SCIENTIFIC COMMITTEE
THIRTEENTH REGULAR SESSION
Rarotonga, Cook Islands
9–17 August 2017

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

WCPFC-SC13-2017/SA-WP-06
Rev1 August 4th

L. Tremblay-Boyer[1], S. McKechnie, G. Pilling, J. Hampton

[1]Oceanic Fisheries Programme, The Pacific Community

Details of changes made during the revision of the original report

Minor changes have been made to the stock assessment report for yellowfin tuna that had previously been uploaded to the WCPFC scientific committee website. These relate to the addition of an extra level to the size data weighting axis in the structural uncertainty grid. Due to time constraints, only the two bounds of the size data weightings (divisors of 10 and 50) that were investigated in the one-off sensitivity model runs were initially included in the grid, due to the large computational overhead, limited computational resources and time constraints imposed on the assessments. Subsequent to the report deadline, we have completed the model runs for the extra level (divisor of 20; the level used in the diagnostic case model) and so modify the stock assessment report to incorporate summaries that include these extra runs. These, along with other minor corrections can briefly be summarised as:

The changes made to the yellowfin stock assessment report can be summarised as:

- An extra 24 models were run as part of the structural uncertainty grid, with all extra models using a divisor of 20 for the size data weighting. These model runs had a small effect on the summaries of the grid as, even though the extra level of the size weighting axis falls between the more extreme divisors of 10 and 50, the resulting model runs behave similarly to that of the divisor of 10, thus making reference points more optimistic by a 2-4 points. The new figures are shown in Appendix Section 11.5 and the figure numbers are set to be directly comparable to the original figures in the main text (e.g. Figures A39–41 are equivalent to Figures 39–41). The new tables are also in Appendix Section 11.5 and the same numbering procedure is used (Tables A6, A15 and A16 are equivalent to Tables 6, 15 and 16 in the main text).
- All references to "spawning biomass" in the tables and figures have been corrected to "spawning potential". These occurred for Table 4, and Figures 13, 32–33 and 41–44.
- The map showing the distribution and magnitude of yellowfin tuna catches for the most recent decade of the assessment (Figure 6) has been updated.

Contents

1 Executive Summary

This paper describes the 2017 stock assessment of yellowfin tuna *Thunnus albacares* in the western and central Pacific Ocean. The model time period now extends to the end of 2015, adding a further three years of data since the last stock assessment was conducted in 2014. New developments to the stock assessment include addressing relevant recommendations of the 2014 yellowfin stock assessment report (Davies et al., 2014), investigation of an alternative regional structure, exploration of uncertainties in the assessment model, particularly in response to the inclusion of additional years of data, and improving diagnostic weaknesses of previous assessments.

This assessment is supported by additional analyses of catch-per-unit-effort data for longline fisheries (Tremblay-Boyer and Pilling, 2017a,b), tagging data (McKechnie et al., 2017b), and the data summaries for fisheries definitions used in the stock assessment (McKechnie et al., 2017b).

Changes made in the progression from the 2014 reference case to 2017 diagnostic case models include:

1. The 2014 reference case model.
2. The 2014 reference case model with the new MFCL executable.
3. A complete update of the 2014 reference case model - all inputs extended from 2012 to 2015 using identical methodology for CPUE, tagging, size frequencies etc, and the same MFCL model settings.
4. The previous model with the same structure and MFCL settings but CPUE indices using the GLM approaches with the updated Pacific-wide operational LL database (McKechnie et al., 2017b).
5. The previous model with the same MFCL settings but with the new regional structure and consequently all fisheries, and input data (including CPUE standardisations), reconfigured based on these new regional definitions.
6. The previous model with two modifications to the recruitment estimates; the change from quarterly to annual recruitments when estimating the spawner-recruit relationship, and the fixed terminal six recruits set at the arithmetic rather than geometric mean of recruitments for the remaining period.

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 the most important axes of uncertainty from the one-off models were included. In comparison to previous assessments, little emphasis is placed on the diagnostic case model. Instead it is recommended that management advice is formulated from the results of the structural uncertainty grid.

Across the range of model runs in this assessment, the key factor influencing estimates of stock status was the size data weighting value. Downweighting the influence of the size data led to more pessimistic stock status estimates.

Based on the results of the model grid, the general conclusions of this assessment are as follows:

1. The grid contains a wide range of models with some variation in estimates of stock status, trends in abundance and reference points. However, biomass is estimated to have declined

throughout the model period for all models in the grid. Those declines are found across most tropical and temperate regions of the model.

2. Across the model grid, the terminal depletion estimated for the majority of runs estimate stock status levels to be above the $20\%SB_{F=0}$. The range of $SB_{latest}/SB_{F=0}$ values was 0.18 to 0.45. Only two runs (<5%) fell below the LRP of $20\%SB_{F=0}$. The median estimate (0.33) is comparable to that estimated from the 2014 assessment grid, noting the differences in grid uncertainty axes used in the two assessments.

3. Corresponding estimates of F_{recent}/F_{msy} ranged from 0.58 to 1.13, with 2 out of the 48 runs (<5%) indicating that $F_{recent}/F_{msy} > 1$. The median estimate (0.75) is also comparable to that estimated from the 2014 asssessment grid.

4. Fishing mortality for adult and juvenile yellowfin tuna is estimated to have increased continuously since the beginning of industrial tuna fishing (seen in the diagnostic case model). In general these have been on average higher for juveniles, but in recent years adult fishing mortality has also increased. A significant component of the increase in juvenile fishing mortality is attributable to the Philippines, Indonesian and Vietnamese surface fisheries, which have the most uncertain catch, effort and size data. The work of the WPEA project to assist in enhancing the current fishery monitoring programme and improving estimates of historical and current catch from these fisheries remains important given the contribution of these fisheries in the overall fishing impact analyses from this assessment.

5. The significance of the recent increased recruitment events and the progression of these fish to the spawning potential component of the stock are encouraging, although whether this is a result of management measures for the fishery or beneficial environmental conditions is currently unclear. It is noteworthy, however, that recent favourable recruitment events have also been estimated for skipjack (McKechnie et al., 2016a) and bigeye (McKechnie et al., 2017a) in the WCPO, and bigeye in the EPO (Aires-da Silva et al., 2017), which may give weight to the favourable environmental conditions hypothesis. Whether these trends are maintained in coming years will help separate these factors and will likely provide more certainty about the future trajectories of the stock.

6. There remains a range of other model assumptions that should be investigated either internally or through directed research. Briefly, the apparent non-linear impact of the weighing on the size composition data on population estimates, and the conflict between the abundance indices and the tagging data for region 8 are worthy of note. Also, biological studies to improve our estimates of growth of yellowfin within the WCPO, for instance through direct ageing of otoliths as was done in bigeye, should be considered a high priority.

2 Introduction

This paper presents the current stock assessment of yellowfin tuna (*Thunnus albacares*; YFT) in the western and central Pacific Ocean (WCPO, west of 150° W). The first assessment for this species was conducted in 1999 and the most recent iterations are documented in Hampton and Kleiber (2003), Hampton et al. (2004, 2005, 2006), (Langley et al., 2007, 2009, 2011) and (Davies et al., 2014).

The methodology used for the assessment is that commonly known as MULTIFAN-CL[2] (MFCL; Fournier et al., 1998; Hampton and Fournier, 2001; Kleiber et al., 2017), which is software that is routinely used by the Pacific Community (SPC) and which implements a size-based, age- and spatially-structured population model. Parameters of the model are estimated by maximizing an objective function consisting both of likelihood (data) and prior information components (penalties). Structural uncertainty is described over a range of key model assumptions to produce a 'grid' with all combinations of the most probable settings for these assumptions.

The current assessment incorporates the most recent data from yellowfin fisheries across the WCPO and maintains the general model structure of the recent assessments. Key new data that were included and/or examined include a Pacific-wide operational longline dataset and the Japanese tagging program data. We also examined the sensitivity of key assessment results to a range of model assumptions, principally related to biological assumptions and uncertainty in the various input data sets. The assessment has also benefited from developments addressing the recommendations from the independent review of the 2011 bigeye assessment (Ianelli et al., 2012), those made in the 2014 assessment report (Davies et al., 2014), and recommendations of the 2017 pre-assessment workshop held in Noumea over 24–27 April, 2017 (PAW; Pilling and Brouwer, 2017). The PAW reviewed the main input data sets and provided recommendations regarding the range of assessment model options and sensitivities to be included within the stock assessment. These recommendations provided the main direction for the current assessment.

This assessment is supported by several other analyses which are documented separately, but should be considered in reviewing this assessment. These include: improved purse seine catch estimates (Hampton and Williams, 2017), reviews of the catch statistics of the component fisheries (Williams and Terawasi, 2017), standardised CPUE analyses of operational level catch and effort data (McKechnie et al., 2017b; Tremblay-Boyer and Pilling, 2017a,b; Bigelow and Garvilles, 2017), and preparation of tagging data and size composition data (McKechnie et al., 2017b).

The objectives of the assessment are to estimate population parameters, such as time series of recruitment, biomass, fishing mortality and depletion, which indicate the status of the stock and impacts of fishing. We summarize the stock status in terms of reference points adopted by the Western and Central Pacific Fisheries Commission (WCPFC), and the distribution of reference point ratios is presented to describe their uncertainty.

3 Background

3.1 Stock structure

For the purpose of yellowfin assessments, the stock within the domain of the model area (essentially the WCPO, west of 150° W) has been considered as a discrete stock unit (Langley et al., 2007, 2009,

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

2011; Davies et al., 2014). This area has been disaggregated into model regions so as to describe to some extent spatial processes (such as recruitment and movement) and fishing mortality within regions. Information about stock structure includes a large amount of tagging data (1989-2015), from which the movement of tagged fish among regions can be used to estimate movement coefficients. This information indicates extensive longitudinal movements among the equatorial regions but also a level of latitudinal movements to and from the sub-tropical latitudes (Figure 1). As in the 2014 assessment, the model domain is disaggregated into 9 regions (Figure 2 and Figure 3), and this is described further in Section 4.1.

Since 2006, yellowfin tuna have been tagged throughout the WCPO by the Pacific Tuna Tagging Programme. From 2008, yellowfin tagging has occurred during bigeye-focussed tagging cruises in the equatorial central Pacific, between 180° and 140° W, while significant numbers of yellowfin have also been tagged in western Pacific and PNG tagging cruises between 2006 and 2013. The new regional structure proposed for the current stock assessment, with region boundaries shifted from 20° N to 10° N, was suggested by the PAW based on few movements between tropical tag release sites and temperate zones for bigeye tuna (McKechnie et al., 2017a). This applies to an extent to yellowfin also, noting that we need to have the same regional and fishery definition for both species to facilitate management analyses. The consequences of the change in region configuration were examined in detail during this assessment.

The results of genetic studies into yellowfin stock structure within the Pacific have provided varying results, with earlier publications suggesting limited variation within the Pacific (Ward et al., 1994). However, recent studies with improved techniques have suggested a finer scale genetic stock structure than indicated by earlier work and current patterns in the tagging data. As an example, Aguila et al. (2015) found that yellowfin tuna within the Philippines were a separate stock from the Bismarck Sea. Similarly, Grewe et al. (2015) identified genetically distinct populations of yellowfin tuna in the Coral Sea, Tokelau and Baja California (see also Grewe et al., 2016). The results of this field of research are being monitored for future assessments.

3.2 Biological characteristics

Yellowfin tuna are distributed throughout the tropical and sub-tropical waters of the Pacific Ocean. As mentioned earlier, there is some indication of restricted mixing between the western and eastern Pacific based on analysis of tagging data, while genetic samples also suggest spatial separation at potentially finer scales (Ward et al., 1994; Grewe et al., 2015). Adults (larger than about 100 cm) spawn, probably opportunistically, in waters warmer than 26° C (Itano, 2000), while juvenile yellowfin are first encountered in commercial fisheries (mainly surface fisheries in Philippines and eastern Indonesia) at several months of age.

Yellowfin tuna are relatively fast growing, and have a maximum fork length (FL) of about 180 cm. The growth of juveniles departs from von Bertalanffy type growth with the growth rate slowing between about 40 and 70 cm FL (Lehodey and Leroy 1999). The first age-class has a mean fork length of around 25 cm and is approximately three months of age according to analysis of daily structures on otoliths (Lehodey and Leroy 1999). It is recognised there are possibly regional differences in growth rate for yellowfin tuna. There is some indication that young yellowfin may grow more slowly in the waters of Indonesia and the Philippines than in the wider area of the WCPO (Yamanaka 1990). This is further supported by the comparison between the growth rates derived from WCPO yellowfin stock assessment (Hampton et al. 2006) and the growth rates derived from a MFCL model

that included only the single western, equatorial region (region 3) (Langley et al. 2007, Figure 2). The growth rates from the western equatorial region alone were considerably lower than from the WCPO, with the former growth rates more consistent with the growth of yellowfin in the southern Philippines waters (Yamanaka 1990) and growth increments from tag release/recovery data. On the other hand, the growth rates from the WCPO MFCL model are more consistent with the growth rates determined from daily growth increments from a collection of otoliths collected from a broad area of the equatorial WCPO (Lehodey and Leroy 1999). However, an examination of region-specific growth was beyond the scope of this assessment and the importance of this feature upon the stock assessment results is unknown.

The natural mortality rate is strongly variable with size, with the lowest rate of around 0.6-0.8 yr^{-1} being for pre-adult yellowfin 50-80 cm FL (Hampton 2000). Tag recapture data indicate that significant numbers of yellowfin reach four years of age. The longest period at liberty for a recaptured yellowfin, tagged in the western Pacific at about 1 year of age, is currently 6.5 years. For the purpose of computing the spawning biomass, we assume a fixed maturity schedule consistent with the observations of Itano (2000).

3.3 Fisheries

Yellowfin tuna, an important component of tuna fisheries throughout the WCPO, are harvested with a wide variety of gear types, from small-scale artisanal fisheries in Pacific Island and southeast Asian waters to large, distant-water longliners and purse seiners that operate widely in equatorial and tropical waters. Purse seiners catch a wide size range of yellowfin tuna, whereas the longline fishery takes mostly adult fish.

The annual yellowfin tuna catch in the WCPO increased from 100,000 mt in 1970 to about 600,000 mt in recent years, with the 2015 catch (the last year of this assessment) being just over 600,000 mt, and 2016 catch at a record of about 650,000 mt (Figure 4; Williams and Terawasi, 2017). Purse seiners harvest the majority of the yellowfin tuna catch (52% in 2015), while the longline fleet accounted for 15-20% of the catch in recent years, primarily in the equatorial regions (Figure 5). The remainder of the catch is dominated by the domestic fisheries of the Philippines and Indonesia, principally catching smaller individuals using a variety of small-scale gear types (e.g. pole-and-line, ringnet, gillnet, handline and seine net) but noting that small to medium sized purse seiners based in those countries also catch fish of sizes more typical of purse seine fisheries elsewhere.

Yellowfin tuna usually represent 17–21% of the overall purse-seine catch in recent years, and may contribute higher percentages of the catch in individual sets. Yellowfin tuna is often directly targeted by purse seiners, especially within unassociated schools which accounted for 50-65% of the recent (2012-2015) yellowfin purse-seine catch.

Since 2000, annual catches by longline vessels have varied between 80,000 and 105,000 mt. This has generally been below the relatively high levels of catch in the late 1970s and 1980s (which peaked at about 110,000 mt), and may reflect declines in CPUE related to reduced abundance and possibly to changes in targeting practices by some of the larger fleets (Figure 4). Annual catches from the domestic fisheries of the Philippines and eastern Indonesia are highly uncertain, particularly prior to 1990. Recent estimates for troll + other gears have reached approximately 140,000 mt in 2015 (Figure 4).

Figure 6 shows the spatial distribution of yellowfin tuna catch in the WCPO for the past 10 years.

Most of the catch is taken in western equatorial areas, with declines in both purse-seine and longline catch towards the east. The east-west distribution of catch is strongly influenced by ENSO events, with larger catches taken east of 160 ° E during El Niño episodes. Catches from outside the equatorial region are relatively minor (5%) and are dominated by longline catches south of the equator and purse-seine and pole-and-line catches in the north-western area of the WCPO (Figure 5 and Figure 6).

4 Data compilation

The data used in the yellowfin tuna assessment using MFCL consist of catch, effort, length-frequency and weight-frequency data for the fisheries defined in the analysis, and tag release-recapture data. The details of these data and their stratification are described below. Improvements in these data inputs are ongoing and more detailed summaries of the analyses and methods of producing the necessary input files are given by Tremblay-Boyer and Pilling (2017a) and Tremblay-Boyer and Pilling (2017b) (CPUE standardisations) and McKechnie et al. (2017b) (tagging, size composition, CPUE standardisations and biological parameter analyses). In addition, a more detailed description of the fisheries definitions, their data summaries and changes in fisheries structures since the 2014 stock assessment, are provided in McKechnie et al. (2017b). The full details of these analyses are not repeated here, we rather provide a brief overview of the key features and direct readers to the relevant papers referenced throughout this section.

4.1 Spatial stratification

The geographical area considered in the 2014 stock assessment (Davies et al., 2014) corresponded to the WCPO from 50° N to 40° S, and from oceanic waters adjacent to the east Asian coast (110° E between 20° N and 10° S; 120° E north of 20° N) to 150° W (Fig. 2). This regional structure comprises nine regions, with two regions north of 20° N (regions 1 and 2) and four equatorial regions between 10° S to 20° N. The western equatorial region covers the area from 110° E to 140° E (region 7) and was established in 2014 to reduce the impact of uncertainty in the catch time-series from Indonesia, Philippines, and Vietnam, and the eastern equatorial region (region 4) is defined from 170° E to 150° W. The central equatorial regions (regions 3 and 8) comprise the area between 140° E and 170° E with the area south of the equator between 140° E and 155° E along with the area south of 5° S between 155° E and 160° E, which includes the Bismark and Solomon Seas (region 8) where considerable tagging effort has occurred, and analyses show more persistent residency compared to the wider western equatorial region.

The southern regions extend from 10° S to 40° S, and from 140° E to 150° W, with the boundary between the eastern (region 5) and western (region 6) regions established at 170° E. Within region 5 a small region (region 9) is present and was established in 2014 to better model the tagging data from the Coral Sea (Davies et al., 2014). The eastern boundary for the assessment regions is 150° W, and as such excludes the WCPFC Convention area component that overlaps with the IATTC area.

The initial stepwise progression of the 2017 stock assessment utilised these region definitions. However, in subsequent steps the northern boundaries between regions 1 and 3, and between 2 and 4 were shifted from 20° N to 10° N for many of the 2017 models (Figure 3). The rationale for this modification and the consequences for fisheries definitions are detailed in McKechnie et al. (2017b). Briefly, this was undertaken based on recent analyses of tagging data from the Central Pacific

(Schaefer et al., 2015) which indicated very limited movement between the equatorial band and more temperate waters for bigeye in particular. This spatial structure also more closely reflects the distribution of purse seine fishing in the Pacific, which is largely restricted to the zone between 10° S and 10° N, as well as oceanographic features such as shifts in the depth of the thermocline. Although the regional shift was driven by bigeye tuna movement, these considerations are directly relevant for yellowfin tuna, and given the similarities between the bigeye and yellowfin tuna assessments and of the fisheries that exploit them, the alternative spatial structure was also adopted here.

4.2 Temporal stratification

The time period covered by the assessment is 1952–2015 which includes all significant post-war tuna fishing in the WCPO. Within this period, data were compiled into quarters (1; Jan–Mar, 2; Apr–Jun, 3; Jul–Sep, 4; Oct–Dec). As agreed at SC12, the assessment does not include data from the most recent calendar year. This is because these data are only finalized very late and often subject to significant revision post-SC, in particular the longline data on which this assessment greatly depends.

4.3 Definition of fisheries

MFCL requires the definition of *fisheries* within the model to consist of relatively homogeneous fishing units. Ideally, the defined fisheries will have selectivity and catchability characteristics that do not vary greatly over time and space, although in the case of catchability some allowance can be made for time series variation. For most pelagic fisheries assessments, fisheries are defined according to gear type, fishing method and region, and sometimes also by vessel flag or fleet. An important consideration for whether multiple fisheries were included in a region was the availability of CPUE and size data. The fisheries definitions used in the diagnostic case model are presented in Table 1 and consist of longline, purse seine, pole and line and various miscellaneous small-fish fisheries in Indonesia and the Philippines for a total of 32 fisheries.

The fisheries for models utilising the 2014 regional structure remained the same as those used in the 2014 assessment, with a total of 33 fisheries defined. The consequences for fisheries definitions of modifying the region boundaries in the 2017 regional structures are presented in detail by McKechnie et al. (2017b) and lead to a reduction to 32 fisheries (there is no longer a US LL fishery in region 4). A graphical summary of the availability of data for each fishery used in the diagnostic case model is provided in Figure 7.

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 (log, FAD, whale, dolphin, and unknown set types) and unassociated (school) sets. Further fisheries were defined for pole-and-line fisheries and miscellaneous fisheries (gillnets, ringnets, 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 Catch and effort data

4.4.1 General features

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

As described in section 3.3, total annual catches by major gear categories for the WCPO are shown in Figure 4 and a regional breakdown is provided in Figure 5. The spatial distribution of catches over the past ten years in provided in Figure 6. Discarded catches are estimated to be minor and were not included in the analysis (typically 1-3% for longline and purse seine based on observer data; (OFP, 2017)). Annual catches in the northern region have been relatively stable over much of the assessment period. Most of the catch occurs in the tropical regions (3, 4, 7, and 8). As noted above, data through 2015 were used in the current assessment, since data for 2016 were not available in time for inclusion.

Within the model, effort for each fishery was normalised to an average of 1 to assist numerical stability. Some longline fisheries were grouped to share common catchability parameters in the various analyses. For such grouped fisheries, the normalisation occurred over the group rather than for the individual fisheries so as to preserve the relative levels of effort between the fisheries. For some fisheries no effort is used–this is typically in cases where effort data is either considered unreliable or the fishery aggregates different 'other' fishing gears such that effort units are not compatible.

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

- The steady increases in total yellowfin catch over most of the assessment period in the equatorial region.
- The relatively stable catches of yellowfin in the northern temperate region by longline vessels, and to a much lesser degree, Japanese pole and line and purse seine vessels in region 1.
- The development of the equatorial purse-seine fisheries from the mid-1970s, and corresponding catches, particularly in equatorial regions 3, 4 and 8.
- 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.

4.4.2 Purse seine

Previous assessments have considered two sets of purse-seine input catch data, but the problems surrounding logbook reporting of catches and grab-sample bias have been clearly demonstrated and only a single set of purse seine catch estimates have been included in the 2014 and current assessments. Details of the analyses, including the independent review and response are provided in Lawson (2013); Cordue (2013); Powers (2013), and McArdle (2013). The procedure now used routinely for estimating purse seine catch by species is documented in Hampton and Williams (2017) as Method 3.

As in previous assessments, effort data units for purse seine fisheries are defined as days fishing and/or searching, and are allocated to set-types based on the proportion of total sets attributed to a specified set type (associated or unassociated sets) in logbook data. Recently it has been discovered that some fleets have changed their reporting practices (SPC-OFP, 2013) such that far fewer searching days are reported and these are instead reported as non-fishing transit days. This practice essentially represents effort creep and we have not yet specifically corrected recent data to ensure consistency of reporting. Therefore the impact of this is not known, but it will be minimized by the practice of estimating frequent time-based changes in catchability.

Catch-per-unit-effort for the Philippines domestic purse seine was analysed using a GLM for CPUE indices (Bigelow and Garvilles, 2017, Figure 8). These indices were applied to estimate the effective effort of the S-PH-7 fishery for the years 2005-2015. Catch-per-unit-effort for all fleets in the purse seine fishery operating largely within the PNG archipelagic waters was analysed for standardised indices using the GLM (McKechnie et al., 2017b), (Figure 8). These indices were applied to estimate the effective effort of the S-ASS-All-8 fishery for the years 1997-2015. Both indices included estimates of time-variant precision.

4.4.3 Longline fisheries

The CPUE indices for the main longline fisheries in each region are one of the most important inputs to the assessment, as they both provide information on trends in abundance over time for each region, and scale relative abundance among the regions. Several sources of standardized CPUE series were used in various stages of the current assessment. Given how influential these indices can be, previous approaches were also reproduced such that the impact of methodological changes could be assessed in isolation of other changes to the assessment.

2014 assessment approach and update The first set of indices was constructed to closely match those used in the 2014 reference case model and were used early in the stepwise progression from the 2014 to the 2017 diagnostic case model. These were originally derived from Japanese operational-level longline data using generalized linear models (GLMs) and a delta-lognormal approach (Hoyle and Okamoto, 2011; McKechnie et al., 2014b). These were only available for the old (2011 assessment regional structure) regions 1-6 and through to 2009 and for some areas the indices for 2009 were very uncertain. In order to have time series that went through until 2012 (for the 2014 assessment) it was necessary to use Japanese aggregate catch and effort data and then "splice" these together. The procedures for this are described in McKechnie et al. (2014b). These indices were used in regions 1 and 2 of the 2014 assessment, and this process was updated for the 2017 assessment (for the early models in the stepwise development) by extending the indices to the end of 2015.

For the remaining regions of the model, the indices used in 2014 were calculated using SPC-held operational LL logsheet data for a range of LL fleets, including both distant-water and domestic-based vessels. The procedures used to standardise these data are again outlined in (McKechnie et al., 2014b). The development of these methods was undertaken in response to the independent review of the bigeye assessment which suggested the spatial contraction of the Japanese fleet (and therefore the indices based on it) and targeting changes as the two major issues to address for standardisation of longline CPUE data (Ianelli et al., 2012). The indices developed for the 2014 assessment attempted to address these issues in two ways: 1) by using data across multiple fleets in order to minimize the spatial/temporal gaps in longline CPUE coverage; and 2) using operational data which allows us to consider vessel effects, clustering based on species composition of the catch, and other operational

details to better account for targeting changes. For these regions, the indices used in the 2017 assessment were recalculated using identical methods on datasets updated to the end of 2015.

Finally, in 2014, the indices in regions 4 and 6 were calculated from datasets supplemented with Chinese Taipei logsheet data that was not held in SPC databases. This allowed estimation of these indices in earlier years before more widespread data become available in the SPC-held operational database. To update these indices for the stepwise progression of model development it was necessary to again "splice" the 2014 indices with indices calculated on the SPC-held data updated up to the end of 2015.

Full longline operational dataset indices Subsequent to the 2014 assessment, an extensive operational LL dataset has been amalgamated from the SPC-held data together with the data held by all the important distant water fishing nations (see McKechnie et al., 2015 for details of this development). For the 2017 assessment it was possible to calculate standardised indices for nearly all regions for most of the assessment time-period. Several types of models were fitted to these data and these are outlined in several ancilliary papers (McKechnie et al., 2017b; Tremblay-Boyer and Pilling, 2017a,b). The indices used in the diagnostic case model were estimated with models that were extremely similar to the 2014 standardisation models and produced indices that were in general congruent with the previous estimates, although in some regions the indices now cover the early time-period which was often missing in the 2014 assessment. These indices were estimated for both the 2014 and 2017 regional structures and these are utilised in the stepwise model progression process.

Alternative indices for these fisheries were utilised in sensitivity analyses that, 1) attempted to overcome the missing vessel identifiers for JP vessels in the early time-period (Tremblay-Boyer and Pilling, 2017a), and 2) utilise newly developed geostatistical techniques to address some of the issues of the changing spatial coverage of fishing effort in the dataset over time (Tremblay-Boyer and Pilling, 2017b).

Coefficients of variation (CVs) for region-specific standardised effort were averaged to 0.2 over the period 1960-1986 following the previous assessment. GLMs fitted to the operational dataset indices produce CVs which were comparable across all regions so it was decided that a similar mean CV be used for all regions. By implementing these methods, MFCL is able to account for the time-varying nature of the CVs such that the CPUE data are given more weight in time-steps with more precise estimates of abundance (typically late in the assessment period when sample sizes in the standardisations are larger).

Another important input for the standardized indices is regional scaling factors incorporated to estimate the relative level of exploitable longline biomass among regions (see McKechnie et al., 2014a). The geostatistical models (Tremblay-Boyer and Pilling, 2017b) produce an abundance surface over the entire stock assessment region from which it is straightforward, even for multiple regional structures, to compute relative estimates of abundance among regions. These *regional weights* or *regional scaling factors* are then used to scale the regions-specific standardised indices. We therefore used the geostatistical model outputs aggregated for the period 1960-1986 to estimate the regional scaling factors. This was consistent with the approach adopted in 2014.

Catch in numbers were used for all LL fisheries in the model. All indices for which catchability was shared and assumed constant, i.e., the L-All fisheries in each region, are presented in Figure 8. Due to conflicts with other data for region 8, particularly the tagging data (discussed later in Section 8), the CPUE index for L-All-8 was not used in the reference case model. For the other longline fisheries, the effort units were defined as the total number of hooks set.

Catch scenarios Recent research (MRAG Asia Pacific, 2016) has indicated the potential for systematic under reporting of bigeye catch for some fleets. An alternative catch history for the LL fisheries in the assessment was constructed at the request of the PAW for bigeye but the PAW did not deem it necessary to have a similar sensitivity for yellowfin tuna at this time.

4.4.4 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, some catch data from the small-fish fisheries in Vietnam have also been included. These improved catch estimates have been incorporated into the current assessment.

For these other fisheries effort is either included as days fished, or more often set to 'missing'. Effort was set to missing for 3 fisheries, the three small-scale miscellaneous fisheries Misc-PH-7, Misc-ID-7 and Misc-VN-7. For these fisheries the model directly computes fishing mortality consistent with the observed catch using a Newton-Raphson procedure. A nominal effort of one was added for the final year of the model to allow the estimation of a catchability coefficient to assist with projection analyses.

Catch-per-unit-effort for the combined Indonesia and Philippines handline fishery was analysed using a GLM for CPUE indices by (Bigelow and Garvilles, 2017). These indices were applied to the catches of the HL-PH-7 fishery for the years 2004-2015. These indices were included as a sensitivity in the 2014 assessment but are part of the diagnostic case in the current assessment.

4.5 Size data

Available length-frequency data for each of the defined fisheries were compiled into 95 2-cm 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 programmes, or through port sampling. Davies et al. (2011) provides more details on the source of the size data.

Each length-frequency record in the model consisted of the actual number of yellowfin tuna measured and Figure 7 provides details of the temporal availability of length and weight (for longline) frequency data and the relative sample sizes. Note that a maximum sample size of 1,000 was implemented in the assessment and the effective sample size was further downweighted with the sensitivity to the magnitude of the divisor investigated in the sensitivity and structural uncertainty analyses.

4.5.1 Purse seine

Length-frequency samples from purse seiners have been collected from a variety of port sampling and observer programmes since the mid-1980s. Most of the early data are sourced from the U.S. National Marine Fisheries Service (NMFS) port sampling programme for U.S. purse seiners in Pago Pago, American Samoa and an observer programme conducted for the same fleet. Since the early

1990s, port sampling and observer programmes on other purse seine fleets have provided additional data. Only data that could be classified by set type were included in the final data set.

Only length frequency samples are used in the yellowfin assessment for purse seine fisheries, and assessments prior to 2014 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 made for the 2014 assessment and are also adopted here. These 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).

4.5.2 Longline

A detailed review of all available length and weight frequency data for yellowfin tuna was undertaken and is described in McKechnie (2014). That paper provides details of the analytical approaches for constructing data inputs for the 2014 assessment, which also apply to 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). Therefore, we considered the coverage and size of samples and typically chose to use weight frequency data where it was available. Japanese weight data were not available for regions 4, 5, and 6 in recent years and had to be supplemented by "all flags" length data.

The general approach used by McKechnie (2014) was that Japanese size data were weighted with respect to the spatial distribution of catch within the region, and the size data from all-fleets data were weighted by the flag-specific catch. For the catch weighting, a moving 11 quarter time window was used to calculate the relative importance of each stratum. The exact methods used in 2014 were adopted in the current assessment, with the addition of three extra years of data.

4.5.3 Other fisheries

Size data were either missing or poor for the Indonesian and Vietnamese small-fish fisheries and the Indonesian-Philippines ex-EEZ purse seine fishery. In the case of the first two, selectivity was assumed to be shared with the Philippines small-fish fishery and in the last case it was shared with the associated purse seine fishery, also in region 7.

Size composition data for the Philippines domestic fisheries (both small-fish fisheries and large-fish handline fisheries) were derived from a number of sampling programmes 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.

As for the most recent yellowfin assessments the length frequency samples from the small fish miscellaneous fishery (MISC PH 7) were adjusted to exclude all reported fish lengths greater than 90 cm from the current assessment. 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 programmes.

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.

4.6 Tagging data

A considerable amount of tagging data were available for incorporation into the current yellowfin stock assessment Table 2. A summary of the data characteristics and the process of constructing the MFCL tagging file are described in detail by McKechnie et al. (2017b). The 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) releases and returns 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 3rd quarter of 2014. Additional data have become available since the 2014 assessment for the Japanese Tagging Programme, conducted by NRIFSF and the Ajinomoto Co. Inc, over the period 2000–2014, which were included in a sensitivity analysis following the recommendations of PAW.

The largest tag data set available for inclusion in the current assessment was that from the recent PTTP which was mainly undertaken in the western tropical Pacific from Indonesia to the Gilbert Islands of Kiribati over the last decade. This data set was expanded since the previous assessment to 14,883 recaptures across 45 release groups (Table 2).

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 the current 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, estimates of base levels of tag shedding/tag mortality, and tag recaptures that occurred outside the model domain. The procedures used in re-scaling the releases to account for these losses are described in detail in Berger et al. (2014) and McKechnie et al. (2016b), but essentially the re-scaling preserves the recovery rates of tags from the individual tag groups that would otherwise be biased low when an often significant proportion of recaptures cannot be assigned to a recapture category in the assessment. The same processing methods were adopted for the current assessment and a more detailed summary of the process and the resulting dataset can be found in McKechnie et al. (2017b).

There is a delay between tagged fish being caught, the tag being reported and the data being entered into 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 second quarter of 2014 were excluded from the assessment (McKechnie et al., 2016b), which is consistent with the approach taken in 2014.

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 78805 effective releases were classified into 83 tag release groups (Table 2). The returns from each

size-class of each tag release group (19104 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.

5 Model description: Structural assumptions, parameterisation, and priors

The model can be considered to consist of several components, (i) the dynamics of the fish population; (ii) the fishery dynamics; (iii) the dynamics of tagged fish; (iv) the observation models for the data; (v) the parameter estimation procedure; and (vi) stock assessment interpretations. Detailed technical descriptions of components (i)–(iv) are given in Hampton and Fournier (2001) and Kleiber et al. (2017). In addition, we describe the procedures followed for estimating the parameters of the model and the way in which stock assessment conclusions are drawn using a series of reference points.

5.1 Population dynamics

The current yellowfin model partitions the population into 9 spatial regions (see Section 4.1) and 28 quarterly age-classes. The last age-class comprises a 'plus group' in which mortality and other characteristics are assumed to be constant. The model follows the population at quarterly time steps, extending through a time window of 1952-2015. The main population dynamics processes are described here for the diagnostic case model only as applied in the current (2017) assessment as follows. Variations to the assumptions and structure that accommodate the developments since the 2014 assessment to achieve the diagnostic case model and one-step sensitivities have been explored and are described in Section 6.1 and Section 6.2.

5.1.1 Recruitment

Recruitment is defined as the appearance of age-class 1 fish in the population. Tropical tuna spawning does not always follow a clear seasonal pattern but occurs sporadically when food supplies are plentiful (Itano, 2000). 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 spawning potential via a Beverton and Holt stock-recruitment relationship (SRR) with a fixed value of steepness (h). Steepness is defined as the ratio of the equilibrium recruitment produced by 20% of the equilibrium unexploited spawning potential to that produced by the equilibrium unexploited spawning potential (Francis, 1992; Harley, 2011). Typically, fisheries data are not very informative about steepness (ISSF, 2011); hence, the steepness parameter was fixed at a moderate value (0.80) and the sensitivity of the model results to the value of steepness was explored by setting it to lower (0.65) and higher (0.95) values.

In the diagnostic case model, it was assumed that annual recruitment was related to annual mean spawning potential, which was recommended by the 2011 Bigeye Tuna Peer Review (Ianelli et al.,

2012) and was previously assumed for the south Pacific albacore and WCPO skipjack assessments (Harley et al., 2015; McKechnie et al., 2016a). An alternative model run was also undertaken to explore the former assumption of estimating the SRR at the quarterly-scale.

The SRR is incorporated in MFCL mainly so that yield analysis and population projections can be undertaken for stock assessment purposes, particularly the determination of equilibrium- and depletion-based reference points. We therefore applied a weak penalty (equivalent to a CV of 7.0) for deviation from the SRR so that it would have negligible effect on recruitment and other model estimates (Hampton and Fournier, 2001), but still allow the estimation of asymptotic recruitment. This approach was recommended (rec. 20) by the 2011 bigeye assessment review (Ianelli et al., 2012). The SRR was calculated over the period from 1965–mid-2014, to prevent the early recruitments (which may not be well estimated), and the terminal recruitments (which are not freely estimated), from influencing the relationship, which is consistent with the approach of the 2014 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 for the 6 terminal recruitments.

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

5.1.2 Initial population

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

5.1.3 Growth

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

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.

Based upon previous analyses assuming a standard von Bertalanffy growth pattern, substantial departures from the model may be indicated, particularly for fish of small sizes (up to about 80 cm for yellowfin). Similar observations have been made on yellowfin growth patterns determined from daily otolith increments and tagging data (Lehodey et al., 2012). We therefore modelled growth by allowing the mean lengths of the first eight quarterly age-classes to be independent parameters, with the remaining mean lengths following a von Bertalanffy growth curve. These deviations attract a small penalty to avoid over-fitting the size data.

5.1.4 Movement

Movement was assumed to occur instantaneously at the beginning of each quarter via movement coefficients that connect regions sharing a common boundary. Note that fish can move between non-contiguous regions in a single time step due to the 'implicit transition' computational algorithm employed (see Hampton and Fournier, 2001 and Kleiber et al., 2017 for details). Movement is parameterised as the proportion of fish in a given region that move to the adjacent region. Across each inter-regional boundary in the model, movement is possible in both directions for the four quarters, each with their own movement coefficients. Thus the number of movement parameters is 2× no.region boundaries (13) × 4 quarters. The seasonal pattern of movement persists from year to year with no allowance for longer-term variation in movement. Usually there are limited data available to estimate age-specific movement and the movement coefficients are considered here invariant with respect to age for this species. A prior of 0.1 is assumed for all movement coefficients, inferring a relatively high mixing rate between regions. A small penalty is applied to deviations from the prior.

5.2 Natural mortality

Natural mortality (M) may be held fixed at pre-determined age-specific levels or estimated as age-specific parameters. M-at-age was recalculated for previous assessments using an approach applied to other tunas in the WCPO and EPO (Harley and Maunder, 2003; Hoyle, 2008; Hoyle and Nicol, 2008). The generally increasing proportion of males 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. The externally-estimated M-at-age function was input to MFCL as fixed values and is shown in Figure 9 for the diagnostic case model.

An alternative to fixed M-at-age is to allow it to be estimated internally in MFCL. This was investigated in the 2014 assessment using a penalisation approach in sensitivity analyses and in the structural uncertainty grid. MFCL has been developed to include functional forms for M estimation subsequent to the 2014 assessment, including a Lorenzen-type (Lorenzen, 1996) M-at-age function, and a parameterisation based on cubic splines. We investigate the Lorenzen parameterisation as a one-off sensitivity analysis and present these in more detail in Section 6.2.

5.3 Sexual maturity

Reproductive output at age, which is used to derive spawning potential, attempts to provide a measure of the relative contribution of fish at different ages to the next generation. This approach was previously applied to albacore (Hoyle, 2008) and bigeye (Hoyle and Nicol, 2008) tunas in the WCPO. The reproductive potential of each age class was assumed to be the product of the proportion female at age, the proportion of females mature at age, the spawning frequency at age of mature females, and the fecundity at age per spawning of mature females (Figure 10). Compared to simply using adult biomass, this results in a slight shift in the age of first maturity and a substantial reduction in the reproductive potential for older age classes. The process of estimating reproductive potential-at-age shares some similarities with calculation of M -at-age (Section 5.2) in that the function is calculated at-length and then backtransformed to at-age using an assumed growth function. Since no new information was available for yellowfin in this region, we used the same parameters as assumed for the previous assessment.

5.4 Fishery dynamics

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

5.4.1 Selectivity

In many stock assessment models, selectivity is modelled as a functional relationship with age, e.g. using a logistic curve to model monotonically increasing selectivity and various dome-shaped curves to model fisheries that select neither the youngest nor oldest fish. Modelling selectivity with separate age-specific coefficients (with a range of 0-1), constrained with smoothing penalties, allows more flexibility but has the disadvantage of requiring a large number of parameters. In most cases we have instead used the same methods as the 2014 assessment which was based on cubic spline interpolation techniques. This is a form of smoothing, but the number of parameters for each fishery is the number of cubic spline *nodes* that are deemed to be sufficient to characterise selectivity over the age range. We use five nodes for most fisheries, which seems to be sufficient to allow for reasonably complex selectivity patterns. The exception is fishery PL ALL 7, where we lowered the number of nodes to four based on model diagnostics.

For particular fisheries alternative functions were employed, including the logistic and monotonically increasing (Table 3). In all cases, selectivity was assumed to be fishery-specific and time-invariant. However, it is possible for a single selectivity function to be "shared" among a group of fisheries that have similar operational characteristics and/or exist in similar areas and with similar length compositions. This grouping facilitates a reduction in the number parameters being estimated and the groupings used are provided in Table 3.

In all fisheries, the selectivity for the last four age-classes, for which the mean lengths are very similar, was constrained to be equal.

The offshore longline fishery (LL-OS-7) has caught consistently larger fish than the other longline fleets in a comparable time period. There are operational differences between the longline fleets that may account for a higher selectivity of larger fish by the Chinese/Taiwanese fleet. These differences in size composition, which were consistent across length- and weight-frequency data, implied that the selectivity of older yellowfin by the L-All fisheries was less than 100%. On this basis, the selectivity of the Chinese/Taiwanese longline fisheries was constrained to have full selectivity for the oldest age classes, while the selectivity of the other longline fisheries (including the principal LL ALL fisheries) was allowed to have declining selectivity for the older age classes. Therefore, the selectivity for the LL-OS-7 fishery was parameterised using a logistic functional form rather than the cubic spline method.

While full length-based selectivity is not currently permitted in MFCL, the age-based selectivity functions are penalised such that selectivity of age-classes that are similar in size will have similar selectivities for a given fishery or group of fisheries.

5.4.2 Catchability

Constant (time-invariant) catchability was assumed for all fisheries that received standardised indices of relative abundance (the LL ALL fisheries in each region, except for region 8; Table 3). This assumption is similar to assuming that the CPUE for these fisheries indexes the exploitable abundance over time. The LL ALL 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. The catchabilities of the other fisheries used as relative abundance indices (HL-PHID-7, PS-PHID-7, PS-ASS-8) were similarly defined.

For all other fisheries, catchability was allowed to vary slowly over time (akin to a random walk) using a structural time-series approach. Random walk steps were taken every two years, and the deviations were constrained by prior distributions of mean zero and variance specified for the different fisheries according to our prior belief regarding the extent to which catchability may have changed. For fisheries having no available effort estimates (e.g. the Philippines and Indonesian surface fisheries), partial fishing mortalities were estimated consistent with the observed catches using a Newton-Raphson procedure. Therefore, catchability deviations (and effort deviations) are not estimated for these fisheries.

