



**SCIENTIFIC COMMITTEE
SEVENTEENTH REGULAR SESSION**

Online meeting
11–19 August 2021

Stock assessment of southwest Pacific swordfish

WCPFC-SC17-2021/SA-WP-04

28 July 2021

**N. Ducharme-Barth¹, C. Castillo-Jordán¹, J. Hampton¹, P. Williams¹, G. Pilling¹, P.
Hamer¹**

¹Oceanic Fisheries Programme, Pacific Community (SPC), Nouméa, New Caledonia

Contents

1	Executive Summary	6
2	Acronyms and Abbreviations	8
3	Introduction	10
4	Background	12
4.1	Stock structure and movement	12
4.2	Vertical movement behavior	13
4.3	Biological characteristics	14
4.4	Fisheries	14
5	Data compilation	15
5.1	General notes	15
5.2	Spatial stratification	15
5.3	Temporal stratification	15
5.4	Definition of fisheries	16
5.4.1	Extraction fisheries	16
5.4.2	Index fisheries	16
5.5	Catch and effort data	18
5.5.1	General characteristics	18
5.6	Size data	20
5.7	Tagging data	20
6	Model description	21
6.1	General characteristics	21
6.1.1	Uncertainty characterization	21
6.2	Population dynamics	22
6.2.1	Recruitment	22
6.2.2	Initial population	23
6.2.3	Growth	23
6.2.4	Movement	24
6.2.5	Natural mortality	25
6.2.6	Reproductive potential	25
6.2.7	Length-weight relationship	26
6.2.8	Sex structure	26
6.3	Fishery dynamics	26
6.3.1	Selectivity	27
6.3.2	Catchability	27

6.3.3	Effort deviations	28
6.4	Likelihood components	28
6.5	Parameter estimation	29
6.6	Diagnostics	29
6.7	Stock assessment interpretation methods	30
6.7.1	Depletion and fishery impact	30
6.7.2	MSY based reference points	31
6.7.3	Majuro and Kobe plots	31
6.7.4	Reference points	32
6.7.5	Ensemble influence	32
7	Model runs	32
7.1	Developments from the last assessment	33
7.2	Sensitivity analyses	34
7.2.1	Sex-structure	34
7.2.2	Alternative catch scenarios	35
7.2.3	Alternative spatial structure	35
7.2.4	Penalty on stock-recruit relationship	36
7.2.5	DWFN CPUE scenario	36
7.2.6	CPUE average CV	36
7.2.7	Size composition scalar	37
7.2.8	Movement	37
7.2.9	Steepness	37
7.2.10	Natural mortality	37
7.2.11	Growth	37
7.2.12	Length-weight relationship	38
7.2.13	Reproductive potential	38
7.3	Model ensemble	38
8	Results	39
8.1	Consequences of key model developments	39
8.2	Diagnostic case: Model fits	41
8.2.1	Catch data	42
8.2.2	Standardized CPUE	42
8.2.3	Weight frequency data	42
8.2.4	Length frequency data	42
8.3	Diagnostic case: Estimated quantities	43
8.3.1	Selectivity	43
8.3.2	Biomass	43
8.3.3	Recruitment	43

8.3.4	Fishing mortality	44
8.4	Diagnostic case: Further analyses of stock status	44
8.4.1	Fishery impacts	44
8.4.2	Dynamic Majuro and Kobe plots	45
8.5	Multimodel inference: Sensitivities	45
8.5.1	Sex-structure	45
8.5.2	Alternative catch scenarios	46
8.5.3	Alternative spatial structure	46
8.5.4	Penalty on stock-recruit relationship	46
8.5.5	DWFN CPUE scenario	47
8.5.6	CPUE average CV	47
8.5.7	Size composition scalar	47
8.5.8	Movement	48
8.5.9	Steepness	48
8.5.10	Natural mortality	49
8.5.11	Growth	49
8.5.12	Length-weight relationship	49
8.5.13	Reproductive potential	49
8.6	Multimodel inference: Model ensemble	50
8.6.1	Ensemble influence	50
8.6.2	Model retention	51
8.6.3	Uncertainty characterization: Model & Estimation uncertainty	52
9	Discussion and conclusions	53
9.1	General remarks	53
9.2	Improvements to the assessment	54
9.3	Challenges	55
9.4	Stock assessment modeling considerations	59
9.5	Main assessment conclusions	60
9.6	Summary of research recommendations	61
10	Tables	70
11	Figures	73
12	Appendix	128
12.1	Diagnostic Case	128
12.1.1	Likelihood profiles	128
12.1.2	Retrospective analyses	131
12.1.3	Age-structured production model	134
12.1.4	Hindcasting	136

12.1.5 Residual plots	138
12.2 Ensemble weighting	145
12.3 Sex-disaggregated model ensemble	150

1 Executive Summary

This paper describes the 2021 stock assessment of swordfish (*Xiphias gladius*) in the southwest Pacific Ocean (SWPO) using a sex-aggregated model structure in MULTIFAN-CL. A further four years of data were available since the last full stock assessment conducted in 2017, and the model extends through to the end of 2019. New developments to the stock assessment include:

- A new approach for developing a model ensemble and characterizing uncertainty in management reference points, capturing both the model and estimation uncertainty based on (Ducharme-Barth and Vincent, 2021).
- Updates to the biological assumptions and definition of reproductive potential.
- Implementation of an “index” fishery approach.

The assessment is supported by the development of a new framework for developing a model ensemble and presenting the uncertainty in reference points (Ducharme-Barth and Vincent, 2021), background analyses of biological parameters, preparation of the length-weight composition data and definition of the fisheries structures (Ducharme-Barth et al., 2021b), re-analysis of tagging data to create a movement prior for the stock assessment (Patterson et al., 2021), and a review of swordfish biology, status and stock structure in the Pacific (Moore, 2020). Key changes made in the progression from the 2017 to 2021 diagnostic case models include:

- Updating all data up to the end of 2019.
- Updating the biological assumptions, including defining reproductive potential as the product of sex-ratio and female maturity at length.
- Implementation of the “index fishery” approach.
- Adding the New Zealand index fishery and removing the Japanese and Chinese Taipei index fisheries from the diagnostic case model.
- Assuming that the population was at an unfished equilibrium at the start of the model period in 1952.

In addition to the diagnostic case model, we report the results of one-off sensitivity models to explore the relative impacts of key data and model assumptions for the diagnostic case model on the stock assessment results and conclusions. These one-off sensitivities included the development of a sex-disaggregated model and a model that did not assume spatial structure. A model ensemble was developed based on Ducharme-Barth and Vincent (2021) for consideration in developing management advice where presented uncertainty is the combination of model (structural) and estimation (statistical) uncertainty. This is a more holistic and transparent approach to presenting uncertainty than has been previously considered in other WCPFC assessments.

As per recent Scientific Committee practice, it is recommended that management advice is for-

culated from the results of the model ensemble. The initial ensemble of 384 models was filtered to 25 final models by excluding models with estimation issues or those that produced nonsensical results. Broadly speaking, the results from the current assessment are consistent with those from the previous stock assessment (Takeuchi et al., 2017) in terms of the central tendency of estimates, albeit more uncertain. This is expected given the change in approach for presenting uncertainty. The most important factors contributing to the uncertainty around the estimated stock status are the assumptions for movement rate, and natural mortality. Choice of distant water fishing nation (DWFN) index, length frequency scalar, and average penalty for the Australian index were also influential but to a lesser extent. Larger assumed values of natural mortality resulted in a smaller, less depleted stock. High rates of movement from the South Pacific to the Tasman region, coupled with low rates of movement in the inverse direction resulted in a larger, less depleted stock.

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

- The model and estimation uncertainty from the final model ensemble indicate that the stock is at a 20% risk of undergoing overfishing according to $F_{recent}/F_{MSY} > 1$ (0.47 median; 0.25 – 1.29 80th quantile), a 13% risk that the stock is overfished according to $SB_{latest}/SB_{F=0} < 0.2$ (0.39 median; 0.18 – 0.79 80th quantile) and a 10% risk that the stock is overfished according to $SB_{latest}/SB_{MSY} < 1$ (2.95 median; 0.99 – 6.78 80th quantile).
- Comparison with the equivalent sex-disaggregated model ensemble indicated a more optimistic stock status in terms of median $SB_{latest}/SB_{F=0}$ (0.45 median; 0.13 – 0.72 80th quantile), with greater uncertainty resulting in a 19% risk that the stock is overfished according to $SB_{latest}/SB_{F=0} < 0.2$.
- The model ensemble predicts the stock to have gradually declined from the 1950s to the mid-1990s before rapidly declining to an overall low point near 2010. Current stock status is estimated to be at a similar level as the overall low with a declining trend in the terminal 4 years of the model.
- Directed longitudinal tagging of swordfish to reduce the uncertainty in movement rates, and a feasibility study to explore applying Close-Kin Mark Recapture (CKMR) techniques to SWPO swordfish are the two most critical research items.

2 Acronyms and Abbreviations

AIC Akaike information criterion

ASPM Age-Structured Production Model

B Total biomass

CASAL C++ Algorithmic Stock Assessment Laboratory

CKMR Close-kin Mark Recapture

CPUE Catch-per-unit-of-effort

DWFN Distant-water fishing nation

EEZ Exclusive Economic Zone

EPO Eastern Pacific Ocean

ESS Effective sample size

EU European Union

F Fishing mortality

GAM Generalized Additive Model

GL(M)M Generalized Linear (Mixed) Model

h Stock-recruitment steepness

IATTC Inter American Tropical Tuna Commission

LJFL Lower-jaw fork length

LRP Limit reference point

M Natural mortality

MASE Mean absolute scaled error

MSE Management strategy evaluation

MSY Maximum sustainable yield

L_{∞} von Bertalanffy asymptotic maximum length

PAW SPC Pre-Assessment Workshop

PICT Pacific Island Country/Territory

PSAT Pop-up satellite archival tag

RMSE Root mean squared error

SB Spawning potential (single sex model), spawning biomass (multi-sex model)

SC WCPFC Scientific Committee

SPC Pacific Community; WCPFC Scientific Services Provider

SRR Swordfish

SWO Swordfish

SWPO Southwest Pacific Ocean

VAST Vector Auto-regressive Spatio-Temporal

WCPFC Western and Central Pacific Fisheries Commission

WCPO Western and central Pacific Ocean

WW Whole weight

3 Introduction

This paper presents the 2021 MULTIFAN-CL stock assessment of swordfish (*Xiphias gladius*; SWO) in the SWPO (south of the Equator between 140°E and 130°W), including the area of overlap between the WCPFC and IATTC convention areas. The first assessment of swordfish in the south Pacific region, conducted in 2006, focused on the area 140°E – 175°W and used MULTIFAN-CL (Kolody et al., 2006). This assessment was updated in 2008 by Kolody et al. again using MULTIFAN-CL, while a CASAL-based assessment was performed for the south-central Pacific area alone (140°E – 130°W; Davies et al. (2008)). In 2013 a new assessment was conducted using MULTIFAN-CL (Davies et al., 2013), which assumed two model regions delineated at 165°E in the WCPFC area south of the equator (140°E – 130°W; Figure 1). This delineation was based upon the results of electronic tagging programs (Evans et al., 2012). The most recent assessment was conducted in 2017 (Takeuchi et al., 2017), using MULTIFAN-CL and applying the same two region structure as the 2013 assessment. Follow up work was undertaken in 2018 to develop a multi-sex model (Takeuchi et al., 2018). However, this model was not used to provide management advice.

The 2017 assessment included a structural uncertainty grid of 72 models incorporating natural mortality, steepness, movement, and the size frequency likelihood weighting. The assessment estimated that biomass had declined throughout the model period for all models in the grid, but the decline was particularly steep in the last 15 years, and more so in region 2 (the eastern region). A number of grid models (23 of 72) estimated that recent fishing mortality was exceeding F_{MSY} and some (8 of 72) estimated that spawning potential was less than SB_{MSY} . While there is no formal LRP for swordfish in the SWPO, none of the models estimated that recent spawning potential was lower than 20% of what would be estimated in the absence of fishing (e.g. $20\%SB_{F=0}$), and thus not breaching the current WCPFC depletion based LRP for tuna species.

As in previous assessments, the objectives of the 2021 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 indicate the stock status and impacts of fishing. As previously mentioned there are no formal reference points adopted by the WCPFC for SWPO swordfish and we summarize the stock status and fishing impacts in terms of yield and depletion based reference points. The methodology used for the assessment is based on the general approach of integrated modeling (Fournier et al., 1998; Maunder and Punt, 2013), which is carried out using the stock assessment framework MULTIFAN-CL² (Fournier et al., 1998; Hampton and Fournier, 2001; Kleiber et al., 2019). MULTIFAN-CL implements a size-based, age- and spatially-structured population model. Model parameters are estimated by maximizing an objective function, consisting of both likelihood (data) and penalties constraining the estimates of certain components.

Each new assessment involves updates to fishery input data, implementation of new features in the

²<http://www.multifan-cl.org>

MULTIFAN-CL modeling software, and consideration of new information on biology, population structure and other population parameters. These updates are an important part of efforts to improve the modeling procedures and more accurately estimate stock status, fishing impacts, biological, and population processes. However, these updates can manifest in changes to the estimated status of the stock and fishing impacts from previous assessments. It is important to recognize that each new assessment represents a new estimation of the historic population dynamics, impacts of fishing and stock status. Advice from the SC on previous assessments, and the annual PAW (Hamer et al., 2021) guide this ongoing process.

A feature of previous stock assessments of SWPO swordfish is the use of an “uncertainty grid”. The uncertainty grid is a group of models that are selected and run to explore the interactions among key axes of uncertainty that relate to biological assumptions, data inputs, and data weighting/treatment. This approach to representing uncertainty can be inefficient as the factorial combinations quickly produce intractable numbers of models, including unrealistic factor combinations, e.g., high M and high L_∞ . Furthermore, the approach can be ad-hoc, and subjective in relation to weighting the models for management advice. For this assessment we apply a novel approach to representing uncertainty in the model derived management quantities. The new approach is described in detail in the supporting paper by Ducharme-Barth and Vincent (2021). In short the approach involves developing a multivariate prior for the key parameters that are fixed in the stock assessment model and then drawing from this prior to develop a correlated parameter set for each model in the ensemble in order to capture structural or *model*-based uncertainty in the derived management quantities. The approach propagates parameter uncertainty and correlation through the stock assessment while providing an implicit model weighting based on the shape of the prior distribution. The approach also incorporates the *estimation* uncertainty of individual models in the ensemble by combining their statistical uncertainty in a parametric bootstrap. The presented uncertainty therefore includes both estimation and model-based uncertainty in the management quantities. This approach is seen as a more appropriate way to capturing uncertainty in model outputs, and received strong support from the 2021 PAW (Hamer et al., 2021), to be applied for the first time in this assessment.

The 2021 SWPO swordfish assessment incorporates data up until and including 2019 from longline fisheries across the SWPO model region (Figure 1) and incorporates the new ensemble approach for incorporating structural and statistical model uncertainty. Preparatory work on data for the assessment is extensive and is described in limited detail in this report. We strongly advise that this assessment report is read in conjunction with several supporting papers:

- Focusing on the front end: A framework for incorporating uncertainty in biological parameters in model ensembles of integrated stock assessments (Ducharme-Barth and Vincent, 2021).
- Background analysis for the 2021 stock assessment of southwest Pacific swordfish (Ducharme-Barth et al., 2021b)

- Biology, stock structure, fisheries, and status of swordfish, *Xiphias gladius*, in the Pacific ocean - a review (Moore, 2020)
- Broadbill swordfish movements and transition rates across stock assessment spatial regions in the western and central Pacific (Patterson et al., 2021)

4 Background

4.1 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 and Ueyanagi, 1974; Nishikawa et al., 1985) and studies of spawning and reproduction (Young et al., 2003; Mejuto et al., 2008). These studies suggest spawning occurs in more localized regions predominantly in tropical and sub-tropical areas, with conspicuous absence from the Western Pacific equatorial region, and the coastal regions of North and South America. Tagging studies provide some confirmation that swordfish undergo directed seasonal migrations between temperate foraging grounds and tropical spawning grounds, with perhaps less longitudinal movement (summarized in Moore (2020)). It remains unclear how much site fidelity or philopatric behavior 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.

Evans et al. (2014) 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 PSAT tagging of larger swordfish off the east coast of Tasmania 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 life due to the deployment duration of the tags. They also provide no information on movement or dispersal of small juveniles and earlier life stages, and are generally inconclusive as to biological stock structure or meta-population 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 \cap shaped connectivity pattern, such that the south west and north west Pacific populations are the most 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 behavior appears complex, and likely differs by sex and size, and there is evidence of fidelity to particular areas ([Burnett et al., 1987](#); [Tracey and Pepperell, 2018](#)). As such there is likely to be sub-structure and migratory contingents within any 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.

4.2 Vertical movement behavior

An important aspect of swordfish behavior 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 >300 m, during the day ([Abascal et al., 2010](#); [Dewar et al., 2011](#)). This typical behavior 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 behavior as swordfish follow the diurnal migrations of prey from the deep scattering layer, primarily squids 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, or deeper during the day. 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](#)).

4.3 Biological characteristics

Swordfish are sexually dimorphic, with females growing larger and faster than males (Young et al., 2003; 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 (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 age-length data-set from the Farley et al. (2016) study is used in this assessment as part of the biological re-analysis to develop the joint prior for the model ensemble (discussed in Ducharme-Barth et al., 2021b). 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 and age at 50% maturity is 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 in Section 6.2.5, 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.

4.4 Fisheries

Due to the diurnal vertical migration behavior 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 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; Hsu et al., 2015; 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 between 9,000 and 11,000 mt (Figure 2). Early catches were primarily by Japanese fleets, 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) fleet increased as it started targeting swordfish in the south-central Pacific Ocean (Figure 2). Since 2014 the catch of swordfish has declined across all fleets in SWPO, with recent estimates indicating an annual catch of between approximately 7,000

and 8,000 tons (Figure 2).

We note the request from the SC16 online discussion forum for a swordfish fishery characterization that clarifies the catches taken as target and bycatch, within EEZs and on the high seas. It was suggested that this characterization could be included in the 2021 stock assessment, however, we feel that this would distract from the purpose of the stock assessment, and we do not separate fisheries along these lines for stock assessment purposes. The requested information is detailed in the paper SPC-OFP (2021). Briefly, since 2000, around 50% of the swordfish catch in the SWPO has been taken in the tropical and sub-tropical waters north of 20°S. Estimates of the proportions of catch taken as targeted versus bycatch over the last 10 years indicate that for EEZs in the SWPO, on average approximately 68% is targeted catch, and for the HS, 26% is targeted catch (SPC-OFP, 2021).

5 Data compilation

5.1 General notes

Data used in the SWPO swordfish assessment consist of catch, effort, length and weight-frequency data for the fisheries (Table 2 & Table 3) defined in the analysis. Detailed summaries of the analyses and methods of producing the necessary input data are detailed in the supporting papers referred to previously. Full details of the inputs and related methods are not repeated here, rather we provide a brief overview of the key features and direct readers to the relevant papers throughout this section.

5.2 Spatial stratification

The broad geographic area considered in the 2021 SWPO swordfish assessment covers the southwest Pacific Ocean from 140°E to 130°W, and from the equator to 50°S, which includes the area of overlap between the WCPFC and IATTC convention areas (Figure 1). Within this area there are two spatial assessment regions and within each region, three fishery strata, consistent with the 2017 assessment. 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). The three fishery strata in each region (northern, central and southern) help account for within region size structure of the catch, and also for fishery areas that have different flag/fleet and operational characteristics.

5.3 Temporal stratification

The time period covered by the assessment is 1952–2019 which includes all significant post-war commercial fishing in the SWPO. Within this period, data were compiled into quarters (1; Jan–Mar, 2; Apr–Jun, 3; Jul–Sep, 4; Oct–Dec). Modeled population dynamics (e.g. periodicity of recruitment) were implemented with an annual time step, while movement between regions occurred at a quarterly scale. The assessment does not include data from the most recent complete calendar

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

5.4 Definition of fisheries

MULTIFAN-CL requires “fisheries” to be defined that consist of relatively homogeneous fishing units. Ideally, the defined fisheries will have selectivity and catchability characteristics that do not vary greatly over time and space. For most pelagic fisheries assessments, fisheries are defined according to gear type, fishing method and region, and, in some cases, also by vessel flag or fleet. The fishery sub-areas of the two model regions are also used to define the spatial boundaries of the specific fisheries.

For this assessment we also apply an “index” fishery approach (Tremblay-Boyer et al., 2018; Ducharme-Barth et al., 2020; Maunder et al., 2020; Vincent et al., 2020). In this approach there are two fishery categories: “index” fisheries and “extraction” fisheries. Extraction fisheries are used to account for removals from the population while “index” fisheries are constructed to provide the trend for the population in each region.

5.4.1 Extraction fisheries

A total of 13 longline extraction fisheries were defined (Table 2), based on the sub-area boundaries, flags and time periods. These fisheries can be summarized as vessels coming from Distant Water Fishing Nations DWFNs, Pacific Island Countries and Territories PICTs, Australia, New Zealand, or the European Union (predominantly Spanish vessels). For this assessment, the Vanuatu flag was grouped with the DWFNs since most vessels are known to be ex-Chinese Taipei flagged vessels, fishing throughout the SWPO and hence have characteristics more comparable to other “DWFN” fleets. The catches by flag and size compositions of the various fisheries listed in Table 2 are summarized in the inputs paper by Ducharme-Barth et al. (2021b).

5.4.2 Index fisheries

The index fisheries approach was not applied in the previous assessment. Defining the index fisheries for each model region for the new assessment was challenging, as a key criterion for grouping flags/fleets into regional index fisheries is that they should have relatively similar and time invariant selectivity and catchability. If catchability changes over space and time this needs to be accounted for by the application of suitable covariates in a CPUE standardization procedure. Unfortunately for most flags, historical collection of detailed data on operational gear settings (e.g. materials, baits, etc.) is lacking and some flags have changed their spatial fishing patterns and targeting practices, all of which have implications for swordfish catchability and encounter probability. In other cases reporting procedures have changed which are thought to have led to differences in data quality that could bias catch rate data, but can not be accounted for in catch rate standardization.

Discussions in the PAW and with experienced fisheries staff in particular DWFNs helped identify time periods when changes in gear configurations, setting practices, reporting procedures and spatio-temporal fishing patterns likely occurred. This information provided a basis for exploring nominal catch rate data and supporting decisions on splitting catch and effort data in space and time blocks to generate the various index fisheries. After considerable data exploration, 13 index fisheries thought to best represent the underlying dynamics of the stock were created as follows (Figure 3):

- Australia: Similar to the previous assessment, a single index fishery was defined for the Australian domestic longline fishery operating in region 1. This index runs from 1998 – 2019, and uses a delta-GLM approach to standardize nominal CPUE. (Campbell, 2019; Dell et al., 2020). Indices were calculated by size category (recruits, pre-spawners, and mature), though an average index across all sizes was used in the current assessment. Comprehensive, consistent logbook coverage throughout the index period allows for a number of fishing practice, environmental, and vessel competitive effects to be included in the standardization the model.
- New Zealand: New for the current stock assessment is the inclusion of a single index fishery defined for the New Zealand domestic longline fishery operating in region 2. Concerns over potential catchability changes brought on by the introduction of a swordfish catch quota in 2004 led the PAW to recommend only including the portion of the index following the introduction of the catch quota. The retained index (2004 – 2019) was developed by applying a GAM to standardize operational level nominal CPUE (Finucci et al., 2021). Similar to the analysis of the Australian longline CPUE data, extensive operational covariate information is available in the New Zealand logbook. Vessel effects, hooks per basket, light stick use, and night time setting all proved to be informative covariates.
- European Union: As in the previous assessment, a single index fishery was defined for the EU (Spanish) longline fishery operating in region 2. This index was created by applying the VAST modeling framework (Thorson, 2019) to EU (Spanish) operational logbook data from 2004 – 2019 (Ducharme-Barth et al., 2021b). Logbook reported operational covariates are unavailable for this fishery which limited the standardization modeling.
- Japan: Six index fisheries were defined for the Japanese longline fishery, where indices were split according to time period and model region. Follow-up investigation of the operational longline data including summaries of operational setting characteristics from the Japanese observer program, as well as discussion with Japanese scientists led to the determination of the following three temporal periods: early (1952 – 1974), middle (1975 – 1993), and late (1994 – 2019). Several factors influenced the choice of these model periods including: 1) changes in logbook reporting procedures that likely improved the quality of catch and effort data from 1975, 2) the transition in mainline material (likely impacting the relationship between hooks per basket and fishing depth) in the mid-1990s from kuralon rope to monofilament, and 3) indication from the observer program data summaries that fishing and targeting prac-

tices differed for the Japanese fleet between regions 1 and 2. To account for these temporal and spatial splits, four separate VAST standardization models were applied to the Japanese operational longline data to generate the indices (Ducharme-Barth et al., 2021b).

- Chinese Taipei: Four index fisheries were defined for the Chinese Taipei longline fishery. Similar to Japan, examination of the operational logbook data indicated apparent differences in targeting, based on observed catch composition, between regions 1 and 2 that could not be standardized out due to a lack of operational gear covariates. Additionally, the switch in mainline material during the 1990s was believed to be universal across all longline fisheries, though the exact timing of the change is uncertain. Accordingly, to account for this and a perceived shift towards bigeye tuna targeting in region 2, the index was split into two time periods for both regions: early (1965 – 1997), and late (1998 – 2019). To account for these temporal and spatial splits, three separate VAST standardization models were applied to the Chinese Taipei operational longline data to generate the indices (Ducharme-Barth et al., 2021b).

In previous WCPO stock assessments (Tremblay-Boyer et al., 2018; Ducharme-Barth et al., 2020; Vincent et al., 2020) the index fisheries also served to inform the scaling of biomass between regions. However, this requires the assumption of shared catchability (and selectivity) between index fisheries operating in the different spatial regions. This assumption could not be made for any two sets of fisheries operating in different spatial regions, and so the index fisheries were unable to directly inform relative biomass scale between regions.

5.5 Catch and effort data

5.5.1 General characteristics

Catch and effort data were compiled according to the fisheries defined in Table 2 & Table 3. For all fisheries, except the EU fleets, catch data were expressed as the number of swordfish captured and fishing effort as the number of hooks set. As catches submitted by the European Union vessels are available in weight per set across the entire time period of activity³, a separate fishery was created for this fleet using weight per set. Catch and effort data for all fisheries were aggregated within the quarterly time intervals.

The internal structure in the current production version of MULTIFAN-CL, a catch-errors formulation without an independent likelihood component for the index data, prevents a direct fit to the standardized CPUE index via a formal likelihood component. Fitting to the index is done by penalizing the estimated effort deviates to fit the input catch and effort⁴ data for a given fishery. Modifications to the input catch for the index fisheries are needed so that these fisheries do not impact the model through removals. This is done by setting the catch for these fisheries to a

³We note that the European Union has submitted catches as both numbers and weight in recent years.

⁴In this case effort is defined as standardized effort $\frac{Catch}{CPUE}$

negligible value, 1 fish per quarter. Likewise, in order to avoid having the non-standardized effort associated with the extraction fisheries influence the population trend, the effort deviate penalties for these fisheries are set to be very small, equivalent to a CV of 20.

As virtually all the catch of swordfish is taken by longline this is the only fishing method considered for this assessment. Although small recreational catches occur around Australia and New Zealand, they have not been quantified, but are likely to be negligible in terms of overall fishing impact. Swordfish discarding rates are not known for most of the catch history. As such, discards are not explicitly included for most of the catch history used in this assessment. We assume, given the value of swordfish, discarding rates would be low and unlikely to represent meaningful unaccounted for fishing mortality. More recently, discard reporting has improved, and increased discarding may occur in quota fisheries (e.g. Australia and New Zealand) if high grading is occurring or quotas are met. For these fisheries, discards are a mandatory reporting requirement since 2016 and where discards are reported in recent years they are included as catch.

Total annual catches for the WCPO area south of the equator and by model region are shown in [Figure 2](#). The spatial distributions of catch and effort over the past two decades are provided in [Figure 4](#). The catch and effort summaries for SWPO swordfish indicate a number of notable trends in the fisheries over the model period, specifically:

- the distant water Asian fleets took most of the catch from the 1950s to the late 1990s, with catches relatively evenly spread across model regions 1 and 2.
- from the late 1980s catches increased, largely due to the Australian fleet in region 1, and the EU and DWFN fleets in region 2, with a smaller contribution from the New Zealand fleet.
- catches since 2000 in region 1 are dominated by the Australian fleet, and in region 2 by the DWFNs, New Zealand and the emergence of the EU fleet.
- The highest catch areas are consistently off the central east coast of Australia, the north east of New Zealand, and the north east corner of the model region 2, particularly around Kiribati and the adjacent high seas. Catches have also increased since 2000 in the central high seas area of region 2 to the east of New Zealand.
- Catches have declined across model regions and all fisheries since 2014.
- Effort in the longline fishery has increased overtime across all model regions, most notably in the tropical region north of 20°S.
- The higher catch areas off eastern Australia and northern New Zealand correspond to relatively moderate effort levels, and the higher catches in the north east of region 2 correspond to higher effort areas.
- The increased catches since 2000 in the central high seas area of region 2 correspond to a moderate increase in effort since 2000.

- Despite the high effort in central-western region from 10–20°S, swordfish catches are low.

5.6 Size data

Swordfish size composition data have been collected from the fishery since the early 1990's. These data are characterized by inconsistent temporal and spatial resolution, may be subject to very small sample sizes and consequently exhibit high variability in some periods (see [Ducharme-Barth et al., 2021b](#)). The standard length measurement for swordfish is the lower-jaw fork length (LJFL) in cm and the standard weight measurement is the whole weight (WW) in kg. Length and weight measurements from the different port, observer and logbook sampling programs that were not measured as LJFL or WW are converted according to the conversion factors listed in [SPC-OFP \(2019\)](#). The previous assessment considered 29 length frequency bins (10 cm width, 30-310 cm lower bin interval) and 31 weight frequency bins (10 kg width, 2-302 kg lower bin interval). Re-examination of the size frequency data indicated that rounding of length measurements to 10 cm intervals precluded considering a smaller length bin resolution for the current assessment. However, there was no such issue found with the weight frequency data, and the bin size was subsequently reduced to 5 kg for 62 total bins (2-307 kg lower bin interval).

Statistical re-weighting of size composition data are required as length and weight samples are collected unevenly in space and time. The methods for re-weighting of the size composition data are detailed in [Ducharme-Barth et al. \(2021b\)](#) and are based on those developed by [McKechnie \(2014\)](#) for longline extraction fisheries, and [Tremblay-Boyer et al. \(2018\)](#) for longline index fisheries. For the extraction fisheries, re-weighting of composition data by the $5^{\circ} \times 5^{\circ}$ spatial catch is required to ensure that sampling biases in space, time, and the fleets providing data, are minimized so that size composition data better reflect the composition of the overall removals. Strata-specific size data samples are therefore re-weighted by catch for the extraction fisheries. For the index fisheries, re-weighting of composition data are required to ensure that the size composition of the abundance indices reflect the size component of the population that is being sampled by the index fisheries through time. Strata specific samples are therefore re-weighted by relative abundance using the CPUE. An exception to this were the index fisheries for Australia and New Zealand for which high resolution spatial standardized CPUE data were unavailable. For these two fisheries the same preparation of composition data was used for both the extraction and the index fisheries. Given that the same composition data were used for both the extraction and index fisheries, the observed number of size-frequency samples input into the assessment was divided by two for each fishery.

5.7 Tagging data

Limited tagging data exists for swordfish in the SWPO. The most recent work was the large-scale, collaborative PSAT study conducted by [Evans et al. \(2014\)](#). Their research suggested limited east/west transfer across the Coral and Tasman Seas in the SWPO, and provided the basis for the current two region assessment structure. This research also appeared to indicate the absence of

mixing between the southern and northern WCPO, or the SWPO and EPO. However, it is worth noting that these conclusions were drawn from a relatively restricted sample of 53 individuals with short times at liberty (30 – 362 days), and predominantly tagged off the east coast of Australia.

Though not used directly in the stock assessment, these tagging data are used to derive fixed, quarterly bulk transfer rates for movement between the two regions. A Hidden Markov Model was used to estimate the transition probability of tagged individuals across the 165°E regional boundary, while also accounting for the uncertainty in the position locations of tagged individuals (Patterson et al., 2021).

6 Model description

6.1 General characteristics

The model can be considered to consist of several components, (i) the population dynamics of the stock; (ii) the fishery dynamics; (iii) the observation models for the data; (iv) the parameter estimation procedure; and (v) stock assessment interpretations. Detailed technical descriptions of components (i)–(iii) are given in Hampton and Fournier (2001) and Kleiber et al. (2019).

Here, we describe the procedures followed for estimating the parameters of the model and the way in which stock assessment conclusions are drawn using a series of reference points.

6.1.1 Uncertainty characterization

The 2021 SWPO swordfish stock assessment introduces a new approach, described in Ducharme-Barth and Vincent (2021), for characterizing the uncertainty in management reference points and quantities of interest. Building off the familiar “structural uncertainty grid”, the new *model ensemble* approach continues to consider the effects of “structural” uncertainty (termed *model uncertainty*) while extending it to also account for the “statistical” (termed *estimation*) uncertainty from each model in the ensemble. This allows for a more holistic and transparent description of the uncertainty in estimates of stock status. Another key difference between the model ensemble and the structural uncertainty grid is the relaxation of the full factorial design. Instead of choosing set levels for certain fixed parameters (e.g. steepness $\in \{0.65, 0.8, 0.95\}$) which can be parameterized using a distribution, a random set of the fixed parameters is drawn from a joint prior distribution for each model in the ensemble. Drawing parameter sets from this joint prior has some notable benefits: 1) sampling from a prior distribution implicitly weights the ensemble to the most likely parameter combinations given the shape of the prior, 2) incorporating parameter correlations in the construction of the joint prior self-censors the ensemble to exclude implausible parameter combinations, and 3) model uncertainty in the choice of fixed parameters can be more efficiently characterized using fewer models in the ensemble. For aspects of model uncertainty that cannot be parameterized using the joint prior (e.g. fitting to alternate CPUE indices), the full factorial approach is still used but now is overlaid on the parameter draws from the joint prior in

a hybrid factorial ensemble.

The estimation uncertainty for each model was determined by calculating the Hessian matrix in order to obtain an estimate of the covariance matrix, which is used in combination with the delta method to compute approximate confidence intervals for quantities of interest (i.e. the biomass and recruitment trajectories). This was done for all models in the model ensemble and the estimation uncertainty was combined across models in a parametric bootstrap similar to the approach used in stock assessments conducted by the International Pacific Halibut Commission (Stewart and Martell, 2014).

Similar to previous WCPO stock assessments, a representative diagnostic case is still used for model development, diagnostic evaluation, and as the baseline configuration for the structural uncertainty grid or model ensemble, in this case. Biological parameters assumed in the diagnostic case, and the range of parameters considered in the model ensemble can be found in Table 4 and Figure 5.

6.2 Population dynamics

The model partitions the population into 2 spatial regions and 20 annual age-classes. The last age-class comprises a “plus group” in which mortality and other characteristics are assumed to be constant. The population is “monitored” in the model at annual time steps, extending through a time window of 1952–2019. The main population dynamics processes are as follows.

6.2.1 Recruitment

Recruitment is defined as the appearance of age-class 1 year fish (e.g. fish averaging ~ 98 – 110 cm given current growth curves) in the population during the first quarter of the calendar year. In the context of a stock recruit relationship, there is a one year lag between the calculation of the average SB that produces recruits entering the model. The concept of SB for sex-aggregated assessments is described further in Section 6.2.6.

Spatially-aggregated (across both spatial regions) recruitment was assumed to have a moderate relationship with annual mean spawning potential via a Beverton and Holt stock-recruitment relationship (SRR) with a fixed value of steepness (h). Deviations in the estimated recruitments from the average recruitment predicted by the SRR attract a penalty in the likelihood. A moderate penalty was assumed in the diagnostic case model equivalent to annual recruitment having a CV of 0.5. Sensitivity to this parameter was explored and three levels were considered, recruitment CV $\in \{0.3, 0.5, 0.97\}$, as a factor in the model ensemble.

Steepness in a stock assessment context is defined as the ratio of the equilibrium recruitment produced by 20% of the equilibrium unexploited spawning potential to that produced by the equilibrium unexploited spawning potential (Francis, 1992; Harley, 2011). Typically, fisheries data are not very informative about the steepness parameter of the SRR parameters (ISSF, 2011); hence, the steepness parameter was fixed in the assessment. As in the previous assessment, the diagnostic

case assumed a moderate steepness value of 0.8. For the model ensemble, a steepness prior was created to account for the limited existing information available for swordfish (Myers et al., 1999). Based on this information, a Beta prior ($\alpha = 17.34333, \beta = 2.65$) was specified to have a median of 0.88 and 95th percentile range (0.69–0.98) similar to what was considered for steepness in the previous stock assessment (Table 4 & Figure 5 b).

The SRR was incorporated mainly so that yield analysis and population projections could be undertaken for stock assessment purposes, particularly the determination of equilibrium- and depletion-based reference points. The SRR was calculated over the full model period. As will later be made evident in consideration of runs that did not include standardized CPUE from the Japanese or Chinese Taipei longline, this had the effect of constraining the recruitments to be close to the average recruitment predicted from the stock recruit relationship since there was no other data in the model to inform them.

In recent WCPO stock assessments the terminal recruitments have often been fixed at the mean recruitment of the rest of the model period. This acknowledges that these estimates are poorly supported by data and if unconstrained can vary widely, with potentially large consequences for stock projections. This approach has been continued here by fixing the 2 terminal years of recruitment at the mean of the recruitments over the rest of the assessment period.

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

6.2.2 Initial population

The population age structure in the initial time period in each region was assumed to be in equilibrium and determined as a function of the average total mortality during the first 2 years. This assumption avoids having to treat the initial age structure, which is generally poorly determined, as independent parameters in the model. Given the negligible catches observed in the first years of the model and the lack of industrial fishing prior to the start of the model in 1952, this essentially assumes that the model commences from an unfished equilibrium state based on the assumed natural mortality.

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

6.2.3 Growth

The standard assumptions for WCPO assessments fitted in MFCL were made concerning age and growth: i) the lengths-at-age are normally distributed for each age-class with standard deviation σ_{Age} ; ii) the standard deviations of length (σ_{Length}) for each age-class are a log-linear function of

the mean lengths-at-age; and 3) the probability distributions of weights-at-age are a deterministic function of the lengths-at-age and a specified weight-length relationship. These processes are assumed to be regionally and temporally invariant. The growth curves assume a von Bertalanffy functional form and are defined in MULTIFAN-CL using L_1 (length at first age-class), L_2 (length at terminal age-class), and growth coefficient k . All parameters were held fixed in the assessment except the two growth variability parameters which were estimated internally.

Growth for the 2021 SWPO swordfish was derived from a re-analysis of the otolith aging data first described by (Farley et al., 2016), where the annual ages were re-calculated as decimal ages using the algorithm developed for bigeye tuna (Farley et al., 2020). Posterior distributions for sex-specific parameter values were estimated using separate Bayesian models, with relatively uninformative priors, to facilitate the consideration of a sex-disaggregated assessment model (Ducharme-Barth et al., 2021b). The “classic” von Bertalanffy growth parameters, L_∞ , k , and t_0 were estimated before being converted to L_1 and L_2 for input into MULTIFAN-CL.

Discussion at the PAW noted that the model output appeared sensitive to the value of t_0 , and that the estimated t_0 value of ~ -2 (i.e. age at hypothetical size 0cm) was far from the theoretical value of 0. This could indicate a possible mis-specification either in the aging algorithm or the choice of growth model. As uninformative priors were used in the initial estimation, it was suggested to re-estimate the parameters with a more informative t_0 centered on 0. Placing a more informative prior on t_0 predictably shifted the estimation of this parameter closer to zero. This resulted in larger assumed values for k , and smaller assumed values for L_1 and L_2 relative to parameters estimated from the model with the uninformative prior (Figure 6).

Sex-aggregated parameters were “averaged” across the sex-disaggregated parameters by re-estimating a set of growth parameters from pairs of sex-specific growth curves. This was done for parameters sets both with the informative and un-informative t_0 prior. These two posterior distributions were combined to create a single distribution accounting for both alternative states of nature. Random draws from this sex-aggregated distribution was used in the model ensemble (Table 4 & Figure 5 a).

6.2.4 Movement

Movement between the two model regions was assumed to occur instantaneously at the beginning of each quarter. Movement in MULTIFAN-CL is parameterized by a pair of bi-directional coefficients for each region boundary, only one in the current case. Movement is possible in both directions across the regional boundary in each of the four quarters, and in this case movement was assumed to be constant across age classes and quarters.

As mentioned previously in Section 5.7, movement rates between regions were derived from the PSAT tagging data. The estimated covariance structure from the Hidden Markov Model developed by Patterson et al. (2021) was used to generate a correlated distribution of bi-directional movement

rates from the “Tasman” region (1) to the “South Pacific” region (2), and vice-versa. Model estimates were very uncertain, but tended to show an asymmetric movement pattern with fewer fish traveling from the Tasman to the South Pacific, than from the South Pacific into the Tasman region. This bivariate distribution was incorporated into the joint prior for the model ensemble (Table 4 & Figure 5 b).

6.2.5 Natural mortality

Natural mortality (M) was assumed to be fixed in the model at age-specific values following the Stock-Synthesis parameterization (Methot and Wetzel, 2013) of the Lorenzen (2000) natural mortality curve. This formulation derives age-specific values of natural mortality using the growth curve, and M at a reference length M_{ref} , where the reference length was chosen to be the length at the maximum age L_2 .

Recent work (Maunder et al. *Submitted*) suggests using either a maximum age based ($\frac{5.4}{A_{max}}$; Hamel and Cope *Submitted*) or life-history based ($4.118k^{0.73}L_{\infty}^{-0.33}$; Then et al., 2015) approach for developing an estimate of M for mature individuals. Both methods were used to develop priors for M_{ref} . The maximum age based prior was defined by a lognormal distribution, $\text{Lognormal}(\mu = \log(\frac{5.4}{20}), \sigma = 0.1)$. To create the prior for the life-history based approach, correlated pairs of L_{∞} and k were drawn from the ensemble joint prior distribution. Additionally, the coefficients from Then et al. (2015) have their own uncertainty and a parametric bootstrap was used following the approach from Lopez-Quintero et al. (2017) to incorporate the uncertainty in the life-history relationship. Keeping track of the L_{∞} , t_0 , and k that were used to derive L_1 , L_2 , and M_{ref} allowed for the creation of a correlated 6-tuple of parameters.

These two distributions for M_{ref} were combined to create a single distribution for the model ensemble. Random draws of M_{ref} input into the Lorenzen (2000) equation allowed us to account for uncertainty in the M -at-age in the model ensemble (Table 4 & Figure 5 a). This process was also repeated using sex-specific growth parameters to produce sex-specific distributions of M_{ref} and M -at-age. Though distributions of M_{ref} showed considerable overlap between the two approaches, the maximum age based approach tended to produce higher values of M_{ref} , and M -at-age than the life-history based approach (Figure 7).

6.2.6 Reproductive potential

The reproductive potential ogive is an important component of the assessment structure as it translates model estimates of total population biomass to the relevant management quantity, in the sex-aggregated case, spawning potential (SB). In order to only account for female reproductive output in a sex-aggregated model, reproductive potential is defined as the product of female maturity at age (or length) and female sex-ratio at age (or length). In a sex-disaggregated model, the reproductive potential is simply equal to the female maturity at age (or length).

Subsequent to the 2017 stock assessment, a new feature in MFCL was developed allowing the reproductive potential ogive to be defined and input into the assessment as a function of length. This length-based ogive is then converted internally to a reproductive potential-at-age using a smooth-spline approximation (Davies et al., 2019). This allows for a more natural definition of reproductive potential as the product of the two length-based processes: proportion females-at-length⁵ (sex-ratio), and proportion of females mature-at-length⁶. An additional benefit to define reproductive potential as a function of length is that it is growth invariant. Previous stock assessments using an age-specific formulation had to redefine the reproductive potential-at-age ogive for each different growth curve included in the assessment.

Parameters for the sex-ratio at length and female maturity at length were estimated using Bayesian models (Ducharme-Barth et al., 2021b) in order to incorporate the sex-ratio and female maturity relationships into the joint prior. Similar to the procedure for the growth and M parameters, a correlated n -tuple of the maturity and sex-ratio parameters could be drawn from the joint posterior in order to propagate uncertainty in the reproductive potential relationship in the model ensemble (Figure 5 a).

6.2.7 Length-weight relationship

Parameters for the length-weight relationship were estimated using sex-specific Bayesian models based on observer samples (Ducharme-Barth et al., 2021b) in order to incorporate uncertainty in this relationship into the joint prior (Table 4 & Figure 5 a).

6.2.8 Sex structure

Swordfish are known to exhibit sexually dimorphic growth (Farley et al., 2016). One of the key recommendations of the previous assessment (Takeuchi et al., 2017) was to more explicitly model these sex-specific characteristics in a sex-disaggregated model, and a follow-up analysis was conducted in 2018 (Takeuchi et al.) to explore this capability in MULTIFAN-CL. The current assessment maintains the development of the sex-aggregated model as the diagnostic case and as the basis for providing management advice. However, equivalent sex-disaggregated diagnostic cases and model ensembles were developed as sensitivities to explore the potential for bias in the management advice from the sex-aggregated model.

6.3 Fishery dynamics

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

⁵For the current assessment, female sex-ratio-at-length was calculated from Regional Observer Program data in SPC's holdings through 2019.

⁶Taken from Farley et al. (2016).

and effort deviations, which are a random effect in the fishing effort - fishing mortality relationship.

6.3.1 Selectivity

Selectivity is fishery-specific and assumed to be time-invariant and length-based, but modeled as age-based (Kleiber et al., 2019). Though a regional structure was assumed in the current stock assessment, differences in selectivities among fisheries operating in sub-areas of the same model region may serve as a proxy for spatial structuring of the population by size.

Selectivity was modeled using cubic splines. This allows for greater flexibility than assuming a functional relationship with age (e.g. logistic curve to model monotonically increasing selectivity or double-normal to model fisheries that select neither the youngest nor oldest fish), and requires fewer estimated parameters than modeling selectivity with separate age-specific coefficients. This is a form of smoothing, and the number of parameters for each fishery is the number of cubic spline “nodes” that are deemed sufficient to characterize selectivity over the age range. The current assessment assumed 3 nodes for extraction fisheries and 5 nodes for index fisheries. Model sensitivity to increasing the number of nodes was investigated but not pursued further as it did not result in perceptible change to the estimated management quantities. Common terminal selectivity was assumed to start from age class 16. Additionally, for the fisheries showing strong domed shaped selectivity patterns (Fishery 8), selectivity was assumed to be fixed at zero beyond the age of shared selectivity in order to prevent spurious selectivity estimation by the cubic spline beyond the range of the observations.

Most fisheries in the assessment model were unrestricted in their estimates of selectivity. However, Fishery 9 (DWFN Region 1S) and all index fisheries were restricted to be non-decreasing. Non-decreasing selectivity was assumed for the index fisheries as testing during model development showed that unrealistic biomass estimates were produced when selectivity was unconstrained. Though size frequency data for Fishery 9 was sparse, it sampled the largest individuals in the assessment area so a non-decreasing selectivity was specified accordingly. Selectivity for Fishery 9 was also assumed to be zero for the first six age classes to prevent spurious estimation of selectivity of non-observed individuals. Similarly, selectivity for the first age class was constrained to be zero for Fishery 2 (DWFN Region 1C), though sensitivity to this assumption was tested and removing this constraint did not improve the model fit.

As in the previous assessment, a selectivity time break was used in 2008 for Fishery 4 (Australia Region 1) to account for a perceived change in selectivity due to a change from J-hooks to circle hooks.

6.3.2 Catchability

Constant (time-invariant) catchability was assumed for the index fisheries that received standardized indices of relative abundance. This assumption is similar to assuming that the CPUE for these

fisheries indexes the exploitable abundance over time. Additionally, a seasonal catchability deviate was estimated to account for seasonal patterns in catch rate likely due to seasonal latitudinal migrations of individuals within regions.

The current assessment assumes a very high penalty on the fit to the catch (SD of residuals on the log scale of 0.002) which is essentially equivalent to treating the catch as “error-free” as done in “catch-conditioned” modeling approaches. Additionally, the effort deviates for the extraction fishery were assigned a very low penalty so that the nominal cpue attributed to these fisheries did not impact the model likelihood. As a result, there was no need to estimate both effort and catchability deviates for the extraction fisheries.

6.3.3 Effort deviations

Effort deviations were used to model the random variation in the effort - fishing mortality relationship, and are constrained by pre-specified penalties (on the log-scale). The fleet-specific CPUE indices implemented through the index fisheries represent the principal indices of stock abundance, and the extent to which the model can deviate from the indices is moderated by the penalty weights assigned to the standardized effort series. For these fisheries the effort deviates were set to have a mean of zero and the CV was allowed to be time-variant and based on the variance estimates derived from the associated CPUE standardization model. An exception to this was for the CVs associated with the Australian longline index which were estimated to be incredibly small. These were rescaled to a median CV of 0.2 so that this index would have a similar scale to the other indices in the model. The index CVs were transformed to an effort deviate penalty for each CPUE observation in MFCL. Penalties are inversely related to variance, such that lower effort penalties are associated with indices having high variance, consequently these indices are less influential in fitting the model. Sensitivity to altering the relative likelihood weights of the different indices (i.e. increasing/decreasing the penalty on the fit to the different index fisheries) was explored and this was accomplished by varying the mean CV associated with each fishery. In the model ensemble, the mean CV for each index fishery was drawn from a uniform distribution $U(0.1,0.8)$.

Lastly, for all the extraction fisheries (Table 2), the effort deviates were given extremely low penalties equivalent to a CV of 20 to prevent the nominal CPUE for these fisheries (derived from the input catch and effort data) from influencing the estimated population dynamics.

6.4 Likelihood components

There are three data components that contribute to the log-likelihood function for this assessment: the total catch data, the length-frequency data, and the weight-frequency data. Fit to the CPUE data does not influence the fit as an explicit likelihood component but rather as a penalty on the effort deviates.

As mentioned previously, the observed total catch data are assumed to be unbiased and relatively

precise, with the SD of residuals on the log scale being 0.002.

The probability distributions for the length- and weight-frequency proportions are assumed to be approximated by robust normal distributions, with the variance determined by the effective sample size (ESS) and the observed length-frequency proportion. Size frequency samples are assigned ESS lower than the number of fish measured. Lower ESS recognize that (i) length- and weight-frequency samples are not truly random (because of non-independence in the population with respect to size) and would have higher variance as a result; and (ii) the model does not include all possible process error, resulting in further under-estimation of variances. The observed sample sizes are capped at 1,000 internal to MULTIFAN-CL, and these were further divided by 20 (i.e. size composition scalar of 20), resulting in a maximum ESS of 50 for each length and weight sample for a fishery. Fisheries 3 & 9 had length and weight composition scalars of 10 which increased the influence of this size frequency data on the likelihood. The sample sizes for these fisheries were quite small in comparison to the others, and increasing their relative likelihood contribution aided in the estimation of the selectivity curves. Alternative scalars for specifying ESS were explored in sensitivity analyses. In the model ensemble, the divisors for both the length and weight composition data were drawn from a uniform distribution $U(10,50)$.

6.5 Parameter estimation

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

6.6 Diagnostics

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

procedure is presented in the Appendix (Section 12.1.1), including examination of the profiles for the individual data components. Additional likelihood profiles were also conducted for average adult biomass defined over the entire model period, average recent (2015–2019) adult biomass, and the adult biomass in the terminal year of the model.

Retrospective analyses are also undertaken as a general test of the stability of the model. A robust model, when rerun with data for the terminal year(s) sequentially excluded (Cadigan and Farrell, 2005), should produce outputs without a systematic pattern in either the scaling or time-series trends. The retrospective analyses for the 2021 diagnostic case model are presented in the Appendix (Section 12.1.2). In conjunction with the retrospective analysis, we conducted additional likelihood profiles over recent (terminal 5 years) and terminal (last year) adult biomass for each retrospective “peel”. These additional likelihood profiles are particularly helpful in understanding which data components are influencing any observed retrospective pattern.

Considerable efforts have been made in recent years to improve the level of diagnostic reporting for WCPO stock assessments. Following the advice developed by Carvalho et al. (2017; 2021) improved residual and goodness of fit diagnostics for the fits to the index and size frequency data have been included. ASPM diagnostics (Minte-Vera et al., 2017) were also used to determine if the production function or recruitments drive stock dynamics, and which components of data inform the estimation of the recruitment deviates (Appendix Section 12.1.3). “Model-free” hindcasting based on the CPUE indices (Kell et al., 2021) was conducted to evaluate the model prediction skill and internal consistency (Appendix Section 12.1.4). Predictive performance from hindcasting, as determined by MASE (Hyndman and Koehler, 2006) has been suggested (Kell et al., 2021) as a way of *objectively* weighting models in an ensemble, and comparing model performance when the data and likelihoods vary between models. Model-free hindcasting based on the CPUE indices was conducted for all models in the model ensemble in order to develop an ensemble weight for each model. These weights were then used to test the sensitivity of management reference points to alternative model weighting schemes (equiprobable and hindcast weighted).

6.7 Stock assessment interpretation methods

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

6.7.1 Depletion and fishery impact

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

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

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

6.7.2 MSY based reference points

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

Fishing mortality-at-age (F_a) for the calculation of the yield based reference points analysis was determined as the mean over a recent period of time (2014–2018). We do not include the terminal year in the average as fishing mortality tends to have high uncertainty for the terminal data year of the analysis and the recruitments in this year are constrained to be the average over the full time-series, which affects F for the youngest age-classes.

