

**INTER-AMERICAN TROPICAL TUNA COMMISSION**

**SCIENTIFIC ADVISORY COMMITTEE**

**16<sup>TH</sup> MEETING**

La Jolla, California (USA)

02 - 06 June 2024

**DOCUMENT SAC-16-03**

**STOCK ASSESSMENT OF YELLOWFIN TUNA IN THE EASTERN PACIFIC  
OCEAN: 2025 BENCHMARK ASSESSMENT**

Carolina Minte-Vera, Mark N. Maunder, Haikun Xu and Rujia Bi

**CONTENTS**

SUMMARY .....	2
EXTENDED SUMMARY .....	3
1. Introduction .....	15
1.1. Background .....	15
1.2. Conceptual model .....	16
2. Data .....	22
2.1. Catch .....	22
2.2. Size composition .....	23
2.3. Indices of abundance and corresponding length composition .....	26
3. assumptions and parameters .....	27
3.1. Growth .....	27
3.2. Natural mortality .....	27
3.3. Reproductive biology .....	28
3.4. Recruitment .....	28
3.5. Movement .....	28
3.6. Fishery Selectivity and data weighting .....	29
3.7. Selectivity for the indices and weighting of the index composition data .....	31
3.8. Initial conditions .....	31
3.9. Model dimensions .....	31
4. Models .....	31
4.1. Ancestral model .....	31
4.2. Reference models .....	31
5. Assessment results .....	33
5.1. Model convergence and diagnostics .....	33
5.2. Estimated trends .....	36
6. Bridging analysis .....	38
7. Risk analysis .....	39
7.1. Definition of reference points .....	39
7.2. Stock status .....	40
8. Future directions .....	41
Acknowledgements .....	41
References .....	42
Tables .....	44
Figures .....	53

## SUMMARY

1. The 2025 benchmark assessment of yellowfin tuna in the eastern Pacific Ocean continues to use a risk analysis approach to provide management advice. Three levels of hypotheses are structured hierarchically to address the main uncertainties: (level 1) the spatial structure; (level 2) effort creep, uncertainty in growth and natural mortality; and (level 3) the steepness of the stock-recruitment relationship. The data used was updated up to 2023 and the results are shown for the start of 2024.
2. The main uncertainty addressed in this benchmark assessment is spatial structure. One of the main advances in the assessment was in the delineation of regions and areas within these regions. A new cluster analysis method using length composition data was used. Extensive exploratory models were developed, including a spatial model with three regions. The spatial model was not used in the final risk analysis as it estimated movement between regions to be very low. Two independent models were developed for separate regions to best represent the localized dynamics, as well as an EPO-wide areas-as-fisheries model. The two regions are: northern and east areas (NE), where core of the purse-seine catches associated with dolphins are taken, and south and west areas (SW), where the distant-water longline fleets operate and where the floating object fisheries have expanded.
3. Four configurations represent the spatial-structure hypotheses: 1) EPO, 2) SW, 3) NE, and 4) NE-short, which is in the same area as NE, but the models start in 2006 instead of 1984, to address patterns in the data. One risk analysis was done for each spatial configuration, with eighteen models each. A “base” scenario was built with new assumptions of growth, natural mortality and reproductive biology based on results of recent research. One-off scenarios were constructed to represent the uncertainty in biological parameters (growth, low and high, and natural mortality, low and high) and effort creep (increase in the catchability of the indices of abundance of 1% a year). Three values of steepness of the Beverton-Holt stock recruitment curve (1.0, 0.9, 0.8) are considered for the third level hypothesis. In the risk analyses, equal weight was used for level 2 hypotheses and expert judgement was used for level 3 (steepness).
4. Indices of abundance and standardized length composition were obtained using spatiotemporal models fit to data from the purse-seine fishery associated with dolphins (to fit EPO, NE, NE-short) or longline fishery (SW). The indices were key to provide the relative trends in abundance but the information on absolute scale came from the length composition data.
5. All four spatial-structure hypotheses estimate the same general stock status. There is zero probability that the IATTC interim spawning biomass or fishing mortality limit reference points (RP) have been breached. The stock(s) is estimated to be well above the target RPs, the spawning biomass correspond to MSY ( $S_{MSY}$ ) and the staff proposed MSY proxy  $S_{30\%}$  (SAC-15-05), with low probability of being below these. The fishing mortality is estimated to be well below the level corresponding to MSY and the MSY proxy  $F_{30\%}$  with low probability of being above these. The EPO model is the most optimistic.

## EXTENDED SUMMARY

A benchmark stock assessment and risk analysis were conducted for yellowfin tuna in the eastern Pacific Ocean covering the period from 1984 through to the start of 2024. The main uncertainty addressed in this benchmark assessment was spatial structure. Advances were made in determining the regions and spatial definitions of fisheries (areas) based on a new cluster analysis method using length composition data. Seventy-two models based on three levels of hypotheses were used in the risk analysis. The hypotheses addressed (level 1) the spatial structure; (level 2) effort creep (changes in catchability over time), uncertainty in growth and natural mortality; and (level 3) the steepness of the stock-recruitment relationship. A model starting in 2006 was also conducted to account for the possibility of change in population or fishery dynamics before and after this period to explain differences in information content between the index of relative abundance and length composition data.

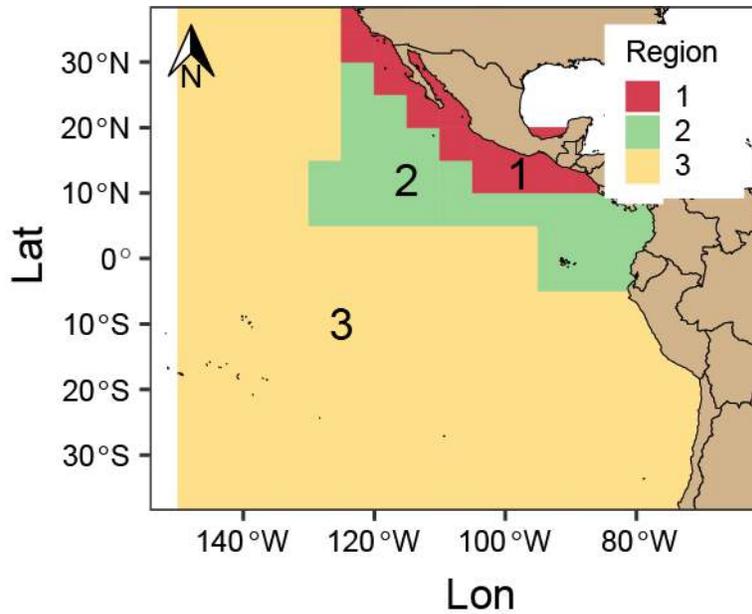
### **Stock and fishery structure**

The data suggests that there are either two or more stocks in the EPO or there is spatial structure in the population. A large recruitment in the 1990s enters the core dolphin associated purse seine fishery north of the equator in a different year than the large recruitment that enters the longline fishery south of the equator. Length composition data from the purse seine fishery associated with dolphins has smaller fish in the northeast and larger fish in the south and west, with intermediate sized yellowfin in the core area.

A three-region spatially-structured model was developed to evaluate the stock structure and movement. The regions (see Figure S-1) were delineated based on clustering length composition data. However, the model estimated limited movement among the regions. Therefore, this approach was abandoned until more information (e.g., improved tagging data) became available and when the assessment platform includes more flexibility for modelling movement.

Two approaches were used to incorporate spatial structure: 1) a single model for the whole EPO using areas-as-fleets to allow flexibility in the representation of spatial structure (EPO model) and 2) separate assessments for a) the northeast region where the core of the catches is taken (NE) and b) for the south and west region (SW). See figure S-1 for spatial definitions. A main difference among the assessments is that the indices of abundance for the EPO-wide and the NE assessments are based on dolphin associated purse seine CPUE and the index of abundance for the SW assessment is based on the longline CPUE.

Fisheries were defined in the model based on gear type (purse seine, longline, pole-and-line), purse seine set type (floating object, unassociated, dolphin associated), and area of operation to represent the different sizes of yellowfin caught. The areas were developed based on clustering length composition data. Some of the fisheries were split into small and large fish fisheries to better represent the size of the fish removed from the stock. Fisheries representing discarded small fish were also defined.



**FIGURE S-1.** Regional divisions obtained using cluster analysis of length composition data from the purse-seine fishery associated with dolphins. The EPO models consider all regions, the NE and NE\_short models include regions 1 and 2, the SW model comprised only region 3. Further subdivisions in areas were made within each region based on the cluster analysis results, which were treated as different fisheries.

#### Data

Catch, length composition, and indices of relative abundance were the main data types used in the assessment. The purse-seine catch, CPUE and length composition of yellowfin tuna are estimated by the staff using several data sources including observer reports, logbooks, and port sampling. The longline catch, CPUE and length composition data are obtained by the CPCs and submitted annually to the IATTC. Longline operational-level CPUE data by Japan, Korea and China and high-resolution vessel-specific CPUE data by Chinese Taipei were made available for this assessment.

The dolphin associated purse seine and longline CPUE-based indices of relative abundance and associated length composition data were developed based on spatio-temporal models. The longline fishery length composition data were also developed using a spatio-temporal model, but weighted by catch rather than relative abundance.

Additional data sets of reproductive biology, daily increments in otoliths and tagging obtained by the staff were used to estimate reproductive output (to define spawning biomass), growth, and natural mortality externally from the assessment models.

#### Model Assumptions

The stock assessment was conducted using Stock Synthesis, an integrated statistical age-structured stock assessment modeling platform. The models started from a fished state in 1984 (or 2006 for NE\_short) and were modelled through to the start of 2024 on a quarterly time step. Thirty age classes were defined from 0 quarters to 29 (7.25 years), with the oldest age used as a plus group. The models are sex-structured, but only natural mortality differs between females and males. The models are conditioned on catches and fit to relative abundance indices and length composition data.

The initial conditions include estimating the initial recruitment, the initial fishing mortality, and 16 recruitments deviations to represent the initial age structure. No penalty associated with initial

equilibrium catches is used. The fishery used to create the initial conditions depends on the spatial structure assumption, but in general it was chosen as a fishery with a wide range of sizes and large catches.

Growth was updated by fitting the growth cessation model to a combination of new otolith daily increment age and length data and tagging data. Information for younger fish (up to 10 quarters of age) came from the otolith data and the information for older fish came from tagged fish with expected age at recovery of 10 or more quarters and with reliable length information.

Natural mortality (M) was updated using a cohort analysis fit to recently collected tagging data and to sex ratio data. M was assumed to vary by age and sex, using the Lorenzen function to model the decline in M with age and assuming an increase in female natural mortality related to maturity. The sex ratio data came from both purse-seine and longline fisheries.

Maturity and fecundity were updated based on new data for maturity, batch fecundity, and frequency of spawning at length.