For the other fisheries with time-series variability in catchability, the catchability deviation priors were assigned a variance approximating a CV of 0.10. Apart from those fisheries for which the data were based on annual estimates, the catchabilities of all other fisheries were allowed to vary seasonally, including those fisheries that received standardised CPUE indices (Table 3).

5.4.3 Effort deviations

Effort deviations were used to model the random variation in the effort - fishing mortality relationship, and are constrained by pre-specified prior distributions (on the log-scale). There were several categories of fisheries with respect to the effort deviation penalties applied and these are outlined in Table 3. The region-specific CPUE indices 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 standardised 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 GLMs fitted to each fishery (McKechnie et al., 2017b). As occurred in the 2014 assessment, the regional differences in the estimated CVs were sufficiently small that the same average CV was assumed for all indices; the average CV for the period 1960-1986 was set to 0.2. The resulting scaled CVs were transformed to an effort deviate penalty for each CPUE observation in MULTIFAN-CL. 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.

The miscellaneous fisheries have very unreliable estimates of effort and to prevent them from influencing population dynamics they are given missing effort at all time-steps except the last four quarters, when they are given an arbitrary effort of one, and receive effort deviation penalties equivalent to a CV of about 0.22. This is done simply to provide a basis for scaling effort so that the effort deviates converge around 0 to allow effort-based population projections.

Finally, for all other fisheries the nominal effort was used, but to prevent the CPUE of these fisheries

from influencing population dynamics they received effort deviation penalties equivalent to a CV of about 0.7 for the average effort (Table 3). The penalities for many of the fisheries are assumed to be scaled according to the square root of the observed effort such that low penalites are applied for low observed effort and higher penalites are applied for high effort. The effort penalties used in the model are shown by region and fishery in Figure 11.

5.5 Dynamics of tagged fish

5.5.1 Tag reporting

In principle, tag-reporting rates can be estimated internally within the model. In practice, experience has shown that independent information on tag-reporting rates for at least some fisheries tends to be required for reasonably precise estimates to be obtained.

In addition to varying by fishery, we allowed reporting rates to also vary among tagging "programmes" implemented at different times in the history of the fishery, or conducted by different agencies, as tag reporting rates may vary considerably between tagging programmes due to changes in the composition and operation of individual fisheries, and different levels of publicity and follow-up. We provided reporting rate priors for all fishery/tagging programme groups that reflect independent estimates of the reporting rates and their variances. These were derived from analyses of tag seeding experiments (Peatman et al., 2016). 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 within a tagging programme were assumed to be time-invariant. Tag recapture and reporting rate groupings are provided in Table 3.

5.5.2 Tag mixing

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

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

Tagged fish are modelled as discrete cohorts based on the region, year, quarter and age at release for the first 12 quarters after release. Subsequently, the tagged fish are pooled into a common group. This is to limit memory and computational requirements.

5.6 Likelihood components

There are four data components that contribute to the log-likelihood function for this assessment: the total catch data, the length-frequency data, the weight-frequency data and the tagging data.

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. Note that this is close to the 'catch conditioned' approach often used in other integrated assessments.

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. Length frequency samples are assigned ESS lower than the number of fish measured. Lower ESS recognise that (i) length- and weight-frequency samples are not truly random (because of non-independence in the population with respect to size) and would have higher variance as a result; and (ii) the model does not include all possible process error, resulting in further under-estimation of variances. We divided the observed sample sizes by 20, and then applied a maximum ESS of 50 to 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 Dirichlet-Multinomial (Thorson et al., 2017). This method allows the estimation of ESS from the data, conditional on the specification of maximum ESS, thus substantially removing the need for arbitrary assumptions about ESS. As this is a new feature of MFCL (Davies et al., 2017), we did not apply it to all of the models developed in this assessment. However, once testing and documentation of the method is complete we expect it will become the standard likelihood function for compositional data for MFCL in future assessments.

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 have employed a parameterisation of the overdispersion parameter such that as it approaches 1, the negative binomial approaches 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 recognise 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. (2017).

5.7 Parameter estimation and uncertainty

The parameters of the model were estimated by maximizing the log-likelihood of all data components plus the log of the probability density functions of the priors and smoothing 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. The values of these parameters are provided in the yft.ini input file.

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. Therefore, 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., 2017). Previous assessments had conducted likelihood profiling for the total population scaling parameter in MFCL. While this was convenient in the sense that this was a model parameter that could be fixed to obtain points along the profile, we realised that there was potential confounding of this parameter with the mean of the recruitment deviations, such that an increase or decrease in the scaling parameter could be compensated at least in part by a shift in the distribution of recruitment deviations in the opposite direction. We therefore opted to use the population biomass averaged over the entire model period as a basis for generating the likelihood profile. This has the slight complication that average biomass is not a model parameter, but is a derived quantity determined by all model parameters. It was therefore necessary to take a penalised approach, whereby the model is penalised for deviations from the "target" biomass for each point on the profile. Initially, the penalties are set to be small, and progressively increased to be very large to obtain a final model fit conditioned on the specified average biomass. 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 11.1), including an 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 2017 diagnostic case model are also presented in the Appendix (Section 11.2).

5.8 Stock assessment interpretation methods

Several ancillary analyses using the fitted model/suite of models were conducted in order to interpret the results for stock assessment purposes. The methods involved are summarized below and further details can be found in Kleiber et al. (2017).

5.8.1 Yield analysis

The yield analysis consists of computing equilibrium catch (or yield) and biomass, conditional on a specified basal level of age-specific fishing mortality (F_a) for the entire model domain, a series of fishing mortality multipliers (*fmult*), the natural mortality-at-age (M_a), the mean weight-at-age (w_a) and the SRR parameters. All of these parameters, apart from *fmult*, which is arbitrarily specified over a range of 0–50 (in increments of 0.1), are available from the parameter estimates of the model. The maximum yield with respect to *fmult* can easily be determined using the formulae given in Kleiber et al. (2017), 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 of interest as potential reference points. These ratios were 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 (2011–2014). We do not include 2015 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 (1952–2015). 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 5.8.4.

5.8.2 Depletion and fishery impact

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

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

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

5.8.3 Reference points

The unfished spawning potential ($SB_{F=0}$) in each time period was calculated given the estimated recruitments and the Beverton-Holt SRR as outlined in Section 5.8.2. 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 yellowfin stock where $SB_{F=0}$ is calculated as the average over the period **2005–2014**. 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 2015, and the mean over 2011–2014, respectively (Table 4).

The other key reference point, $F_{recent}/F_{\text{MSY}}$ (Table 4), is the estimated average fishing mortality over the full assessment area over a recent period of time (F_{recent}; 2011–2014 for this stock assessment) divided by the fishing mortality producing MSY which is produced by the yield analysis and has been detailed in Section 5.8.1.

5.8.4 Majuro and Kobe plots

For the standard yield analysis (Section 5.8.1), the fishing mortality-at-age, F_a, is determined as the average over some recent period of time (2011–2014 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 (1952–2014, with no value calculated for the terminal year) by repeating the yield analysis for each year in turn. This enabled temporal trends in 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

The progression of model development from the 2014 reference case to the model proposed as the diagnostic case in 2017 was incremental, with stepwise changes to the model made in-turn to ensure that the consequences of each modification could be ascertained. Changes made to the previous assessment model include additional input data for the years 2013–2015, modified methods in producing the input files (e.g. McKechnie et al., 2017b; Tremblay-Boyer and Pilling, 2017a), new regional structures (McKechnie et al., 2016b), and implementation of additional, or new features of MFCL (Davies et al., 2015). An outline, and basis for, the progression through the models is as follows:

1. The 2014 reference case model [*Ref14*].

2. The 2014 reference case model with the new MFCL executable [*Ref14-NewEx*].
3. A complete update of the 2014 reference case model - all inputs extended from 2012 to 2015 using identical methodology for CPUE, tagging, size frequencies etc, and the same MFCL model settings [*Ref14Update*].
4. The previous model with the same structure and MFCL settings but CPUE indices using the GLM approaches with the updated Pacific-wide operational LL database (McKechnie et al., 2017b) [*Update-CPUE*].
5. The previous model with the same MFCL settings but with the new regional structure and consequently all fisheries, and input data (including CPUE standardisations), reconfigured based on these new regional definitions [*NewAreas*].
6. The previous model with two modifications to the recruitment estimates; the change from quarterly to annual recruitments when estimating the spawner-recruit relationship, and the fixed terminal six recruits set at the arithmetic rather than geometric mean of recruitments for the remaining period [*Diagnostic*].

6.2 Sensitivity analyses

More than a hundred runs were undertaken in conducting the 2017 yellowfin assessment, but in terms of presenting information on the bounds of plausible model sensitivity we have focused on a small set of uncertainty axes which are described in further detail below. These axes were used for "one-off" changes from the diagnostic case model and several of these sensitivity models were used in the structural sensitivity analyses (after Hoyle et al., 2008). The latter process involves constructing a grid of model runs where all-possible combinations of the assumptions are explored (see Section 6.3).

The recommendations of the 2017 PAW formed the basis for several of the one-off sensitivity analyses undertaken from the diagnostic case, but several other runs were undertaken in order to provide a better understanding of the impact of some of the changes in modelling assumptions. The 'key' sensitivity runs for which full details of management quantities have been calculated are provided in Table 7 to 10. Each of the one-off sensitivity runs was carried out by making a single change to the diagnostic case, and it should again be reinforced that these model runs are not carried out to provide absolute estimates of management quantities but to demonstrate the *relative* changes that result from the various changed assumptions.

6.2.1 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 value may have significant influence on most reference points used in management. As was the case in other tropical tuna and albacore tuna assessments and following discussions in SC7 where 0.8 steepness was agreed (WCPFC, 2011), 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.

6.2.2 Tag mixing period [*Mix1*]

The tag mixing period is imposed to allow tagged fish to distribute themselves throughout the region of tagging after they are first tagged, although it is somewhat difficult to ascertain how long this period should be. The tagging data indicate high levels of mixing among the equatorial regions, and in combination with a larger number of regions, the possibility of more rapid mixing of tagged fish in the population was explored by reducing the mixing period to 1 quarter. 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*). This sensitivity also serves to explore to some extent an increased relative weight assigned to the tagging data as it increases the effective number of observed tag recaptures (but see section 6.2.5).

6.2.3 Relative weighting of length and weight frequency data [*Size10, Size50*]

The difficulties in assigning weighting to the length frequency data were discussed in Section 5.6. To assess the sensitivity of model results to the weighting of these data, two 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 reference case model (*Size10*); and one where those data were down-weighted (corresponding to a maximum effective sample size of 20 fish) relative to the reference case (*Size50*).

6.2.4 Dirichlet multinomial likelihood for the size frequency data [*Dch*]

The Dirichlet multinomial likelihood is a new development in MFCL (Davies et al., 2017), 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 Dirichlet multinomial in a sensitivity analysis in this assessment, assuming a maximum effective sample size of 100, pending further testing and evaluation of its performance.

6.2.5 Overdispersion in the tagging likelihood [*OD2*]

The negative binomial likelihood for the tagging data includes a parameter (τ) that determines the variance of the negative binomial distribution in relation to the Poisson distribution (where the variance is equal to the mean). Therefore, the τ parameter effectively determines the weighting receved by the tagging data in the overall model likelihood. Initial attempts to estimate τ (assumed common for all fisheries and tag release groups) saw it converge to a value close to 1, i.e. Poisson. This setting was therefore used in the Diagnostic Case. It is likely, however, that there is overdispersion in the tagging data resulting from non-independent tagged fish, lack of mixing, and other unaccounted-for process error. Therefore, this sensitivity analysis assumed a negative binomial variance of twice the Poisson, i.e. $\tau = 2$, in order to determine the impact of down-weighting the tagging data in the overall likelihood.

6.2.6 Quarterly stock recruitment relationship [*SRRqtrly*]

To assess the sensitivity of the model to the new procedure that deviations in total recruitment from the SRR are computed annually (rather than quarterly in the 2014 assessment), an alternative model was run reverting to the previous quarterly procedure for computing recruitment deviations.

6.2.7 Estimate natural mortality M-at-age [*Lorenzen*]

While the M-at-age function in the diagnostic case model was fixed at the values estimated externally for MFCL (Hampton, 2000), a sensitivity model was constructed to estimate natural mortality using recent developments in MFCL that allow biologically reasonable functional forms for M-at-age (Kleiber et al., 2017). We utilise the Lorenzen approach (Lorenzen, 1996) which assumes a monotonically declining relationship between M and the mean length of fish in successively older age classes a, l_a, such that $M_a = c(l_a)^b$, with two estimable parameters, b and c.

6.2.8 Alternative standardised CPUE indices [*CPUE-Proxy, CPUE-Geostat*]

The model CPUE-Proxy assesses the effects of using alternative GLM-standardised CPUE indices calculated by Tremblay-Boyer and Pilling (2017a). These indices were an attempt to account for some of the problems associated with missing identification of JP vessels in the operational dataset, which makes standardising out vessel effects difficult. A further sensitivity model (*CPUE-Geostats*) was fitted that included the geostatistical CPUE indices estimated by Tremblay-Boyer and Pilling (2017b). These standardisations attempted to account for some of the issues that are encountered when the spatial distribution of fishing effort in the CPUE dataset changes over the model period.

6.2.9 Inclusion of tags from the Japanese Tagging Programme [*JPtag*]

The Japanese Tagging Programme provided additional tagging data for the 2017 assessments of bigeye and yellowfin. However, there were very few recaptures outside of region 1 where most of the release events originated, and there was no prior information on reporting rates for the JPTP in tropical areas. Therefore we elected for now to include these data in a sensitivity run to the diagnostic case model to assess their impact on model output.

6.2.10 Relaxation of reporting rate parameter bounds [*RR0.99*]

There has been concern that estimates of reporting rate parameters at their bounds (typically 0.9 by default) have affected previous assessments of bigeye and other species by acting as a limit on biomass estimates. This sensitivity relaxes the bound such that it occurs at a value of 0.99 rather than 0.9 in the diagnostic case model.

6.2.11 Sensitivity to alternative CPUE series in region 8 [*inclLL8*]

In the 2014 yellowfin assessment, the conflict between the standardised longline CPUE data and tagging data within region 8 was noted and resolved by using the purse-seine index for this region

instead. We took a similar approach for the current assessment and included a sensitivity to examine the impact of using the longline standardised CPUE series in this region in place of the series used within the diagnostic case model (i.e. PS ASS 8).

6.2.12 Regional structure [*2014Reg*]

The progression from the 2014 reference case model to the 2017 diagnostic case model (Section 6.1) led to a slight change in biomass, depletion and stock status estimates during the progression from the 2014 to 2017 regional structures. A sensitivity model was therefore constructed where the diagnostic case model settings were maintained but reverting to the previous 2014 regional structure and fisheries configuration.

6.3 Structural uncertainty

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

The structural uncertainty grid for the 2017 assessment was constructed from 5 axes: steepness (3 settings), tagging data overdispersion (2), tag mixing (2), size data weighting (2) and regional structure (2), with the settings used directly comparable to Section 6.2 through identical notation. The final grid thus consisted of 48 models (Table A5). Note that due to the very large computational load of running the grid with each model requiring significant run time, the size frequency data weighting level for the diagnostic case model (scalar of 20) was not included. Instead the upper (scalar of 10 *Size10*) and lower (scalar of 50 *Size50*) weightings of the sensitivities were used to bound the uncertainty from this axis thus also reducing the number of models that had to be run.

7 Results

7.1 Consequences of key model developments

In order to examine the impacts of stepwise developments from the 2014 yellowfin assessment (Davies et al., 2014) to the diagnostic case run of the current assessment described in Section 6.1, for each of the stepwise development runs the resulting temporal trajectories of spawning potential and spawning potential depletion ($SB_{[t]}/SB_{F=0[t]}$) are compared (Figure 12 and 13). A summary of the consequences of this progression through the models is as follows:

2014 reference case with new executable *Ref2014* → *Ref2014–newexec* Use of the latest version of MULTIFAN-CL software (version 2.0.3.1) had a negligible effect on the trajectories of spawning potential biomass and depletion (see grey and dark red lines overlap in Figure 12 and 13).

Updated with 2015 data but oldstyle CPUE *Ref2014–newexec → Ref2014Update* Extending the data series through to 2015, but using the same CPUE standardisation approach at that used in the 2014 assessment led to a lower estimate of spawning potential in the earlier years of the assessment (up to around 1990) but the decreasing trend in the time series was flatter, so that in the most recent years a slightly higher spawning potential was estimated (Figure 12). The spawning potential depletion trajectory was comparable to that from the 2014 reference case with new executable run, until the 1990s when the updated data run indicated a reduced rate of depletion. Terminal depletion was around 10 percentage points higher.

switch to new CPUE (no-vessel GLMs with PW operational dataset) *Ref2014Update → Update–CPUE* Incorporation of the standardised CPUE time series developed from the full Pacific-wide operational data set led to a notable increase in the spawning potential in the earlier period of the time series, with spawning potential estimated to be around 30% higher in the first year (Figure 12). Noting that this data set provided greater operational information in the historical period, these data were clearly influential. Following 1960, however, the trajectory was closer to that of the 2014 reference case, with terminal spawning potential estimates lower than those from the model run with updated data and 'old-style' CPUE. The spawning potential depletion trajectory suggested the stock was slightly less depleted than in previous runs until around 1990, when the trajectory was comparable to that of the 2014 reference case, and indicated a slightly lower depletion level (Figure 13).

Switch to 2017 regional structure *Update–CPUE → NewAreas* The change to the 2017 regional structure led to an increase in the estimated spawning potential throughout the time series, converging with the other trajectories towards the end of the time series (Figure 12). The trajectory of the depletion ratio was slightly higher through most of the time series, and terminal estimates indicated slightly less depletion (Figure 13).

Switch to annual SRR *NewAreas → Diagnostic* Incorporation of the new assumption that annual recruitment is related to spawning biomass scaled down very slightly the estimated spawning potential from the previous step to the 2017 regional structure, but resulted in a similar declining trajectory over time (Figure 12). Resulting estimates of depletion also followed a similar trajectory (Figure 13).

Overall, the effects of the stepwise developments from the 2014 reference case assessment model to the 2017 diagnostic case run were a notable increase in historical spawning potential (of around 25-30% in the 1950s) but a less than 15% increase in recent spawning potential. Estimates of stock depletion indicated a slightly less depleted stock when compared to the 2014 reference case model, with terminal instantaneous depletion estimates of about 42% (Figure 13; see Table 7 for other reference points; note estimate of $SB_{latest}/SB_{F=0}$ from the diagnostic case was 0.40). This last model was selected as the diagnostic case.

7.2 Model fit for the diagnostic case model

7.2.1 Catch data

As very high penalties were applied to the catch data for all fisheries, residuals of the observed and model-predicted catches were very small (Figure 14).

7.2.2 Standardized CPUE

The temporal variability in the standardised CPUE indices used in the assessment was well fitted by the model-predicted CPUE (Figure 15). In general, the model captures the declines seen in the longline CPUE trends, as well as the increase in CPUE noted in the most recent period in many of the model regions. Increased variability early in the time series leads to some lack of fit, but the overall trend is still well captured. The early variability cannot be explained by seasonal effects on their own, and may reflect lower sample sizes (and generally greater uncertainty) early in the operational dataset together with a small proportion of the region area actually being fished. Consequently, these values have very high CVs, which allows the model to deviate away from them to an extent (Figure 8), as shown by the point colours in Figure 15 (deviations from observed CPUEs in blue are less penalized than those in yellow and red).

Region 7 is the one instance where there are multiple indices, reflecting the diversity of fisheries operating therein. While the long time-series of longline CPUE is well explained, the model does not capture the variability in the other two other time series as well (PS PHID 7 especially, but also HL PHID 7 to an extent), but these are more variable and have somewhat lower penalties due to their effort. The overall trend in HL PHID 7 appears well modelled, and while the magnitude of the decline in the early period of the PS PHID 7 series cannot be fully captured, the general tendency is captured for the later two thirds of the series when the penalties are higher (Figure 15).

Region 8 also has multiple candidate CPUE indices but we elected to only retain that of the purse seine associated CPUE series (PS ASS 8) to reduce conflict with the tagging returns for this region. This time-series, while showing considerable fluctuation since 1997, is well captured by the model. A one-off sensitivity explored the impact of using the other standardized time-series available for this region (LL ALL 8) and is further discussed in Section 7.5.1.

These overall patterns are also reflected in the time-series of effort deviations Figure 16 where there is a tendency for positive and negative deviations early in the time series in regions 5, and 1 and 2 respectively. In several other regions the observed data displayed more pronounced variation than the model predictions although in some cases the observed indices reflect the challenge of capturing somewhat competing trends, as discussed above (e.g. region 7). The CPUE of fisheries not receiving standardised CPUE indices have far less influence in estimating temporal trends in biomass as much lower effort deviation penalties are applied and catchability is allowed to vary over time for these fisheries. Examination of the estimated effort deviations for these fisheries did not reveal any serious issues, with deviations centred around zero and no major temporal trends that might indicate misspecification of temporal changes in catchability (Figure 17).

7.2.3 Size frequency data

Length-frequency data were fit adequately by the model where sample sizes were sufficiently high (e.g. tropical purse-seine, Phillipines' handline HL PHID 7)(Figure 18). In general high frequencies of lower sizes were underfitted by the model, which might be linked to the diversity of sources underlying these data resulting, for instance, in different mode locations across regions, especially for tropical associated purse-seine fisheries. Some of the fisheries have relatively poor fit (e.g. PL ALL 3) but are also poorly sampled. Temporal trends are generally fitted well (Figure 20), the model slightly overestimates length for PS ASS 4 for part of the time series but the selectivity for this fishery is also grouped with those of the other tropical associated purse-seines.

Time-aggregated model fits to the weight-frequency data for the longline fisheries were generally good (Figure 19). There was a slight overestimation of large fish frequencies for fisheries LL US 2, LL ALL 3, LL ALL 7 and LL ALL 8, an underestimation for LL ALL 6 and some bimodal features for the American and Australian longline fisheries (LL US 2 and LL AU 5) not fully captured, similar to the 2014 assessment (Davies et al., 2014). The fit to LL ALL 6 was poor but its selectivity is grouped with LL ALL 3, 4 and 5 which haver much larger sample sizes. Temporal fits were also good (Figure 21), except in instances where there were abrupt temporal shifts in the sampled weights which also tended to match with low sampling sizes (e.g. LL ALL 2 and LL OS 7). While the aggregated fit to weight frequencies for LL OS 3 was very good, the model did not capture a slight decline in sizes over time for this fishery which might be related to the stable-to-increasing CPUE trend for this region for the period where weight data are available.

7.2.4 Tagging data

The model fit to the overall tagging data was good (Figure 22), with some slight under- and over-estimation of returns noted for the PTTP. The estimation was also good at the finer scale of the individual tag recapture groupings (Figure 23). It is noted that returns in the mixing period are included within this diagnostic, which are fitted exactly and hence may exaggerate the goodness of fit. This period will be removed for the diagnostic plots of future tuna assessments.

The observed and model-predicted recapture numbers show large increases concurrent with the three tagging programmes that released the most fish (RTTP; early 1990's, PTTP; 2006-present; Figure 22, Figure 23). Observed and modelled values for most time periods are closely matched, even during periods of particularly high recaptures late in the assessment period.

The good fit to the tagging data is complemented by the predicted recaptures in respect of time at liberty, which closely matching the observations (Figure 24). The model estimates tag attrition due to fishing and natural mortality therefore appear to adequately describe those observations over all tag release programmes. A steep decline in recaptures was observed in the first 5 quarters following release, but a sustained number of tagged fish were recaptured up to 13 quarters at liberty (i.e. the plus-group defined in the assessment model).

7.3 Model parameter estimates (diagnostic case)

7.3.1 Catchability

Time-series catchability estimates were derived for fisheries that did not receive a standardised CPUE index. There is substantial variation in the dynamics of catchability among fisheries (Figure 25), although those trends are influenced by the limited catch and effort time-series available for some fisheries, in particular those whose effort has reduced (such as several PL fisheries), or that are very low volume fisheries. Examining the PS fisheries in more detail, given there is a potential for notable effort creep when compared to some other gear-types, catchability has generally increased over time. However, in recent years there is substantial among-fishery variation in trends, with those in region 3 indicating some decline in catchability in recent years, while those in region 4 have stabilised (with variation from year to year).

7.3.2 Selectivity

Given the range of fisheries included within the yellowfin assessment model, the selectivity functions estimated also show considerable between-fishery variability (Figure 26). Many of the longline fisheries were estimated to have selectivity curves that were, or were very close to, asymptotic, while the selectivity of other grouped longline fisheries (LL ALL 3, 4, 5, 6 and 9) peaked around age-class 10 (i.e. 2.5 years). The large fish handline fishery in region 7 also showed LL–like selectivity, with very few fish younger than about 10 quarters of age being estimated to be caught by this fishery.

The PS, PL and miscellaneous small fish fisheries in region 7 displayed a variety of relatively complex selectivity functions (Figure 26), and the selectivities of some of these fisheries was clearly focused on younger and smaller fish (e.g. the PH and ID miscellaneous fisheries), although purse-seines are still predicted to catch larger fish to an extent, but only for unassociated fisheries that target free schools. This trend of associated purse-seines catching a greater proportion of smaller fish than the corresponding unassociated fisheries is generally maintained across regions.

7.3.3 Movement

Figure 27 displays the source (the region the fish were originally recruited to) of the equilibrium yellowfin biomass in each model region. From these it is evident that the model estimates significant movements of fish between the majority of regions, in particular those surrounding the Equator. In general, there is a tendency for the more temperate regions to receive most of their fish as local recruits (particularly regions 1 and 2) but in several regions (e.g. regions 3 and 4) a significant proportion of fish ending up in a region were actually recruited to nearby regions 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.

7.3.4 Tag Reporting Rates

The estimated tag reporting rates by fishery recapture groups (see groupings in Table 3) are displayed in Figure 28. Reporting rate estimates differed among fisheries groups and across tagging programmes. In most cases, the reporting rate estimates for those groupings that received higher penalties were relatively close to the prior mean. As was the case in the 2014 assessment, reporting rate estimates for some groups were on the upper bound of 0.9, including for one of the larger reporting group (PTTP PS in region 8). This is indicative of a conflict with the CPUE trends in this region, a feature that was present in the 2014 assessment and will be discussed further below. This problem remained even when the upper bound on the reporting rate parameter was relaxed to 0.99 (see Section 7.5.1).

7.3.5 Growth

The Von Bertalanffy function of the diagnostic case model specifies most rapid growth for the youngest age-classes, starting from a mean length of about 20-30cm for the youngest age-class, while the non-von Bertalanffy growth of juvenile yellowfin is clearly evident, with irregular growth occurring in the 25–75 cm size range showing slower growth than predicted by the von Bertalanffy function (Figure 29). The diagnostic case estimate suggests an L2 of 153 cm. The fitted length-at-age

relationship closely matches the one estimated in the 2014 assessment, with a similar trend of non-VB growth for earlier stages and a mean length of the oldest age class estimated in the current assessment to be less than 0.5cm below the 2014 estimate (153.0 *vs.* 153.4cm).

7.4 Stock assessment results

7.4.1 Recruitment

As MFCL has the ability to use both movement and regional recruitment to distribute the population between regions to maximise the total objective function, some care must be taken when interpreting the results of the estimated regional recruitment distribution.

The diagnostic case model recruitment estimates for each region and the total assessment region are shown in Figures 30 and 31. The overall trends in recruitment for the WCPO (Figure 30) are challenging to interpret due to the greater uncertainty in the historical estimates. However, the overall pattern is comparable to that estimated in previous assessments, but with the declining trend seen previously further reduced in the current assessment. Overall average annual recruitment as a proportion of the total across regions is highest within region 7, followed by region 1, while moderate levels of recruitment also occur within regions 5, 6 and 8. Low recruitments were estimated for the other regions. The regional estimates display notable interannual variability and variation on longer time scales. Recruitment is estimated to be higher in most of the key regions up to 1965, with the exception of region 1, and these estimates are highly uncertain (Figure 30). This uncertainty in the early recruitments underpinned the decision to assume the period 1965-2014 as the basis for the long-term average recruitment and estimation of the BH-SRR. Over this period, recruitment on average displays very little trend, and the uncertainty decreases substantially since the mid 1965's.

The estimate of the SRR is presented in Figure 32. Recruitment levels have been maintained as stock size has declined. As seen in the overall WCPO time series of estimates (Figure 30), the estimated inter-annual variability has decreased over time as the stock has declined. Some relatively large recruitments are estimated to have occurred over the recent period, and this pulse of fish is consistent with the up-turn in CPUE indices seen in recent years. This is seen within the trends for regions 5 and 6 in particular (Figure 30).

7.4.2 Biomass

The relative pattern in spawning potential by region is presented in Figure 33 and Figure 31. Following some variability earlier in the time series, biomass is estimated to have declined steadily over the model period, with notable declines seen in all regions except region 1 and to a lesser extent regions 2 and 9 (Figure 33), regions that account for generally smaller proportions of the overall WCPO biomass. There have been several short term increases in abundance throughout the assessment period but these were never sustained more than a few years. In the most recent years, that decline has slowed, and shows a small increase in the last two years, mirroring the trends shown in particular in regions 3, 6 and 8. Trends in biomass are more variable within regions 1, 2, 5 and 9. While the distribution of fish for the diagnostic case model is slightly different from the 2014 assessment due to the changes in the regional structures, this is less notable than in the bigeye assessment.

As for recruitment, uncertainty in the biomass estimates is substantially higher in the early part of the calculation period. This uncertainty decreases after 1960, and uncertainty is relatively low after 2000 (Figure 33). Consistent with the recruitment patterns, spawning potential at the start of the model is estimated to be substantially higher than that estimated for period 1960 onwards. This feature was also present in previous assessments. This substantial decline occurs despite the low and relatively similar catch levels in the periods up to 1975 (Figure 4). Therefore the high early biomass and subsequent decline around 1965 is influenced by the uncertain and high recruitment estimates preceding 1965.

7.4.3 Fishing mortality

Average fishing mortality rates for juvenile and adult age-classes increases strongly from 1970. While that for juveniles is estimated to have plateaued since the late 1990s, with notable variability since that time, adult mortality is estimated to have continued to increase and is at the highest level in the most recent years (Figure 34). Estimates are comparable for juvenile and adult age-classes throughout the model period, although the juvenile rates are generally slightly higher, except for the most recent years.

Changes in fishing mortality-at-age and population age structure are shown for decadal time intervals in Figure 35. Since the 1970's exploitation rates have increased on the youngest age classes through to the most recent decade, associated with increased catches by the purse-seine fishery in tropical regions and the Philippines, Indonesian and Vietnamese fisheries in region 7 (Figure 36). There is also a high exploitation rate on the older age classes (6 and older), which coincides with the peak in the selectivity of the unassociated purse-seine fisheries and longline selectivity. Overall, there has been a decline in the proportion of old (greater than age class 15) fish in the population since the 1970s (Figure 35). Fishing mortality-at-age estimates by region (Figure 36) show expected results, with temperate regions dominated by LL fishing showing highest fishing mortality for the older age classes while in the equatorial regions where the PS and miscellaneous small fish fisheries occur show a higher mortality for the youngest age-classes.

7.5 Multimodel inference–stepwise model development, sensitivity analyses and structural uncertainty

7.5.1 One-off changes from the structural uncertainty analysis

Comparisons of the spawning potential and depletion trajectories for the diagnostic model case and one-change sensitivity runs from the structural uncertainty analysis are provided in Figure 37 and 38. The key reference points are compared in Appendix Tables 7 to 10 (one-off sensitivity models) and Table 6 (structural uncertainty grid) and the likelihood components in Appendix Tables 11 to 14.

Steepness [*h0.65, h0.95***]** The low penalties on the SRR relationship resulted in the assumed value of steepness having a relatively small effect on model fit and time-series estimates of spawning potential (Figure 37). Both alternative steepness values led to very slightly lower estimates of spawning potential. 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 38). 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 (F_{recent}/F_{MSY} for the diagnostic case, h0.65 and h0.95 of 0.59, 0.65 and 0.54, respectively) while the depletion-based reference points tend to be less sensitive to assumed steepness, $SB_{latest}/SB_{F=0}$ ranging from 0.43 to 0.50 (Appendix Table 7).

Models with tag mixing period of 1 Qtr [*Tagmix1*] The model with the mixing period reduced to one quarter (*tagmix1*) from the two quarters in the diagnostic case model, scaled spawning potential downwards notably, by approximately 30% at the start of the time series ('Tag' panel in Figure 37) and estimated the stock to be more depleted than the diagnostic case model (about 4 percentage points more depleted; Figure 38; Appendix Table 10). Both the spawning potential and depletion time series appeared to recover slightly in the most recent years in comparison to the diagnostic case model.

Weighting of the length frequency data [*Size10, Size50*] Two alternative models were run with differing weightings given to the length- and weight-frequency data, one up-weighting the data relative to the reference case (Size10; divisor of 10), and the other down-weighting the data relative to the reference case (Size50; divisor of 50), compared to the diagnostic case with a divisor of 20. These models produced interesting results as both the upscaling and the downscaling of these data led to lower spawning biomass and higher depletion estimates compared to the diagnostic case, indicating that some non-linearities might be present in the impact of composition data on the assessment model.

The estimated spawning potential under Size10 weighting was comparable in scale to the diagnostic case model run, with some relative change in trajectory early in the time series ('Size' panel in Figure 37), where the model appeared able to fit the CPUE data slightly better. The estimates of fisheries depletion were very similar (Figure 38), as were the depletion reference points ($SB_{latest}/SB_{F=0}$ = 0.46 for both models; Appendix Table 9).

Downweighting the size data (Size50) down-scaled the spawning potential estimates notably, although the relative change in trajectory was similar to the reference case. However, there was no increase in the spawning potential in the most recent period that was seen under the alternative size weightings (Figure 37). Fisheries depletion was estimated to be notably lower and diverging from the diagnostic case over the assessment period, with estimates of terminal depletion close to 0.2 (Figure 38). The depletion reference points showed a comparable trend ($SB_{latest}/SB_{F=0}$ = 0.46 for the diagnostic case, 0.24 for *sizecomp50*; Appendix Table 9).

Dirichlet multinomial likelihood for the size frequency data [*Dch*] The model with the Dirichlet multinomial likelihood for the size frequency data (Dch) estimated a lower level of spawning potential throughout the time series (Figure 37) and a greater level of fishing depletion (Figure 38). The trajectory from this run lay between that of the two size-weighting model runs discussed above, and showed similar temporal patterns in both the spawning potential and depletion quantities. Depletion reference point $SB_{latest}/SB_{F=0}$ was notably lower than that of the diagnostic case model, at 0.38 compared to 0.46, respectively.

2014 regional structure [*2014Regions*] Using the same MFCL model settings (though adjusted for the different fisheries structures) but input files relating to the 2014 regional structures resulted in the model scaling stock size downwards slightly (Figure 37), although the relative time-series changes in abundance remained similar to the diagnostic case model. The estimates of fisheries depletion for this model were similar to the diagnostic case model early in the time-series but steadily diverged such that the terminal estimates of dynamic depletion (Figure 38) and the depletion-based reference points (Appendix Table 8) were slightly lower, about 5 percentage points more pessimistic than the

diagnostic case model. The depletion reference point showed a similar pattern (for the 2014regions: $SB_{latest}/SB_{F=0} = 0.44$). Our interpretation of this is that the new regional structure more accurately delineates the higher-exploitation tropical area from the lower exploitation sub-tropical area to the north, effectively attributing the higher exploitation to a smaller area of the overall stock, and the lower exploitation to a larger area of the overall stock. This results in lower overall stock-level depletion. However, the impact is much less significant than seen in the bigeye assessment.

Overdispersion parameter of the negative binomial tagging likelihood [*TagOD2, TagOD4***]** When less weight was given to the tagging dataset (OD2), the estimates of spawning potential were scaled downwards slightly, most notably at the beginning of the time series (Figure 37). Similarly, the estimates of fishing depletion for this model were also slightly more pessimistic, with depletion estimates diverging from the diagnostic case model results, leading to terminal estimates of dynamic depletion approximately 5 percentage points lower. Further down-weighting of the tagging data set (OD4) exacerbated this trend, with further down-scaling of the spawning potential estimates throughout the time series. Depletion estimates further diverged from the diagnostic case model results, leading to terminal estimates of dynamic depletion approximately 10 percentage points lower. The depletion reference points showed less extreme trends ($SB_{latest}/SB_{F=0} = 0.46$ for *TagOD2*, 0.41 for *TagOD4*; Appendix Table 10).

Quarterly SRR estimation [*SRRQtrly***]** The model with the SRR fitted using quarterly (consistent with the 2014 assessment), rather than annual recruitments (assumed by the diagnostic case model), led to a very slight decrease in estimates of spawning potential for the majority of the time series in comparison to the diagnostic case model results, with the two time series converging in the most recent period (Figure 37). In comparison, the stock depletion trajectory was virtually identifical to the diagnostic case, with the quarterly SRR model suggesting very slightly more depletion in the terminal year (Figure 38). The depletion reference point $SB_{latest}/SB_{F=0}$ value was slightly higher, however (0.48).

Estimate natural mortality [*Lorenzen***]** The natural mortality at age estimated *via* the Lorenzen parameterisation had little impact on the spawning potential but slightly more optimistic depletion estimates (Figure 37 and Figure 38). The estimated shape for the relationship was much flatter than the fixed one used in the model, predicting lower mortalities for juvenile classes and higher mortalities for years 1 and 2 (Figure 9).

Alternative standardised CPUE indices [*Proxy, Geostats***]** The model with the alternative GLM-based CPUE indices (CPUE-Proxy) produced estimates of spawning potential that showed some minor variation across the time-series compared to the diagnostic case model, but the scale of the estimates was very similar ('CPUE' panel in Figure 37). The main divergence was in the period before 1980. The estimates of fishing depletion were also very similar for most of the assessment period, with this CPUE series suggesting a slightly higher level of terminal depletion (Figure 38). The depletion reference point $SB_{latest}/SB_{F=0}$ estimate was similar to that from the diagnostic case, being 0.44 (Appendix Table 8).

The model with the geostatistical CPUE indices (CPUE-Geostat) also produced estimates of spawning potential that showed some minor differences in variation across the time-series compared to the diagnostic case model, with similarly scaled estimates but a slightly steeper decline (see 'CPUE' panel in Figure 37). The main divergence was in the period before the early 1970s. The trajectory of fishing depletion (Figure 38) was slightly more pessimistic than the diagnostic case model, but became slightly more pessimistic than both the diagnostic case model and CPUE-Proxy model in about the last 15 years. Estimate of $SB_{latest}/SB_{F=0}$ was 0.46, however.

LL8 fishery: The sensitivity where the region 8 abundance was indexed by the longline instead of the associated purse-seine fishery (CPUE.LL8) showed a slight increase in spawning potential throughout the time-series and a depletion lower by about 3% points (see 'CPUE' panel in Figure 37 and 38, Appendix Table 8). The region 8 longline CPUE was not fitted very well, especially compared to the region 8 PS index, and the estimated reporting rate for the region 8 reporting group was still at the 90% bound, indicating that, similar to the 2014 assessment, there is a conflict between the predicted increase in CPUE from the longline and the rates of tag returns in this region.

Inclusion of tags from the Japanese Tagging Programme [*JPTags*] Inclusion of the Japanese tagging programme tags in the model led to comparable estimates of spawning potential to the diagnostic case model run, but a slightly steeper decline in spawning potential over time, leading to lower estimates after 1980 (Figure 37). The estimates of fishing depletion were very similar for most of the assessment period but gradually became more pessimistic than the diagnostic case model, diverging from around 1980, with terminal depletion estimates about 5 percentage points lower than the diagnostic case model in the terminal years (see 'Tag' panel in Figure 38). The estimate of $SB_{latest}/SB_{F=0}$ was slightly lower, at 0.45. Movement rates of recruits out of region 1 was predicted to be much lower than in the diagnostic case (not shown), presumably because there are very few tags being caught and reported for the JPT outside of this region.

Relaxation of reporting rate parameter bounds [*RR0.99*] This sensitivity model scaled spawning potential upwards marginally, but the trajectories over the remainder of the period were comparable to that of the diagnostic case model (Figure 37). The estimates of fisheries depletion were also comparable over the majority of the time series, with some divergence from the diagnostic case model causing slightly less depletion, especially in the last 10 years (Figure 38; Appendix Table 10).

7.5.2 Structural uncertainty analysis

The results of the structural uncertainty analysis are summarised in several forms – time-series plots of fisheries depletion for all models in the grid (Figure 39), boxplots of $F_{recent}/F_{\text{MSY}}$ and $SB_{latest}/SB_{F=0}$ for the different levels of each of the five axes of uncertainty (Figure 40), Majuro plots showing the estimates of $F_{recent}/F_{\text{MSY}}$ and $SB_{latest}/SB_{F=0}$ (and $SB_{recent}/SB_{F=0}$ for comparison) across all models in the grid (Figure 41), and averages and quantiles across the full grid of 48 models (and specific subsets of models) for all the reference points and other quantities of interest (Table 6, as well as 15 and 16) 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 2017 assessment and these have significant consequences for summarising stock status and deriving management recommendations.

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

- The grid contains a wide range of models with different estimates of stock status, trends in abundance and reference points. The terminal depletion estimated for the majority of runs was above the $20\%SB_{F=0}$, with the range across the grid being 0.18 to 0.45. Overall, only two runs (<5%) fall below the LRP of $20\%SB_{F=0}$.
- The most influential axis with respect to model output was size weighting. Up-weighting the size data led to a cluster of more optimistic model outputs, where terminal depletion

($SB_{latest}/SB_{F=0}$) was generally above 0.39, ranging between 0.29 and 0.49. In contrast, down-weighting of the size data led to a cluster of trajectories indicating lower estimates of biomass and higher levels of depletion, with a median of 0.27 and a range between 0.16 and 0.42 $SB/SB_{F=0}$ (Figure 39).

- The next most influential axes related to the assumptions made for tagging data. Assumption of a shorter mixing time tended to result in higher levels of depletion (Figure 39), although there was considerable overlap in trajectories between the two mixing period values assumed, driven by the other uncertainty axes. Down-weighting the tagging data in the overall likelihood led to more optimistic model outputs, with median terminal depletion of just under 0.4, and F_{recent}/F_{msy} around 0.7.
- The dynamics for the regional structure axis were more complex. Within the one-off sensitivity analysis, the 2014 regional structure led to slightly higher levels of depletion overall. Within the grid, however, the range of results under the 2014 regional structure fell within the range encompassed by that of runs under the assumption of the 2017 regional structure. There is also a clear interaction with the size weighting axis: under the up-weighting setting (10) the median for the 2014 region runs is lower than that for the 2017 region runs (0.38 *vs.* 0.42) whereas under the down-weighting scenario (50) the median for the 2017 region runs is the lower one (0.23 median for 2017 regions *vs.* 0.32 for the 2014 regions). This also highlights a higher variability for reference points for runs under the 2017 region structure.
- 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, although this effect is not as pronounced as for other more influential axes, with median of 0.31, 0.36 and 0.39 for runs under the 65%, 80% and 90% steepness scenarios, respectively. Accordingly, lower steepness maps to higher fishing mortality with respect to F_{MSY}, also corresponding to a higher median MSY (Figure 39; Appendix Table 6).

7.5.3 Further analyses of stock status

There are several ancillary analyses related to stock status that are typically undertaken on the reference case model (dynamic Kobe/Majuro analyses, fisheries impacts analyses, etc.). The shift towards relying more on multimodel inference for the 2017 assessment makes it more difficult to present these results over a large number of model runs. In this section we rely heavily on both fisheries impact analyses from the diagnostic case and the tabular results of the structural uncertainty grid (Table 6 and Table 15 and 16 for a split between regional structures). Where space allows, diagnostic plots are presented for models from the grid with different estimates of stock status.