6.7.3 Majuro and Kobe plots

For the standard yield analysis (Section 6.7.2), the fishing mortality-at-age, F_a , is determined as the average over some recent period of time (2015–2018 herein). In addition to this approach the MSY-based reference points (F_t/F_{MSY} , and SB_t/SB_{MSY}) and the depletion-based reference point ($SB_t/SB_{F=0[t]}$) were also computed using the average annual F_a from each year included in the

model (1952–2019) by repeating the yield analysis for each year in turn. This enabled temporal trends in the reference point variables to be estimated taking account of the differences in MSY levels under varying historical patterns of age-specific exploitation. This analysis is presented in the form of dynamic Kobe plots and “Majuro plots”, which have been presented for all WCPO stock assessments in recent years.

6.7.4 Reference points

The unfished spawning potential ($SB_{F=0}$) in each time period was calculated given the estimated recruitments and the Beverton-Holt SRR as outlined in Section 6.7.1. This offers a basis for comparing the exploited population relative to the population subject to natural mortality only. As the WCPFC has currently not yet adopted a formal LRP for billfish, both depletion based (i.e. $SB_{latest}/SB_{F=0}$) and yield based (i.e. SB_{latest}/SB_{MSY} & F_{recent}/F_{MSY}) reference points are presented.

Though there is no formal LRP for billfish, the WCPFC adopted $20\%SB_{F=0}$ as a limit reference point (LRP) for the tropical tunas, where $SB_{F=0}$ is calculated as the average over the period 2009–2018. Stock status was referenced against these points by calculating $SB_{latest}/SB_{F=0}$, where SB_{latest} is the estimated spawning potential in 2019 (Table 1).

The key fishing mortality reference point, F_{recent}/F_{MSY} (Table 1), is the estimated average fishing mortality over the assessment area in the recent time period (F_{recent} ; 2015–2018 for this stock assessment) divided by the fishing mortality producing MSY which is produced by the yield analysis and has been detailed in Section 6.7.2.

6.7.5 Ensemble influence

The new method used for constructing the model ensemble necessitates a new approach for identifying which fixed parameters, or model configurations correspond higher or lower values of the key management reference points. We used GLMs to regress reference point estimates (e.g. $SB_{latest}/SB_{F=0}$) against all fixed parameters (e.g. M_{ref}) and varying factor levels (e.g. choice of DWFN CPUE index to fit to) from all models in the ensemble. Extracting the coefficients from these GLMs allowed us to identify the magnitude and direction of influence of the different components of the ensemble. These results can be an important tool in identifying areas to focus research in order to reduce model uncertainty.

7 Model runs

Model development from the 2017 diagnostic case to the 2021 diagnostic case occurred incrementally via successive changes. These stepwise changes were done in order to identify the impact of each modification to the assessment outcomes. Changes made to the previous assessment model include: additional input data for the years 2016–2019, refreshing the biological assumptions, moving to an

index fisheries approach, updating the preparation of the size frequency data, and incorporation of new features to MULTIFAN-CL developed since the last assessment. These changes occurred in the following sequence of steps:

7.1 Developments from the last assessment

1. **2017 Diagnostic:** The 2017 SWPO swordfish diagnostic case.
2. **mfcl v2080:** Updating the MULTIFAN-CL executable to version 2.0.8.0, and aligning the default model configuration with recent WCPO stock assessments.
3. **update biology:** Updating the biological assumptions to be more representative of the potential parameter sets considered in the joint prior for the model ensemble, including defining reproductive potential as a function of length and assuming a Lorenzen M-at-age relationship.
4. **remove nominal effort:** Remove the influence of the nominal effort on the model fit for all fisheries without standardized effort, and remove the estimation of the newly redundant time-varying catchability parameters for these fisheries.
5. **index fishery:** Define separate fisheries for the extraction and index fleets.
6. **add NZ index:** Include the New Zealand standardized CPUE in the model.
7. **update AU & EU index:** Update the standardized CPUE for the Australian and European Union index fisheries.
8. **int. age structure $\sim M$:** Make the assumption that the model starts from an unfished equilibrium, where the starting age structure is just based on M .
9. **remove JP & TW idx:** Remove the Japanese and Chinese Taipei index fisheries from the model.
10. **estimate SDB:** Estimate the second growth variability parameter (σ_{Length}), which scales the standard deviations of length for each age-class as a log-linear function of the mean lengths-at-age.
11. **new extraction fleet:** Define a new fisheries structure for the extraction fleets where the EU longline fisheries in regions 1 & 2 are broken out from the others to define their catches in metric tons. The DWFN extraction fisheries in region 2C were combined since the move to an index fisheries structure no longer necessitated accounting for a catchability change in this fishery. The previous assessment noted conflict between the Australian length and weight composition data. The length-composition data for the Australian fisheries were also excluded based on discussions at the PAW which intimated that the weight frequency data which comes from port-sampling would be more representative.

12. *update to 2019*: Update all data in the model to the end of the current assessment period, 2019.
13. *5 kg wt bins*: Increase the resolution of the weight frequency data from 10 kg bins to 5 kg bins.
14. *Peatman composition*: Re-weight the composition data for the extraction fisheries so that it is more representative of fishery removals, and for the index fisheries so that it is more representative of the underlying stock.
15. *2021 diagnostic*: Small changes to selectivity including: grouping the selectivity for the EU longline fisheries, assuming constant selectivity after age class 16, removing the non-decreasing penalty on Fishery 3 selectivity, and removing the restriction that selectivity was 0 at first age for most fleets.

7.2 Sensitivity analyses

During the course of model development for the 2021 SWPO swordfish stock assessment, sensitivity models were run to explore the effects of changing the assumptions governing the population dynamics, fisheries dynamics, and weighting of data components in the likelihood. The presentation of the results focuses on the subset of analyses that were most influential to the stock assessment outcomes and/or those that were identified as questions of interest either in the previous assessment or the 2021 PAW (Hamer et al., 2021).

One-off sensitivities were conducted as single stepwise changes from the 2021 diagnostic case. The purpose of these sensitivity runs was not to provide absolute estimates of management quantities but to assess the *relative* change that resulted from the various assumptions. For the one-off sensitivities that also factor into the model ensemble, these can be interpreted as the marginal effect of that component conditioned on the other parameters being held at the values fixed in the diagnostic case.

7.2.1 Sex-structure

As mentioned earlier in the report, a parallel sex-structured model was developed to mirror the sex-aggregated diagnostic case. This model assumed sex-disaggregated biology (growth, length-weight relationship, natural mortality), defined reproductive potential as female maturity, and assumed a 50:50 sex-ratio at recruitment. Building upon the work of Takeuchi et al. (2018), the current model also included new capability within MULTIFAN-CL to properly share the fishery effort deviates across the two sexes. Selectivity was modeled as shared selectivity at length between the two sexes, to account for the observed sexually dimorphic growth. All catch and size composition data was shared evenly between the two sexes, therefore differences in the sex-specific catch and fishing mortality are driven by the relationship between the growth curves and the size composition data. This relationship was very sensitive to the assumed variance parameters (σ_{Age} & σ_{Length}), so a

sensitivity to this assumption was evaluated.

7.2.2 Alternative catch scenarios

Though current scientific evidence suggests that the SWPO swordfish stock is separated from stocks in the eastern Pacific Ocean and the north Pacific Ocean, there is uncertainty relates to the specific boundaries of the SWPO swordfish stock, and the level of connectivity that exists. This uncertainty is relevant in the context of the hot-spot of high catches around the north-east corner of the assessment region in recent years. The connectivity of swordfish between this high catch region and the rest of the SWPO has implications for the overall status of the stock. Five alternative catch scenarios were developed to test the sensitivity of the location of the northern and eastern-most assessment boundaries. Maintaining the same model structure as assumed in the diagnostic case:

1. ***add 5N***: Catches occurring up to 5°N, from 140°E to 130°W, were added to the model.
2. ***subtract 5S***: Catches occurring between the equator and 5°S, from 140°E to 130°W, were subtracted from the model.
3. ***subtract 10S***: Catches occurring between the equator and 10°S, from 140°E to 130°W, were subtracted from the model.
4. ***subtract 15S***: Catches occurring between the equator and 15°S, from 140°E to 130°W, were subtracted from the model.
5. ***subtract 165W***: Catches occurring between 165°W and 130°W, from 50°S to the equator, were subtracted from the model.

All catch scenarios assumed the same size composition data.

7.2.3 Alternative spatial structure

There are several reasons why it is desirable to model populations in a spatially explicit manner. Conducting a stock assessment in such a way can more realistically represent complex dynamics, and allow for the separation of selectivity into spatial availability (modeled by movement) and size/age based vulnerability to the gear. It can also partition the population into quasi-distinct components subject to different levels of fishing mortality. Despite these benefits, it is important to get the movement dynamics correct.

Though based on the best available science, the estimated movement rates between the two assessment regions remain very uncertain. In such a case, it may be more prudent and parsimonious to assume a single, well-mixed stock given that the assumptions made for movement are likely to be influential on the estimates of stock status. To explore this, a single region model that assumed the same fisheries structure as the diagnostic case was developed for SWPO swordfish.

7.2.4 Penalty on stock-recruit relationship

The variability of recruitment around the stock-recruit relationship is fixed in MULTIFAN-CL. Though the degree of recruitment variability may not have a large effect on reference points based on dynamic depletion (i.e. $SB/SB_{F=0}$), it can have an important impact on the scaling of the estimated stock-recruit relationship and the dependent yield based reference points. As a result, sensitivity to the assumed recruitment variability around the stock-recruit relationship was explored. Five levels of recruitment variability were considered, roughly equivalent to CVs of 2.2, 0.7, 0.5, 0.3 and 0.2.

7.2.5 DWFN CPUE scenario

In developing the index fisheries for the assessment model, it was hoped that simultaneously fitting to the multiple, overlapping index fisheries would impose continuity on the estimated population dynamics given the temporal discontinuities that existed for the indices. However, perhaps unsurprisingly, testing during model development indicated that the model had difficulty simultaneously fitting to the 13 different index fisheries, given some observed divergences between indices. This lack of coherence could in part be due to the paucity of operational gear characteristics available for standardization, and the non-stationarity in spatial sampling for the three DWFN fleets (Japanese, Chinese Taipei, and European Union). In order to minimize data conflicts, it was decided to only fit to indices from a single DWFN fleet at a time, and sensitivity to the assumption of DWFN fleet choice (or not including indices from a DWFN fleet) was explored. The European Union fishery was retained for the diagnostic case as its selectivity was the most reliably estimated given the comparatively larger sample size of size frequency data relative to Japan and Chinese Taipei.

7.2.6 CPUE average CV

As discussed in the previous section, it is likely that not all indices are equally reliable. Differences in their ability to accurately index the underlying stock could arise due to the choice of standardization approach, spatial sampling pattern and available standardization covariates. Given these potential differences, it is preferable to give greater influence to more “reliable” indices, especially in cases where conflicting trends exist. Iterative re-weighting of indices based on the predicted goodness of fit and input CVs can be used to address this issue. This approach is relatively straightforward to implement for a single model, however can sometimes result in the adjusted CVs alternating between multiple stable states from iteration to iteration. The lack of a robust stopping criteria and the need for analyst intervention makes implementation challenging in the context of a large model ensemble and it was not feasible to do so for the current assessment. In spite of this, it remained important to evaluate the sensitivity of stock assessment outcomes to the relative influence of each index, and sensitivity to the average CV attributed to each index as evaluated in the context of the diagnostic case. Thirty models were considered, and for each run the time-varying CVs for each fishery were rescaled to a mean value drawn from a uniform distribution $U(0.1,0.8)$.

7.2.7 Size composition scalar

The difficulties in assigning weighting to the size-frequency data were discussed in Section 6.4. To assess the sensitivity of model results to the weighting of these data, two sets of 30 models were considered. The first set of models held the length composition scalar value fixed at the level assumed in the diagnostic case, and allowed the scalar for the weight composition data for each model to be drawn from a uniform distribution $U(10,50)$. This was repeated for the second set of models except in this case it was the scalar for the weight composition data that was held constant.

7.2.8 Movement

As described in Section 7.2.3, the movement assumption is a critical component of spatially explicit stock assessment models. Using the diagnostic case model as a baseline, sensitivity of the stock status to the movement assumptions was explored. A set of 30 models was considered, where the bi-directional movement parameters were given by a random draw from the bivariate movement prior.

7.2.9 Steepness

Steepness is a particularly difficult parameter to estimate in stock assessment models, but if it is fixed in the model, the choice of value may have an important influence on yield based reference points used for management. Sensitivity to the steepness assumption was investigated by running a set of 30 models based on the diagnostic case where steepness for each model was fixed at a value drawn from the steepness prior.

7.2.10 Natural mortality

Similar to steepness, the choice of natural mortality can be highly influential on yield based reference points. Additionally, given that the stock is assumed to start at an unfished equilibrium, choice of M can also influence initial population scale. Sensitivity to the assumed M_{ref} value was investigated by running a set of 30 models based on the diagnostic case where M_{ref} for each model was fixed at a value drawn from the M_{ref} prior.

7.2.11 Growth

As experienced in recent tropical tuna assessments (McKechnie et al., 2017; Vincent et al., 2020); changes in the growth relationship, particularly L_2 , can have dramatic impacts on the estimated stock status. Sensitivity of the stock status to the growth parameters was explored. A set of 30 models based on the diagnostic case was considered, where the growth parameters were given by a correlated random draw from the joint prior.

7.2.12 Length-weight relationship

Though uncertainty in this relationship is not often considered, the length-weight relationship is important in reconciling fits between the length and weight composition data, as well as converting catch in numbers to biomass. Sensitivity of the stock status to the length-weight parameters was explored using a set of 30 models based on the diagnostic case, where the length-weight parameters were provided from a correlated random draw from the joint prior.

7.2.13 Reproductive potential

Uncertainty in the reproductive potential relationship could result in differential stock status. If the length at 50% maturity is increased, a greater reproductive burden is placed on larger individuals, which are typically thought to be fully selected by longline fisheries, and translate to more pessimistic stock status. Sensitivity of the stock status to the differences in the reproductive potential ogive was explored. A set of 30 models based on the diagnostic case was considered, where the parameters defining the reproductive potential ogive were specified by a correlated random draw from the joint prior.

7.3 Model ensemble

The model ensemble used as the basis for management advice for the SWPO swordfish stock was constructed as the combination of a factorial grid of discrete factors, overlaid on correlated parameter sets drawn from the joint prior. Relevant factors, and dimensions of the joint prior were decided based on the sensitivity analyses and feedback received as a part of the 2021 PAW (Hamer et al., 2021). The factorial component of the ensemble consisted of four “axes”: a DWFN CPUE index scenario axis with four levels (Japanese index, Chinese Taipei index, European Union index, or no DWFN index), a recruitment variability axis with three levels ($CV \in \{0.7, 0.5, 0.3\}$), a growth axis with two levels (informative and uninformative prior on t_0), and a natural mortality axis with two levels accounting for the two alternative approaches used to define M_{ref} . This resulted in 48 unique factorial combinations.

Model testing indicated that the results were quite sensitive to the choice of movement parameters, with a large proportion of runs having to be discarded due to implausible biomass estimates or poor model convergence. Even though Ducharme-Barth and Vincent (2021) showed that a model ensemble with as few as 30 models could effectively characterize the same level of uncertainty as a much larger ensemble, this efficiency would likely not be realized in the current case given the failure rate indicated in the movement sensitivity. To account for this, the unique 48 factorial combinations were replicated eight times resulting in a starting model ensemble of 384 models, where 384 was a balance between a large starting ensemble size and computational tractability. For each of the 384 models in the ensemble a unique parameter set was drawn from the joint prior.

8 Results

Development of the 2021 SWPO swordfish diagnostic case, model ensemble and associated diagnostics was a significant computational undertaking, totaling close to 33,350 hours of computation time and 9,300 total runs. For the sake of brevity, only the most relevant results are discussed and presented in this report.

8.1 Consequences of key model developments

The progression of model development from the 2017 diagnostic case model to the 2021 diagnostic case model is outlined in [Section 7.1](#) and results are displayed with respect to both the change in spawning potential (SB) and the level of depletion of spawning potential relative to the unfishied condition ($SB_{F=0}$). A summary of the consequences of this progression through the models is as follows:

1. **2017 Diagnostic:** The 2017 SWPO swordfish diagnostic case.
2. ***mfc1 v2080*** ([Step 2](#); [Figure 8](#)) – Updating to the v2.0.8.0 MULTIFAN-CL executable resulted in a slightly more pessimistic estimated depletion, due to marginally higher initial and lower terminal spawning potential estimates.
3. ***update biology*** ([Step 3](#); [Figure 8](#)) – Relative to the previous step, the depletion estimate was marginally more optimistic following updating of the biological assumptions (growth, natural mortality, length-weight, and reproductive potential). Spawning potential was estimated to be noticeably lower, likely due to a shift in the reproductive potential ogive to larger individuals.
4. ***remove nominal effort*** ([Step 4](#); [Figure 8](#)) – Removing the influence of the nominal effort on the model resulted in a lower estimates of spawning potential and a more pessimistic level of depletion.
5. ***index fishery*** ([Step 5](#); [Figure 8](#)) – Transitioning to the index fishery approach resulted in the lowest estimates of spawning potential and the most pessimistic depletion estimates. Such a large change was unexpected given that the underlying data (standardized CPUE & size composition) did not materially change. This large difference is likely attributed to a change in the selectivity curves associated with the standardized CPUE, particularly for the DWFN fishery in region 2C. When this fishery was split into equivalent extraction and index fisheries, the standardized CPUE went from being associated with a dome shaped selectivity curve to a non-decreasing selectivity curve which likely drove the more pessimistic estimate.
6. ***add NZ index*** ([Step 6](#); [Figure 8](#)) – Including the New Zealand standardized CPUE as an index fishery led to higher estimates of spawning potential and more optimistic depletion, likely due to the model attempting to fit the increasing trend seen in the New Zealand index from 2004 – 2015.

7. *update AU & EU index* (Step 7; Figure 8) – Updating the Australian index in the model resulted in negligible changes to stock status, given the consistency in standardization approach and trend between the two assessments. Updating the European Union index using the VAST modeling approach resulted in standardized CPUE trend that tended to increase from 2004 – 2015, whereas the previous index was largely flat over the same period. This again resulted in an upwards correction of both biomass and depletion.
8. *int. age structure $\sim M$* (Step 8; Figure 8) – Assuming that the stock was at an unfished equilibrium at the start of the model period (1952) did not meaningfully change terminal depletion, or terminal spawning potential. However, this change did result in lower initial population numbers, and lower initial spawning potential.
9. *remove JP & TW idx* (Step 9; Figure 8) – As discussed earlier in Section 7.2.5, the model had particular difficulty simultaneously fitting to the three updated DWFN indices, resulting in some odd estimated selectivity curves and poor fits to the length frequency data. The decision was made at this step to only attempt to fit to a single DWFN index at a time in order to minimize data-conflict but to consider the choice of index fishery to fit to as a component of the model ensemble. Retaining only the European Union index, on the basis of producing consistent fits to its length frequency data had implications for the spawning potential trajectory prior to the mid-1990s. As there is no size-frequency data in the model prior to this point, all of the trend in the spawning potential in the early part of the model period was driven by the recruitment variability needed to fit the Japanese index, and to a lesser degree the Chinese Taipei index. Removing the Japanese and Chinese Taipei CPUE indices from the model resulted in pre-1990 recruitment estimates that did not deviate greatly from the mean level of recruitment predicted under the assumed stock-recruit relationship. This resulted in a larger initial population size, and a more optimistic level of depletion.
10. *estimate SDB* (Step 10; Figure 8) – In the process of attempting to simultaneously fit the three DWFN indices a number of model configurations were considered in attempts to improve the estimation of the selectivity curves and the fit to the associated length composition data. One of the configurations considered was to relax the assumption of constant variability around the growth curve as a function of age, and to allow variability to scale as a function of length via the estimation of the σ_{Length} parameter. The estimation of this additional parameter was retained on the basis of stepwise AIC model selection, and it resulted in marginally lower estimates of biomass and depletion.
11. *new extraction fleet* (Step 11; Figure 8) – There were three notable changes to the extraction fleet definitions from the previous stock assessment: 1) defining catches for the European Union fishery in terms of metric tons, since this is their original unit of record; 2) aggregating the “early” and “late” DWFN fisheries in region 2C now that the use of index fleets renders a catchability break meaningless for the extraction fisheries; and 3) the exclusion of the Australian length composition data. These changes resulted in a higher

estimate of initial biomass which translated to a more pessimistic terminal depletion relative to the previous step.

12. *update to 2019* (**Step 12**; Figure 8) – All input data were updated through the 2019 calendar year to correspond to the new stock assessment period. The additional four years of data shifted the spawning potential trend down, but resulted in identical depletion through 2015. The standardized CPUE indices showed a decline following 2015, and both depletion and spawning potential declined accordingly.
13. *5 kg wt bins* (**Step 13**; Figure 8) – Increasing the resolution of the weight frequency data in order to provide more information for the estimation of recruitments did not meaningfully change the stock status. Both spawning potential and depletion shifted up slightly.
14. *Peatman composition* (**Step 14**; Figure 8) – Re-weighting the composition data for the extraction fisheries so that it is more representative of fishery removals, and for the index fisheries so that it is more representative of the underlying stock resulted in a slight reduction of the biomass and depletion estimates slightly lower.
15. *2021 diagnostic* (**Step 15**; Figure 8) – A number of small changes to selectivity were considered on the basis of improving fits to the size-composition data including: grouping the selectivity for the EU longline fisheries, assuming constant selectivity after age class 16, removing the non-decreasing penalty on Fishery 3 selectivity, and removing the restriction that selectivity was 0 at first age for most fleets. These changes were all retained on the basis of stepwise AIC model selection, however both spawning potential and depletion were essentially unchanged from the previous step.

Relative to the diagnostic case of the previous assessment, the diagnostic case for the current assessment produced similar estimates of depletion for the SWPO swordfish stock through the late-2000s. Beginning in 2010 the two models diverged, with the current diagnostic case estimating higher terminal spawning potential and more optimistic levels of depletion.

8.2 Diagnostic case: Model fits

This section discusses model fits to the main input data components for the diagnostic case model, defined by the final step in the stepwise model development (model *2021 diagnostic*; Figure 8), the final step described in Section 8.1. Additional diagnostics (likelihood profiles, retrospectives, ASPM, hindcasting, and residual plots) are presented in Appendix Sections 12.1.1 – 12.1.5.

In terms of model convergence, the 2021 SWPO swordfish diagnostic case model achieved both a low final gradient (7.12×10^{-7}) and a positive definite Hessian solution.

8.2.1 Catch data

High penalties were applied to the catch data for all fisheries and so the catch is fit very closely, with a SD of residuals on the log scale of 0.002 (Figure 9).

8.2.2 Standardized CPUE

There was substantial temporal variability in the standardized CPUE indices used in the assessment, but despite this, the model-predicted CPUE fit the indices very well (Figure 10), with low RMSE (< 0.3) for all index fisheries. In general, the model captures most of the seasonal variation and is generally able to capture the long-term trend for all fisheries, with few estimated outliers (Figure 11). However, there is some residual trend as the model tends to overestimate the decline in the European Union index in terminal years, and underestimate the initial decline in the Australian index. There is indication that the model is over-fitting the Australian index as the residual CV is less than the mean input CV. This is likely due to the reinforcing signal from the large numbers of weight frequency samples for this fishery which show a similar temporal pattern in median observed size.

8.2.3 Weight frequency data

The weight frequency is generally well fit by the model, particularly for fisheries such as the Australian longline (Fisheries 4 & 14) where there are substantial numbers of samples (Figure 12). However, the model does appear to underestimate the modal size of encountered individuals for the two PICT fisheries in region 2 (Fisheries 12 & 13). Since these fisheries have substantially more length frequency data, the model appears to sacrifice a fit to the weight frequency data in favor of the length frequency data. This indicates a data-conflict driven either by model misspecification (e.g. inappropriate length-weight relationship, or inappropriate processed length-standard length/processed weight-standard weight conversion factors), or differences in the sampling programs that generated the length and weight observations. This mis-fit to the weight frequency data for Fisheries 12 & 13 is even more apparent in the context of the temporal trend in mean weight (Figure 13). For the other fisheries with observed weights, the model is able to better track the temporal trend in average size. Additionally, we see that for the Australian fisheries, the trend in mean size largely mirrors the trend in the standardized index.

8.2.4 Length frequency data

Fits to the length frequency data were generally found to be satisfactory when sample sizes were high (Figure 14). However, fits to the DWFN longline fisheries tended to be poor, especially for those operating in region 1 C & S. For these fisheries the model had difficulties fitting a large number of small observed fish, smaller than the assumed L_1 , which could indicate a mis-specification in the growth curve or selectivity specification. Tracing the source of these small fish indicates that they came from a few observer records in recent years. Given the low levels of longline observer coverage

for DWFN fleets and ephemeral presence in the record of observations, these small fish may not be representative of the overall fishery selectivity patterns.

Few fisheries had enough consistency in temporal length frequency sampling to discern real trends (Figure 15). While the overall trend appears to have been captured reasonably well for the New Zealand fisheries, the predicted temporal trend for the PICT Fishery in Region 2N appears to be biased high, particularly in recent years. As mentioned in the context of the fits to the weight frequency data, there appears to be a conflict between the two types of size frequency data for this particular fishery.

8.3 Diagnostic case: Estimated quantities

8.3.1 Selectivity

As mentioned previously in Section 6.3.1, selectivity curves were relatively unconstrained, though most were estimated to fully select individuals of the oldest ages (Figure 16). Given the rapid rate of growth leading to large size at recruitment to the modeled population, many fisheries estimated moderate levels of selectivity even for the youngest age. Additionally, though a selectivity time-block was estimated for Fishery 4 (Australian longline), it does not appear that the hook-type change resulted in a meaningful change in selectivity given the similarity in the estimated curves.

8.3.2 Biomass

Trend and magnitude for the estimated biomass are largely similar to the previous stock assessment (Figure 17). The stock is predicted to slowly decline over the first 40 years of the model period. Beginning in the mid-1990s, biomass and spawning potential decrease rapidly to a low point at the end of the 2000s. The stock showed signs of recovery up to the mid-2010s before declining through to the end of the model period. Additionally, the magnitude of biomass and spawning potential is larger in region 2 than region 1. As mentioned in Section 5.4.2 this result is achieved independent of the index fisheries, and is driven primarily by the magnitude of catches between the two regions. However, the larger population levels in region 2 are corroborated by the VAST CPUE standardization models used to standardize the early period Japanese and Chinese Taipei data (Ducharme-Barth et al., 2021b).

8.3.3 Recruitment

The overall trend in estimated recruitment was fairly flat though this is largely a product of the lack of informative data in the early part of the time period and the decision to estimate mean levels of recruitment for this period (Figure 17 & 18). However, beginning in the mid-1990s, the influence of the index and composition data leads to a period of high variability with an average recruitment slightly higher than the early period. This recent period of higher recruitment may contaminate our interpretation of yield based reference points given that the yield simulations are based on a

stock-recruit relationship that is defined over the entire model period. As such, $SB_{latest}/SB_{F=0}$ may be a more robust measure of stock status.

The proportions of recruitment in each region are freely estimated parameters, and region 1 is estimated to have a higher level of recruitment for almost all model years. This however still translates to a larger biomass in region 2 given the asymmetric movement assumptions in the diagnostic case model.

8.3.4 Fishing mortality

Though total fishing mortality was estimated to be almost zero at the start of the model period, a reflection of the almost non-existent initial catches, it gradually increased through to the mid-1990s before dramatically increasing through the 2000s and leveling off in the last decade (Figure 19). This dramatic increase in total fishing mortality appears to be driven mainly by the increase in region 2 over the same period. Fishing mortality was always estimated to be comparatively higher in region 1 given the larger initial catches and smaller estimated population size.

8.4 Diagnostic case: Further analyses of stock status

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

8.4.1 Fishery impacts

Fishery impact is measured at each time step as the ratio of the estimated spawning potential relative to the spawning potential that is estimated to have occurred in the historical absence of fishing. This is a useful quantity to monitor, as it can be computed both at the region level, and for the SWPO as a whole. This information is plotted in two ways, firstly as the fished and unfished spawning potential trajectories (Figure 20), and secondly as the depletion ratios themselves (Figure 21). The latter is relevant for the agreed reference points and example plots of these values are displayed for the diagnostic case model.

The diagnostic case model estimates that steady declines have occurred in both model regions (Figure 22). In the early period, fishing impacts were greatest in region 1 as DWFN fleets expanded south across the equator and into the Coral and Tasman seas. The bulk of this impact was concentrated in the middle latitudes (20°S – 40°S). The emergence of domestic longline fisheries for Australia and New Zealand had a large fisheries impact in region 1 at the turn of the millennium. In region 2, and to a lesser degree region 1, large impacts have been made by an increase in equatorial (0 – 10°S) fishing activity by DWFN fleets such that this is the dominant sector in terms of impact

in recent years. Despite developing and increasing capacity in the last two decades, PICT longline fisheries do not have a meaningful impact on the SWPO swordfish stock. Fishing south of 40°S is also estimated to have a negligible impact on the stock.

It is worth noting that the general conclusions drawn from this analysis are conditioned on the historical swordfish targeting, selectivity, and catchability patterns for these fisheries. These observed impacts may not hold into the future if fisheries shift target species resulting in more directed or by-caught swordfish (e.g. switch from albacore tuna to bigeye tuna targeting, given the similarities in vertical diving/feeding behavior between bigeye tuna and swordfish).

8.4.2 Dynamic Majuro and Kobe plots

The section summarizing the model ensemble (Section 8.6) presents terminal estimates of stock status in the form of Majuro and Kobe plots. Further analyses can estimate the time-series of stock status in the form of dynamic Majuro or dynamic Kobe plots, the methods of which are presented in Section 6.7.3. An example of the dynamic Majuro plot for the diagnostic case is presented in Figure 23. At the start of the assessment period, the SWPO swordfish stock was predicted to essentially be in a virgin unfished condition. The F/F_{MSY} was estimated to be nearly zero indicating that initial fishing pressure at the start of the model period was negligible. Each of these reference points progressively shifted towards the overfishing and overfished definitions over the remaining period. The diagnostic case model shown never reaches $20\%SB_{F=0}$ or an F/F_{msy} of one. These results are mirrored in the dynamic Kobe plot (Figure 24).

8.5 Multimodel inference: Sensitivities

8.5.1 Sex-structure

Sensitivity of the stock status to the sex-structure of the model was in itself sensitive to the assumptions made in terms of the variability around the growth curve. A mini-factorial grid was created to explore this where all combinations of $\sigma_{Age} \in \{5, 10, 15, 30\}$ and $\sigma_{Length} \in \{0, 0.1, 0.2, 0.3\}$ were tested. Results showed that stock status was robust to the choice of σ_{Length} but not σ_{Age} (Figure 25). When $\sigma_{Age} = 5$, stock status in terms of $SB/SB_{F=0}$ was virtually identical between the sex-disaggregated and sex-aggregated models. As σ_{Age} increased stock status became increasingly pessimistic.

Investigation of the estimated sex-specific selectivities and fits to the sex-aggregated size composition split by sex for the most pessimistic scenario ($\sigma_{Age} = 30$) indicated the large growth variability skewed the selectivity estimation to the larger females for fisheries that typically caught small fish (more likely males given dimorphic growth). This resulted in very low estimated selectivity for males given the shared selectivity at length and very poor fits to the “male” size composition data. Given this estimated selectivity curve the model assumed males to be “quasi”-invulnerable, resulting in almost all of the sex-aggregated catch, and thus fishing mortality, to be attributed to females.

This result seemed unlikely given that these fisheries tended to catch smaller individuals.

These results indicate that if variability in growth is assumed to be low, sex-aggregated and sex-disaggregated models produce comparable results despite the large differences in growth. Assuming a larger variability in growth skewed the selectivity estimation towards larger sizes, leading to more pessimistic outcomes. To further explore the impacts that this has on management advice, an equivalent sex-disaggregated model ensemble was developed. The results of which are presented in Appendix [Section 12.3](#).

8.5.2 Alternative catch scenarios

Estimated initial population scale was sensitive to the assumed catch scenarios ([Figure 26 a](#)). Scenarios which assumed greater levels of total removals from the system (*add 5N*), relative to the diagnostic case, showed higher estimated population scale. Conversely, as the level of fisheries removals decreased (*subtract 5S - subtract 165W*) population scale was estimated to be lower. Model convergence also appeared sensitive to the assumed catch scenario and level of estimated population scale. Models with the three lowest estimated population scales (*subtract 10S*, *subtract 15S*, and *subtract 165W*) all failed to achieve a positive definite Hessian solution. In terms of the estimated depletion, subtracting catches occurring north of 5°S & 10°S did not greatly change the impact from the 2021 diagnostic case ([Figure 26 b](#)), though only considering catches occurring south of 15°S lead to a noticeably more optimistic status. Assuming that catches occurring just outside of the SWPO swordfish assessment region (to 5°N) were from the same stock resulted in a much more pessimistic outcome. The removal of high catches east of 165°W unexpectedly resulted in a more pessimistic depletion trajectory. This was driven by the more pessimistic estimated depletion for region 1 given that catches in this region were unchanged, but the total population scale was estimated to be lower.

8.5.3 Alternative spatial structure

The model appeared to be fairly robust to simplifying the spatial structure of the stock assessment model. The estimated spawning potential trajectory was shifted lower, and resulted in a more pessimistic estimation of fisheries depletion ([Figure 26: *single region*](#)). However, both estimates were well within the range of stock statuses estimated as a part of the movement sensitivity ([Section 8.5.8](#)).

8.5.4 Penalty on stock-recruit relationship

Model estimates of both spawning potential and fisheries depletion over the last 25 years were robust to the choice of penalty on the stock-recruit relationship ([Figure 27](#)), as were depletion based reference points. However lowering the penalty, relative to the diagnostic case level (CV = 0.5) allowed for a greater variability in recruitment and trend in the estimated recruitment during the early, uninformative period ([Figure 28](#)). This resulted in an estimated stock-recruit relationship

with lower asymptotic recruitment. Accordingly, yield based reference points were sensitive to the assumed penalty value.

8.5.5 DWFN CPUE scenario

Estimated biomass and depletion appeared quite variable to the choice of DWFN index fishery included in the model with the clear outlier being the Japanese index (Figure 29). This is most notably manifested by the highly variable patterns in recruitment that are estimated during the early period which are needed to fit the different observed patterns for the Japanese and Chinese Taipei indices (Figure 30). Recruitment in recent years appears to be relatively insensitive to the choice of DWFN index apart from the Japan index, indicating that information from the domestic longline indices (Australia/New Zealand) or the size frequency data are driving the recruitment estimates.

8.5.6 CPUE average CV

Model outputs, biomass and depletion, appeared relatively insensitive to the relative influence of either the Australian or European Union index. In the context of the diagnostic case model, the New Zealand index was the most influential. This is most obvious by looking at model outputs according to the ratio of the mean Australian index CV to the mean New Zealand index CV (Figure 31). When this ratio is large it means that the model fits more closely to the New Zealand index given that a smaller CV translates to a larger penalty. When the ratio is small the inverse occurs, and the Australian index is fit more closely. When the New Zealand index has a greater influence, the model has to estimate a larger, more productive population in order to match the almost two-fold increase in the New Zealand standardized CPUE from 2004 - 2014. This larger estimated population, given that input catch levels were the same across all models, translates into more optimistic levels of depletion. Models with more influence placed on the Australian index conversely resulted in more pessimistic stock status.

8.5.7 Size composition scalar

Sensitivity to the size composition scalars for length and weight confirmed the conflicting signals indicated by these two data types. When the length composition data were assumed to have a greater influence on the likelihood (lower scalar), spawning potential was estimated to be higher and depletion was more optimistic (Figure 32). However, when the influence of the weight frequency data was evaluated, the opposite pattern emerged. A lower weight frequency scalar (greater influence on the likelihood) resulted in a lower estimated spawning potential, and more pessimistic depletion (Figure 33). It is also worth noting that the model seemed to have difficulty converging for models that estimated lower biomass levels, similar to what was noted in Section 8.5.2.

8.5.8 Movement

Parameterizing models using correlated pairs of movement rates from the joint prior developed by [Patterson et al. \(2021\)](#) resulted in a very broad range of both estimated spawning potential and fisheries depletion. Uncertainty in movement rates were asymmetrical, with the rate of movement from the South Pacific into the Tasman region predicted to have a much greater variance and strongly right skewed distribution. Many of these large values for movement from the South Pacific into the Tasman region simply could not be reconciled by the data and either resulted in implausible estimates or model convergence failure. Some models with larger rates of movement from the South Pacific into the Tasman region were able to converge, however these models also assumed larger movements from the Tasman into the South Pacific region. As seen in [Figure 34](#), in order to reconcile the larger ratio for the movement rate from the South Pacific into the Tasman region given the available data, the model had to estimate high levels of spawning biomass. These larger spawning potential levels and coincidentally more optimistic levels of depletion are not necessarily plausible in the context of the relative biomass scale between regions.

Relative biomass scale between regions was unconstrained in the model, so the model had considerable flexibility to manipulate biomass and recruitment in order to make the various movement scenarios work given the available data. There was almost perfect, 1:1 correlation between the ratio of spawning potential between regions ($\overline{SB_{Tasman}}/\overline{SB_{S.Pacific}}$) and the ratio of movement from the South Pacific into the Tasman region. However, it would seem unrealistic for swordfish biomass to be greater in the Tasman region given the discrepancy in region size. Indeed a larger South Pacific region biomass is supported by the VAST CPUE standardization models used to standardize the early period Japanese and Chinese Taipei data ([Ducharme-Barth et al., 2021b](#)). Additionally, the 2017 stock assessment ([Takeuchi et al., 2017](#)) also estimated biomass and spawning potential to be larger in the South Pacific rather than Tasman regions and this result was considered reasonable by SC13. Constraining the interpretation of results accordingly, we see that movement scenarios where a greater *proportion* of individuals move from the South Pacific into the Tasman region than from the Tasman into the South Pacific are not supported by the other sources of data in the stock assessment ([Figure 35](#)), and that the remaining scenarios tend to estimate lower spawning potential and more pessimistic depletion. This does not mean that the *net movement* of individuals is similarly constrained. In fact, the net movement of individuals from the South Pacific region into the Tasman region can still be positive given that; a) the population is estimated to be larger in the South Pacific region than the Tasman region, and b) more recruitment is predicted to occur in the Tasman region.

8.5.9 Steepness

Model outputs appeared insensitive to the assumed value for steepness, though models assuming a higher steepness values resulted in marginally lower spawning potential estimates and marginally more optimistic estimates of terminal depletion ([Figure 36](#)). Inspection of the model total likeli-

hoods showed that there was very little information in the data to support one particular steepness value over another, though the model with the best objective function value assumed the highest steepness value. However, models assuming higher steepness values showed convergence issues which manifested as the lack of a positive definite Hessian solution.

8.5.10 Natural mortality

Assumed natural mortality had a clear impact on the estimated trends of depletion and spawning potential. Given that the population was assumed to be at an unfished equilibrium at the start of the model period, lower values of M_{ref} corresponded to larger less productive stocks which showed steep declines in spawning potential and more pessimistic depletion trajectories (Figure 37). Higher values of assumed M_{ref} resulted in more gradually declines in spawning potential and more optimistic depletion trajectories. Given the other sources of data included in the stock assessment, there appears to be some information available in the CPUE indices, included in the diagnostic case, to inform an estimate of M_{ref} (Figure 38). The available data appear to support a value of $M \sim 0.35\text{--}0.45$ which is towards the upper bound of the range of M_{ref} in the joint prior, and more in line with the maximum age based prior.

8.5.11 Growth

Estimated fishing depletion did not appear to be overly sensitive to changes in the three fixed growth parameters (L_1 , L_2 , or k) considering the range of each parameter in the joint prior (Table 4). However, the magnitude of estimated spawning potential appeared to be quite sensitive to the assumed parameter values. Of the three parameters, L_1 showed the clearest trend between the assumed value and stock status (Figure 39). L_1 was also the parameter that showed the most contrast in total likelihood over the range of values considered (Figure 40) with the most likely value ~ 105 cm. The data did not appear to support the use of an informative t_0 prior as the L_1 values estimated from that model tended to be less than 100 cm. Large values of L_1 typically correlated to model runs with the lowest estimated spawning potential and the most pessimistic depletion. This is likely due to the large size at initial recruitment, meaning that recruits were immediately subject to higher levels of fishing pressure.

8.5.12 Length-weight relationship

Model outputs in terms of both spawning potential and depletion appeared insensitive to the assumed values for α & β (Figure 41).

8.5.13 Reproductive potential

Model outputs in terms of both spawning potential and depletion appeared insensitive to the different reproductive potential ogives developed from correlated parameter draws from the joint prior (Figure 42).

8.6 Multimodel inference: Model ensemble

A hybrid ensemble of 384 initial models was considered as the basis for providing management advice for SWPO swordfish. As seen from the one-off sensitivities (Section 8.5), several dimensions of the model ensemble either propagated a lot of uncertainty into the estimated stock assessment outcomes or resulted in models that were poorly determined (i.e. lack of convergence as determined by a non-positive definite Hessian solution).

8.6.1 Ensemble influence

Generalized linear models were used to regress reference point estimates against all fixed parameters and varying factor levels from all 384 models in the ensemble in order to identify which parameters were most influential to key management reference points. Regression against $SB_{latest}/SB_{F=0}$ indicated that the most influential components of the ensemble were M_{ref} , the two movement rate parameters, inclusion of Japanese CPUE, and to a lesser extent the scalar on the length frequency and average CV for the Australian CPUE index (Figure 43). Models with a large M_{ref} , high movement rate from the South Pacific into the Tasman region, and/or that included the Japanese CPUE index tended to show more optimistic levels of depletion, whereas models with high rates of movement from the Tasman to the South Pacific region, a reduced influence of the length frequency data on the likelihood, and greater influence on the Australian CPUE index typically showed more pessimistic depletion levels.

Repeating this analysis for the yield based reference points, F_{recent}/F_{MSY} and SB_{latest}/SB_{MSY} identified a similar set of influential components. For F_{recent}/F_{MSY} , the most influential components of the ensemble were M_{ref} , the two movement rate parameters, steepness, and to a lesser extent the scalar on the length frequency and average CV for the Australian CPUE index (Figure 44). Larger values for steepness, M_{ref} , and movement rate from the South Pacific into the Tasman region contributed to a reduced chance of overfishing according to F_{recent}/F_{MSY} . Higher movement from the Tasman into the South Pacific region, reduced influence of the length frequency data, and increased influence of the Australian CPUE index led to higher chance that overfishing was occurring according to F_{recent}/F_{MSY} .

In terms of SB_{latest}/SB_{MSY} , the same components of the ensemble were influential and led to similar effects, except in the risk of the stock being classified as overfished according to SB_{latest}/SB_{MSY} (Figure 45). However, additional factors were also found to have a statistically significant effect on SB_{latest}/SB_{MSY} : inclusion of either the Japanese or Chinese Taipei CPUE typically led to the stock being estimated to have a high spawning potential relative to SB_{MSY} , while assuming high penalty on the stock-recruitment relationship resulted in lower values of SB_{latest}/SB_{MSY} .

This GLM approach for identifying which components of the ensemble are most influential to management quantities can also be used to identify which components of the ensemble are more likely to result in poorly estimated results (e.g. large maximum gradient component or non-positive definite

solution), or implausible biological estimates. In the context of this ensemble, all models converged to a fairly small gradient ($\leq 7.59 \times 10^{-6}$), however certain combinations of growth parameters tended to result in a greater chance of a non-positive definite Hessian solution (Figure 46). These were typically models that drew parameters from the component of the joint prior that assumed an informative, zero-centered prior on t_0 , and/or those that assumed a small t_0 , large k and large L_∞ . Similarly, aspects of the ensemble leading to unrealistic biomass estimates could also be identified. As discussed in Section 8.5.8, it does not appear reasonable for the model to estimate a larger average spawning potential in region 1 than region 2. Unsurprisingly, the two movement rate parameters (high movement from the South Pacific into the Tasman, and low reciprocal movement from the Tasman back to the South Pacific) are the main drivers behind models with this result (Figure 47). Additionally, given the stock is assumed to start from an un-fished equilibrium, and was subject to 68 years of direct and in-direct industrial fishing pressure, it is reasonable to assume that the latest spawning biomass estimate is less than it was in 1952. Models that fail to meet this assumption are again those with unrealistic assumed movement, and those that fit to either the Japanese or Chinese Taipei CPUE indices (Figure 48). Both of these indices showed large increases in the early model period (~ 1970 for Japan and ~ 1980 for Chinese Taipei), which the model interpreted as a basis for increasing biomass. However, neither of these increases is seen in the index of the other DWFN fishery operating in the same time period. This conflict between the two DWFN indices in the early period indicates that either one index or the other (or both) is not representative of the overall stock dynamics. This lack of coherence in the contemporaneous indices could be either due to picking up a localized trend in abundance due to a bias in spatial sampling, or poor standardization due to a lack of operational covariates.

8.6.2 Model retention

A post-hoc filter was applied to remove models that showed clear issues in the estimation or unrealistic estimates of spawning potential. The following steps were applied, and the effect of applying each filter to management metrics (Figure 49) or components of the joint prior (Figure 50) are summarized.

1. $\overline{SB_{Tasman}}/\overline{SB_{S.Pacific}} < 1$: Applying this filter removed 232 models from the ensemble, with 152 models remaining. Removed models typically estimated larger spawning potential, lower values of fishing mortality, and more optimistic levels of terminal depletion. These were also models that usually assumed high rates of South Pacific to Tasman movement, and low rates of Tasman to South Pacific movement.
2. $SB_{2019}/SB_0 < 1$: Applying this filter⁷ removed an additional 42 models (one model fitting to the European Union index, 27 models fitting to the Japanese index, and 14 models fitting the Chinese Taipei index), with 110 models remaining. Removed models tended to have similar

⁷Omitting this filtration step would result in a similar final model ensemble, as models that fail to meet this criteria are largely removed in the following steps.

characteristics in terms of management reference points as those removed in the previous step.

3. $SB_0 < 250,000$ mt: Applying this filter removed a single model (109 remaining) which had a terminal spawning potential estimate $> 2.5\times$ the maximum estimated SB_0 from the previous stock assessment. This filter did not appreciably change the distribution of the estimated management quantities or of the retained component of the joint prior.
4. *Remove models with effort deviates on bound*: Applying this filter removed 4 models, with 105 models remaining. This filter did not appreciably change the distribution of the estimated management quantities or of the retained component of the joint prior.
5. *Remove models with selectivity parameters on bound*: Applying this filter removed 50 models, with 55 models remaining. This filter did result in a marginally more pessimistic distribution of depletion, and retained component of the joint prior that contained growth parameter combinations which predominantly came from the model that assumed an uninformative prior on t_0 .
6. *Retain models with $\sigma_{Age} < 45$ cm*: A large standard deviation around the growth curve was needed in order to accommodate the discrepancy between the sex-disaggregated estimates of L_2 and the sex-aggregated estimate of L_2 . An upper limit of 45 cm for σ_{Age} was imposed. This value was broad enough to allow for both the median female L_2 and the median male L_2 to be within one standard deviation from the median sex-aggregated L_2 . This step removed 14 models, with 41 models remaining. Removed models were exclusively those that assumed an informative, zero-centered prior for t_0 .
7. *Remove models that do not have positive definite hessian solution*: Models that did not satisfy the convergence criteria as defined by achieving a positive definite hessian solution were removed. This resulted in 16 models being removed for a final ensemble of 25 retained models. This filter did not appreciably change the distribution of the estimated management quantities or of the retained component of the joint prior.

This post-hoc filtering process resulted in 25 models being retained for analysis. After the first two filters were applied, the distributions for the management metrics appeared robust to the remaining filtration steps. The elements of the joint prior that were pruned most heavily were: parameter combinations with a high movement from the South Pacific to the Tasman, low movement from the Tasman to the South Pacific, parameters from the model that assumed an informative, zero-centered prior for t_0 , and models that fit to either the Japanese or Chinese Taipei indices.

8.6.3 Uncertainty characterization: Model & Estimation uncertainty

Uncertainty in management quantities is characterized as the combination of model uncertainty (different aspects of the joint prior), and the statistical estimation uncertainty from each of the

retained models in the joint prior. A parametric bootstrap of the dependent variables, using the estimated variance from each model’s covariance matrix, was used to generate a distribution of 500 predictions for each management quantity of interest. These distributions were combined across models to simultaneously capture the model and estimation uncertainty. Though uncertainty in management quantities appear larger than in previous assessments, it is important to emphasize that uncertainty in previous assessments was likely an under-representation given that it did not present the statistical uncertainty across the ensemble of models. This is clearly shown in [Figure 51](#).

Large uncertainty aside, there is a clear trend that stock status, as determined by $SB/SB_{F=0}$, declined sharply from the mid-1990s to the late 2000’s ([Figure 52 a](#)). Relative to that low point, the current status is estimated to be similar but the prevailing trend over the last five years is one of decline. This aggregate trend in stock status is largely repeated at the regional level, though there is greater uncertainty in the stock status in region 1 than region 2 ([Figure 52 b](#)). At this stage it is difficult to identify if the recent decreasing trend in spawning potential and depletion will continue into the future. The longer term decline in total biomass appears to level-off in recent years but this may be due to the two terminal years of recruitments being fixed at the mean level ([Figure 53](#)). However, prior to being fixed, median recruitment did appear to increase slightly from a low point ~ 2015 .

The SWPO swordfish stock lacks formal, agreed upon reference points. Accordingly, stock status is summarized in terms of both depletion and maximum-yield based reference points ([Table 5](#)). On average, the SWPO stock is not estimated to be in an overfished state according to either $SB_{latest}/SB_{F=0}$ (0.39 median; 0.18 – 0.79 80th quantile) or SB_{latest}/SB_{MSY} (2.95 median; 0.99 – 6.78 80th quantile), and is not currently estimated to be undergoing overfishing according to F_{recent}/F_{MSY} (0.47 median; 0.25 – 1.29 80th quantile). However, it is misleading to characterize the stock status only based on measures of central tendency and it is important to consider the uncertainty in those estimates. This is perhaps best conveyed by examining the Majuro ([Figure 54](#)) and Kobe ([Figure 55](#)) plots. Considering the full range of uncertainty, we see that there is a non-trivial risk that the stock is currently undergoing overfishing (20% $F_{recent}/F_{MSY} > 1$) and a risk that the stock is in an overfished state (13% $SB_{latest}/SB_{F=0} < 0.2$ & 10% $SB_{latest}/SB_{MSY} < 1$).

9 Discussion and conclusions

9.1 General remarks

The 2021 SWPO swordfish stock assessment estimates the median stock status, in terms of spawning potential depletion ($SB_{latest}/SB_{F=0}$), to be comparable to the 2017 assessment ([Takeuchi et al., 2017](#)) used to provide management advice for the stock. Stock status as assessed by yield based reference points (F_{recent}/F_{MSY} & SB_{latest}/SB_{MSY}) was more optimistic than the 2017 assessment. Yield based reference points are particularly sensitive to the assumptions surrounding the production function so a more optimistic yield based outlook is not unexpected given that the current

model ensemble generally assumed higher values of steepness and natural mortality than the 2017 stock assessment. Compared to the previous stock assessment, stock status is characterized to be less certain and with a non-trivial chance of being considered overfished or undergoing overfishing. Again, this broader uncertainty is not unexpected as the previous stock assessment only considered model uncertainty, and did not consider estimation uncertainty.

Taking a historical perspective of the trajectory of the stock, depletion appears to have been relatively low through the first 40 years of the model period, and a decline in spawning potential appears to be largely driven by the observed removals. Uncertainty in this early period is large given the lack of informative data during this period. What appears more certain is that beginning in the mid-to-late 1990s, the stock appears to decline quite rapidly to a low-point at the end of the 2000s. This coincides with the development of directed domestic longline fisheries in both Australia and New Zealand, the emergence of a targeted fishery by the European Union and also a considerable increase in the catch of typically smaller swordfish in tropical waters in the north-east of the assessment region. The stock appeared to rebound slightly following this low-point as catches stabilized at a high level, though the recent trend indicates a decline in both stock status and catches.

Recruitment in the early period prior to availability of composition data or the start of the Australian CPUE index was very uncertain, and restricted to be near the average value as predicted by the stock-recruit relationship ($\sim 500,000$ individuals). Starting in the mid-1990s when the information content in the model increased, recruitment estimates were predicted to be more certain and increasing. However, the last years of *estimated* recruitment were predicted to be considerably lower, below both recent and average levels.

9.2 Improvements to the assessment

The process of conducting a stock assessment is one of iterative and continued improvement, picking up the development where the previous assessment left off and incorporating new model features, updated data, and an improved understanding of biological processes. The 2021 swordfish stock assessment was no exception. Relative to previous assessments, the most notable improvement was the approach used to more fully, and transparently capture the uncertainty in stock status following the framework developed by [Ducharme-Barth and Vincent \(2021\)](#). As a part of this process, biological assumptions were updated in the creation of the joint prior. This included an improvement to the calculation of decimal age for the otolith age data, following [Farley et al. \(2020\)](#), without making a common birth date assumption, and modeling reproductive potential as a function of length. The use of the joint prior to underpin the model ensemble also allowed for the consideration of additional sources of uncertainty (e.g. length-weight relationship, reproductive potential, and relative weighting between input CPUE indices) not typically considered in the full-factorial structural uncertainty grids.