Recruitment was assumed to occur quarterly and follow a Beverton-Holt stock-recruitment relationship. The recruitment variability was implemented using a penalty function. An iterative process was used to set the standard deviation of the logarithm of the recruitment deviations and the lognormal bias correction factor. Three levels of steepness of the stock-recruitment relationship were used in the risk analysis: 1.0, 0.9, and 0.8.

Selectivity was specified using a decision tree based on the magnitude of the catch, reliability of the length composition data, and ability of a double normal selectivity curve to represent the length composition. Fleets with high catch volumes, reliable composition data, and a good fit to composition data, had time blocks in selectivity, the parameters of a double-normal selectivity curve estimated, and Francis weighting for the fit. Other fleets had no time blocks, fixed selectivity, lower data weighting, and/or not fit to the composition data. Asymptotic selectivity was used for the fisheries and surveys that catch the largest individuals (longline fisheries and/or purse-seine fisheries associated with dolphins depending on the model).

A risk analysis approach is used in this benchmark assessment. The approach starts by identifying alternative “states of nature” (i.e. hypotheses) that are considered plausible for describing the population dynamics of yellowfin tuna. The identification of those hypotheses is done in a hierarchical way, with the higher-level hypotheses representing the most important uncertainty (level 1) and lower-level hypotheses nested within the higher level to represent other uncertainties (level 2), and are crossed with the level 3 hypotheses, which encompass parameters for which there is little or no information in the data. The three levels of hypotheses in the risk analysis for yellowfin tuna are: level 1 - the spatial structure (EPO, NE, NE\_short, SW), level 2 - scenarios constructed to represent the uncertainty in biological parameters (growth, natural mortality) and effort creep (1% increase per year in the catchability of the indices of abundance); and level 3 - the steepness of the stock-recruitment relationship. The level 2 hypotheses were implemented by changing one assumption at a time in the base reference model of each spatial structure. The low and high scenarios for growth and natural mortality were based on the uncertainty of the external estimates (values that have approximately half the likelihood as the maximum likelihood estimate). Three values of steepness of the Beverton-Holt stock recruitment relationship ( $h=1.0$ ,  $h=0.9$ ,  $h=0.8$ ) are considered for the third level hypothesis. The combination of the three levels of hypotheses results in  $4*6*3 = 72$  reference models.

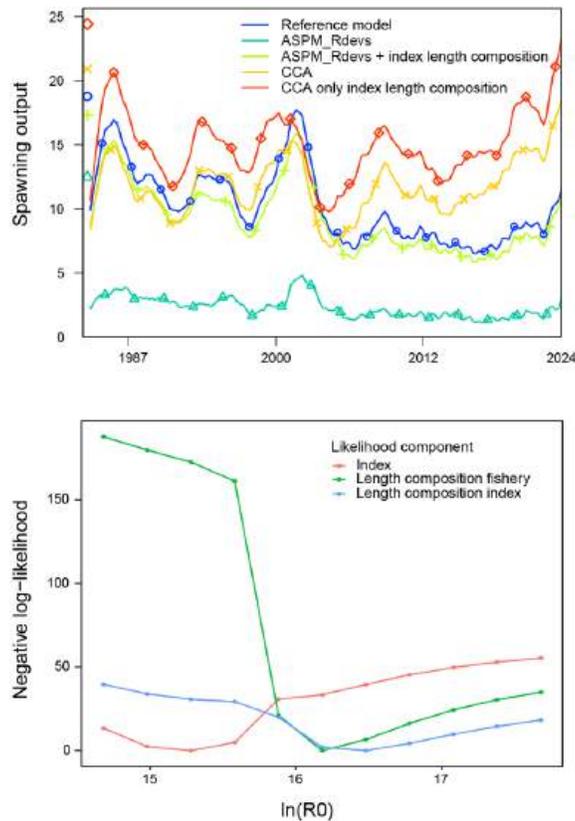
The models were fit by minimizing a penalized negative log-likelihood function (NLL). To ensure that the models obtained the global minima, a series of jitter analyses, which randomly change the initial parameter values to test convergence, were performed until the model passed the jitter test (the initial parameters estimates were ones that produced the lowest NLL among all jittered models). The fits were

also evaluated using residual analysis. Integrated model diagnostics were used to understand the information content of the data. The diagnostics used were the age-structure production model with estimated recruitment deviations (ASPM\_dev), ASPM\_dev also fit to the index length composition, catch curve analysis (CCA), CCA only fit to the index length composition, likelihood profile on the scale parameter ( $\log_{10} R_0$ ), and retrospective analysis.

One risk analysis was done for each of the four level 1 hypotheses by combining 18 reference models. Equal weight was used for all level 2 hypotheses. The weights for three values of steepness (level 3 hypotheses) were based on expert judgement from the risk analysis done for the last benchmark assessment:  $P(h=1.0) = 0.46$ ,  $P(h=0.9) = 0.32$ ,  $P(h=0.8) = 0.22$ .

### Assessment results

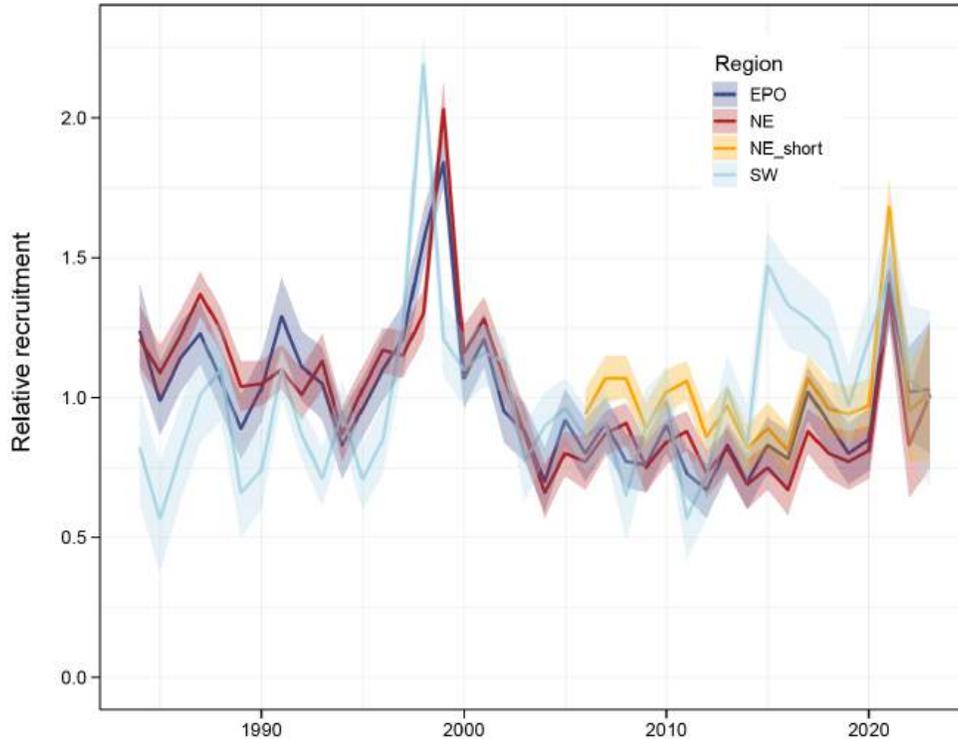
All 72 models converged and are used in the stock assessment and risk analysis. The integrated model diagnostics indicated that the indices and the catches alone are not enough to estimate the absolute scale of the models, and the length composition data provides the information on scale (Figure S-2). The indices provide information on relative trends. The length composition data for the NE and EPO models support higher absolute biomass levels in the second half of the time series (Figure S-2). For this reason, the NE\_short model that starts in 2006 was developed to represent possible changes in the dynamics of the stock or the fishery that are not understood or accurately modelled in the NE and EPO models. In addition, the tagging, otoliths and reproductive biology data come from recent years and may best represent the period in the NE\_short models.



**FIGURE S-2.** Integrated model diagnostics (Top panel: catch curve analysis (CCA) and age structured production model diagnostic (ASPM); Bottom panel:  $R_0$  likelihood component profile) for EPO base model with steepness 1 to illustrate the main information content of the data.

## Recruitment

The recruitment trends show patterns of similarities and differences among spatial structure hypotheses (Figure S-3). All models that compose the ensemble for each spatial structure hypothesis have similar trends in recruitment. Models for all four spatial structure assumptions estimate two peaks in recruitment, but for the SW model the largest peak occurs in 1998, while in the others it occurs in 1999. The second peak is in 2021. The SW models also estimated high recruitment in 2015-2017. The EPO and NE models estimate a regime shift in recruitment to a lower level after this peak, while the SW model does not.



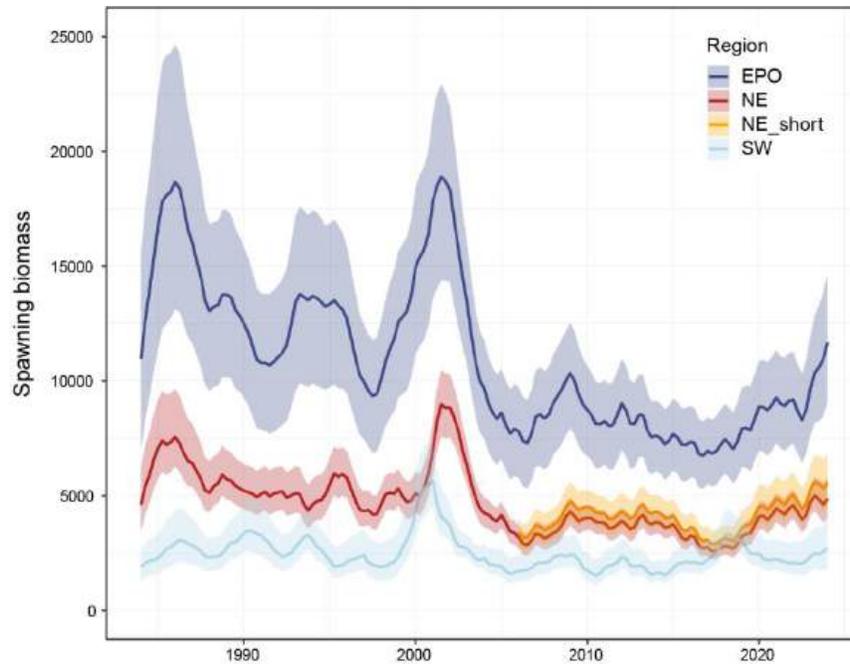
**FIGURE S-3.** Comparison of multi-model estimates of median relative annual recruitment and 80% confidence interval of yellowfin tuna for each hypothesis of spatial structure. The multi-model estimates include all level 2 and level 3 uncertainty scenarios.

