Western and Central Pacific Fisheries Commission

SCIENTIFIC COMMITTEE
TWENTY-FIRST REGULAR SESSION

Nuku'alofa, Tonga
13–21 August 2025

Stock assessment of swordfish in the southwest Pacific Ocean: 2025

WCPFC-SC21-2025/SA-WP-05
Revision 3, August 18, 2025

J. Day[1], C. Castillo-Jordán[1], A. Magnusson[1], K. Kim[1], T. Teears [1], N. Davies[2], J. Hampton[1], S. McKechnie[1], T. Peatman[3], T. Vidal[1], P. Hamer[1]

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

Author contribution

Jemery Day lead analyst, performed the analysis, results interpretation, and primary author of the report. **Claudio Castillo-Jordán** provided technical support and advice and contributed to report production. **Arni Magnusson** technical support and advice, dealt with the ensemble model outputs and key results tables and plots for the paper. **Kyuhan Kim** technical support, engineered the code for running large SS3 model ensembles on HTCondor, processing ensemble model outputs and creating related figures. **Thom Teears** technical advice, assistance with data preparation. **Nick Davies** technical support and advice. **John Hampton** technical support and advice. **Sam McKechnie** produced the catch summary plots for the paper. **Tom Peatman** conducted the size data preparatory work and provided the size data inputs. **Tiffany Vidal** support with data provisions. **Paul Hamer** oversight of the assessment, size data input review, results interpretation and report writing.

Revision 1:

This version includes two additional tables, describing the uncertainty grid (Table 2) and definitions of reference points and quantities used in the analysis (Table 3). Some minor typographical errors are also corrected.

Revision 2:

This version incorporates two new Majuro plots, a static plot for the uncertainty grid, without estimation uncertainty, (Figure 81) and a dynamic time-series plot for the diagnostic model, again without estimation uncertainty (Figure 82).

Revision 3:

This version updates the y-axis for Figure 40, which was previously plotted (and mislabelled) in terms of F/F_{MSY}, rather than plotted as annual fishing mortality, F. This figure now also includes additional juvenile and adult fishing mortality time series. Some minor typographical errors are also corrected.

Contents

1 Executive summary

This report describes the development of the 2025 stock assessment of southwest Pacific Ocean swordfish. It represents the seventh assessment of swordfish in the southwest Pacific since 2006, with all previous assessment except one being conducted in MULTIFAN-CL, the other was conducted in CASAL. At the request of SC20, the 2025 assessments was transitioned from MULTIFAN-CL to Stock Synthesis in line with the objectives of WCPFC Project 123: Scoping the Next Generation of Tuna Stock Assessment Software. While the 2025 Stock Synthesis assessment used aspects of the previous MULTIFAN-CL assessment to guide the initial assessment model set-up, the objective was to develop a new benchmark assessment in Stock Synthesis, utilising features of the Stock Synthesis software, rather than try to replicate the previous MULTIFAN-CL assessment. The 2025 assessment makes a fundamental structural change to develop a two-sex assessment and various other changes and alternative approaches are applied in the quest to improve the assessment. The key changes include:

- Change of stock assessment platform from MULTIFAN-CL to Stock Synthesis.
- Conversion from a 1-sex model to a 2-sex model.
- Major revision of the size data inputs (both length and weight), stronger filtering methods and additional downweighting of unreliable size data.
- Separation of all fisheries into distinct fisheries in each model subregion.
- Changing from age-based selectivity to length-based selectivity.
- Constraining selectivity options to logistic or double normal, rather than complex spline-based selectivities.
- Inclusion of sex-specific conditional age-at-length data within the assessment and using these data to contribute to internal growth estimation.
- Estimating the scale of the Lorenzen mortality form, using more recent published approaches of Hamel and Cope (2022); Hoyle (2022).
- Switching to an updated length-weight relationship based on a new dataset and with more careful filtering of the old data (Macdonald et al., 2025).
- The use of a quarterly model time step, rather than an annual time step, to allow greater resolution in modelling the growth curve.
- The adoption of variable bin widths for the weight composition data, using a feature in Stock Synthesis that is not presently available in MFCL.

Due to the time spent focusing on these substantial structural changes, a standard grid approach was used to characterise uncertainty in management quantities. This is in contrast to the more

sophisticated approach pioneered in 2021 of creating a model ensemble based on joint prior distributions to characterise uncertainty. The ensemble approach appears to be a more suitable approach, and could be reinstated for the next swordfish assessment.

An additional four years of data were incorporated into this assessment, from 2020–2023, including new catch, CPUE and size composition data. We are confident that the changes implemented in 2025 have advanced this assessment. However, there are some key concerns. Firstly, growth continues to be a major uncertainty for this assessment, and no suitable resolution was found to problems discovered when applying fixed external growth, as was used successfully in the 2021 diagnostic model. As a result, growth uncertainty is not incorporated in the uncertainty grid. Von Bertalanffy growth parameters were estimated internally for females, with sex specific conditional age-at-length data as the major source of information for these growth estimates, and with fixed offsets used for male growth parameters. Secondly, model outputs and diagnostics indicated that the uncertainty in population scale is large. This key output, population scale, required an appropriate contribution from the weight composition data to counter the tendency of the CPUE data to support unrealistically high estimates of population scale. The conflict in these data sources is problematic and independent validation of the population size is important to increase confidence in the management advice for this integrated assessment. Any bias in estimated population scale will bias key management metrics. Thirdly, we note the influence of weight composition data, which has cyclic variation and recent decreases, most notably for Australian longline weight composition data in subregion 1C. A better understanding of the processes (biological, sampling or fishing practices) influencing the variation in the weight composition time series is important to ensure that these data are treated appropriately in the model. Finally, spatial structure remains a key uncertainty, the current spatial structure may be mis-specified and the treatment of the large catch area in the northeast of the model region, and adjacent waters outside the model region, needs review with regards to appropriate spatial extent and structure of the assessment region. Acknowledging these concerns, the result the assessment can be summarised as:

The 2025 SWPO swordfish stock assessment estimates that, at the stock-wide scale, for all models in the uncertainty grid, the median $SB_{\text{recent}}/SB_{\text{MSY}}$ is well above 1.0, with a value of 2.37, showing that SB is well above SB_{MSY}, (Figure 77, Table 4), and the median $F_{\text{recent}}/F_{\text{MSY}}$ is well below 1.0, with median value of 0.27, showing that F_{recent} is also well below F_{MSY} (Figure 74, Table 4). These two key reference points are calculated incorporating both structural and estimation uncertainty.

The estimated stock status, relative to SB_{MSY} across the whole model region shows a slow, gradual decline until the 1990s, with an initial increase, followed by a period of faster decline, through to about 2009, followed by a brief recovery to 2015 and over the last decade the stock has continued to decline (Figure 77).

Overall, the outcomes of this assessment suggest that the swordfish stock in the SWPO is not overfished and is not undergoing overfishing.

We invite the SC21 to:

- Note the considerable amount of work and changes to develop the new two-sex assessment in Stock Synthesis.
- Note the results of the assessment in terms of the default MSY based reference points.
- Note the concerns raised on the uncertainty in population scale, limited conditional age-at-length data and ongoing uncertainty in the model spatial extent and structure.
- Consider the application of Close Kin Mark Recapture (CKMR) as a means to obtain a more reliable independent estimate of the population size for validation of the model estimates.

2 Introduction

2.1 Overview

This paper presents the 2025 stock assessment of swordfish (*Xiphias gladius*) in the southwest Pacific Ocean (SWPO) (south of the Equator between 140°E and 130°W), including the area of overlap between the Western and Central Pacific Fisheries Commission (WCPFC) convention area and the Inter-American Tropical Tuna Commission (IATTC) convention areas. Unlike previous assessments that were conducted in MULTIFAN-CL (Fournier et al., 1998), this assessment was conducted in Stock Synthesis (Methot and Wetzel, 2013). The switch from MULTIFAN-CL to Stock Synthesis was driven by the request from SC20 to transition billfish assessments from MULTIFAN-CL to Stock Synthesis in line with the objectives of WCPFC Project 123:*Scoping the Next Generation of Tuna Stock Assessment Software* (Anon, 2024):

- **4.1.2 Review of Project 123 outcomes**
- **para 20.** The report from the Informal Small Group 09 (Project 123: Scoping the next generation of tuna stock assessment software) is in Attachment 1. The development of new software was discussed but not prioritized by ISG-09. The ISG-09 provided the following prioritization of proposed project 123 activities:
  - Move the SW Pacific swordfish assessment to Stock Synthesis;
  - Move the next SW Pacific striped marlin assessment to Stock Synthesis, if the successor software is not available;

2.2 Historical SWPO swordfish assessments

The first assessment of swordfish in the south western Pacific region (Kolody et al., 2006) was conducted in 2006 using MULTIFAN-CL and focused on the area 140°E – 175°W. This first assessment noted the considerable uncertainty in stock status and that the assessment uncertainty will not be substantially reduced until there is considerable additional data collected, ideally including improved interpretation of catch rates, direct observations of movement from tagging (conventional, electronic or genetic), direct aging from hard parts, and improved size data and sex sampling. The general conclusion was that the biomass (total and spawning) are probably above levels that would sustain MSY and fishing mortality is probably below F_{MSY}. This initial assessment was updated in 2008 by Kolody et al. (2008), again using MULTIFAN-CL, and drew similar conclusions, with some reduction in the uncertainty compared to the 2006 assessment. In the same year, an alternative assessment (Davies et al., 2008) was also conducted using CASAL (Bull et al., 2012), focusing on the south-central Pacific area alone (140°E – 130°W). The 2008 CASAL assessment produced quite different outcomes to the 2008 MULTIFAN-CL assessment, and concluded there was insufficient information available to estimate population abundance. Consequently, the status of the swordfish stocks in the south-central Pacific Ocean could not be determined from this assessment.

In 2013, a new assessment was conducted using MULTIFAN-CL (Davies et al., 2013), which assumed two model regions, separated at 165°E in the WCPFC area south of the equator (140°E – 130°W). This spatial structure was based upon the results of electronic tagging programs (Evans et al., 2012). The conclusions from this assessment depended on the assumed growth curve; Australian or Hawai'ian. It was concluded that under the Hawai'ian growth overfishing was not occurring, while under the Australian growth over overfishing was occurring, but under either growth scenario the stock was not considered to be overfished. In 2017, an assessment was conducted in MULTIFAN-CL (Takeuchi et al., 2017), using the same two region spatial structure used in the 2013 assessment. This assessment concluded that it was highly likely that the stock was not in an overfished condition and that the stock was not experiencing overfishing. Follow up work was undertaken in 2018 to develop a two-sex model (Takeuchi et al., 2018). However, this model was not used to provide management advice.

The most recent assessment was conducted using MULTIFAN-CL in 2021 (Ducharme-Barth et al., 2021a). This assessment employed a novel approach to characterising uncertainty in management quantities by using an ensemble combination of models capturing model (structural) and estimation (statistical) uncertainty (Ducharme-Barth and Vincent, 2022). This assessment concluded that the stock was unlikely to be experiencing overfishing and unlikely to be in an overfished state.

2.3 2025 assessment in Stock Synthesis

As with previous assessments, the objectives of the 2025 SWPO swordfish assessment are to estimate population parameters, such as time series of recruitment, spawning potential, spawning potential depletion and fishing mortality. These model estimated quantities are used to assess the stock status and impacts of fishing. There are no formal reference points adopted by the WCPFC for SWPO swordfish and the stock status and fishing impacts are summarised in terms of WCPFC default MSY based reference points, F/F_{MSY} and SB/SB_{MSY} (Table 3). Additional stock status indicators are provided including spawning biomass (SB), summary biomass, and depletion relative to dynamic unfished levels, $SB/SB_{F=0}$, (Table 3). The methodology used for the assessment is based on the integrated assessment framework (Fournier and Archibald, 1982; Maunder and Punt, 2013), and is conducted using Stock Synthesis (Methot and Wetzel, 2013). Model parameters are estimated by maximising an objective function, consisting of both a likelihood component (based on fits to data) and penalties constraining the estimates of certain parameters.

The transition of the 2025 assessment from MULTIFAN-CL to Stock Synthesis is not presented as a typical stepwise development to produce a new benchmark assessment by building from the previous assessment, typically using the same software. While the development of an initial model in Stock Synthesis drew on the structure and assumptions of the previous MULTIFAN-CL assessment, there was no attempt to replicate the 2021 assessment results from the MULTIFAN-CL model using Stock Synthesis. This was due to some initial changes to the fishery structure, significant downweighting of the length-based composition data, a change from age-based selectivity to length-based selectivity,

and a change from using a spline shape for selectivity. The 2025 assessment should be considered as the development of a new benchmark assessment in Stock Synthesis, with the freedom to take advantage of features available in Stock Synthesis to optimize the model structure and settings, and the fit to the data. This assessment also makes a fundamental structural change compared to the previous assessment by moving to a two-sex model, as recommend by previous assessments and supported through the 2025 SPC Pre-assessment workshop (PAW) (Hamer, 2025).

The 2025 SWPO swordfish assessment incorporates data up until and including 2023 from longline fisheries across the SWPO model region (Figure 1). Preparatory work on data for the assessment is extensive and is described in limited detail in this report. We advise that this assessment report is read in conjunction with several supporting papers:

- Analysis on longline size frequency data (Peatman et al., 2025)
- Analysis of New Zealand longline CPUE (Finucci and Moore, 2025)
- Analysis of Australian longline CPUE (Tremblay-Boyer and Williams, 2024)
- Analysis of Spanish longline fishery CPUE (Kim et al., 2025)
- Analysis of longline CPUE from Pacific Islands and Territories (PICTs) observer program data (Neubauer, 2025)
- SPC Pre-assessment workshop summary report (Hamer, 2025)

3 Background

3.1 Stock structure and movement

Swordfish are a highly mobile fish occurring in tropical and temperate waters in all global oceans and large seas. In the Pacific Ocean they are one of the most widely distributed pelagic species found from 50°N to 50°S and at all longitudes (Moore, 2020). Longline catch rate distributions suggest three large, relatively high density areas, the North-West, South-West and Eastern Pacific, which is in contrast to spawning distributions inferred from larval surveys, (Nishikawa et al., 1985; Ijima and Jusup, 2023) and studies of spawning and reproduction (Young and Drake, 2002; Mejuto et al., 2008). These studies suggest spawning occurs in more localized regions predominantly in tropical and sub-tropical areas. In the southwest Pacific, larval distributions suggest spawning occurs in latitudes between around 5°S to 30°S, with likely key spawning areas in the Coral Sea, and around Tonga and French Polynesia (Ijima and Jusup, 2023). Tagging studies provide some confirmation that swordfish undergo directed seasonal migrations between temperate foraging grounds and tropical spawning grounds, with perhaps less longitudinal movement (summarised in Moore (2020)). It remains unclear how much site fidelity or philopatric behaviour occurs, although it has been demonstrated for some satellite tagged fish (Tracey and Pepperell, 2018). The degree to which individuals migrate and sub-populations mix has implications for fisheries management and stock assessment (i.e. localized depletion risk), but stock and meta-population structure in the South Pacific still remains poorly understood.

Both Evans et al. (2014) and Patterson et al. (2021) analysed movement data from swordfish tagged with pop-up satellite tags (PSATs) in the waters off eastern Australia, Cook Islands, and northeast of New Zealand between Fiji and French Polynesia, as well as northern New Zealand (Holdsworth et al., 2007), and northern Chile (Abascal et al., 2010). The analysis suggested a lack of movement between the southern and northern regions of the WCPO, and between the WCPO and the far eastern EPO. Evans et al. (2014) also identified cyclical latitudinal movements of fish tagged at multiple locations, but found limited movement between the eastern and western areas of the Tasman and Coral Seas, delineated at approximately 165°E. This study led to the recommendation of the two model regions east and west of 165°E that have been adopted since the 2013 assessment (Davies et al., 2013). More recent pop-off satellite tagging (PSAT) of larger swordfish off the east coast of Tasmania however showed that some fish do move large distances from west to east across the 165°E regional boundary (Tracey and Pepperell, 2018). It was suggested that the difference between the observed movements in the two studies may relate to the size of tagged fish, with the smaller fish in the Evans et al. (2014) study showing more restricted movements.

PSAT tagging studies however, typically involve very few fish and do not capture long-periods of fish life-cycle, due to the deployment duration of the tags. They also provide no information on movement or dispersal of juveniles and earlier life stages, and are generally inconclusive in providing information on biological stock structure or metapopulation structure. Other approaches

to studying stock structure; such as genetics, otolith chemistry, parasites and tissue stable isotopes have all been applied to swordfish in the Pacific region and more broadly (reviewed in Moore (2020)). Most studies applying these approaches have had limitations in either the method and or the sample coverage, although there is now strong evidence from genetics that populations in the Indian and Pacific Oceans are separated (Grewe et al., 2020). For the Pacific Ocean, Reeb et al. (2000) suggest a broad ⊃ shaped connectivity pattern, such that the south west and north west Pacific populations are the most genetically distinct from each other, with central and eastern populations intermediate between the two. Alvarado Bremer et al. (2006) suggested that the south east Pacific population was also genetically distinct from the north east and south west populations, and there was additional evidence to suggest that the south-central Pacific may represent a population intermediate between the southwest and southeast, but this was inconclusive due to the low sample sizes. More recent genetic studies have, however, found less evidence for population structure across the Pacific (Lu et al., 2016). Overall, the stock structure remains poorly understood for swordfish, movement behaviour appears to be complex extensive and to cross management boundaries (Sepulveda and Aalbers, 2025), and likely differs by sex and size, and there is evidence of fidelity to particular areas (Burnett et al., 1987; Tracey and Pepperell, 2018; Sepulveda and Aalbers, 2025). As such there is likely to be sub-structure and migratory contingents within a more broadly resolved stock structure. More research is required in this area to understand the implications of spatio-temporal movement dynamics and regional fidelity for stock assessment spatial structure.

3.2 Vertical movement behaviour

An important aspect of swordfish behaviour for understanding their ecology and how they interact with fishing gear is vertical movement. Swordfish show diurnal vertical migration, moving up to the surface waters at night and returning to mesopelagic depths, typically deeper than 300 m, during the day (Abascal et al., 2010; Dewar et al., 2011). This typical behaviour may be modified by spending more time closer to the surface during the day where dissolved oxygen levels are lower, or at night during the new moon phase when lunar illumination is low (Abascal et al., 2010). This vertical migration is related to feeding behaviour as swordfish follow the diurnal migrations of prey from the deep scattering layer, primarily squid that are in-turn feeding on the deep scattering layer fish community, to the surface at night. Targeting of swordfish with longlines can therefore be more effective at night by setting lines closer to the surface, often with the use of light sticks. Gear setting practices, especially factors affecting depth of fishing (hooks between floats, branch line length, float line length, and/or line setting speed), bait type (e.g. squid v fish), and use of light sticks can be influential on swordfish catch rates (Dell et al., 2020). Environmental factors such as degree of lunar illumination or proximity to oceanographic fronts and convergence zones have also been shown to influence swordfish catch rates (Bigelow et al., 1999; Poisson et al., 2010).

3.3 Biological characteristics

Swordfish are sexually dimorphic, with females growing larger and faster than males (Young and Drake, 2002; Mejuto and García-Cortés, 2014; Moore, 2020). Potential sexual differences in other life history characteristics are less well known (e.g. migration patterns, natural mortality, etc.). There have been a number of studies on swordfish growth rates and maturity in the Pacific (DeMartini et al., 2000; Young et al., 2003; Mejuto et al., 2008; Valeiras et al., 2008) providing a range of estimates for these key biological parameters that have contributed to stock assessment uncertainty. WCPFC SC recommended that additional work on age, growth, age validation, and reproductive biology be undertaken for the southwest Pacific swordfish stock (WCPFC Project 71). That research and its results are described in Farley et al. (2016), which indicated that swordfish lived longer and grew slower than previously estimated. The conditional age-at-length dataset from the Farley et al. (2016) study is used in this assessment, and there has been no addition of conditional age-at-length data since then. The key findings of Farley et al. (2016) include: southwest Pacific swordfish can live for at least 21 years, male and female growth is similar until about age 3 years after which females reach larger sizes than males for the same age, the size at 50% maturity for females is estimated to be around 160 cm eye orbital fork length, which is around 175 cm lower jaw-fork length and age at 50% maturity occurs at around 4–5 years. The spawning period in the southwest Pacific is from August to May with a peak in December–January. Natural mortality is discussed further later in this paper, but the review by Moore (2020) shows that annual M used in previous assessments in the Pacific Ocean range from 0.22–0.48 depending on age and sex.

3.4 Fisheries

Due to the diurnal vertical migration behaviour of swordfish, they are predominantly caught using longline gear, with targeted fishing closer to the surface at night, or as bycatch in deeper longline sets during the day, that are likely to be targeting bigeye tuna. Squid bait is preferred over fish baits and attaching light sticks periodically along the longline is a common tactic employed to increase targeted catch rates (Bigelow et al., 1999; Dell et al., 2020).

Fisheries for swordfish have operated in the SWPO since the early 1950s. Historically, the majority of swordfish catches represented a valuable bycatch from the tuna-target longline fisheries. While some of the recent catches are still considered valuable bycatch, particularly in deeper set longlines, targeted catches slowly increased in the SWPO from the early 1970s up until the late 2000s-early 2010s, when catches peaked at around 9,000–10,000 t (Figure 4, Figure 5). Early catches were primarily taken by Japanese fisheries, but from the mid-1990s to the early 2000s catch levels of other nations increased rapidly, as more targeted fishing of swordfish in Australian and New Zealand waters developed, influenced by the introduction of quotas for the species in these jurisdictions. From the mid-2000s, catches by the European Union (Spanish) fishery increased as this fishery started targeting swordfish in the south-central Pacific Ocean (Figure 4, Figure 5, Figure 6). Since 2014, the catch of swordfish has declined across all fisheries in the SWPO, with recent estimates

indicating an annual catch of between 5,000 and 6,000 t (Figure 4).

Since 2000, around 50% of the swordfish catch in the SWPO has been taken in the tropical and sub-tropical waters (Figure 6), in sub-regions 2N, 1C and 2C. In recent years the major catches have occurred: off the central east coast of Australia by the Australian fishery, as targeted catch; around northern New Zealand, and to the northeast of New Zealand by the Spanish fishery, as targeted catch in the high seas; and in the northeastern tropical area, subregion 2N, primarily by vessels from Chinese Taipei, as bycatch, while targeting bigeye tuna (Figure 10).

4 Data and methods

4.1 The data and model inputs

The 2025 stock assessment of SWPO swordfish uses an age- and size-structured model implemented in the generalized stock assessment software package, Stock Synthesis (Version 3.30.23.1, Methot and Wetzel (2013); Methot et al. (2025)).

The methods utilised in Stock Synthesis are based on the integrated analysis paradigm. Stock Synthesis can incorporate multiple seasons, multiple areas and multiple fisheries, although most applications are based on a single season and area. Recruitment is governed by a stochastic Beverton-Holt stock-recruitment relationship, parameterised in terms of the steepness of the stock-recruitment function (h), the expected average recruitment in an unfished population (R_0), and the degree of variability about the stock-recruitment relationship (σ_R). Stock Synthesis allows the user to choose among a large number of age- and length-specific selectivity patterns. The values for the parameters of Stock Synthesis are estimated by fitting to data on catches, catch-rates, catch length-frequencies, catch weight-frequencies, and conditional age-at-length data. The population dynamics model and the statistical approach used in fitting the model to the various data types are given in the Stock Synthesis technical documentation (Methot et al., 2025).

The diagnostic model includes two regions, with fixed quarterly movement between these regions, assuming a single reproductive stock, with 19 fisheries (Table 1), each operating in one of 6 sub-regions. Selectivity is modelled separately for fisheries with length composition data and weight composition data, to avoid any potential conflict induced by potentially inaccurate conversion factors between lengths and weights. Selectivity is estimated for a restricted set of the 19 fisheries, with these estimated selectivities shared amongst groups of fisheries. For some fisheries, with less consistent and reliable sampling, selectivity is initially estimated, following some initial filtering of this composition data. For these fisheries, selectivity is subsequently fixed at the estimated values, and the composition data is then further downweighted, as it is not considered to be reliable enough to influence estimated population processes, such as population scale and recruitment. Selectivity patterns are assumed to be length-specific for all fisheries. Selectivity is assumed to be logistic for the fisheries operating in the two southernmost sub-regions and double normal for all other fisheries. The model accounts for males and females separately by fitting separate growth curves for each sex. The size composition data is not reported separately by sex, so these are all treated as aggregated data. The models uses a quarterly time step, and most data sources are entered in quarterly intervals (catch, size composition data, age data), although notably the CPUE data is entered with an annual time step. Recruitment is assumed to occur only in the first quarter of every year, as this is most likely to be seasonal, but movement between the two regions is assumed to have the same movement rate in each quarter. The initial and final years are 1952 and 2023.

4.1.1 Biological parameters

Growth In contrast to the 2021 swordfish assessment (Ducharme-Barth et al., 2021a), a two-sex model is used for the diagnostic model in this assessment, using conditional age-at-length data as the sole data source available which is separated into male and female components. This allows growth to be modelled separately for male and female fish, which is important given that it is known that female swordfish grow much larger than males, and live to an older maximum age.

Natural mortality Natural mortality (M) was modelled using a two-step approach following (Hoyle, 2022). First, a target level of instantaneous natural mortality was estimated using the oldest fish observed in the population ($A_{\max}$), based on methods developed by (Hamel and Cope, 2022) that improve upon earlier approaches (Then et al., 2015). Using a maximum observed age of 21.15 years from 300 aged otoliths, the target mortality was calculated as $5.4/A_{\max}$.

Second, this target mortality was distributed across age classes using the Lorenzen functional form (Lorenzen et al., 2022), which describes the inverse relationship between natural mortality and body size. The Lorenzen curve was converted to a function of age using the growth equation, then rescaled so that the average mortality across all age classes, weighted by the proportion mature at each age, equals the target mortality:

$$M_t = M_{\text{target}} \frac{n\lambda_t}{\sum_{i=t_0}^{t_{\max}} p_i \lambda_i}$$

where t is age, λ_t is the Lorenzen mortality at age t, p is the proportion mature at age t, $t_{\max}$ is the oldest age class, and n is the number of age classes. This weighting ensures that the average instantaneous mortality rate for mature fish equals the target mortality.

To produce sex-specific mortality curves, given minimal sexual dimorphism in growth at young ages, instantaneous mortality at age one was assumed equal between sexes, consistent with the expectation that size-dependent mortality processes (e.g., predation) would affect both sexes similarly at this life stage. For older ages, sex-specific mortalities were calculated using the Lorenzen functional form, while maintaining the target average mortality for mature fish across both sexes combined.

To incorporate this form of mortality within Stock Synthesis, natural mortality was implemented using the Lorenzen mortality option (option 2), with female mortality fixed at age seven (when 98% of females are mature) to anchor the Lorenzen curve. Male mortality was calculated using growth parameter offsets while maintaining the target average mortality for mature fish across both sexes. For sensitivity analyses, alternative mortality levels were explored using the 25^{th} and 75^{th} percentiles from a lognormal distribution centered on the target value, with a coefficient of variation of 0.31 (Hamel and Cope, 2022).

In the diagnostic model, sufficient information was available to estimate the key mortality parameter

directly, eliminating the need for the rescaling procedure described above. When mortality was estimated, a lognormal prior was used, centred on the target mortality with a coefficient of variation of 0.31 (Hamel and Cope, 2022).

Length-weight relationship The parameters of the length-weight relationship are obtained from Macdonald et al. (2025). These parameters have been revised from the length weight relationship used in the 2021 assessment, and are appropriately bias corrected and also derived from an updated dataset with additional data sources and data quality control, using only data that were measured using lower-jaw fork lengths and whole weights to obtain standard length and weight measurements and avoid possibly confounding effects from potentially unreliable conversion factors Macdonald et al. (2025). The length-weight parameters for females are $a = 6.565533 \times 10^{-6}$ and $b = 3.139509$ and for males are $a = 8.942387 \times 10^{-6}$ and $b = 3.063015$.

Sexual maturity Swordfish are assumed to be sexually mature at around 175 cm (lower-jaw fork length) or 4-5 years of age (Farley et al., 2016), using the same age-maturity matrix as used in the 2021 assessment (Ducharme-Barth et al., 2021a). Fecundity-at-length is assumed to be proportional to weight-at-length.

4.1.2 Spatial stratification

The spatial structure used in the 2025 SWPO swordfish assessment is identical to the structure used in the 2017 (Takeuchi et al., 2017) and 2021 (Ducharme-Barth et al., 2021a) assessments, and covers the southwest Pacific Ocean from 140°E to 130°W, and from the equator to 50°S, including the area of overlap between the WCPFC and IATTC convention areas (Figure 1). The delineation of these two regions at 165°E followed Davies et al. (2013), based on the tagging analysis by Evans et al. (2012, 2014). Movement between region 1 and region 2 is explicitly modelled in this assessment (Patterson et al., 2021), with any spatial structure within regions treated in an areas-as-fleets manner, with catches and composition data from each fishery fully separated by subregion for each fishery, in a structural change from the assumptions made in the 2021 assessment. This is particularly important for the southern fisheries where it is known that large fish are caught. In particular, the New Zealand fishery clearly catches larger fish south of 40°S, generally caught off the west coast of the South Island, compared to the New Zealand fishery operating further north, which generally catch fish offshore from the North Island. The 2021 assessment treated the New Zealand fishery as one fishery, and did not separate these length composition data by latitude. The separation of all fisheries into distinct subregional fisheries with respect to catches and size composition data allocated to a single subregion, allows for any latitudinal changes in size distribution and selectivity to be more accurately modelled.

Recent increases in effort and catch close to the north eastern boundaries of subregion 2N, especially in the last 25 years, and the size of catches reported near to, but on the other side of, this boundary suggest that the stock structure and spatial boundaries are unlikely to be biological boundaries. The

variation in fishery behaviour and the vast spatial scale of region 2, and the operation of fisheries from different flags across the current subregional boundaries suggest that this spatial structure is in need of a comprehensive review. However, to do this thoroughly and comprehensively, seeking input from all interested parties was beyond the scope of this assessment cycle. The current structure has been used for several previous assessment, and appears to have been based on recommendations taking account of known tagging movement, and apparent wide mixing within region 2, but failing to account for other important considerations in a stock assessment context, including political boundaries, and the fleet structure, fishing behaviour and operations of fisheries from different flags. The current spatial structure combines at least three disparate and significant fisheries with different operating characteristics in region 2, all with an increase in catches starting in the mid 1990s. The New Zealand fishery is essentially a domestic fishery operating near New Zealand in the south west, the European (largely Spanish) fishery operates broadly in central areas of region 2, with large spatial and temporal changes in areas of focus from year to year, and with operations apparently targeting swordfish (Kim et al., 2020). Finally, there is a third mixed flag fishery, considered to be largely a bycatch fishery, operated by distant water fishing nations in the far north eastern corner of region 2. To fully restructure region 2 into smaller regions, rather than subregions, would require some information on movement between any new regions, which is not currently available, and possibly some additional biological information, or else some speculation or assumptions on movement and any biological differences. While there is some information on movement between regions 1 and 2 (Patterson et al., 2021), this is based on a relatively small number of satellite tag releases of fish with restricted age ranges, the current movement assumptions obtained from this satellite tag data, could be strengthened with more data collected, over a wider age range of satellite tagged fish, and with additional tagging studies. The length composition data over the full area of the assessment is not consistent enough in time or space to give reliable information on population structure from a regression tree analysis, an approach which is often used to provide information on spatial structure to be used in an assessment.

4.1.3 Fisheries

Extraction fisheries The number of extraction fisheries has increased from 13 fisheries in the 2021 assessment, to 19 fisheries in the 2025 assessment. This increase was required to ensure that all fisheries are now separated into northern, central and southern components, where appropriate, with each fishery now covering only a single subregion Table 1. Only some fisheries were separated by subregion in the 2021 assessment, namely the distant water fishing nations in both regions 1 and 2, and for the PICT fisheries in region 2.

These 19 fisheries can be classified as fisheries operated by one of the following groups: distant water fishing nations (DWFN); Pacific Island Countries and Territories (PICTs); Australia (AU); New Zealand NZ); and the European Union (EU, but predominantly Spanish vessels). For this assessment, vessels fishing under the Vanuatu flag were grouped with the DWFNs as they are likely to be vessels from Chinese Taipei fishing under special access arrangements in the Vanuatu EEZ.

Index fisheries and CPUE series In addition to the extraction fisheries, there are two index fisheries in the diagnostic model, with associated CPUE series, with only one index per region in the diagnostic model. The index fishery for region 1 is based on the Australian CPUE series (Tremblay-Boyer and Williams, 2024) and for region 2 is based on the New Zealand CPUE series (Finucci and Moore, 2025). Neither of these CPUE indices are based on a spatio-temporal CPUE analysis based on VAST or sdmTMB, as used in recent WCPFC tuna assessments, so it was not possible to create separate index size composition data, weighted by the index instead of the catch data, as is the usual practice for index fisheries (Peatman et al., 2025). Instead, the selectivity for these fisheries are shared with the selectivities estimated for the equivalent extraction fisheries, which are catch weighted rather than index weighted, from the appropriate region corresponding to where the CPUE data were collected: fishery 05.AU.1C for the Australian index fishery (20.IDX.AU.1) and fishery 13.NZ.2C for the New Zealand index fishery (21.IDX.NZ.2).

The Australian CPUE index was standardised using General Additive Models and data from 1998–2023, and using operational datasets including use of light sticks and bait type (Tremblay-Boyer and Williams, 2024). The annual index was chosen for all size classes, for use in the 2025 swordfish stock assessment, so that the assessment model did not have to attempt to fit to seasonal variation in the CPUE, which may reflect seasonal changes in selectivity, availability or fishing practices, rather than seasonal changes in abundance.

The NZ CPUE index was standardised using General Additive Models and data from 2004–2023, and also using operational datasets including use of light sticks and bait type (Finucci and Moore, 2025). The annual index was chosen for the core set of experienced vessels, for use in the 2025 swordfish stock assessment.

The two CPUE abundance indices used in the diagnostic model, from Australia and New Zealand, are plotted on the same normalised scale for easy comparison in Figure 2. While there is a roughly similar pattern, possibly showing some similar periodicity in both indices, the New Zealand index shows larger absolute changes than the Australian index.

Two alternative index fisheries were considered for region 2, for use in sensitivities and in the uncertainty grid, with the New Zealand index being replaced by either an EU based CPUE index (Kim et al., 2025) or by a PICT observer index (Neubauer, 2025). As the PICT observer index was not a spatio-temporal analysis, this index fishery did not have separate index composition data and the selectivity was shared with the shared selectivities from the extraction fisheries from subregion 2C with shared selectivity (11.DW.2C, 16.EU.2C and 19.PICT.2C). While there was a spatio-temporal index for the EU index, there were some concerns about the validity of that index, so again this index fishery simply shared selectivity with the same extraction fisheries from subregion 2C with shared selectivity (11.DW.2C, 16.EU.2C and 19.PICT.2C).

The annual PICT observer index was standardised using GLMM and data from 2001–2022 (Neubauer, 2025), with spatio-temporal splines, oceanographic predictors (NINA4) and operational covariates

including gear characteristics. The PICT series including data from New Caledonia, Fiji, Tonga and French Polynesia to get an index that was most likely to be representative of the whole of region 2. While a small part of this data from New Caledonia was from fishing operations in region 1, this was overlooked for simplicity, and the index, including that small quantity of data from region 1, was assigned to region 2 instead.

The annual EU index was standardised using VAST (Kim et al., 2025), with a similar approach as used in the 2021 assessment (Ducharme-Barth et al., 2021b). This analysis showed a very sharp decline in CPUE in the last two years. This decline looked unrealistic and closer inspection showed it was associated with a shift in fishing operations to a more southerly region and a substantial increase in blue shark catches. As such, the last two years of the time series were exclude, over the period when the blue shark catches increased, as these were unlikely to be indicative of swordfish abundance.

The coefficient of variance (CV) was arbitrarily set to 0.2, for the whole time series, for the Australian index and to 0.1 for both the NZ and PICT indices, but set to the values obtained from the CPUE analysis for each year for the EU index, with values ranging from 0.9–0.23 in this case. Often the estimated CV is unrealistically small, for stock assessment purposes, due to the apparent statistical power from using a large number of data points used in CPUE analysis. Using these CVs fails to account for other sources of uncertainty and can place too much weight on a CPUE series. A range of alternatives approaches are possible, including fitting some smoother (e.g. loess, or by using a fit to an earlier model) and estimating the CV from the root mean square error taken from the smoother or fit chosen. This typically gives a more realistic starting CV, and is often set to a fixed value for all points in the CPUE series, making the assumption that there is no time variation in the uncertainty for the index. However, it is good practice to estimate an additional CV parameter for each index, which is a standard option available in Stock Synthesis, to effectively balance, or re-weight, the input and output variances internally. This approach of estimating additional variance on the CPUE was adopted for all CPUE indices in this assessment.

While the practice adopted here, in starting from an arbitrary CV for each index, could (and should) be questioned, there were many other more important dragons to try to fight in this assessment, and this detail ended up being rather lost in a storm, while those dragons were being fought. That said, additional CV was estimated internally for each CPUE index in this assessment, from these arbitrary starting points, so hopefully this is one of several "lesser evils" present in this assessment.

Other indices used in the 2021 stock assessment, using data from Japanese and Chinese Taipei fisheries, were not incorporated into this assessment, partly due to constraints on time and resources required to reanalyse these series, but also because these series were noisy and largely flat, came from fisheries that were largely bycatch fisheries and appeared to contain little signal in previous assessments. Index fisheries were not created for these indices in 2025.

The three alternative standardised CPUE indices for region 2 (NZ, PICT and EU), are shown in

the same plot in Figure 3, with each CPUE series shown using a different coloured circle, NZ (blue), PICT (red) and EU (green). The fits to these indices are also shown, from sensitivities which were conducted later, all using the AU index for region 1, but with only a single index from region 2 for each alternative CPUE option. The EU index was truncated at the end of 2021, due to the concerns noted above, which could potentially represent a change in fishing practices for that fishery, rather than a change in abundance for the stock (Kim et al., 2025).

4.1.4 Catch data

The model uses year and quarter for all catch data. All EU fishery catch data is reported in tonnes and all other fisheries report catch in number of fish caught (Table 1). These catch data are aggregated at 5° × 5° cells and year-quarter resolution, with the aggregation process either conducted by SPC, where operational data is available to inform this, or by the particular countries, following statistical procedures that are reported to the WCPFC. Landed catches, aggregated up to annual totals by flag and converted to tonnes where necessary, are shown in Figure 4, with the same catches separated by region in Figure 5 and by subregion in Figure 6. These same aggregated annual catches, from both regions combined and converted to numbers of fish rather than tonnes, are shown by fishery in Figure 7 and in stacked form in Figure 8.

The spatial distribution of catches across the model region are shown in Figure 9, aggregated by 5° × 5° cells. This spatial catch data by major flag, summed over all time periods, shows some strong patterns amongst the flags.

Some flags feature quite limited spatial distribution of catches, including the New Zealand fishery, and also the Australian and Japanese fisheries, operating off the east coast of Australia, albeit during different time periods. The New Zealand catch, largely taken in NZ territorial waters, demonstrates that a smaller proportion of the total New Zealand catch is taken south of 40°S (Figure 6). The catch from the PICT fisheries is relatively small, but is spread over a wide region, mostly either side of the boundary between subregions 2N and 2C. The European fishery (mostly Spanish vessels) is widely spread throughout region 2C, with quite large catches. There is a band of 5° × 5° cells with relatively small catches, separating the areas fished heavily by the European fishery in subregion 2C and the north eastern part of subregion 2N, which has been fished particularly heavily in recent years, mostly by distant water fishing nations such as China, Korea and Chinese Taipei. The spatial distribution of catches, show large variation within regions and subregions, with distinctive patterns showing areas of high and low catches (Figure 9).

The temporal patterns in the spatial distribution of catches across the model region are shown in Figure 10, again aggregated by 5° × 5° cells. This spatio-temporal catch data, by major flag, shows additional information as to the development of various fisheries over time, with catches up until 2000 dominated by the Japanese fisheries operating over a broad area, and the Korean fisheries also operating over a broad scale in subregion 2N. The Australian and New Zealand fisheries rapidly developed in the late 1990s and became locally dominant in their EEZs since then. The European

fisheries operations increased from around 2000, especially over a wide range of subregion 2C that was previously very lightly fished. The catches in the north eastern part of subregion 2N have increased substantially since 2000, initially with the Chinese fishery becoming active, but also with fisheries from Korea increasing their fishing and more recently from Chinese Taipei. This catch increase in subregion 2N is particularly concentrated in the north east of this subregion, but it continues beyond the north eastern boundary of subregion 2, with high catches from Chinese Taipei, Korea, China and other distant water fishing nations, outside the regional boundaries used for this assessment.

4.1.5 Size composition data

Size composition data from each fishery is comprised of quarterly samples, with frequency distributions of either length samples or weight samples, and for some fisheries, sometimes both length and weight samples. Sampling is sometimes conducted by observers on board vessels and sometimes in port or at fish markets, and different measurement standards for lengths and fish processing standards are applied in different jurisdictions and by different authorities (Hamer et al., 2025).

For most fisheries, temporal coverage of the size frequency data is limited or patchy (Peatman et al., 2025). Within some fisheries, the contribution from each flag can also be patchy and vary temporally, and also spatially, within a subregion. Variation in targeting practices between separate sampling events may also lead to inconsistencies. Examination of the raw size composition data for several fisheries indicated greater variation than would be expected for a representative sampling program, so this prompted considerable investigation of the size composition data, and of conversion factors.

Measurement standards differ between sampling programs and flags, so all length composition data were converted to lower jaw-fork length (LJFL) using the best available conversion factors (Hamer et al., 2025). Similarly, weights were converted to whole weights, from a range of varying weight measurement standards or fish processing states.

Length data were provided based on three main length measurement methods: eye orbital–fork length, lower jaw–fork length, or pelvic fin–fork length as is standard practice for billfish measurements (Hamer et al., 2025). Weight data is provided in various forms depending on the source, for example, Japanese processed weights (gilled, gutted, head and tail left on, fins trimmed, bill removed at a few cm in from the tip of the lower jaw), and gutted and headed (Australia), or processed weights (New Zealand) (Hamer et al., 2025). All length measurements were converted to LJFL, and weight measurements were converted to the equivalent whole (unprocessed) weight using the conversion factors listed in Hamer et al. (2025). While efforts were made to ensure the conversion factors used were the best available, there still appears to be some uncertainty associated with some conversion factors. This may help to explain inconsistencies seen when length and weight data are used from the same fishery or from the same subregion. These two data types can show conflict in the model, where the model can overfit one of these data sources and simultaneously

underfit the other. While it was not possible to resolve all of these problems, this highlighted a need to review the standard length and weight measurements used over the range of the fishery, and the conversion factors that are required, which may also need to vary regionally. This should perhaps constitute a separate project, requiring further work and resources, if it is desirable to improve the quality of the size composition data used in subsequent swordfish and striped marlin stock assessments. Standardising the length composition data, especially for the mixed flag fisheries, to appropriately counter any spatial or temporal variation in either fishing or sampling practices, may provide more useful length composition data for use in future assessments.

Port based and observer length composition data were separated this year for the PICT fisheries (Peatman et al., 2025), and this revealed clear differences in the distributions from these two distinct data sources. These two components of the length composition data have simply been aggregated in previous billfish assessments, so this difference, and the conflict apparent between these two sampling protocols, has not been known or demonstrated to exist previously. A decision was made to use the observer length composition data for these fisheries, and to exclude, or filter out, the port or market based observations, as the observer data was considered to be the more reliable data source, in terms of representing the catches.

Another feature of size composition data for a number of swordfish fisheries was the appearance of small fish, sampled in certain time periods (quarters), in a manner that appeared to be episodic, unpredictable, transient and irregular, with little to no consistent evidence that these represented large cohorts appearing in the population. The flow through, that you would expect from a large cohort as it ages and grows, and as the population continues to be fished, was not consistently observed, and often was notable in its complete absence. This suggest that this sampling of occasional spikes of small fish may not be representative of the population as a whole, and may be an artefact of fishing practices, such as changes in location, or targeting or simply the composition of flags conducting sampling within a fishery. In any case, this size composition data appears to have limited information to provide to the stock assessment about population processes. To address some of these issues, and other problems with the size composition data, additional filtering steps were considered and implemented for all fisheries for swordfish this year (Peatman et al., 2025).

Detailed information on the size data by fisheries is available in Peatman et al. (2025). Size data were aggregated by fishery and year-quarter. For each year-quarter, length data were combined into 29 equal width 10 cm size classes, or bins, ranging from 30–310 cm, with all fish larger than 310 cm placed in the last bin. Traditionally, weight data has also used equal sized bin widths in swordfish assessments. However, given that the relationship between length and weight is roughly cubic, this gives a mismatch between the proportions of samples in the lower weight bins, compared to those in the lower length bins. Lower weight bins typically are filled with samples for small fish which would be equivalent to several length bins, had these same fish been recorded with length measurements rather than weights, giving little resolution to any potential recruitment signal that may be observed in small fish. One weight bin in the smallest size range is equivalent to multiple

length bins for the same fish. For large fish, which typically feature many weight bins with very few fish in them, producing a very long right hand tail, the opposite is true, with multiple weight bins "fitting" into a single length bin, for the largest of fish, giving a very high resolution for the largest weight-measured fish, compared to those large fish with lengths measured.

To address this mismatch, a variable bin width was applied to the weight composition data. This started with a bin width of 1 kg, for the lightest fish measured (3 kg), then transitioning to a 2 kg bin width, starting at 12 kg fish, transitioning again to a 5 kg bin width at 20 kg fish, to a 10 kg bin width at 40 kg fish, to a 20 kg bin width at 100 kg fish, to a 40 kg bin width at 160 kg, to a 50 kg bin width at 200 kg and then finally with a plus group for any fish heavier than 350 kg. This resulted in a total of 30 weight bins, compared to 29 length bins, covering the same sized fish. The 2012 swordfish assessment used 63 weight bins, all with a constant 5 kg bin width. MULTIFAN-CL does not currently have the option to incorporate variable size bin widths, so it was not possible to better balance the number of weight bins with the length bins in a MULTIFAN-CL assessment. The use of variable weight bin widths allows for resolution of the data to give information on the smaller fish that can help inform the recruitment estimates in the model. The number of length bins, and the length bin width, used in the 2025 assessment was unchanged from the length bin definitions used in the 2021 assessment.

Through the course of the development of the 2025 assessment, a range of approaches to the use of the size composition data were investigated, and some subjective decisions were eventually made to completely exclude length composition data from some fisheries, and to heavily downweight others. Despite multiple explorations and approaches tested, it was not possible to get stable model behaviour when these data sources were allowed to have too much influence on the model. The weight data were generally considered to be more reliable than the length data, and the Australian length composition data was excluded for that reason, given there was also weight composition data available for that fishery. For the New Zealand composition data, which has both length and weight data for fisheries in both subregions, the weight data was chosen instead of the length composition data for fishery 13.NS.2C, partly because the data collection was also ongoing for the weight data, and the ongoing collection of length composition data, collected by onboard observers, has been discontinued in New Zealand. In contrast, for the southern New Zealand fishery 14.NZ.2S, the length data was used instead of the weight data, as there were more length samples per quarter typically, and these were collected over a longer time series than the weight data. This was considered an important fishery, given the propensity to catch larger fish and the asymptotic selectivity assumed for this fishery, and the historical length data was considered to be more informative to the assessment. This was considered more important than the fact that observer based length data is no longer being collected in this part of the fishery. The lack of ongoing collection of length composition data for this southern component of this New Zealand fishery will need to be carefully considered for future assessments.

Ideally the size composition data would be modelled to essentially standardise this data to allow for

factors such as the composition of flags from which it was collected, the finer scale spatial location of fishing operations within these large subregions, any changes to targeting practices and potentially other factors that could be useful to smooth out this data source, following approaches suggested by Maunder et al. (2020).

4.1.6 Conditional age-at-length data

In another change from the 2021 assessment, conditional age-at-length data by sex was incorporated in the 2025 assessment. These samples were collected from the Australian fishery, 05.AU.1C, between October 1999 and November 2002 (unpublished data provided by Jess Farley, CSIRO, Hobart) and the otoliths were read to give fractional ages (Farley et al., 2016). These data were provided with lengths measured as eye-orbital-fork length rather than LJFL, the standard length measurement used in the swordfish assessment. These length data were converted to LJFL using the standard SPC held conversion factors. Again we suggest review and improvements of length to length conversion factors, so the conversion factor used to convert these to LJFL is appropriate for this critically important data source. Ideally, a consistent standard for measuring the length of swordfish would be agreed and adopted in the region. It is not ideal to have age-length data provided as eye orbital-fork length, given the assessment uses LJFL, as the need to apply conversion factors can introduce unnecessary bias. While it is good to have some age data included in this assessment, it is notable that this age data only comes from one subregion of the assessment, so the implicit assumption is that this data from subregion 1C is representative of all subregions in the assessment. Without any other ageing data available, this seemed to be a good first step, although clearly it would be preferable to have more ageing data, preferably with length measured using the standard LJFL, used in this assessment, and to have that age data collected over the full spatial range of the fishery, and with improved temporal sampling. These otoliths were only read once, so unfortunately it is not possible to estimate age-reading error.

In lieu of any estimate of ageing error, an assumed ageing error vector, similar to an ageing error vector from a species of similar maximum age, with otoliths read at the same laboratory, was used with the standard deviation of the age reading error at 0.25 for the youngest fish, and gradually increasing up to 0.75 for the oldest age class (19 years). The conditional age-at-length data has the youngest fish with a decimal age of around 0.4 years and the oldest fish with a decimal age of about 21.1 years. While it would clearly be preferable to have a data informed measure of the age reading error, this approach of making an assumption about this error seems preferable to assuming there is no age reading error. Ideally, sensitivities to the assumed values of the standard deviation of the age reading error would be tested. Hopefully this is another of those "lesser evils" featuring in this assessment. Remember those dragons!

4.1.7 New data

New data featuring in this assessment, compared to the 2021 assessment, include an additional four years of: catch data; CPUE data (including a new PICT CPUE index); length composition data; weight composition data, in addition to all of the conditional age-at-length data, collected from the period 1999–2002. Apart from the new PICT CPUE index and the conditional age-at-length data, all the other new data came from the last four years, 2020–2023. A summary of the available data sources, indicating the years that each data source is available is provided in Figure 11 and Figure 12.

4.2 Stock assessment methods

4.2.1 Population dynamics model and parameter estimation

A two-sex stock assessment for SWPO swordfish was conducted using the software package Stock Synthesis (version SS-V3.30.23.1, Methot and Wetzel (2013); Methot et al. (2025)). Stock Synthesis is a statistical age- and size-structured model which can allow for multiple fishing fleets, or fisheries, and can be fitted simultaneously to the data sources available for swordfish. The population dynamics model, and the statistical approach used in fitting the model to the various data types, are described in the Stock Synthesis technical documentation and are not reproduced here.

Some key features and assumptions of the diagnostic model are:

1. Swordfish constitute a single stock within the area of the fishery (SWPO)
2. The population was at its unfished biomass with the corresponding equilibrium (unfished) age-structure at the start of 1952.
3. Nineteen extraction fisheries are modelled.
4. Selectivity was assumed to be length-based, time invariant and to vary among fisheries, or at least between groups of fisheries. A logistic form of selectivity was assumed for the southern fishery groups, those fisheries operating in subregions 1S and 2S, and double normal selectivity assumed for all other fisheries. The two parameters of the logistic selectivity functions for each southern fishery group were estimated within the assessment. Five parameters of the double normal selectivities, assumed for all other fishery groups from central and northern subregions, were estimated within the assessment, for each fishery subgroup. However, for some fisheries subgroups, this double normal selectivity was only estimated in an initial preliminary stage, with these selectivities subsequently fixed at these preliminary estimated parameter values, with the corresponding length composition data further downweighted. The one fishery with time varying selectivity, both implemented and estimated, in the 2021 assessment, showed very small differences in selectivity for those two time periods in the 2021 assessment, so this time-block approach was not adopted for the selectivity estimated for that fishery in 2025.
5. The rate of natural mortality, M, was estimated within the assessment for the diagnostic

model, using a Lorenzen form for natural mortality with a prior. Natural mortality was assumed to be time invariant.

6. Recruitment to the stock is assumed to follow a Beverton-Holt type stock-recruitment relationship, parameterised by the average recruitment at unexploited spawning biomass, R_0, and the steepness parameter, h. A very weak prior was placed on R_0, a beta prior with a standard deviation of 0.1, to prevent the biomass from reaching unrealistically high levels and to prevent R_0 from being estimated too close to a boundary. Early development versions of this model had a tendency to estimate unrealistically high population scale, which was a major issue to be addressed and solved in the development stages of this assessment. Steepness is fixed to 0.8 for the diagnostic model. Deviations from the average recruitment at a given spawning biomass (recruitment residuals) are estimated from 1990 to 2021. Early recruitment deviations are estimated from 1969 to 1989. Recruitment deviations are not estimated prior to 1969 or after 2021 because there are insufficient data to permit reliable estimation of recruitment residuals outside of this time period.

7. The value of the parameter determining the magnitude of the process error in annual recruitment, σ_R, is set equal to 0.5 in the diagnostic model. The magnitude of bias-correction depends on the precision of the estimate of recruitment. Time-dependent bias-correction factors were estimated following the approach of Methot and Taylor (2011).

8. A plus-group is modelled at age 19 years.

9. Growth of swordfish is assumed to be time-invariant, meaning there is no change over time in mean size-at-age, with the distribution of size-at-age being estimated within the assessment, along with the remaining von Bertalanffy growth parameters. Separate growth curves are estimated for female swordfish, with the male growth parameters set from fixed offsets from the growth, based on offsets obtained from an externally estimated growth curve. Variances on the growth curve are fixed, based on values obtained from the externally estimated growth curve, with this variance set at two defined ages, young (aged 1 year) and old (aged 15 years).

10. Movement was assumed to occur at a fixed quarterly rate between region 1 and region 2. This movement rate was based on analysis of satellite tagged fish Patterson et al. (2021), and was noticably different to the movement specified in the diagnostic model, as described in the 2021 assessment paper (user beware), although this movement is consistent with the movement values actually used in the uncertainty ensemble used in the the 2021 swordfish assessment.

11. Recruitment proportions in the diagnostic model were fixed at a ratio of 1:3 between regions 1 and 2. When this recruitment proportion parameter was estimated within the model, the much larger recruitment proportion estimated in region 1 was considered unlikely. The diagnostic model assumption of a fixed recruitment proportion per region, roughly proportional to the area of each region, implies that spawning is equally likely through the assessment area, rather

than assuming (or estimating) that recruitment was much more concentrated in region 1 than in region 2. Catches in region 2 in the last 20 years or so have been around four times higher than in region 1. While it is possible that large numbers of recruits move from region 1 to region 2, to account for these larger catches in region 2, or recruitment spills over from adjacent areas, and recruitment in region 1 is much larger than recruitment in region 2, this was considered to be unlikely, as larval data show that larvae are found widely across both region 1 and region 2. A sensitivity to this assumption could be tested, but there was limited time and resources to explore all possible options in the 2025 assessment. Perhaps this possibility could be given further consideration in the next assessment. The assumed population and spatial structure, implicitly assumes no interaction with any "separate swordfish stock" outside the assessment region, and no movement across the assessment boundaries, in either direction. This is also an assumption that warrants much greater scrutiny in future assessments.

12. The input sample sizes for size composition data were adjusted for each fleet in a complex multi-step process which was largely subjective. This subjective process was developed and undertaken due to concerns arising out of the representativeness and influence of length composition data in previous billfish assessments, and in an attempt to achieve plausible model behaviour. The quarterly length frequencies were assessed for reliability, based on the number of quarters sampled, the number of years sampled, sample sizes of measured fish, and consistency of quarterly distributions, in particular looking for unusual and irregular spikes. At this stage, selectivity was shared between groups of fisheries, in an attempt to get plausible selectivity estimates, and all length fisheries were subjectively downweighted, with some data excluded completely (or downweighted to zero), allowing the subjectively adjudicated "more reliable" length composition data to have more influence on the selectivity for each fishery group. By this stage, most of the fits to the assessment input data and assessment outputs were looking reasonable, with the notable exception of the estimate of population scale, and the CPUE indices did not appear to be influential in estimating the population scale. The next step in the process was to upweight the size composition data for individual fisheries, with one fishery selected from each of five fishery groups. These five key fisheries included the three fisheries with weight composition data, those considered most reliable, 04.AU.1N, 05.AU.1C and 13.NZ.2C, and the two southern fisheries with the most reliable length composition data, and asymptotic selectivity, fisheries 03.DW.1S and 14.NS.2S. Some additional filtering on the length composition data for two of these length-composition fisheries, leaving only length samples from the most regularly sampled season from these southern fisheries, season 2. Selectivities estimated for all other fishery groups where then fixed, after the model was run with these initial data weightings, and the sample sizes, already subjectively downweighted or eliminated, were further downweighted by an additional factor of 0.001, so these "less reliable" fisheries with fixed selectivity effectively had no effect on estimates of population scale or recruitment. However, for these fisheries, the selectivity was fixed in a manner that ensured that fish were removed from the population, at approximately the correct size. After

this step, the remaining five fishery groups were the only fishery groups with selectivity estimated. Francis weighting (two iterations) was applied to adjust the input sample sizes on these five key fisheries.

13. The sample sizes for the conditional age-at length data were also iteratively re-weighted, using the method described in Punt (2017).