Additionally, a number of other small improvements were made to the model including applying

spatiotemporal modeling techniques to the standardization of DWFN CPUE data, the move to an index fishery approach, and re-weighting the composition data to be more representative to either the extracted catch or the indexed population (Ducharme-Barth et al., 2021b). Moving to the index fishery approach also allowed us to completely remove the impact of the non-standardized effort on the estimated population trajectory.

It is worth noting that the relatively rapid run-time of the SWPO swordfish stock assessment model⁸ allowed considerable freedom to explore a large number⁹ of different model configurations, sensitivities, and diagnostic approaches. The most computationally demanding component was the calculation of ensemble model weights using hindcast predictions of CPUE as this required an additional 10 models for each grid model to be run. In order for such an approach to be feasible for application to other WCPFC assessments, two conditions must be met¹⁰: 1) the model ensemble should be kept small and 2) models should run quickly. Applying the framework outlined by Ducharme-Barth and Vincent (2021) is a good place to start for keeping ensemble size manageable. The equivalent full factorial ensemble, assuming 3 levels for each main component of the joint prior, would have required at least 944,784 models!! It has already been identified that at least the WCPO bigeye tuna & yellowfin tuna stock assessments are likely overly complex and in need of simplification (Ducharme-Barth et al., 2020; Vincent et al., 2020). While simplification of these assessment models to a more appropriate model structure will hopefully lead to more well determined solutions with better diagnostics, it should also have the added benefit of reducing model run time. Naturally there is a trade-off between these two factors such that faster run-time will allow for a larger ensemble to remain tractable. Assuming both of these conditions are met, there is no reason why such an approach could not be applied to all other WCPFC stock assessments.

9.3 Challenges

Development of the 2021 SWPO swordfish stock assessment was not without its challenges, and a number of questions were raised related to the quality of the input data, assumptions made, and modeling approaches taken.

One of the key assumptions made in most stock assessments is that input data are un-biased and representative of the underlying stock, and assessment outcomes are conditioned on that assumption being met. While this assumption may be safe to make in assessments with well designed fisheries independent sampling programs, it could be problematic in assessments such as SWPO swordfish where most, if not all, input data come from fisheries-dependent data sources which have the potential for bias. In the context of the relative abundance indices which drive the population dynamics, these potential biases can be corrected through the application of standardization models.

⁸~3 hours run time for model convergence to a small maximum gradient component ($\sim \leq 1 \times 10^{-5}$) and Hessian calculation.

⁹~9,300 models

¹⁰An unstated assumption is that the analyst has access to some basic level of scientific computing capabilities (e.g. dedicated high-performance multi-core machine/server, locally hosted or cloud-based like Amazon Web Services; or a distributed computing network like HTCCondor).

Though advancements have been made in the standardization procedures applied to generate indices of relative abundance, the quality of the indices will always be limited by the underlying data. In the case of SWPO swordfish, there was evidence of conflict between the different considered indices. Similarly, conflicts also existed between the observations of size-frequency data between and within fisheries. Conflict in data components is generally believed to arise as a product of model misspecification, and this possibility should be thoroughly explored. However, data conflicts can also arise if the input data do not representatively characterize the underlying population in an unbiased way. Though it may not be realistic to develop a fisheries independent data collection program for swordfish across the entire SWPO, efforts should be made to impose a statistically robust sampling design on the collection of fisheries dependent data for use in stock assessment.

Given the rigorous standardization approaches applied to the Australian and New Zealand abundance indices, made possible by extensive and comprehensive operational characteristics recorded in the logbook, these indices were considered to be the most representative indicators of change in relative abundance for the sampled swordfish stock. However, these fisheries operate in a relatively restricted range and may not completely capture the totality of stock dynamics. Additional indices were considered which sampled the stock across a broader geographic distribution, though these were also not without issue. The European Union index showed strong seasonal confounding with spatial location of sampling. This non-stationarity in spatial sampling can pose problems even for spatiotemporal standardization approaches which explicitly seek to account for the spatial location of observations (Ducharme-Barth et al. *Submitted*). Seasonal-spatial confounding could indicate a fishery which tracks a high population density of the target species along a migration pathway, and be prone to hyper-stability in CPUE. Non-stationarity in spatial sampling was also an issue for the Japanese and Chinese Taipei indices, particularly as fishing effort has contracted spatially over the last decade. Alternative sources of relative abundance information will likely need to be considered in future assessments. Lastly, the comprehensive operational covariate information available for Australia and New Zealand, and the importance of some of these covariates in the standardization model underscores the need to enhance the minimum data reporting requirements for longline vessels operating in the WCPFC convention area for this purpose.

The high catches seen in recent years off the coasts of both Australia and New Zealand, despite the relatively low effort, appears to be a clear confirmation of the directed swordfish targeting of the domestic longline fisheries operating in the region. Conversely, the lack of high swordfish catches in the high effort, western-central region from 10–20°S coincides with our belief that longline vessels in this area typically target south Pacific albacore and are unlikely to encounter swordfish as a result. The explanation for the very high catches seen in recent years in the waters around the north-east corner of region 2, despite moderate levels of effort, is less clear. These high catches could be due to a number of reasons: 1) environmental factors in the south central and eastern Pacific Ocean (thermocline/oxycline depth) leading to swordfish foraging at depths where they could be more vulnerable to longline gear; 2) increases in swordfish targeting by longlines operating in this sector,

either directly or indirectly through increased targeting of bigeye tuna which have a similar diurnal diving behavior; 3) a genuine increase in the abundance of the stock in this area. Directed research is needed to resolve this issue, as it has implications for the reliability of fisheries dependent CPUE data coming from this region used to index abundance. This directed research could take the form of a fisheries independent survey, which is sorely needed for all WCPO species faced with uncertain fisheries dependent CPUE. At a minimum, expanded reporting of operational covariates in the logbook would greatly enhance our ability to account for this in a CPUE standardization. Useful covariates for swordfish would include: *a priori* target species, light stick use, bait type, setting time (or fraction of night time soak), and gear settings that influence fishing depth (e.g. hooks between floats, branch line length, float line length, and/or line setting speed). Cognizant of the reporting requirements already placed on vessel operators, some of the burden related to expanding the reporting of operational characteristics could potentially be alleviated via more widespread adoption of electronic logbooks or electronic monitoring systems.

Sensitivity explorations to the inclusion of additional catches just outside of the WCPFC assessment boundaries for SWPO swordfish or the simplification of the model to a single region, indicate that the stock status summarized in [Table 5](#) could be too optimistic. However, the more pessimistic conclusions drawn from both *add 5N* and the *single region* scenario are contingent on the assumed stock structure, and movement assumptions. In the *add 5N* model, the increase in fishing pressure caused by the inclusion of additional catches is applied to all of region 2 leading to the more depleted state. A different regional structure which assumes that this high catch area is a spatially distinct component of the stock (e.g. a three region model with an additional longitudinal boundary at 165°W) could result in a different outcome. The lack of understanding on this area of high swordfish catches at the northeast of the model region and how it relates to the greater SWPO stock in terms of genetics and movement, does not allow us to validate or refute such a spatial hypothesis. Assuming a single region and a well-mixed stock as done in the *single region* model could be a reasonable starting point for the next SWPO swordfish stock assessment given the uncertainty in movement estimates ([Patterson et al., 2021](#)) and our current understanding that the SWPO swordfish stock is a single genetic population ([Evans et al., 2021](#)). We note that if there is significant as-of-yet undetected spatial structure, such a simplification could bias results since fishing pressure is not equal across the spatial domain. In future assessments, implications for management advice could be more thoroughly explored by including catch scenarios and/or spatial structure in the model ensemble.

A re-occurring theme in this report has been the sensitivity of model outputs to fundamental aspects of the stocks dynamics or biology given the lack of information in the model to properly estimate population scale. Lack of precision in movement estimates and natural mortality clearly translates to large uncertainties in the model ensemble, and/or implausible estimates ([Section 8.6.1](#)). Ambiguity about stock structure and population connectivity could lead to inappropriate management advice if the assessment model is mis-specified with respect to the unobserved “true” state of na-

ture. Such fundamental uncertainties can be addressed in one of two ways: 1) additional directed research on stock structure, movement dynamics, and natural mortality; 2) use of a Management Strategy Evaluation (MSE) procedure to develop a harvest strategy for the SWPO that is robust to such uncertainty. While the implementation of a harvest strategy should enable the achievement of fishery objectives, it will not reduce the uncertainty in the stock assessment. Prior to the next SWPO stock assessment, directed longitudinal genetic sampling and tagging of individuals across the SWPO (e.g. French Polynesia, Cook Islands, Kiribati, Tonga, Fiji, New Zealand, New Caledonia, and Australia) is needed to properly determine: 1) the connectivity of the high exploitation zone in the northeast of the SWPO to the rest of the SWPO¹¹, 2) if the 165°E boundary needs revising given the genetic similarity between Australian and New Zealand individuals, 3) better estimates of movement between assessment regions, if a spatially explicit model is still the most appropriate. If possible, tagging studies should endeavor to produce long term estimates of movement (1+ years at liberty), and be structured so as to be representative of the movement dynamics of the population as a whole and not a particular age or size based subset of the population.

Exploration of a sex-disaggregated model showed that changes in the assumed growth variability resulted in quite different estimated selectivities. Models with smaller assumed values for σ_{Age} appeared to produce more reasonable shared selectivity estimates. However, in most fisheries, males were never fully selected due to the larger female size and non-decreasing nature of most estimated selectivity curves. Given the dearth of sex-specific catch or composition information, the shared selectivity at length assumption is required in order to partition the aggregated catch and composition data by sex. The lack of full vulnerability to longline fisheries for large 200 cm males could potentially be justifiable from a behavioral standpoint, but becomes more difficult to justify strictly on the basis of length especially when we generally consider that other longline vulnerable species (e.g. tropical tunas & albacore) are fully vulnerable at sizes less than 200 cm. Even though depletion is based only on female biomass, male selectivity is still important to get right as it has implications for dis-aggregating the aggregated catch by sex, which in turn impacts the sex-specific fishing mortalities. Substantial collection of sex-specific catch and composition data are needed to improve the quality of any future sex-disaggregated stock assessment.

A common thread through all of these issues is that genetic sampling and techniques can offer a way to address these challenges, reduce model uncertainty, and improve the structure of the stock assessment. As previously discussed, concerted genetic sampling across the spatial domain of the SWPO could contribute to solving unanswered questions on stock structure. Application of Close-Kin Mark-Recapture (CKMR) techniques could provide badly needed fisheries independent information on population scale and natural mortality (Bravington et al., 2016). Fisheries independent estimates of total population scale from CKMR could be fit to directly within MULTIFAN-CL if the structure of the model were modified, or it could be used as an independent external estimate to identify implausible models. Lastly, investigation of the swordfish genome for a sex-specific seg-

¹¹As well as connectivity of this region to the other swordfish stocks in the norther and eastern Pacific Ocean.

ment could aid in catch and size composition by sex reconstruction from longline port sampling. However, before engaging in any new sampling program it is important that it be well thought out and designed so that it can meet the desired objectives. On-board swordfish processing techniques (“trunking”) may make it difficult to collect otoliths to age genetically sampled individuals for a CKMR model, which may necessitate the development of an alternative ageing protocol (i.e. epigenetic ageing). The spatial structure of the population in terms of locations of juveniles and adults available for sampling will also need to be assessed to determine if CKMR is even feasible. In order to improve the quality of the stock assessment it is likely worthwhile to conduct a feasibility study, similar to [Bravington et al. \(2020, 2021\)](#), on the application of CKMR and other genetic techniques for SWPO swordfish.

9.4 Stock assessment modeling considerations

The previous section identified many challenges related to the current assessment, but addressing those issues may take time as many require the collection of new data. In the interim, there are a number of modeling based approaches which were identified as fertile areas of future investigation, but were unable to be explored in the current assessment.

- Composition data for many of the fisheries were sparse, and in the mixed-flag fisheries came from a variety of different sampling programs. Aggregating all of these different sources of composition together could be a source for some of the data conflict seen in the assessment. It may be worth re-exploring these fisheries definitions, using a data-driven approach such as the one used to define the EPO fisheries in the 2021 South Pacific albacore stock assessment ([Lennert-Cody et al., 2010](#); [Vidal et al., 2021](#)).
- The composition data re-weighting approach that is described in [Ducharme-Barth et al. \(2021b\)](#) made assumptions as to the threshold value applied for excluding composition data that corresponded to low levels of the total catch for a fishery. This was done to remove temporal variation as a result of limited or unbalanced sampling. Alternative model runs assuming different applied threshold values should be explored.
- For the index fisheries, it is desirable that the size composition be standardized such that it is representative of changes in abundance and not changes in spatial sampling ([Maunder et al., 2020](#)). While it may not be possible to apply such an approach for DWFN fisheries where there are low levels of size frequency samples or the data are not spatially resolved, this approach could be considered for standardizing the composition data associated with the domestic and PICT longline fisheries.
- The current approach allowed for uncertainty in the data weighting of the different components, and represented a step forward from previous assessments. Results indicated that the model outputs were not overly influenced by changes to the data weighting relative to other aspects of the ensemble. However this approach to incorporating uncertainty in data weight-

ing is overly simplistic and more work is needed to develop an efficient approach for dealing with data weighting in a model ensemble context. If iterative re-weighting approaches are to be applied in an ensemble, robust stopping criteria are needed; this would improve the automation, transparency, and repeatability for applying iterative type approaches.

- Related to data-weighting, the Self-Scaling Multinomial with Auto-correlated Random Effects likelihood for the size composition data should be tested prior to the next stock assessment to see if it is feasible for use in a production stock assessment, and to determine how sensitive it is to different fleet groupings and configurations.
- Exploration of marginal likelihood profiles for M and L_1 indicated that there may be sufficient information content in the data to freely estimate these parameters. Whether these are estimable simultaneously, in conjunction with the other growth parameters, and/or across the different model configurations in the ensemble needs further evaluation. The new decimal age estimates from the otolith data should be incorporated into the stock assessment if serious attempts at estimating these parameters are made.
- Given the new approach for characterizing uncertainty as model and estimation uncertainty, there is a need to make sure all relevant management quantities and reference points are included in the variance estimation routine for dependent variables in MULTIFAN-CL. Scrupulous readers of this report will notice that fewer reference points and management quantities of interest are reported in [Table 5](#) than in years past, and those that are missing are those that we were unable to calculate the statistical uncertainty for given the current production version of MULTIFAN-CL. Deciding on a common, key set of reference points that should be presented will be useful in ensuring that future versions of MULTIFAN-CL report the uncertainty for the relevant quantities.

9.5 Main assessment conclusions

The main conclusions of this assessment are summarized as follows:

- The model and estimation uncertainty from the final model ensemble indicate that the stock is at a 20% risk of undergoing overfishing according to $F_{recent}/F_{MSY} > 1$ (median 0.47; 0.25 – 1.29 80th quantile), a 13% risk that the stock is overfished according to $SB_{latest}/SB_{F=0} < 0.2$ (median 0.39; 0.18 – 0.79 80th quantile) and a 10% risk that the stock is overfished according to $SB_{latest}/SB_{MSY} < 1$ (median 2.95; 0.99 – 6.78 80th quantile).
- Comparison with the equivalent sex-disaggregated model ensemble ([Appendix Section 12.3](#)) indicated a more optimistic stock status in terms of median $SB_{latest}/SB_{F=0}$ (median 0.45; 0.13 – 0.72 80th quantile), with greater uncertainty resulting in a 19% risk that the stock is overfished according to $SB_{latest}/SB_{F=0} < 0.2$.
- The model ensemble predicts the stock to have gradually declined from the 1950s to the

mid-1990s before rapidly declining to an overall low point near 2010. Current stock status is estimated to be at a similar level as the overall low with a declining trend in the terminal 4 years of the model.

- Uncertainty in the movement rates, and natural mortality are major contributors to the overall assessment uncertainty.

9.6 Summary of research recommendations

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

- Development of a statistically robust sampling plan for the collection of fisheries dependent biological samples, including but not limited to size frequency data (by sex), and genetic samples.
- Expand minimum reporting requirements for longline operational characteristics to include: *a priori* target species, light stick use, bait type, setting time (or fraction of night time soak), and gear settings that influence fishing depth (e.g. hooks between floats, branch line length, float line length, and/or line setting speed).
- Directed longitudinal genetic sampling and tagging of individuals across the SWPO (e.g. French Polynesia, Cook Islands, Kiribati, Tonga, Fiji, New Zealand, New Caledonia, and Australia) is needed to properly determine: 1) the connectivity of the high exploitation zone in the northeast of the SWPO to the rest of the SWPO, 2) if the 165°E boundary needs revising given the genetic similarity between Australian and New Zealand individuals, 3) better estimates of movement between assessment regions given the high uncertainty in movement and the impact on management quantities.
- Improved collection of sex-specific catch and length composition to improve estimates of sex specific selectivity in a sex-disaggregated model.
- Conduct a feasibility study to see if a Close-Kin Mark-Recapture approach can be applied to swordfish to provide information on total population scale and natural mortality, two components of the model that have large uncertainty.

Acknowledgments

We thank the various fisheries agencies for the provision of the catch, effort and size frequency data used in this analysis, and in particular J. Farley at CSIRO for facilitating analysis of biological data for the creation of the joint prior. We are deeply appreciative to the many sacrifices made by onboard fisheries observers who play an instrumental role in collecting data that comprise critical inputs to the stock assessment. We thank F. Bouyé for his tireless work in increasing and maintaining SPC-OFP's scientific computing capabilities which allowed us to unleash the awesome

computing power of a fully armed and operational HTCCondor flock for this stock assessment. Lastly, without S. McKechnie's hard work and effort pulling together text and figures, the *WCPFC-SC17-2021/SA-IP-07: Background analysis for the 2021 stock assessment of southwest Pacific swordfish* report likely would not have been finished. This stock assessment was supported by funding from the Western and Central Pacific Fisheries Commission.

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.
- 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.
- Bravington, M., Farley, J., Hampton, J., and Nicol, S. (2020). Preliminary analyses for a close kin mark recapture feasibility study in WCPO. Technical Report WCPFC-SC16-2020/SA-IP-15.
- Bravington, M., Nicol, S., Anderson, G., Farley, J., Hampton, J., Castillo Jordan, C., and Macdonald, J. (2021). South Pacific Albacore Close-Kin Mark-Recapture: update on design (Project 100b). Technical Report WCPFC-SC17-2021/SA-IP-14.
- Bravington, M. V., Skaug, H. J., and Anderson, E. C. (2016). Close-Kin Mark-Recapture. *Statistical Science*, 31(2).
- 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.
- Cadigan, N. G. and Farrell, P. J. (2005). Local influence diagnostics for the retrospective problem in sequential population analysis. *ICES Journal of Marine Science: Journal du Conseil*, 62(2):256–265.
- Campbell, R. (2019). Aggregate and size-based standardised CPUE indices for longline target species caught within the ETBF - 2019 Update. Technical Report Working paper presented to the 24th meeting of the Tropical Tuna RAG, held 17-18 July 2019, Mooloolaba.
- Carvalho, F., Punt, A. E., Chang, Y.-J., Maunder, M. N., and Piner, K. R. (2017). Can diagnostic tests help identify model misspecification in integrated stock assessments? *Fisheries Research*, 192:28–40.
- 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., Wetzell, 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.
- Davies, N., Bian, R., Kolody, D., and Campbell, R. (2008). CASAL stock assessment for south-west-central Pacific broadbill swordfish 1952-2007. Technical Report WCPFC-SC4-SA-WP-7.

- Davies, N., Fournier, D., and Hampton, J. (2019). Developments in the MULTIFAN-CL software 2018-2019. Technical Report WCPFC-SC15-2019/SA-IP-02, Pohnpei, Federated States of Micronesia.
- Davies, N., Pilling, G., Harley, S., and Hampton, J. (2013). Stock assessment of swordfish (*Xiphias gladius*) 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., Gruss, A., Vincent, M., Kiyofuji, H., Aoki, Y., Pilling, G., Hampton, J., and Thorson, J. (2021a). Impacts of fisheries-dependent spatial sampling patterns on catch-per-unit-effort standardization: A simulation study and fishery application. *Fisheries Research*.
- Ducharme-Barth, N., Peatman, T., and Hamer, P. (2021b). Background analysis for the 2021 stock assessment of southwest Pacific swordfish. Technical Report WCPFC-SC17-2021/SA-IP-07.
- Ducharme-Barth, N. and Vincent, M. (2021). Focusing on the front end: A framework for incorporating uncertainty in biological parameters in model ensembles of integrated stock assessments. Technical Report WCPFC-SC17-2021/SA-WP-05.
- Ducharme-Barth, N., Vincent, M., Hampton, J., Hamer, P., Williams, P., and Pilling, G. (2020). Stock assessment of bigeye tuna in the western and central Pacific Ocean. Technical Report WCPFC-SC16-2020/SA-WP-03.
- 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., Grewe, P., Foster, S., Gunasekera, R., Lansdell, M., Meredith, S., Sarau, S., Tracey, S., and Wichman, M. (2021). Connectivity of broadbill swordfish targeted by Australian Eastern Tuna and Billfish Fishery with the broader Western Pacific Ocean. Technical Report WCPFC-SC17-2021/SA-IP-12.

- 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.
- Farley, J., Krusic-Golub, K., Eveson, P., Clear, N., Rouspard, F., Sanchez, C., Nicol, S., and Hampton, J. (2020). Age and growth of yellowfin and bigeye tuna in the western and central Pacific Ocean from otoliths. Technical Report WCPFC-SC16-2020/SC16-SA-WP-02.
- Finucci, B., Griggs, L., Sutton, P., Fernandez, D., and Anderson, O. (2021). Characterisation and CPUE indices for swordfish (*Xiphias gladius*) from the New Zealand tuna longline fishery, 1993 to 2019. Technical Report New Zealand Fisheries Assessment Report 2021/07.
- Fournier, D., Hampton, J., and Sibert, J. (1998). MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. *Canadian Journal of Fisheries and Aquatic Sciences*, 55:2105–2116.
- Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A., and Sibert, J. (2012). AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, 27(2):233 – 249.
- Francis, R. I. C. C. (1992). Use of risk analysis to assess fishery management strategies: A case study using orange roughy (*Hoplostethus atlanticus*) on the Chatham Rise, New Zealand. *Canadian Journal of Fisheries and Aquatic Science*, 49:922–930.
- Grewe, P., 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. and Cope, J. (2021). Considerations for developing a longevity-based prior for the natural mortality rate. *Fisheries Research*, Submitted.
- Hamer, P., Castillo Jordon, C., Xu, H., Ducharme-Barth, N., and Vidal Cunningham, T. (2021). Report from the SPC Pre-assessment Workshop - March 2021. Technical Report WCPFC-SC17-2021/SA-IP-02, Pacific Community.
- Hampton, J. and Fournier, D. (2001). A spatially-disaggregated, length-based, age-structured population model of yellowfin tuna (*Thunnus albacares*) in the western and central Pacific Ocean. *Marine and Freshwater Research*, 52:937–963.

- Harley, S. J. (2011). A preliminary investigation of steepness in tunas based on stock assessment results. Technical Report WCPFC-SC7-2011/SA-IP-08, Pohnpei, Federated States of Micronesia, 9–17 August 2011.
- 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.
- Hsu, A. C., Boustany, A. M., Roberts, J. J., Chang, J., and Halpin, P. N. (2015). Tuna and swordfish catch in the U.S. northwest Atlantic longline fishery in relation to mesoscale eddies. *Fish. Oceanogr.*, 24(6):508–520.
- 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.
- Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688.
- ISSF (2011). Report of the 2011 ISSF stock assessment workshop. Technical Report ISSF Technical Report 2011-02, Rome, Italy, March 14–17.
- Kell, L. T., Sharma, R., Kitakado, T., Winker, H., Mosqueira, I., Cardinale, M., and Fu, D. (2021). Validation of stock assessment methods: is it me or my model talking? *ICES Journal of Marine Science*, page fsab104.
- Kleiber, P., Fournier, D., Hampton, J., Davies, N., Bouye, F., and Hoyle, S. (2019). MULTIFAN-CL User’s Guide. Technical report.
- Kolody, D., Campbell, R., and Davies, N. (2006). A Multifan-CL stock assessment for south-west Pacific Swordfish 1952-2004. Technical Report WCPFC-SC2 ME-WP-3.
- 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.
- Lee, H. H., Piner, K. R., Methot, R. D., and Maunder, M. N. (2014). Use of likelihood profiling over a global scaling parameter to structure the population dynamics model: An example using blue marlin in the Pacific Ocean. *Fisheries Research*, 158:138–146.
- Lennert-Cody, C. E., Minami, M., Tomlinson, P. K., and Maunder, M. N. (2010). Exploratory analysis of spatial-temporal patterns in length–frequency data: An example of distributional regression trees. *Fisheries Research*, 102(3):323–326.
- Lopez-Quintero, F. O., Contreras-Reyes, J. E., and Wiff, R. (2017). Incorporating uncertainty into a length-based estimator of natural mortality in fish populations. *Fishery Bulletin*.

- Lorenzen, K. (2000). Allometry of natural mortality as a basis for assessing optimal release size in fish-stocking programs. *Canadian Journal of Fisheries and Aquatic Sciences*, 57:2374 – 2381.
- 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.
- Maunder, M., Lee, H., Piner, K., Hamel, O., Cope, J., Punt, A., Ianelli, J., and Methot, R. (2021). A review of estimation methods for natural mortality and their performance in the context of fishery stock assessment. *Fisheries Research*, Submitted.
- 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.
- McKechnie, S. (2014). Analysis of longline size frequency data for bigeye and yellowfin tunas in the WCPO. Technical Report WCPFC-SC10-2014/SA-IP-04, Majuro, Republic of the Marshall Islands, 6–14 August 2014.
- McKechnie, S., Pilling, G., and Hampton, J. (2017). Stock assessment of bigeye tuna in the western and central Pacific Ocean. Technical Report WCPFC-SC13-2017/SA-WP-05, Rarotonga, Cook Islands, 9–17 August 2017.
- 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.
- 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.
- 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.
- Minte-Vera, C. V., Maunder, M. N., Aires-da Silva, A. M., Satoh, K., and Uosaki, K. (2017). Get the biology right, or use size-composition data at your own risk. *Fisheries Research*, 192:114–125.
- 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 gladius*, in the Pacific ocean - a review. Technical Report WCPFC-SC17-2021-SA-IP-08.
- Myers, R. A., Bowen, K. G., and Barrowman, N. J. (1999). Maximum reproductive rate of fish at low population sizes. *Canadian Journal of Fisheries and Aquatic Sciences*, 56(12):2404–2419.
- 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.
- Nishikawa, Y. and Ueyanagi, S. (1974). The distribution of the larvae of swordfish, *Xiphias gladius*, in the Indian and Pacific Oceans. Technical Report U.S. Department of Commerce, NOAA Technical Report, NMFS, SSRF-675: 261-264.
- 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.
- 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.
- Reeb, C. A., Arcangeli, L., and Block, B. A. (2000). Structure and migration corridors in Pacific populations of the Swordfish *Xiphias gladius*, as inferred through analyses of mitochondrial DNA. *Marine Biology*, 136(6):1123–1131.
- SPC-OFP (2019). Project 90 Update: Better data on fish weights and lengths for scientific analyses. Technical Report WCPFC-SC15-2019/ST-WP-03, Oceanic Fisheries Programme, The Pacific Community.
- SPC-OFP (2021). Catch and effort data summaries to support discussions on the new swordfish CMM. Technical Report WCPFC-SC17-2021/MI-IP12.
- Stewart, I. and Martell, S. (2014). Assessment of the Pacific Halibut stock at the end of 2013. Technical Report IPHC Report of Assessment and Research Activities 2013.
- 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., Pilling, G., 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.

- Thorson, J. T. (2019). Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. *Fisheries Research*, 210:143–161.
- 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., Hampton, J., McKechnie, S., and Pilling, G. (2018). Stock assessment of South Pacific albacore tuna. Technical Report WCPFC-SC-14-2018/SA-WP-05, Busan, South Korea, 8-16 August 2018.
- 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.
- Vidal, T., Castillo Jordan, C., Peatman, T., Ducharme-Barth, N., Xu, H., Williams, P., and Hamer, P. (2021). Background analyses and data inputs for the 2021 South Pacific albacore tuna stock assessment. Technical Report WCPFC-SC17-2021/SA-IP-03.
- Vincent, M., Ducharme-Barth, N., Hampton, J., Hamer, P., Williams, P., and Graham, P. (2020). Stock assessment of yellowfin tuna in the western and central Pacific Ocean. Technical Report WCPFC-SC16-2020/SA-WP-04.
- Winker, H., Carvalho, F., and Kapur, M. (2018). JABBA: Just Another Bayesian Biomass Assessment. *Fisheries Research*, 204:275–288.
- 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.

10 Tables

Table 1: Description of symbols used in the yield and stock status analyses. For the purpose of this assessment, “recent” is the average over the period 2015–2018 for F related metrics and 2016–2019 for metrics related to SB while “latest” is 2019.

Symbol	Description
C_{latest}	Catch in the last year of the assessment (2019)
YF_{recent}	Equilibrium yield at average fishing mortality for a recent period (2015–2018)
MSY	Equilibrium yield at F_{MSY}
F_{recent}/F_{MSY}	Average fishing mortality-at-age for a recent period (2015–2018) relative to F_{MSY}
SB_0	Equilibrium unexploited spawning potential
SB_{latest}	Spawning potential in the latest time period (2019)
SB_{recent}	Spawning potential for a recent period (2016–2019)
SB_{MSY}	Spawning potential that will produce the maximum sustainable yield (MSY)
$SB_{latest}/SB_{F=0}$	Spawning potential in the latest time period (2019) relative to the average spawning potential predicted to occur in the absence of fishing for the period 2009–2018
SB_{latest}/SB_{MSY}	Spawning potential in the latest time period (2019) relative to that which will produce the maximum sustainable yield (MSY)
SB_{recent}/SB_{MSY}	Spawning potential for a recent period (2016–2019) relative to that which will produce the maximum sustainable yield (MSY)
$20\%SB_{F=0}$	WCPFC adopted limit reference point for tropical tunas – 20% of spawning potential in the absence of fishing averaged over years $t - 10$ to $t - 1$ (2009–2018)

Table 2: Fishery definitions, selectivity groupings, and increasing selectivity with age constraint for extraction fisheries used in the 2021 SWPO swordfish stock assessment. Flag/fleets: AU = Australia; DWFN = Distant Water Fishing Nations; EU = European Union; NZ = New Zealand; PICT = Pacific Islands Countries and Territories.

Fishery	Nationality	Region	Group	Increasing
01_DW_1N	DWFN	1	1	N
02_DW_1C	DWFN	1	2	N
03_DW_1S	DWFN	1	3	N
04_AU_1	AU	1	4	N
05_EU_1	EU	1	5	N
06_Other_1	PICT	1	6	N
07_DW_2N	DWFN	2	7	N
08_DW_2C	DWFN	2	8	N
09_DW_2S	DWF	2	9	Y
10_NZ_2	NZ	2	10	N
11_EU_2	EU	2	5	N
12_Other_2N	PICT	2	11	N
13_Other_2C	PICT	2	12	N

Table 3: Fishery definitions, selectivity groupings, and increasing selectivity with age constraint for index fisheries used in the 2021 SWPO swordfish stock assessment. Flag/fleets: AU = Australia; EU = European Union; JP = Japan; NZ = New Zealand; TW = Chinese Taipei.

Fishery	Nationality	Region	Group	Increasing
14_idx_AU	AU	1	13	Y
15_idx_NZ	NZ	2	14	Y
16_idx_EU	EU	2	15	Y
17_idx_JPearly_1	JP	1	16	Y
18_idx_JPearly_2	JP	2	17	Y
19_idx_JP_mid_1	JP	1	16	Y
20_idx_JP_mid_2	JP	2	17	Y
21_idx_JP_late_1	JP	1	16	Y
22_idx_JP_late_2	JP	2	17	Y
23_idx_TW_early_1	TW	1	18	Y
24_idx_TW_early_2	TW	2	19	Y
25_idx_TW_late_1	TW	1	18	Y
26_idx_TW_late_2	TW	2	19	Y

Table 4: Summary of fixed biological assumptions made in both the diagnostic case and model ensemble. The minimum, maximum, median and 80th percentile are given for the ensemble parameters.

	Diagnostic	Min	10	Median	90	Max
Steepness	0.8	0.603	0.771	0.883	0.95	0.987
α	1.35e-05	1.02e-05	1.14e-05	1.27e-05	1.42e-05	1.62e-05
β	3	2.96	2.98	3	3.02	3.04
L_1	109	93.5	98	103	110	113
L_2	241	217	224	233	243	250
k	0.191	0.157	0.174	0.228	0.274	0.298
σ_{Age}	27.9	24.4	26.7	38.8	60	60
σ_{Length}	0.297	0.145	0.28	0.763	1.27	1.39
\bar{M}	0.295	0.103	0.189	0.295	0.361	0.49
<i>Tasman</i> \Rightarrow <i>S. Pacific</i>	0.075	0.00268	0.00851	0.0213	0.0489	0.127
<i>S. Pacific</i> \Rightarrow <i>Tasman</i>	0.025	0.00225	0.00826	0.03	0.0932	0.507

Table 5: Summary of reference points (measures of central tendency & relevant percentiles) including model and estimation uncertainty from the 25 models in the final ensemble. Models were equally weighted in the ensemble.

	Mean	Median	01	10	90	99
C_{latest}	7,772	7,723	7,364	7,524	8,259	8,453
YF_{recent}	6,558	6,608	3,351	4,964	8,106	9,347
MSY	9,922	9,543	3,869	5,470	14,738	22,278
F_{recent}/F_{MSY}	0.67	0.47	0.16	0.25	1.29	2.34
SB_0	83,853	69,390	16,491	31,472	145,944	334,518
SB_{latest}	38,287	31,517	10,588	16,096	69,370	125,681
SB_{recent}	41,916	38,106	14,975	18,956	68,550	99,304
SB_{MSY}	12,507	11,480	2,427	5,212	21,722	29,297
SB_{latest}/SB_{MSY}	3.7	2.95	0.44	0.99	6.78	18
SB_{recent}/SB_{MSY}	4.1	3.61	0.64	1.23	7.39	16
SB_{latest}/SB_0	0.59	0.46	0.1	0.2	1.09	2.49
$SB_{latest}/SB_{F=0}$	0.45	0.39	0.08	0.18	0.79	1.42

11 Figures

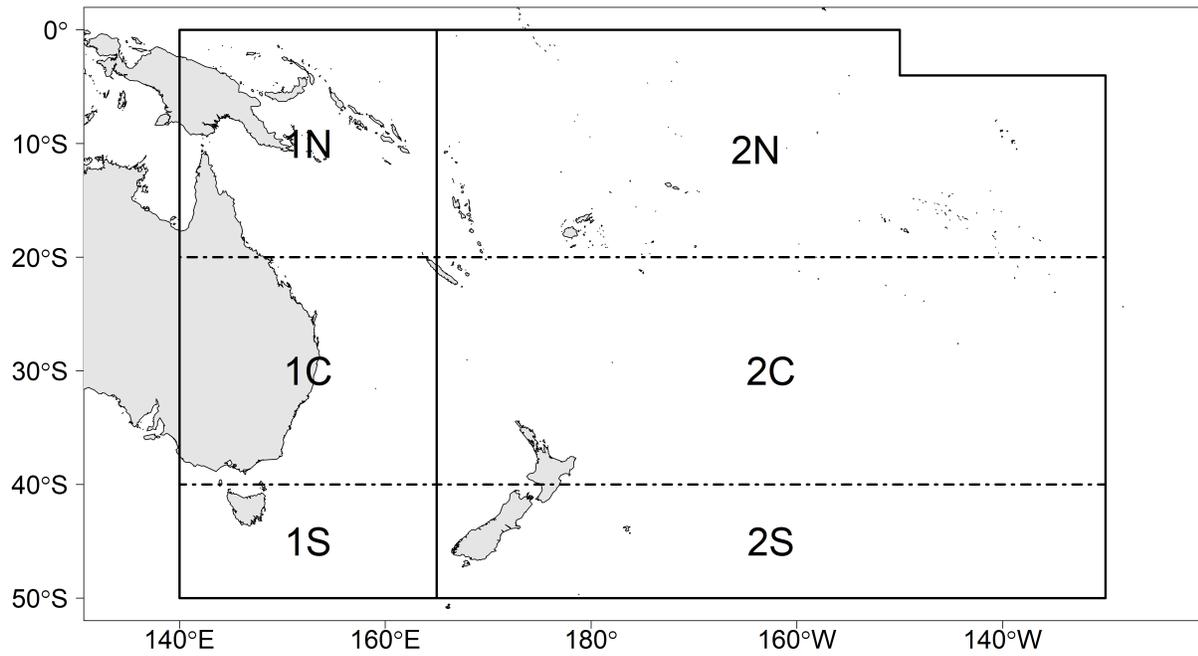


Figure 1: Spatial structure for the 2021 SWPO swordfish stock assessment. Sub-regions used to differentiate fisheries are shown with the dotted lines.

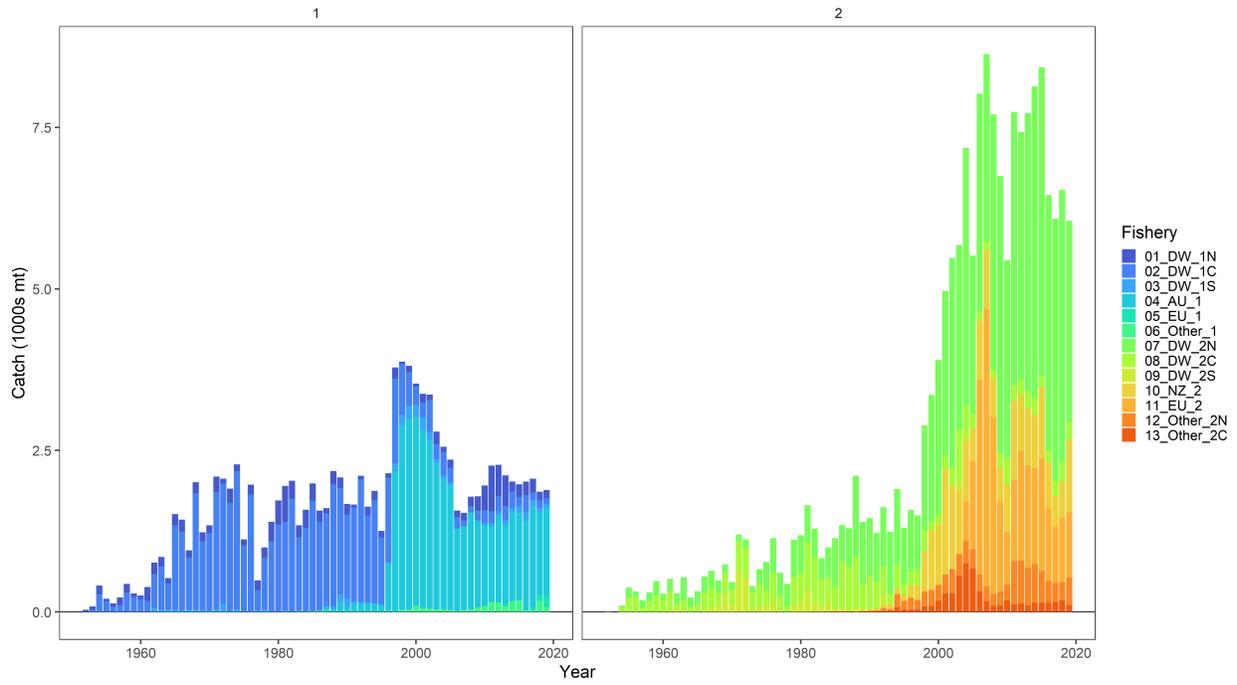


Figure 2: Catch (in mt) of the extraction fisheries for each assessment region.

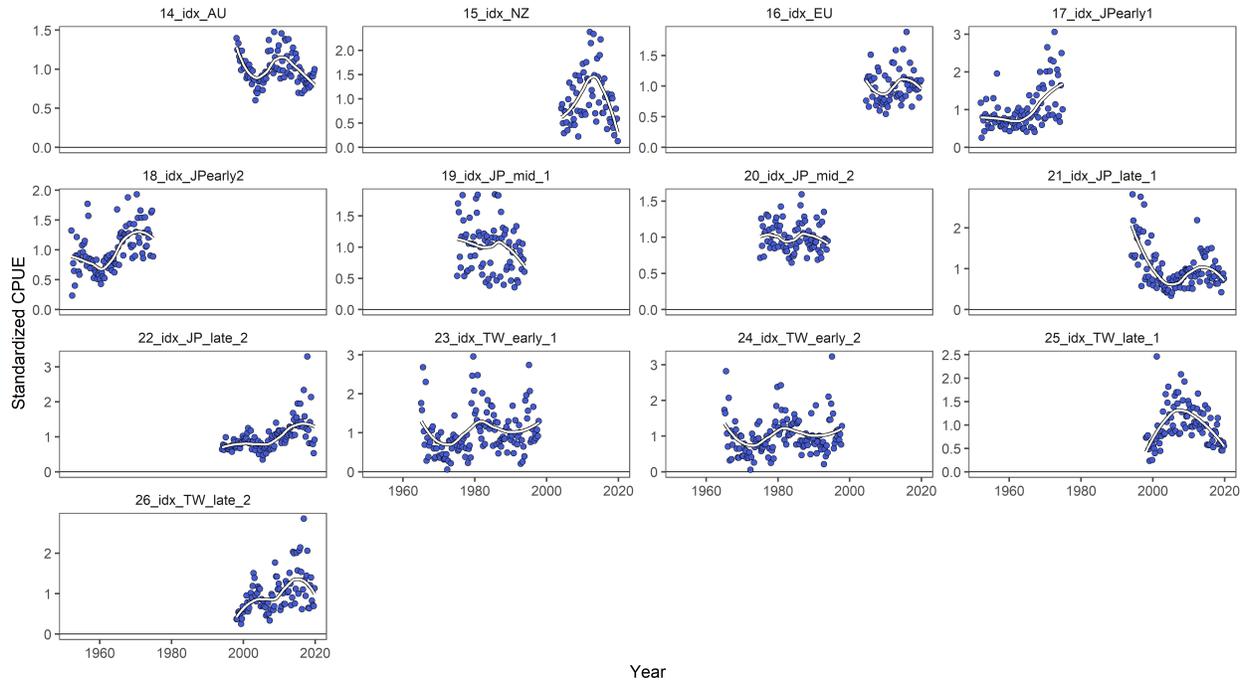
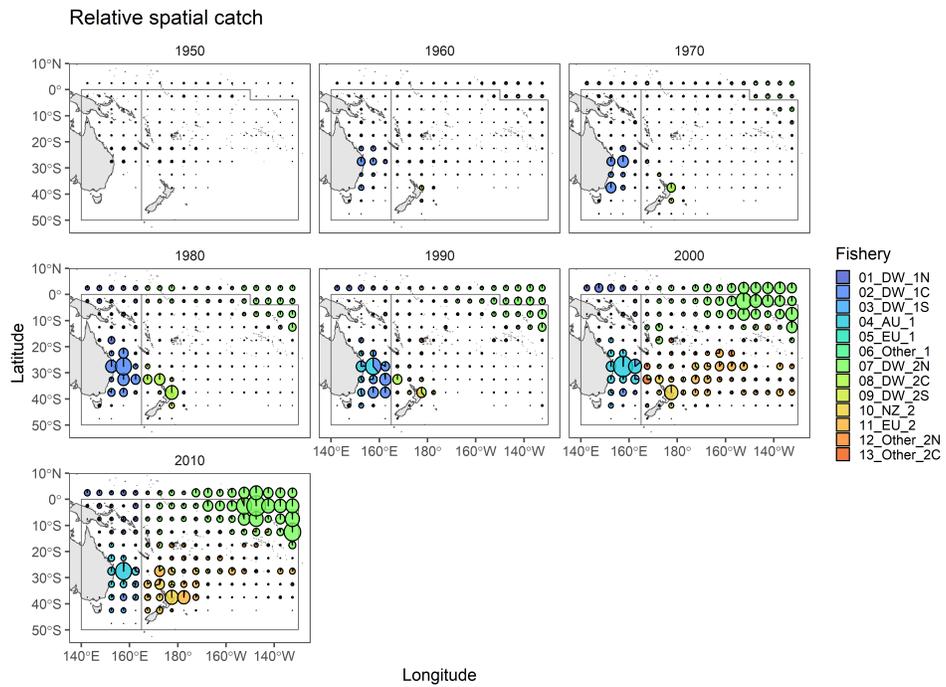
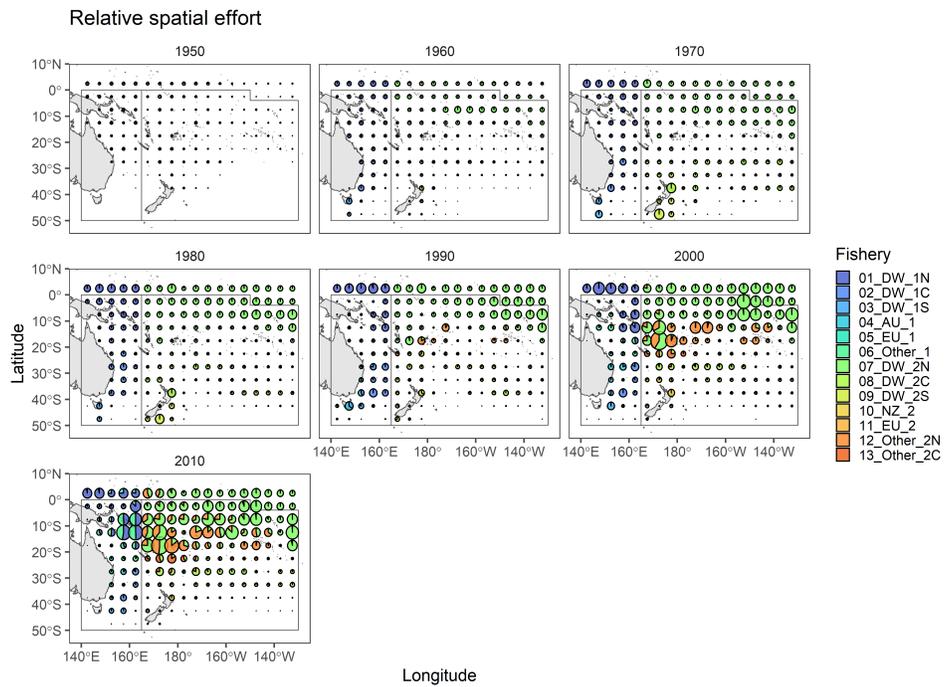


Figure 3: Standardized CPUE by index fishery, scaled to a mean of 1. The colored points indicate the observed quarterly CPUE, while the white line indicates the smoothed trend through the data.

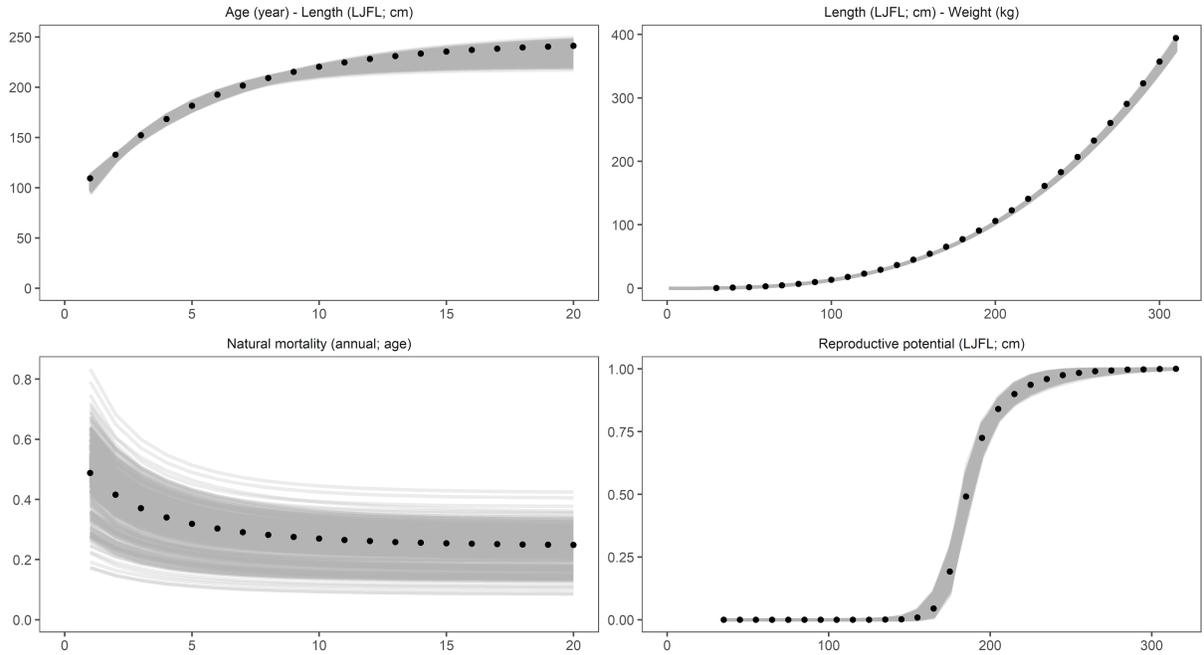


(a) Relative spatial catch by fishery.

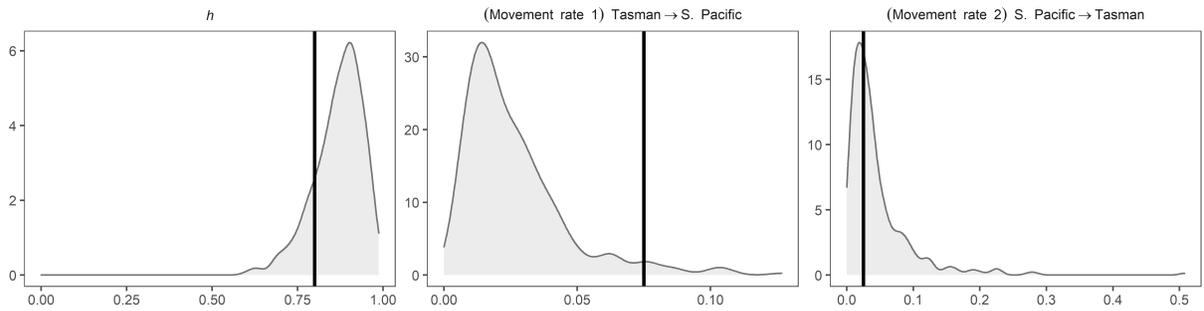


(b) Relative spatial effort by fishery.

Figure 4: Spatial patterns in relative swordfish catch and longline fishing effort by extraction fishery and decade for the SWPO swordfish assessment region and immediately surrounding waters.



(a) Uncertainty in assumed biological relationships.



(b) Uncertainty in steepness and regional movement rates.

Figure 5: Uncertainty in the assumed biological relationships, steepness, and movement rates considered in the model ensemble is shown by the gray lines. The values assumed in the diagnostic case are shown either with the black points in panel a) or the vertical black line in panel b).