## Biomass

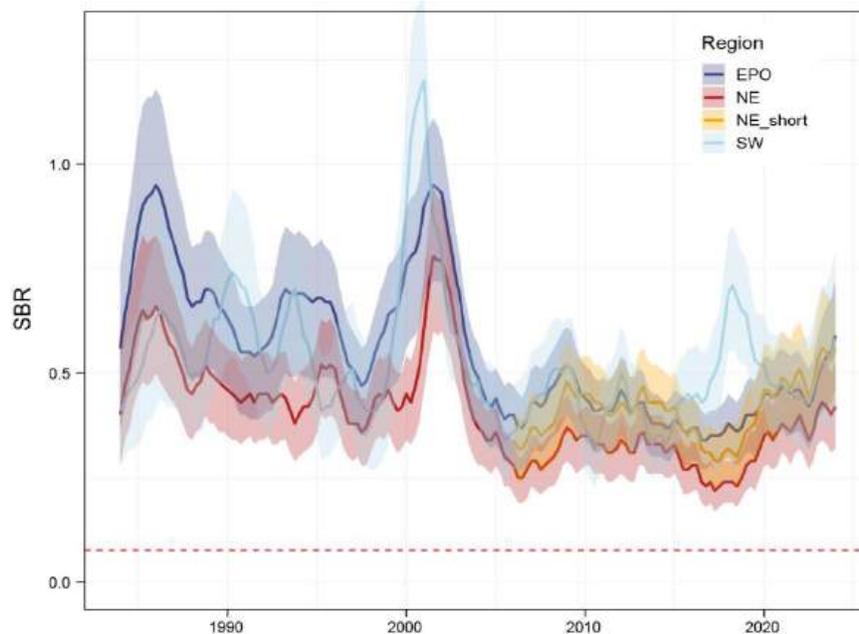
The spawning biomass in the NE is estimated to be about twice the level of that estimated for the SW. The estimate for the EPO is larger than the sum of the estimates for the two component stocks. The biomass trends (Figure S-4) generally follow the recruitment trends. Large spawning biomasses are a result of strong recruitment 2 or 3 years prior. The strong cohorts of 1998 and 1999 in the NE and SW regions show up as large spawning biomasses in 2001 and 2002 in the two regions, respectively. The trends in biomass since 2010 are diametric for the NE and SW regions.

The EPO-wide models, which use the areas-as-fleets approach to model spatial structure, estimates larger and more uncertain spawning biomass levels than the NE and SW combined, indicating that the EPO-wide models have difficulty fitting data with incompatible signals.

The NE and NE\_short models estimate very similar spawning biomasses.



**FIGURE S-4A.** Comparison of multi-model estimated spawning biomass of yellowfin tuna for each hypothesis of spatial structure with 80% confidence intervals.



**FIGURE S-4B.** Comparison of multi-model estimated spawning biomass ratio (spawning biomass over equilibrium virgin spawning biomass) of yellowfin tuna for each hypothesis of spatial structure with 80% confidence intervals. The red dashed line (at 0.077) indicates the SBR at the limit reference point  $S_{LIMIT}$ .

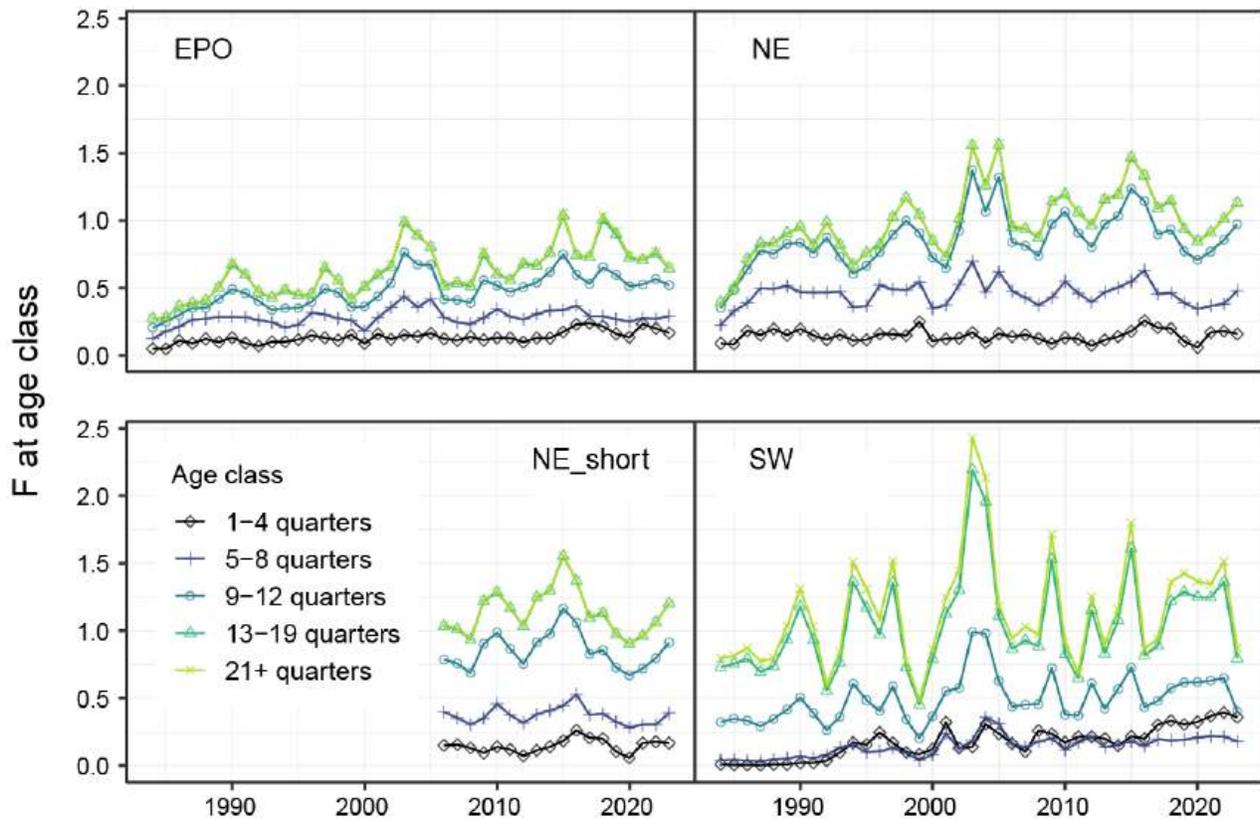
**Fishing mortality**

The relative distribution of fishing mortality at age is similar for the EPO, NE and NE\_short models: the fishing mortality is much higher for the older age classes. The magnitude of the fishing mortality, however, is lower for the EPO model, which is a consequence of its biomass being estimated higher than the sum of

the biomasses for the NE and SW regions. The relative distribution of fishing mortality at age of the SW region follows a different pattern. The fishing mortality on the intermediate aged yellowfin (9-12 quarters of age) is lower since the unassociated catches are lower and the purse-seine fishery associated with dolphins generally catches larger yellowfin. The fishing mortality on the youngest yellowfin (1-4 quarters of age) has steadily increased following the expansion of the FAD fishery in the mid 1990's. After 2015 the fishing mortality of this age group surpasses the 5-8 age class.

The trends in fishing mortality are similar between the NE and the NE\_short models, indicating that starting the model later does not change the perception of the effects of fishing in recent years. For those two hypotheses, there is a general increase in fishing mortality in all age classes after the year 2006, decline after 2015, with the lowest at the start of the covid19 pandemic, in 2020. After that, the fishing mortality increases, particularly for older yellowfin.

The increase in fishing mortality noticed in the last five years in the NE area is not shared by the EPO model. This may be due to the influence of the SW area, which has stable fishing mortality followed by a sharp decline in 2023. This indicates that using an EPO-wide model may underestimate and mask regional trends in fishing mortality.

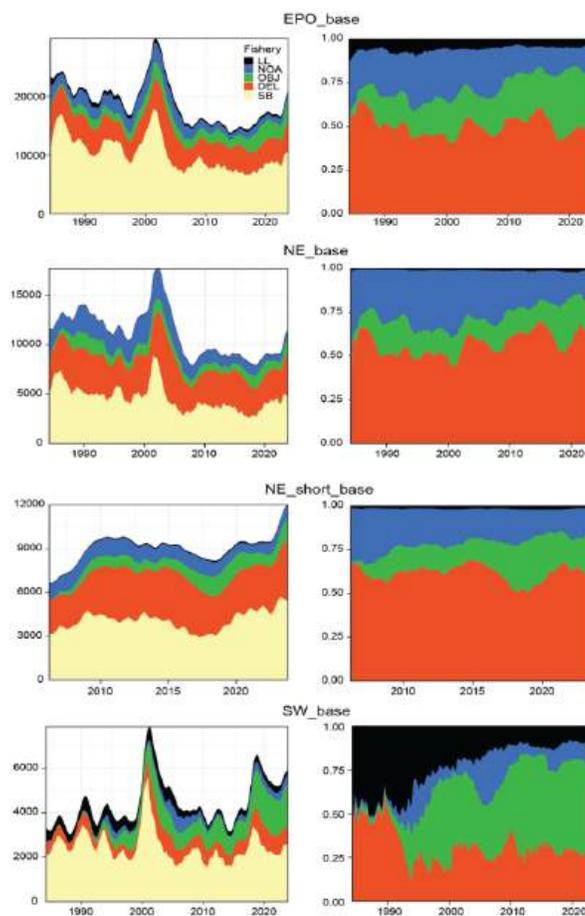


**FIGURE S-5.** Annual fishing mortality at age (sum of the four quarterly estimates within a year) of yellowfin by age group for each hypothesis of spatial structure (level 1). The values for each age group are weighted across level 2 and level 3 hypotheses.

## Fisheries impact

The EPO, NE and NE\_short models estimate similar impacts of the different types of fisheries (Figure S-6). The longline fisheries have the smallest impact, while the purse-seine fisheries associated with dolphins have the greatest impact during most of the modelled period. The unassociated fisheries had the second largest impact in the early years, but in the 1990s the impact of the floating-object fisheries started to increase and surpassed that of the unassociated fisheries around 2008.

For the SW models, the impact of the different purse seine set type has changed considerably over time. The longline fishery and the purse-seine associated with dolphins had the largest impact until mid-1990's, when there was an expansion of the floating object fishery, which steadily increased its impact and became the fishery with the largest impact in this region, larger than all other fisheries combined. The longline fishery has decreased both its effort and its impact on yellowfin in that area. The fishery associated with dolphins has slowly increased its absolute impact in this region, but in proportion it has stayed stable since the year 2000.



**FIGURE S-6.** Impact of the different fishing methods on the spawning biomass. Left panels: comparison of spawning biomass trajectory of a simulated population of yellowfin tuna that was never exploited (colored area) and that predicted by the stock assessment model (SB, yellow shaded area), and the impact of each fishing method (purse-seine on floating objects OBJ, also includes sorting discards and pole and line, purse-seine associated with dolphins DEL, purse-seine unassociated NOA and longline LL fisheries) for each stock structure hypothesis calculated from the base reference models with steepness of 1. Right panels: Proportional impacts.