Fishery impacts by region: We measure fishery impact at each time step as the ratio of the estimated spawning potential relative to the spawning potential that which would 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 42), and secondly as the depletion ratios themselves (Figure 43). The latter is relevant for the agreed reference points and are discussed in more detail below.

The diagnostic case model estimated that steady declines in spawning potential have occurred in all

regions (Figure 43), although the rate and extent of decline differs among regions. Tropical regions (regions 3, 4, 7, 8) show the more pronounced depletion, with region 8 especially around 20%. With the exception of region 4, all of these regions show an increase in both unfished and fished adult biomass since the last assessment, resulting, in the case of region 8 for instance, in an increase of almost 10% to less depleted levels. This recent increase is not seen for more temperate regions, which are the least depleted overall with northern regions (1 and 2) at around 80% and slightly better than their southern counterparts (regions 5 and 6) at around 60-65%. Finally, region 9 (i.e. the Coral Sea, and also the smallest region in the model), is estimated to have experienced virtually no depletion.

Fishery impact by gear and across regions: 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 may have the most impact on spawning potential (Figure 44). The relative contributions of different fishing gears are very similar to that of the 2014 assessment, with very early impacts on the population driven primarily by longline effort, but as the PS fishery expanded from the 1980's the impact of associated PS fishing has steadily increased, such that it has led to that gear having considerably more impact than LL fishing in all the equatorial regions. 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 neighbouring regions that are linked by movement. Unassociated PS fishing has a lower impact than associated PS in most regions but is still substantial. In region 8, the impact by this fishery compared to associated sets has been increasing and is larger in the recent years, particularly since 2010, the only region for which that is the case.

Yield analysis and equilibrium estimates across the grid: The yield analyses conducted in this assessment incorporate the spawner recruitment relationship (Figure 32) 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 1,193,336 and 2,813,584mt (Median = 2,113,938 mt; Table 6), and the spawning potential that would support the MSY (SB_{MSY}) was estimated to be between 186,800 and 946,800 mt. The ratio of SB_{MSY} to SB_0 was estimated to be between 0.15 and 0.34 (mean = 0.24). The ratio of SB_{MSY} to $SB_{F=0}$ was estimated to be between 0.16 and 0.35 (median = 0.25). Levels relative to the $20\%SB_{F=0}$ and F_{msy} have been discussed above (see section 7.5.2).

A plot of the yield distribution under different values of fishing effort relative to the current effort are shown in Figure 45 for two models representing the levels of the key axis of grid uncertainty, being the level of size data weighting. These support the general observations for other model output. The model with the upweighting of the size data (*Size10*) estimates that MSY would be achieved with moderately higher fishing effort, and the model with down-weighting of the size data (*Size50*) indicate that current effort is close to that giving MSY (Table 9).

The yield analysis also enables an assessment of the MSY level that would be theoretically achievable under the different patterns of age-specific fishing mortality observed through the history of the fishery. We present a plot for the diagnostic case model in Figure 46. Prior to 1970, the WCPO yellowfin fishery was almost exclusively conducted using LL gear, with a low exploitation of small yellowfin. Fisheries other than longline were known to operate in the region before 1970, but no catch estimates are available. The associated age-specific selectivity pattern resulted in a much

higher MSY in the early period compared to the recent estimates. This pronounced decline occurred after the expansion of the small-fish fisheries in region 7 and, soon after, the rapid expansion of the PS fishery which shifted the age composition of the catch dramatically towards much younger fish.

Dynamic Majuro plots and comparisons with Limit Reference Points: The section summarising the structural uncertainty grid (Section 7.5.2) presents terminal estimates of stock status in the form of Majuro plots. Further analyses can estimate the time-series of stock status in the form of Majuro plots, the methods of which are presented in Section 5.8.4. The large number of model runs in the structural uncertainty grid precludes undertaking and presenting this process for all runs, however an example for each of the levels of the key uncertainty axis (weighting on size composition) is included in Figure 47 and 48 for the 2017 region structure. These examples produce intuitive results with respect to the terminal results already presented in Section 7.5.2. All models at the start of the assessment period were close to an $SB/SB_{F=0}$ of one and an F/F_{msy} approaching zero, but each progressively tracked towards the overfishing and overfished definitions over the remaining period, with the *Size50* model much closer to both $20\%SB_{F=0}$ and an F/F_{msy} of 1 in final years than the *Size10* model. The diagnostic case model never reaches $20\%SB_{F=0}$ or an F/F_{msy} of 1 (Table 7).

8 Discussion and conclusions

8.1 Changes to the previous assessment

The 2017 stock assessment introduces a number of changes from the 2014 assessment that have had a large influence on the resulting estimates of stock status. Three additional years of data (tagging, catch, effort, size frequencies) were included in the assessment, covering a period of strong El Nino conditions, and a period of generally increasing catch levels. Within this period there has been an increase in several of the standardised CPUE indices in the terminal years of the model. The model attributes this to a period of slightly higher recruitments in some regions before the upturn in the CPUE (which in most cases indexes the abundance of older fish vulnerable to longline gear). This has also resulted in an increase in stock status indicators in the last several years (e.g. diagnostic case model compared to 2014 reference case model).

Other changes made to the model included implementing minor developments to MFCL that have become available since the 2014 assessment. These included developments in the modelling of recruitment (annual SRR, arithmetic rather than geometric mean of other recruitments), a trial of the Dirichlet multinomial likelihood for the size frequency data and estimation of a Lorenzen-type relationship between natural mortality and the size of fish. The values selected for these model settings had notable impacts on the estimates of stock status for the current assessment, and are discussed further in Section 8.4.

8.2 Investigation of a modified regional structure

The investigation of a new regional structure for yellowfin tuna was influenced by discussions related to bigeye tuna tagging, as well as approaches to better capture the spatial patterns of key fishing methods. A recommendation of the PAW in April 2017 (Pilling and Brouwer, 2017) was to investigate a new regional structure to better reflect the limited movement seen for bigeye between the equatorial zone (10° S–10° N) and more temperate areas that has been observed for fish tagged throughout the

Pacific (Schaefer et al., 2015) a pattern of movement that yellowfin appears to share to an extent (but see Section 8.6). This structure was also proposed as a means to better reflect the spatial division of the equatorial purse seine fishing zone and areas where fishing is more dominated by longline fishing, to ensure consistent regional configurations between species to faciliate management analyses, and also, incidentally, better delineates some oceanography features that impact vertical habitat for fish and, consequently, their vulnerability to fishing gear. The assumed regional structure had, by comparison to bigeye tuna, a relatively small impact on estimated stock status which was observed in the stepwise model development, the one-off sensitivity model runs and when the two regional structures were included in the structural uncertainty grid. The 2017 regional structure leads to slightly more optimistic results, which can be attributed to the model redistributing the stock biomass from the more heavily exploited equatorial regions to the northern temperate regions, which are less exploited. Consequently, the general relative pattern in sub-regional depletion trajectories in the 2014 and 2017 regional structures are relatively similar (higher in equatorial regions, lower in northern temperate regions), though because more of the population is assumed to be present in the northern regions under the 2017 regional structure (which now show less depletion), the net depletion estimated for the WCPO is slightly lower. In comparison to other axes of uncertainty examined within this assessment, however, the impact of the alternative spatial structure on stock status estimates was relatively low.

The potential impact of oceanography on recruitments, *via* the pronounced ENSO events of recent years, might also contributed to the impact of the change in regional structure, as the abrupt improvement in population numbers in the recent years is only seen in tropical regions (where the effects of ENSO on oceanographic variables likely to impact yellowfin recruitment, like sea surface temperature, are also the strongest). Giving increased weight to temperate regions under the 2017 regional structure will therefore limit the increase in recent years.

8.3 Biological considerations

Within the 2014 and the current assessment, growth was estimated as part of the MULTIFAN-CL model fit. The resulting growth curve fitted the size data relatively well. The estimated growth appeared plausible as a result, and matched almost exactly that estimated in the 2014 assessment. However, some uncertainty in growth and variability within the WCPO is noted within the introductory biology section (Section 3.2), and the potential impact on stock assessments of uncertainty in growth and maturity has been clearly demonstrated within the 2017 bigeye stock assessment. A recent review of YFT underscored the relative paucity of biological knowledge for this species given its economical importance (Pecoraro et al., 2016). **Pursuing biological studies to improve the understanding of the growth of yellowfin within the WCPO through direct ageing of otoliths would help reduce potential uncertainty in this area**. With respect to maturity, and as noted within the bigeye report, **re-parameterising MFCL to allow maturity-at-length, rather than maturity-at-age, would more seamlessly allow the impacts of changes in growth on maturity schedules to be captured within the model itself**.

As noted in Section 3.1, recent genetic studies have suggested a finer geographic stock structure than considered within this assessment. While it does not appear that those studies are yet sufficiently informative to drive changes in the assessment regional structure, the findings of the latest study, supported by the WCPFC Tissue Bank (Project 35b) should be monitored, especially given that the latest regional structure was driven by understanding about bigeye movement, which was assumed to translate into yellowfin population dynamics in the absence of alternative evidence.

Two further biological uncertainties are the value of steepness to be used within the stock recruitment relationship, and the value of natural mortality. For steepness, there is little additional information available with which to estimate this parameter, and hence the approach of examining the influence of alternative values of steepness on management quantities, as used in all other recent WCPO tuna stock assessments, has been used here. For natural mortality, alternative potential formulations of *M-at-age* were examined within this assessment as one-off sensitivities. The recent additional functionality in MULTIFAN-CL that allows alternative functional forms to be assumed will allow this to be pursued further in future WCPO tuna assessments. Analysis of available tagging data to examine the potential for estimating natural mortality rates of yellowfin may also be warranted.

8.4 Issues with the weighting of data components

This assessment is the first for yellowfin in the WCPO that quantifies the influence of the different data components on the scaling of the biomass estimates, and represents an area of ongoing research. The new method of conducting likelihood profiles with respect to the derived parameter of mean total biomass has proven to be successful, and avoids some of the problems associated with more traditional profiles on the population scaling parameter (total population scaling parameter in MFCL, which was performed in previous assessments). However, as noted within the bigeye assessment, the results indicate conflict between the various data components and more especially within the length and weight component of the size composition data (Figure 49). Now that these conflicts have been assessed, a major task in the intersessional period will be investigating means to reduce conflict and/or develop further methods of objectively weighting the various data components, continue to improve methods to develop likelihood profiles that are informative of model behaviour, and understand the potential cause of the conflict between the available length and weight data. In the meantime, the approach of investigating a range of weightings for components, as part of the one-off sensitivity analyses and structural uncertainty grid, as pursued in this assessment, have shown the sensitivity, and robustness, of the estimates to the assumptions made in the assessment. As noted in the bigeye assessment report, the Dirichlet-multinomial (and similar self-scaling likelihoods for compositional data) provides the potential to estimate relatively objective effective sample sizes for length- and weight-frequency data, and potentially conditional age-at-length data. Following further testing and evaluation, it is likely that one of these approaches will be used as a future standard for modelling compositional data in MFCL. **This task should be considered a high priority for yellowfin given the sensitivity of population status estimates to this model input**.

8.5 Transition from single to multi-model inference

This assessment was the first for yellowfin where population status shifted away from that of a single reference case model to a summary statistic aggregated from a number of model runs over key uncertainty axes. The current results highlight the importance of carefully selecting these axes to balance, on the one hand, the chosen levels to be representative of key model uncertainties and on the other hand, the more practical consideration that computing time increases rapidly as more grid axes are added. Given that in the current assessment individual model runs took on average 2 to 3 days to converge, this latter consideration is not trivial. The non-linearity of the impact of a key grid component, size composition, on depletion estimates, with the weighting of 20 used in the diagnostic case suggesting less depletion than both the 10 and 50 weightings used with the size composition axis, further underscores the importance of selecting grid axes and the levels within

them carefully. In the current case, the average population status would have probably been higher had a size data weighting level of 20 also been included in the grid, but this would also have implied the inclusion of levels with more weight to the composition data. In addition, interaction among some key grid axis (e.g. region configuration and size weighting) means that more complex dynamics can emerge from grid selection.

8.5.1 Trends in Region 8

The 2014 assessment noted a key conflict between the potential standardised time series and tagging data for region 8. This conflict is still present in the assessment, and was resolved, similar to 2014, with the use of the available purse seine time series which appeared to reduce the conflict but also had both a smaller temporal span and a more fluctuating trend in time. This was used within the diagnostic case model and the structural uncertainty grid runs, while an alternative model was run with longline time series (LL ALL 8) as a a one-step sensitivity run. It is noted that potential shifts in longline targeting within the region may not yet be fully captured within the CPUE standardisation of this fishery. A challenge appeared to be reconciling the trend seen within the longline time series with the PTTP tagging information, in particular the large tagging events in the more recent period specifically for that region. This tended to lead to high reporting rate estimates for the purse seine fisheries in this region, which were responsible for the bulk of PTTP tag returns (over 75%). This may be influenced by tagging activities being on the anchored FADs used by the fleet in this area, which may influence mixing rates. While the impact of this conflict on overall population reference points appeared relatively minimal (Table 8), investigation of the potential reporting rate for these tagging events is warranted, through tag seeding work. This is due to both the extent of tagging activity in this region, and region 8 also being predicted by the model to be the most depleted (noting the estimated improvement in recent years).

8.6 Data sources

This assessment has been supported by access to two new sources of information. The first is the operational data set that has offered a much more extensive source of information for CPUE analysis, including an extended time series that has clearly led to improvements in the analysis. Alternative CPUE analysis approaches have been feasible, which has helped evaluate the uncertainties that arise from these input time series. The second is the availability of information from the Japanese tagging programme. This was examined as a one-off sensitivity in the current assessment, but further collaborative work would be welcome to investigate the pattern of movement, and the potential changes in reporting rate that may occur as fish move from model region to model region. This latter aspect is especially important and applies to the PTTP as well. In recent years “range contraction" has been mentioned a number of times in SC discussions, initially raised by southern countries for yellowfin in the 2010s and more recently for skipjack. For this discussion to move forward, knowledge on the movement of individuals between tropical and temperate regions is required but is largely lacking given the practical obstacles of tracking larger components of populations through large areas (*vs.* satellite tags which are expensive and can only be put on a few large individuals, limiting the scope for large scale conclusions about population movement).

The WPEA project has led to improvements in the understanding of key fisheries operating within region 7 of the model. Continued extension and improvement of this data set will increasingly benefit future yellowfin assessments given the relatively high proportion of the total yellowfin catch that

occurs there. Improved statistics for fisheries in this region should continue to be a high priority for yellowfin management given the predicted depletion of biomass in this region and the high and evolving impact of non-traditional fisheries in this area.

8.7 Main assessment conclusions

The main conclusions of the current assessment below are based upon the total grid of 48 model runs. Scientific Committee decisions on the grid runs to be used for management advice may, in part, be influenced by decisions made regarding the regional structure used within the bigeye assessment, although we note that the yellowfin assessment uncertainty grid results appear relatively robust to that assumption and is instead more sensitive to the weighting on the size composition data. Separate tables for formulating such advice are presented in Appendix (Section 11). The general conclusions of this assessment can be summarised as follows:

1. The grid contains a wide range of models with some variation in estimates of stock status, trends in abundance and reference points. However, biomass is estimated to have declined throughout the model period for all models in the grid. Those declines are found across most tropical and temperate regions of the model.

2. Fishing mortality for adult and juvenile yellowfin tuna is estimated to have increased continuously since the beginning of industrial tuna fishing (seen in the diagnostic case model). In general these have been on average higher for juveniles, but in recent years adult fishing mortality has also increased. A significant component of the increase in juvenile fishing mortality is attributable to the Philippines, Indonesian and Vietnamese surface fisheries, which have the most uncertain catch, effort and size data. The work of the WPEA project to assist in enhancing the current fishery monitoring programme and improving estimates of historical and current catch from these fisheries remains important given the contribution of these fisheries in the overall fishing impact analyses from this assessment.

3. Across the model grid, the terminal depletion estimated for the majority of runs estimate stock status levels to be above the LRP of $20\%SB_{F=0}$. The range of $SB_{latest}/SB_{F=0}$ values was 0.18 to 0.45. Only two runs (<5%) fell below $20\%SB_{F=0}$. The median estimate (0.33) is comparable to that estimated from the 2014 assessment grid, noting the differences in grid uncertainty axes used in that assessment.

4. Corresponding estimates of F_{recent}/F_{msy} ranged from 0.58 to 1.13, with 2 out of the 48 runs (<5%) indicating that $F_{recent}/F_{msy} > 1$. The median estimate (0.75) is also comparable to that estimated from the 2014 asssessment grid.

5. The significance of the recent increased recruitment events and the progression of these fish to the spawning potential component of the stock are encouraging, although whether this is a result of management measures for the fishery or beneficial environmental conditions is currently unclear. It is noteworthy, however, that recent favourable recruitment events have also been estimated for skipjack (McKechnie et al., 2016a) and bigeye (McKechnie et al., 2017a) in the WCPO, and bigeye in the EPO (Aires-da Silva et al., 2017), which may give weight to the favourable environmental conditions hypothesis. Whether these trends are maintained in coming years will help separate these factors and will likely provide more certainty about the future trajectories of the stock.

6. The current assessment investigated the impact of a wider range of uncertainties than previous

yellowfin assessments. Nonetheless, issues within the model have been highlighted, and there remains a range of other model assumptions that should be investigated either internally or through directed research. These are noted in the main text, but briefly, the apparent non-linear impact of the weighing on the size composition data on population estimates, and the conflict between the abundance indices and the tagging data for region 8 are worthy of note. Also, biological studies to improve our estimates of growth of yellowfin within the WCPO, for instance through direct ageing of otoliths as was done in bigeye, should be considered a high priority.

7. The addition of new data sources, including the use of the operational longline data set for CPUE standardisation purposes, and the Japanese tagging data, were very helpful in improving the current assessment and the understanding of yellowfin biology. Additional collaborative work using these data sets would further improve the assessment, especially if we can get an improved understanding of reporting rates for the Japanese tagging data.

8.8 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 Keisuke Satoh and Takayuki Matsumoto, for the provision of the Japanese tagging data. Peter Williams and his team from SPC provided exhaustive assistance in collating and interpreting catch statistics for this species.

References

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

Aguila, R., Perez, S., Catacutan, B., Lopez, G., Barut, N., and Santos, M. (2015). Distinct yellowfin tuna (*Thunnus albacares*) stocks detected in western and central pacific ocean (wcpo) using dna microsatellites. *PLoS ONE*, 10(9):e0138292.

Aires-da Silva, A., Minte-Vera, C., and Maunder, M. N. (2017). Status of bigeye tuna in the eastern Pacific Ocean in 2016 and outlook for the future. IATTC, Scientific Advisory Committee, Eighth Meeting, SAC-08-04a.

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. WCPFC-SC10-2014/SA-IP-06, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

Bigelow, K. and Garvilles, E. (2017). Relative abundance of yellowfin tuna for the purse seine and handline fisheries operating in the Philippines Moro Gulf (Region 12) and High Seas Pocket #1. WCPFC-SC13-2014/SA-IP-07, Rarotonga, Cook Islands, 9-17 August 2017.

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.

Cadrin, S. and Vaughan, D. (1997). Retrospective analysis of virtual population estimates for Atlantic menhaden stock assessment. *Fishery Bulletin*, 95(3):256–265.

Cordue, P. L. (2013). Review of species and size composition estimation for the western and central pacific purse seine fishery. WCPFC-SC9-2013/ST-IP-02, Pohnpei, Federated States of Micronesia, 6–14 August 2013.

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

Davies, N., Fournier, D. A., Hampton, J., Hoyle, S., Bouye, F., and Harley, S. (2017). Recent developments in the MULTIFAN-CL stock assessment software. WCPFC-SC13-2017/SA-IP-05, Rarotonga, Cook Islands, 9–17 August 2017.

Davies, N., Harley, S., Hampton, J., and McKechnie, S. (2014). Stock assessment of yellowfin tuna in the Western and Central Pacific Ocean. 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. WCPFC-SC7-2011/SA-WP-02, Pohnpei, Federated States of Micronesia, 9–17 August 2011.

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.

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. (Unpublished report held in NIWA library, Wellington.).

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

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

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

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

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

Hampton, J., Kleiber, P., Langley, A., and Hiramatsu, M. (2004). Stock assessment of yellowfin tuna in the western and central Pacific Ocean. WP SA-1, SCTB 17, Majuro, Republic of Marshall Islands, 9-18 August 2004.

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

Hampton, J., Langley, A., and Kleiber, P. (2006). Stock assessment of yellowfin tuna in the western and central pacific ocean including an analysis of management options. WCPFC SC2 SA WP-1, Manila, Philippines, 7-18 August 2006.

Hampton, J. and Williams, P. (2017). Annual estimates of purse seine catches by species based on alternative data sources. WCPFC-SC13-2017/SA-IP-03, Rarotonga, Cook Islands, 9–17 August 2017.

Harley, S. J. (2011). A preliminary investigation of steepness in tunas based on stock assessment results. WCPFC-SC7-2011/SA-IP-08, Pohnpei, Federated States of Micronesia, 9–17 August 2011.

Harley, S. J., Davies, N., Tremblay-Boyer, L., Hampton, J., and McKechnie, S. (2015). Stock assessment of south Pacific albacore tuna. WCPFC-SC11-2015/SA-WP-06, Pohnpei, Federated States of Micronesia, 5–13 August 2015.

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., Langley, A. D., and Hampton, J. (2008). General structural sensitivity analysis for the bigeye tuna stock assessment. WCPFC-SC4-2008/SA-WP-03, Port Moresby, Papua New Guinea, 11–22 August 2008.

Hoyle, S. and Nicol, S. (2008). Sensitivity of bigeye stock assessment to alternative biological and reporductive assumptions. 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. WCPFC-SC4-2008/ME-WP-02, Port Moresby, Papua New Guinea, 11–22 August 2008.

Hoyle, S. D. and Okamoto, H. (2011). Analyses of Japanese longline operational catch and effort for Bigeye and Yellowfin Tuna in the WCPO. WCPFC-SC7-2011/SA-IP-01, Pohnpei, Federated States of Micronesia, 9–17 August 2011.

Ianelli, J., Maunder, M. N., and Punt, A. E. (2012). Independent review of the 2011 WCPO bigeye tuna assessment. 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.

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

Langley, A., Hampton, J., Kleiber, P., and Hoyle, S. D. (2007). Stock assessment of yellowfin tuna in the Western and Central Pacific Ocean, including an analysis of management options. WCPFC SC3 SA WP-1, Honolulu, Hawaii, 13-24 August 2007.

Langley, A., Harley, S., Hoyle, S., Davies, N., Hampton, J., and Kleiber, P. (2009). Stock assessment of yellowfin tuna in the Western and Central Pacific Ocean. WCPFC SC7 SA WP-3, Pohnpei, Federated States of Micronesia, 9-17 August 2011.

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

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

Lawson, T. (2013). Update on the estimation of the species composition of the catch by purse seiners in the Western and Central Pacific Ocean, with responses to recent independent reviews. WCPFC-SC9-2013/ST-WP-03, Pohnpei, Federated States of Micronesia, 6–14 August 2013.

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.

Lehodey, P., Hampton, J., and Leroy, B. (2012). Preliminary results on age and growth of bigeye tuna (thunnus obesus) from the western and central pacific ocean as indicated by daily growth increments and tagging data. Technical Report SCTB-12-1999/WP-BET-2, Papeete, French Polynesia, 16–23 June 1999.

Lorenzen, K. (1996). The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystem and aquaculture. *Journal of Fish Biology*, 42:627–647.

McArdle, B. (2013). To improve the estimation of species and size composition estimation of the western and central Pacific purse seine fishery from observer-based sampling of the catch. WCPFC-SC9-2013/ST-IP-04, Pohnpei, Federated States of Micronesia, 6–14 August 2013.

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

McKechnie, S., Hampton, J., Pilling, G. M., and Davies, N. (2016a). Stock assessment of skipjack tuna in the western and central Pacific Ocean. WCPFC-SC12-2016/SA-WP-04, Bali, Indonesia, 3–11 August 2016.

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

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

McKechnie, S., Ochi, D., Kiyofuji, H., Peatman, T., and Caillot, S. (2016b). Construction of tagging data input files for the 2016 skipjack tuna stock assessment in the western and central Pacific Ocean. 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. WCPFC-SC13-2017/SA-WP-05, Rarotonga, Cook Islands, 9–17 August 2017.

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

McKechnie, S., Tremblay-Boyer, L., and Pilling, P. (2017b). Background analyses for the 2017 stock assessments of bigeye and yellowfin tuna in the western and central Pacific Ocean. WCPFC-SC13-2017/SA-IP-06, Rarotonga, Cook Islands, 9–17 August 2017.

MRAG Asia Pacific (2016). *Towards the quantification of illegal, unreported and unregulated (IUU) fishing in the Pacific Islands Region.* MRAG Asia Pacific, 101p.

OFP (2017). Estimates of annual catches in the WCPFC Statistical Area. Technical Report SC13-ST–IP–01, Rarotonga, Cook Islands, 9–17 August 2017.

Peatman, T., Caillot, S., Leroy, B., McKechnie, S., Roupsard, F., Sanchez, C., Nicol, S., and Smith, N. (2016). Analysis of tag seeding data and reporting rates. WCPFC-SC12-2016/SA-IP-13, Bali, Indonesia, 3–11 August 2016.

Pecoraro, C., Zudaire, I., Bodin, N., Murua, H., Taconet, P., Díaz-Jaimes, P., Cariani, A., Tinti, F., and Chassot, E. (2016). Putting all the pieces together: integrating current knowledge of the biology, ecology, fisheries status, stock structure and management of yellowfin tuna (thunnus albacares). *Reviews in Fish Biology and Fisheries*.

Pilling, G. and Brouwer, S. (2017). Report from the spc pre-assessment workshop, noumea, april 2017. Technical Report WCPFC-SC13-2017/SA-IP-02, Rarotonga, Cook Islands, 9–17 August 2017.

Powers, J. E. (2013). Review of SPC estimation of species and size composition of the western and central Pacific purse seine fishery from observer-based sampling of the catch. WCPFC-SC9-2013/ST-IP-03, Pohnpei, Federated States of Micronesia, 6–14 August 2013.

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.

SPC-OFP (2013). Purse seine effort: a recent issue in logbook reporting. WCPFC-TCC9-2013-18, Pohnpei, Federated States of Micronesia.

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

Tremblay-Boyer, L. and Pilling, G. (2017a). Geo-statistical analyses of operational longline CPUE data. WCPFC-SC13-2017/SA-WP-03, Rarotonga, Cook Islands, 9–17 August 2017.

Tremblay-Boyer, L. and Pilling, G. (2017b). Use of operational vessel proxies to account for vessels with missing identifiers in the development of standardised CPUE time series. WCPFC-SC13-2017/SA-WP-04, Rarotonga, Cook Islands, 9–17 August 2017.

Ward, R., Elliott, N., and Grewe, P. (1994). Allozyme and mitochondrial dna variation in yellowfin tuna (thunnus albacares) from the pacific ocean. *Marine Biology*, 118:531–539.

WCPFC (2011). Report of the Scientific Committee seventh regular session. Summary report, Commission for the Conservation and Management of Highly Migratory Fish Stocks in the Western and Central Pacific Ocean., Pohnpei, Federated States of Micronesia. 9-17 August 2011.

Williams, P. and Terawasi, P. (2017). Overview of tuna fisheries in the Western and Central Pacific Ocean, including economic conditions–2016. WCPFC-SC13-2017/GN-WP-01, Rarotonga, Cook Islands, 9–17 August 2017.

9 Tables

Table 1: Definition of fisheries for the MULTIFAN-CL yellowfin analysis. Gears: PL = pole and line; PS = purse seine unspecified set type; LL = longline; DOM = the range of artisanal gear types operating in the domestic fisheries of Philippines and Indonesia. Flag/fleets: JPN = Japan; PH = Philippines; ID = Indonesia; ALL = all nationalities.

Fishery	Nationality	Gear	Region
F1 L-ALL-1	ALL	LL	1
F2 L-ALL-2	ALL	LL	2
F3 L-US-2	US	LL	2
F4 L-ALL-3	ALL	LL	3
F5 L-OS-3	OS	LL	3
F6 L-OS-7	OS	LL	7
F7 L-ALL-7	ALL	LL	7
F8 L-ALL-8	ALL	LL	8
F9 L-ALL-4	ALL	LL	4
F10 L-AU-5	AU	LL	5
F11 L-ALL-5	ALL	LL	5
F12 L-ALL-6	ALL	LL	6
F13 S-ASS-ALL-3	ALL	PS	3
F14 S-UNA-ALL-3	ALL	PS	3
F15 S-ASS-ALL-4	ALL	PS	4
F16 S-UNA-ALL-4	ALL	PS	4
F17 Z-PH-7	PH	Dom	7
F18 Z-ID.PH-7	ID.PH	Dom	7
F19 S-JP-1	JP	PS	1
F20 P-JP-1	JP	PL	1
F21 P-ALL-3	ALL	PL	3
F22 P-ALL-8	ALL	PL	8
F23 Z-ID-7	ID	Dom	7
F24 S-ID.PH-7	ID.PH	PS	7
F25 S-ASS-ALL-8	ALL	PS	8
F26 S-UNA-ALL-8	ALL	PS	8
F27 L-AU-9	AU	LL	9
F28 P-ALL-7	ALL	PL	7
F29 L-ALL-9	ALL	LL	9
F30 S-ASS-ALL-7	ALL	PS	7
F31 S-UNA-ALL-7	ALL	PS	7
F32 Z-VN-7	VN	Dom	7

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

Prog		CSTP			PTTP			RTTP	
Years		1991–1995			2006–2014			1989–1992	
Category	Grps	Rel	Rec	Grps	Rel	Rec	Grps	Rel	Rec
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	0	0	0	12	5066	1086	7	1518	239
4	0	0	0	9	1672	372	5	1819	180
5	0	0	0	5	3087	460	4	820	105
6	0	0	0	1	7	0	3	221	8
7	0	0	0	3	6456	1401	7	8239	2378
8	0	0	0	15	39600	11564	9	7957	1241
9	3	2343	70	0	0	0	0	0	0
Total	3	2343	70	45	55888	14883	35	20574	4151

Table 3: Summary of the groupings of fisheries within the assessment for estimation of selectivity, catchability (used for the implementation of regional weights), tag recaptures, and tag reporting rates. Note that effort is missing for all L and Z fisheries and so effort deviation penalties only apply to the last four quarters (see Section 5.4.3). See Table 1 for further details on each fishery.

Fishery	Region	Selectivity	SeasCat	TimVarCat	TimVarCatCV	EffPen	EffPenCV	Recaptures	Reporting
F1 L-ALL-1	1	1	Y	N	NA	time-variant	0.20	1	1
F2 L-ALL-2	2	1	Y	N	NA	time-variant	0.20	2	1
F3 L-US-2	2	2	Y	Y	0.1	scaled	0.41	3	20
F4 L-ALL-3	3	3	Y	N	NA	time-variant	0.20	4	1
F5 L-OS-3	3	4	Y	Y	0.1	scaled	0.41	5	1
F6 L-OS-7	7	5	Y	Y	0.1	scaled	0.41	6	1
F7 L-ALL-7	7	6	Y	N	NA	time-variant	0.20	7	1
F8 L-ALL-8	8	7	Y	N	NA	constant	0.71	8	1
F9 L-ALL-4	4	3	Y	N	NA	time-variant	0.20	9	1
F10 L-AU-5	5	8	Y	Y	0.1	scaled	0.41	10	21
F11 L-ALL-5	5	3	Y	N	NA	time-variant	0.20	11	22
F12 L-ALL-6	6	3	Y	N	NA	time-variant	0.20	12	1
F13 S-ASS-ALL-3	3	9	Y	Y	0.71	scaled	0.41	13	23
F14 S-UNA-ALL-3	3	10	Y	Y	0.71	scaled	0.41	13	23
F15 S-ASS-ALL-4	4	9	Y	Y	0.71	scaled	0.41	14	24
F16 S-UNA-ALL-4	4	10	Y	Y	0.71	scaled	0.41	14	24
F17 Z-PH-7	7	11	N	N	NA	constant	0.41	15	25
F18 Z-ID.PH-7	7	12	Y	Y	0.1	time-variant	0.20	16	26
F19 S-JP-1	1	13	Y	Y	0.1	scaled	0.41	17	27
F20 P-JP-1	1	14	Y	Y	0.1	scaled	0.41	18	28
F21 P-ALL-3	3	15	Y	Y	0.1	scaled	0.41	19	29
F22 P-ALL-8	8	16	Y	Y	0.1	scaled	0.41	19	30
F23 Z-ID-7	7	11	N	N	NA	constant	0.41	20	31
F24 S-ID.PH-7	7	9	Y	N	NA	time-variant	0.20	21	32
F25 S-ASS-ALL-8	8	9	Y	N	NA	time-variant	0.20	22	33
F26 S-UNA-ALL-8	8	17	Y	Y	0.71	scaled	0.41	22	33
F27 L-AU-9	9	18	Y	Y	0.1	scaled	0.41	23	34
F28 P-ALL-7	7	19	N	Y	0.1	scaled	0.41	24	35
F29 L-ALL-9	9	3	Y	N	NA	constant	0.71	25	36
F30 S-ASS-ALL-7	7	9	Y	Y	0.71	scaled	0.41	26	23
F31 S-UNA-ALL-7	7	10	Y	Y	0.71	scaled	0.41	26	23
F32 Z-VN-7	7	20	N	N	NA	constant	0.41	27	37

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 2011–2014 and 'latest' is 2015

Symbol	Description
C_{latest}	Catch in the last year of the assessment (2015)
F_{recent}	Average fishing mortality-at-age for a recent period (2011–2014)
F_{MSY}	Fishing mortality-at-age producing the maximum sustainable yield (MSY)
MSY	Equilibrium yield at F_{MSY}
$F_{recent}/F_{\mathrm{MSY}}$	Average fishing mortality-at-age for a recent period (2011–2014) relative to F_{MSY}
SB_0	Equilibrium unexploited spawning potential
SB_{latest}	Spawning potential in the latest time period (2015)
SB_{recent}	Spawning potential for a recent period (2011–2014)
$SB_{F=0}$	Average spawning potential predicted to occur in the absence of fishing for the period 2005–2014
SB_{MSY}	Spawning potential that will produce the maximum sustainable yield (MSY)
$SB_{latest}/SB_{F=0}$	Spawning potential in the latest time period (2015) relative to the average spawning potential predicted to occur in the absence of fishing for the period 2005–2014
$SB_{latest}/SB_{\mathrm{MSY}}$	Spawning potential in the latest time period (2015) relative to that which will produce the maximum sustainable yield (MSY)
$SB_{recent}/SB_{F=0}$	Spawning potential in for a recent period (2011–2014) relative to the average spawning biomass predicted to occur in the absence of fishing for the period 2005–2014
$20\%SB_{F=0}$	WCPFC adopted limit reference point – 20% of spawning potential in the absence of fishing average over years $t-10$ to $t-1$ (2005–2014)

Table 5: Description of the structural sensitivity grid used to characterise uncertainty in the assessment.

Axis	Levels	Option
Steepness	3	0.65, 0.80, or 0.95
Tagging overdispersion	2	Default level, Fixed (moderate) level
Tag mixing	2	1 or 2 (default) quarters
Size frequency weighting	2	sample sizes divided by 10, or 50
Regional structure	2	2017 regions, 2014 regions

Table 6: Summary of reference points over all 48 individual models in the structural uncertainty grid

	Mean	Median	Min	25%	75%	Max
C_{latest}	612764	613660	606762	612889	614273	615350
MSY	662583	666800	539200	627700	690700	754400
$Y_{F_{recent}}$	642192	639200	534400	613200	677000	720400
f_{mult}	1.30	1.34	0.88	1.18	1.44	1.73
F_{MSY}	0.12	0.12	0.07	0.11	0.12	0.16
$F_{recent}/F_{\text{MSY}}$	0.79	0.75	0.58	0.70	0.85	1.13
SB_{MSY}	546433	578500	186800	375675	683000	946800
SB_0	2215167	2376500	1197000	1775000	2633250	3105000
SB_{MSY}/SB_0	0.24	0.24	0.15	0.22	0.26	0.34
$SB_{F=0}$	2113938	2273717	1193336	1716208	2438297	2813584
$SB_{MSY}/SB_{F=0}$	0.25	0.26	0.16	0.22	0.28	0.35
SB_{latest}/SB_0	0.32	0.33	0.18	0.27	0.38	0.45
$SB_{latest}/SB_{F=0}$	0.34	0.35	0.16	0.27	0.41	0.49
$SB_{latest}/SB_{\text{MSY}}$	1.38	1.39	0.80	1.15	1.61	1.91
$SB_{recent}/SB_{F=0}$	0.31	0.32	0.15	0.24	0.37	0.44
$SB_{recent}/SB_{\text{MSY}}$	1.37	1.39	0.81	1.26	1.55	1.81

10 Figures

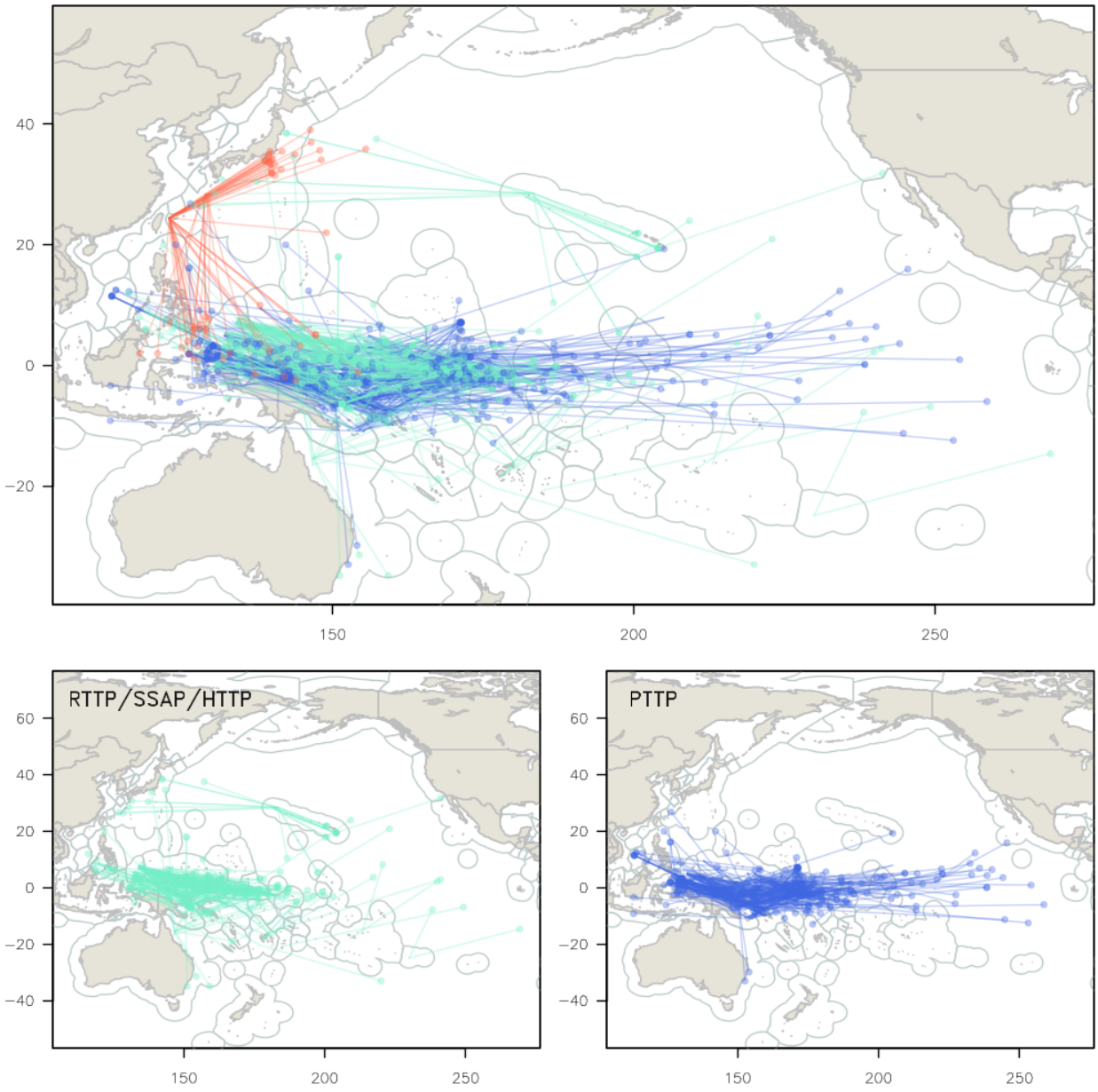

Figure 1: Map of the movements of tagged yellowfin released in the Pacific Ocean and subsequently recaptured more than 1,000 nautical miles from their release site.

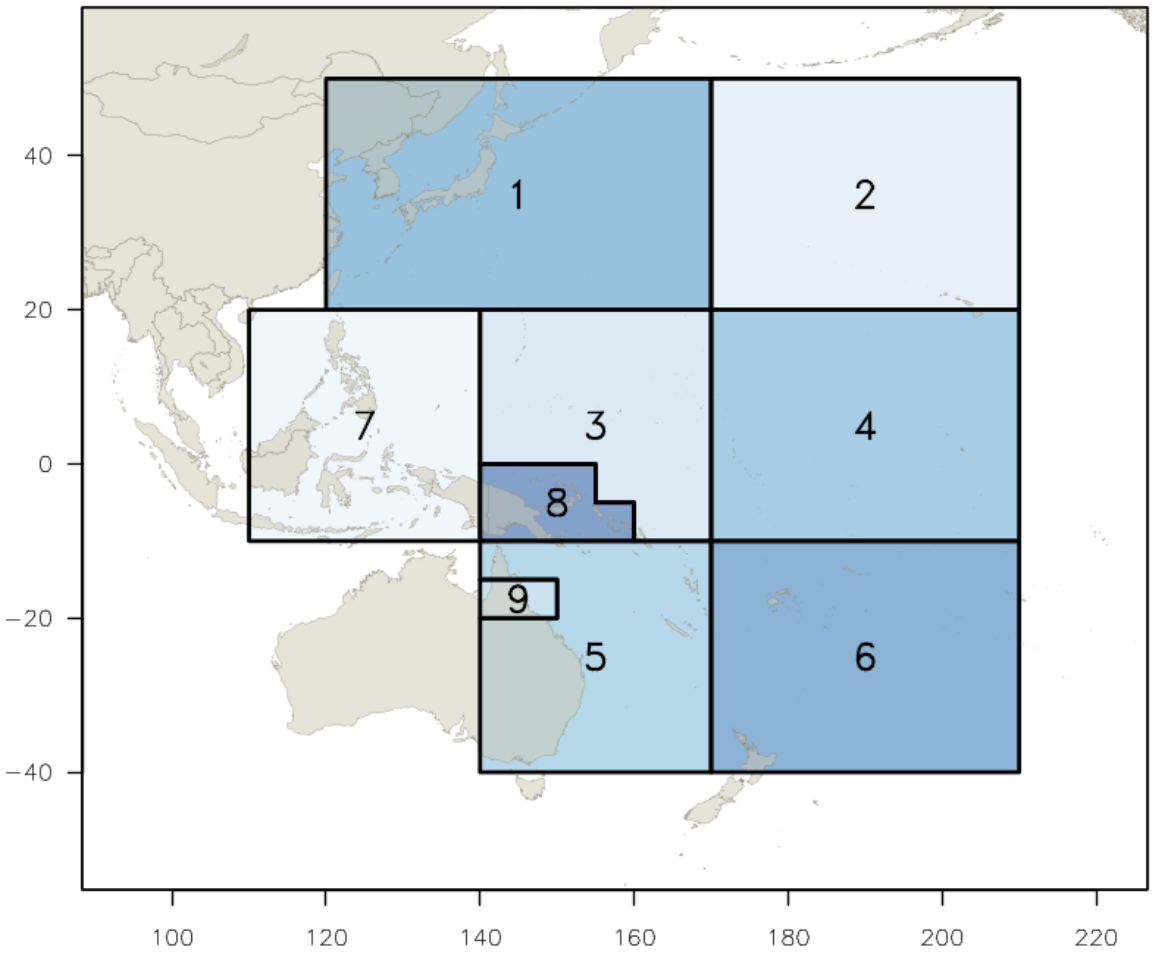

Figure 2: The geographical area covered by the stock assessment and the boundaries for the 9 regions when using the "2014 regional structure".