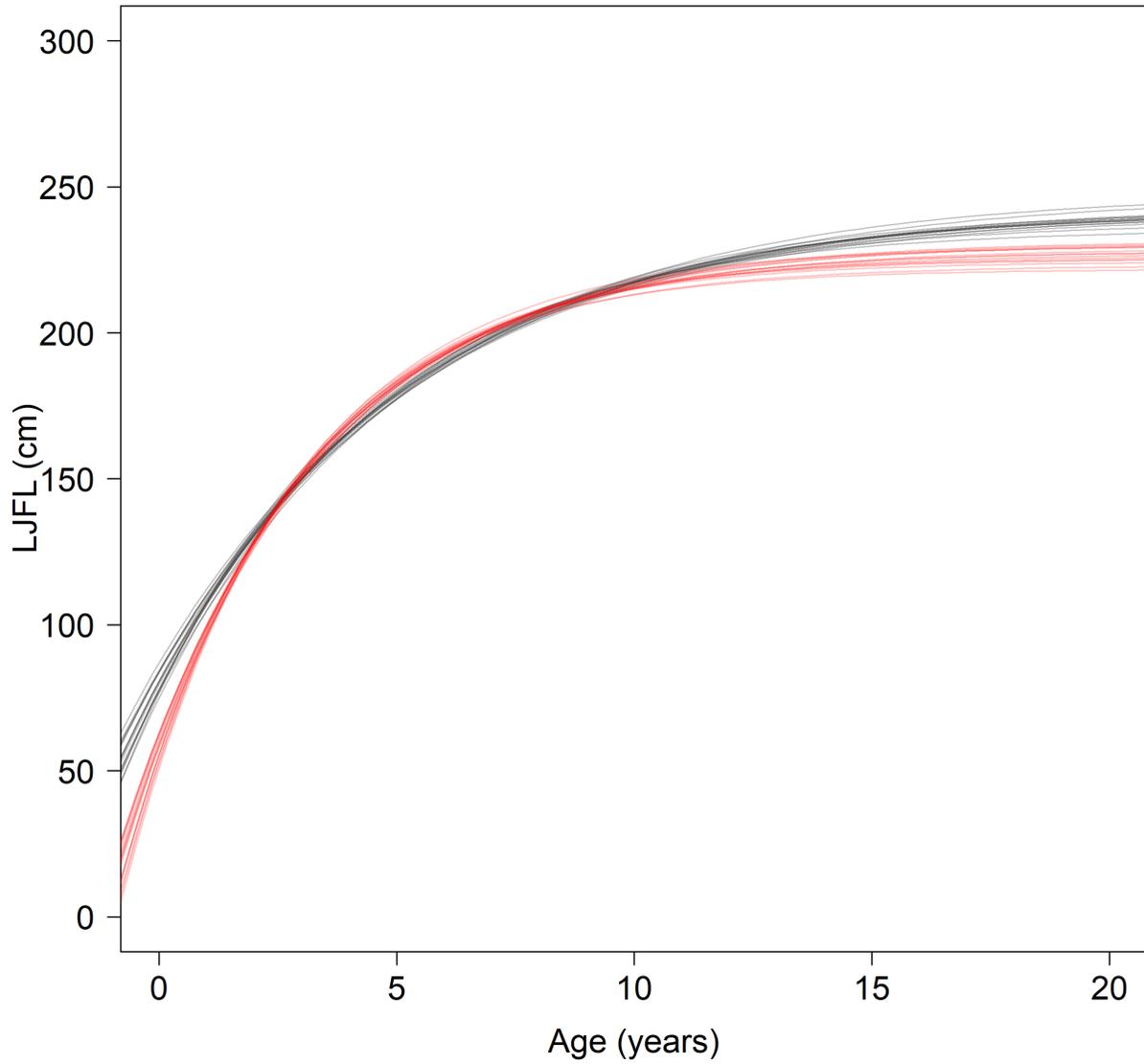
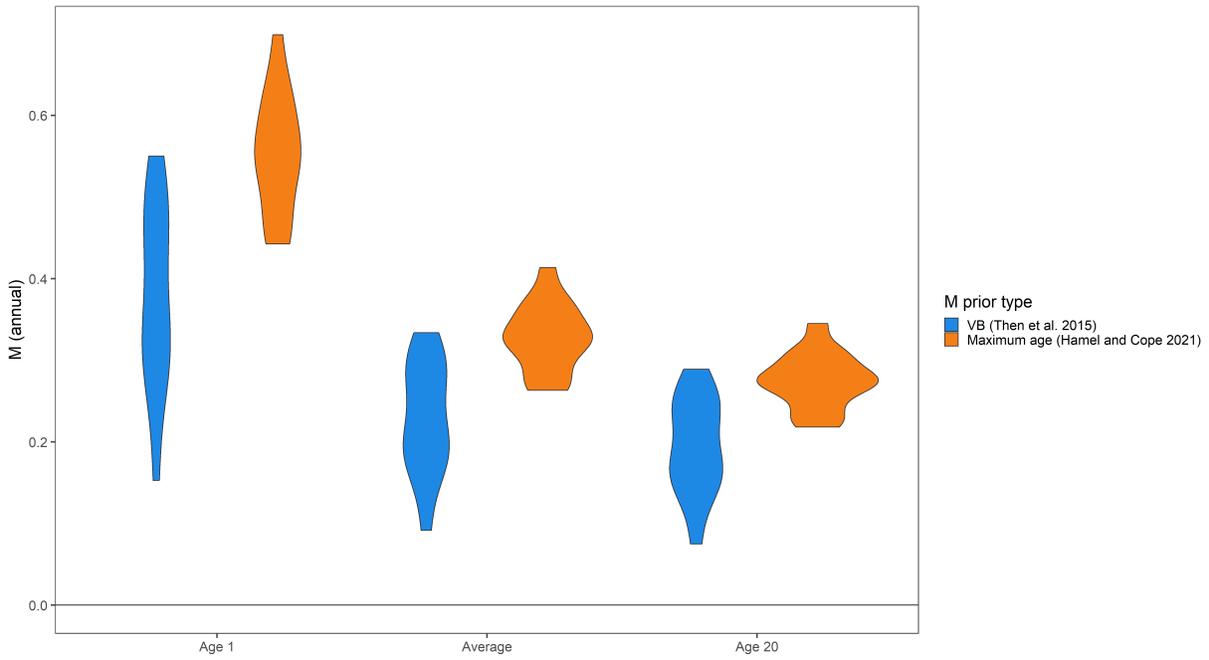
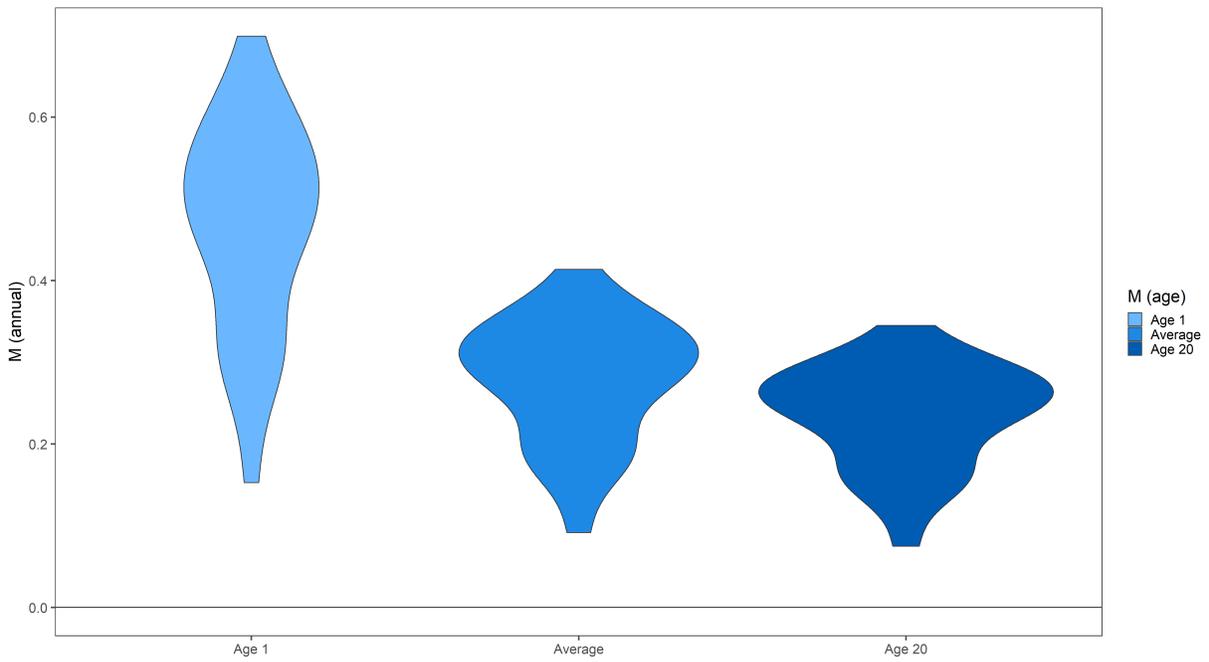


Figure 6: Growth curves considered in the model ensemble based on an un-informative prior for t_0 (black), and a more informative prior, centered on zero, for t_0 (red).

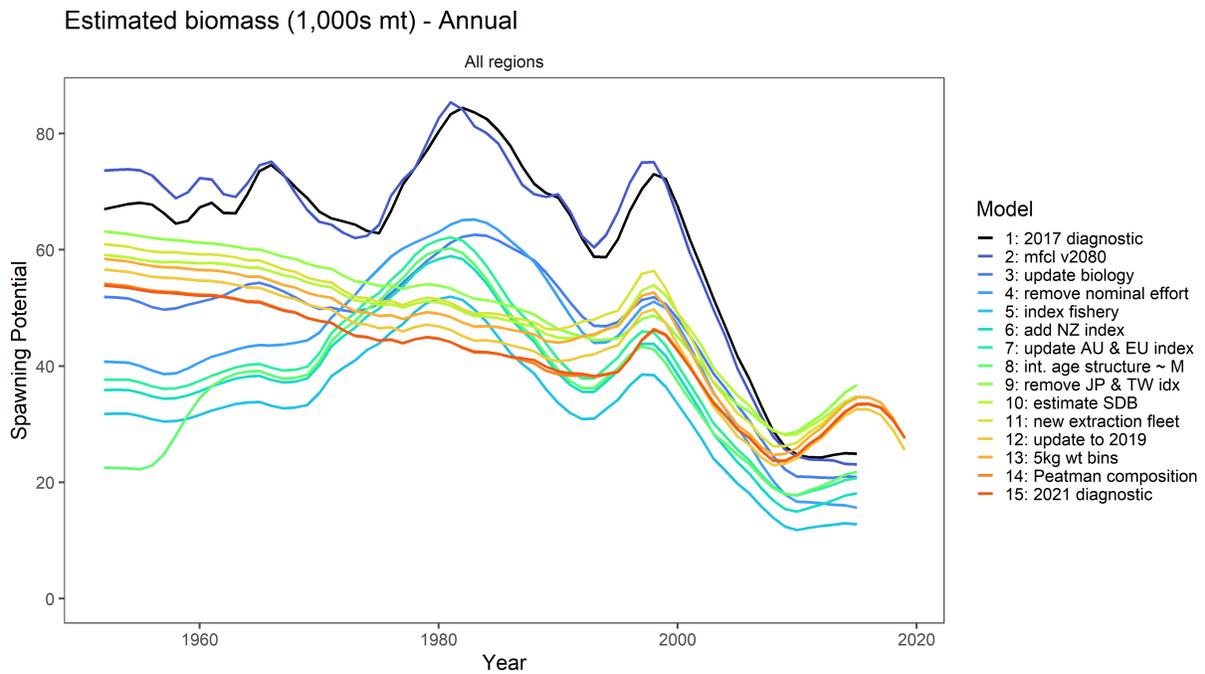


(a) M prior by approach: maximum-age (orange) and life-history based (blue) .

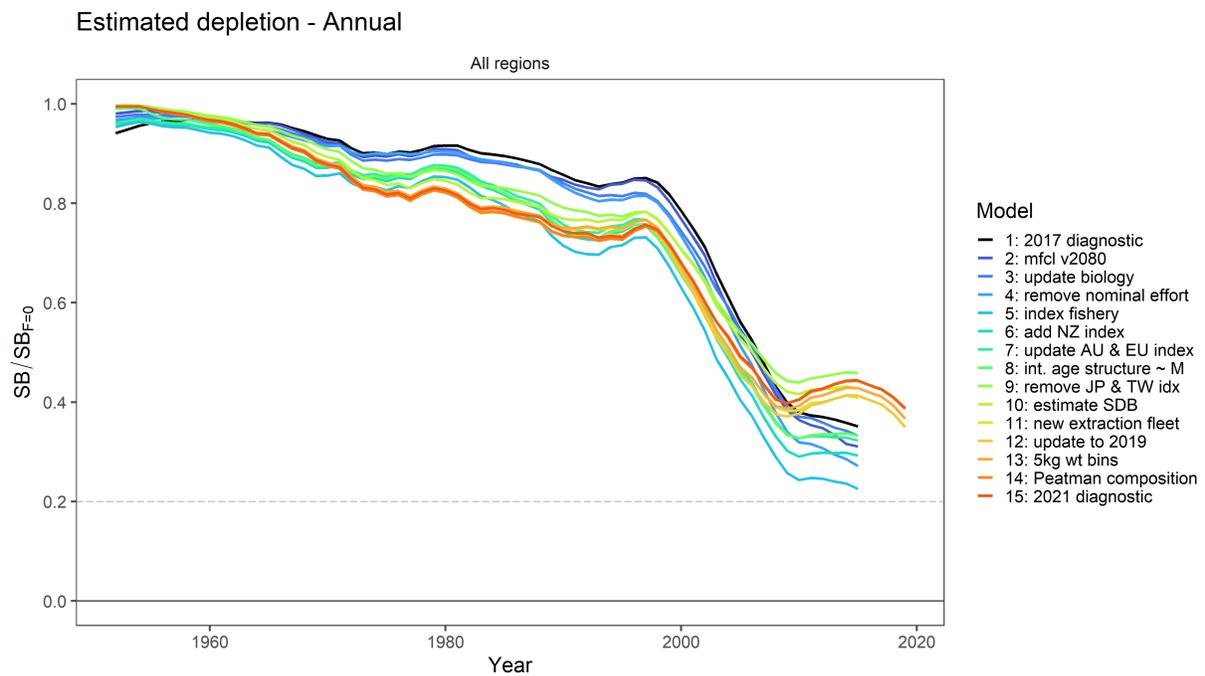


(b) M prior combining the two approaches: maximum-age and life-history based.

Figure 7: Prior distributions for natural mortality (M) considered in the model ensemble with breakdown by M at age 1, 20 (M_{ref}) and average over all ages.



(a) Stepwise change in spawning potential from 2017 diagnostic model to 2021 diagnostic model.



(b) Stepwise change in depletion from 2017 diagnostic model to 2021 diagnostic model.

Figure 8: Model progression from the 2017 to 2021 diagnostic cases in terms of spawning potential and depletion.

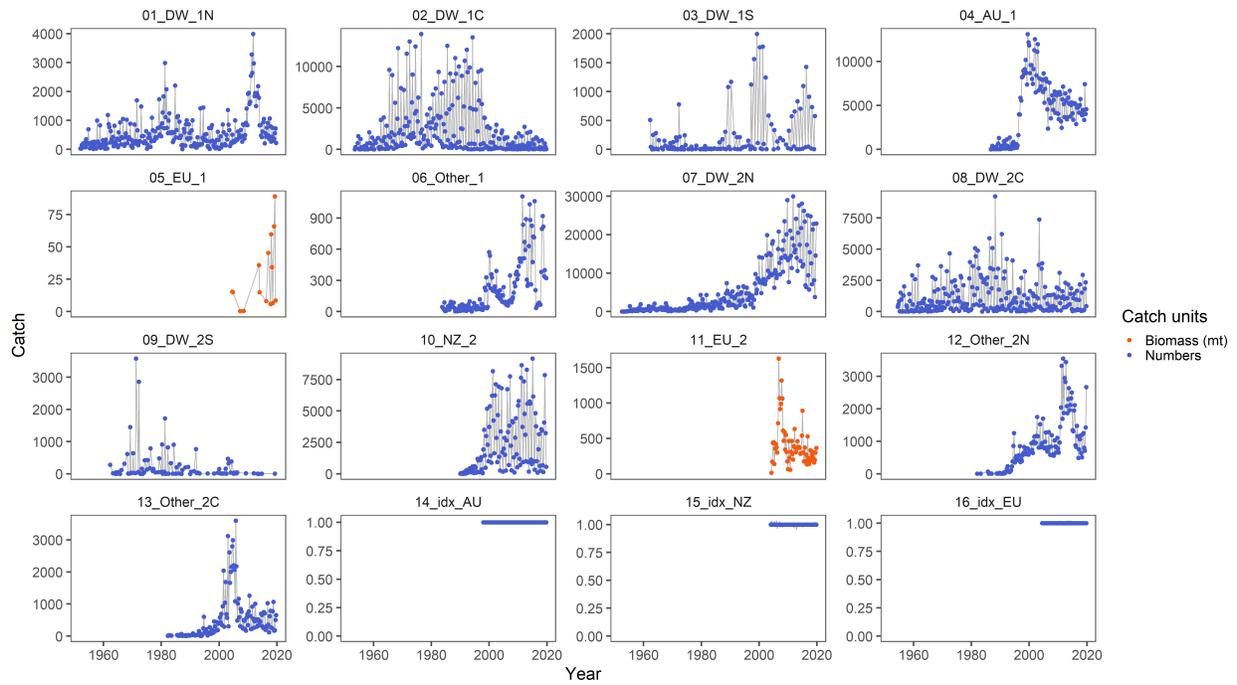


Figure 9: Model fit (gray line) to the observed catch (points) by fishery. The color of the point denotes the unit of catch: orange (biomass) and blue (numbers).

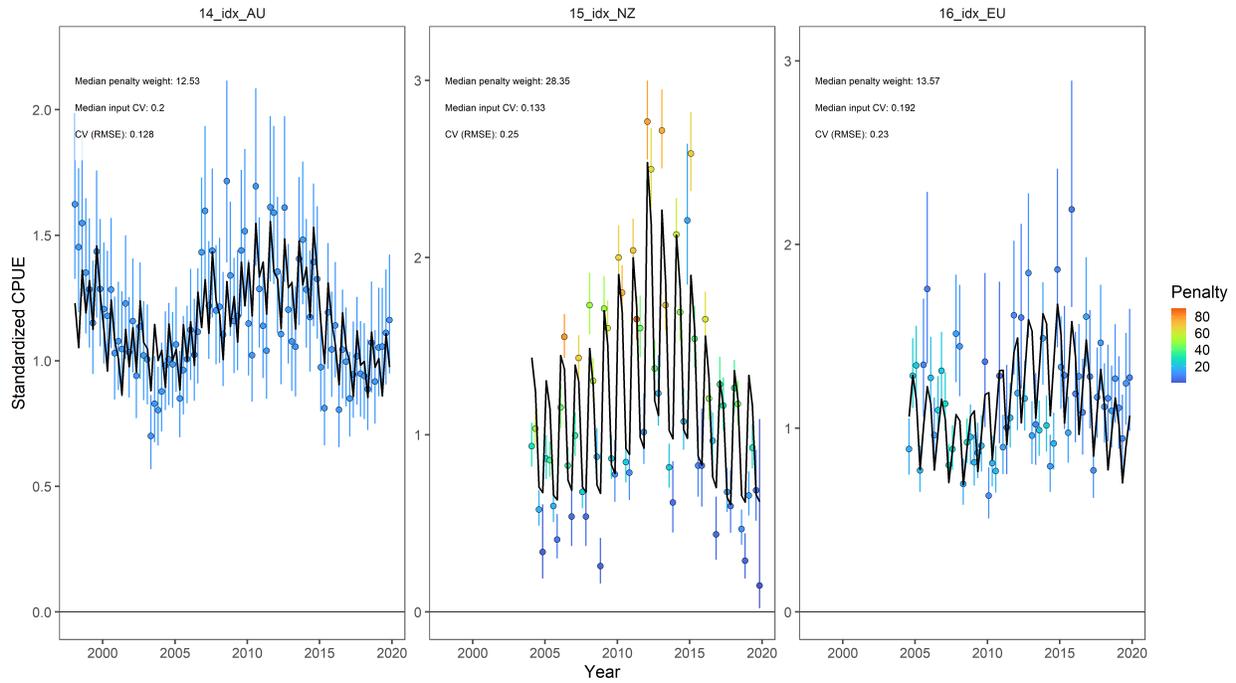


Figure 10: Model fit (black line) to the observed standardized CPUE (points) by index fishery. The color of the point denotes the input penalty associated with that observation. A large penalty indicates greater precision in the estimate which should pull the estimate closer to the observation. Within each panel the median input penalty weight, equivalent median input CV, and the RMSE of the fit is shown for each fishery.

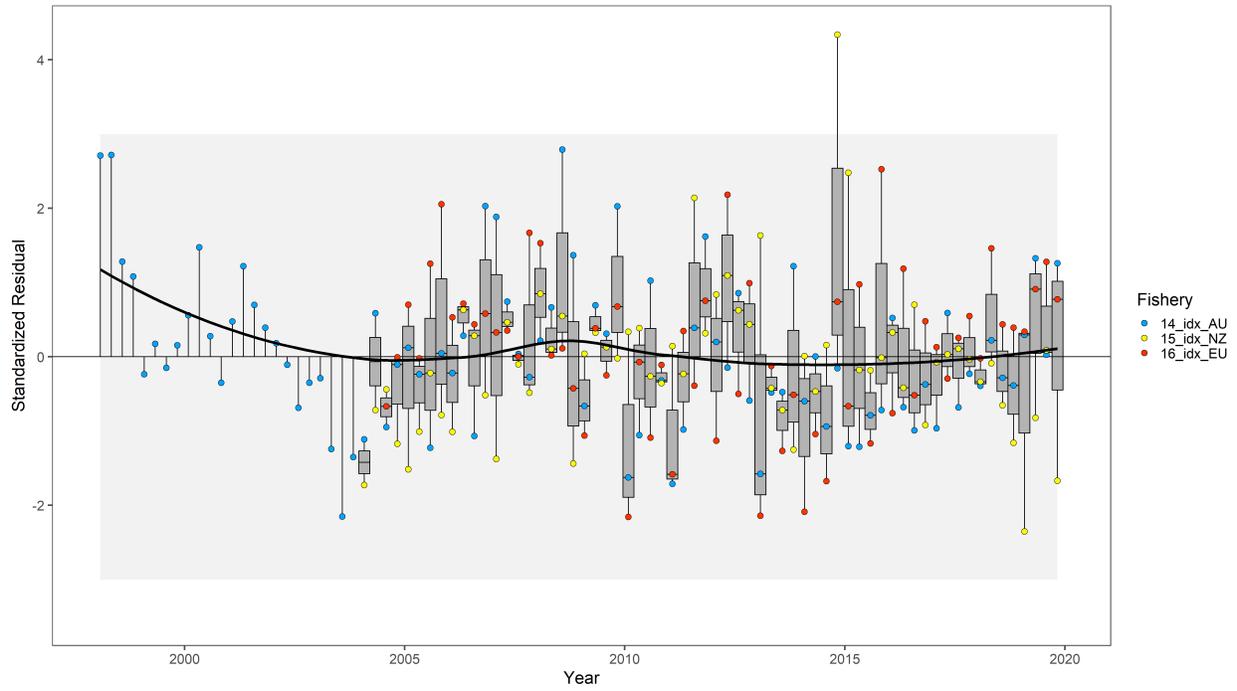


Figure 11: Joint residual plot following [Winker et al. \(2018\)](#) combining standardized residuals of the fit to the CPUE where color denotes the fishery, boxplots indicating the median and quantiles across all residuals by time step, and a smoothed spline through all residuals showing the overall trend of the fit. Additionally, the light gray background box indicates the 3σ rule used to identify outliers.

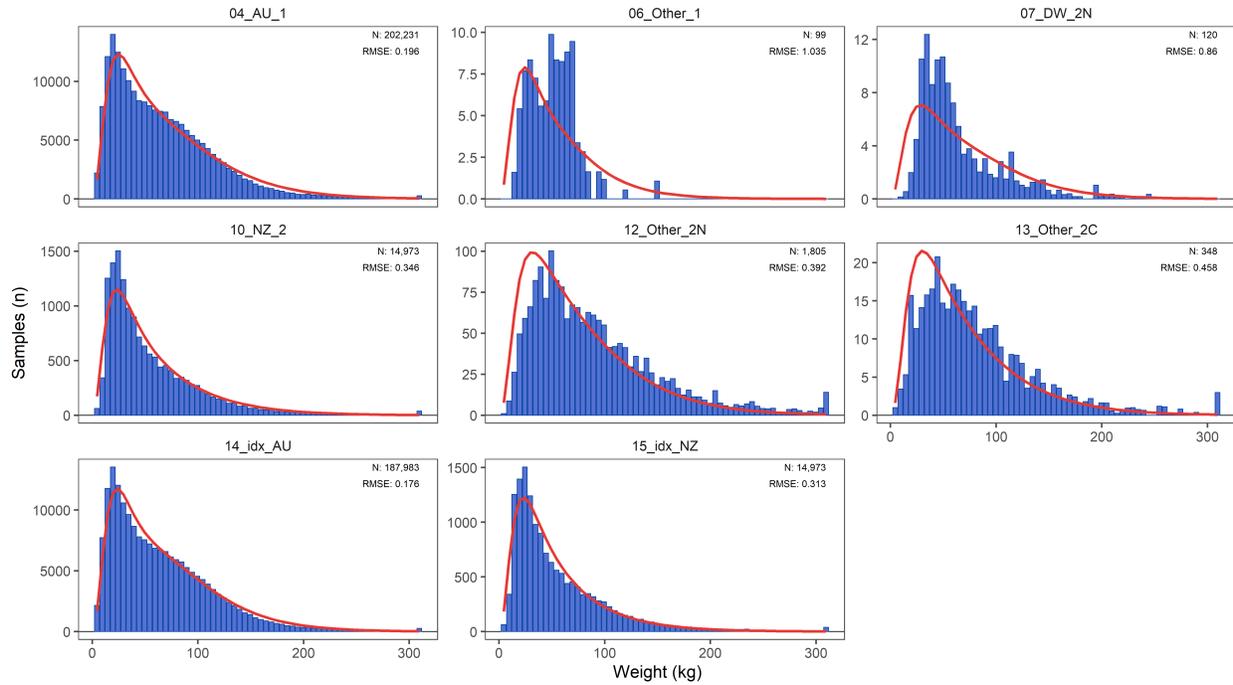


Figure 12: Histograms of the observed weight samples by 5kg bin, where the predicted numbers are shown by the red line. For each fishery the input sample size is given along with the RMSE of the fit calculated on the basis of observed and predicted numbers.

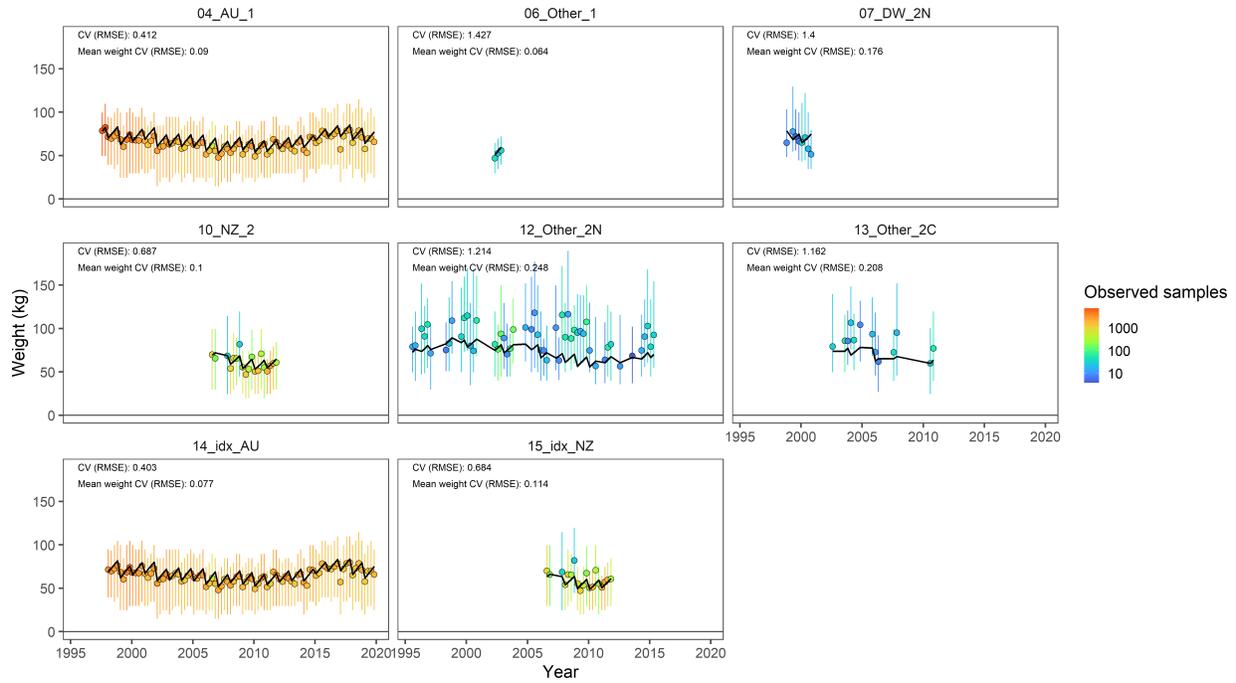


Figure 13: Temporal trend in observed (points) and predicted (line) average weight. The average observed weight is given by the colored point (color indicates input sample size), with the vertical line segment indicating the 50th quantile of observations. The RMSE calculated on the basis of the observed and predicted proportions at weight, along with the mean weight RMSE calculated on the basis of observed and predicted average weight are given for each fishery.

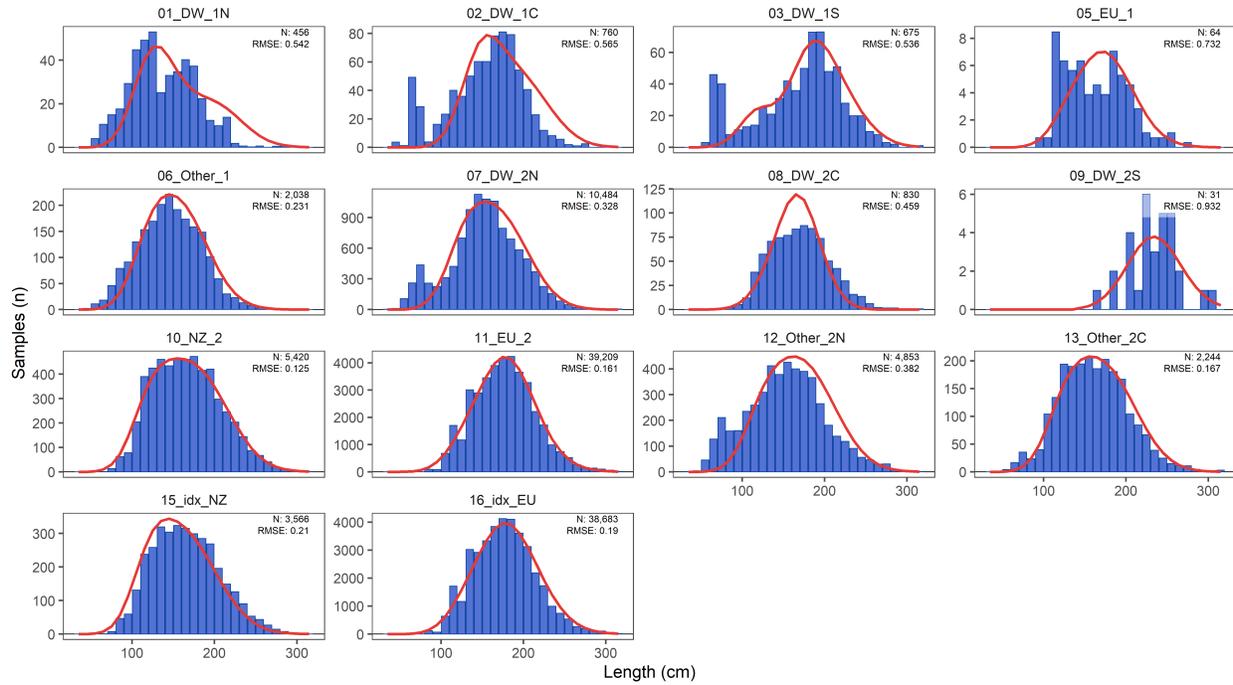


Figure 14: Histograms of the observed length samples by 10cm bin, where the predicted numbers are shown by the red line. For each fishery the input sample size is given along with the RMSE of the fit calculated on the basis of observed and predicted numbers.

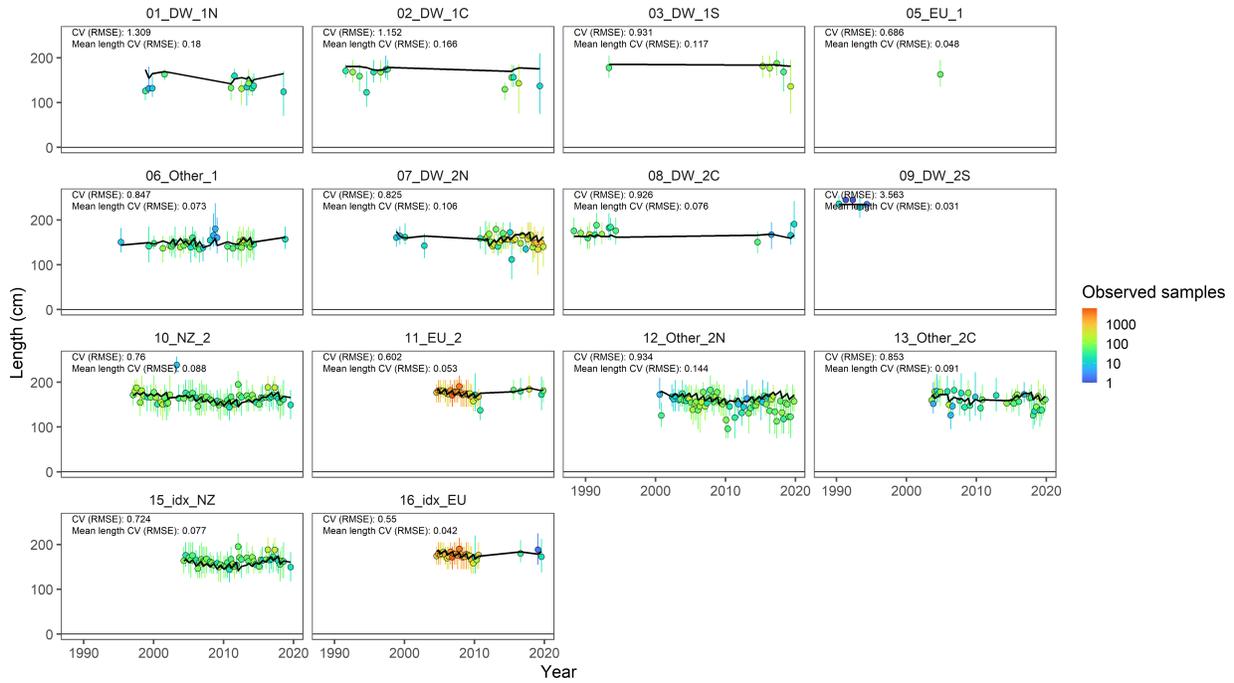
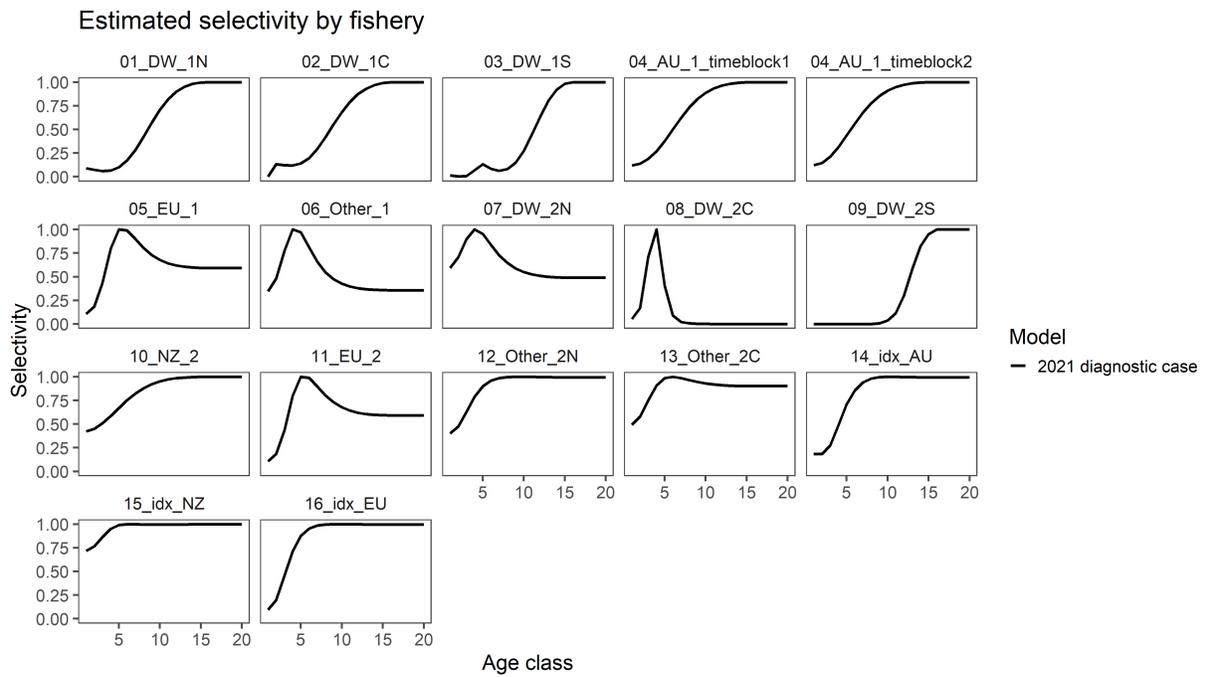
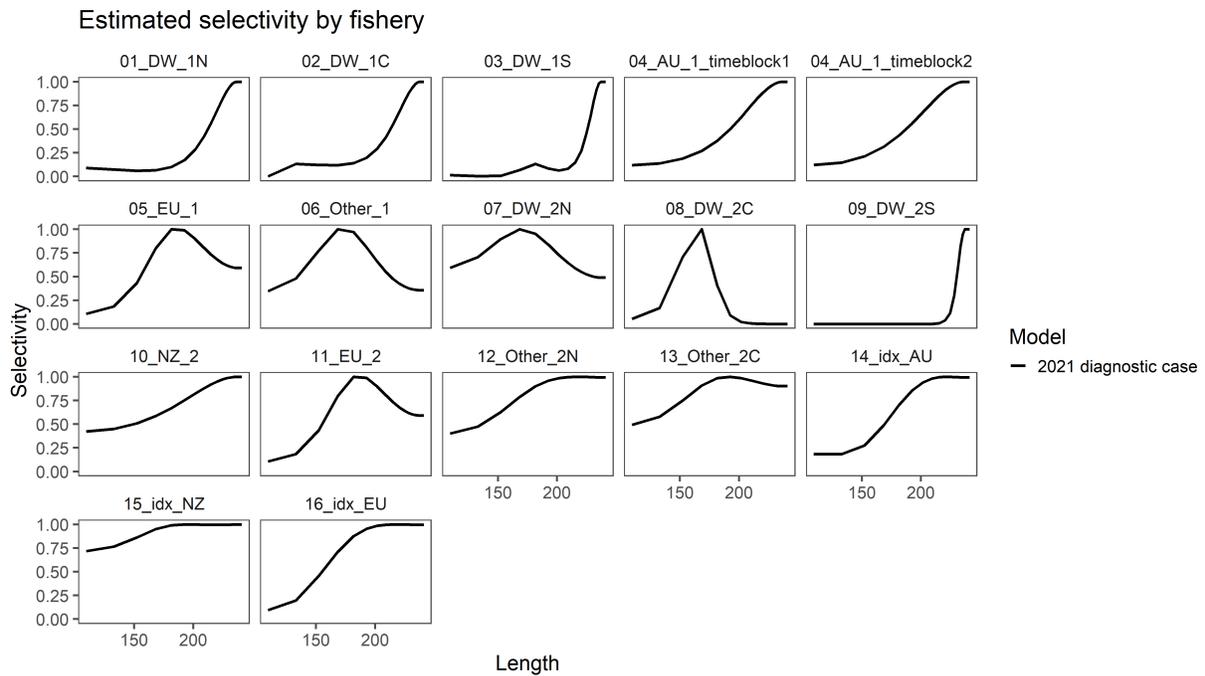


Figure 15: Temporal trend in observed (points) and predicted (line) average length. The average observed length is given by the colored point (color indicates input sample size), with the vertical line segment indicating the 50th quantile of observations. The RMSE calculated on the basis of the observed and predicted proportions at length, along with the mean length RMSE calculated on the basis of observed and predicted average length are given for each fishery.



(a) Selectivity-at-age by fishery.



(b) Selectivity-at-length by fishery.

Figure 16: Estimated selectivity curves by fishery, presented as a function of (a) age and (b) length for the diagnostic case model.

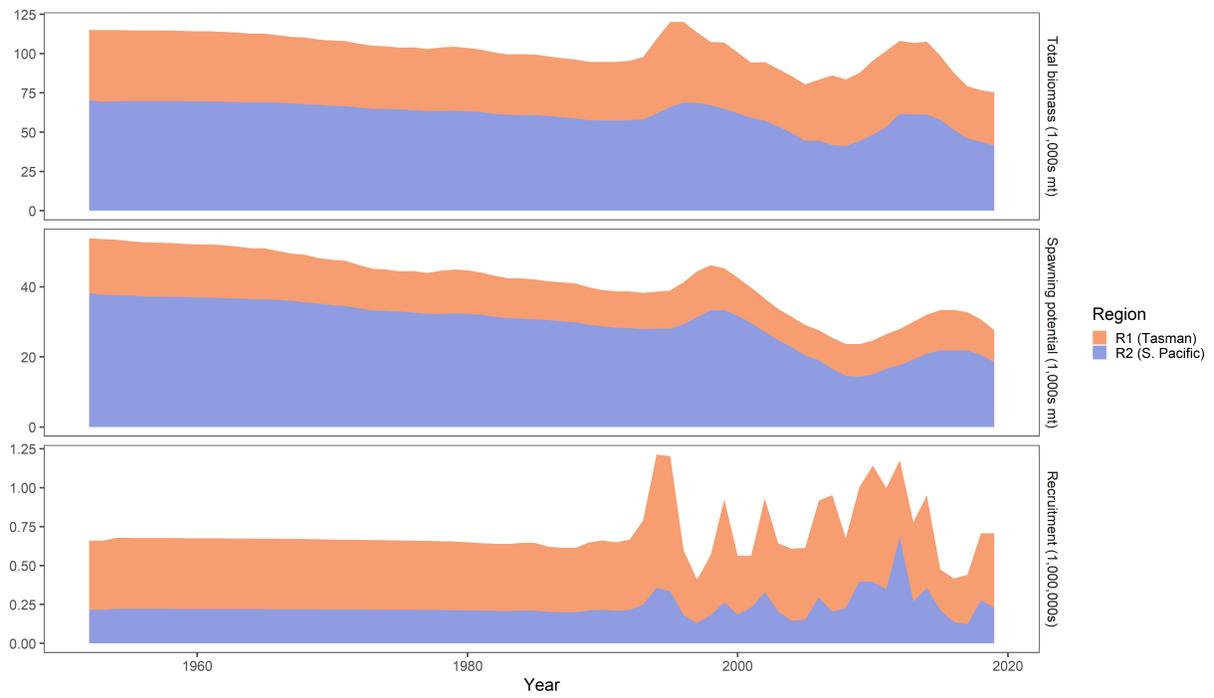
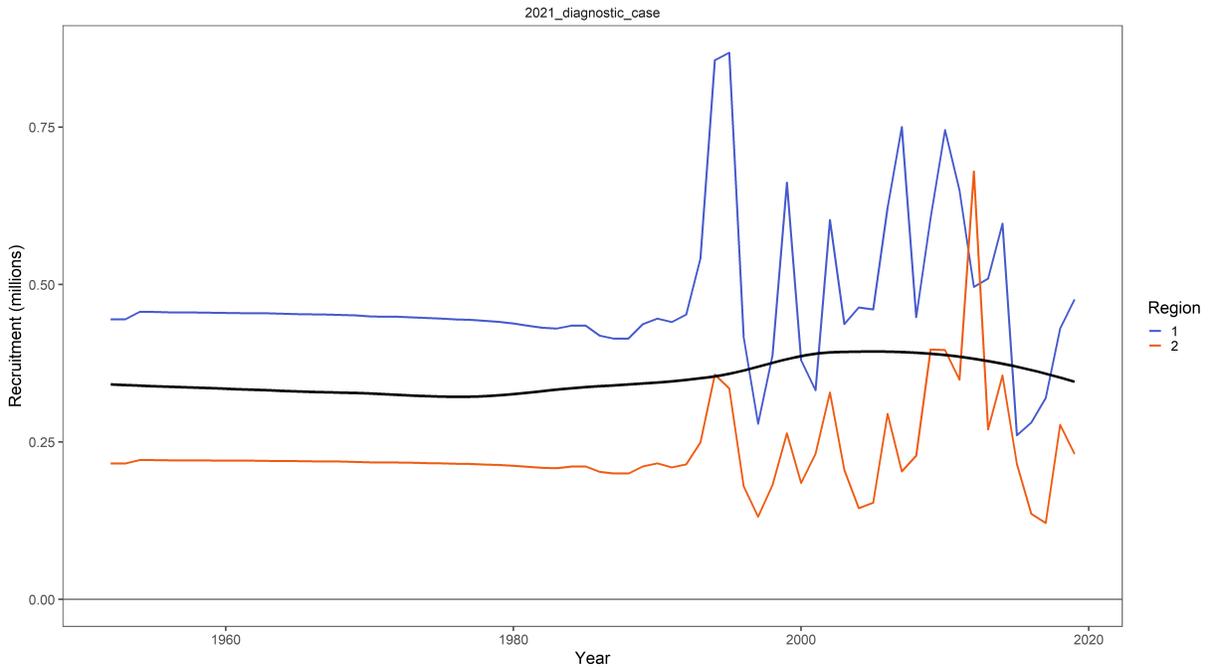
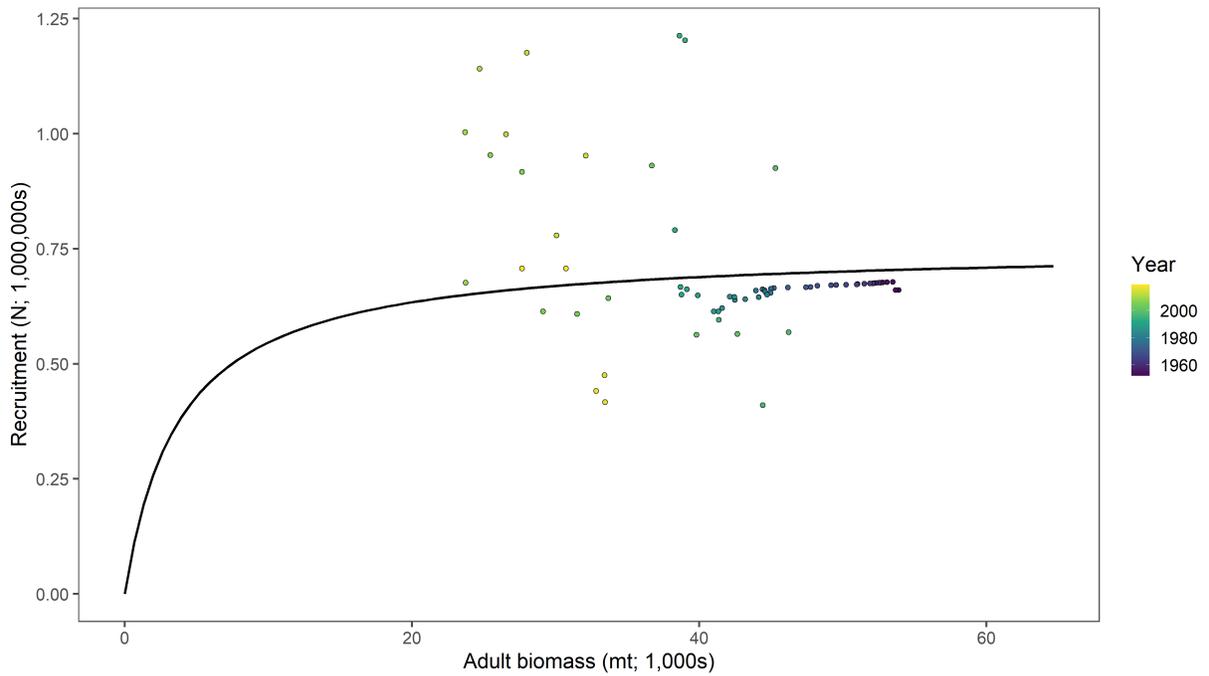


Figure 17: Estimated total biomass (top panel), spawning potential (middle panel), and recruitment (lower panel) for the diagnostic case model. Color indicates the model region: region 1 (orange) and region 2 (blue).



(a) Total recruitment in region 1 (blue) and region 2 (orange) with aggregate trend (black).



(b) Beverton-Holt stock recruitment relationship.

Figure 18: Total estimated regional recruitment (a) and the stock-recruit relationship (b) for the diagnostic case model.

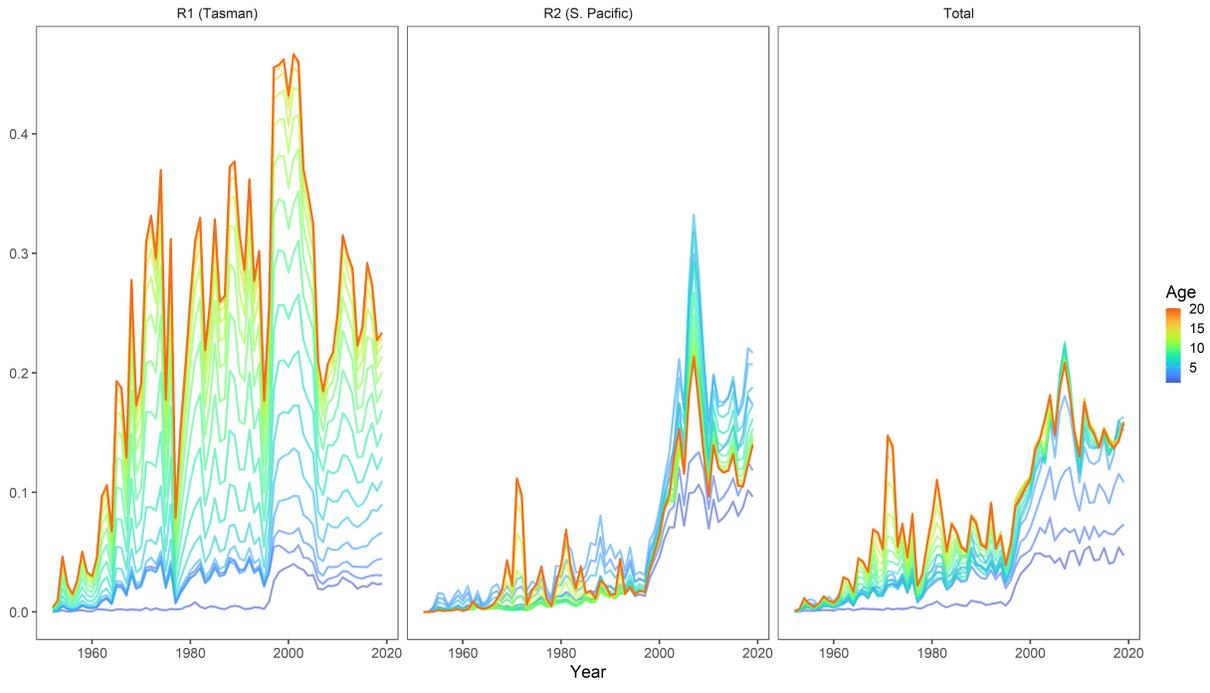


Figure 19: Fishing mortality by age (color) and region (panel: region 1 - left, region 2 - center, and total - right).