## Stock status

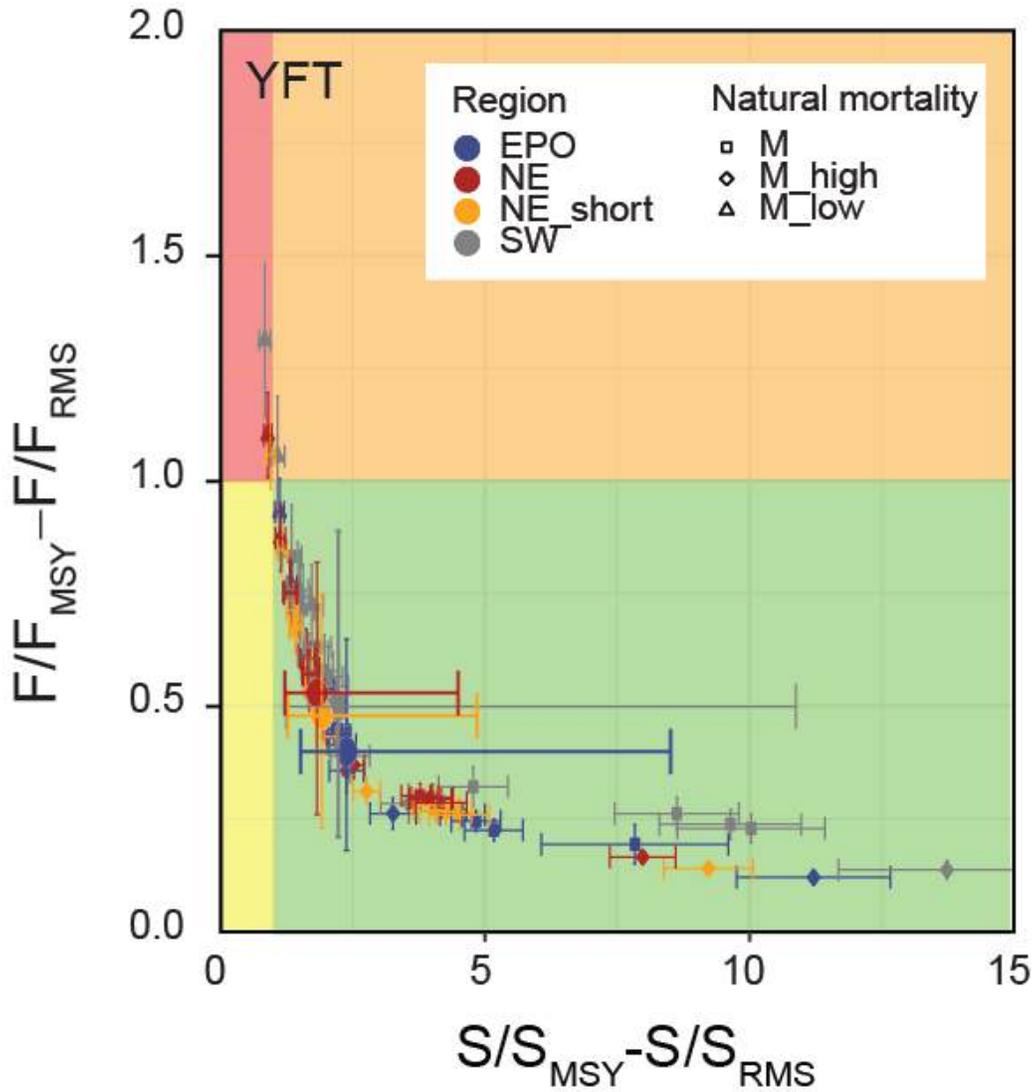
With respect to the IATTC interim target and limit reference points, all four spatial-structure hypotheses estimate the same stock status (Table S-1). The stock(s) is estimated to be well above the spawning biomass correspond to MSY ( $S_{MSY}$ ) and the staff proposed MSY proxy  $S_{30\%}$  (SAC-15-05) with low probability of being below these. The fishing mortality is estimated to be well below the level corresponding to MSY and the MSY proxy  $F_{30\%}$  with low probability of being above these. The assessment estimates zero probability that the spawning biomass or fishing mortality limit reference points have been breached. The EPO model is the most optimistic.

The most pessimistic models are those with low natural mortality (Figures S-7A, S-7B and S-8). Some of these models estimate that the spawning biomass is below the  $S_{30\%}$  level and the fishing mortality is above the  $F_{30\%}$  level. The high natural mortality levels are generally the most optimistic.

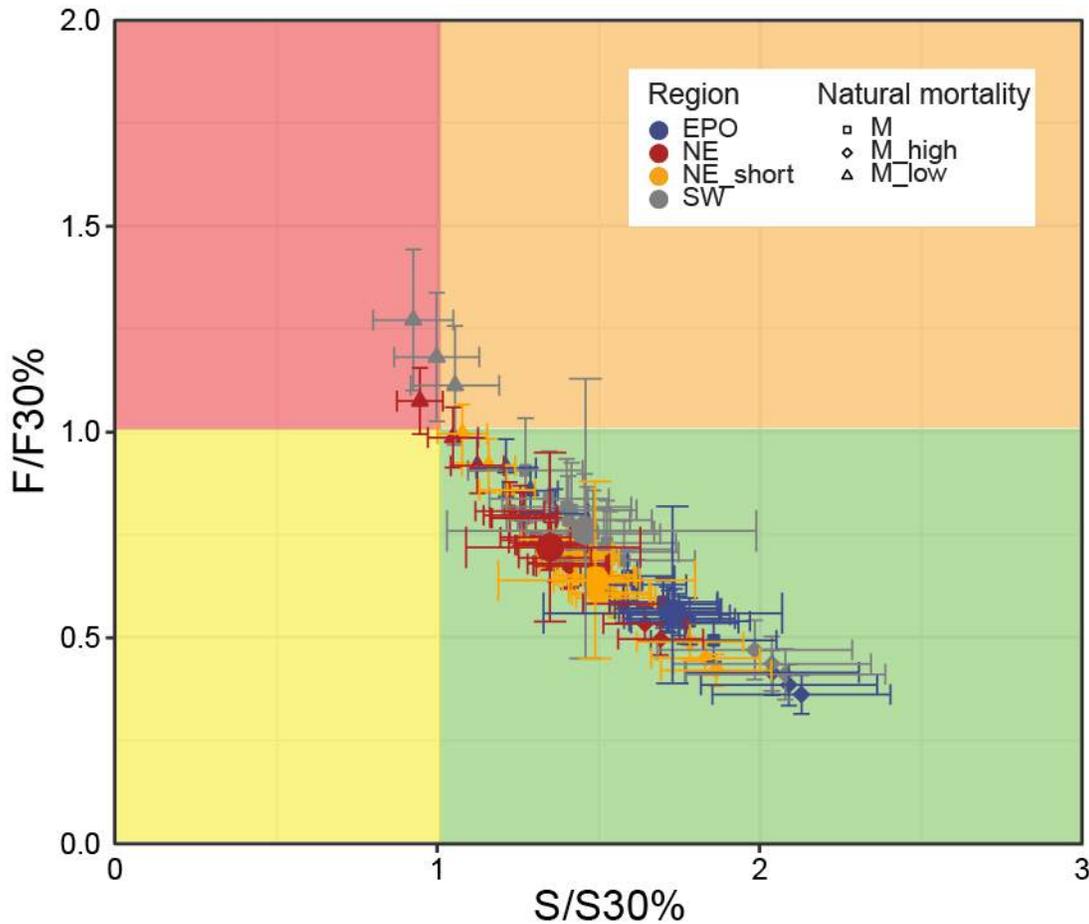
The estimates of the SBR (the ratio of the spawning biomass to the virgin spawning biomass) corresponding to MSY are low (generally below 20%, Table S-1) even though the highest fishing mortality is on older yellowfin. The value is higher with lower steepness of the stock-recruitment relationship and lower natural mortality. For example, the SW model with no relationship between stock size and recruitment (steepness equals 1) and high natural mortality has a value of 5% while the NE\_short model with steepness equal to 0.8 and low natural mortality has a value of 32%. The low level of SBR corresponding to MSY might be due to the assumptions about natural mortality declining with age (i.e., high M for juveniles).

**TABLE S-1.** Management quantities for yellowfin tuna in the EPO for each spatial structure hypothesis. The medians (or expected values \*) and probabilities were obtained from the joint probability distributions across models.

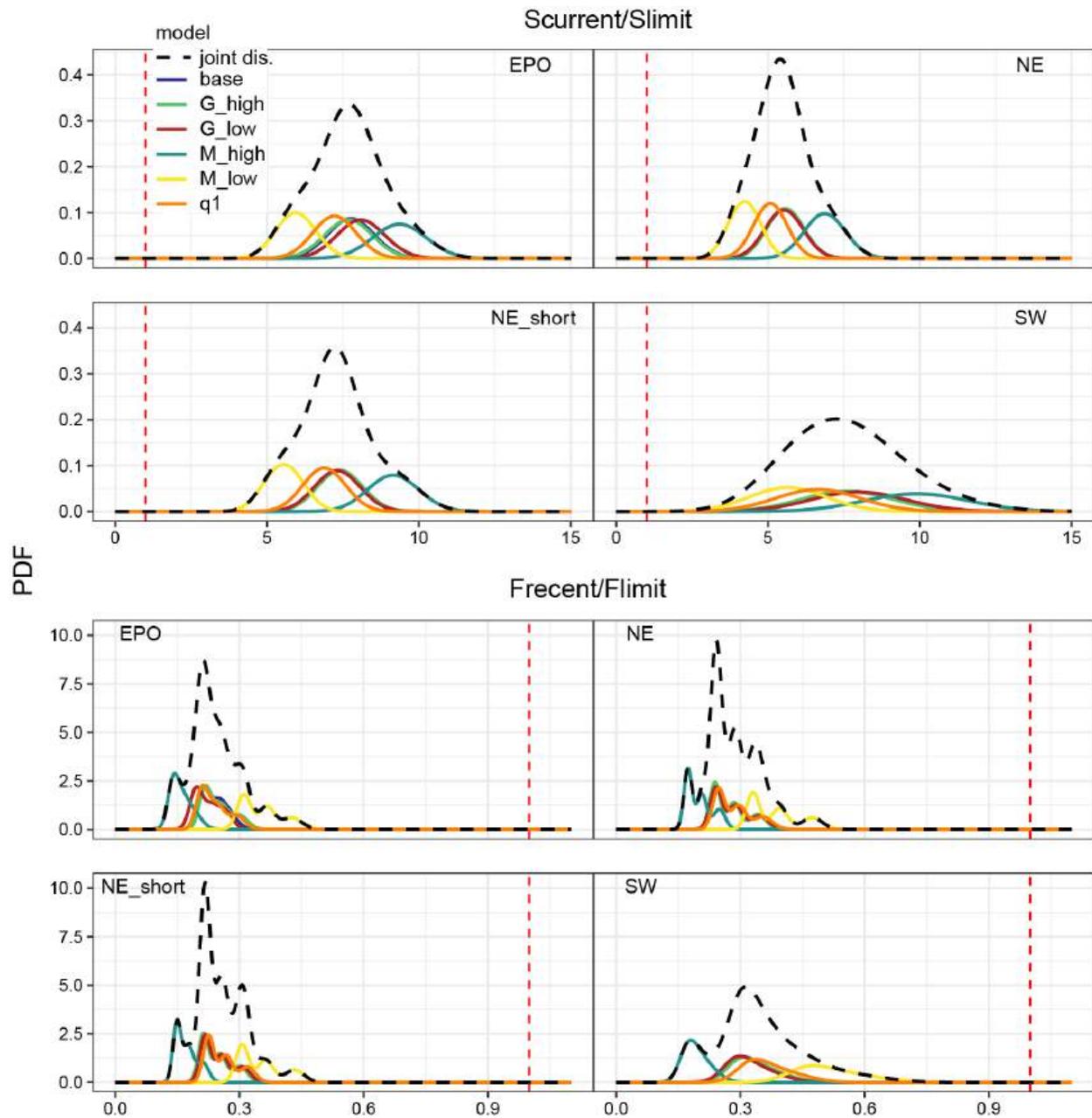
	EPO	NE	NE_short	SW
$S_{MSY}/S_0$ *	0.180	0.189	0.194	0.162
$S_{MSY\_d}/S_{0\_d}$ *	0.190	0.192	0.201	0.170
$F_{current}/F_{30\%S_{0\_d}}$	0.559	0.718	0.643	0.757
$p(F_{current} > F_{30\%S_{0\_d}})$	0.002	0.059	0.020	0.161
$F_{current}/F_{MSY}$	0.397	0.532	0.484	0.502
$p(F_{current} > F_{MSY})$	0.004	0.034	0.031	0.075
$F_{current}/F_{LIMIT}$	0.232	0.272	0.243	0.330
$p(F_{current} > F_{LIMIT})$	0.000	0.000	0.000	0.000
$S_{current}/30\%S_{0\_d}$	1.73	1.35	1.49	1.46
$p(S_{current} < 30\%S_{0\_d})$	0.0000588	0.044	0.004	0.081
$S_{current}/S_{MSY\_d}$	2.38	1.82	1.91	2.22
$p(S_{current} < S_{MSY\_d})$	0.000	0.000	0.000	0.000
$S_{current}/S_{LIMIT}$	7.67	5.43	7.23	7.48
$p(S_{current} < S_{LIMIT})$	0.000	0.000	0.000	0.000