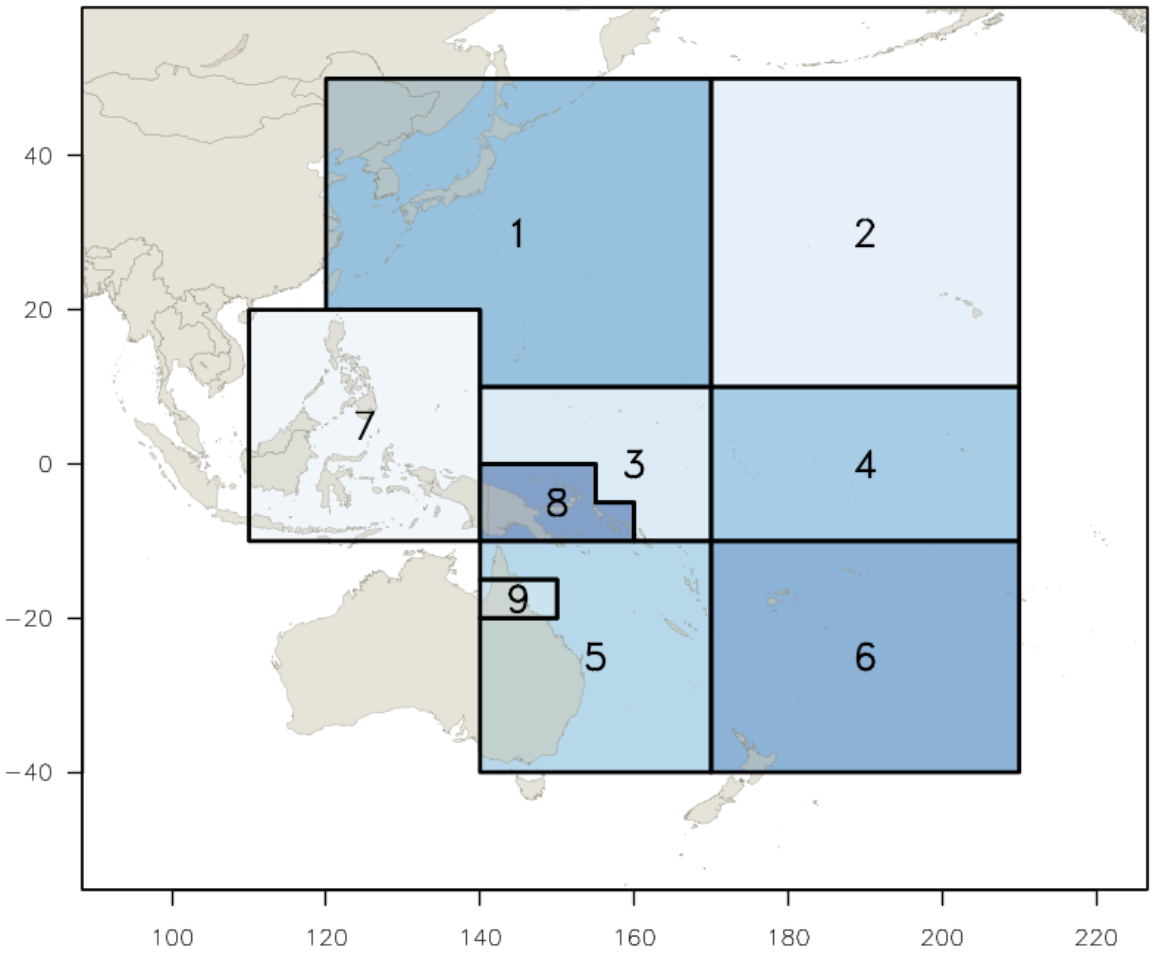

Figure 3: The geographical area covered by the stock assessment and the boundaries for the 9 regions when using the "2017 regional structure".

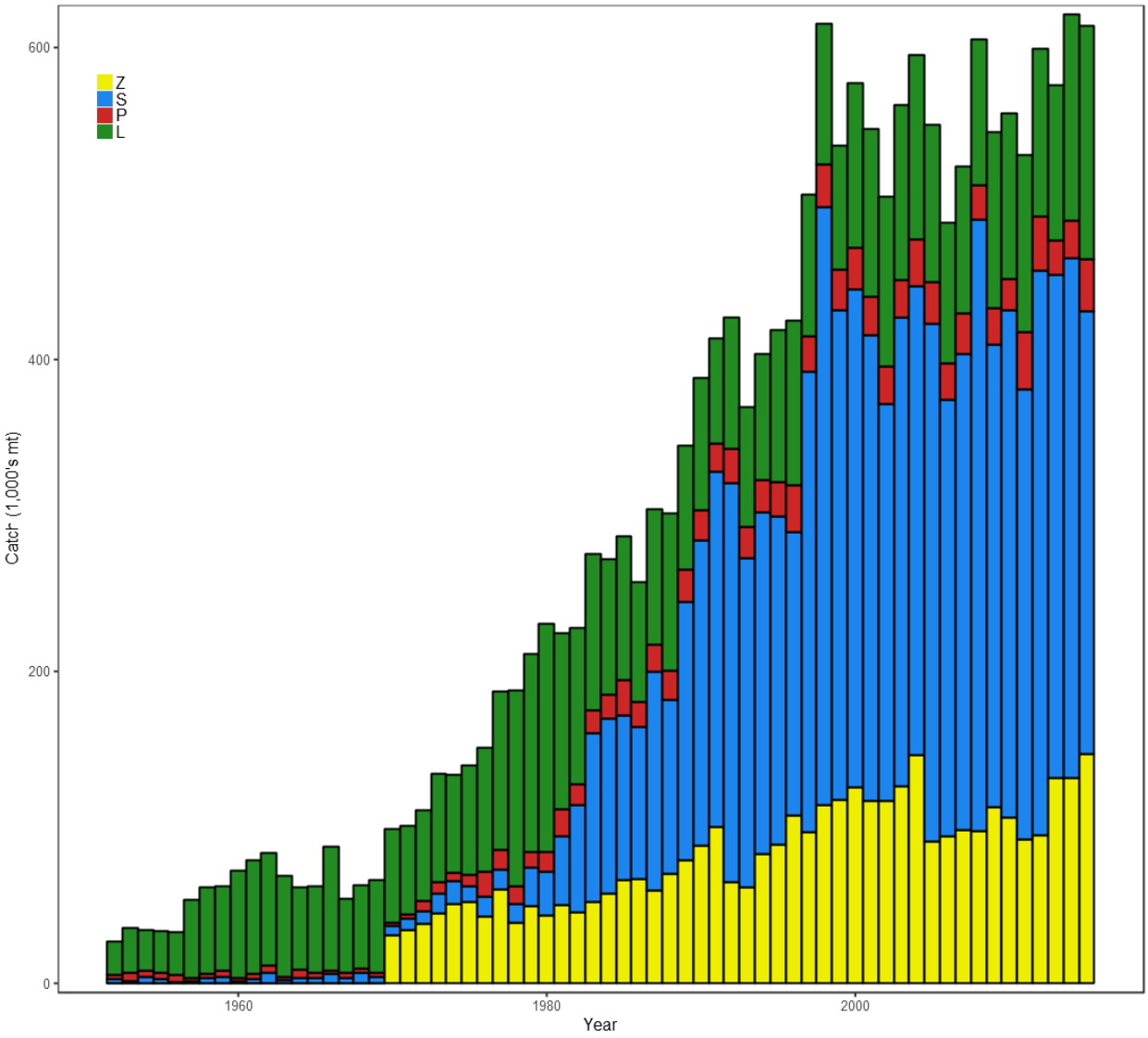

Figure 4: Time series of total annual catch (1000's mt) by fishing gear for the diagnostic case model over the full assessment period. The different colours refer to longline (green), pole-and-line (red), purse seine (blue) and miscellaneous (yellow). Note that the catch by longline gear has been converted into catch-in-weight from catch-in-numbers and so estimates differs from the annual catch estimates presented in (Williams and Terawasi, 2017), however these catches enter the model as catch-in-numbers.

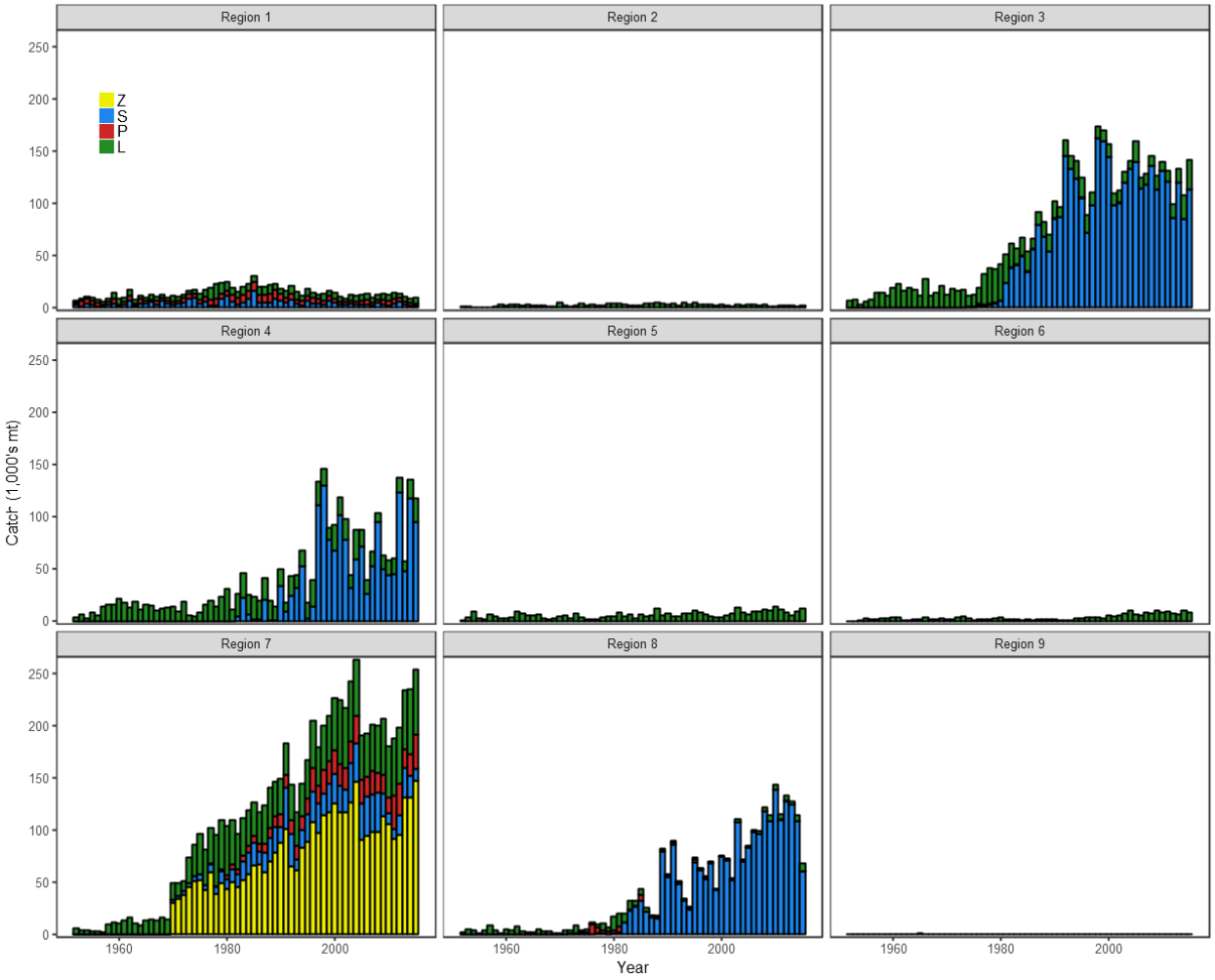

Figure 5: Time series of total annual catch (1000's mt) by fishing gear and assessment region from the diagnostic case model over the full assessment period. The different colours denote longline (green), pole-and-line (red), purse seine (blue) and miscellaneous (yellow).

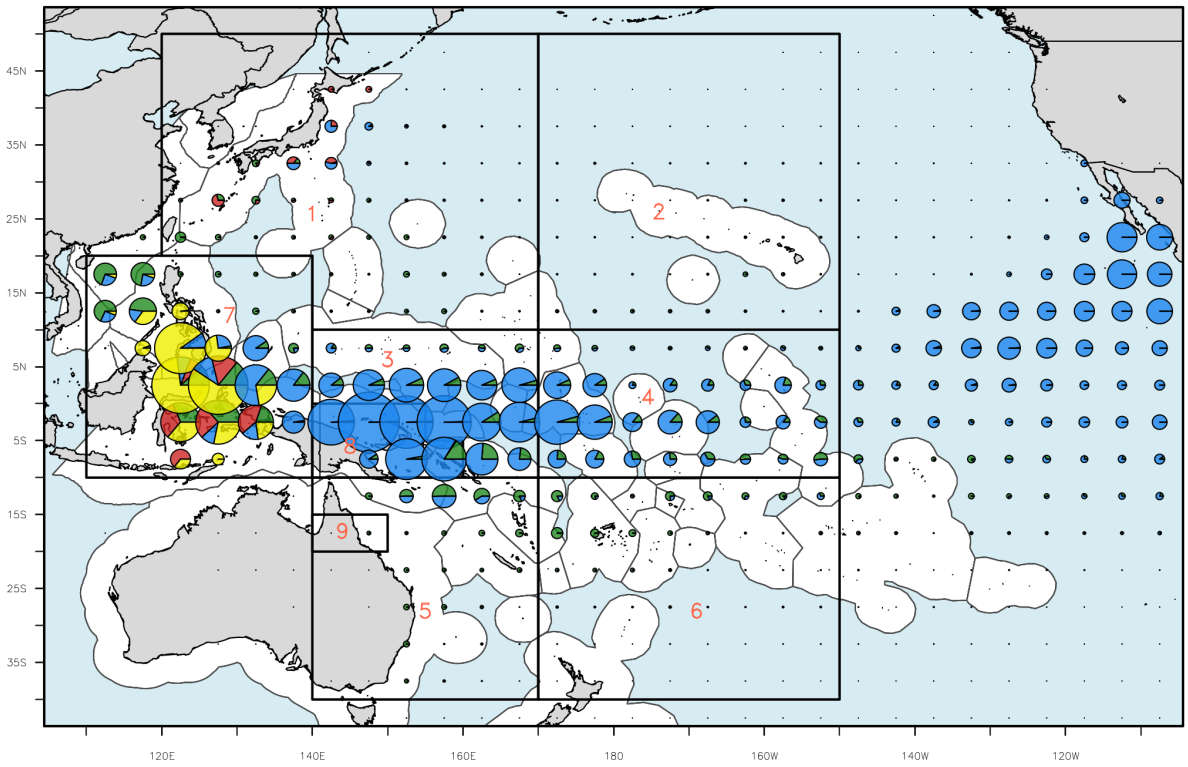

Figure 6: Distribution and magnitude of yellowfin tuna catches for the most recent decade of the stock assessment (2006-2015) by 5° square and fishing gear: longline (green), pole-and-line (red), purse seine (blue) and miscellaneous (yellow), for the WCPO and part of the EPO. Overlayed are the regional boundaries for the stock assessment (2017 regional structure).

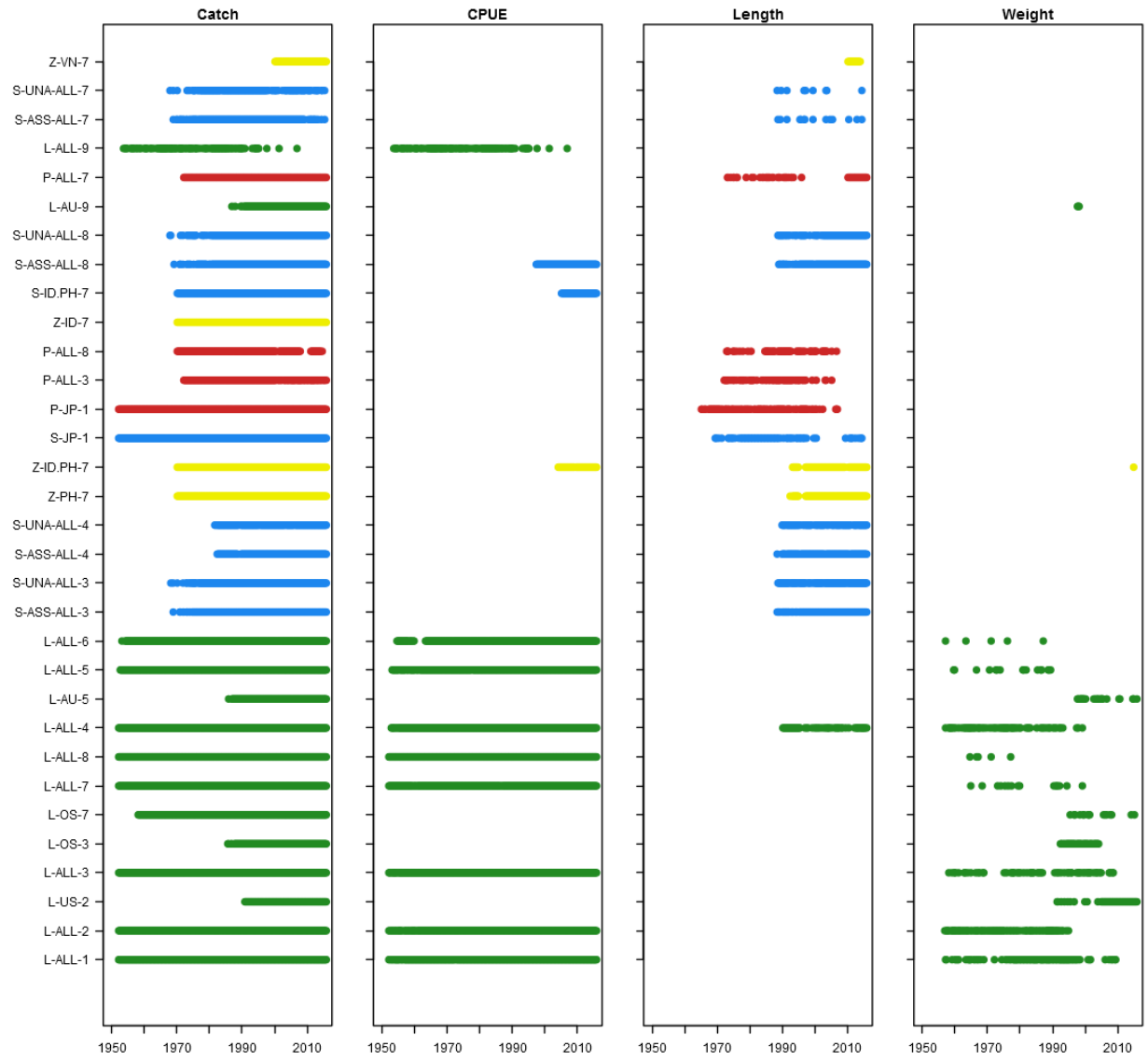

Figure 7: Presence of catch, standardised CPUE, length frequency and weight frequency data by year and fishery for the diagnostic case model (2017 regional structure, hence 32 fisheries). The different colours denote gear-type of the fishery: longline (green); pole-and-line (red); purse seine (blue); and miscellaneous (yellow).

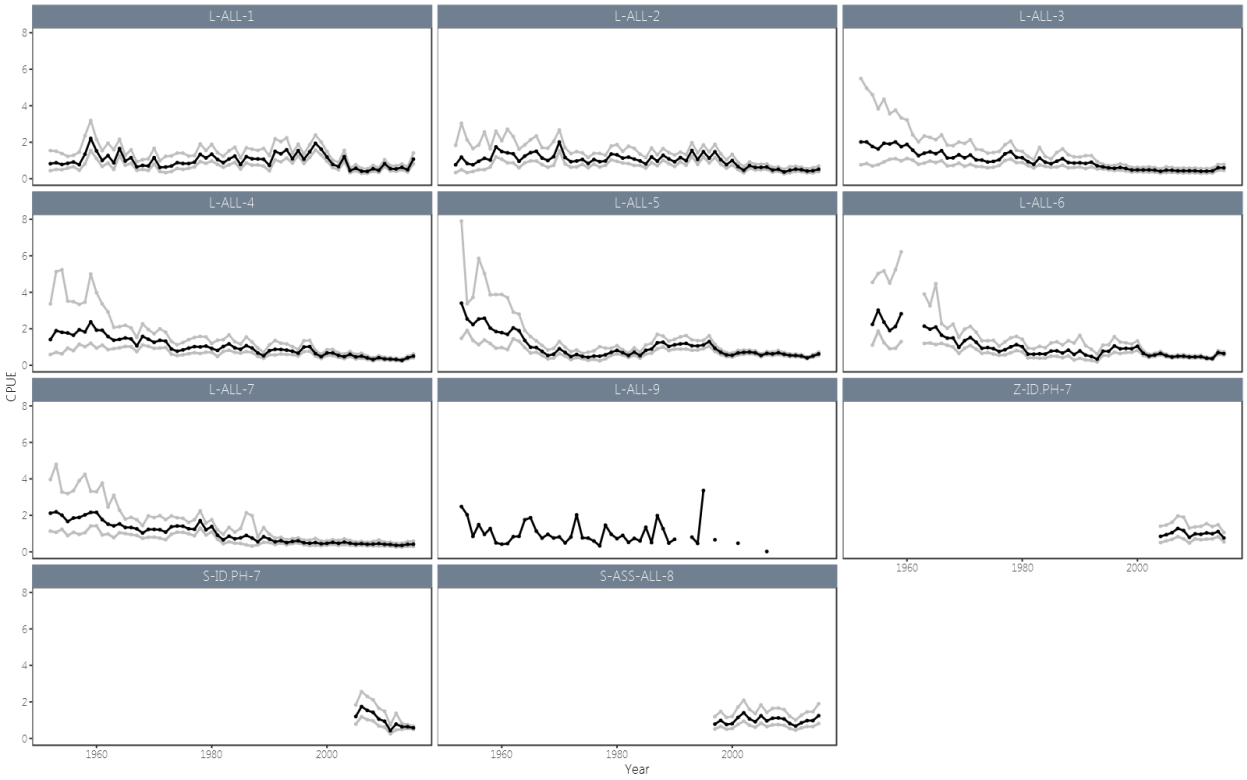

Figure 8: Standardised catch-per-unit-effort (CPUE) indices for the longline fisheries in regions 1–9 used in the diagnostic case model. See McKechnie et al. (2017a) for further details of the estimation of these CPUE indices. The light grey lines represent the 95% confidence intervals derived from the effort deviation penalties used in the diagnostic case model.

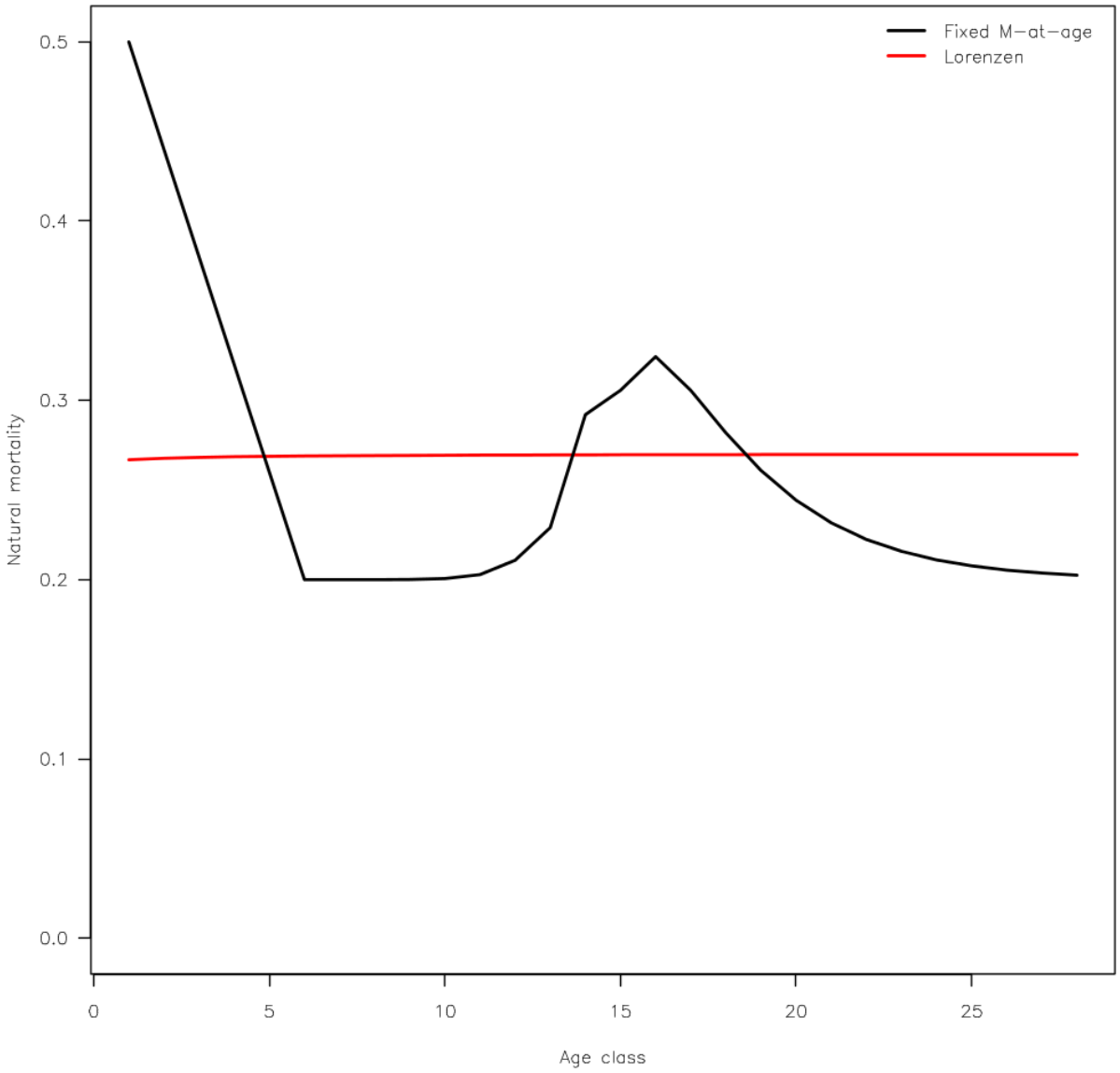

Figure 9: Quarterly natural mortality-at-age as used the diagnostic case model (black line). The red line shows the estimated m-at-age function for the sensitivity model assuming a Lorenzen-type function (*Lorenzen*) between mortality and fish size (presented in Section Section 5.2).

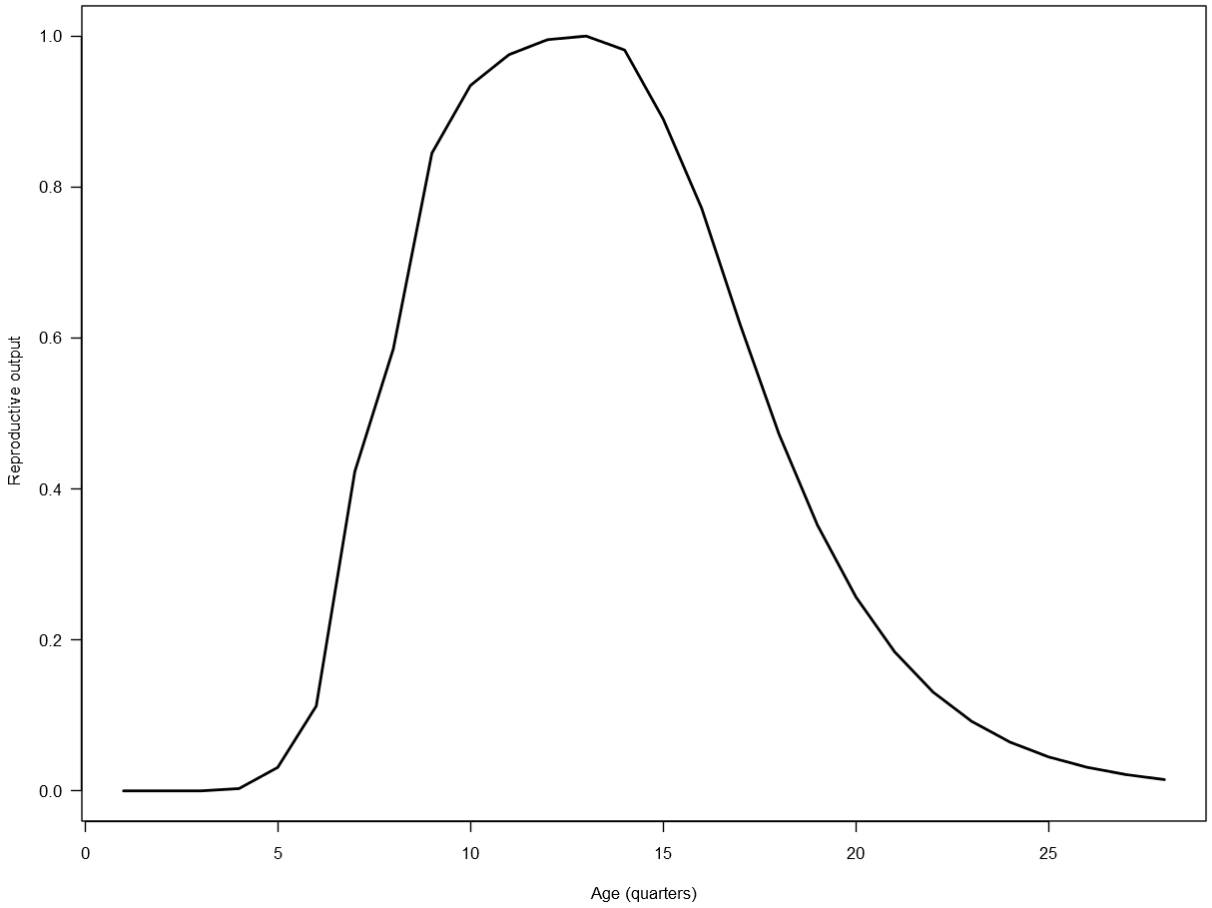

Figure 10: Maturity-at-age as used in the diagnostic case model.

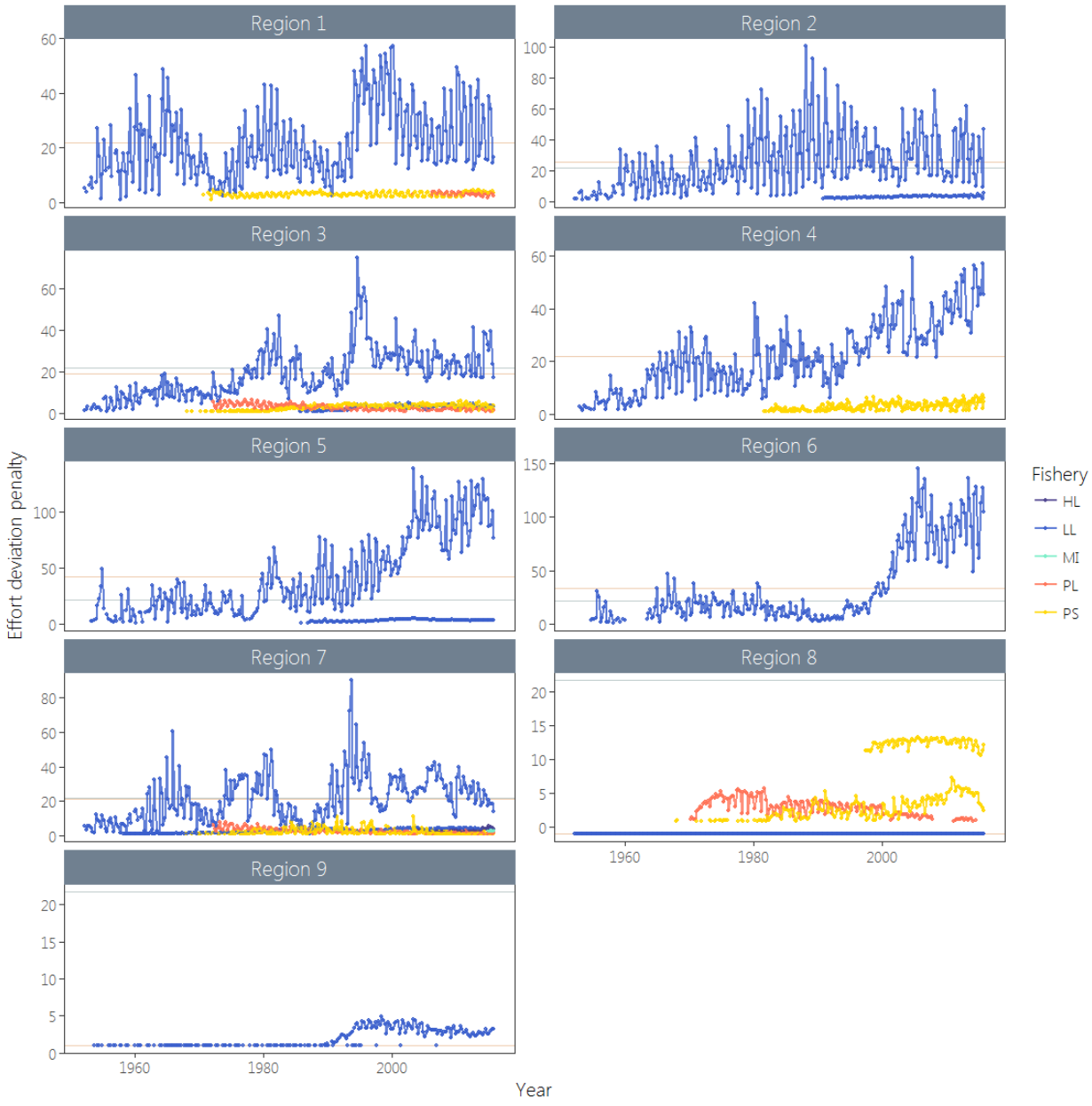

Figure 11: Plot of the effort deviation penalties applied to each fishery, by region, with the colours of the lines representing the gear type. In several cases there is more than one fishery for a given gear-type in a region (e.g. regions 7 and 8). A higher penalty gives more weight to the CPUE of that fishery and so the high weightings applied to the standardised CPUE indices are evident.

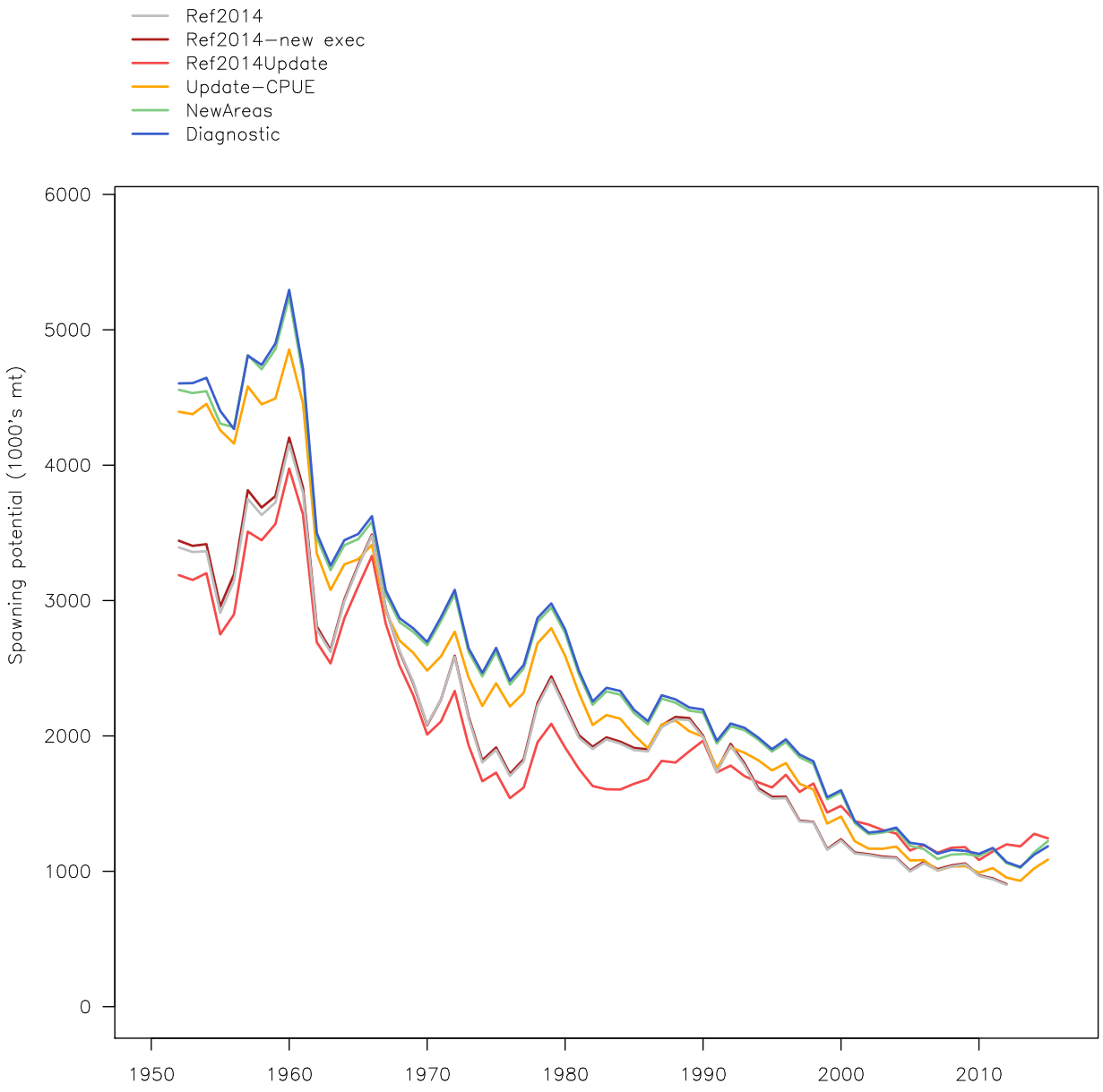

Figure 12: Stepwise changes in spawning potential from the 2014 reference case model through to the 2017 diagnostic case model.

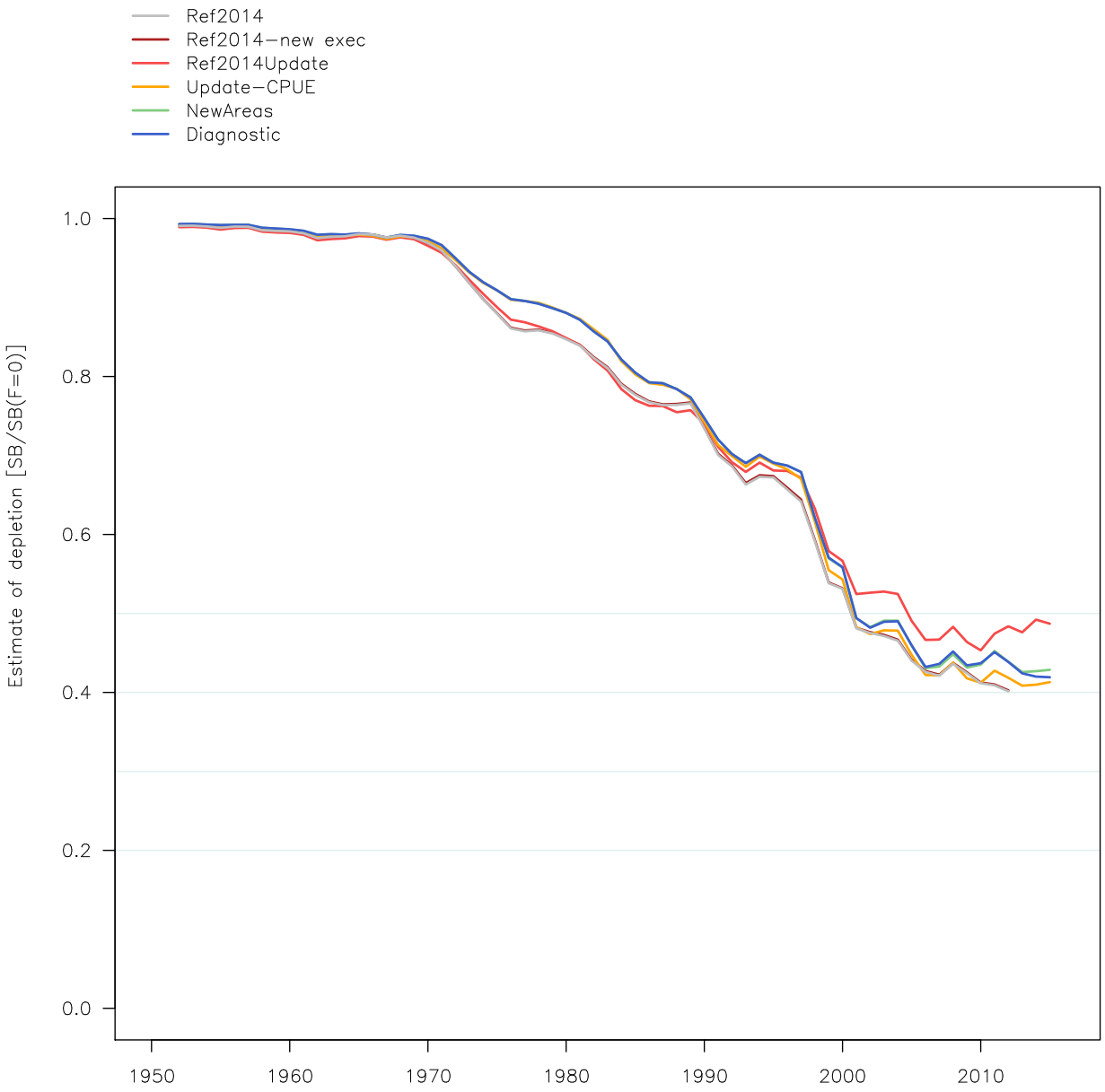

Figure 13: Stepwise changes in fishing depletion from the 2014 reference case model through to the 2017 diagnostic case model.