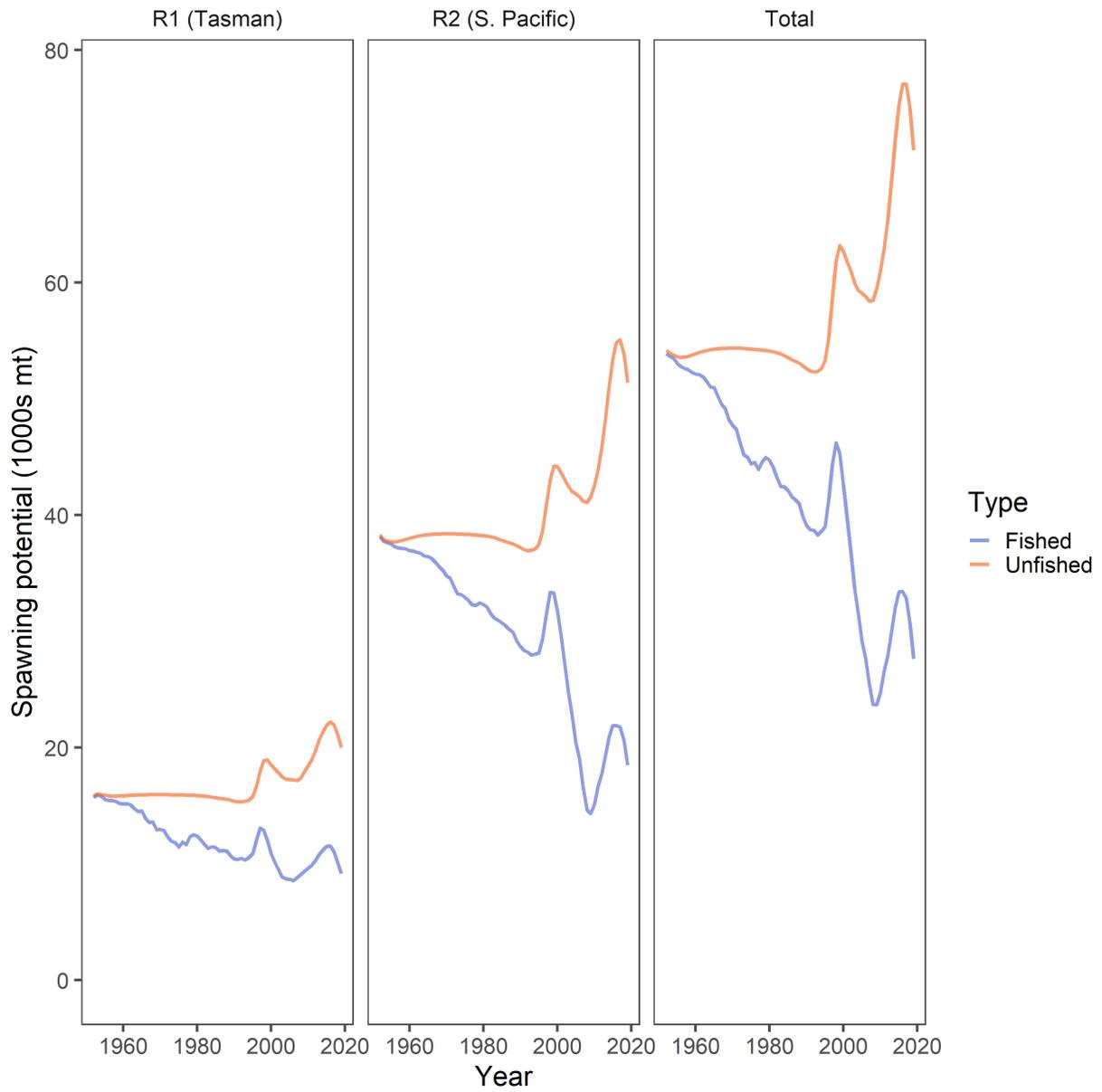
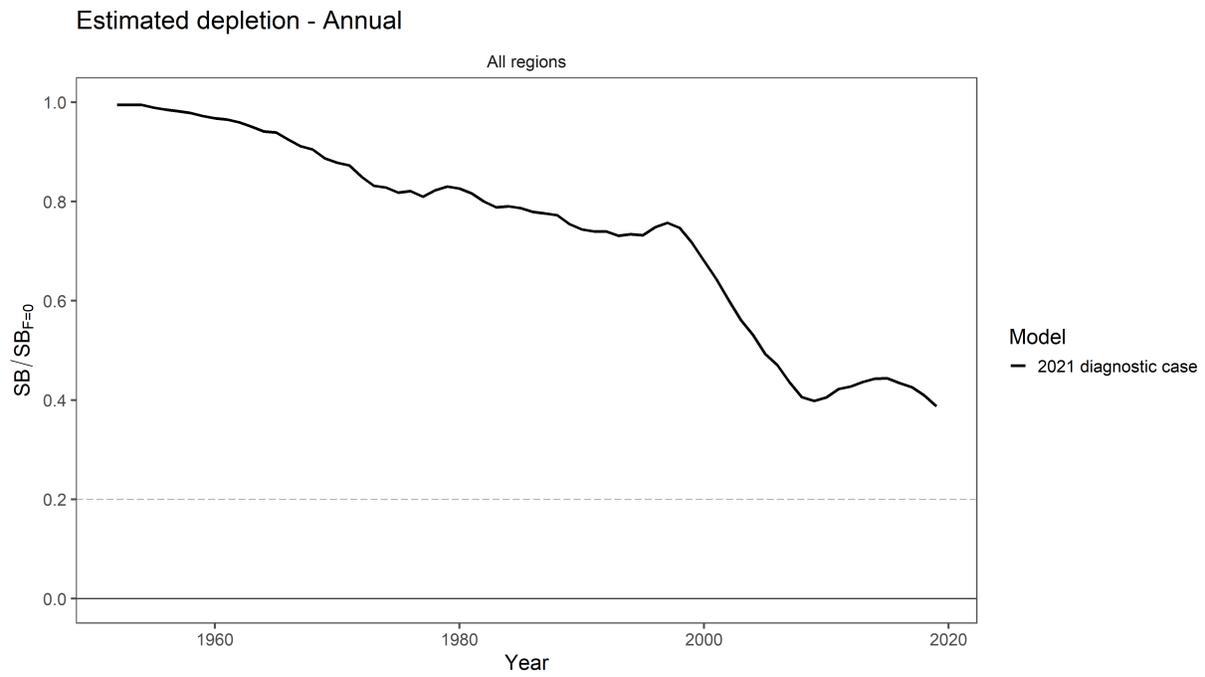
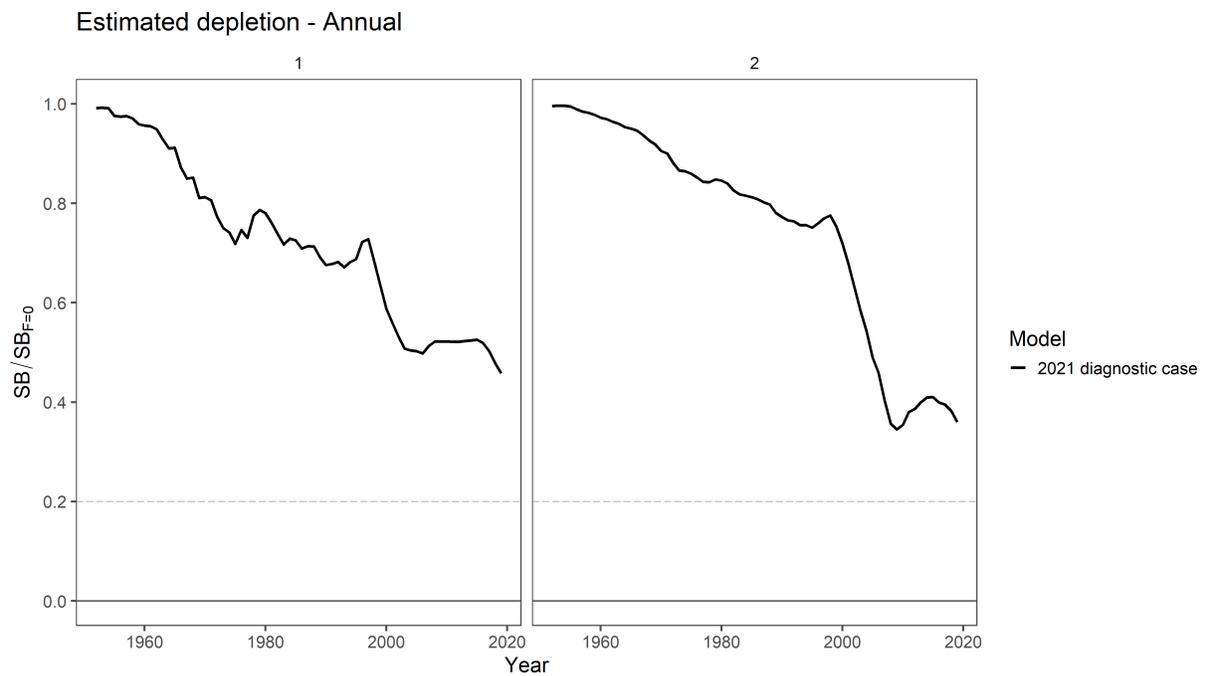


Figure 20: Comparison of the estimated fished (blue) and unfished (orange) trajectories for the total assessment area and by region.



(a) Total depletion.



(b) Depletion by region.

Figure 21: Estimated depletion in spawning potential $SB_t/SB_{t,F=0}$. The dotted line indicates $20\%SB_{F=0}$.

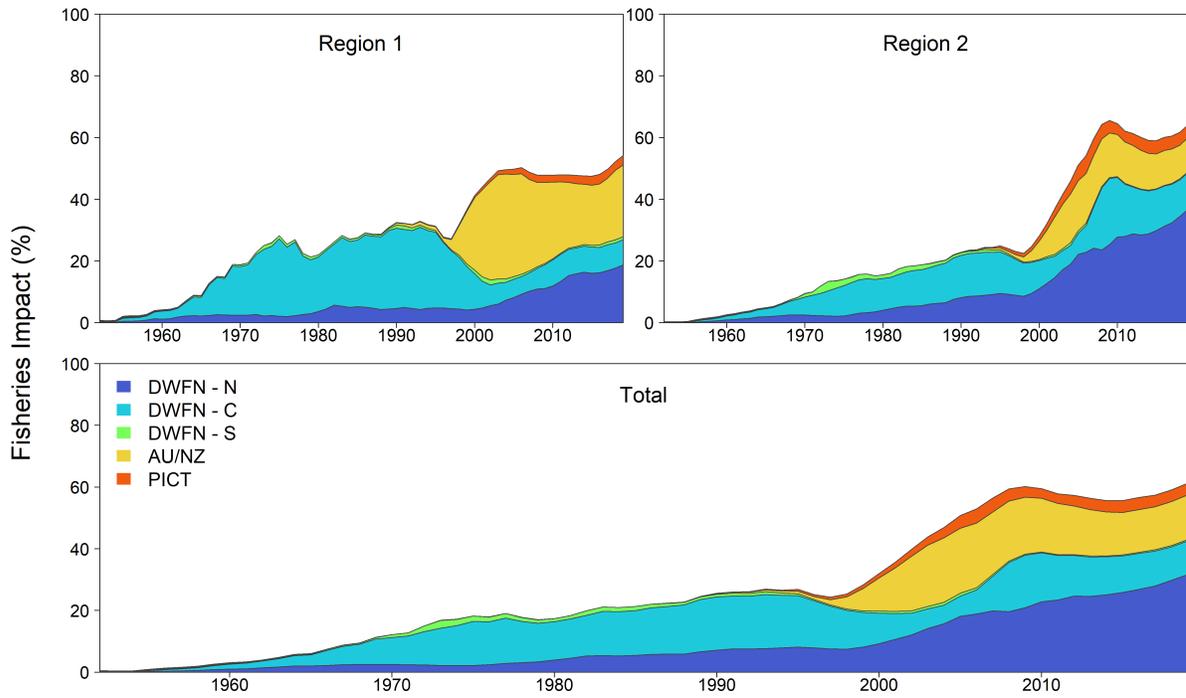


Figure 22: Estimates of reduction in spawning potential due to fishing (fishery impact = $1 - SB_t/SB_{t,F=0}$) for the total assessment area and by region. Impact is attributed to the DWFN fisheries (operating in the northern, central, and southern regions), domestic longline (Australia and New Zealand) fisheries, and PICT longline fisheries.

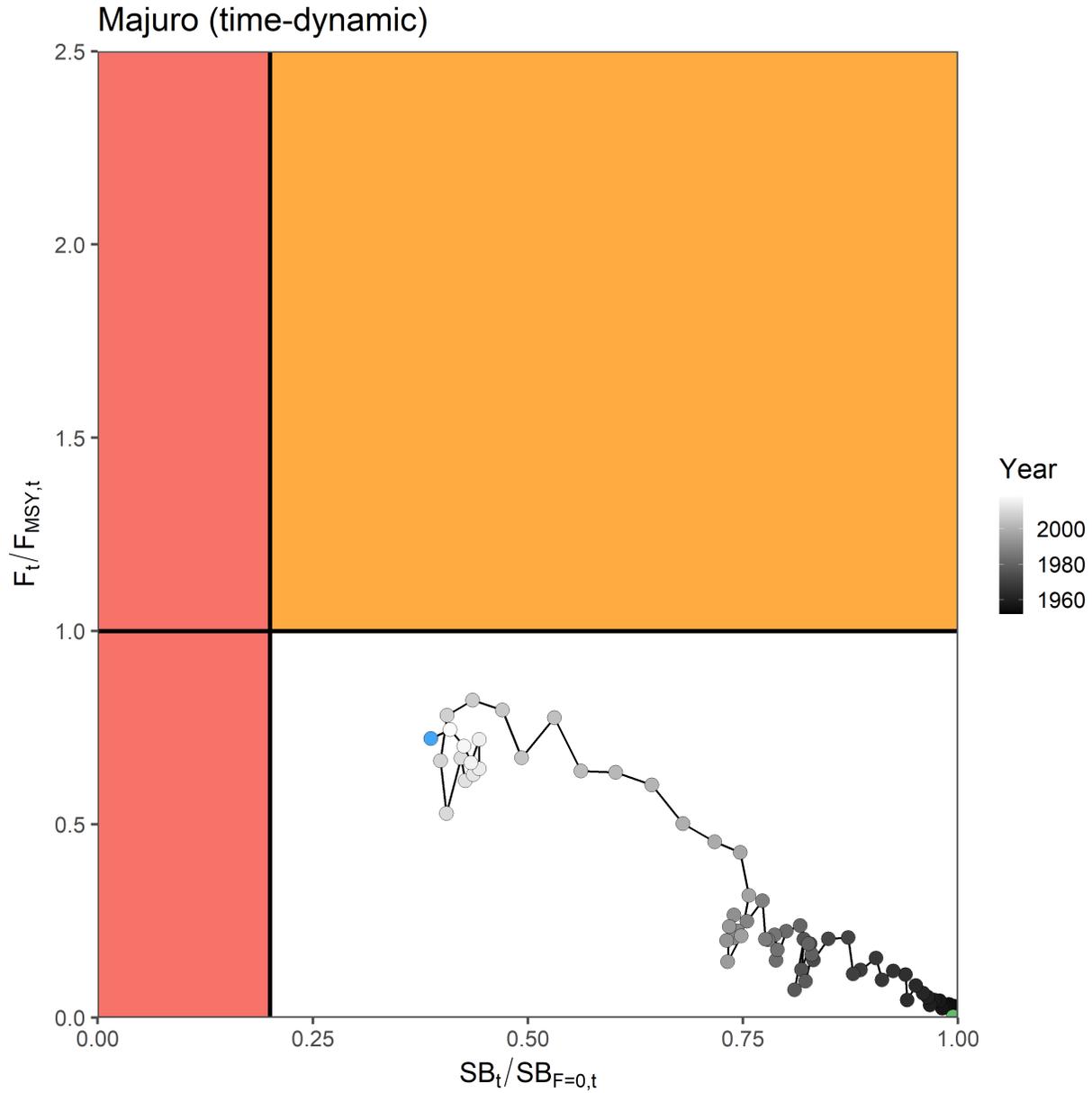


Figure 23: Time-dynamic Majuro plot for the diagnostic case model representing stock status in terms of spawning potential depletion and fishing mortality. The red zone indicates spawning levels lower than $20\%SB_{F=0}$. The orange region is for F_{recent}/F_{MSY} levels greater than 1. The green point indicates the condition in 1952, with points getting progressively lighter until the blue point indicates the status in the terminal year (2019).

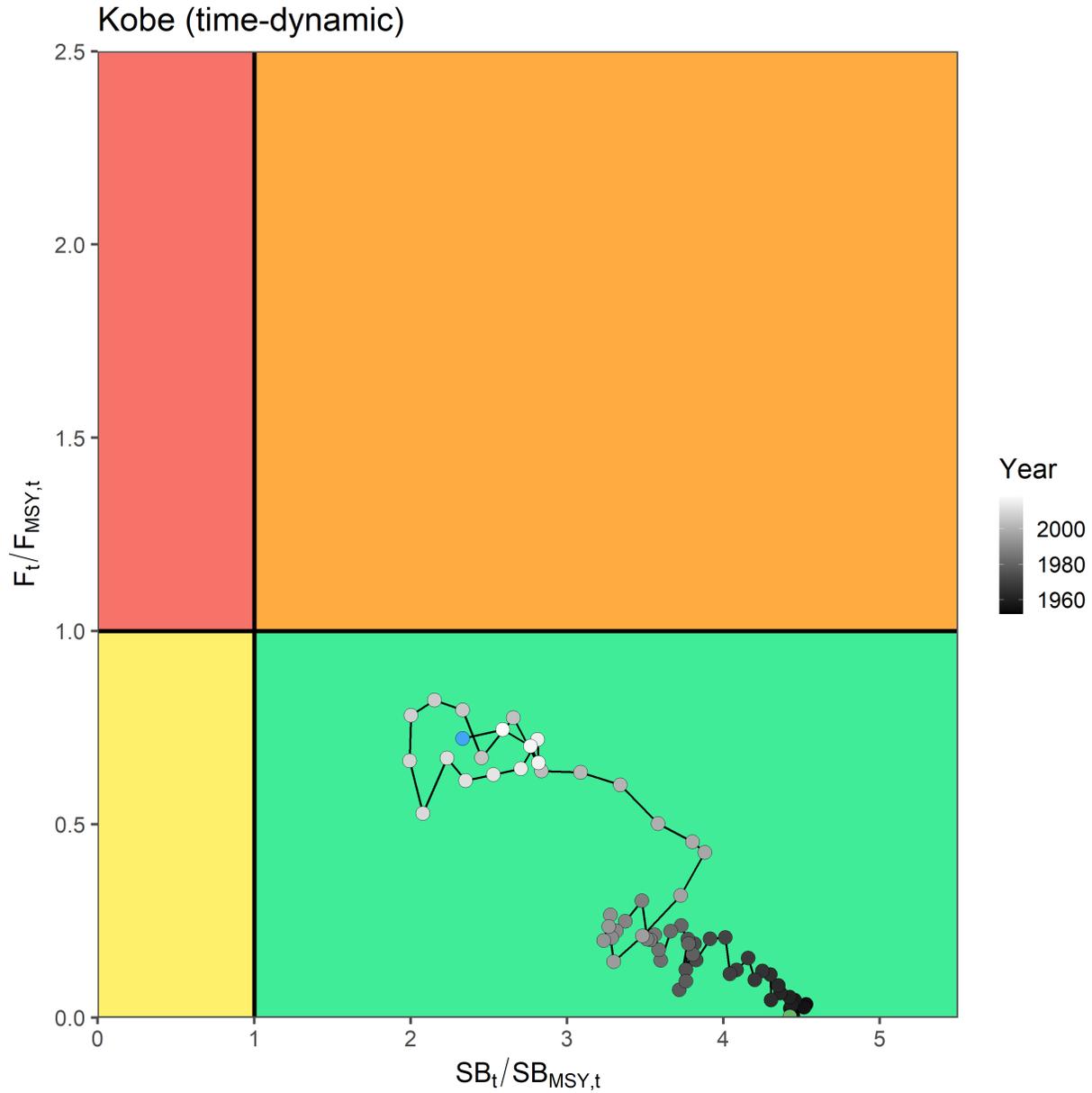


Figure 24: Time-dynamic Kobe plot for the diagnostic case model representing stock status in terms of spawning potential depletion and fishing mortality. The red zone indicates $SB_t/SB_{t,MSY}$ less than 1 and F_{recent}/F_{MSY} levels greater than 1. The green point indicates the condition in 1952, with points getting progressively lighter until the blue point indicates the status in the terminal year (2019).

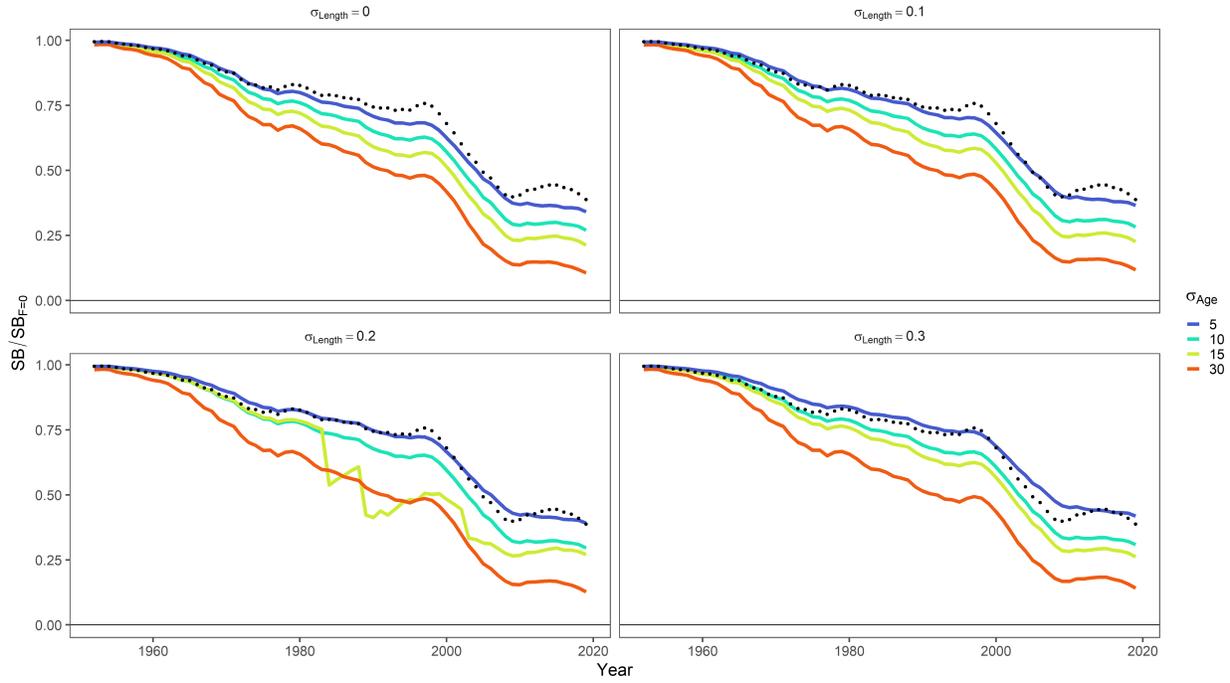
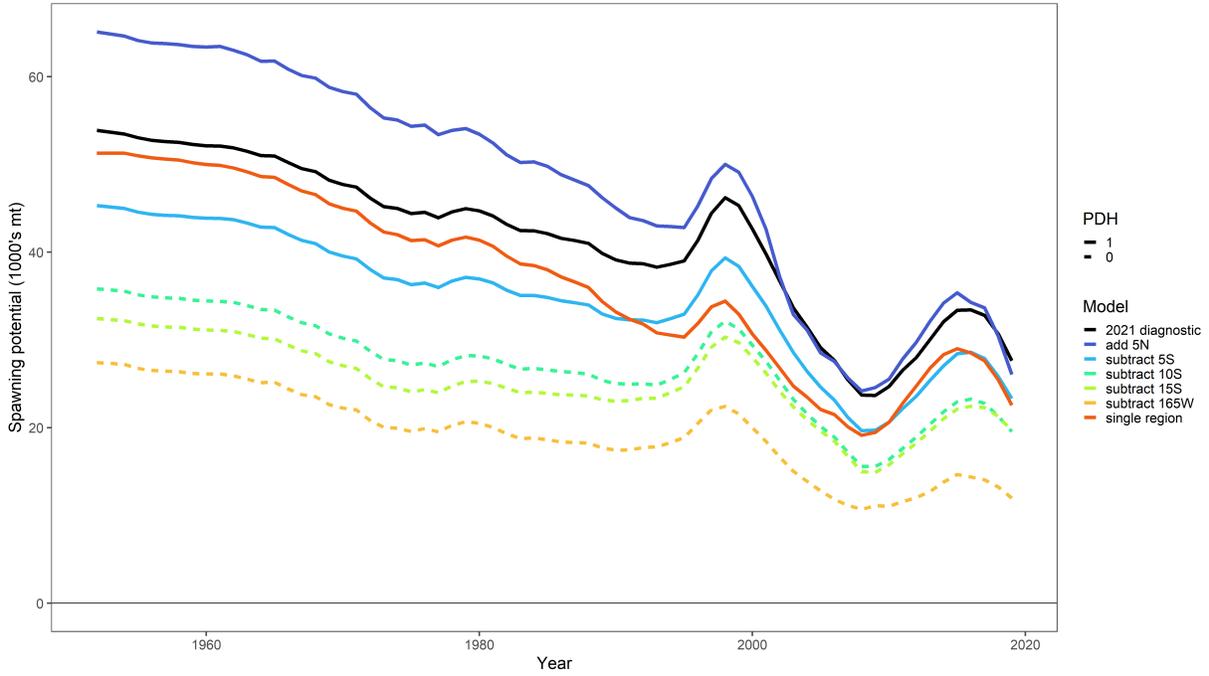
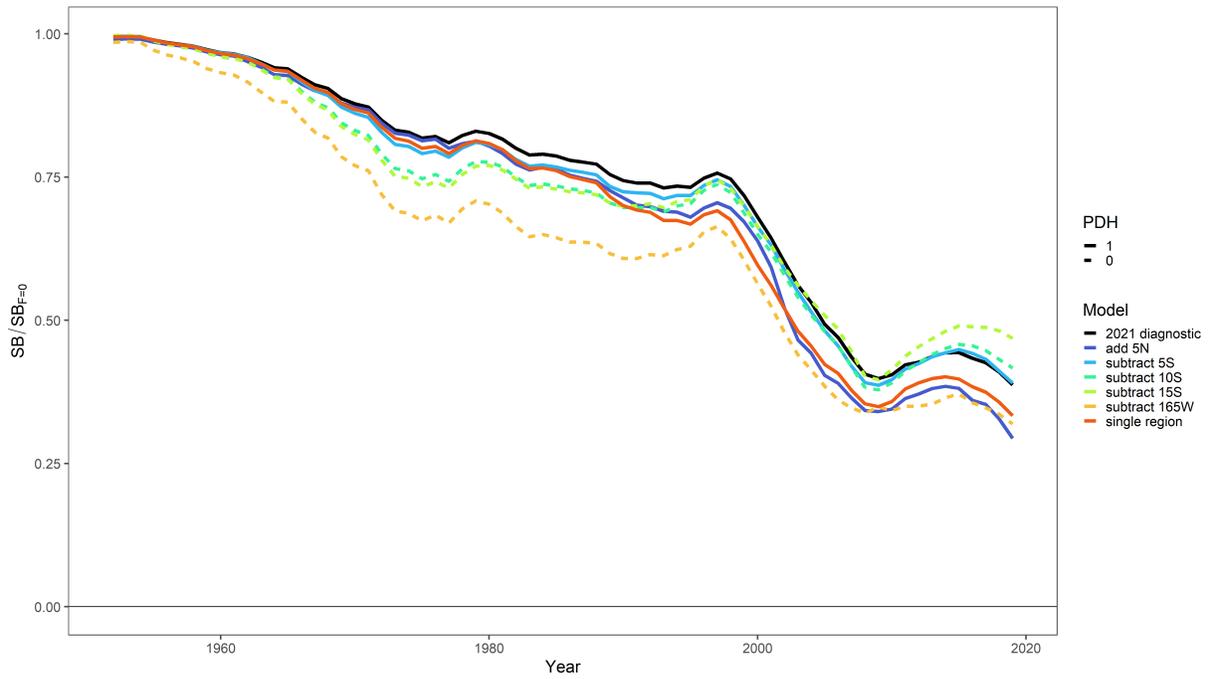


Figure 25: Depletion trajectories for female biomass from the sex-disaggregated model sensitivity (lines) where the color indicates the assumed σ_{Age} . The panels specify the assumption for σ_{Length} . In each panel the depletion trajectory for spawning potential from the sex-aggregated 2021 SWPO swordfish diagnostic case is shown for comparison (black points).

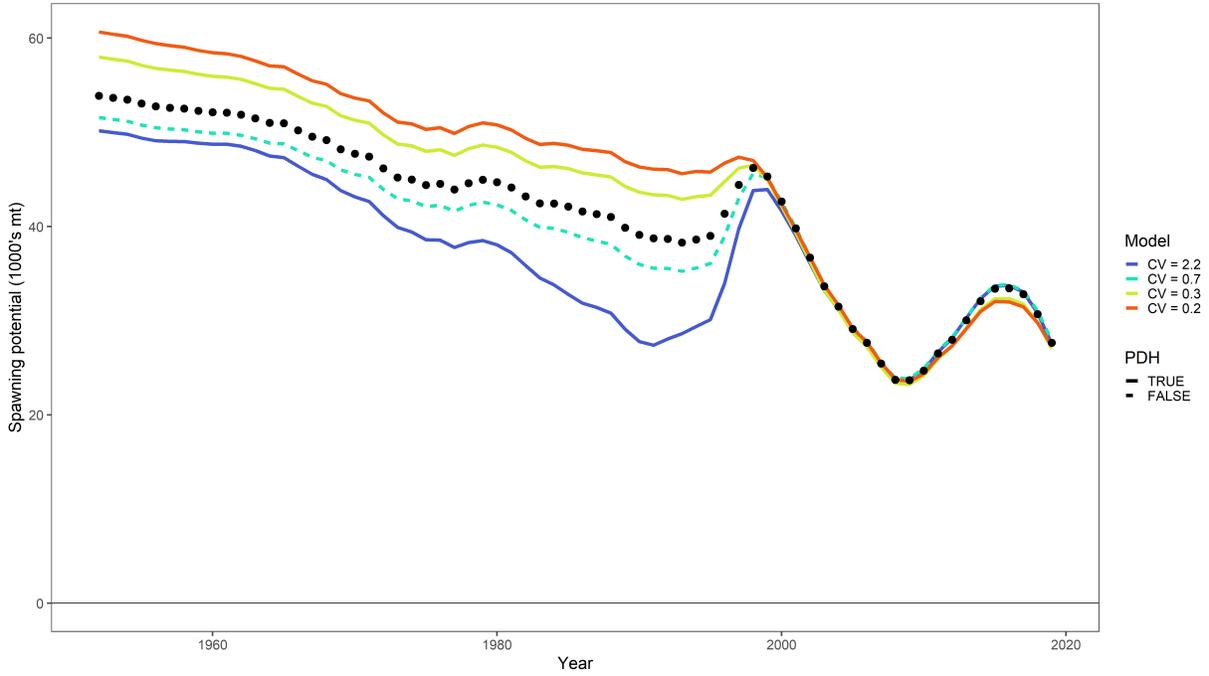


(a) Spawning potential.

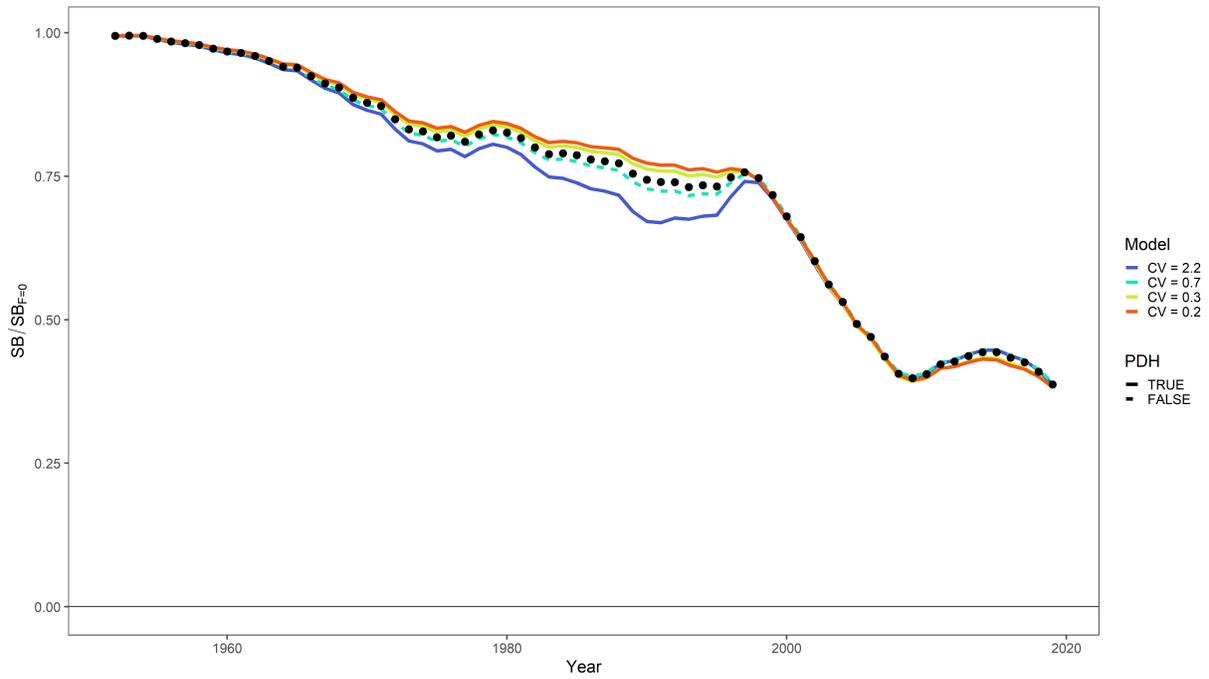


(b) Depletion.

Figure 26: Estimated spawning potential and depletion for the alternative catch & spatial scenarios, where the color indicates the scenario. Line type indicates model convergence (solid), or lack thereof (dashed), according to the presence of a positive definite Hessian solution. The 2021 diagnostic case is given by the black line.



(a) Spawning potential.



(b) Depletion.

Figure 27: Estimated spawning potential and depletion for the sensitivity of the penalty on the stock-recruit relationship. Line type indicates model convergence (solid), or lack thereof (dashed), according to the presence of a positive definite Hessian solution. The 2021 diagnostic case (CV = 0.5) is given by the black points.

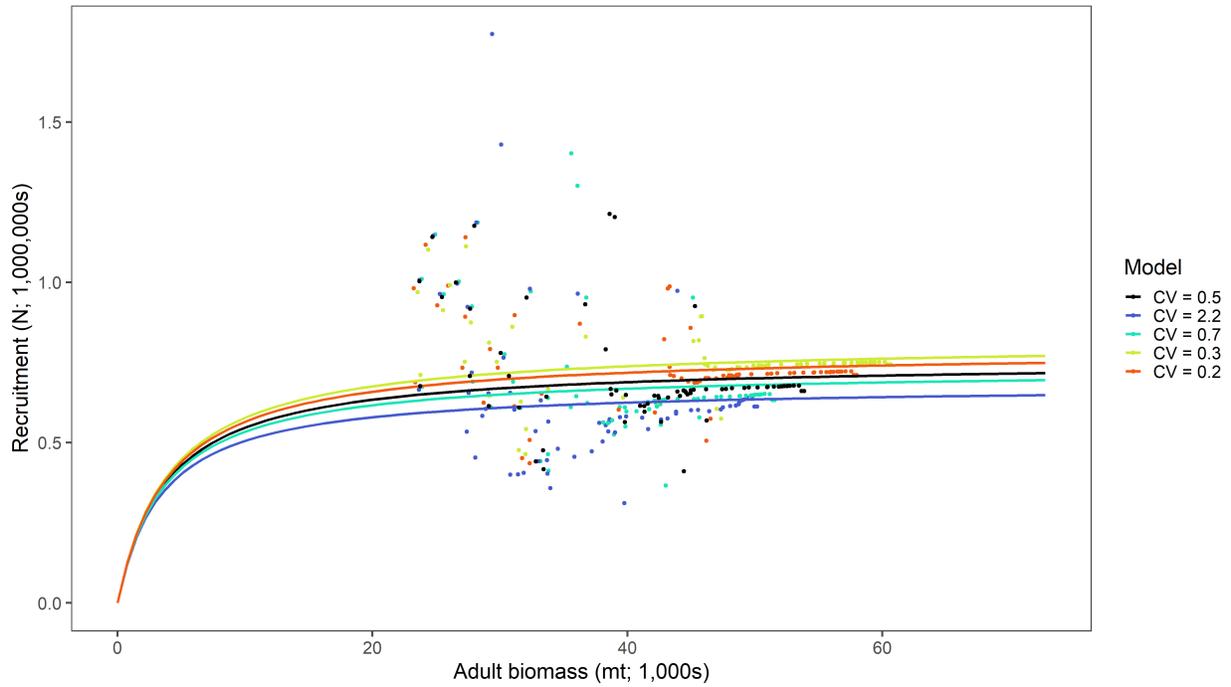
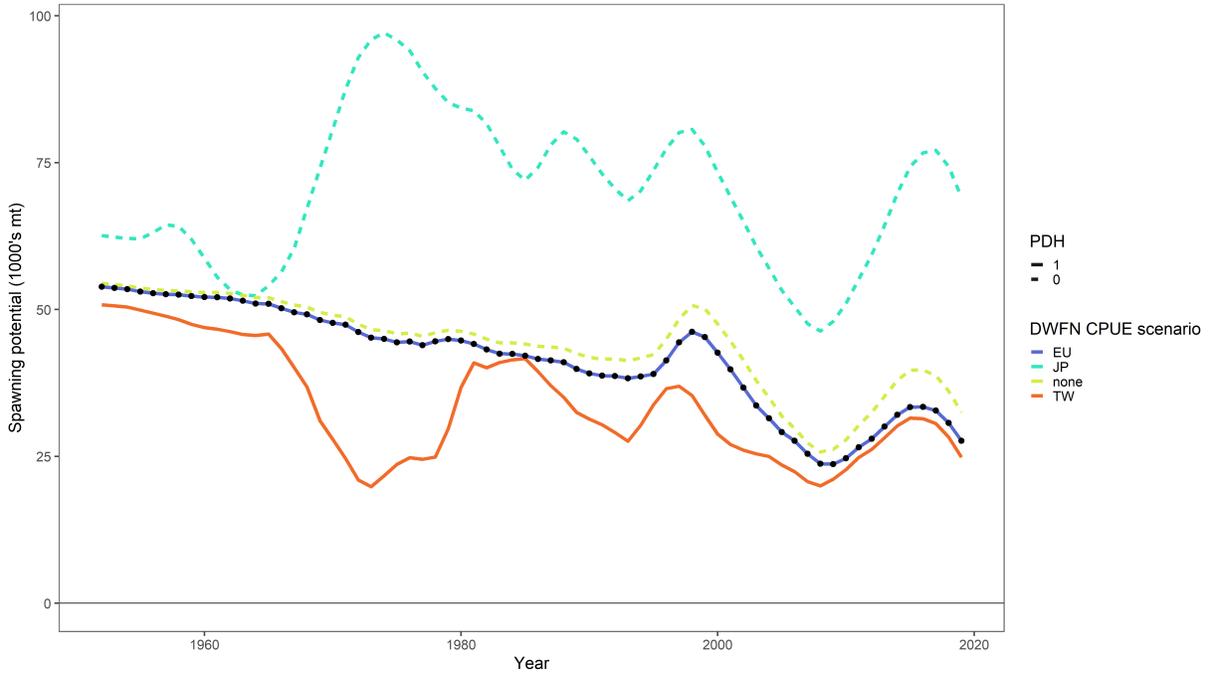
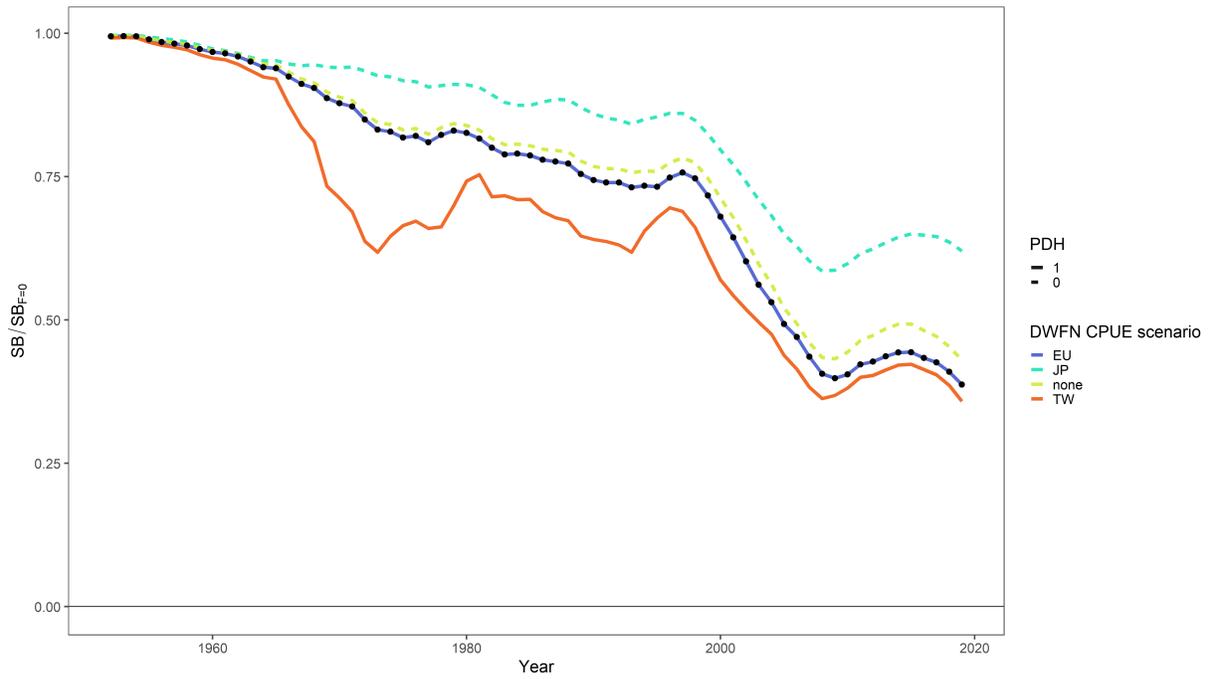


Figure 28: Estimated Beverton-Holt stock recruit relationship for the sensitivity of the penalty on the stock-recruit relationship, where color indicates the magnitude of the penalty. The 2021 diagnostic case ($CV = 0.5$) is given by the black line.



(a) Spawning potential.



(b) Depletion.

Figure 29: Estimated spawning potential and depletion for the considered DWFN index scenarios. Line type indicates model convergence (solid), or lack thereof (dashed), according to the presence of a positive definite Hessian solution. The 2021 diagnostic case is given by the black points.

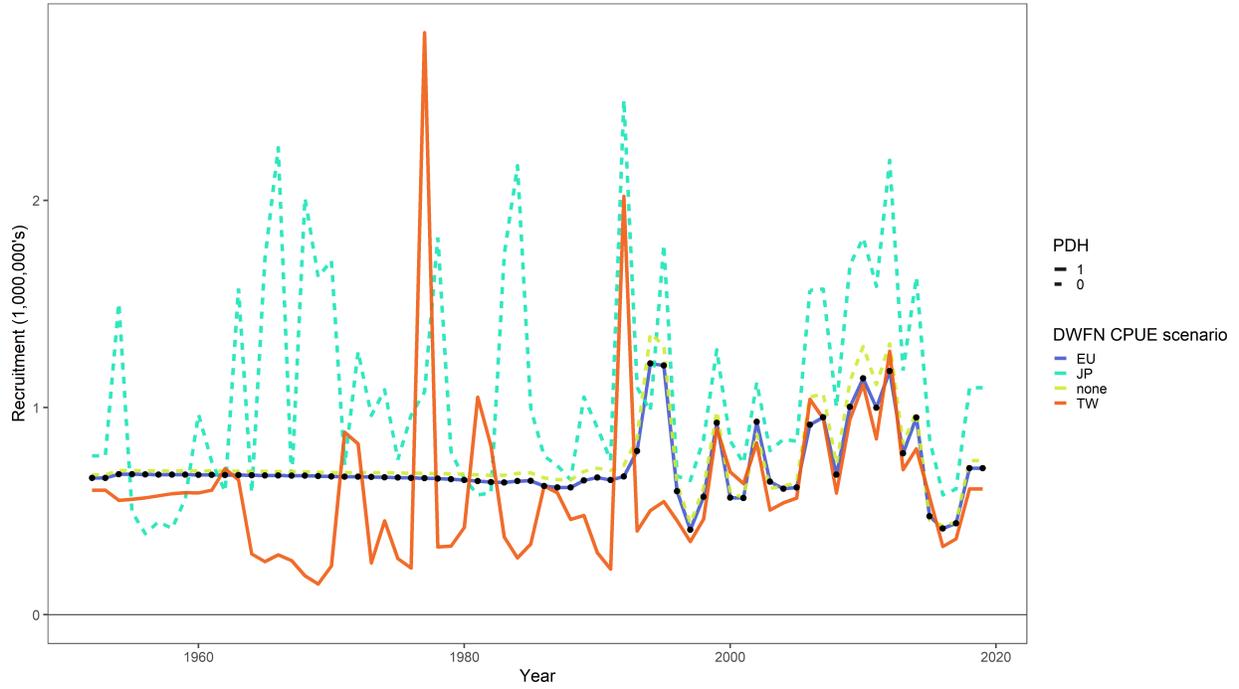
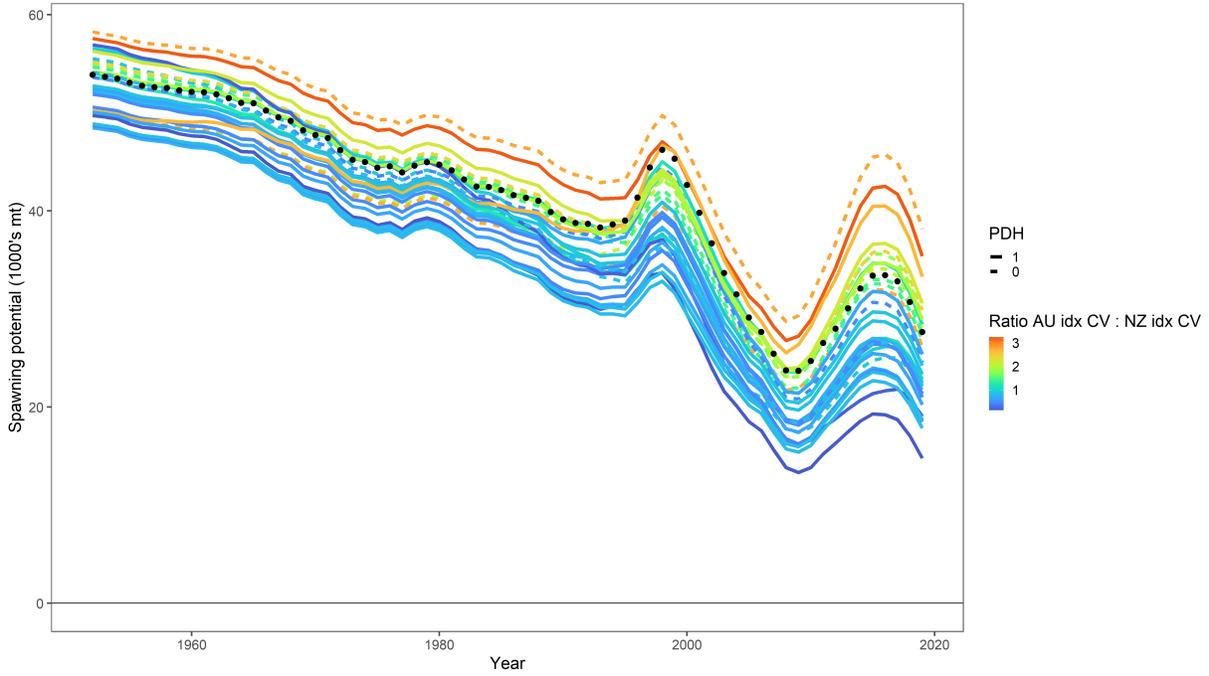
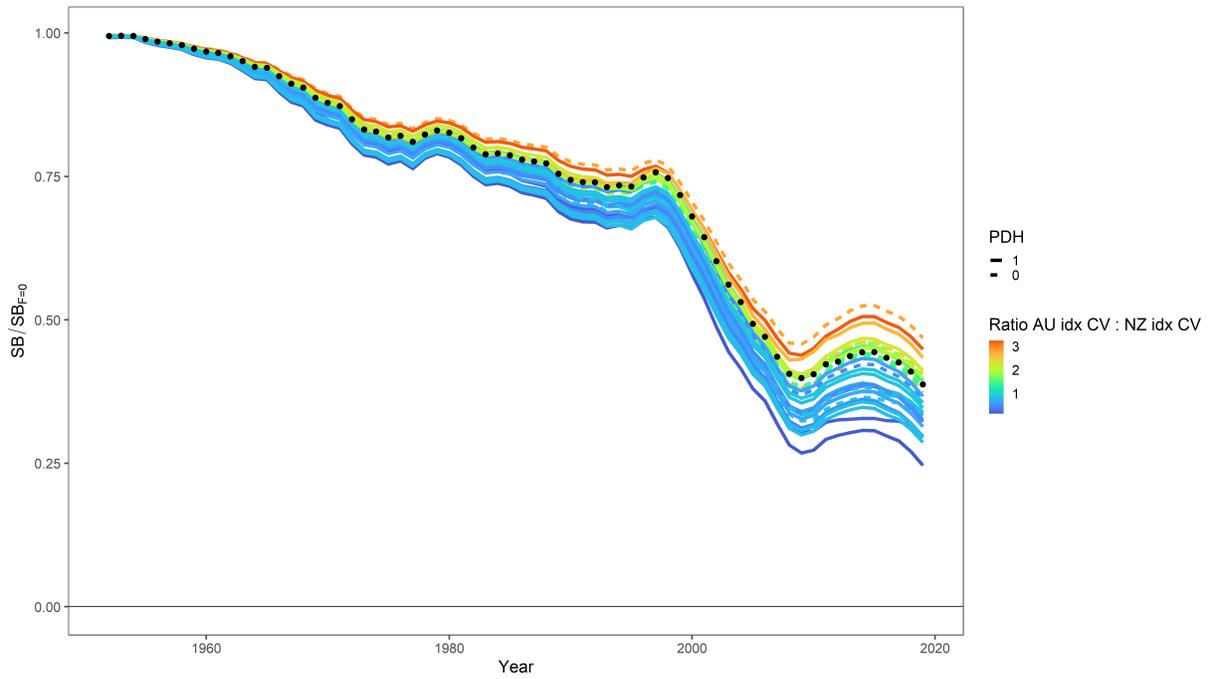


Figure 30: Estimated recruitment for the considered DWFN index scenarios. Line type indicates model convergence (solid), or lack thereof (dashed), according to the presence of a positive definite Hessian solution. The 2021 diagnostic case is given by the black points.

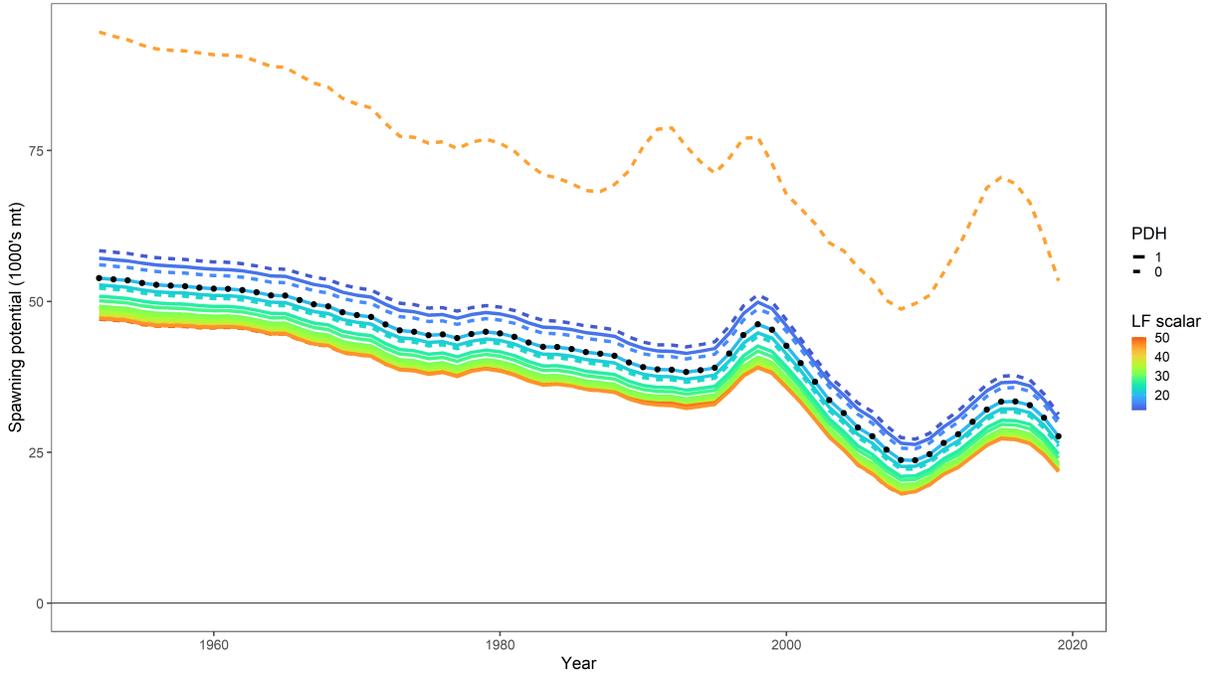


(a) Spawning potential.

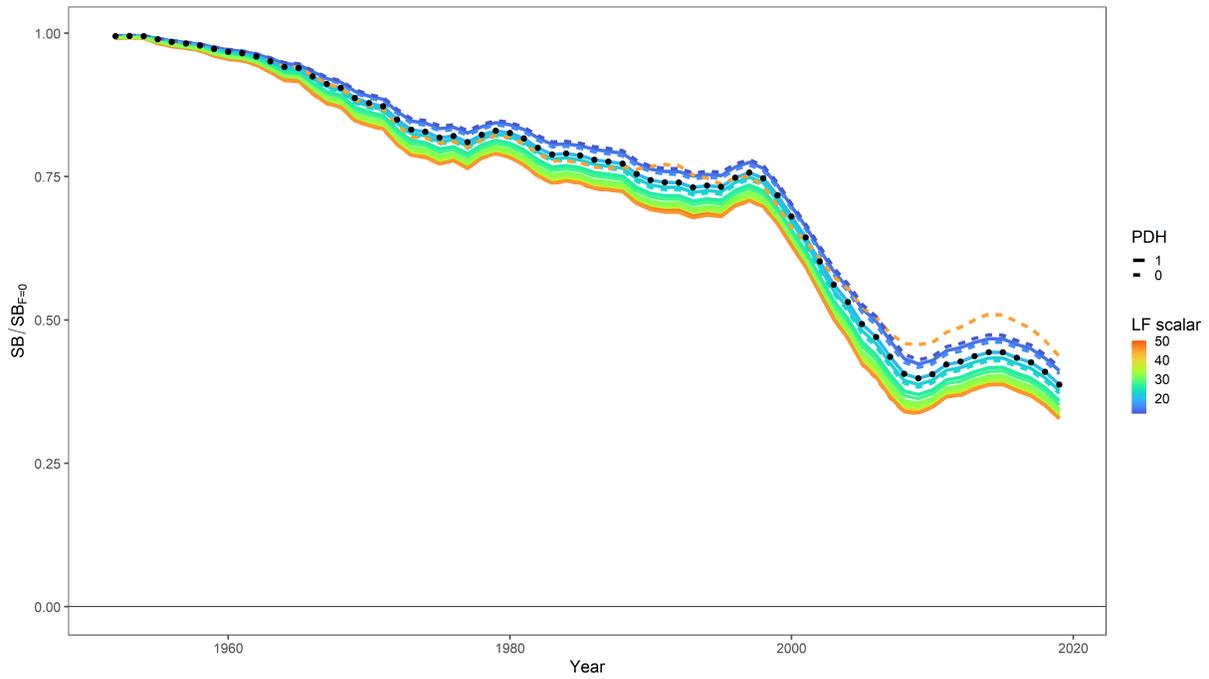


(b) Depletion.

Figure 31: Estimated spawning potential and depletion for the sensitivity to average CV associated with the indices included in the diagnostic case. The color of the line indicates the ratio of the average CV assumed for the Australian index relative to the average CV assumed for the New Zealand index. Line type indicates model convergence (solid), or lack thereof (dashed), according to the presence of a positive definite Hessian solution. The 2021 diagnostic case is given by the black points.