**FIGURE S-7A.** Kobe plot of the most recent estimates of spawning biomass ( $S$ ) and fishing mortality ( $F$ ) relative to their target reference points ( $S_{MSY_d}$  and  $F_{MSY}$ ) for each hypothesis of spatial structure. Each dot is based on the average  $F$  over the most recent three years, 2021-2023, and the  $S$  for the first quarter of 2024 and the error bars represent the 80% confidence interval of model estimates. The larger dots represent the combined result for each spatial structure hypothesis.



**FIGURE S-7B.** Kobe plot of the most recent estimates of spawning biomass ( $S$ ) and fishing mortality ( $F$ ) relative to their proxy target reference points ( $30\%S_d$  and  $F_{30\%S_d}$ ) for each hypothesis of spatial structure. Each dot is based on the average  $F$  over the most recent three years, 2021-2023, and the  $S$  for the first quarter of 2024 and the error bars represent the 80% confidence interval of model estimates. The larger dots represent the combined result for each spatial structure hypothesis.



**FIGURE S-8.** The joint probability distributions for spawning biomass ( $S$ ) in the first quarter of 2024 and average fishing mortality ( $F$ ) in 2021-2023 relative to their limit reference points ( $S_{Limit}$  and  $F_{Limit}$ ). The distributions are provided for each of the four spatial structure hypotheses separated into different components (level 2 hypotheses). The level 3 hypotheses (steepness values) were integrated out.

## 1. INTRODUCTION

This report presents the results of a benchmark stock assessment<sup>1</sup> of yellowfin tuna (*Thunnus albacares*) in the eastern Pacific Ocean (EPO), conducted using Stock Synthesis (version V3.30.23.1), an integrated statistical age-structured stock assessment modeling platform. As the previous benchmark assessment, this assessment forms the foundation of a risk analysis, which explicitly takes uncertainty into account when determining stock status and formulating management advice. Extensive research was done to address the main uncertainties in the assessment, which relate to spatial structure, biological parameters and indices of abundance. The model development phase included the implementation of spatial models with movement. However, these models estimated movement to be very low. The final set of reference models addressed spatial structure using both areas-as-fisheries and independent stocks. The reference models incorporate the most recent results of biological research including a new growth curve and new natural mortality estimates which are both derived from the tagging data collected under the Regional Tuna Tagging Program in the EPO (RTTP-EPO 2019-2020, Project E.4.a) and previous studies conducted since the year 2000. All model input files and output results for this benchmark assessment are available in html and pdf formats.

### 1.1. BACKGROUND

The previous 2020 benchmark assessment ([SAC-11-07](#)) and external reviews ([YFT-02-Rep](#), [RVDTT-01](#), [RVMTT-01](#)) highlighted uncertainty about spatial structure of yellowfin in the EPO. The data suggests that there are either two or more stocks in the EPO or there is spatial structure in the population. A large recruitment in the 1990s enters the core dolphin associated purse seine fishery north of the equator in a different year than the large recruitment that enters the longline fishery south of the equator. Length composition data from the purse seine fishery associated with dolphins has smaller fish in the northeast and larger fish in the south and west, with intermediate sized yellowfin in the core area.

The 2020 benchmark assessment considered a set of overarching hypotheses concerning stock structure. Due to the practical need for an assessment of the whole EPO and the absence a satisfactory method to inform the spatial structure, the assessment model was focused on the data for the “core” dolphin associated fishery area but included catch for the whole EPO.

Since the 2020 benchmark assessment, substantial research was done to further address the spatial structure of the yellowfin tuna in the EPO ([SAC-14-06](#), [SAC-15-03](#)). In 2024, the staff presented exploratory assessment models that further highlighted the spatial structure ([SAC-15-03](#)). While the exploratory models had new fisheries definitions defined spatially, all models were EPO-wide, with spatial differences in length frequencies modeled through differences in selectivity (areas-as-fisheries approach). The length composition associated with the EPO-wide index of abundance, however, had a multimodal distribution because it contained data from regions with distinct average lengths. Fish from the region were most of the catches from the purse-seine fishery associated with dolphins occur were of intermediate size. Fish caught in the same fishery but north of 20°N were smaller. Fish from the western and southern areas were larger. The stock assessment models were unable to reconcile the joint length composition distribution for the index of abundance with a regular (i.e., double normal) selectivity curve indicating that either the standardization of the length composition should be improved, or spatial/stock structure was confounding the patterns. Because of the unresolved spatial patterns, two types of exploratory assessments were put forward. A core area model, restricted to the region of operation of the main dolphin associated purse seine fisheries, which comprises most of the catch of yellowfin in the EPO, and an EPO-wide model, which

---

<sup>1</sup> “Benchmark” stock assessments are a full analysis of model assumptions, methodologies and/or data sources, whereas in an “update” assessment only the data used in the assessment are updated.

simply added the catches of the other regions to the core area model, while still fitting only to the core area data. This second approach is similar to the one used in the 2020 benchmark assessment.

One of the main limitations of the exploratory models was that spatial definitions of fisheries were constrained to be rectangles defined along latitude and longitude lines. The approach used to define the areas was a regression tree method with latitude, longitude, quarter and cyclic quarter as explanatory variables to define splits. The resulting length compositions still showed some multimodality and were not able to be represented by regular selectivities (i.e., double-normal selectivity). Some of the issues in defining fisheries with the regression tree method may be that the spatial structure could have diagonal boundaries or be irregular in shape as can be expected from physical structure or environmental drivers. In this assessment, irregular areas with homogeneous length composition were delineated using a newly developed flexible methodology based on cluster analysis. These areas were used to define fisheries and spatial domains for several models.

The length frequencies of the catches are the result of contact selectivity, availability (e.g., inhabit a different depth than the gear) and density. It is expected that contact selectivity is constant in space, but availability may vary in space, due to several factors. Likewise, density may vary in space and the spatial variation in availability and density may be difficult to disentangle. Both may be affected by environmental conditions. Density may also be affected by local productivity, movement and stock structure. To account for the potential causes of the patterns in length frequencies, several population dynamics models were implemented in this assessment, based on hypotheses related to spatial and stock structure.

As in previous benchmark assessments for yellowfin tuna in the EPO, and similarly to most tropical tuna stock assessments in the world, the indices of abundance are derived from fisheries catch and effort (CPUE) data. In this benchmark assessment, two important advances were made regarding the development of indices of abundance. First, a new spatiotemporal model was implemented that allows for more flexibility when standardizing the CPUE and length composition data of the purse-seine fishery associated with dolphins. Second, for the first time, a multi-fleet longline index was obtained based on the standardization of the operational-level set-by-set data, as result of a collaboration with CPCs that have distant-water longline fleets. Both the purse-seine and the longline indices are used following different to the spatial structure hypotheses.

The environment is a key forcing function in the stock of yellowfin tuna in the EPO. Recruitment exhibits large variability and long-term trends in high and low recruitment periods. Seasonality or longer-term fluctuations in oceanographic conditions may account for seasonal changes in fish and fleet spatial distribution and large-scale oceanographic events, such strong El Niño's, may cause structural changes in the ecosystem. In this assessment, these influences are taken into account in several ways: (1) as in previous assessments, four recruitments are estimated for each year, allowing for within and between year variability, (2) the indices of abundance based on purse-seine data include changes in the spatial domain related to the Oceanic Niño Index (ONI) index<sup>2</sup>, accounting for potential changes in spatial distribution due to oceanographic conditions (3) one group of models takes into account the apparent structural changes in the ecosystem that happen after the 1998-1999 El Niño-La Niña.

## **1.2. CONCEPTUAL MODEL**

### **1.2.1. STOCK AND SPATIAL STRUCTURE**

Spatial structure can be caused by several factors, including stock structure, and these may not be directly related to latitude and longitude ([SAC-14-06](#)). This may result in spatial structures that have diagonal

---

<sup>2</sup> [https://origin.cpc.ncep.noaa.gov/products/analysis\\_monitoring/ensostuff/ONI\\_v5.php](https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php)

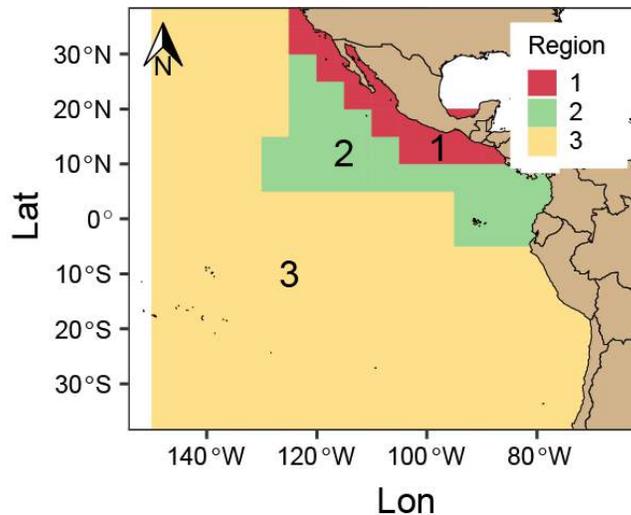
borders or have other irregular shapes. Ideally stock and spatial structure should be investigated using a multidisciplinary approach that considers data from a wide range of sources, in particular data from a well-designed and extensive tagging study. The available tagging data for yellowfin tuna, although still insufficient to allow for delineation of spatial structure, indicates that movement of yellowfin is limited (Schaefer and Fuller, 2022a) and that isolation by distance may be occurring even within the area of distribution of dolphin associated fisheries, which was previously thought could be modelled as one group (SAC-14-06). This suggests that there may be stock structure relevant for management and local depletion may occur.