4.2.2 Relative data weighting and re-weighting

Iterative re-weighting of input and output coefficients of variation, or input and effective sample sizes is an imperfect but objective method for ensuring that the expected variation is comparable to the input (Pacific Fishery Management Council, 2024). This makes the model internally consistent, although some argue against this approach, particularly if it is believed that the input variance is well measured and potentially accurate. It is not necessarily good to down weight a data series just because the model does not fit it, if in fact, that series is reliably measured. On the other hand, most of the indices we deal with in fisheries underestimate the true variance by only reporting measurement error and not process error.

Data series with a large number of individual measurements, such as length or weight frequencies, tend to overwhelm the combined likelihood value, with poor fits to noisy data when fitting is highly partitioned by area, time or fishing method. These misfits to small samples mean that apparently simple series, such as CPUE, might be almost completely ignored in the fitting process. This model behaviour is not optimal, because we know, for example, that the CPUE values are in fact derived from a very large number of observations.

Length compositions were initially weighted using input sample sizes, based on the number of fish measured, obtained from the processing and analysis by Peatman et al. (2025), with these input sample sizes further capped at 1000, with any samples where the original sample size (OSS from Peatman et al. (2025)) was less than 100 fish also excluded, and some additional filtering applied for some specific fisheries. Some subjective decisions were made to downweight many fisheries with length composition data, due to concerns about representativeness and data quality. However, more objective data weighting (Francis, 2011) was used for the length composition data in the two southern fisheries, 03.DW.1S and 14.NZ.2S, and for the three fisheries that used weight composition data, 04.AU.1N, 05.AU.1C and 13.NZ.2C, with the approach of Punt (2017) for conditional age-at-length data. Two iterations of this iterative re-weighting were applied to each of the length, weight and conditional age-at-length data for the diagnostic model. These two steps resulted in the re-weighted input sample size coming close to the effective sample size, for the diagnostic model. It was not feasible to iteratively re-weight all models in the uncertainty grid, and the tuned weightings obtained from the diagnostic model were applied to all models in the uncertainty ensemble.

In Stock Synthesis, it is possible to internally estimate a separate additional variance parameter for the CVs associated with each CPUE time series, so that the input and output variances are balanced.

This additional variance parameter was estimated for all CPUE series, both in the diagnostic model and for all models in the uncertainty grid.

The initial value of the parameter determining the magnitude of the process error in annual recruitment, σ_R, was set to 0.5, reflecting the variation in recruitment for swordfish. The magnitude of bias-correction depends on the precision of the estimate of recruitment and the time-dependent bias-correction factors were estimated iteratively, following the approach of Methot and Taylor (2011), for the diagnostic model only, with these estimated bias correction factors from the diagnostic model applied to all models in the uncertainty grid.

4.2.3 Model development

The decision to transition to Stock Synthesis for the 2025 swordfish assessment was made for a number of reasons, including to take advantage of a number of features that were either unavailable, or not yet fully tested in MULTIFAN-CL in 2024, including using a two-sex model and using length-based selectivity, as well as taking advantage of some well established tools and diagnostics to develop and analyse results within the Stock Synthesis framework. Some of the features that were not available in MULTIFAN-CL in 2024 have subsequently been incorporated into more recent versions of MULTIFAN-CL and tested.

While the estimated selectivities from the 2021 assessment appeared to fit the time aggregated length composition data well, after recent experience with other billfish assessments (Castillo-Jordan et al., 2024), it was considered that much of the swordfish length composition data may not be reliable enough to allow them to have so much influence in the 2025 swordfish assessment. It is clear that the multi-modal time-aggregated length composition data, as seen in some swordfish fisheries, have some unusual temporal patterns, and simply fitting these with time invariant complex spline-based selectivity may not be the best approach.

Given the structural changes made to the 2025 assessment, with additional fisheries based on subregional structure, using a two sex-model, estimating growth internally and using length based selectivity, a traditional stepwise transition from the previous model was not conducted. This was partly because of the considerable difficulties encountered in developing this model, including the decision to give less emphasis to the length composition data. Some of the time aggregated length compositions for some fisheries are bimodal, and the age-based selectivity using splines used in 2021 produce some bimodal selectivities, which appear to fit the aggregated data reasonably well. However, careful scrutiny of the quarterly size composition data suggests that the individual quarters of data are rarely bimodal. Some have a peak of small fish and others a peak of large fish. The spline based approach to selectivity used in the 2021 assessment allowed a compromise, time invariant selectivity to be estimated, which would suggest that fish of an intermediate age are less likely to be caught than small or large fish associated with these two modes and, in any particular quarter, peaks of both large and small fish should be expected. However, this is not a pattern generally observed in the quarterly length composition data for these fisheries.

There appears to be no obvious reason why longline fishing gear would be less likely to catch these intermediate sized fish, so either there is a strange pattern in availability, or an alternative approach could be investigated. This led to further questioning of the quality of the length composition data and a decision to impose length-based selectivity, featuring either a logistic or a double normal shape, and to explicitly prevent the model from estimating a bimodal selectivity shape. This also led to much greater scrutiny of the length composition data and experiments in increasing the filtering and quality of these data, which led to downweighting this data source. At this stage, following a traditional stepwise model development approach, using a bimodal selectivity, first as age-based and then converted to length-based selectivity to try to reproduce the MULTIFAN-CL assessment from 2021 was not productive. No attempt was made to try to replicate the age-based selectivity used in the MULTIFAN-CL model, as the use of length-based selectivity was considered to be more reasonable, given that fishing mortality from longline gear is expected to act largely on the size of a fish, rather than its age.

During the development of the model, the general approach taken was to try to ensure that the selectivities being estimated were reasonable biologically, to allow for the known size variation with latitude, to fit the composition data reasonably well, to fit to the CPUE data, which were changed from quarterly indices to annual indices, and, when growth was estimated, gave plausible growth curves. While it was possible to achieve most of these goals at various stages of model development, and with some judicious downweighting of length composition data for some fisheries and sharing of selectivities within regions, it became apparent that the early versions of this model were struggling to estimate a plausible scale to the populations size.

4.3 Diagnostics

4.3.1 Model behaviour and convergence

The following criteria were considered and examined to assess model behaviour and convergence.

1. Minimum gradient sufficiently small, preferably less than 0.0001.
2. Hessian matrix was positive definite.
3. Models were not sensitive to initial conditions, so no models were found with improved likelihood through conducting jitter analyses, where starting values for all estimated estimate parameters are randomly and independently adjusted by up to 10%.
4. The number of parameters estimated on or close to bounds was small, preferably zero.
5. The number and composition of pairs of highly correlated estimated parameters.

The first three of these criteria were considered essential for the diagnostic model and, in fact, all five criteria were satisfactorily achieved for the diagnostic model. There were no parameters estimated close to bounds for the diagnostic model. Highly correlated parameters are difficult to avoid, especially with selectivity parameters. For the diagnostic model there were no pairs of

estimated parameters with correlation higher than 0.95, in absolute value. Further there were only three pairs of estimated parameters with correlation between 0.9 and 0.95, in absolute value. These three exceptions include two separate pairs of selectivity parameters from fishery 13.NZ.2C, both featuring the parameter describing the width of the right-hand limb of the double normal selectivity, as a member of these highly correlated pairs, along with another selectivity parameter from this same fishery. The remaining pair of highly correlated parameters was R_0 and the female mortality at age seven, used to scale the Lorenzen mortality curve.

For the uncertainty grid, it was not feasible to jitter all grid members, and the results of jittering the diagnostic model suggest that may not be necessary, given the apparent stability of this diagnostic model. The first criterion was used as a filter for models in the uncertainty grid, with any uncertainty grid models without a positive definite Hessian, excluded from the uncertainty grid. The second criterion was relaxed for the uncertainty grid, allowing models with final gradient less than 0.005 to remain in the uncertainty grid.

4.3.2 Jitter analysis

Jitter analysis is a technique used to test the optimality, robustness and stability of the maximum likelihood estimate obtained for a particular model. This involves randomly changing the starting values used for all estimated parameters and re-running the model, to test what alternative solutions may be found by the optimisation algorithm from different initial parameter values, or a formal process for testing the sensitivity to initial conditions. Two diagnostics are of interest with a jitter analysis, initially a check on whether a better "optimal solution" may be found, with a lower negative log-likelihood value, and also to see how frequently the optimal solution is found. As all estimated parameters are randomly modified, or "jittered", simultaneously, this can sometimes result in a model either failing to converge or finding a local minimum in a different (suboptimal) part of the multi-dimensional parameter space. A jitter analysis was conducted with 50 randomised replicates, modifying initial values by up to a factor of 0.1, or 10% change from the alternative initial conditions, used in the diagnostic model.

4.3.3 Likelihood profiles

Likelihood profiles are a standard component of the toolbox of applied statisticians and are most often used to obtain a 95% confidence interval for a parameter of interest. Many stock assessments "fix" key parameters such as natural mortality and steepness based on a priori considerations. Likelihood profiles can be used to evaluate whether there is evidence in the data to support fixing a parameter at a chosen value. If the parameter is within the range of the 95% confidence interval of the total likelihood profile, this provides no support from the data to change that fixed value. If the fixed value is outside the 95% confidence interval, and there is evidence that the data holds information about this parameter, it would be reasonable for a review panel to ask why the parameter was fixed and not estimated, and if the value is to be fixed, on what basis should inconsistency with

the data be ignored. Integrated stock assessments include multiple data sources (e.g., commonly for billfish, catch-rates, length-composition data, weight-composition data, and possibly conditional age-at length data) that may be in conflict, due to inconsistencies in sampling, but more commonly owing to incorrect assumptions or model misspecification (e.g., assuming that catch-rates are linearly related to abundance). Likelihood profiles can be used as a diagnostic to identify these data conflicts, and can be applied to parameters that are either fixed or estimated in the assessment. Likelihood profiles were constructed for the diagnostic model for $\log(R_0)$ and for natural mortality, M, or, more precisely, the female natural mortality at age 7, used to scale the Lorenzen form of mortality used in this assessment.

4.3.4 Retrospective analysis

A retrospective analysis (Mohn, 1999) is conducted to investigate the information that is being provided to the assessment by the last few years of data, by removing these data one year at a time, and examining the impact on the resulting assessment.

The retrospective analysis was undertaken using the following procedure.

1. The last year of data was removed, sequentially, from the 2025 diagnostic model
2. Time dependent model parameters (e.g. the last year of estimated recruitment) were changed to one year earlier
3. The model was re-run with one year less data available
4. Steps 1–3 were repeated for five years, removing one year of data at each step.

Trends in spawning biomass, estimated recruitment and key quantities of management interest, such as F/F_{MSY}, can then be examined to help understand how reliable the most recent few years of estimated recruitment deviations and spawning biomass are in the current assessment and, in association with likelihood profiles, can identify particular components of recent data that may be influential for the assessment.

The severity of retrospective patterns can be quantified using Mohn's rho, a statistic which is defined as the average of the relative differences between an estimate obtained from an assessment with a truncated time series, and an estimate of the same quantity from an assessment using the full time series (Hurtado-Ferro et al., 2015).

Mohn's rho can be calculated to measure the degree of the retrospective bias, for the following time series: recruitment, spawning biomass, relative biomass (as a ratio of unfished biomass) and fishing mortality. Mohn's ρ values have been interpreted using suggested thresholds, such as $|\rho| < 0.15$ indicating no significant bias, $0.15 \leq |\rho| < 0.20$ indicating minor bias, $0.20 \leq |\rho| < 0.30$ indicating moderate bias, and $|\rho| \geq 0.30$ indicating strong bias, indicative of problematic retrospective patterns. As suggested, these thresholds may be best treated as a guide, rather than treated as a pass-fail diagnostic, and may be better used as indicative of the need to look more closely at the

retrospective patterns, and to identify, interpret and explain any anomalies and unusual patterns found.

Step 2 in the retrospective analysis is not always carried out correctly by automated retrospective code, especially when the last year that recruitment are estimated is not the same year as the last year of data. When using the retros function in the R package r4ss, some care needs to be taken, and some changes to the interim outputs made to ensure that this retrospective analysis is correct.

4.3.5 Age structured production model

An age-structured production model (ASPM) diagnostic for the diagnostic model was estimated by:

1. Fixing growth and selectivity parameters at their estimated values
2. Removing the size composition and age data from the model, leaving the CPUE indices as the only data to be fitted
3. Setting all recruitment deviations to a fixed value of 0
4. Re-fitting the model estimating only the population scaling parameter, R_0

A comparison of biomass and depletion scaling and trends estimated by the ASPM and the full model gives an indication of the extent to which these estimates are informed by the CPUE indices only. If the ASPM can produce a reasonable fit to the indices of abundance, assuming they have some contrasting trends in the time series, it suggests there is a production function, and the indices can provide information about absolute abundance in the integrated assessment. If there is a poor fit to the indices, then the catch data alone cannot explain the trends/dynamics in the indices of relative abundance (Carvalho et al., 2021). Poor fit on an ASPM to the abundance indices may indicate that: (a) the stock is recruitment-driven; (b) the stock has not yet declined to the point at which catch is a major factor influencing abundance; (c) the model has mispecification issues; or (d) the indices of relative abundance are poorly representative of abundance trends.

A second ASPM variant can also be used as a diagnostic. For this variant, referred to as ASPM-R, recruitment deviations are freely estimated, rather than being fixed at zero, in addition to estimating the population scaling parameter, R_0.

4.4 Sensitivity analyses and structural uncertainty

4.4.1 Sensitivities

One-off sensitivity models were explored to understand the sensitivity of the diagnostic model estimations to structural and data uncertainties. Each one-off sensitivity model was created by making a single change to the diagnostic model and these were used to inform the construction of

an uncertainty grid (Table 2), with the diagnostic model values listed in bold below. These following sensitivities were explored, with values for the diagnostic model indicated in bold:

1. region 2 CPUE index used (**NZ**, PICT and EU);
2. steepness ($h = 0.7$, **0.8** and 0.9);
3. proportion of recruitment to region 1 compared to region 2, to roughly match the ratio of relative area of the two regions, or to match the ratio of the relative catch by region in the last 20 years (**1:3**, 1:4);
4. movement, (**diagnostic movement rates**, based on Patterson et al. (2021), diagnostic movement rates halved between regions 1 and 2, diagnostic movement rates halved between regions 2 and 1);
5. data weighting on the weight composition data, for the three fisheries with estimated selectivity and weight composition data, (**Francis weighting**, doubling and halving the Francis weights);
6. data weighting on the length composition data, for the two southern fisheries with estimated selectivity and length composition data, (**Francis weighting**, doubling and halving the Francis weights);
7. growth (**estimated** (internal), fixed (external));
8. natural mortality (**estimated (internal)**, fixed (external) using the approach of Hamel and Cope (2022)); and
9. all data from subregion 2N was removed from the assessment.

4.4.2 Structural uncertainty

Stock assessments of pelagic species in the WCPO use an approach to assess the structural uncertainty in the assessment model by running a "grid" of models that explore the interactions among selected "axes" of uncertainty. The grid contains all combinations of levels of several model quantities, or assumptions, and allows the sensitivity of stock status and management quantities to this uncertainty to be determined and factored into management advice. The axes are generally selected from factors explored in the one-off sensitivities with the aim of providing an approximate understanding of variability in model estimates due to assumptions in model structure not accounted for by statistical uncertainty estimated in a single model run, or over a set of one-off sensitivities.

The 2021 swordfish assessment used a new approach to uncertainty characterisation, involving drawing parameter sets from a multi-dimensional joint prior distribution, combined with a factorial component with four axes; CPUE indices, recruitment variability, growth (t_0), and alternative approaches to defining the reference natural mortality. The joint prior included uncertainty on

reproductive potential, length-weight relationship, growth, natural mortality, steepness and movement between regions 1 and 2. It also incorporated model filtering for poorly converged models and those with implausibly high biomass and included estimation uncertainty on the retained models. The approach was the first ensemble type approach to characterising uncertainty for an SPC billfish or tuna assessment and has several benefits over the standard grid approach (see Ducharme-Barth and Vincent (2022)).

Due to the other structural developments undertaken in this new assessment, including incorporating; sex structure, detailed reviews of size data/preparation methods, changing selectivity forms and changing software platform, it was not possible to replicate this improved approach to uncertainty characterisation in the time available. However, the uncertainty grid in this assessment still covers key uncertainties, including; CPUE, steepness, recruitment distribution, movement between region 1 and 2, data weighting of both length and weight data, and natural mortality, but under-represents some important uncertainties, such as growth. An attempt was made to include an alternative fixed external growth option in the uncertainty grid, but these models produced implausible results and there was insufficient time to fully diagnose the problems. Estimation uncertainty is included in the grid approach used in 2025, and this does provide a substantial range of uncertainty in management quantities. However, applying an ensemble approach as in the 2021 assessment (Ducharme-Barth et al., 2021a) would likely expand the uncertainty range, and this approach is recommended for future assessments.

The structural uncertainty grid was constructed from a subset of the sensitivities investigated, using seven axes of uncertainty with 2–5 levels for each (listed below), resulting in a total of 360 models (Table 3). The values for the diagnostic model are shown in bold:

1. region 2 CPUE index used (**NZ** and PICT);
2. steepness ($h = 0.7$, **0.8** and 0.9);
3. proportion of recruitment to region 1 compared to region 2, to roughly match the ratio of relative area of the two regions, or to match the ratio of the relative catch by region in the last 20 years (**1:3**, 1:4);
4. movement, (**diagnostic movement rates**, based on Patterson et al. (2021), diagnostic movement rates halved between regions 1 and 2, diagnostic movement rates halved between regions 2 and 1);
5. data weighting on the weight composition data, for the three fisheries with estimated selectivity and weight composition data, (**Francis weighting**, doubling and halving the Francis weights);
6. data weighting on the length composition data, for the two southern fisheries with estimated selectivity and length composition data, (**Francis weighting**, doubling and halving the Francis weights); and

7. natural mortality (**estimated (internal)**, fixed (external) using the approach of Hamel and Cope (2022)).

4.4.3 Integrated model and estimation uncertainty for key management quantities

For a full picture of uncertainty for the key management quantities adopted for billfish in the WCPFC, $SB_{\text{recent}}/SB_{\text{MSY}}$ and $F_{\text{recent}}/F_{\text{MSY}}$ (Table 2), structural uncertainty was integrated with estimation uncertainty for individual grid models. The procedure adopted was:

1. Obtain the estimates of the key management quantities for each of the 360 grid models;
2. Obtain Hessian-based estimates of the standard deviations for these quantities;
3. Generate 1,000 random draws from normal distributions with mean and standard deviation specified in steps 1 and 2, for each of the 360 grid models;
4. Compute the mean, median, and 10^{th}, 25^{th}, 75^{th} and 90^{th} percentiles of each management quantity, from the 360,000 values.

Note that this procedure implicitly gives equal weight to each of the 360 grid models, unlike the more sophisticated approach pioneered in Ducharme-Barth et al. (2021a), which may be a better approach to use. However, adopting an uncertainty grid rather than an uncertainty ensemble was a pragmatic decision taken for this assessment. Different relative weights could be given by varying the number of random draws of the management quantities from each grid model.

5 Results for the diagnostic model

5.1 Overview of convergence

The diagnostic model had a final gradient of < 0.00003, a positive definite Hessian and with no parameters estimated to be close to bounds. None of the 50 jitters obtained a solution with an improved likelihood value, which gives a strong indication that the model has found the optimal solution.

5.2 Biological parameters

5.2.1 Growth

In the 2021 assessment, for the one-sex model, and during the initial model development in 2025 for a two-sex model, growth was estimated externally, with three von Bertalanffy growth parameters and two CVs for young and old fish (set at ages 1 and 15 years respectively) estimated for each sex in the two-sex model. These fixed external growth curves, were updated slightly from the curves used in the 2021 assessment, with the growth curve parameters re-estimated after first converting the raw conditional age-at-length data to LJFL. This was one of two growth options proposed for use in the uncertainty grid. For the diagnostic model, some of these parameters were estimated internally, allowing the model to incorporate the effects of selectivity on these growth estimates, and estimate growth in an integrated manner. However, due to limitations in the number of conditional age-at-length data points available, it was only possible to estimate three of these ten growth parameters, which were the three von Bertalanffy parameters for females. Pragmatically, the four CV parameters were fixed at the CV estimates obtained from the external growth estimates, and the male parameters were given fixed offsets from the estimated female von Bertalanffy parameters. The estimated growth curves for females and the male growth curve, using fixed offsets from the female curve, are shown in Figure 13.

While it would be preferable to estimate more then three of these 10 growth parameters internally, attempting to do so resulted in poor model behaviour and implausible growth curves. The approach of fixing the male offsets allowed the male growth to vary relative to the estimated female growth, and seemed to offer the best compromise of flexibility and model stability.

Using conditional age-at-length data and estimating sex-specific growth parameters internally was a new approach, compared to the 2021 assessment, and this required considerable care and attention to detail, to ensure the resulting growth estimates were plausible and reasonable. Some decisions on the plausibility of growth and variance estimates were somewhat subjective, but this was considered to be worthwhile given the importance of using a biologically realistic growth form in the model, and to be able to account for the impact of gear selectivity on the age-at-length data collected from the fishery and the impact of ageing error. Unfortunately there were only single reads of the otoliths, so it was not possible to make an estimate of ageing error, and the values for ageing error were fixed at "roughly plausible" values for a fish of this age. Ideally, a sensitivity to these assumed

values for ageing error would be conducted, but there was insufficient time to explore this aspect of the assessment this year. Remember the dragons?

5.2.2 Natural mortality

The Lorenzen form of natural mortality used in the diagnostic case, with female mortality estimated at age seven years, using the methods outlined in Section 4.1.1 in Figure 14.

5.2.3 Length-weight relationship

The length-weight relationships for male and female swordfish used in the diagnostic model are shown in Figure 15.

5.2.4 Sexual maturity

The age based sexual maturity and fecundity by length relationships for the diagnostic model is shown in Figure 16

5.3 Fits to the data

5.3.1 Estimated selectivity curves

While some size data composition showed periodic and episodic catches of small fish, attempting to fit this mode of small fish in the aggregated distributions, in addition to fitting a second larger mode, caused problems and instability for the model. Rather than allowing the model to overfit this noisy data, and potentially switching between fitting either the larger or the smaller mode, using a flexible but somewhat complex spline selectivity, which is difficult to explain in terms of the gear being less likely to select fish of intermediate size, it was decided to fit only two simpler functional forms for selectivity. These fisheries catching the largest fish, in subregions 1S and 2S, were assumed to have logistic selectivity, while all other fisheries from the other subregions, in the central and northern zones, were assumed to have double normal selectivity, which did not preclude a logistic shape, but allowed for less chance of catching larger fish, through a combination of lower selectivity of the gear and/or lower availability of larger fish in these subregions.

In the diagnostic model, sharing selectivity between fisheries resulted in only seven different selectivities, for each of seven different groups of fisheries, with each selectivity shared amongst the different fisheries in each group. These are shown as length-based selectivity in Figure 17, which is the form in which they were estimated, and with the equivalent age-based selectivities shown in Figure 18.

Two of these selectivities were initially estimated, in a preliminary run of the model, and then fixed, at these estimated values, in all subsequent model runs, with further downweighting of the associated length composition data, through an additional multiplier of 0.001 on these already downweighted initial input sample sizes, applied to any fisheries that contributed to this initial

estimation of selectivity for each group. One of these groups of fisheries with fixed selectivity (Figure 19), was the group of fisheries with length data in subregion 2C, fisheries 11.DW.2C, 16.EU.2C and 19.PICT.2C, with fishery 16.EU.2C (multiplier 0.00001) having five times as much weight as the other two fisheries, 11.DW.2C and 19.PICT.2C (multipliers 0.000002). The other group of fisheries with fixed selectivity (Figure 20), comprised all of the fisheries with length data from both subregions 1N and 2N, fisheries 01.DW.1N, 08.PICT.1N, 10.DW.2N, 15.EU.2N and 18.PICT.2N, with fishery 08.PICT.1N (multiplier 0.00001) having 100 times the weight of fishery 10.DW.2N (multiplier 0.0000001), while the other three fisheries were assigned zero weight, both when estimating the selectivity and when used in the model.

Selectivity is estimated for the remaining five groups of fisheries, using Francis (2011) weighting for one main fishery in each group, and a subjective low weighting for any other fisheries in each group, with input sample size weighting as described below. These five groups comprised the following fishery groups with three weight-based selectivities and two length-based selectivities:

1. The fishery with weight composition data in subregion 1N, 04.AU.1N, Francis weighted with multiplier 0.2891 (Figure 21).

2. The fishery with weight composition data in subregion 1C, 05.AU.1C, Francis weighted with multiplier 0.2555, but also sharing selectivity with the downweighted length-based fisheries in subregion 1C, and 02.DW.1C, 07.EU.1C (multiplier of 0.000005 for both) and with all length composition samples from fishery 09.PICT.1C removed by the initial filtering steps (Figure 22).

3. The fishery with weight composition data in subregion 2C, 13.NZ.2C, Francis weighted with multiplier 0.05214 (Figure 23).

4. The fishery group in subregion 1S, dominated by the length composition data from fishery 03.DW.1S, Francis weighted with multiplier 0.2054, but also sharing selectivity with the downweighted length-based fisheries in subregion 1S, 06.AU.1S (with fewer samples and much smaller multiplier of 0.00001). Fishery 03.DW.1S also had anomalous smaller fish removed from the composition sample (fish less than 90 cm), as they were considered unrepresentative and unusual samples from this fishery, and they confounded the estimation of the time invariant selectivity for this fishery when left in the length composition samples (Figure 24).

5. The fishery group in subregion 2S, dominated by the length composition data from fishery 14.NZ.2S, Francis weighted with multiplier 0.05952, but also sharing selectivity with the downweighted length-based fisheries in subregion 2S, 12.DW.2S and 17.EU.2S, both of which had multipliers of zero, as they had so few samples (Figure 25).

In subregions 1N and 2C, there were separate selectivities estimated for length and weight data. In subregion 2N, the selectivity was shared with the selectivity estimated in sunregion 1N, but dominated by the most consistent of the length composition data from the northern subregions, from

fishery 08.PICT.1N, as many of the length composition samples from these subregions were inconsistent and irregular and did not appear to be reliable, or useful to inform estimates of selectivity in those northern subregions.

In subregion 1N, the separately estimated selectivities for length (Figure 20) and weight (Figure 21), have a similar general shape, with a similar peak, but a broader top, and steeper left and right hand ascending and descending limbs for the weight based selectivity. In subregion 2C, the separately estimated selectivities for length (Figure 19) and weight (Figure 23), also have similar shape, but suggest that the weight based fisheries in 13.NZ.2C reaches maximum selectivity at slightly smaller size than the equivalent length-based selectivities from the same subregion. In subregion 1C, where there was both length and weight composition data, a single shared selectivity was estimated using both composition data types (Figure 22).

5.3.2 Fits to CPUE

The fits to the Australian and New Zealand CPUE indices used in the diagnostic model are shown in Figure 26, plotted on a log scale, to reflect the space in which these indices are fitted. These indices are balanced by estimating an additional variance parameter within Stock Synthesis, which is negative (-0.0617) for the Australian index, suggesting the models fit well with less variance (total of 0.138) than the initial values of 0.2, and positive (0.124) for the New Zealand index, suggesting that the model requires additional variance (total of 0.224) for this index, compared to the initial value used of 0.1. This total variance required for each index suggests that the model is more easily able to fit the Australian index, than the larger absolute variation seen in the New Zealand index.

While a formal runs test analysis was not conducted for these fits, the pattern seen in the residuals in Figure 27 suggest an acceptable residual pattern for the Australian CPUE index, but with some residual pattern concerns for the New Zealand CPUE series, with the model unable to fit the magnitude of the variation seen in this index. This pattern shows a six year run of consistent overestimates until 2010, followed by nine years of consistent underestimates until 2019 for the fits to the New Zealand CPUE series. However, on balance these CPUE fits can be considered to be adequate, as the model shows a reasonable fit to the trends seen in both indices.

5.3.3 Fits to length data

The diagnostic model fits to the aggregated length-frequency distributions are shown in Figure 28. The fits are good for the two southern fisheries, featuring logistic selectivity, fisheries that are very influential in this model, fisheries 03.DW.1S and 14.NZ.2S. These are the fisheries with length composition data that are most heavily weighted, and for which good fits are required.

Many of the remaining aggregated fits are for information only, indicating fits to data which is either heavily downweighted, or given zero weight to the model, so indicative of virtual fits to the length data for each fishery from the fixed selectivity curves used. Of these remaining fisheries with

length composition data, those of importance have quite good fits, even with highly downweighted data, including the following more heavily weighted out of the "downweighted length composition fisheries": 08.PICT.1N; fisheries 02.DW.1C and 07.EU.1C in subregion 1C; 16.EU.2C.

The "virtual fits" to the fisheries from region 2S, 12.DW.2S and 17.EU.2S are poor, but these are from very small samples and they are given zero weighting in the model. In contrast, the fit to the heavily downweighted fishery 06.AU.1S looks reasonable. The fits to the downweighted fisheries in the subregion 1N and 2N shared group are poor, but the only fisheries with non zero weight, in this shared group, are fisheries 08.PICT.1N (which is given the most weight) and 10.DW.2N (which has much lower weight, and features a poor fit to a bimodal distribution). The other three fisheries in this group, with zero weight in the model, include 01.DW.1N (distribution and fits are not too bad), 15.EU.2N (bimodal distribution, poor fit but notably only from two quarters of data) and 18.PICT.2N (bimodal distribution with a poor fit). In subregion 2C, which is dominated by fishery 16.EU.2C, with many samples over a six year period, the other "further downweighted" fisheries are 11.DW.2C, (reasonable fit to the data), and fishery 19.PICT.2C, (poor fit to an unusual looking aggregated distribution). The length based fisheries in subregion 1C, which share selectivity with the much more heavily weighted and dominant weight composition fishery 05.AU.1C, include two fisheries with non-zero (but very small) weighting, fisheries 02.DW.1C, (featuring a suspicious looking spike of small fish), and 07.EU.1C, (passable length fits), and fishery 09.PICT.1C which had all length samples filtered out (Peatman et al., 2025).

A time series of quarterly fits to the length data by fishery is also shown in Figure 30, which gives an overview of the number of quarters of length composition data included in the diagnostic model, post-filtering. These plots show that the temporal fits to the means of these quarterly length distributions are good for the most important length composition fisheries, 03.DW.1S (row 1, column 3) and 14.NZ.2S (row 3, column 2), despite some time variability in the data. Others, from the more influential and important length based fisheries include: 08.PICT.1N (row 2, column 2, generally good fits); 02.DW.1C (row 1, column 2, adequate fits); 07.EU.1C (row 2, column 1, adequate fits, but with only two quarters of data); and 16.EU.2C (row 3, column 4, poor fits, generally underfitting the data). In this last case, with underfits to the data, the other two contributors from this group have better fits, namely fisheries 11.DW.2C (row 2, column 4, adequate fits) and 19.PICT.2C (row 4, column 3, adequate fits).

5.3.4 Fits to weight data

The diagnostic model fits the aggregated weight-frequency distributions very well, Figure 29. These weight fits look a little unusual, due to the variable bin width, which results in artificial spikes, looking a bit like a punk haircut, whenever the bin width changes. A more conventional plot would scale the y-axis by the width of the weight bin, to give a smoother looking distribution. Regardless, when properly interpreted, these plots show excellent fits to the data, for the very small fish with 1 kg bin widths, whenever there is a step in bin size, and also for the largest weights, with wide

bin widths. The sample sizes for fishery 05.AU.1C are much larger than for fisheries 04.AU.1N and 13.NZ.2C, but for all three of these fisheries, the fits to the weight data are excellent. This is important to the model, given that all three of these fisheries have high data weighting.

A time series of quarterly fits to the weight data by fishery are shown in Figure 31, which gives an overview of the number of quarters of weight composition data included in the diagnostic model. These fits range from good to excellent, with some suggestion that incorporating a seasonal component to selectivity could be beneficial. There was insufficient time to incorporate seasonal selectivity in the diagnostic model in 2025, but it may be an interesting extension to explore in future models. Remember the dragons? The ability of the model to capture the time variation in median weight in fishery 05.AU.1C is encouraging.

5.3.5 Fits to conditional age-at-length data

Fits to the conditional age-at-length data are shown, on occasions infamously referred to as "Andre plots" in Figure 32, Figure 33 and Figure 34, with observations (black points) and predicted ages at length (blue line) in the left hand panels. The right hand panel shows observed and expected variances, although these are not very meaningful given the low numbers of samples from each quarter and the high resolution of length measurements used, meaning that most lengths do not have replicate samples to estimate a reliable variance on the age for each length. These plots indicate that the conditional age-at-length data, which is entered on a monthly time scale, is being well fitted by this model, and the blue line fits through the black points fairly well in each panel.

The fits to the mean age, aggregated by year, are show in Figure 35, demonstrating that the model is fitting these data well and providing an indication that the conditional age-at length is appropriately weighted, using the data weighting method of Punt (2017).

5.4 Diagnostic model outcomes

5.4.1 Female spawning biomass

The estimated time series of female spawning biomass is shown in Figure 36, aggregated over regions, with approximate 95% asymptotic intervals, and with separate female spawning biomass estimates for regions 1 and 2. This indicates a steady decline in spawning biomass as the fishery develops, with an accelerated decline in the late 1990s, especially in region 2, as catches increase, with a slight recovery in the late 2000s followed by a further decline in the mid 2010s, ending at around half the estimated female spawning biomass that was estimated in 1952.

Figure 37

5.4.2 Fished versus unfished spawning biomass

To interpret the trends in depletion, the individual trends in spawning biomass, SB, needs to be compared with the predicted spawning biomass in the absence of fishing (unfished) $SB_{F=0}$

(Figure 37, top). $SB_{F=0}$ is stable until the early 1990s and then starts to oscillate with a roughly 20 year period of increasing and then decreasing. $SB/SB_{F=0}$ is also shown in (Figure 37, bottom) over the same time period. Because the periodic oscillation occur simultaneously in the fished an unfished time series, these patterns do not appear in the $SB/SB_{F=0}$ time series.

5.4.3 Estimated recruitment

The estimated recruitment series features a period of early recruitment deviations, from 1969 to 1989, with full recruitment deviations estimated from 1990 to 2021 are shown in Figure 38. Early recruitment deviations, estimated with high standard error, are used to estimate recruitment deviations prior to the start of the size composition data, as the first years of size composition data contain some information about earlier recruitment events in the older cohorts, which were recruited years before that data was collected. These full recruitment deviations appear to show some longer term cyclic behaviour, with some correlation apparent in successive estimates. The recruitment variance checks and recruitment bias adjustments are shown in Figure 39, showing that the asymptotic standard errors are comparatively high in the early recruitment period and then decline as the full recruitment period starts, and remaining low until the end of the estimated recruitment period, indicating that the last of the estimated recruitment deviations is (relatively) well informed by the data. The bias adjustment plot Figure 38 shows that the bias ramp is adjusted appropriately according to methods outlined by Methot and Taylor (2011).

5.4.4 Summary fishing mortality

Fishing mortality will vary with the relative effort between different fisheries, each with their own selectivity, so will vary with the fishery dynamics, as well as by size and age of the fish, depending on the size distribution of the population in any one year. Summary fishing mortality was calculated as an annual fishing mortality by year, using a numbers weighted mean F, calculated across the age class ranges 3–12, 1–3 and 8–15 (Figure 40). This summary fishing mortality is low (< 0.025) until the late 1990s, when there is a steady increase to around 0.08 by about 2005, and then it stabilises around 0.06, with some annual variation, until the end of the time series. This matches the general pattern seen in the 2021 assessment for fishing mortality, noting the MULTIFAN-CL calculates fishing mortality using different methods to those used Stock Synthesis.

5.4.5 Spawning potential ratio

Spawning potential ratio (SPR) is defined as the spawning output at a given fishing rate, relative to that of an unfished stock. Figure 41 shows a phase plot of SB/SB_{MSY} plotted against SPR, relative to the spawning potential at MSY, $(1-\text{SPR})/(1-\text{SPR}_{\text{MSY}})$, which provides a measure of relative fishing intensity, relative to F_{MSY}. Figure 42 shows the time series of SPR. This shows an increase in fishing intensity for about a 10 year period from 1995, and is then relatively stable, with some annual variation, from about 2005 onwards, at a moderate level of fishing intensity, showing the

same general fishing intensity pattern as seen in the summary fishing mortality (Figure 40).

5.4.6 Stock recruitment relationship

Figure 43 shows the stock recruitment curve, with estimated recruitments, and Figure 44 shows the log recruitment deviations from the stock recruitment curve.

5.5 Diagnostics

5.5.1 Jitter analysis

The 50 jitter runs all successfully converged, with 48 of these runs obtaining the same solution and same minimum negative log-likelihood as the diagnostic model. The remaining two jitter runs converged to alternative solutions, local minima, with negative log-likelihood values worse than the likelihood value achieved in the diagnostic model by several log likelihood units. This suggests that there is no local instability in the solution obtained for the diagnostic model, and there appears to be no sensitivity to the initial starting conditions. While not sufficient to claim the model is converged, these results from a jitter analysis are a necessary condition for a converged model.

5.5.2 Likelihood profile: population scale, log(R_0)

The likelihood profile conducted over log(R_0), the scaling parameter for recruitment at the start of time series, is shown in Figure 45. The index data, or CPUE, (plotted in dark blue with triangles) support a higher value for this population scaling parameter, with the largest influence from the New Zealand CPUE index, 21.IDX.NZ.2, as shown in the Piner plots (Figure 46). In contrast, the weight data, (plotted in light blue with + symbols in Figure 45) support a lower value for this population scaling parameter, almost entirely through the weight data from the Australian fishery 05.AU.1C, as shown in the Piner plots (Figure 46). The length data are not highly influential in estimating population scale, which is appropriate given the level of downweighting and the paucity of informative length data available to use in the assessment.

The estimate of population scale is relatively uncertain, as can be seen by the range of values of log(R_0), for which the change in total likelihood is less than 1.92 likelihood units, and hence not considered statistically different, with this cutoff value plotted with a dashed horizontal line in Figure 45. This relatively high degree of uncertainty in the estimate of population scale is also demonstrated through in the 95% asymptotic intervals shown on the time series of spawning biomass (Figure 36), which suggests that estimates of initial female spawning biomass could range between 60,000 t to slightly more than 100,000 t.

This likelihood profile suggests that the two most important data components that influence this population scale estimate are the New Zealand CPUE index and the Australian weight composition data from fishery 05.AU.1C, with conflict between these two data sources.

5.5.3 Likelihood profile: scale for Lorenzen M

The likelihood profile conducted over the scaling parameter for Lorenzen mortality, M for females at age seven years, is shown in Figure 47. As with the likelihood profile on $\log(R_0)$, the index data, or CPUE, (plotted in this M profile in light blue with + symbols) support a higher value for this M scaling parameter, with the dominant index influence coming from the New Zealand CPUE index, 21.IDX.NZ.2, as shown in the Piner plots (Figure 48). In contrast, the weight data, (plotted in dark blue with triangles in Figure 47) support a lower value for this M scaling parameter, almost entirely through the weight data from the Australian fishery 05.AU.1C, as shown in the Piner plots (Figure 48). Once again, the length data (and other data sources) are not highly influential in estimating scale for M, which is appropriate given the level of downweighting and the paucity of informative length data available to use in the assessment. The relatively minor influence of the length composition data is driven by length composition data from fishery 14.NZ.2S, which supports both a lower value for M and a lower population scale.

As can be seen in both likelihood profiles, (Figure 45 and Figure 47), the dashed black curve, showing the total likelihood without priors, is almost identical to the total likelihood with priors, the solid black curve. Indeed the likelihood component from the priors, with priors incorporated on both M and R_0 in the diagnostic model, is explicitly shown in both profiles, the red curve with squares, and is almost flat, in both plots, indicating that the influence of the priors on the diagnostic model is almost inconsequential.

This likelihood profile suggest that the scaling parameter for Lorenzen mortality, M for females at age seven, can be estimated, and is likely to be in the range between 0.22 and 0.255. The prior on M is not very influential on the diagnostic model

Again, this likelihood profile suggests that the two most important data components that influence the Lorenzen M scale estimate are the New Zealand CPUE index and the Australian weight composition data from fishery 05.AU.1C, with conflict between these two data sources. Given the quality of these two data sources, an increase in effort in collecting better data from these two sources is unlikely to resolve this conflict. This suggests that there may be a better model structure to use for this stock assessment, to avoid this conflict, such as alternative spatial structure assumptions or possibly different stock structure assumptions.

The smooth shape seen in the overall likelihood in both of these likelihood profiles, and the individual components of the likelihood suggest that the model is converging well, (apparently no dragons here?) with no indication of erratic behaviour or convergence issues even some distance from the optimal parameter value. While this may not be sufficient to be absolutely certain that there are no convergence issues with this model, it is suggestive of a well-behaved, stable converged model. However, these likelihood profiles also show some conflict in the data and confirm the broad uncertainty bounds for some of the estimated parameters (so perhaps some baby dragons, after all?).

5.5.4 Retrospective analysis

The results of the retrospective analysis for absolute female spawning biomass are shown in Figure 49, showing the full data (the diagnostic model, in dark blue), then the successive models with all data removed after: 2022 (light blue), 2021 (green), 2020 (yellow), 2019 (orange), and 2018 (red). The same plots are shown for: fishing intensity, measured by the SPR, (Figure 50); fishing intensity, measured by the F/F_{MSY}, (Figure 51); recruitment, measured by numbers of age-0 recruits, (Figure 52); recruitment, measured by estimated recruitment deviations, (Figure 53); recruitment deviations, demonstrated by a squid plot (Figure 54); fits to the Australian CPUE data from fishery 20.IDX.AU.1, (Figure 55); and fits to the the New Zealand CPUE data from fishery 21.IDX.NZ.2, (Figure 56).

Mohn's rho is probably best used as an indicative statistic, but a value of 0.628 for spawning biomass suggests there is a problem for the retrospective pattern on female spawning stock biomass. Figure 49 shows the relatively large change to estimates of initial female spawning biomass in removing single years of data, reinforcing the conclusion already reached, that the scale of the population size is not well determined by the data in this assessment. The first peel, removing data from 2023 only, also suggests that the data from 2023 is influential in scaling down the recent female spawning biomass estimates, in the last few years of the time series.

The value of Mohn's rho for F of -0.398 is also large in magnitude, indicating potential issues with fishing intensity. Figure 50 and Figure 51 show a clear pattern with the first peel, removing data from 2023 only, resulting in the estimate of fishing intensity being revised up early in the time series, 2000–2010, and revised down in more recent years. This indicates that the last year of data used in the assessment is influential in estimating fishing intensity. In other words, a single year of data can change these estimates considerably, indicating that fishing intensity is being estimated with considerable uncertainty, given it can change this much with the addition of a single extra year of data.

The value of Mohn's rho for recruitment is 0.816, also suggesting a poor retrospective pattern. Careful analysis of Figure 52 and Figure 53 show a "one-way trip" in recruitment estimates, with lower, or more negative, recruitment deviations estimated from the more recent data, the recruitment deviations tend to get more negative moving from warm to cooler colours, or from more peels to less peels. The pattern in absolute recruitment estimates is confounded by changes in estimates of the spawning biomass and the influence of the stock recruitment relationship in changing the absolute numbers of recruits, so the retrospective analysis showing recruitment deviations is simpler to interpret. This change in the pattern in recruitment deviations, is demonstrated clearly by the one-sided squid plots (Figure 54). This plot describes the size and direction of the change in the recruitment deviations, stepping back in time from when the last recruitment deviation was estimated (2021, in this case). In this plot, each "cohort age/peel" starts with a different colour, and moving from right to left along each string, shows the impact of successive peels on each cohort, sowing more years of deleted data by moving to the left on each coloured string. The variation along each

string indicates how the recruitment deviation estimate changes as each year of successive data is added (moving to the right) or removed (moving to the left).

The squid plot follows changes in the recruitment deviations for particular cohorts, as each of the last five years of data are successively removed. Each coloured string, corresponding to a particular cohort, includes a maximum of six points, one for the diagnostic model using data up to 2023, and then one point for each of the five different retrospective peels. Each string, when followed from right to left, represents removal of successive years of data. The changes in the estimates of recruitment deviations, as each year of data is removed, are measured by changes in the y-axis, along each string, relative to the final recruitment deviation for that cohort from the diagnostic model. A negative value indicates a revision downwards and a positive value indicates a revision upwards. This plot is clearly unbalanced, as it features revisions upwards, compared to the diagnostic model in every case, with a single exception for cohort 2017, after 3 peels.

Large changes on the y-axis indicate large revisions, and if all the changes have the same sign (positive or negative) this indicates consistent changes in the same direction, compared to the diagnostic model, indicating some bias, rather than revisions in each direction which would be expected to occur with roughly equal probability. In this case, most of the change (vertically, in the y-axis) is in the first point (moving from right to left on each string), at least for the most recent six cohorts, indicating that most recent year of data is having the largest influence on these revisions, but with some evidence of a dip after peel three when the revisions change direction.

In this case the revisions are consistently in the same direction over five peels, indicating that the more recent years of data result in these recruitment deviations being consistently revised down, certainly compared to peel 5, resulting in a "poor" recruitment pattern. The last few years of data are not expected to always revise these cohorts in the same direction, and a "good" squid plot looks more like a squid, with threads either side of the $y = 0$ line, typically with larger amplitude on the left, for the cohorts with fewer years informing the recruitment deviation, and smaller amplitude on the right, representing older cohorts. These older cohorts have more years of data to inform, or stabilise, the estimated recruitment deviation for that cohort, so are less affected by losing up to five years of data, compared to the younger, more recent cohorts.

The retrospective patterns in the CPUE are shown in Figure 55 for the Australian CPUE and Figure 56 for the New Zealand CPUE. In both cases, it is not the last year of the CPUE that is dragging the fit to the index down (the blue line in Figure 56), for the diagnostic model at the end of the times series, as in both CPUE series, the final CPUE data point is higher than the fitted value, so the last CPUE data point would be expected to discourage this downwards revision on the retrospective plot.

The earlier likelihood profiles suggest that the CPUE and the weight data has the most influence on the model, and the analysis above suggests that the CPUE is not responsible for this retrospective pattern in the CPUE series. The most likely alternative scenario is for the weight data to be

contributing to this recent change, and (Figure 46) suggests that the weight data from fishery 05.AU.1C, is the most likely candidate.

The y-axis scale used, and the plotting of points from every quarter in Figure 30 make this a little hard to visualise, so each year of weight data was re-plotted and aggregated annually, for fishery 04.AU.1N (Figure 58), fishery 05.AU.1C (Figure 58) and for fishery 13.NZ.2C (Figure 59) to examine this more closely.

The weight data from fishery 04.AU.1N finishes in 2019, so this component of the weight data has limited influence on this retrospective pattern, as it can only have any impact in peel 5 (Figure 57). The weight data from fishery 05.AU.1C (Figure 58) demonstrate that the last five years of data indicate a drop in mean weight, which is well reflected in the fit to these data, so these data are influential in the retrospective pattern seen in the CPUE. The weight data from fishery 13.NZ.2C (Figure 59) also show a recent decline (compared to earlier years) in mean weight in 2020, 2022 and 2023, the only years with data collected in that fishery since 2017, again with that change in mean weight being fitted reasonably well by the diagnostic model.

It seems that the recent weight data in fisheries 05.AU.1C and 13.NZ.2C are contributing to this poor retrospective pattern. It may be prudent to seek further information on this recent decline in mean weight recorded in these fisheries in the last five years, from those involved in those fisheries, particularly the drop in mean weight in 2023 for fishery 05.AU.1C, to see if there are any operational reasons or changes in data collection procedures that may explain this change, or whether this is a real signal from the fishery that is providing important information to the assessment. The weight data from fishery 05.AU.1C is quite influential, and there seem to be some recent changes in these data, which could be related to an increase in small fish, through good recruitment, or a decrease in larger fish in the population, that perhaps warrant further enquiries and investigation.

These retrospective analyses reinforce the suggestion that the diagnostic model estimates of population scale is quite uncertain, and the diagnostic model is quite sensitive to particular data sources. While the model is well behaved in some respects, with apparently stable converged solutions this uncertainty on scale estimation and the sensitivity to some of the more recent data suggest the model lacks some robustness, given the structural model assumptions made and the data available to inform the model.

The size data are important for estimating scale in this model, as the CPUE data seem to be relatively uninformative for population scale. Given this, having a reliable growth curve is very important for this assessment, with regional variation in growth potentially important, and currently unknown given most of the data informing the estimated growth comes from a small subregion of the assessment region, fishery 05.AU.1C. The size data needs to be prepared adequately, ideally with some model based standardisation to correct for patchy sampling. More age data is urgently needed as this is likely to be a more reliable source of information on growth and on population scaling. Finding an alternative growth form that gives plausible model results would also help to

categorise the model uncertainty, and the single estimated growth curve used in the uncertainty grid, does not allow for any uncertainty on growth in this assessment, or in the limited data used to inform growth in this model.

5.5.5 Age structured production model

The results from the ASPM for absolute and relative female spawning biomass are shown in Figure 60. The ASPM indicated a wide range of scale and depletion outcomes for these three alternate models. Recruitment is also shown for these three alternate models in Figure 61, both in terms of absolute numbers and recruitment deviations. The recruitment deviations for the ASPM are set to zero, so the variation in absolute recruitment for this model is solely due to the reduction in spawning biomass, over the time series. The diagnostic model shows more variation in the early recruitment deviations, as this model attempts to fit the population age structure that is observed as soon as the size composition data enters the model. The ASPM-R is only using recruitment to fit to the CPUE series, and is not constrained by also needing to fit to the size composition data.

Fits to the Australian and New Zealand CPUE series for these same three alternate models are shown in Figure 62. The ASPM is unable to fit these indices well, estimating large additional standard error on the CPUE and fitting a straight line through each of these indices. This model estimates such a large population scale that the catches have little impact on the spawning biomass, and with little overall trend in either series, this suggests that the ASPM cannot get sufficient information from the CPUE alone to give much information about population scale. In contrast, the ASPM-R shows an excellent fit to the Australian CPUE, with very small total standard error estimated for this index, but produces a less impressive fit to the New Zealand CPUE, with large additional standard error estimated to achieve any kind of fit to this index, giving the worst fit to this index of the three alternate models. The diagnostic model falls midway between these two extremes, with better fits than the ASPM to both series, and, compared to the ASPM-R, produces an improved fit to the New Zealand index, ansd simultaneously, a worse fit to the Australian index.

Estimates of the recruitment scale parameter and of initial spawning biomass, are uncertain and quite different for all three alternate models (Figure 63), with the uncertainty particularly high for the initial spawning biomass for the ASPM.

The ASPM analysis again confirms that the CPUE indices used in the diagnostic model are not informative in estimating the population scale. The ASPM fit does not reproduce the trend in the index, suggesting that the combination of catch data and production dynamics alone is insufficient to explain the observed changes in relative abundance. The ASMP-R fits the Australian index very well, but is only able to do that with a very low estimation of population scale, probably too low to be plausible for a fishery of this size.

This suggests that more weight should be placed on size composition data, or other data sources, for trend estimation, as is done to an extent already in the diagnostic model, or that perhaps the

quality and spatial representation of the indices used should be reconsidered.

5.6 Sensitivities and structural uncertainty

5.6.1 Sensitivities

Results of the one-off sensitivities that were incorporated into the uncertainty grid, are shown in figures listed below, separated by each axis of the uncertainty grid. These figures comparing estimates of initial female spawning biomass, the time series of F/F_{MSY}, the female spawning biomass time series, and the relative female spawning biomass time series, relative to SB_0. The full uncertainty grid (Table 2), with sensitivity figures, included:

1. region 2 CPUE index used (**NZ** and PICT) (Figure 64);
2. steepness ($h = 0.7$, **0.8** and 0.9) (Figure 65);
3. proportion of recruitment to region 1 compared to region 2, to roughly match the ratio of relative area of the two regions, or to match the ratio of the relative catch by region in the last 20 years (**1:3**, 1:4) (Figure 66);
4. movement, (**diagnostic movement rates**, based on Patterson et al. (2021), diagnostic movement rates halved between regions 1 and 2, diagnostic movement rates halved between regions 2 and 1) (Figure 67);
5. data weighting on the weight composition data, for the three fisheries with estimated selectivity and weight composition data, (**Francis weighting**, doubling and halving the Francis weights) (Figure 68);
6. data weighting on the length composition data, for the two southern fisheries with estimated selectivity and length composition data, (**Francis weighting**, doubling and halving the Francis weights) (Figure 68); and
7. natural mortality (**estimated (internal)**, fixed (external) using the approach of Hamel and Cope (2022)) (Figure 69).

A third CPUE sensitivity was conducted using the EU CPUE replacing, the NZ index as the only index used for region 2. Given some issues discovered during the construction of this index (Kim et al., 2025), and concerns about the reliability of the last two years of this index, the data points from the years 2022 and 2023 were dropped from the EU index in this sensitivity. The resulting model had very poor fits to the EU index used, with the model achieving the best likelihood by fitting the Australian index well, at the expense of fits to the EU index. The resulting spawning biomass trajectories from the diagnostic model and from the EU sensitivity were very also similar, (Figure 65), despite the different trends in the NZ and EU indices, shown in the blue and green dots in Figure 3. In light of this outcome, showing very little difference between two potential categories for the CPUE index grid axis, and given the concerns about the reliability of the EU index, the

sensitivity of using the EU index was not included in the uncertainty grid.

For some of these sensitivities, the impact on the estimated spawning biomass trajectory and the fishing mortality was fairly minor, with the largest effects coming from the relative data weighting, especially the change to the weighting on the weight composition data, the fixed mortality and the change to steepness, with the model apparently least sensitive to changes in the movement rates between regions 1 and 2. While alternative ranges for some of these parameters could have been explored, computational limitations on the size of the grid and time constraints to explore these sensitivities in more detail required some pragmatic decisions to be made. Hopefully, the resulting grid covers the structural or model uncertainty adequately. Yes, there were more dragons here too.

An additional sensitivity with fixed, externally estimated growth (Figure 70) was included as an axis in an initial uncertainty grid, but this sensitivity consistently produced very high female spawning biomass estimates and low estimates of F/F_{MSY} and with poor fits and worse likelihood components for the weight composition data. This sensitivity was not considered to give plausible results, so this initial axis was removed from the final uncertainty grid. Ideally, growth uncertainty would be incorporated in the grid, but the available alternative external growth option did not seem plausible and there was insufficient time to explore a plausible alternative growth axis.

One final sensitivity was conducted, in which all data from subregion 2N was completely removed from the assessment (Figure 71). While this sensitivity estimated a much smaller overall scale to the population, this is not surprising given the catches taken in subregion 2N, especially in the last 25 years. Perhaps surprisingly, the time series of F/F_{MSY} and the relative population time series was very similar for the diagnostic model and this sensitivity of excluding this subregion from the model.

5.7 Structural uncertainty results

After taking considerable care and attention in developing a stable, well behaved diagnostic model, all 360 models in the structural uncertainty grid were found to have positive definite Hessians. Of these 360 grid models, 336 models produced a final gradient < 0.001. Of the 24 models with gradients > 0.001 the largest gradient was 0.0041. These models were checked to see if they were concentrated in one section of the grid, all with similar grid options selected, or on one axis of the grid. In fact, they were widely spread among several axes simultaneously, and showed no sign of any obvious pathological behaviour. Given that the largest gradient was still < 0.005, this gradient selection criterion was relaxed to leave all 360 model in the grid. Hence it was unnecessary to remove any models from the uncertainty grid, on convergence grounds. Results of the structural uncertainty grid are summarised in a number of figures and tables.

All models in the grid were run with the maximum function evaluations argument set to 500, (`> ss3 -maxfn 500`), to increase the default number of function evaluations at each phase and to increase the opportunity for the model to converge. In practice, setting this value to 200 would

probably have been sufficient, on inspection of the outputs from the uncertainty grid runs.

5.7.1 Grid axes with model uncertainty only

The grid outputs were initially scrutinised, to investigate the effects of model uncertainty, without considering estimation uncertainty. These results are summarised with box and whiskers plots, for F/F_{MSY} (Figure 72) and $SB_{\text{recent}}/SB_{\text{MSY}}$ (Figure 73), for all 360 models in the uncertainty grid, with each subplot showing the range of all of models in the uncertainty grid, separated by grid axis components, and with the diagnostic model highlighted in dark grey. In this standard box and whiskers plot, the black line indicates the median, the boxes give the interquartile, percentile ranges from 25% to 75%, and the whiskers reach up to 1.5 times the interquartile range, or to the largest (or smallest) data point, whichever produces shorter whiskers. Any further outlying data points are plotted individually.

The most influential axes of the grid are steepness, mortality and data weighting, particularly for the weight composition data, (Figure 72) and $SB_{\text{recent}}/SB_{\text{MSY}}$ (Figure 73). Lower values for steepness suggest higher values for $SB_{\text{recent}}/SB_{\text{MSY}}$ and lower values of F/F_{MSY}. The external fixed mortality option also suggests for $SB_{\text{recent}}/SB_{\text{MSY}}$ and lower values of F/F_{MSY}, as does halving the data weighting given to the weight composition data.

The mean from the CPUE axes of the uncertainty grid is almost unchanged between the two CPUE options for F/F_{MSY}, (Figure 72), although the NZ CPUE, as used in the diagnostic model, has slightly more uncertainty for F/F_{MSY} than the PICT CPUE. The mean of the CPUE axes of the uncertainty grid is marginally higher for the diagnostic CPUE (NZ) than for the PICT CPUE for $SB_{\text{recent}}/SB_{\text{MSY}}$ (Figure 73). The effect of the change to recruitment proportion is relatively small. Halving the movement rate from region 2 to 1 doesn't change the mean of F/F_{MSY}, but increases the variance slightly for F/F_{MSY} and reduces the variance slightly for $SB_{\text{recent}}/SB_{\text{MSY}}$. Halving the movement rate movement from region 1 to 2 increases the mean of F/F_{MSY} marginally.

