Figure 36; by region Figure 37–Figure 39), time dynamic percentiles of depletion (Overall Figure 40; by region Figure 41), box-plots of $F_{recent}/F_{\mathrm{MSY}}$ and $SB_{latest}/SB_{F=0}$ for the different levels of each of the 3 axes of uncertainty (Figure 42), Majuro plots showing the estimates of $F_{recent}/F_{\mathrm{MSY}}$ and $SB_{latest}/SB_{F=0}$ (and $SB_{recent}/SB_{F=0}$ for comparison) across all models in the grid (Figure 43), and averages and percentiles across the full grid of 24 models for all the reference points and other quantities of interest (Table 6) that have also been presented for the diagnostic case model and one-off sensitivity models.

Many of the results of the structural uncertainty analysis are consistent with the results of previous assessments of tuna stocks in the WCPO that used the same uncertainty axes. However, additional axes have been included in the 2020 assessment and these have consequences for summarizing stock status and deriving management recommendations.

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

- Uncertainty in the grid remains comparable to the previous stock assessments for bigeye tuna showing a ∼30 point spread in the estimates of $SB_{latest}/SB_{F=0}$, and median values appear

similar, though slightly more optimistic than those estimated in the previous assessment (McKechnie et al., 2017a; Vincent et al., 2018).

- All models predict bigeye tuna to be above $20\%SB_{F=0}$, with individual model runs estimating depletion at the end of the assessment period being between about 0.30 and 0.47 in terms of $SB_{latest}/SB_{F=0}$, with a slightly wider range for $SB_{recent}/SB_{F=0}$ (0.27–0.52;Table 6).
- However, there is evidence to suggest that the overall stock status is being "buffered" by the temperate regions (1, 2, 5, & 6). The most pessimistic predictions of overall stock status correspond to models where depletion in these temperate regions is predicted to be high and in some cases approach regional $20\%SB_{F=0}$.
- The equatorial regions (3, 4, 7, & 8) appear to show consistent levels of regional depletion approaching $20\%SB_{F=0}$ across models.
- Patterns in regional depletion are driven by the increased down-weighting of the size-frequency data-weighting axis (Figure 37). This could be due to a reduction in conflict between the regional trends indicated by the CPUE versus the size-frequency data or through a change in the estimated movement dynamics given the importance of the size-frequency data to the estimation of these parameters (Section 7.5.5) as the size-frequency data becomes increasingly down-weighted. Indeed, models with a down-weighting of 500 show a different pattern of movement to that of the diagnostic case, with more biomass flowing from the temperate regions to equatorial regions (particularly from region 1).
- The most influential grid axis is the size-frequency data-weighting axis, with decreased weights leading to more pessimistic outcomes of depletion (as shown for the regional depletion estimates). Within this axis, two of the levels show strong overlap, and relatively optimistic outcomes (20 & 60; Figure 42). The level of 200 shows some overlap with the lower range of the two preceding levels, and the extreme down-weighting of the size composition, 500, shows the greatest separation from the other axis levels.
- The next most influential axis was the natural mortality axis. This axis effectively splits each data-weighting level in two, with the high natural mortality scenarios sitting on top of the mean natural mortality level that was assumed in the diagnostic case, and showing more optimistic outcomes.
- The steepness axis displayed largely predictable results, with steepness of 0.65 and 0.95 producing more pessimistic and optimistic estimates, respectively, than the 0.8 assumed in the diagnostic case model. The lower the steepness the more depleted the stock, and the higher the fishing mortality with respect to fishing mortality at MSY.
- Across all axis levels, only the extreme size data down-weighting axis predicts fishing mortality in the recent period to exceed the fishing mortality that produces MSY ($F_{recent}/F_{\text{MSY}}$).
- Most models across the grid show more pessimistic stock status in recent years, which appears

to match the pattern seen in the standardized CPUE. The grid runs corresponding to the extreme data down-weighting show the opposite pattern, possibly as a result of the different movement dynamic observed for these models.

7.7 Further analyses of stock status

There are several ancillary analyses related to stock status that are typically undertaken on the diagnostic case model (dynamic Majuro analyses, yield analyses, etc.). The use of multi-model inference, defining management reference points based on the ensemble of grid runs, complicates the presentation of these results for each individual model. These analyses are presented with respect to the diagnostic case.

7.7.1 Fishery impacts

Fishery impact is measured at each time step as the ratio of the estimated spawning potential relative to the spawning potential that is estimated to have occurred in the historical absence of fishing. This is a useful quantity to monitor, as it can be computed both at the region level, and for the WCPO as a whole. This information is plotted in two ways, firstly as the fished and unfished spawning potential trajectories (Figure 44), and secondly as the depletion ratios themselves (Figure 45). The latter is relevant for the agreed reference points and example plots of these values are displayed for the diagnostic case model.

The diagnostic case model estimated that steady declines in spawning potential have occurred in all regions (Figure 44), although the rate and extent of decline differs among regions. The equatorial regions (3, 4, 7, and 8) show a rapid, increased rate of decline beginning in the 1970s, with most decline attributed to fishing impacts. In the more temperate regions by contrast, most of the declines in spawning potential are estimated to have occurred due to declining recruitment, rather than the impacts of fishing in these regions. With respect to overall depletion, the temperate regions appear to be "buffering" the effects of fishing pressure within the tropics. These patterns are reflected in the regional depletion plots (Figure 45), with similar sustained declines in all regions for both models, except for region 9 which shows more volatility, likely due to its small size.

It is also possible to attribute the fishery impact with respect to depletion levels to specific fishery components (grouped by gear-type), in order to estimate which types of fishing activity have the most impact on spawning potential (Figure 46). The early impacts on the population were primarily driven by longline fishing, but as the purse seine fishery expanded from the 1980's the impact of associated purse seine fishing has rapidly increased. Within the equatorial regions it is the dominant driver of fishery impacts, eclipsing the impacts of the longline fisheries which have continued to gradually increase over time. The substantial increase in fishing effort by miscellaneous gears in region 7 has also had a significant impact on spawning potential in that region, and also in neighboring regions that are linked by movement. Unassociated purse seine fishing has comparatively little impact due to far fewer bigeye being caught in these sets.

While the overall depletion levels of the 2020 diagnostic case model are similar to the 2017 diagnostic case model, there are some differences in the regional contributions to depletion by fishing gear (Figure 46). In the current assessment, the northern temperate regions are predicted to be less depleted than in the 2017 assessment, while the southern temperate regions show greater levels of depletion relative to the previous assessment. This is likely due to a differences in the estimated movement dynamics. Additionally, despite not having purse seines operating in any of the temperate regions, they are increasingly impacted by the associated purse seines. This could be a product of reduced immigration as recruits from the equatorial regions that would be expected to immigrate to the temperate regions in the absence of fishing are being caught by the associated purse seine fisheries.

7.7.2 Yield analysis

The yield analyses conducted in this assessment incorporate the stock recruitment relationship (Figure 27) into the equilibrium biomass and yield computations. Importantly, in the diagnostic case model, the steepness of the SRR was fixed at 0.8 so only the scaling parameter was estimated. Other models in the one-off sensitivity analyses and structural uncertainty analyses assumed steepness values of 0.65 and 0.95.

Across the structural uncertainty grid the equilibrium virgin spawning potential in the absence of fishing ($SB_{F=0}$) was estimated to be between 903,708 and 1,908,636 mt (Table 6), and the spawning potential that would support the MSY (SB_{MSY}) was estimated to be between 192,500 and 482,700 mt. Relative to the previous stock assessment, the current analysis predicts a marginally less productive stock. The ratio of SB_{MSY} to $SB_{F=0}$ was estimated to be between 0.19 and 0.26 (mean = 0.23).

A plot of the yield distribution under different values of fishing effort relative to the current effort are shown in Figure 47, for models in the structural uncertainty grid. Across most grid axis–level combinations, the yield analysis indicates that current fishing pressure is close to producing MSY. More productive combinations, higher natural mortality and/or higher steepness, expectedly predict the highest potential yields, and that these would be achieved under higher fishing pressure.

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 48). 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.

7.7.3 Dynamic Majuro plots and comparisons with Limit Reference Points

The section summarizing the structural uncertainty grid (Section 7.6) 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 dynamic Majuro plots, the methods of which are presented in Section 5.7.4. An example for the diagnostic case is presented in Figure 49. These examples produce intuitive results with respect to the terminal results already presented in Section 7.6. At the start of the assessment period, the WCPO bigeye tuna stock was predicted to be close to an $SB/SB_{F=0}$ of one or virgin unfished condition. The F/F_{msy} was estimated to be nearly zero indicating that initial fishing pressure at the start of the model period was negligible. Each of these reference points progressively shifted towards the overfishing and overfished definitions over the remaining period. The diagnostic case model shown never reaches $20\%SB_{F=0}$ or an F/F_{msy} of one. The equivalent dynamic Kobe plots are displayed in Figure 50.

8 Discussion and conclusions

8.1 General remarks

The 2020 WCPO bigeye tuna stock assessment estimates the median stock status to be comparable to the *10 ° N–Updated New growth* model from the 2018 reanalysis (Vincent et al., 2018) used to provide management advice for the stock. Taking a historical perspective of the trajectory of the stock, depletion appears to have been relatively low through the first 20 years of the model period. The apparent declines in biomass during this period appear to be due to several large cohorts from high recruitment events in the early 1950's, moving through and out of the population. The level of fisheries depletion increases in the 1970's as fishing mortality on juveniles slowly begins to increase. This coincides with the expansion of DWFN & PICT pole-and-line fisheries targeting skipjack tuna, of which juvenile bigeye are a bycatch species. In the 1980's, coinciding with the development of industrial purse-seining, an increase in juvenile fishing mortality results in an increasing trajectory of depletion. Erosion of the stock structure begins to manifest itself in the 1990's, and is likely a result of the increased fishing pressure on juvenile bigeye tuna. Additionally, it is important to note that throughout this time of increased fishing mortality on the juveniles, fishing mortality on adults captured via the longline fisheries has also been estimated to increase though at a less rapid rate. The stock was estimated to decline further through the turn of the century before stabilizing at the beginning of the current decade. However, the "bump" in biomass seen in 2014–2015 due to the predicted high recruitment at the beginning of the decade appears to have passed through the population, with the stock showing slight decline since the last assessment. Fishing mortality appears to have stabilized over the last decade for both the adult and juvenile age classes, though inter-annual variation is high, particularly for the juveniles. Throughout this period, outside of the two large early recruitment events, annual recruitment appears to have been fairly stable ($\sim$ 50 million individuals per year) across a wide range of population sizes.

8.2 Improvements to the assessment

The process of conducting a stock assessment is one of iterative and continued improvement, picking up the development where the previous assessment left off and incorporating new model features, updated data, and an improved understanding of biological processes. The 2020 bigeye tuna stock assessment is no exception. Building off the previous WCPO tuna assessments, the current assessment made several notable improvements: (i) spatiotemporal CPUE (Ducharme-Barth and Vincent, 2020); (ii) length-based reproductive potential ogive (Davies et al., 2018); and (iii) updated growth information from otolith and tag-recapture data-sets (Eveson et al., 2020; Farley et al., 2020). The use of a spatiotemporal CPUE model allowed for a "gap-free" standardized index to be available in all regions and all time-steps for the first time. Additionally, modeling the operational longline data across the Pacific-basin allowed for the relative uncertainty between regional indices to be captured in the assessment. Using a length-based reproductive potential ogive provided an improved definition for a key biological process, and since growth was estimated internally in a subset of runs, allowed for the correctly specified reproductive potential-at-age ogive to be used. The new growth curves for the assessment are the latest steps in a process initiated in 2017 (Farley et al., 2017; McKechnie et al., 2017a), and incorporates improvements to: the ageing methodology; calculation of decimal age for the conditional age-at-length without making a common birth date assumption; inclusion of otoliths from small 15–20 cm bigeye tuna; and makes use of information contained in the tagging data to define growth.

Additionally, a number of other small improvements were made to the model including updating the processing models for the tagging and composition data inputs, and refreshing the assumed length-weight relationship. Also of note for this assessment, is that in the aggregate, the net effect of the changes made in the assessment resulted in a much improved state with respect to tag reporting rate estimation. Only two groups were estimated to be impacted by the upper bound of the bounded parameter space, and these belonged to fisheries where the tag reporting rate is presumed to be high.

8.3 Uncertainty

The range of uncertainty presented in the structural uncertainty grid is comparable to that presented in the previous stock assessment (McKechnie et al., 2017a; Vincent et al., 2018), although perhaps broader than the final subset of models used as the basis for management advice. The key driver with regards to uncertainty in the current assessment is the size-frequency data-weighting axis which used a larger range of down-weighting than previous assessments. Although it is generally thought that the size-composition has too large of an influence on the likelihood (for the reasons mentioned in Section 5.5), it is challenging to objectively determine what the appropriate weighting should be. Development of the SSMULT-RE likelihood for size-frequency data can hopefully overcome this deficiency, however further investigation is needed before this feature is ready for use as the baseline likelihood structure. In the meantime, the range of values selected

for the data-weighting axis cover the range of outcomes between the data-weighting that has traditionally been assumed for bigeye tuna and the results shown from the one-off sensitivity using the SSMULT-RE. Though this range of down-weighting is more extreme than has been considered in the past, the SC will need to consider the trade-off between reducing the fit to the size composition data and the impact on model dynamics from the resulting improvement in the fit to the other main components of the likelihood: CPUE (via the effort deviates on the index fishery) and the tagging data.

The absence of growth as an axis in the structural uncertainty grid is conspicuous given the influence it was shown to have in previous assessments and in some of the one-off sensitivities for the current assessment. The *Oto-Only* growth model predicted very high levels of biomass and corresponding low level of depletion despite visually showing similar patterns in growth to the *Tag-Int* model. These results appear implausible given the poor fit to the other data components (Section 7.5.3) and when comparing the scale of estimated biomass to those from other tropical tunas (Section 8.4.1). Considerable effort was devoted to uncovering the mechanism for this dramatically different solution set, but as yet we do not have a resolution. Fitting to the conditional age-at-length data showed estimates of biomass and depletion that appeared more reasonable. However, given the lack of agreement in the estimated growth curves and the disparity between model outcomes driven solely by the parameter starting conditions, these models were determined to be poorly estimated and thus less reliable. For these reasons, both the *Oto-only* and conditional age-at-length estimation models were excluded from the grid. This potentially results in an under-representation of the total structural uncertainty. However given the generally more optimistic nature of these models, based on the results from the one-off sensitivities, ignoring this potential added source of uncertainty is not likely to contribute to an underestimate of the risk of breaching $20\%SB_{F=0}$.

8.4 MFCL and other modeling considerations

Model development of the 2020 stock assessment was not without its challenges. These were identified at various stages of the development process and broadly fall into two categories: those related to the biological assumptions and data-inputs; and those that relate to model complexity.

8.4.1 Biological assumptions and data-inputs

Bigeye growth again proved to be problematic and challenging, though in a way that was different from the previous assessments. Previous issues related to growth evolved from different ageing methods being applied to produce growth curves that visually appeared to be very different from one another. The peculiarity of this particular case is that the same set-of otoliths, all aged the same way, were utilized in the construction of the different growth options. These growth options look very similar, and indeed all estimate L2 parameters within ∼1–2 cm. The differences lie more at smaller sizes/lower ages, with some models predicting an L1 of ∼20 cm and fitting the newly collected small-fish otoliths quite well. The models including other sources of data, either

tag-increment data or modal progression information from the size composition, tend to estimate higher L1 parameters indicating some conflict between the different data sources. These subtle differences in the L1 parameters and the rapid initial growth that is able to be achieved using the Richards growth curve appear to be pushing model estimates in diverging directions, the exact mechanism for which requires further exploration. It is possible that in attempting to fit the size-frequency data, the quarterly age increments are too far apart to properly capture the variability in sizes given the rapid growth achieved by younger individuals in a short amount of time. Rather than increase the resolution of the age classes, the relationship governing variability around the growth curve could be relaxed to accommodate the fit to these young, small fish. Additionally, a proper length-based selectivity function could aid in capturing these modes that lie between age-classes.

A number of critical biological assumptions that under-pin the bigeye tuna stock assessment are informed by studies with fairly small sample sizes. The fecundity-at-length relationship used in the reproductive potential ogive is based on a sample of 129 individuals collected from a geographically small region around the Philippines and Chinese Taipei (Sun et al., 2006). The maturity-at-length relationship is based on samples covering a broader spatial area but is still based on relatively few individuals (n = 343; Farley et al., 2017). Perhaps most critically, the conversion factor used to convert longline caught individuals ("gilled-and-gutted" weight to whole weight) from the fresh fish fleets of many of the PICTs is based on 79 samples from longline vessels operating in the Solomon Islands and the Federated States of Micronesia (Langley et al., 2006). Since this conversion factor is applied to all longline caught fish not processed using the Japanese style of gilling (and removing the operculum), gutting, and tailing the fish, small changes to this conversion factor could resonate within the model. This deficiency has previously been identified through Project 90 (SPC-OFP, 2019), however emphasize the importance of allocating resources to collect additional samples across a number of fleets to improve this conversion. Additionally, we recommend collection of additional reproductive samples across the extent of the WCPO and processing of existing samples within the SPC tissue bank in order to improve confidence in key biological assumptions around reproduction and sex ratio with length.

A number of research recommendations to improve the standardized CPUE were outlined in Ducharme-Barth and Vincent (2020). These are repeated here as the standardized CPUE index from the longline fisheries is a critical input to the assessment, and improvements to that index would benefit the assessment.

- Data reconstruction to improve the number of covariates available across fleets for the duration of the model period.
- Interviews and surveys with vessel operators and captains to identify species specific targeting practices and relevant covariates to standardize on.
- Create a recent period index to take advantage of improved covariate coverage using either the observer data or a combination of observer and operational data with an expanded covariate

set.

- Revisit previous analyses of catch-per-effective-effort which account for the vertical distribution of tropical tuna biomass and longline hooks in the water column.
- Allocate funding to produce a spatiotemporal sub-surface oceanographic product which includes dissolved oxygen for the Pacific Ocean for the entire assessment period.
- The analysis of archival tagged, longline-vulnerable bigeye and yellowfin tuna needs to be continued and supplemented by additional releases in multiple locations across the WCPO in order to complete our understanding of how vertical position in the water column is a function of fish length, time of day, and oceanographic conditions.
- Basic research on the depth distribution of longline hooks across vessels from DWFNs and PICTs.

If we could choose a sampling program for the bigeye tuna stock in the WCPO it would be a spatiotemporally stratified fisheries independent sampling program where the same stations are assessed inter- and intra-annually. Given the massive scale of the WCPO, such a program is just not tractable without collaborating with the industry. The feasibility of implementing a quasi-fisheries-independent sampling program should be assessed. A brief proposal for such a program is described in the Appendix (Section 11.1).

8.4.2 Model complexity

Despite the large amounts of data available for this assessment, there is some evidence to suggest that the current structure of the bigeye tuna stock assessment is overly-complex and over-parameterized. Run times are long and the model is slow to converge. Though inconvenient from a development and diagnostic stand-point this in and of itself is not a major concern, and there are several aspects of MFCL in active or recent development that can help address the complexity problem by drastically reducing the number of estimated parameters: i.e. the orthogonal polynomial parameterization for recruitment and catch-conditioning of the assessment model. More concerning is the model's flexibility to shift biomass across the 9 regions using a combination of movement coefficients and recruitment parameters; and the potential for confounding between these two sets of parameters. As an example, in the course of development, a model was run which turned off the 2,403 regional recruitment deviates (n regions $\times t-1$ quarterly time-steps in the model). This simplified model was able to produce spawning potential and depletion time series that closely matched the diagnostic case by manipulating the estimated proportion of recruitment in each region and the quarterly movement rates. Granted, the ensuing distribution of recruitments did not match our current understanding of bigeye tuna spatial recruitment patterns, the model was flexible enough to arrive at a similar solution through a different set of parameters.

Model complexity and spatial structure are inextricably linked, with the movement dynamics play-

ing an important role in how the model fit to the data is achieved. Explicit definitions of spatial structure in an assessment can provide for a more realistic representation of complex dynamics. Explicit spatial structure is useful because the selectivity function can be separated into spatial availability driven by regional movement, and vulnerability to the gear as a function of size or age. It also allows for the compartmentalization of the population into quasi-distinct components of variable size, subject to different levels of fishing mortality between regions. However, the difference between the aforementioned benefits and improper interpretation of the selectivity function or spatial regions inappropriately "buffering" the effects of more depleted regions hinge, in part, on the ability to get the movement dynamics correct. The information from tagging data should be an informative component to the estimation of the movement dynamics as it is the only data source that explicitly informs movement between regions, all other data sources (such as size-frequency data) only infer it. At least for the current bigeye tuna assessment, this does not appear to be the case. Removal of the tagging data did not appear to greatly change the estimated biomass and depletion trajectories or the movement dynamics. This is potentially problematic on two fronts: (i) if the estimation of movement is driven by regional differences in size-frequency it could contaminate the estimated movement rates if there is spatial variability in growth; and/or (ii) that given the current spatial footprint of the tagging programs and recapture locations, there is not enough information content in the data to properly estimate the 104 diffusion coefficients required for the 9 region model. Fortunately, moving forward, there are ways to correct for either of these issues. A model explicitly accounting for spatial heterogeneity in growth (though this raises additional challenges with regards to implementation and model complexity) or down-weighting of the size-frequency data (via the sample size or choice of likelihood) can reduce the influence of the size-frequency data on the movement rates. With regards to the tagging data, if the existing spatial structure is maintained, the spatial sampling design of the tagging program could be redesigned to provide better estimates of movement across regional boundaries. Alternatively, the spatial structure could be reconfigured to maximize the existing information content of the tagging data. Another possibility is for the tagging data to be used in an integrated ecosystem model such as SEAPODYM (Senina et al., 2020) to estimate region-specific movements, and input these external movement estimates into MFCL as informative "priors".

Given the probable over-complexity of the current model structure (e.g. quarterly recruitment dynamics and movement across 9 spatial regions), the next bigeye tuna stock assessment should take steps to reduce model complexity. This should begin with investigating and making full use of the orthogonal polygon recruitment parameterization, and should also completely remove the effort deviates from the model (8,438 in the current diagnostic case) through the catch-conditioned parameterization. In order to remove the effort deviates associated with the index fishery, the index fisheries should be removed and a proper likelihood for CPUE should be implemented in MFCL. One of the challenges associated with pruning an integrated assessment model is how to make decisions as to which components are appropriately complex and those that should be simplified so as not to

sacrifice too much biological realism on the "altar of parsimony[8]". Data information theory (AIC or BIC) and likelihood ratio-tests can be used in this process so long as the data/likelihood structure is unchanged and that the models are nested in the case of likelihood ratio-tests. Objective decision making is more challenging when considering between alternatives with incomparable likelihoods. An external peer-review and/or WCPFC stock assessment modeling workshop should be conducted prior to the next WCPO bigeye tuna stock assessment to provide feedback on the development of an appropriate model structure. To this effect, and to address some of the issues identified as related to the tagging, movement, and regional parameter estimation; a simplified, single region assessment model was developed in tandem with the spatially structured diagnostic case model. This model is presented briefly in the Appendix (Section 11.8) and can serve as a starting point for discussions on model complexity.

8.5 Main assessment conclusions

The main conclusions of this assessment are summarized as follows:

- All models in the structural uncertainty grid show WCPO bigeye tuna to be above $20\%SB_{F=0}$, which is consistent with the previous assessments following the adoption of the *Updated New* growth and the *10° N* regional structure.
- However, there is evidence to suggest that the overall stock status is being "buffered" by the temperate regions (1, 2, 5, & 6). The most pessimistic predictions of overall stock status correspond to models where depletion in these temperate regions is predicted to be high and in some cases approach regional $20\%SB_{F=0}$.
- The equatorial regions (3, 4, 7, & 8) appear to show consistent levels of regional depletion approaching $20\%SB_{F=0}$ across models.
- A substantial decline in bigeye abundance was estimated by all models in the assessment. With respect to the difference in recent levels of depletion and those from the beginning of the model period, the impacts of fishing pressure appear to be persistent and meaningful, at least on a multi-year scale.
- The effects of the large recruitments seen at the end of the previous stock assessment do not appear to have continued in recent years. Most models across the grid show a downturn in stock status in recent years, which appears to match the pattern seen in the standardized CPUE.

8.6 Summary of research recommendations

A brief summary of the research and modeling recommendations made elsewhere in this report:

[8]The "altar of parsimony" is not to be confused with that of speed, though a more parsimonious model is likely to also be a faster model.

- Continued research on the implementation of new features to MFCL such as the: censored-gamma tag likelihood (Section 7.2.4), Self-Scaling Multinomial with Random Effects likelihood for size-frequency data (Section 5.5).
- Model structure (complexity) should be critically evaluated via an external peer-review or WCPFC modeling workshop, and the model structure should be revised accordingly (Section 8.4.2). Further, reductions in model complexity should be investigated via the re-parameterization of recruitment using orthogonal polynomials and conditioning the assessment model on the catch.
- The spatial sampling design for the tagging programs needs to be revisited to ensure that releases are made in areas that will assist in the estimation of regional movement coefficients (Section 8.4.2). Additionally, efforts should be made to quantify the movements of older, longline vulnerable bigeye tuna through electronic tagging as the majority of the movement information from the tagging data comes from conventional tags placed on juvenile bigeye tuna.
- Investigate the feasibility of constructing and incorporating informative movement priors using an external integrated ecosystem model such as SEAPODYM.
- Development of a formal likelihood component for the fit to the CPUE to further reduce the number of parameters estimated when implementing the "index" fisheries (Section 8.4.2).
- Continue work on Project 90 in order to base conversion factors on a larger number of samples (Section 8.4.1). If possible, sampling should be stratified by fishery so fleet specific conversion factors can be developed.
- Continue research on important biological processes such as growth and maturity (Section 8.4.1). Further otoliths should be collected in a spatially stratified manner to help identify regional differences in growth. Additional biological samples should also be collected in a spatially stratified manner to capture possible variations in maturity-at-length and fecundity-at-length across the WCPO, and develop more representative inputs for the stock assessment. Existing biological samples stored in the SPC Tissue Bank should be processed to further supplement the additional sampling.
- The variability around the growth curve in MFCL should be restructured to allow for increased variability around the growth curve for the youngest ages to account for the rapid growth at these ages (Section 8.4.1).
- Development of a truly length-based selectivity function within MFCL could improve the fit to the data of these younger smaller individuals (Section 8.4.1).
- Data reconstruction to improve the number of covariates available for CPUE standardization across fleets for the duration of the model period (Ducharme-Barth and Vincent, 2020).

- Interviews and surveys with vessel operators and captains to identify species specific targeting practices and relevant covariates to use for CPUE standardization (Ducharme-Barth and Vincent, 2020).
- Creation of a recent period CPUE index to take advantage of improved covariate coverage using either the observer data or a combination of observer and operational data with an expanded covariate set (Ducharme-Barth and Vincent, 2020).
- Revisit previous analyses of CPUE which account for the vertical distribution of tropical tuna biomass and longline hooks in the water column (Bigelow et al., 2002; Ducharme-Barth and Vincent, 2020).
- Allocate funding to produce a spatiotemporal sub-surface oceanographic product which includes dissolved oxygen for the Pacific Ocean for the entire assessment period to account for important oceanographic effects in the spatiotemporal CPUE standardization (Ducharme-Barth and Vincent, 2020).
- The analysis of archival tagged, longline-vulnerable bigeye and yellowfin tuna needs to be continued and supplemented by additional releases in multiple locations across the WCPO in order to complete our understanding of how vertical position in the water column is a function of fish length, time of day, and oceanographic conditions (Ducharme-Barth and Vincent, 2020).
- Basic research on the depth distribution of longline hooks across vessels from DWFNs and PICTs (Ducharme-Barth and Vincent, 2020).
- Assess the feasibility of funding a quasi-fisheries-independent sampling program to improve collection of important biological samples, serve as a platform for conducting experiments to explore the selectivities and catchabilities of the gear, and generate a fisheries-independent index of CPUE (Section 8.4.1).

Acknowledgements

We thank the various fisheries agencies for the provision of the catch, effort and size frequency data used in this analysis, and in particular NRIFSF and especially K. Satoh and T. Matsumoto, for the provision of the JPTP tagging data. We are deeply appreciative to the many sacrifices made by onboard fisheries observers who play an instrumental role in collecting data that comprise critical inputs to the stock assessment. We thank S. McKechnie for his insights into the inner workings of the model dynamics and his discerning eye when it comes to visualizing data. This stock assessment was supported by funding from the Western and Central Pacific Fisheries Commission.

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.

Allain, V., Pilling, G. M., Williams, P. G., Harley, S., Nicol, S., and Hampton, J. (2016). Overview of tuna fisheries, stock status and management framework in the Western and Central Pacific Ocean. In Pauwels, S. and Fache, E., editors, *Fisheries in the Pacific*, pages 19–48. pacific-credo Publications.

Alvarado Bremer, J. R., Stequert, B., Robertson, N. W., and Ely, B. (1998). Genetic evidence for inter-oceanic subdivision of bigeye tuna (Thunnus obesus) populations. *Marine Biology*, 132(4):547–557.

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.

Bigelow, K. A., Hampton, J., and Miyabe, N. (2002). Application of a habitat-based model to estimate effective longline fishing effort and relative abundance of Pacific bigeye tuna (*Thunnus obesus*). *Fisheries Oceanography*, 11(3):143–155.

Bousquet, N., Dortel, E., Chassot, E., Million, J., Eveson, J., and Hallier, J.-P. (2012). Preliminary assessments of tuna natural mortality rates from a Bayesian Brownie-Petersen model. Technical Report IOTC-2012-WPTT14-41.

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

Cass-Calay, S., Tetzlaff, J., Cummings, N., and Isley, J. (2014). Model diagnostics for StockSynthesis 3: Examples from the 2012 assessment of cobia in the U.S. Gulf of Mexico. Technical Report Collect.Vol. Sci. Pap. 70(5): 2069-2081, ICCAT.

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

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

Davies, N., Fournier, D. A., Takeuchi, Y., Bouye, F., and Hampton, J. (2018). Developments in the MULTIFAN-CL software 2017-2018. Technical Report WCPFC-SC14-2018/SA-IP-02, Busan, South Korea, 9–17 August 2018.

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

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

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

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

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 WCPFC-SC16-2020/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 WCPFC-SC16-2020/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., Nielson, A., and Sibert, J. (2012). AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, 27(2):233–249.

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

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

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.

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

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

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

Hampton, J. and Williams, P. (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. (2017). Annual Estimates of Purse Seine Catches by Species based on Alternative Data Sources. Technical Report WCPFC-SC13-ST-IP03, Rarotonga, Cook Islands, 9–17 August 2017.

Harley, S., Hoyle, S., Langley, A., Hampton, J., and Kleiber, P. (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 reproductive 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.

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 (*Thunnus albacares*) in Hawaiian waters and the western tropical Pacific Ocean: Project summary. JIMAR Contribution 00-328 SOEST 00-01.

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

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

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

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

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

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

McKechnie, S., Pilling, G., and Hampton, J. (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.

Morita, Y. (1973). Conversion factors for estimating live weight from gilled-and-gutted weight of bigeye and yellowfin tunas. *Bulletin of the Far Seas Fisheries Research Lab*, 9:109–121.

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.

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

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

Peatman, T., Fukofuka, S., Park, T., Williams, P., Hampton, J., and Smith, N. (2019). Better purse seine catch composition estimates: progress on the Project 60 work plan. Technical Report WCPFC-SC15-2019/ST-WP-02, Pohnpei, Federated States of Micronesia.

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

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

Rooker, J. R., David Wells, R. J., Itano, D. G., Thorrold, S. R., and Lee, J. M. (2016). Natal origin and population connectivity of bigeye and yellowfin tuna in the Pacific Ocean. *Fisheries Oceanography*, 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 (*Thunnus obesus*) tagged and released in the equatorial Central Pacific Ocean, with conventional and archival tags. *Fisheries Research*, 161:336–355.

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

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

Scutt Phillips, J., Leroy, B., Peatman, T., Escalle, L., and Smith, N. (2019). Electronic tagging for the mitigation of bigeye and yellowfin tuna juveniles by purse seine fisheries. Technical Report WCPFC-SC15-2019/EB-WP-08, Pohnpei, Federated States of Micronesia.

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

Senina, I., Lehodey, P., Sibert, J., and Hampton, J. (2020). Integrating tagging and fisheries data into a spatial population dynamics model to improve its predictive skills. *Canadian Journal of Fisheries and Aquatic Sciences*, 77(3):576–593.

SPC-OFP (2019). Project 90 Update: Better data on fish weights and lengths for scientific analyses. Technical Report WCPFC-SC15-2019/ST-WP-03, Oceanic Fisheries Programme, The Pacific Community.

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

Thorson, J. T. (2019). 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., Shelton, A. O., Ward, E. J., and Skaug, H. J. (2015). Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. *Ices Journal of Marine Science*, 72(5):1297–1310.

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

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

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

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

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

WCPFC (2018). WCPFC-Scientific Committee Fourteenth Regular Session Summary Report. Technical report, Busan, South Korea, 8-16 August 2018.

WCPFC (2019). WCPFC-Scientific Committee Fifteenth Regular Session Summary Report. Technical report, Pohnpei, Federated States of Micronesia.

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

Williams, P., Reid, C., and Ruaia, T. (2020). Overview of tuna fisheries in the WCPO, including economic conditions – 2019. Technical Report WCPFC-SC16-2020/GN-IP-01.

9 Tables

Table 1: Fishery definitions and groupings for extraction fisheries used in the 2020 bigeye tuna stock assessment. The tag reporting rate groups for the Regional Tuna Tagging Program (RTTP), Pacific Tuna Tagging Program (PTTP), and the Japanese Tagging Program (JPTP) are also shown. Gears: DOM = artisanal gear types used in domestic fisheries; LL = longline; PL = pole and line; PS = purse seine unspecified; PS-ASS = associated purse seine; PS-UNA = unassociated purse seine. Flag/fleets: ALL = all nationalities; AU = Australia; ID = Indonesia; JP = Japan; OS = Offshore; PH = Philippines; US = United States; VN = Vietnam.

Fishery	Nationality	Gear	Region	Selectivity	Increasing	Recaptures	RTTP	PTTP	JPTP
F1 L-ALL-1	ALL	LL	1	1	Y	1	1	1	27
F2 L-ALL-2	ALL	LL	2	2	N	2	1	1	27
F3 L-US-2	US	LL	2	3	Y	3	2	2	28
F4 L-ALL-3	ALL	LL	3	4	N	4	1	1	27
F5 L-OS-3	OS	LL	3	8	Y	5	1	1	27
F6 L-OS-7	OS	LL	7	9	Y	6	1	1	27
F7 L-ALL-7	ALL	LL	7	10	N	7	1	1	27
F8 L-ALL-8	ALL	LL	8	11	Y	8	1	1	27
F9 L-ALL-4	ALL	LL	4	5	Y	9	1	1	27
F10 L-AU-5	AU	LL	5	12	Y	10	3	3	29
F11 L-ALL-5	ALL	LL	5	7	N	11	1	1	27
F12 L-ALL-6	ALL	LL	6	6	Y	12	1	1	27
F13 SA-ALL-3	ALL	PS-ASS	3	13	N	13	4	16	30
F14 SU-ALL-3	ALL	PS-UNA	3	16	N	13	4	16	30
F15 SA-ALL-4	ALL	PS-ASS	4	14	N	14	5	17	31
F16 SU-ALL-4	ALL	PS-UNA	4	17	N	14	5	17	31
F17 Z-PH-7	PH	Dom	7	19	N	15	6	18	32
F18 Z-ID.PH-7	ID.PH	Dom	7	20	N	15	6	18	32
F19 S-JP-1	JP	PS	1	21	N	16	7	19	33
F20 P-JP-1	JP	PL	1	22	N	17	8	20	34
F21 P-ALL-3	ALL	PL	3	23	N	18	9	21	35
F22 P-ALL-8	ALL	PL	8	24	N	19	10	22	36
F23 Z-ID-7	ID	Dom	7	25	N	15	6	18	32
F24 S-ID.PH-7	ID.PH	PS	7	19	N	15	6	18	32
F25 SA-ALL-8	ALL	PS-ASS	8	15	N	20	11	23	37
F26 SU-ALL-8	ALL	PS-UNA	8	18	N	20	11	23	37
F27 L-AU-9	AU	LL	9	12	Y	21	3	3	29
F28 P-ALL-7	ALL	PL	7	26	N	22	12	24	38
F29 L-ALL-9	ALL	LL	9	7	N	23	1	1	27
F30 SA-ALL-7	ALL	PS-ASS	7	13	N	24	13	25	39
F31 SU-ALL-7	ALL	PS-UNA	7	16	N	24	13	25	39
F32 Z-VN-7	VN	Dom	7	27	N	25	14	26	40

Table 2: Fishery definitions and groupings for the index fisheries used in the 2020 bigeye tuna stock assessment. The tag reporting rate groups for the Regional Tuna Tagging Program (RTTP), Pacific Tuna Tagging Program (PTTP), and the Japanese Tagging Program (JPTP) are also shown. Gears: I = index fishery. Flag/fleets: ALL = all nationalities.