Spatial structure has been considered in the assessments of yellowfin tuna in the EPO by applying the areas-as-fleets approach (Waterhouse et al., 2014), in which “fisheries” are defined not only by gear type and set type, but also by their geographical area of operation. This approach allows for spatial differences in length composition to be considered without explicitly constructing a spatial model.

The regression tree approach previously used by the staff to define fisheries was limited to defining rectangular areas. For this benchmark assessment, the staff applied a new hierarchical clustering method (SAC-16-INF-F) to analyze the length frequency of yellowfin tuna and define irregular-shaped fisheries spatially. The method is based on a new clustering algorithm to aggregate distributions such as length composition (Minami and Lennert-Cody, 2024). This approach represents a considerable advance in relation to the previously used tree analysis method that defines areas only along latitude and longitude lines, which were not enough to capture the spatial structure shown in the length frequency of yellowfin tuna in the EPO.

For the yellowfin tuna application, the cluster analyses were used for two purposes: (i) defining regions that could potentially represent spatial or stock structure and (ii) defining fisheries by geographical area of operation. In addition to the cluster analysis, tagging data and a spatiotemporal model of catch per set of small fish in the purse-seine floating object fishery were used to investigate areas that could have connectivity of juveniles (<60 cm). Details of the methods and results can be found in SAC-16-INF-F.

Four areas were delimited based on the cluster analysis of the length frequency of catches from purse-seine sets associated with dolphins. Area 1 (Northern Coastal) has smaller fish, area 2 (core) has a wide range of intermediate sizes, area 3 (offshore) has the largest fish, and area 4 (Galapagos) has larger fish than the core area, but smaller than area 3. The catch per set and tagging data suggest that fish <60 cm in the Galapagos area may be more likely associated with the core area than areas to the south or west (i.e. the high-density patches of small fish in the Galapagos area tended to continue northeast towards the core area). The final spatial structure assumptions delimited 3 regions, and areas within those regions to define fisheries (Figure 1).



**FIGURE 1.** Regional divisions considered in the stock assessment of yellowfin tuna in the EPO. The Galapagos area is added to region 2.

Given the uncertainty in stock structure, several hypotheses were examined for consideration in this assessment:

**Hypothesis 1 - full mixing:** one mixed stock for the whole EPO. The differences in length composition are due to selectivity or availability. Clearly, the spatial differences in the dolphin associated purse seine composition data complicate the implementation of an index from this fishery. Conceptually, the index would have different selectivity for each area. This hypothesis is not supported by the tagging data (Schaefer and Fuller, 2022a), or by the reproductive biology data (Schaefer and Fuller, 2022b) and it was not considered. Separate stocks or ontogenetic movement are considered more likely explanations for the composition data from the purse seine fishery associated with dolphins.

**Hypothesis 2 - regional dynamics:** one stock with spatial dynamics among regions. The differences among regions may be due to spatial structure, movement, catchability/selectivity/availability and/or different exploitation rates. This hypothesis was implemented in two ways. First, EPO-wide areas-as-fleets model, where catches taken from the three regions and areas within were treated as separate fisheries and the differences in size composition were modeled with different selectivity curves (Table 1). Second, a spatial model with three regions. Hypothesis H2 was implemented as a three-region spatial model (Table 1), considering movement between adjacent regions. Further subdivisions in areas were made within each region based on the cluster analysis results, which were treated as different fisheries. The movement rates were estimated to be very low between regions, perhaps due to the limitations on how movement is parametrized in the SS3 platform. This indicated that independent regional models would be more appropriate to use.

**Hypothesis 3 – Independent stocks:** it is unlikely that the fish of different regions would be completely independent but given the estimation in the spatial model of movement rates to be very low, independent models for different regions are justified. In addition, modelling the regions independently will allow for understanding of the dynamics in different regions without influence from data outside the region. Two separate assessments were done. One for a northeast region (NE) that combines region 1 and region 2 in Figure 1 and is where most of the yellowfin catches are taken and another for region 3, which encompasses the south and west of the EPO (SW). Region 1 and region 2 were combined in one assessment for

simplicity, as the indices of abundance for region 1 showed a similar general pattern as the region 2 indices. An exploration model was done for region one that showed strong patterns in recruitment and that the growth assumptions, which are based on fish from the core area, are not consistent with the fish caught off Baja California, in region 1.

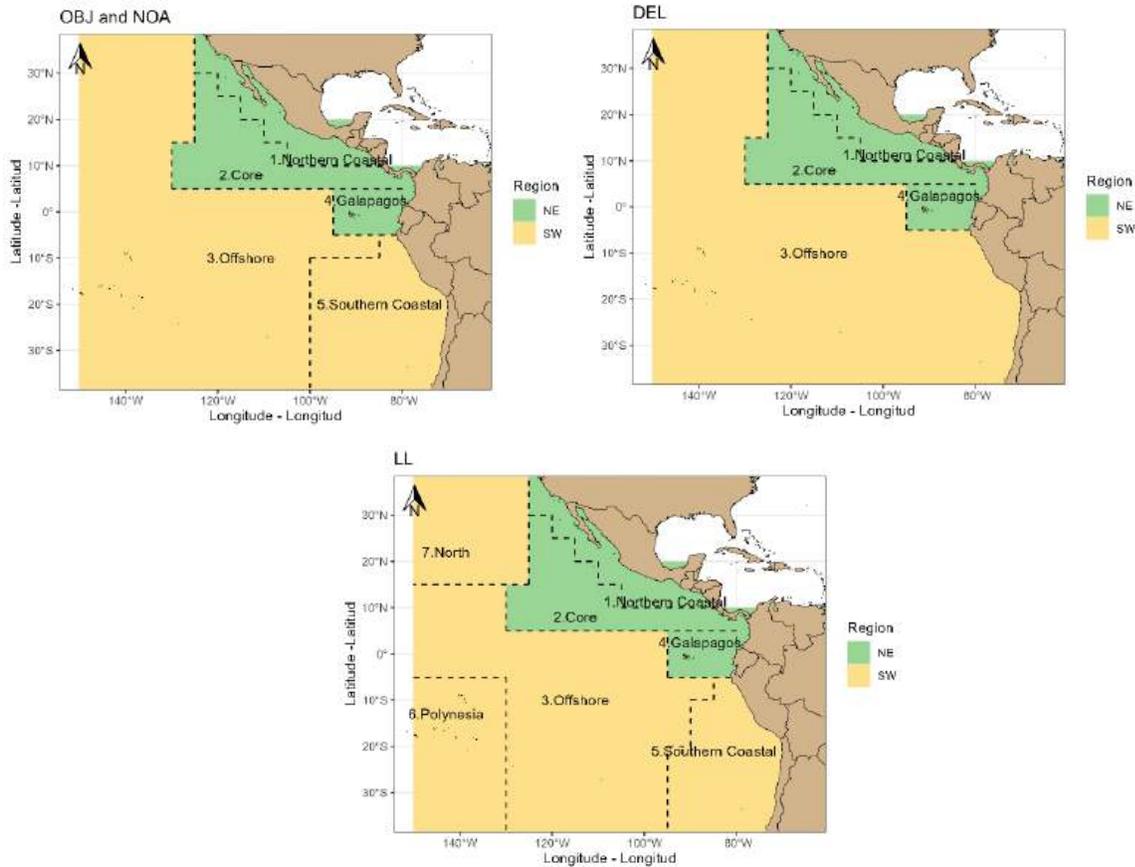
**TABLE 1.** Hypotheses considered to address the spatial structure of yellowfin tuna in the EPO with indication of assumptions about regions within spatial models and spatial domain on indices of abundance. The allocation of geographic areas definition of fisheries by gear type is indicated: NC- Northern coastal, CO- Core, O- Offshore, GP – Galapagos, SC – Southern Coastal, N – North, P – Polynesia. The shaded cells indicate the areas included in each hypothesis. The regions and areas are shown in Figures 1 and 2. The spatial version of H2 was not used in the risk analysis.

Hypothesis			H2 Regional dynamics	H2 Regional dynamics	H3 Independent regions	H3 Independent regions
Type of model			areas-as- fleets	spatial	areas-as- fleets	areas-as- fleets
Spatial domain			EPO	EPO	NE	SW
Indices of abundance			EPO index	3 indices, one for each region	NE index	LL index
Gear	Area	#	risk analysis	exploratory	risk analysis	risk analysis
Purse-seine	NC	1	1	1	1	
	CO	2	1	2	1	
	O	3	1	3		1
	GP	4	1	2	1	
	SC	5	1	3		1
Longline areas	NC	1	1	1	1	
	CO	2	1	2	1	
	O	3	1	3		1
	GP	4	1	2	1	
	SC	5	1	3		1
	N	6	1	3		1
	P	7	1	1	3	

### 1.2.2. FISHERY DEFINITIONS

Fisheries were defined in the models based on gear type (purse seine, longline, pole-and-line), purse seine set type (floating object, unassociated, dolphin associated), and geographic area of operation to represent the different sizes of yellowfin caught (Table 2). Two types of longline fisheries were also defined, one for catches reported in numbers and the other for catches reported in weight (the conversion from weight to numbers is done inside the stock assessment model to ensure consistency). Fisheries representing discarded small fish (sorting discards) were also defined. Purse-seine in unassociated sets were split by size category, one fishery for small and one for large fish because the length composition showed a

bimodal pattern, which is difficult to model adequately in a single fishery. The catches were split using the proportion of catch in each size category recorded by the observers on board. The length composition for each size category was obtained from port sampling data from wells that contained only one set. The resulting length compositions had unimodal distributions. Details of the splitting procedure are in SAC-16-INF-F. The classification of fisheries by geographic areas of operation was conducted using the cluster analysis on length composition data weighted by the catches (SAC-INF-F). The areas defined for purse-seine and longline are shown in Figure 2. Finally, “surveys” are defined in the stock synthesis platform as fisheries that do not have catches associated with them and are used to model the indices of abundance and corresponding length composition (see section 2.3 on indices of abundance).