5.7.2 Time series with model and estimation uncertainty

Combining model uncertainty and estimation uncertainty for each grid model, a range of key output quantities and metrics were plotted with ribbon plots. These show the median, along with 50%, 80% and 90% quantile ranges (showing the spread of values between 25%–75%, 10%–90% and 5%–95% respectively), with the quantiles calculated from the full range of 360,000 values for each time point in the time series.

Ribbon plots were constructed for the following quantities of management interest:

- F/F_{MSY} (Figure 74)
- recruitment (Figure 75)
- SB (Figure 76)

- SB/SB_{MSY} (Figure 77)

Kobe plots were produced showing the estimates of $F_{\text{recent}}/F_{\text{MSY}}$ and SB/SB_{MSY}, across all models in the structural uncertainty grid, summarising the results for each of the 360 models in the structural uncertainty grid, with no estimation error, for the recent period (2020-2023) (Figure 78). Another Kobe plot was also produced, with contour plots summarising the full uncertainty grid from 360,000 model runs, including structural and estimation uncertainty, with 50%, 80% and 90% quantile ranges, and with the median of all results shown as the red cross (Figure 79). Finally a dynamic Kobe plot with a time series was produced, summarising the results for the 2025 diagnostic model over the model period, from 1952-2023 (Figure 80). The larger red point is the estimated 2023 status, with the trajectory through this phase space indicated by labelled years, arrows and the changing colour of the trajectory line. All three of these Kobe plots indicate that the stock is estimated to be in the "not overfished" and "not overfishing" green zone of the Kobe plot, and to have been in this quadrant for the whole history of the fishery.

Majuro plots were produced showing the estimates of $F_{\text{recent}}/F_{\text{MSY}}$ and $SB_{\text{recent}}/SB_{F=0}$, across all models in the structural uncertainty grid, summarising the results for each of the 360 models in the structural uncertainty grid, with no estimation error, for the recent period (2020-2023) (Figure 81). A dynamic Majuro plot was also with a time series was produced, summarising the results for the 2025 diagnostic model over the model period, from 1952-2023 (Figure 82). The larger red point is the estimated 2023 status, with the trajectory through this phase space indicated by labelled years, arrows and the changing colour of the trajectory line.

The means and quantiles across the 360 models in the grid for all of the reference points and other quantities of interest are shown in Table 4. For key management quantities ($SB_{\text{recent}}/SB_{\text{MSY}}$ and $F_{\text{recent}}/F_{\text{MSY}}$), the combination of structural and estimation uncertainty, using 360,000 model runs, was used to include the additional estimates of estimation uncertainty for management advice, and also included in Table 4.

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

- The grid contains 360 models with a moderate range of estimates of stock status and reference points, which suggest that the stock is slightly less depleted (higher $SB_{\text{recent}}/SB_{F=0}$) and with a higher female spawning biomass than the estimates from the 2021 assessment (Table 4). Both $F_{\text{recent}}/F_{\text{MSY}}$ and $SB_{\text{recent}}/SB_{\text{MSY}}$ are estimated to be lower in the 2025 stock assessment, $SB_{\text{recent}}/SB_{\text{MSY}}$ has a median value of 2.37 in the 2025 assessment, compared with a median value of 3.61 in the 2021 assessment.
- The most influential axes of the grid are steepness, mortality and data weighting, particularly for the weight composition data (Figure 72).
- Estimates of fishing mortality were relatively insensitive to the CPUE option used, recruitment proportion or movement option assumed (Figure 72).

- Both SB and $SB_{\text{recent}}/SB_{\text{MSY}}$ and show a notable decline since the late 1990s through to 2010, followed by a slight recovery until 2015 and a decline since then (Figure 76 and Figure 77).
- None of the models in the structural uncertainty grid have $SB_{\text{recent}}/SB_{\text{MSY}} < 1$, with a minimum value of 1.54, and the quantile range from the 10^{th} to the 90^{th} percentile is 1.88–3.34 (Table 4). When estimation uncertainty is included, the resulting distribution has a negligible probability 0.01% of $SB/SB_{\text{MSY}} < 1$, and the quantile range from the 10^{th} to the 90^{th} percentile is 1.80–3.37 (Table 4).
- All models in the structural uncertainty grid show exploitation to be below F_{MSY}. This also holds when estimation uncertainty is included, and the median $F_{\text{recent}}/F_{\text{MSY}}$ is 0.27, and the quantile range from the 10^{th} to the 90^{th} percentile is 0.16–0.41 (Table 4).

6 Discussion

6.1 General remarks

The 2025 SWPO swordfish stock assessment features a change of stock assessment software, from MULTIFAN-CL to Stock Synthesis, and multiple changes to the structure and assumptions used in the 2021 stock assessment. This results in a simpler and more robust model, which appropriately takes account of variation in the size of fish caught by longitude, incorporates some age data directly into the assessment and is less reliant on length composition data. While there are still features that could be improved in future assessments, these recent changes all appear to lead to improvements in this SWPO swordfish stock assessment. However, there are still difficult questions to address in future, relating to stock structure, regional boundaries and the stock assessment boundaries. The key weakness in this model is an inability to estimate the population scale without large uncertainty bounds.

In general terms relating to management advice, the outcomes of this assessment are broadly consistent with the 2021 stock assessment, and are also consistent with the series of SWPO swordfish stock assessments that were developed in preceding years. The CPUE indices do not appear to contain a strong signal, either on population scale, or a signal which suggests a large change in spawning biomass. The catch history in the last 30 years, the relative stability in these CPUE indices and the continued relatively high catches, compared to historical levels, all suggest that there is little sign of the stock being heavily depleted, which also matches the results produced by the full stock assessment.

6.2 Changes to the previous assessment

The major changes from the 2021 assessment assessment include the following.

1. Conversion from a 1-sex model to a 2-sex model.
2. Major revision of the length data inputs, including more scrutiny and filtering and separation of port collected and onboard observer collected data for PICT fisheries, and subsequent filtering of port data, for these fisheries where both data types existed.
3. Separation of all fisheries into distinct fisheries in each subregion, so no fisheries were allowed to span multiple subregions. This allows separate selectivity to be estimated latitudinally, which is especially important for the southern subregions, which typically catch larger fish, but where these samples often were "lost" in fisheries merged with central subregions (with much higher catch of smaller fish, typically) in the 2021 assessment.
4. Changing from age-based selectivity to length-based selectivity.
5. Constraining selectivity options to logistic or double normal, in particular preventing bimodal selectivity patterns, as were allowed to be estimated, and were actually estimated for some

fisheries in the 2021 assessment using age-based selectivity with splines.

6. Very large and somewhat subjective initial downweighting of length composition data for most fisheries. Followed up by fixing selectivity, and setting the length composition weighting to zero, for some fisheries and further downweighting of length composition data for other fisheries.
7. Inclusion of sex specific conditional age-at-length data within the assessment and using these data to help estimate growth.
8. Estimating the scale of the Lorenzen mortality form, using recent approaches of Hamel and Cope (2022); Hoyle (2022).
9. Switching to an updated length-weight relationship based partly on a new dataset and also more careful filtering of old data (Macdonald et al., 2025).
10. The use of a quarterly model time step, rather than an annual time step to allow greater resolution in modelling the growth curve, especially for the younger faster growing fish.
11. The adoption of variable bin widths for the weight composition data, using a feature in Stock Synthesis which is not presently available in MULTIFAN-CL.
12. The use of parameters to estimate additional variance to the indices, to balance the input and output variances for the CPUE indices.
13. The successful application of iterative reweighting to both the size composition data and the conditional age-at-length data.
14. Conversion of stock assessment platform from MULTIFAN-CL to Stock Synthesis.

Due to the focus on these substantial structural changes, a standard grid approach was used to characterise uncertainty in management quantities. This is in contrast to the more sophisticated approach pioneered in 2021 (Ducharme-Barth et al., 2021a; Ducharme-Barth and Vincent, 2022), of creating a model ensemble based on prior distributions to characterise uncertainty. The ensemble approach appears to be a more suitable approach, albeit a little complex and more difficult to implement technically.

An additional four years of data were incorporated into this assessment, from 2020–2023, including new catch, CPUE and size composition data. The simplification of the assessment structure resulted in a model that only estimated 21 main parameters, plus an additional 53 recruitment deviations. Attempts to estimate more parameters led to model instability or implausible model outcomes. Even with such a small number of parameters to estimate, model run times were considerable longer than the MULTIFAN-CL model developed in 2021, with run times of up to four hours not unusual for many models tested and run during the development of this assessment.

While the treatment of uncertainty was perhaps a step backward from 2021, in reverting from an

uncertainty ensemble to an uncertainty grid, the other model developments should be considered a minor revolution for this assessment. These were considered necessary due to the apparent lack of information about population scale in the index data, the quality of the length data available to this assessment, the limited age data, the problems associated with conversion factors required to convert the input data into a common form, the mix of catch records in numbers and weight, and the long term collection of weight data, when good length data is likely to be more useful to the assessment, especially when using length based selectivity.

As with any change, there was a certain amount of associated chaos, making it impossible to conduct an orderly transition from the 2021 assessment through a standard stepwise approach, starting with the 2021 diagnostic model and, in a series of well defined steps, arriving at the 2025 diagnostic model. It took much longer than expected to achieve a viable diagnostic model and this contributed to the late delivery of this assessment. Hopefully the thorough diagnostic model development was worthwhile.

The new model is remarkably stable, including an uncertainty grid with 360 models which contained positive definite Hessians, in contrast to the uncertainty ensemble used in 2021, where many models had to be filtered out due to convergence issues. While the 2025 model is perhaps simpler, it appears to be more robust. Jitter analysis failed to find any improvement to the solution obtained for the diagnostic model.

6.3 Challenges

There are clearly still plenty of issues to explore with this model, particularly in relation to model structure, stock structure, information on spatial variation in biology (such as growth, recruitment, reproduction, movement), spatial structure, including a very large and disparate region 2, and heavy fishing on either side of the NE corner of the assessment area. Hopefully this assessment provides a useful direction for the model development, and even some advances and improvements. Interestingly the main conclusions are not that different to conclusions drawn in the first SWPO swordfish assessment developed 20 years ago (Kolody et al., 2006).

While the assessment has some interesting features, including CPUE which is uninformative on population scale, there are some intriguing features. This is highlighted by some of the diagnostics and the likelihood profile, indicating that the signal in the weight data, especially from fishery 05.AU.1C was critical in getting a model that was presentable. The recent signal in these data, indicating that the average weight of fish caught has reduced consistently in the last five years perhaps raise more questions than it answers. Is that a fish down of the population, with many of the older fish apparently now either less available, or possibly dead, or a large recruitment and an influx of young fish dominating the older larger fish. Were there changes to fishing or sampling practices that may explain a feature that the model assumes is a change in the population?

The length data in the southern fisheries appears to be particularly important in this assessment,

with larger fish sampled in these subregions than further north. However sample sizes from these fisheries are very small, and one of the more important fisheries, 14.NZ.2S is apparently no longer collecting any length samples, having switched from observer sampling of lengths to sampling of weights in fish markets. Determining if this length sampling program can be reinstated, or not, may be important to future assessments of this stock.

The appearance of spikes of small fish is observed in size data for many fisheries in the model, so often that it suggested these small fish are real samples. However, the lack of follow through of any potential large cohorts presents substantial problems for this assessment, and possibly any locally sampled recruitment spikes that represent new recruitment could disperse quickly into other areas that may not be fished as heavily, or possibly outside the assessment boundaries, which may even be fished more heavily, making it very difficult to track these "potential large cohorts" in the data or in the model.

Likelihood profiles suggest there is some conflict in the data sources used in this model. The ASPM analysis supports the conclusion that there is little information on population scale contained in the index data. The retrospective analysis indicates that the model is still sensitive to changes in recent data, which indicates either the model is sensitive to data inputs, or some recent data is unusual and may warrant greater scrutiny, or there have been some recent systemic changes (fishing practices, population processes, environmental process), or some combination of all of the above. These diagnostics indicate that there is fertile ground to explore for whoever is so lucky as to conduct the next SWPO swordfish stock assessment.

6.4 Main assessment conclusions

The 2025 SWPO swordfish stock assessment estimates that, at the stock-wide scale, for all models in the uncertainty grid, the median $SB_{\text{recent}}/SB_{\text{MSY}}$ is well above 1.0, with a value of 2.37, showing that SB is well above SB_{MSY}, (Figure 77, Table 4), and the median $F_{\text{recent}}/F_{\text{MSY}}$ is well below 1.0, with median value of 0.27, showing that F_{recent} is also well below F_{MSY} (Figure 74, Table 4). These two key reference points are calculated incorporating both structural and estimation uncertainty.

The estimated stock status, relative to SB_{MSY} across the whole model region shows a slow, gradual decline until the 1990s, with an initial increase, followed by a period of faster decline, through to about 2009, followed by a brief recovery to 2015 and over the last decade the stock has continued to decline (Figure 77).

Overall, the outcomes of this assessment suggest that the swordfish stock in the SWPO is not overfished and is not undergoing overfishing.

6.5 Recommendations for further research

1. Increase the biological sampling of otoliths to build greater coverage of conditional age-at-length data and application or otolith-based methods to explore population structure (i.e.

otolith chemistry). Increase the quantity of tissue samples collected for expanded genetic studies. Age data are some of the most informative data for this model and the current age dataset is limited spatially and temporally. More age data would likely be very informative for this model.

2. Collection of appropriately sampled close kin mark recapture data. This potential data source and modelling tool could potentially offer far more information to inform this assessment than all of the other data sources, that seemingly are unable to give a great deal of confidence in the assessment of the population scale.
3. Robust age validation procedures to enhance accuracy and confidence in the conditional age-at-length data.
4. Duplicate readings of otoliths should be obtained to enable age reading error to be estimated.
5. Further research and data collections to inform growth parameters, and any regional variation across different Pacific regions is essential to refine the assessment.
6. Continued and increased collection of key representative size composition data could help improve the model outcomes.
7. Explore the influence of environmental factors on recruitment. By exploring these connections, we can gain valuable insights into population dynamics and identify potential vulnerabilities to environmental changes.
8. Review the length and weight composition data and explore further models to standardise these composition inputs to assessments.
9. Collaborative work with colleagues from the EU to diagnose potential problems with the current EU CPUE index.
10. Additional exploration of alternative abundance indices, CPUE or otherwise.
11. Regional agreements on standard length and weight measurements to be applied to best inform the stock assessment. Minimising the use of conversion factors will minimise the propagation in any errors associated with conversion factors.
12. Full review of conversion factors used in the assessment, and consideration of the devlopment of regional or flag-based conversion factors. Conversions from processed length or weight to the standarised units used in the assessment (whole weight and lower jaw-fork length) are important.
13. A review of data collection preferences for stock assessment purposes would be helpful, especially composition data. Long term data collection programs are useful, but typically age data is more useful than length data, which is more useful than weight data.

14. Investigation into stock structure and connectivity with adjacent regions in the north east corner of the assessment region.

6.6 GitHub repository for diagnostic model and r4ss plots

A publicly accessible GitHub repository has been created, for the 2025 swordfish diagnostic model, containing a single zip file, `ss3.zip`, which contains the following standard Stock Synthesis files.

- input files (four files: `starter.ss, control.ss, swo2025.ctl, swo2025.dat`)
- output files (multiple files)
- `r4ss` plots sub-directory, containing an easy to navigate .html summary file: `_SS_output.html`
- likelihood profile subdirectory with plots for $\log(R_0)$
- likelihood profile subdirectory with plots for M

Edits and other additions will be made to this repository, but the `ss3.zip` file will remain unchanged in that location.

The link to this GitHub repository is:
https://github.com/PacificCommunity/ofp-sam-swo-2025-diagnostic

To increase transparency, to explore model input setting, outputs and to examine more plots than could be included in this report, please read this paper in conjunction with files available through this link.

7 Acknowledgments

We thank the various fisheries agencies and regional fisheries observers for their support with data collection, provision and preparatory analysis. We thank participants at the preparatory stock assessment workshop (PAW) for their contributions to the assessment. We especially thank the SPC data management team for their hard work and support to provide the data fuel for the assessment. We are grateful to Fabrice Bouyé for ensuring smooth operation of the computing resources required to do the assessment.

Many people contributed by answering questions and giving their time. Listing some of the contributors include: Dale Kolody (the first person I ever talked swordfish with, about 20 years ago, when he was developing the first SWPO swordfish assessment); current CSIRO staff including, Rich Hillary, Paige Eveson, Jess Farley, Scott Cooper, Brad Moore, Ashley Williams; the Stock Synthesis user community, including Rick Methot, Ian Taylor, Mark Maunder, Michael Schirripa; members and attendees at the PAW (in person and on-line), all of them, but notably Simon Hoyle, Nicholas Ducharme-Barth, Leyla Knittweis, James Larcombe, Felipe Carvalho, Meg Oshima, Michelle Sculley. I apologise to others who I may have inadvertently left off this list.

Particularly noteworthy contributions came from: Mark Maunder, whose small but vital contribution helped turn this assessment on its head when all looked gloomy and the assessment looked like it would never see the light of day, offering sage advice on simplifying the assessment and further downweighting length composition data; Simon Hoyle whose assistance and help on details with the New Zealand fishery and technical advice and suggestions were critical, especially in dealing with Lorenzen mortality, but on many other aspects as well; Nicholas Ducharme-Barth for sharing the journey and his experience and advice as the last assessment author; Meg Oshima for helping me set up and run the ASPM analysis in Stock Synthesis; my incredibly supportive colleagues from Zone 259, Jo Potts and Jed Macdonald - Zone 259 rocks, you guys; the wider SAM and FEMA teams and the administrative support teams at SPC.

Thanks are due to so many people who contributed to this assessment, formally and informally. For some, this was there first time working on swordfish and for others it may well be their last. All involved showed energy and enthusiasm for working on this stock assessment after a range of previous experience with this species and in this region. A collaborative and co-operative hard working team is a wonderful thing, but there are numerous support staff who make this possible and have tolerated way too much discussion about swordfish. Many friends and colleagues have helped support this work by just being there, going for a run, mountain biking, hiking, hanging out, shooting the breeze, talking about dragons and revolutions - you certainly know who you are. Others have contributed in the background, either through data collection and processing or administration. To the team who are formally authors, it has been a great journey and your input and enthusiasm has been enlightening and contagious. I hope we are all richer as a result. As lead author, I can confidently say that this is the best (swordfish) stock assessment I have ever worked

on. I hope you think so too. It is also my first stock assessment, or perhaps some much bigger number - but certainly my first for swordfish anyway. However, as a team, I think we have again broken new ground and hopefully set some new standards, that teams to come can build on, just as we have started from a solid basis built by those who preceded us. While I am sure there are many aspects of this assessment that we could have improved, the one aspect that don't think I can fault any of my co-authors on is their good humour, patience, resilience, tenaciousness and plain hard work. It has been a pretty long journey at times – but with a much more satisfying end product than I expected two months ago! I hope others find it at least half as good as we like to think it is!

8 References

Abascal, F. J., Mejuto, J., Quintans, M., and Ramos-Cartelle, A. (2010). Horizontal and vertical movements of swordfish in the Southeast Pacific. *ICES Journal of Marine Science*, 67(3):466–474.

Alvarado Bremer, J. R., Hinton, M. G., and Greig, T. W. (2006). Evidence of spatial genetic heterogeneity in Pacific swordfish (Xiphias gladius) revealed by the analysis Of ldh-A sequences. *Bulletin of Marine Science*, 79:493–503.

Anon (2024). Summary Report: Twentieth regular session of the Scientific Committee. Technical report, The Commission for the Conservation and Management of Highly Migratory Fish Stocks in the Western and Central Pacific Ocean, Manila, Philippines, 14–21 August 2024.

Bigelow, K. A., Boggs, C. H., and He, X. (1999). Environmental effects on swordfish and blue shark catch rates in the US North Pacific longline fishery. *Fisheries Oceanography*, 8(3):178–198.

Bull, B., Francis, R., Dunn, A., McKenzie, A., Smith, M., Bian, R., and Fu, D. (2012). CASAL(C++ algorithmic stock assessment laboratory) CASAL User Manual v2.30-2012/03/21. Technical Report NIWA Technical Report 135.

Burnett, C., Beckett, J., Dickson, C., Hurely, P., and Iles, T. (1987). A summary of releases and recaptures in the Canadian large pelagic fish tagging program 1961–1986. *Canadian Data Report of Fisheries and Aquatic Sciences*, 673:99 p.

Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., Kitakado, T., Yemane, D., Piner, K. R., Maunder, M. N., Taylor, I., Wetzel, C. R., Doering, K., Johnson, K. F., and Methot, R. D. (2021). A cookbook for using model diagnostics in integrated stock assessments. *Fisheries Research*, 240:105959.

Castillo-Jordan, C., Day, J., Teears, T., Davies, N., Hampton, J., McKechnie, S., Magnusson, A., Peatman, T., Vidal, T., Williams, P., and Hamer, P. (2024). Stock Assessment of Striped Marlin in the Southwest Pacific Ocean: 2024. Technical Report WCPFC-SC20-2024/SA-WP-03_Rev1.

Davies, N., Bian, R., Kolody, D., and Campbell, R. (2008). CASAL Stock Assessment for South-West-Central Paciific Broadbill Swordfish 1952-2007. Technical Report WCPFC-SC4-2008/SA-WP-07.

Davies, N., Pilling, G., Harley, S., and Hampton, J. (2013). Stock assessment of swordfish (Xiphias gladius) in the in the southwest Pacific Ocean. Technical Report WCPFC-SC9-2013/SW-WP-05, Pohnpei, Federated States of Micronesia.

Dell, J., Campbell, R., Hillary, R., and Preece, A. (2020). Standardised CPUE indices for the target species in the Eastern Tuna and Billfish fishery. Technical Report Working Paper to the 29th meeting of Tropical Tuna Resource Assessment Group held, 10-11 September 2020.

DeMartini, E. E., Uchiyama, J. H., and A, W. H. (2000). Sexual maturity, sex ratio, and size composition of swordfish, Xiphias gladius, caught by the Hawaii-based pelagic longline fishery. *Fisheries Bulletin*, 98:489–506.

Dewar, H., Prince, E. D., Musyl, M. K., Brill, R. W., Sepulveda, C., Luo, J., Foley, D., Orbesen, E. S., Domeier, M. L., Nasby-Lucas, N., Snodgrass, D., Michael Laurs, R., Hoolihan, J. P., Block, B. A., and Mcnaughton, L. M. (2011). Movements and behaviors of swordfish in the Atlantic and Pacific Oceans examined using pop-up satellite archival tags: Swordfish movements in the Atlantic and Pacific Oceans. *Fisheries Oceanography*, 20(3):219–241.

Ducharme-Barth, N., Castillo-Jordan, C., Hampton, J., Williams, P., Pilling, G., and Hamer, P. (2021a). Stock assessment of swordfish in the southwest Pacific Ocean. Technical Report WCPFC-SC17-2021/SA-WP-04.

Ducharme-Barth, N., Peatman, T., and Williams, P. (2021b). Background analysis for the 2021 stock assessment of southwest Pacific swordfish. Technical Report WCPFC-SC17-2021-SA-IP-07.

Ducharme-Barth, N. D. and Vincent, M. T. (2022). Focusing on the front end: A framework for incorporating uncertainty in biological parameters in model ensembles of integrated stock assessments. *Fisheries Research*, 255:106452.

Evans, K., Abascal, F., Kolody, D., Sippel, T., Holdsworth, J., and Maru, P. (2014). The horizontal and vertical dynamics of swordfish in the South Pacific Ocean. *Journal of Experimental Marine Biology and Ecology*, 450:55–67.

Evans, K., Kolody, D., Abascal, F., Holdsworth, J., Maru, P., and Sippel, T. (2012). Spatial dynamics of swordfish in the South Pacific Ocean inferred from tagging data. Technical Report WCPFC-SC8-2012-SA-IP-05.

Farley, J., Clear, N., Kolody, D., Krusic-Golub, K., Eveson, P., and Young, J. (2016). Determination of swordfish growth and maturity relevant to the southwest Pacific stock. Technical Report WCPFC-SC12-2016/ SAWP-11, Bali, Indonesia, 3–11 August 2016.

Finucci, B. and Moore, B. (2025). Characterisation and CPUE indices for swordfish (*xiphias gladius*) from the New Zealand surface longline fishery 1993 to 2023. Technical Report WCPFC-SC21-SA-IP-11.

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

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

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

Grewe, P., Feutry, P., Fostser, S., Aulich, J., Lansdell, M., Cooper, S., Clear, N., Eveson, P., Fernando, D., Darnaude, A., Nikolic, N., Fahmi, Z., Marsac, F., Farley, J., and Davies, C. (2020). Genetic population structure of sailfish, striped marlin, and swordfish in the Indian Ocean from the PSTBS-IO project. Technical Report IOTC-2020-WPB18-09, 18th Working Party on Billfish, Indian Ocean Tuna Commission.

Hamel, O. S. and Cope, J. M. (2022). Development and considerations for applications of a longevity-based prior for the natural mortality rate. *Fisheries Research*, 256:106477. Publisher: Elsevier BV.

Hamer, P. (2025). Summary report from the SPC Pre-assessment Workshop – April 2025. Technical Report WCPFC-SC21-SA-IP-01.

Hamer, P., Schneiter, E., Vidal, T., and Williams, P. (2025). Progress report: Review and reconciliation of size data collected in the WCPFC-CA for stock assessment purposes (WCPFC Project: 127). Technical Report WCPFC-SC21-ST-WP-02.

Holdsworth, J. C., Sippel, T. J., and Saul, P. (2007). An investigation into swordfish stock structure using satellite tag and release methods. Technical Report WCPFC-SC3-2007-BI SWG/WP-3.

Hoyle, S. (2022). Natural mortality ogives for the Indian Ocean bigeye tuna stock assessment. Technical Report IOTC-2022-WPTT24(DP)-1y.

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

Ijima, H. and Jusup, M. (2023). Tuna and billfish larval distributions in a warming ocean.

Kim, J., Na, H., Park, Y.-G., and Kim, Y. H. (2020). Potential predictability of skipjack tuna (Katsuwonus pelamis) catches in the Western Central Pacific. *Sci Rep*, 10(1):3193.

Kim, K., Magnusson, A., and Hamer, P. (2025). Analysis of swordfish CPUE from the Spanish fleet in the Southwest Pacific Ocean. Technical Report WCPFC-SC21-2025/SA-IP-12.

Kolody, D., Campbell, R., and Davies, N. (2006). A Multifan-CL stock assessment for south-west Pacific swordfish 1952-2004. Technical Report WCPFC-SC2-2006/ME-WP-03.

Kolody, D., Campbell, R., and Davies, N. (2008). A Multifan-CL stock assessment of south-west Pacific swordfish 1952-2007. Technical Report WCPFC-SC4-2008/SA-WP-6.

Lorenzen, K., Camp, E., and Garlock, T. (2022). Natural mortality and body size in fish populations. *Fisheries Research*, 252:106327. Publisher: Elsevier BV.

Lu, C.-P., Smith, B. L., Hinton, M. G., and Alvarado Bremer, J. R. (2016). Bayesian analyses of Pacific swordfish (*Xiphias gladius*) genetic differentiation using multilocus single nucleotide polymorphism (SNP) data. *Journal of Experimental Marine Biology and Ecology*, 482:1–17.

Macdonald, J., Bell, L., Cunningham, M., Schneiter, E., Ghergariu, M., and Judd, P. (2025). Project 90 update: Better Data on Fish Weights and Lengths for Scientific Analyses. Technical Report WCPFC-SC21-ST-IP-03.

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

Maunder, M. N., Thorson, J. T., Xu, H., Oliveros-Ramos, R., Hoyle, S. D., Tremblay-Boyer, L., Lee, H. H., Kai, M., Chang, S.-K., Kitakado, T., Albertsen, C. M., Minte-Vera, C. V., Lennert-Cody, C. E., Aires-da Silva, A. M., and Piner, K. R. (2020). The need for spatio-temporal modeling to determine catch-per-unit effort based indices of abundance and associated composition data for inclusion in stock assessment models. *Fisheries Research*, 229:105594.

Mejuto, J., Garcia-Cortes, B., and Ramos-Cartelle, A. (2008). Reproductive activity of swordfish (it Xiphias gladius) in the Pacific Ocean on the basis of different macroscopic indicators. Technical Report WCPFC-BI-2008-WP-06.

Mejuto, J. and García-Cortés, B. (2014). Reproductive activity of swordfish *Xiphias gladius*, in the Atlantic Ocean inferred on the basis of macroscopic indicators. *Revista de Biología Marina y Oceanografía*, 49(3):427–447.

Methot, Richard D., J. and Taylor, I. G. (2011). Adjusting for bias due to variability of estimated recruitments in fishery assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, 68(10):1744–1760.

Methot, Richard D., J. and Wetzel, C. R. (2013). Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research*, 142:86–99.

Methot, Richard D., J., Wetzel, C. R., Taylor, I. G., Doering, K. L., Perl, E. L., and Johnson, K. F. (2025). Stock Synthesis User Manual Version 3.30.23.2. Technical Report NOAA Fisheries, Seattle, WA.

Mohn, R. (1999). The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. *Ices Journal of Marine Science*, 56(4):473–488.

Moore, B. (2020). Biology, stock structure, fisheries, and status of swordfish, Xiphias gladdius, in the Pacific ocean - a review. Technical Report WCPFC-SC17-2021-SA-IP-08.

Neubauer, P. (2025). Exploring the potential for observer CPUE for southwest Pacific swordfish (Xiphias gladius) and striped marlin (Kajikia audax). Technical Report WCPFC-SC21-SA-IP-13.

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

Pacific Fishery Management Council, P. (2024). Terms of Reference for the Groundfish Stock Assessment Review Process for 2025-2026. Technical Report Pacific Fishery Management Council, Portland, Oregon.

Patterson, T., Evans, K., and Hillary, R. (2021). Broadbill swordfish movements and transition rates across stock assessment spatial regions in the western and central Pacific. Technical Report WCPFC-SC17-2021/SA-IP-17.

Peatman, T., Day, J., Castillo-Jordan, C., and Hamer, P. (2025). Analysis of longline size frequency data for the 2025 southwest Pacific swordfish and striped marlin assessments. Technical Report WCPFC-SC21-SA-IP-14.

Poisson, F., Gaertner, J.-C., Taquet, M., Durbec, J.-C., and Bigelow, K. (2010). Effects of lunar cycle and fishing operations on longline-caught pelagic fish: fishing performance, capture time, and survival of fish. *Fisheries Bulletin*, 108:268–281.

Punt, A. E. (2017). Some insights into data weighting in integrated stock assessments. *Fisheries Research*, 192:52–65.

Reeb, C. A., Arcangeli, L., and Block, B. A. (2000). Structure and migration corridors in Pacific populations of the Swordfish Xiphius gladius , as inferred through analyses of mitochondrial DNA. *Marine Biology*, 136(6):1123–1131.

Sepulveda, C. A. and Aalbers, S. A. (2025). Swordfish horizontal movements in relation to stock structure within the eastern north pacific. *ICES Journal of Marine Science*, 82(7):fsaf098.

Takeuchi, Y., Davies, N., Fournier, D., Pilling, G., and Hampton, J. (2018). Testing MULTIFAN-CL developments for multispecies/multi-sex assessments, using SW Pacific swordfish. Technical Report WCPFC-SC14-2018/SA-IP-10, Busan, South Korea, 8-16 August 2018.

Takeuchi, Y., Graham, P., and Hampton, J. (2017). Stock assessment of swordfish (Xiphias gladius) in the southwest Pacific Ocean. Technical Report WCPFC-SC13-2017/SA-WP-13, Rarotonga, Cook Islands, 9-17 August 2017.

Then, A. Y., Hoenig, J. M., Hall, N. G., and Hewitt, D. A. (2015). Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. *Ices Journal of Marine Science*, 72(1):82–92.

Tracey, S. and Pepperell, J. (2018). Understanding the movement, behaviour and post-capture survival of recreationally caught swordfish from southeast Australia - a pilot study. Technical Report Final Report: FRDC Project No 2015–022, Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania.

Tremblay-Boyer, L. and Williams, A. (2024). Standardised CPUE indices for the target species in the eastern tuna and billfish fishery – 1998 to 2023. Technical Report Technical Report Working Paper presented to the 41st meeting of the Tropical Tuna Resource Assessment Group held 16-17 July 2024, Brisbane, CSIRO, Hobart., CSIRO, contact Ashley.Williams@csiro.au.

Valeiras, X., Mejuto, J., and Ruiz, M. (2008). Age and growth of swordfish (Xiphias Gladius) in the North Pacific. Technical Report WCPFC-SC4-2008-BI-WP-1.

Young, J. and Drake, A. (2002). Age and growth of broadbill swordsh (Xiphias gladius) from Australian waters. Technical Report Final Report: FRDC Project 2001/014.

Young, J., Drake, A., Brickhill, M., Farley, J., and Carter, T. (2003). Reproductive dynamics of broadbill swordfish, Xiphias gladius, in the domestic longline fishery off eastern Australia. *Mar. Freshwater Res.*, 54(4):315.

9 Tables

Table 1: Definition of fisheries, with labels indicating: fishery number; flag; and subregion, as well as selectivity and data availability information for the 2025 swordfish stock assessment. The Catch column indicates whether catch is recorded in numbers of fish (nos) or in tonnes (t). The Years column indicates the years in which composition data is available for each fishery, after the initial filtering of data. For years appearing in square brackets, [], this indicates data from this fishery was further downweighted to zero in the model, so effectively removing these data from the model entirely, with this data not even used to estimate an initial selectivity for that fishery, nor as part of a group of similar fisheries with shared selectivity. In this case, selectivity was obtained from the fishery listed in the mirrored column (Mirr). For years appearing in parentheses, (), this indicates data that were downweighted, due to data quality concerns, and largely used to estimate an initial selectivity, with these data then further downweighted by an additional factor of 0.001, with this composition data left in the model, albeit with very small weight, to minimise the influence of these data on other model parameter estimates. In some cases, when the word "fix" appears in the Selectivity column, this initial estimated selectivity is subsequently entered as a fixed selectivity in the model. In other cases, this selectivity is still estimated, but is shared with a selectivity estimated from a different fishery using data with much higher weighting, so the fisheries with Years in parentheses makes a largely insignificant contribution to the estimated selectivity. Qtrs indicates the number of quarters of size composition data available, after initial data filtering, and before any further downweighting to zero. In some cases, with years in square brackets, the composition data from these fisheries has no influence, and the available data is only used in the model to show the fit, or lack of fit, to this data from the (imposed) selectivity. The Mirr (or mirror) column lists the other fishery number that selectivity is shared with for this fishery, linking to the either the first or the most influential fishery within each group with shared selectivity. As well as indicating whether selectivity is fixed or estimated, the Selectivity column lists the shape of the selectivity curve, either asymptotic (logistic) or potentially dome-shaped (double normal).

Fishery	Catch	Years	Qtrs	Comp	Selectivity	Mirr
01.DW.1N	nos	[1998-2023]	12	len	fix dbl nrml	08
02.DW.1C	nos	(1991-2016)	7	len	fix dbl nrml	
03.DW.1S	nos	1991-2023	13	len	logistic	
04.AU.1N	nos	1997-2019	69	wt	dbl nrml	
05.AU.1C	nos	1997-2023	106	wt	dbl nrml	
06.AU.1S	nos	(1993,1997)	2	len	logistic	03
07.EU.1C	t	(2004)	2	len	fix dbl nrml	02
08.PICT.1N	nos	(2000-2014)	24	len	fix dbl nrml	
09.PICT.1C	nos	[]	0	len	fix dbl nrml	02
10.DW.2N	nos	(2010-2023)	46	len	fix dbl nrml	08
11.DW.2C	nos	(1988-2023)	10	len	fix dbl nrml	
12.DW.2S	nos	[1990-1994]	5	len	logistic	14
13.NZ.2C	nos	2006-2023	54	wt	dbl nrml	
14.NZ.2S	nos	1996-2021	26	len	logistic	
15.EU.2N	t	[2004]	2	len	fix dbl nrml	08
16.EU.2C	t	(2004-2009)	21	len	fix dbl nrml	11
17.EU.2S	t	[2005-2006]	3	len	logistic	14
18.PICT.2N	nos	[2006-2023]	40	len	fix dbl nrml	08
19.PICT.2C	nos	(2004-2021)	9	len	dbl nrml	16
Index fisheries						
20.AU.IDX.1	nos	1998–2023		wt	dbl nrml	05
21.NZ.IDX.2	nos	2004–2023		wt	dbl nrml	13

Table 2: Structural uncertainty grid for the 2025 SWPO swordfish stock assessment. Bold values indicate settings for the diagnostic model.

Axis	Levels	Option 1	Option 2	Option 3	Option 4	Option 5
CPUE	2	**AU & NZ**	AU & PICT			
Steepness	3	0.7	**0.8**	0.9		
Proportion recruitment by region	2	**1:3**	1:4			
Movement	3	**diagnostic**	Halve from 1 to 2	Halve from 2 to 1		
Data weighting	5	**Francis**	Halve weight	Double weight	Halve length	Double length
Natural mortality	2	**Estimated**	Fixed			

Table 3: Description of symbols used in the yield and stock status analyses.

Symbol	Description
C_{latest}	Catch in the last year of the assessment (2023)
F_{latest}	Average fishing mortality-at-age in the latest time period (2023)
F_{recent}	Average fishing mortality-at-age for a recent period (2019–2022)
F_{MSY}	Fishing mortality-at-age producing the maximum sustainable yield (MSY)
MSY	Equilibrium yield at F_{MSY}
$F_{\text{recent}}/F_{\text{MSY}}$	Average fishing mortality-at-age for a recent period (2019–2022) relative to F_{MSY}
$F_{\text{latest}}/F_{\text{MSY}}$	Average fishing mortality-at-age in the latest time period (2023) relative to F_{MSY}
SB_{latest}	Spawning biomass in the latest time period (2023)
SB_{recent}	Spawning biomass for a recent period (2020–2023)
TB_{latest}	Total biomass in the latest time period (2023)
TB_{recent}	Recent biomass for a recent period (2020–2023)
$SB_{F=0}$	Average spawning biomass predicted in the absence of fishing for the period 2011–2020
SB_{MSY}	Spawning biomass that will produce the maximum sustainable yield (MSY)
	spawning biomass predicted to occur in the absence of fishing for the period 2011–2020
$SB_{\text{latest}}/SB_{F=0}$	Spawning biomass in the latest time period (2023) relative to the average spawning biomass predicted to occur in the absence of fishing for the period 2011–2020
$SB_{\text{latest}}/SB_{\text{MSY}}$	Spawning biomass in the latest time period (2023) relative to that which will produce the maximum sustainable yield (MSY)
$SB_{\text{recent}}/SB_{F=0}$	Spawning biomass for a recent period (2020–2023) relative to the average spawning biomass predicted to occur in the absence of fishing for the period 2011–2020
$SB_{\text{recent}}/SB_{\text{MSY}}$	Spawning biomass for a recent period (2020–2023) relative to the spawning biomass that produces maximum sustainable yield (MSY)

Table 4: Summary of reference points over the 360 individual models in the uncertainty model ensemble, along with results incorporating estimation uncertainty. Note that these values do not include estimation uncertainty, except for the key management quantities (bottom two rows of the table).

Metric	Mean	Median	Min	10%ile	90%ile	Max
C_{latest}	5922	5926	5758	5846	5994	6071
SB_{latest}	47080	43738	26110	35742	62301	96686
SB_{recent}	48523	44994	27255	36729	64013	99654
TB_{latest}	118832	110466	65944	91676	155510	234896
TB_{recent}	118023	109628	65577	90625	154507	234147
F_{latest}	0.06	0.06	0.03	0.04	0.07	0.10
F_{recent}	0.06	0.06	0.03	0.04	0.07	0.10
SB_{MSY}	20039	19502	11521	13580	26557	38811
MSY	12078	11560	8189	9708	15339	22310
F_{MSY}	0.21	0.20	0.15	0.16	0.27	0.27
$F_{\text{recent}}/F_{\text{MSY}}$	0.28	0.28	0.11	0.18	0.38	0.53
$F_{\text{latest}}/F_{\text{MSY}}$	0.28	0.27	0.11	0.18	0.38	0.53
$SB_{\text{recent}}/SB_{\text{MSY}}$	2.48	2.33	1.54	1.88	3.34	4.10
$SB_{\text{latest}}/SB_{\text{MSY}}$	2.41	2.27	1.47	1.82	3.24	3.98
$SB_{\text{recent}}/SB_{F=0}$	0.52	0.50	0.37	0.46	0.58	0.67
$SB_{\text{latest}}/SB_{F=0}$	0.50	0.49	0.36	0.45	0.57	0.65
Including estimation uncertainty						
	Mean	Median	Min	10%ile	90%ile	Max
$F_{\text{recent}}/F_{\text{MSY}}$	0.28	0.27	0.00	0.16	0.41	0.93
$SB_{\text{recent}}/SB_{\text{MSY}}$	2.48	2.37	0.48	1.80	3.37	5.37

10 Figures

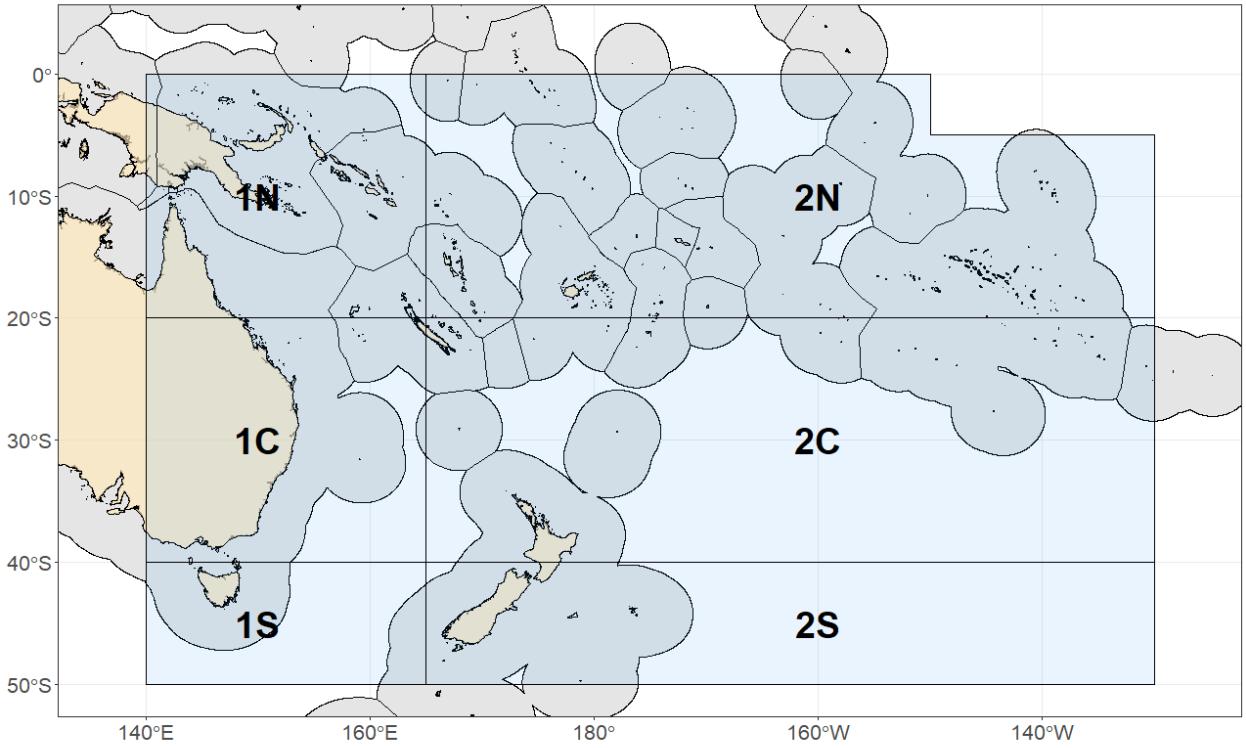


Figure 1: Spatial structure for the 2025 SWPO swordfish stock assessment. Region 1 (in the west) and region 2 (in the east) are the two main regions in the model, with each region further subdivided into northern, central, and southern subregions, which is used for fishery definitions in an area-as-fleets context.

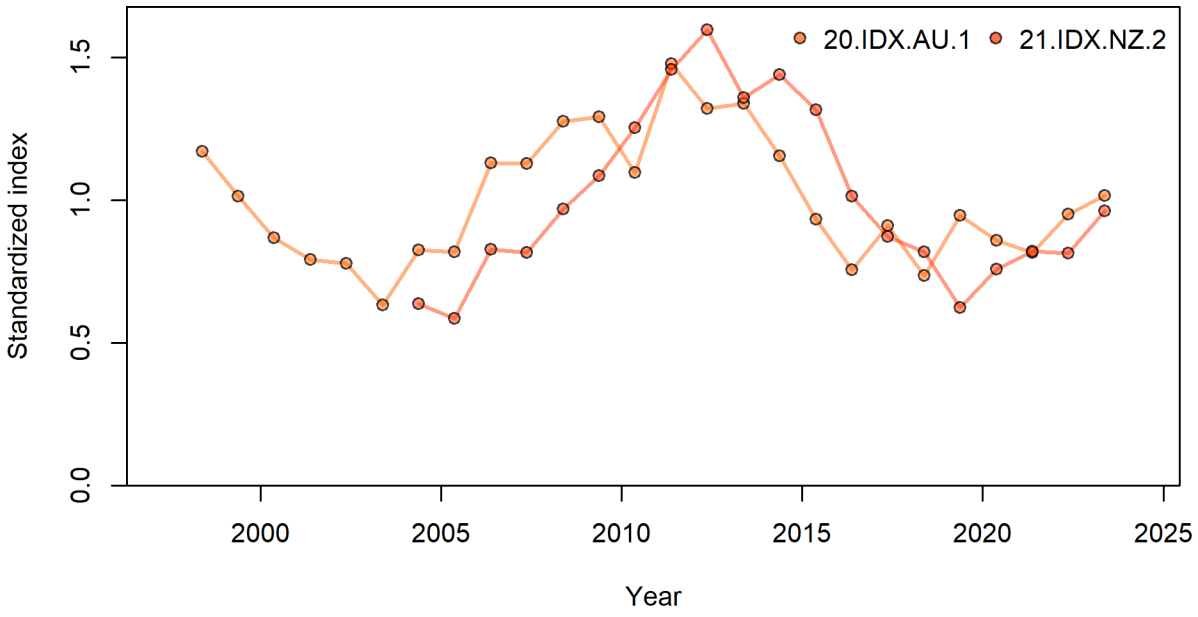


Figure 2: The two CPUE indices from the diagnostic model plotted on a normalised scale (mean of each series equals 1), enabling comparison of trends between these two time series. The Australian CPUE, from region 1 (20.IDX.AU.1), is shown in orange and the slightly shorter New Zealand CPUE, from region 2 (21.IDX.NZ.2), is shown in red.

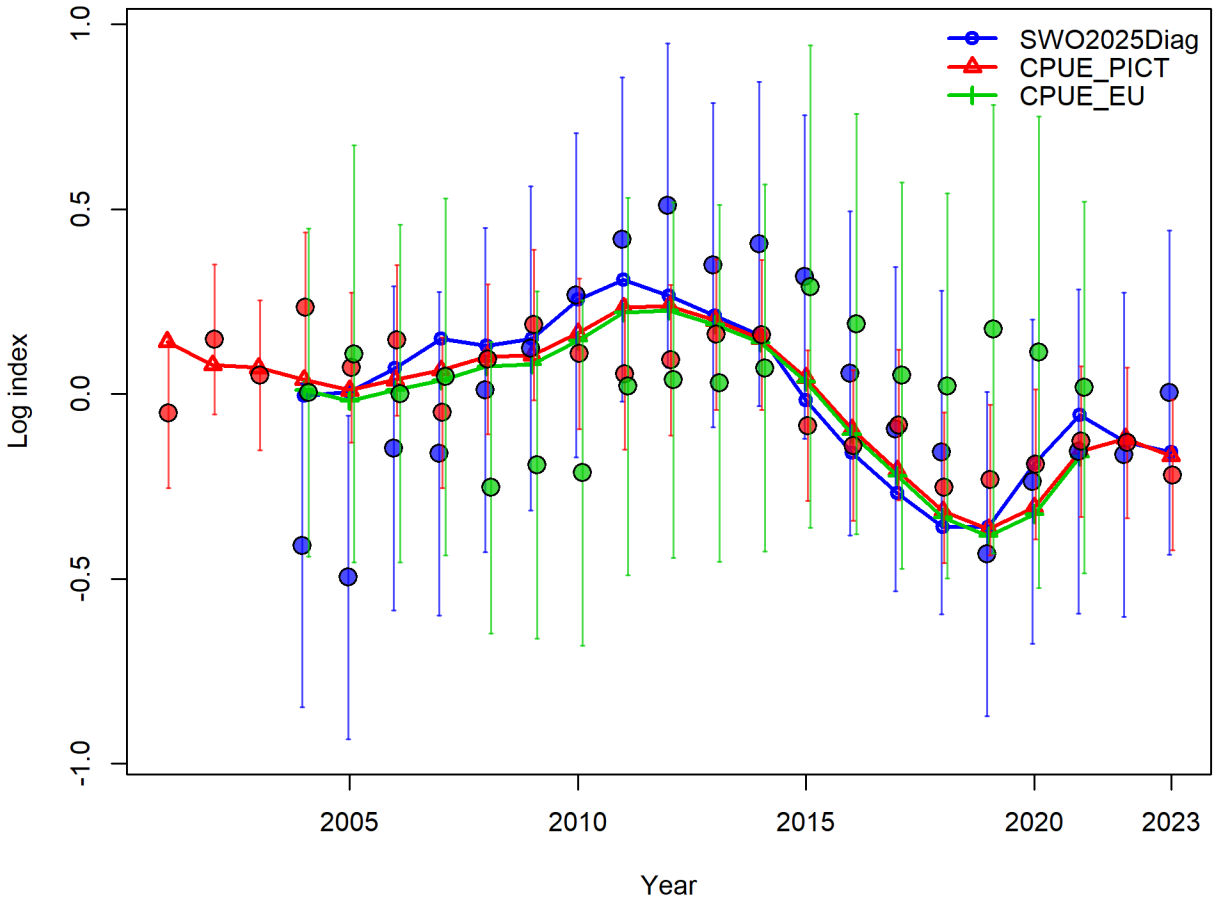


Figure 3: Comparisons of three alternate CPUE indices for region 2, and fits to these different indices, with the diagnostic model including indices from Australia and New Zealand (blue), the sensitivity to swap the NZ CPUE index with the PICT index (red) and the sensitivity to swap the NZ CPUE index with the EU index (green), with estimated variances required to fit these different indices shown with colour-coded vertical lines.

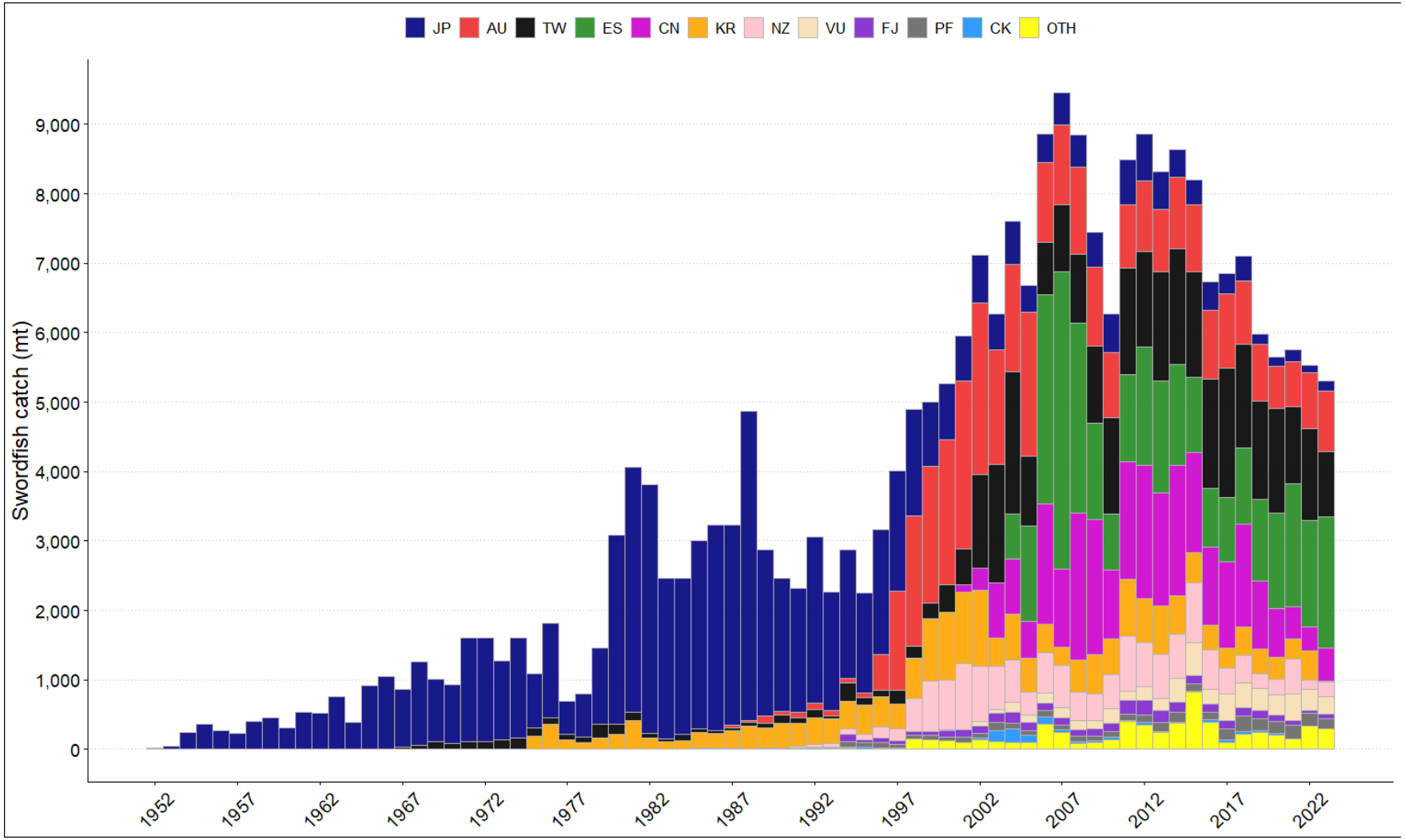


Figure 4: Annual catch (t) of swordfish by flag from 1952–2023, aggregated over both regions.

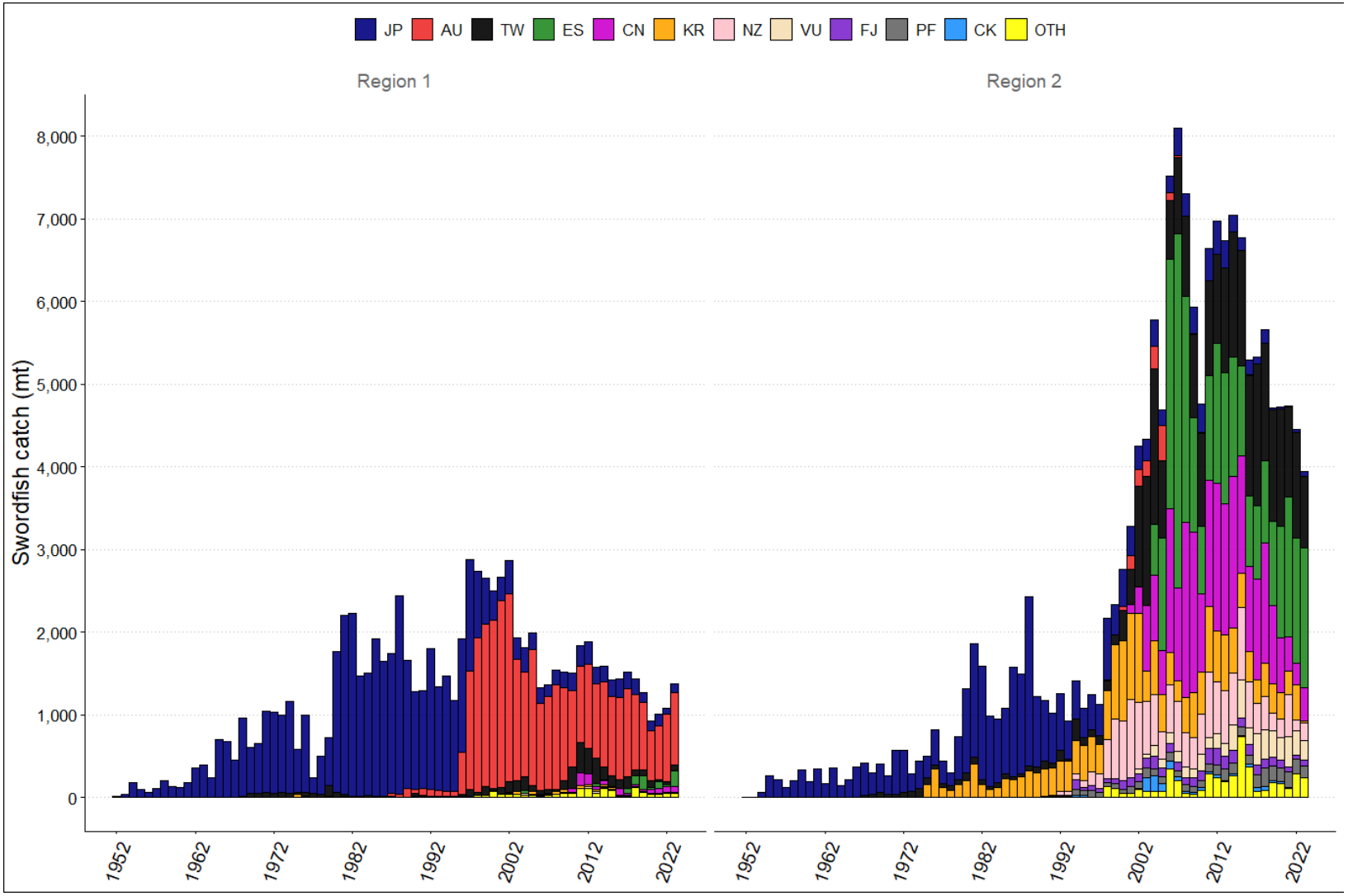


Figure 5: Annual catch (t) of swordfish by flag from 1952–2023, separated by region.