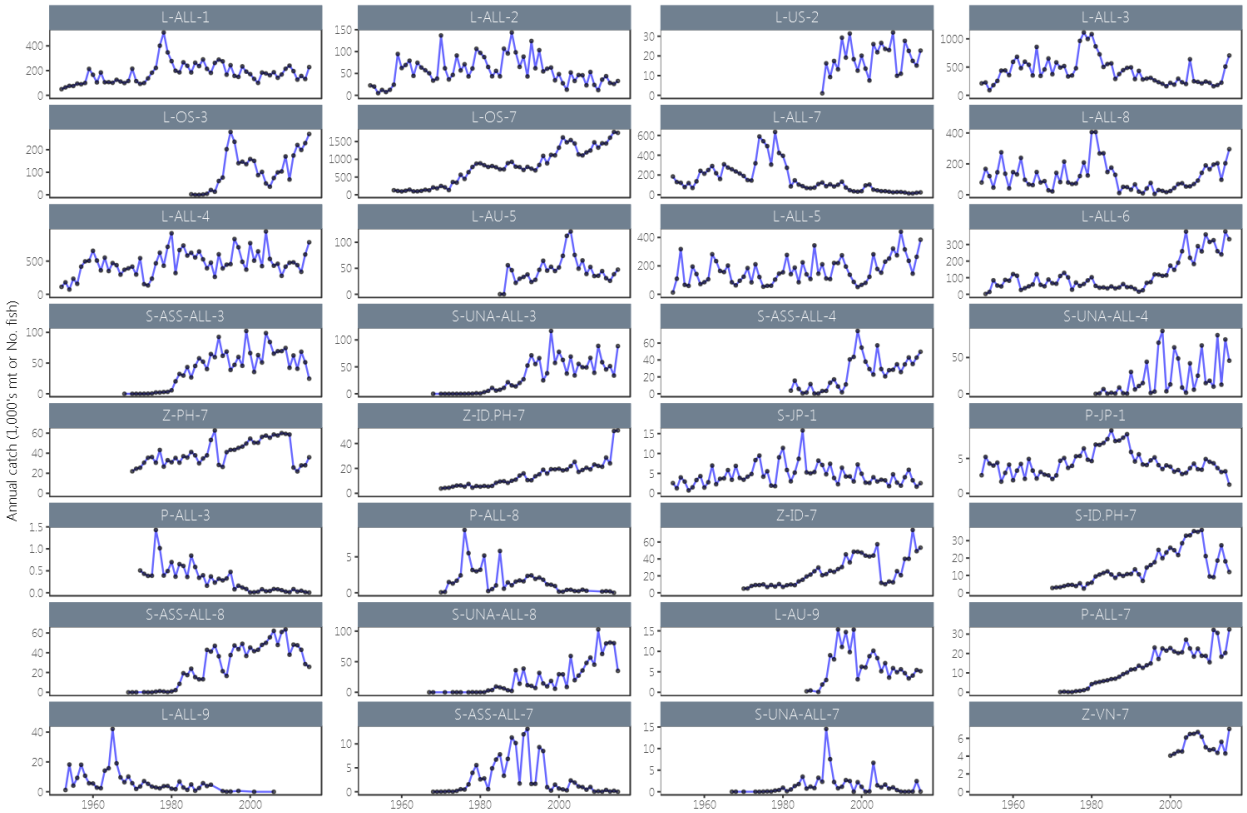

Figure 14: Observed (black points) and model-predicted (blue lines) catch for the 32 fisheries in the diagnostic case model. The y-axis is in catch-in-numbers for the longline fisheries and catch-in-weight for the other fisheries, both divided by 1,000.

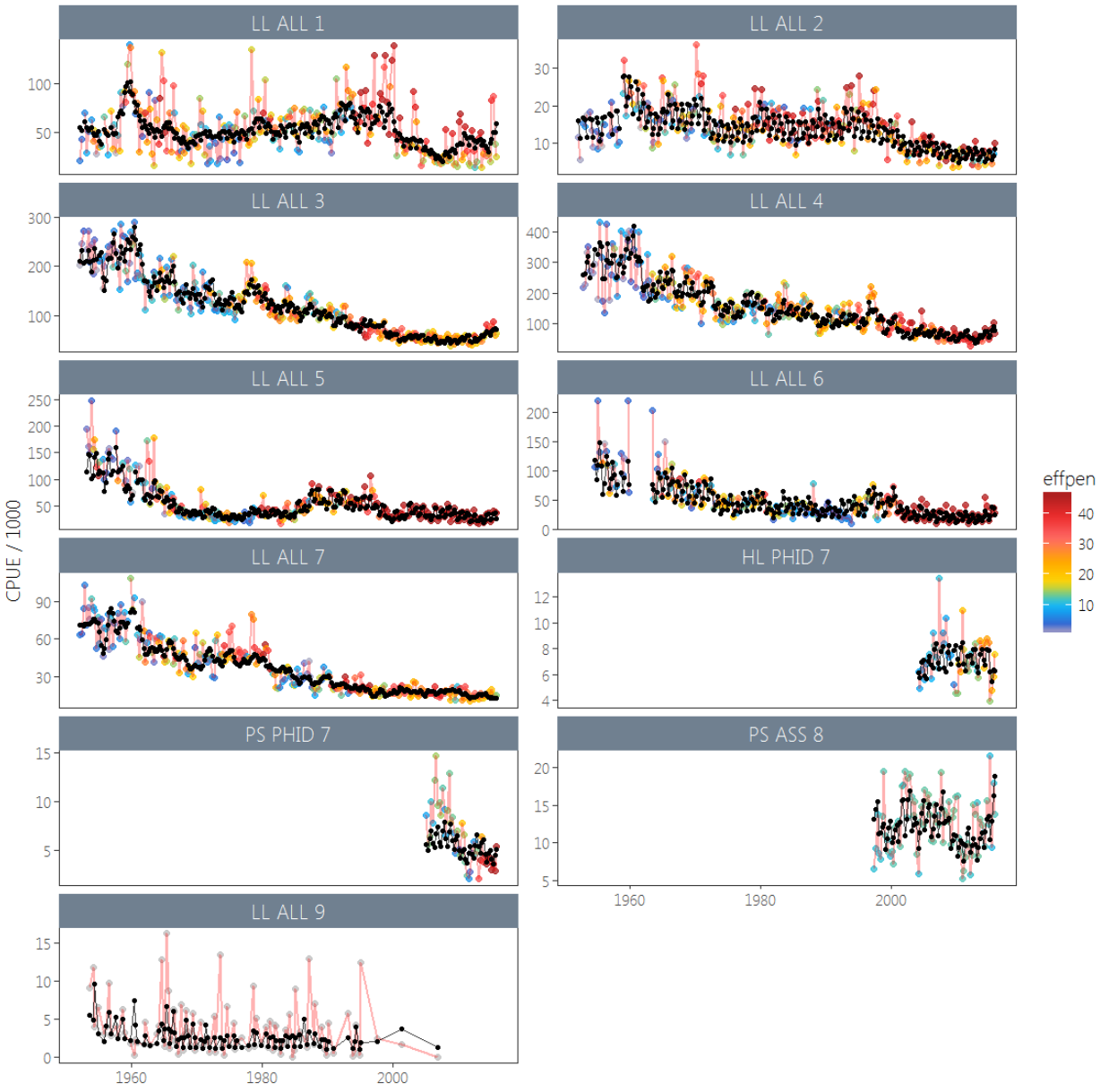

Figure 15: Observed (coloured points and red lines) and model-predicted (black points and lines) CPUE for the 10 fisheries which received standardised CPUE indices in the diagnostic case model, and the nominal CPUE index used for region 9. Observed points are coloured as a function of the penalty weight applied on them.

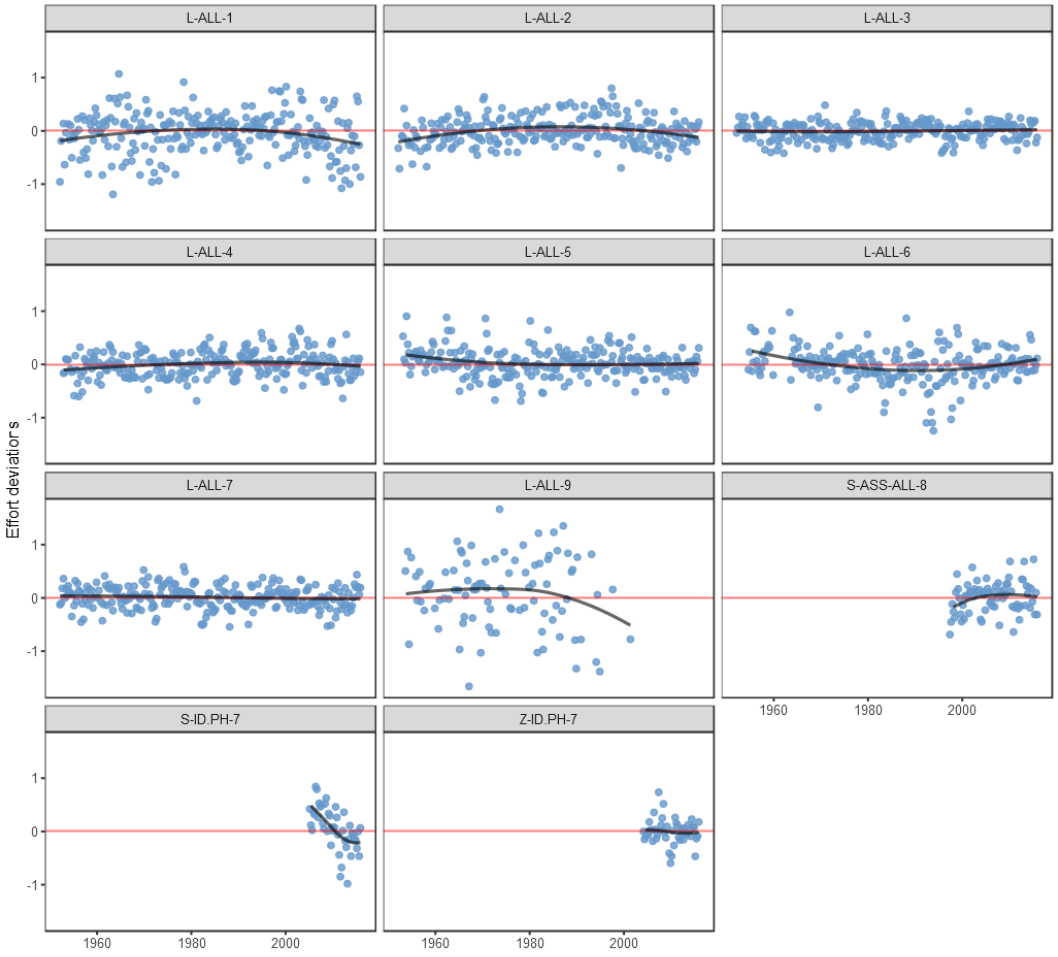

Figure 16: Effort deviations by time period for each of the fisheries receiving standardised CPUE indices in the diagnostic case model. The dark line represents a lowess smoothed fit to the effort deviations.

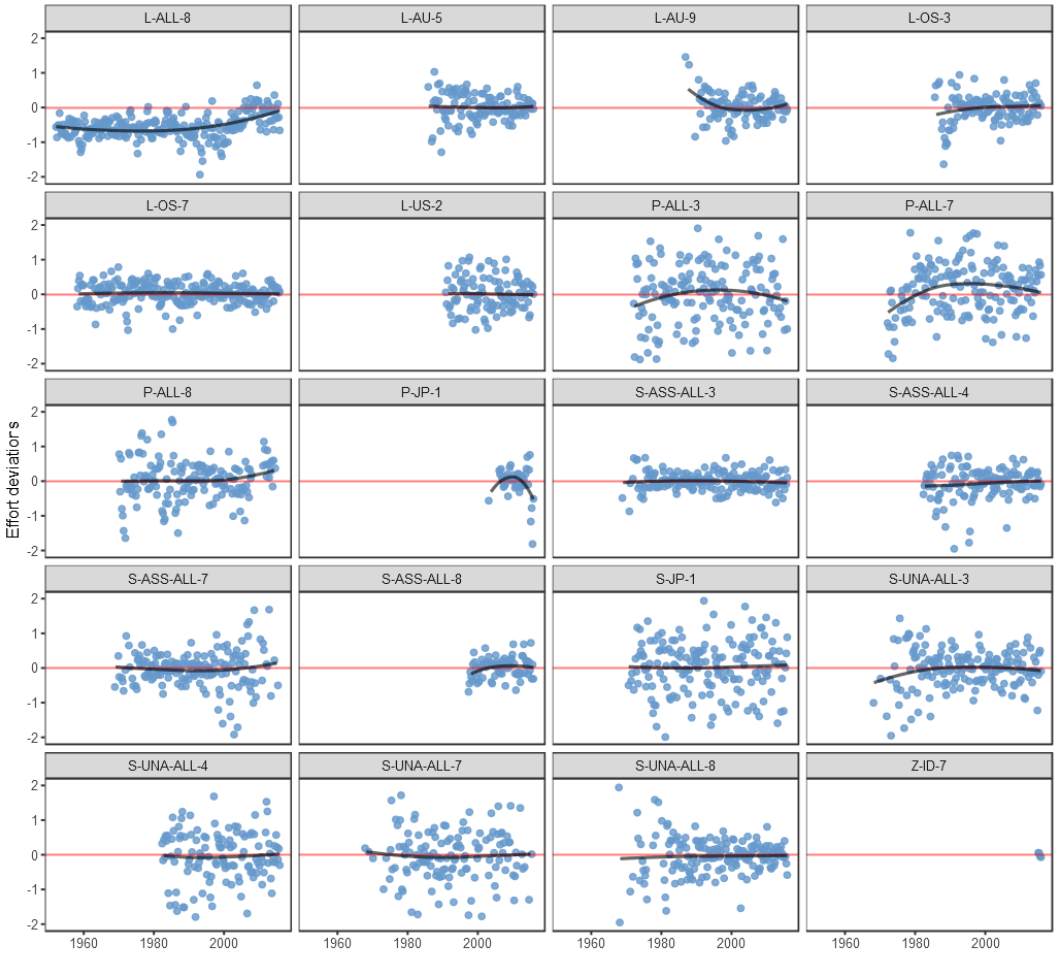

Figure 17: Effort deviations by time period for each of the fisheries that did not receive standardised CPUE indices in the diagnostic case model. The dark line represents a lowess smoothed fit to the effort deviations.

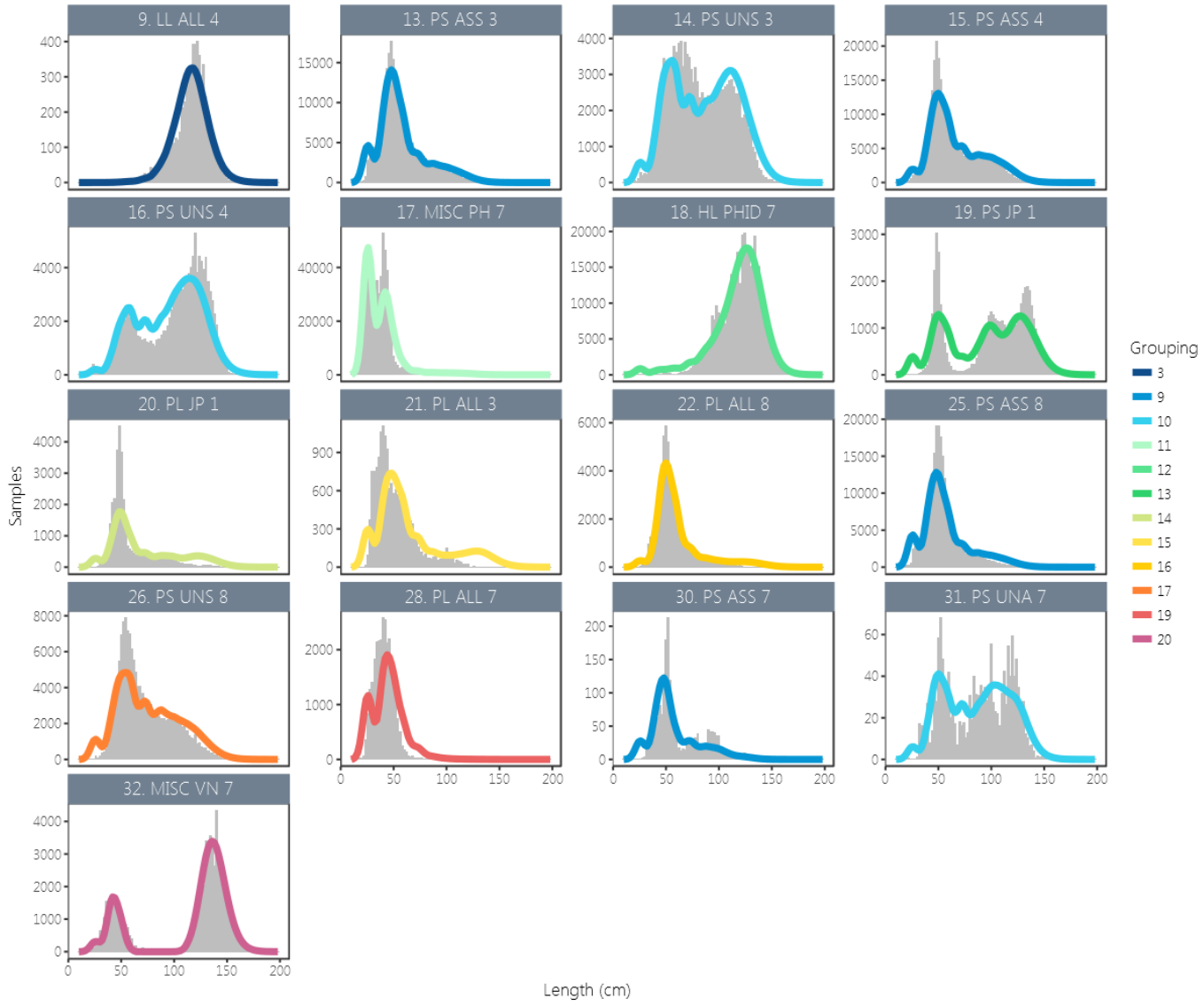

Figure 18: Composite (all time periods combined) observed (grey histograms) and predicted (lines, coloured by fishery grouping) catch-at-length for all fisheries with samples for the diagnostic case model.

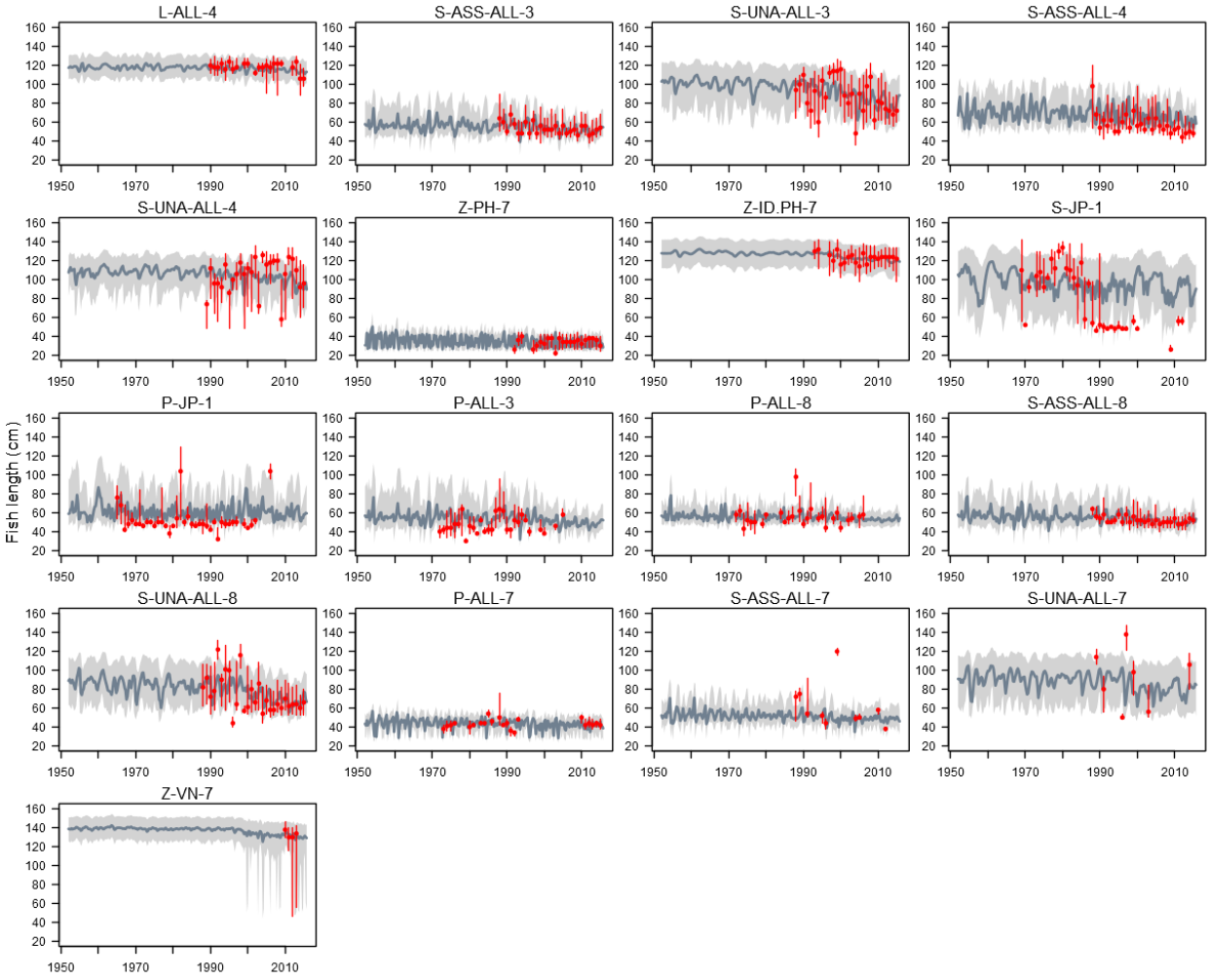

Figure 19: A comparison of the observed (red points) and predicted (grey line) median fish length (FL, cm) for all fisheries with samples for the diagnostic case model. The uncertainty intervals (grey shading) represent the values encompassed by the 25% and 75% quantiles. Sampling data are aggregated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

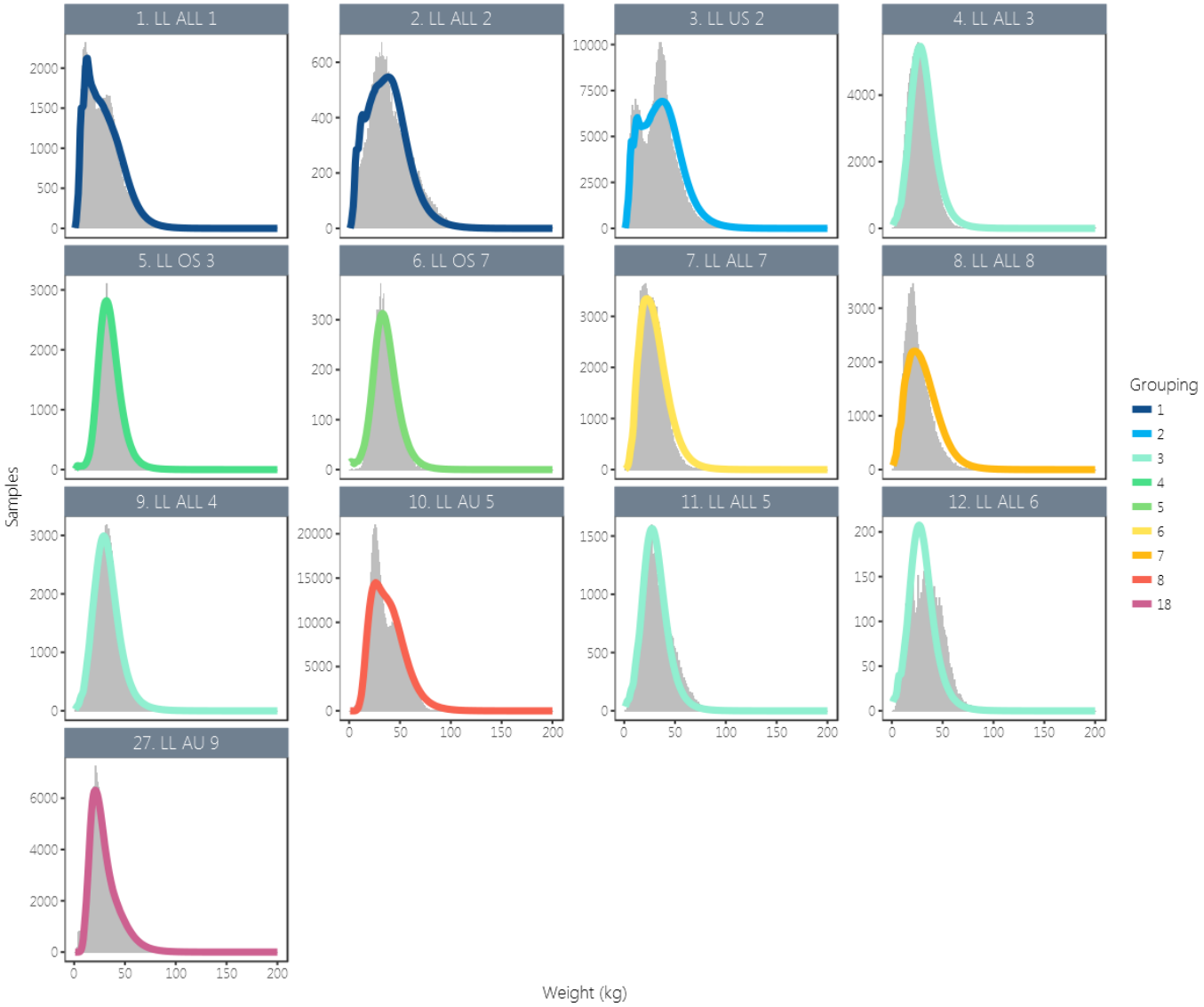

Figure 20: Composite (all time periods combined) observed (grey histograms) and predicted (line, coloured by fishery grouping) catch-at-weight for all fisheries with samples for the diagnostic case model.

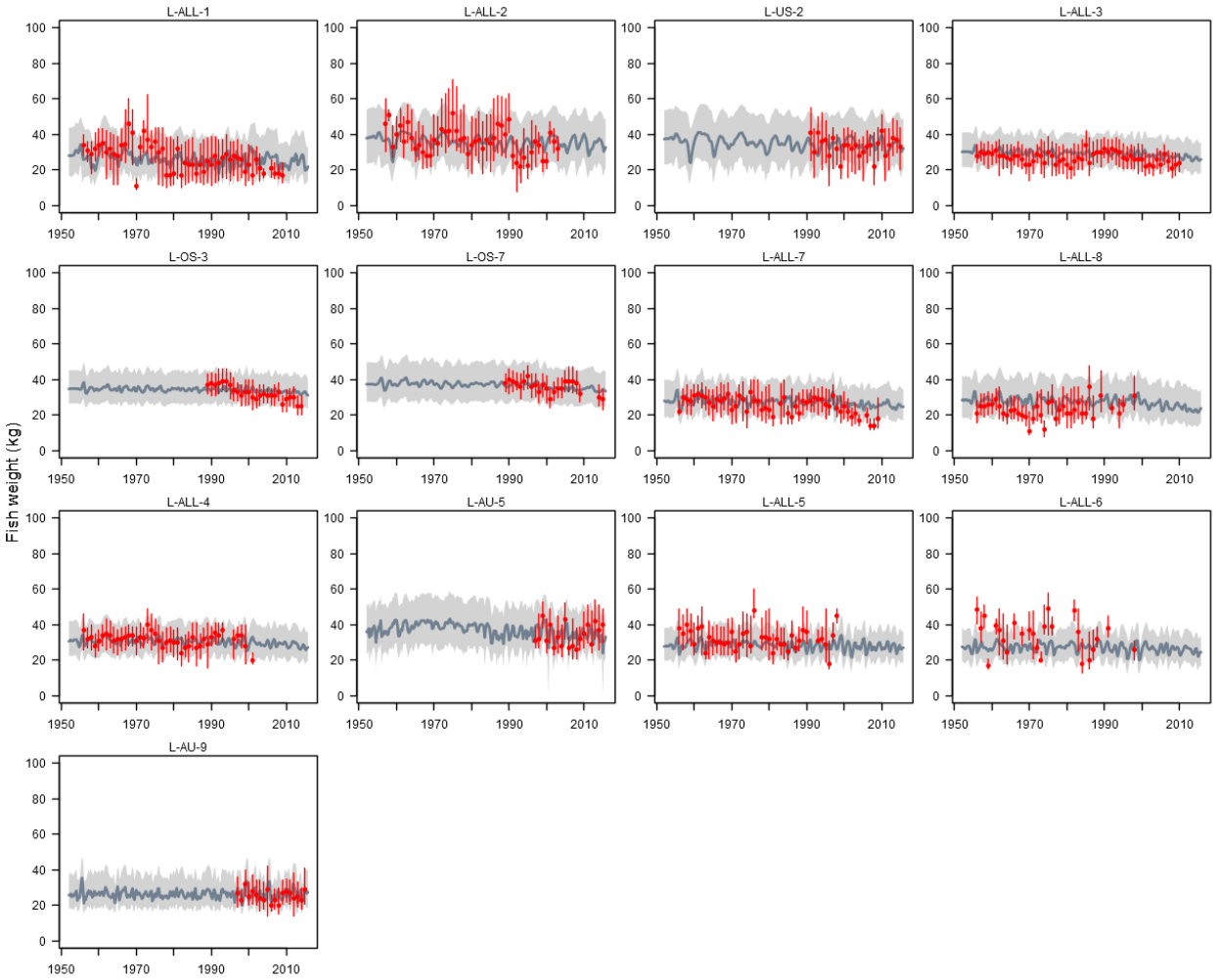

Figure 21: A comparison of the observed (red points) and predicted (grey line) median fish weight (kg) for all fisheries with samples for the diagnostic case model. The uncertainty intervals (grey shading) represent the values encompassed by the 25% and 75% quantiles. Sampling data are aggregated by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

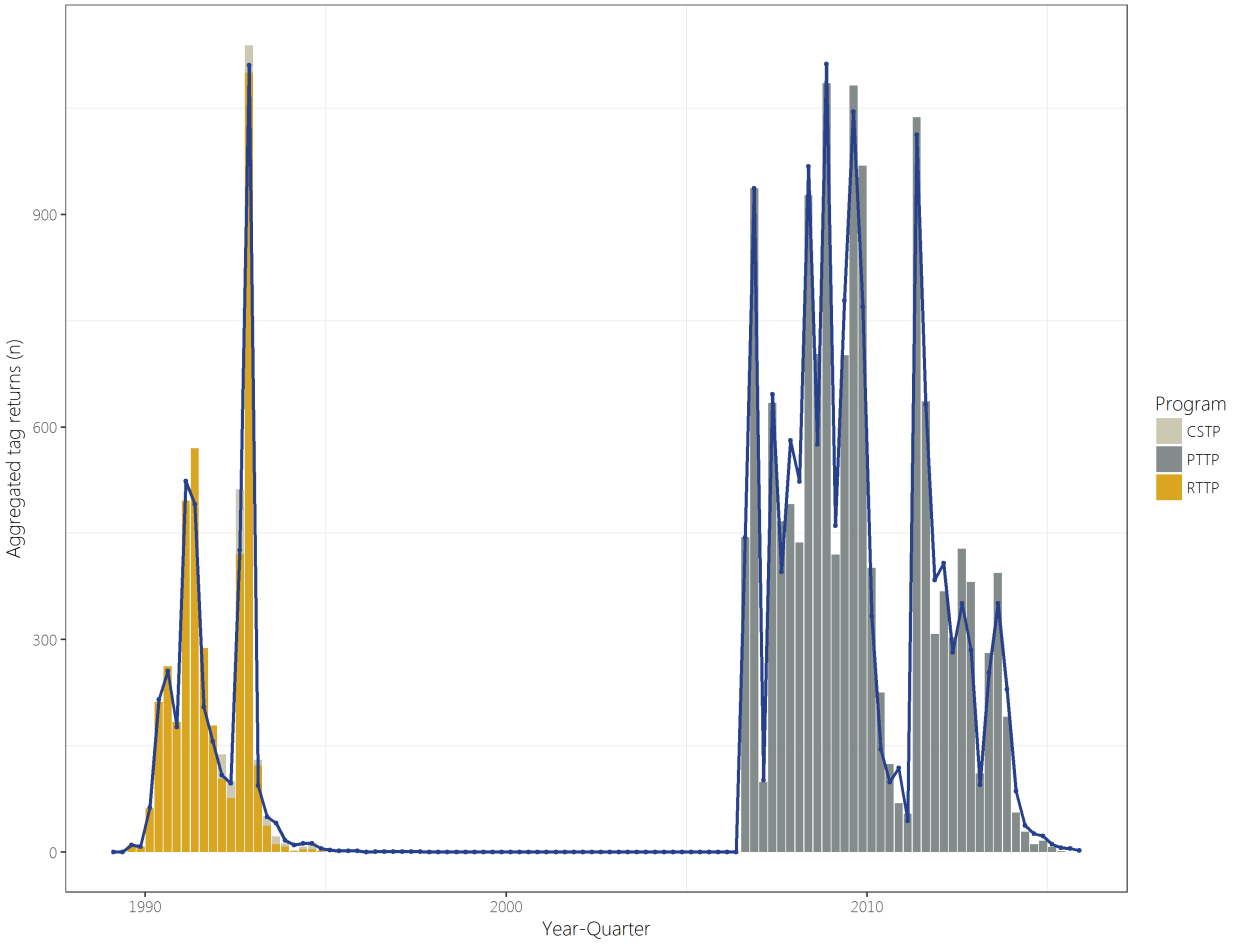

Figure 22: Observed (coloured bars) and model-predicted (blue line) tag returns over time for the diagnostic case model across all tag release events with all tag recapture groupings aggregated. The colour of the bars denotes the tagging programme from which the recaptured fish were released.

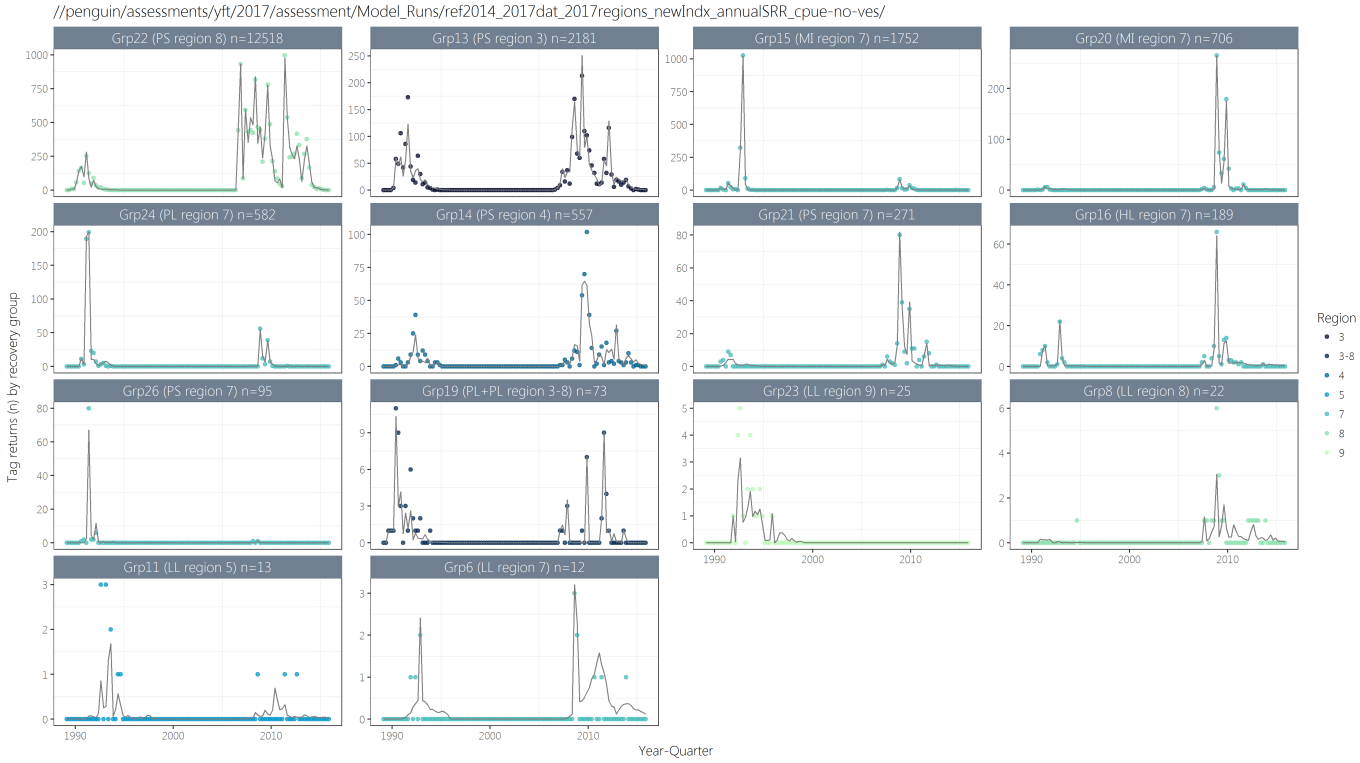

Figure 23: Observed (points) and model-predicted (black line) tag returns over time for the diagnostic case model by recapture group for groups with at least 10 recaptures. Groups with extremely low numbers of recaptures are uninformative and not shown. The colours of the points denote the region where the recapture groups are located.

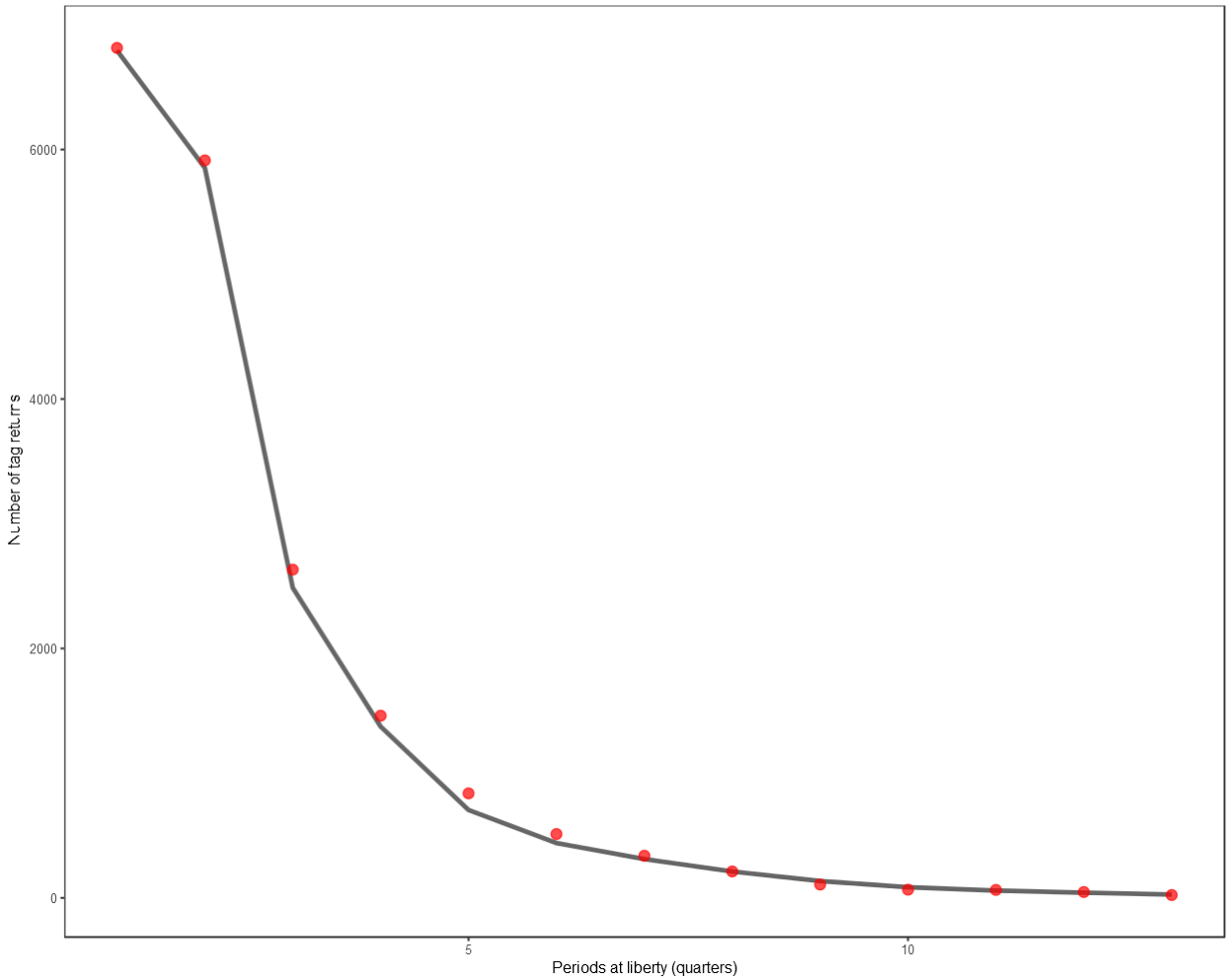

Figure 24: Observed and model-predicted tag attrition across all tag release events for the diagnostic case model.

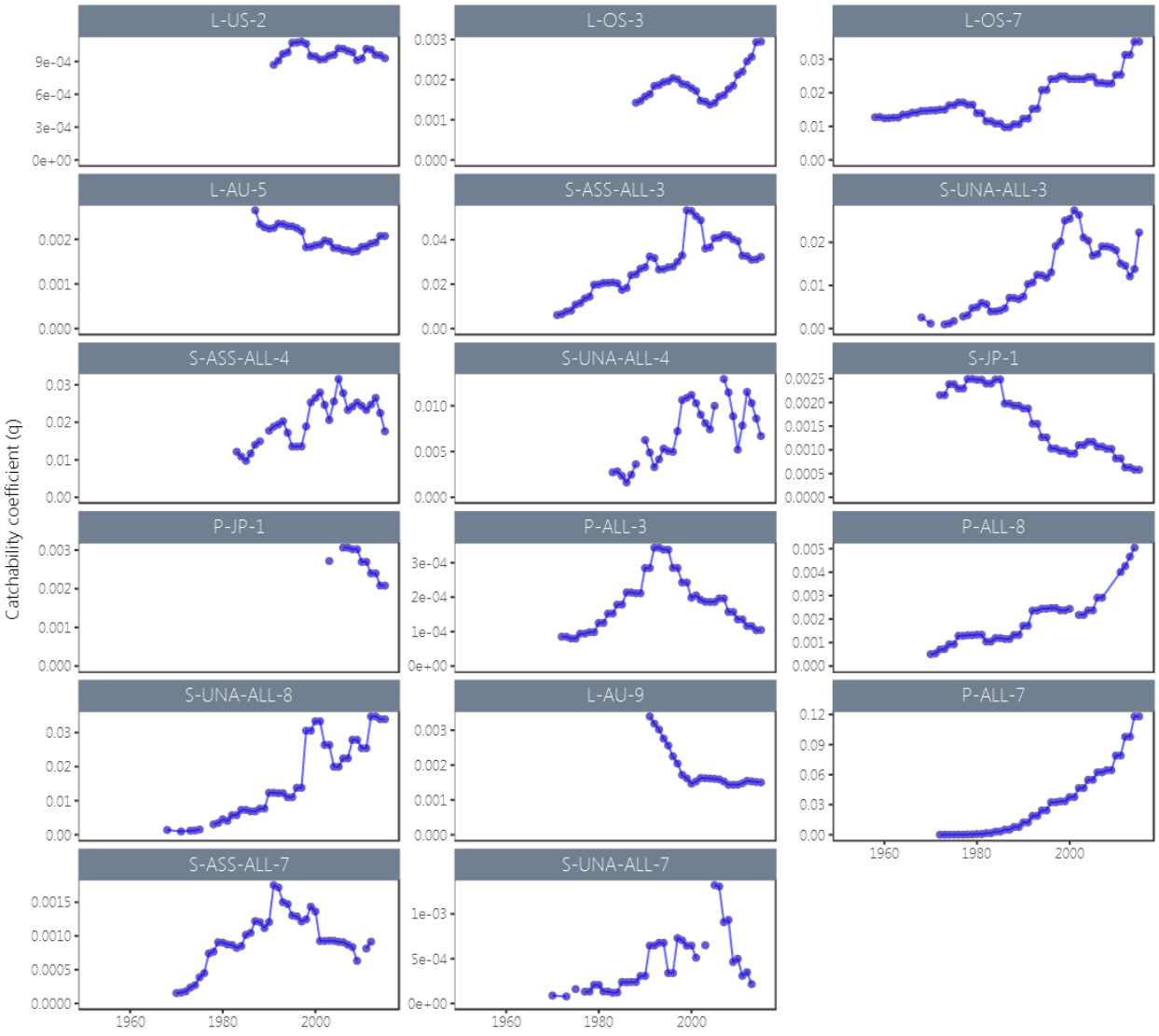

Figure 25: Estimated time series of catchability for those fisheries assumed to have random walk in these parameters. Values shown are the annual means which removes seasonal variability.