**FIGURE 2.** Definitions of regions and areas within regions for the models implemented in the EPO yellowfin tuna stock assessment and risk analysis. Areas used to define fisheries spatially for the purse-seine with sets on floating objects (OBJ), unassociated (NOA), and associated with dolphins (DEL), and longline (LL) area based on cluster analysis of length composition. The pole-and-line fishery is assumed to take place in area 1. The EPO models include both NE and SW regions. Independent models for NE and SW were also implemented.

**TABLE 2.** Fleets defined for the 2025 benchmark stock assessment of yellowfin tuna in the EPO. Gear: PS: purse seine; LP: pole and line; LL: longline; PS set type: OBJ: floating object; NOA: unassociated; DEL: dolphin associated; see Figure 2 for area definitions.

Fleet number	Fleet type	Fleet name	Gear	Set type	Area-size class	Catch data	unit
1	Fishery	F1_PS_OBJ_North_coastal	PS	OBJ	1 - North coastal	Retained catch + discards (inefficiencies)	t
2		F2_PS_OBJ_Core	PS		2 - Core		t
3		F3_PS_OBJ_Offshore	PS		3 - Offshore		t
4		F4_PS_OBJ_Galapagos	PS		4- Galapagos		t
5		F5_PS_OBJ_South_coastal	PS		5- South coastal		t
6		F6_PS_NOA_North_coastal	PS	NOA	1 - North coastal		t
7		F7_PS_NOA_Core	PS		2 - Core		t
8		F8_PS_NOA_Offshore_small	PS		3 – Offshore - small		t
9		F9_PS_NOA_Offshore_large	PS		3 – Offshore - large		t
10		F10_PS_NOA_Galapagos_small	PS		4- Galapagos - small		t
11		F11_PS_NOA_Galapagos_large	PS		4- Galapagos - large		t
12		F12_PS_NOA_South_coastal_small	PS		5- South coastal - small		t
13		F13_PS_NOA_South_coastal_large	PS		5- South coastal - larger		t
14		F14_PS_DEL_North_coastal	PS	DEL	1 - North coastal		t
15	F15_PS_DEL_Core	PS	2 - Core		t		
16	F16_PS_DEL_Offshore_South	PS	3 - Offshore + South coastal		t		
17	F17_PS_DEL_Galapagos	PS	4- Galapagos		t		
18	F18_PS_DIS_small_North_coastal	PS	OBJ - discards	1 - North coastal	Sorting discards	t	
19	F19_PS_DIS_small_Core	PS		2 - Core		t	
20	F20_PS_DIS_small_Offshore	PS		3 - Offshore		t	
21	F21_PS_DIS_small_Galapagos	PS		4- Galapagos		t	
22	F22_PS_DIS_small_South_coastal	PS		5- South coastal		t	
23	F23_LP	LP	-	1 - North coastal	Retained catch	t	
24	F24_LL_North_coastal_n	LL		1 - North coastal		1,000	
25	F25_LL_Core_n	LL		2 - Core		1,000	
26	F26_LL_Offshore_n	LL		3 - Offshore		1,000	
27	F27_LL_Galapagos_n	LL		4 - Galapagos		1,000	
28	F28_LL_South_coastal_n	LL		5- South coastal		1,000	
29	F29_LL_North_n	LL		6 - North		1,000	
30	F30_LL_Polynesia_n	LL		7 - Polynesia		1,000	
31	F31_LL_North_coastal_w	LL		1 - North coastal		t	
32	F32_LL_Core_w	LL		2 - Core		t	
33	F33_LL_Offshore_w	LL		3 - Offshore		t	
34	F34_LL_Galapagos_w	LL		4 - Galapagos		t	
35	F35_LL_South_coastal_w	LL		5- South coastal		t	
36	F36_LL_North_w	LL		6 - North		t	
37	F37_LL_Polynesia_w	LL	7 - Polynesia	t			
Number		Fleet name	Gear	Set type	Area	Size	
38	Survey	S1_EPO	PS	DEL	1 - North coastal	Standardized length compositions	t
39		S2_PS_NCoastal	PS		2 - Core		t
40		S3_PS_Core	PS		3 - Offshore		t
41		S4_PS_Offshore	PS		4 - Galapagos		t
42		S5_LL_Offshore	LL	-	5- South coastal		1,000
43	S6_PS_Echo_east	PS	OBJ	2 - Core	As F2	t	
44	S7_PS_Echo_west	PS	Echosounder	3 - Offshore	As F3	t	

## 2. DATA

### 2.1. CATCH

#### 2.1.1. PURSE-SEINE

Total EPO purse seine catches by species were estimated by catch strata and then aggregated across geographical area (fishery definitions) by quarter. The catch strata are defined as the combination of area, month, set type, and vessel fish-carrying capacity. The method used to estimate the species composition of the catch depends on the sources of information available. Estimates prior to 2000 are based on the recorded species totals in the unloading, observer or logbook data, as applicable. To correct for underestimated bigeye catches, a factor that adjusts the catches of all three tropical tuna species, based on the port-sampling data from 2000-2004, is applied. The adjusted species totals are prorated to catch strata using ancillary information from the observer and logbook databases. Since 2000, the port-sampling data have been used to determine the species composition of the total catch. The total catch of all three species combined (from unloading, observer and logbook data) is prorated to catch strata, using the information in the observer and logbook databases. The port-sampling data on the species and size composition of the catch are then used to estimate the catch of each species by catch stratum. Detailed explanations of the sampling and estimators can be found in the appendix of Suter (2010) and in [WSBET-02-06](#). This catch estimation methodology, which is a design-based approach, is used to obtain the fleet-level Best Scientific Estimates (BSEs) of species composition of the catches for each purse-seine fishery. The methodology is integrated into the R package *BSE* (<https://github.com/HaikunXu/BSE>).

Bias-adjustment was made for the estimated OBJ catches derived from the BSE algorithm for the two years affected by the COVID-19 pandemic (2020 and 2021). The pandemic disrupted the collection of species and size composition data by IATTC port-samplers, leading to a systematic loss of port-sampling data from ports where much of the EPO bigeye catch is unloaded ([SAC-13 INF-L](#)). Because the BSE algorithm relies on the estimates of species composition of purse-seine catches derived from port-sampling data, it is likely that the purse-seine catches by species for the two COVID-19 years are biased (Majumdar et al., 2023). The yellowfin catches obtained from the BSE algorithm in the OBJ fishery were likely underestimated in 2000 and overestimated in 2021. Correction factors following table 3 in [SAC-14 INF-D](#) are applied in this benchmark assessment. The BSE quarterly OBJ yellowfin tuna catch for 2020 was increased by 22% and for 2021 was decreased by 9%, respectively.

#### 2.1.2. LONGLINE

The IATTC staff does not collect data on longline catches directly. Instead, catches are reported annually to the IATTC by individual Members and Cooperating Non-Members (CPCs), according to Resolution C-03-05 on data provision. Catches are reported by species, but the availability and format of the data vary among fleets: the main longline fleets report catch, and effort data aggregated by 5° latitude and 5° longitude by month. IATTC databases include data on the spatial and temporal distributions of longline catches in the EPO by the fleets of distant-water CPCs (China, Chinese Taipei, French Polynesia, Japan, Korea, and Vanuatu) and coastal CPCs (mainly Mexico and the United States).

For this assessment, longline catch data are aggregated by fishery defined on the area of operation (Figure 2). Because catches may be reported in numbers or in weight, two longline fishery fleets are defined for each area, so that the catches can be included in the assessment model in their original units (1,000s of fish and metric tons, the), and the conversion between numbers and weight is done internally by the stock assessment. Updated and new catch data for the longline fishery fleets are incorporated into the current assessment. If catch data for a recent year or years were unavailable, catches were set equal to the last year for which data were available. For fleets that reported catch aggregated by year and 5° by 5° cell, the

data were split into quarters, using the proportion of catches by quarter and area for the closest year for which data were available. The catches of coastal CPCs that reported aggregated catches were added to the area that covers the CPC's Exclusive Economic Zone (EEZ). The algorithm to calculate the catch by longline fishery fleet is described in WSBET-02-03, and the associated R code is available at [https://github.com/HaikunXu/IATTCassessment/blob/master/R/ll\\_catch.R](https://github.com/HaikunXu/IATTCassessment/blob/master/R/ll_catch.R).

### **2.1.3. DISCARDS**

Two types of discards are considered in this benchmark assessment: those resulting from inefficiencies in the fishing process and those related to catch sorting. Examples of inefficiency are catches from a set exceeding the remaining storage capacity of the fishing vessel or dumping unwanted bycatch species, and catch sorting is assumed to occur when fishers discard tuna that are under a certain size.

For the purse-seine fishery, the amount of yellowfin discarded, regardless of the reason, is estimated with information collected under the on-board observer program of the Agreement on the International Dolphin Conservation Program (AIDCP), using the methods in Maunder and Watters (2003). No observer data is available to estimate discards before 1993, and it is assumed that there were no discards before that time. Also, there are periods for which observer data are not sufficient to estimate the discards, in which case it is assumed that the discard rate (discards/retained catches) is equal to the discard rate for the same quarter in the previous year or, if quarterly data are not available, a proximate year. Total catch by purse-seine fisheries (Fleets 1 to 17) represents retained catch plus discards resulting from inefficiencies in the fishing process. Sorting discards are rare in NOA and DEL fisheries and are also added to the total catch. Sorting in OBJ fisheries was a problem in the early 2000's, and decreased after regulations were put in place. Five discards' fisheries are also defined (Fleets 18-22) following the rationale of Watters and Maunder (2001), and correspond to the same areas of the OBJ fisheries. The OBJ sorting discards are assumed to be composed of 1-3 quarters old yellowfin.

Discards by the longline fisheries are not available so the retained catch is assumed to represent the total catch.

## **2.2. SIZE COMPOSITION**

### **2.2.1. PURSE-SEINE**

The size composition of the catch, in numbers of fish by 1-cm length interval, is estimated by stratum and then aggregated across strata to obtain quarterly estimates for each fishery. The estimated number of fish is then converted to proportion of fish at length for the assessment. The estimated numbers at length for each stratum are obtained by multiplying the well-level estimates of the proportion at length, combined across sampled wells, by the estimated total catch in numbers for the species in the stratum. Since 2000, the well-level estimates of proportions at length make use of both the species counts and the length-measurement data. Details of the estimators can be found in [WSBET-02-06](#).

For some purse-seine unassociated fisheries that showed bimodal length distributions, the well-level data was used to represent the sizes of small and large fish and the catches split into two fisheries, one for each size class. Details of this procedure are described in SAC-16-INF-F.

### **2.2.2. LONGLINE**

In the 2020 benchmark assessment, although length composition data was available for the longline fleets, it was not used to fit the final models that entered the risk assessment because the assessment emphasized the core area of the yellowfin catches, which are taken by the purse-seine fleet associated with dolphins (DEL). The dolphin associated fishery has little overlap with the longline fleets. Because the current benchmark assessment also considers spatial structure, the longline length composition data is

important in some model scenarios, such as the spatial model and the SW models.