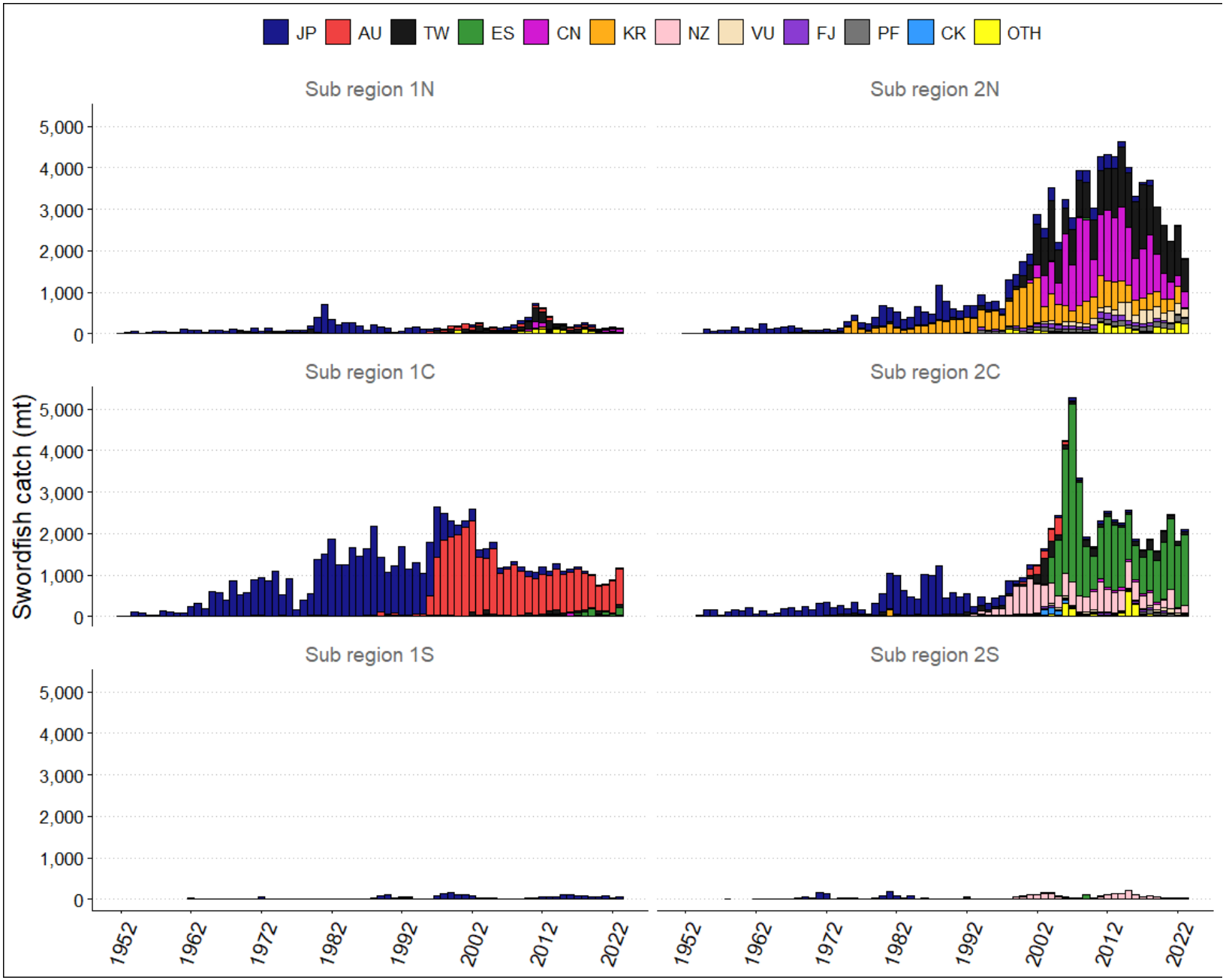


Figure 6: Annual catch (t) of swordfish by flag from 1952–2023, separated by subregion.

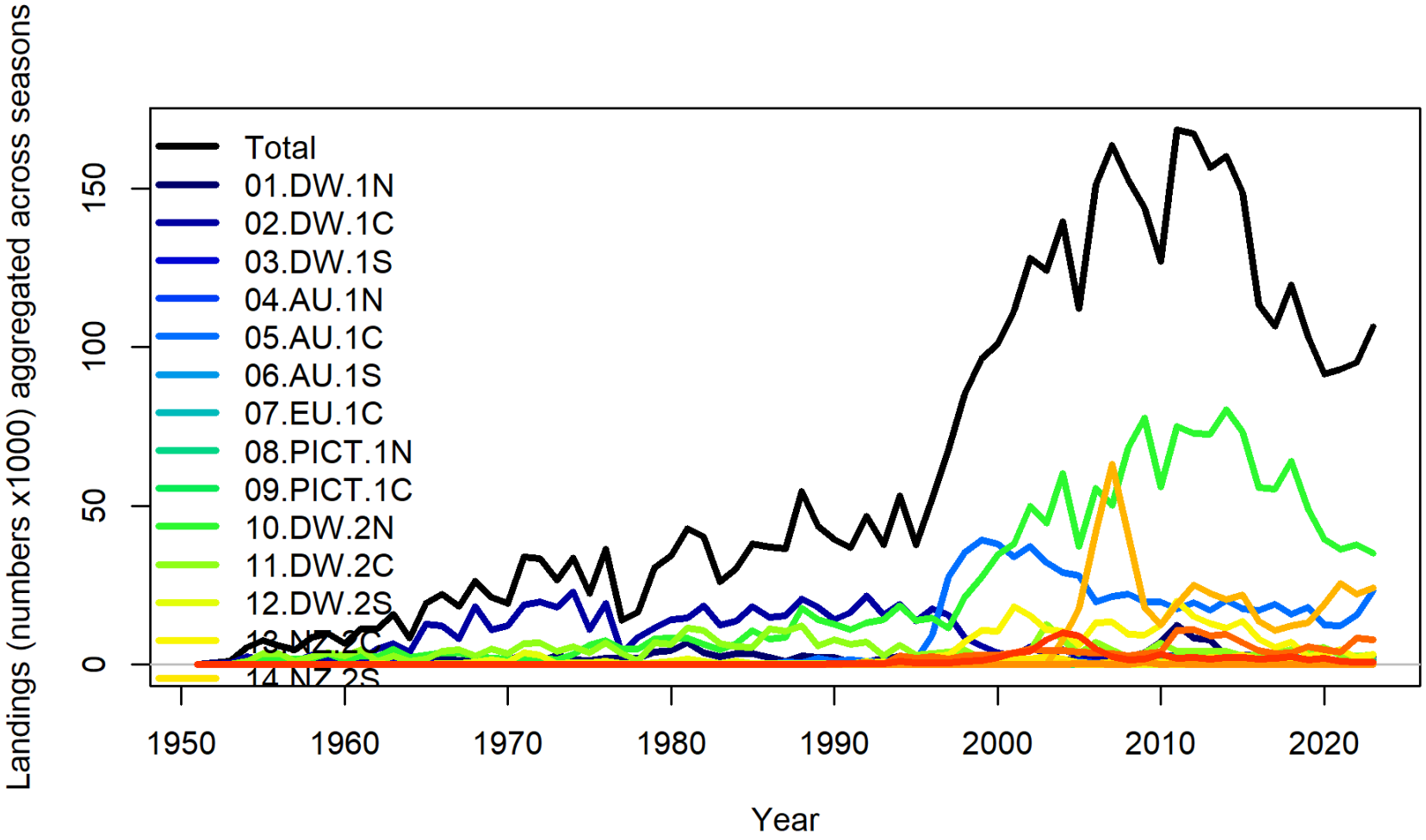


Figure 7: Annual catch (numbers of fish x 1000) of swordfish by fishery from 1952–2023, aggregated over both regions.

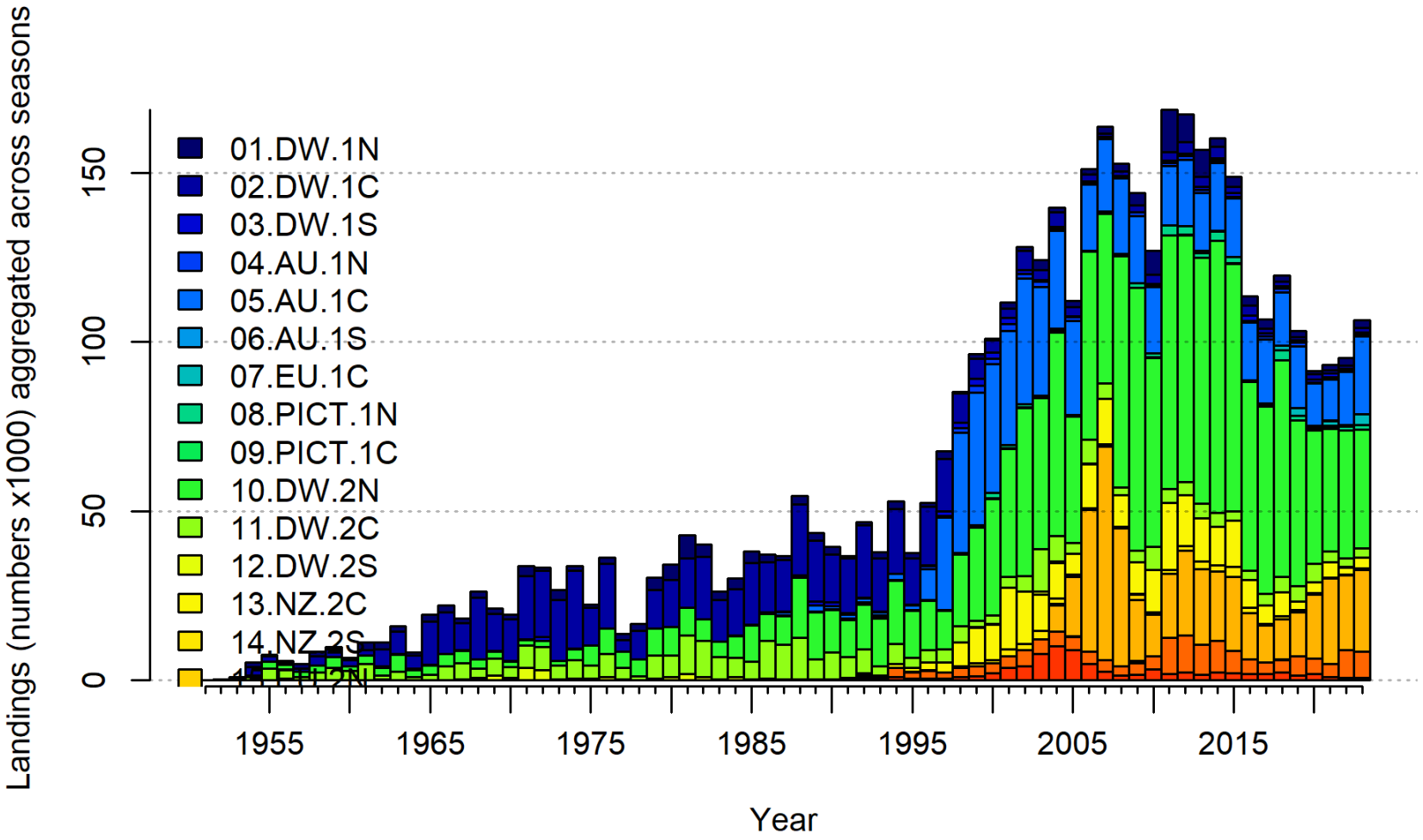


Figure 8: Stacked annual catch (numbers of fish x 1000) of swordfish by fishery from 1952–2023, aggregated over both regions.

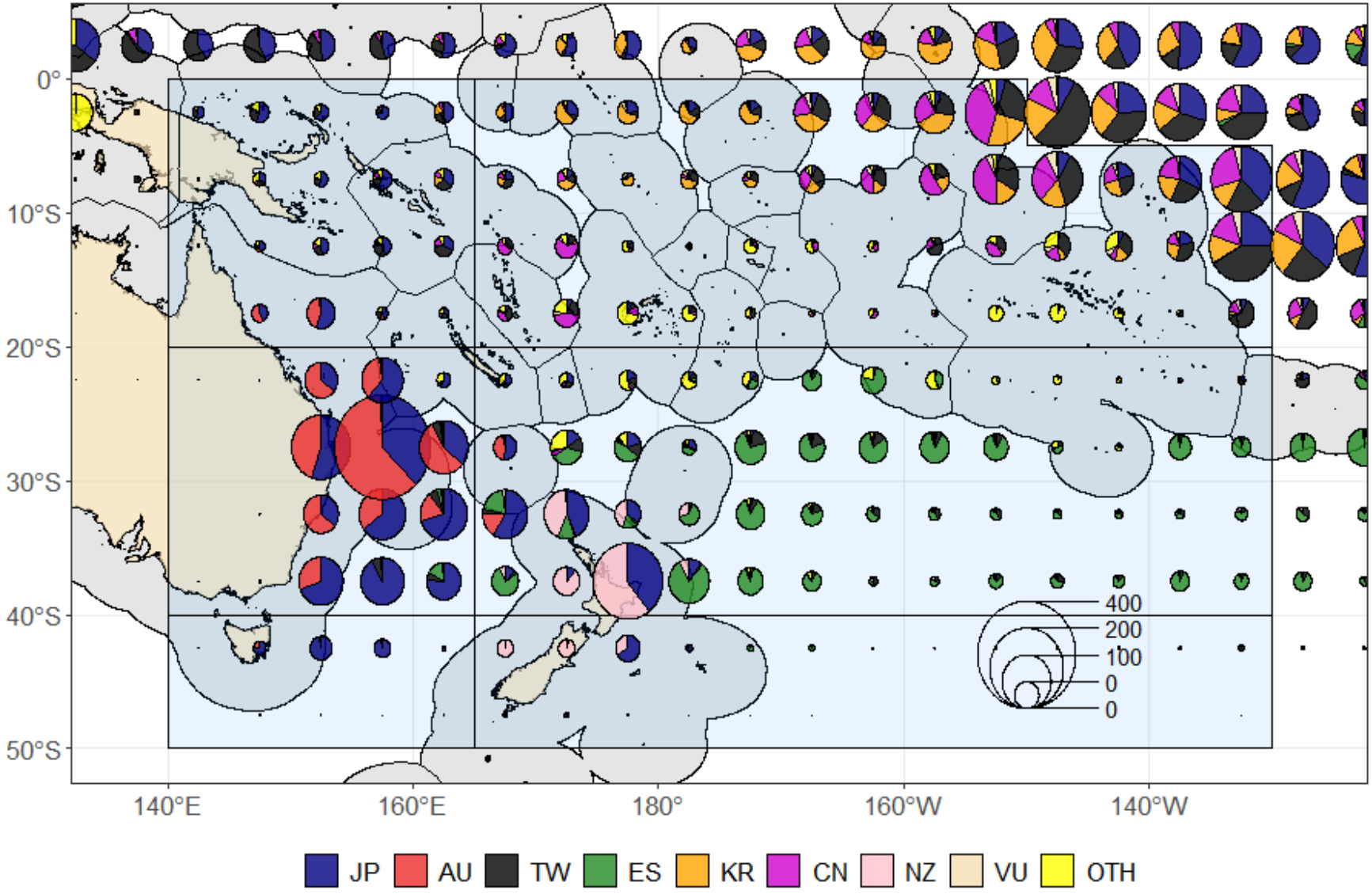


Figure 9: Spatial distribution and magnitude of swordfish catches (t) by flag, aggregated over all years, 1952–2023, for each 5° × 5° cell.

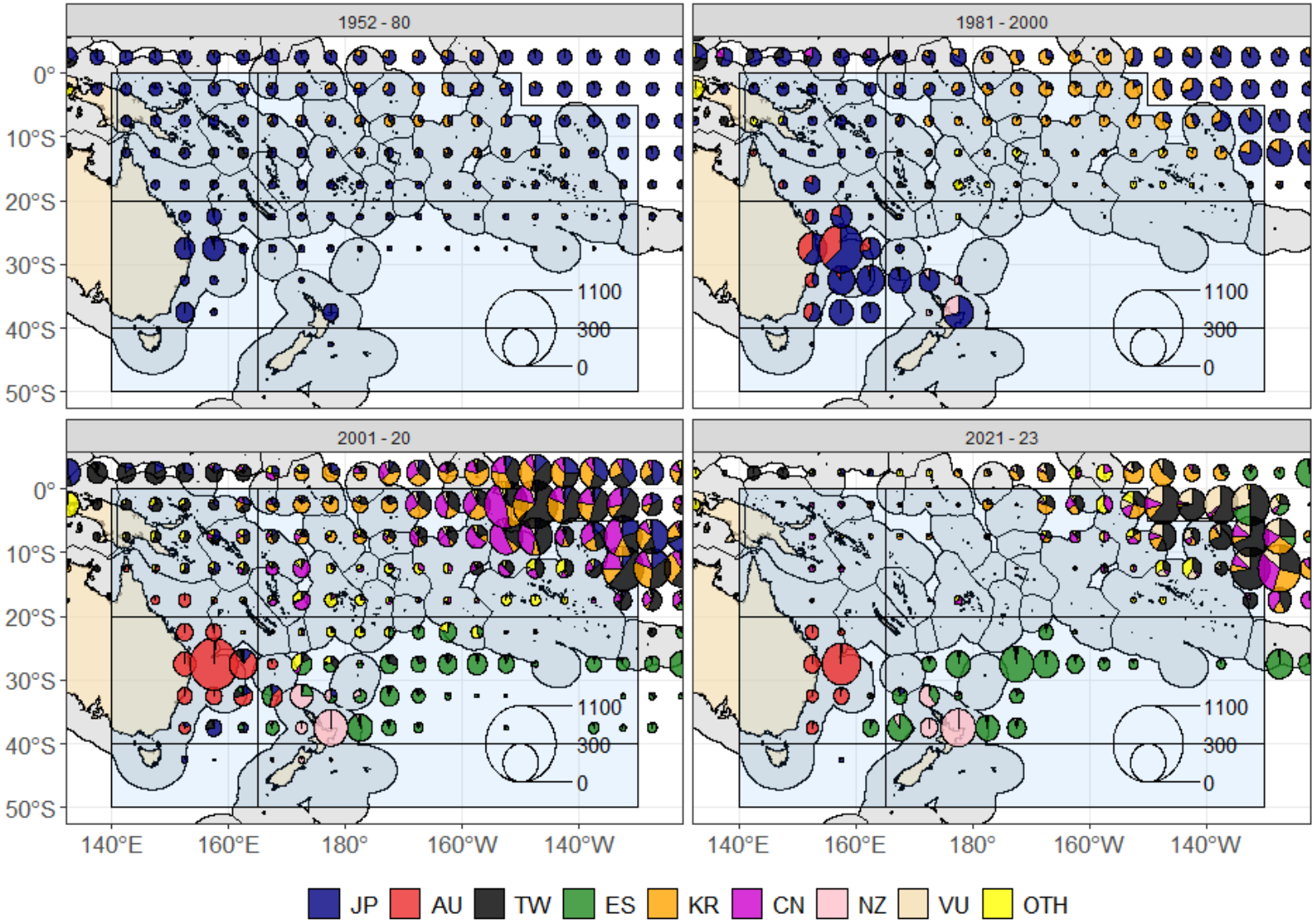


Figure 10: Temporal and spatial distribution and magnitude of swordfish catches (t) by flag, for the first 29 years (1952–1980, top left) two subsequent 20 year time periods (1981–2000, top right); and 2001–2020, bottom left; and then for the last three years (2021–2013, bottom right)), for each 5° × 5° cell.

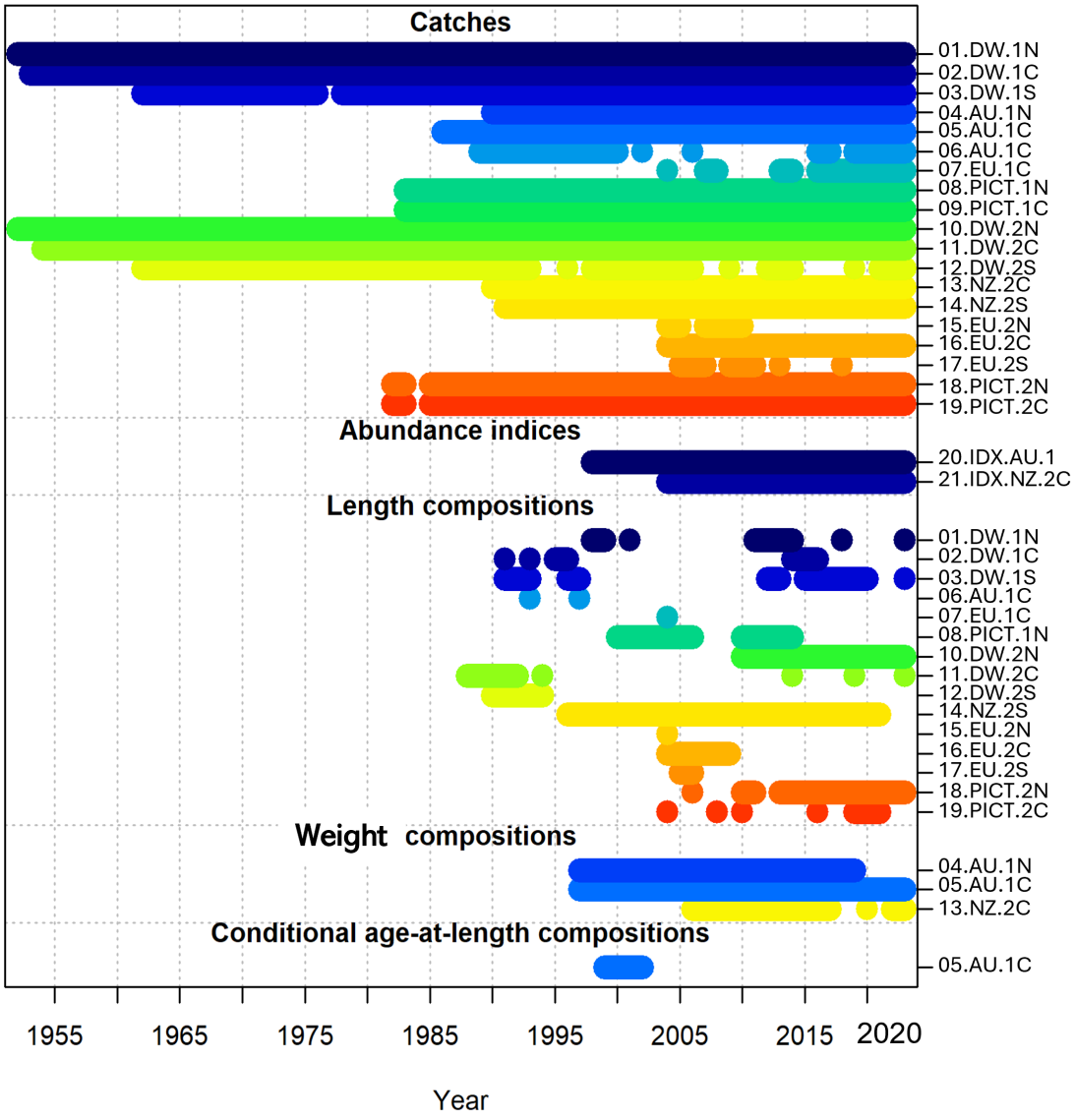


Figure 11: Summary of input data used for the 2025 swordfish assessment, indicating which years the various data types were available.

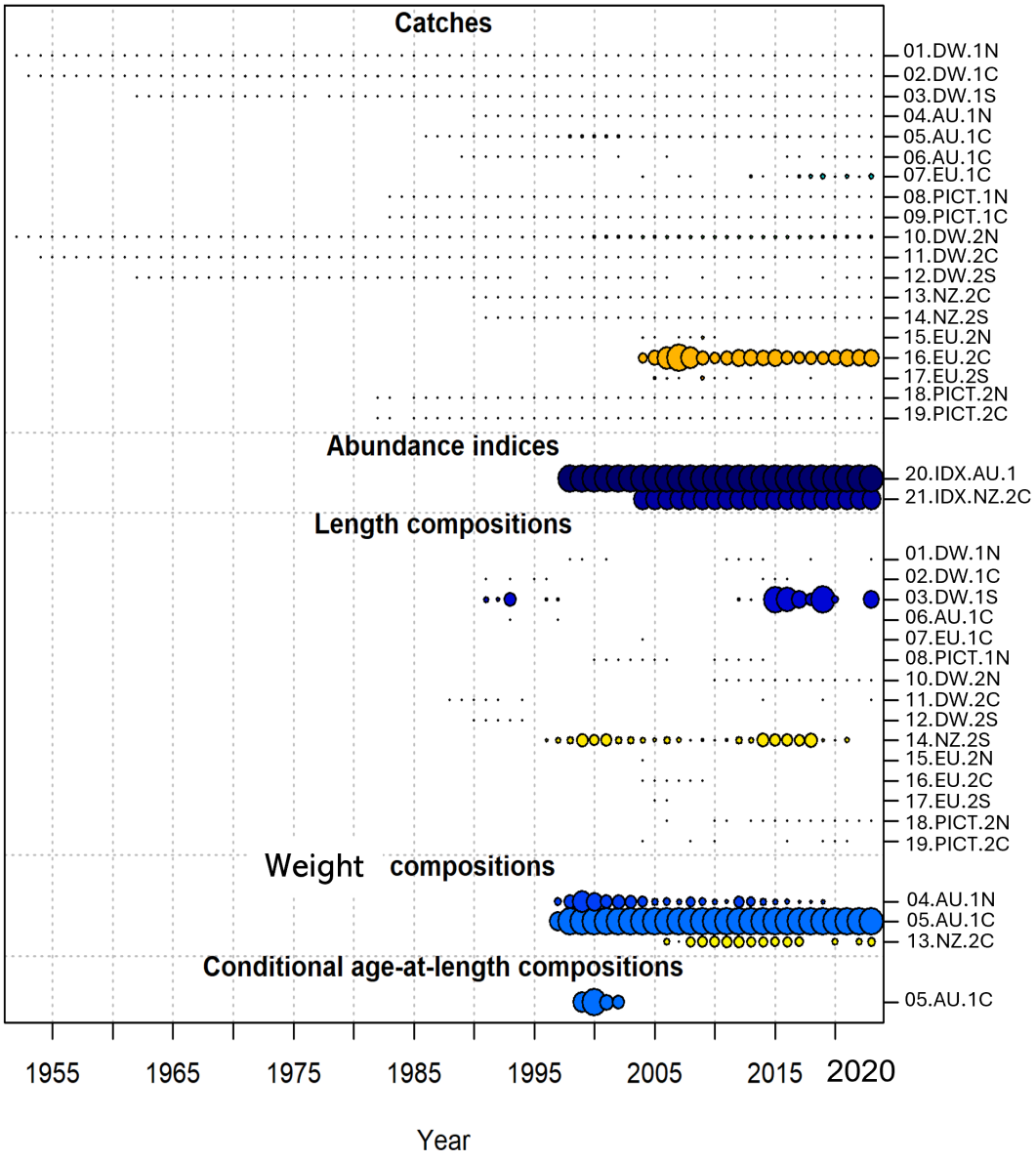


Figure 12: Summary of input data used for the 2025 swordfish assessment, indicating which years the various data types were available, with circle area relative to precision of each data type (with data weighting applied), or the actual size, for catch data. The catch data sizes are not representative as the catches for the EU fleet (t) have not been "corrected" to make them comparable to the catches for all other fisheries, where catches are reported as numbers of fish caught.

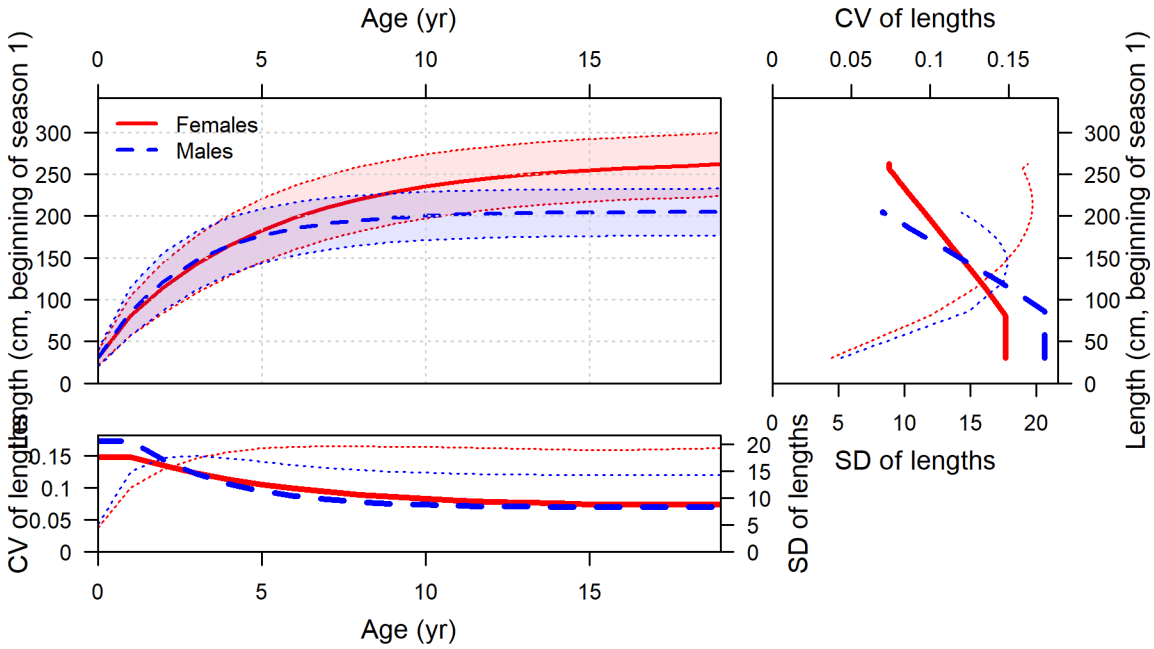


Figure 13: Estimated growth from the diagnostic model, with three von Bertalanffy parameters estimated for females (red), fixed CV parameters for young and old ages taken from the values obtained from external growth, and fixed offsets from the female parameters used for the male von Bertalanffy parameters (blue), fixed at values of the offsets obtained from the external growth estimates. Shaded area indicates 95% distribution of length at age around the estimated growth curve for each sex. Length at age (top-left panel) with CV (thick line) and SD (thin line) of length at age shown in top-right and lower-left panels.

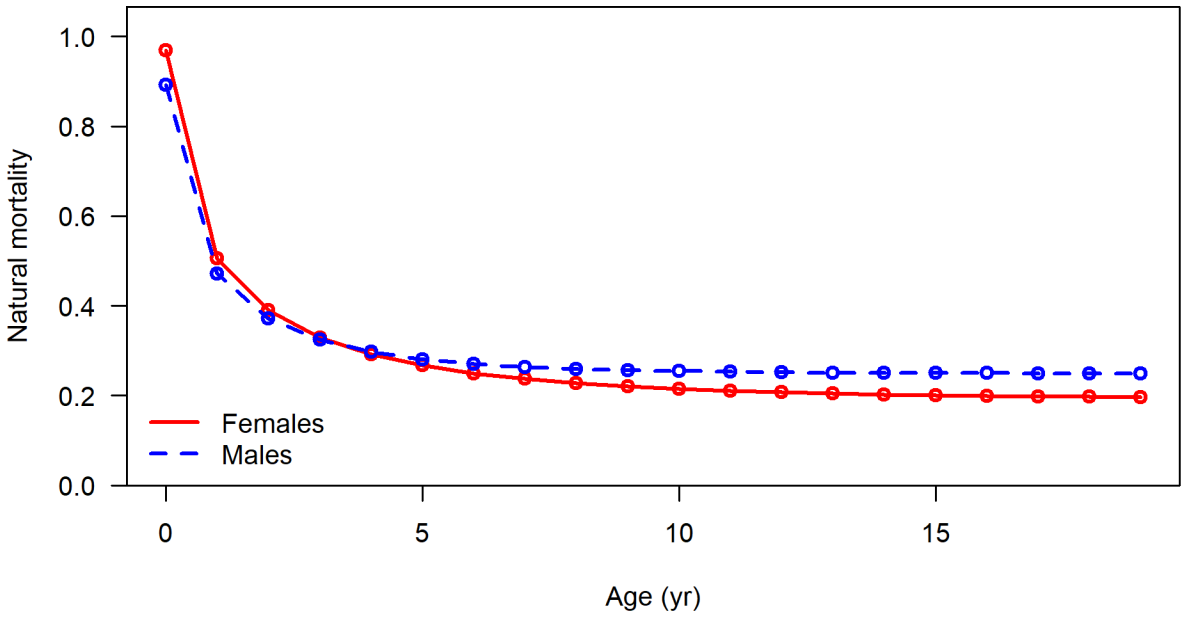


Figure 14: Lorenzen form of natural mortality for male and females, for the diagnostic model, with female mortality at age 7 estimated internally.

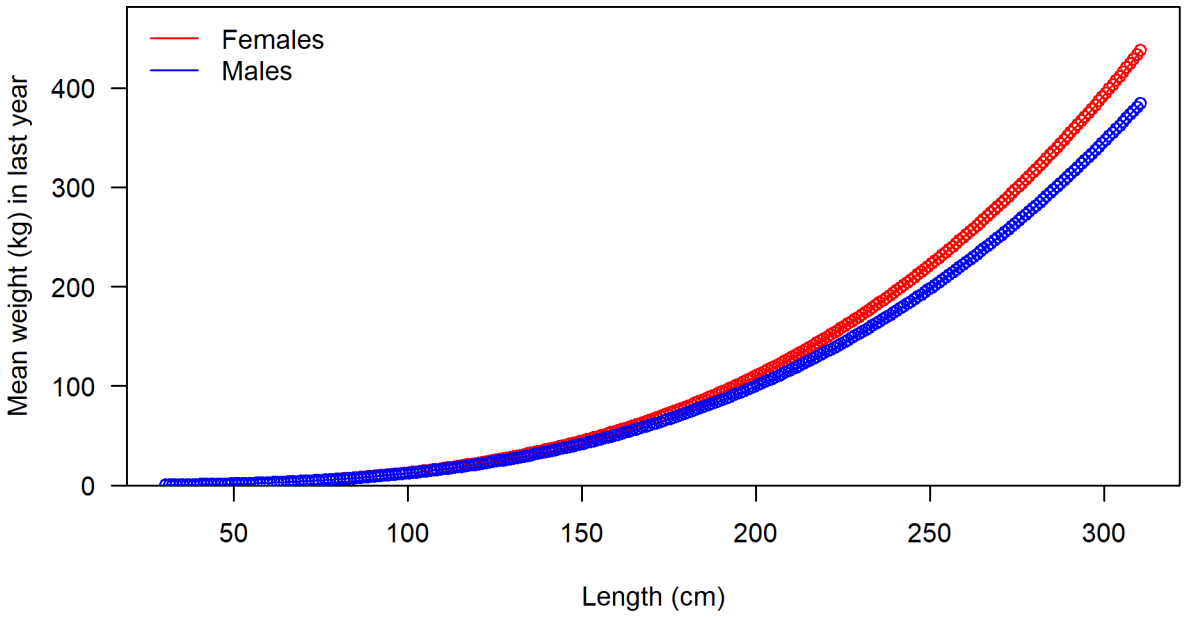
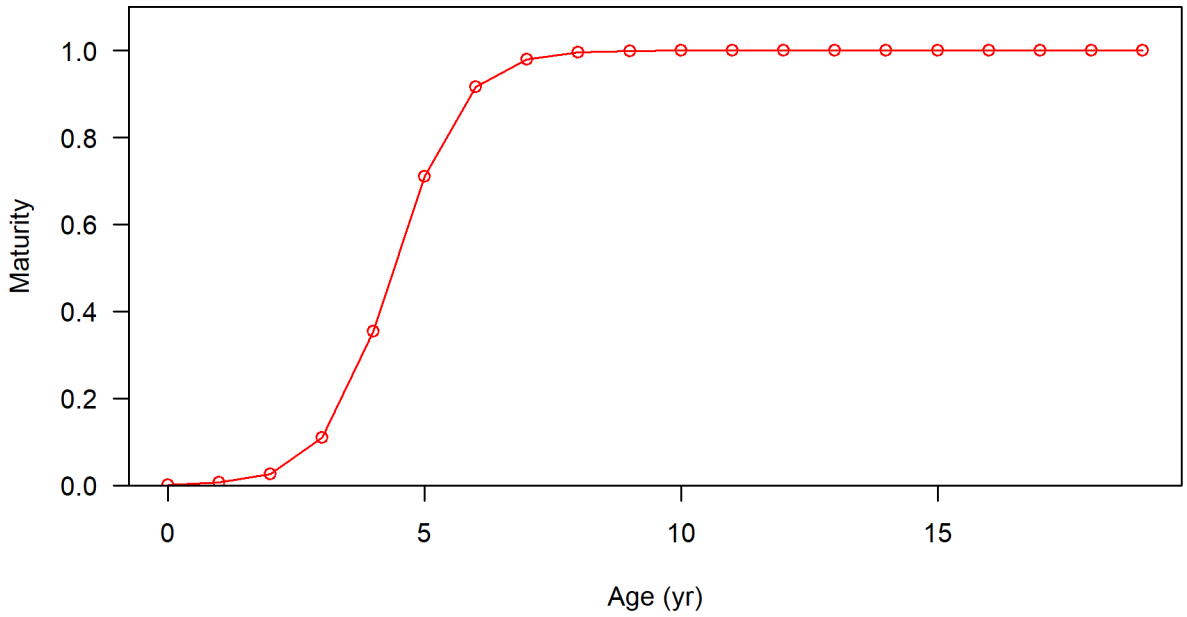


Figure 15: Weight at length relationship by sex for the diagnostic model with females shown in red and males in blue.



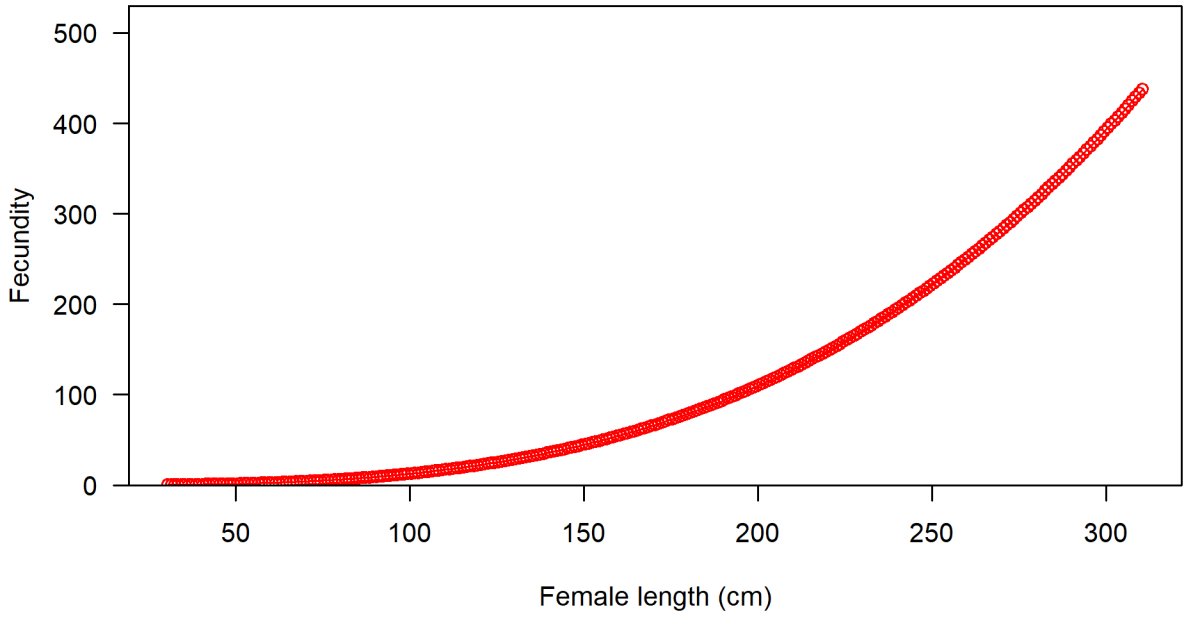


Figure 16: Maturity at age relationship for the diagnostic model (top) and fecundity by length relationship (bottom).

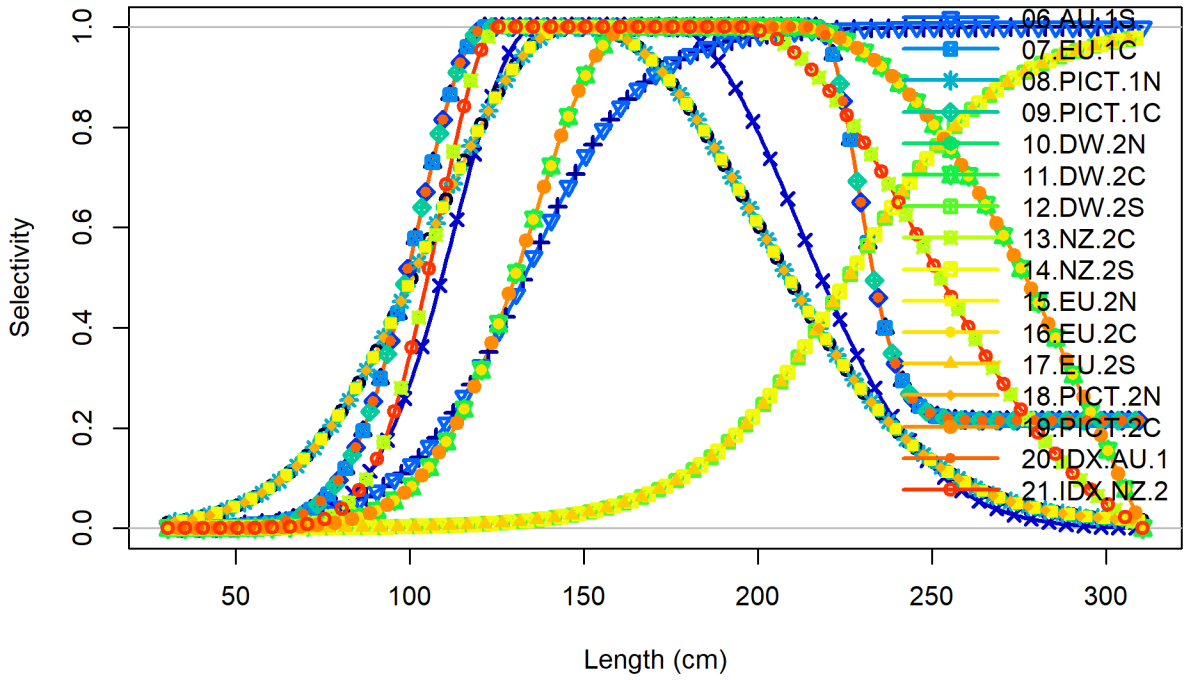


Figure 17: The seven different length-specific selectivity curves, some fixed and some estimated, for the seven distinct fishery groups with different selectivities.

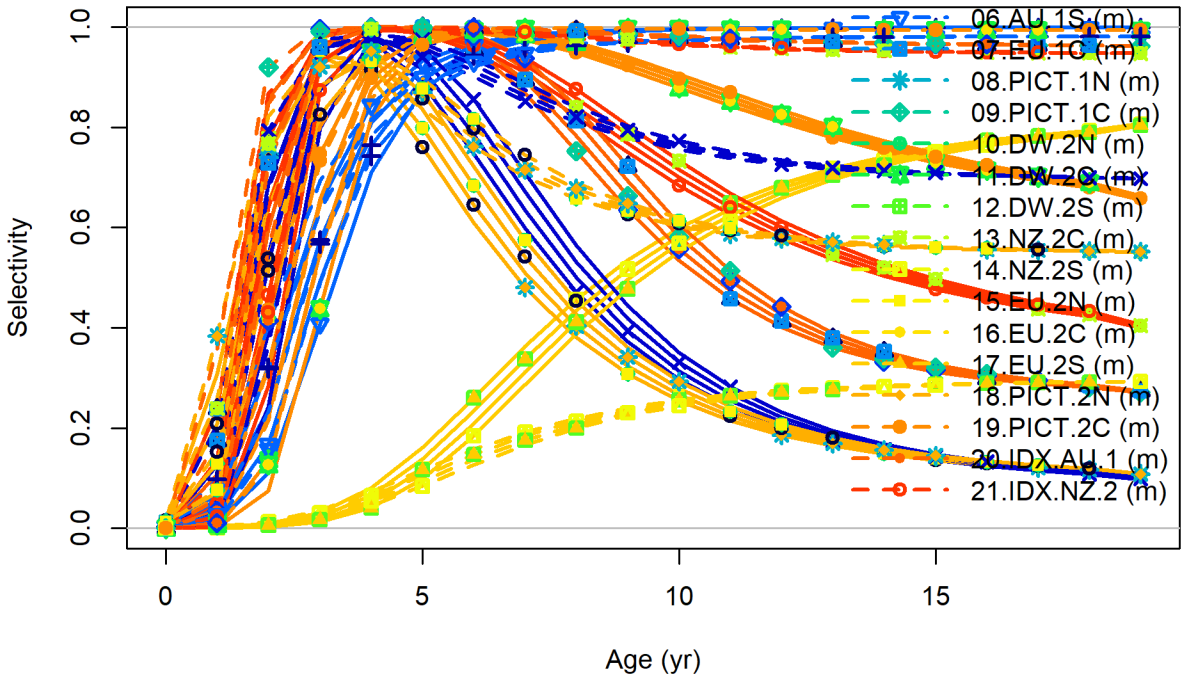


Figure 18: The seven different equivalent age-specific selectivity curves (plotted separately for males and females, due to different sex specific growth curves), derived from the length-based selectivity curves, some fixed and some estimated, for the seven distinct fishery groups with different selectivities.

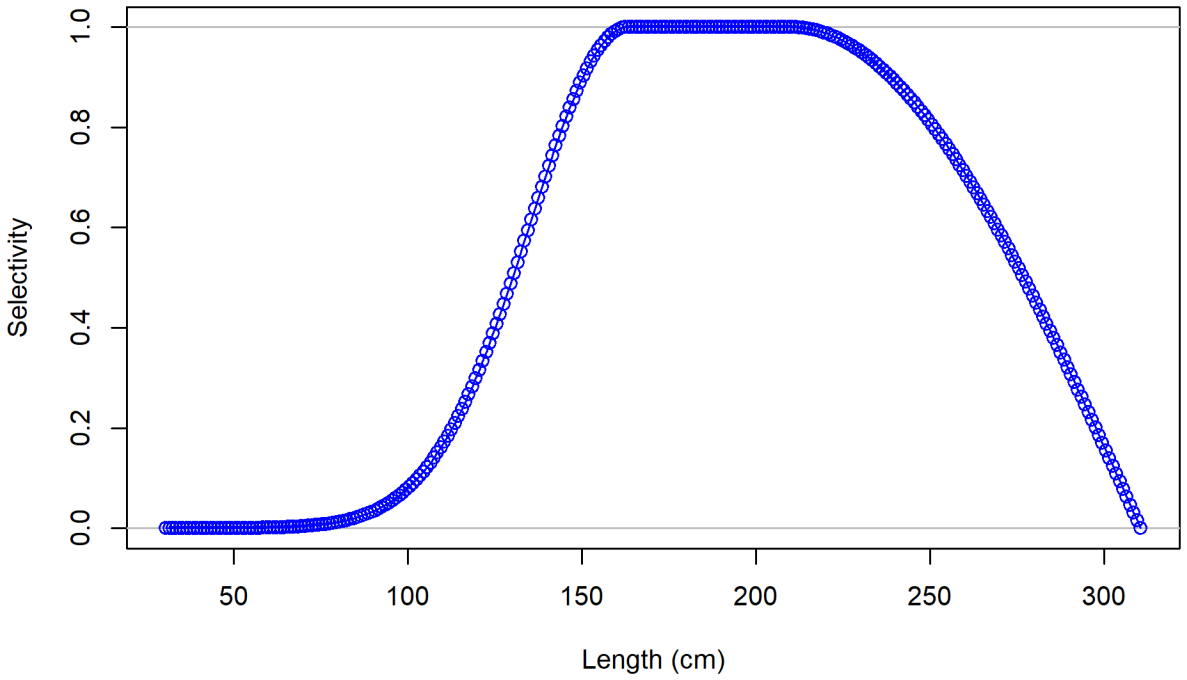


Figure 19: The fixed length-specific selectivity curve for the group of fisheries, from subregion 2C, dominated by fishery 16.EU.2C, but with some influence (a factor of five times less weight) from fisheries 11.DW.2C and 19.PICT.2C. This selectivity was initially estimated in a preliminary run of the model, and then fixed, at these estimated values, in all subsequent model runs, with further downweighting of the associated length composition data, for all three fisheries in this group, through an additional multiplier of 0.001 on the already downweighted initial input sample sizes.

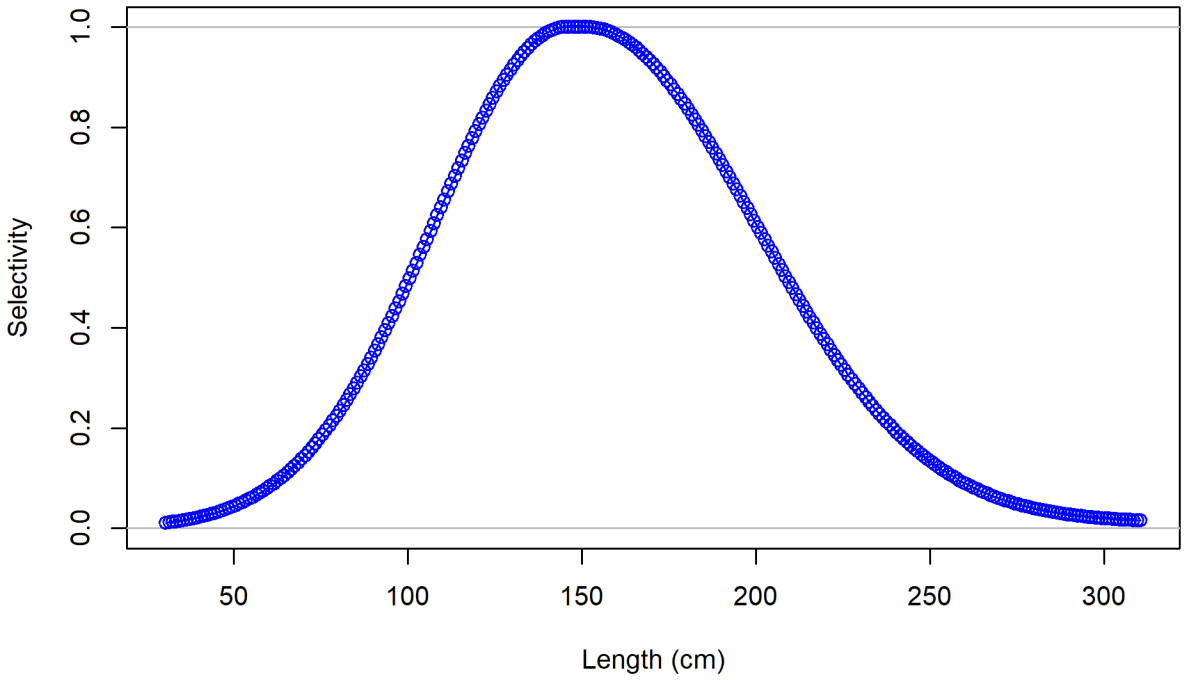


Figure 20: The fixed length-specific selectivity curve for the group of fisheries, from subregion 1N and 2N, dominated by fishery 08.PICT.1N, but with a small influence (a factor of 100 times less weight) from fishery 18.PICT.2N, and zero weight from fisheries 01.DW.1N, 10.DW.2N and 15.EU.2N (all with small or unreliable samples). This selectivity was initially estimated in a preliminary run of the model, and then fixed, at these estimated values, in all subsequent model runs, with further downweighting of the associated length composition data, through an additional multiplier of 0.001, for fisheries 08.PICT.1N and 18.PICT.2N, on the already downweighted initial input sample sizes.

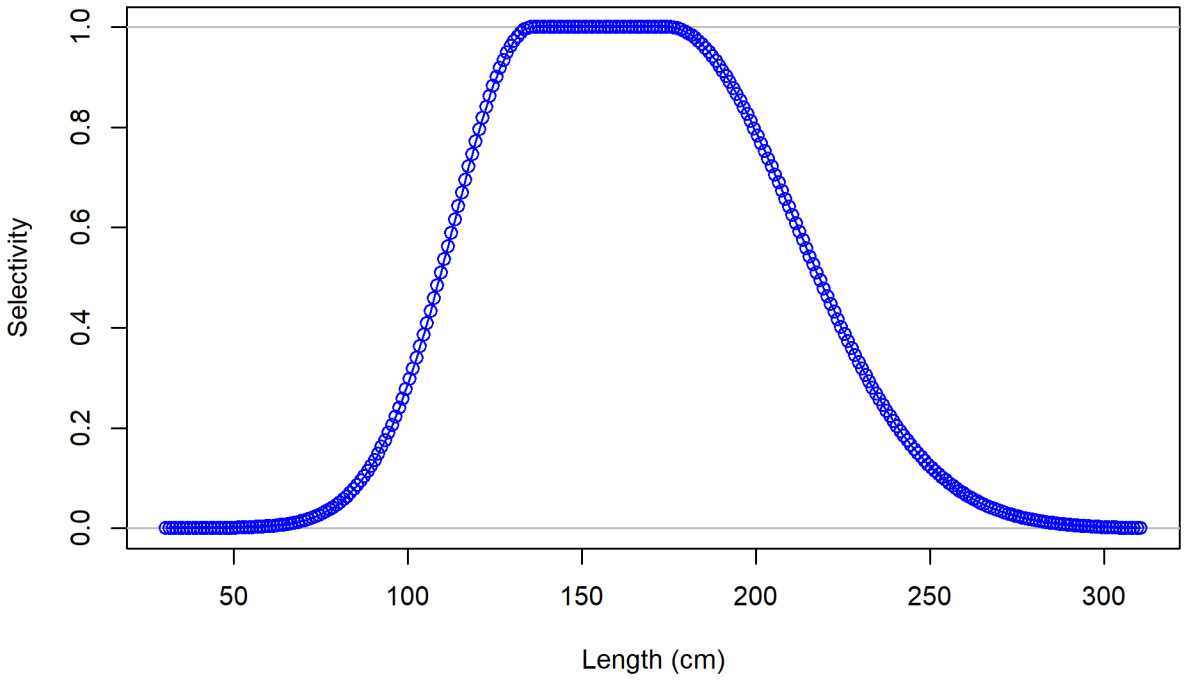


Figure 21: The estimated length-specific selectivity curve with weight composition data in subregion 1N, from fishery 04.AU.1N, Francis weighted with multiplier 0.2891.

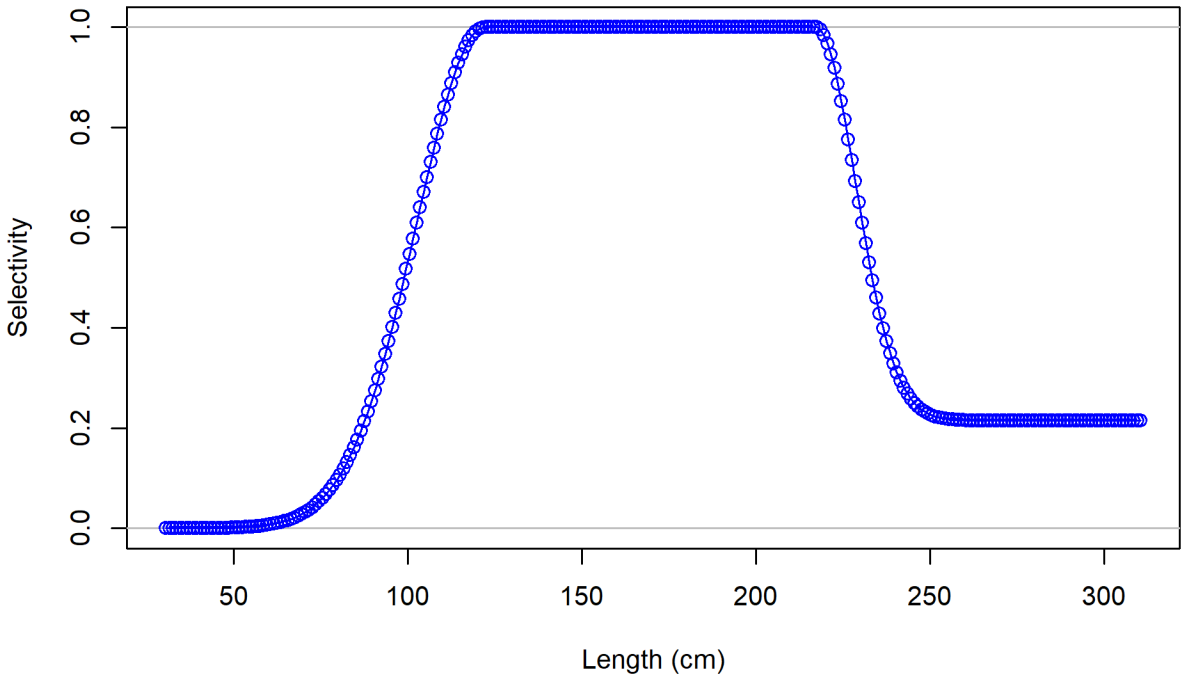


Figure 22: The estimated length-specific selectivity curve from the fishery group in subregion 1C, dominated by the weight composition data from fishery 05.AU.1C, Francis weighted with multiplier 0.2555, but also sharing selectivity with the downweighted length-based fisheries in subregion 1C, and 02.DW.1C, 07.EU.1C (multiplier of 0.000005 for both) and with all length composition samples from fishery 09.PICT.1C removed by the initial filtering steps.