Fishery	Nationality	Gear	Region	Selectivity	Increasing	Recaptures	RTTP	PTTP	JPTP
F33 I-I-1	ALL	I	1	28	Y	26	15	15	15
F34 I-I-2	ALL	I	2	28	Y	26	15	15	15
F35 I-I-3	ALL	I	3	28	Y	26	15	15	15
F36 I-I-4	ALL	I	4	28	Y	26	15	15	15
F37 I-I-5	ALL	I	5	28	Y	26	15	15	15
F38 I-I-6	ALL	I	6	28	Y	26	15	15	15
F39 I-I-7	ALL	I	7	28	Y	26	15	15	15
F40 I-I-8	ALL	I	8	28	Y	26	15	15	15
F41 I-I-9	ALL	I	9	28	Y	26	15	15	15

Table 3: Summary of the number of release events, tag releases and recoveries by region and program. Note that for the purposes of tag reporting rates, the CSTP and RTTP were considered to be the same program and are thus shown combined in this table.

Prog	JPTP			PTTP			RTTP		
Years	2000–2017			2006–2017			1989–2001		
Category	Grps	Rel	Rec	Grps	Rel	Rec	Grps	Rel	Rec
1	36	3699	426	0	0	0	0	0	0
2	3	50	1	0	0	0	0	0	0
3	0	0	0	13	2180	491	5	231	27
4	3	189	18	14	20344	6240	3	628	94
5	0	0	0	1	33	7	1	124	32
6	0	0	0	0	0	0	1	69	4
7	0	0	0	3	500	126	5	1207	375
8	0	0	0	14	2784	912	5	472	60
9	0	0	0	0	0	0	5	3976	443
Total	42	3938	445	45	25841	7776	25	6708	1035

Table 4: Description of symbols used in the yield and stock status analyses. For the purpose of this assessment, "recent" is the average over the period 2014–2017 for F related metrics and 2015–2018 for metrics related to SB while "latest" is 2018.

Symbol	Description
C_{latest}	Catch in the last year of the assessment (2018)
F_{recent}	Average fishing mortality-at-age for a recent period (2014–2017)
YF_{recent}	Equilibrium yield at average fishing mortality for a recent period (2014–2017)
fmult	Fishing mortality multiplier at maximum sustainable yield (MSY)
F_{MSY}	Fishing mortality-at-age producing the maximum sustainable yield (MSY)
MSY	Equilibrium yield at F_{MSY}
$F_{recent}/F_{\text{MSY}}$	Average fishing mortality-at-age for a recent period (2014–2017) relative to F_{MSY}
SB_0	Equilibrium unexploited spawning potential
SB_{latest}	Spawning potential in the latest time period (2018)
SB_{recent}	Spawning potential for a recent period (2015–2018)
$SB_{F=0}$	Average spawning potential predicted in the absence of fishing for the period 2008–2017
SB_{MSY}	Spawning potential that will produce the maximum sustainable yield (MSY)
$SB_{latest}/SB_{F=0}$	Spawning potential in the latest time period (2018) relative to the average spawning potential predicted to occur in the absence of fishing for the period 2008–2017
$SB_{latest}/SB_{\text{MSY}}$	Spawning potential in the latest time period (2018) relative to that which will produce the maximum sustainable yield (MSY)
$SB_{recent}/SB_{F=0}$	Spawning potential for a recent period (2015–2018) relative to the average spawning potential predicted to occur in the absence of fishing for the period 2008–2017
$20\%SB_{F=0}$	WCPFC adopted limit reference point – 20% of spawning potential in the absence of fishing averaged over years $t-10$ to $t-1$ (2008–2017)

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

Axis	Levels	Option
Steepness	3	0.65, 0.8* or 0.95
Natural mortality	2	Diagnostic* or M-hi
Size frequency weighting	4	sample sizes divided by 20*, 60, 200 or 500

Table 6: Summary of reference points over the 24 models in the structural uncertainty grid.

	Mean	Median	Min	10	90	Max
C_{latest}	159,738	159,288	157,297	157,722	162,033	162,271
YF_{recent}	136,568	134,940	117,800	124,668	149,424	161,520
fmult	1.45	1.38	0.83	0.98	2.03	2.33
F_{MSY}	0.05	0.05	0.04	0.04	0.07	0.07
MSY	146,715	140,720	117,920	125,628	179,164	187,520
$F_{recent}/F_{\text{MSY}}$	0.74	0.72	0.43	0.49	1.02	1.21
$SB_{F=0}$	1,395,173	1,353,367	903,708	982,103	1,780,138	1,908,636
SB_{MSY}	320,162	321,550	192,500	219,810	443,730	482,700
$SB_{MSY}/SB_{F=0}$	0.23	0.23	0.19	0.2	0.26	0.26
$SB_{latest}/SB_{F=0}$	0.38	0.38	0.23	0.3	0.47	0.51
$SB_{latest}/SB_{\text{MSY}}$	1.7	1.67	0.95	1.23	2.15	2.6
$SB_{recent}/SB_{F=0}$	0.4	0.41	0.21	0.27	0.52	0.55
$SB_{recent}/SB_{\text{MSY}}$	1.78	1.83	0.87	1.18	2.32	2.84

10 Figures

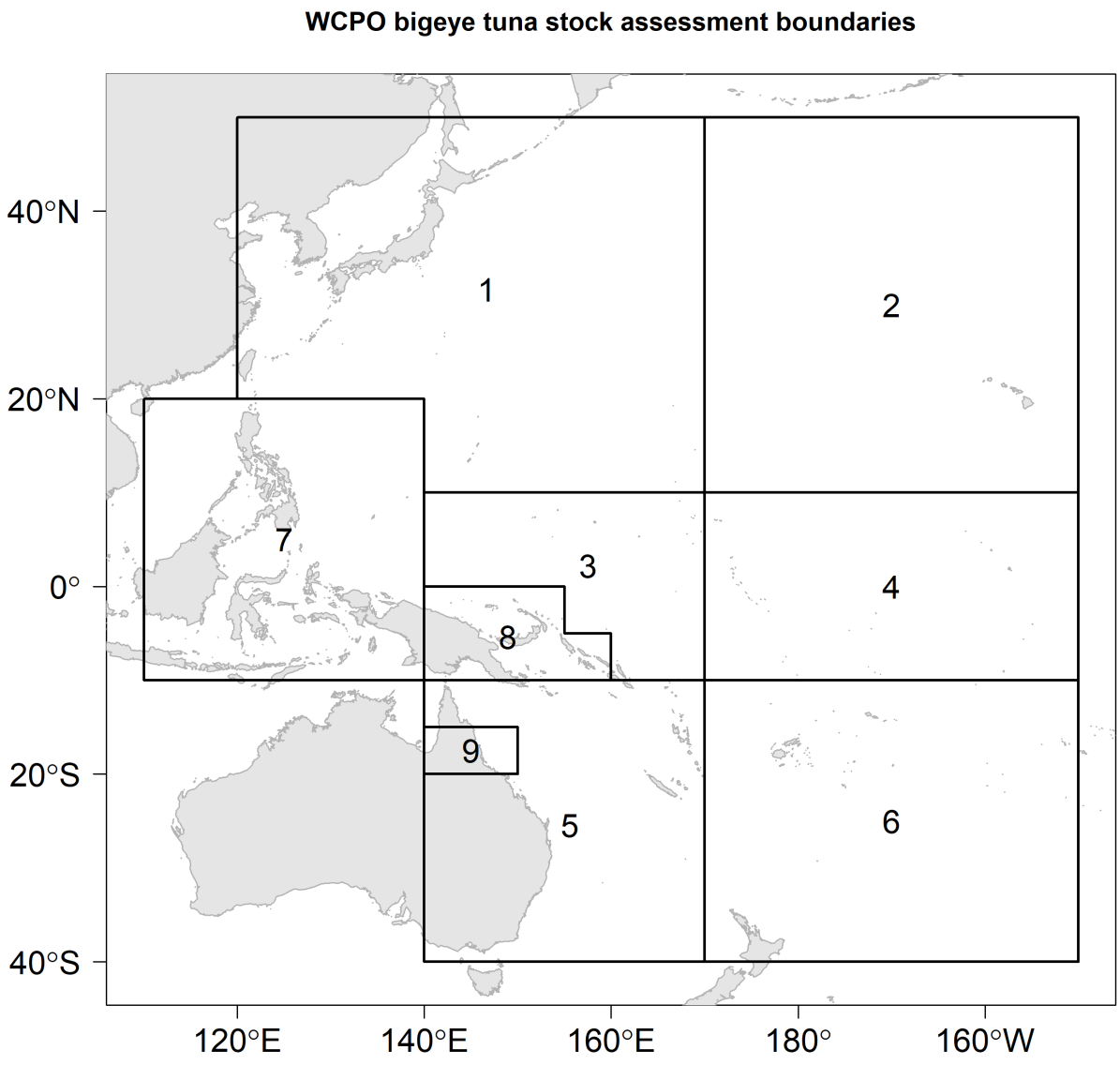


Figure 1: Spatial structure for the 2020 bigeye tuna stock assessment.

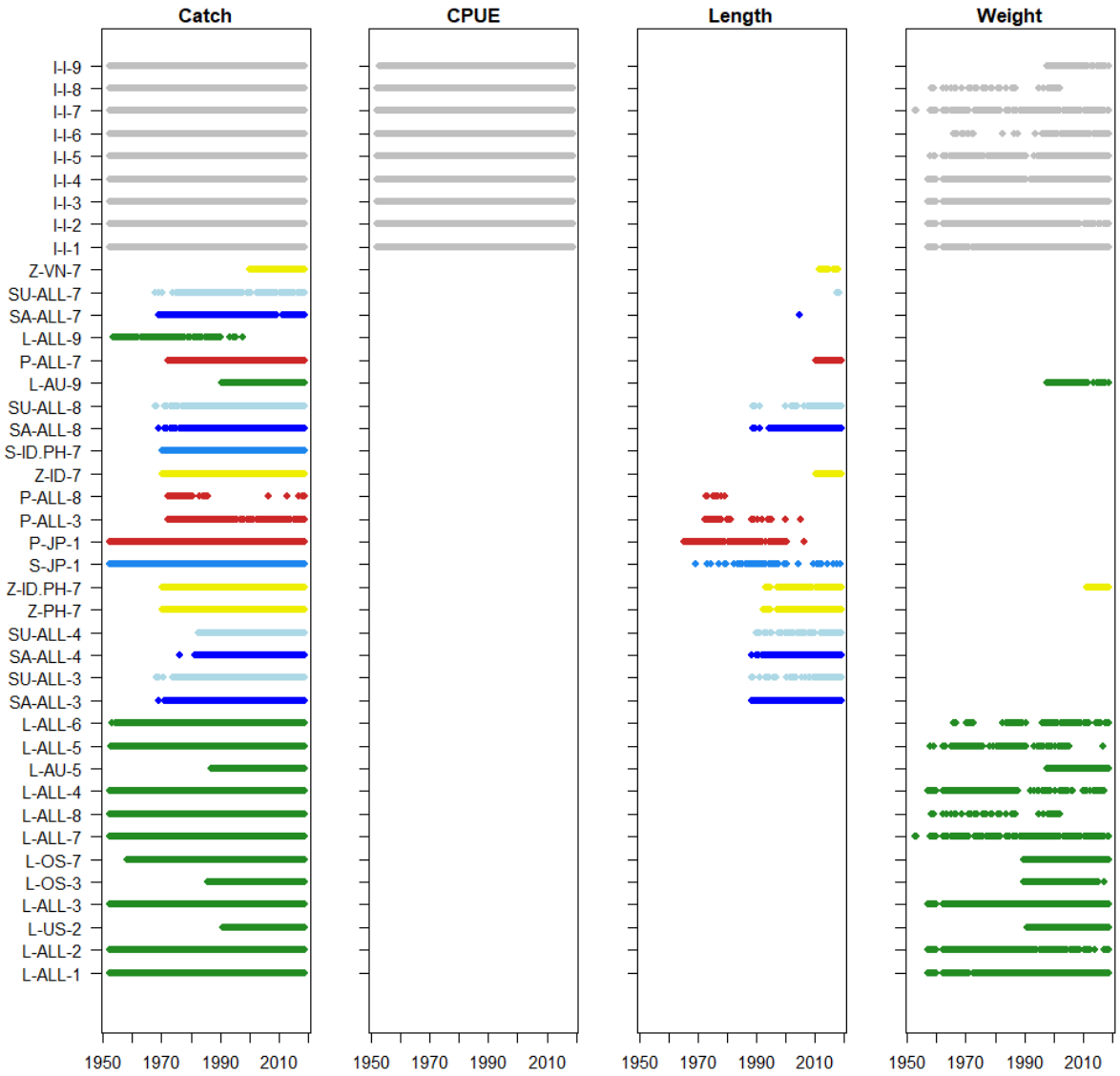


Figure 2: Presence of catch, standardised CPUE, length frequency and weight frequency data by year and fishery for the diagnostic case model. The different colors denote gear-type of the fishery: longline (green); pole-and-line (red); purse seine (blue); miscellaneous (yellow); and index (gray).

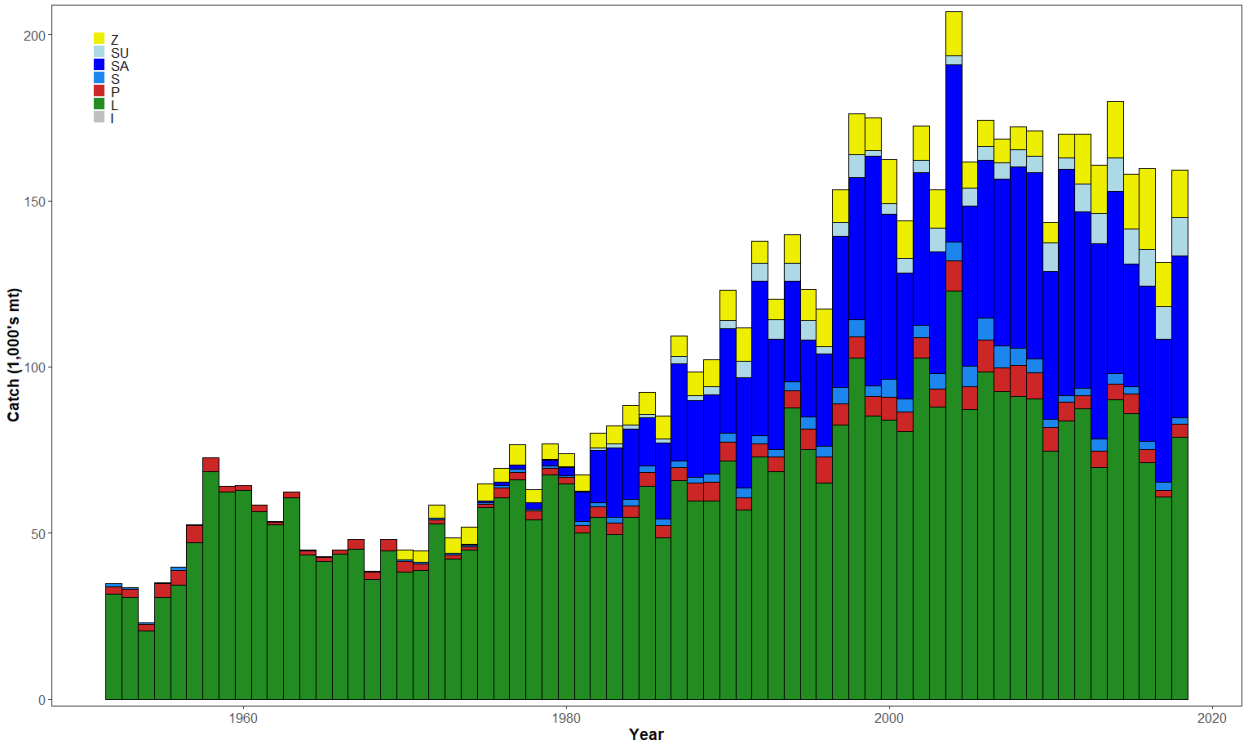


Figure 3: Time series of total annual catch (1000's mt) by fishing gear for the diagnostic case model over the full assessment period. The different colours refer to longline (green), pole-and-line (red), purse seine (blue), purse seine associated (dark blue), purse seine unassociated (light blue), miscellaneous (yellow), and index (gray). Note that the catch by longline gear has been converted into catch-in-weight from catch-in-numbers and so may differ from the annual catch estimates presented in (Williams et al., 2020), however these catches enter the model as catch-in-numbers.

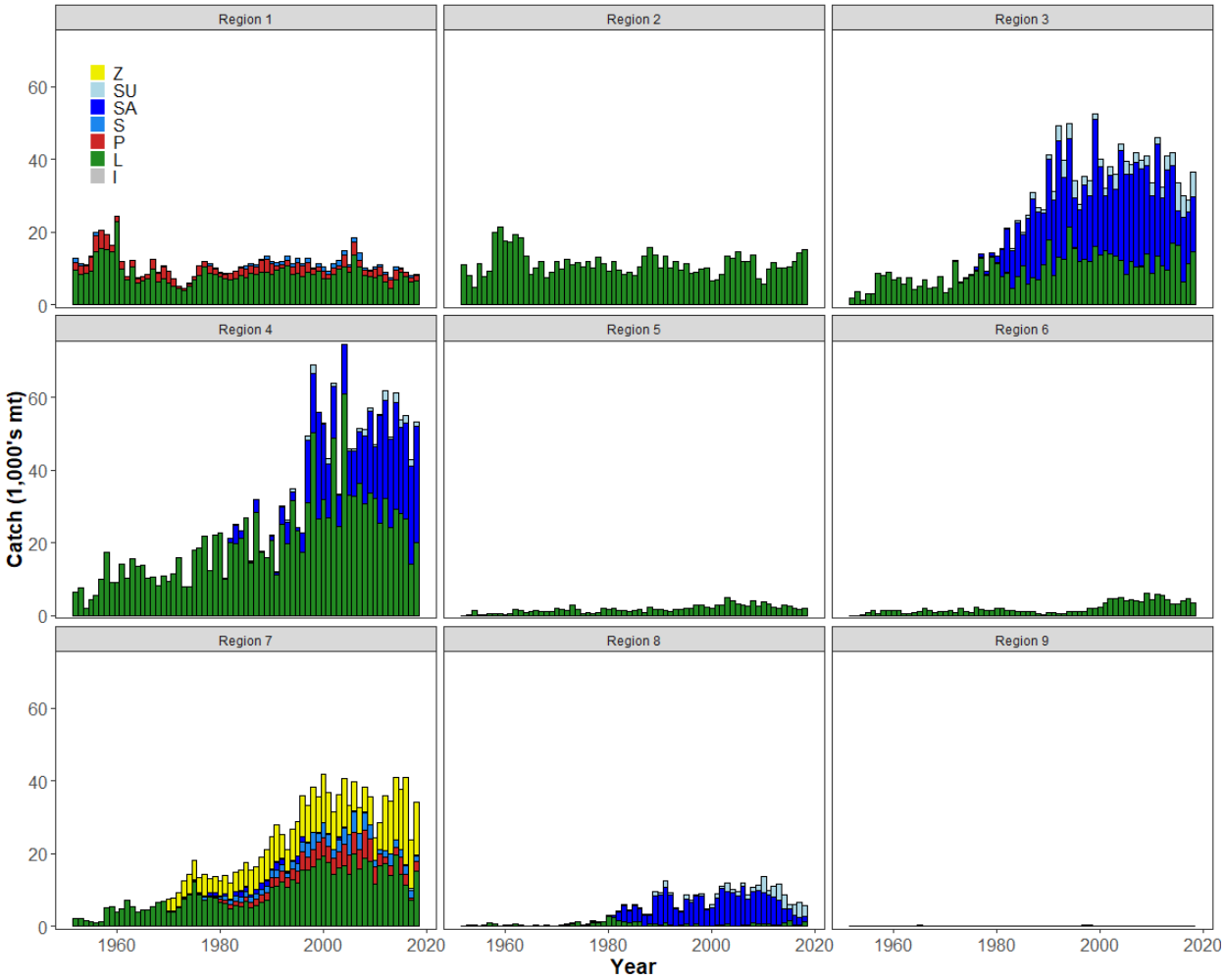


Figure 4: Time series of total annual catch (1000's mt) by fishing gear and assessment region for the diagnostic case model over the full assessment period. The different colours refer to longline (green), pole-and-line (red), purse seine (blue), purse seine associated (dark blue), purse seine unassociated (light blue), miscellaneous (yellow), and index (gray).

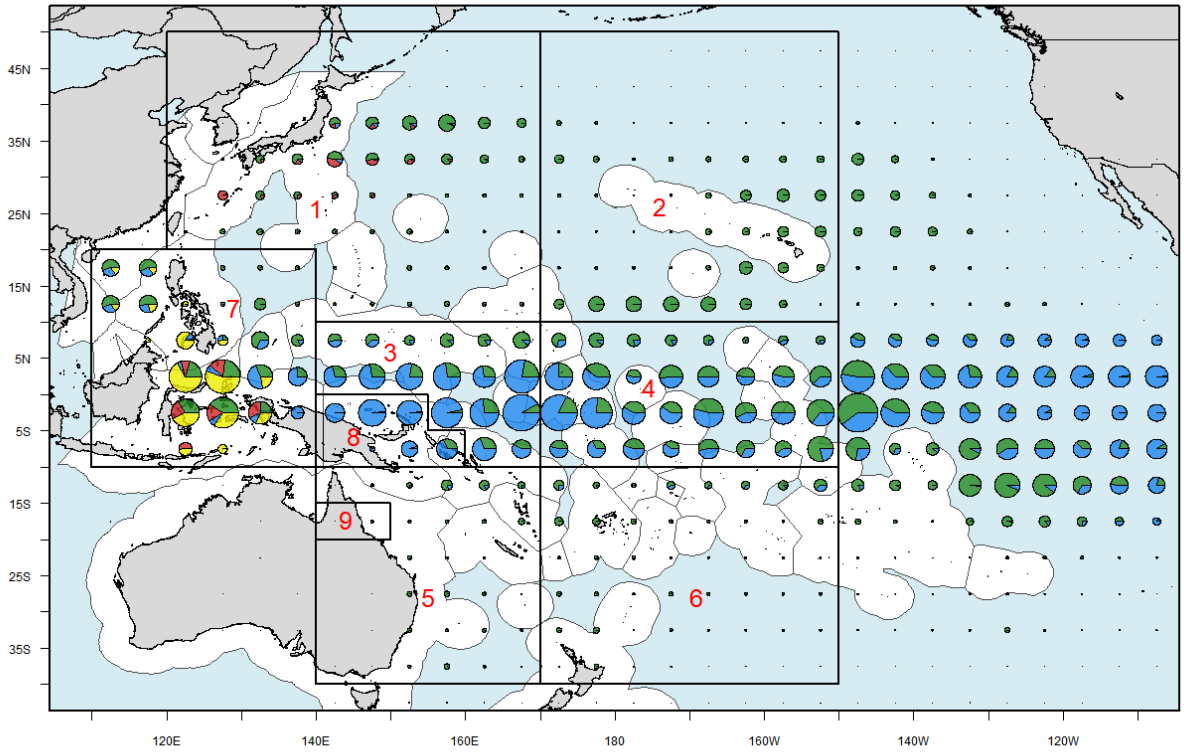


Figure 5: Distribution and magnitude of bigeye tuna catches for the most recent decade of the stock assessment (2009-2018) by 5° square and fishing gear: longline (green), pole-and-line (red), purse seine (blue) and miscellaneous (yellow), for the WCPO and part of the EPO. Overlayed are the regional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

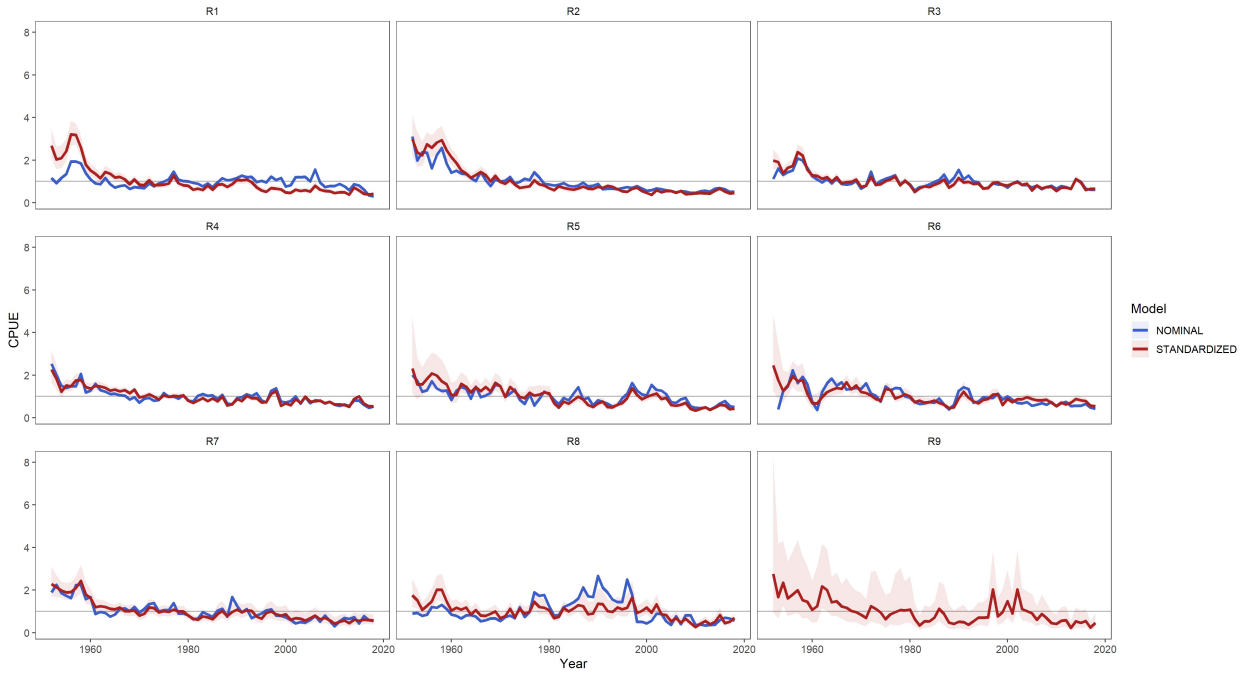


Figure 6: Standardized catch-per-unit-effort (CPUE) indices for the index fisheries in regions 1–9 used in the diagnostic case model (red) relative to the nominal CPUE (blue). See Ducharme-Barth and Vincent (2020) for further details of the estimation of these CPUE indices. The shaded polygon represents the 95% confidence intervals derived from the effort deviation penalties used in the diagnostic case model. Note that these indices have been rescaled to a mean of 1 within region and do not show the associated regional scaling.

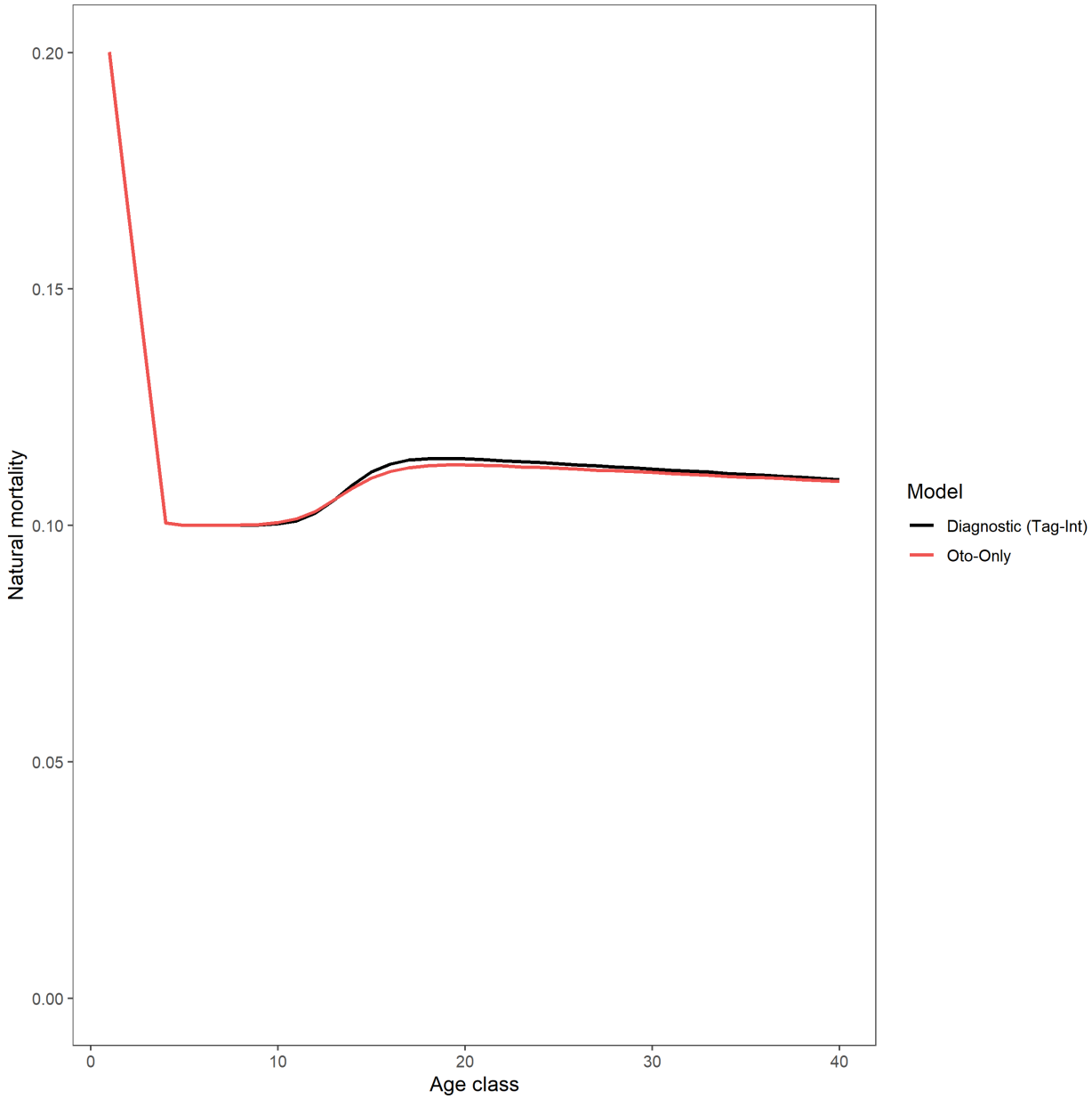


Figure 7: Quarterly natural mortality-at-age as used the diagnostic case model (*Tag-Int*; black line) and the sensitivity model *Oto-only* (red line).

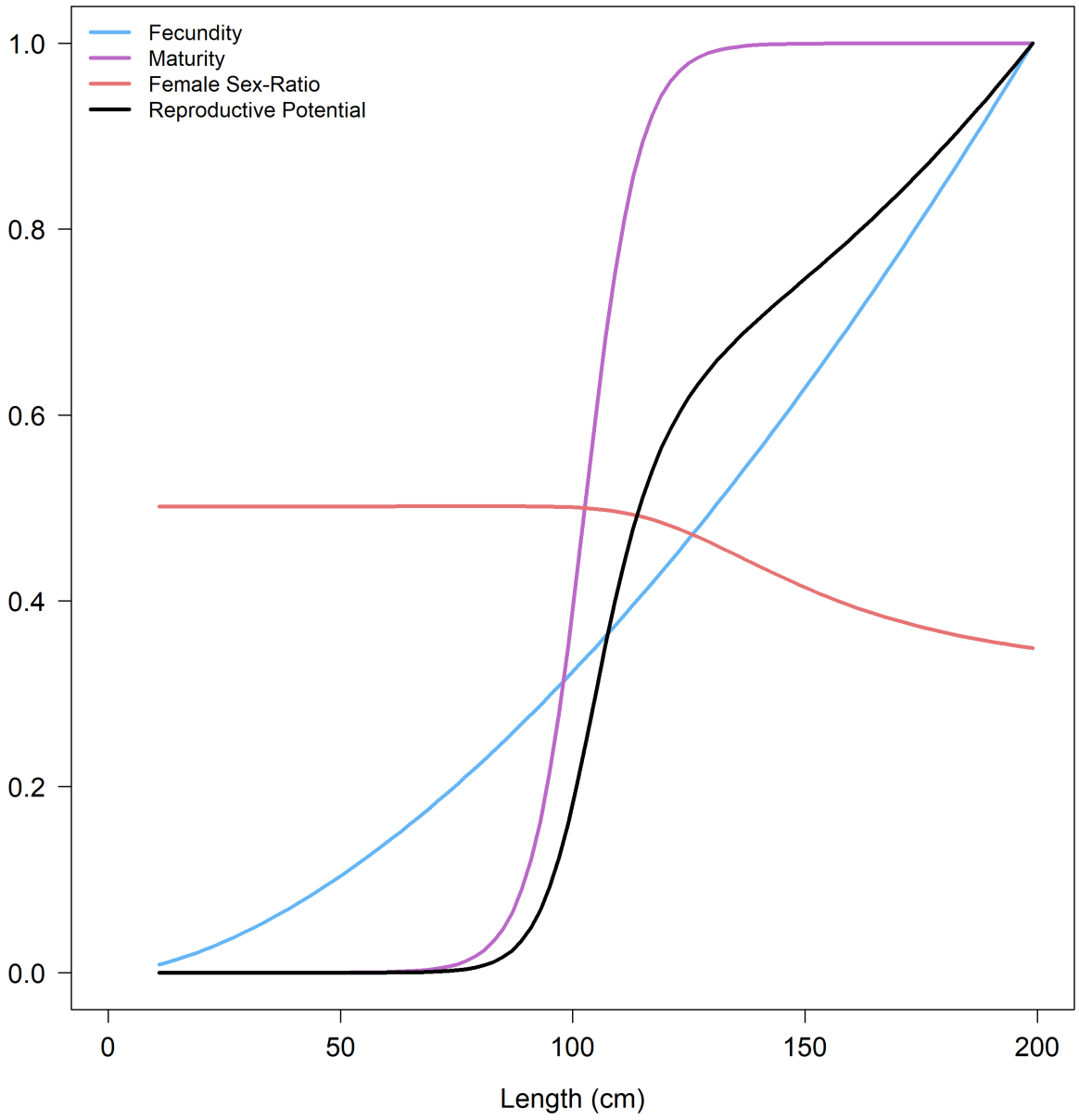


Figure 8: Individual components of reproductive potential-at-length (relative fecundity–blue;female maturity–purple; and female sex-ratio–red) and the resulting reproductive potential-at-length ogive (black).

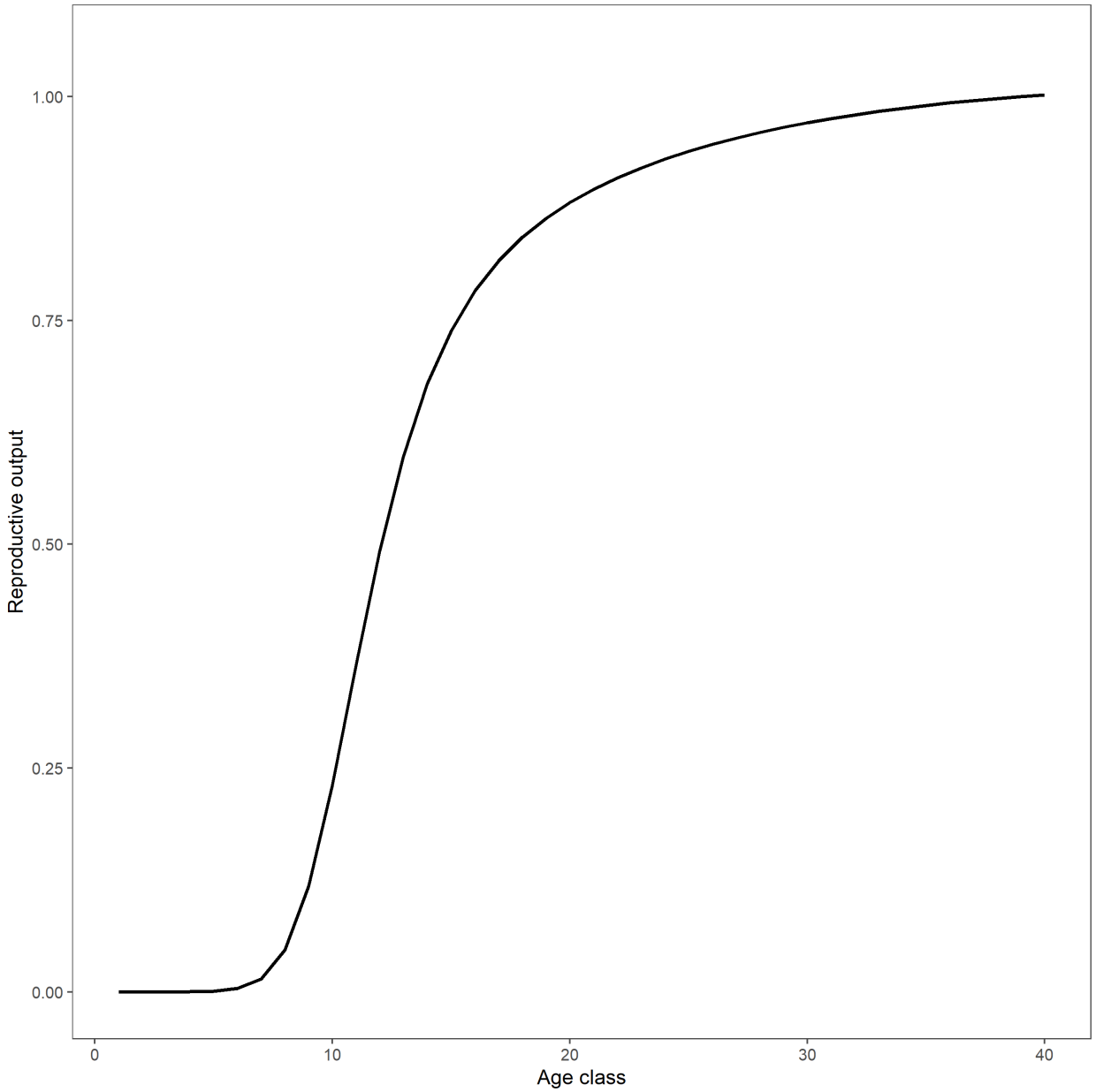


Figure 9: Reproductive potential-at-age converted from reproductive potential-at-length (Figure 8) within the diagnostic case.