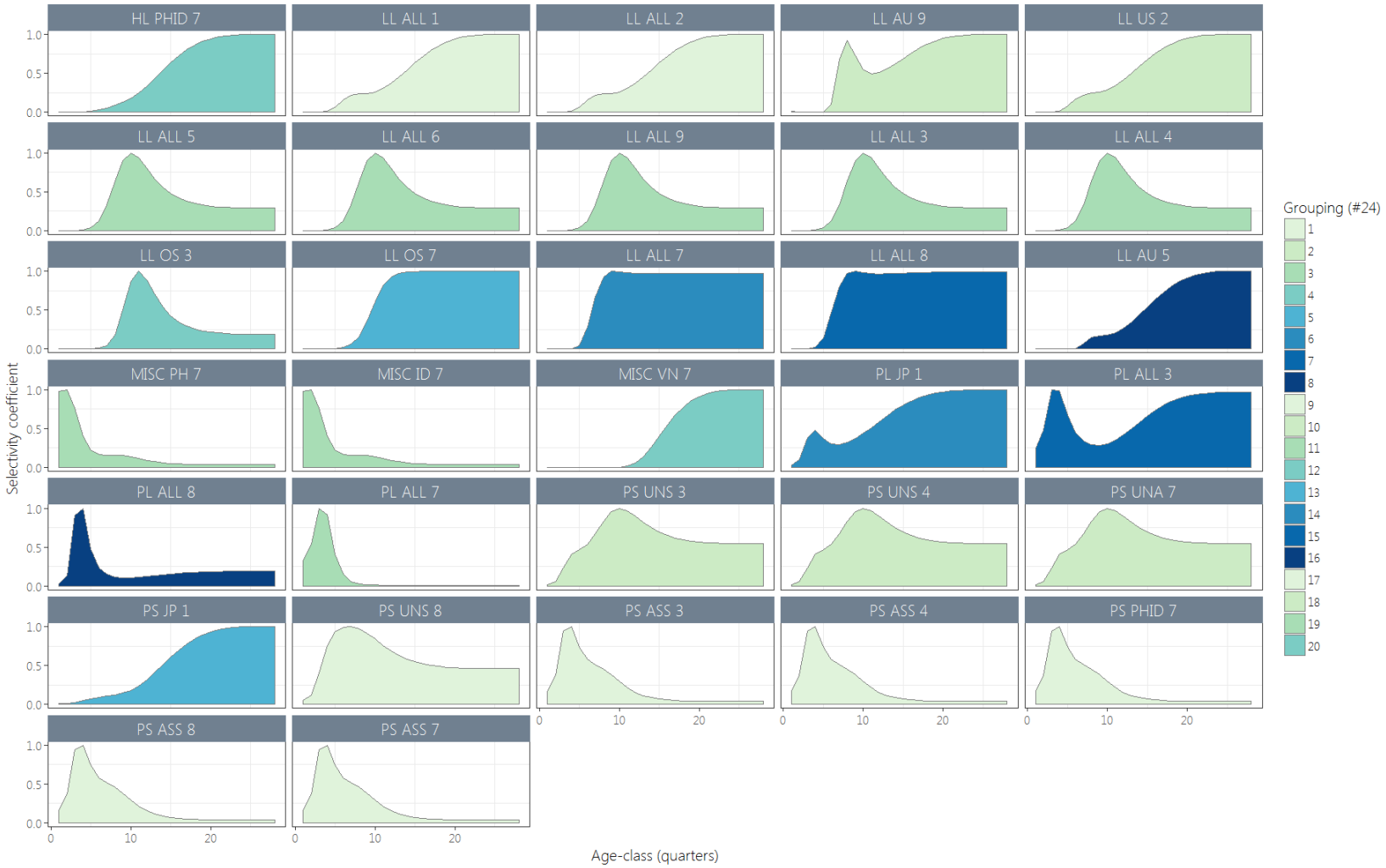

Figure 26: Estimated age-specific selectivity coefficients by fishery for the diagnostic case model. Colours reflect fishery groupings.

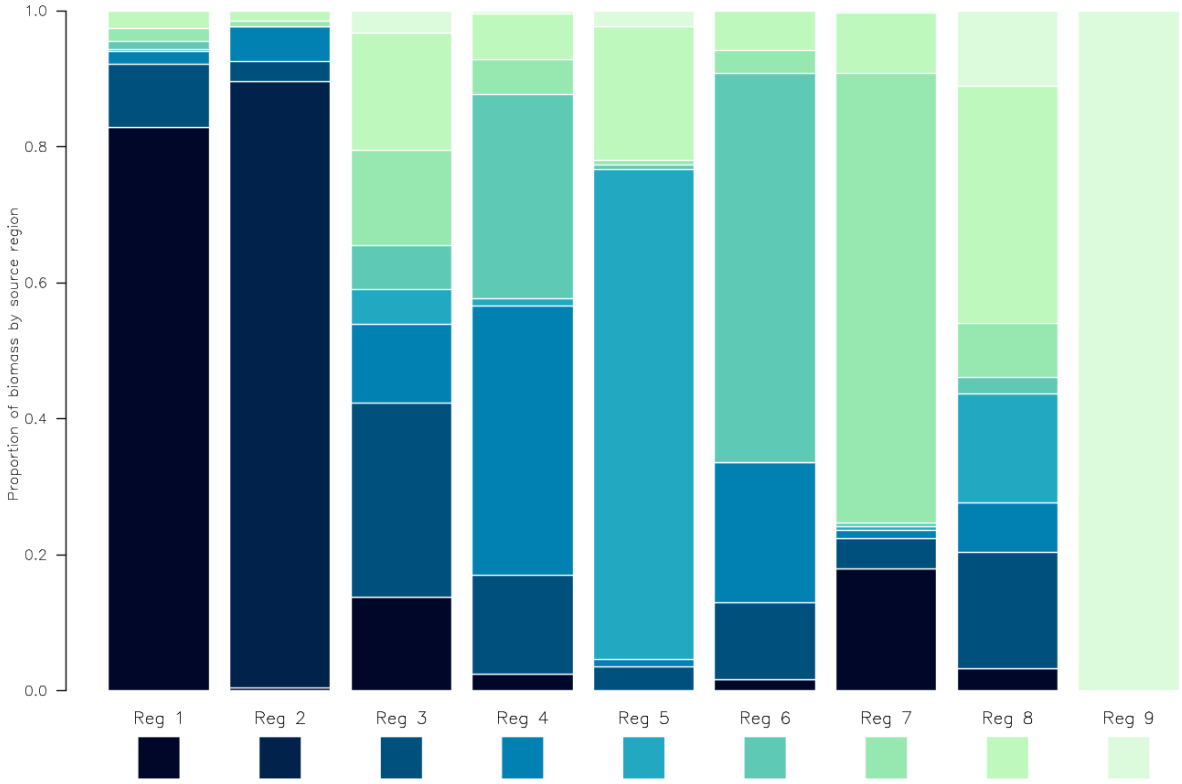

Figure 27: Proportional distribution of total biomass (by weight) in each region apportioned by the source region of the fish, for the diagnostic case model. The colour of the home region is presented below the corresponding label on the x-axis. The biomass distributions are calculated based on the long-term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

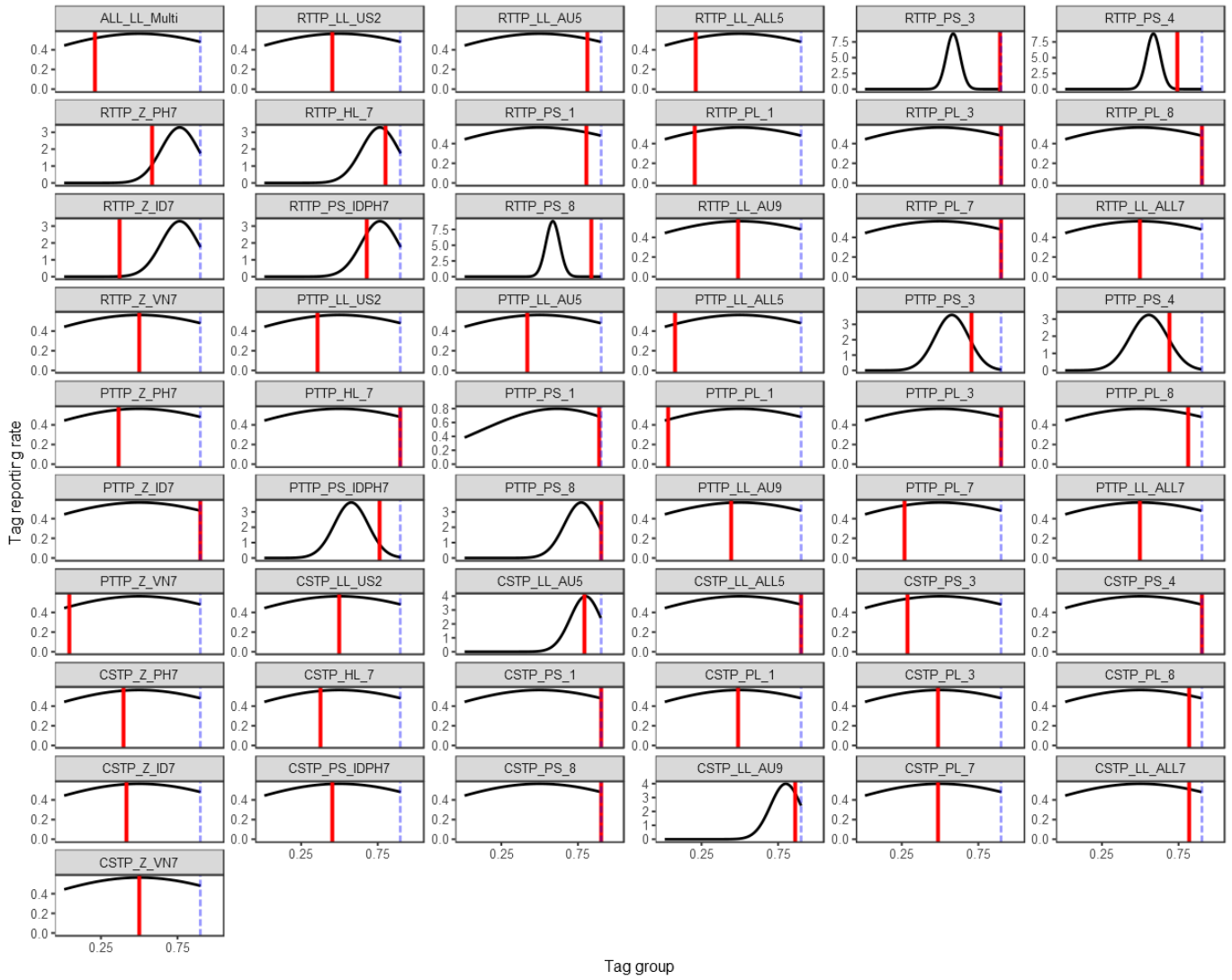

Figure 28: Estimated reporting rates for the diagnostic case model (red lines) and the prior distribution (black lines) for each reporting rate group. The imposed upper bound (0.9) on the reporting rate parameters are show as a blue dashed line. Reporting rates can be estimated separately for each release program and recapture fishery group but in practice are aggregated over some recapture groups to reduce dimensionality.

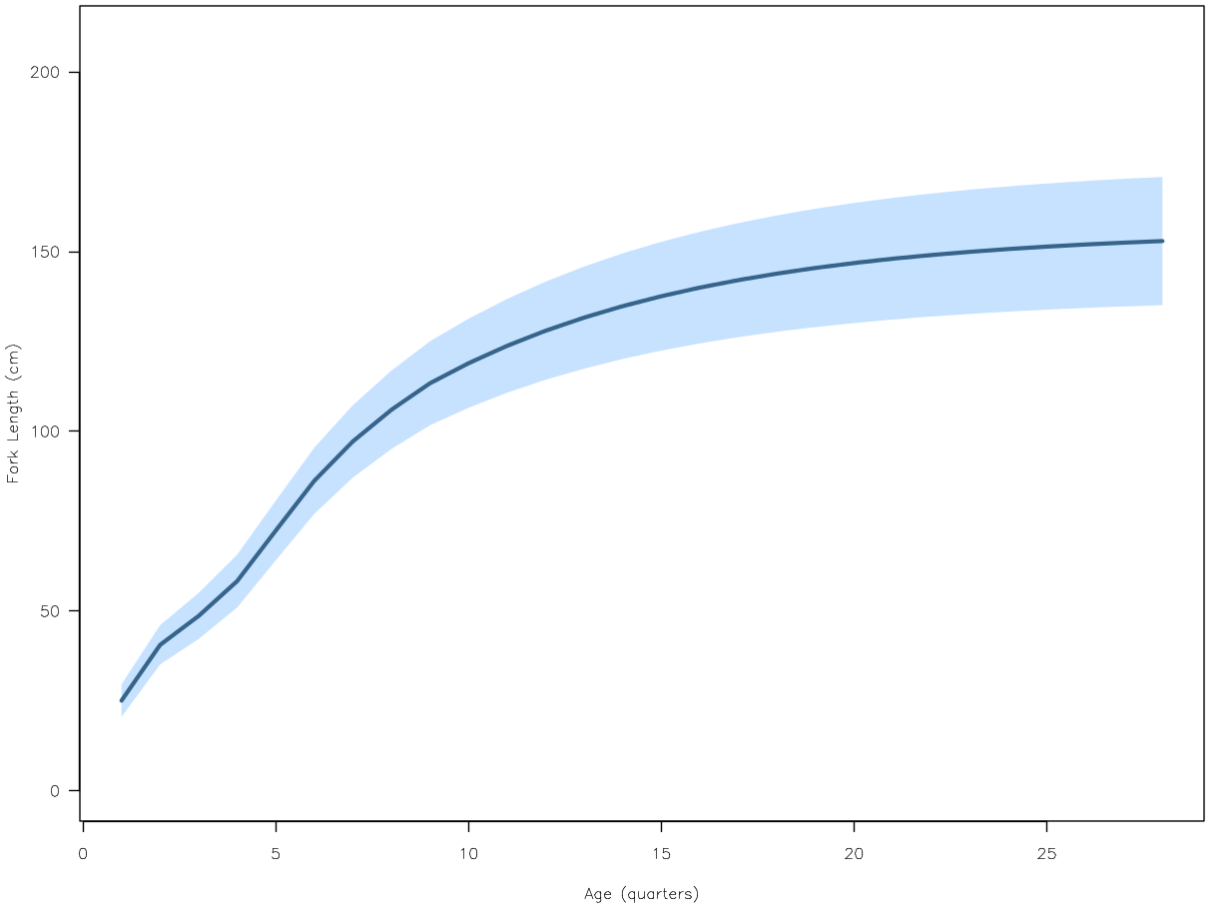

Figure 29: Estimated growth for the diagnostic case model. The blue line represents the estimated mean fork length (cm) at-age and the blue region represents the length-at-age within one standard deviation of the mean, for the diagnostic case model.

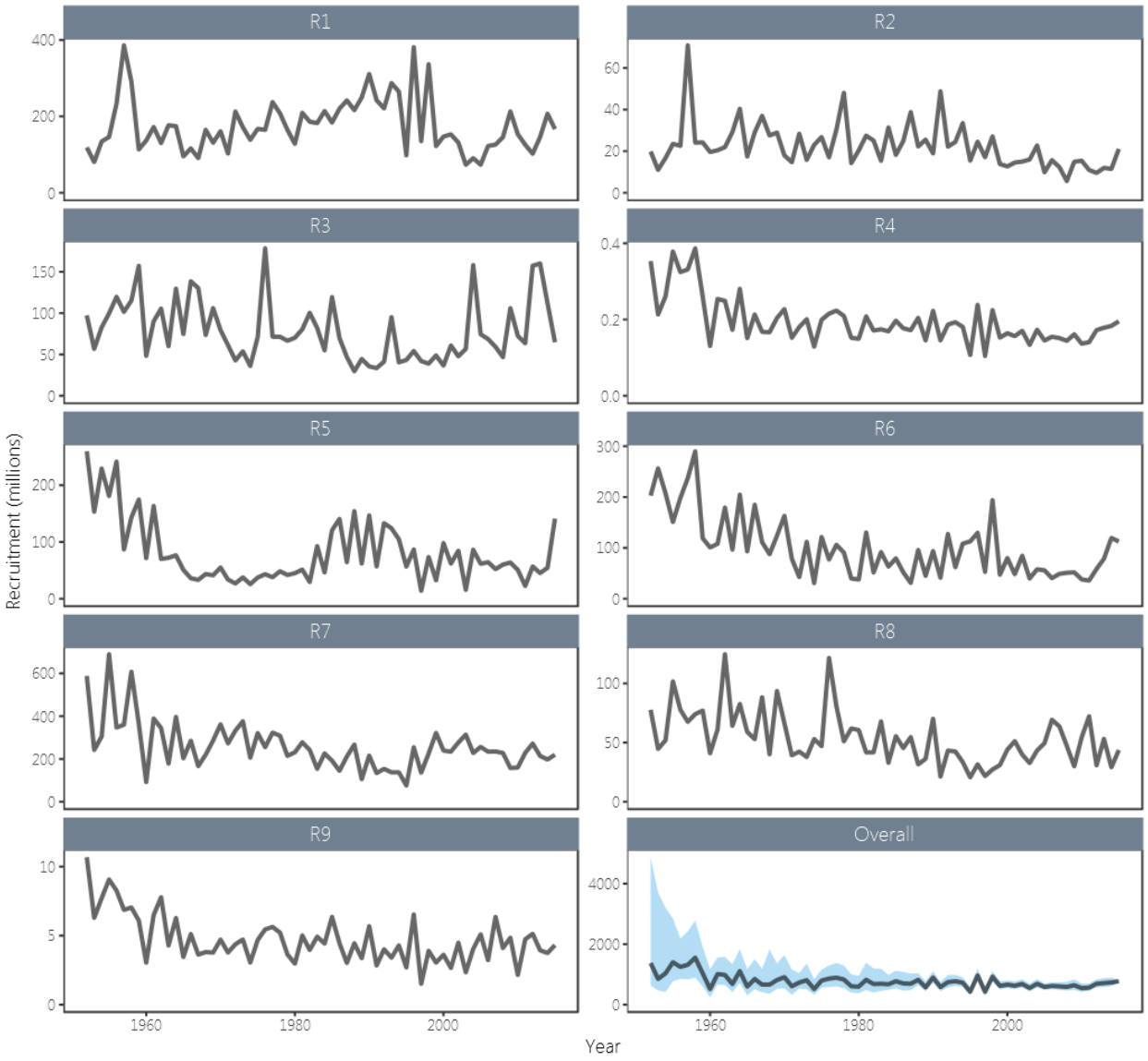

Figure 30: Estimated annual, temporal recruitment by model region for the diagnostic case model. The figure for the overall recruitments shows the estimated 95% confidence intervals as the blue shaded regions. Because of the wide uncertainty in the early years, the y-axis scale makes it difficult to interpret trends. Note that the scale of the y-axis is not constant across regions.

Figure 31: Estimated annual average spawning potential (a), total biomass (b) and recruitment (c) by model region for the diagnostic case model, showing the relative sizes among regions. Equatorial regions are in warm colours.

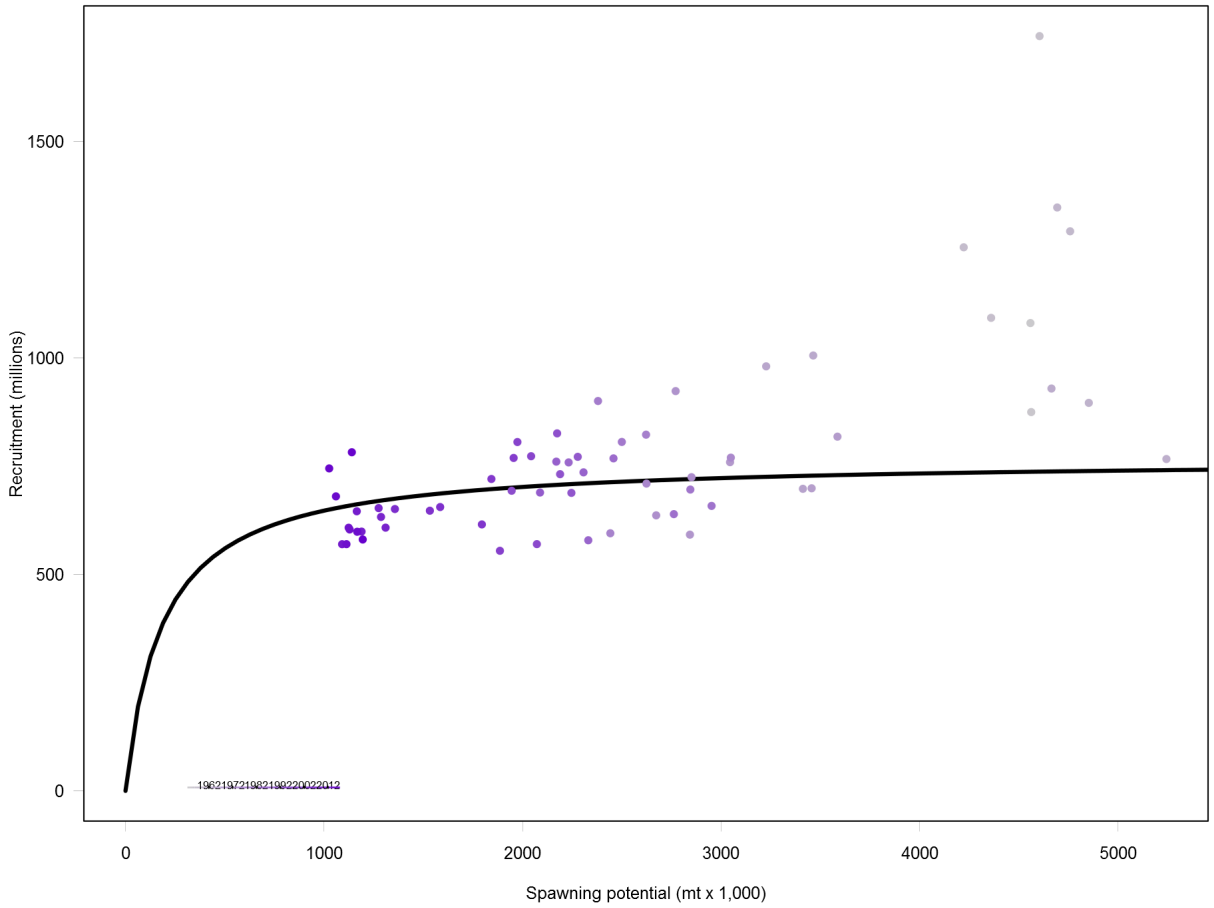

Figure 32: Estimated relationship between recruitment and spawning potential based on annual values for the diagnostic case model. The darkness of the circles changes from light to dark through time.

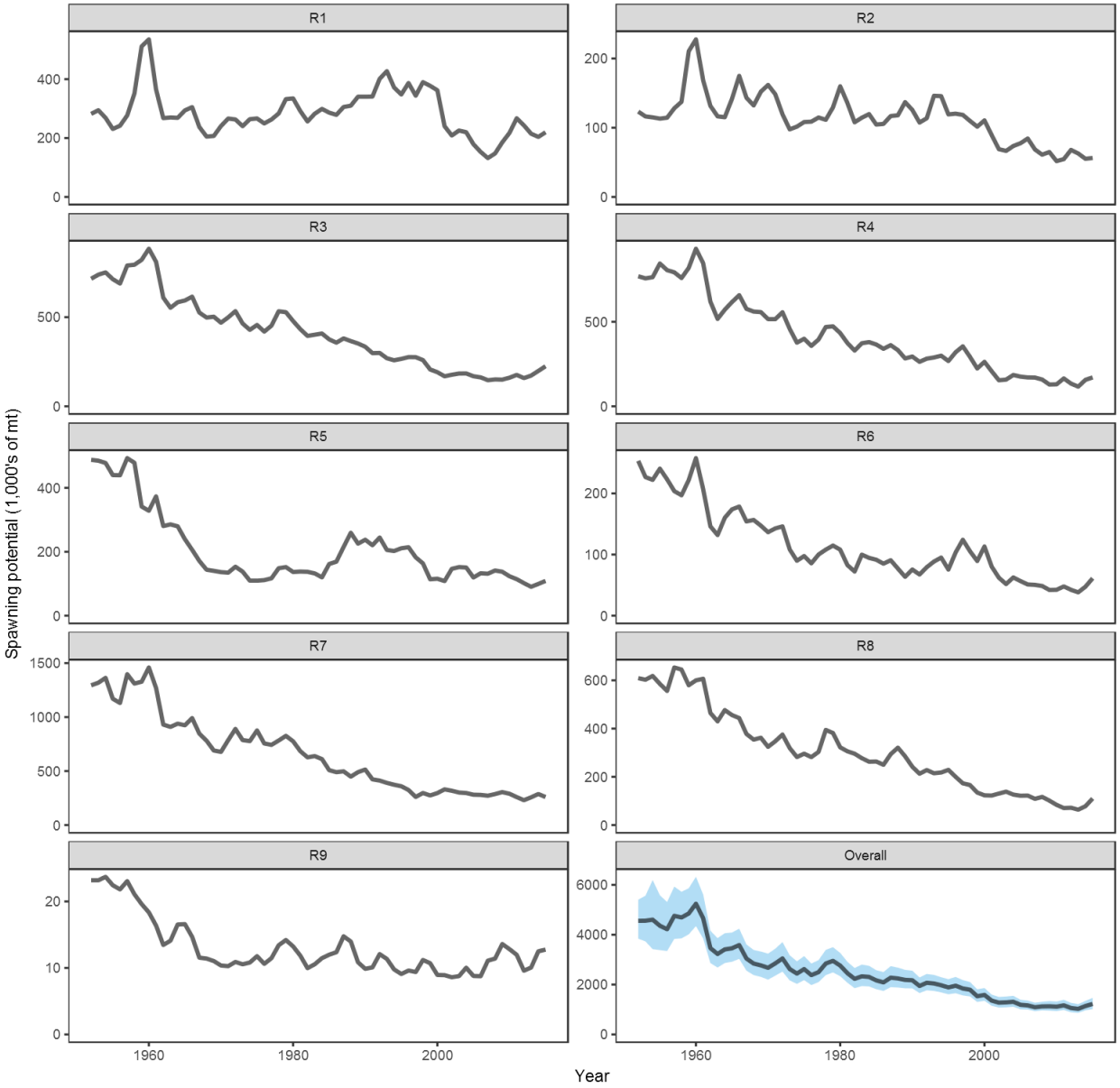

Figure 33: Estimated temporal spawning potential by model region for the diagnostic case model. The figure for the overall recruitments shows the estimated 95% confidence intervals as the blue shaded regions. Note that the scale of the y-axis is not constant across regions.

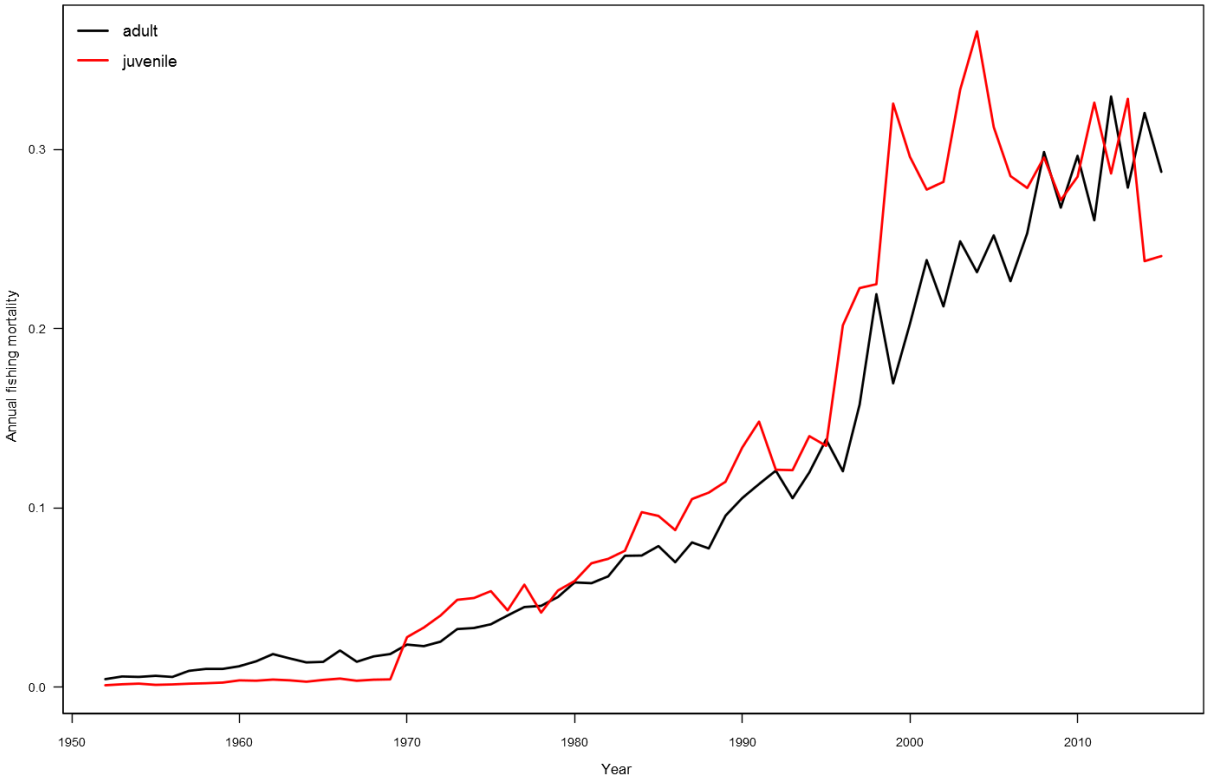

Figure 34: Estimated annual average juvenile and adult fishing mortality for the diagnostic case model.

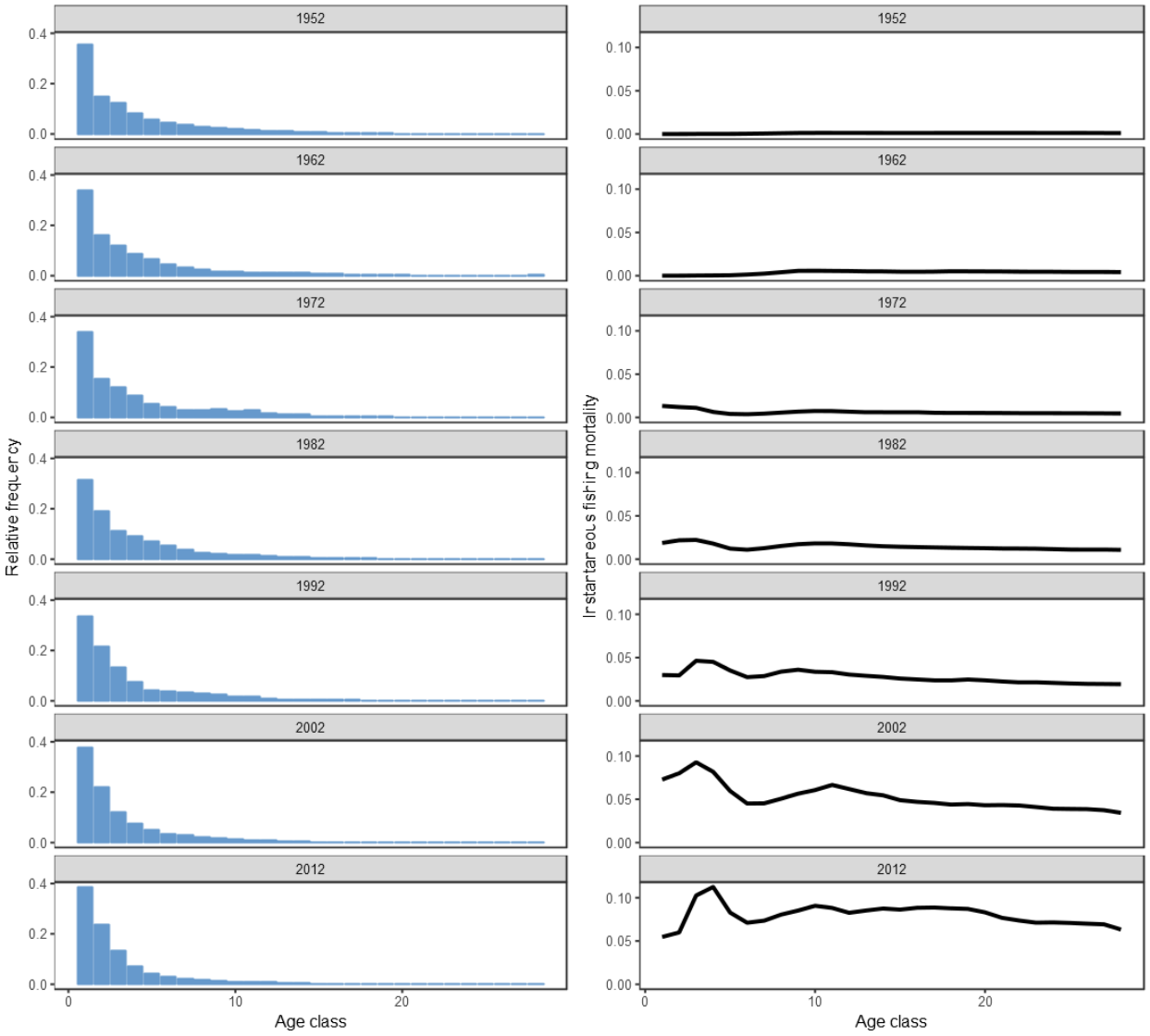

Figure 35: Estimated proportion of the population at-age (quarters; left panels) and fishing mortality-at-age (right panels), at decadal intervals, for the diagnostic case model.

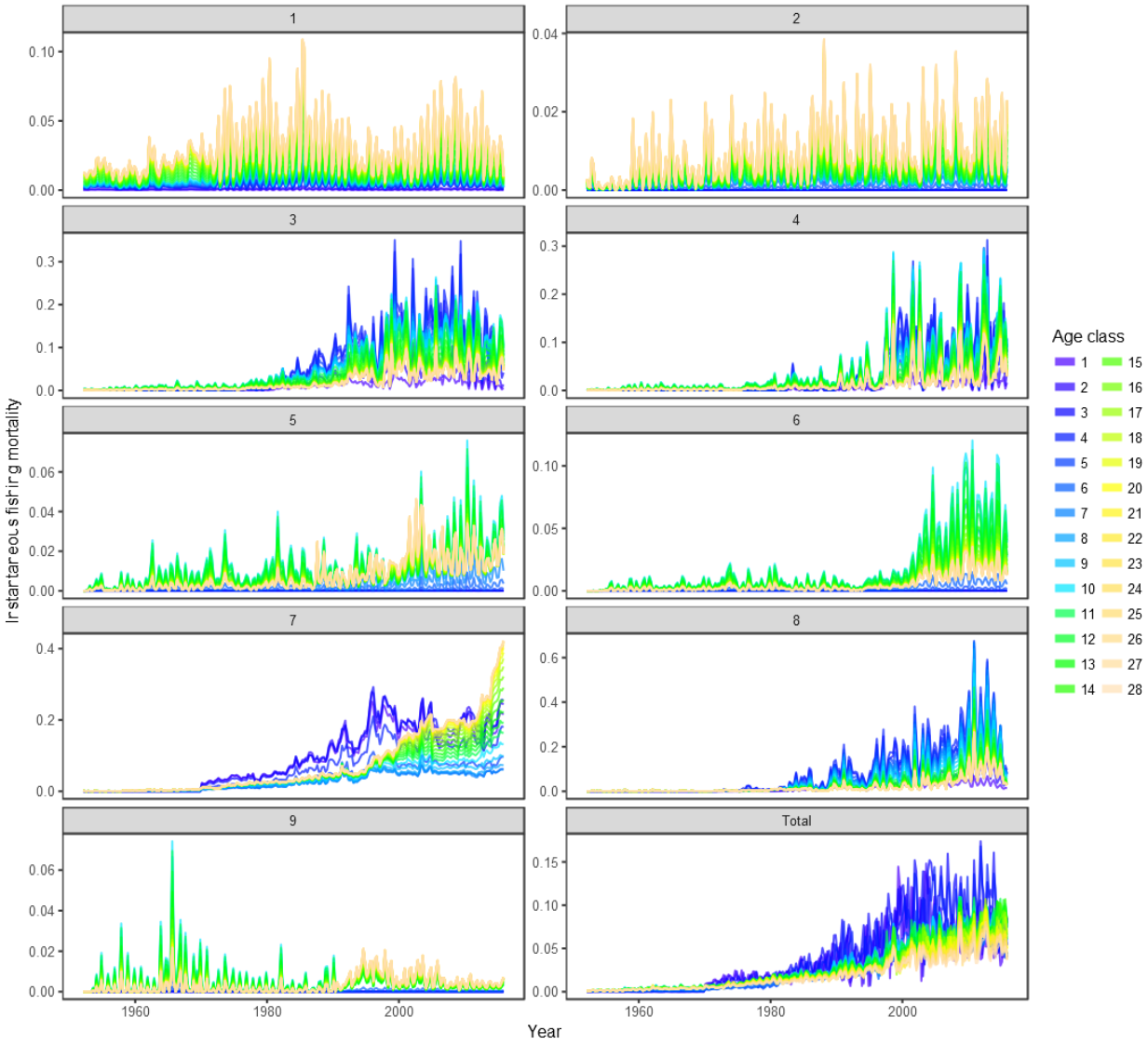

Figure 36: Estimated age-specific fishing mortality for the diagnostic case model, by region and overall.

Figure 37: Estimated spawning potential for each of the one-off sensitivity models investigated in the assessments. The models are separated into 5 groups split by theme for clarity: *SRRQtrly*, *h0.65*, *h0.95* and *Lorenzen* under Bio; *CPUE.Proxy*, *CPUE.Geostats* and *CPUE.LL8* under CPUE; *2014Regions* under Structure, *Size10*, *Size50*, *Dch* under Size.Comp; *Tagmix1*, *TagOD2*, *TagOD4*, *JPTags* under Tag. Details of the models can be found in Section 6.2.

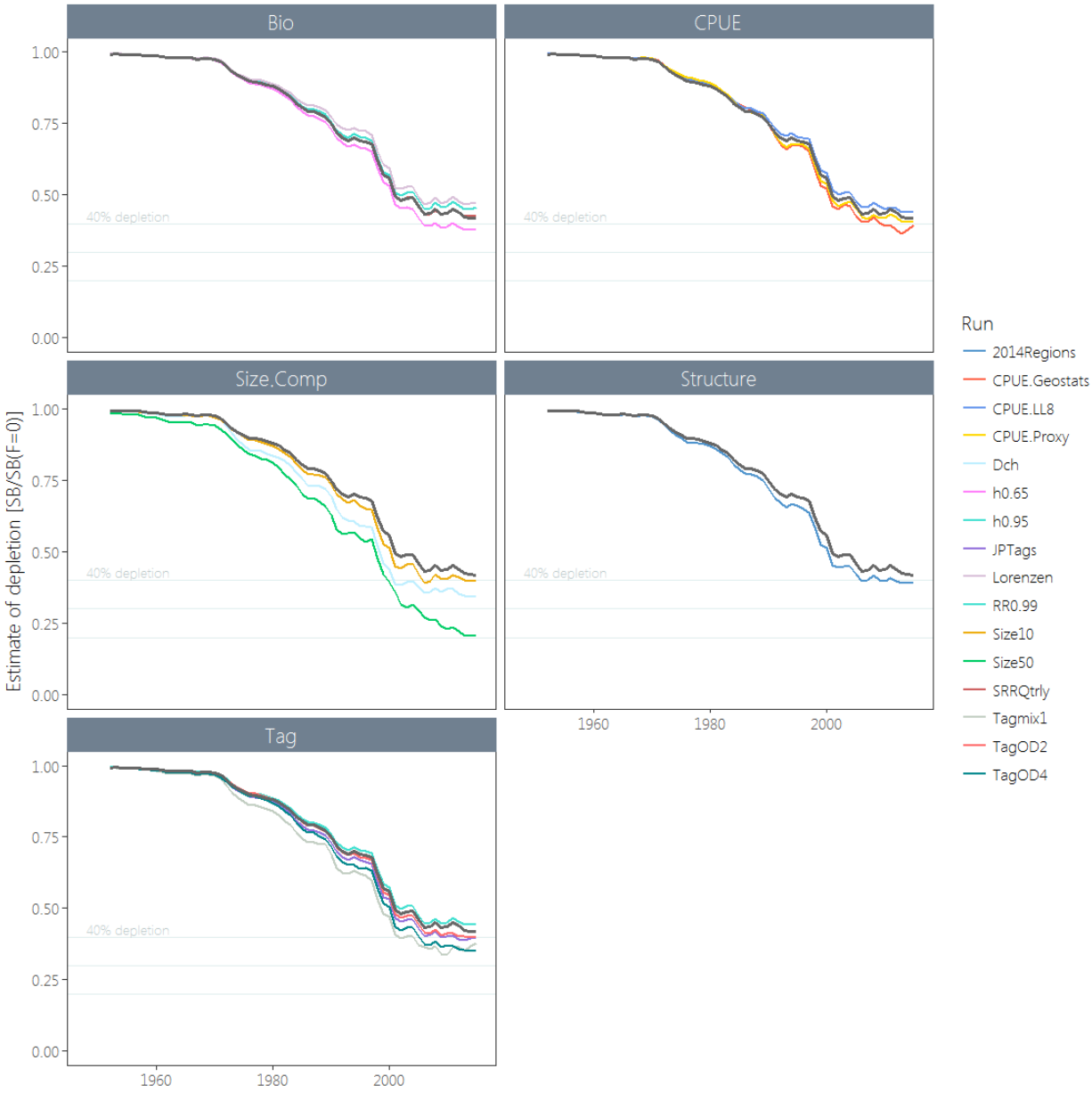

Figure 38: Estimated fishing depletion for each of the one-off sensitivity models investigated in the assessments. The models are separated into 5 groups split by theme for clarity: : *SRRQtrly*, *h0.65*, *h0.95* and *Lorenzen* under Bio; *CPUE.Proxy*, *CPUE.Geostats* and *CPUE.LL8* under CPUE; *2014Regions* under Structure, *Size10*, *Size50*, *Dch* under Size.Comp; *Tagmix1*, *TagOD2*, *TagOD4*, *JPTags* under Tag. Details of the models can be found in Section 6.2.