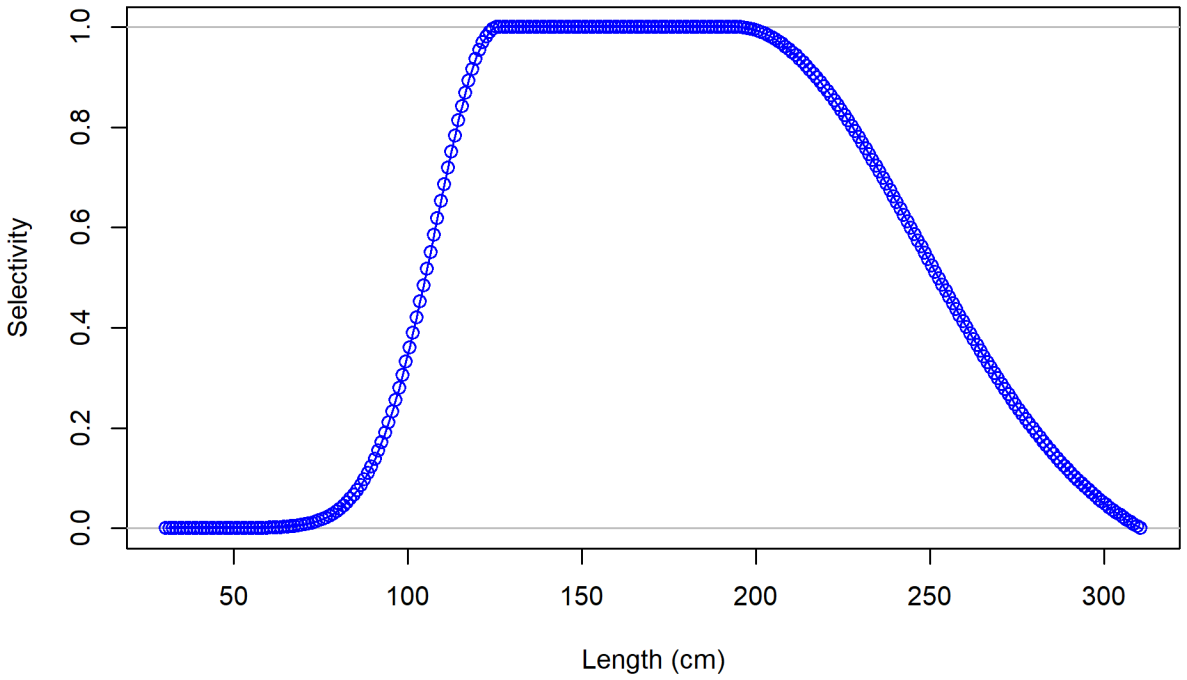


Figure 23: The estimated length-specific selectivity curve from the fishery group in subregion 2C, dominated by the weight composition data from fishery 13.NZ.2C, Francis weighted with multiplier 0.05214, but also sharing selectivity with the downweighted length-based fisheries in subregion 2C, 16.EU.2C (multiplier of 0.00001), which has weighting 5 times larger than the other two length-based fisheries in this region, 11.DW.2C and 19.PICT.2C (multiplier of 0.000002 for both).

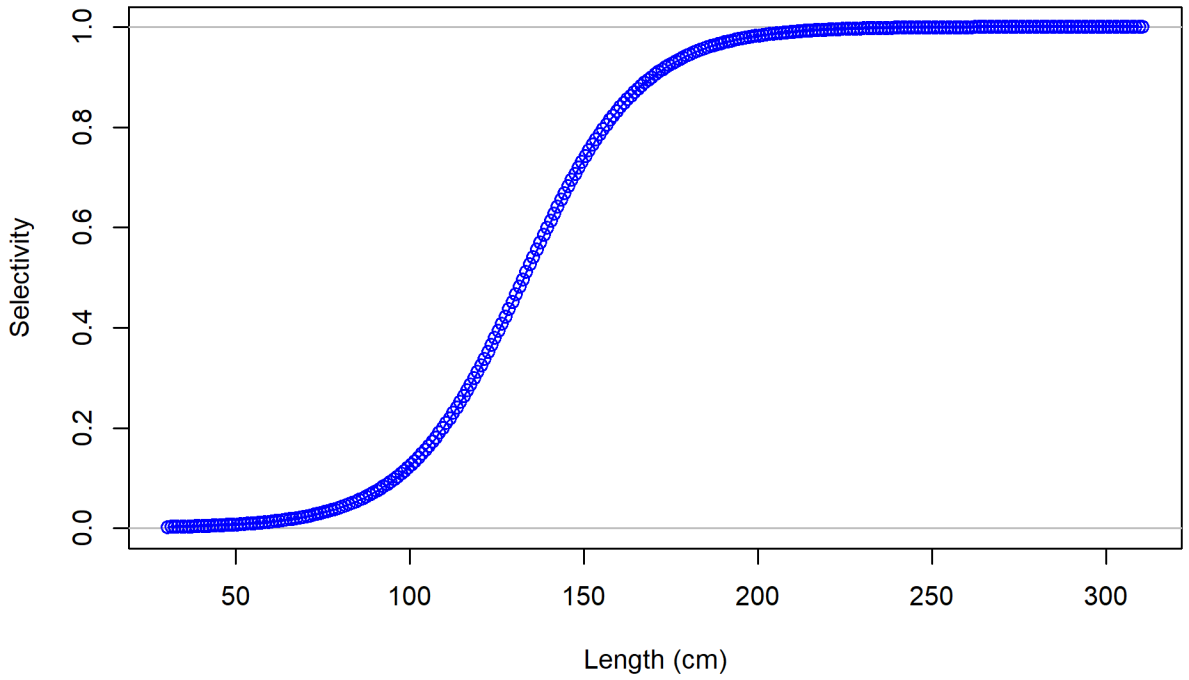


Figure 24: The estimated length-specific selectivity curve from the fishery group dominated by the length composition in subregion 1S, dominated by the length composition data from fishery 03.DW.1S, Francis weighted with multiplier 0.2054, and 06.AU.1S (with fewer samples and much smaller multiplier of 0.00001).

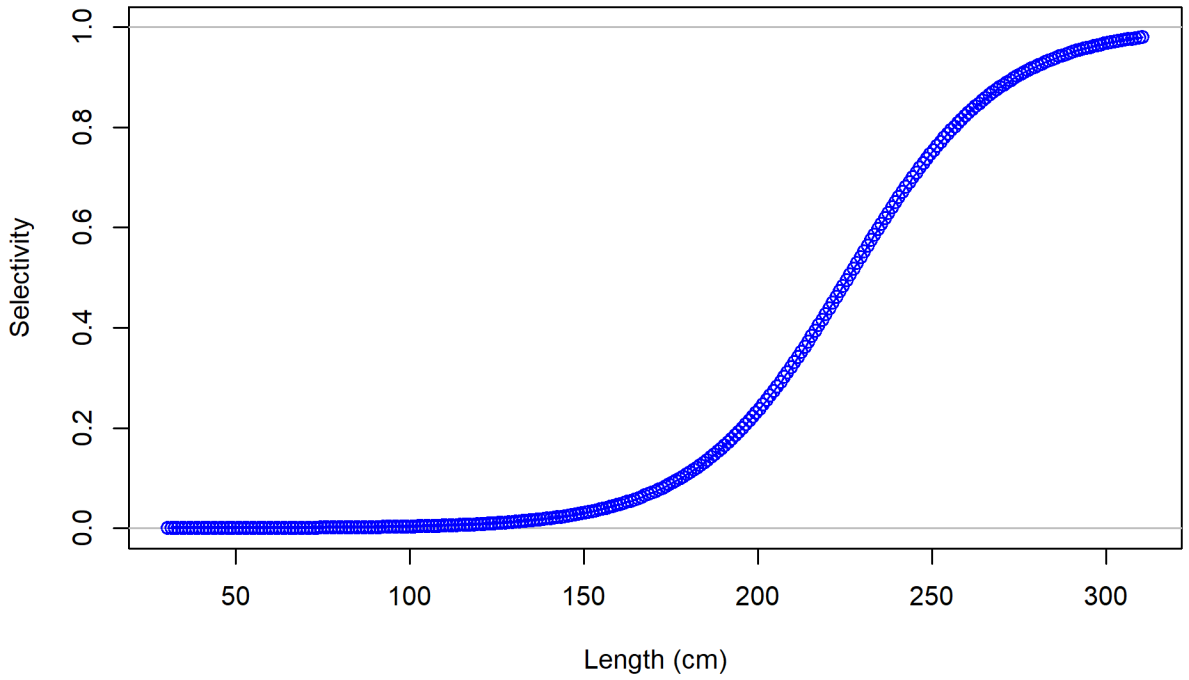


Figure 25: The estimated length-specific selectivity curve from the fishery group dominated by the length composition in subregion 2S, dominated by the length composition data from fishery 14.NZ.2S, Francis weighted with multiplier 0.05952, but also sharing selectivity with the down-weighted length-based fisheries in subregion 2S, 12.DW.2S and 17.EU.2S, both of which had multipliers of zero, as they had so few samples.

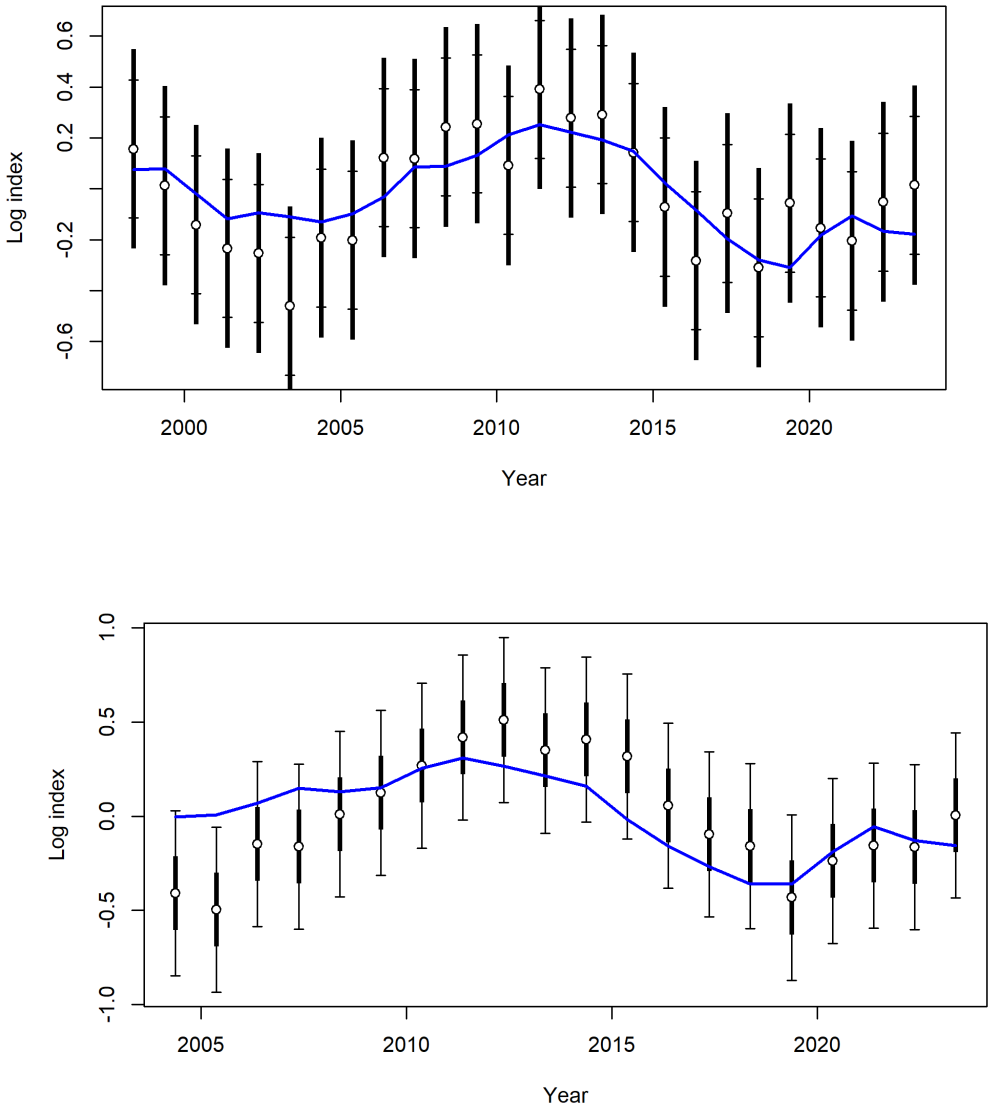


Figure 26: The fits to the two CPUE indices from the diagnostic model, with both plots shown on a log scale, to reflect the space in which these indices are fitted. Observed (circles) and model-estimated (blue line) CPUE indices vs year, with approximate 95% asymptotic intervals for the Australian index (20.IDX.AU.1, top) and for the New Zealand index (21.IDX.NZ.2, bottom). The thin lines with capped ends should match the thick lines for a balanced model. These indices are balanced by estimating an additional variance parameter within Stock Synthesis, which is negative for the Australian index, suggesting the models fit well with less variance than the initial values of 0.2, and positive for the New Zealand index, suggesting that the model requires additional variance to fit this index, compared to the initial value used of 0.1.

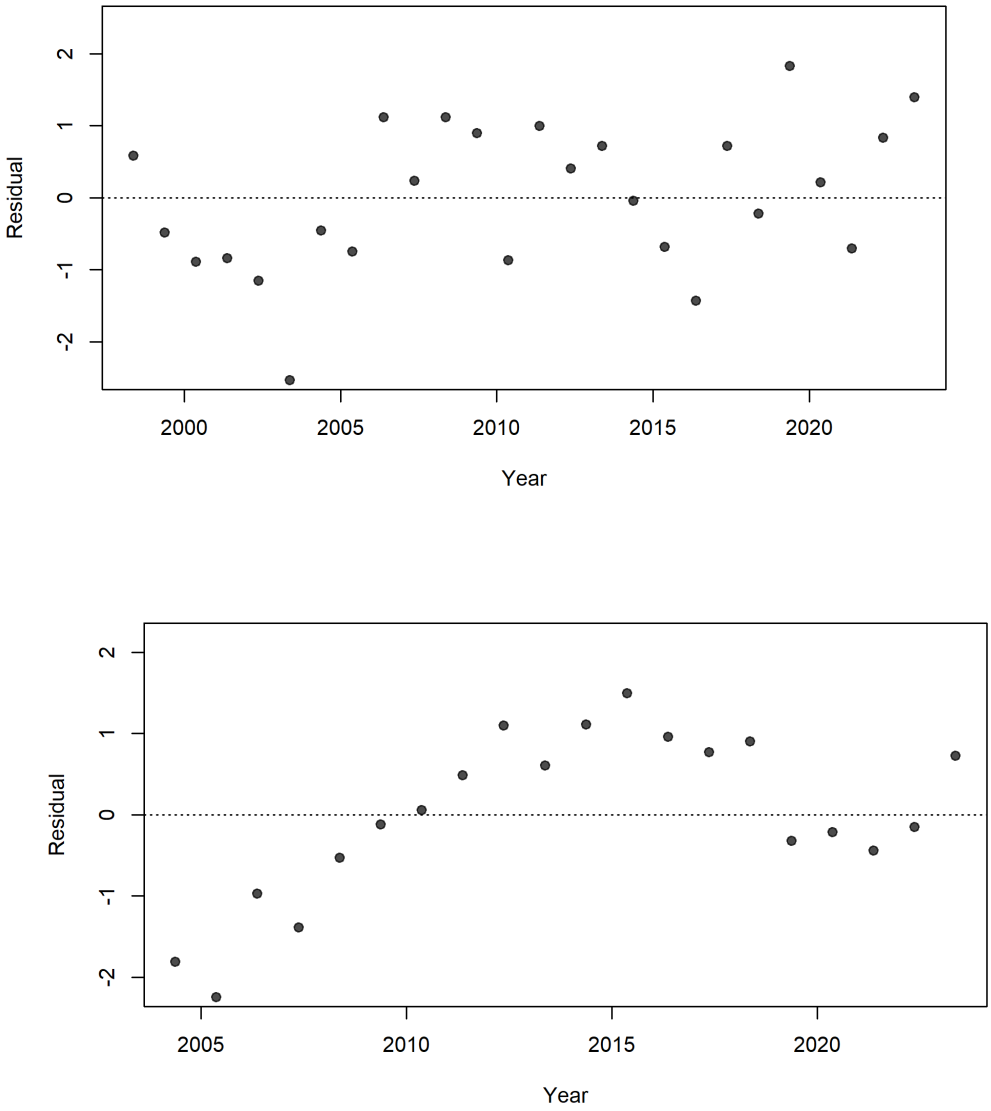


Figure 27: The residual patterns for the fits to the two CPUE indices from the diagnostic model. The residual pattern for the Australian index (20.IDX.AU.1, top) looks acceptable, but there is a poor residual pattern seen in the fits to and the New Zealand index (21.IDX.NZ.2, bottom).

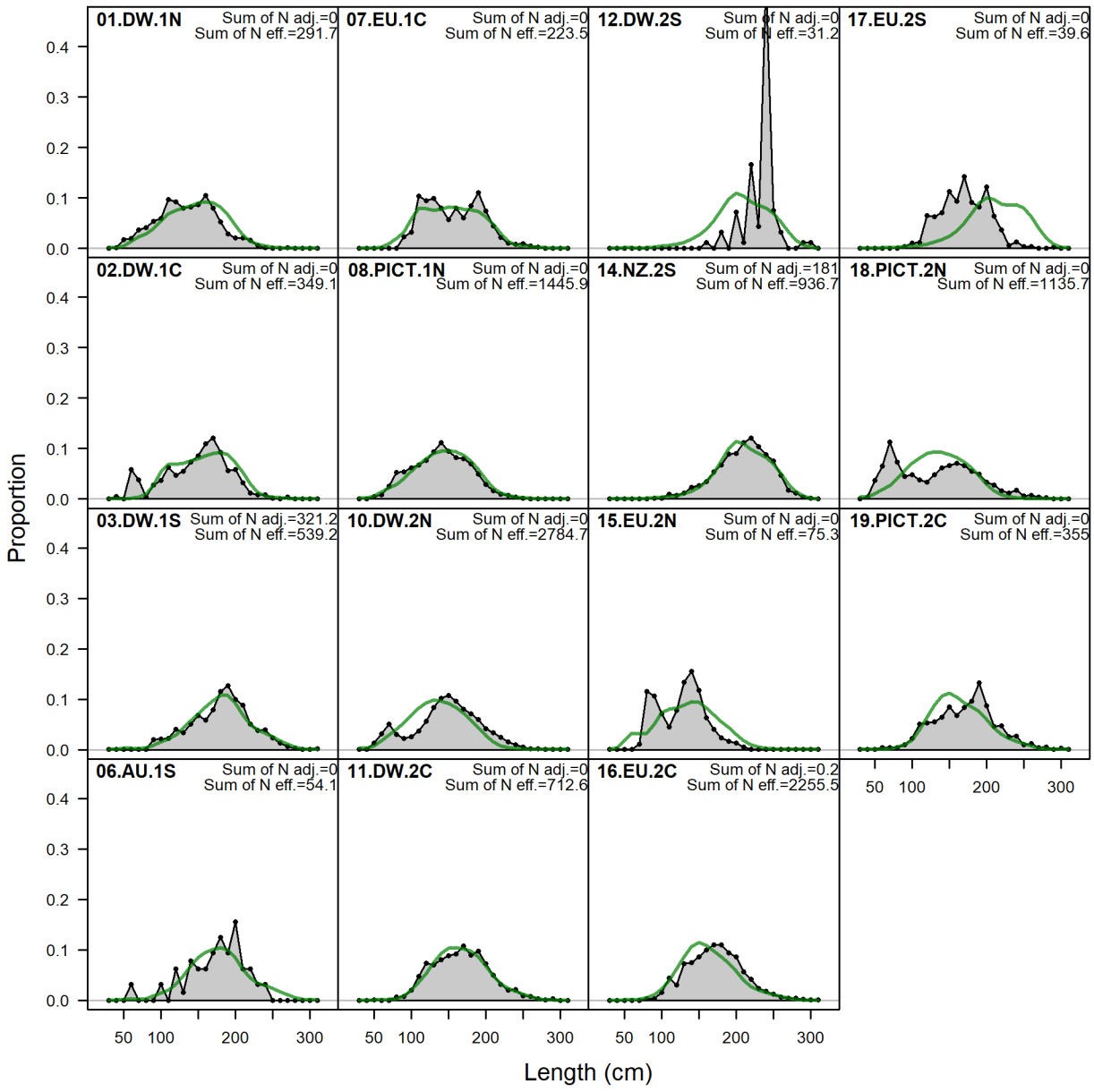


Figure 28: Fits to length compositions by fishery, aggregated across all years, for the 2025 diagnostic model. Observed data are grey and the fitted value is the green line. "N adj." is the input sample size after data-weighting adjustments. "N eff." is the calculated effective sample size used in the McAllister-Ianelli tuning method.

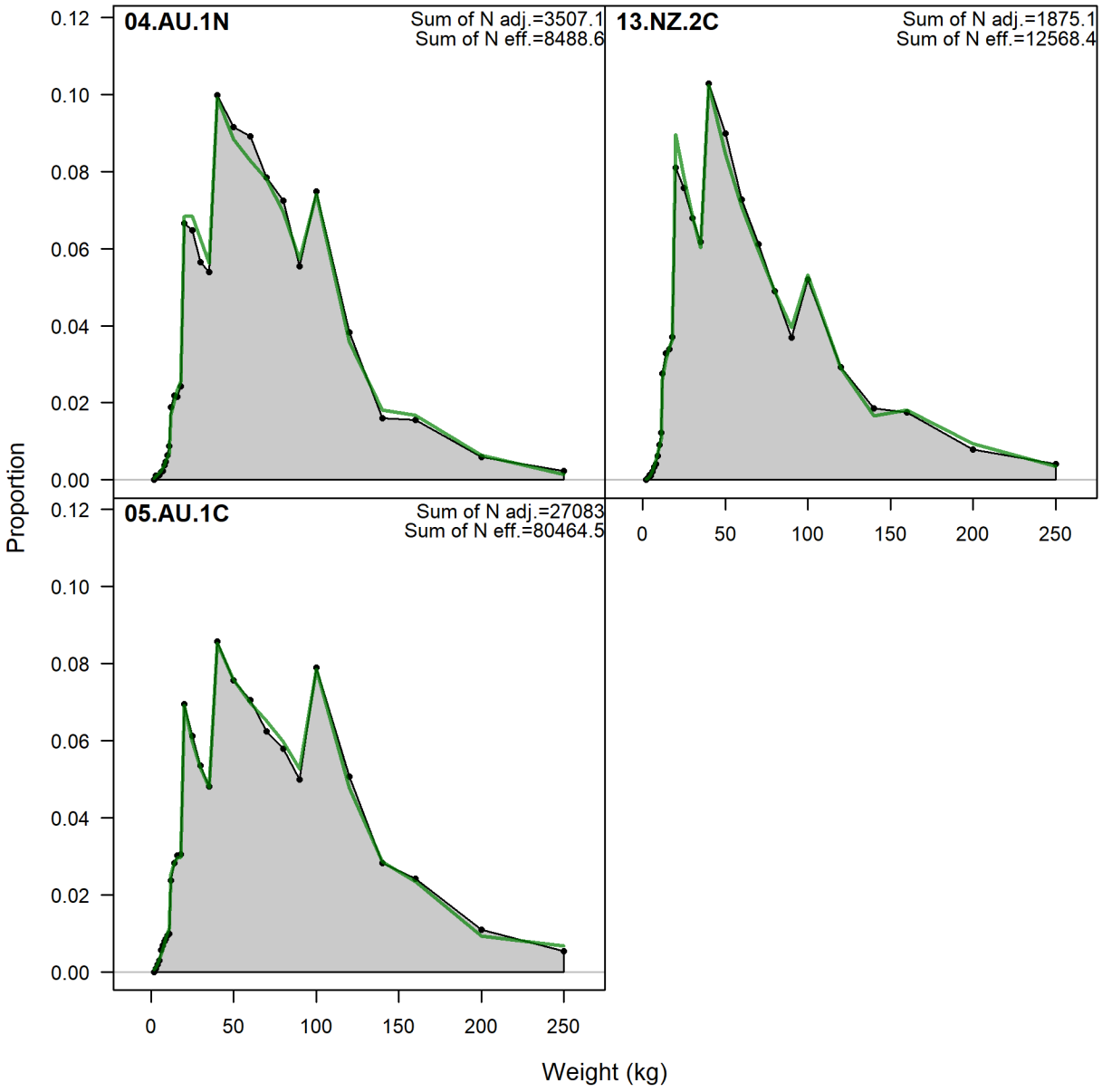


Figure 29: Fits to weight compositions by fishery, aggregated across all years, for the 2025 diagnostic model. Observed data are grey and the fitted value is the green line. "N adj." is the input sample size after data-weighting adjustments. "N eff." is the calculated effective sample size used in the McAllister-Ianelli tuning method.

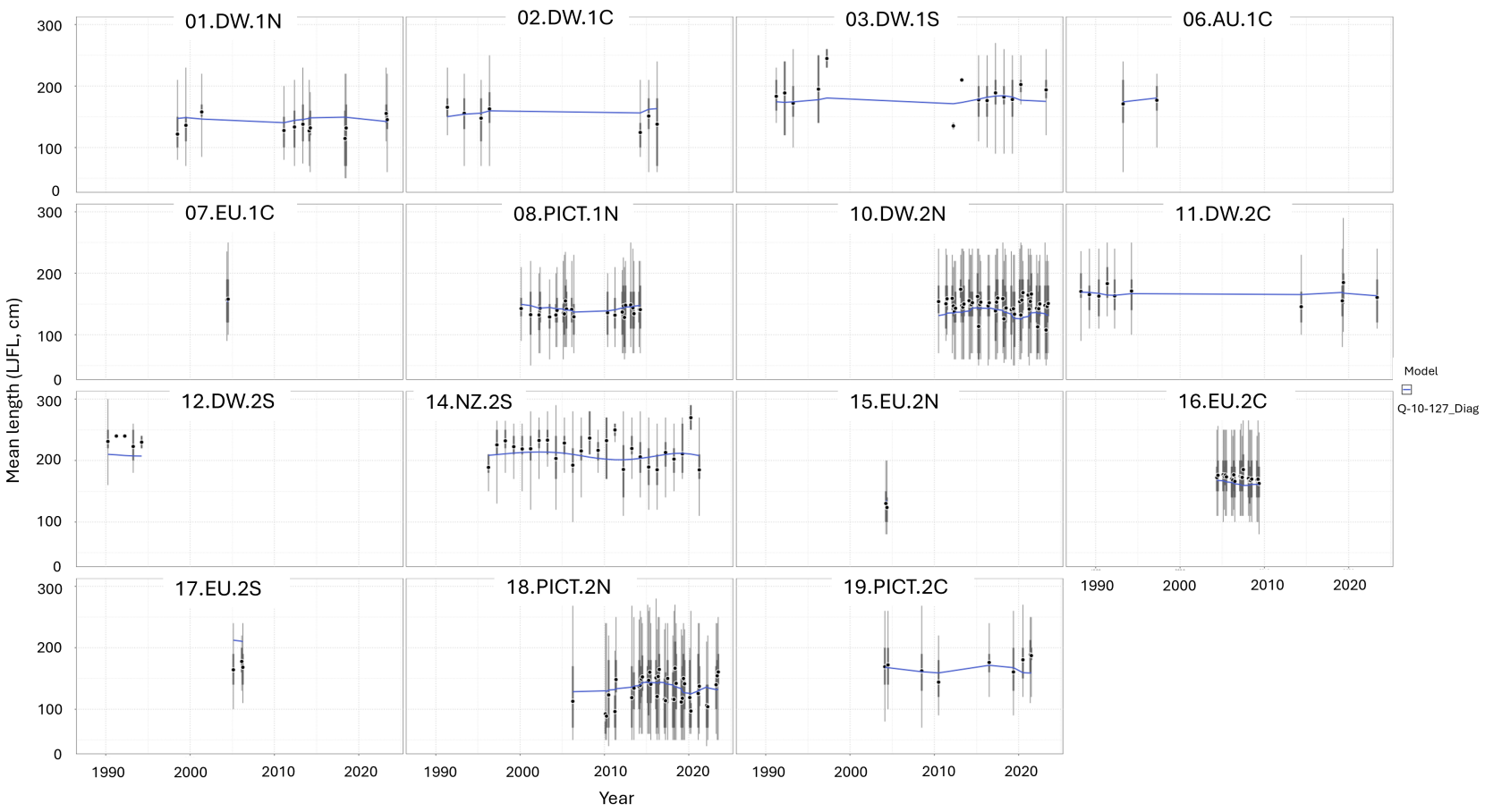


Figure 30: Observed (black points) and predicted (blue line) median fish lengths (LJFL, cm) for the fisheries with length data for the 2025 diagnostic model. The uncertainty intervals for the observed data are represented by boxes (thick black vertical lines) represent the values encompassed by the 25% and 75% quantiles for each quarter of data, while the whiskers (thin black vertical lines) represent the values encompassed by the 5% and 95% quantiles for each quarter of data. Sampling data are by quarter and only filtered length samples included in the diagnostic model are plotted.

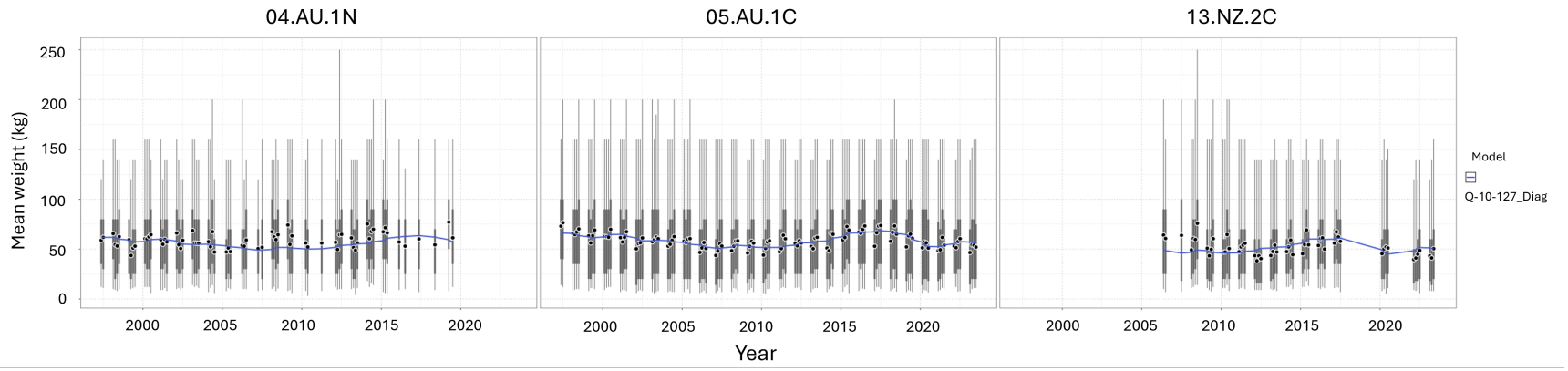


Figure 31: Observed (black points) and predicted (blue line) median fish weights (kg) for the fisheries with weight data for the 2025 diagnostic model. The uncertainty intervals for the observed data are represented by boxes (thick black vertical lines) represent the values encompassed by the 25% and 75% quantiles for each quarter of data, while the whiskers (thin black vertical lines) represent the values encompassed by the 5% and 95% quantiles for each quarter of data. Sampling data are by quarter and only filtered weight samples samples included in the diagnostic model are plotted.

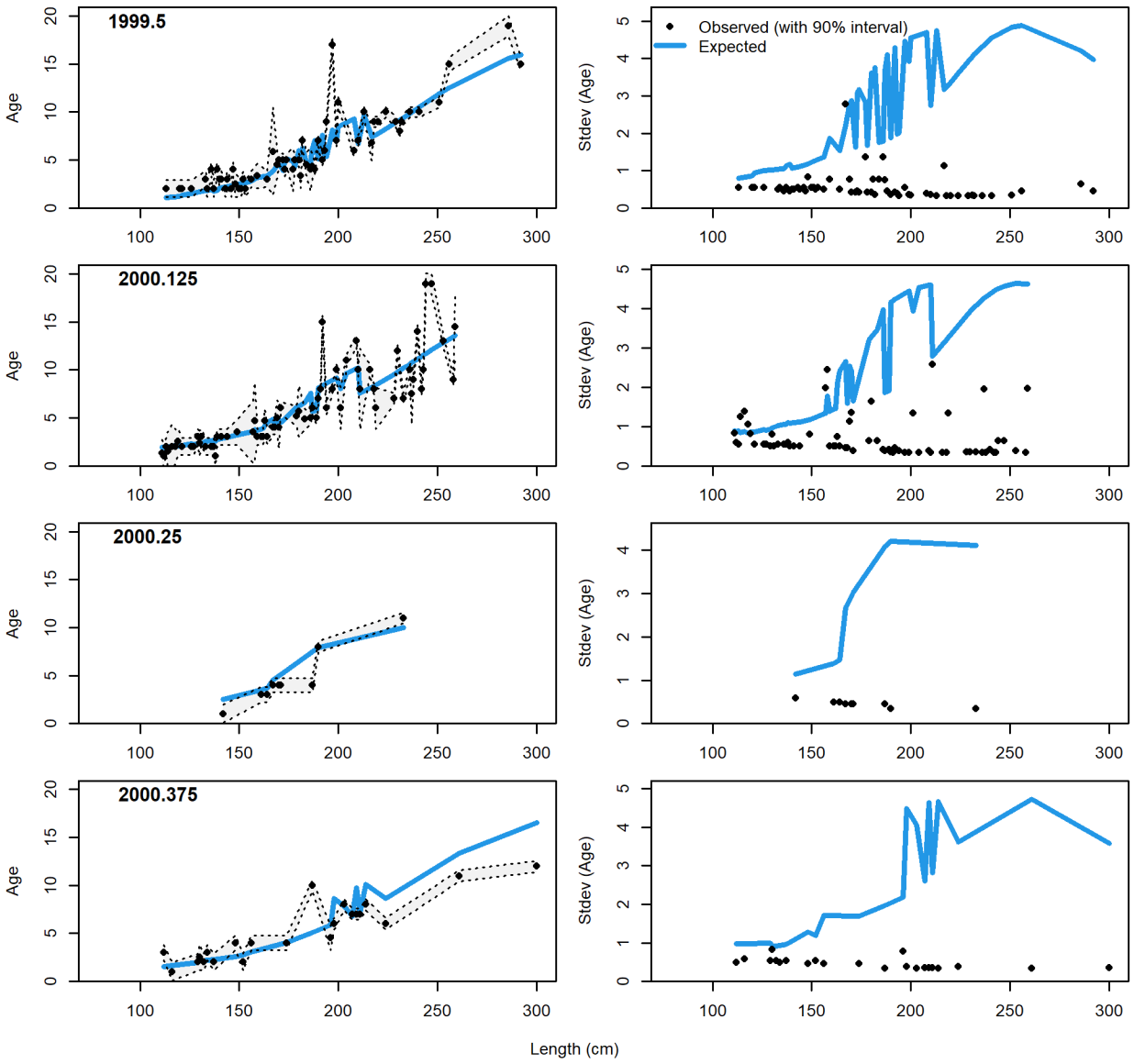


Figure 32: Fits to conditional age-at-length data from fishery 05.AU.1C by quarter with observations (black points, with 90% confidence intervals) and predicted ages at length (blue line) in the left hand panel. The right hand panel shows observed and expected variances, which are not very meaningful with the low numbers of samples each quarter and the high resolution of lengths used, meaning that most lengths do not have replicate samples to estimate a reliable variance on the age for each length (1 of 3).

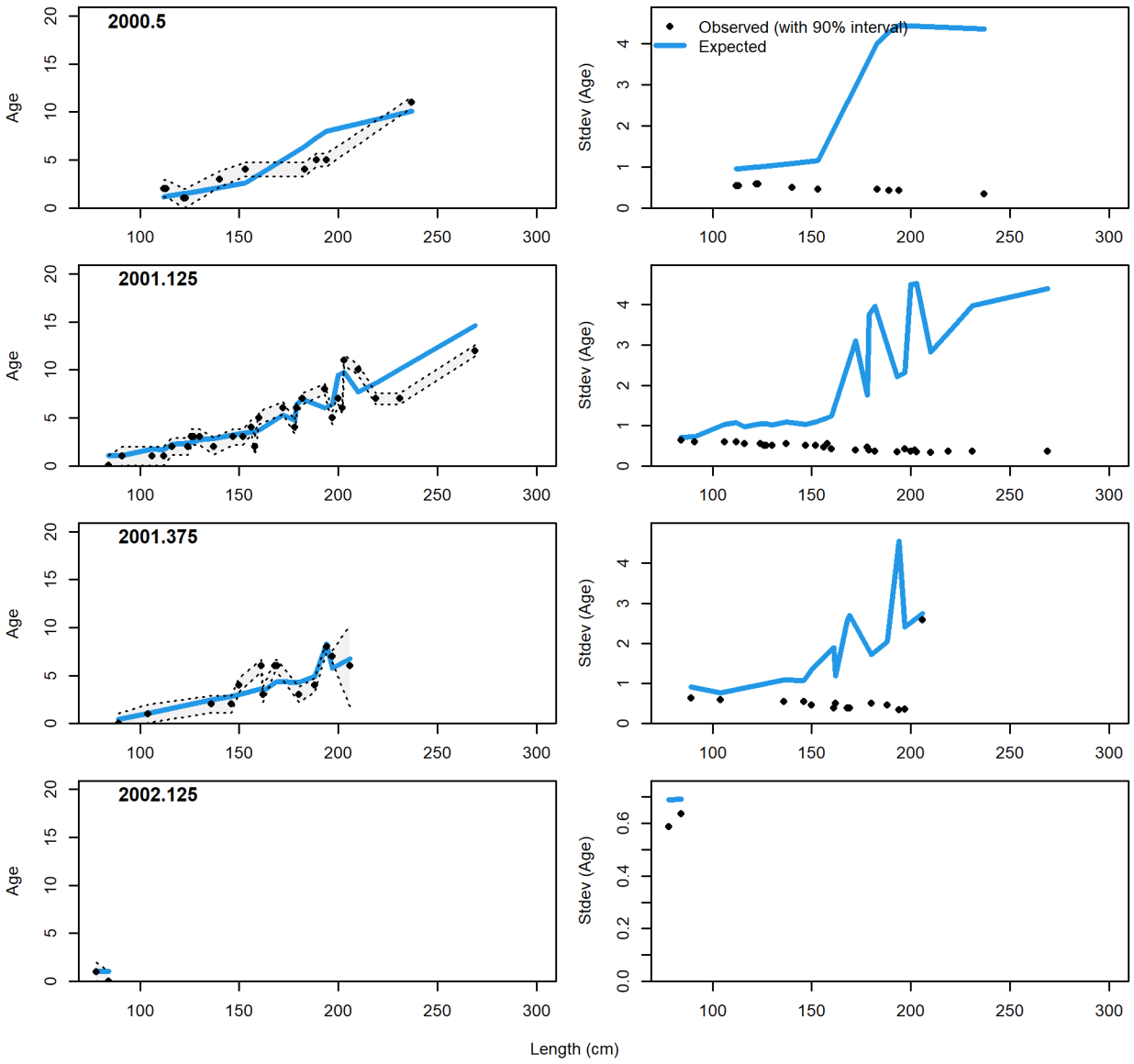


Figure 33: Fits to conditional age-at-length data from fishery 05.AU.1C by quarter with observations (black points, with 90% confidence intervals) and predicted ages at length (blue line) in the left hand panel. The right hand panel shows observed and expected variances, which are not very meaningful with the low numbers of samples each quarter and the high resolution of lengths used, meaning that most lengths do not have replicate samples to estimate a reliable variance on the age for each length (2 of 3).

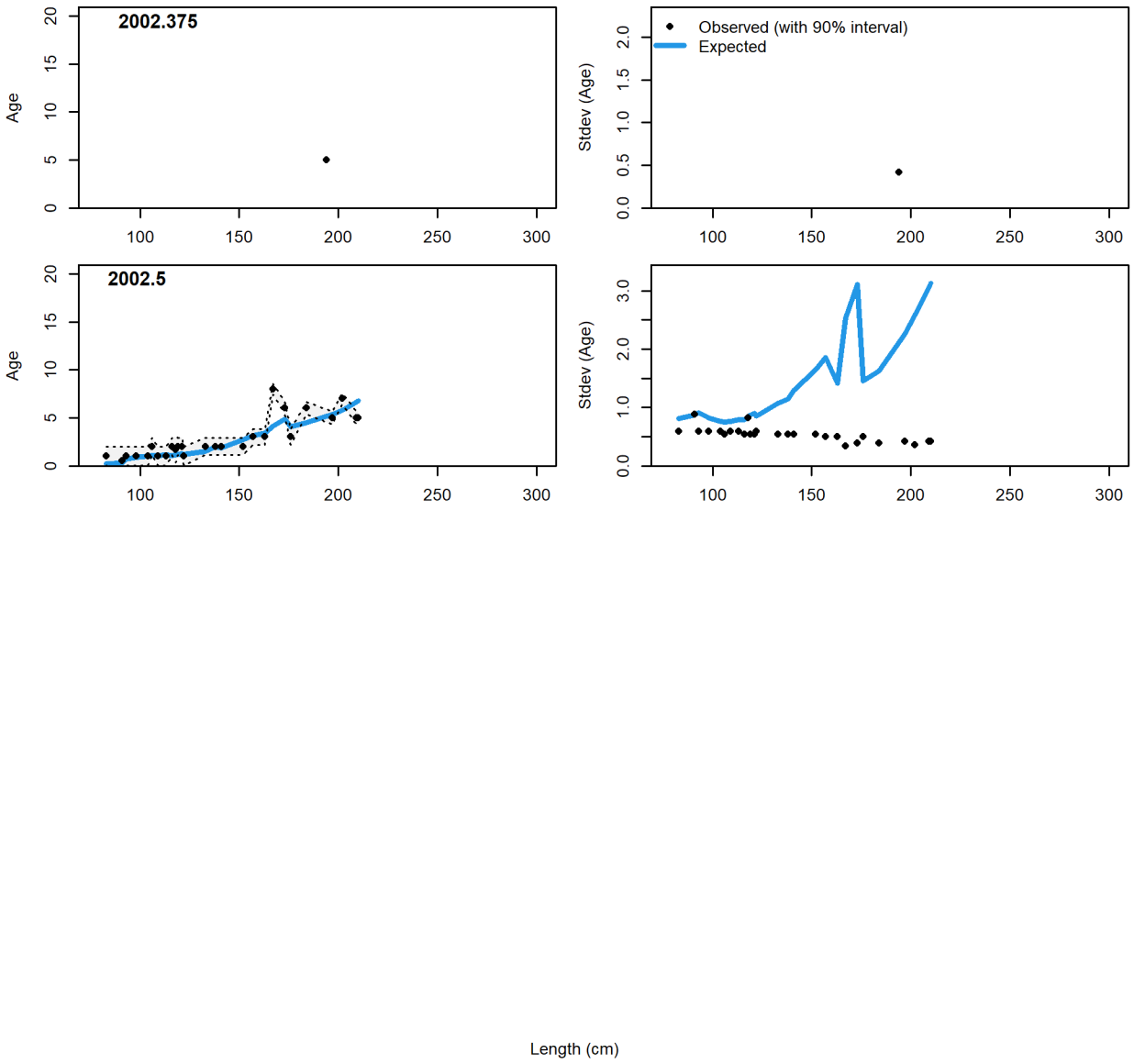


Figure 34: Fits to conditional age-at-length data from fishery 05.AU.1C by quarter with observations (black points, with 90% confidence intervals) and predicted ages at length (blue line) in the left hand panel. The right hand panel shows observed and expected variances, which are not very meaningful with the low numbers of samples each quarter and the high resolution of lengths used, meaning that most lengths do not have replicate samples to estimate a reliable variance on the age for each length (3 of 3).

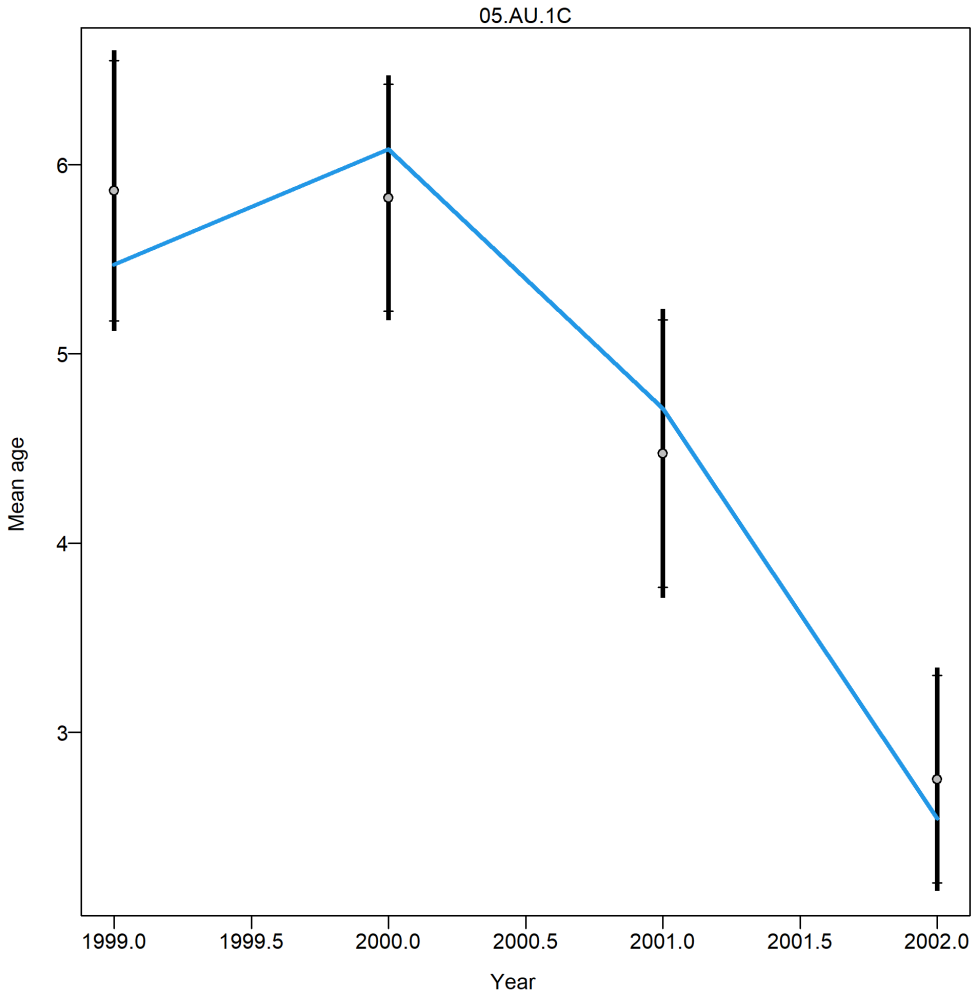


Figure 35: Mean age from conditional age-at-length data (aggregated across length bins) for 05.AU.1C with 95% confidence intervals based on current samples sizes. Francis data weighting method TA1.8: thinner intervals (with capped ends) show result of further adjusting sample sizes based on suggested multiplier (with 95% interval) for conditional age-at-length data from 05.AU.1C: 1.0442 (0.9744-578.9995). For more info, see Francis (2011) and Punt (2017).

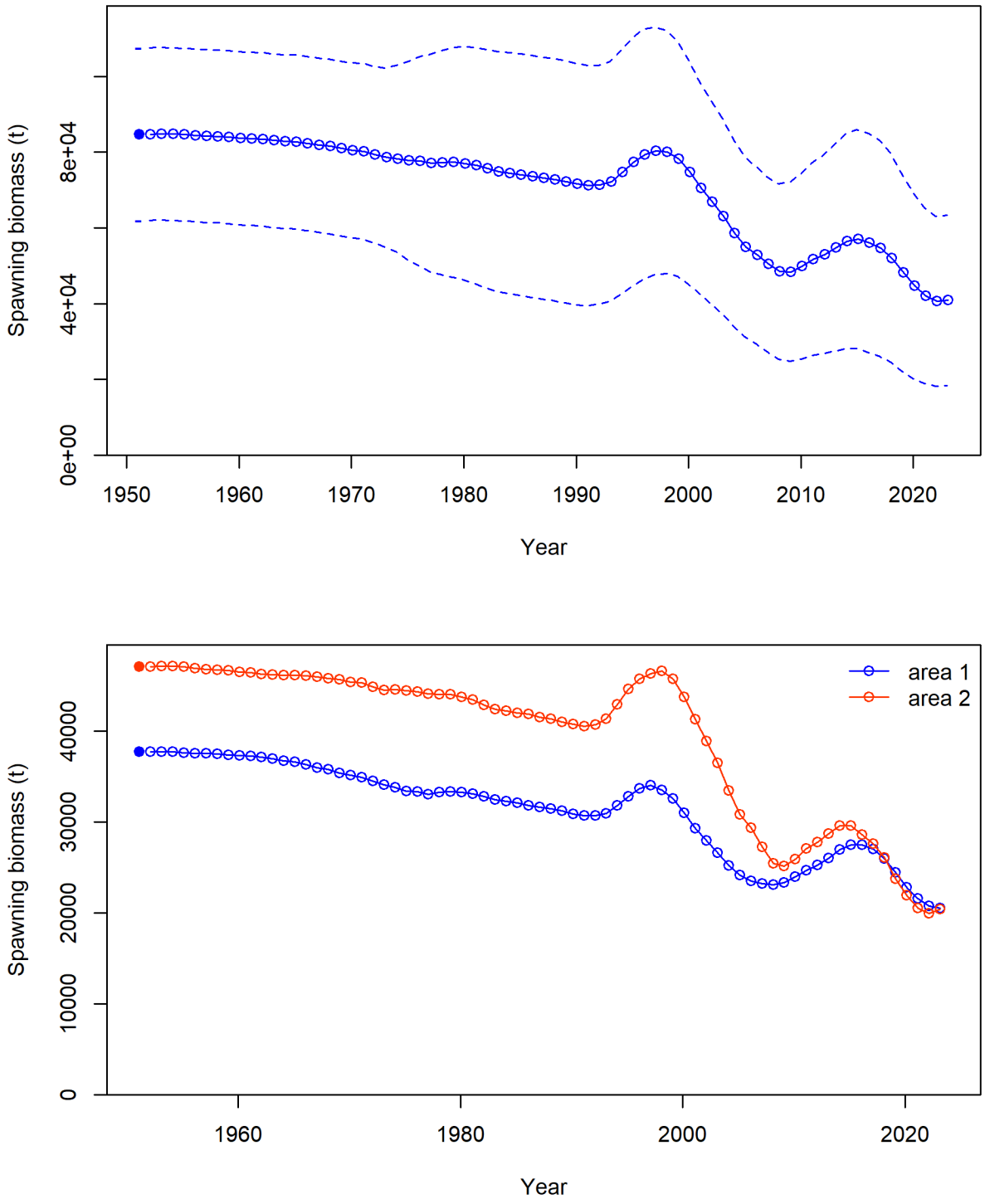


Figure 36: Time-trajectory of estimated female spawning biomass (t) for the diagnostic model, with approximate 95% asymptotic intervals, aggregated over all regions (top) and female spawning biomass (t) by region (bottom), with region 1 shown in blue and region 2 shown in red.

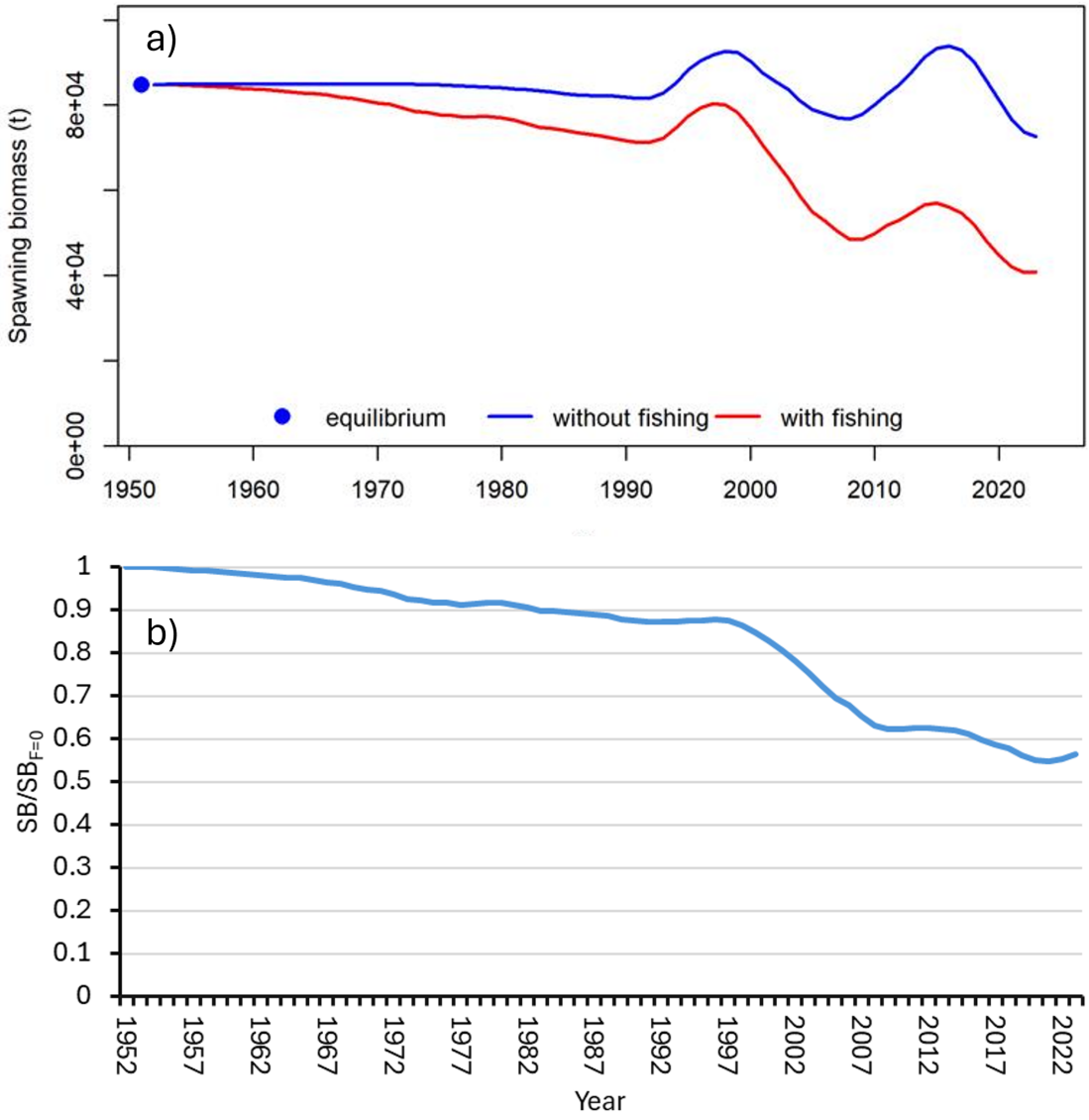


Figure 37: Time-trajectory of fished (red) versus unfished (blue) female spawning biomass (top) and $SB/SB_{F=0}$ (bottom), for the diagnostic model.