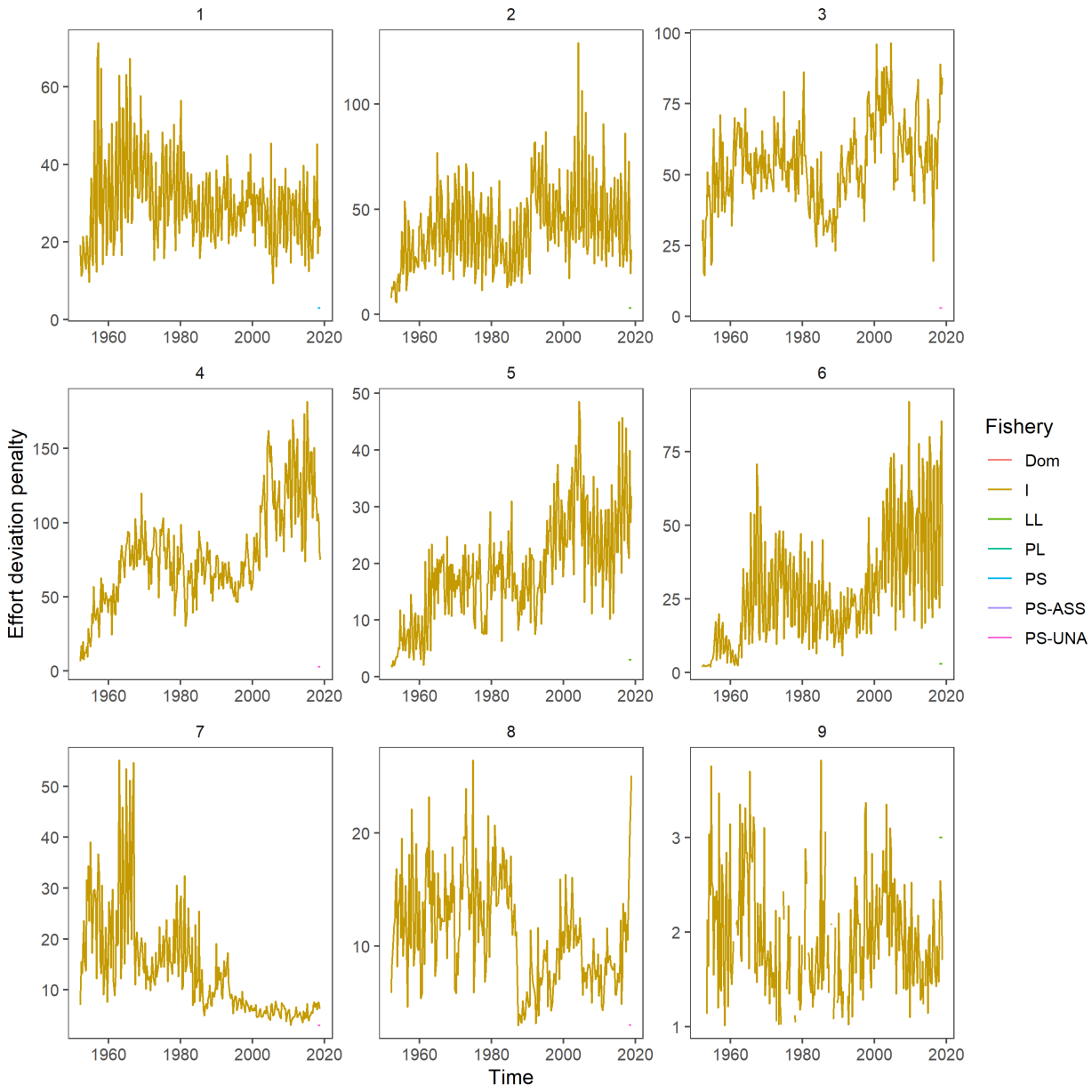
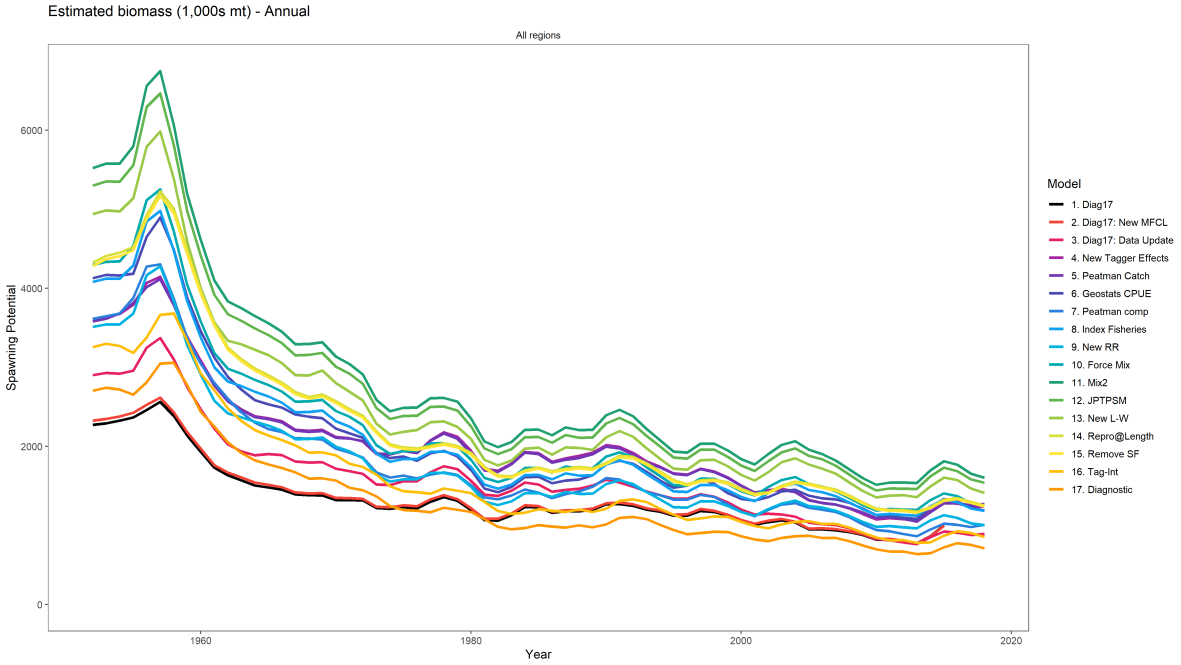
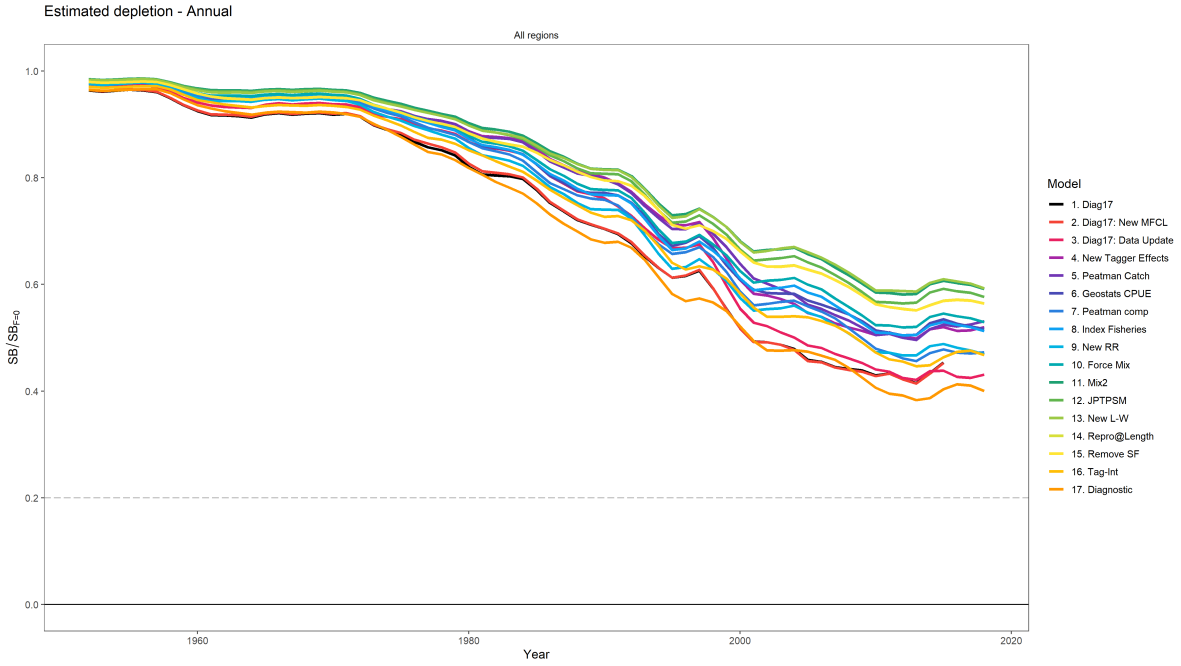


Figure 10: Plot of the effort deviation penalties applied to each fishery, by region, with the colors of the lines representing the gear of the fishery. Only penalties for the index fisheries are shown since effort was removed from all other fisheries. A higher penalty gives more weight to the CPUE of that fishery. Note that the y-axis scales are different between panels.



(a) Progression from 2017 to 2020 diagnostic model - spawning potential



(b) Progression from 2017 to 2020 diagnostic model - depletion

Figure 11: Stepwise changes in spawning potential (a), and fishing depletion (b) from the 2017 reference case model through to the 2020 diagnostic case model.

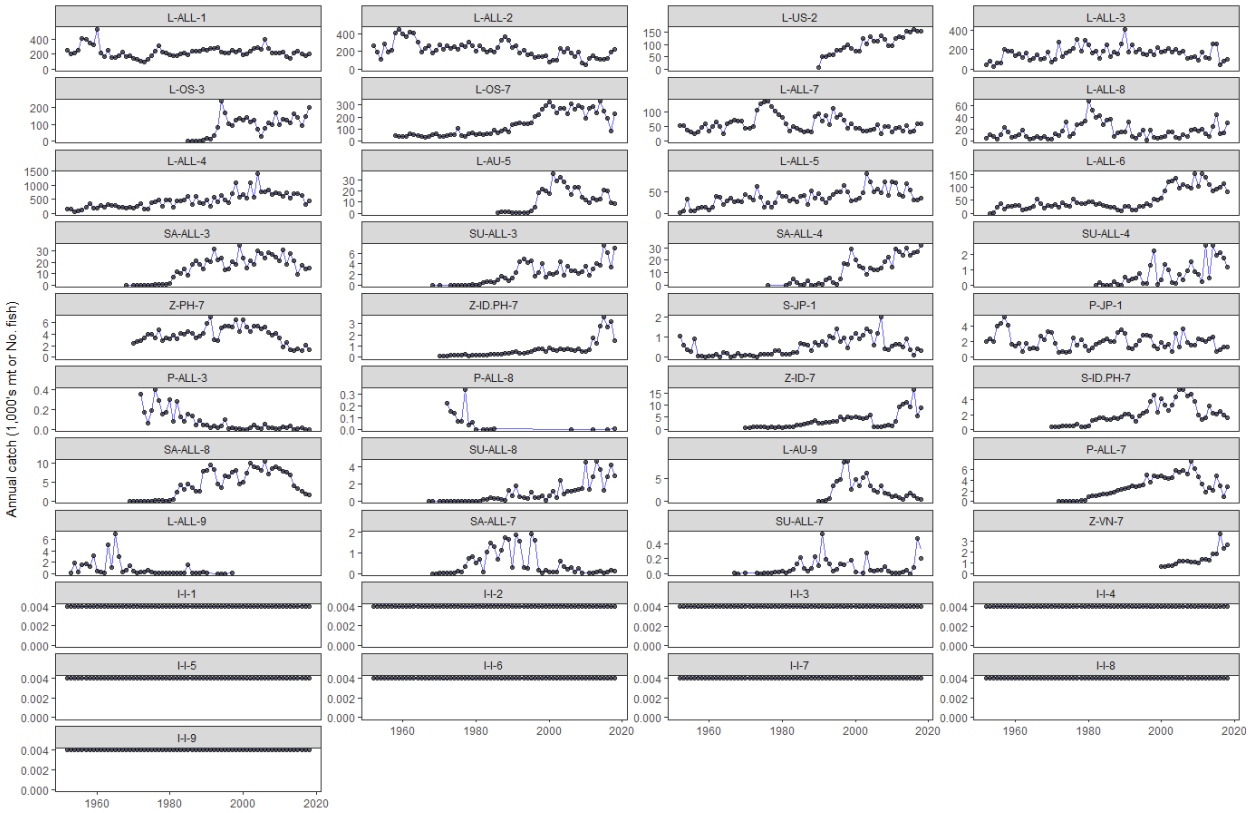


Figure 12: Observed (black points) and model-predicted (blue lines) catch for the 41 fisheries in the diagnostic case model. The y-axis is in catch-in-numbers for the longline fisheries and catch-in-weight for the other fisheries, both divided by 1,000.

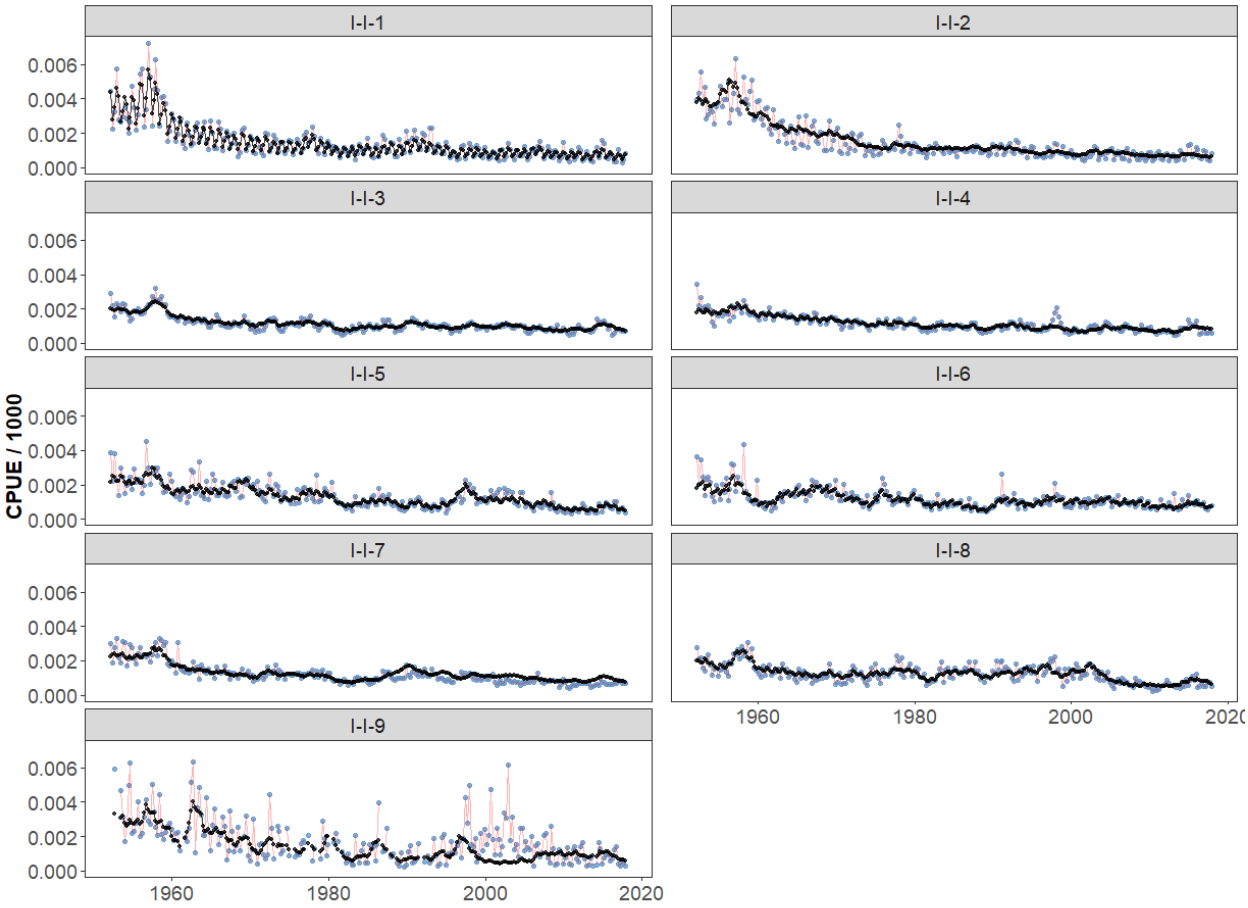


Figure 13: Observed (blue points and red lines) and model-predicted (black points and lines) CPUE for the nine index fisheries which received the standardized CPUE.

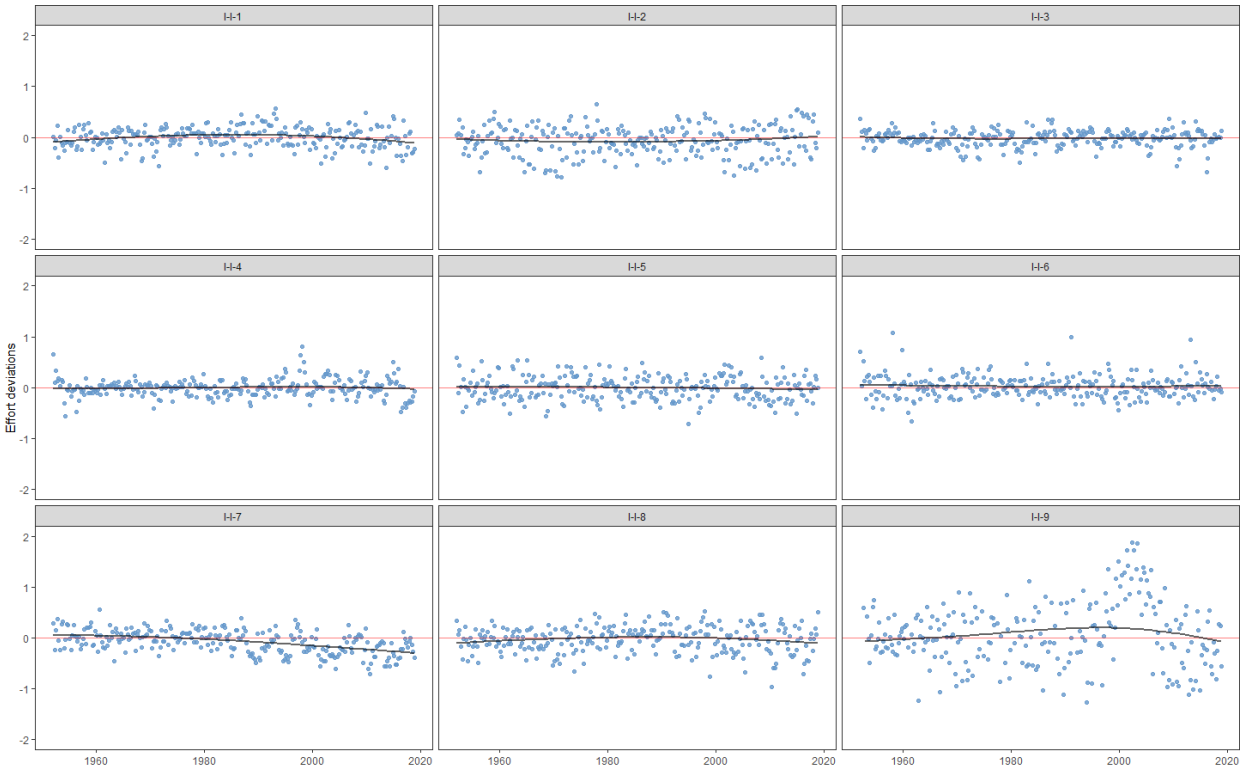


Figure 14: Effort deviations by time period for each of the fisheries receiving standardised CPUE indices in the diagnostic case model. The dark line represents a lowess smoothed fit to the effort deviations.

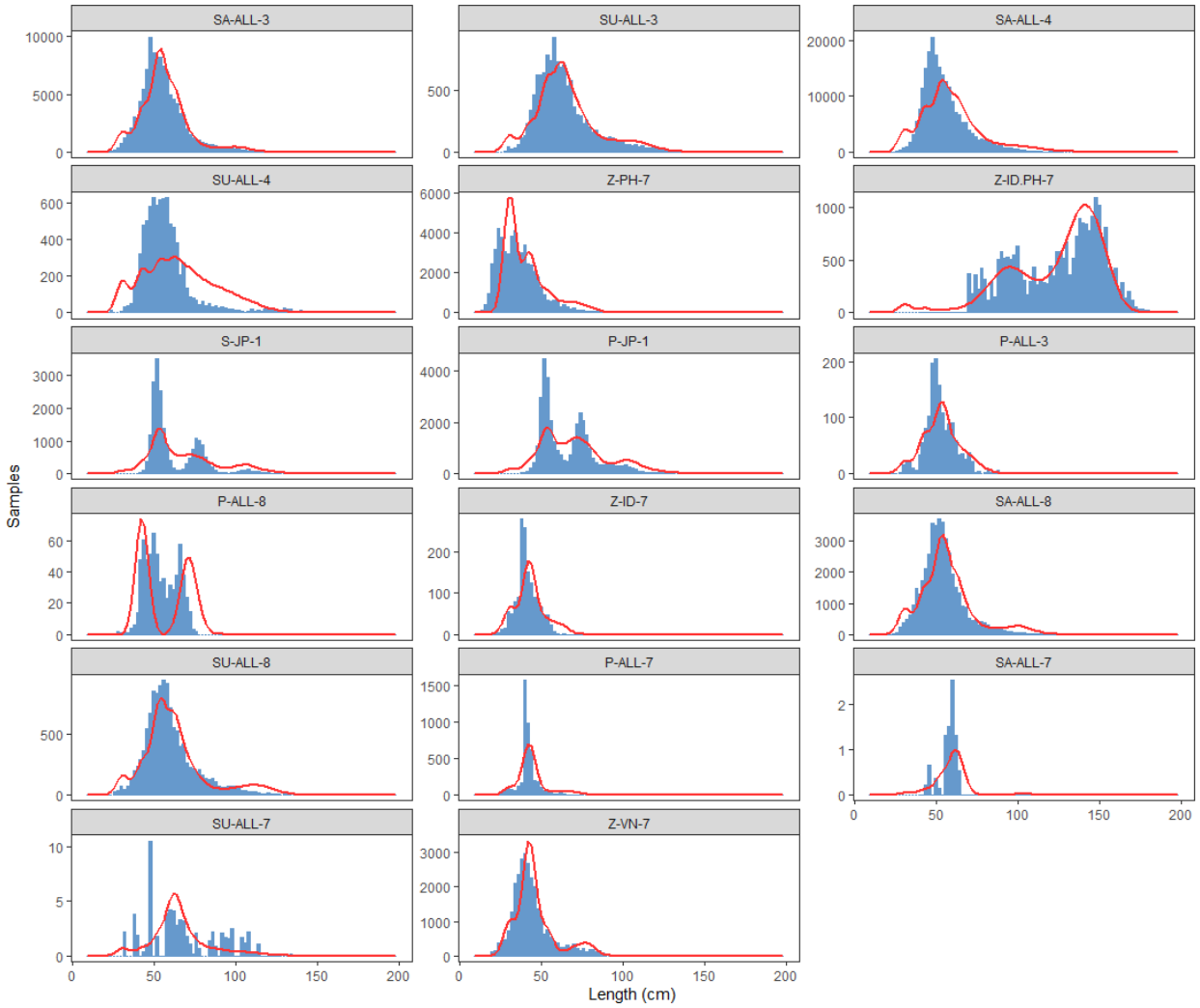


Figure 15: Composite (all time periods combined) observed (histograms) and predicted (lines) catch-at-length for all fisheries with samples for the diagnostic case model.

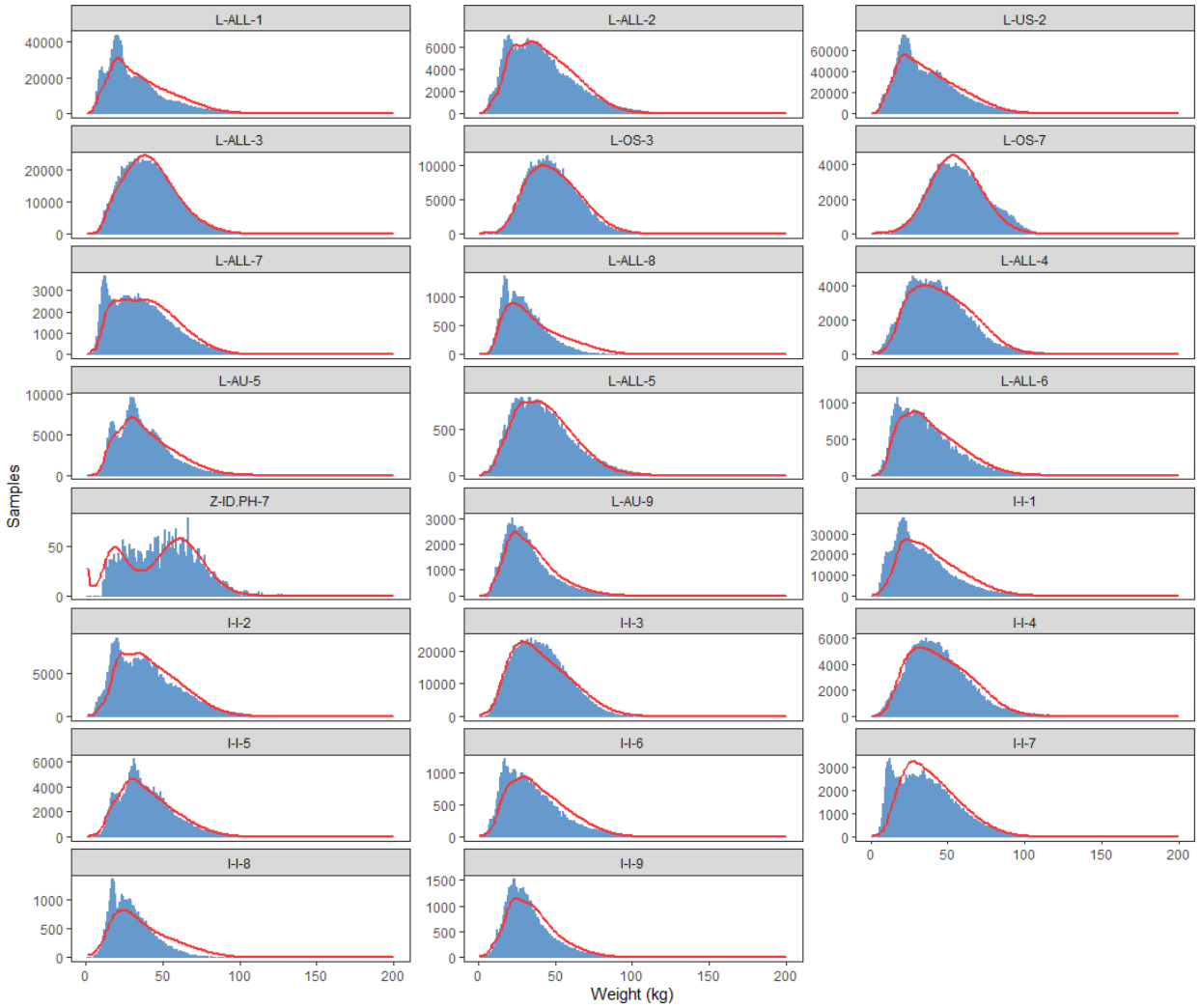


Figure 16: Composite (all time periods combined) observed (histograms) and predicted (lines) catch-at-weight for all fisheries with samples for the diagnostic case model.

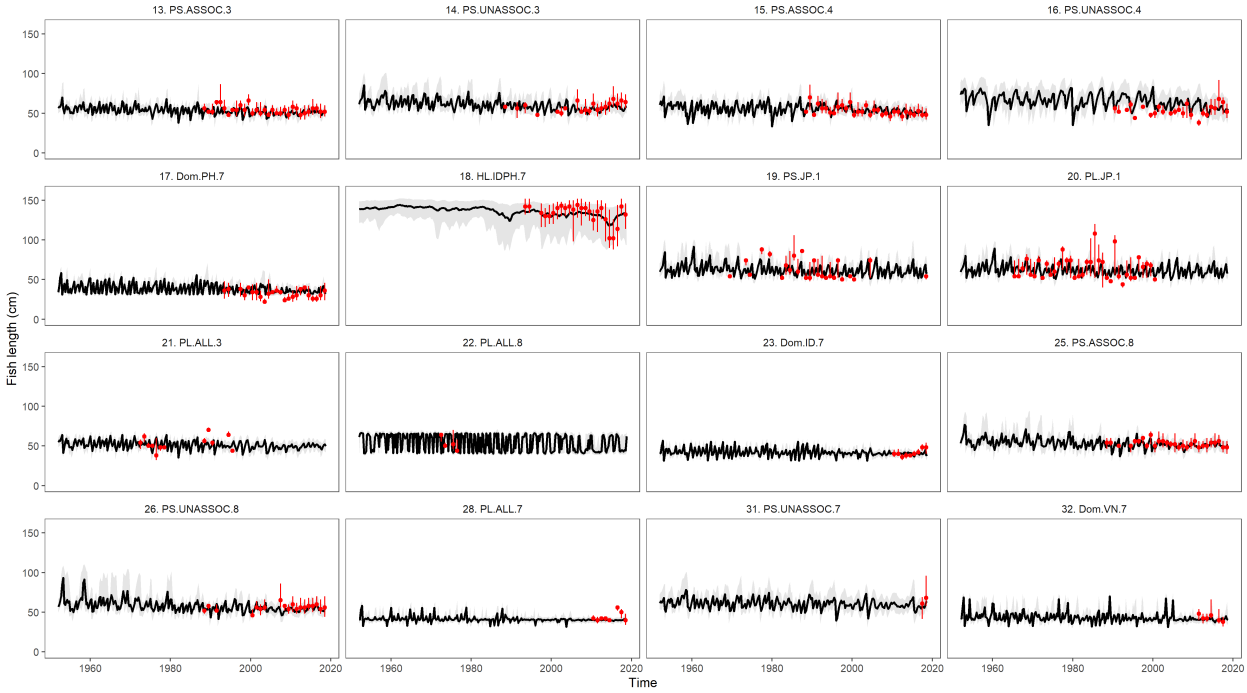


Figure 17: A comparison of the observed (red points) and predicted (grey line) median fish length (FL, cm) for all fisheries with samples for the diagnostic case model. The uncertainty intervals (grey shading) represent the values encompassed by the 25% and 75% quantiles. Sampling data are aggregated by year and only length samples with a minimum of 30 fish per year are plotted.

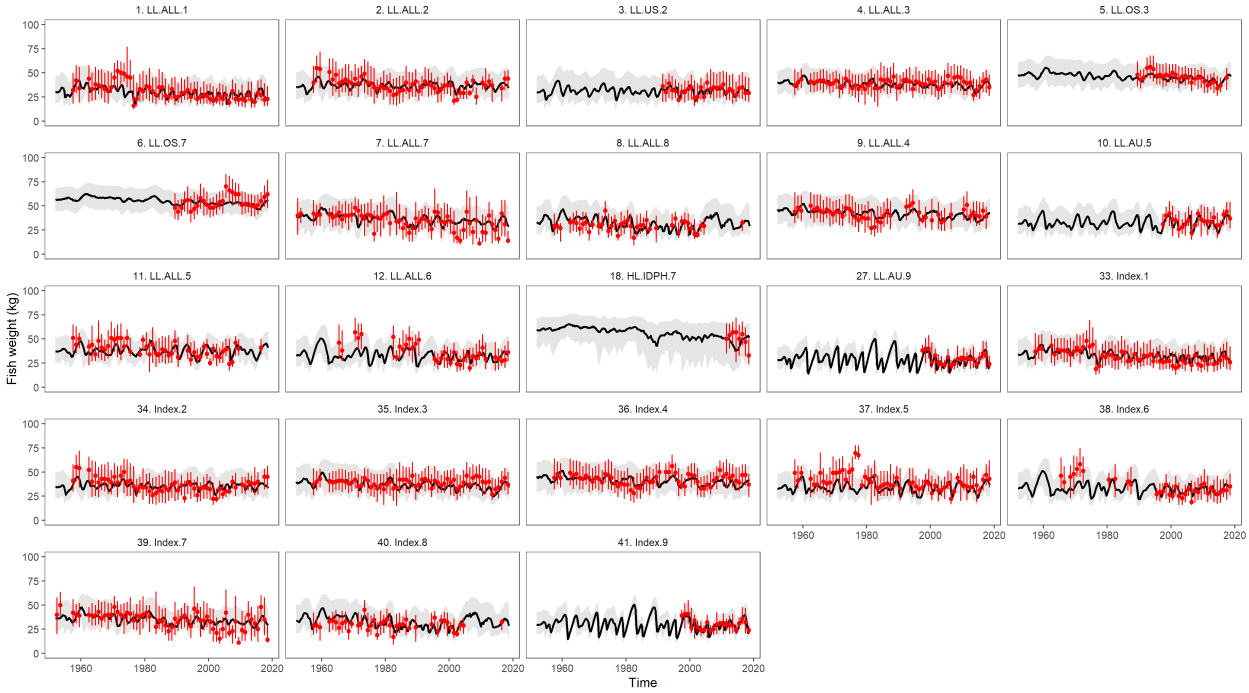
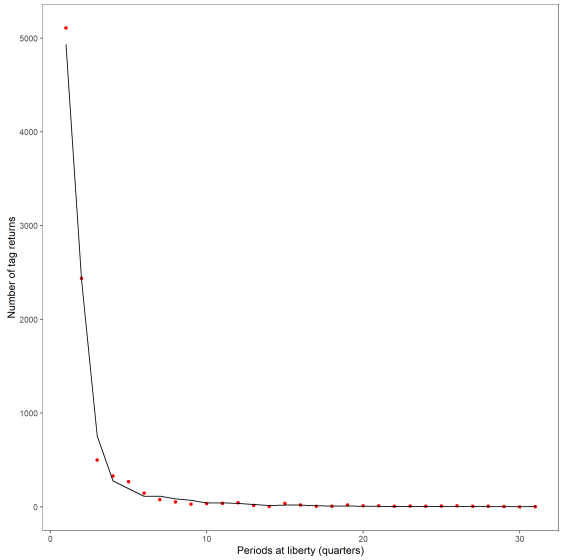
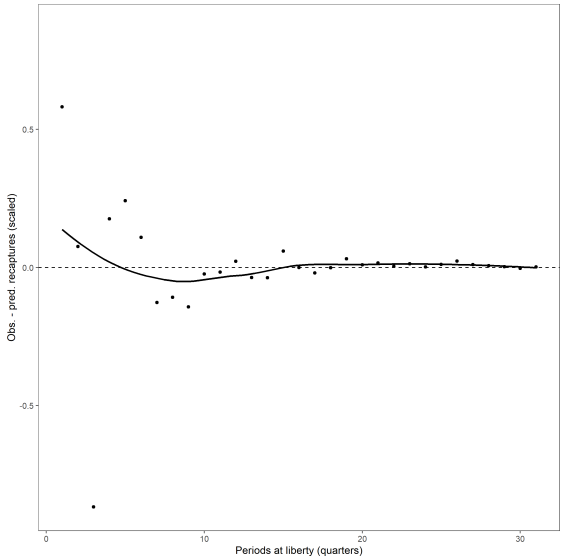


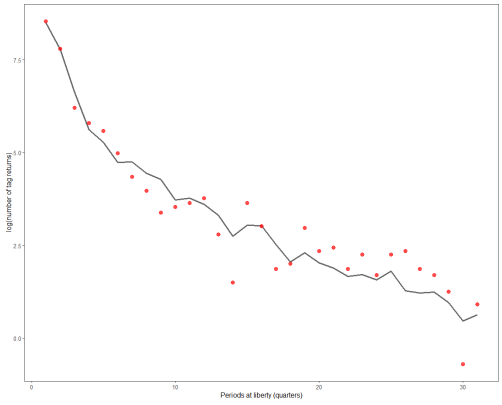
Figure 18: A comparison of the observed (red points) and predicted (grey line) median fish weight (WW, kg) for all fisheries with samples for the diagnostic case model. The uncertainty intervals (grey shading) represent the values encompassed by the 25% and 75% quantiles. Sampling data are aggregated by year and only weight samples with a minimum of 30 fish per year are plotted.