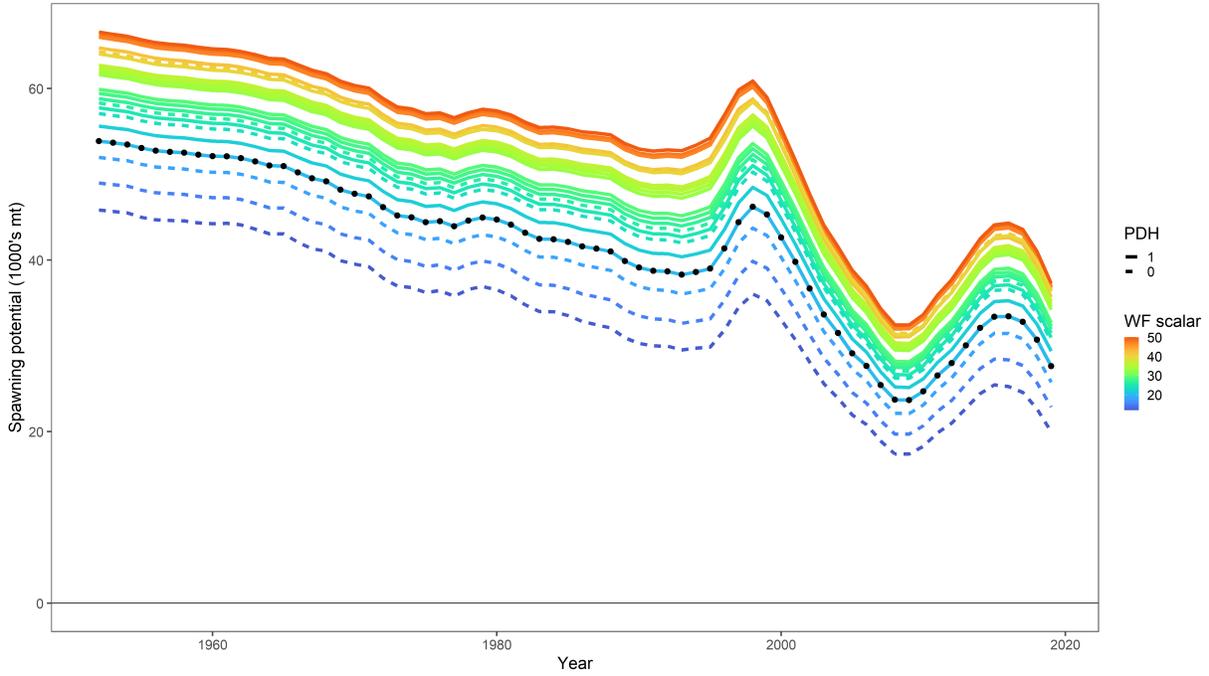


(a) Spawning potential.

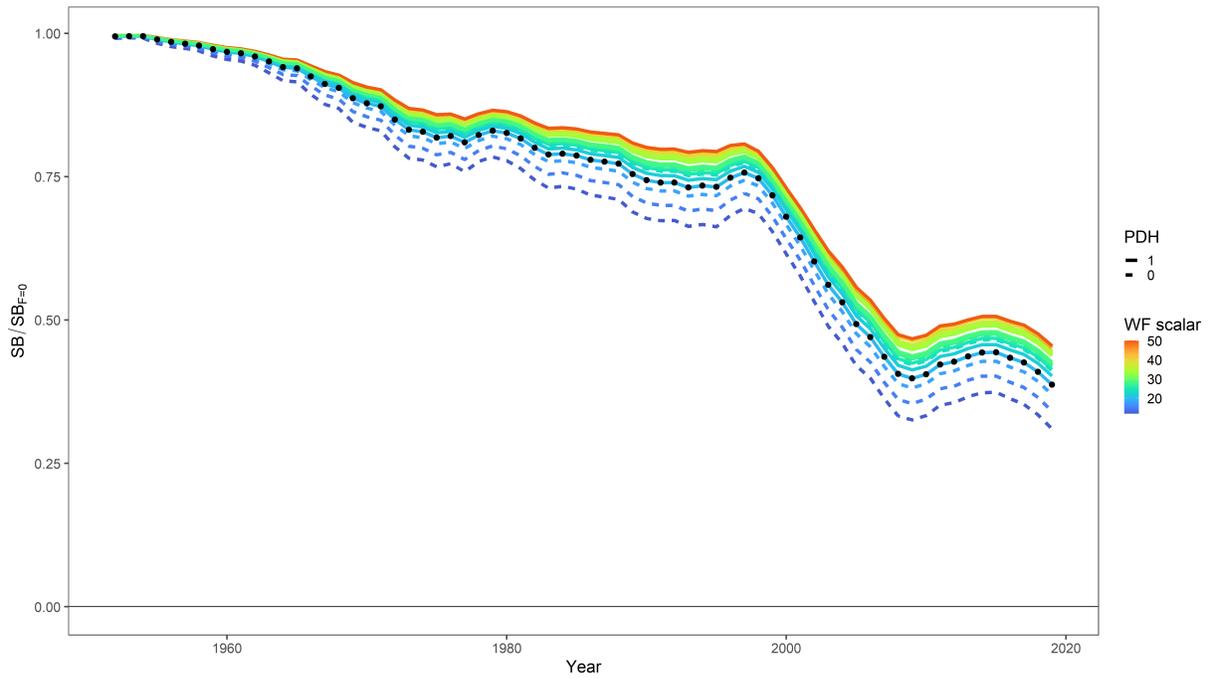


(b) Depletion.

Figure 32: Estimated spawning potential and depletion for the sensitivity to the assumed scalar for the length frequency data. The color of the line indicates the scalar assumed. Line type indicates model convergence (solid), or lack thereof (dashed), according to the presence of a positive definite Hessian solution. The 2021 diagnostic case is given by the black points.

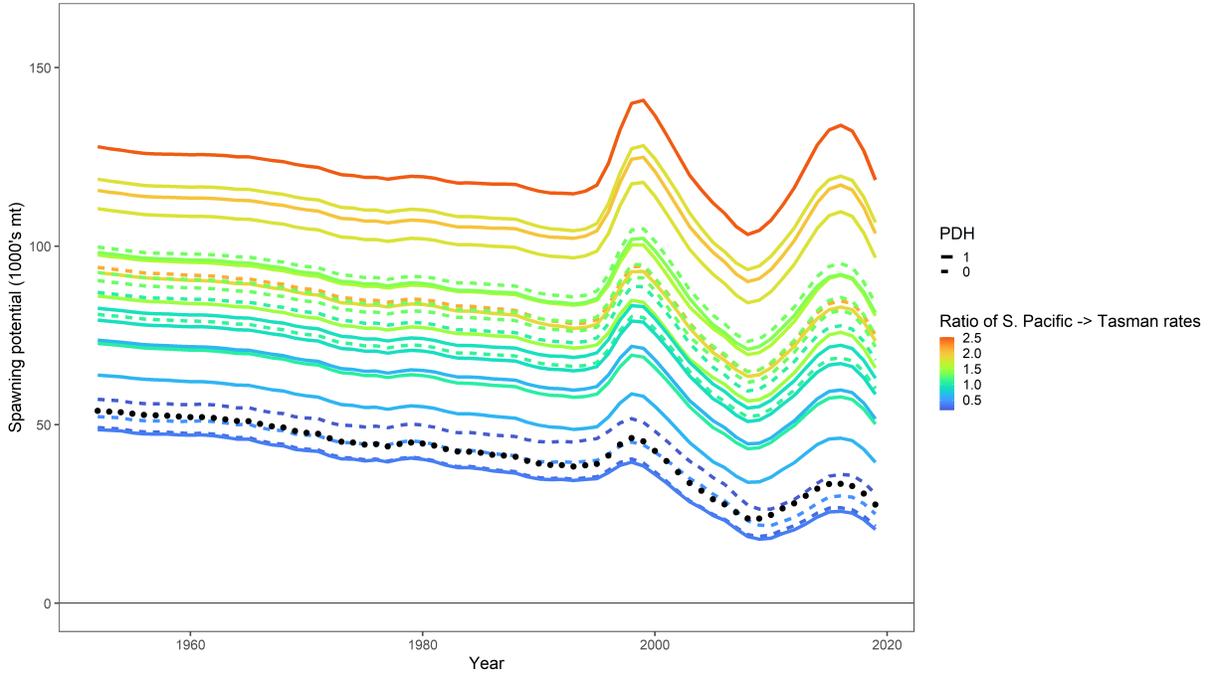


(a) Spawning potential.

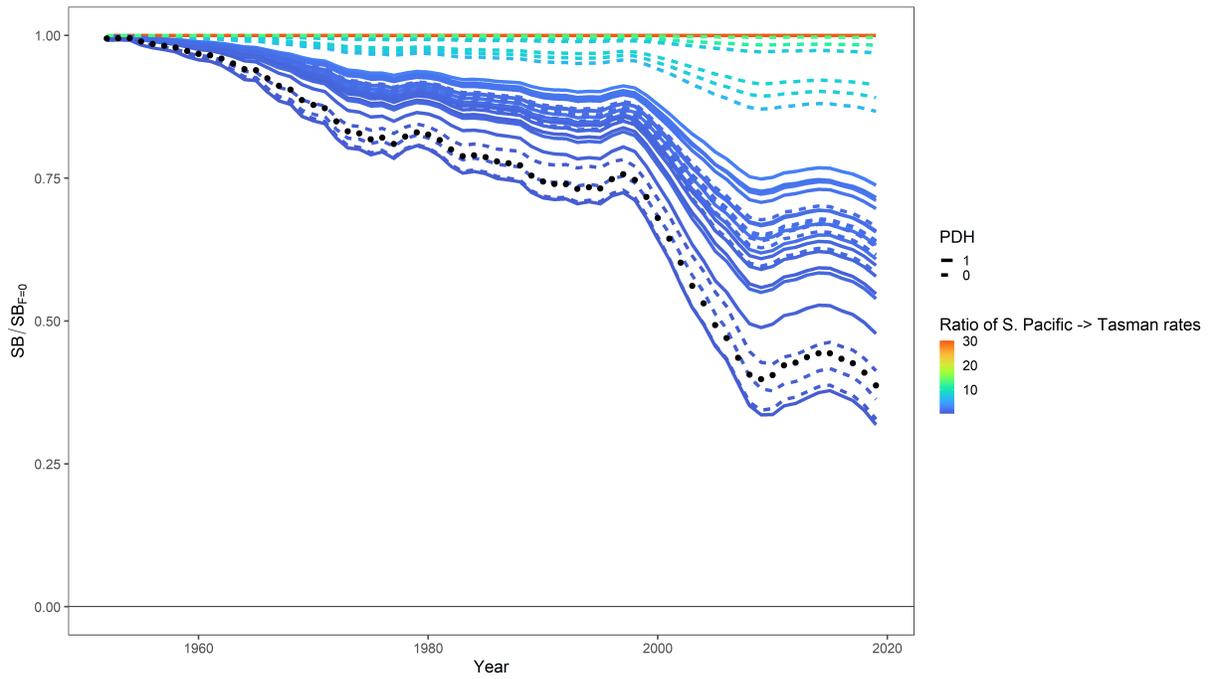


(b) Depletion.

Figure 33: Estimated spawning potential and depletion for the sensitivity to the assumed scalar for the weight frequency data. The color of the line indicates the scalar assumed. Line type indicates model convergence (solid), or lack thereof (dashed), according to the presence of a positive definite Hessian solution. The 2021 diagnostic case is given by the black points.



(a) Spawning potential.



(b) Depletion.

Figure 34: Estimated spawning potential and depletion for the sensitivity to the assumed regional movement. The color of the line indicates the ratio of movement from the South Pacific region into the Tasman region. Line type indicates model convergence (solid), or lack thereof (dashed), according to the presence of a positive definite Hessian solution. The 2021 diagnostic case is given by the black points.

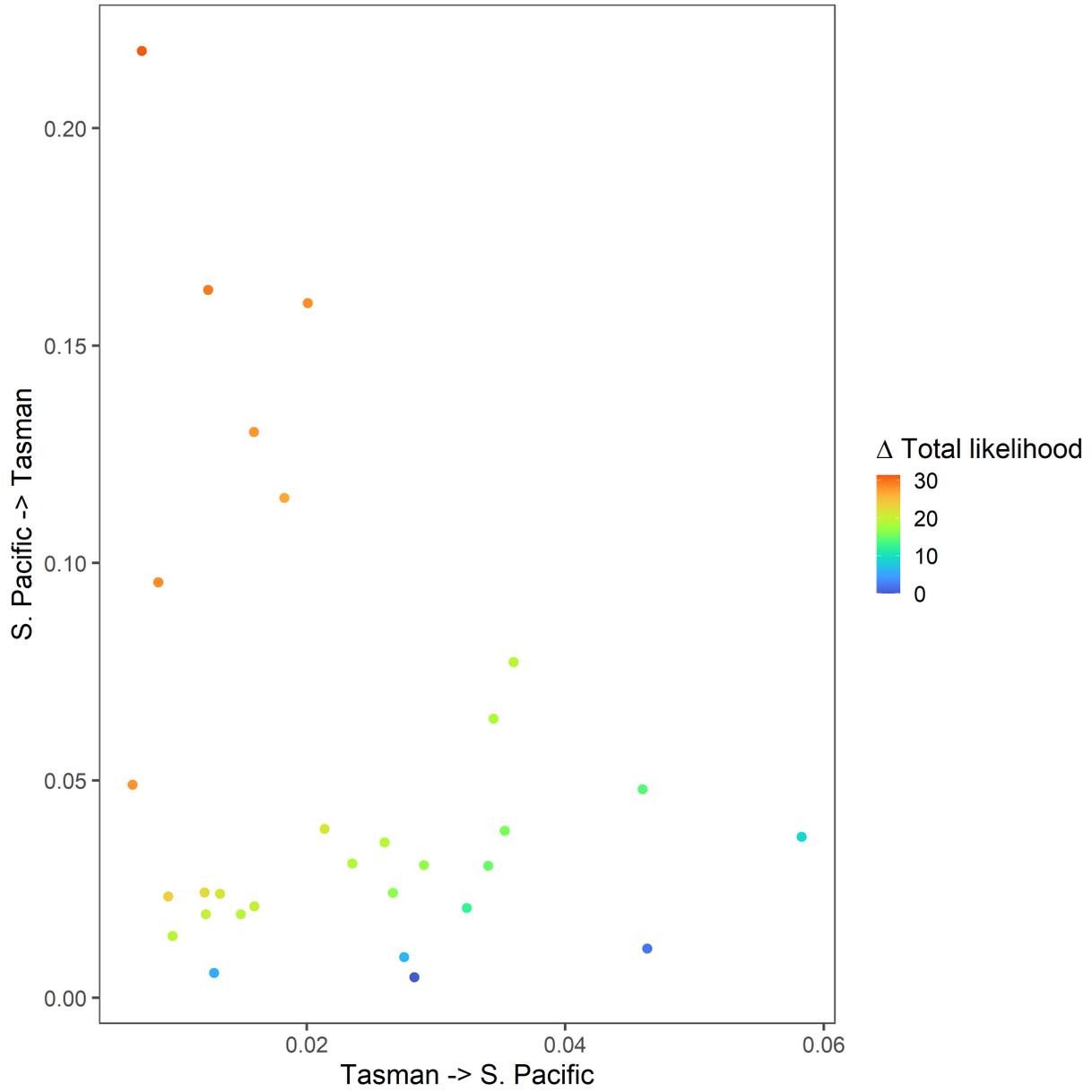
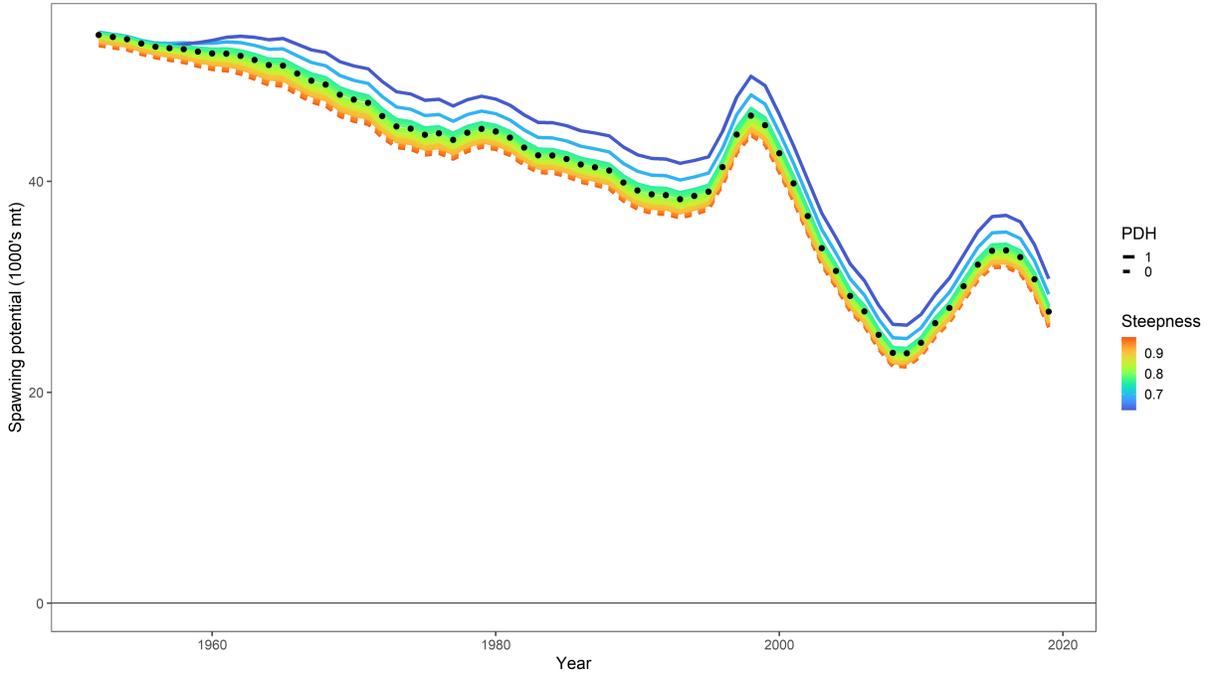
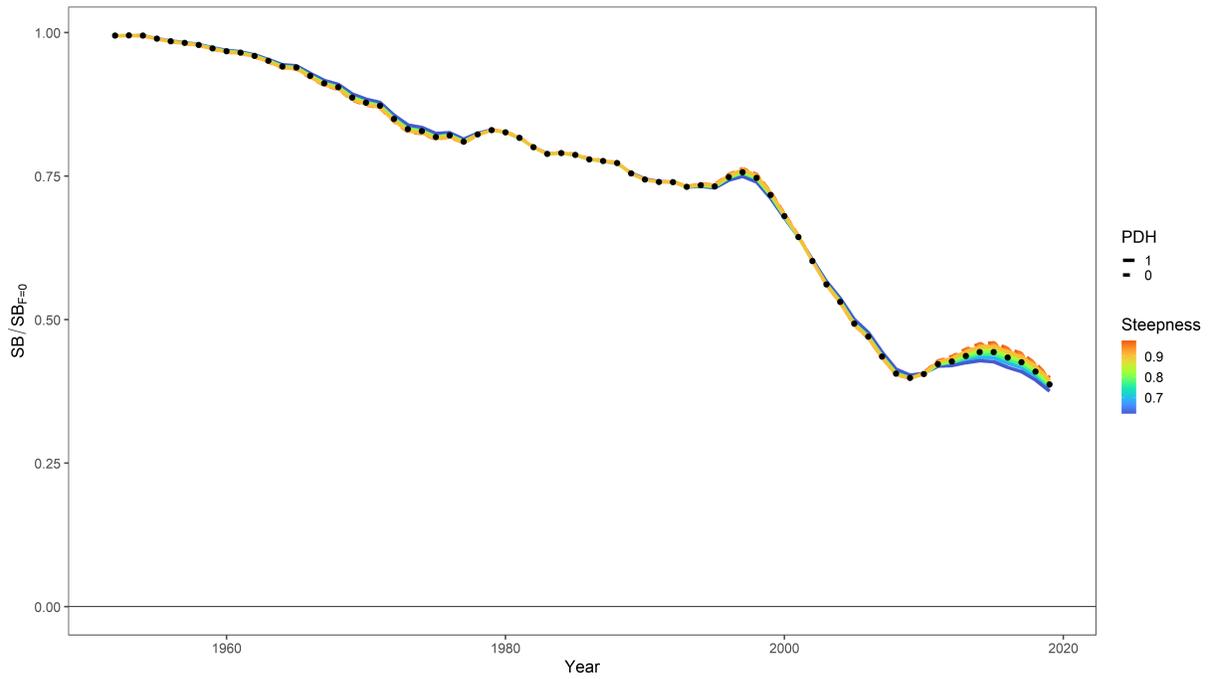


Figure 35: Bivariate plot showing the change (Δ) in Total likelihood given the assumed bi-directional movement rates. Better fits to the data are indicated by smaller Δ Total likelihood values (cooler) colors.

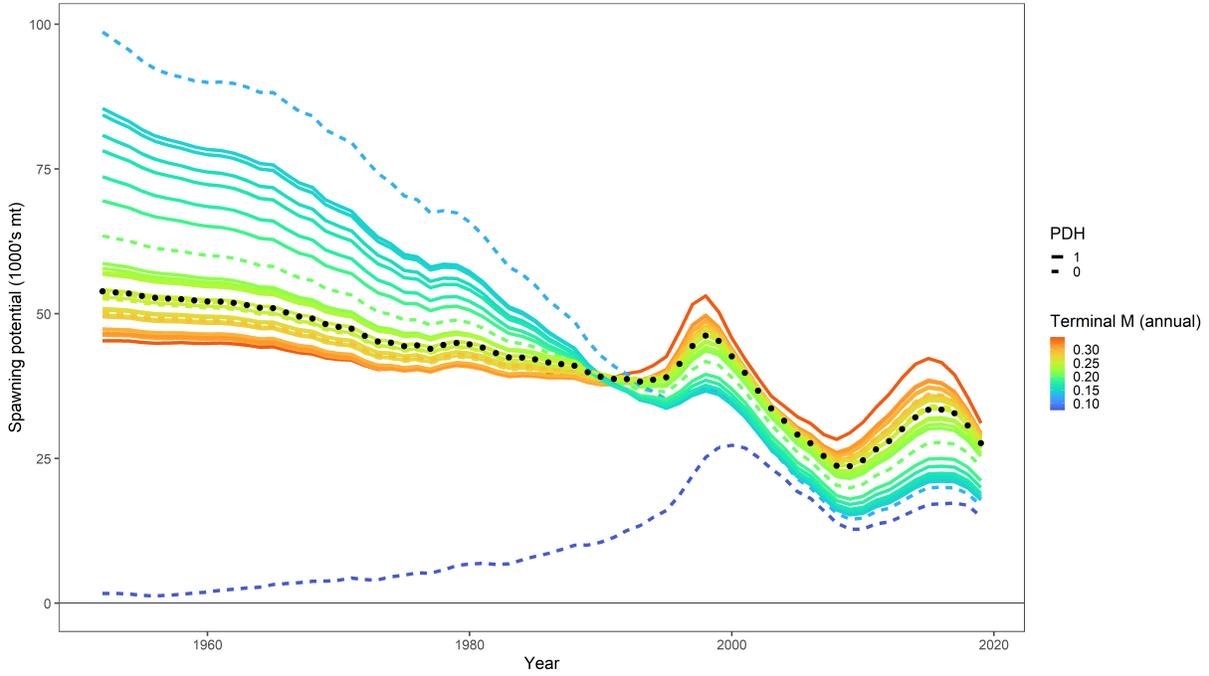


(a) Spawning potential.

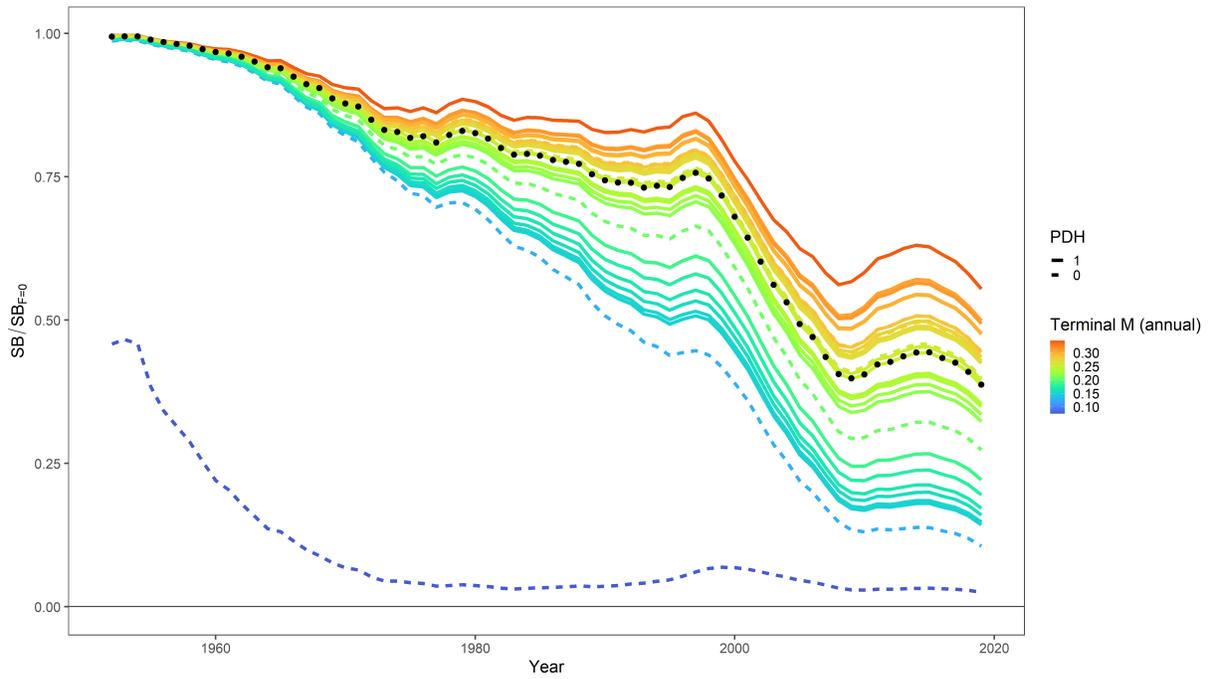


(b) Depletion.

Figure 36: Estimated spawning potential and depletion for the sensitivity to the assumed steepness. The color of the line indicates the assumed steepness value. Line type indicates model convergence (solid), or lack thereof (dashed), according to the presence of a positive definite Hessian solution. The 2021 diagnostic case is given by the black points.



(a) Spawning potential.



(b) Depletion.

Figure 37: Estimated spawning potential and depletion for the sensitivity to the assumed M_{ref} value. The color of the line indicates the assumed M value for the oldest age. Line type indicates model convergence (solid), or lack thereof (dashed), according to the presence of a positive definite Hessian solution. The 2021 diagnostic case is given by the black points.

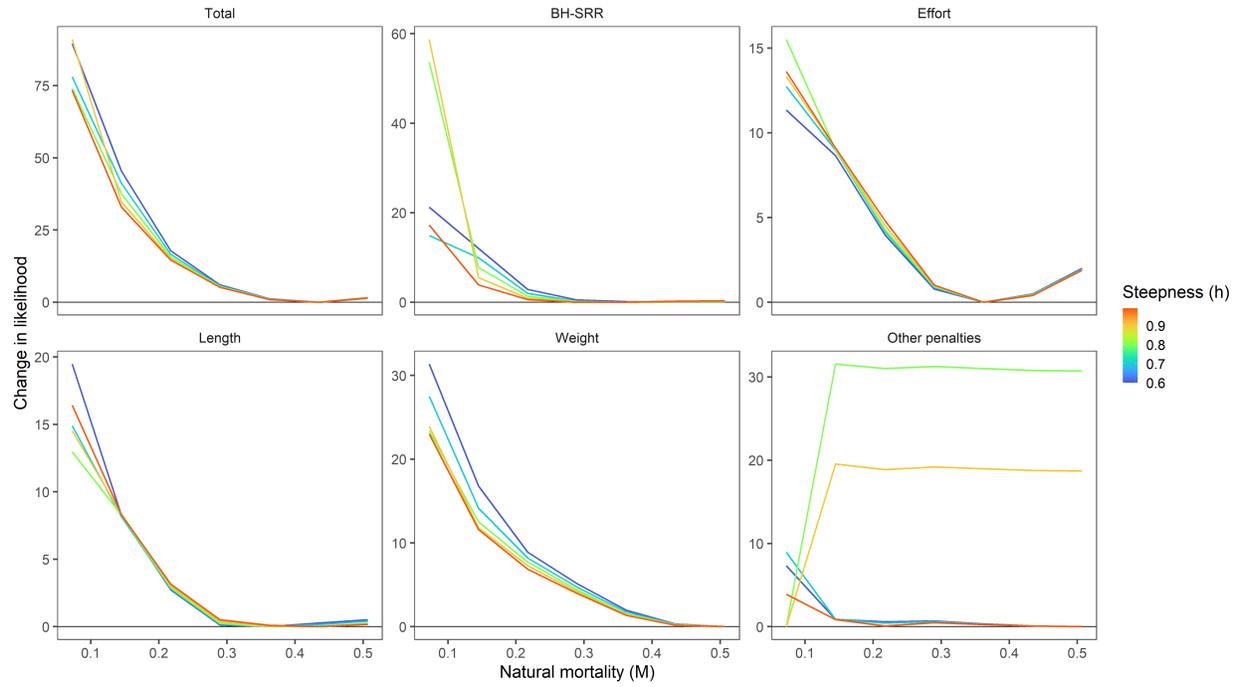
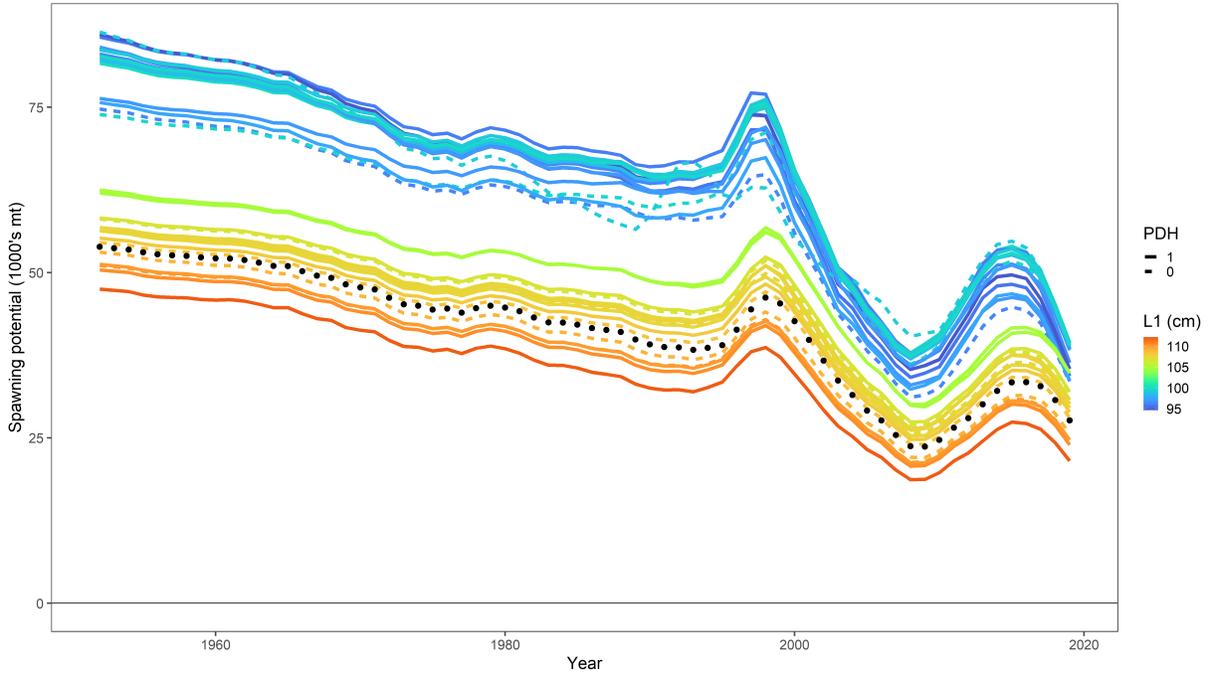
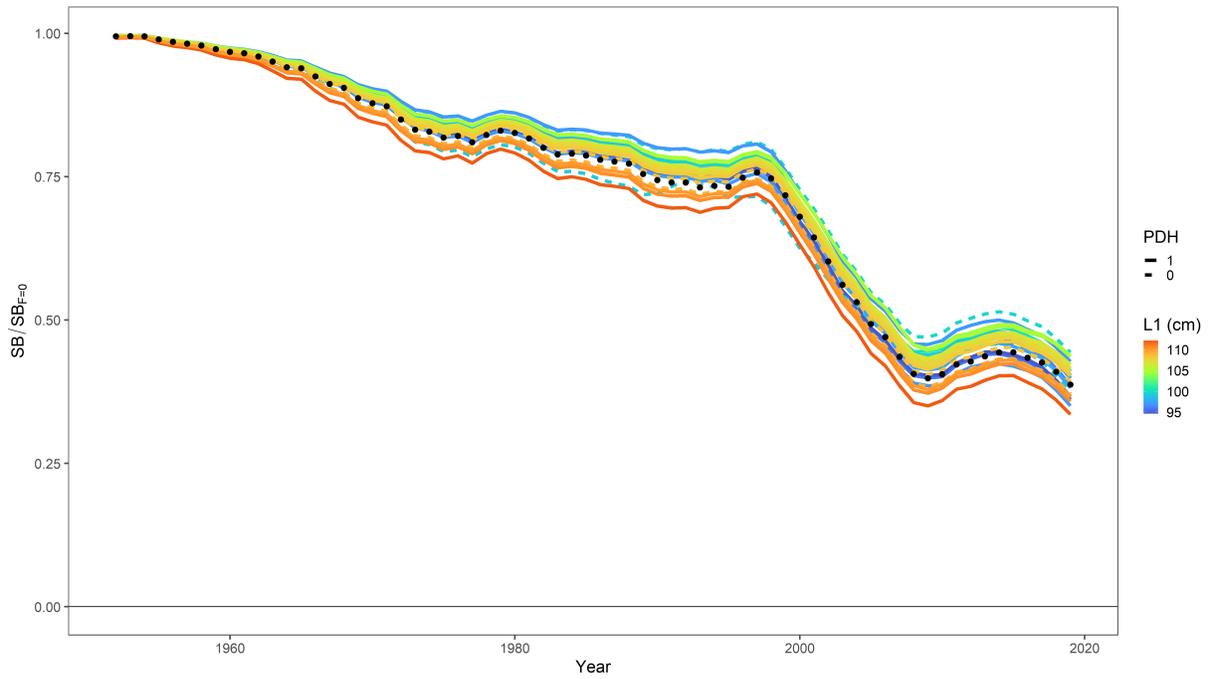


Figure 38: Likelihood profile on natural mortality (M) by likelihood component (panel) and assumed steepness value (line color).



(a) Spawning potential.



(b) Depletion.

Figure 39: Estimated spawning potential and depletion for the sensitivity to the assumed growth parameters: L_1 , L_2 , and k . The color of the line indicates the assumed L_1 value. Line type indicates model convergence (solid), or lack thereof (dashed), according to the presence of a positive definite Hessian solution. The 2021 diagnostic case is given by the black points.

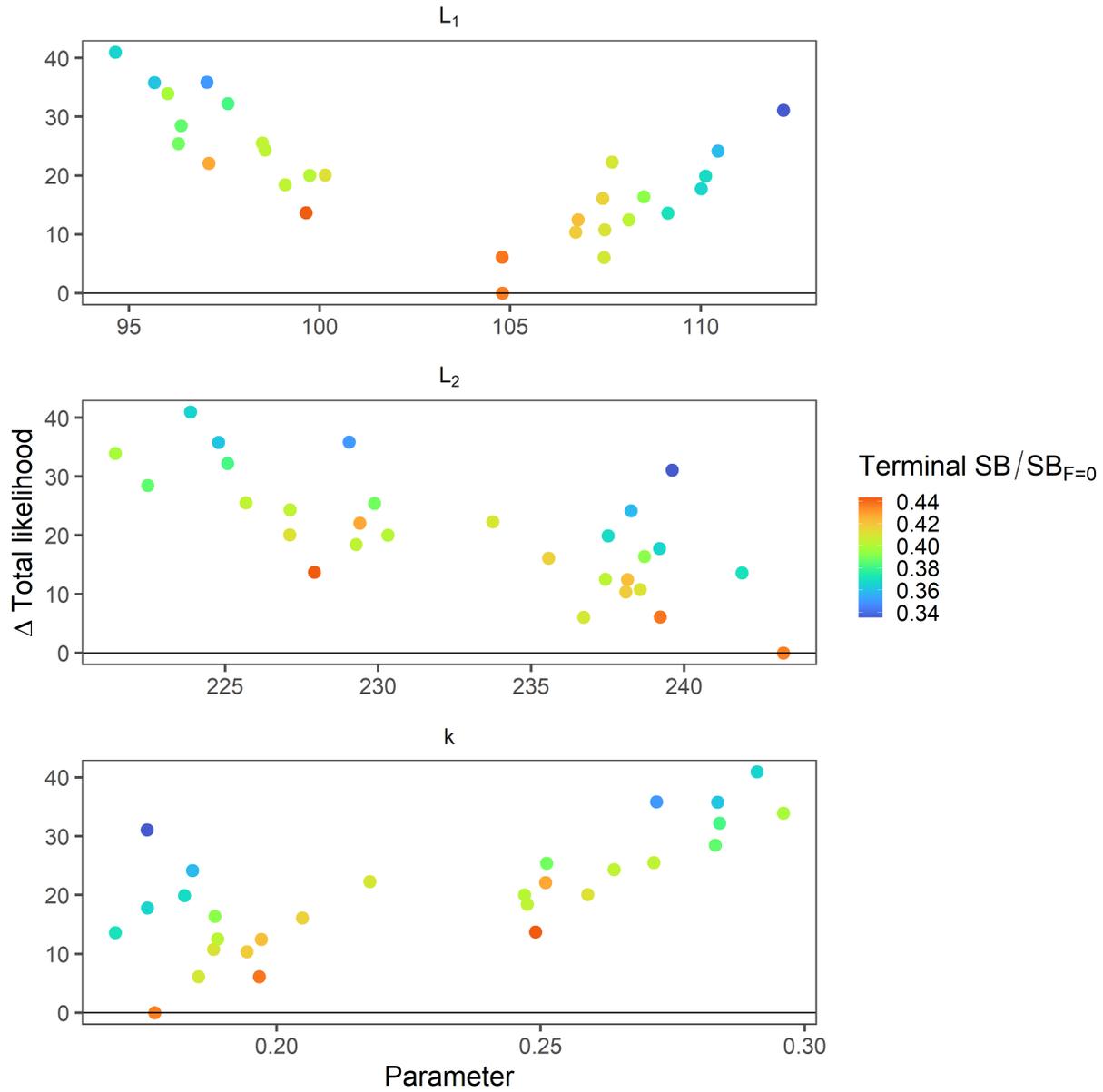
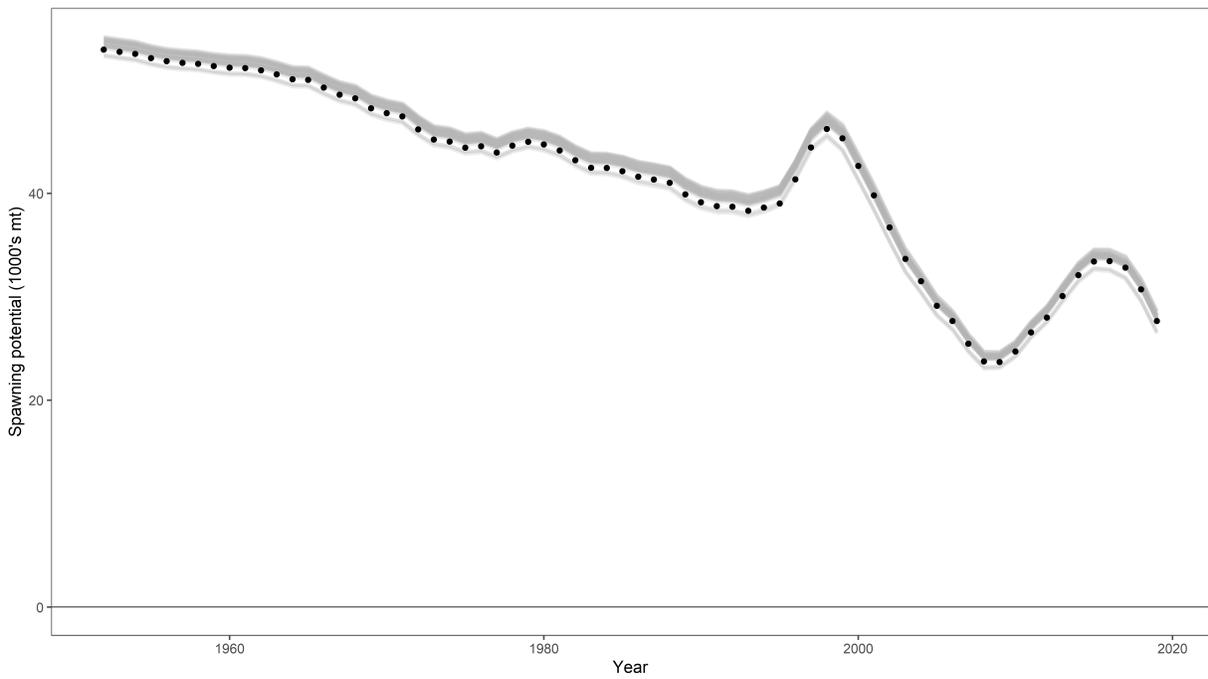
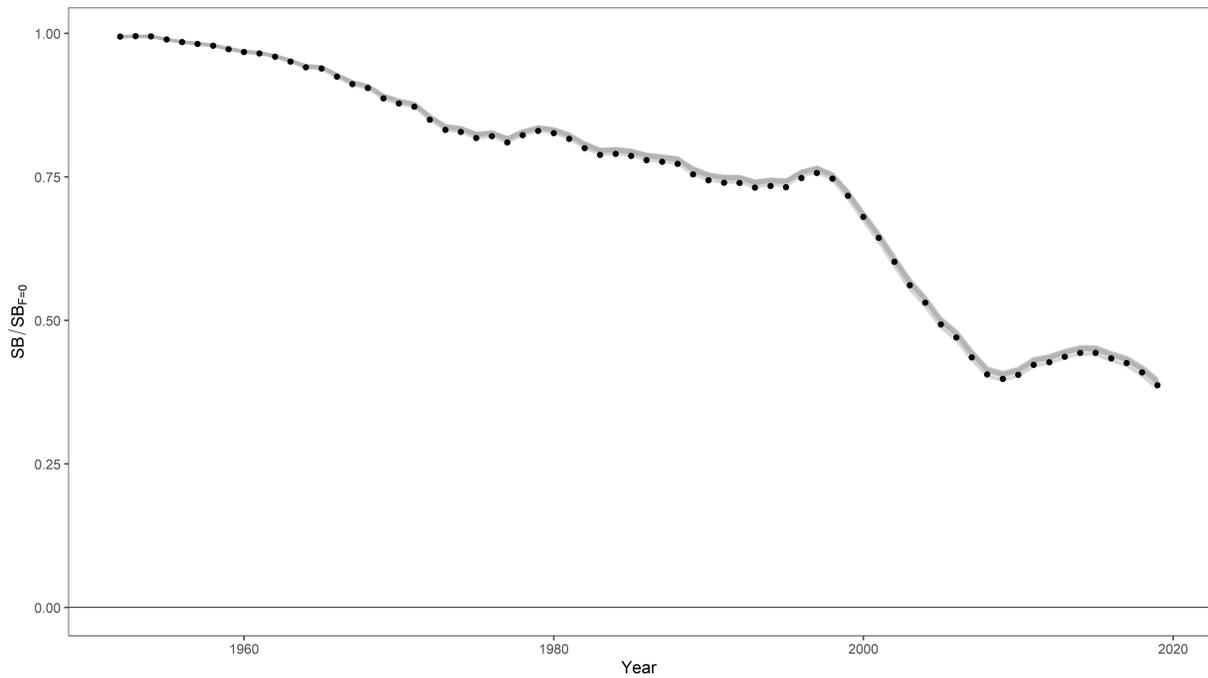


Figure 40: The change (Δ) in Total likelihood given the assumed growth parameters: L_1 , L_2 , and k . Better fits to the data are indicated by smaller Δ Total likelihood values. The color of the point indicates the terminal estimated $SB/SB_{F=0}$ for each model.

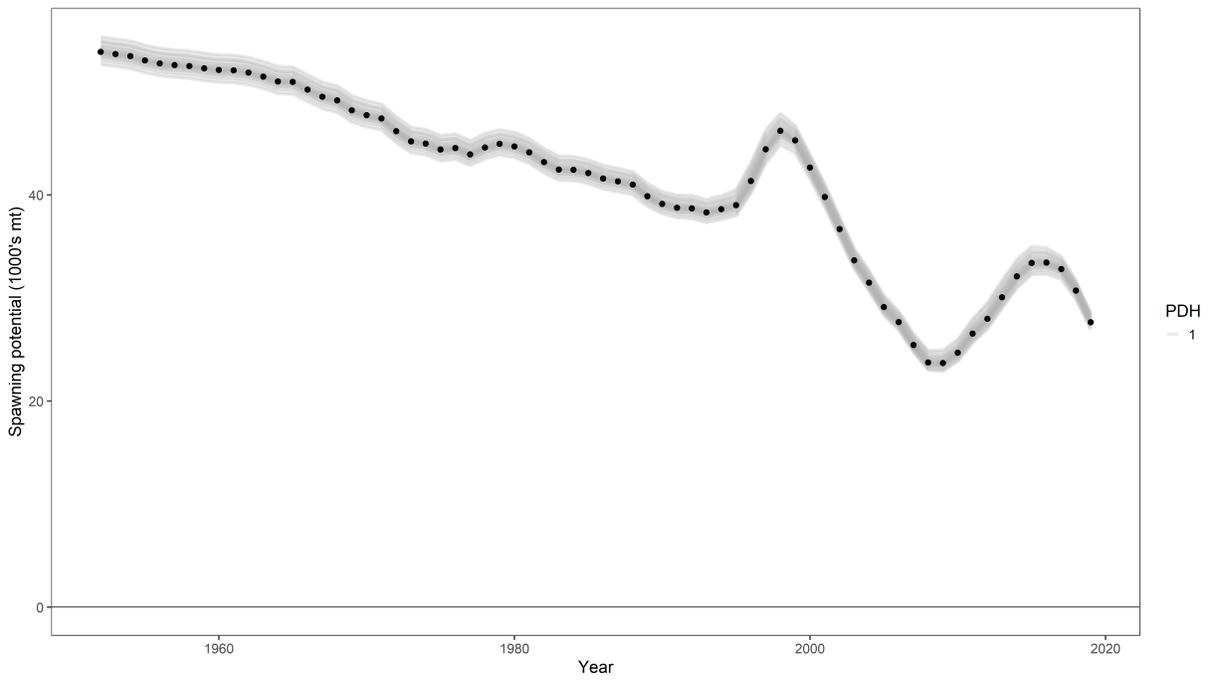


(a) Spawning potential.

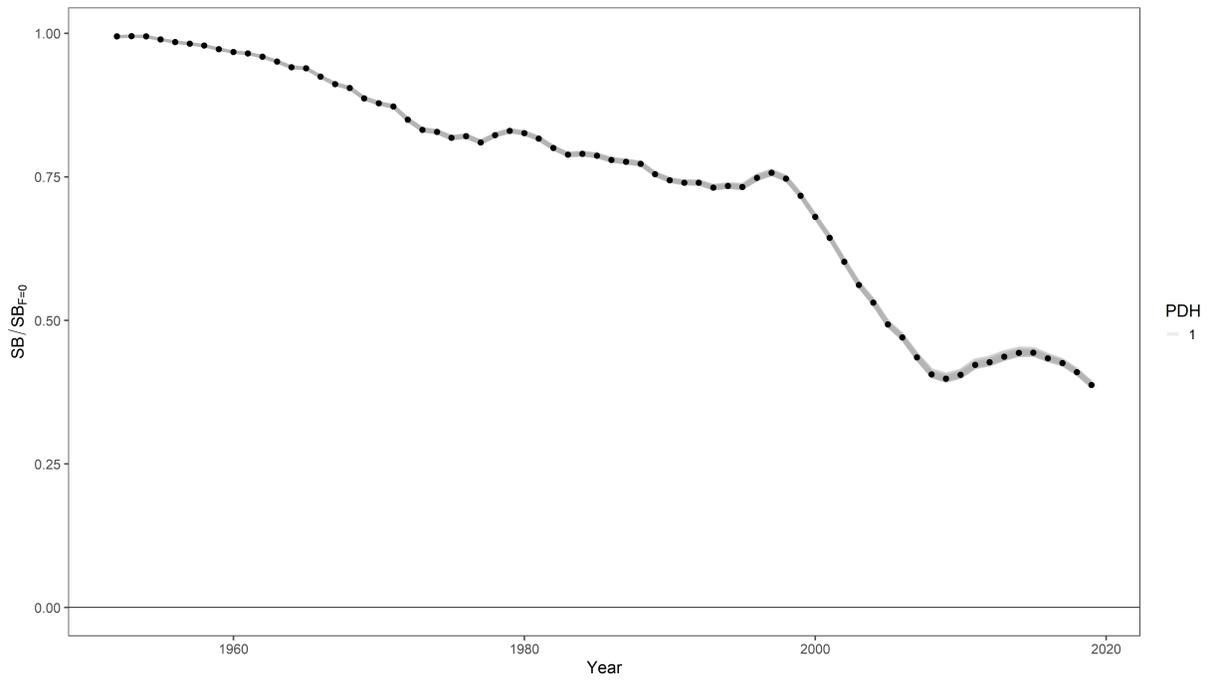


(b) Depletion.

Figure 41: Estimated spawning potential and depletion for the sensitivity to the assumed length-weight relationship parameters: α and β . The 2021 diagnostic case is given by the black points.



(a) Spawning potential.



(b) Depletion.

Figure 42: Estimated spawning potential and depletion for the sensitivity to different assumptions for the reproductive potential ogive. The 2021 diagnostic case is given by the black points.

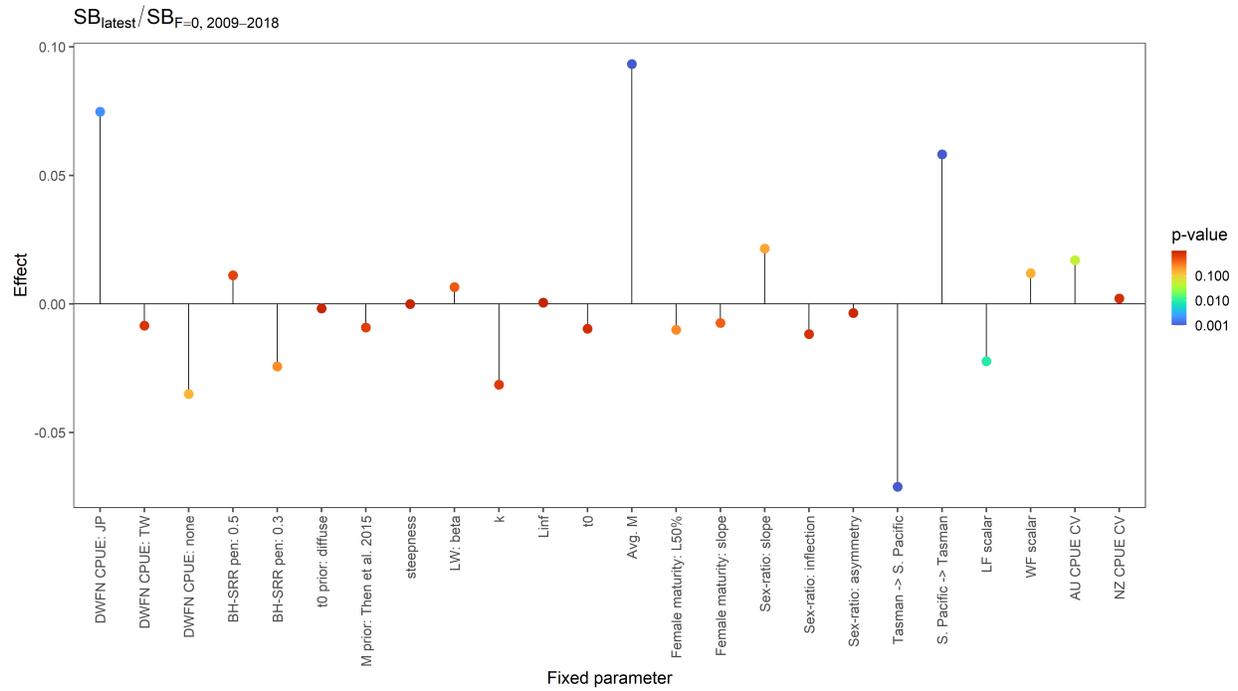


Figure 43: Estimated covariates from a generalized linear model of biological parameters against $SB_{latest}/SB_{F=0}$ where the color of the point is the p-value of the effect and cooler colors indicate greater significance.