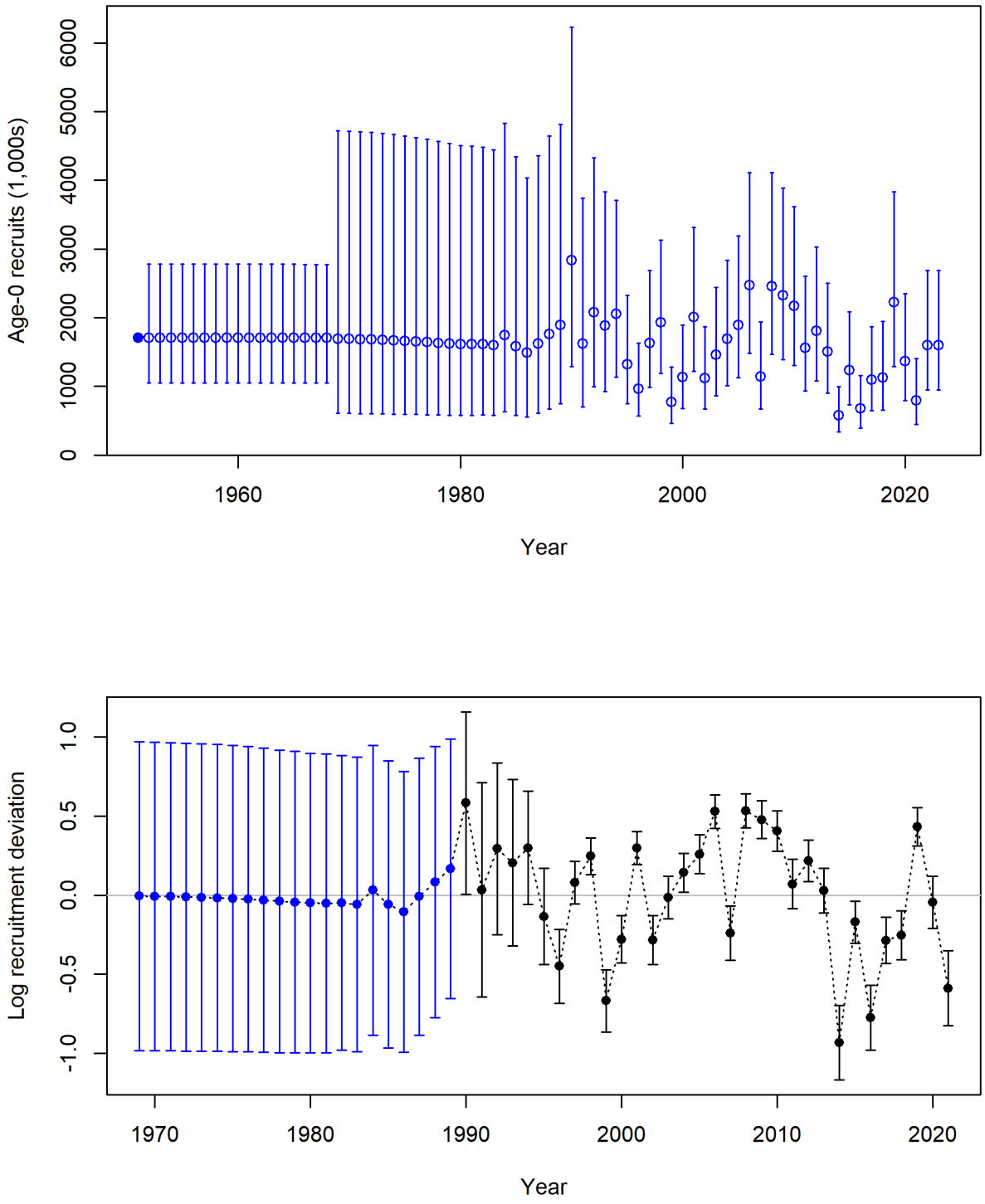


Figure 38: Recruitment estimation for the diagnostic model, with estimated age zero recruits, in 1,000s, with approximate 95% asymptotic intervals (top) and estimated recruitment deviations, with approximate 95% asymptotic intervals (bottom).

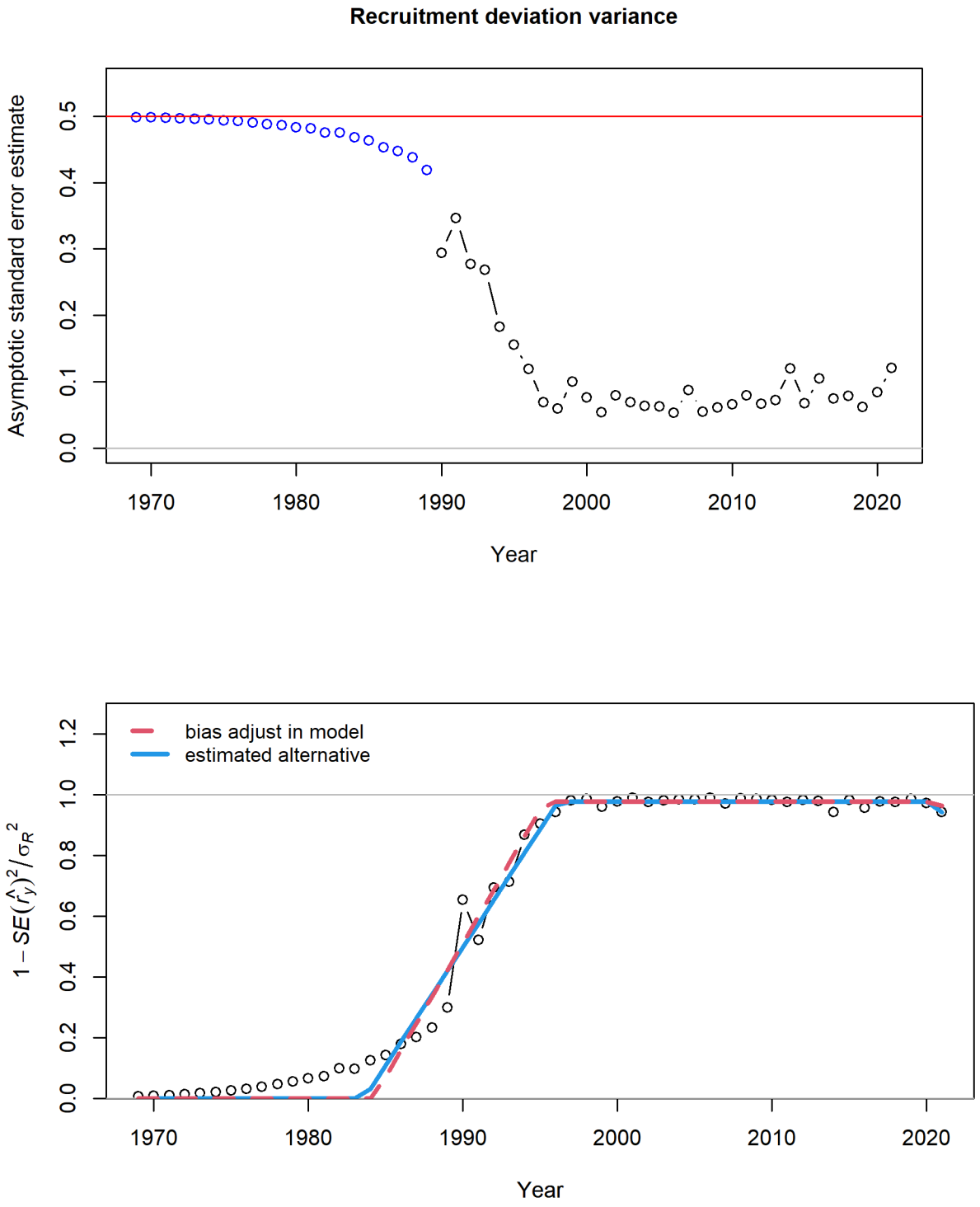


Figure 39: Recruitment variance estimation for the diagnostic model, showing the standard errors of recruitment deviation estimates (top) and the bias adjustment (bottom).

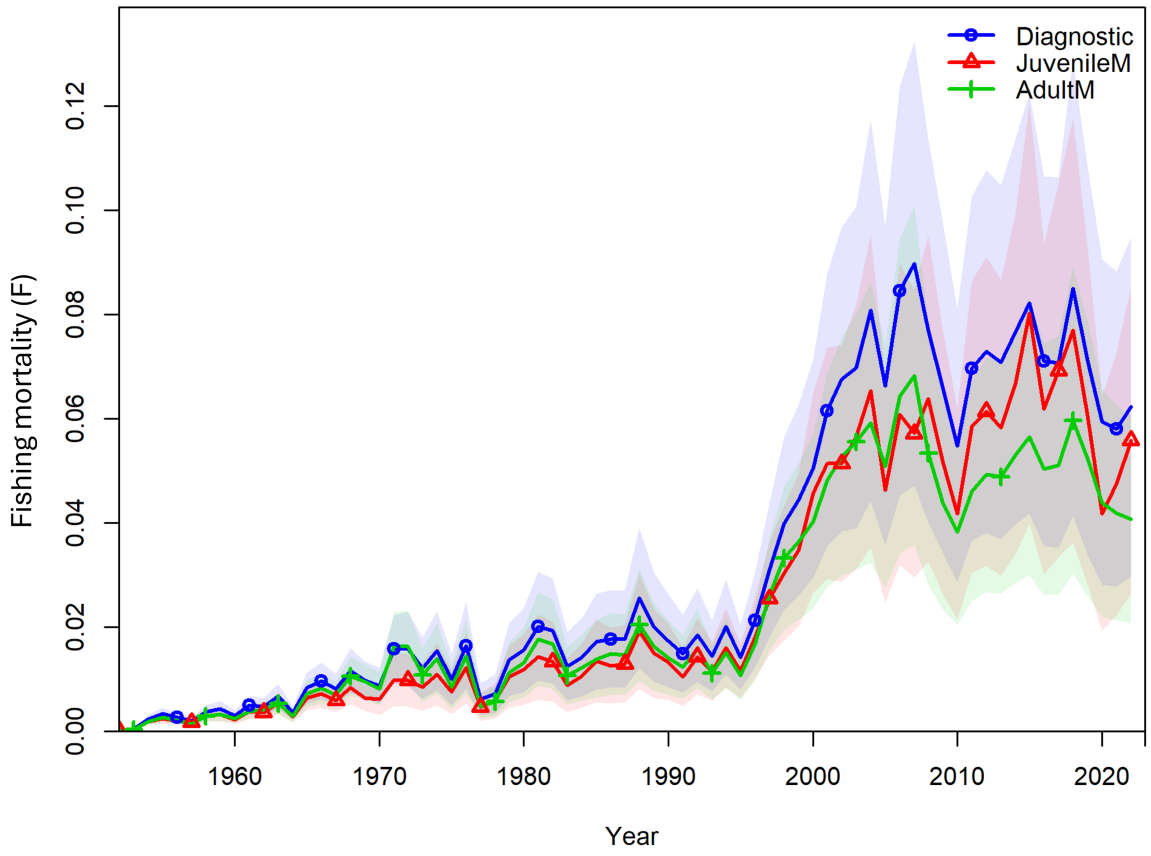


Figure 40: Annual fishing mortality summary by year, summarised by a numbers weighted mean F, calculated across age classes 3–12 (Diagnostic, blue), for juveniles, age range 1–3 years (JuvenileM, red) and for adults, age range 8–15 years (AdultM, green) for the diagnostic model, with approximate 95% asymptotic intervals.

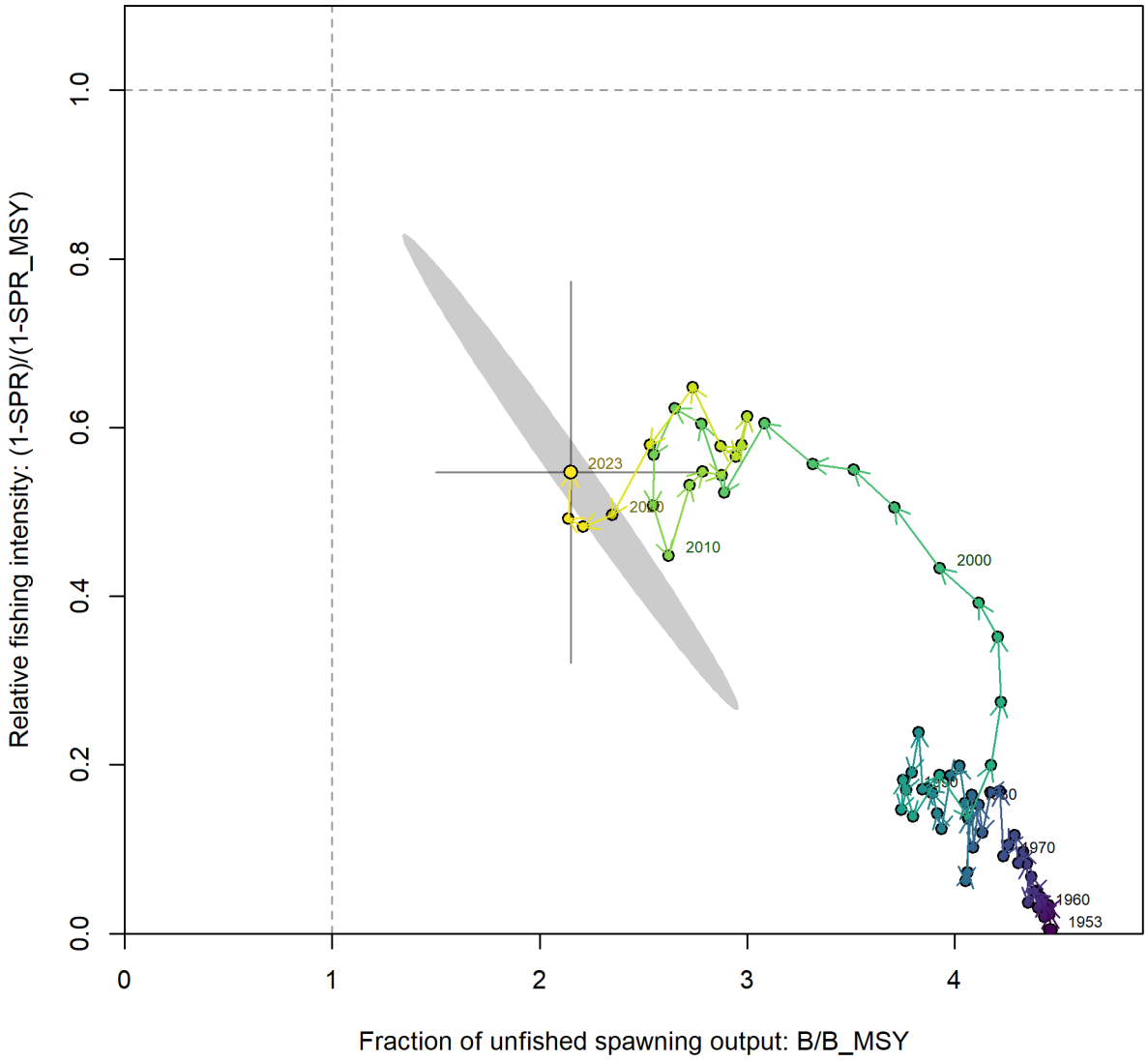


Figure 41: Phase plot of biomass ratio (SB/SB_{MSY}) vs. SPR, $(1-\text{SPR})/(1-\text{SPR}_{\text{MSY}})$. Each point in the phase plot represents the biomass ratio at the start of the year and the relative fishing intensity in that same year. The colours on the blue end of the spectrum represent early years and the yellow colours represent the most recent years. Lines through the final point show 95% intervals based on the asymptotic uncertainty for each dimension. The shaded ellipse is a 95% region which accounts for the estimated correlation between the two quantities: -0.991.

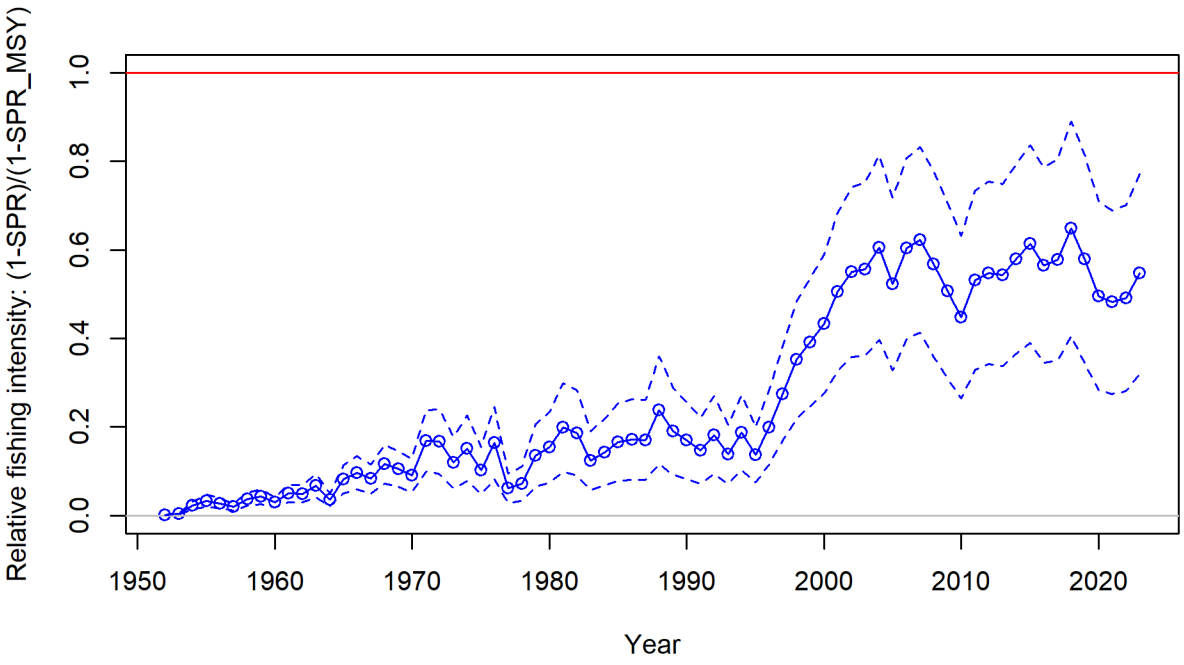


Figure 42: Time series of $(1 - \mathrm{SPR})/(1 - \mathrm{SPR}_{\mathrm{MSY}})$ with asymptotic uncertainty.

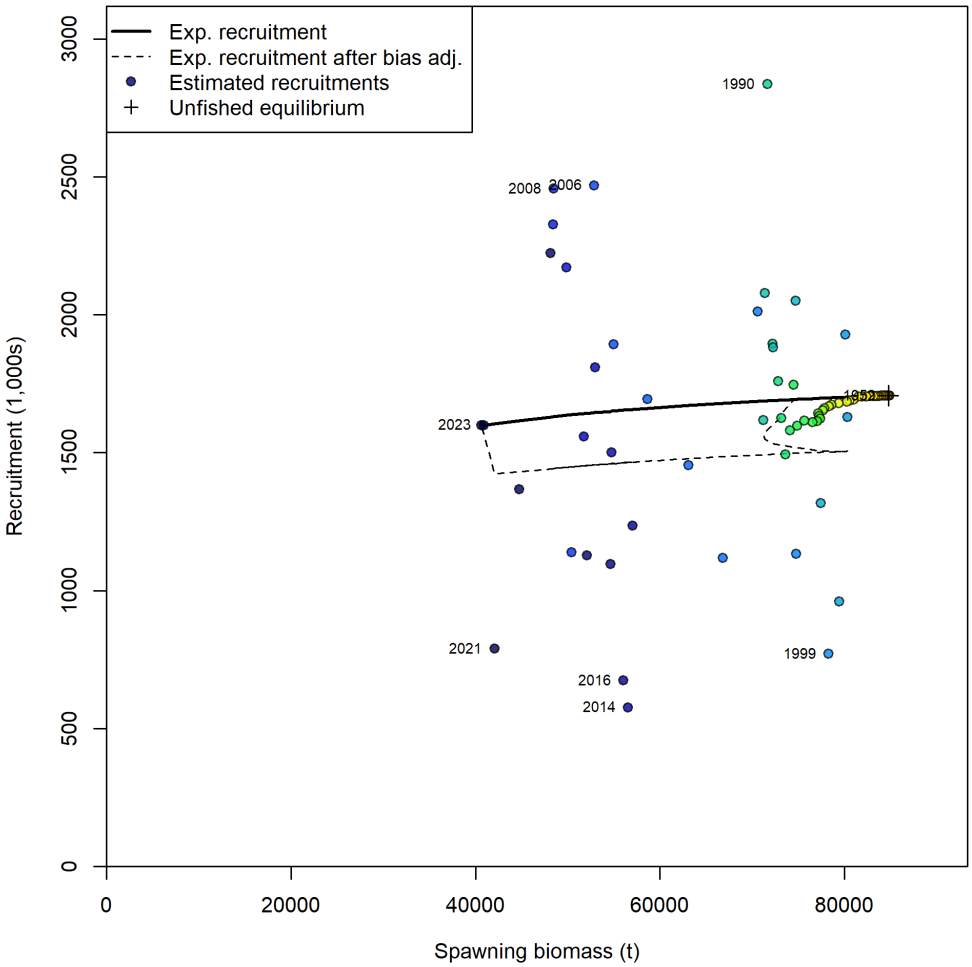


Figure 43: The stock-recruit curve and estimated recruitments for the diagnostic model. Labels are shown on first, last, and years with (log) deviations > 0.5, with point colours indicating year, with warmer colours indicating earlier years and cooler colours showing later years.

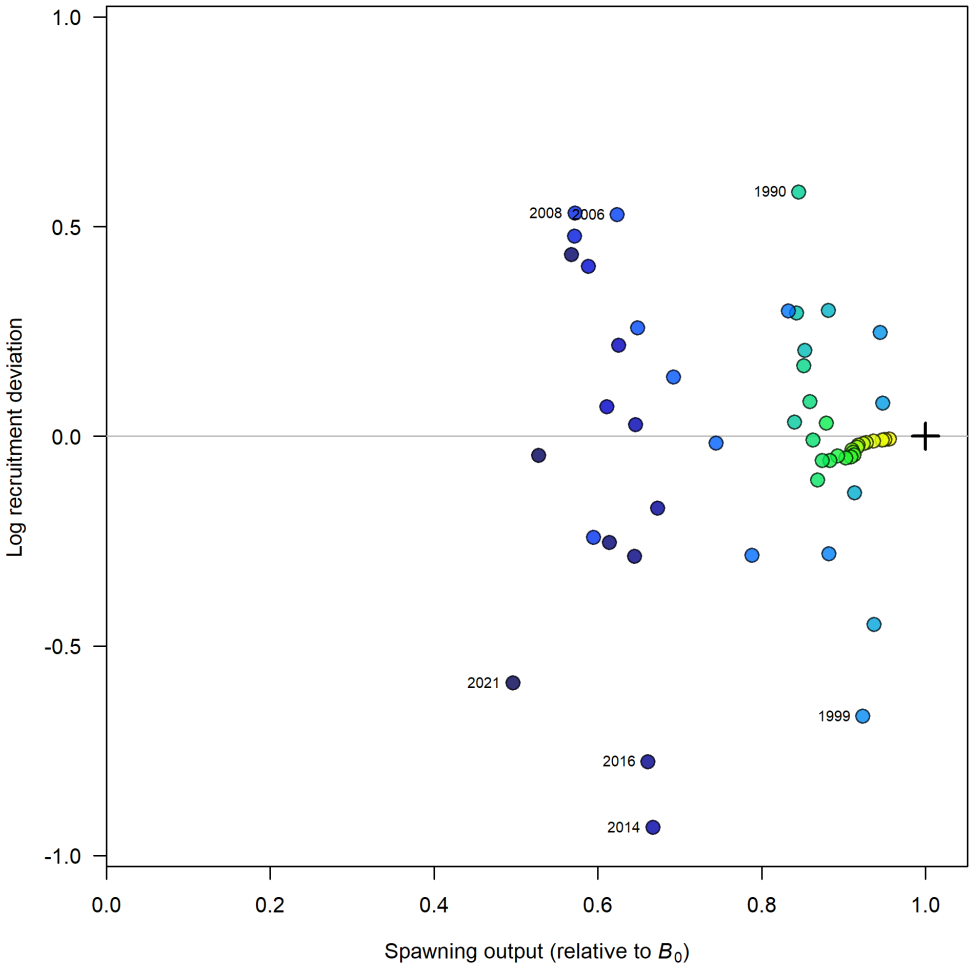


Figure 44: Log recruitment deviations from the stock recruitment curve for the diagnostic model. Labels are shown on first, last, and years with (log) deviations > 0.5, with point colours indicating year, with warmer colours indicating earlier years and cooler colours showing later years.

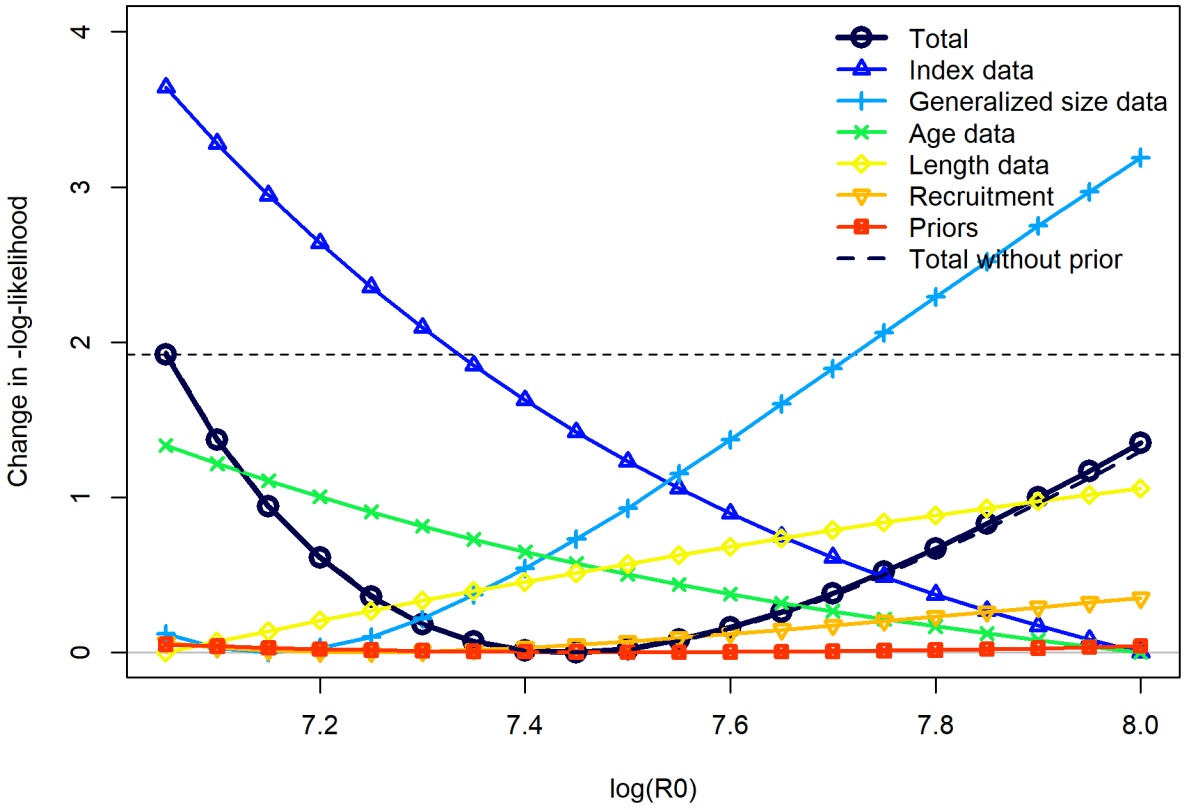


Figure 45: The likelihood profile for $\log(R_0)$ from the diagnostic model, with $\log(R_0)$ ranging between 7 and 8. The total negative log-likelihood is shown in black, with other components of the likelihood shown in different colours. The index data, or CPUE (dark blue) has the largest influence, supporting a higher population scale, followed, in contrast, by the weight composition data (light blue), which supports a lower population scale.

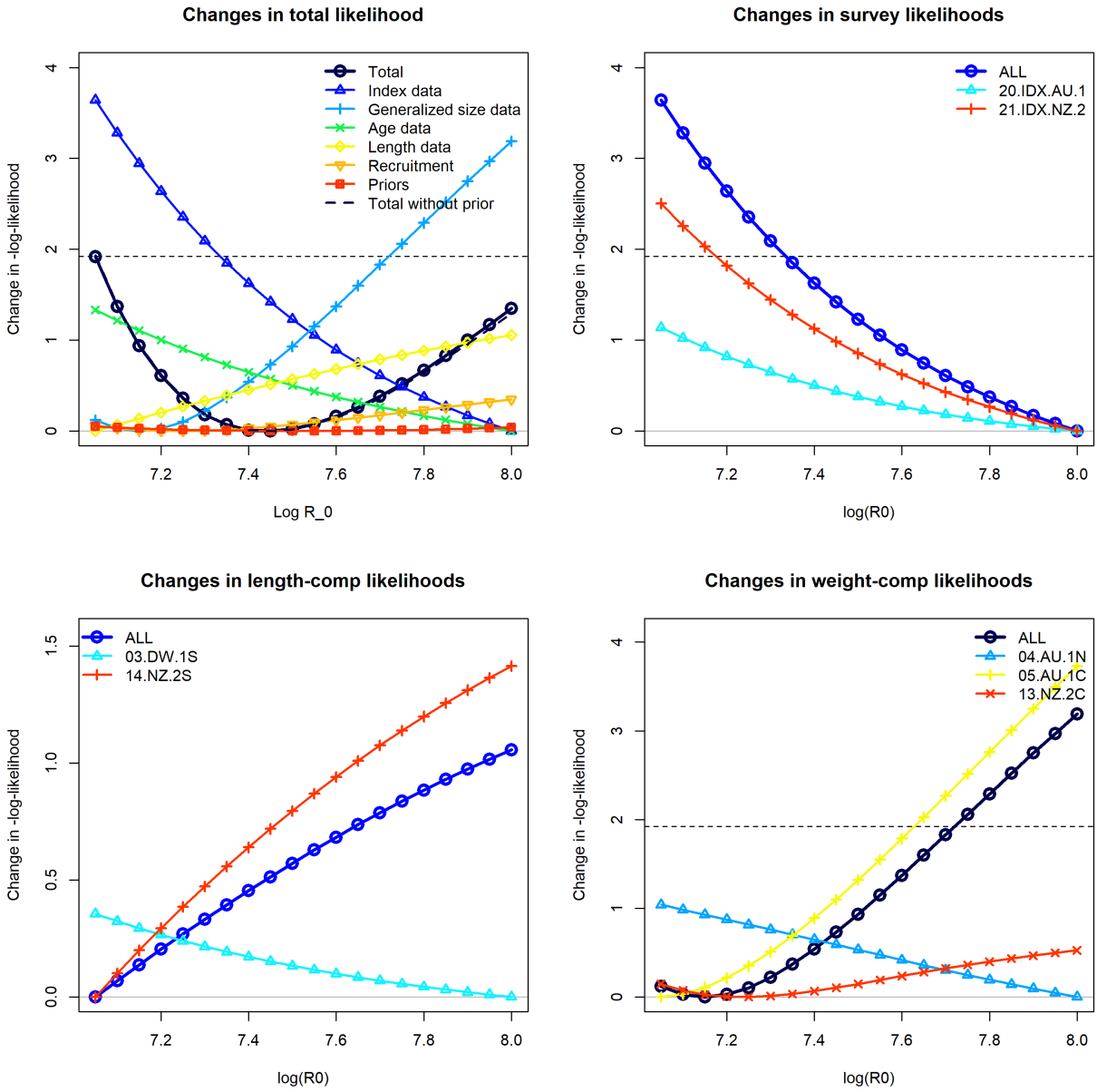


Figure 46: Piner plot for the likelihood profile for $\log(R_0)$ from the diagnostic model, showing components of the change in likelihood for length, weight and surveys (CPUE) by fishery. The total negative log-likelihood is shown either in black (top left and bottom right plots) or in dark blue (top right and bottom left plots). Other components of the likelihood, or the contribution to the likelihood from individual fisheries are shown in different labelled colours. The major contribution from the CPUE data supporting a higher population scaling parameter comes from the New Zealand CPUE data (top right panel). The major contribution from the weight data supporting a lower scaling population parameter comes from the Australian fishery 05.AU.1C (bottom right panel). The length composition data has a smaller contribution, but this does show some difference in the influence of these datasets between length data from the two southern fisheries, 03.DW.1S and 14.NZ.2S.

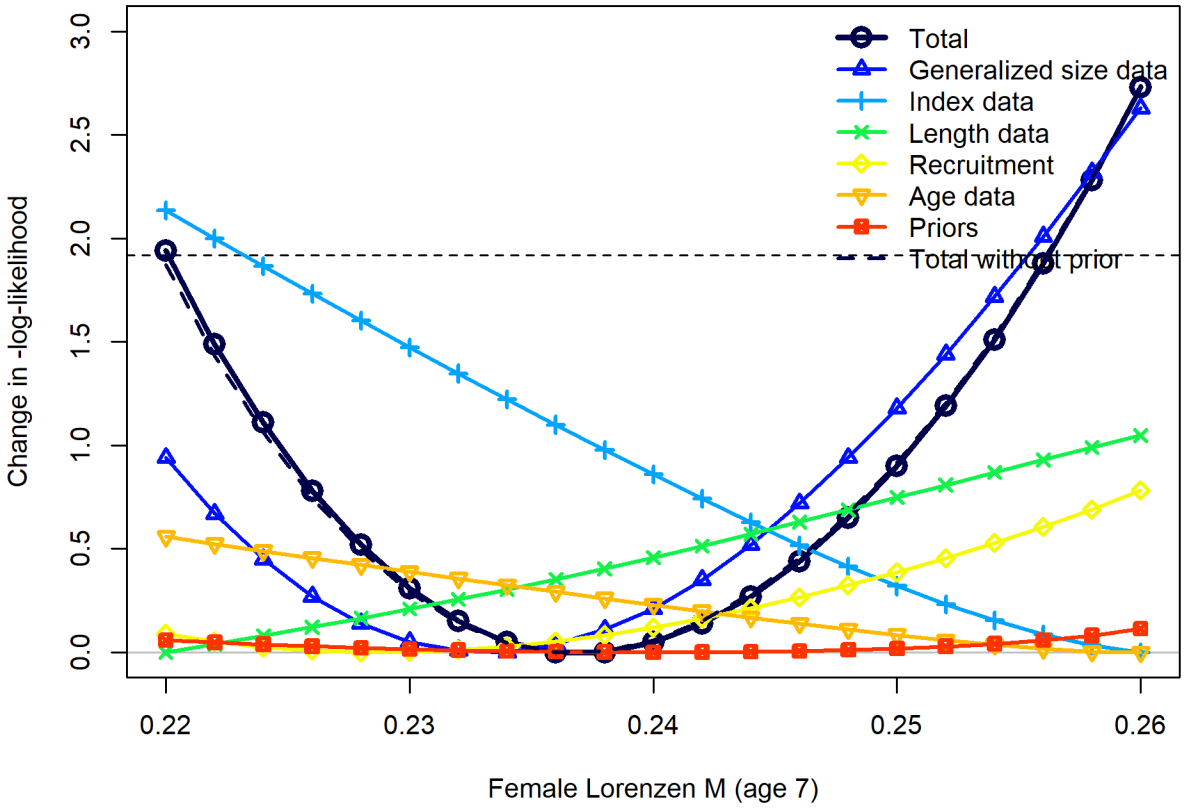


Figure 47: The likelihood profile for the scaling parameter for Lorenzen M from the diagnostic model, ranging between 0.22 and 0.26. The total negative log-likelihood is shown in black, with other components of the likelihood shown in different colours. The index data, or CPUE (light blue) and weight composition data (dark blue) have the largest influence, and are in conflict, with the CPUE data supporting a higher M scale and the weight composition data supporting a lower M scale.

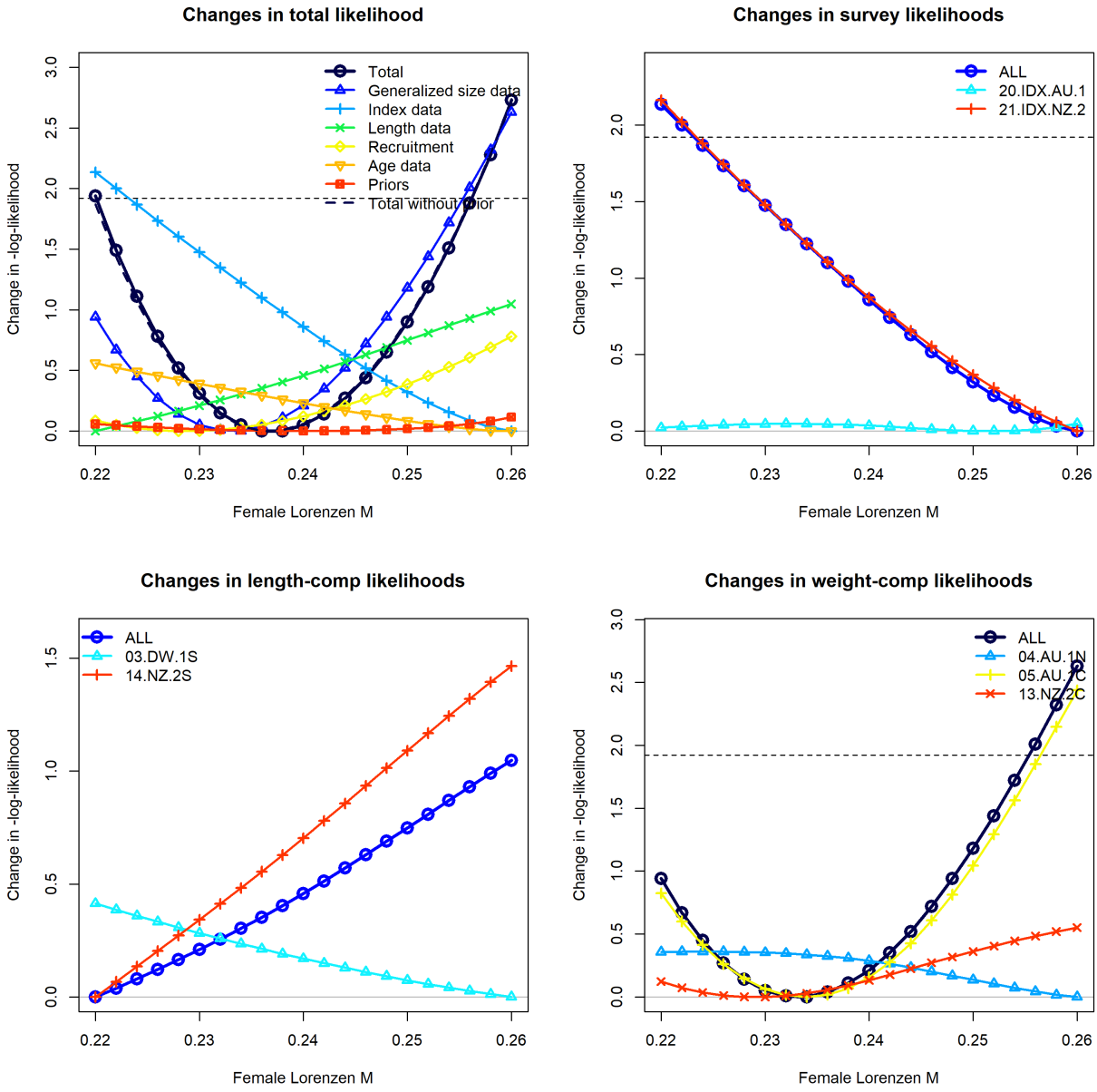


Figure 48: Piner plot for the likelihood profile for the scaling parameter for Lorenzen M from the diagnostic model, showing components of the change in likelihood for length, weight and surveys (CPUE) by fishery. The total negative log-likelihood is shown either in black (top left and bottom right plots) or in dark blue (top right and bottom left plots). Other components of the likelihood, or the contribution to the likelihood from individual fisheries are shown in different labelled colours. The major contribution from the CPUE data supporting a higher M scale comes from the New Zealand CPUE data (top right panel). The major contribution from the weight data supporting a lower M scale comes from the Australian fishery 05.AU.1C (bottom right panel). The length composition data has a smaller contribution, but this does show some difference in the (small) influence of these datasets between length data from the two southern fisheries, 03.DW.1S and 14.NZ.2S.

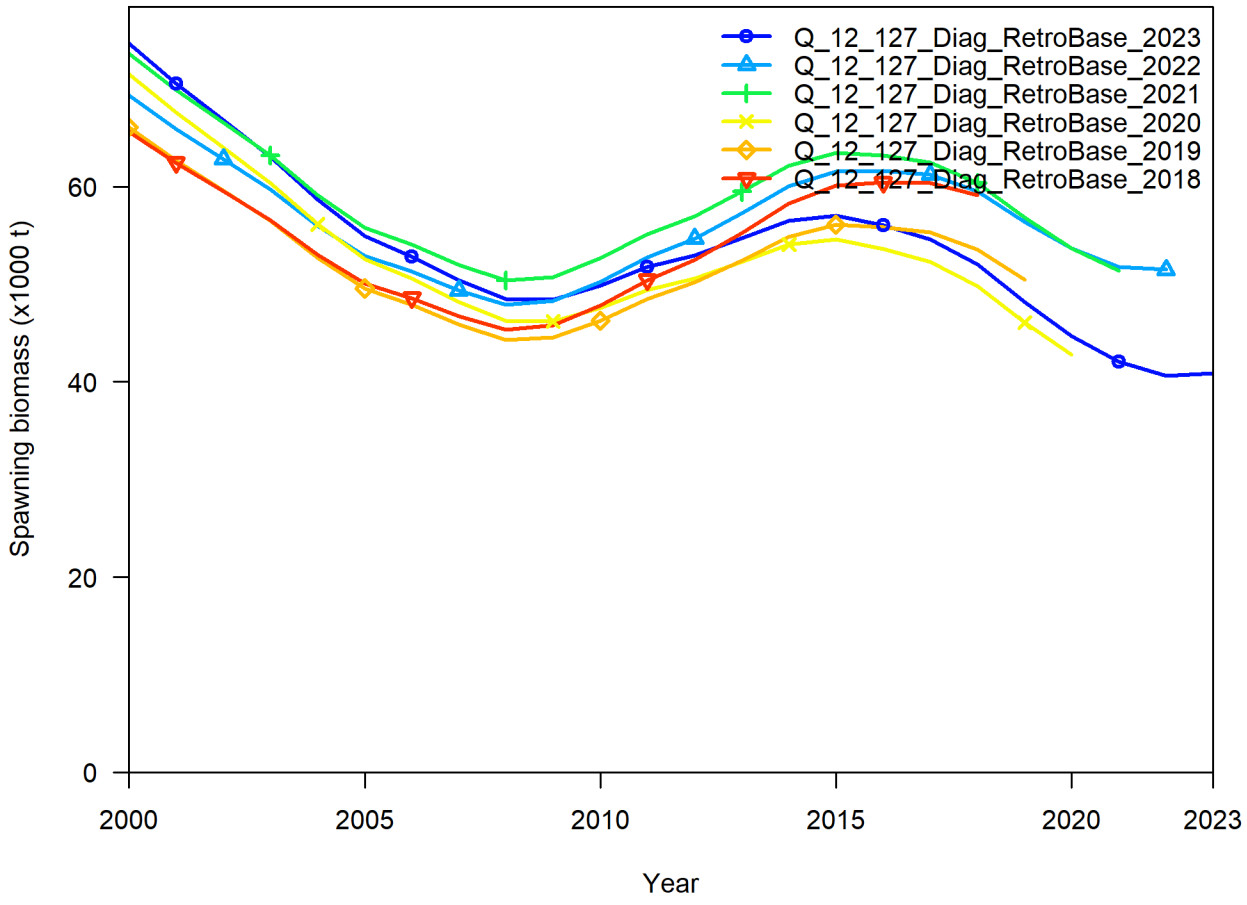


Figure 49: Retrospectives for absolute female spawning biomass, with data included to 2023 (dark blue), followed by five peels, each with the successive "last year of data" removed, so with data included respectively, only to 2022, 2021, 2020, 2019 and 2018 (red).

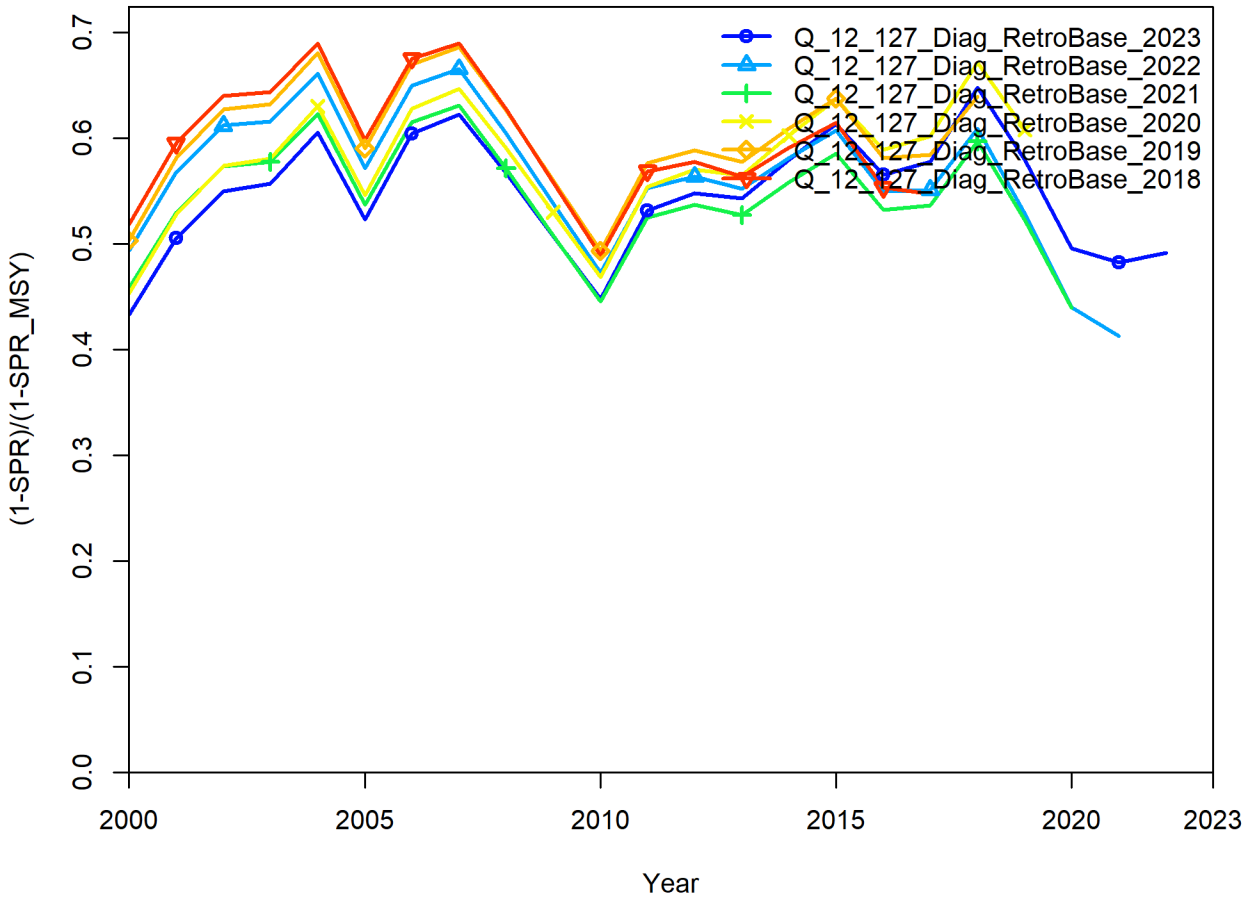


Figure 50: Retrospectives for fishing intensity, measured by the SPR, with data included to 2023 (dark blue), followed by five peels, each with the successive "last year of data" removed, so with data included respectively, only to 2022, 2021, 2020, 2019 and 2018 (red).

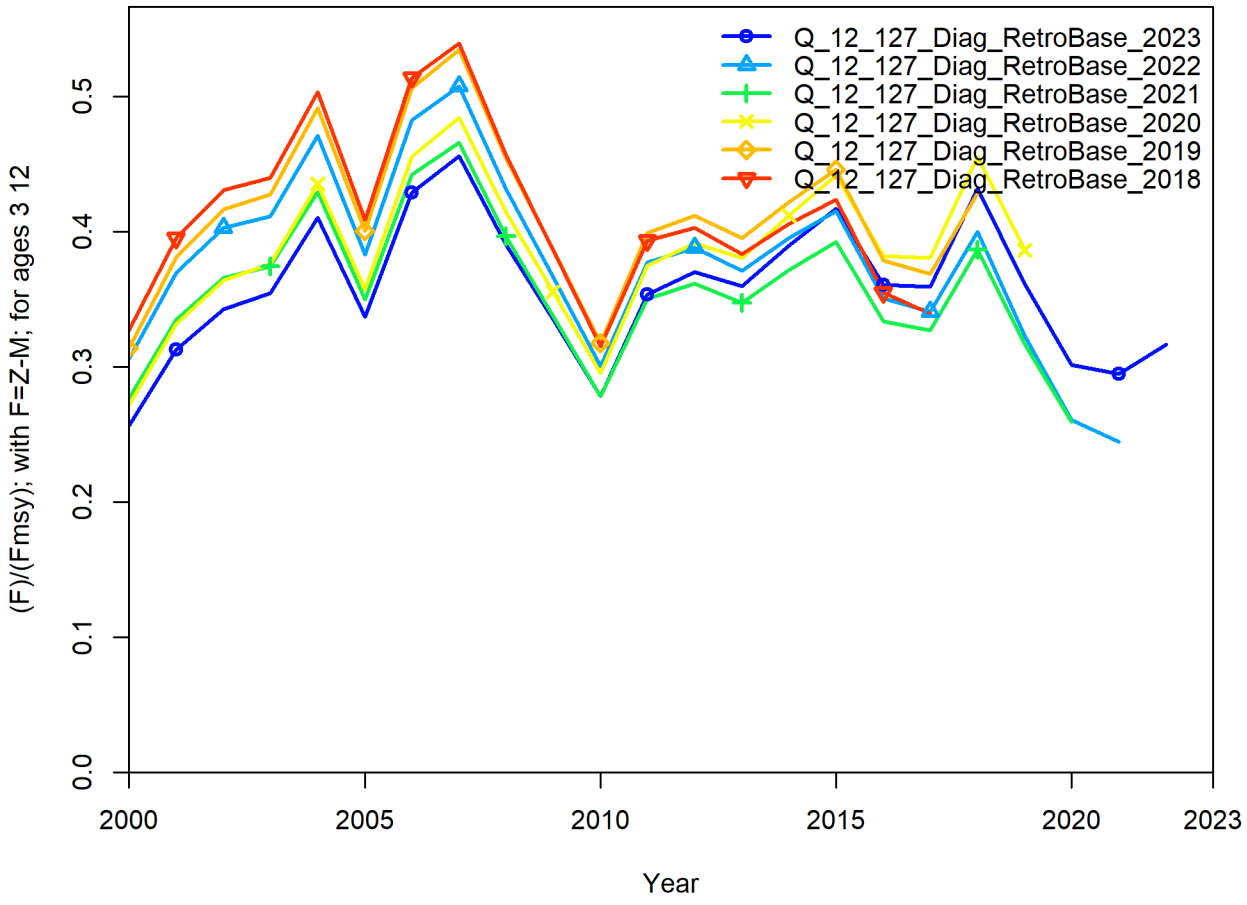


Figure 51: Retrospectives for fishing intensity, measured by F/F_{MSY}, with data included to 2023 (dark blue), followed by five peels, each with the successive "last year of data" removed, so with data included respectively, only to 2022, 2021, 2020, 2019 and 2018 (red).

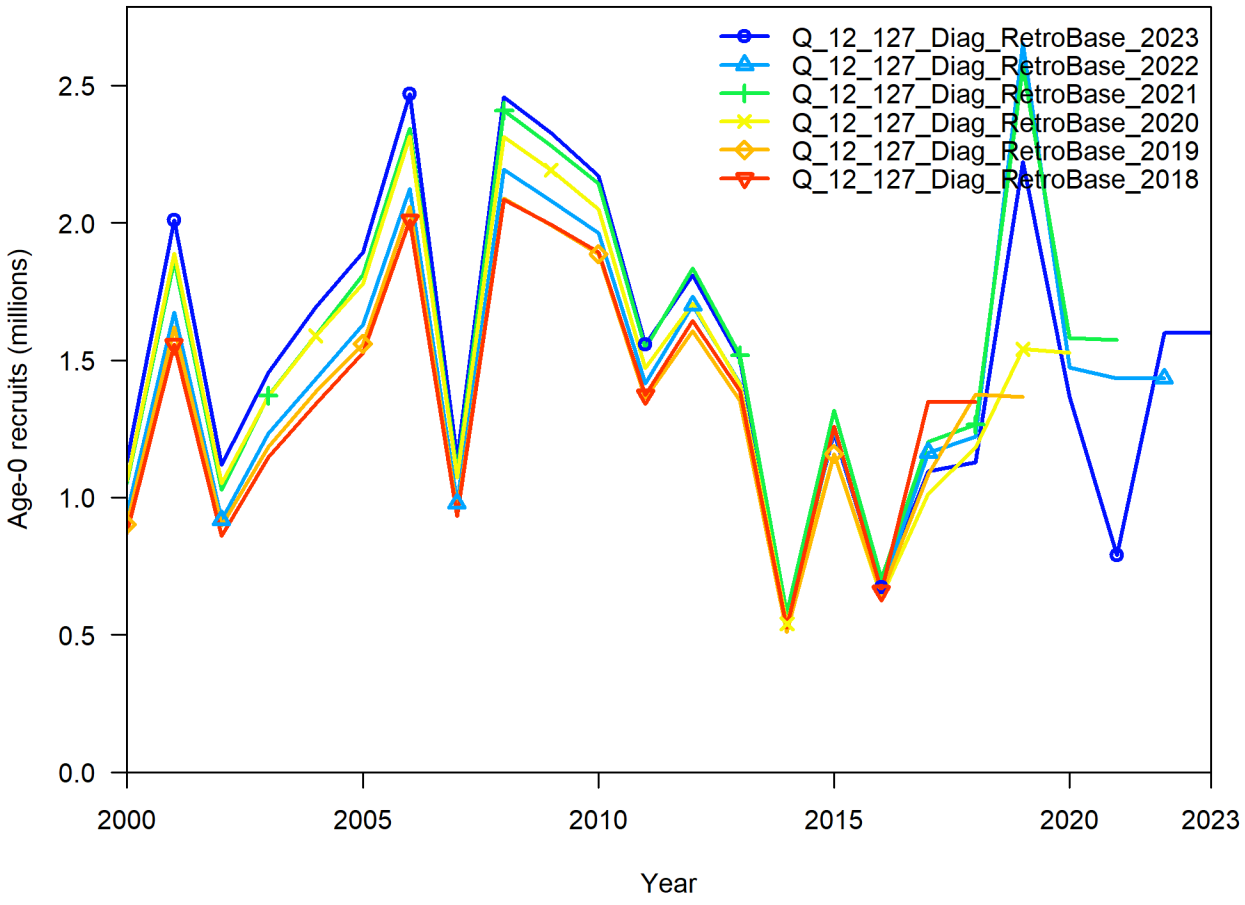


Figure 52: Retrospectives for recruitment, measured by numbers of age-0 recruits, with data included to 2023 (dark blue), followed by five peels, each with the successive "last year of data" removed, so with data included respectively, only to 2022, 2021, 2020, 2019 and 2018 (red).

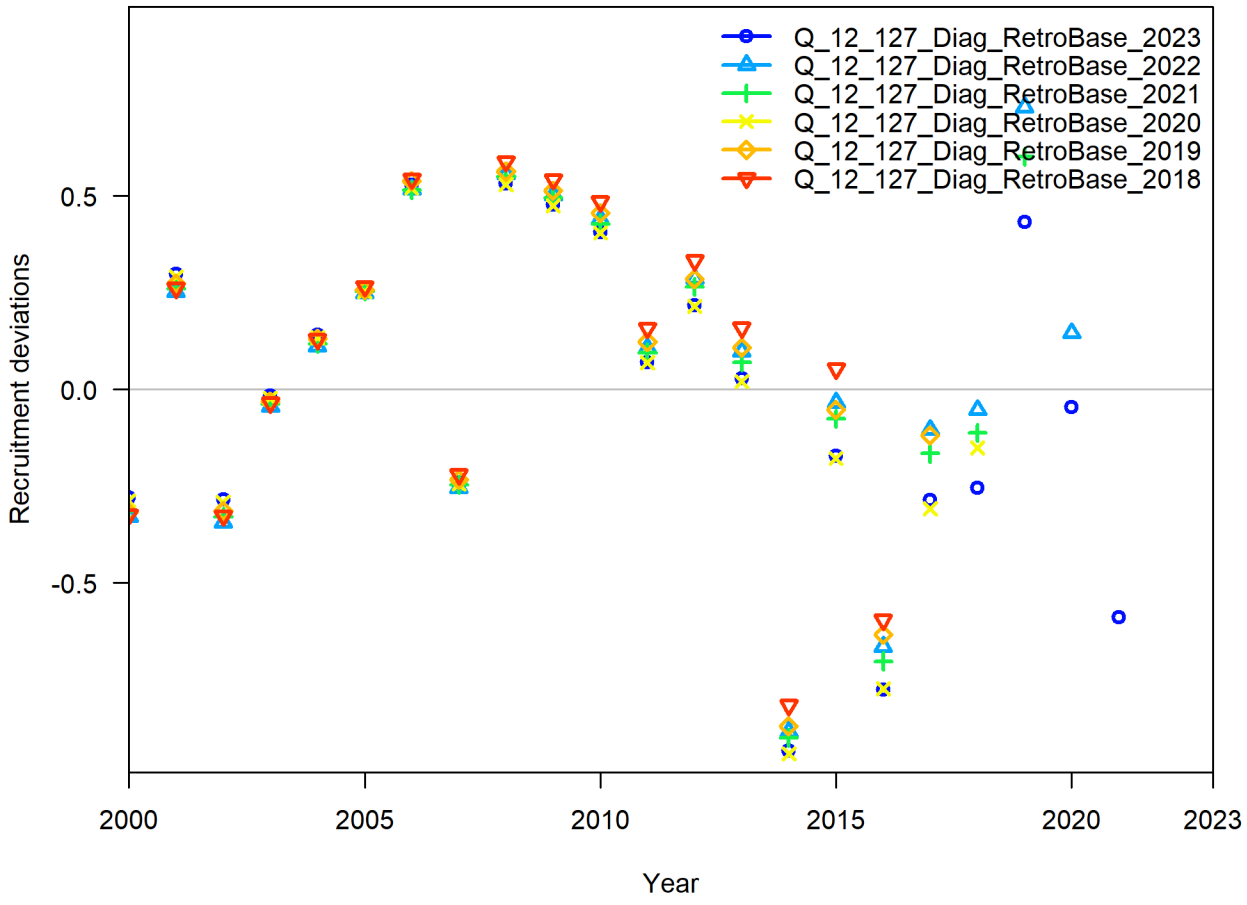


Figure 53: Retrospectives for recruitment, measured by estimated recruitment deviations, with data included to 2023 (dark blue), followed by five peels, each with the successive "last year of data" removed, so with data included respectively, only to 2022, 2021, 2020, 2019 and 2018 (red).