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

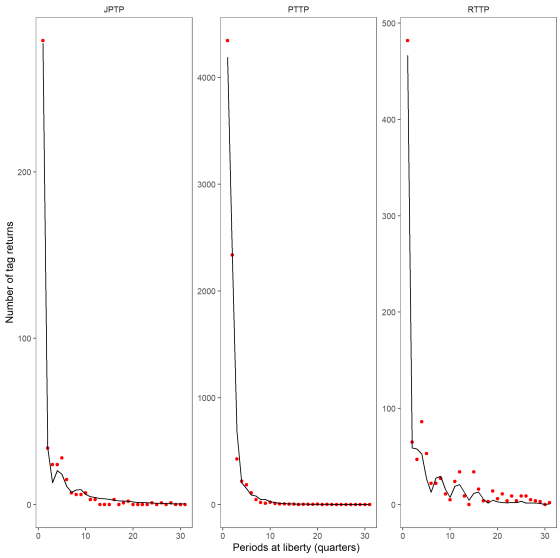


(b) Scaled residuals.

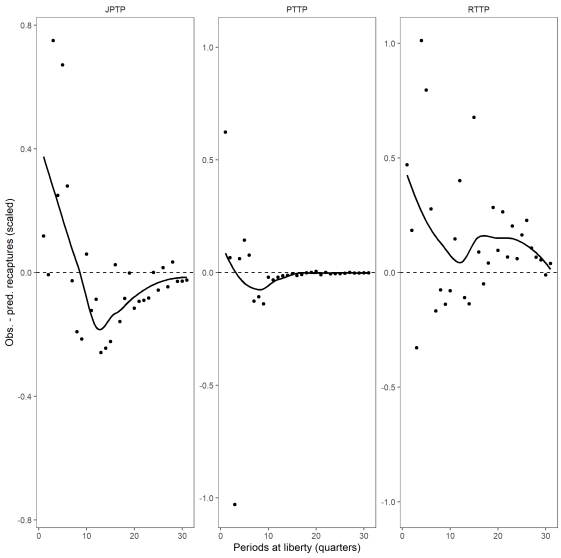


(c) Observed and model-predicted tag attrition across release events. (log-scale)

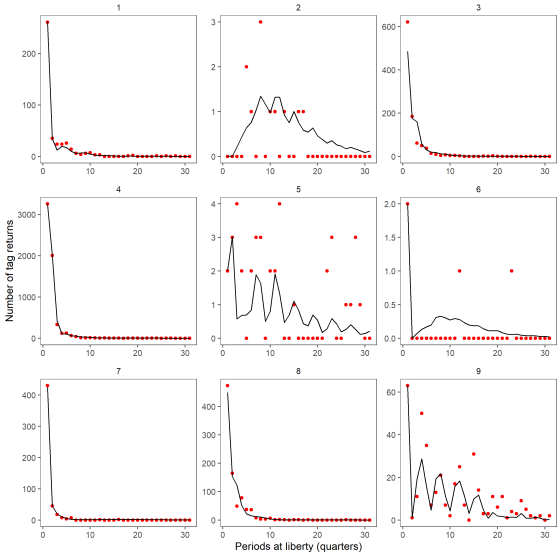
Figure 19: Observed and model-predicted (a) tag attrition, (b) scaled residuals, and (c) log-scale across all tag release events for the diagnostic case model.



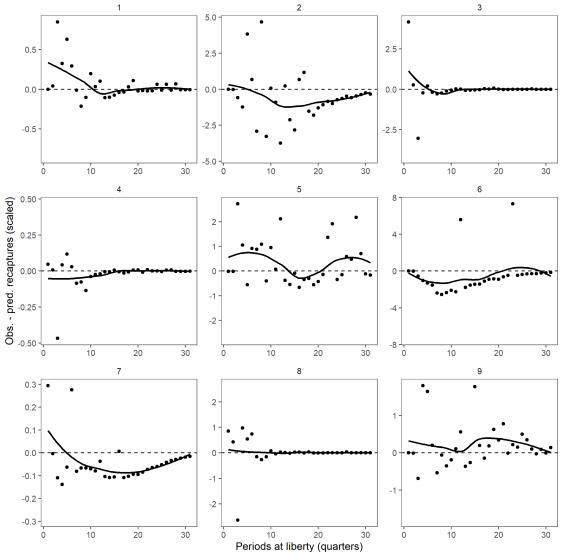
(a) Observed and model-predicted tag attrition across release events by program.



(b) Scaled residuals by program.

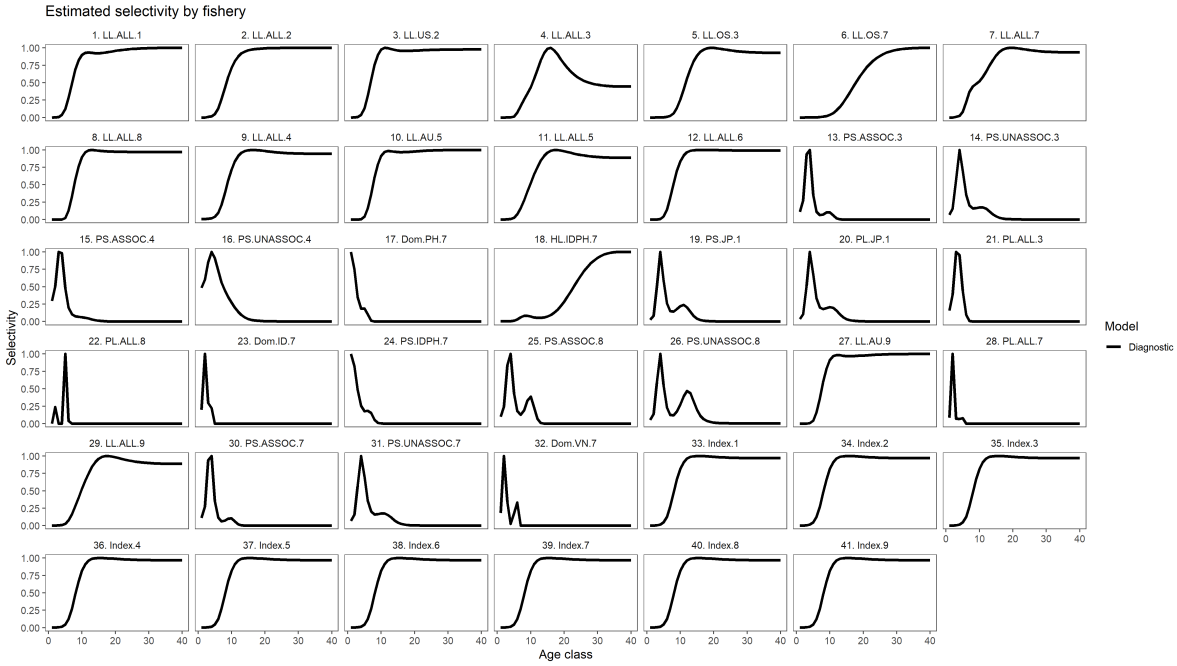


(c) Observed and model-predicted tag attrition across release events by region.

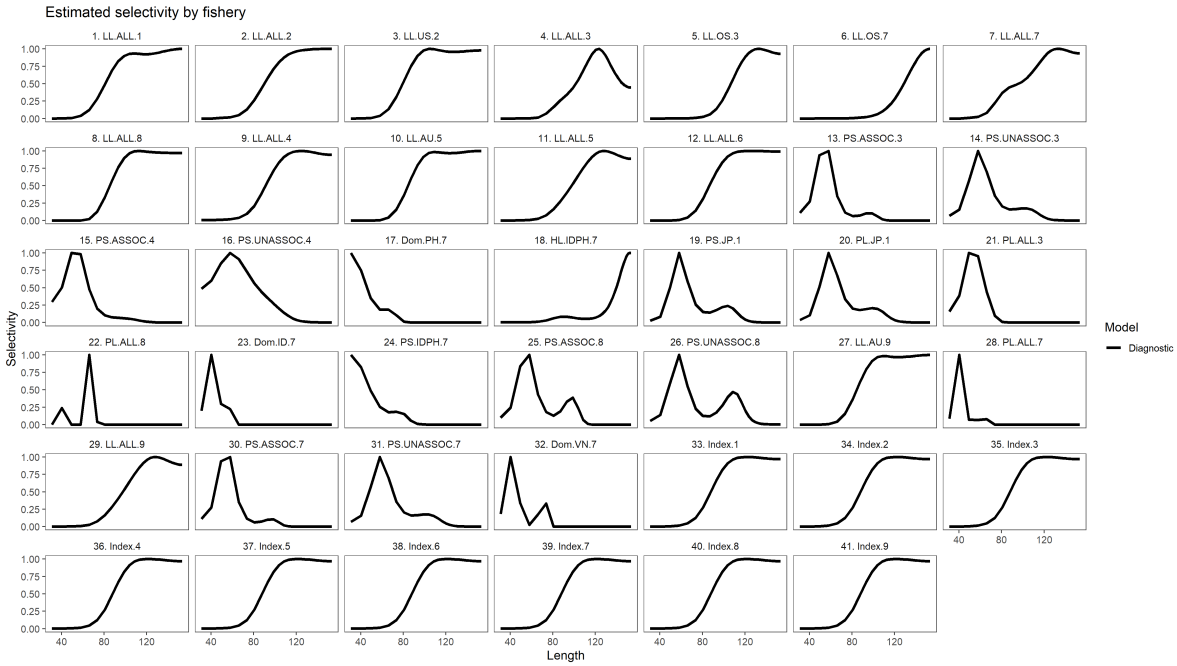


(d) Scaled residuals by region.

Figure 20: Observed and model-predicted tag attrition by (a) program, (b) program scaled residuals, (c) by region and (d) region residuals across all tag release events for the diagnostic case model. Note the differences in y-axis scale by panel.



(a) Selectivity-at-age.



(b) Selectivity-at-length.

Figure 21: Estimated (a) age-specific and (b) length-specific selectivity curves by fishery for the diagnostic case model.

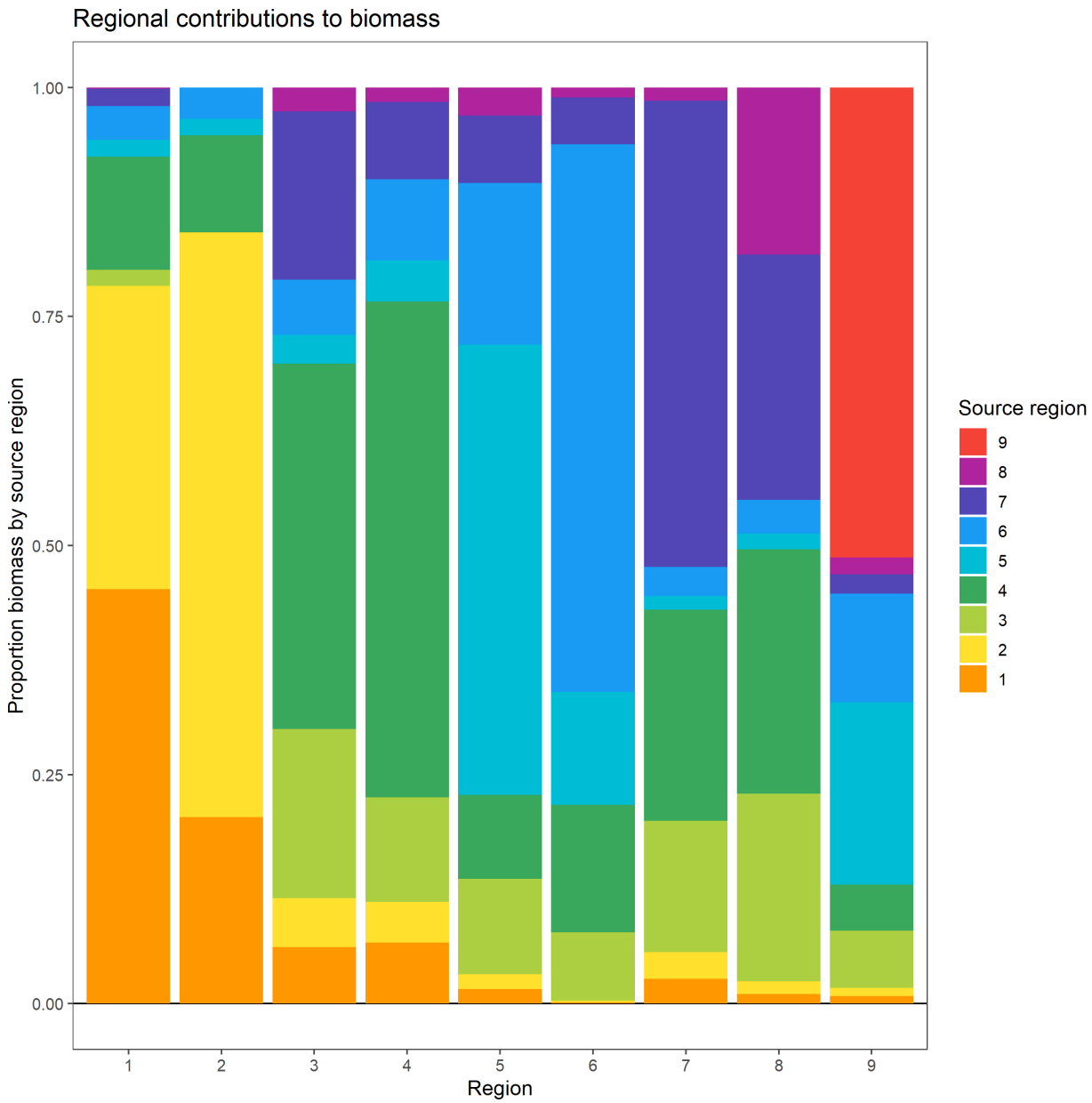


Figure 22: Proportional distribution of total biomass (by weight) in each region apportioned by the source region of the fish, for the diagnostic case model. The color of the source region is presented next to the figure. The biomass distributions are calculated based on the long-term average distribution of recruitment between regions, estimated movement parameters, and natural mortality.

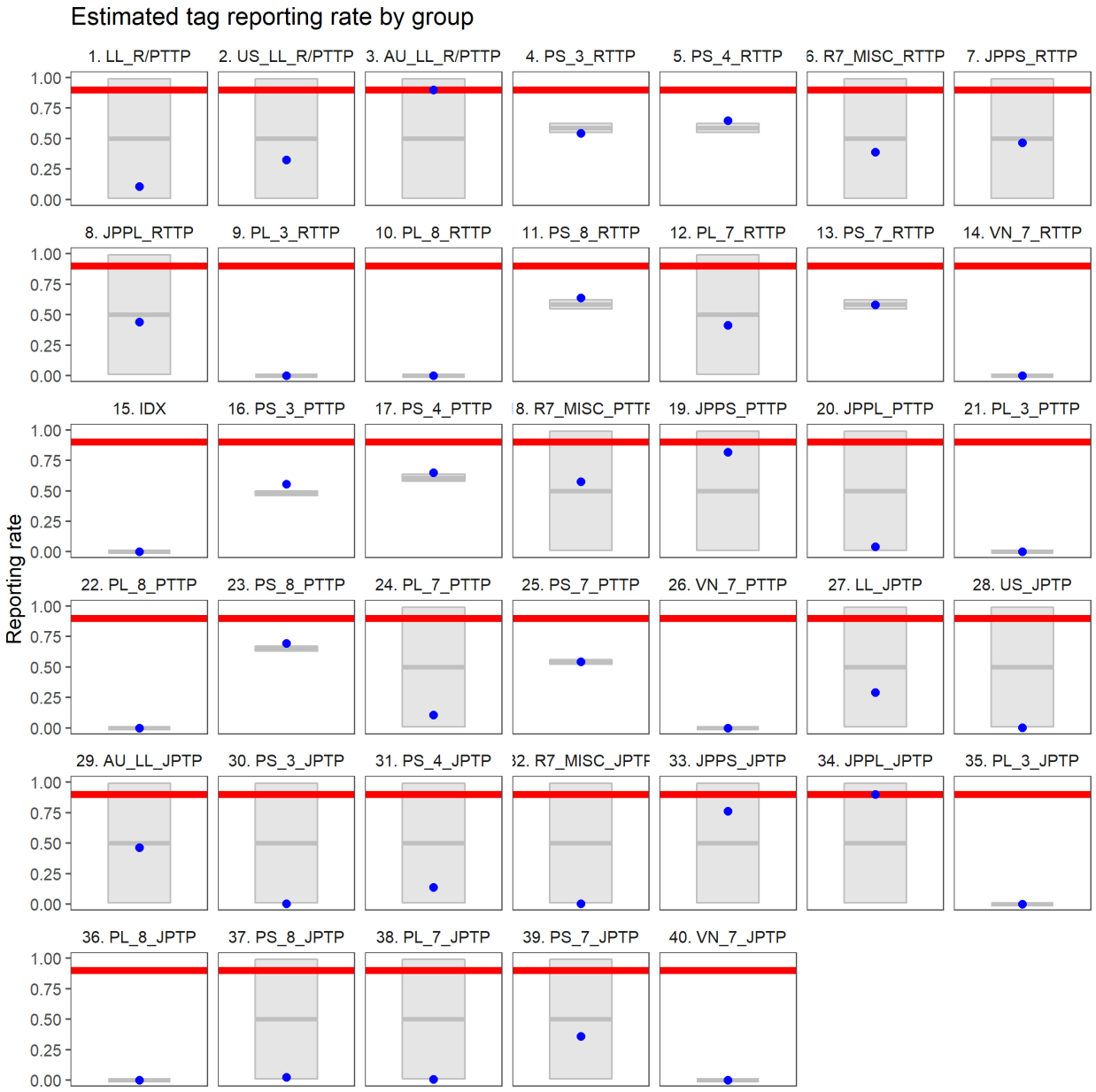


Figure 23: Estimated reporting rates (blue) for the diagnostic case model. The "prior" is shown with the gray bar, and the upper bound by the red line.

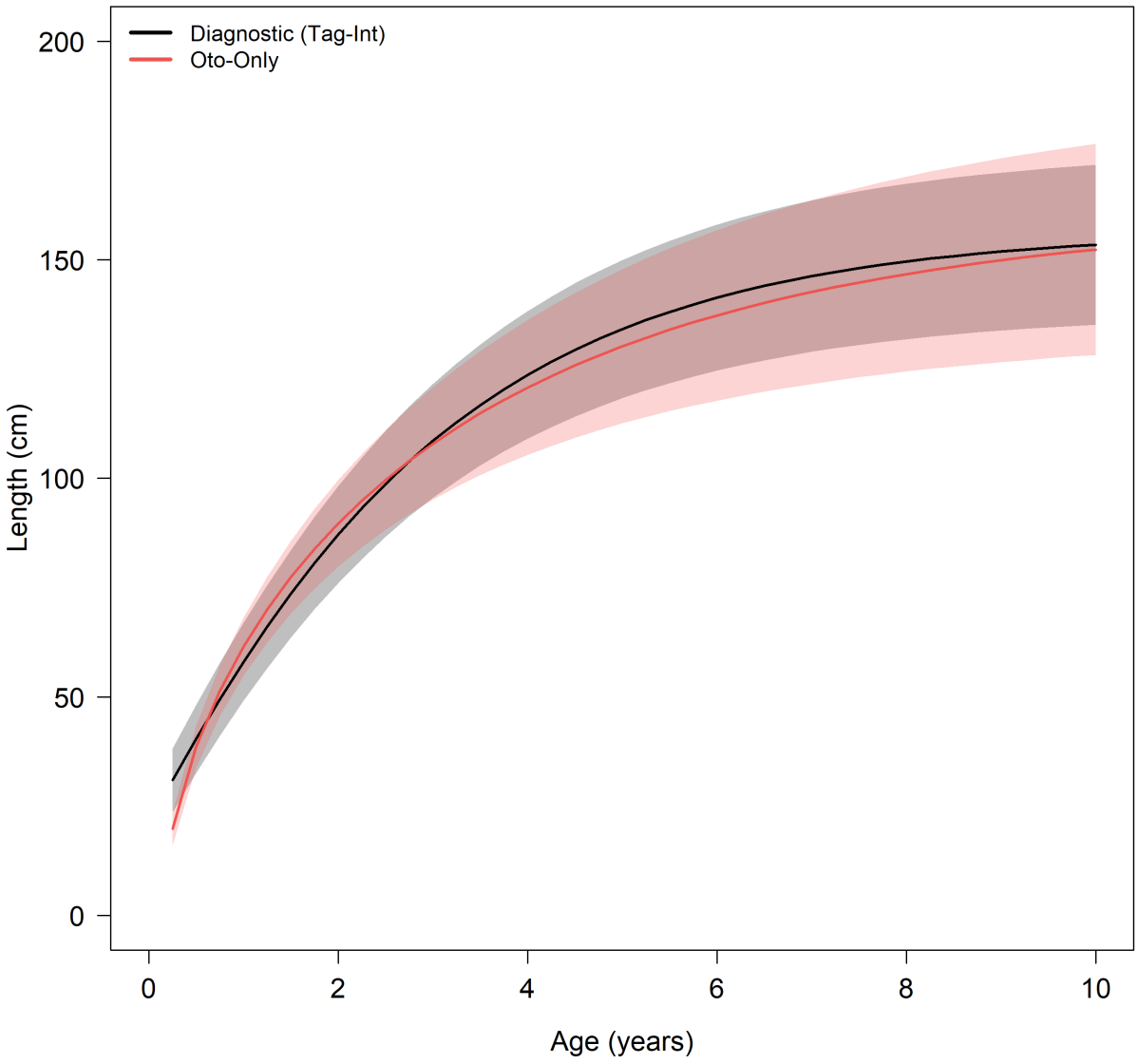


Figure 24: The two fixed growth curves and estimated variability considered in the one-off sensitivities. The uncertainty shown as 2 standard deviations away from the mean is shown by the corresponding shaded region.

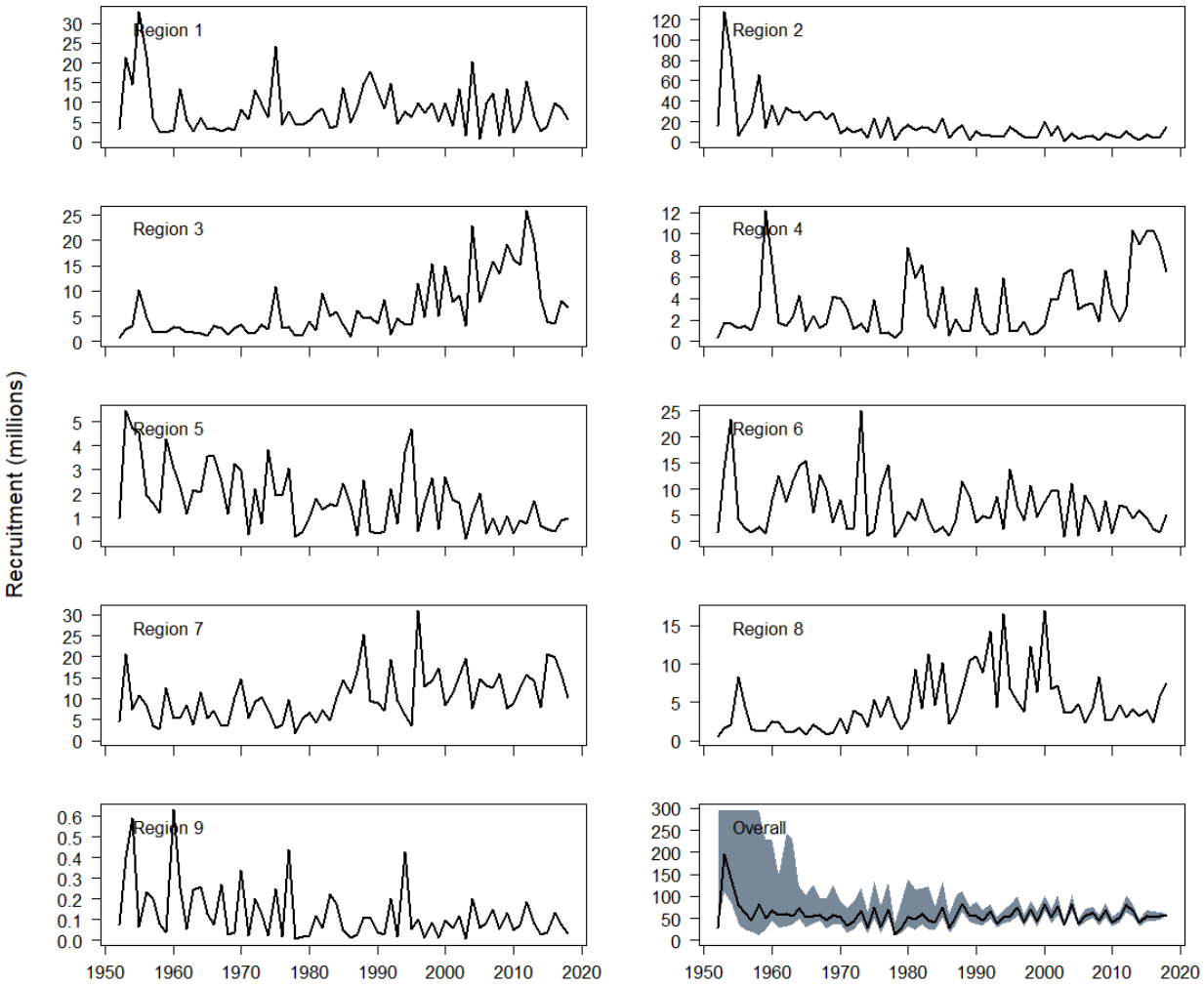
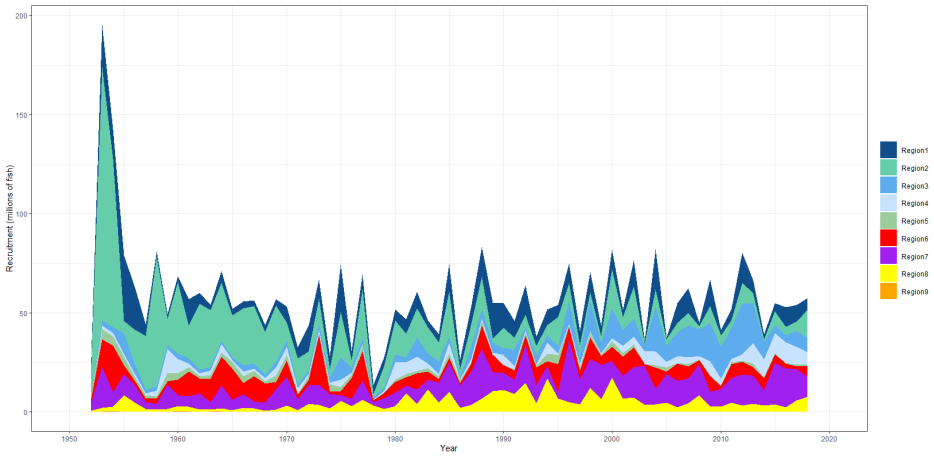
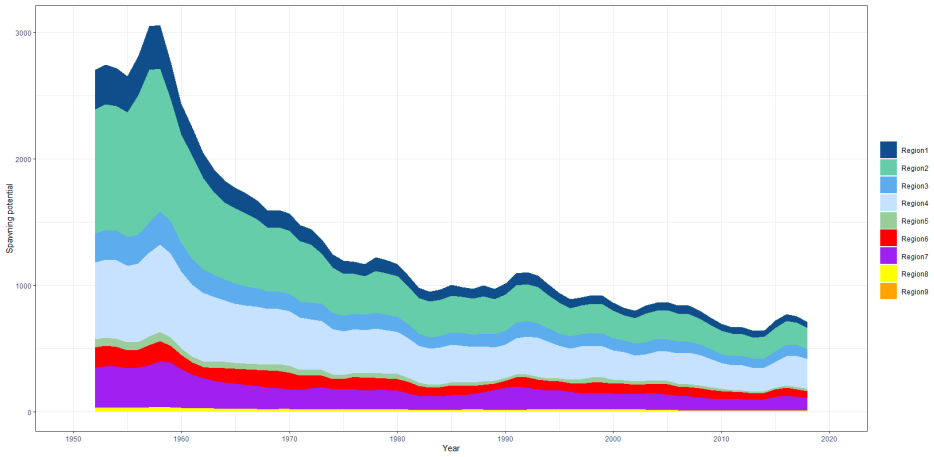


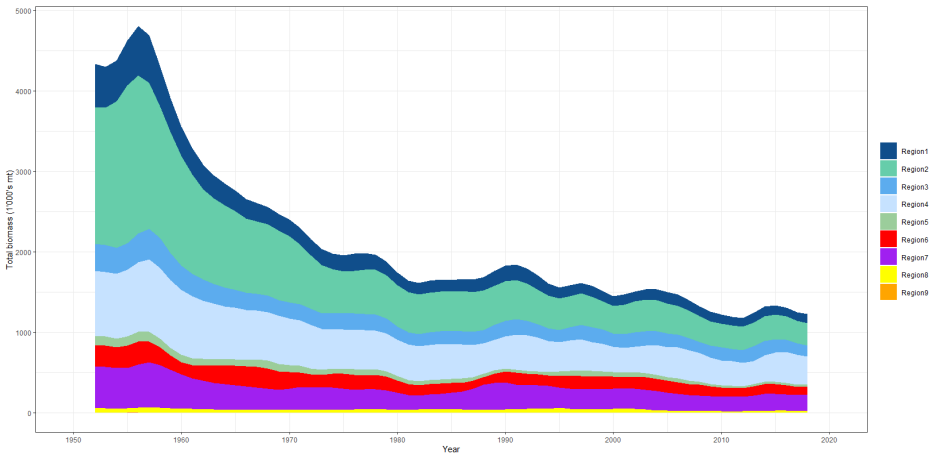
Figure 25: Estimated annual, temporal recruitment by model region for the diagnostic case model. The asymptotic 95% confidence interval as calculated using the delta-method is shown for the "Overall" region. Note that the scale of the y-axis is not constant across regions, and that the y-axis for the "Overall" region has been truncated in order to show the overall trend in recruitment.



(a) Regional recruitment



(b) Spawning potential



(c) Total biomass

Figure 26: Estimated (a) annual average recruitment, (b) spawning potential and (c) total biomass by model region for the diagnostic case model, showing the relative sizes among regions.

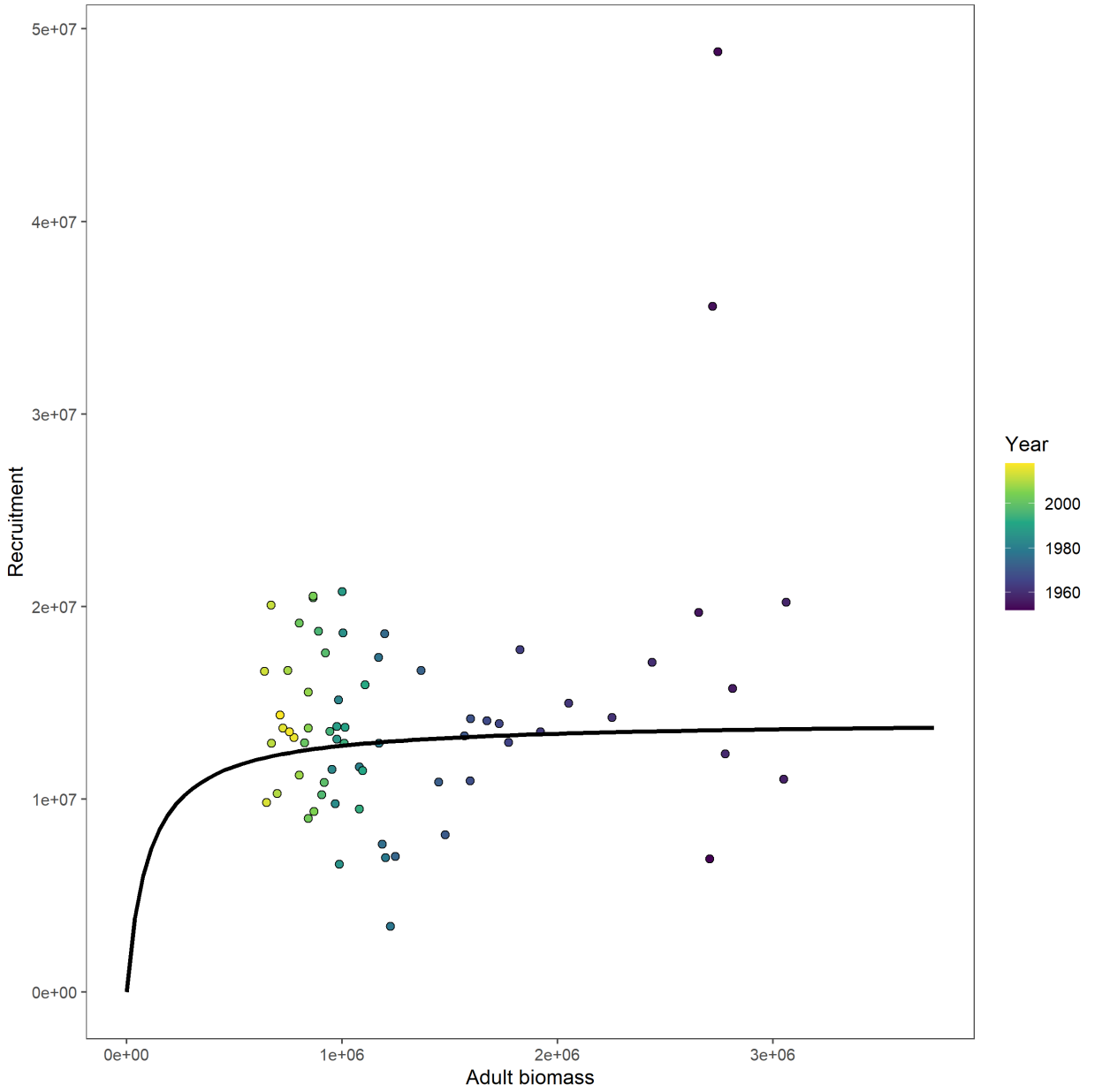


Figure 27: Estimated relationship between recruitment and spawning potential based on annual values for the diagnostic case model. The color of the circles change from dark blue to yellow through time. Note that though the recruitments prior to 1962 are not used to define the stock recruit relationship, they are shown in this figure.

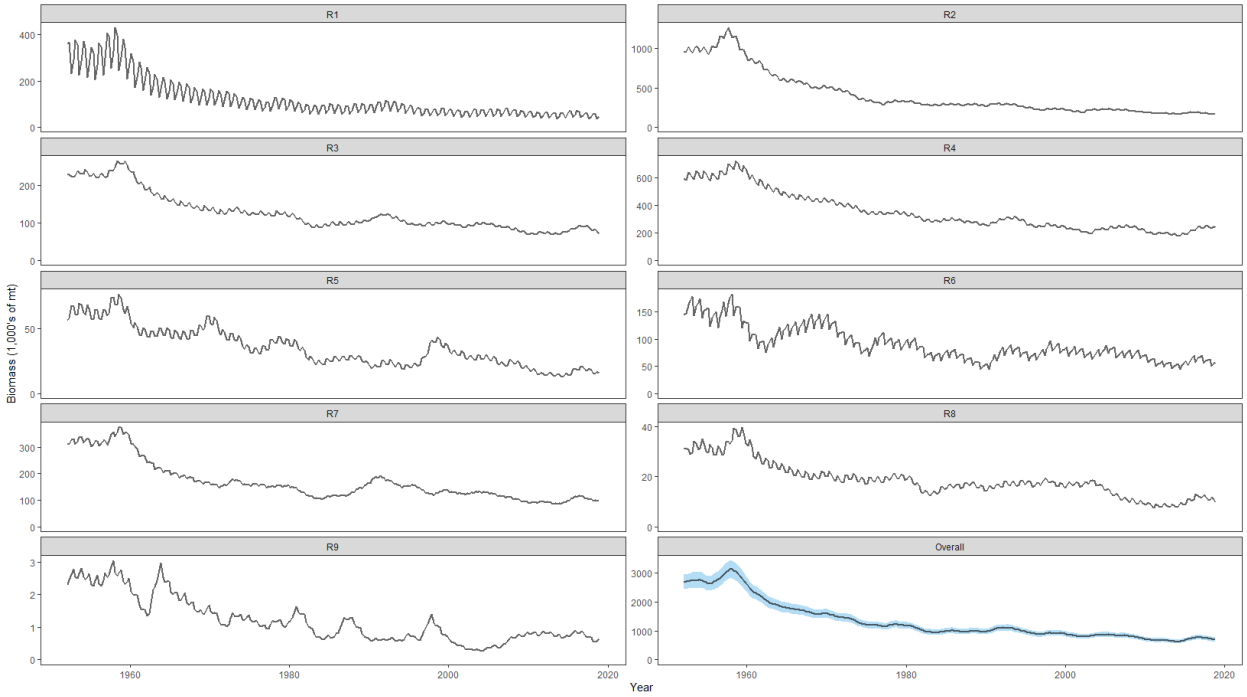


Figure 28: Estimated seasonal, temporal spawning potential by model region for the diagnostic case model. The asymptotic 95% confidence interval as calculated using the delta-method is shown for the "Overall" region. Note that the scale of the y-axis is not constant across regions.