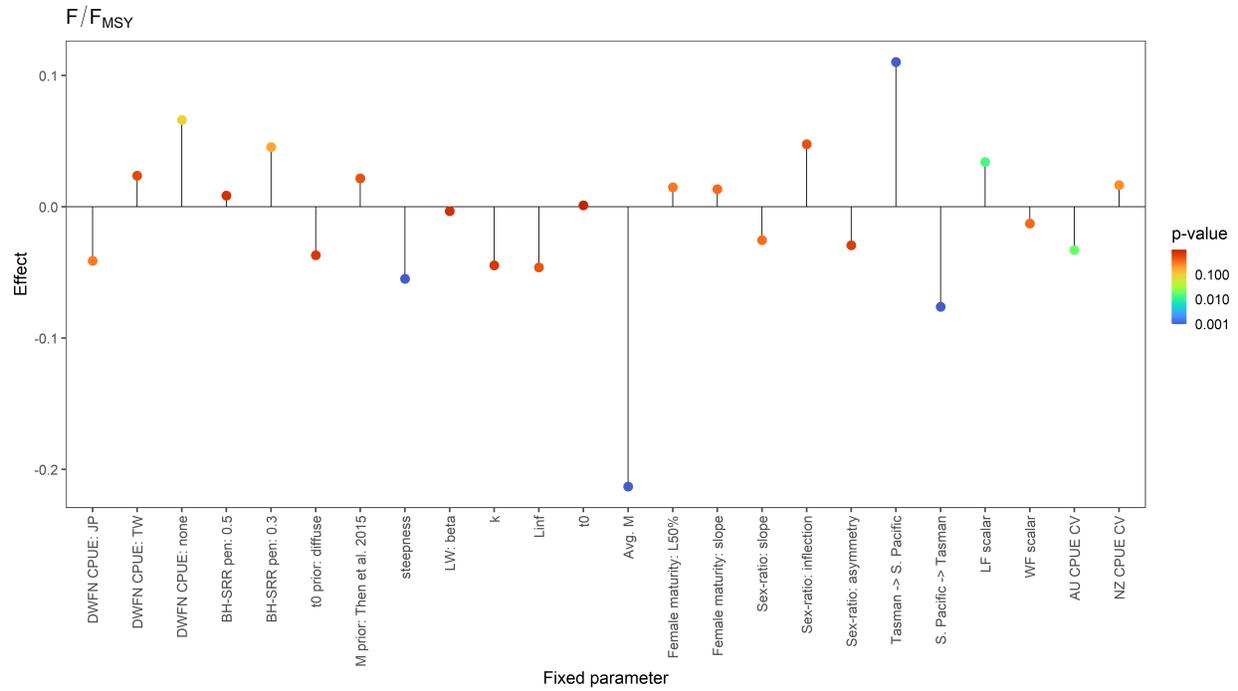


Figure 44: Estimated covariates from a generalized linear model of biological parameters against F_{recent}/F_{MSY} where the color of the point is the p-value of the effect and cooler colors indicate greater significance.

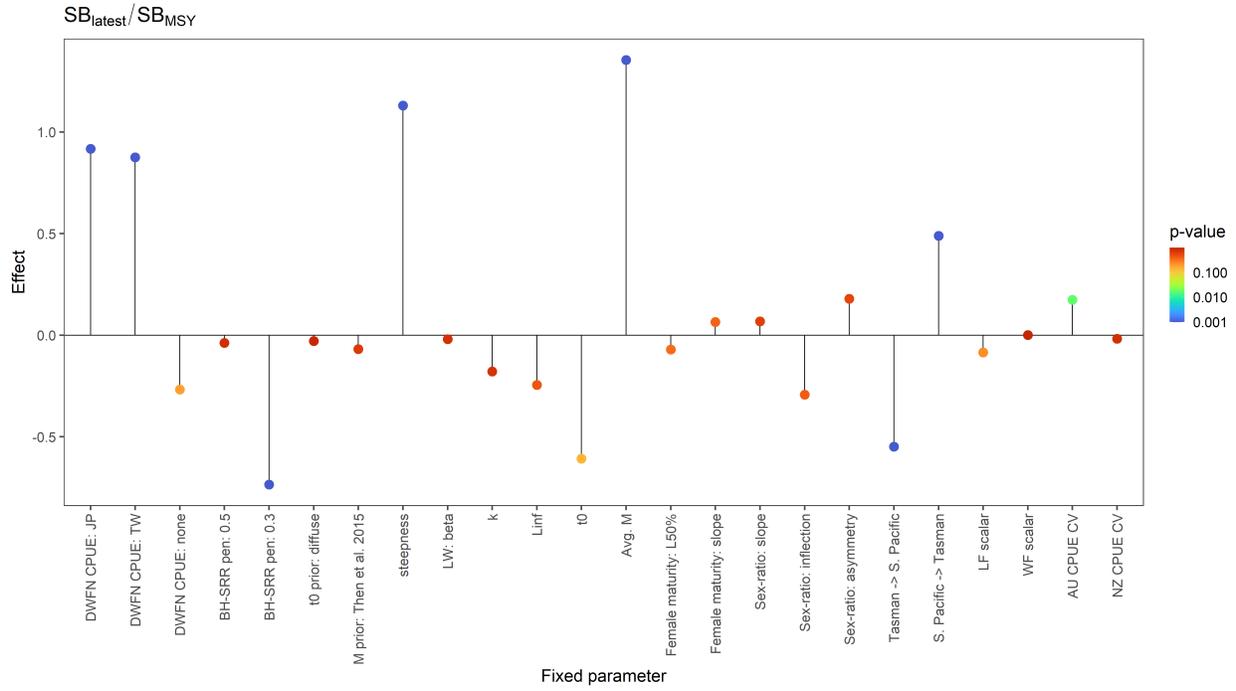


Figure 45: Estimated covariates from a generalized linear model of biological parameters against SB_{latest}/SB_{MSY} where the color of the point is the p-value of the effect and cooler colors indicate greater significance.

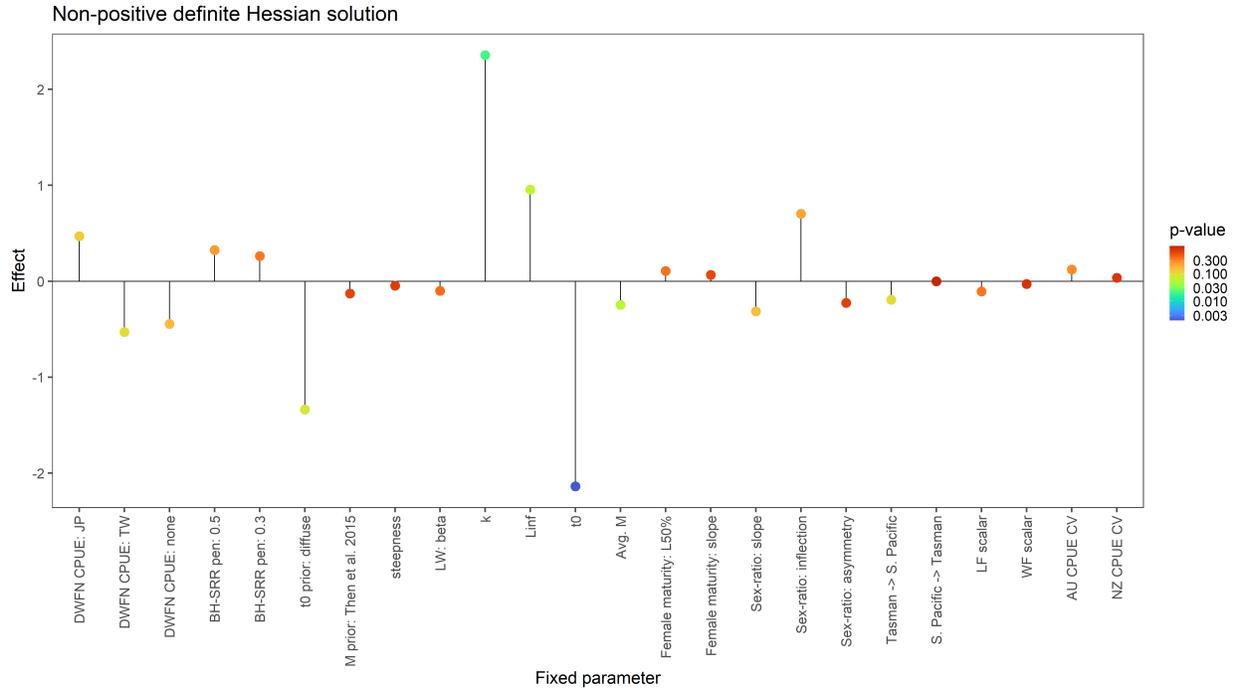


Figure 46: Estimated covariates from a binomial generalized linear model of biological parameters against a non-positive definite Hessian solution where the color of the point is the p-value of the effect and cooler colors indicate greater significance.

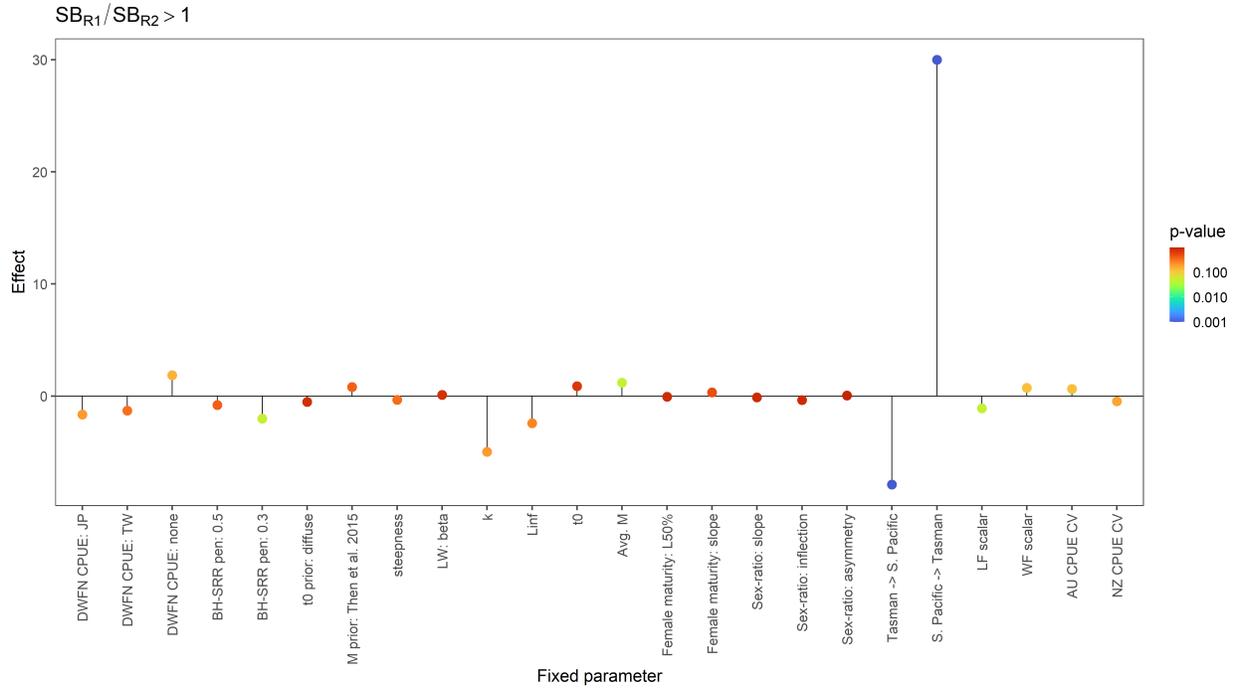


Figure 47: Estimated covariates from a binomial generalized linear model of biological parameters against a solution where $\overline{SB}_{Tasman} / \overline{SB}_{S.Pacific} > 1$. The color of the point is the p-value of the effect and cooler colors indicate greater significance.

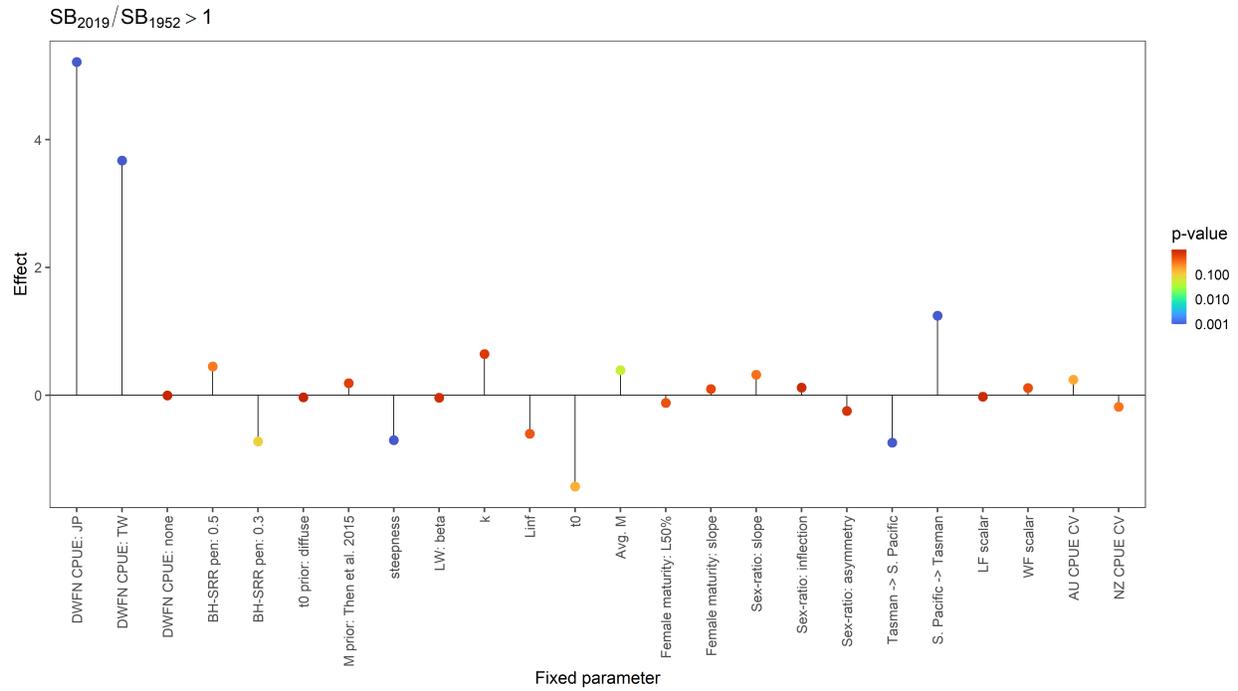


Figure 48: Estimated covariates from a binomial generalized linear model of biological parameters against a solution where $SB_{2019}/SB_0 > 1$. The color of the point is the p-value of the effect and cooler colors indicate greater significance.

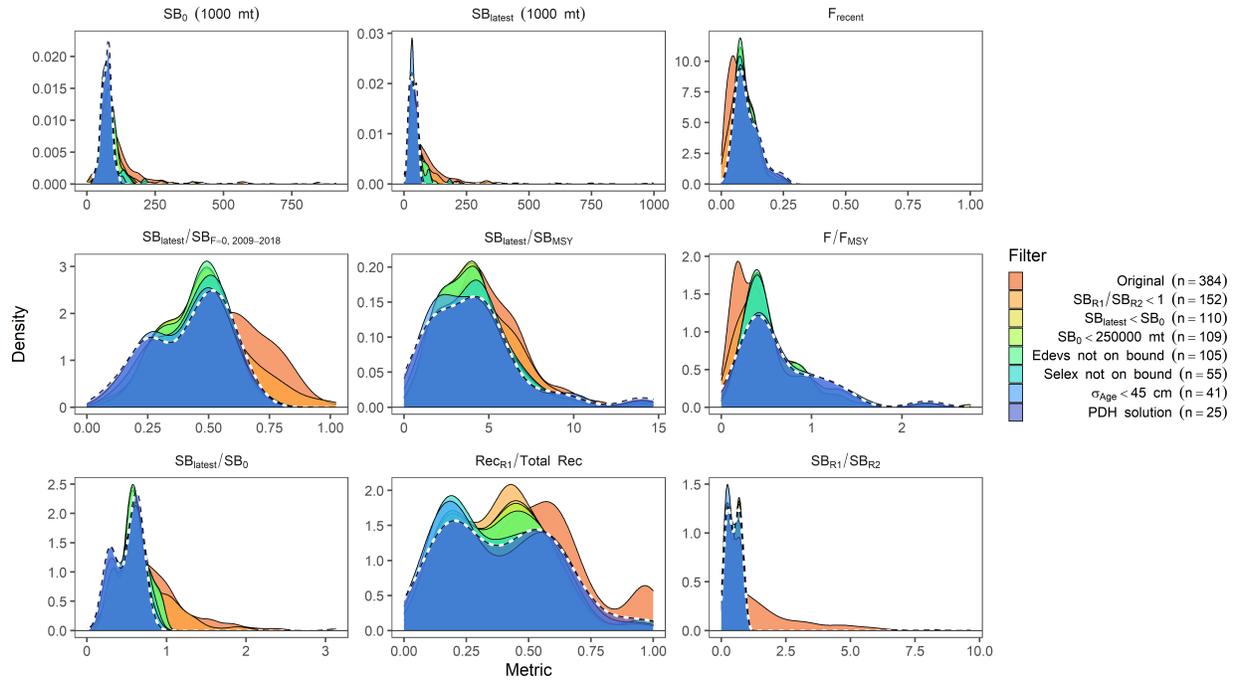


Figure 49: Distributions of key management quantities showing the effects of the post-hoc filtration leading to the final models retained in the ensemble. The final distribution is outlined by the white dotted line.

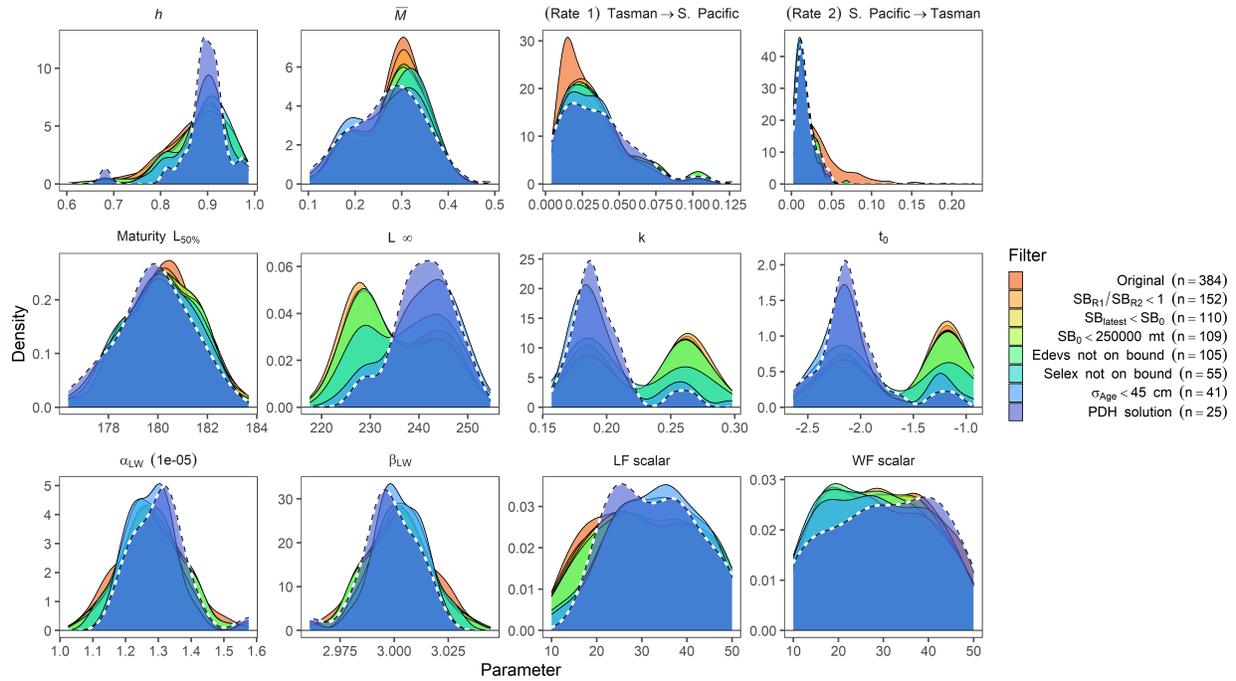


Figure 50: Distributions of key components of the joint prior showing the effects of the post-hoc filtration leading to the final models retained in the ensemble. The final distribution is outlined by the white dotted line.

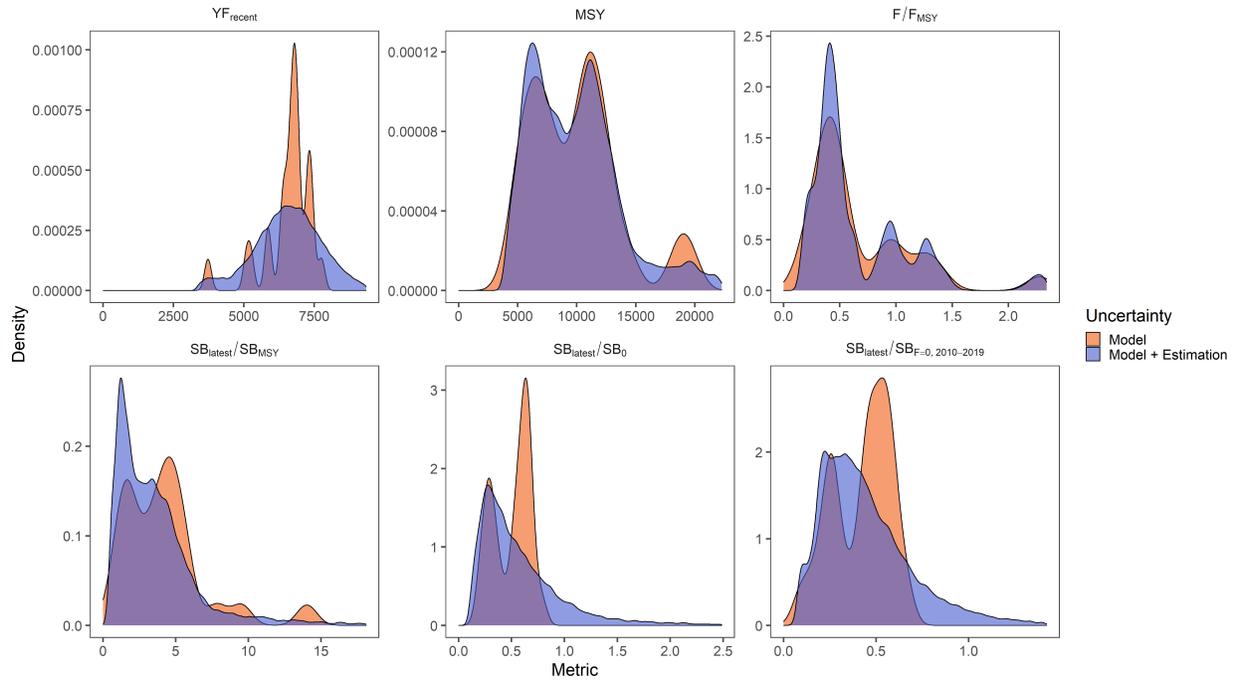
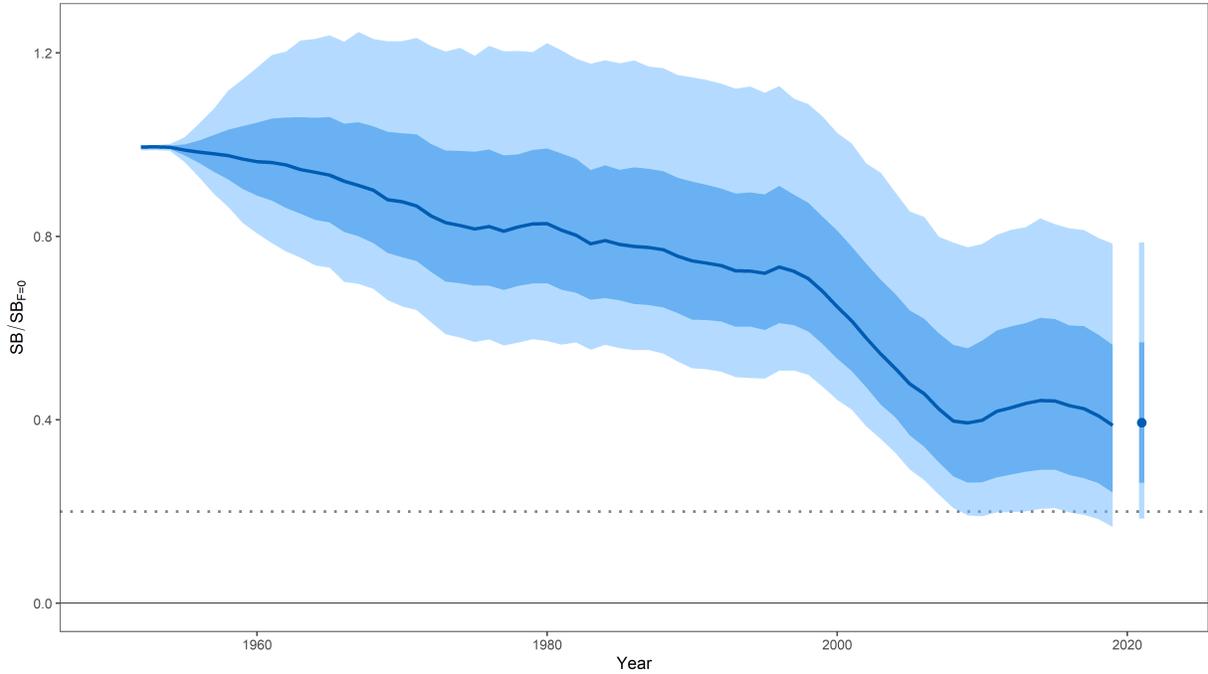
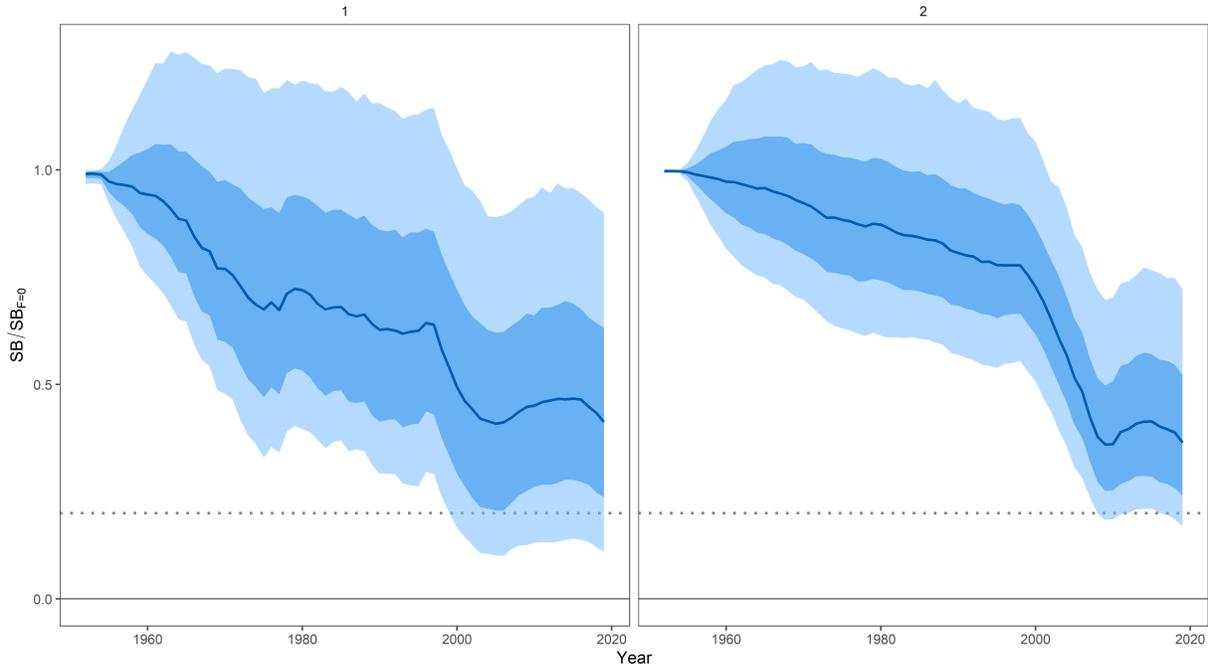


Figure 51: Distributions key management quantities contrasting the uncertainty captured only by the ensemble (model; orange) and the more holistic definition of uncertainty (model + ensemble; blue)



(a) Total depletion.



(b) Regional depletion.

Figure 52: Uncertainty in depletion where uncertainty is characterized as model + estimation uncertainty. The median is showed by the dark line, the 50th quantile showed by the dark band, and the 80th quantile showed by the light band. The median and quantiles for total $SB_{latest}/SB_{F=0}$ are shown in panel (a). For reference, the WCPFC tropical tuna LRP 20% $SB_{F=0}$ is shown with the dotted line.

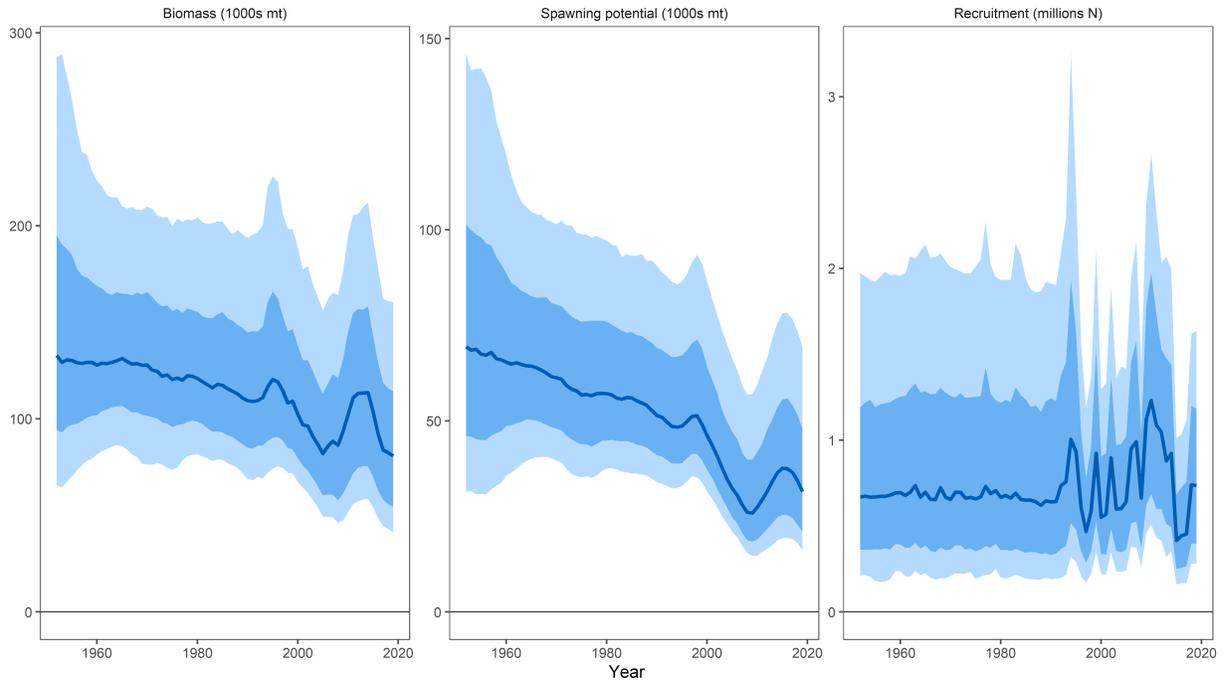


Figure 53: Temporal uncertainty in total biomass (left), spawning potential (center) and recruitment (right). Uncertainty is characterized as model + estimation uncertainty. The median is showed by the dark line, the 50th quantile showed by the dark band, and the 80th quantile showed by the light band.

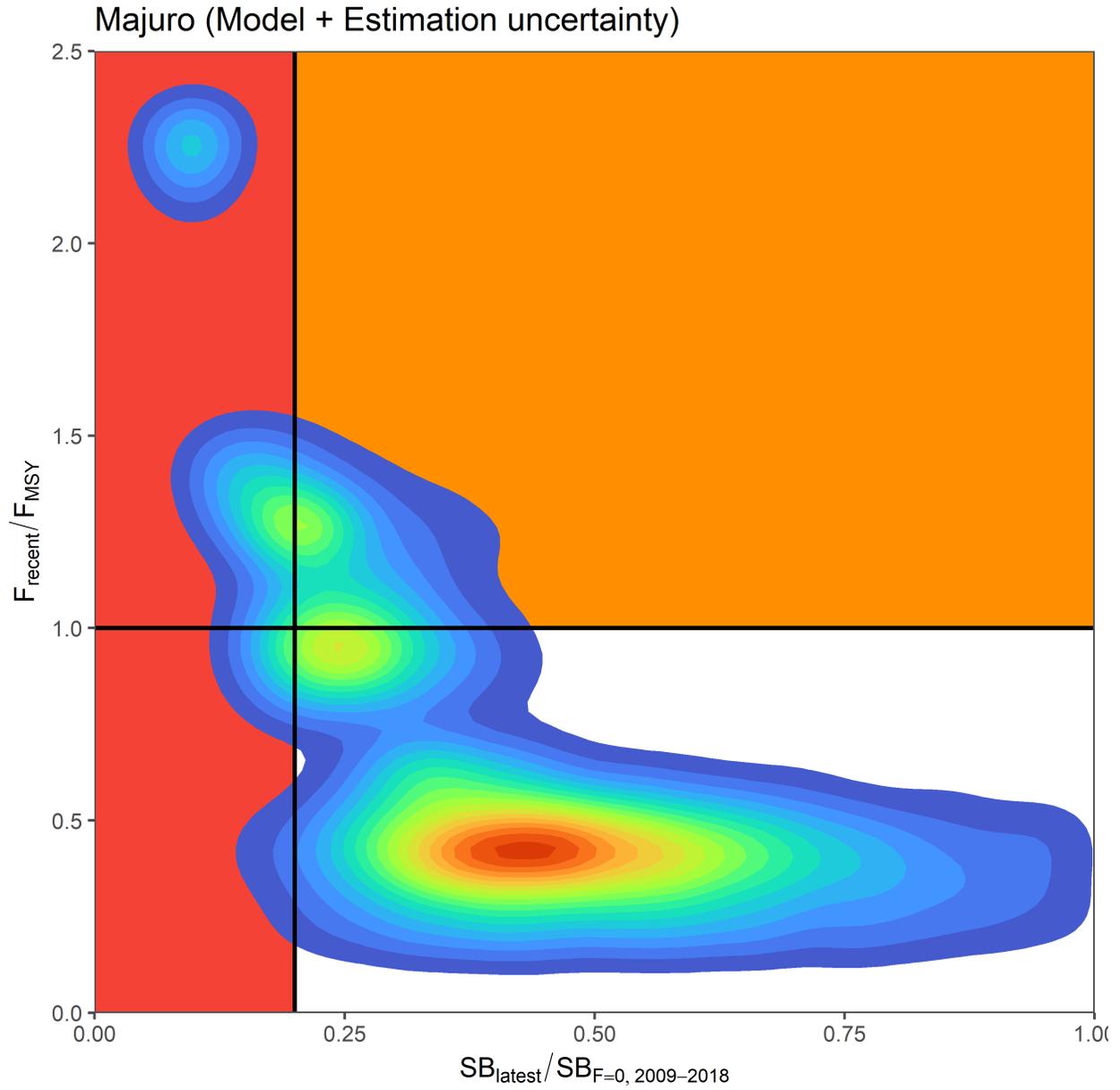


Figure 54: Uncertainty in terminal stock status, based on the 12,500 bootstrap samples characterizing the model + estimation uncertainty. Warmer colors indicate a greater density of samples, while cooler colors show the fringe of the distribution.

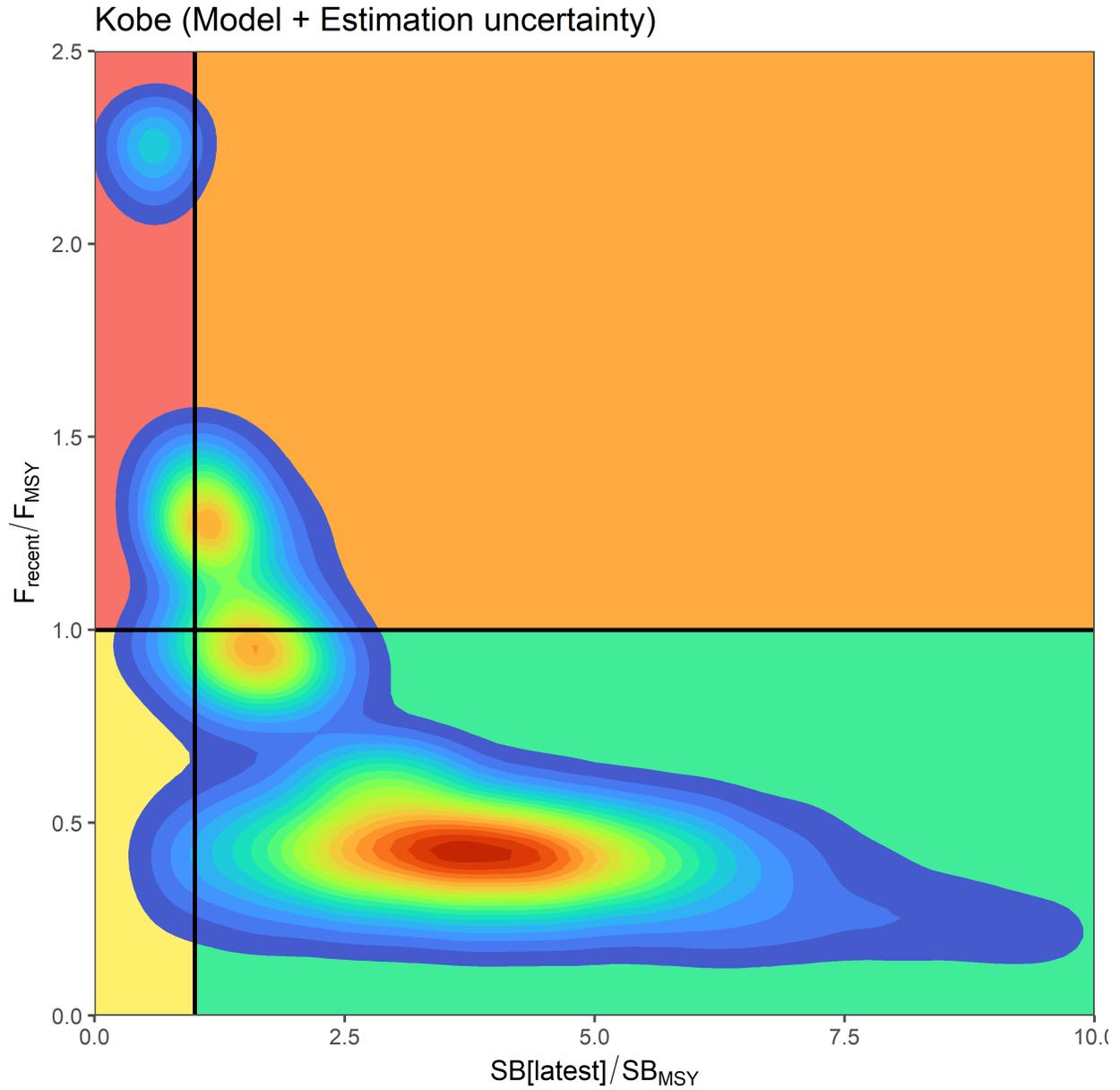


Figure 55: Uncertainty in terminal stock status, based on the 12,500 bootstrap samples characterizing the model + estimation uncertainty. Warmer colors indicate a greater density of samples, while cooler colors show the fringe of the distribution.

12 Appendix

12.1 Diagnostic Case

12.1.1 Likelihood profiles

Calculation of the likelihood profile (Figure 56) on average adult biomass for the diagnostic case indicated conflict between the different data components, with the overall fit resulting from a compromise between the different data sets. The standardized CPUE, length composition, and stock-recruitment penalty indicated a preference for the highest estimated adult biomass ($\sim 70,000$ mt) while the weight composition data and other penalty components indicated a lower estimate of average adult biomass ($\sim 15,000$ mt). An examination of the fisheries specific breakdown of the likelihood profile (Figure 57) shows that this fit is largely driven by data from only a few fisheries. The Australian weight frequency and standardized CPUE data prefer a lower estimate of average adult biomass, the New Zealand standardized CPUE a higher adult biomass, and the EU standardized index an intermediate level of adult biomass. Length composition across all fleets generally prefer higher estimated biomass levels. These patterns are consistent when considering likelihood profiles for average total biomass, recent adult (total) biomass, and latest adult (total) biomass.

Conflict between the different data components is indicative of potential model or data misspecification. Disagreement between the two types of size-frequency data could be an indicator that there is an issue with the conversion factors used to prepare the length frequency data. It is also possible that this disagreement is a product of the length and weight frequency data coming from distinct groups of fisheries targeting different segments of the population or with different data collection programs (e.g. weight frequency data are more likely to come from port sampling for some fisheries). The effect of recruitment variability around the estimated stock-recruitment relationship was non-negligible on estimated stock status. This is consistent with the higher penalty assumed in the diagnostic case relative to WCPO tropical tuna stock assessments, though uncertainty in the assumed penalty value is considered in the model ensemble.

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

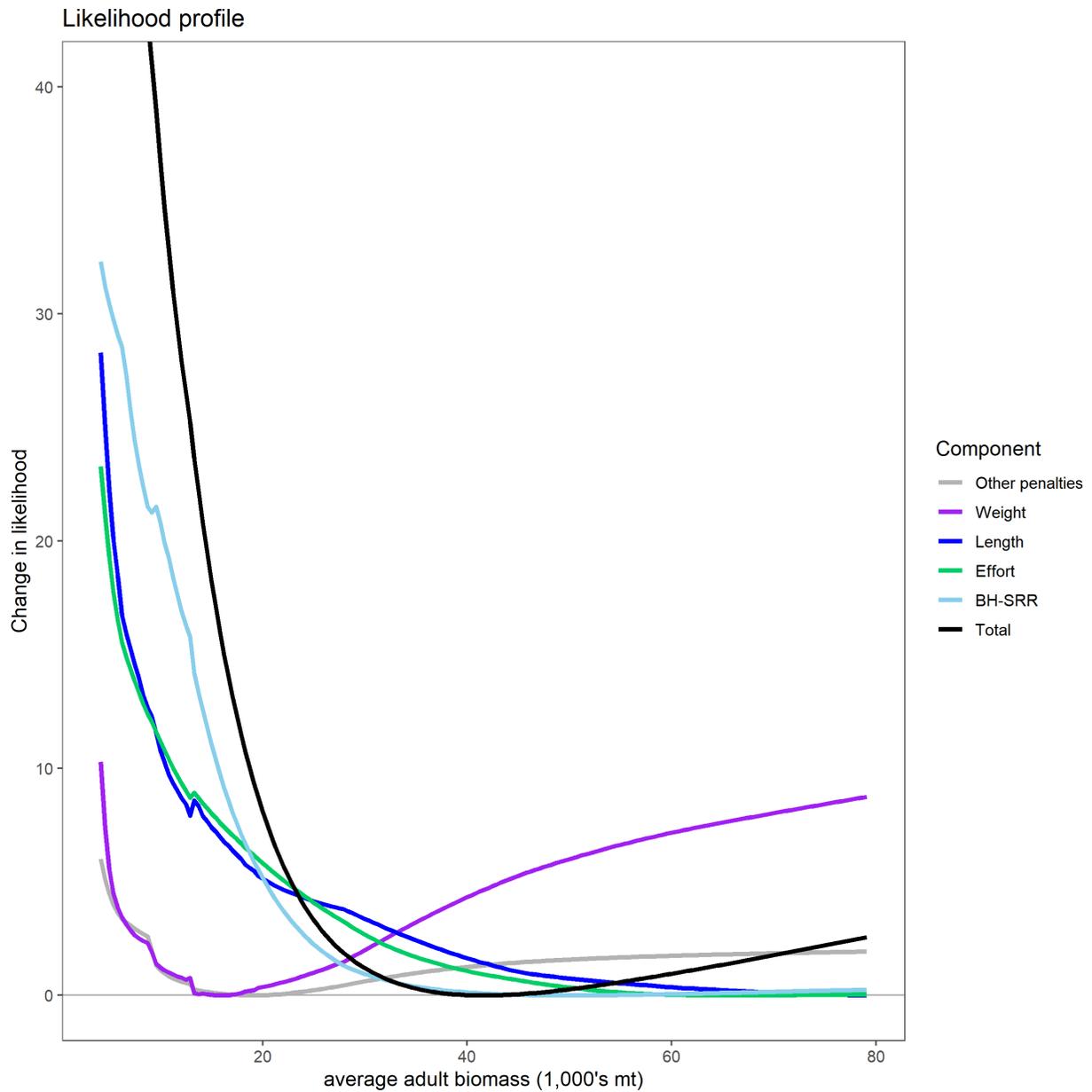


Figure 56: Likelihood profile with respect to average adult biomass for the diagnostic case. The total likelihood (black) is shown along with the influence of each likelihood component: weight frequency (purple), length frequency (blue), CPUE/effort deviates (green), the recruitment variability (light blue), and other penalties including the catch likelihood (gray).

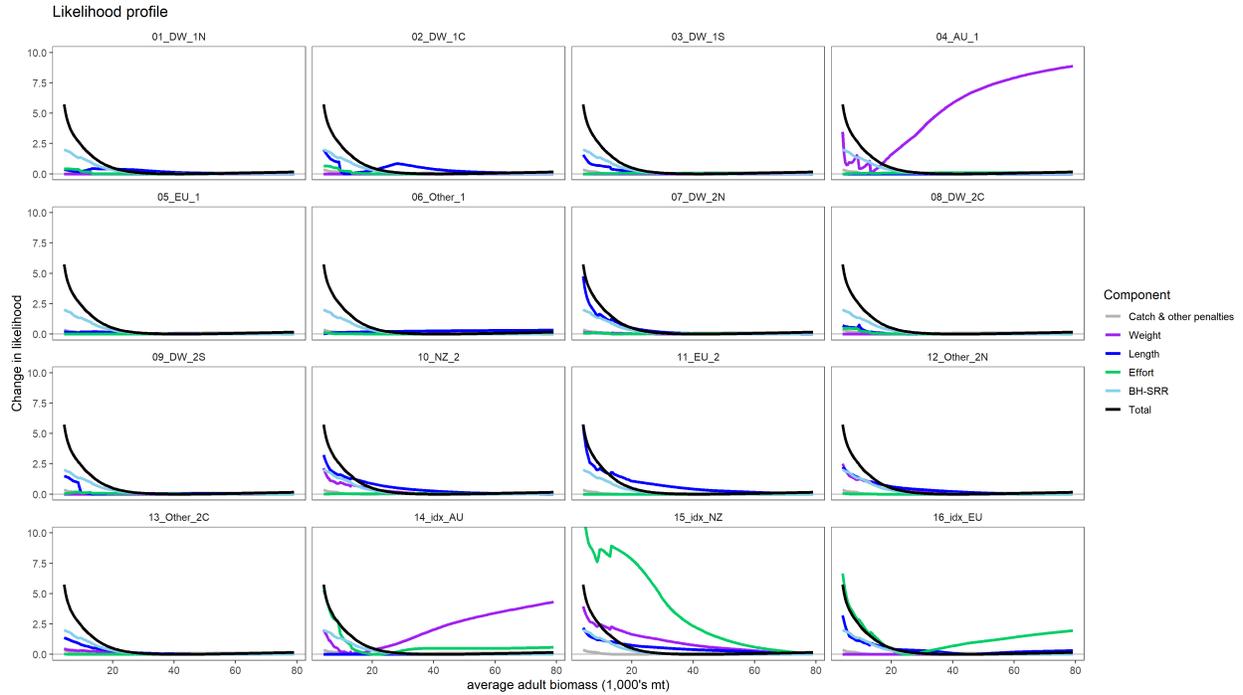
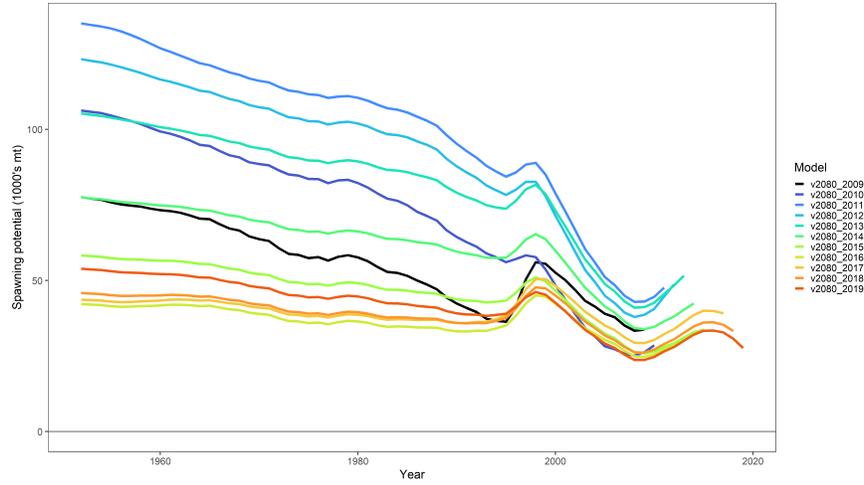


Figure 57: Likelihood profile by fishery with respect to average adult biomass for the diagnostic case. The total likelihood (black) is shown along with the influence of each likelihood component: weight frequency (purple), length frequency (blue), CPUE/effort deviates (green), the recruitment variability (light blue), and other penalties including the catch likelihood (gray).

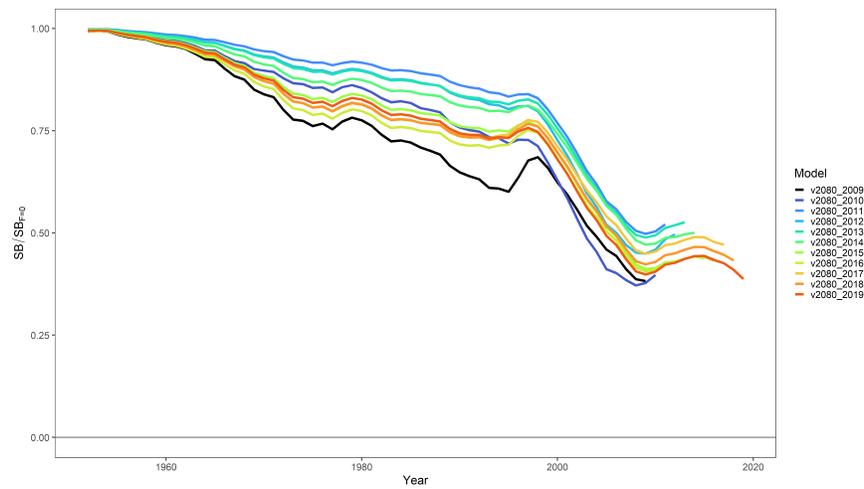
12.1.2 Retrospective analyses

An additional diagnostic, the retrospective analysis, was used to assess the overall stability of the 2021 diagnostic case model and to identify any persistent biases in estimated quantities as a result of possible model mis-specification. A series of 9 additional models were fitted starting with the removal of one year of data from the model, followed by models with that sequentially “peeled” away all input data for the years 2018-2009. The models are named below by the final year of data included (e.g. 2009-2018). A comparison of the spawning biomass, depletion, and recruitment trajectories are shown in [Figure 58](#). Additionally, Mohn’s ρ was calculated to ascertain the degree of retrospective bias ([Hurtado-Ferro et al., 2015](#)). Values of ρ for $SB_t/SBt, F = 0$ & recruitment¹² ($SB_t/SBt, F = 0 = 0.08156$; recruitment = 0.08651) were less than ± 0.2 , and did not indicate strong retrospective bias. However, there did appear to be a large retrospective pattern in SB ($\rho = 0.34396$) which indicated that the model lacked strong information to inform total population scale, and was sensitive to the excluded CPUE and/or size composition data in terminal years. Indeed, likelihood profiling on adult terminal biomass for each retrospective “peel” ([Figure 59](#)) indicated that a shift in the profile for the CPUE component was driving the higher/lower estimates of terminal adult biomass. An examination of the fishery specific components indicated that this shift was driven by the fit to the New Zealand CPUE index which increased, and subsequently decreased drastically over the retrospective period. Predictably, when terminal New Zealand CPUE was high or close to its peak, the model estimated a larger total population scale to accommodate the high values from the New Zealand index. Repeating the retrospective likelihood analysis on average adult biomass indicated a similar pattern in influence from the New Zealand CPUE, but also showed influence to changes in the Australian weight composition data as well.

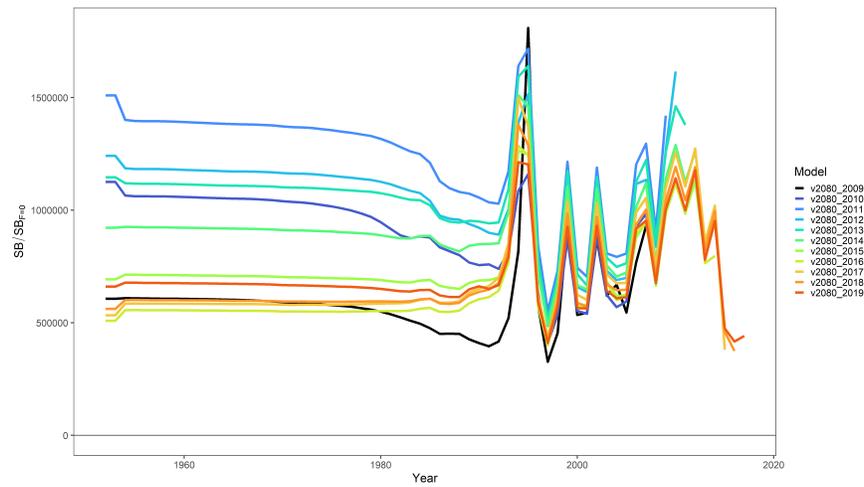
¹²Since recruitments in the final 2 years were fixed at the mean level, Mohn’s ρ was calculated based off of the year of last estimated recruitment, and as such the two most recent years for each model are not shown in [Figure 58](#).