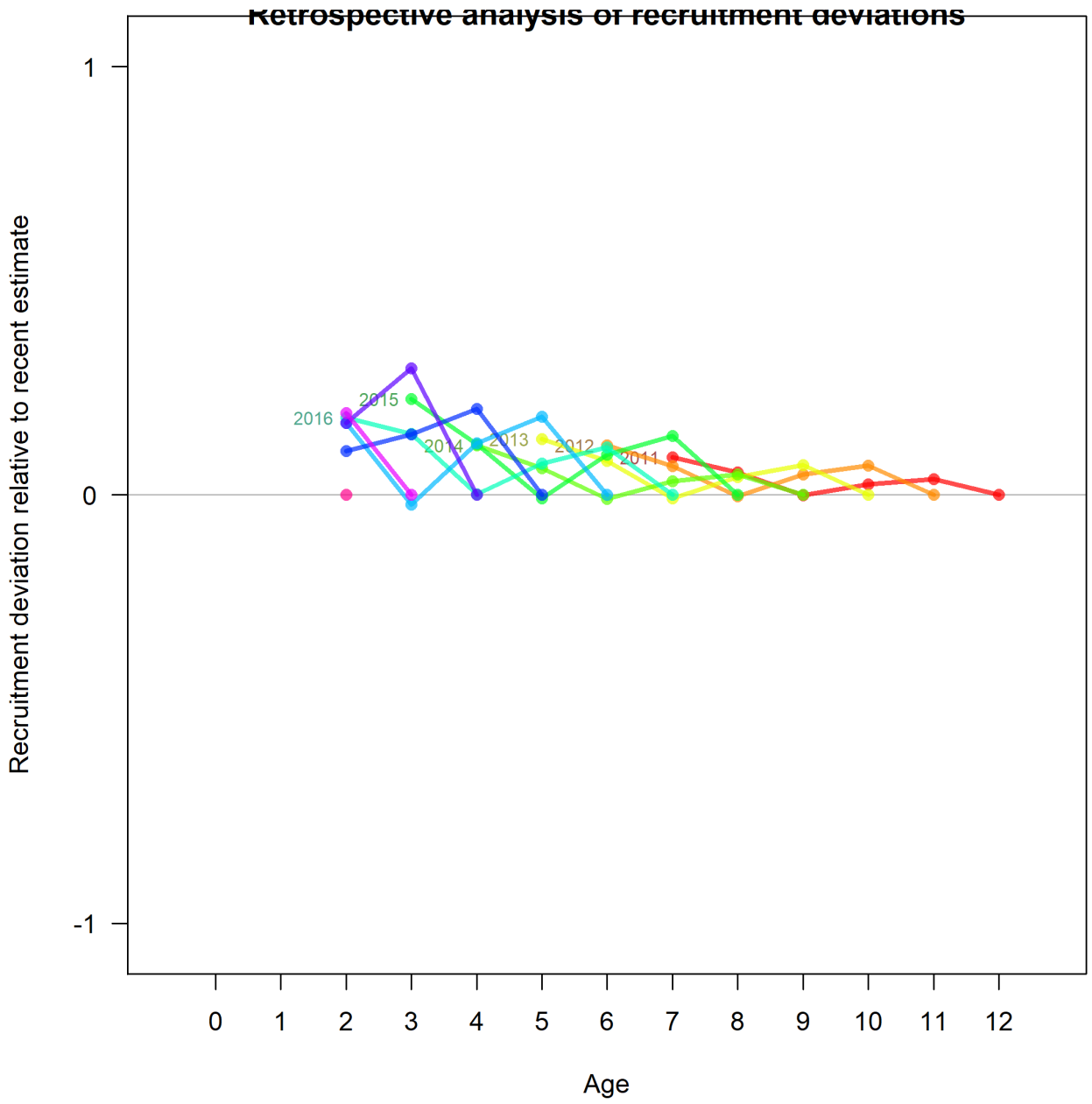


Figure 54: Recruitment deviations from retrospectives, as a squid plot, with data included to 2023, and five retrospective peels, respectively, including data only to 2022, 2021, 2020, 2019 and 2018. This shows changes in the recruitment deviations for particular cohorts, as each of the last five years of data are successively removed, with the the age of that cohort at each peel on the x-axis. The red string (labelled 2011) is the cohort from 2011, aged 12 years old in 2023, and 7 years in 2018, the last peel. Each coloured string, corresponding to a particular cohort, includes a maximum of six points, one for the diagnostic model (on the right of each string, set to $y = 0$), and then one additional point for each of the five retrospective peels. Each string, when followed from right to left, represents removal of successive years of data. The changes in the estimates of the recruitment deviations for each cohort, as each year of data is removed, are measured by changes in the y-axis, along each string, relative to the final recruitment deviation for that cohort from the diagnostic model. A negative value indicates a revision downwards and a positive value indicates a revision upwards. This plot is clearly unbalanced, as it features revisions upwards, compared to the diagnostic model in every case, with a single exception, for cohort 2017, after 3 peels.

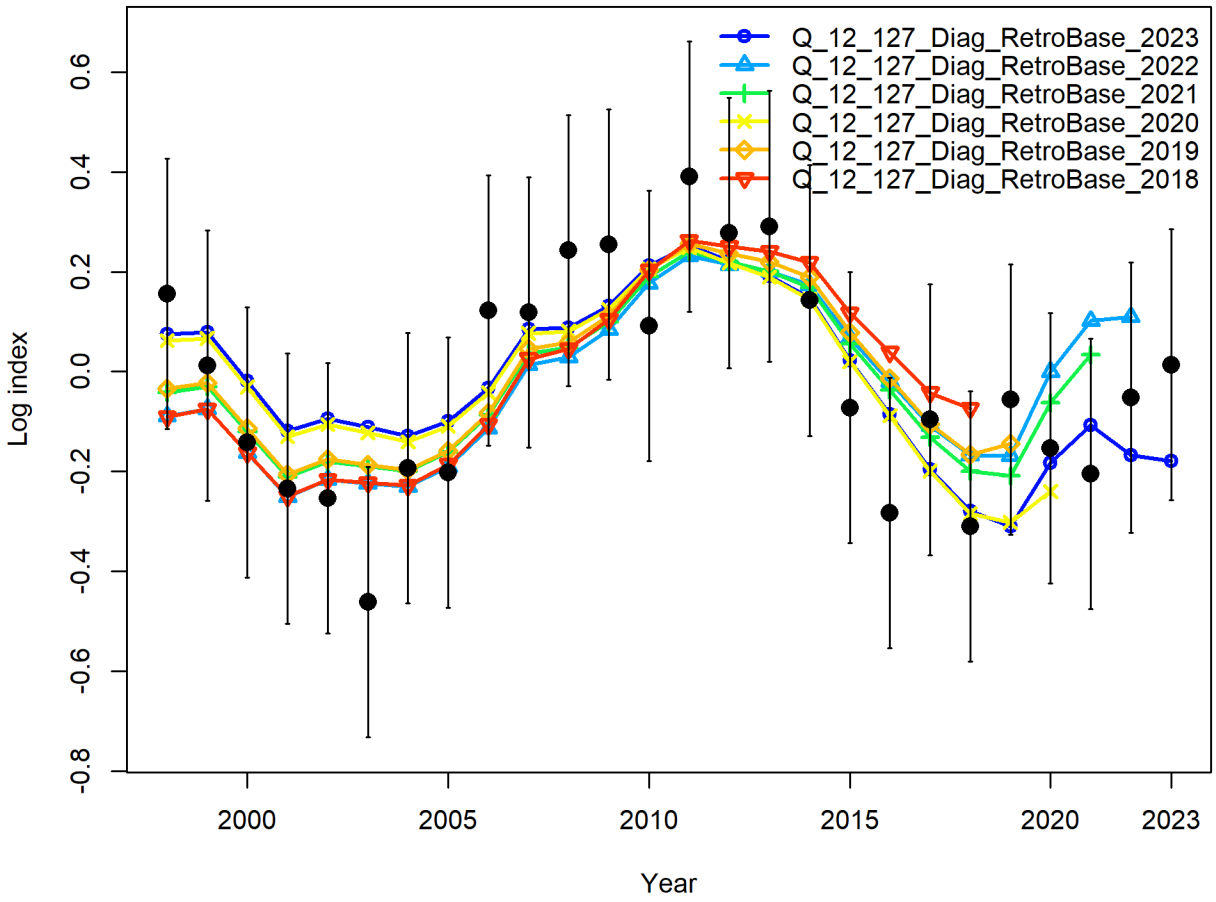


Figure 55: Retrospectives plotted for CPUE fits for the Australian CPUE data from fishery 20.IDX.AU.1, with data included to 2023 (dark blue), followed by five peels, each with the successive "last year of data" removed, so with data included respectively, only to 2022, 2021, 2020, 2019 and 2018 (red).

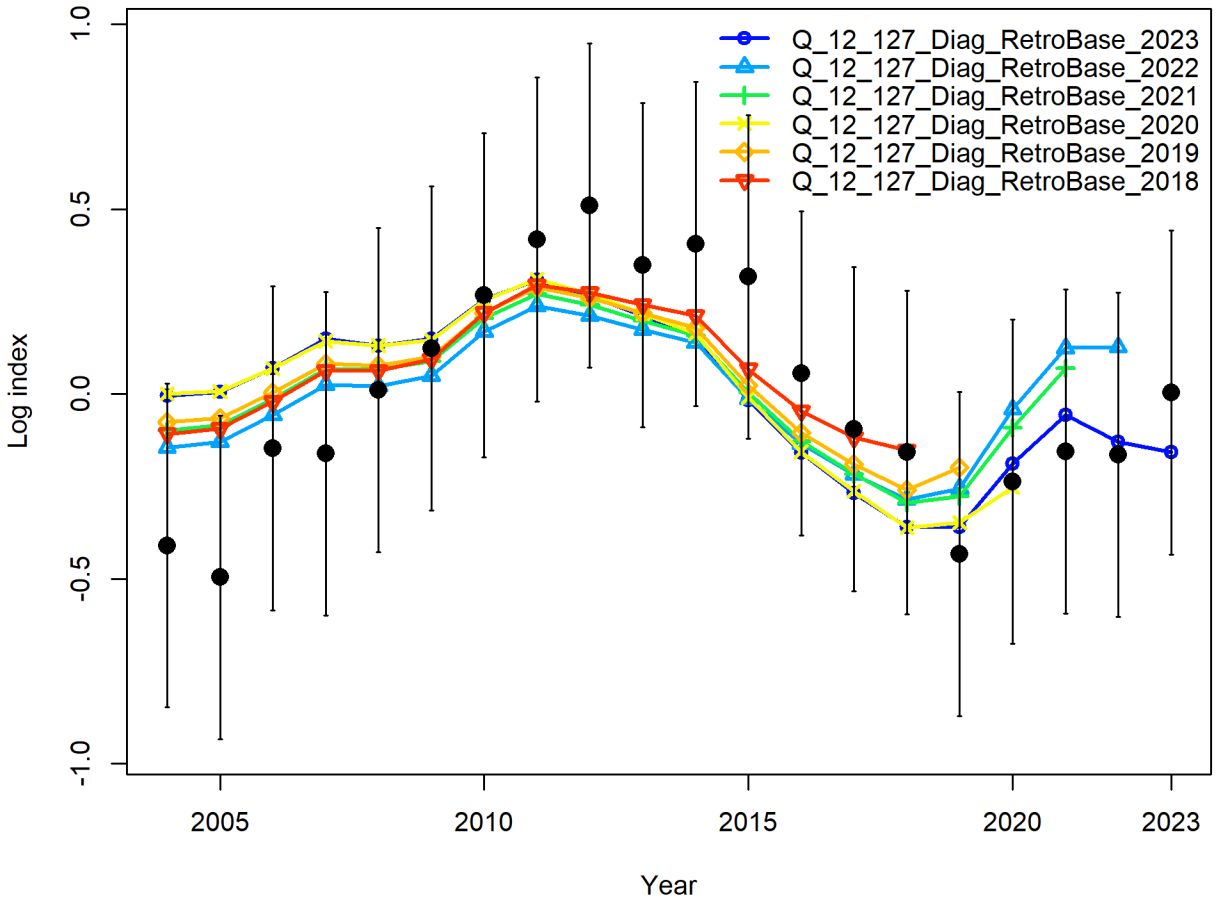


Figure 56: Retrospectives plotted for CPUE fits for the New Zealand CPUE data from fishery 21.IDX.NZ.2, with data included to 2023 (dark blue), followed by five peels, each with the successive "last year of data" removed, so with data included respectively, only to 2022, 2021, 2020, 2019 and 2018 (red).

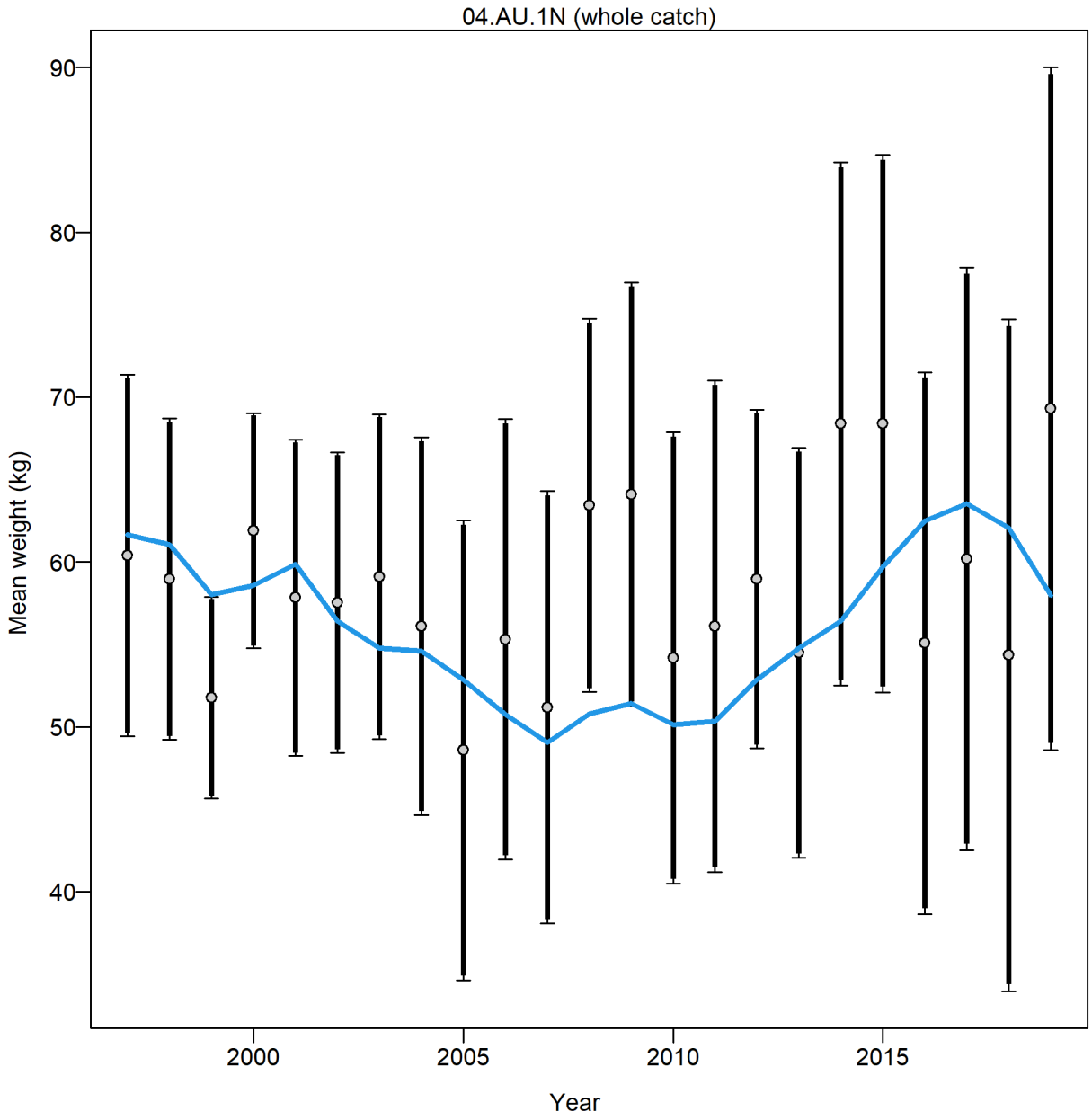


Figure 57: Annually aggregated observed (grey dots) and predicted (blue line) median fish weights (kg) for fishery 04.AU.1N.

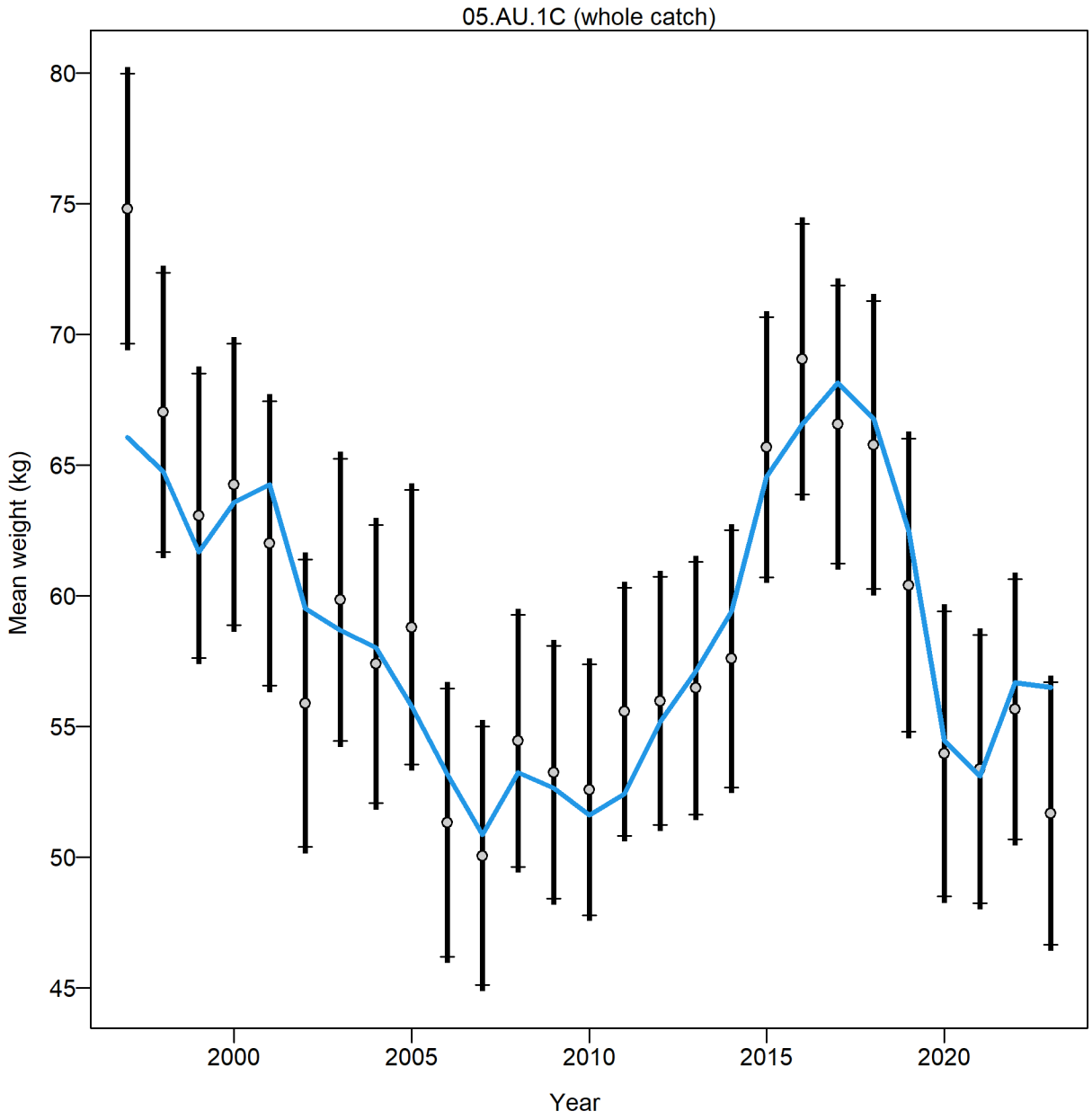


Figure 58: Annually aggregated observed (grey dots) and predicted (blue line) median fish weights (kg) for fishery 05.AU.1C.

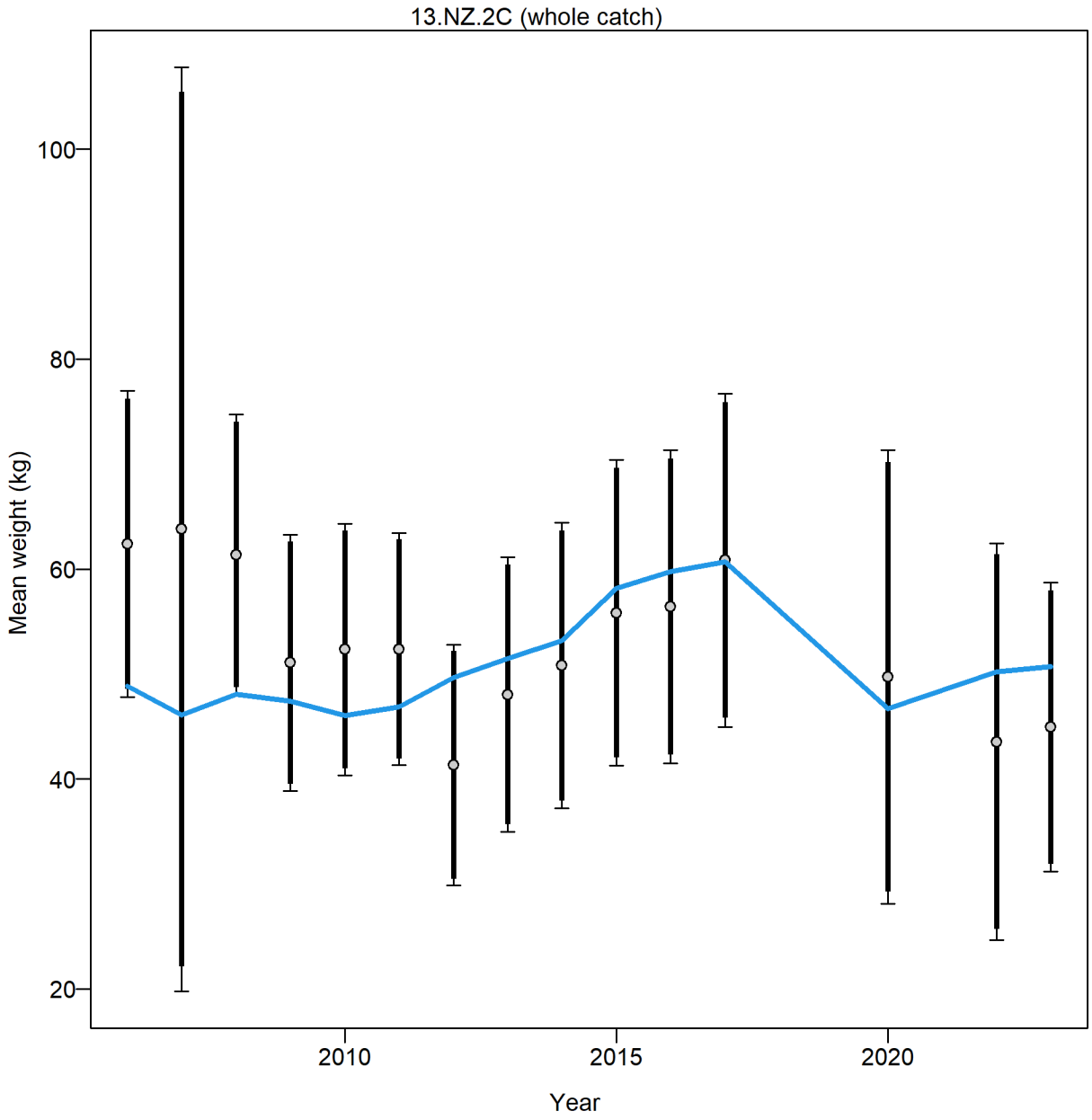


Figure 59: Annually aggregated observed (grey dots) and predicted (blue line) median fish weights (kg) for fishery 13.NZ.2C.

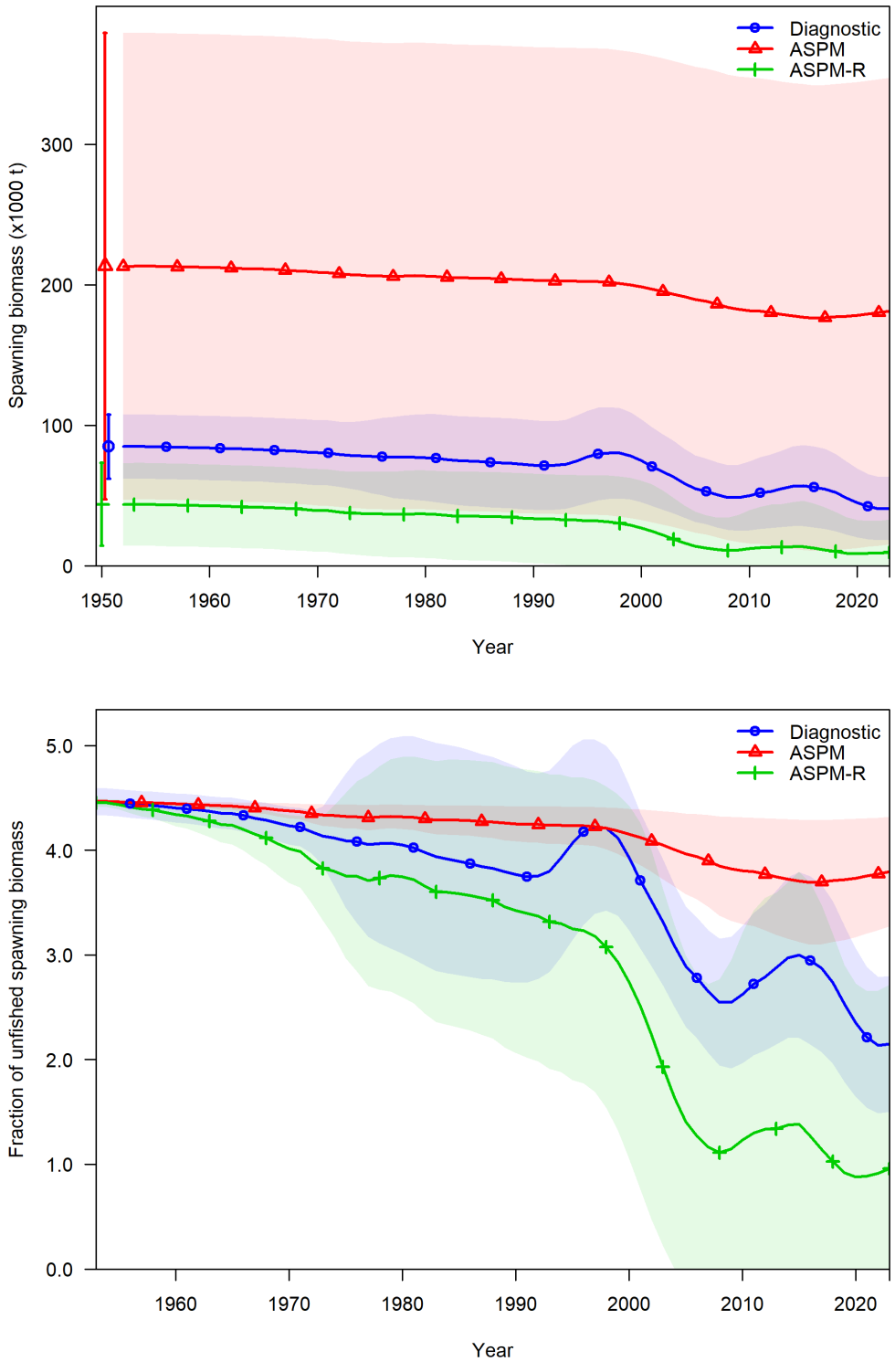


Figure 60: Estimated female spawning biomass, in 1000s of t (top), and relative spawning biomass (bottom), both with confidence bounds, for the diagnostic model (blue), the ASPM (red) and the ASPM-R (green).

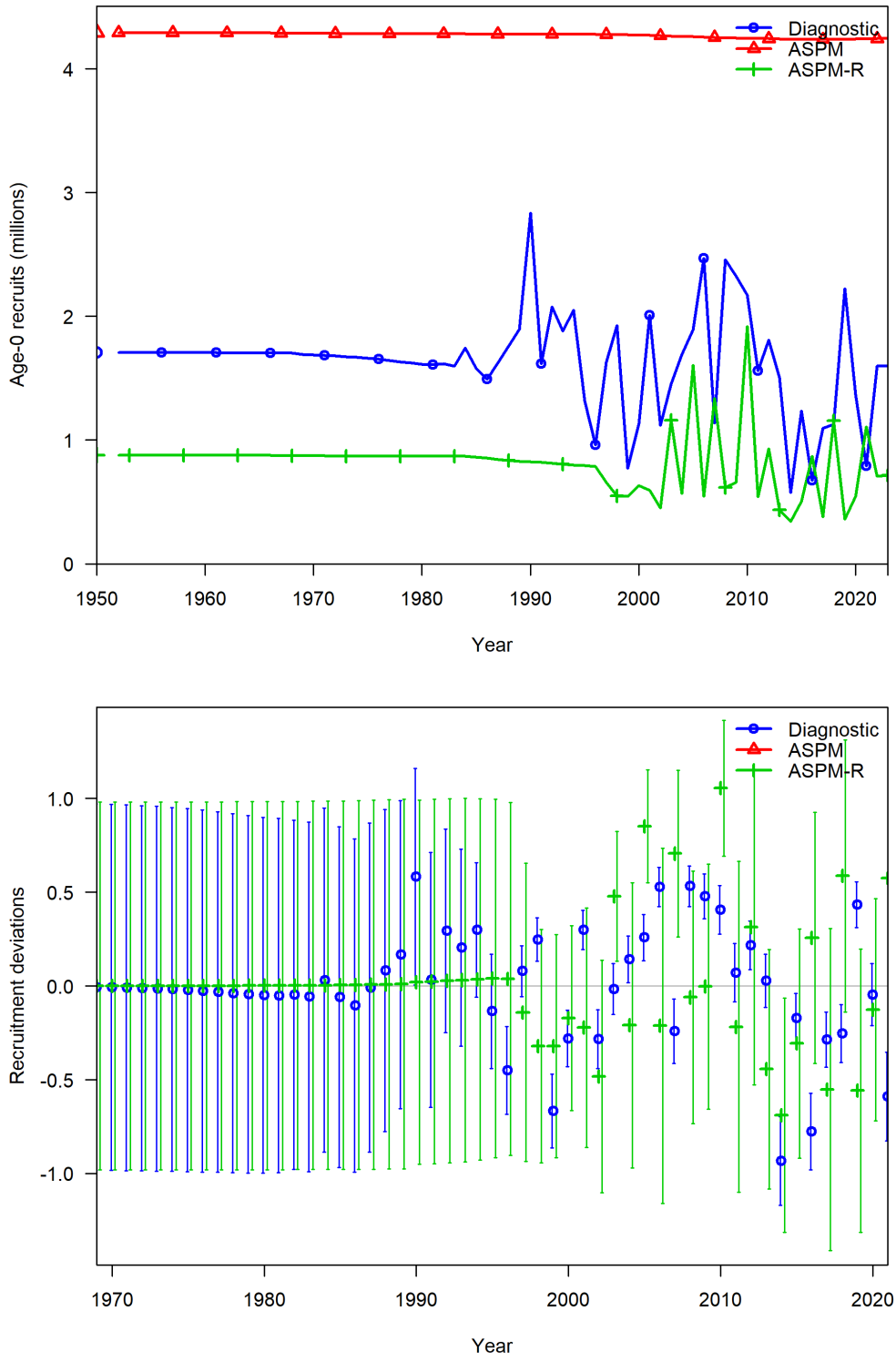


Figure 61: Estimated absolute recruitment (top), and recruitment deviations, with confidence bounds (bottom), for the diagnostic model (blue), the ASPM (red) and the ASPM-R (green).

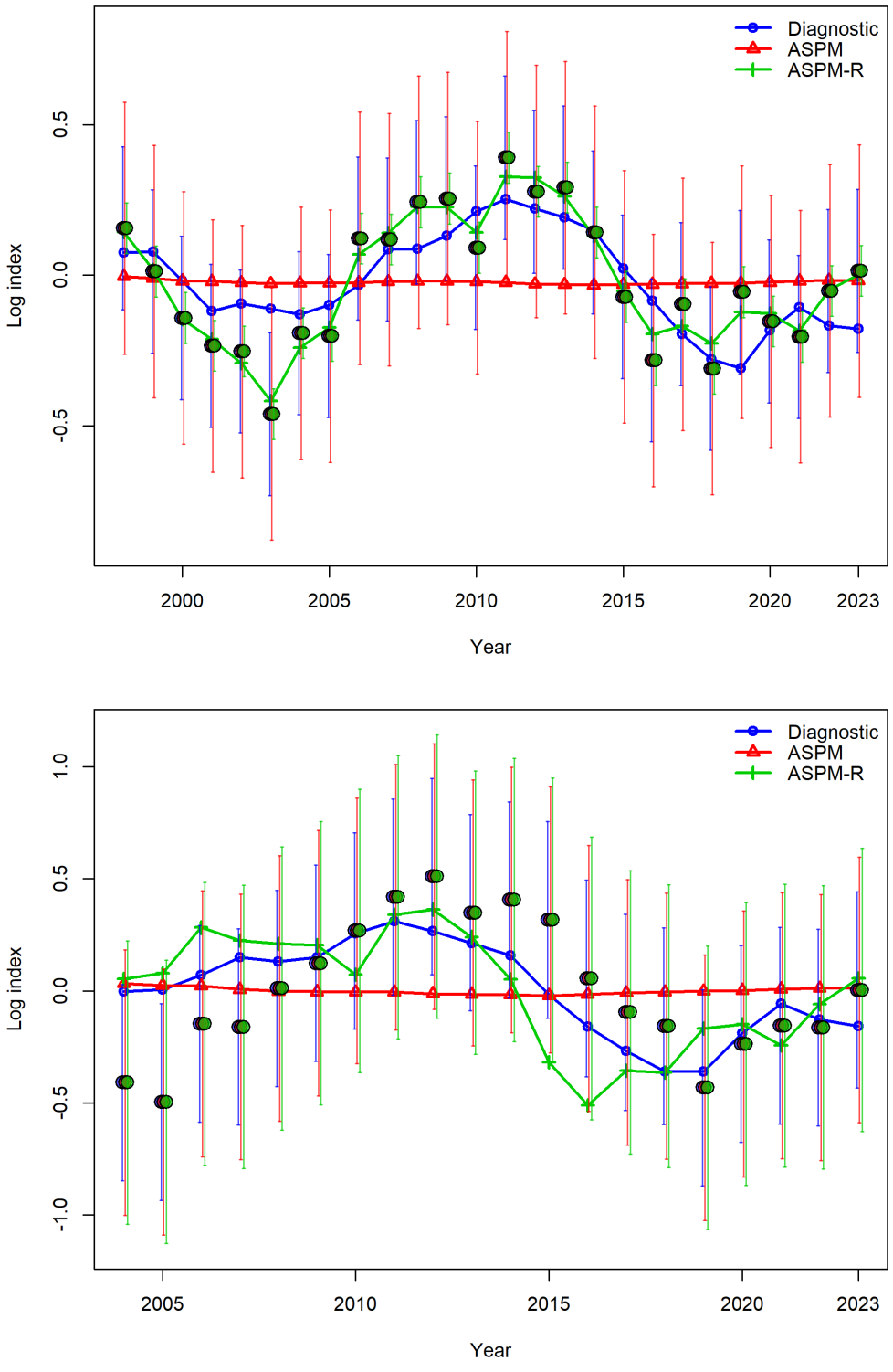


Figure 62: Fits to the Australian CPUE series (top), and the New Zealand CPUE series (bottom), for the diagnostic model (blue), the ASPM (red) and the ASPM-R (green).

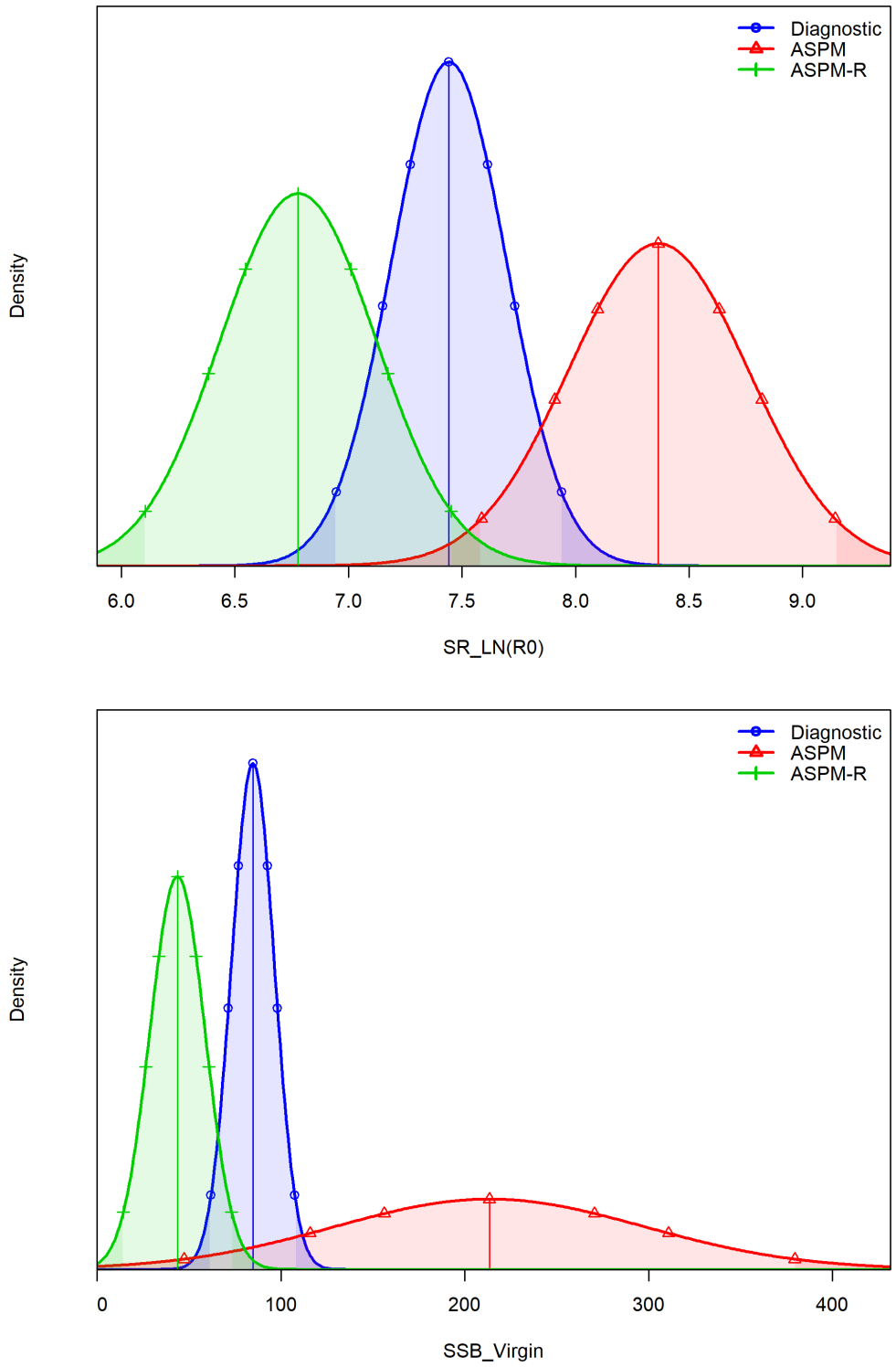


Figure 63: Estimates of $\log R_0$ (top), and initial female spawning biomass, in 1000s of t (bottom), both with asymptotic uncertainty distributions, and for the diagnostic model (blue), the ASPM (red) and the ASPM-R (green).

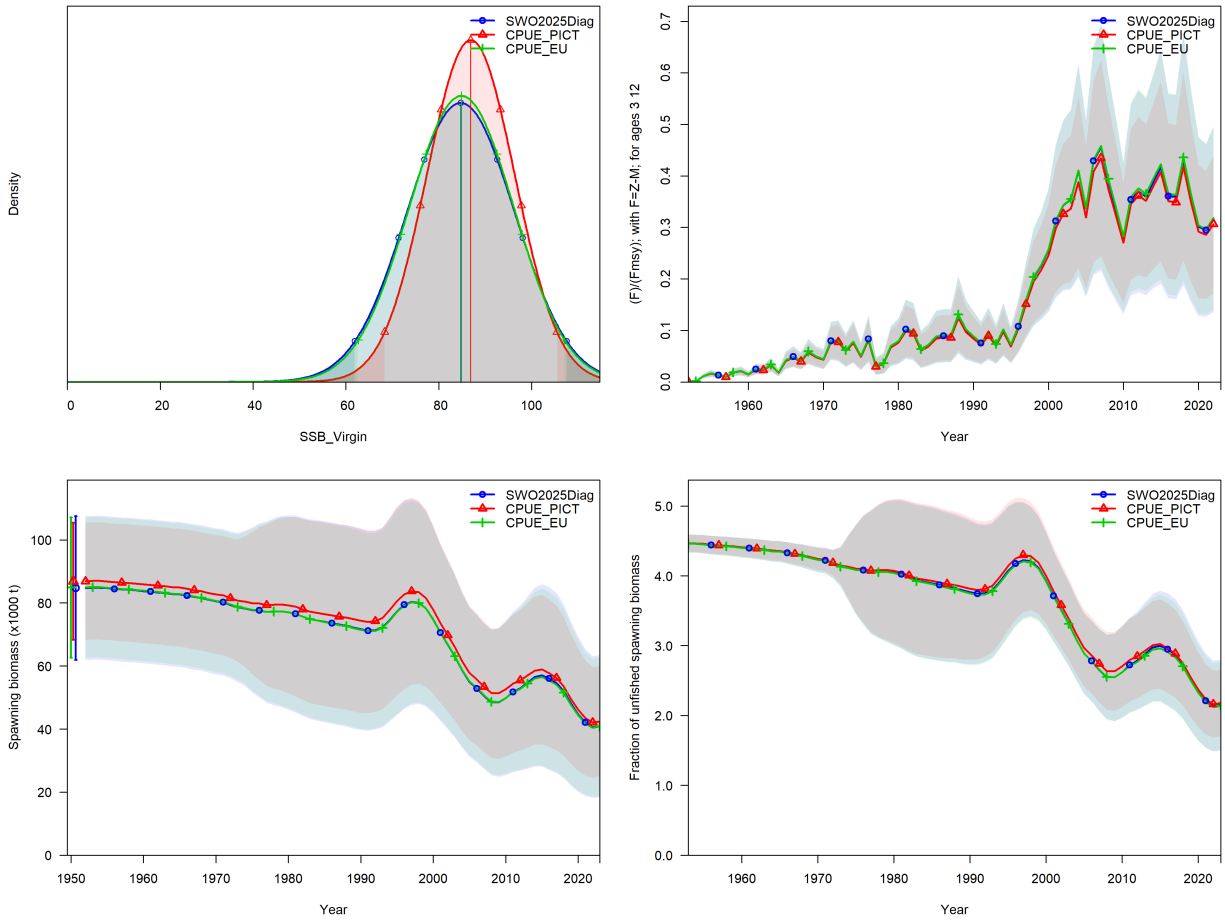


Figure 64: Comparisons between the diagnostic model with indices from Australia in region 1 and New Zealand in region 2 (blue) and the sensitivities to swap the region 2 index from NZ to the PICT index (red) and to swap the region 2 index from NZ to the EU index (green), with asymptotic uncertainty bounds: the estimated initial female spawning biomass, in 1000s of t (top left); the time series of F/F_{MSY} (top right); the female spawning biomass time series (bottom left); and the relative female spawning biomass time series, relative to SB_0 (bottom right).

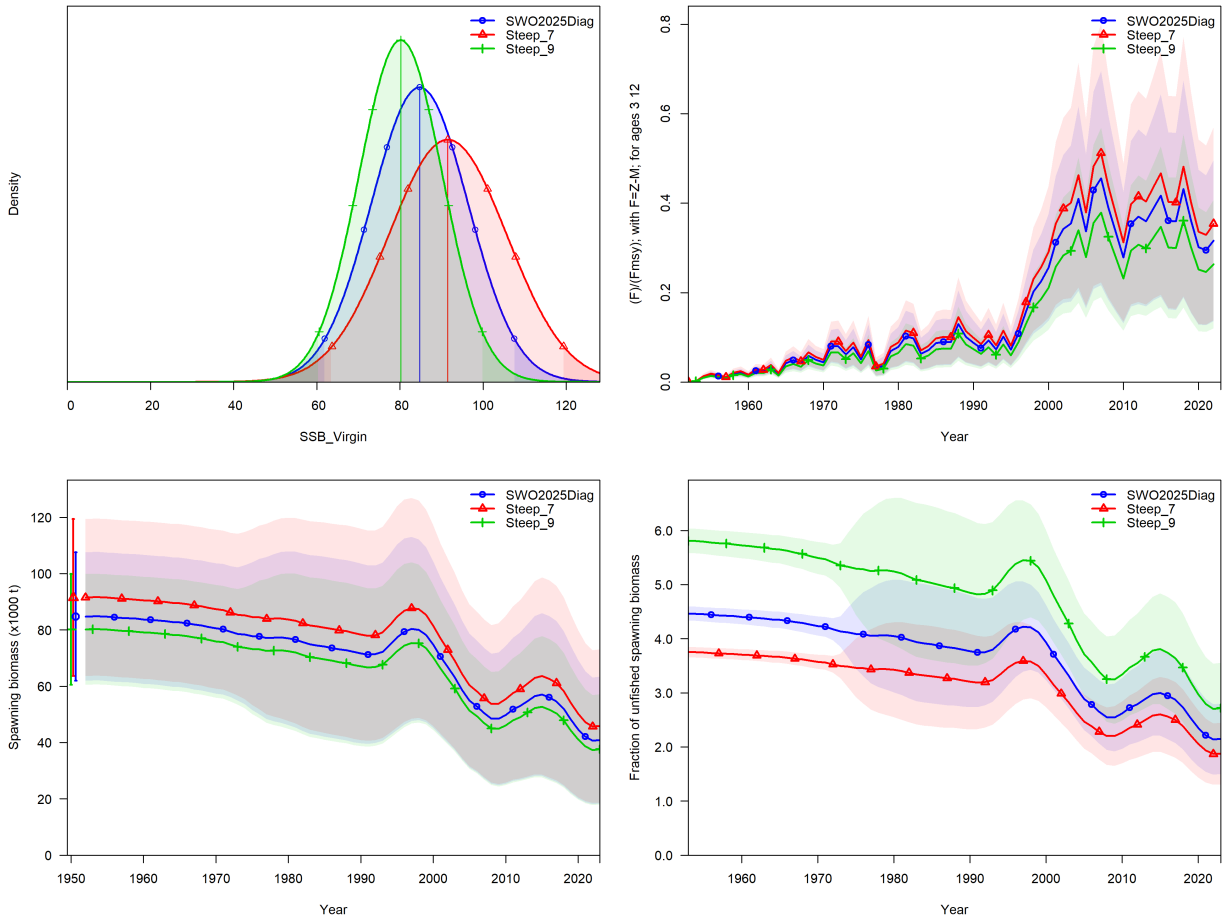


Figure 65: Comparisons between the diagnostic model with $h = 0.8$ (blue) and the sensitivities to steepness with $h = 0.7$ (red) and $h = 0.9$ (green), with asymptotic uncertainty bounds: the estimated initial female spawning biomass, in 1000s of t (top left); the time series of F/F_{MSY} (top right); and the female spawning biomass time series (bottom left); and the relative female spawning biomass time series, relative to SB_0 (bottom right), although this last plot is less informative with varying values for h, as the starting point is not scaled to roughly the same value.

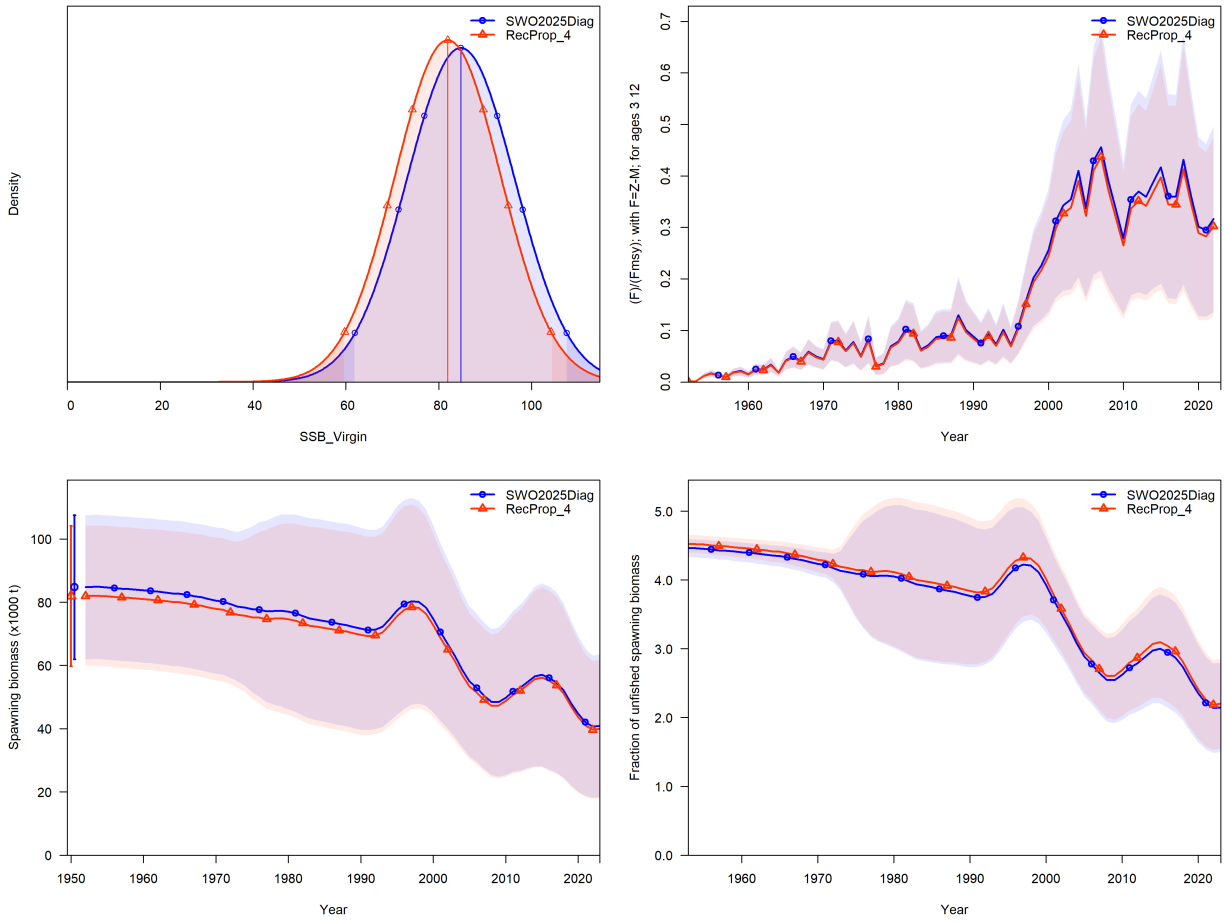


Figure 66: Comparisons between the diagnostic model with recruitment proportion 1:3 (blue) and the sensitivity with recruitment proportion 1:4 (red), with asymptotic uncertainty bounds: the estimated initial female spawning biomass, in 1000s of t (top left); the time series of F/F_{MSY} (top right); and the female spawning biomass time series (bottom left); and the relative female spawning biomass time series, relative to SB_0 (bottom right).

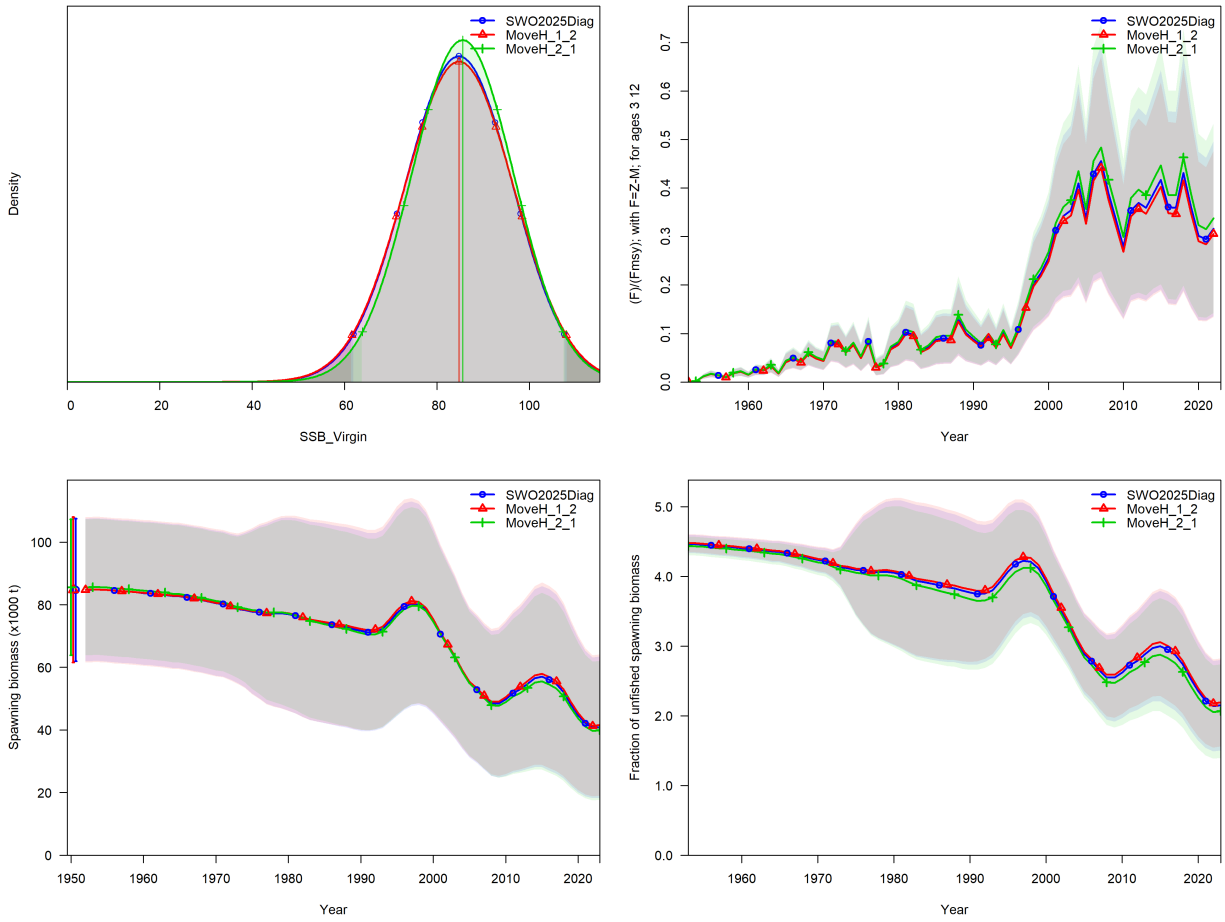


Figure 67: Comparisons between the diagnostic model with (blue) and the sensitivities halving the movement rate from region one to region two (red) and halving the movement rate from region one to region two (green), with asymptotic uncertainty bounds: the estimated initial female spawning biomass, in 1000s of t (top left); the time series of F/F_{MSY} (top right); and the female spawning biomass time series (bottom left); and the relative female spawning biomass time series, relative to SB_0 (bottom right).