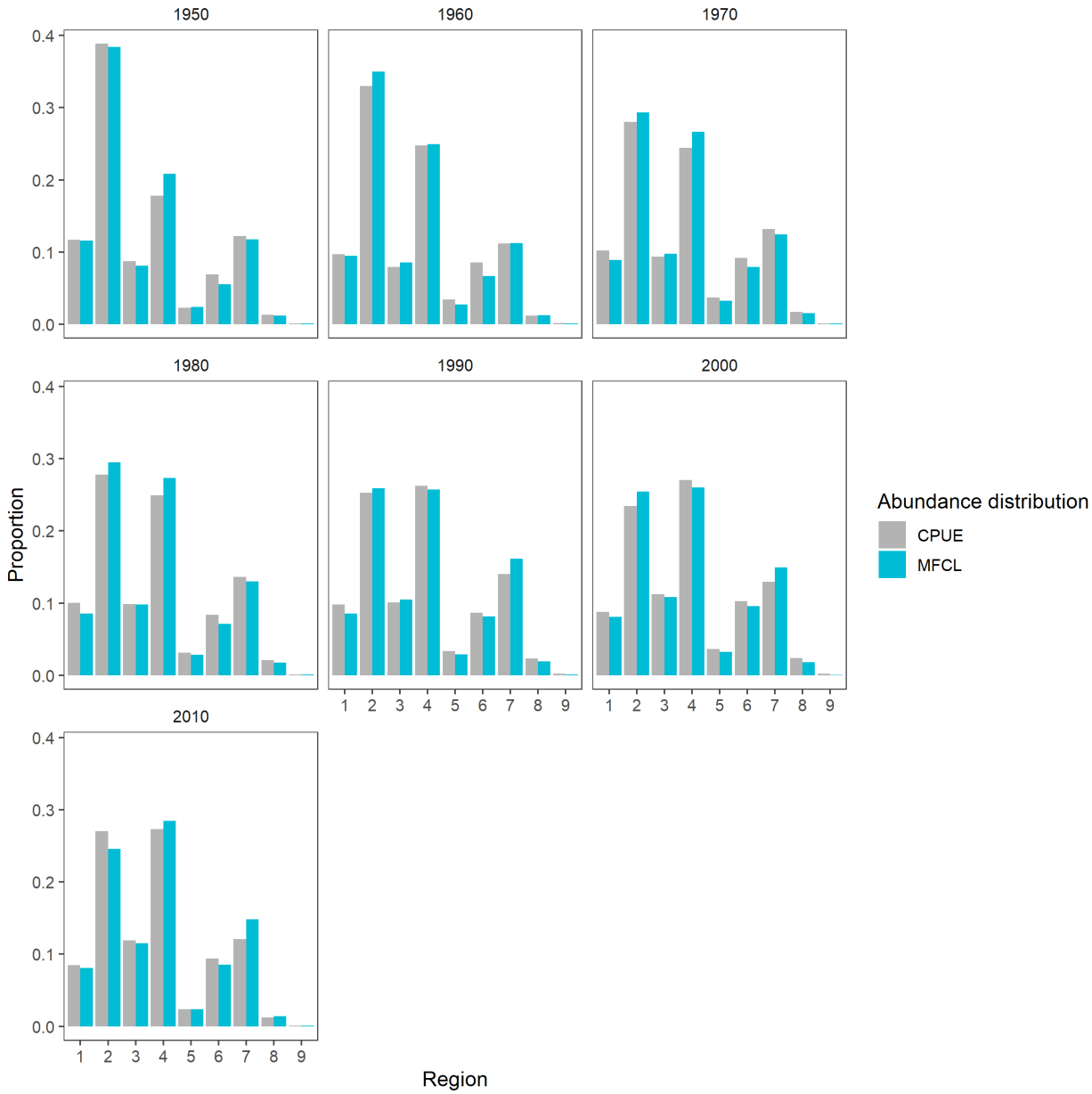


Figure 29: Estimated abundance distribution by region and decade from the spatiotemporal CPUE standardization model and the diagnostic case.

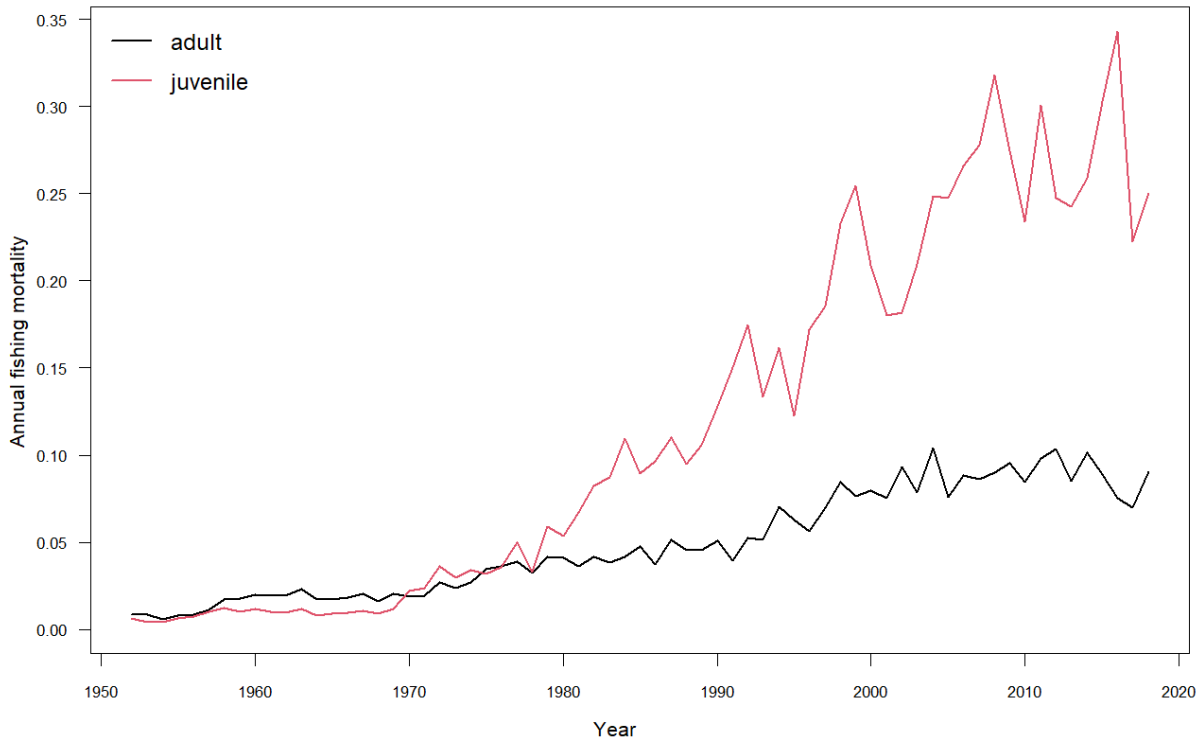


Figure 30: Estimated annual average juvenile and adult fishing mortality for the diagnostic case model.

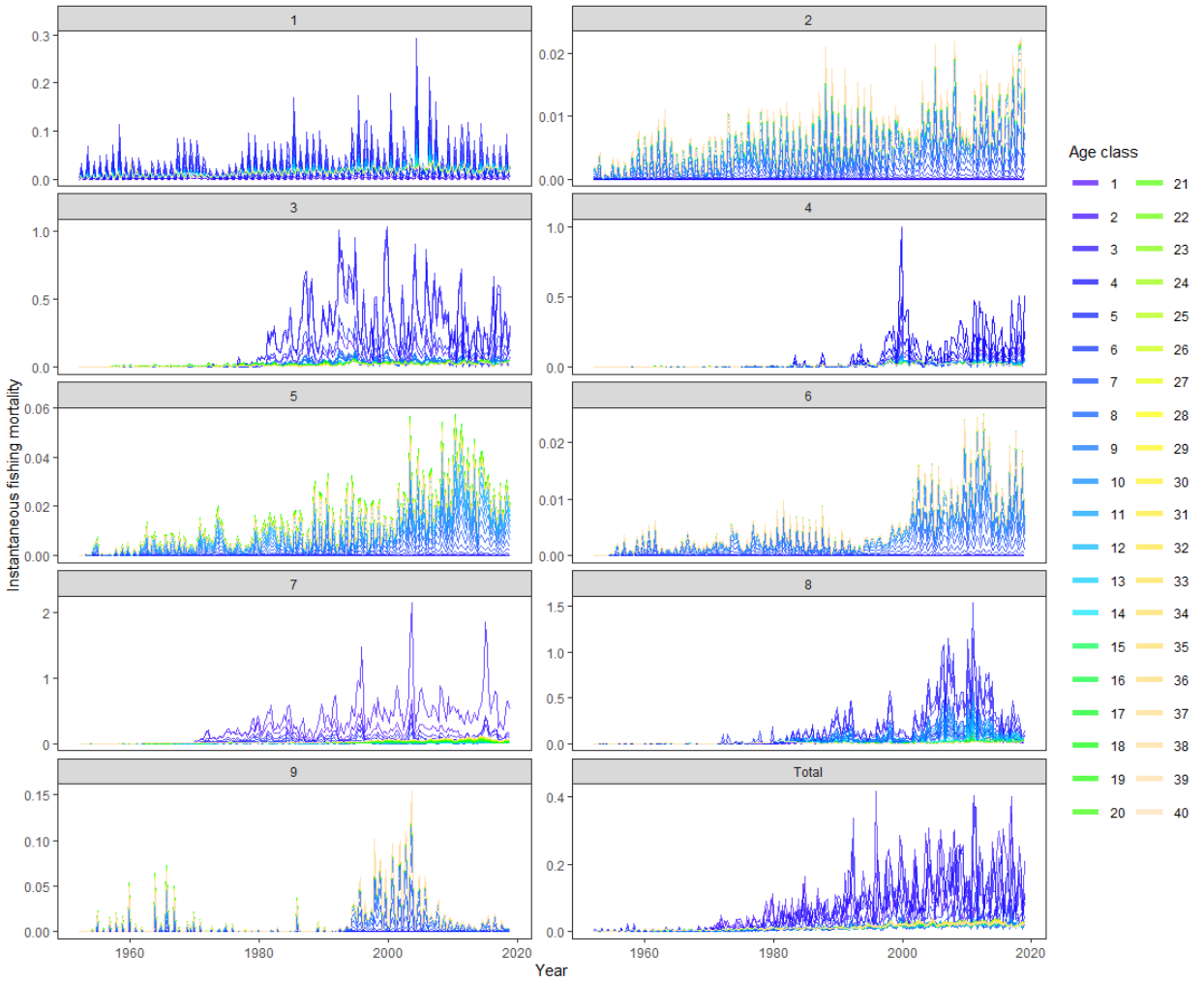


Figure 31: Estimated age-specific fishing mortality for the diagnostic case model, by region and overall.

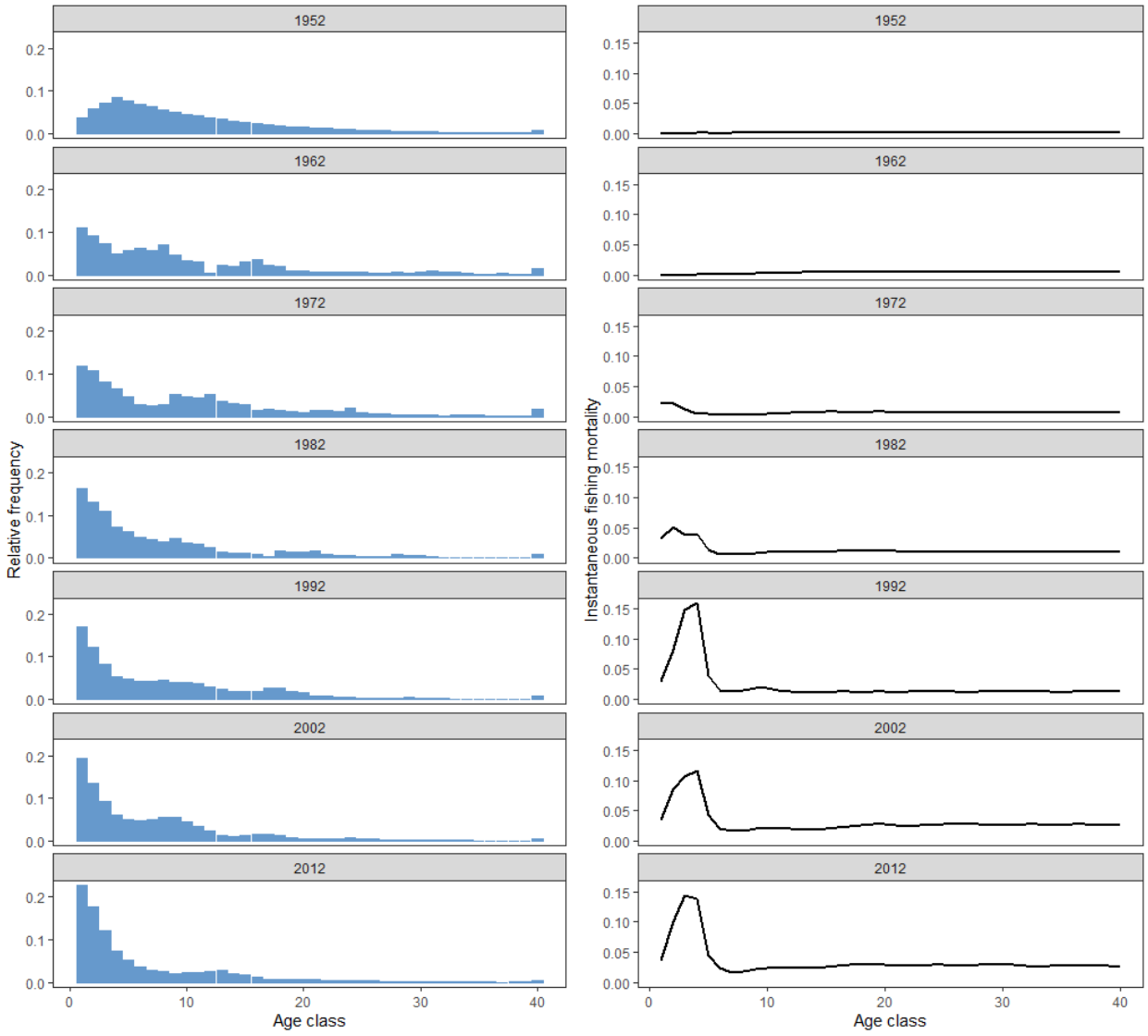
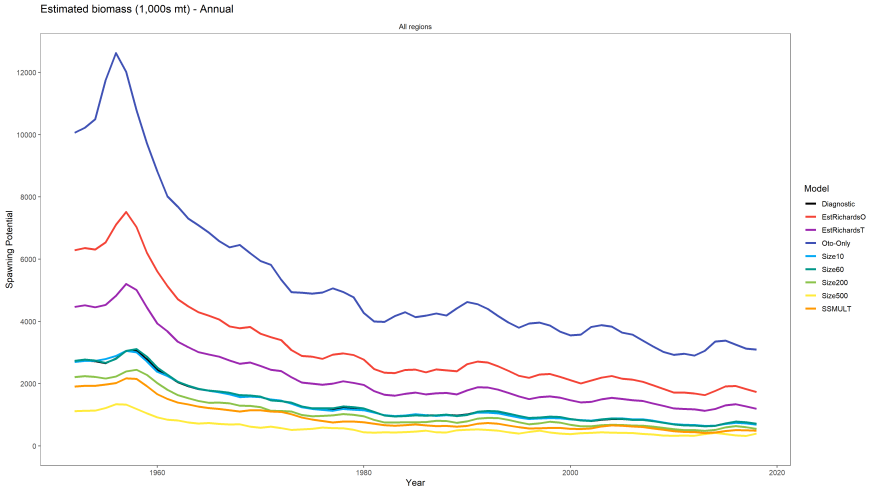
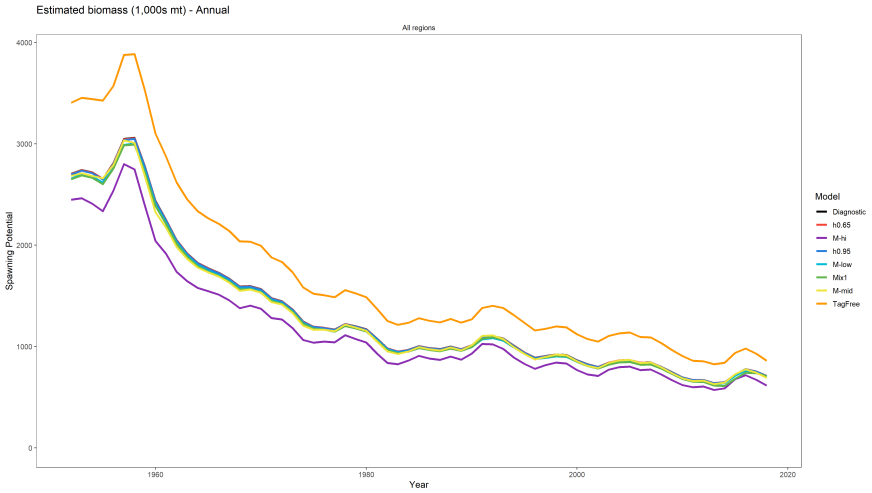


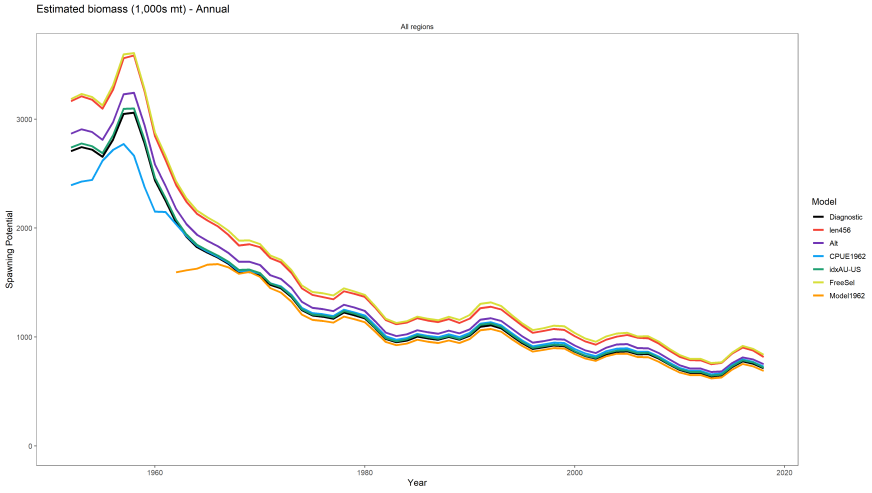
Figure 32: Estimated proportion of the population-at-age (quarters; left panels) and fishing mortality-at-age (right panels), at decadal intervals, for the diagnostic case model.



(a) First set of sensitivities

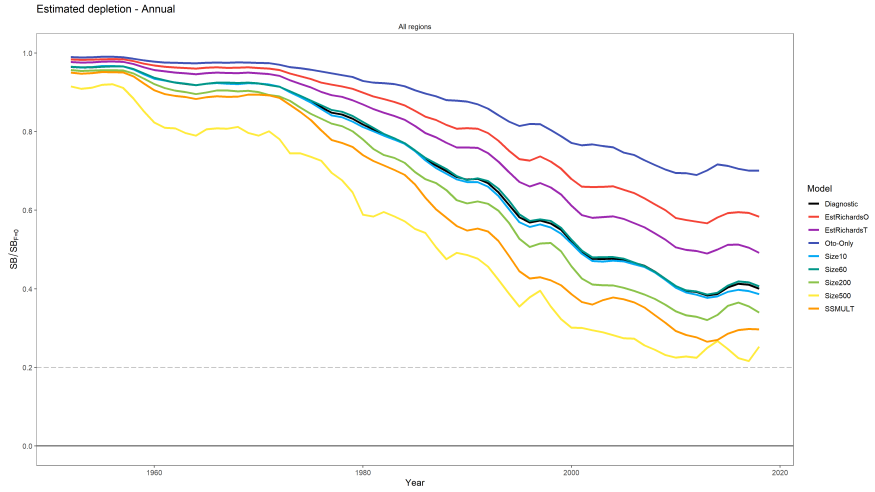


(b) Second set of sensitivities

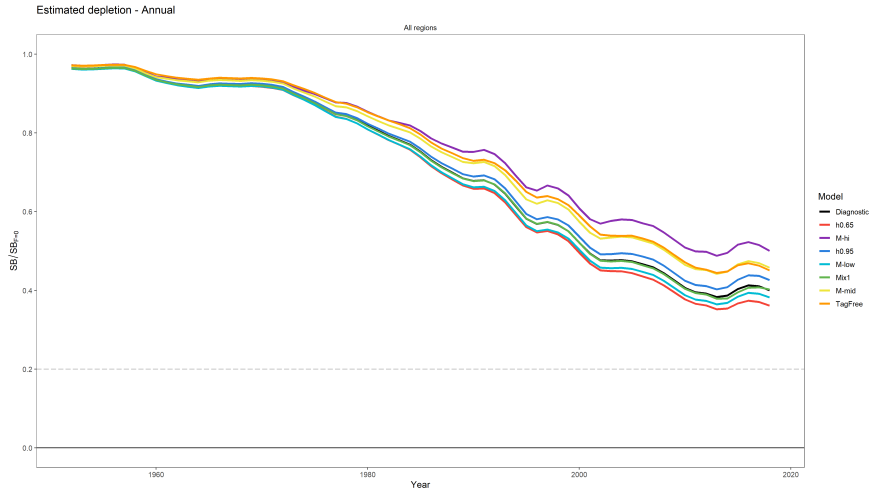


(c) Third set of sensitivities

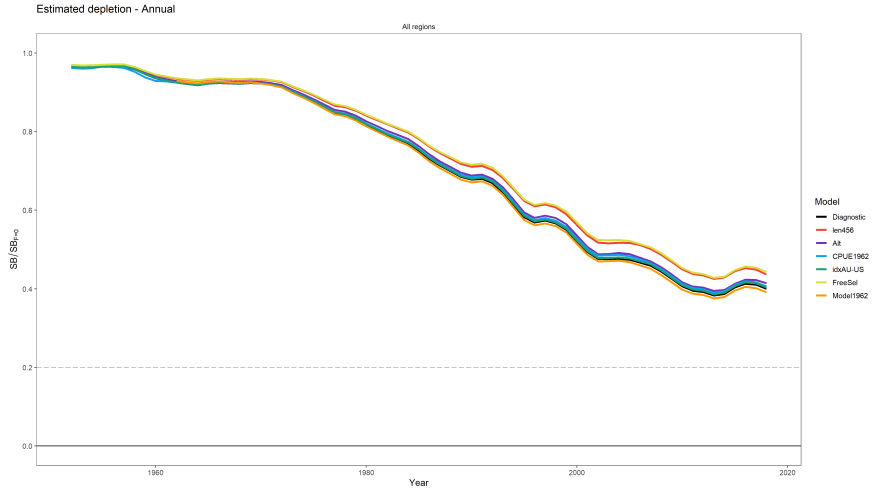
Figure 33: Estimated spawning potential for each of the one-off sensitivity models investigated in the assessment. The models are shown in three groups for ease of interpretation.



(a) First set of sensitivities



(b) Second set of sensitivities



(c) Third set of sensitivities

Figure 34: Estimated depletion for each of the one-off sensitivity models investigated in the assessment. The models are shown in three groups for ease of interpretation.

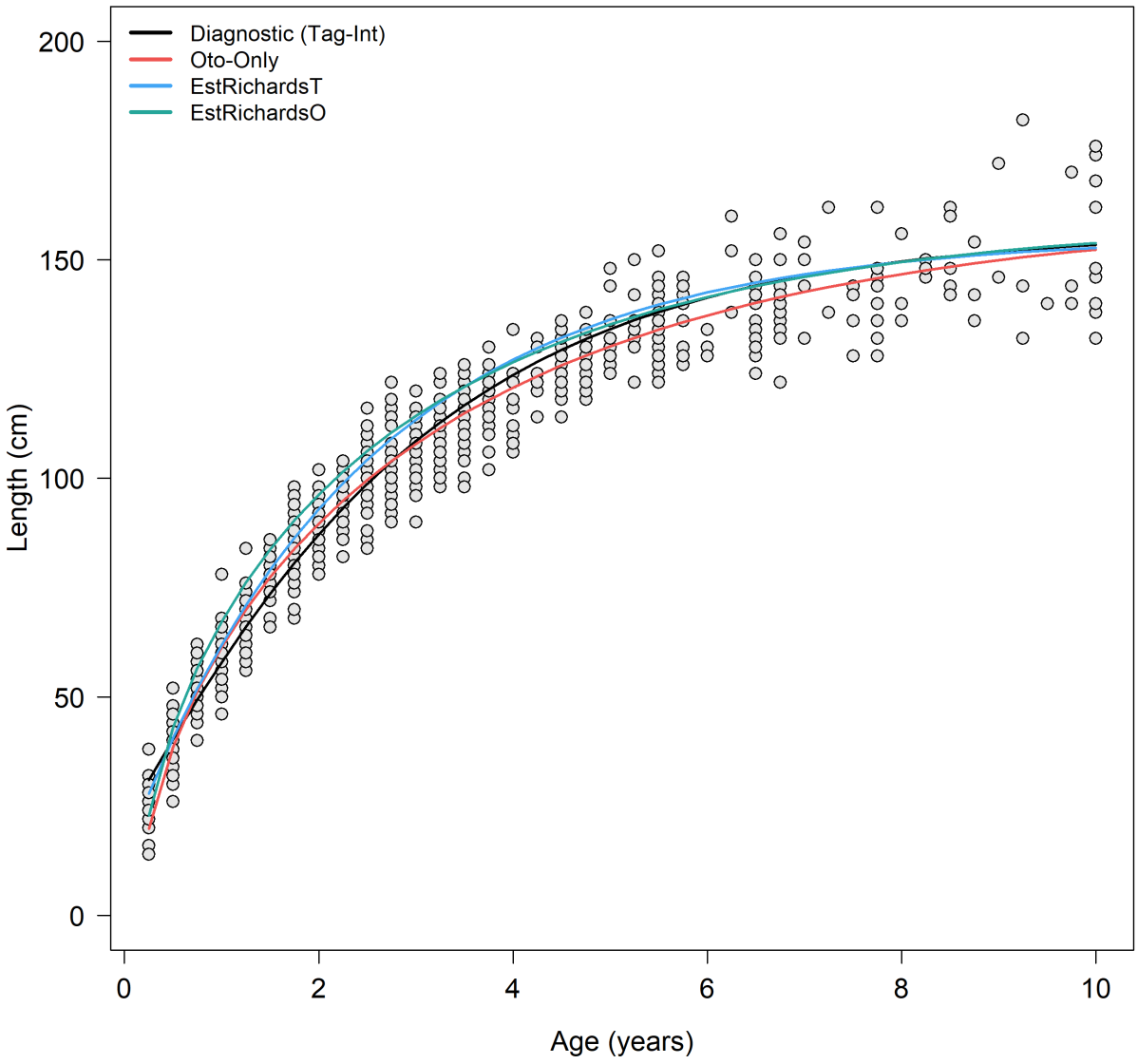
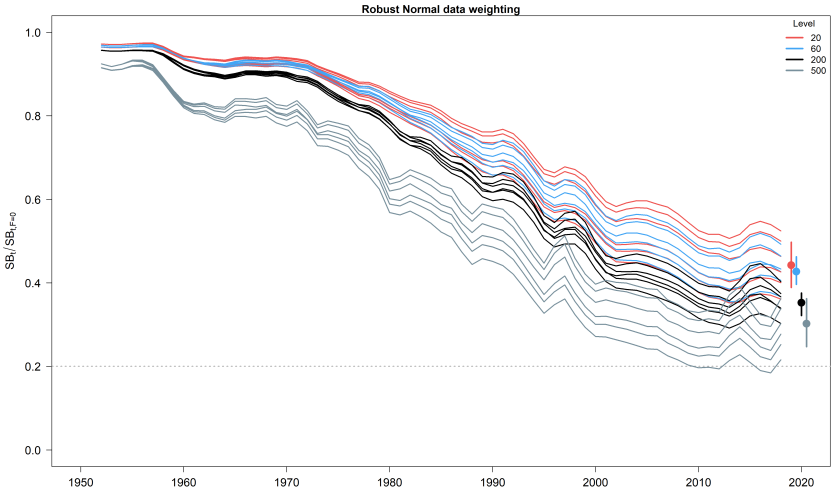
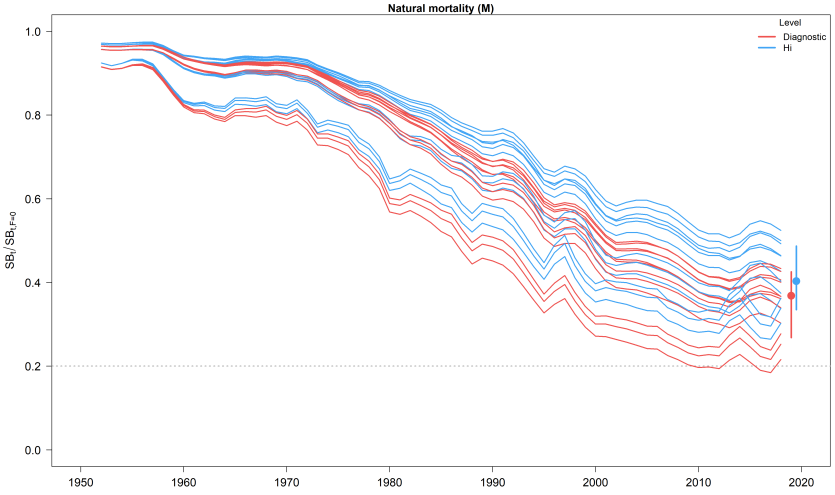


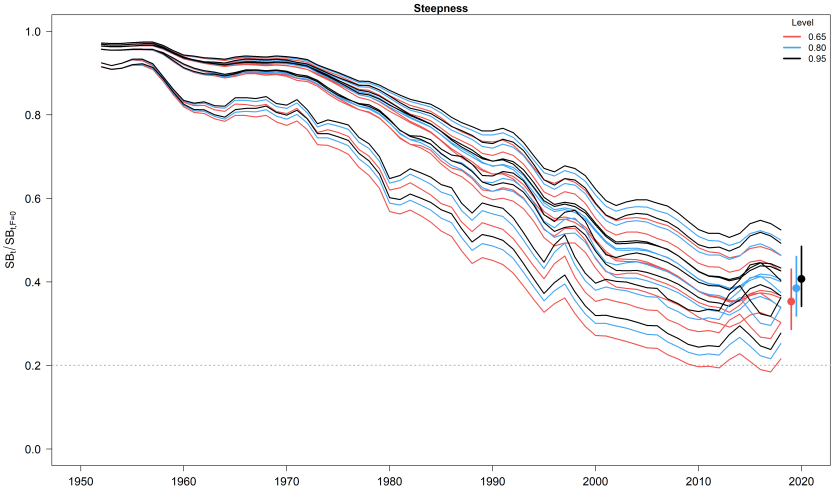
Figure 35: Growth curves considered in the one-off sensitivities: two fixed at external values *Diagnostic: Tag-Int (black)* and *Oto-Only* (red) and two estimated internally to MFCL using the conditional age-at-length data from the starting values of the fixed curves *EstRichardsT* (blue) and *EstRichardsO* (green). The otoliths comprising the conditional age-at-length data input into the assessment are shown in gray.



(a) Data-weighting axis



(b) Natural mortality axis



(c) Steepness axis

Figure 36: Estimated depletion ($SB_t/SB_{t,F=0}$) for each of the axes in the structural uncertainty grid: (a) data-weighting, (b) natural mortality, and (c) steepness. The median $SB_{latest}/SB_{F=0}$ and 80^{th} percentile for each axis level are shown on the right.

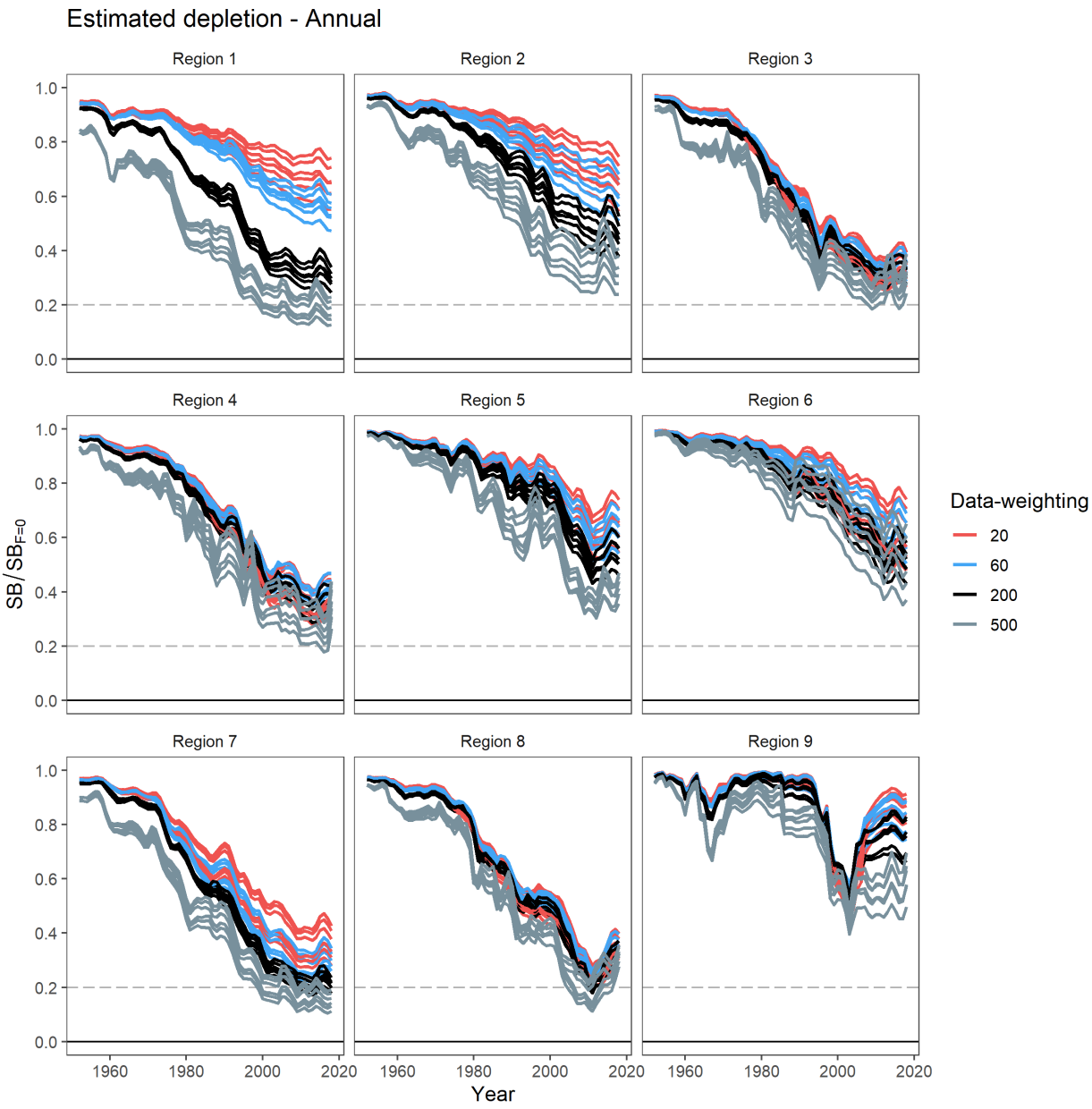


Figure 37: Estimated region specific depletion ($SB_t/SB_{t,F=0}$) for the data-weighting axis.