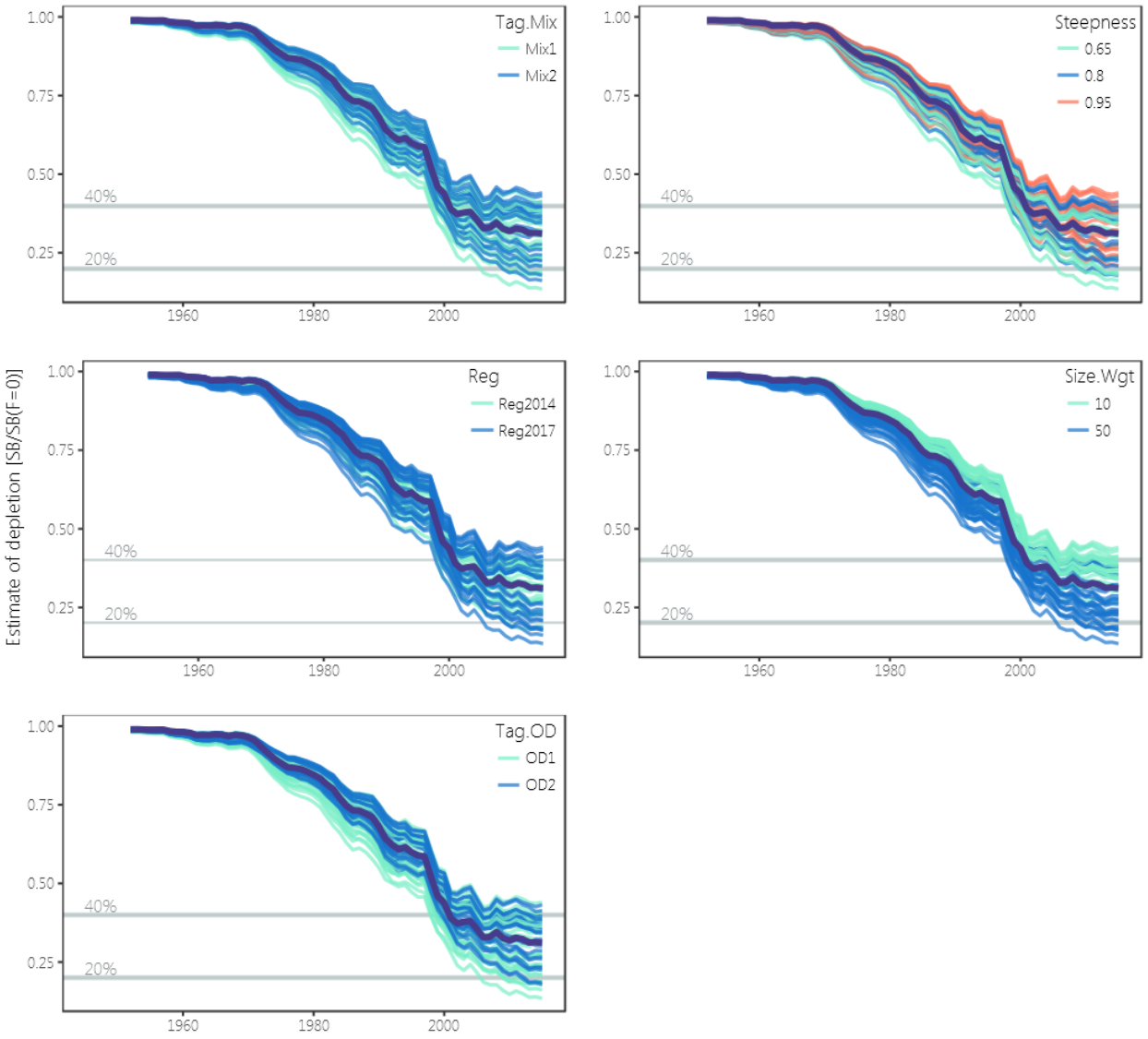

Figure 39: Plots showing the trajectories of fishing depletion (of spawning potential) for the model runs included in the structural uncertainty grid (see Section 6.3 for details of the structure of the grid models). The five panels show the models separated on the basis of the fives axes used in the grid, with the colour denoting the level within the axes for each model.

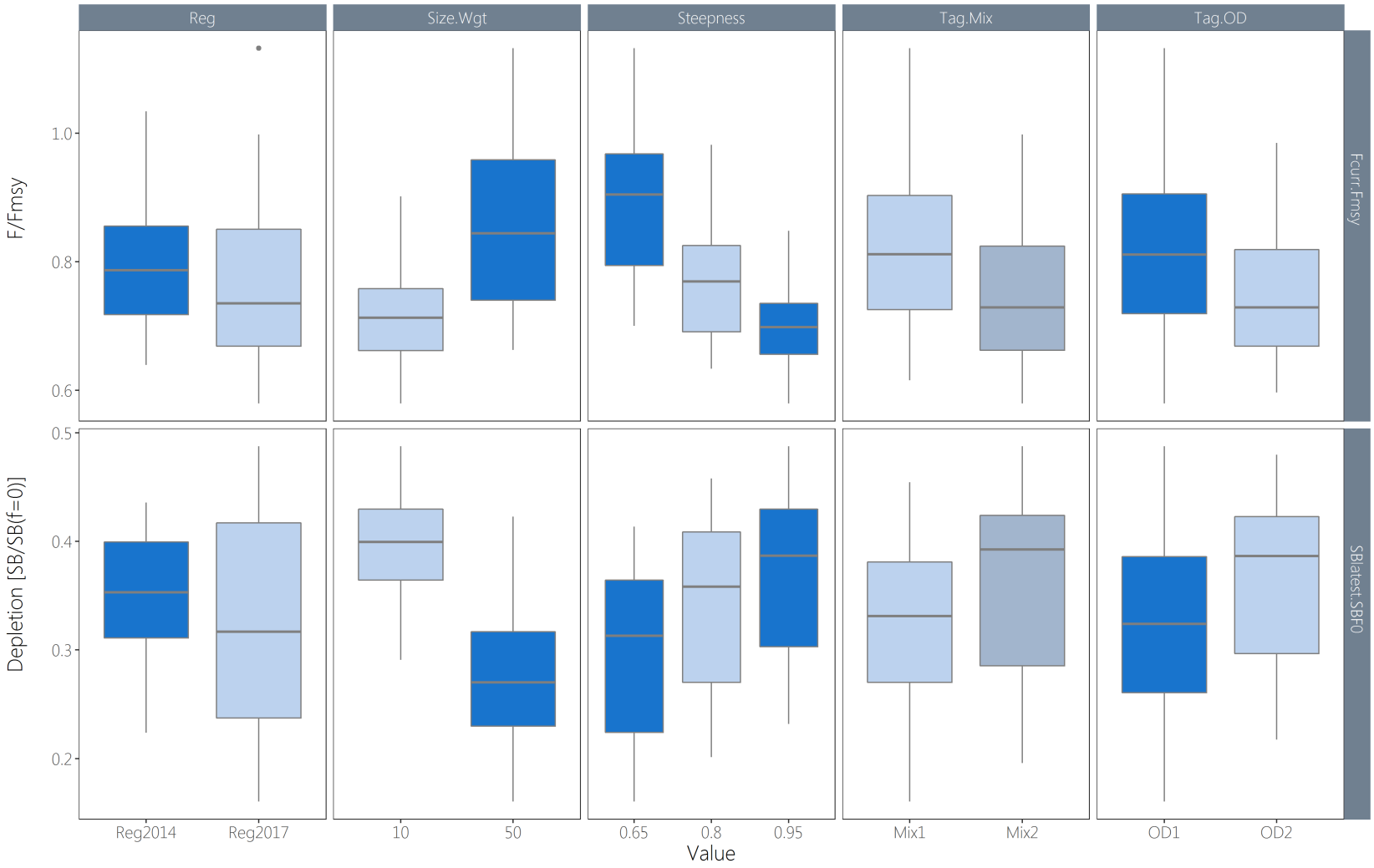

Figure 40: Boxplots summarising the results of the structural uncertainty grid with respect to the fishing mortality reference point F_{recent}/F_{MSY} (upper panels), and the spawning potential reference point (lower panels). The colours indicate the level of the model with respect to the uncertainty axis.

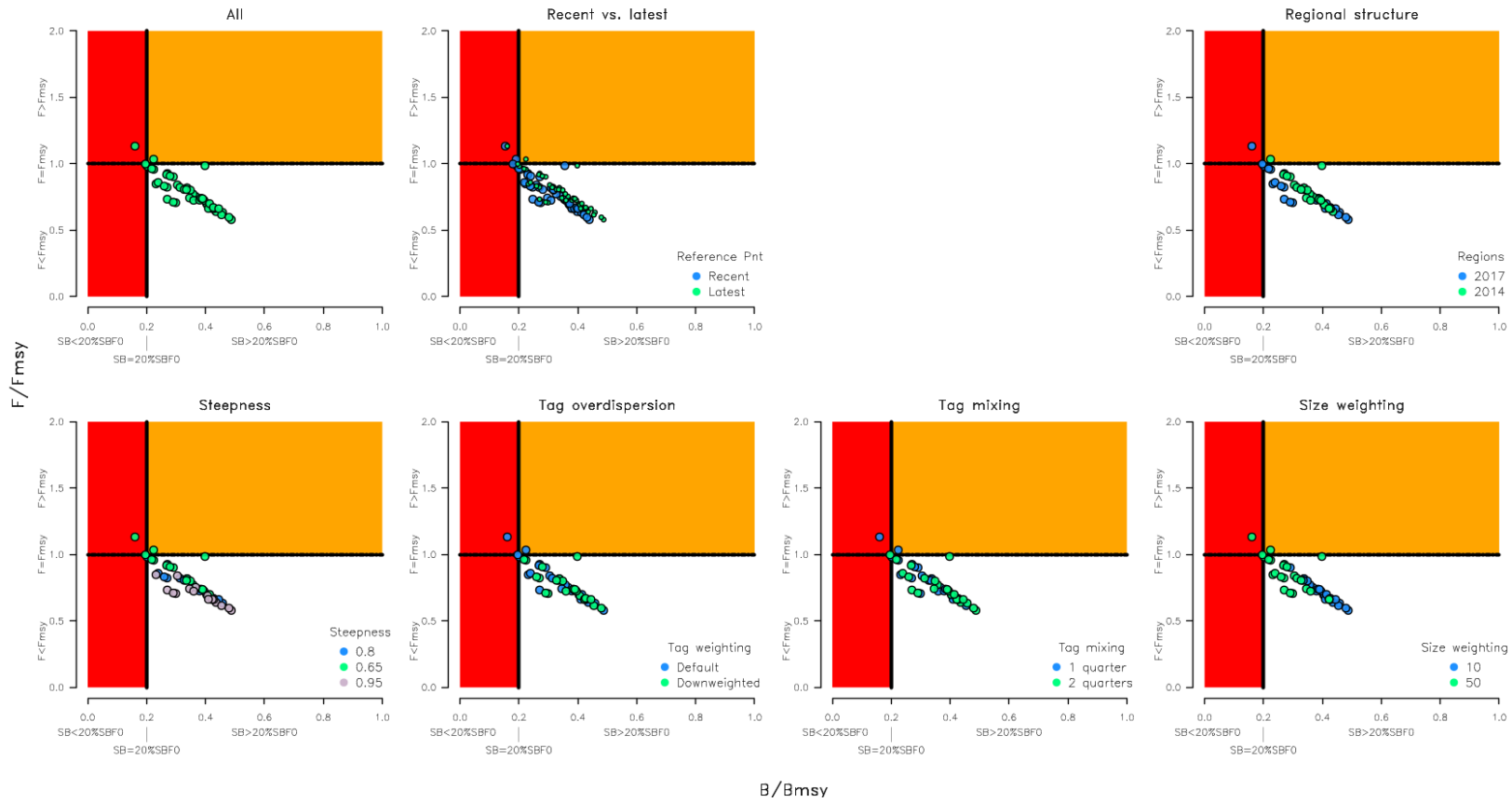

Figure 41: Majuro plots summarising the results for each of the models in the structural uncertainty grid. The plots represent estimates of stock status in terms of spawning potential depletion and fishing mortality. The red zone represents spawning potential levels lower than the agreed limit reference point, which is marked with the solid black line. The orange region is for fishing mortality greater than F_{MSY} (F_{MSY} is marked with the black dashed line). The points represent $SB_{latest}/SB_{F=0}$ for each model run except in panel (b) where $SB_{recent}/SB_{F=0}$ is also displayed. Panels (c)–(g) show the estimates for the different levels for the five axes of the grid.

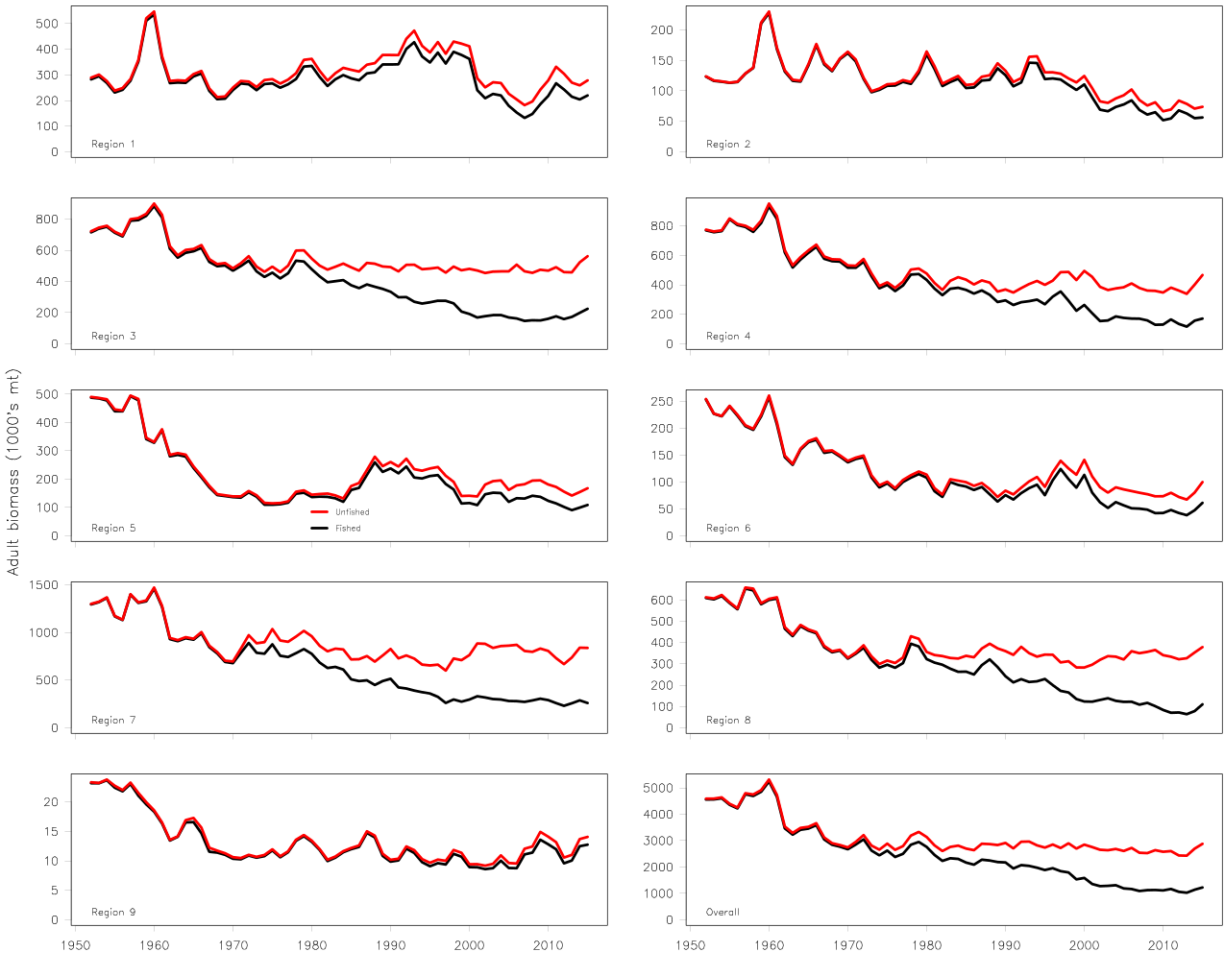

Figure 42: Comparison of the estimated annual spawning potential trajectories (lower solid black lines) with those trajectories that would have occurred in the absence of fishing (upper dashed red lines) for each region, and overall, for the diagnostic case model.

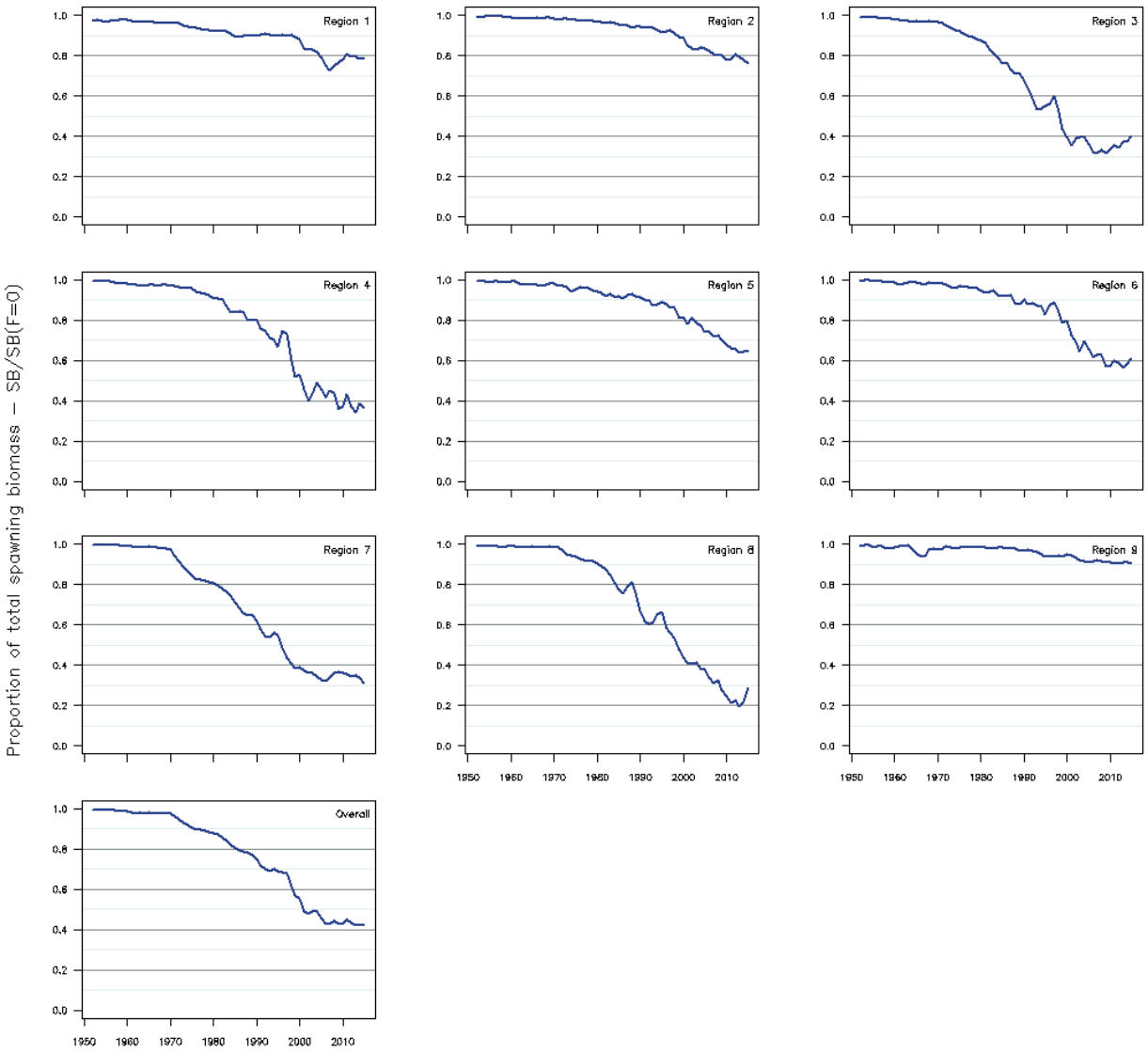

Figure 43: Ratio of exploited to unexploited spawning potential $SB_{latest}/SB_{F=0}$, for each region, and overall, for the diagnostic case model.

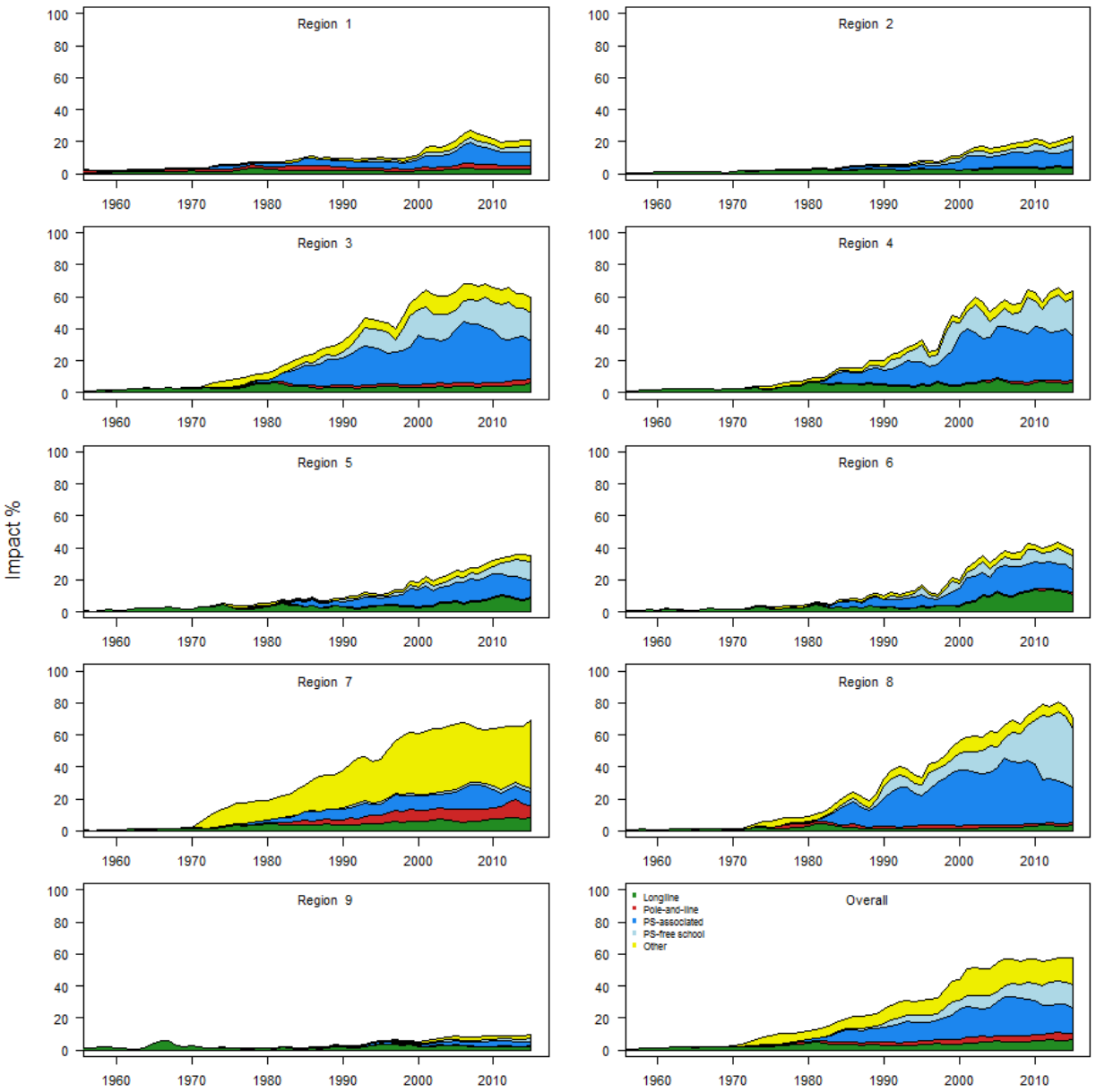

Figure 44: Estimates of reduction in spawning potential due to fishing (fishery impact = $1-SB_{latest}/SB_{F=0}$) by region, and over all regions (lower right panel), attributed to various fishery groups for the diagnostic case model.

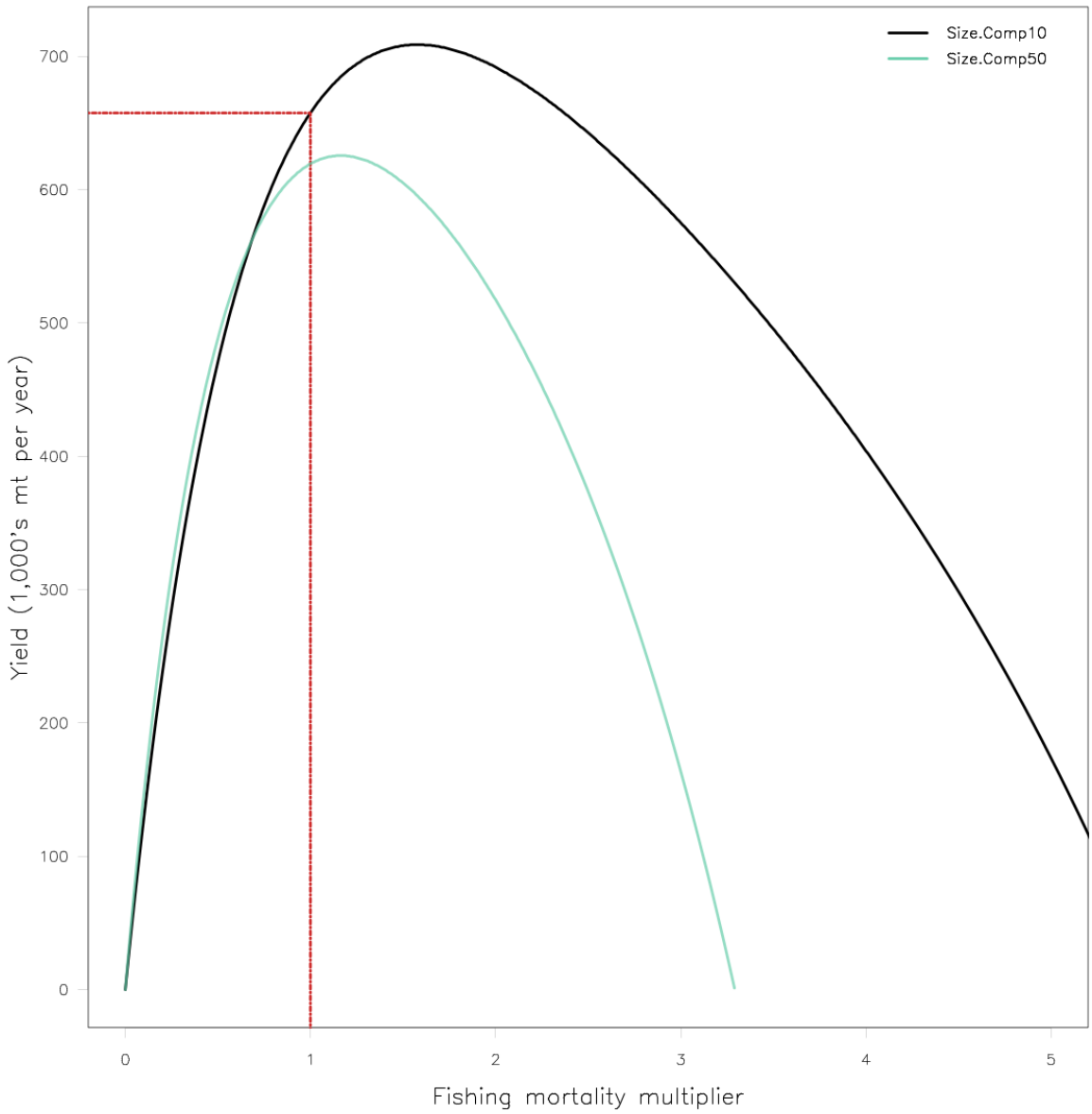

Figure 45: Estimated yield as a function of fishing mortality multiplier for two example models from the grid for 2017 regions with diagnostic case settings except as follow: the black line is the estimated yield curve for the grid run with the upweighting of size data (Size10), the green line is the estimated yield curve for the grid run with the downweighting of size data (Size 50). The red dashed line indicates the equilibrium yield at current fishing mortality.

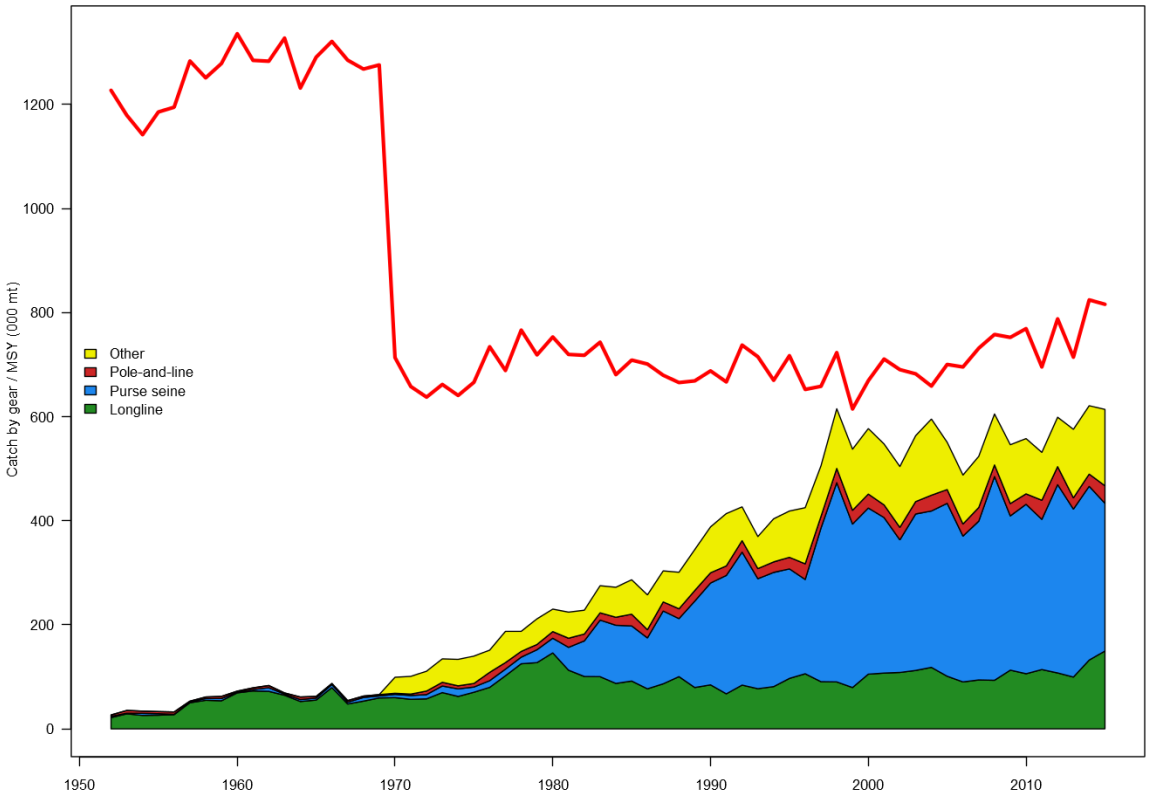

Figure 46: History of the annual estimates of MSY (red line) for the diagnostic case model compared with annual catch by the main gear types. Note that this is a "dynamic" MSY which is explained further in Section 5.8.4.

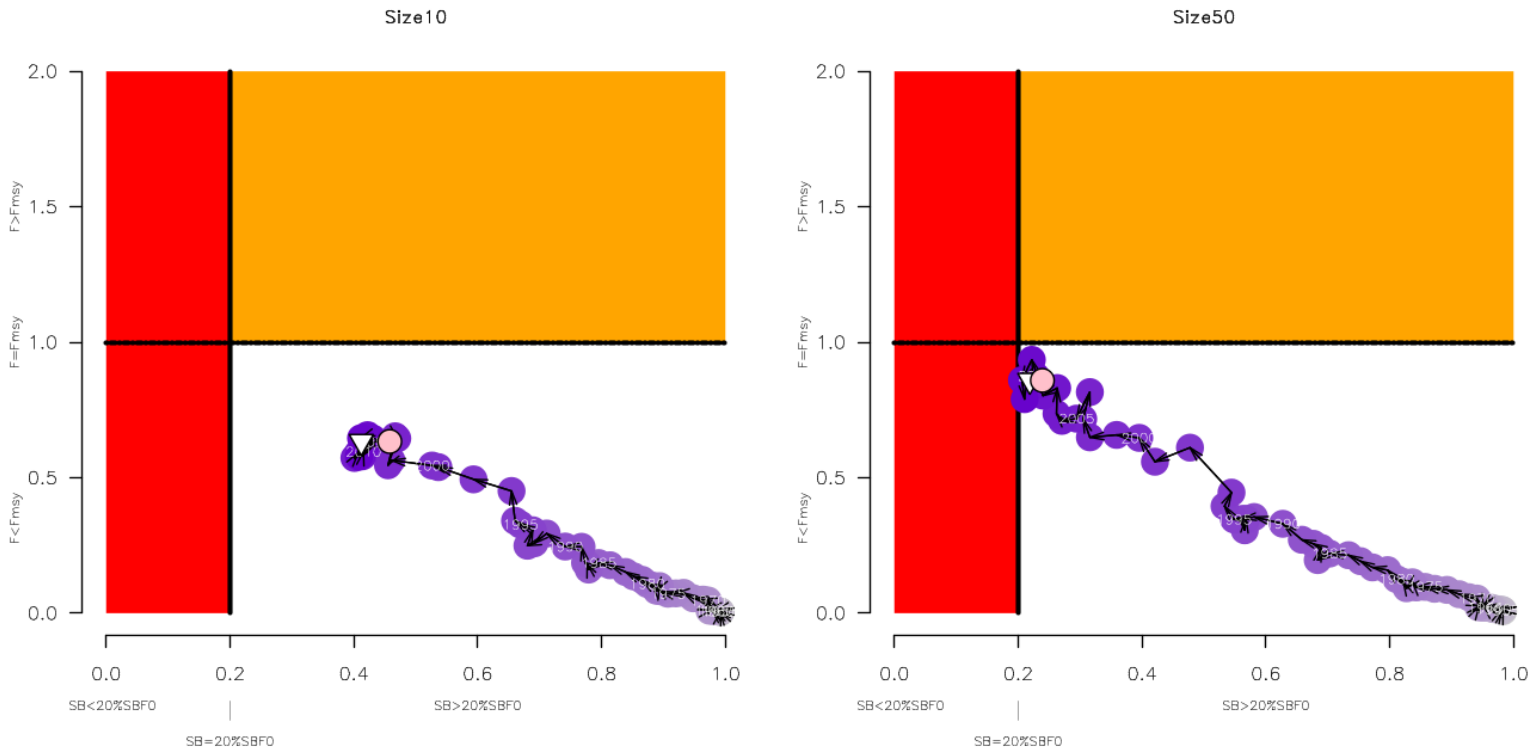

Figure 47: Estimated time-series (or "dynamic") Majuro plots for two example models from the assessment (one from each of the levels of the size weighting axis). These plots are interpreted in the same manner as the description in Figure 41 except that they show the temporal change in stock status with respect to the reference points F_{recent}/F_{MSY} and $SB_{latest}/SB_{F=0}$, rather than the terminal estimates presented in previous figures. Note that the process of estimating a "dynamic" Majuro plot is explained further in Section 5.8.4.

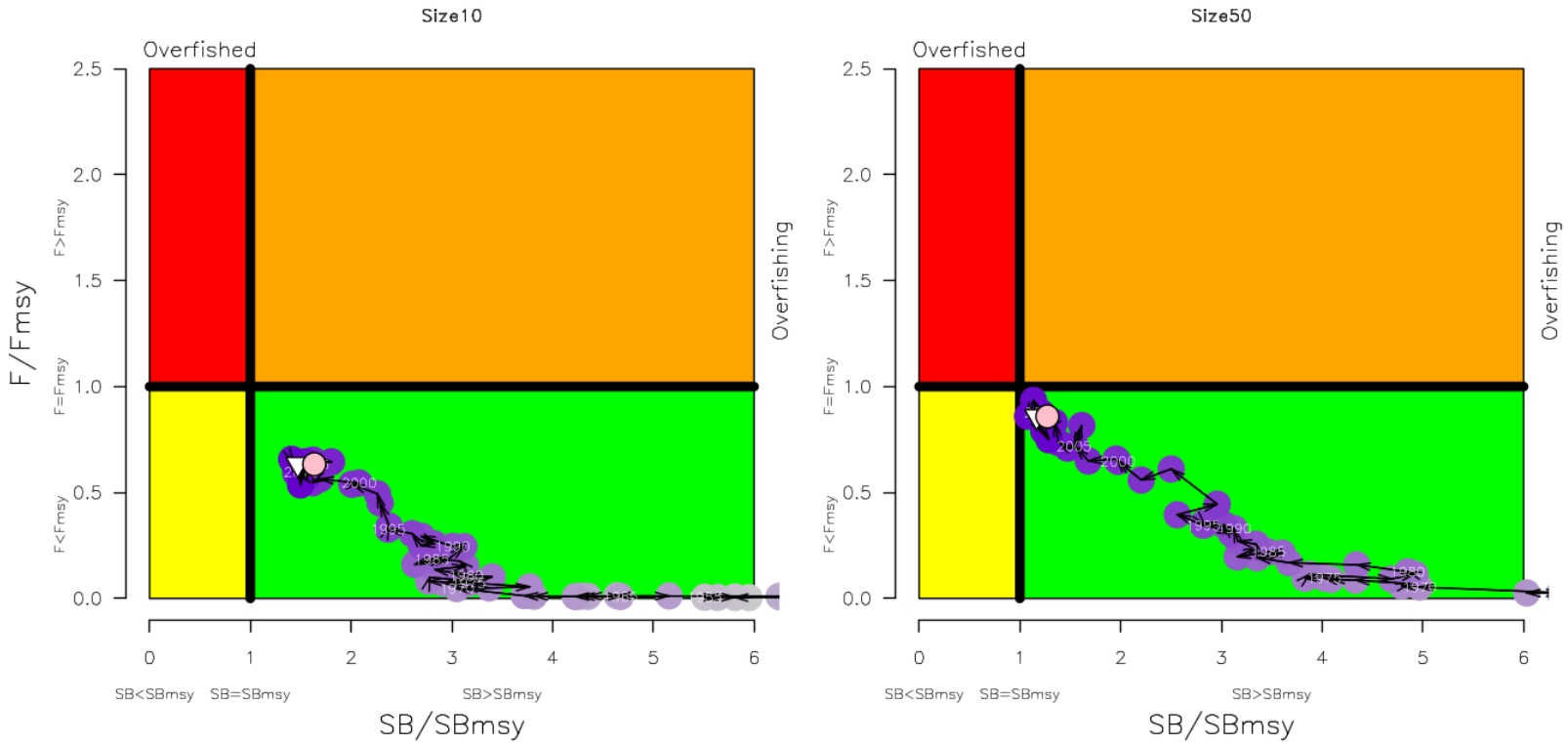

Figure 48: Estimated time-series (or "dynamic") Kobe plots for two example models from the assessment (one from each of the levels of the size weighting axis). Note that the process of estimating a "dynamic" Kobe plot is explained further in Section 5.8.4.

11 Appendix

11.1 Likelihood profile

The approach for calculating a likelihood profile of the derived parameter, mean total biomass over the assessment period (to represent scale of the stock size) is outlined in Section 5.7. The profile was constructed by sequentially moving from the MLE in either direction while progressively penalising the mean total biomass at increasingly high and low values until it was determined that the minimum value had been reached for all data components. The profile reflects the loss of fit over all the data, i.e. the overall objective function value, and the individual data components, caused by changing the population size from that of the maximum likelihood estimated value. The change in likelihood relative to the maximum likelihood estimate is shown for the total likelihood (black line) and the individual data components (coloured lines) in Figure 49 and displays significant declines as the parameter moves further away from the maximum value of the diagnostic case model, although the curves for the individual components display different values of support for the mean total biomass.

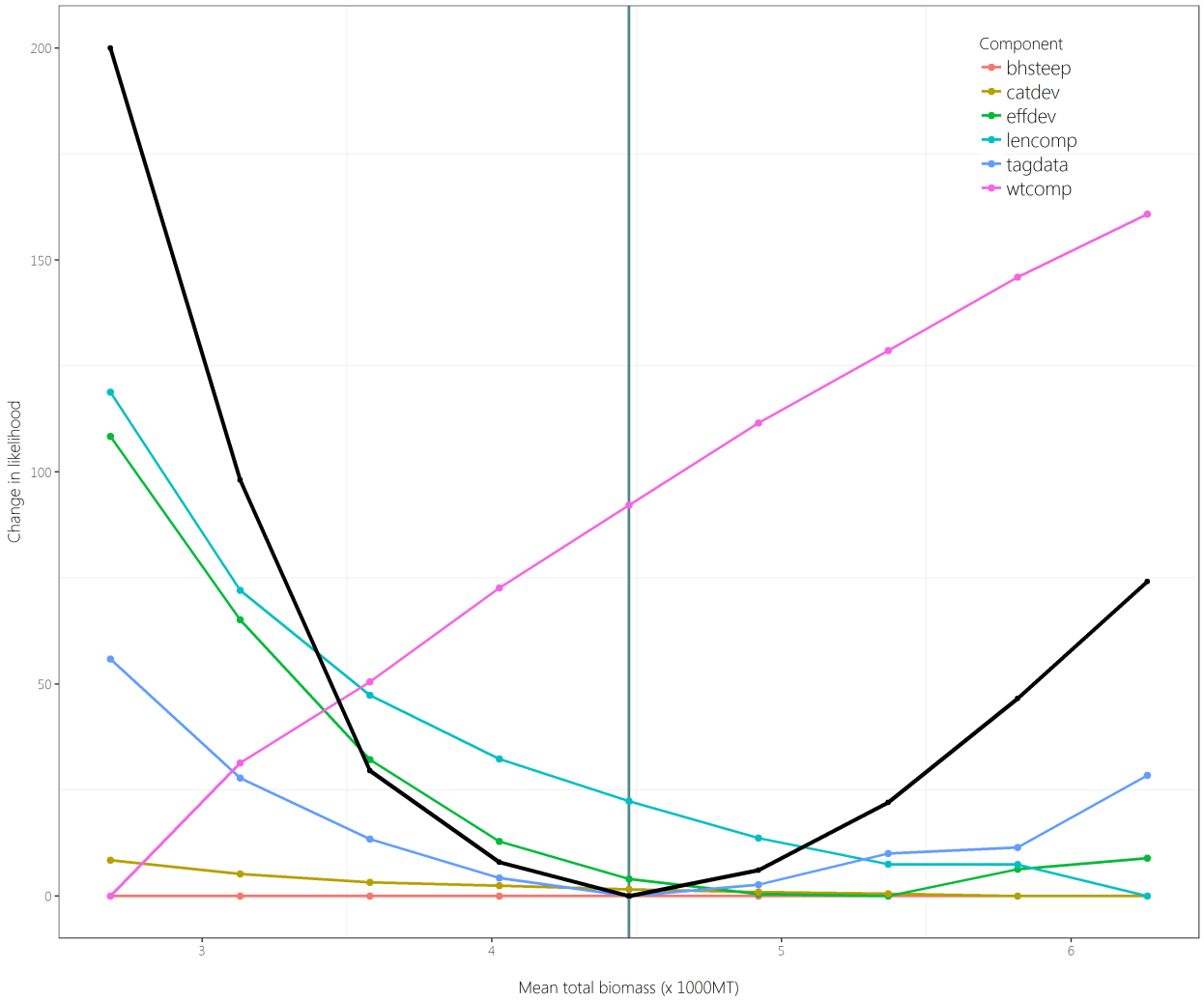

Figure 49: Change in the total (black line), and individual data component log-likelihoods with respect to the derived parameter, mean total biomass over the assessment period, across a range of values at which this parameter was penalised to fit, for the diagnostic case model.