(a) Spawning potential retrospective



(b) Depletion retrospective



(c) Recruitment retrospective

Figure 58: Estimated (a) spawning potential, (b) fishery depletion ($SB_t/SB_{t=0}, F = 0$), and (c) recruitment for each of the retrospective models.

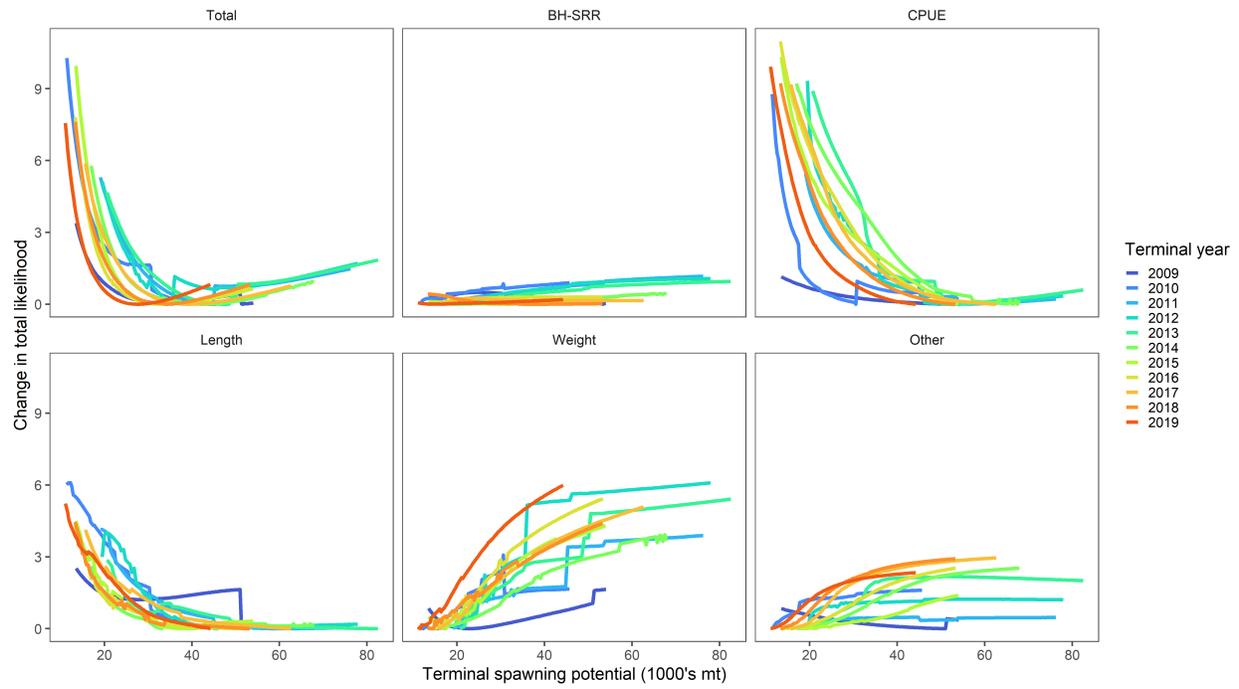


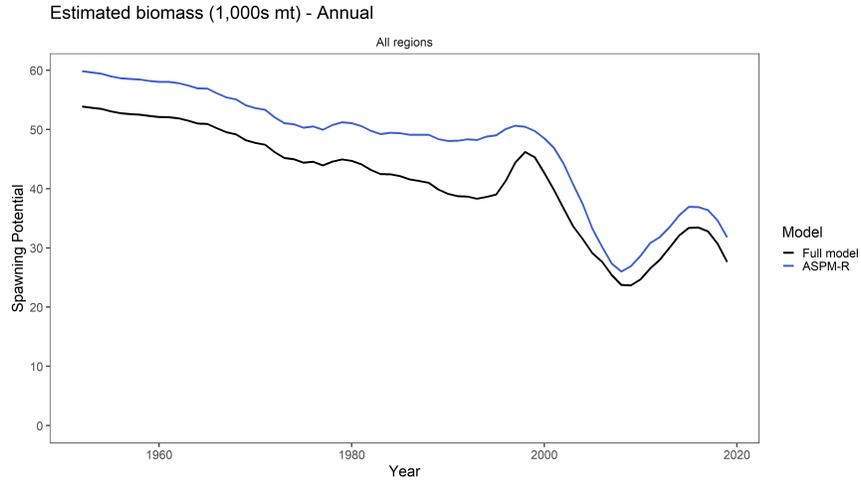
Figure 59: Likelihood profile on terminal adult biomass by retrospective “peel” (color) and likelihood component (panel).

12.1.3 Age-structured production model

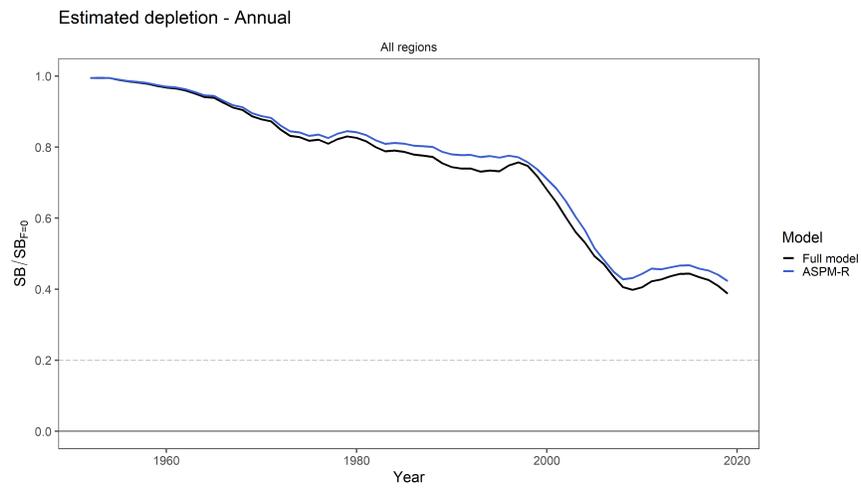
An Age-structured production model (ASPM) analysis was explored for the diagnostic case. This was accomplished by removing the influence of the size data on the model likelihood, fixing selectivity to be equal to those estimated from the full model, and not estimating recruitment deviations. Additionally, regional recruitments were estimated to be the identical for both regions. A variant of the ASPM, the ASPM-R allows for full estimation of regional recruitment deviates and proportion of recruitments in each region, and was also examined here.

The basic ASPM did not return sensible biomass estimates indicating that the CPUE and catch for the diagnostic case did not contain enough information to estimate a production function. However, this could also be due to the assumption of identical recruitment by region which was likely incompatible with the assumed movement dynamics. Further testing should explore the ASPM either with a single region formulation for the model or with regional recruitment partitioning estimated to see if this improves the estimation.

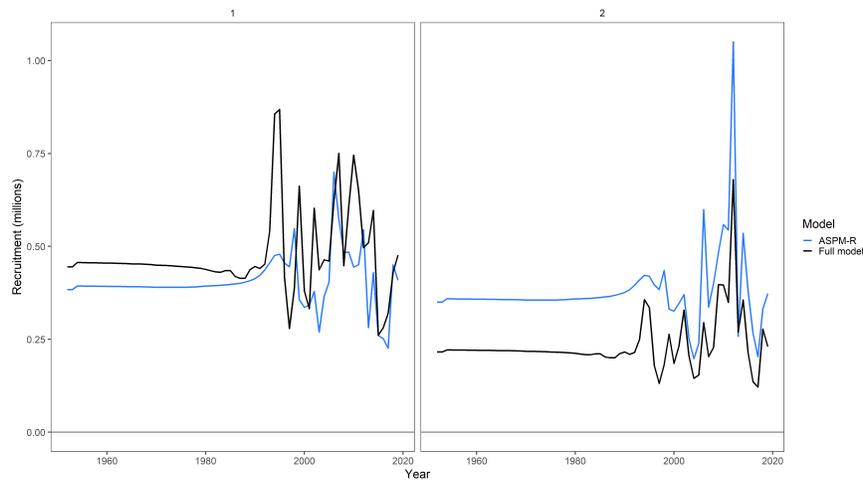
The ASPM-R version produced results that were remarkably similar to the full diagnostic case model (Figure 60). This indicated that recruitment variability is strongly informed by the CPUE data, not the size composition data, and recent trends are driven by patterns in recruitment.



(a) Spawning potential retrospective



(b) Depletion retrospective



(c) Recruitment retrospective

Figure 60: Estimated (a) spawning potential, (b) fishery depletion ($SB_t/SB_{t=0}, F = 0$), and (c) recruitment from the ASPM analysis.

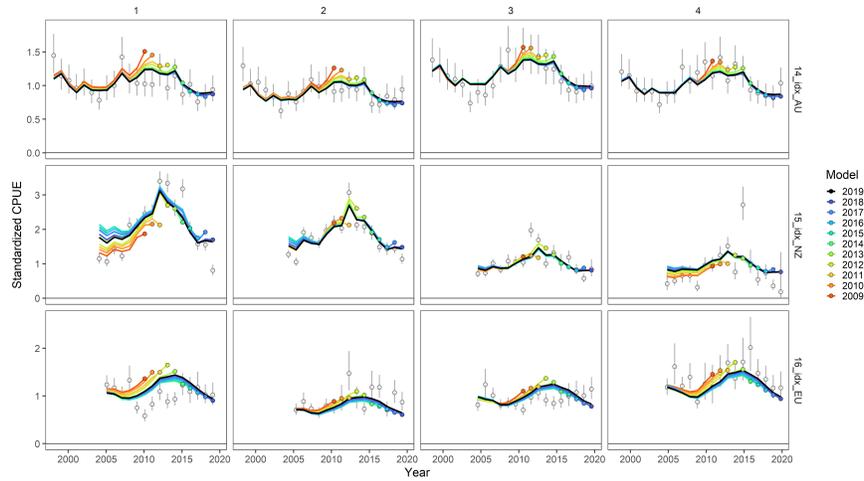
12.1.4 Hindcasting

“Model-free” hindcasting was done according to the approach outlined in [Kell et al. \(2021\)](#), where the influence of all the CPUE on the model likelihood was retrospectively removed (by simultaneously applying an extremely low penalty to the fit to the CPUE indices for terminal years), with all other aspects of the model unchanged. This is equivalent to physically removing the CPUE from the assessment model though it makes the calculations of *predicted* CPUE in the hindcast period considerably easier. Hindcasting was done over a 10 year interval (2009-2018), and CPUE was predicted ahead 1 – 5 (h_{window}) years from the last year of the CPUE “peel”. Predictive performance and internally consistency were assessed using Mean Absolute Square Error (MASE) ([Hyndman and Koehler, 2006](#)) between the hindcast predicted CPUE and the observed CPUE value. Larger values of MASE indicate poor predictive ability and internal consistency, and MASE values less than 1 indicate that the model has greater predictive ability than a naive-predictor. The predicted CPUE for the naive predictor was assumed to be equal to the last observed value from the hindcast (i.e $CPUE_{pred,2019} = CPUE_{obs,2018}$ for a one year ahead naive prediction).

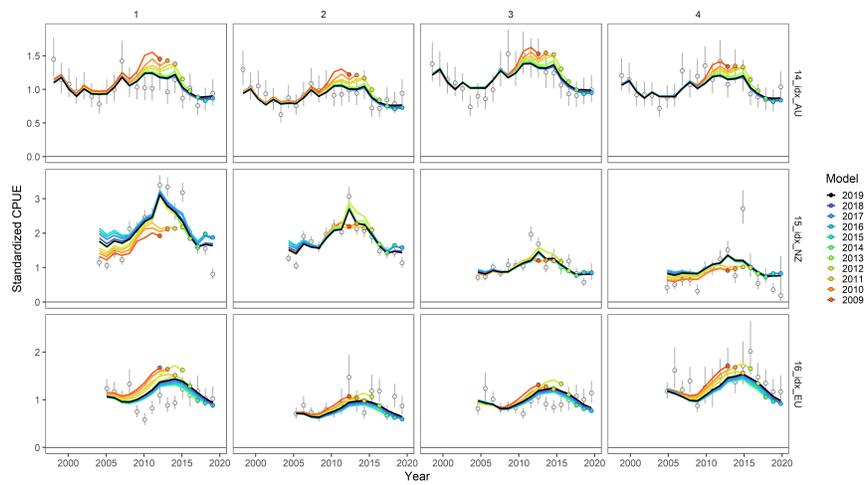
Hindcast prediction tended to improve across larger h_{window} prediction intervals ([Table 6](#) & [Figure 61](#)) indicating that as observed CPUE changed, the other components of the model (catch & size composition) were “internally consistent” with the CPUE data that were excluded from the model likelihood. Additionally, MASE tended to indicate better performance than the naive-predictor as the prediction window increased. On average, the observed CPUE for Australia ($\overline{MASE} = 0.92$) and New Zealand ($\overline{MASE} = 0.54$) were well predicted by the other data components in the model. The European Union index was found to be less consistent with the other data components in the model.

Table 6: MASE from “Model-free” hindcasting with 1–5 year prediction periods (h_{window}) for each index fishery in the 2021 diagnostic case model

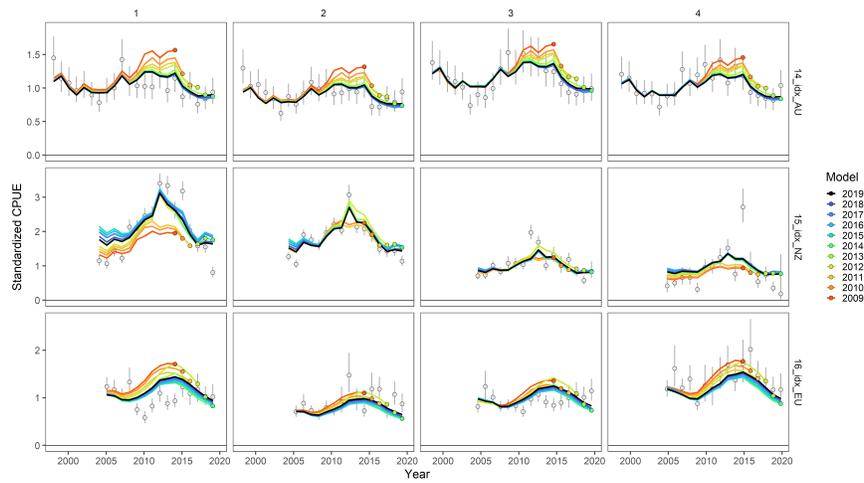
	Index fishery	h_{window}	MASE
1	14 AU idx	1	1.04
2	15 NZ idx	1	0.77
3	16 EU idx	1	1.293
4	14 AU idx	2	0.986
5	15 NZ idx	2	0.749
6	16 EU idx	2	1.271
7	14 AU idx	3	0.747
8	15 NZ idx	3	0.585
9	16 EU idx	3	1.227
10	14 AU idx	4	0.782
11	15 NZ idx	4	0.388
12	16 EU idx	4	0.996
13	14 AU idx	5	1.05
14	15 NZ idx	5	0.256
15	16 EU idx	5	0.915



(a) One year hindcast predictions



(b) Three year hindcast predictions



(c) Five year hindcast predictions

Figure 61: Hindcast predictions of CPUE for (a) one, (b) three, and (c) five year prediction windows. Indices are broken into quarterly panels for interpretability. Line color indicates last year of CPUE influence on model estimates.

12.1.5 Residual plots

Additional residual diagnostics for the fit to the standardized CPUE, weight composition, and length composition data for the 2021 SWPO swordfish diagnostic case model (Section 8.2).

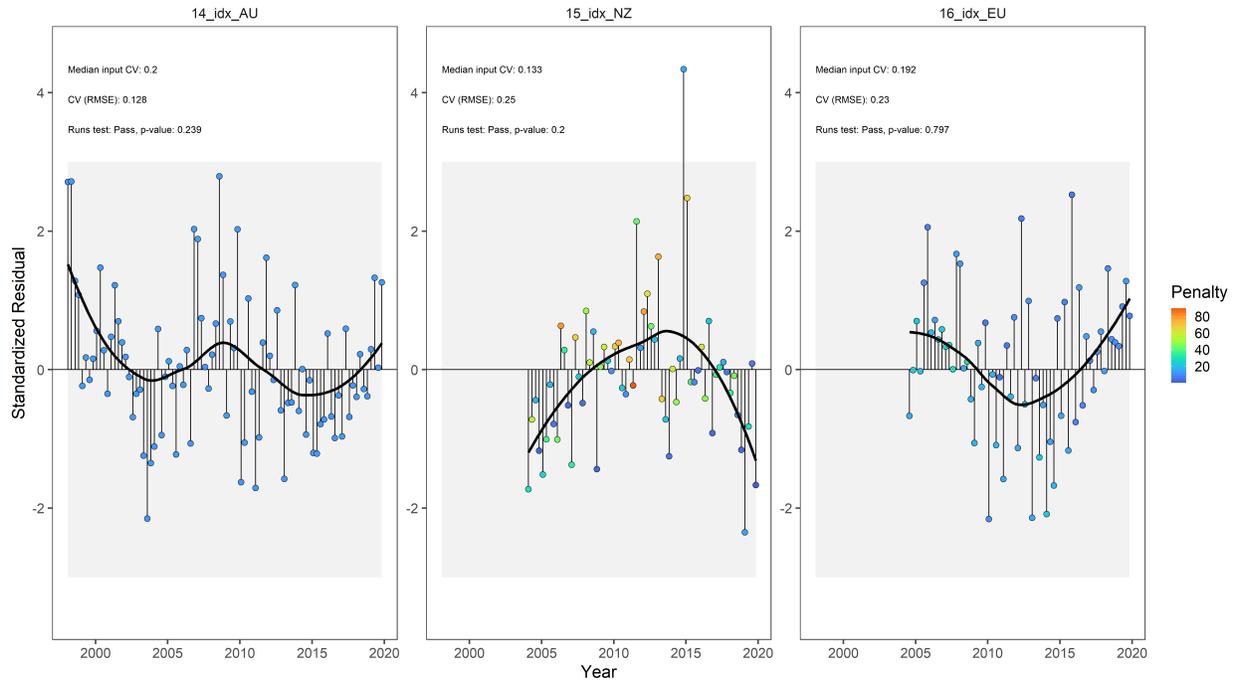


Figure 62: Standardized residuals of the fit to the index fishery CPUE data. The color of the point denotes the input penalty associated with that observation. A large penalty indicates greater precision in the estimate which should pull the estimate closer to the observation. Within each panel the median input CV, the RMSE of the fit, and the results of a “runs”-test (Carvalho et al., 2021) for non-random residual variation is shown for each fishery. Additionally, the light gray background box indicates the 3σ rule used to identify outliers.

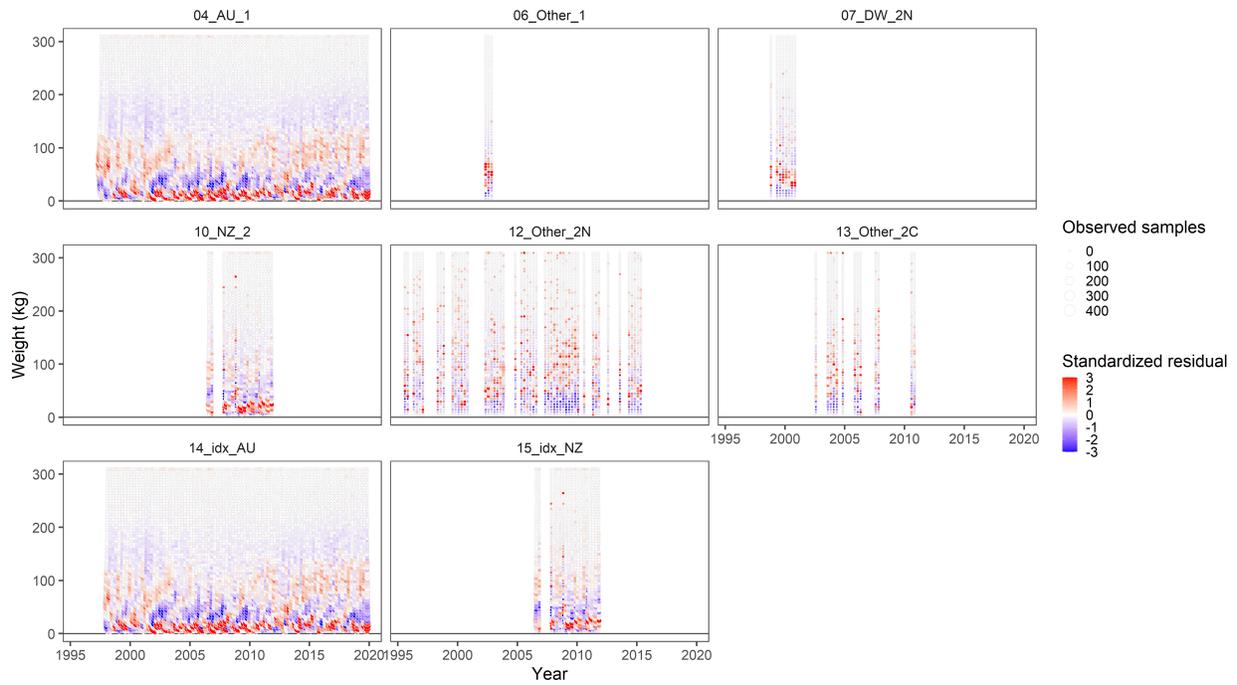


Figure 63: Residual bubble plot of the fit to the weight composition data by fishery. Circle size denotes the number of observations in a given composition bin and temporal strata. The color indicates an underestimate (blue) or an overestimate (red).

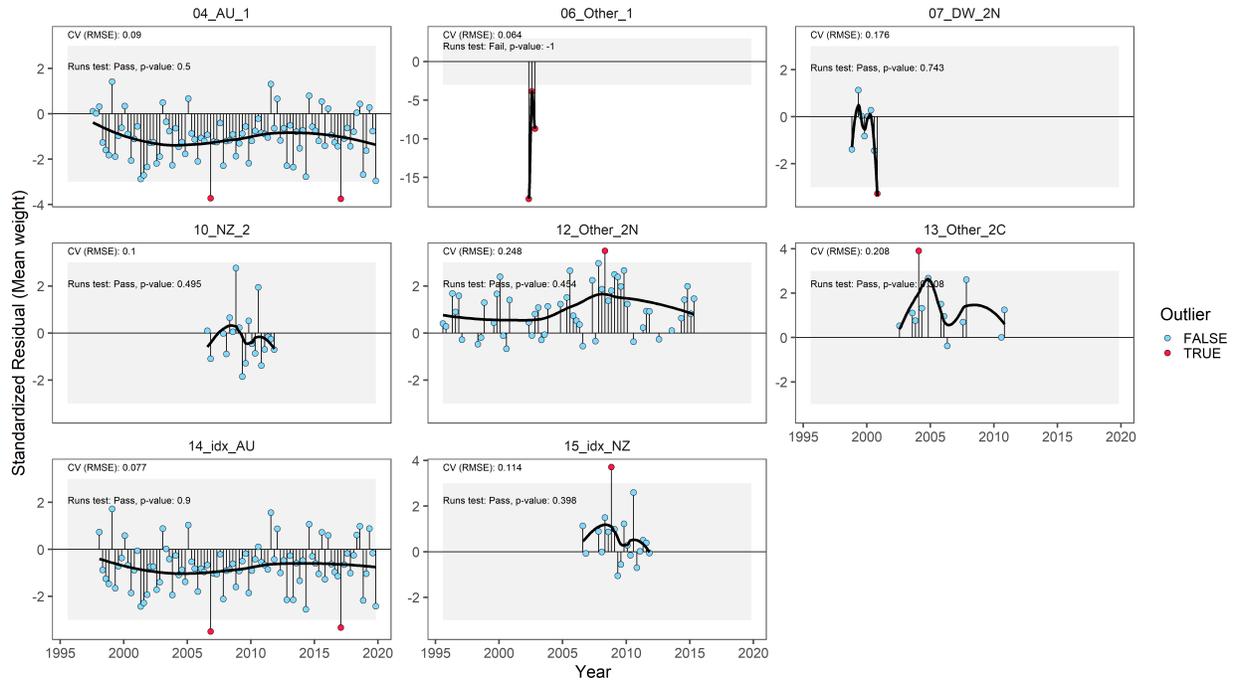


Figure 64: Standardized residuals of the fit to the mean weight frequency data by fishery. The color of the point denotes whether or not the observation is an outlier according to the 3σ rule. Within each panel the RMSE calculated on the basis of the observed and predicted proportions at weight, and the results of a “runs”-test (Carvalho et al., 2021) for non-random residual variation is shown for each fishery.

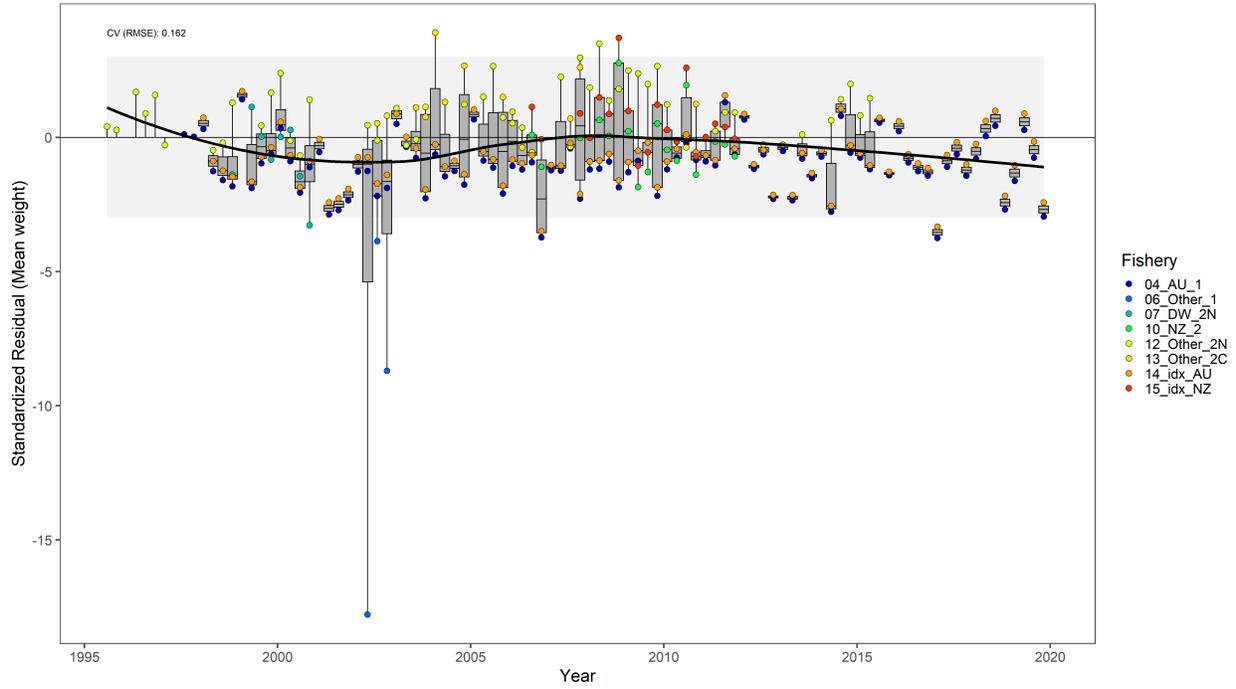


Figure 65: Joint residual plot following [Winker et al. \(2018\)](#) combining standardized residuals of the fit to the average weight where color denotes the fishery, boxplots indicate the median and quantiles across all residuals by time step, and a smoothed spline through all residuals shows the overall trend of the fit. Additionally, the light gray background box indicates the 3σ rule used to identify outliers.

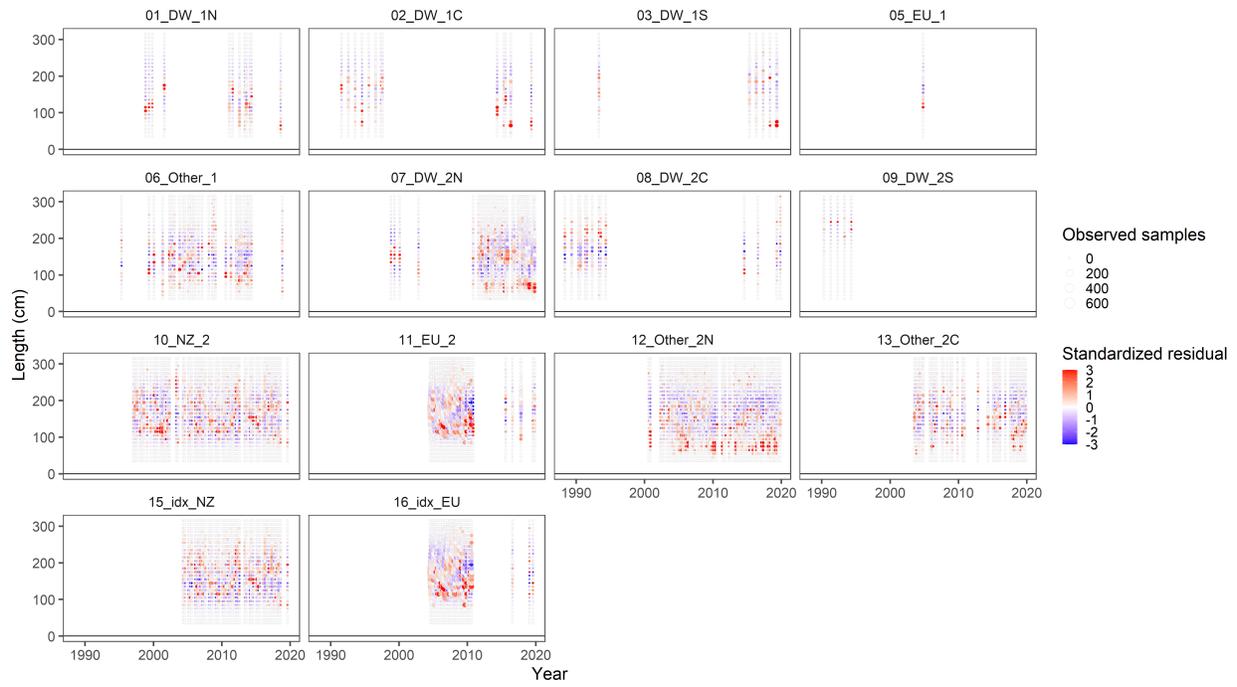


Figure 66: Residual bubble plot of the fit to the length composition data by fishery. Circle size denotes the number of observations in a given composition bin and temporal strata. The color indicates an underestimate (blue) or an overestimate (red).

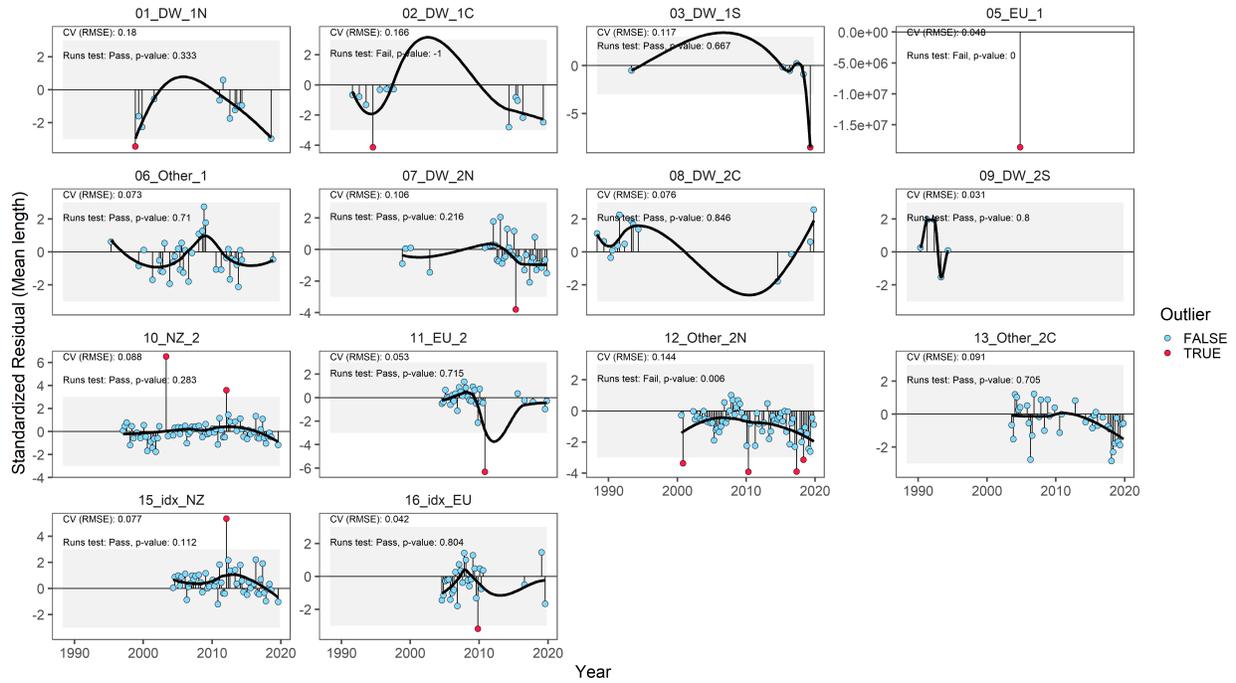


Figure 67: Standardized residuals of the fit to the mean length frequency data by fishery. The color of the point denotes whether or not the observation is an outlier according to the 3σ rule. Within each panel the RMSE calculated on the basis of the observed and predicted proportions at length, and the results of a “runs”-test (Carvalho et al., 2021) for non-random residual variation is shown for each fishery.

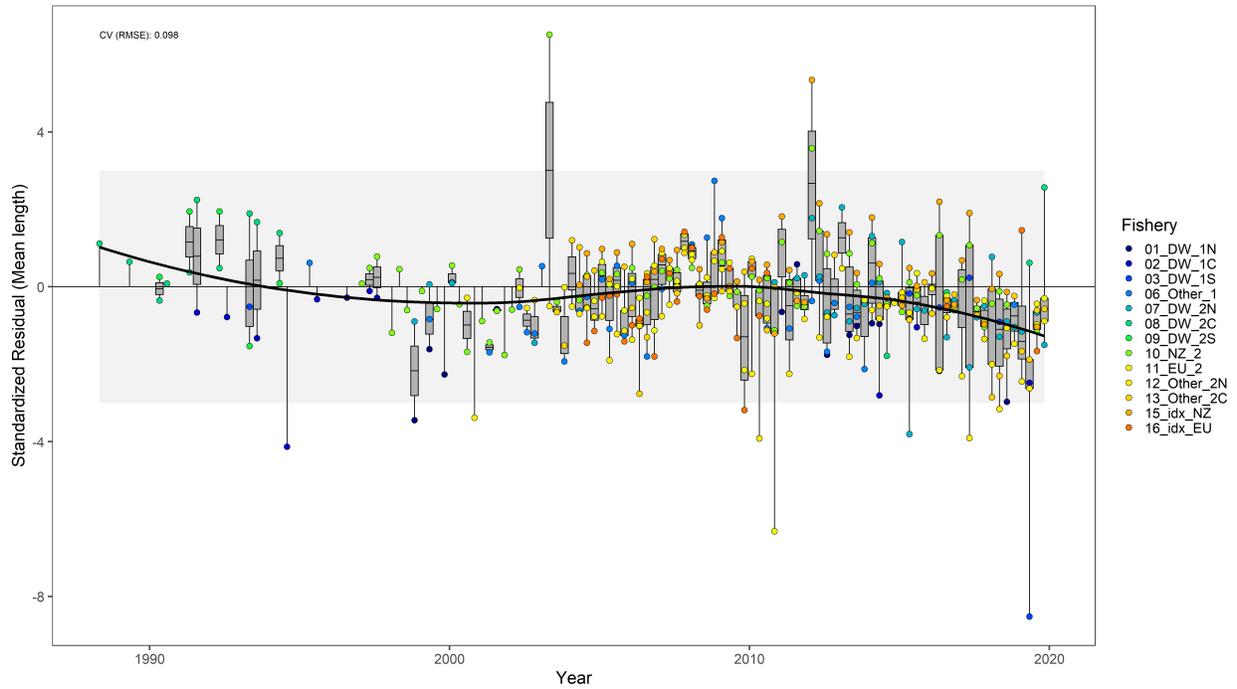


Figure 68: Joint residual plot following Winker et al. (2018) combining standardized residuals of the fit to the average length where color denotes the fishery, boxplots indicate the median and quantiles across all residuals by time step, and a smoothed spline through all residuals shows the overall trend of the fit. Additionally, the light gray background box indicates the 3σ rule used to identify outliers.

12.2 Ensemble weighting

Sensitivity in management reference points relative to choice of ensemble weighting scheme was evaluated. An objective, diagnostic based approach for weighting the ensemble was developed using hindcast predictions of CPUE. Given that the data and likelihoods changed from model to model in the grid, other likelihood based approaches such as Sample-Importance Resampling would have been inappropriate to consider. Since CPUE was found to be an important driver of model dynamics via the predicted recruitments (Section 12.1.3), good hindcast predictability of observed CPUE was felt to be a more appropriate indicator of model performance and internal consistency than the ability to predict other data components of the model such as the size frequency data.

Hindcast predictions of CPUE were conducted across all models in the structural uncertainty grid in the manner shown in Appendix Section 12.1.4. Ensemble weights were then constructed as the inverse of the average MASE across all CPUE indices and prediction windows for a single model. This construction ensured that models with lower average MASE (better predictive performance and internal consistency) received a higher ensemble weighting. Model and estimation uncertainty were combined as described in Section 6.1.1 but the number of bootstrap samples used to propagate the estimation uncertainty for each model was adjusted in proportion to the derived model weight. Sensitivity to the weighting construction was also explored, and weights were constructed by taking the inverse of MASE^n . Raising the MASE to larger powers of n had the effect of increasing model weights for models with better performance and decreasing models with poorer performance.

Using the final 25 model ensemble, results were robust to the weighting scheme used unless n was very large (i.e $n = 10$; Figure 69). The most extreme $n = 10$ weighting scheme resulted in distributions that were more dramatically skewed towards the outcomes of models that had good MASE scores. Analysis of MASE scores relative to model configuration (Figure 70) indicated that these models typically were those that downweighted the New Zealand CPUE index (similarly up weighting the Australian index & not fitting to either Japan or Chinese Taipei CPUE), increased the weighting on the composition data, assumed larger values of M_{ref} , larger values of steepness, and lower recruitment variability around the stock-recruit relationship. Increasing the weighting towards these models typically resulted in a more pessimistic stock status.

Ensemble weighting using strong weights ($n = 10$) also corroborated the post-hoc filtration scheme applied to retain models for the final model ensemble (Figure 71). Applying a strong re-weighting to the full unfiltered ensemble (384 models) resulted in similar distributions of the management metrics (except F_{recent}/F_{MSY}) to the un-weighted but filtered final ensemble (25 models). This suggests that the filtration scheme applied is also justifiable on the basis that it predominantly excluded models that have poor hindcast predictive performance. Using the full unfiltered model ensemble with strong weighting as the basis for management advice would result in a more pessimistic stock status (Table 7).

More research is needed into the most appropriate ensemble weighting method. As demonstrated

here, even using an objective diagnostic based approach such as hindcast performance still required a subjective decision as to how to best construct the weight (i.e choice of n). Additionally, choice of h_{window} to consider, or even to conduct hindcast predictions on a different component of the model (i.e. length composition) could lead to a different weighting and perhaps to different management advice.

Table 7: Summary of reference points (measures of central tendency & relevant percentiles) including model and estimation uncertainty from the full 384 model ensemble. Models were weighted inversely proportional to the average Mean Absolute Square Error (MASE) from model-free hindcasting of the relative abundance indices.

	Mean	Median	01	10	90	99
C_{latest}	7,750	7,810	7,364	7,560	7,831	8,444
YF_{recent}	6,598	6,550	3,953	5,207	8,058	9,373
MSY	8,961	7,281	4,344	5,514	15,020	21,746
F_{recent}/F_{MSY}	0.75	0.84	0.15	0.26	1.14	2.14
SB_0	97,018	87,773	18,975	38,729	157,805	283,242
SB_{latest}	37,238	31,645	11,163	17,365	63,803	116,686
SB_{recent}	40,015	35,666	15,401	20,867	64,977	94,616
SB_{MSY}	14,688	15,857	3,101	6,514	21,404	25,589
SB_{latest}/SB_{MSY}	3.07	1.99	0.49	1.02	6.36	14
SB_{recent}/SB_{MSY}	3.43	1.99	0.72	1.29	6.89	16
SB_{latest}/SB_0	0.48	0.35	0.11	0.18	0.88	2.21
$SB_{latest}/SB_{F=0}$	0.4	0.33	0.1	0.19	0.7	1.33

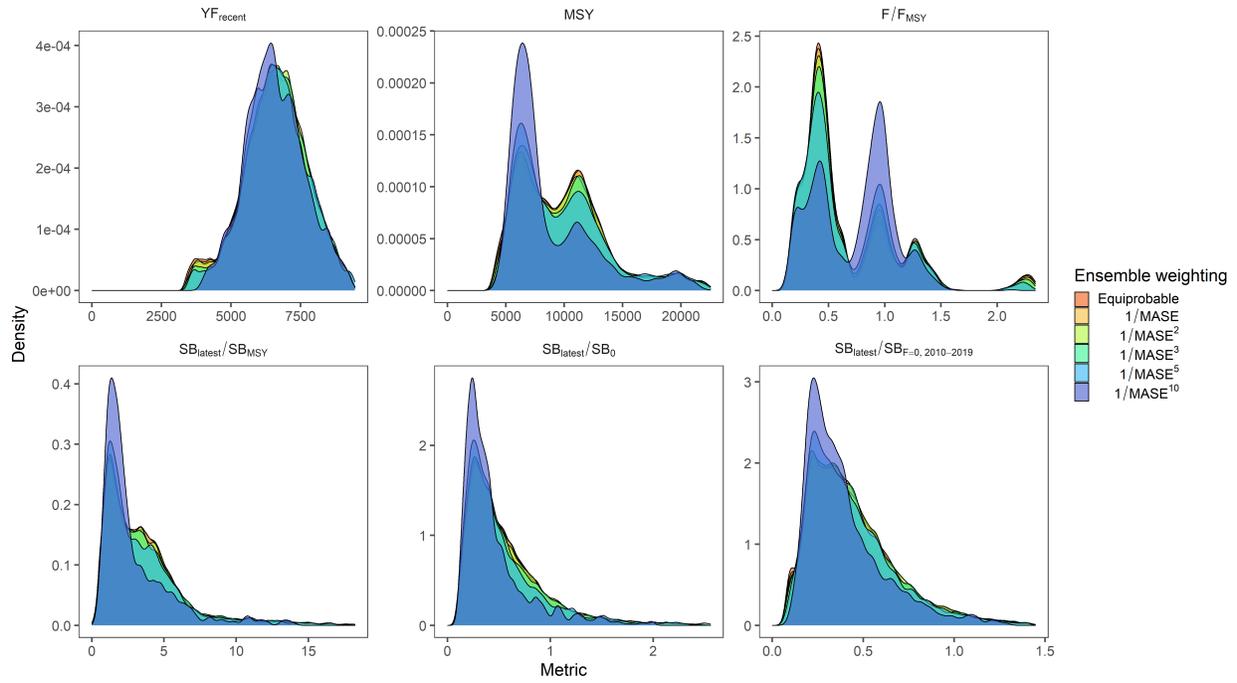


Figure 69: Distributions of key management quantities showing the effects of different assumed ensemble weights for the final models retained in the ensemble.

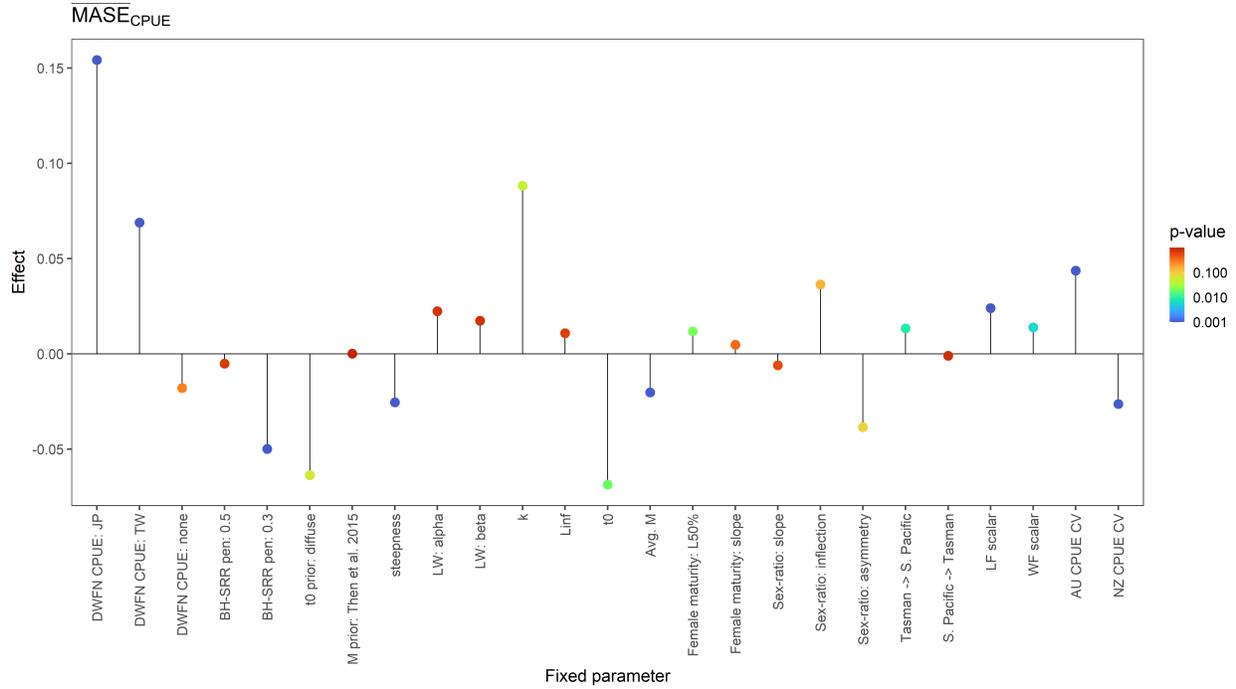


Figure 70: Estimated covariates from a generalized linear model of biological parameters against MASE where the color of the point is the p-value of the effect and cooler colors indicate greater significance.

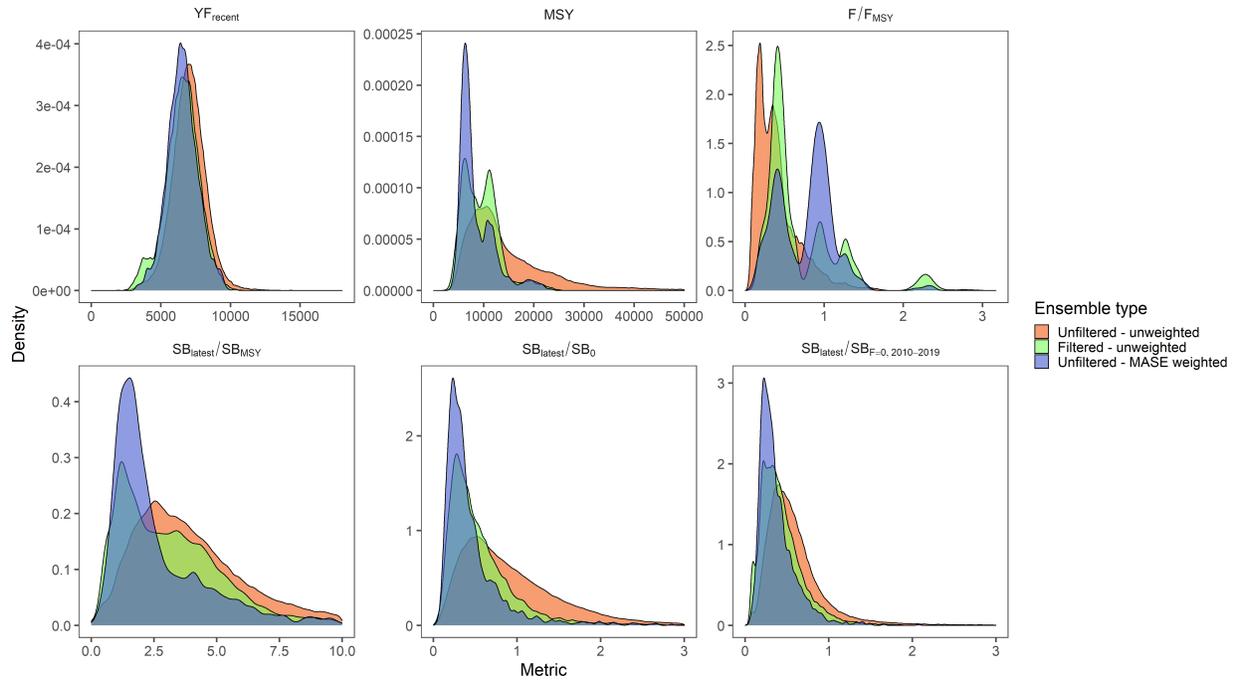


Figure 71: Distributions of key management quantities showing the effects of applying filtering and/or re-weighting the ensemble.

12.3 Sex-disaggregated model ensemble

To explore the effects of model mis-specification in terms of the sex-structure on the management advice, the full 384 model ensemble was replicated assuming a sex-disaggregated structure following the approach described in [Section 7.2.1](#). In this ensemble, each sex assumed distinct biological characteristics drawn from sex-specific joint priors and the growth variability parameters were fixed since these were shown to be poorly estimated given the lack of sex-disaggregated composition data. For σ_{Age} these were held fixed at the estimated standard deviation from the Bayesian analysis used to construct the prior ([Ducharme-Barth et al., 2021b](#)), and for σ_{Length} these were assumed to come from a uniform distribution $U(0,0.3)$. The sex-disaggregated ensemble was subsequently filtered applying the same criteria as for the sex-aggregated ensemble, though omitting the filter for σ_{Age} as this parameter was fixed for each sex. This resulted in a final sex-disaggregated ensemble of 58 models. The distributions (incorporating model and estimation uncertainty) of $SB_{latest}/SB_{F=0}$ for this ensemble, the sex-disaggregated version of the 25 models retained for the sex-aggregated ensemble, and the final sex-aggregated ensemble were compared ([Figure 72](#)). These showed similar distributions and central tendencies, though the sex-disaggregated distributions appeared to be more uncertain relative to the sex-aggregated ensemble.

This application is the furthest that the multi-sex formulation has been pushed in MULTIFAN-CL, and venturing into relatively uncharted areas of the source code is not without surprises. Unfortunately, it was found that estimation uncertainty for yield based reference points was not being calculated properly, thus precluding the use of a sex-disaggregated model for management advice at this time. However, our limited comparison shows that on the basis of $SB_{latest}/SB_{F=0}$, the sex-aggregated model ensemble does not appear to be producing substantially different advice to the advice from the sex-disaggregated ensemble.

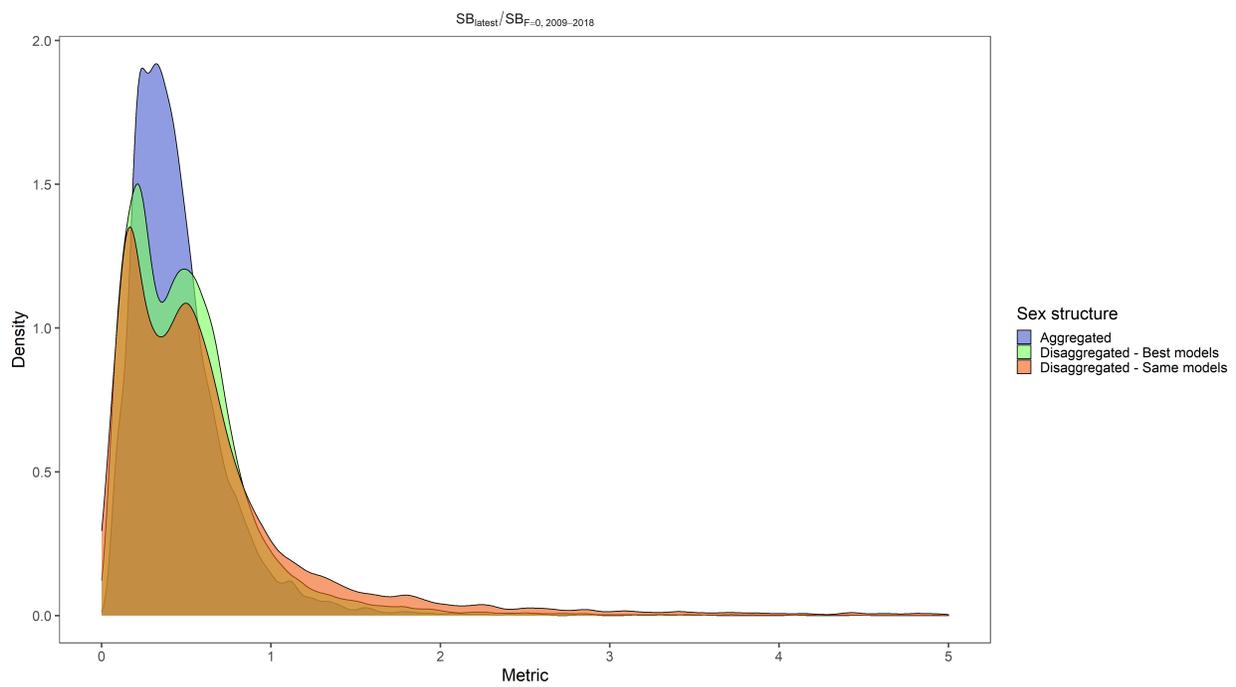


Figure 72: Distributions of $SB_{latest}/SB_{F=0}$ showing the effects of using either a sex-aggregated or sex-disaggregated structure in the model ensemble.