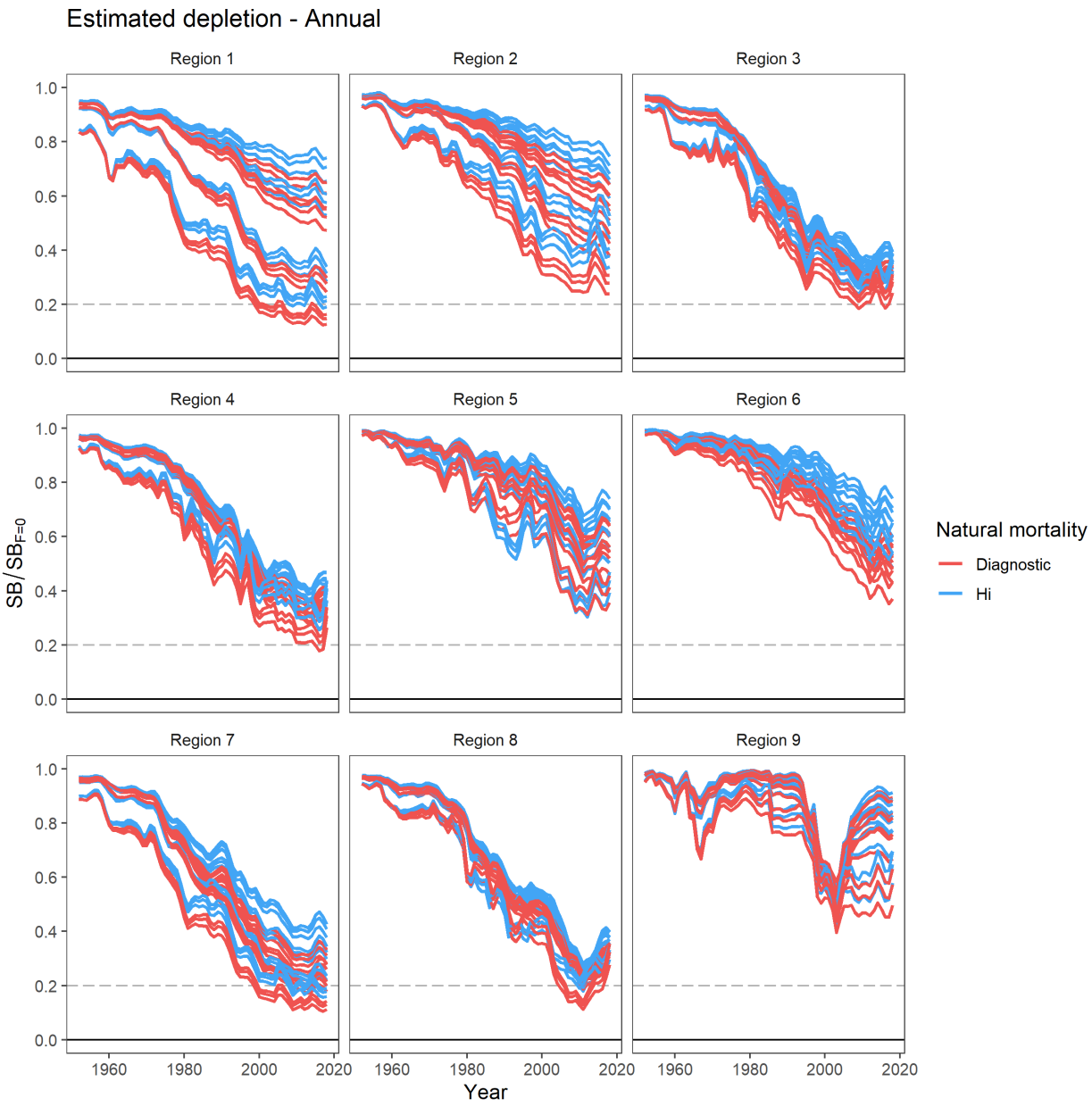


Figure 38: Estimated region specific depletion ($SB_t/SB_{t,F=0}$) for the natural mortality axis.

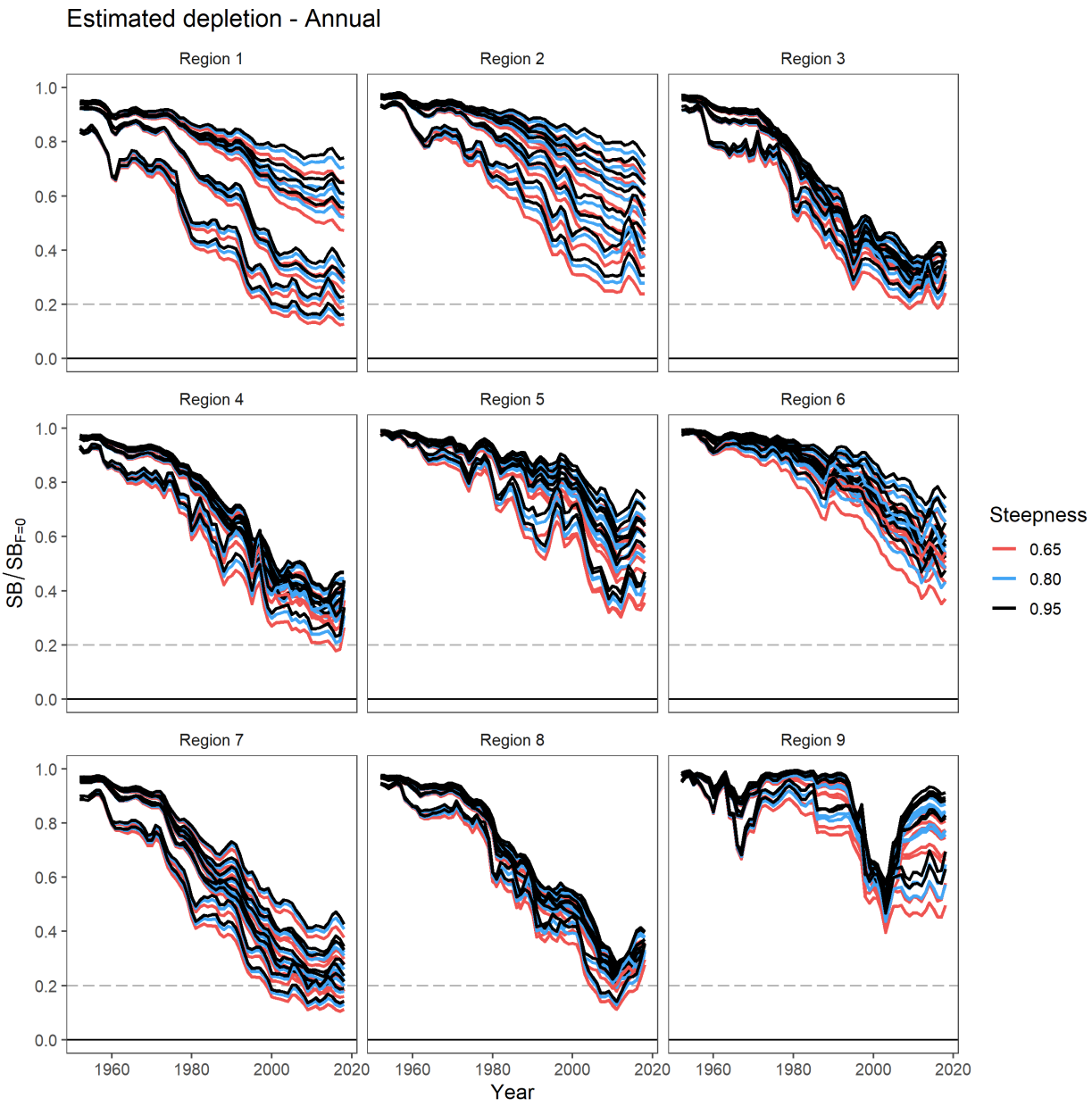


Figure 39: Estimated region specific depletion ($SB_t/SB_{t,F=0}$) for the steepness axis.

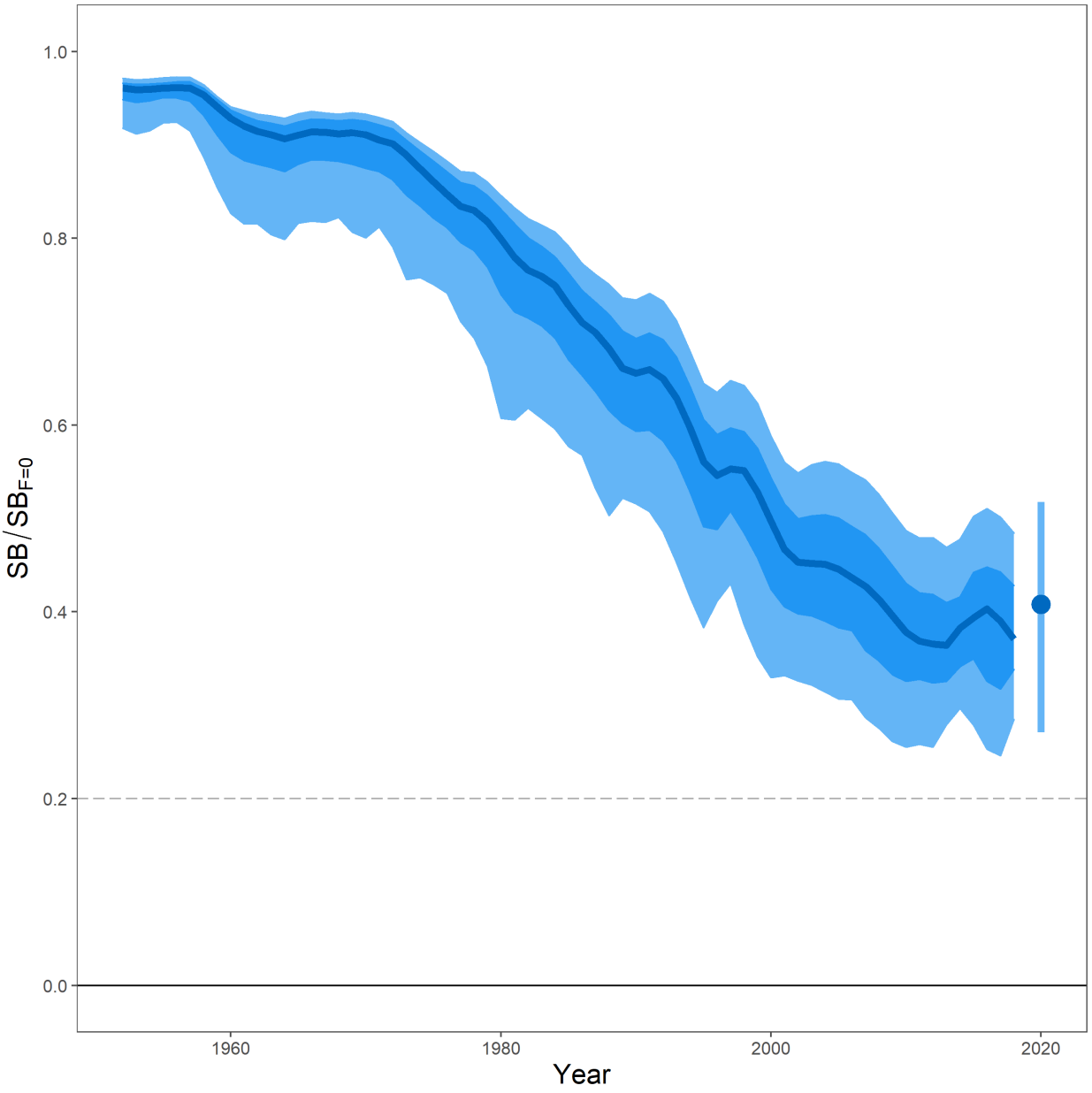


Figure 40: Time-dynamic percentiles of depletion ($SB_t/SB_{t,F=0}$) and median (dark line) across all 24 models in the structural uncertainty grid. The lighter band shows the 80^{th} percentile around the median, and the dark band shows the 50^{th} percentile around the median. The median $SB_{recent}/SB_{F=0}$ and 90^{th} percentile is shown on the right by the dot and line.

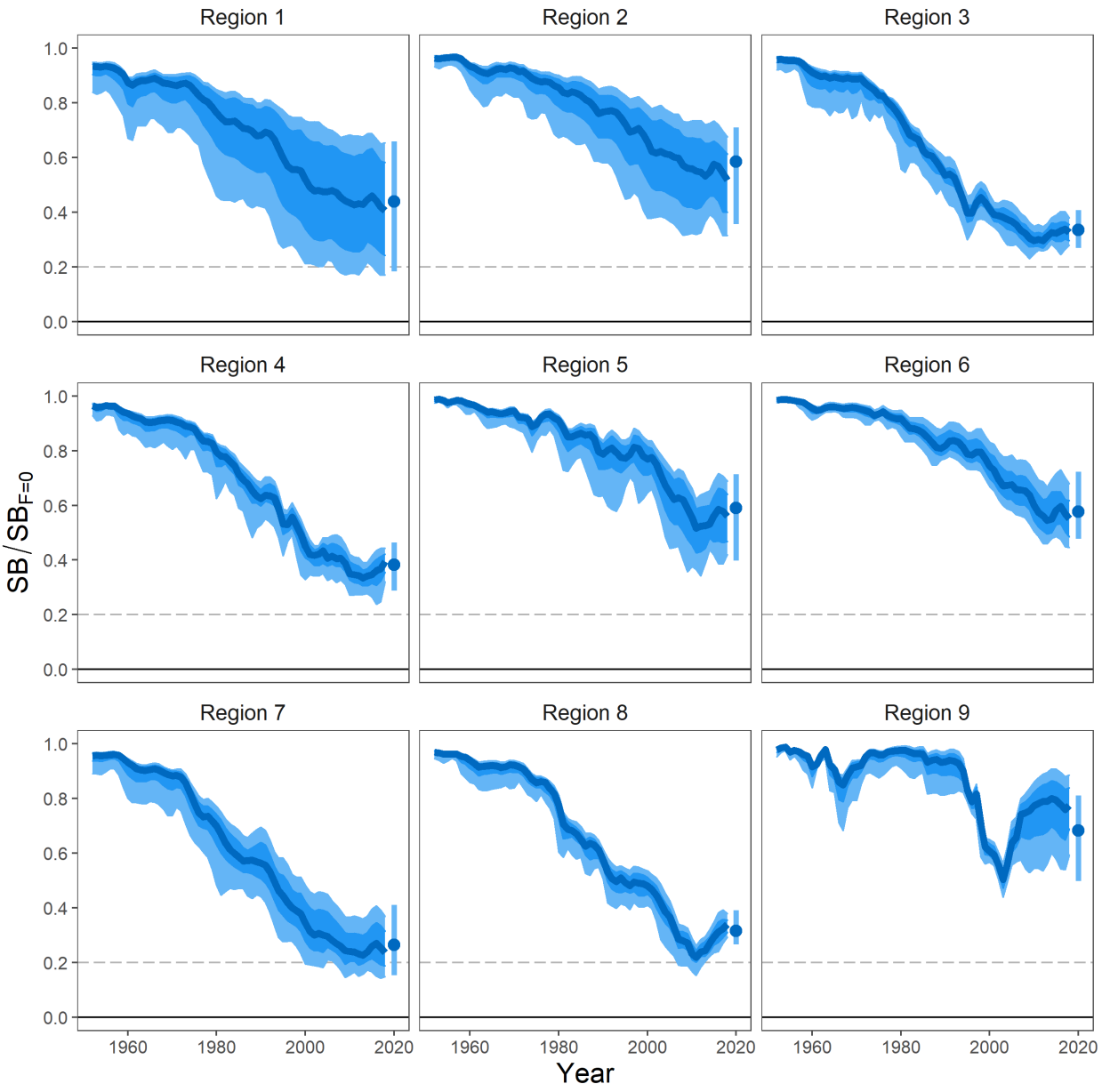
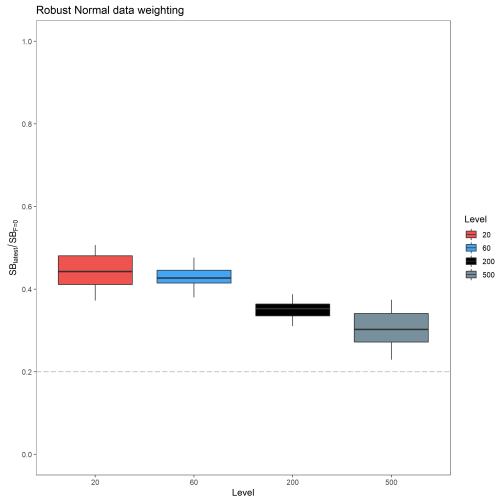
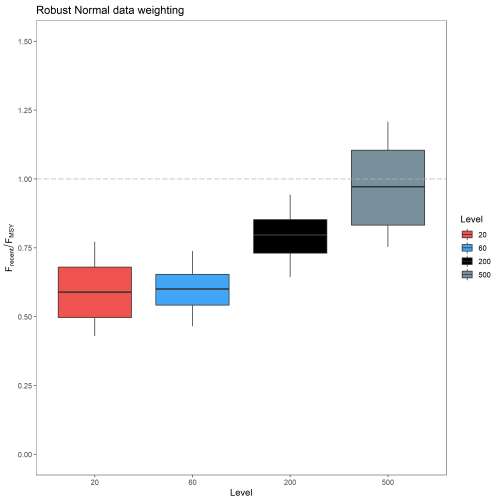


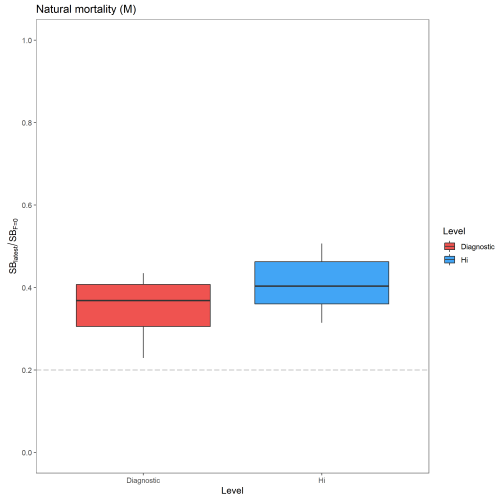
Figure 41: Time-dynamic percentiles of depletion ($SB_t/SB_{t,F=0}$) and median (dark line) by region across all 24 models in the structural uncertainty grid. The lighter band shows the 80^{th} percentile around the median, and the dark band shows the 50^{th} percentile around the median. The median $SB_{recent}/SB_{F=0}$ and 90^{th} percentile is shown on the right by the dot and line.



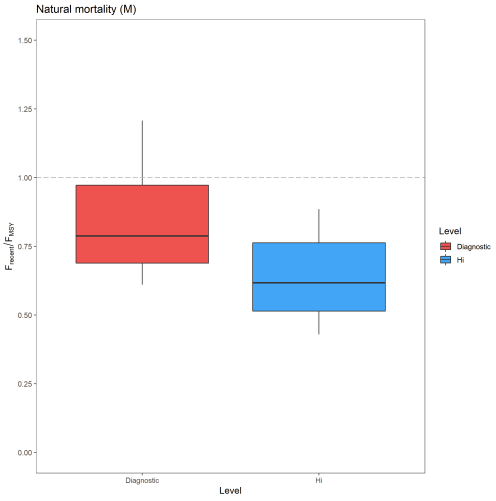
(a) Data-weighting $SB_{latest}/SB_{F=0}$



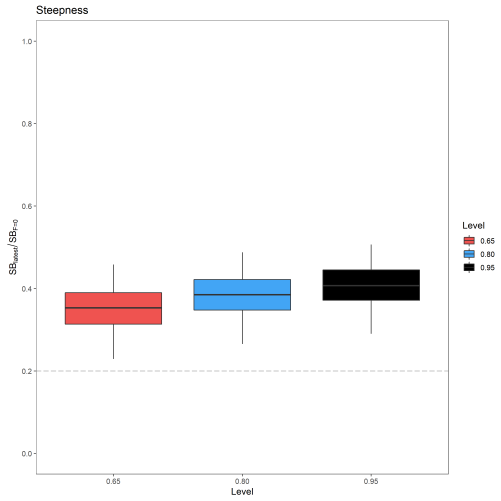
(b) Data-weighting $F_{recent}/F_{\text{MSY}}$



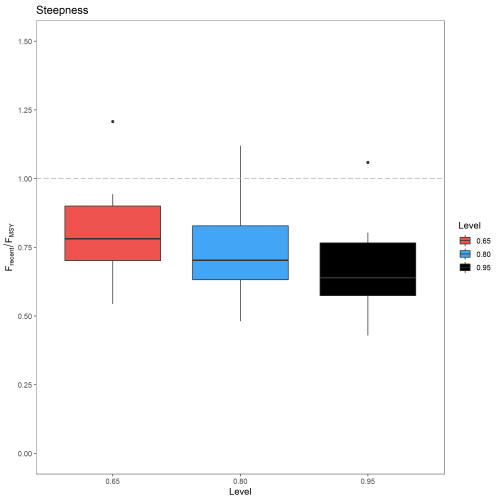
(c) Natural mortality $SB_{latest}/SB_{F=0}$



(d) Natural mortality $F_{recent}/F_{\text{MSY}}$

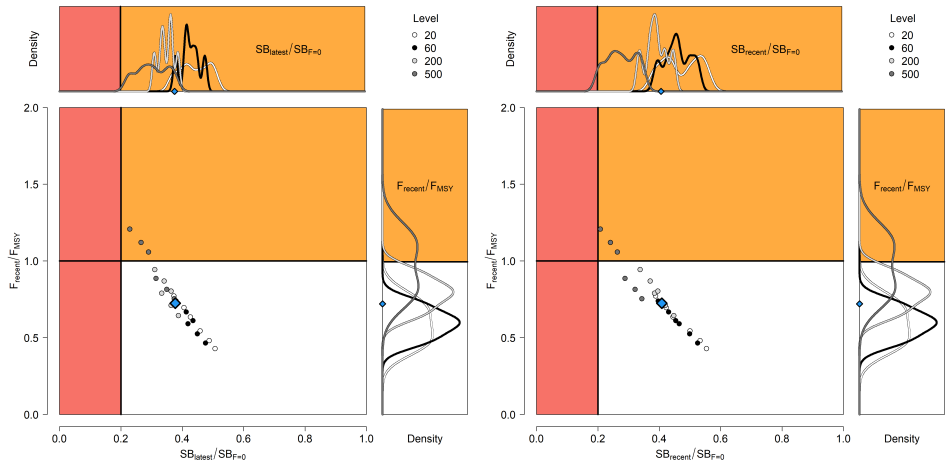


(e) Steepness $SB_{latest}/SB_{F=0}$



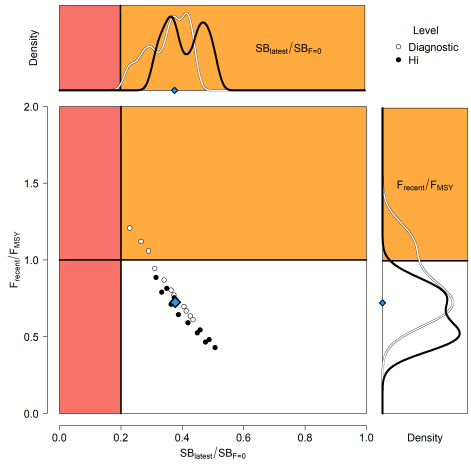
(f) Steepness $F_{recent}/F_{\text{MSY}}$

Figure 42: Estimated $SB_{latest}/SB_{F=0}$ and $F_{recent}/F_{\text{MSY}}$ for each of the axes in the structural uncertainty grid: (a–b) data-weighting, (c–d) natural mortality, and (e–f) steepness.

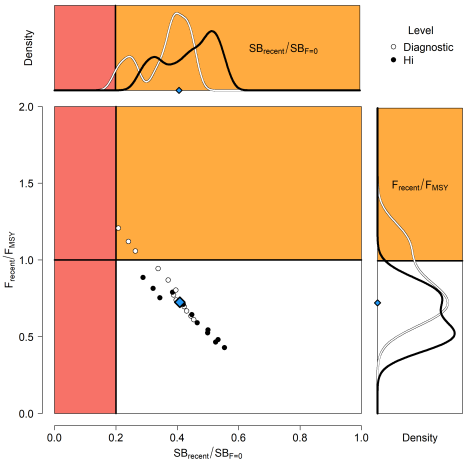


(a) Data-weighting $SB_{latest}/SB_{F=0}$

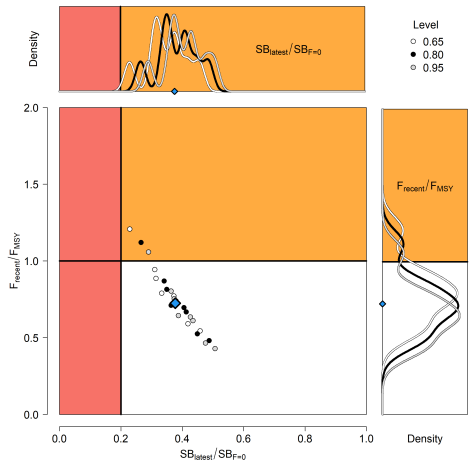
(b) Data-weighting $SB_{recent}/SB_{F=0}$



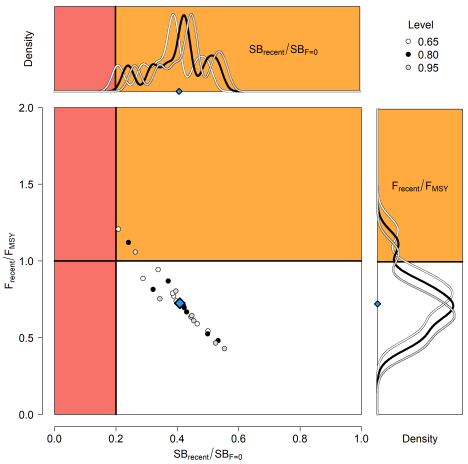
(c) Natural mortality $SB_{latest}/SB_{F=0}$



(d) Natural mortality $SB_{recent}/SB_{F=0}$



(e) Steepness $SB_{latest}/SB_{F=0}$



(f) Steepness $SB_{recent}/SB_{F=0}$

Figure 43: Majuro plots summarizing the results for each of the models in the structural uncertainty grid (left $SB_{latest}/SB_{F=0}$; right $SB_{recent}/SB_{F=0}$). The plots represent estimates of stock status in terms of spawning potential depletion and fishing mortality. The red zone represents spawning potential levels lower than the agreed limit reference point $20\%SB_{F=0}$. The orange region is for fishing mortality greater than F_{MSY}.The points represent each model in the grid, and the blue diamond represents the median estimate.

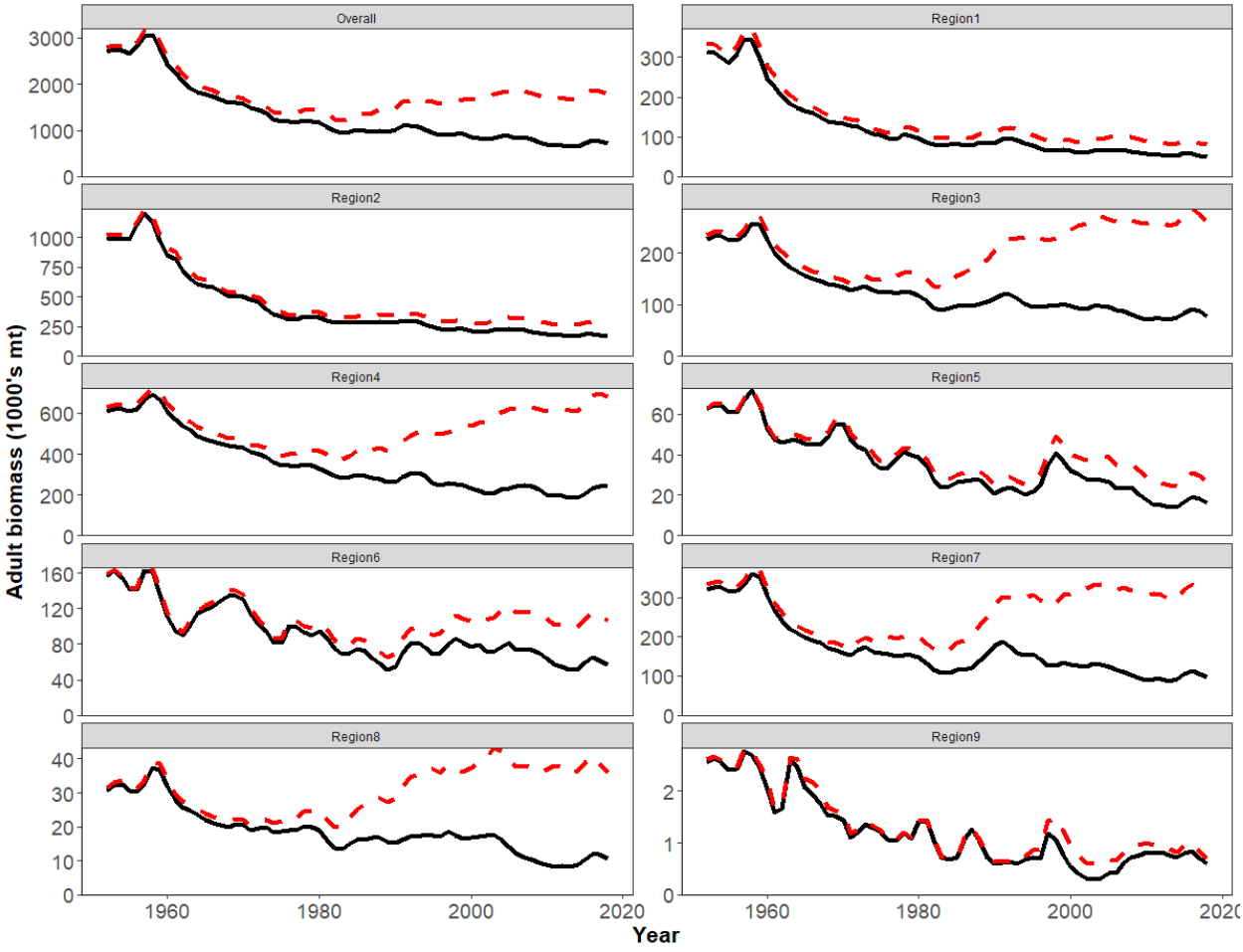


Figure 44: Comparison of the estimated annual spawning potential trajectories (lower solid black lines) with those trajectories that would have occurred in the absence of fishing (upper dashed red lines) for each region, and overall, for the diagnostic case model.

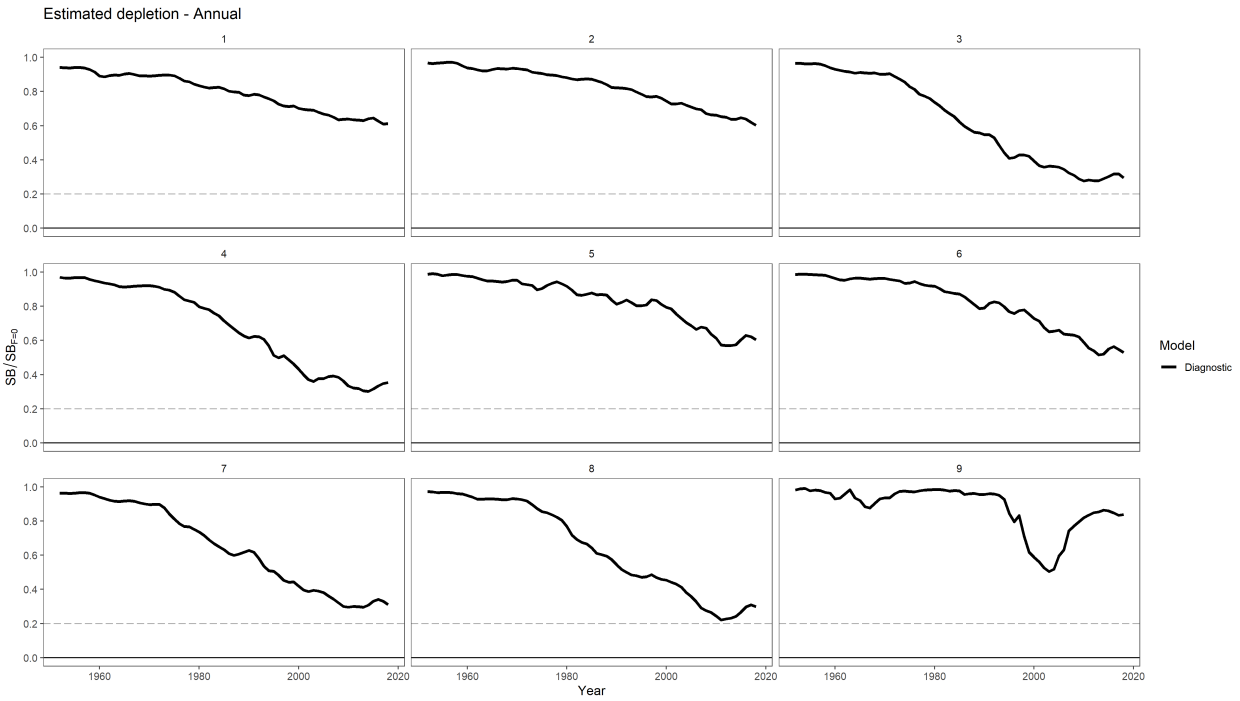


Figure 45: Ratio of exploited to unexploited spawning potential, $SB_t/SB_{t,F=0}$, for each region for the diagnostic case model.

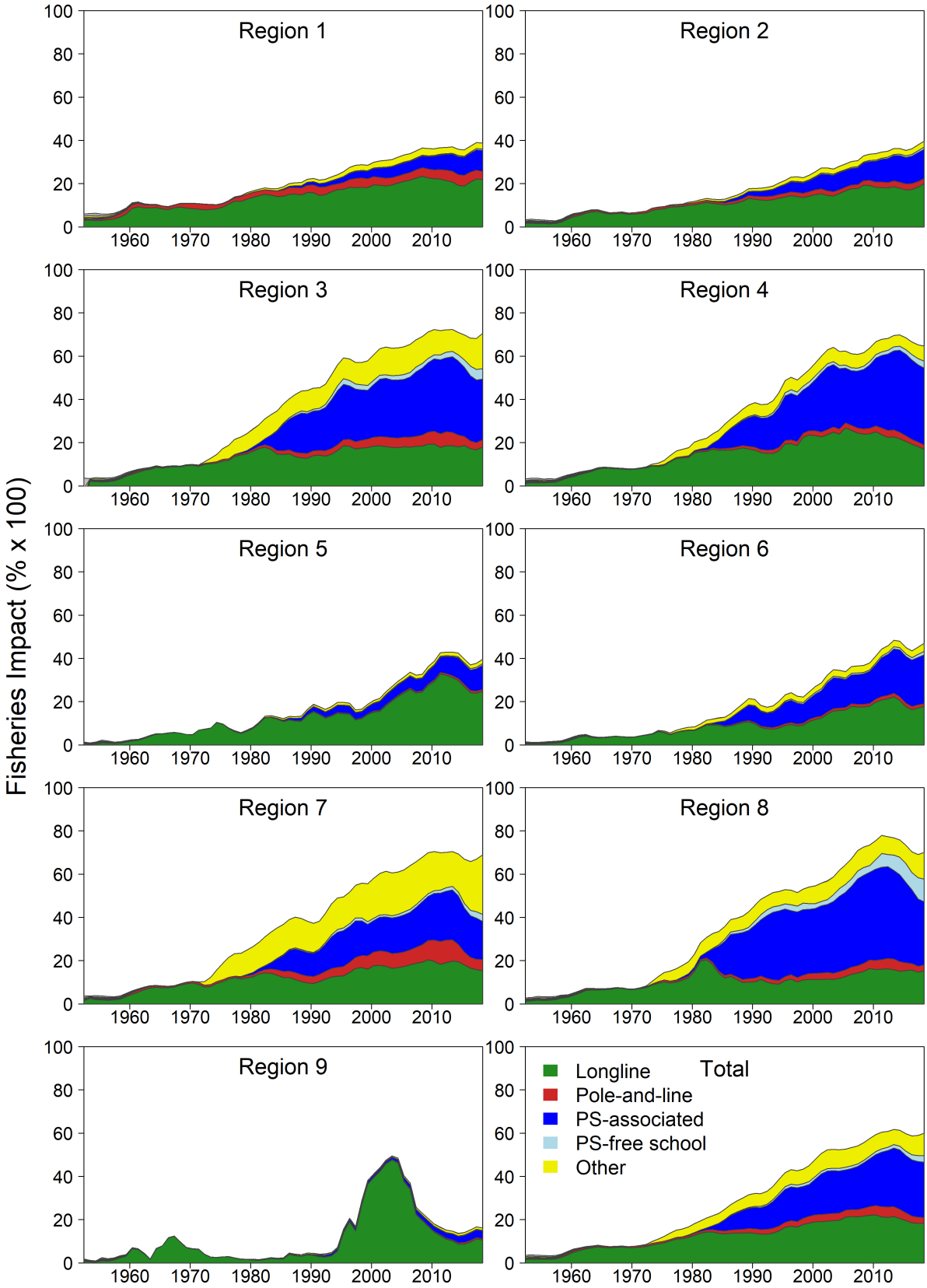


Figure 46: Estimates of reduction in spawning potential due to fishing (fishery impact = $SB_t/SB_{t,F=0} \times 100\%$) by region, and over all regions (lower right panel), attributed to various fishery groups for the diagnostic case model.