11.2 Retrospective analyses

Retrospective analyses involve rerunning the selected model by consecutively removing successive years of data to estimate model bias (Cadrin and Vaughan, 1997; Cadigan and Farrell, 2005). A series of five additional models were fitted starting with the full data-set (through 2015), followed by models with the retrospective removal of all input data for the years 2015-2011 sequentially. The models are named below by the final year of data included (e.g., 2010-2015). A comparison of the spawning potential, recruitment and depletion trajectories are shown in Figure 50, 51 and 52.

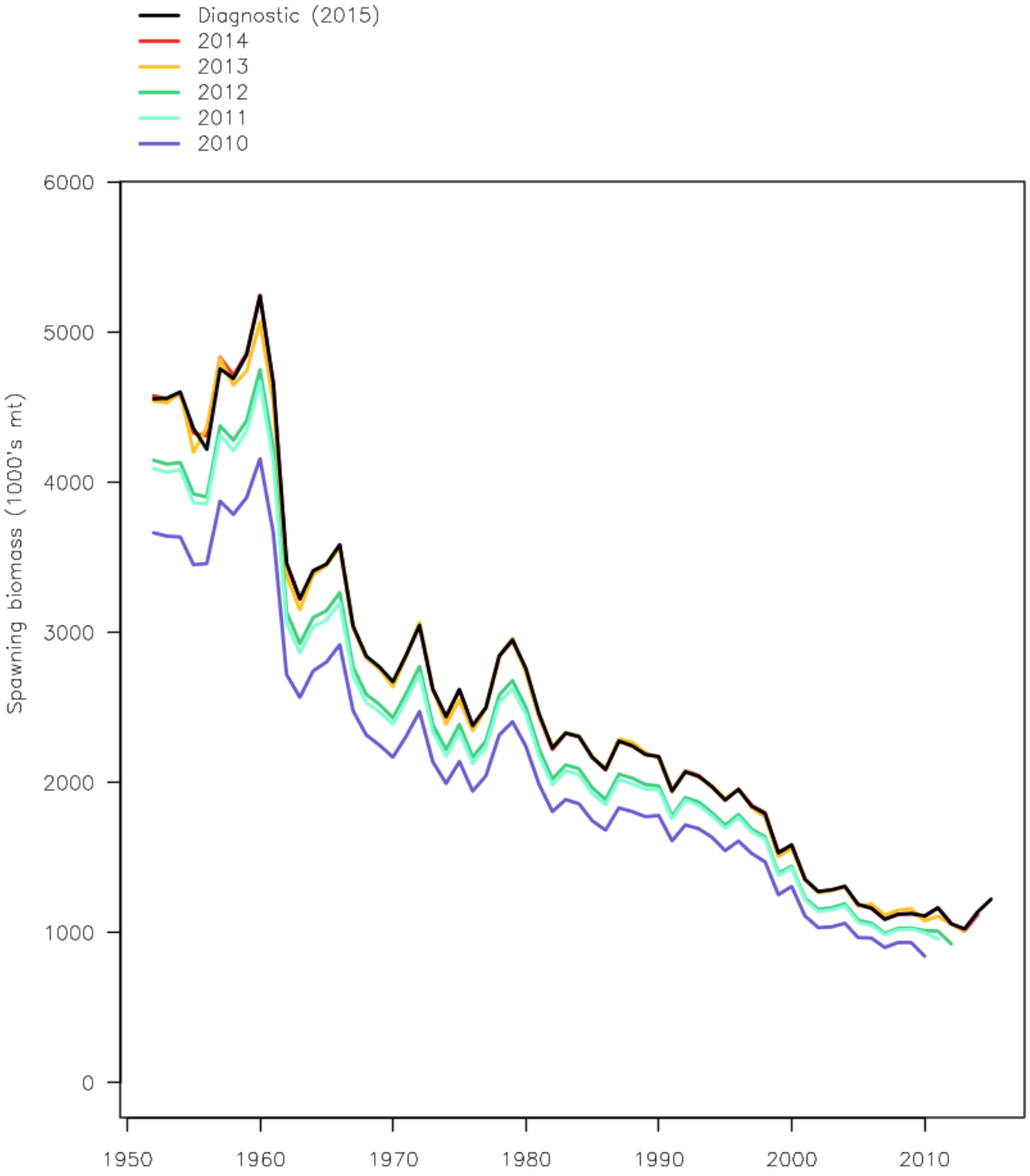

Figure 50: Estimated spawning potential for each of the retrospective models

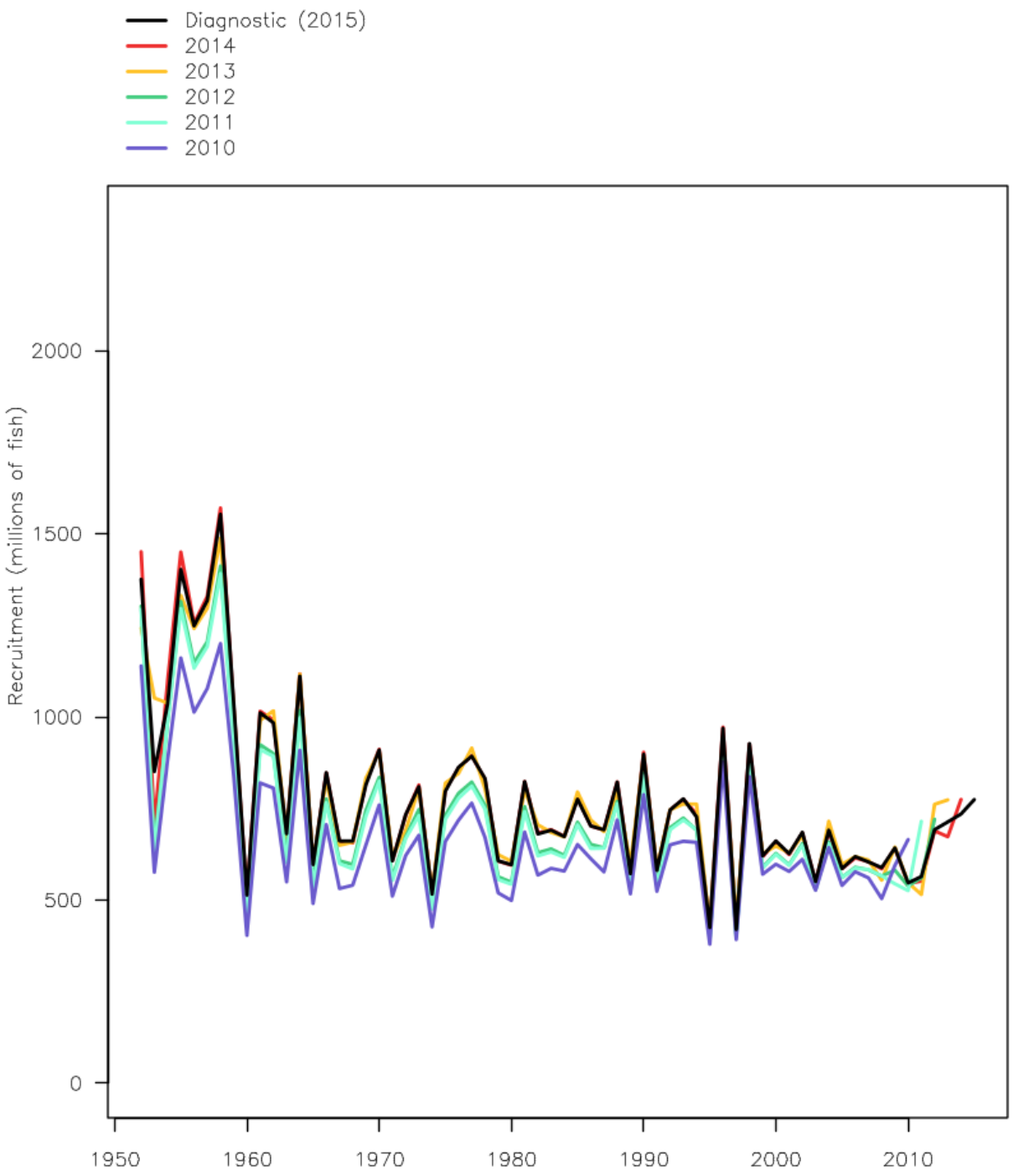

Figure 51: Estimated recruitment for each of the retrospective models

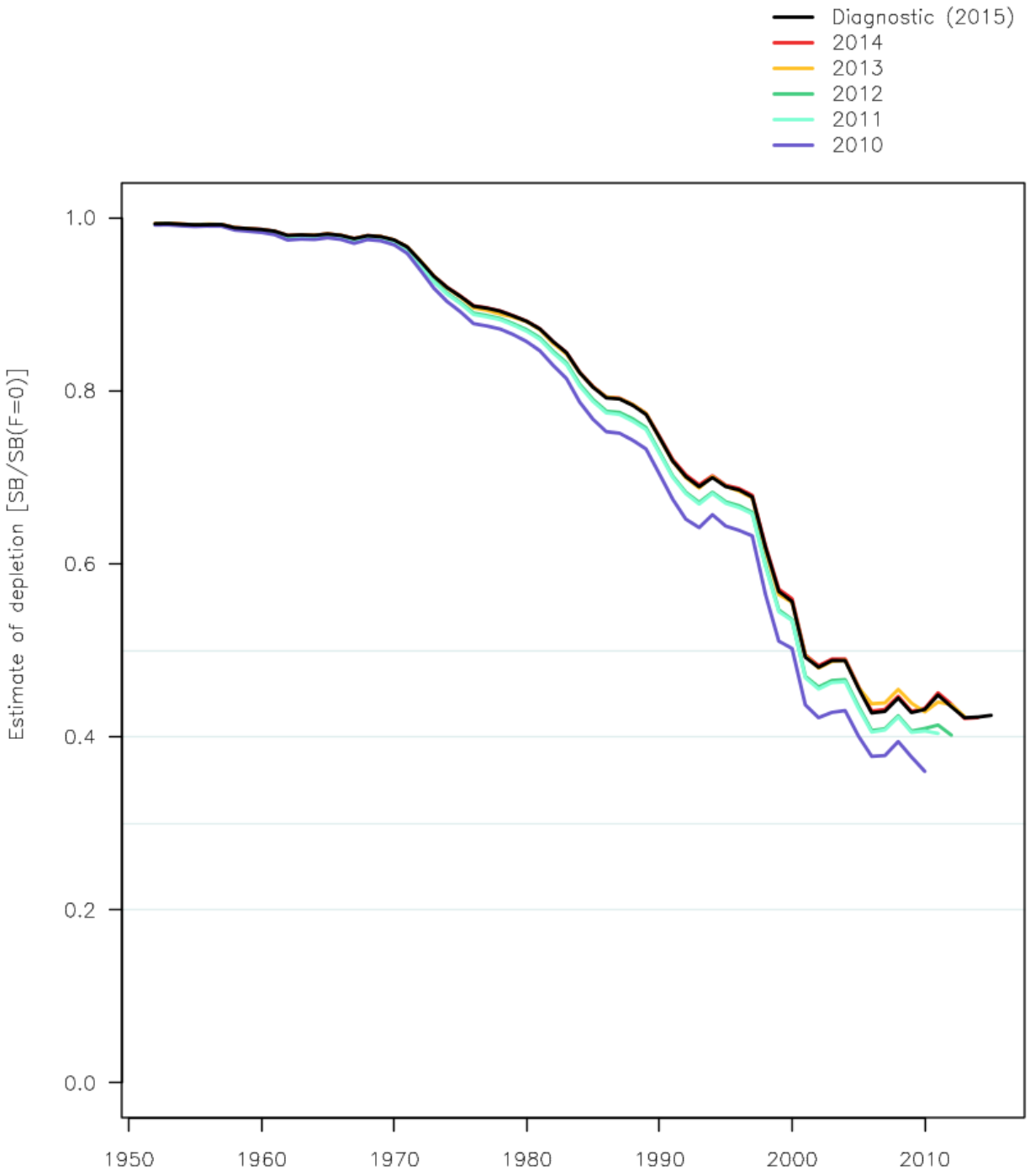

Figure 52: Estimated fishery depletion (SB/SB F=0) for each of the retrospective models

11.3 Sensitivity analyses – reference points and likelihood values

In the current assessment, inferences about stock status and recommendations for management advice are placed on the structural uncertainty grid, rather than the diagnostic case model and the one-off sensitivity model runs, which have received more focus in recent assessments. To this end, 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. The set of focal reference points for these models are presented below in four sets split by theme, along with those of the diagnostic case model:

Table 7: Reference points for the diagnostic case model and the one-off sensitivity models

Quantity	diagno	qtrlySRR	steep65	steep95	Lorenzen
C_{latest}	613581	614024	614012	614022	612919
MSY	757200	651600	796400	731600	817600
$Y_{F_{recent}}$	686000	621200	740000	646800	708400
f_{mult}	1.70	1.39	1.53	1.87	1.91
F_{MSY}	0.11	0.09	0.10	0.12	0.11
$F_{recent}/F_{\text{MSY}}$	0.59	0.72	0.65	0.54	0.52
SB_{MSY}	750100	841000	903400	638500	748200
SB_0	2954000	2824000	3262000	2742000	2873000
SB_{MSY}/SB_0	0.25	0.30	0.28	0.23	0.26
$SB_{F=0}$	2592702	2555626	2848046	2421662	2429001
$SB_{MSY}/SB_{F=0}$	0.29	0.33	0.32	0.26	0.31
SB_{latest}/SB_0	0.40	0.43	0.37	0.45	0.45
$SB_{latest}/SB_{F=0}$	0.46	0.48	0.43	0.50	0.53
$SB_{latest}/SB_{\text{MSY}}$	1.58	1.45	1.35	1.92	1.74
$SB_{recent}/SB_{F=0}$	0.42	0.43	0.38	0.45	0.47
$SB_{recent}/SB_{\text{MSY}}$	1.46	1.31	1.21	1.72	1.52

Table 8: Reference points for the diagnostic case model and the one-off sensitivity models

Quantity	diagno	reg2014	cpue.proxy	cpue.geostats	cpue.LL8
C_{latest}	613581	614661	615175	611516	614531
MSY	757200	726800	774800	794400	788800
$Y_{F_{recent}}$	686000	677200	705200	744000	698400
f_{mult}	1.70	1.56	1.67	1.54	1.80
F_{MSY}	0.11	0.11	0.11	0.11	0.11
$F_{recent}/F_{\text{MSY}}$	0.59	0.64	0.60	0.65	0.55
SB_{MSY}	750100	708900	747400	736800	781400
SB_0	2954000	2733000	2988000	2843000	3088000
SB_{MSY}/SB_0	0.25	0.26	0.25	0.26	0.25
$SB_{F=0}$	2592702	2457147	2560105	2387759	2674748
$SB_{MSY}/SB_{F=0}$	0.29	0.29	0.29	0.31	0.29
SB_{latest}/SB_0	0.40	0.39	0.38	0.39	0.43
$SB_{latest}/SB_{F=0}$	0.46	0.44	0.44	0.46	0.49
$SB_{latest}/SB_{\text{MSY}}$	1.58	1.51	1.52	1.50	1.68
$SB_{recent}/SB_{F=0}$	0.42	0.39	0.41	0.37	0.44
$SB_{recent}/SB_{\text{MSY}}$	1.46	1.35	1.40	1.19	1.50

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

Quantity	diagno	sizecomp10	sizecomp50	Dirichlet
C_{latest}	613581	615117	614037	611486
MSY	757200	708800	624800	654000
$Y_{F_{recent}}$	686000	657600	618800	633200
f_{mult}	1.70	1.58	1.16	1.34
F_{MSY}	0.11	0.11	0.14	0.11
$F_{recent}/F_{\text{MSY}}$	0.59	0.63	0.86	0.74
SB_{MSY}	750100	708100	253700	664800
SB_0	2954000	2774000	1383000	2641000
SB_{MSY}/SB_0	0.25	0.26	0.18	0.25
$SB_{F=0}$	2592702	2528295	1352523	2528886
$SB_{MSY}/SB_{F=0}$	0.29	0.28	0.19	0.26
SB_{latest}/SB_0	0.40	0.42	0.23	0.37
$SB_{latest}/SB_{F=0}$	0.46	0.46	0.24	0.38
$SB_{latest}/SB_{\text{MSY}}$	1.58	1.64	1.27	1.46
$SB_{recent}/SB_{F=0}$	0.42	0.41	0.22	0.34
$SB_{recent}/SB_{\text{MSY}}$	1.46	1.47	1.16	1.31

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

Quantity	diagno	tagmix1	tag.OD2	tag.OD4	tag.wJPTP	rrbound99
C_{latest}	613581	612167	612361	611390	613068	614414
MSY	757200	658400	720800	658800	728000	770800
$Y_{F_{recent}}$	686000	623200	671200	635600	677200	687600
f_{mult}	1.70	1.48	1.57	1.37	1.57	1.77
F_{MSY}	0.11	0.11	0.11	0.11	0.11	0.11
$F_{recent}/F_{\text{MSY}}$	0.59	0.67	0.64	0.73	0.64	0.56
SB_{MSY}	750100	633700	682200	621100	706700	759700
SB_0	2954000	2425000	2661000	2403000	2780000	2999000
SB_{MSY}/SB_0	0.25	0.26	0.26	0.26	0.25	0.25
$SB_{F=0}$	2592702	2345225	2385165	2262555	2480980	2609359
$SB_{MSY}/SB_{F=0}$	0.29	0.27	0.29	0.27	0.28	0.29
SB_{latest}/SB_0	0.40	0.42	0.41	0.38	0.40	0.43
$SB_{latest}/SB_{F=0}$	0.46	0.43	0.46	0.41	0.45	0.49
$SB_{latest}/SB_{\text{MSY}}$	1.58	1.59	1.60	1.48	1.57	1.68
$SB_{recent}/SB_{F=0}$	0.42	0.36	0.40	0.36	0.39	0.44
$SB_{recent}/SB_{\text{MSY}}$	1.46	1.34	1.39	1.30	1.37	1.51

Table 11: Likelihood components for the diagnostic case model and the one-off sensitivity models

Component	diagno	qtrlySRR	steep65	steep95	Lorenzen
Beverton Holt	3.8	4.2	4.0	4.1	6.4
Effort devs	3635.6	3510.7	3510.8	3511.2	3517.8
Catch devs	180.0	175.9	175.7	175.7	167.2
Length comps	-281116.3	-281115.5	-281116.9	-281115.9	-281186.8
Wgt comps	-935902.2	-935911.2	-935912.1	-935911.4	-936056.3
Tagging	9593.1	9596.9	9598.7	9597.2	9584.8
Total	1202570.0	1202706.6	1202705.9	1202706.3	1202926.7

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

Component	diagno	reg2014	cpue.proxy	cpue.geostats	cpue.LL8
Beverton Holt	3.8	3.8	3.8	3.8	3.8
Effort devs	3635.6	3968.4	3834.8	3360.1	3631.8
Catch devs	180.0	177.7	178.0	180.0	175.1
Length comps	-281116.3	-279377.2	-281033.8	-280963.2	-281074.7
Wgt comps	-935902.2	-974810.4	-935707.5	-935624.3	-935758.8
Tagging	9593.1	9587.8	9514.8	9200.6	9683.7
Total	1202570.0	1239354.9	1202036.2	1203025.7	1202394.5

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

Component	diagno	sizecomp10	sizecomp50	Dirichlet
Beverton Holt	3.8	3.8	3.8	3.8
Effort devs	3635.6	3805.1	2795.1	3531.9
Catch devs	180.0	181.0	169.5	184.3
Length comps	-281116.3	-311317.6	-236063.3	171549.3
Wgt comps	-935902.2	-1036965.5	-795353.8	269287.2
Tagging	9593.1	10327.7	7437.7	10603.3
Total	1202570.0	1332538.4	1020375.4	-236928.9

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

Component	diagno	tagmix1	tag.OD2	tag.OD4	tag.wJPTP	rrbound99
Beverton Holt	3.8	3.8	3.8	3.8	3.8	3.8
Effort devs	3635.6	3744.9	3460.2	3357.6	3481.8	3494.1
Catch devs	180.0	178.1	177.1	179.2	174.6	176.4
Length comps	-281116.3	-280950.5	-281378.9	-281514.0	-281068.2	-281113.3
Wgt comps	-935902.2	-935636.9	-936031.3	-936136.7	-935920.4	-935893.0
Tagging	9593.1	10853.0	9758.2	10466.4	10428.5	9528.3
Total	1202570.0	1200930.8	1203015.2	1202595.9	1201858.0	1202766.4

11.4 Structural uncertainty analysis – reference points and likelihood values by regional structure

Table 15: Summary of reference points over the 24 models in the structural uncertainty grid within the grid subset using 2014 regions

	Mean	Median	Min	25%	75%	Max
C_{latest}	611914	614237	606762	607858	614688	615350
MSY	666317	666800	573200	635300	713400	738800
$Y_{F_{recent}}$	648933	649200	573200	617200	678000	720400
f_{mult}	1.28	1.27	0.97	1.17	1.39	1.56
F_{MSY}	0.11	0.11	0.07	0.11	0.12	0.14
$F_{recent}/F_{\text{MSY}}$	0.79	0.79	0.64	0.72	0.86	1.03
SB_{MSY}	610892	609000	381400	517175	698000	946800
SB_0	2370958	2391000	1816000	2172500	2676000	2935000
SB_{MSY}/SB_0	0.25	0.26	0.21	0.24	0.26	0.34
$SB_{F=0}$	2273813	2306726	1704671	2093189	2431993	2800716
$SB_{MSY}/SB_{F=0}$	0.27	0.26	0.22	0.25	0.28	0.35
SB_{latest}/SB_0	0.34	0.33	0.24	0.30	0.38	0.40
$SB_{latest}/SB_{F=0}$	0.35	0.35	0.22	0.31	0.40	0.44
$SB_{latest}/SB_{\text{MSY}}$	1.33	1.37	0.90	1.15	1.47	1.68
$SB_{recent}/SB_{F=0}$	0.31	0.32	0.19	0.27	0.36	0.40
$SB_{recent}/SB_{\text{MSY}}$	1.33	1.32	0.96	1.22	1.49	1.64

Table 16: Summary of reference points over the 24 models in the structural uncertainty grid within the grid subset using 2017 regions

	Mean	Median	Min	25%	75%	Max
C_{latest}	613614	613627	612040	613254	614031	615164
MSY	658850	664200	539200	625400	687000	754400
$Y_{F_{recent}}$	635450	628200	534400	609700	674900	716000
f_{mult}	1.33	1.36	0.88	1.18	1.50	1.73
F_{MSY}	0.12	0.12	0.10	0.11	0.13	0.16
$F_{recent}/F_{\text{MSY}}$	0.78	0.74	0.58	0.67	0.85	1.13
SB_{MSY}	481975	462300	186800	260250	671675	863200
SB_0	2059375	2023000	1197000	1412500	2610000	3105000
SB_{MSY}/SB_0	0.22	0.23	0.15	0.18	0.26	0.28
$SB_{F=0}$	1954062	1995736	1193336	1359035	2438297	2813584
$SB_{MSY}/SB_{F=0}$	0.23	0.24	0.16	0.19	0.28	0.31
SB_{latest}/SB_0	0.31	0.30	0.18	0.23	0.39	0.45
$SB_{latest}/SB_{F=0}$	0.33	0.32	0.16	0.24	0.42	0.49
$SB_{latest}/SB_{\text{MSY}}$	1.42	1.39	0.80	1.24	1.67	1.91
$SB_{recent}/SB_{F=0}$	0.30	0.30	0.15	0.22	0.39	0.44
$SB_{recent}/SB_{\text{MSY}}$	1.41	1.42	0.81	1.28	1.63	1.81

11.5 Summaries of the structural uncertainty grid with the addition of an extra level of size data weighting

This section presents an update of the structural uncertainty grid for extra models run subsequent to the submission of the previous version of the yellowfin stock assessment report uploaded to the WCPFC SC website. These relate to the addition of an extra level to the size data weighting axis in the structural uncertainty grid. Due to time constraints, only the two bounds of the size data weightings (divisors of 10 and 50) that were investigated in the one-off sensitivity model runs were

originally included in the grid, due to the large computational overhead, limited computational resources and time constraints imposed on the assessments. Subsequent to the report deadline, we have completed the model runs for this extra level and so modify the stock assessment figures and tables to include summaries from these extra runs. These are briefly presented in this section in Figures A39 to A41 and Tables A6, A15 and A16. These figures and tables are directly comparable to the original figures and tables *via* the matching numbering. As this section presents the summaries of the most complete structural uncertainty grid, we recommend referring to the following figures and tables when formulating recommendations from this stock assessment. The original figures, tables and main body text are left unmodified only to prevent confusion for those that have previously downloaded the stock assessment report.

The addition of the extra level of the size data weighting uncertainty axis had a small effect on the estimates of stock status and model outputs of the assessment. In general, reference points were shifted by a few points in the optimistic direction (e.g. the mean of $SB_{latest}/SB_{F=0}$ went from 0.34 to 0.37 and the mean of $F_{recent}/F_{\text{MSY}}$ went from 0.79 to 0.75, Table A6). This is because even though the divisor of 20 falls between the more extreme levels (10 and 50) originally included, the new models produce estimates very similar, if slightly more variable, than those predicted under the more optimistic divisor of 10. Therefore, including the divisor of 20 runs shifts the balance of the grid towards more optimistic estimates. This is also seen when the grid metrics are calculated by regional structure, but the effects are slightly more pronounced under the 2017 region structure (e.g. the mean of $SB_{latest}/SB_{F=0}$ goes from 0.33 to 0.37 while it goes from 0.35 to 0.37 for the 2014 region structure). The 2017 region structure is also the one with the most variable outcomes in the grid.

Table A5: Description of the updated structural sensitivity grid used to characterise uncertainty in the assessment.

Axis	Levels	Option
Steepness	3	0.65, 0.80, or 0.95
Tagging overdispersion	2	Default level, Fixed (moderate) level
Tag mixing	2	1 or 2 (default) quarters
Size frequency weighting	3	sample sizes divided by 10, 20 or 50
Regional structure	2	2017 regions, 2014 regions

Table A6: Summary of reference points over all 72 individual models in the structural uncertainty grid

	Mean	Median	Min	25%	75%	Max
C_{latest}	612742	613430	606762	612107	614237	615350
MSY	673589	674400	539200	635300	713400	795200
$Y_{F_{recent}}$	647239	644000	534400	614200	681300	739600
f_{mult}	1.36	1.37	0.88	1.22	1.51	1.86
F_{MSY}	0.12	0.11	0.07	0.11	0.12	0.16
$F_{recent}/F_{\text{MSY}}$	0.75	0.73	0.54	0.66	0.82	1.13
SB_{MSY}	589514	609000	186800	501800	718650	946800
SB_0	2335931	2438500	1197000	2065250	2731000	3256000
SB_{MSY}/SB_0	0.25	0.26	0.15	0.23	0.27	0.34
$SB_{F=0}$	2207825	2301517	1193336	2034075	2509122	2845244
$SB_{MSY}/SB_{F=0}$	0.26	0.26	0.16	0.25	0.29	0.35
SB_{latest}/SB_0	0.35	0.36	0.18	0.30	0.40	0.45
$SB_{latest}/SB_{F=0}$	0.37	0.39	0.16	0.30	0.43	0.50
$SB_{latest}/SB_{\text{MSY}}$	1.42	1.41	0.80	1.24	1.62	1.91
$SB_{recent}/SB_{F=0}$	0.33	0.35	0.15	0.27	0.39	0.45
$SB_{recent}/SB_{\text{MSY}}$	1.42	1.43	0.81	1.28	1.59	1.93

Table A15: Summary of reference points over the 36 models in the structural uncertainty grid with the 2014 region structure

	Mean	Median	Min	25%	75%	Max
C_{latest}	612269	613582	606762	610311	614687	615350
MSY	670978	668400	573200	635300	713400	776800
$Y_{F_{recent}}$	649811	649200	566000	617200	680400	739600
f_{mult}	1.32	1.33	0.97	1.22	1.45	1.71
F_{MSY}	0.11	0.11	0.07	0.11	0.12	0.14
$F_{recent}/F_{\text{MSY}}$	0.77	0.75	0.58	0.69	0.82	1.03
SB_{MSY}	627036	611400	381400	536425	718650	946800
SB_0	2410167	2425000	1816000	2235750	2676000	3073000
SB_{MSY}/SB_0	0.26	0.26	0.21	0.24	0.27	0.34
$SB_{F=0}$	2298919	2301517	1704671	2129469	2474667	2800716
$SB_{MSY}/SB_{F=0}$	0.27	0.27	0.22	0.25	0.29	0.35
SB_{latest}/SB_0	0.35	0.35	0.24	0.32	0.38	0.42
$SB_{latest}/SB_{F=0}$	0.37	0.37	0.22	0.33	0.41	0.47
$SB_{latest}/SB_{\text{MSY}}$	1.36	1.40	0.90	1.17	1.52	1.76
$SB_{recent}/SB_{F=0}$	0.33	0.34	0.19	0.29	0.37	0.42
$SB_{recent}/SB_{\text{MSY}}$	1.37	1.40	0.96	1.26	1.50	1.77

Table A16: Summary of reference points over the 36 models in the structural uncertainty grid with the 2017 region structure

	Mean	Median	Min	25%	75%	Max
C_{latest}	613214	613430	611018	612354	614010	615164
MSY	676200	683000	539200	638000	711800	795200
$Y_{F_{recent}}$	644667	641000	534400	613100	681500	739600
f_{mult}	1.4	1.41	0.88	1.24	1.57	1.86
F_{MSY}	0.12	0.12	0.10	0.11	0.12	0.16
$F_{recent}/F_{\text{MSY}}$	0.73	0.71	0.54	0.64	0.81	1.13
SB_{MSY}	551992	595050	186800	345375	717000	901800
SB_0	2261694	2464000	1197000	1589500	2748500	3256000
SB_{MSY}/SB_0	0.23	0.24	0.15	0.22	0.26	0.28
$SB_{F=0}$	2116731	2291150	1193336	1618096	2539821	2845244
$SB_{MSY}/SB_{F=0}$	0.25	0.26	0.16	0.21	0.29	0.32
SB_{latest}/SB_0	0.34	0.37	0.18	0.26	0.42	0.45
$SB_{latest}/SB_{F=0}$	0.37	0.40	0.16	0.27	0.45	0.5
$SB_{latest}/SB_{\text{MSY}}$	1.48	1.51	0.80	1.29	1.74	1.91
$SB_{recent}/SB_{F=0}$	0.33	0.36	0.15	0.25	0.40	0.45
$SB_{recent}/SB_{\text{MSY}}$	1.47	1.54	0.81	1.34	1.67	1.93

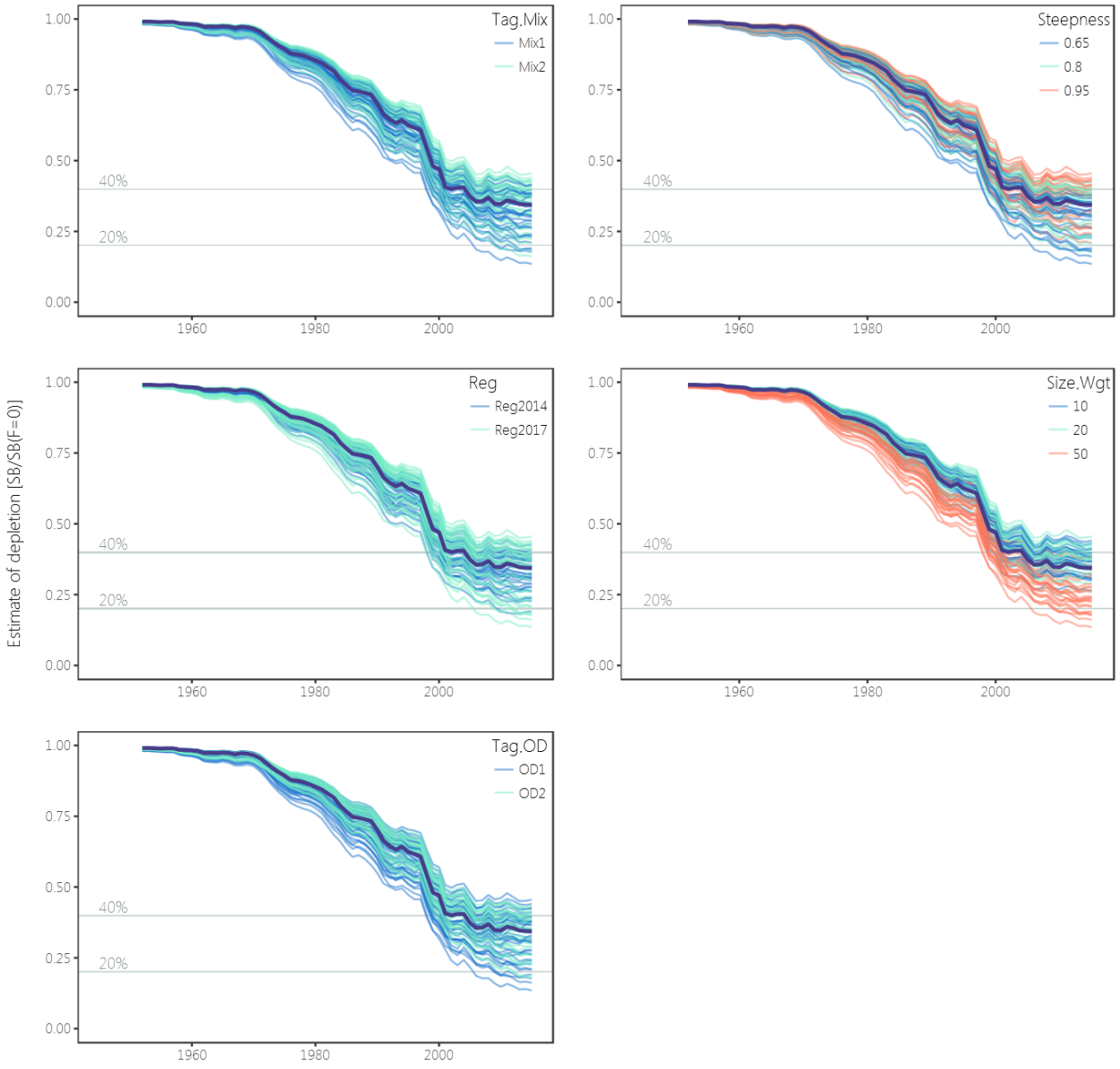

Figure A39: Plots showing the trajectories of fishing depletion (of spawning potential) for the model runs included in the structural uncertainty grid (see Section 6.3 for details of the structure of the grid models). The five panels show the models separated on the basis of the fives axes used in the grid, with the colour denoting the level within the axes for each model.

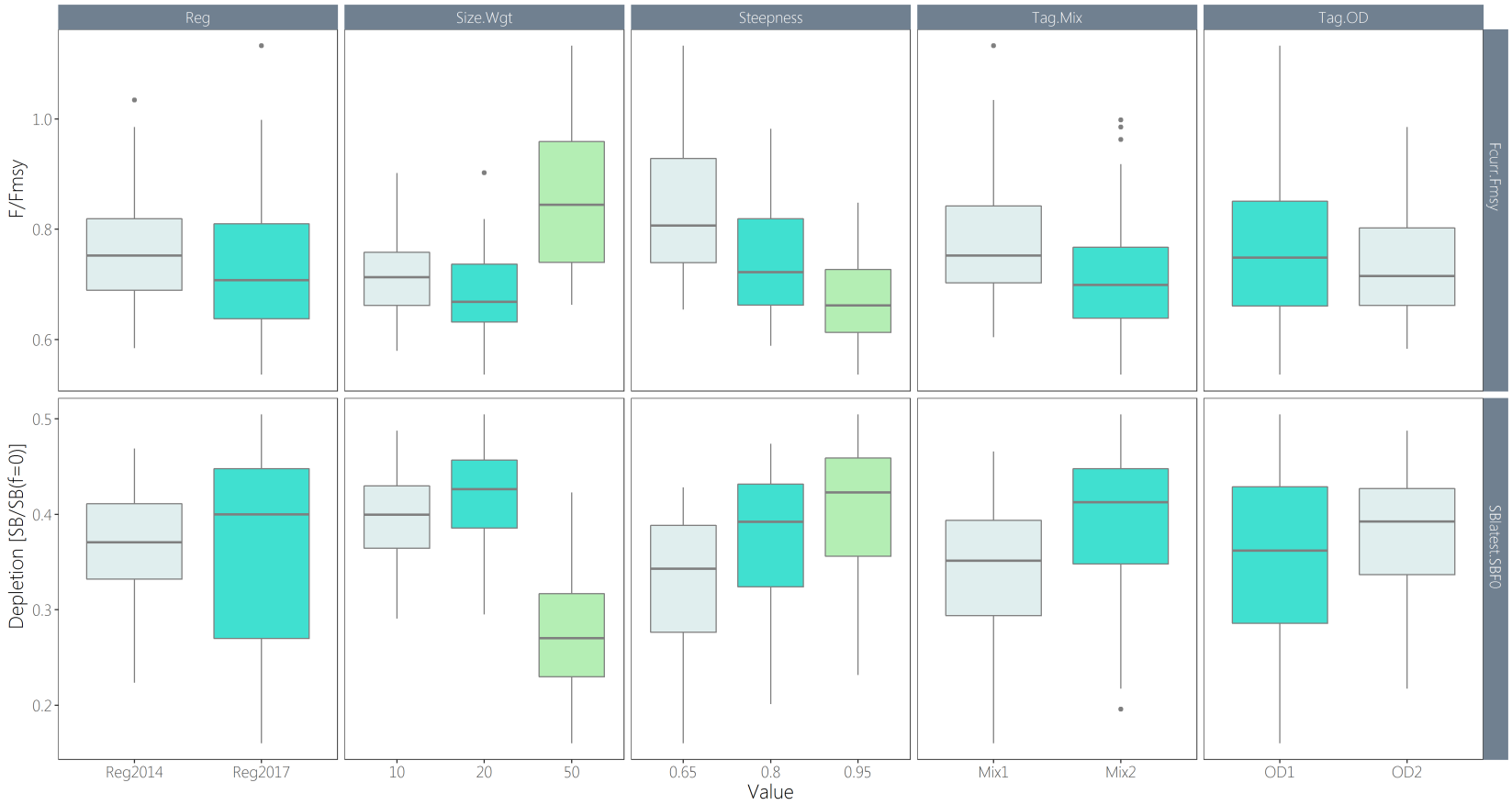

Figure A40: Boxplots summarising the results of the structural uncertainty grid with respect to the fishing mortality reference point F_{recent}/F_{MSY} (upper panels), and the spawning potential reference point (lower panels). The colours indicate the level of the model with respect to the uncertainty axis, with the setting used for the diagnostic case in turquoise.

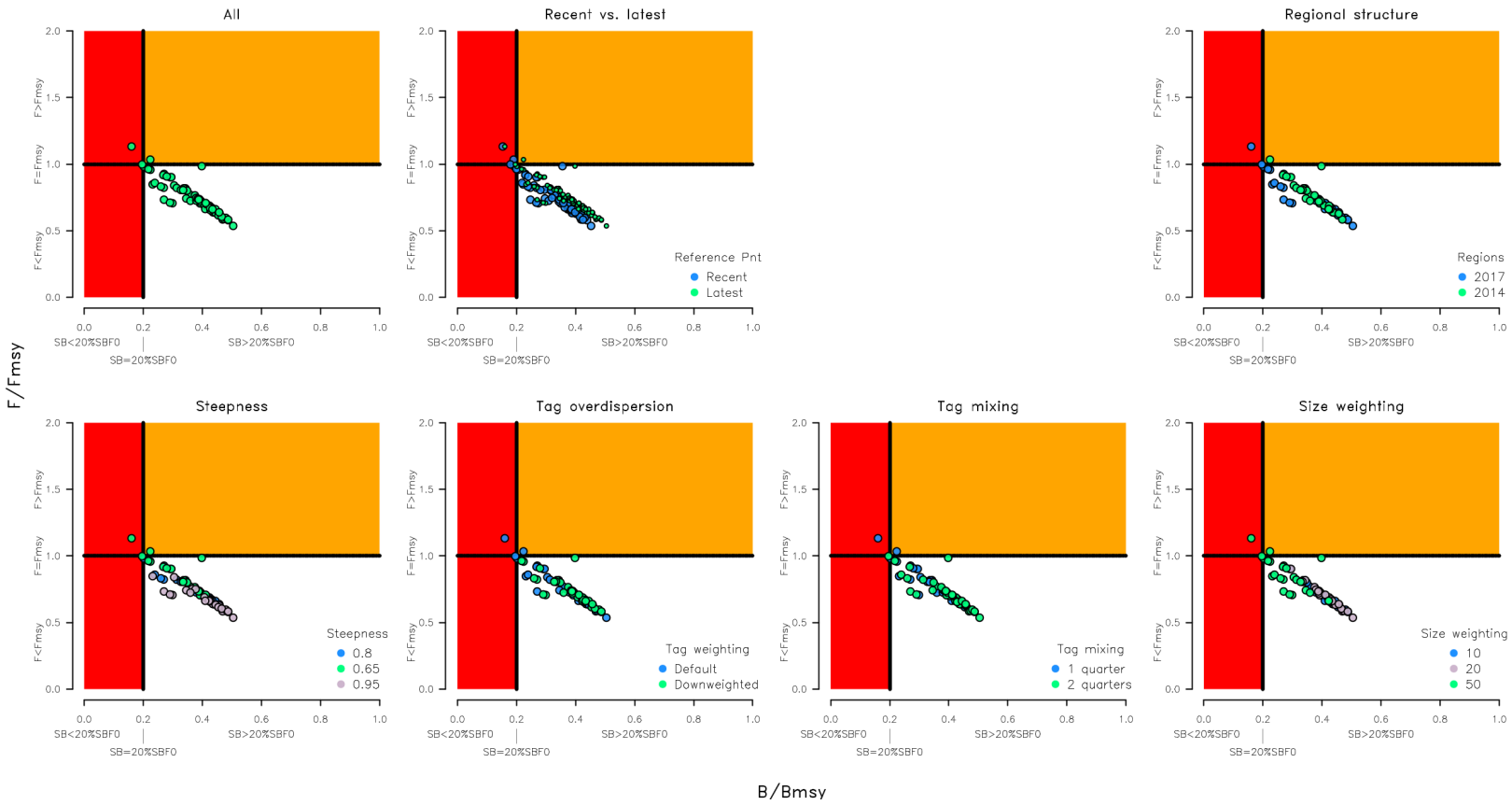

Figure A41: Majuro plots summarising the results for each of the models in the structural uncertainty grid. The plots represent estimates of stock status in terms of spawning biomass depletion and fishing mortality. The red zone represents spawning biomass levels lower than the agreed limit reference point, which is marked with the solid black line. The orange region is for fishing mortality greater than F_{MSY} (F_{MSY} is marked with the black dashed line). The points represent $SB_{latest}/SB_{F=0}$ for each model run except in panel (b) where $SB_{recent}/SB_{F=0}$ is also displayed. Panels (c)–(g) show the estimates for the different levels for the five axes of the grid.