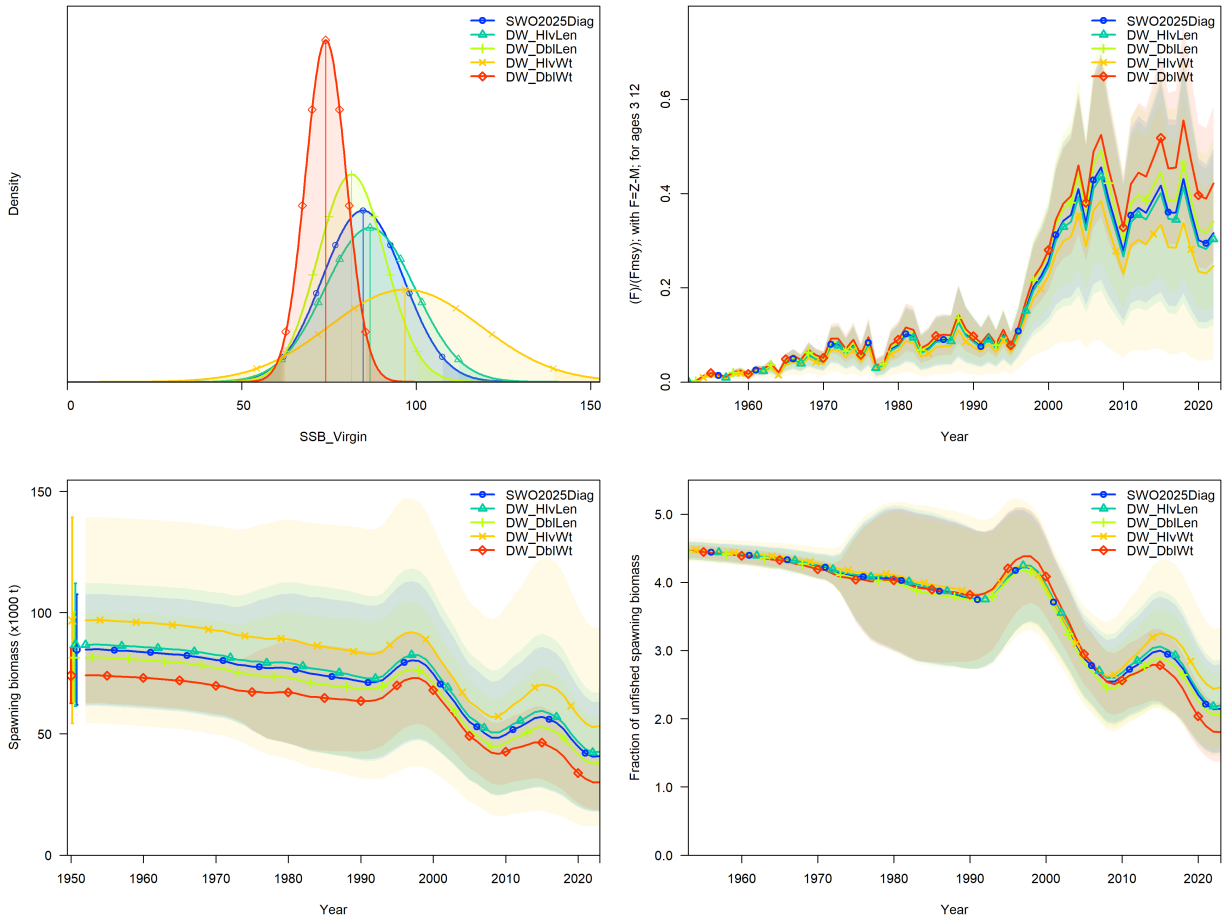


Figure 68: Comparisons between the diagnostic model with (blue) and the sensitivities with a range of data weighting options, halving the weighting on the length composition data (green), doubling the weighting on the length composition data (light green), halving the weighting on the weight composition data (orange), doubling the weighting on the weight composition data (red), with asymptotic uncertainty bounds: the estimated initial female spawning biomass, in 1000s of t (top left); the time series of F/F_{MSY} (top right); and the female spawning biomass time series (bottom left); and the relative female spawning biomass time series, relative to SB_0 (bottom right).

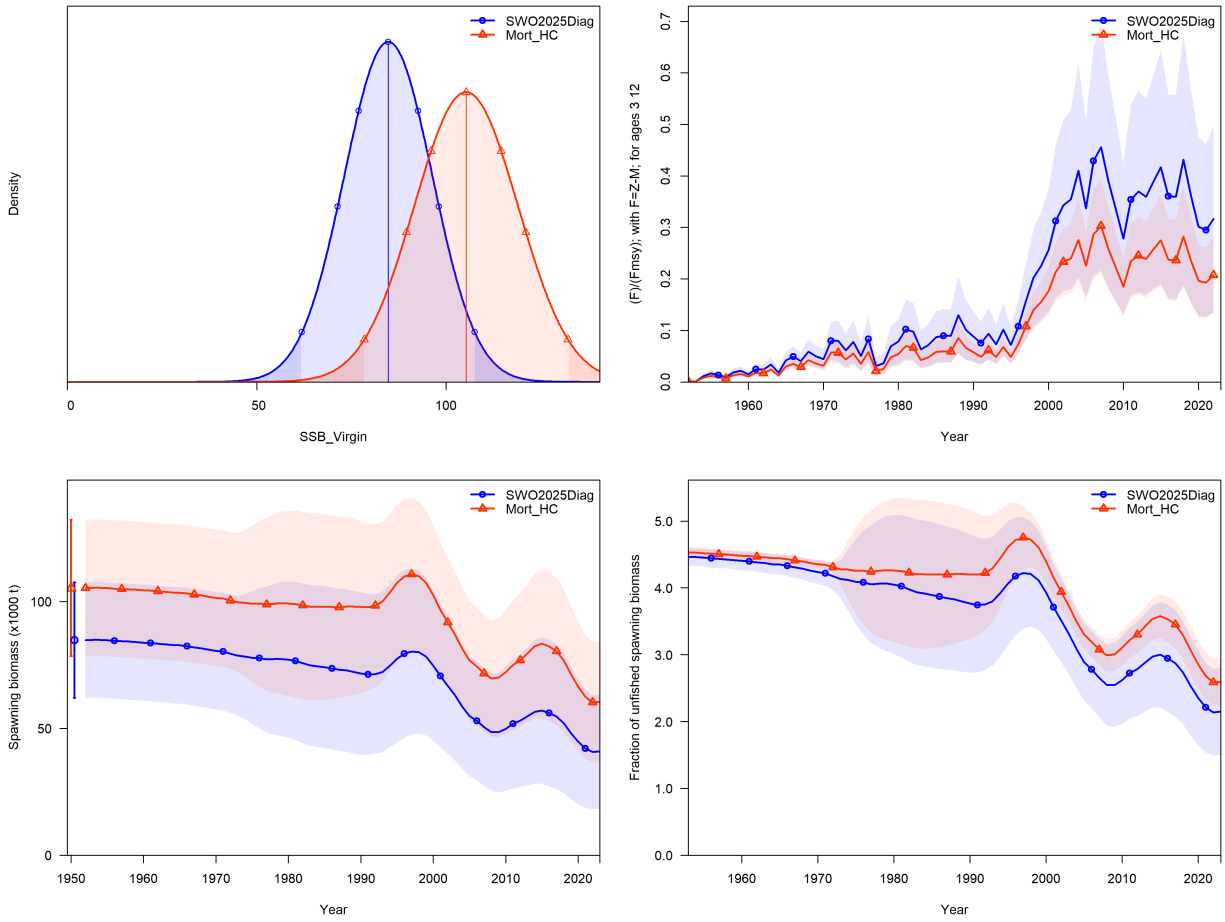


Figure 69: Comparisons between the diagnostic model with the scale of the Lorenzen mortality parameter estimated (blue) and the fixed mortality estimate, using the Hamel and Cope (2022) approach (red), with asymptotic uncertainty bounds: the estimated initial female spawning biomass, in 1000s of t (top left); the time series of F/F_{MSY} (top right); and the female spawning biomass time series (bottom left); and the relative female spawning biomass time series, relative to SB_0 (bottom right).

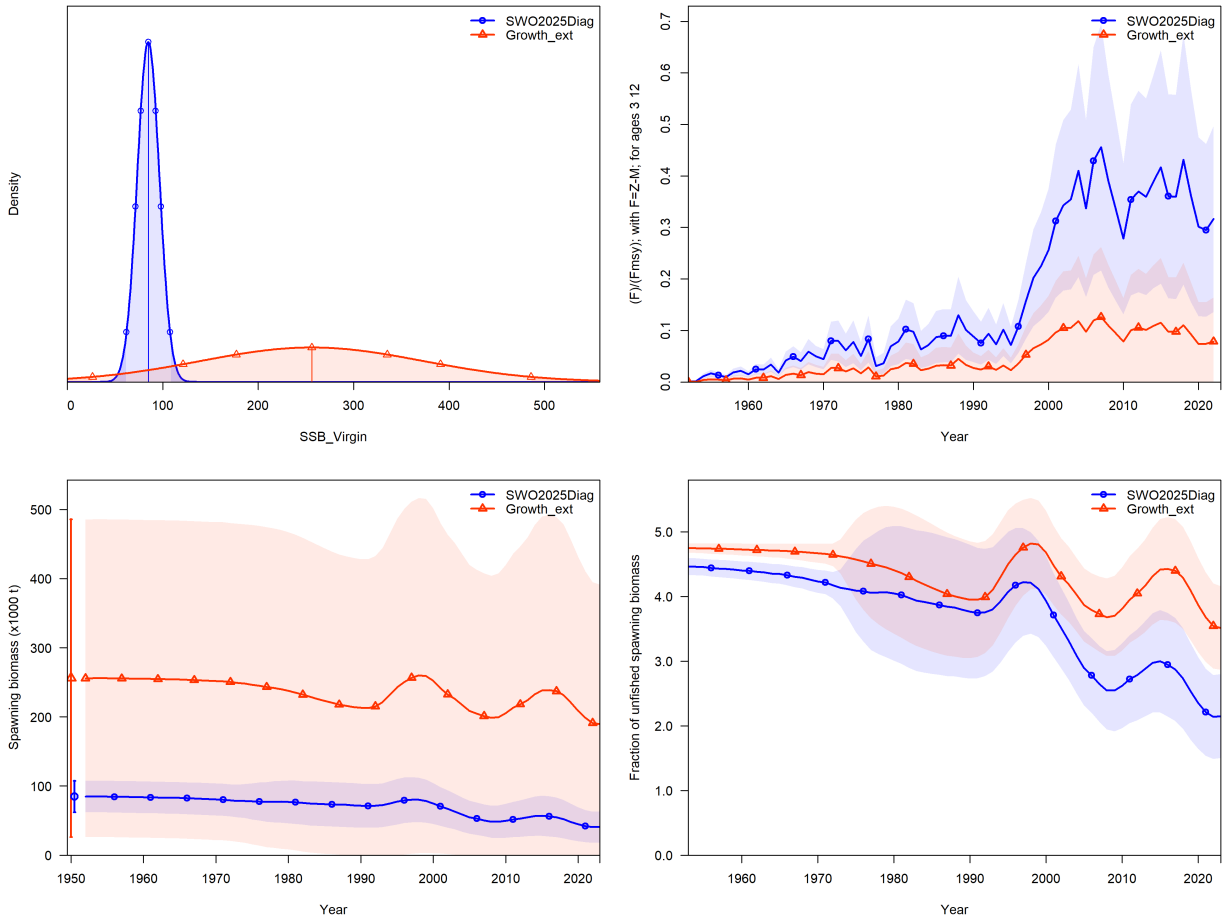


Figure 70: Comparisons between the diagnostic model (blue) and the sensitivity to fixed external growth (red), with asymptotic uncertainty bounds: the estimated initial female spawning biomass, in 1000s of t (top left); the time series of F/F_{MSY} (top right); and the female spawning biomass time series (bottom left); and the relative female spawning biomass time series, relative to SB_0 (bottom right).

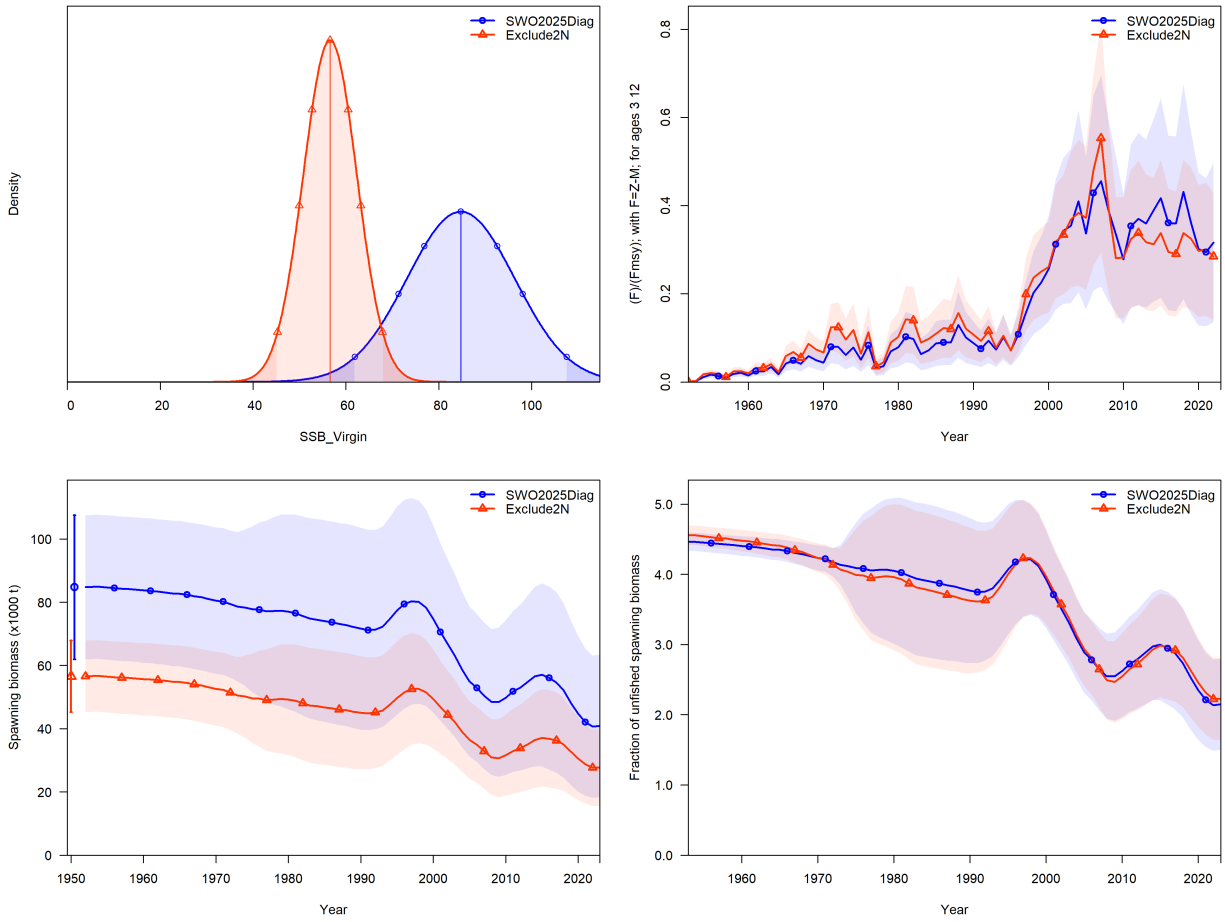


Figure 71: Comparisons between the diagnostic model (blue) and the sensitivity to excluding subregion 2N (red), with asymptotic uncertainty bounds: the estimated initial female spawning biomass, in 1000s of t (top left); the time series of F/F_{MSY} (top right); the female spawning biomass time series (bottom left); and the relative female spawning biomass time series, relative to SB_0 (bottom right).

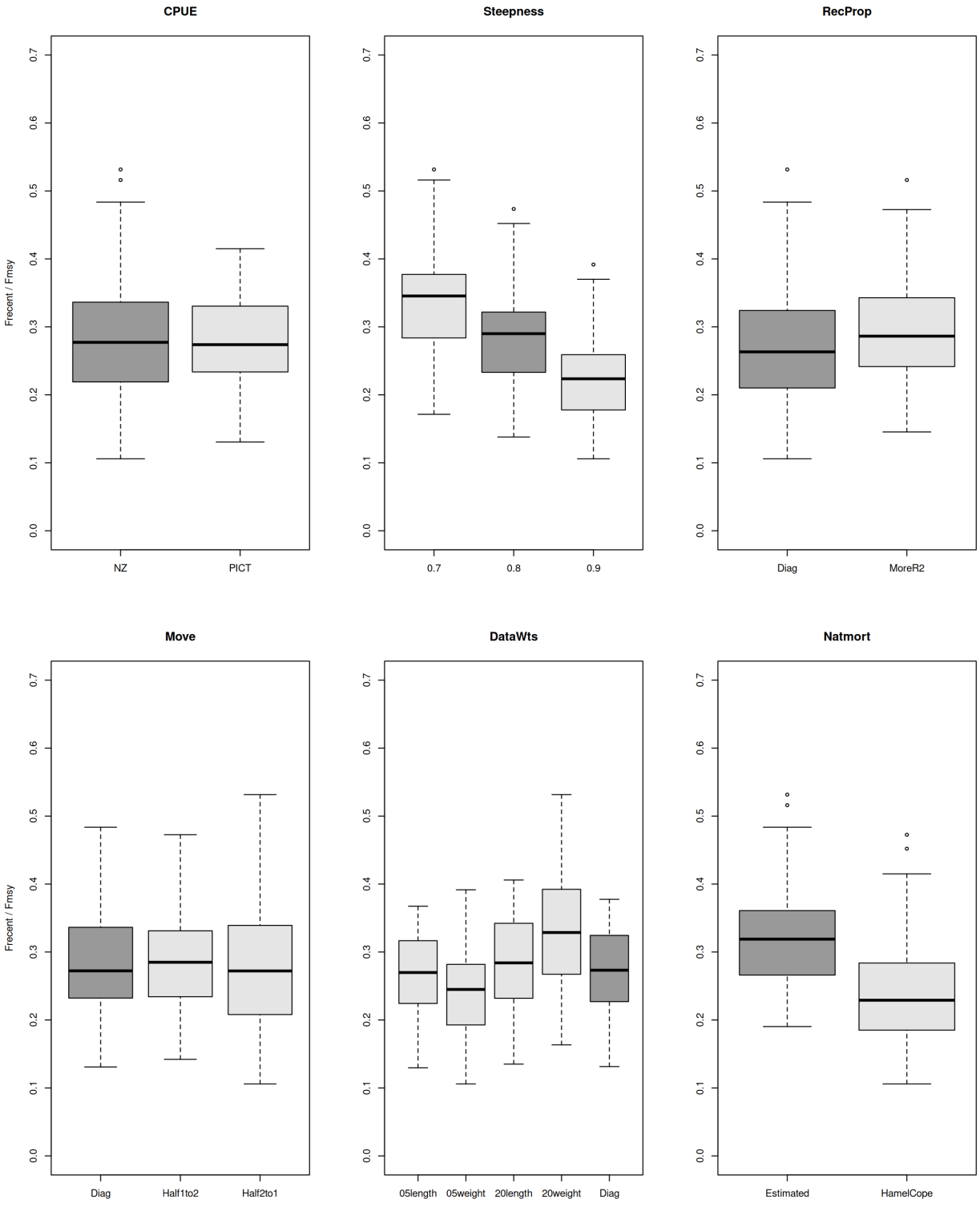


Figure 72: Standard box and whiskers plots of F/F_{MSY} by grid axis for all 360 models in the uncertainty grid, with each subplot separated into the components of each grid axis, with the diagnostic model highlighted in dark grey in each subplot. This plot only incorporates model uncertainty, without considering estimation uncertainty. The most influential axes of the grid are steepness, mortality and data weighting, particularly for the weight composition data.

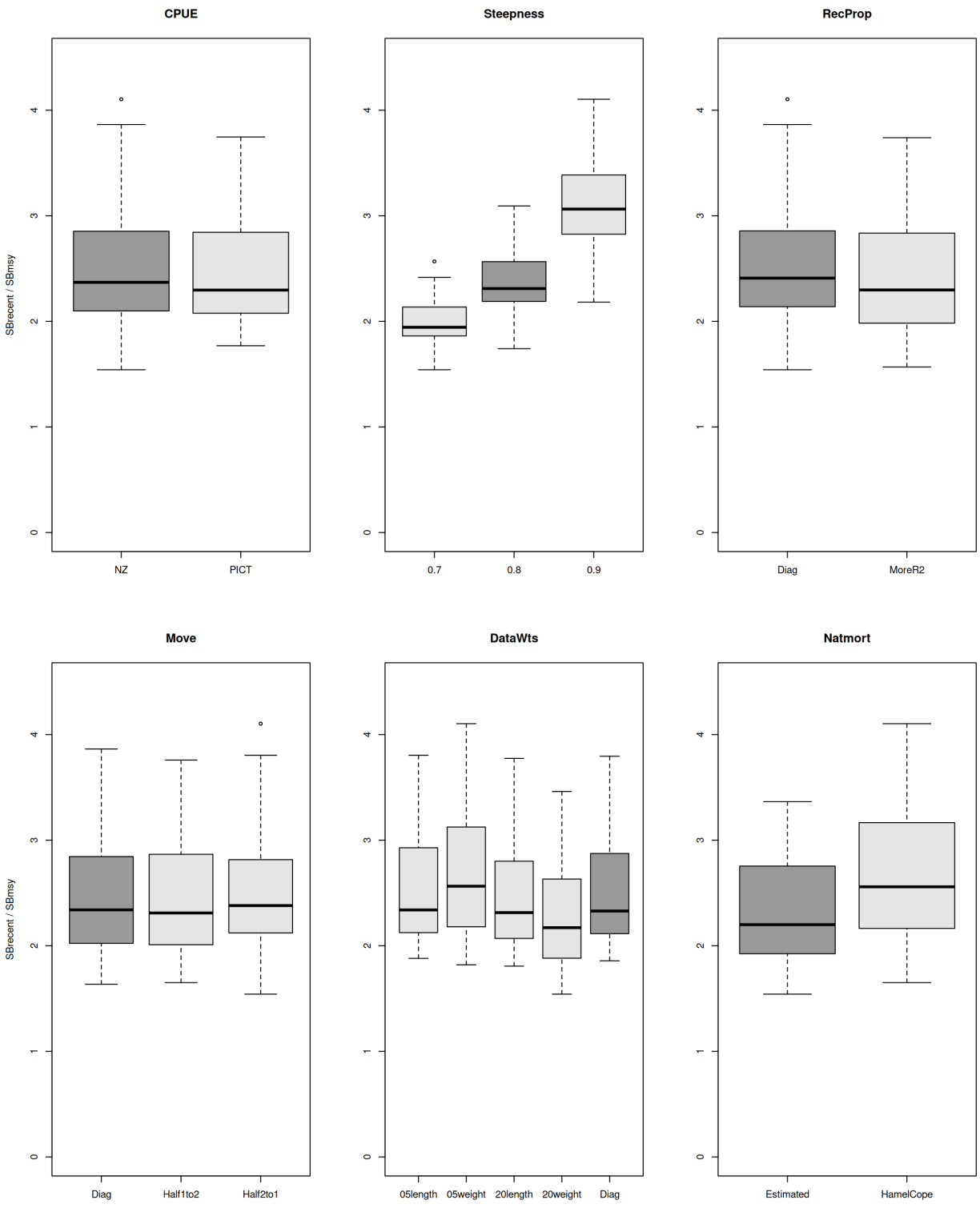


Figure 73: Standard box and whiskers plots of $SB_{\text{recent}}/SB_{\text{MSY}}$ by grid axis for all 360 models in the uncertainty grid, with each subplot separated into the components of each grid axis, with the diagnostic model highlighted in dark grey in each subplot. This plot only incorporates model uncertainty, without considering estimation uncertainty. The most influential axes of the grid are steepness, mortality and data weighting, particularly for the weight composition data.

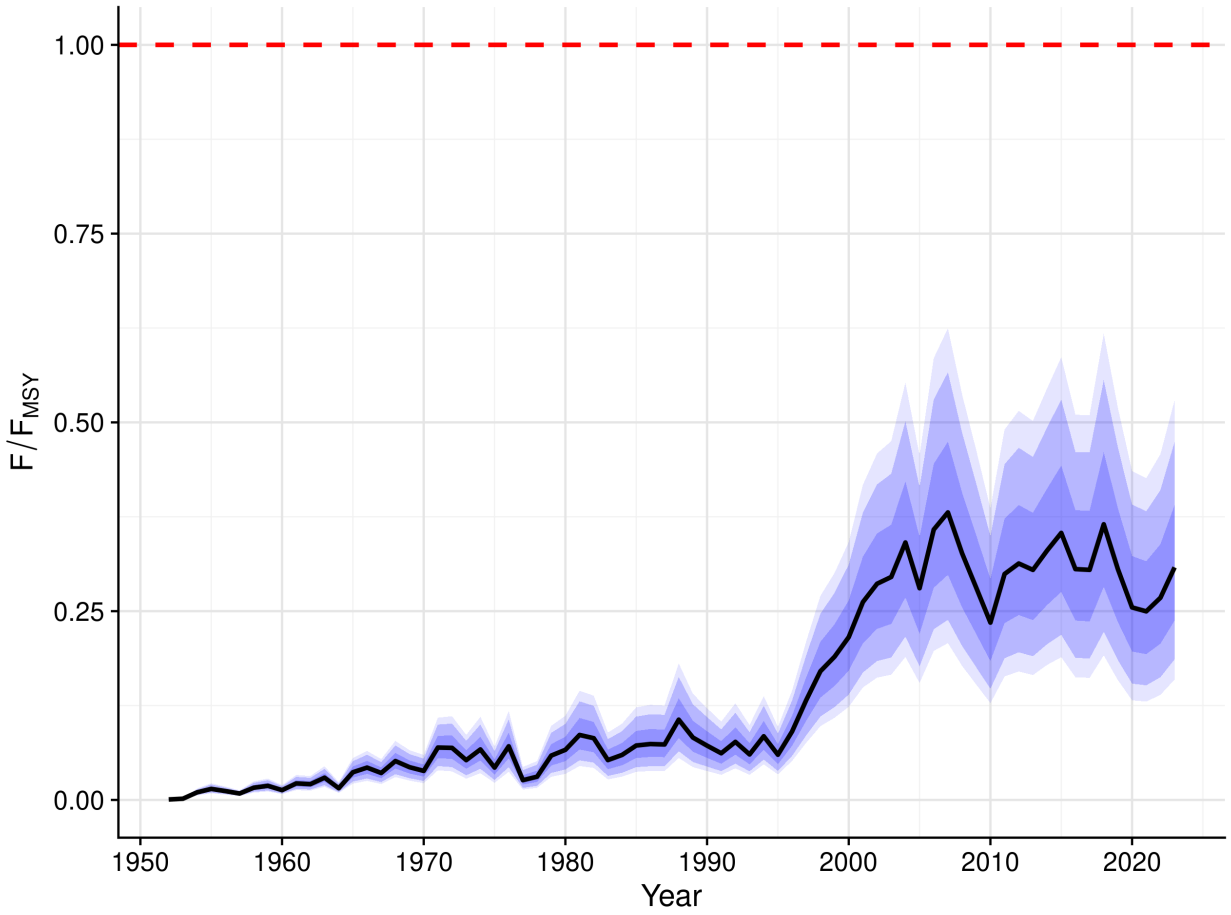


Figure 74: Ribbon plots from the uncertainty grid with estimation error, showing the median of all trajectories, along with 50%, 80% and 90% quantile ranges: F/F_{MSY}.

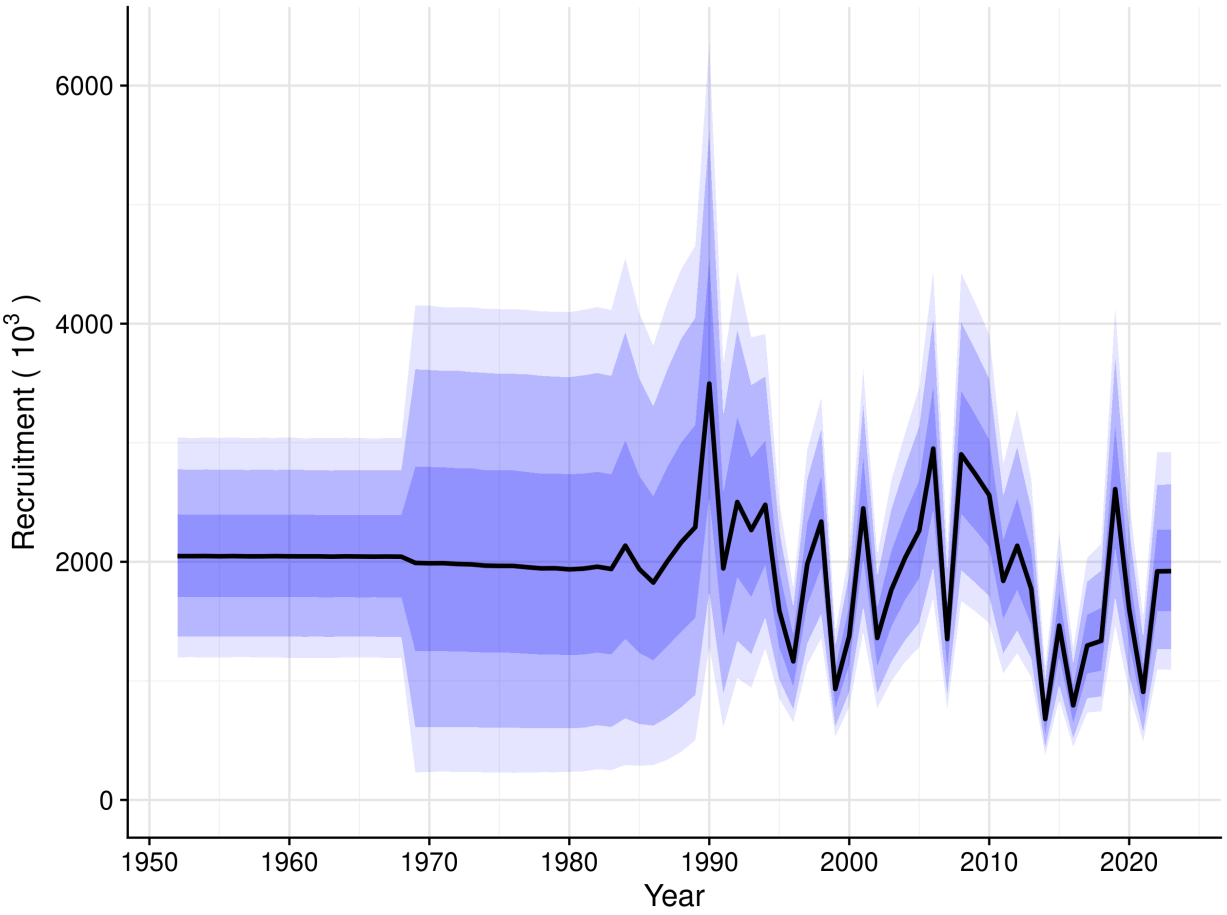


Figure 75: Ribbon plots from the uncertainty grid with estimation error, showing the median of all trajectories, along with 50%, 80% and 90% quantile ranges: Recruitment (in 1000s).

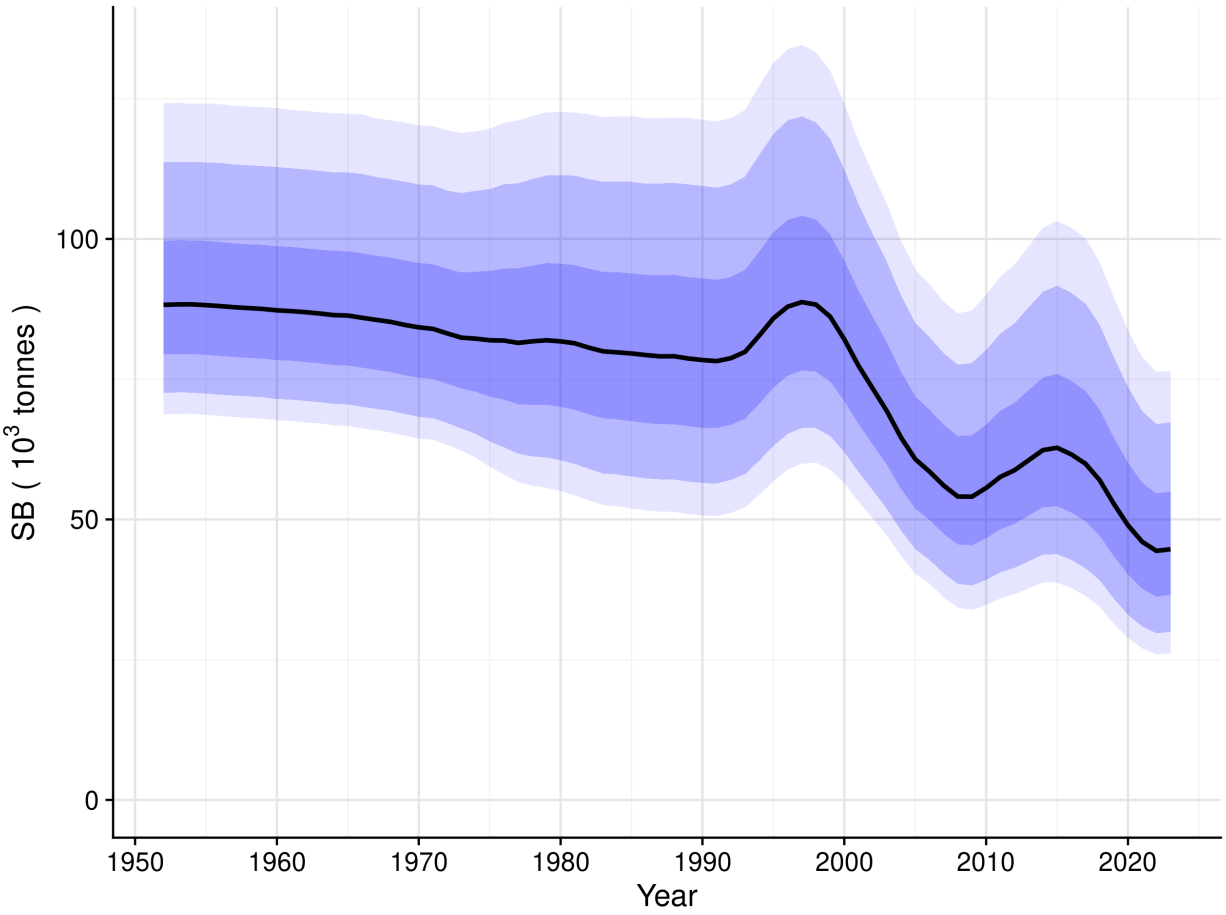


Figure 76: Ribbon plots from the uncertainty grid with estimation error, showing the median of all trajectories, along with 50%, 80% and 90% quantile ranges: *SB* (in 1000s of t).

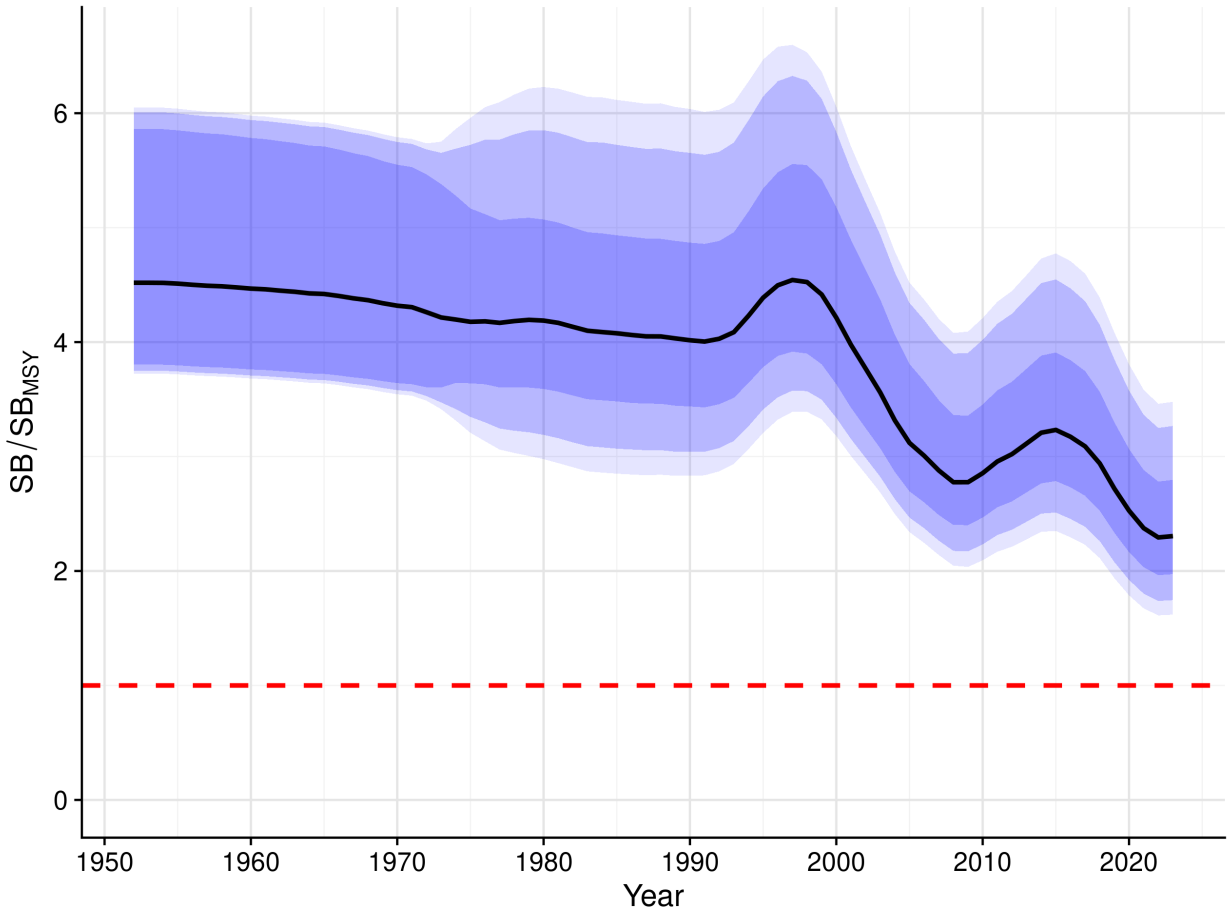


Figure 77: Ribbon plots from the uncertainty grid with estimation error, showing the median of all trajectories, along with 50%, 80% and 90% quantile ranges: $SB_{\text{recent}}/SB_{\text{MSY}}$.

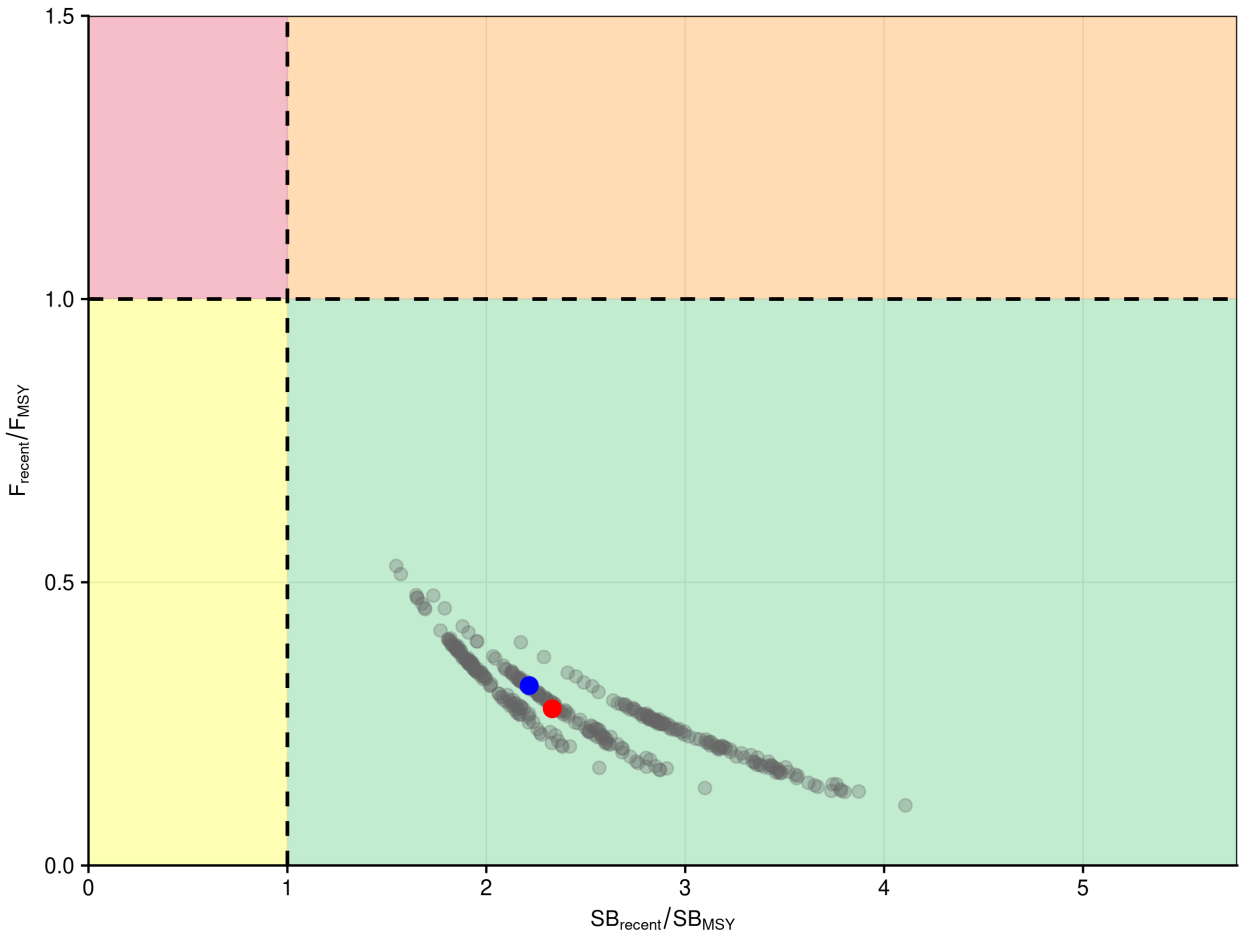


Figure 78: Kobe plot summarising the results for each of the 360 models in the structural uncertainty grid, with no estimation error, for the relevant recent period, which are 2020–2023 for $SB_{\text{recent}}/SB_{\text{MSY}}$ and 2019–2022 for $F_{\text{recent}}/F_{\text{MSY}}$ respectively. The red dot indicates the median and the blue dot the diagnostic model.

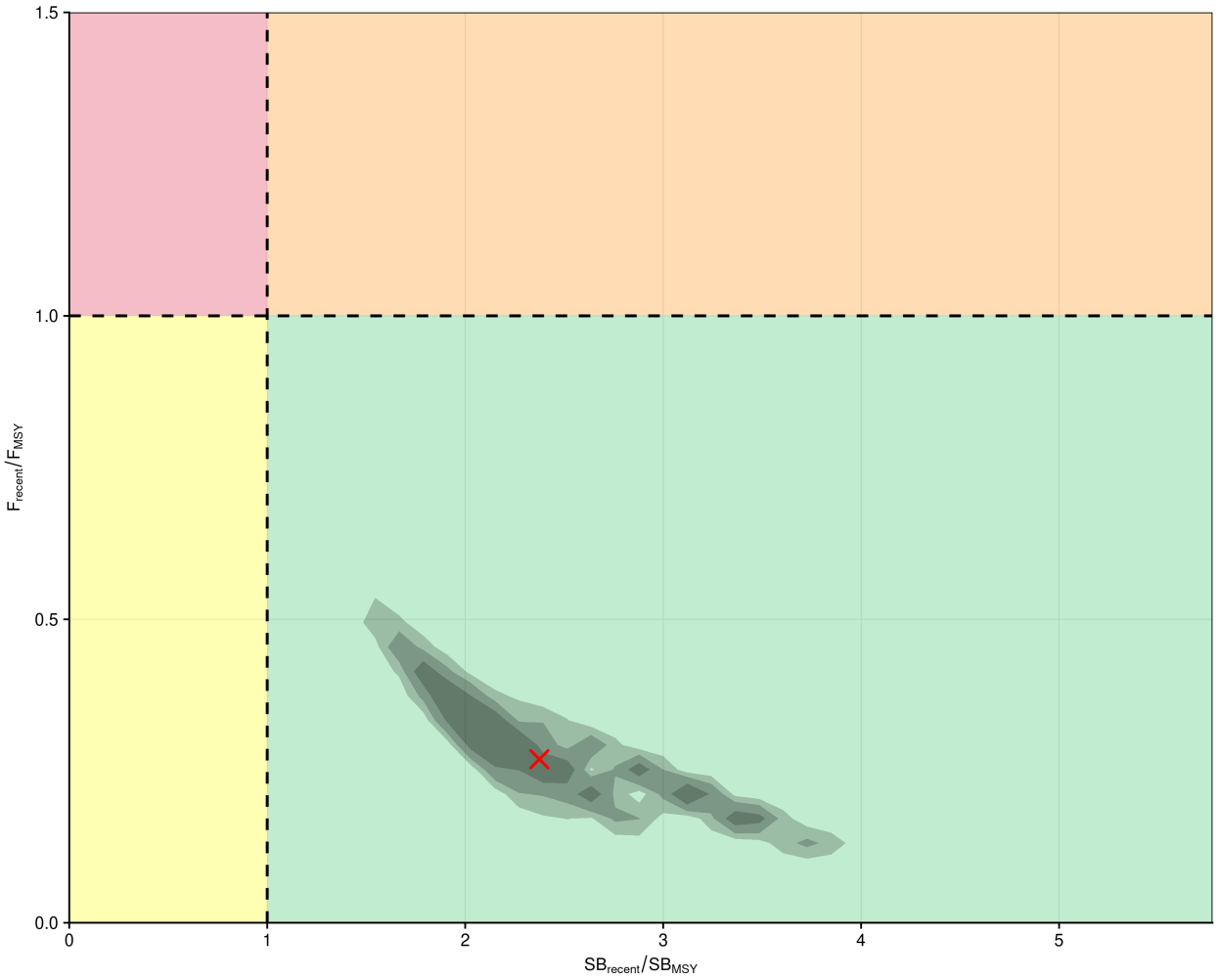


Figure 79: Kobe with contour plots, summarising the full uncertainty grid from 360,000 model runs, including structural and estimation uncertainty, with 50%, 80% and 90% quantile ranges. The red cross shows the median of all models.

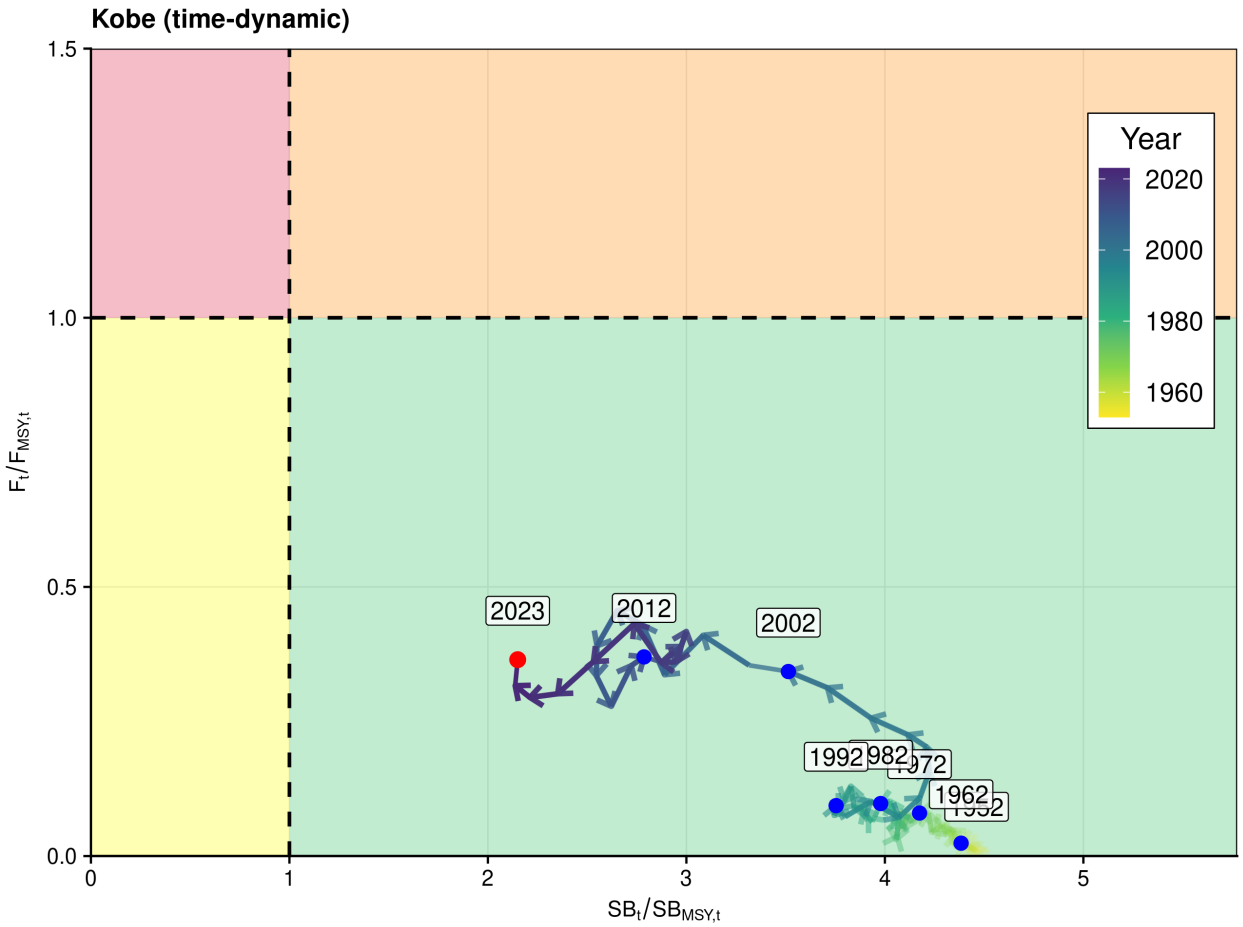


Figure 80: Time dynamic Kobe plot summarising the results for the 2025 diagnostic model over the model period. The larger red point is the estimated 2023 status, with the trajectory through this phase space indicated by labelled years (blue points), arrows and the changing colour of the trajectory line.

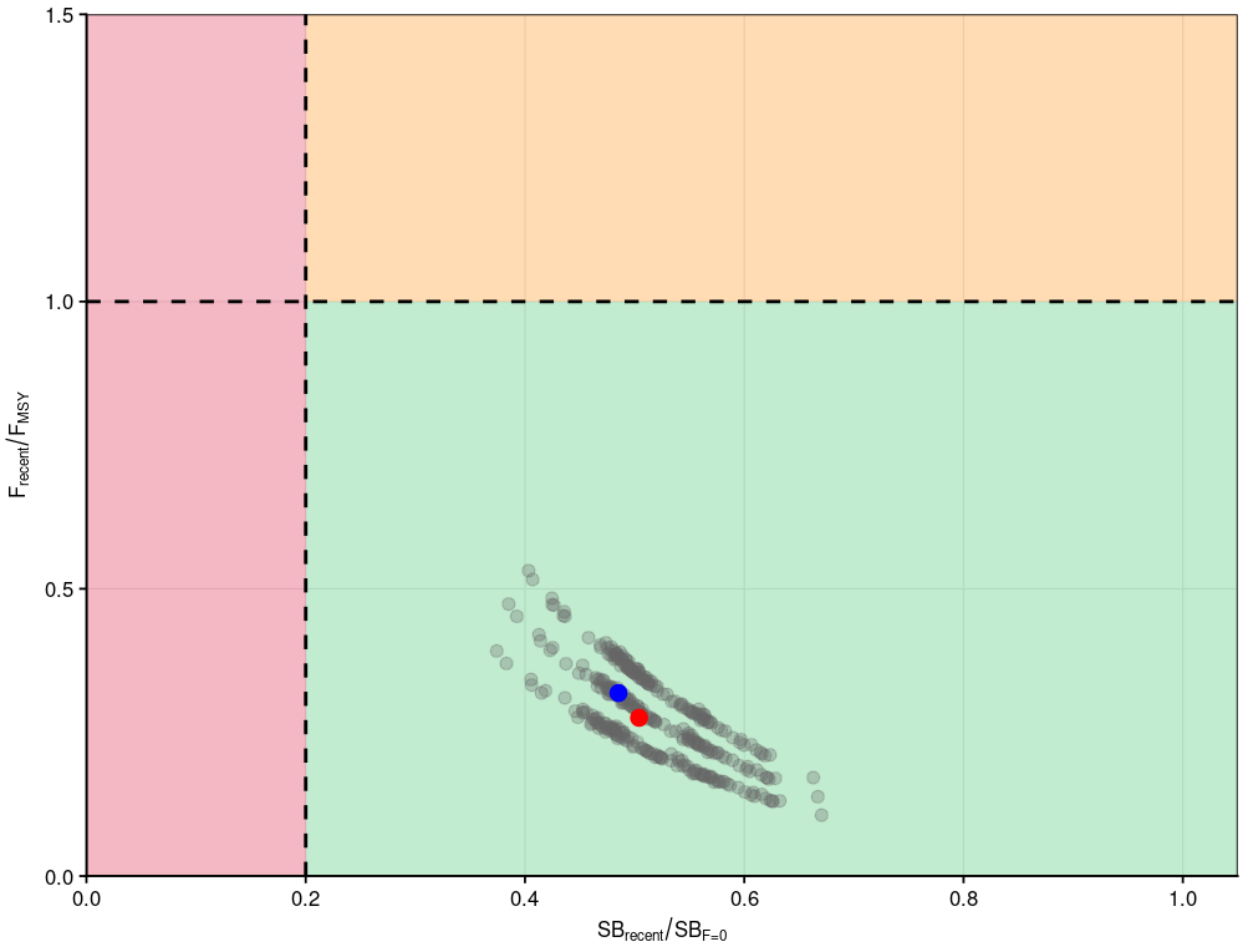


Figure 81: Majuro plot summarising the results for each of the 360 models in the structural uncertainty grid, with no estimation error, for the relevant recent period, which are 2020–2023 for $SB_{\text{recent}}/SB_{F=0}$ and 2019–2022 for $F_{\text{recent}}/F_{\text{MSY}}$ respectively. The red dot indicates the median and the blue dot the diagnostic model.

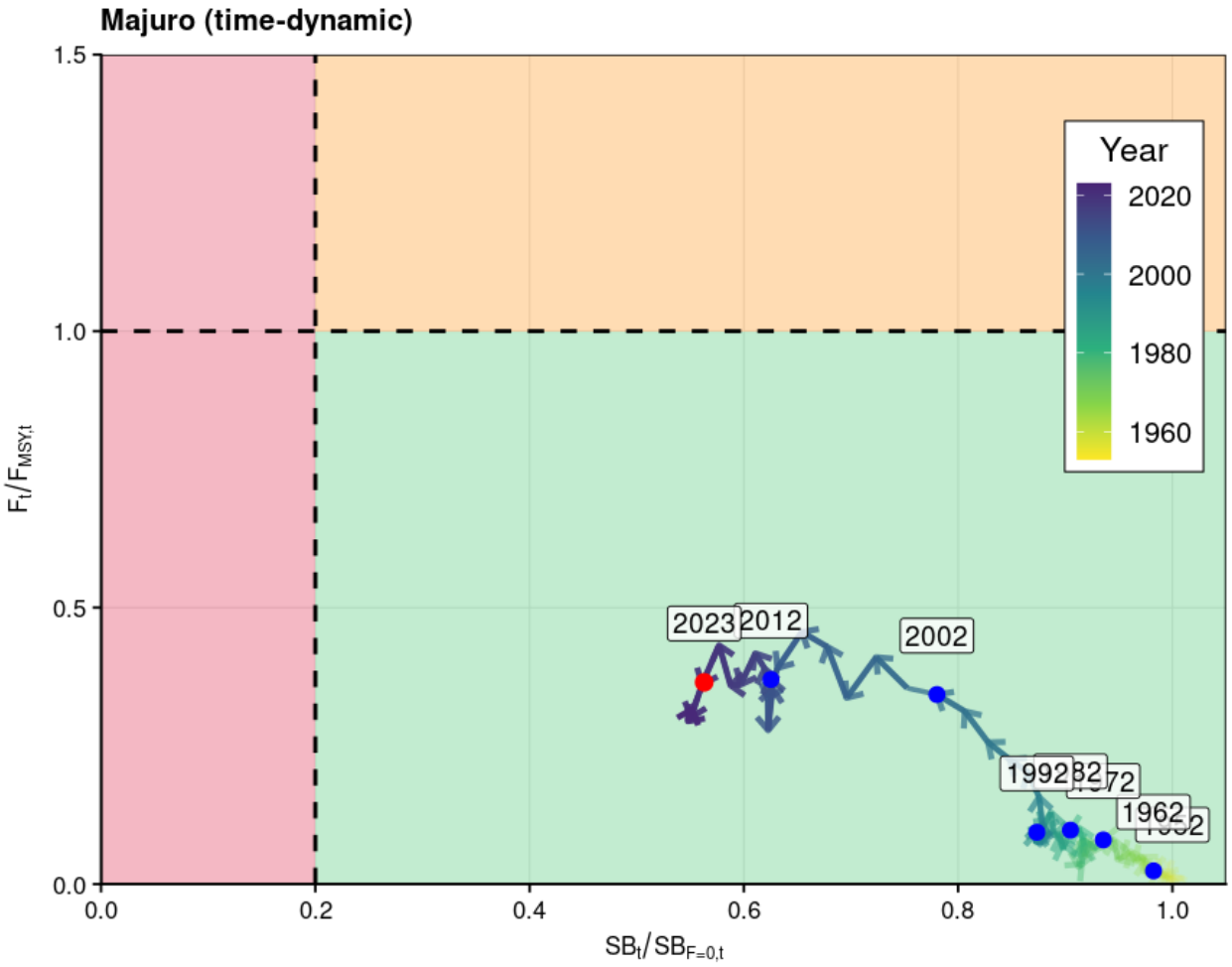


Figure 82: Time dynamic Majuro plot summarising the results for the 2025 diagnostic model over the model period. The larger red point is the estimated 2023 status, with the trajectory through this phase space indicated by labelled years (blue points), arrows and the changing colour of the trajectory line.