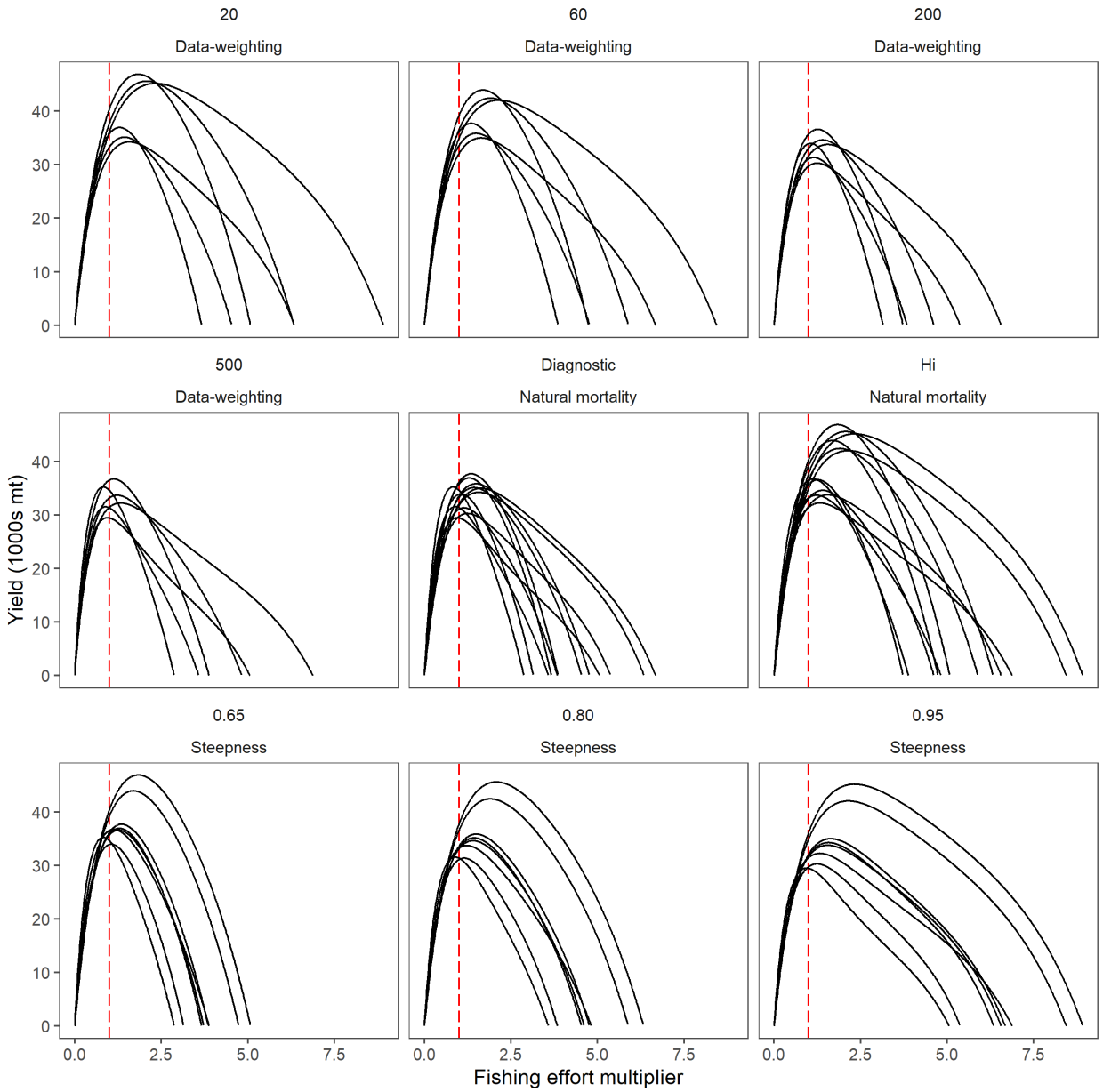


Figure 47: Estimated yield as a function of fishing mortality multiplier across the grid. The red line represents current fishing mortality, and can be used as a reference to determine the current yield relative to the maximum of the yield curve.

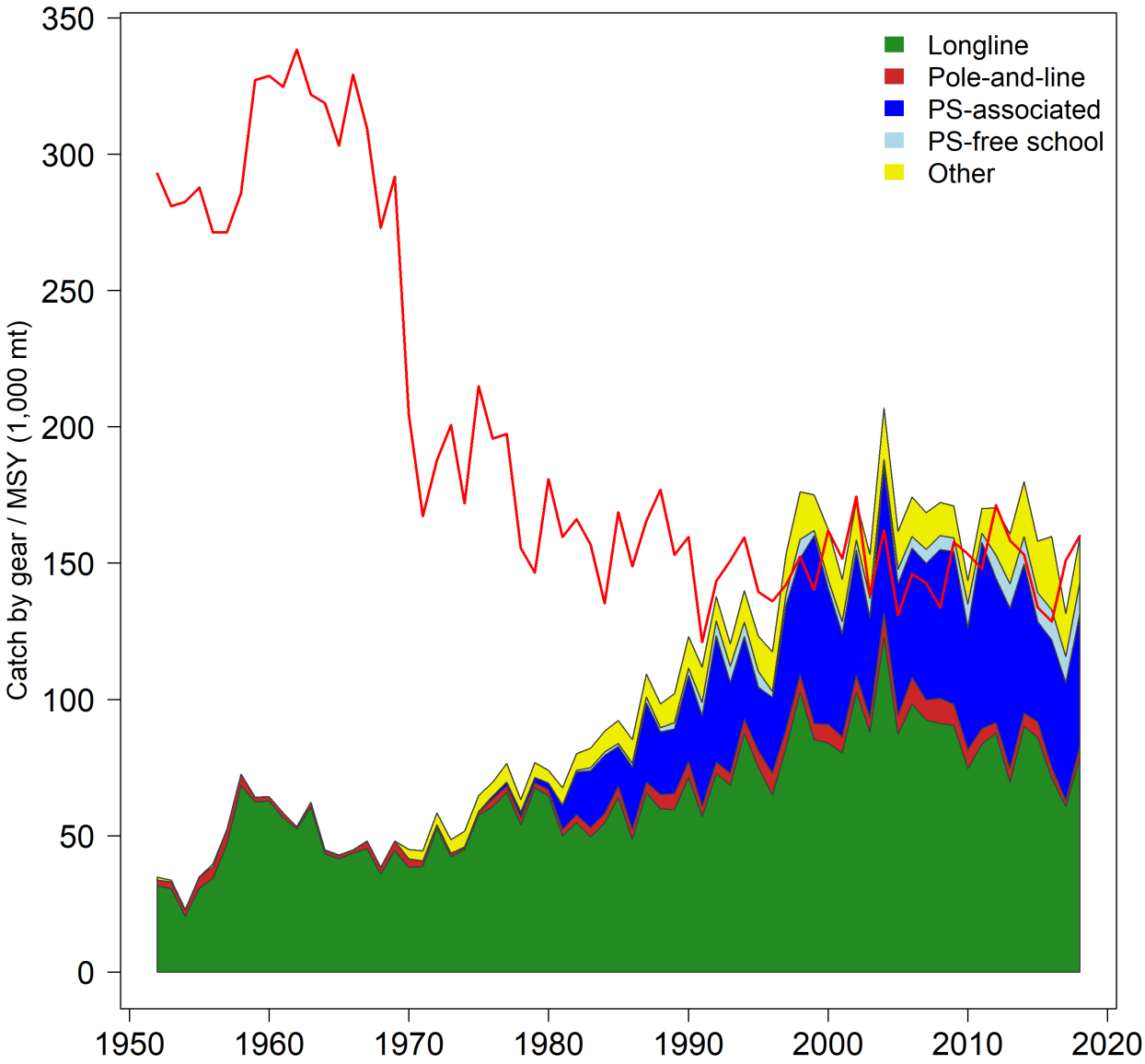


Figure 48: History of the annual estimates of MSY (red line) for the diagnostic model compared with annual catch by the main gear types. Note that this is a "dynamic" MSY which is explained further in Section 5.7.1.

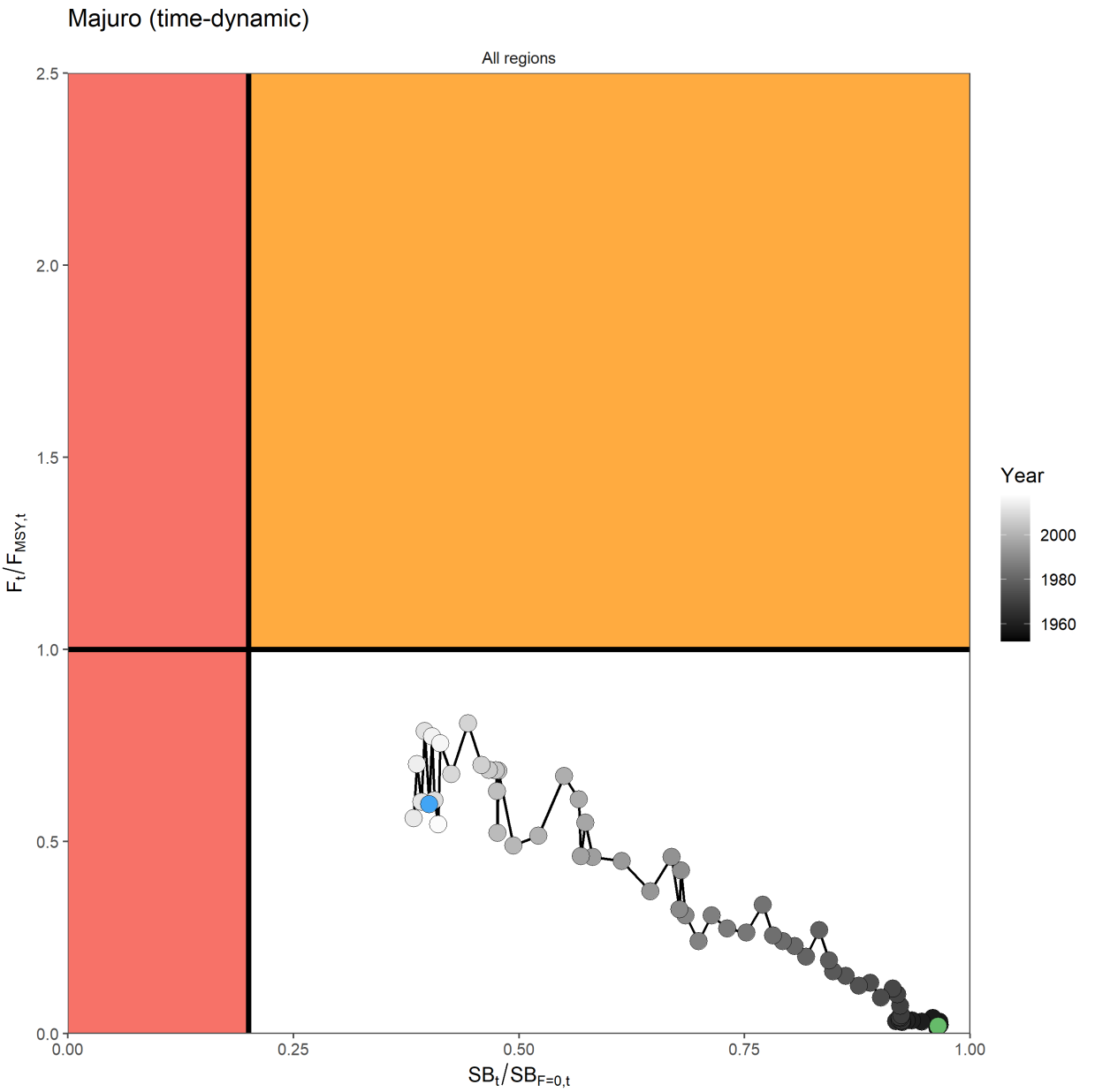


Figure 49: Estimated time-series (or "dynamic") Majuro plot for the diagnostic case. This plot is interpreted in the same manner as the description in Figure 43 except that it shows the temporal change in stock status with respect to the reference points $F_{recent}/F_{\text{MSY}}$ and $SB_{latest}/SB_{F=0}$, rather than the terminal estimates presented in previous figures. The shading of the points goes from dark (model start) to white (model end) with the initial and terminal conditions for 1952 and 2018 shown in green and blue, respectively. Note that the process of estimating a "dynamic" Majuro plot is explained further inSection 5.7.4.

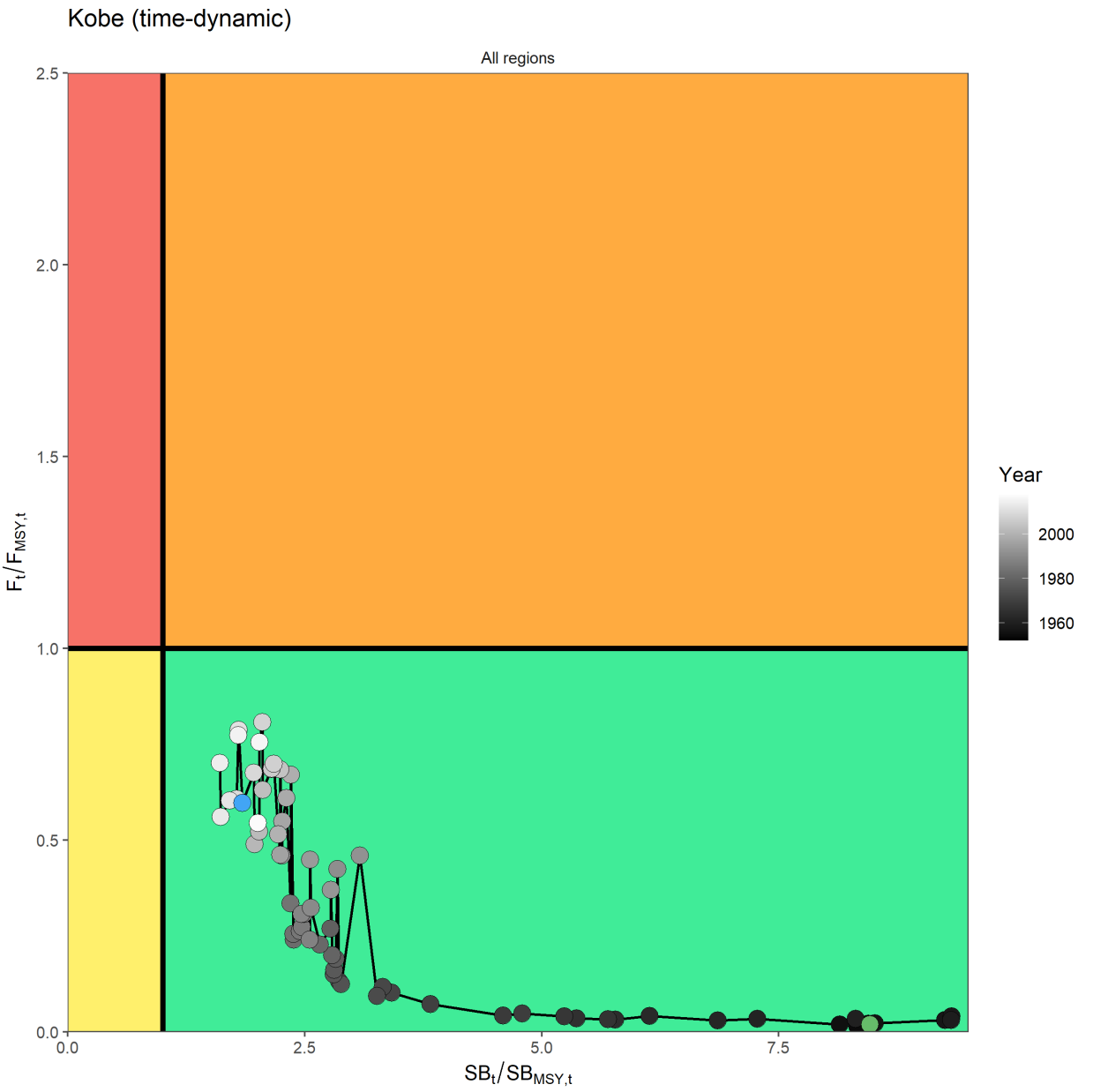


Figure 50: Estimated time-series (or "dynamic") Kobe plot for the diagnostic case. This plot shows the temporal change in stock status with respect to $F_{recent}/F_{\text{MSY}}$ and $SB_{latest}/SB_{\text{MSY}}$, rather than terminal estimates. The shading of the points goes from dark (model start) to white (model end) with the initial and terminal conditions for 1952 and 2018 shown in green and blue, respectively. Note that the process of estimating a "dynamic" Kobe plot is explained further inSection 5.7.4.

11 Appendix

11.1 Proposal for quasi-fisheries-independent sampling program

A quasi-fisheries-independent sampling program can be used to address many of the aforementioned issues related to the input data: improve collection of important biological samples (otoliths, genetics, reproduction, etc.), serve as a platform for conducting experiments to explore the selectivities and catchabilities of the fishing gear, and generate a fisheries-independent index of CPUE. A proposed implementation for an annual (or quarterly) survey would be to charter commercial longline vessels operating out of DWFN and PICT ports across the WCPO. In order to simultaneously capture the dynamics across the WCPO, vessels would each be assigned sampling stations within reasonable proximity to their home port and requested to fish within the same temporal window. Each of these vessels would be accompanied by an observer (or multiple observers to assist in the biological sampling), and would be instructed to fish according to a pre-determined protocol that would attempt to standardize vessel specific differences in targeting as much as possible. To minimize the cost of such a chartering arrangement, vessels participating in the program would be allowed to keep the revenues from selling any commercially valuable species caught during sampling. If this is not enough to offset the operating costs of the sampling trip, these costs would be covered through the chartering agreement.

11.2 Likelihood profile

Calculation of the likelihood profile (Figure 51) for the diagnostic case indicated conflict between the different data components, with the overall fit resulting from a compromise between the different data sets. The standardized CPUE indicated a preference for the lowest estimated total biomass ($\sim$ 1.3 million mt) while the length composition data and tagging components indicating higher estimates of average total biomass ($\sim$ 2.6 and 3.1 million, respectively). The weight composition indicates an estimate of average total biomass, in between the other components, of just under 2 million mt.

Conflict between the different data components is indicative of potential model or data misspecification. Disagreement between the two types of size-frequency data could be an indicator that there is an issue with the conversion factors used to prepare the length frequency data. It is also possible that this disagreement is a product of the length and weight frequency data coming from distinct groups of fisheries targeting different segments of the population (weight frequency data comes from longline fisheries targeting older, larger individuals while length frequency data typically comes from purse seine fisheries targeting younger, smaller individuals). Additionally, the high estimate of average total biomass indicated by the tagging data could be linked to the two tag reporting groups with reporting rates still estimated to be on the upper bound. The effect of recruitment variability around the estimated stock-recruitment relationship has a negligible influence on estimated stock status.

The asymmetric nature of the total likelihood profile indicates that the model is more certain about the lower bound on the estimate of average total biomass, and less certain on the higher end given the structural assumptions made in the diagnostic case.

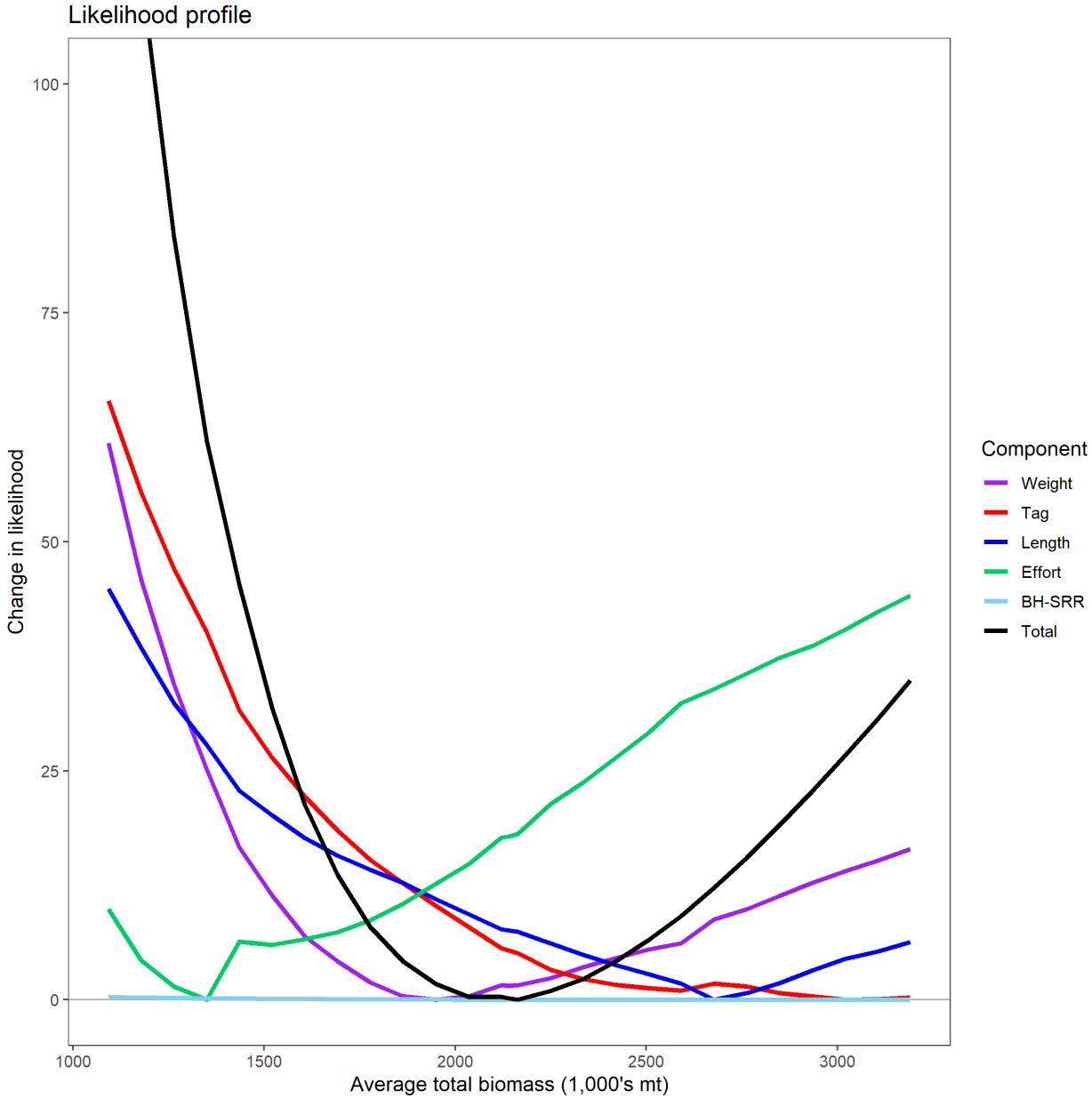
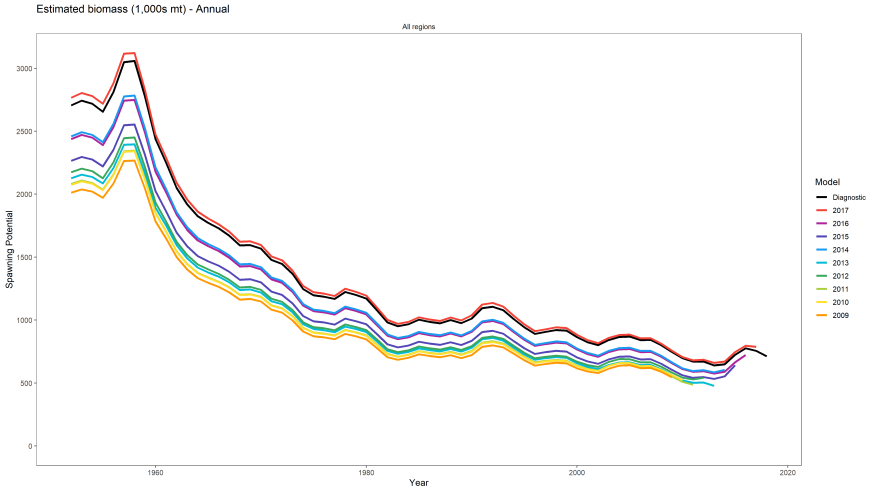


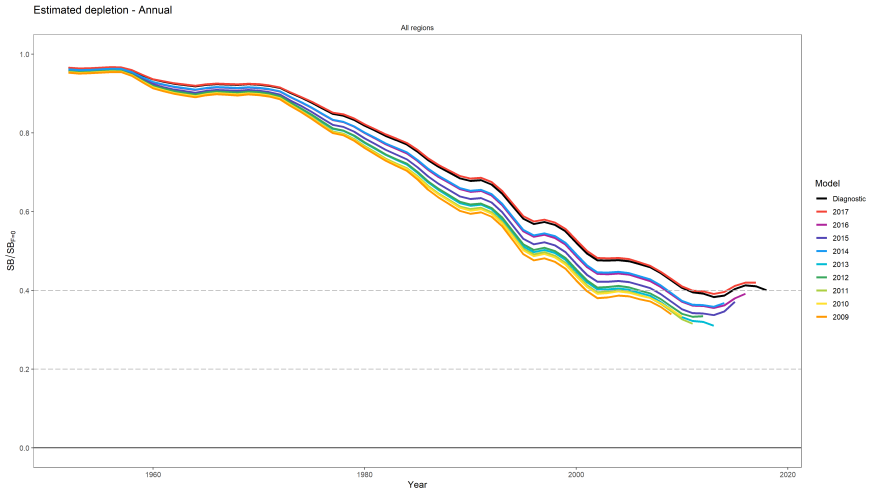
Figure 51: Likelihood profile with respect to average total biomass for the diagnostic case. The total likelihood (black) is shown along with the influence of each likelihood componenent: weight frequency (purple), tagging data (red), length frequency (blue), CPUE/effort deviates (green), and the recruitment variability (light blue).

11.3 Retrospective analyses

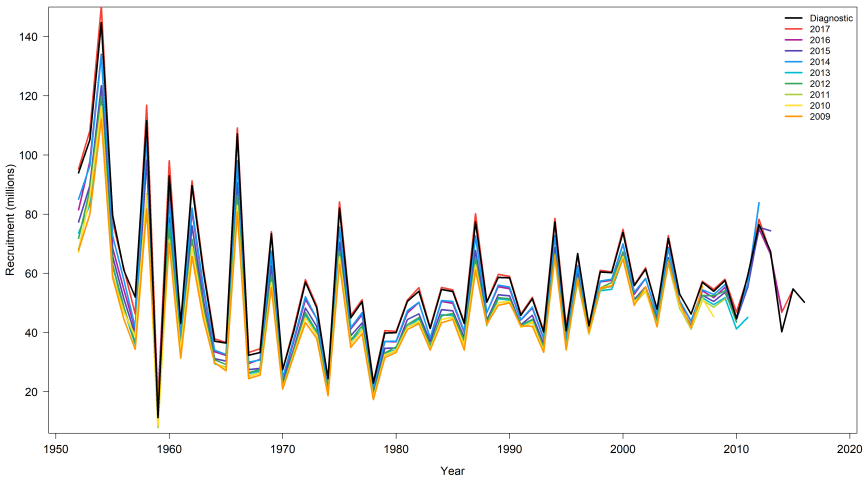
An additional diagnostic, the retrospective analysis, was used to assess the overall stability of the 2020 diagnostic case model and to identify any persistent biases in estimated quantities as a result of possible model mis-specification. A series of 9 additional models were fitted starting with the removing one year of data from the model (through 2017), followed by models with that sequentially "peeled" away all input data for the years 2018-2009. The models are named below by the final year of data included (e.g. 2009-2018). A comparison of the spawning biomass, depletion, and recruitment trajectories are shown in Figure 52. Additionally, Mohn's ρ was calculated to ascertain the degree of retrospective bias (Hurtado-Ferro et al., 2015). All values of ρ (SB = -0.161;$SB_{latest}/SB_{F=0}$ = -0.12; recruitment = -0.02) were less than $\pm$0.2 so there is not strong indication of retrospective bias. Note, that since recruitments in the final 6 quarters were fixed at the mean level Mohn's ρ was calculated based off of the year of last estimated recruitment, and as such the two most recent years for each model are not shown in Figure 52.



(a) Spawning potential retrospective



(b) Depletion retrospective



(c) Recruitment retrospective

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

11.4 Jitter

Using the final parameter estimates from the diagnostic case model as a starting point, 20 additional runs were launched. For each of these runs, a zero-centered random deviate was added to the original parameter (or a one-centered random deviate was multiplied against the original parameter depending on the link function used) to perturb the model from its initial solution. All major estimated parameter groups (movement, selectivity, recruitment, effort deviates, tag reporting rates, etc) were perturbed. The results of this analysis are shown in Figure 53. All models converged to approximately the same estimate of total depletion. However some models found a marginally better fit to the data relative to the diagnostic case in the process (and one a substantially worse fit), indicating some room for improvement with additional model iterations. When the diagnostic case was run for an additional 5,000 function evaluations of the minimizer, a marginally superior likelihood was achieved (exceeding all those of the jittered runs) but stock assessment outcomes and model estimated quantities were virtually unchanged. This provides some confidence that the model has converged to a stable solution.

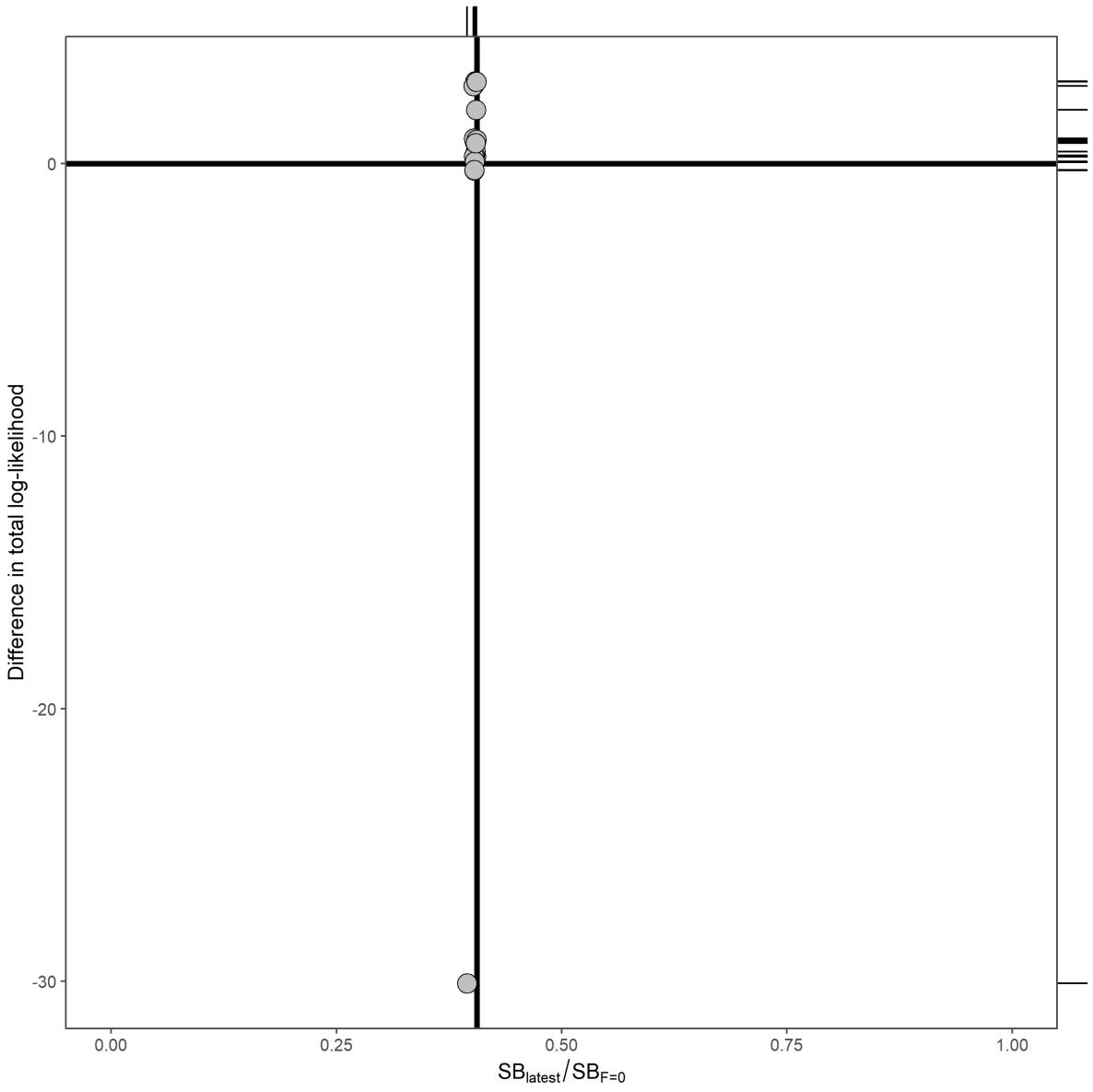


Figure 53: Results from the jittering analysis. Points above the horizontal line indicate better relative likelihood values. The vertical line indicates the estimate of $SB_{latest}/SB_{F=0}$ from the diagnostic plot. A rug plot is included in the margins to show the positions of points that are plotted on top of one-another.

11.5 Hessian diagnostic

The Hessian matrix for the diagnostic case model was not found to be positive definite. The Choleski decomposition of the Hessian matrix indicated 3 out of 11,421 eigenvalues were negative (-6.0072e-08, -4.9418e-08, -2.1068e-09). A non-positive definite Hessian is an indication that the model is likely over-parameterized and some parameters of the model poorly estimated, further supporting the model complexity issues raised in Section 8.4.2. To this end, investigation of the associated eigenvectors for these negative eigenvalues indicated that 4 regional recruitment deviates (Region 5 1958 Q1 and 1975 Q2; Region 6 1993 Q1; and Region 7 1997 Q3) were influential to the 3 negative eigenvalues indicating that these parameters were poorly estimated. As mentioned in Section 8.4.2, implementation of the orthogonal-polynomial parameterization feature in MFCL could help to resolve this issue.

11.6 Stochastic projections

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

- Stochastic 30 year projections were conducted from each assessment model within the uncertainty grid developed for the 2020 bigeye assessment.
- For each model, 40 stochastic projections were conducted, with future recruitments randomly sampled from historical deviates, resulting in 960 projections across the model grid.
- Two scenarios were considered for future recruitment in the projection period: long-term, based upon the period 1962 to 2017, and short-term, based upon the period 2008 to 2017.
- Purse seine fisheries were projected forwards assuming 2018 effort levels. All other fisheries were projected on catch, using the recent average value (2016 to 2018).
- The outputs of the projections (median $SB_{2048}/SB_{F=0}$ and F_{recent}/F_{MSY} , and risk $SB_{2045}/SB_{F=0} < \text{LRP}$) were calculated across the 24 model grid.
- Catchability (which can have a trend in the historical component of the model) was assumed to remain constant in the projection period at the level estimated in the terminal year of the assessment model.

Results of the projections are summarized in Tables 7 & 8 and Figures 54 & 55.

Table 7: Summary of bigeye tuna stock outcomes under 2016–2018 average fishing and short-term recruitment scenario.

$SB_{2025}/SB_{F=0}$	$SB_{2035}/SB_{F=0}$	$SB_{2048}/SB_{F=0}$	Risk $SB_{2048}/SB_{F=0} <$ LRP	F_{recent}/F_{MSY}
0.468	0.487	0.493	0%	0.619

Table 8: Summary of bigeye tuna stock outcomes under 2016–2018 average fishing and long-term recruitment scenario.

$SB_{2025}/SB_{F=0}$	$SB_{2035}/SB_{F=0}$	$SB_{2048}/SB_{F=0}$	Risk $SB_{2048}/SB_{F=0} <$ LRP	F_{recent}/F_{MSY}
0.415	0.435	0.445	5%	0.806

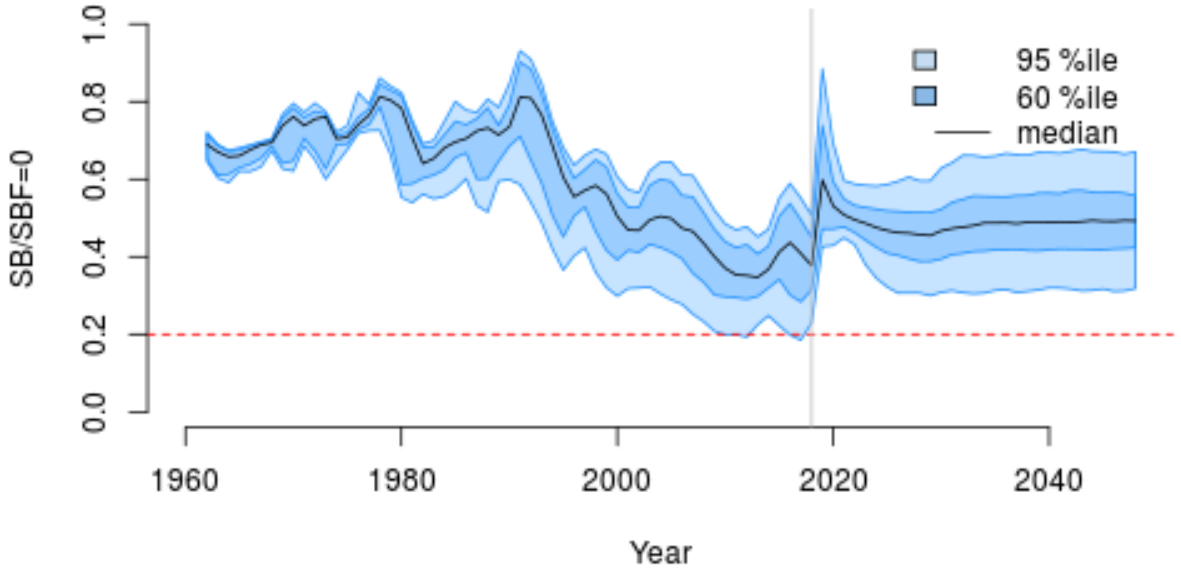


Figure 54: Time series of bigeye tuna spawning potential $SB_t/\overline{SB}_{F=0}$, where $\overline{SB}_{F=0}$ is the average SB from t-10 to t-1, from the uncertainty grid of assessment models for the period 2000 to 2018, and stochastic projection results for the period 2019 to 2048 assuming 2016–2018 average/2018 fishing levels continue. Vertical gray line at 2018 represents the last year of the assessment. During the projection period (2019-2048) levels of recruitment variability are assumed to match those over the short-term period (2008-2017). The red horizontal dashed line represents the agreed limit reference point.

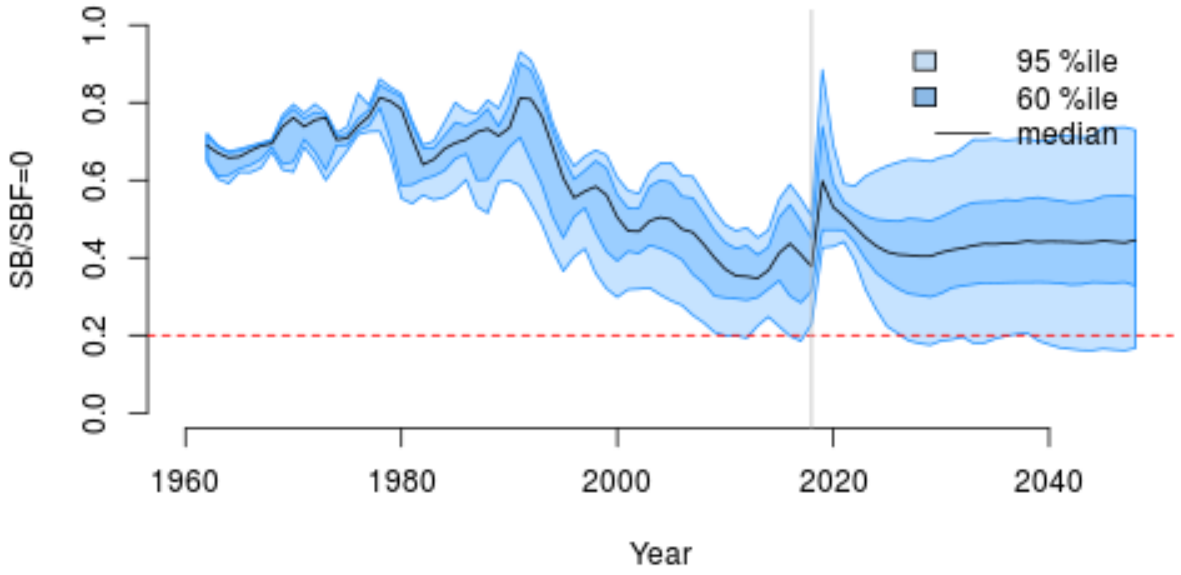


Figure 55: Time series of bigeye tuna spawning potential $SB_t/\overline{SB}_{F=0}$, where $\overline{SB}_{F=0}$ is the average SB from t-10 to t-1, from the uncertainty grid of assessment models for the period 2000 to 2018, and stochastic projection results for the period 2019 to 2048 assuming 2016–2018 average/2018 fishing levels continue. Vertical gray line at 2018 represents the last year of the assessment. During the projection period (2019-2048) levels of recruitment variability are assumed to match those over the long-term period (1962-2017). The red horizontal dashed line represents the agreed limit reference point.

11.7 Sensitivity analyses reference points and likelihood values

In the current assessment, inferences about stock status and recommendations for management advice are based on the structural uncertainty grid, rather than the diagnostic case model and the one-off sensitivity model runs. The estimates of reference points for the one-off sensitivity model runs are presented here in the appendix for relative comparisons against the diagnostic case model and among these models, rather than focusing on the absolute estimates that they provide. Care should be taken when comparing the likelihoods as in many cases, either the input data or likelihood structure has changed making valid comparisons impossible. The set of focal reference points for these models are presented below in three sets, along with those of the diagnostic case model:

Table 9: Reference points for the diagnostic case model and the one-off sensitivity models.

	Diagnostic	*EstRichardsO*	*EstRichardsT*	*Oto-Only*	*Size10*	*Size60*	*Size200*	*Size500*	*SSMULT*
C_{latest}	159,248	159,624	160,372	160,553	159,000	159,158	161,449	162,060	158,309
YF_{recent}	133,160	213,680	169,920	314,120	133,720	134,280	124,320	125,480	128,760
fmult	1.44	2.65	1.84	4.62	1.37	1.5	1.15	0.89	1.18
F_{MSY}	0.05	0.05	0.05	0.07	0.05	0.05	0.04	0.05	0.05
MSY	140,480	297,120	196,120	624,800	139,360	143,360	125,400	126,160	130,280
$F_{recent}/F_{\text{MSY}}$	0.7	0.38	0.54	0.22	0.73	0.67	0.87	1.12	0.85
$SB_{F=0}$	1,754,439	3,056,435	2,447,121	4,431,303	1,777,641	1,731,563	1,609,983	1,498,470	1,641,889
SB_{MSY}	394,100	751,400	597,400	978,900	393,600	397,700	371,400	340,600	331,400
$SB_{MSY}/SB_{F=0}$	0.22	0.25	0.24	0.22	0.22	0.23	0.23	0.23	0.2
$SB_{latest}/SB_{F=0}$	0.41	0.57	0.49	0.7	0.39	0.41	0.34	0.27	0.3
$SB_{latest}/SB_{\text{MSY}}$	1.81	2.31	2	3.16	1.74	1.8	1.48	1.17	1.48
$SB_{recent}/SB_{F=0}$	0.42	0.61	0.52	0.73	0.4	0.43	0.37	0.24	0.3
$SB_{recent}/SB_{\text{MSY}}$	1.88	2.46	2.14	3.29	1.82	1.88	1.61	1.06	1.5

Table 10: Reference points for the diagnostic case model and the one-off sensitivity models.

	Diagnostic	*h0.65*	*h0.95*	*M-low*	*M-mid*	*M-hi*	*Mix1*	*TagFree*
C_{latest}	159,248	159,252	159,236	159,430	158,346	163,550	158,454	160,658
YF_{recent}	133,160	143,160	127,280	130,360	143,040	150,800	134,720	145,400
fmult	1.44	1.29	1.58	1.36	1.76	2.08	1.5	1.58
F_{MSY}	0.05	0.04	0.05	0.04	0.05	0.06	0.05	0.05
MSY	140,480	147,680	137,000	135,560	161,400	182,360	143,720	157,920
$F_{recent}/F_{\text{MSY}}$	0.7	0.77	0.63	0.73	0.57	0.48	0.67	0.63
$SB_{F=0}$	1,754,439	1,908,636	1,662,166	1,807,438	1,513,517	1,257,092	1,708,763	1,954,192
SB_{MSY}	394,100	480,100	334,600	402,800	344,400	284,000	389,200	439,700
$SB_{MSY}/SB_{F=0}$	0.22	0.25	0.2	0.22	0.23	0.23	0.23	0.23
$SB_{latest}/SB_{F=0}$	0.41	0.37	0.43	0.39	0.46	0.49	0.42	0.44
$SB_{latest}/SB_{\text{MSY}}$	1.81	1.48	2.12	1.75	2	2.16	1.82	1.95
$SB_{recent}/SB_{F=0}$	0.42	0.39	0.44	0.4	0.48	0.53	0.42	0.47
$SB_{recent}/SB_{\text{MSY}}$	1.88	1.54	2.21	1.81	2.13	2.36	1.84	2.11

Table 11: Reference points for the diagnostic case model and the one-off sensitivity models.

	Diagnostic	*len456*	*idxAU-US*	*CPUE1962*	*Model1962*	*Alt*	*FreeSel*
C_{latest}	159,248	160,211	159,843	159,406	159,253	159,246	158,737
YF_{recent}	133,160	138,760	133,800	133,760	131,600	137,080	139,240
fmult	1.44	1.62	1.45	1.45	1.39	1.5	1.61
F_{MSY}	0.05	0.05	0.05	0.05	0.05	0.05	0.05
MSY	140,480	152,000	141,600	141,560	137,680	146,320	152,440
$F_{recent}/F_{\text{MSY}}$	0.7	0.62	0.69	0.69	0.72	0.67	0.62
$SB_{F=0}$	1,754,439	1,862,959	1,761,436	1,770,778	1,733,509	1,807,864	1,877,848
SB_{MSY}	394,100	432,200	397,100	399,000	386,800	406,900	431,700
$SB_{MSY}/SB_{F=0}$	0.22	0.23	0.23	0.23	0.22	0.23	0.23
$SB_{latest}/SB_{F=0}$	0.41	0.44	0.41	0.41	0.4	0.42	0.45
$SB_{latest}/SB_{\text{MSY}}$	1.81	1.89	1.81	1.83	1.78	1.85	1.94
$SB_{recent}/SB_{F=0}$	0.42	0.46	0.43	0.43	0.41	0.43	0.47
$SB_{recent}/SB_{\text{MSY}}$	1.88	1.99	1.89	1.9	1.86	1.92	2.03

Table 12: Likelihood components for the diagnostic case model and the one-off sensitivity models.

	Diagnostic	EstRichardsO	EstRichardsT	Oto-Only	Size10	Size60	Size200	Size500	SSMULT
BH Steepness	0.53	0.39	0.46	0.61	0.51	0.64	1.18	2.71	0.67
Effort devs	2,309	2,529	2,387	2,684	2,664	1,957	1,558	1,694	2,676
Length composition	-193,515	-194,259	-193,998	-194,429	-218,432	-150,001	-99,593	-60,570	0
Weight composition	-1,946,245	-1,945,716	-1,946,012	-1,944,560	-2,175,256	-1,575,297	-1,164,290	-850,432	0
Tag data	6,053	5,862	5,904	6,351	6,374	5,763	5,569	5,551	6,051
Total	2,131,398	2,131,583	2,131,718	2,129,953	2,384,650	1,717,577	1,256,754	903,755	8,728

Table 13: Likelihood components for the diagnostic case model and the one-off sensitivity models.

	Diagnostic	h0.65	h0.95	M-low	M-mid	M-hi	Mix1	TagFree
BH Steepness	0.53	0.57	0.52	0.55	0.5	0.5	0.53	0.54
Effort devs	2,309	2,309	2,309	2,399	2,321	2,281	2,349	2,163
Length composition	-193,515	-193,515	-193,515	-193,549	-193,517	-193,518	-193,440	-193,800
Weight composition	-1,946,245	-1,946,245	-1,946,246	-1,946,225	-1,946,281	-1,946,285	-1,946,234	-1,946,438
Tag data	6,053	6,053	6,049	6,055	6,042	6,045	6,427	0
Total	2,131,398	2,131,399	2,131,402	2,131,320	2,131,434	2,131,476	2,130,898	2,138,075

Table 14: Likelihood components for the diagnostic case model and the one-off sensitivity models.

	Diagnostic	len456	idxAU-US	CPUE1962	Model1962	Alt	FreeSel
BH Steepness	0.53	0.54	0.55	0.52	0.52	0.52	0.52
Effort devs	2,309	2,490	2,292	2,127	2,125	2,306	2,317
Length composition	-193,515	-291,200	-193,547	-193,572	-193,535	-193,432	-193,513
Weight composition	-1,946,245	-1,783,560	-1,915,312	-1,908,096	-1,871,101	-1,946,253	-1,946,359
Tag data	6,053	6,085	6,045	6,045	6,053	6,241	6,064
Total	2,131,398	2,066,184	2,100,521	2,093,495	2,056,457	2,131,138	2,131,490

11.8 Simplified assessment structure

A single region model was developed in parallel to the 2020 bigeye tuna stock assessment given the aforementioned issues with model complexity (Section 8.4.2), and also as a response to model sensitivity to the treatment of the tagging data that was observed in the stepwise model development (Section 7.1). As mentioned in the discussion (Section 8.4.2), there are number of reasons where it is desirable to have a spatial structure in a stock assessment model one of which is to allow for components of the stock to be partitioned in terms of fishing pressure and relative abundance. Under the current assessment structure, MFCL has considerable flexibility to fit the data by manipulating regional biomass through the interaction of seasonal movement and quarterly recruitment deviates. If these are poorly estimated or mis-specified it could result in spatial regions inappropriately "buffering" the effects of more depleted regions. The results presented in this section show the outcomes from a model that makes a drastically different set of assumptions on the population dynamics of WCPO bigeye tuna, specifically that the stock is well-mixed, and there is no capacity for spatial buffering and/or spatial partitioning of abundance and fishing mortality. This simplifying assumption could introduce a pessimistic bias to model estimates of stock status.

The simplified single-region structure for bigeye tuna took the 2020 diagnostic case model and turned it into a simplified fleets-as-areas approach. The 41 fisheries were combined into 15 for the single region model: northern longline (Fisheries 1 & 2), US longline (Fishery 3), offshore longline (Fisheries 5 & 6), equatorial longlines (Fisheries 4, 8 & 9), western longline (Fishery 7), southern longline (Fisheries 11, 12 & 29), Australian longline (Fisheries 10 & 27), associated purse seines (Fisheries 13, 15, 25, 30, & 24), unassociated purse seines (14, 16, 26, & 31), domestic miscellaneous fisheries (Fisheries 17, 23, & 22), domestic handline (Fishery 18), northern Japanese purse-seine (Fishery 19), northern Japanese pole-and-line (Fishery 20), equatorial pole-and-line (Fisheries 21, 22, & 28), and the index fishery (Fisheries 33 - 41). For each of the 15 new fisheries, catch and size frequency data were aggregated together from the appropriate fisheries. The standardized CPUE for the index fishery was calculated over the entire WCPO assessment region using the same model as described in Ducharme-Barth and Vincent (2020). To account for the new interpretation of selectivities, all non-decreasing penalties were removed from the longlines except for the index fishery. It was assumed that since the index fishery covered the entire WCPO it would "sample" the largest fish if they were in the system. The domestic handline fishery was assumed to have non-decreasing selectivity as well since this fishery was unchanged from the 2020 diagnostic case and regularly caught the largest fish in that model. To aide in the estimation of some of the longline fishery selectivities, the selectivity of the first 3 quarterly ages (prior to interaction with the gear) were fixed to zero. The estimated selectivities for these 15 fisheries in the simplified single region model are shown in Figure 56. In the 2020 diagnostic case the seasonal dynamics were driven using the quarterly movements. To compensate for the lack of movement, a seasonal catchability component was estimated for the index fishery.

Predictably, a more parsimonious model was able to more efficiently achieve better convergence[9] through a reduction in the complexity. Additionally, though simplifications were made, the model was able to achieve a reasonable fit to the CPUE index (Figure 57) and the size composition data (Figures 58 & 59). It was also able to estimate similar trends in spawning potential and depletion. However, due to the lower estimated scale of the population the depletion estimates were considerably more pessimistic than the 2020 diagnostic case. Results from the single region model structure are shown for some of the same key one-off sensitivities tested in the 2020 bigeye tuna stock assessment in terms of spawning potential (Figure 60) and depletion (Figure 61).

[9]Choleski decomposition of the Hessian matrix for the simplified assessment structure showed that the Hessian was positive definite.

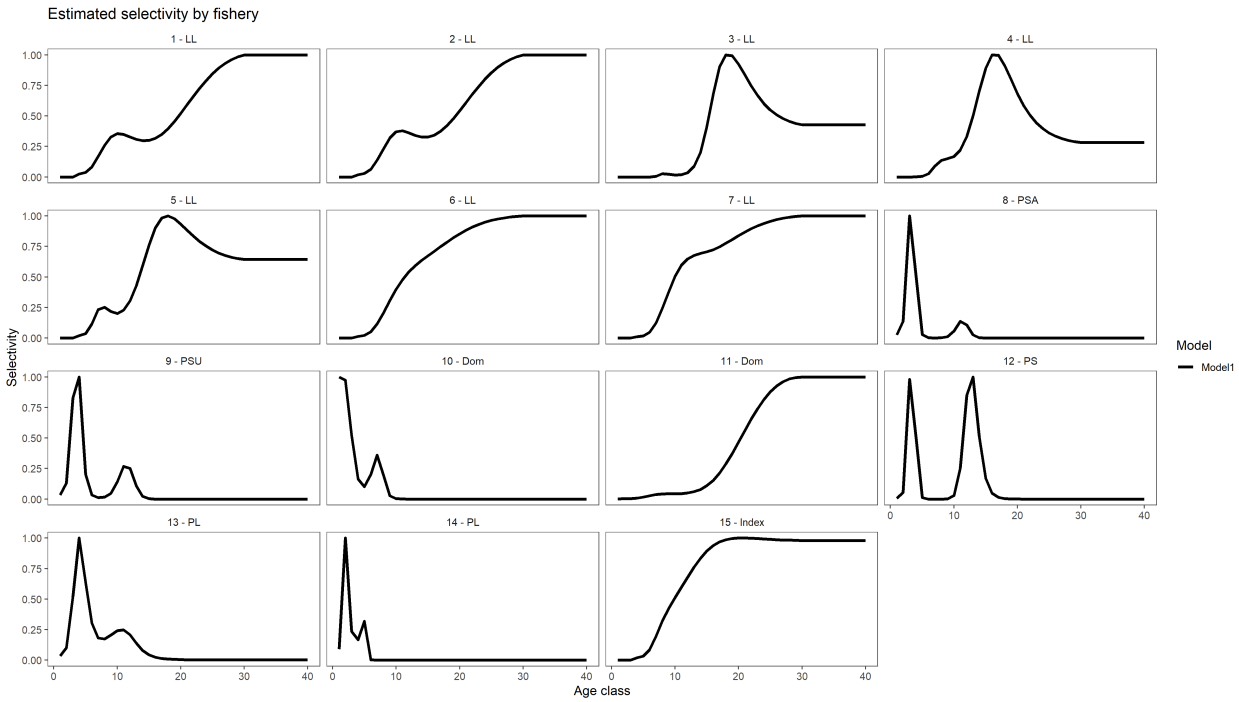


Figure 56: Selectivity at age for estimated for the 15 fisheries in simplified single region model.

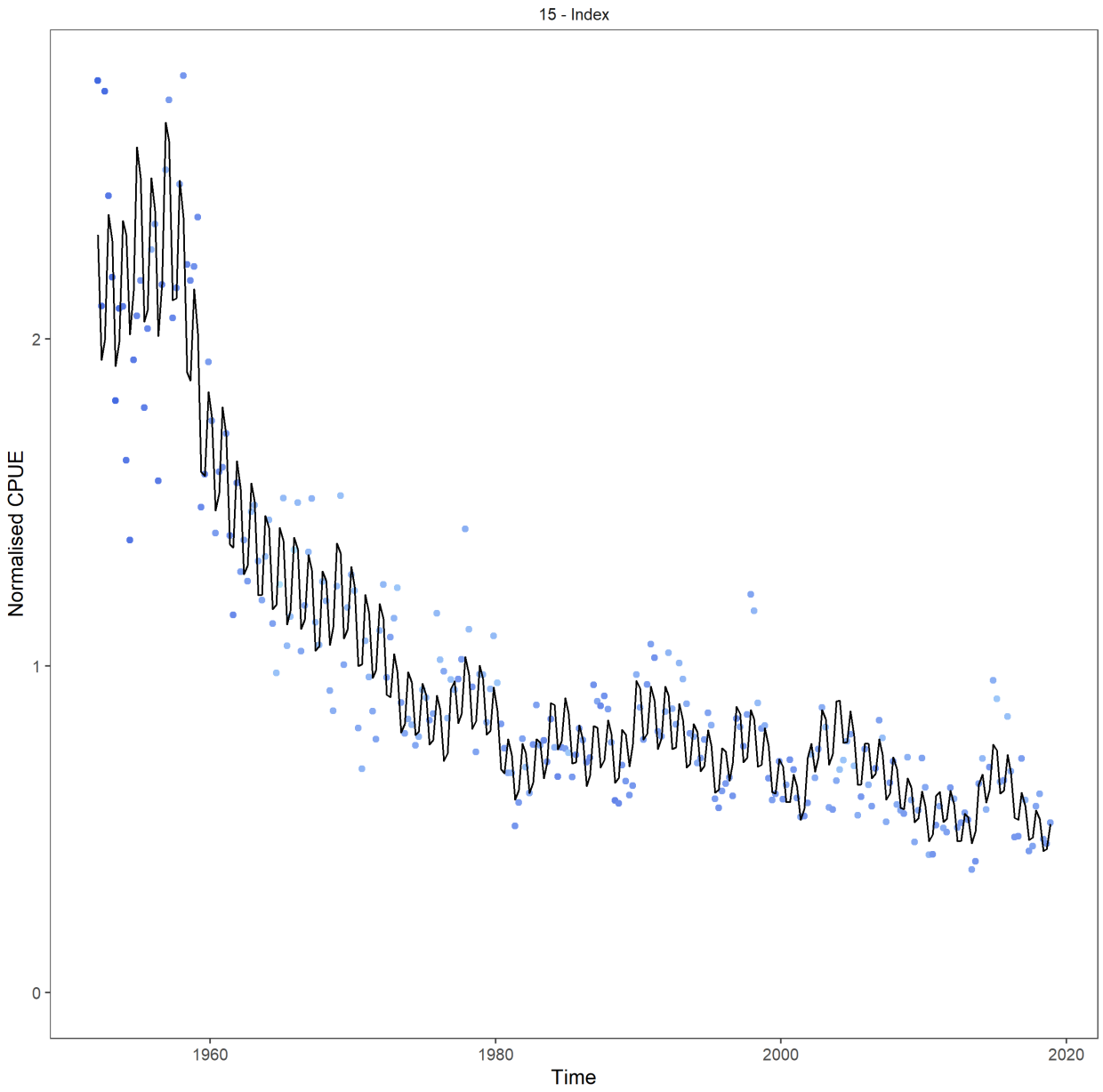


Figure 57: Fit to the standardized CPUE from the simplified single region model. Observed CPUE in the blue dots, and the model prediction in the black line.

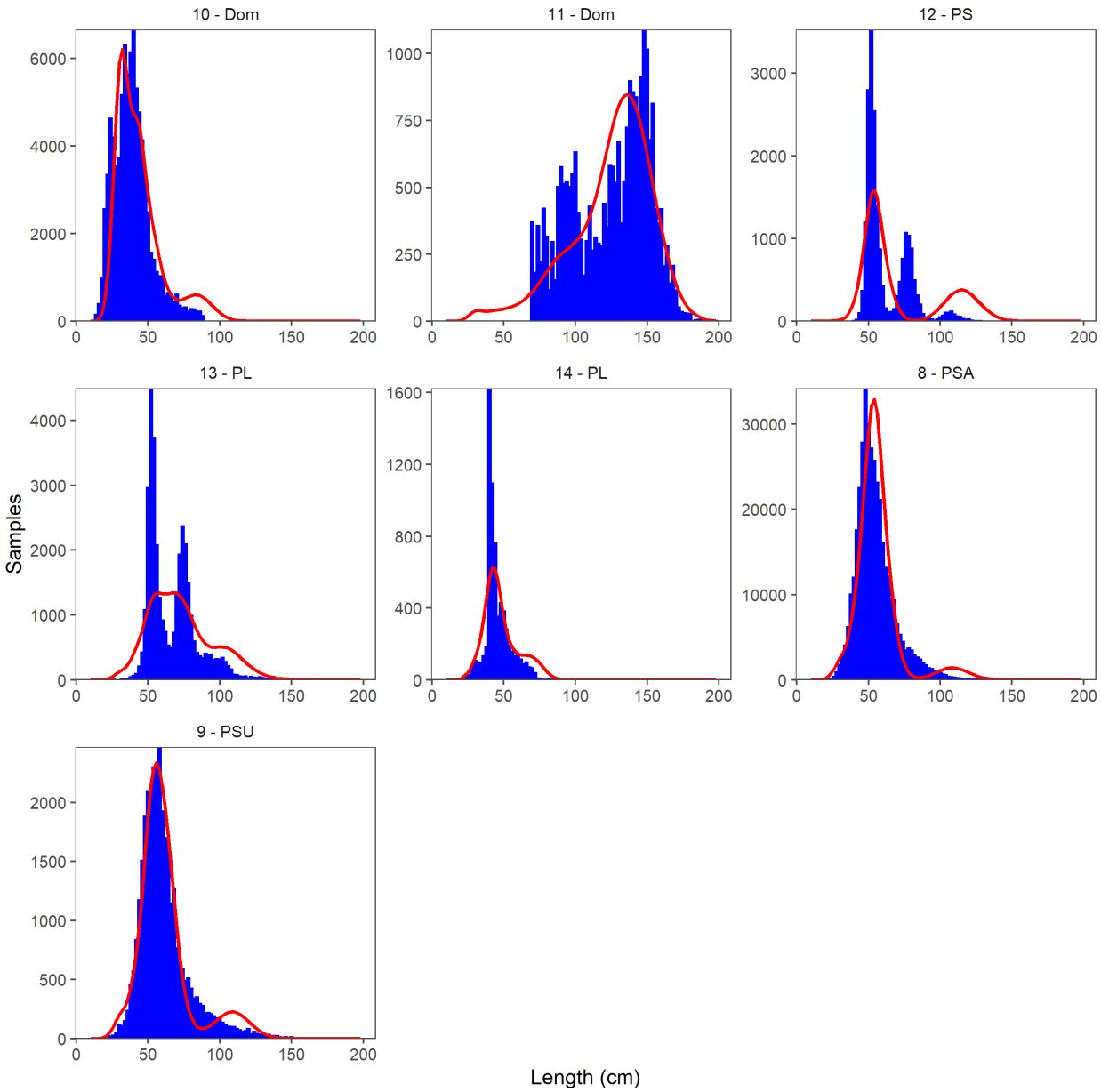


Figure 58: Fit to the length composition for the simplified single region model.

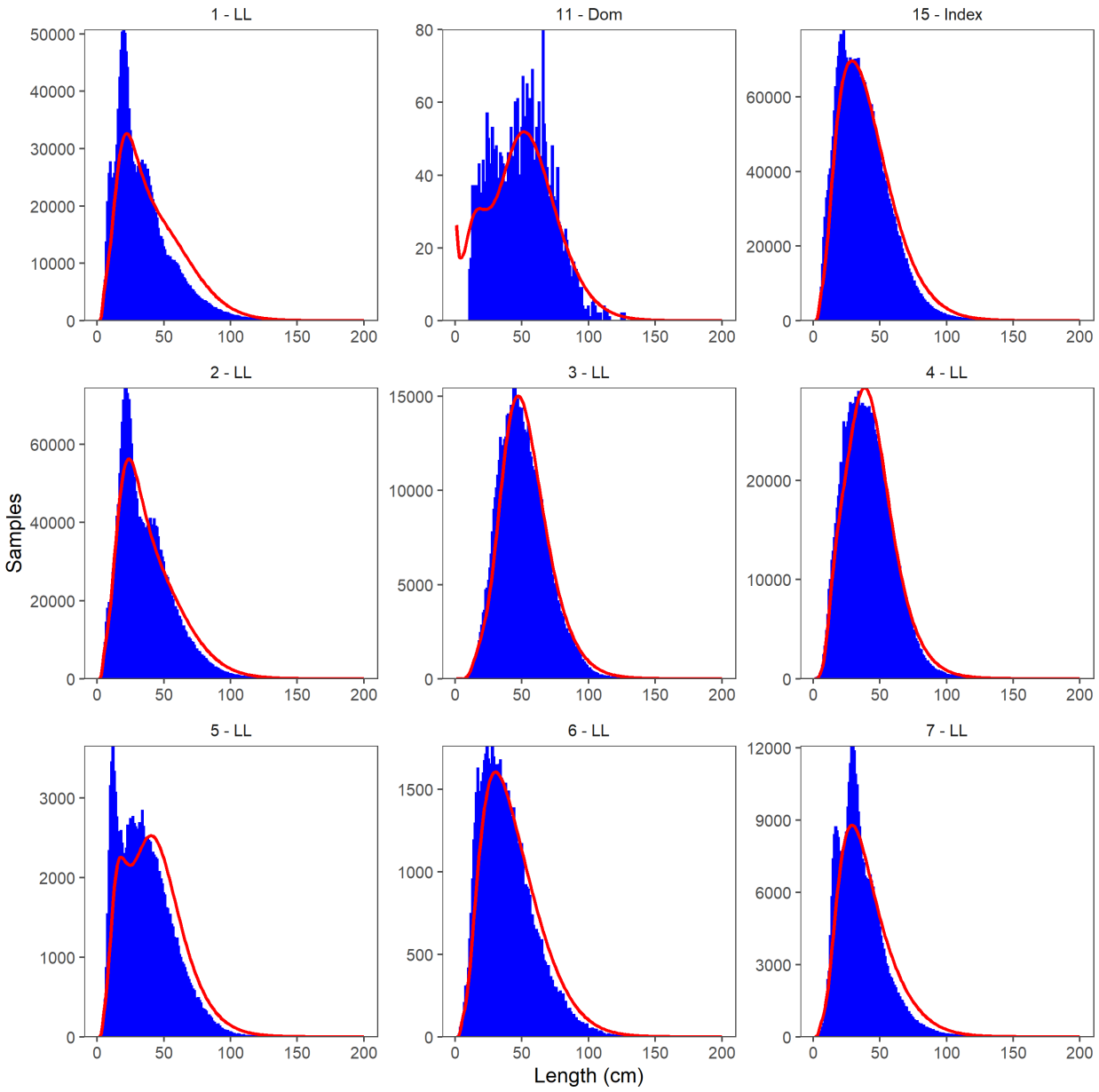


Figure 59: Fit to the weight composition for the simplified single region model.

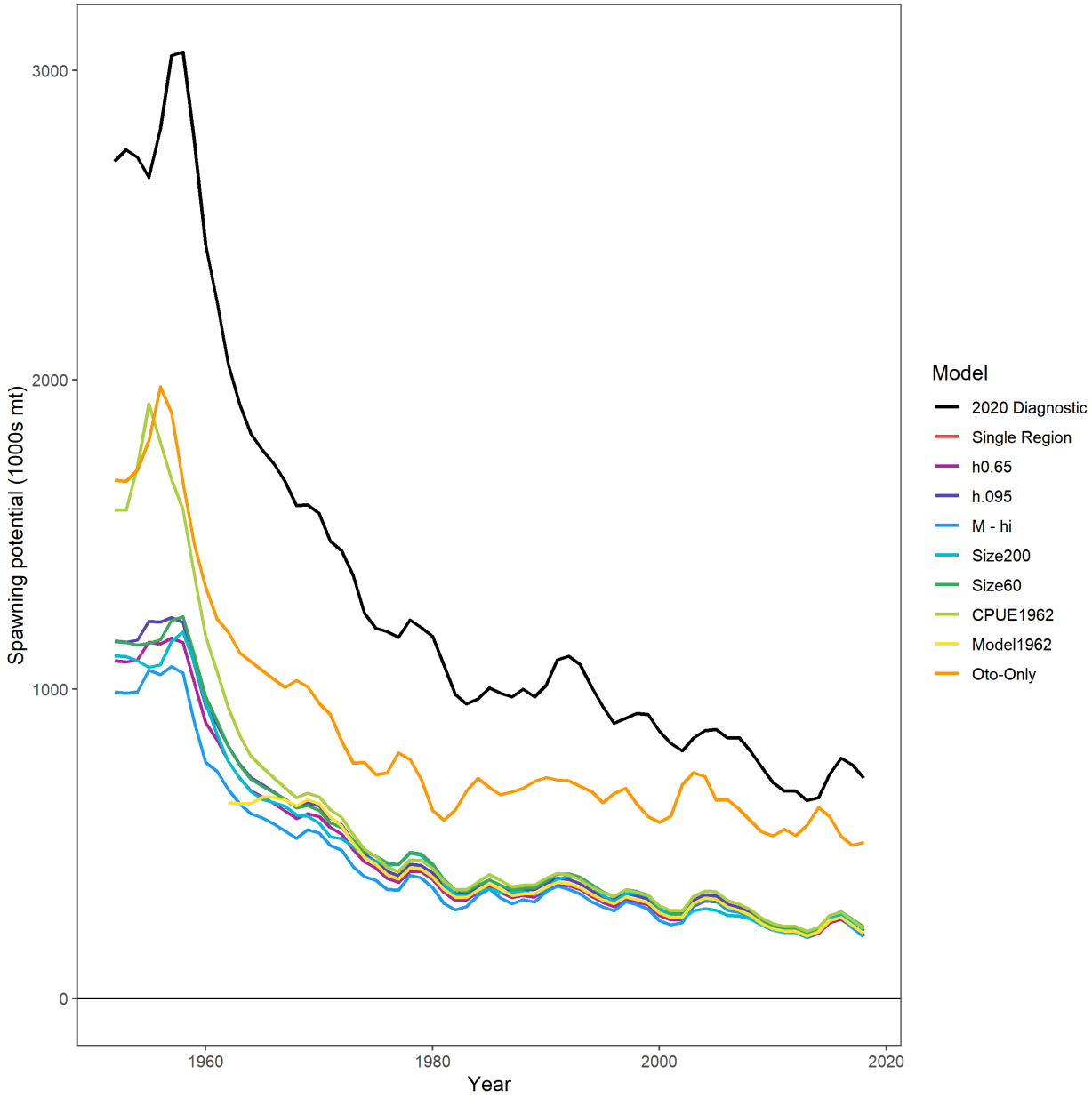


Figure 60: Estimated spawning potential for the simplified single region model and associated one-off sensitivities from that model. The 2020 diagnostic case is shown in black for reference.

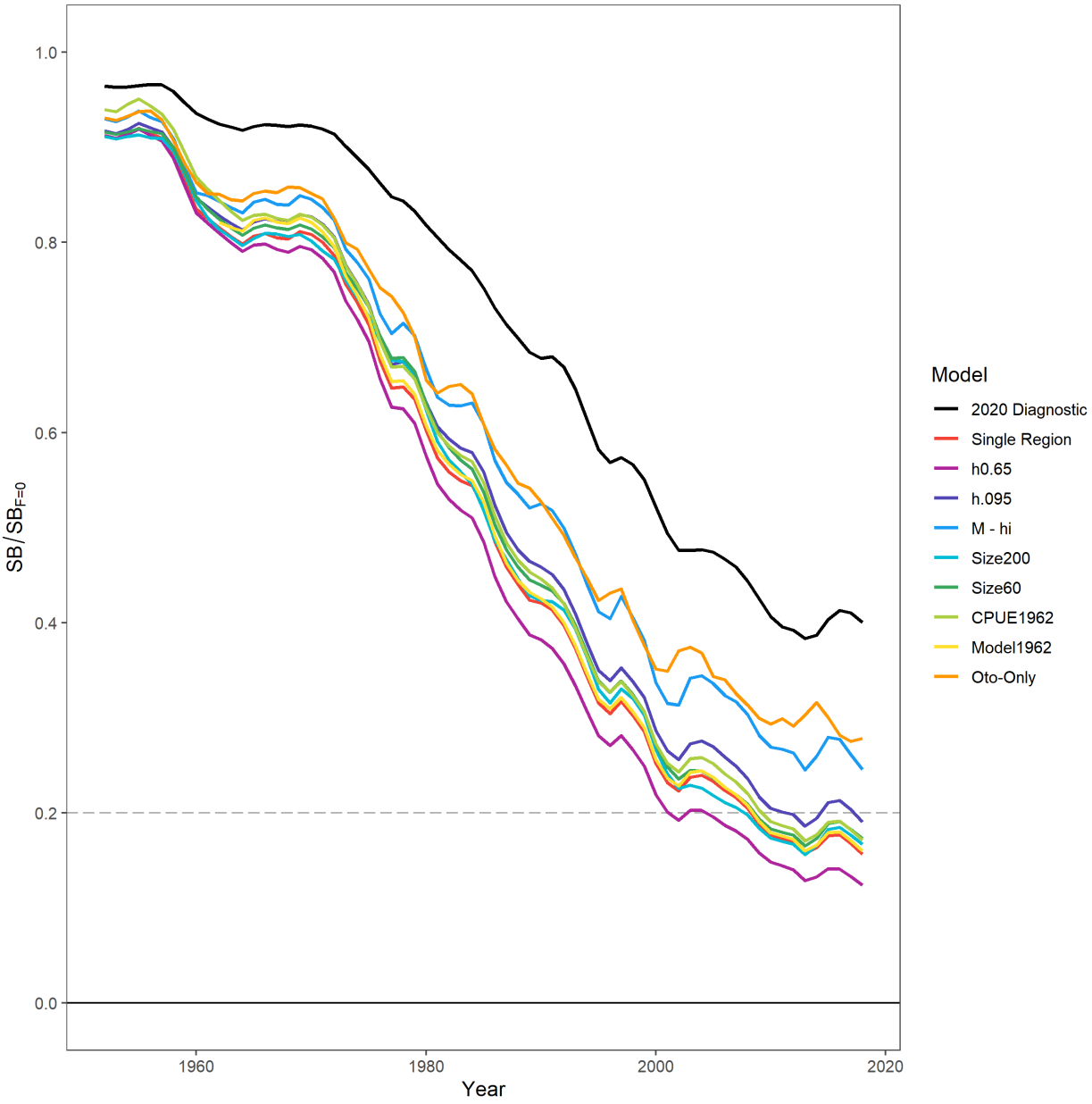


Figure 61: Estimated depletion ($SB_{latest}/SB_{F=0}$) for the simplified single region model and associated one-off sensitivities from that model. The 2020 diagnostic case is shown in black for reference.