The length composition data for longline fishery fleets comes from length composition data from Japanese commercial longline vessels measured by fishers. The Japanese fleet was for many decades the dominant distant water longline fleet fishing in the EPO ([OTH-30-RPT](#)). In recent years there has been a contraction of the area of operation of the Japanese fleet, as well as a marked decrease in effort, catch, and proportion of the total longline catch. The contribution of Japanese longline catch to the total yellowfin longline catch has continuously decreased over time and since 2017 has been less than 25% of what it was in 1985. The Chinese fleet has expanded in the same period. Concerns have been raised about how representative the Japanese length composition data are of the other longline fleets. The composition data for each fishery, defined spatially, should represent all the longline catch for that fishery. Therefore, the data for other CPCs should also be considered. For recent years, however, the data for all CPCs has low coverage, as it comes exclusively from observers (less than 5% coverage). Data from observers differs from data coming from fishers in several aspects. Thus, to represent the fisheries only data from the Japanese fleet measured by fishers was to represent the fisheries in the final models.

The Japanese longline length composition data for yellowfin tuna in the EPO covers the period between 1986 and 2023. All length compositions before 2011 and after 2015 were collected by fishers and on-board observers, respectively. Between 2010 and 2015, there was a rapid transition of the data source from 100% fishers to 100% on-board observers. Length measurements from the Japanese longline fleet were recorded at various spatial resolutions and bin sizes. This benchmark assessment includes only those collected at a spatial resolution of 1° x 1° and a bin size of 1, 2, or 5 cm. The longline length composition data, collected by Korean observers at a spatial resolution of 1° x 1° and a bin size of 1cm, covers the period between 2013 and 2023. The longline length composition data collected by Korean observers were considered in addition to the data obtained by Japanese observers. Korea recently replaced Japan as the fleet with the largest longline effort and catches of bigeye tuna, their main target species. No difference was found between the data obtained by Korean and Japanese observers in the same period. The Korean data complements the Japanese data as some Korean vessels operate in areas not covered by Japanese ones ([SAC-15-02](#)).

Both the data measured by fishers and the data measured by observers were standardized using spatiotemporal models, then raised to the catches or to the density, similarly to what was done in the benchmark assessment of bigeye tuna in 2024 ([SAC-15-02](#)). However, unlike bigeye tuna, two separate models were estimated, one for each data type, as there were differences detected in the data collected by fishers and observers. There is no clear understanding of how the data was collected and whether the methods used were different between fishers and observers. The difference may also be a result of temporal changes in the population, and because the observer data collection replaced the fisher data collection for the most part, temporal effects and sampling methods are confounded. Also, the observer data has smaller fish not present in the fisher data. It is not clear whether the observers record fish that are discarded, the selection criteria for fish that are measured differ between observers and fishers, or some other factor is causing these differences.

The predicted length compositions in a 5° by 5° by quarter resolution were multiplied either by the catches on the same spatial scale or by the estimated abundance (density\*area) then aggregated within the boundaries of spatial definition of the fisheries or the spatial domain of the index, to represent the catches or the index, respectively.

#### **2.2.2.1. STANDARDIZATION PROCEDURE**

The standardization of the length composition data to represent the catches is an improvement over what has been done previously. For the 2020 Benchmark assessment, longline length composition data was not

used to fit the models because the assessment focused on the area where the core of the catches are taken. Previous assessments simply used the nominal length composition data raised to longline catch. Because the spatial distribution of the length composition data differed from the catch spatial distribution, there were areas for which no length composition data was available and those catches were not considered when raising the nominal length composition data to the catches, implying that the resulting length composition may not adequately represent fishery removal. This may not be an issue for the EPO-wide model, or models for the NE region, as the longline catches represent a small proportion of the total catch of yellowfin tuna. This is not the case for the SW for which the longline fishery contributes a large proportion of the catch of adults. To allow for informed imputation of length composition data for areas with no samples, length-specific spatiotemporal models were used. Models similar to those implemented for bigeye tuna were used (SAC-15-02). The models were implemented in VAST (Thorson and Barnett 2017), an open-source R package (<https://github.com/James-Thorson-NOAA/VAST>). The data used was the length frequency, aggregated across vessel by year, month, 1° latitude, and 1° longitude.

The spatiotemporal model used logit and log link functions for the linear predictors of encounter probability and positive catch rate, respectively, for each length bin. Both linear predictors include an intercept (year-quarter) term, a time-invariant spatial term, and a time-varying spatiotemporal term. All three terms are assumed to be independent and identically distributed among length bins. Of the three terms, the intercept term is estimated as fixed effects and the other two terms are estimated as random effects. The spatial and spatiotemporal random effects are both assumed to be autocorrelated in space according to the Matérn function. Neither the catchability covariate (hooks between floats) term nor the vessel effects term is included in this model because they are not available in this dataset.

Preliminary runs of the model failed to converge. This is most likely due to the sparsity of the data, which may not have enough information to estimate the autocorrelation or other parameters. In addition, for the 2024 indicator paper (SAC-15-INF-F), the model had an estimate of spatiotemporal correlation that was not credible (it was the opposite of that estimated for the CPUE model, that is anisotropy in the SW - NE axis rather than NW-SE axis as estimated in the CPUE model). Thus, the spatial autocorrelation parameters for the standardization of the length frequency data were fixed at values estimated in the spatiotemporal model developed to obtain the joint index of abundance using operational-level data from Japan and Korea (SAC-INF-U).

Due to the high dimensions of the length-specific spatiotemporal model, several simplifications are made to make the model computationally more feasible: 1) only 40 spatial knots are used to estimate the spatial and spatiotemporal random effects in the EPO; 2) length bins are regrouped from the original resolution to 10 cm; 3) length frequencies for < 60 cm are negligible and are assumed 0 (length bins in the model: 60-70 cm, 70-80 cm, ..., 170+ cm); and 4) all hyperparameters are assumed to be shared among length bins. It should be noted that the predicted length frequencies ( $lf$ ) for each knot and time do not necessarily sum to 1 across length bins, as the spatiotemporal field of length frequency is predicted for each 10 cm length bin without a multinomial constraint. To solve this problem, we scale the predicted length frequencies to have a sum of 1 for each knot and time.

The length compositions of a fishery fleet are raised to the catch within the spatial domain of the fishery. Specifically, the length frequency for a fishery fleet ( $LF(F)$ ) in time  $t$  and length  $l$  is computed as:

$$LF(F)_{t,l} = \frac{\sum_s (c_{s,t} \times lf_{s,t,l})}{\sum_l \sum_s (c_{s,t} \times lf_{s,t,l})} \quad (\text{Equation 1})$$

where  $c_s$  is the fleet-specific total catch in cell  $s$  and time  $t$ , and  $lf_{s,t,l}$  is the length frequency in cell  $s$ , time  $t$ , and length  $l$  predicted by the length-specific spatiotemporal model. The fleet-specific total catch, reported in the number of fish, is extracted from the IATTC's database and has a spatial resolution of 5° x

5°. To match this spatial resolution, we aggregate the predicted length frequencies from the length-specific spatiotemporal model from 1° x 1° to 5° x 5°. The longline length composition data are spatiotemporal model-based, so to be consistent we also use model-based input sample size for the longline length composition data. Specifically, the input sample size is calculated by the length-specific spatiotemporal model to approximate the estimated imprecision for predicted length frequency (Thorson and Haltuch 2018)..

### **2.3. INDICES OF ABUNDANCE AND CORRESPONDING LENGTH COMPOSITION**

In this benchmark assessment, purse-seine and longline indices are key pieces of information in different reference models. The indices of abundance used depend on the hypothesis of stock/spatial structure considered (Table 1). The purse seine index is used in NE and EPO models (and in the spatial model), and the longline index is used in the SW model.

The weighting of the indices of abundance is determined from the variability estimated by the spatiotemporal model for each time step plus an additional variance component to account for process error not modelled by the assessment models. The variance component was estimated by fitting an age-structure production model with estimated recruitment deviations (ASPM\_dev). Because the ASPM\_dev is fit only to the index, it is the best fit possible to this data, and the estimated variance component is a minimum. However, this follows the philosophy of prioritizing information about absolute abundance and abundance trends from indices of abundance over composition data.

The purse-seine indices of abundance were obtained using the newly developed Integrated Spatiotemporal Model (ISAM,SAC-16-INF-F). This new methodology improves previous procedures of standardizing by using variable extrapolation grids that approximate seasonal changes and changes due to large scale oceanographic conditions (El Niño Southern Oscillation) in yellowfin distribution and fishing grounds. ISAM also allows for the construction of regional indices of abundance that share vessel random effects, which are used to model difference in fishing power by vessel. Several ISAM indices were obtained with different spatial domains depending on the spatial structure hypotheses (Table 1).

The longline index of abundance was obtained by fitting a spatiotemporal model to operational CPUE data from both Japanese and Korean longline fleets (SAC-16-INF-U). VAST, a delta-generalized linear mixed model, models separately encounter probability and positive catch rate to account for zero-inflated catch rate observations. VAST was specified to use the logit link for the linear predictors of encounter probability and the lognormal link for the positive catch rate. Both the linear predictors of encounter probability and positive catch rate include a year-quarter fixed effect, a time-invariant spatial random effect, a time-varying spatiotemporal random effect, a catchability fixed effect of hooks-between-floats, a vessel random effect, and a flag (Japan vs. Korea) fixed effect. Vessels and spatial grid cells that had at least 40 quarters of data were selected. This selection is done to remove the eastern part of the EPO and the EEZ of French Polynesia, both of which have sparse data and low CPUE for yellowfin. Preliminary analysis showed that filtering the CPUE data as described above results in an index of abundance with reduced CV. The longline index of abundance was used when modelling the SW region (Table 2)

Size compositions for fisheries are spatially weighted by catch within their respective operational areas, whereas abundance index compositions are weighted by fish abundance across the EPO. This distinction ensures that index selectivity is treated as primarily gear-based and approximately constant over time, while fishery selectivity accounts for spatial differences in abundance and fleet distribution. For the purse-seine index of abundance, the spatiotemporal model for standardization of length composition and the raising procedure is described in SAC-16-INF\_F. For the longline index, the VAST model for length composition is described in section 2.2.2.1, and both the fishers and observer data are used to represent the index.

### 3. ASSUMPTIONS AND PARAMETERS

#### 3.1. GROWTH

Growth was updated by fitting the growth cessation model to a combination of new otolith daily increment age and length data and tagging data (SAC-16-INF-F). Information for younger fish (up to 10 quarters of age) came from the otolith data and the information for older fish came from tagged fish with expected age at recovery of 10 or more quarters and with reliable length information. Sex information for recoveries is limited. Based on the available data, no difference in growth between males and females was found. Both the average size at age and the variability of size at age were estimated, taking into consideration measurement error and bias from freezing related shrinkage of tagged fish.

Three assumptions about growth were used (Figure 3). The “base” assumption is the maximum likelihood estimates (MLE) obtained in the external model (Asymptotic length of 167.1 cm). A “G\_low” and “G\_high” assumptions were also included to represent the uncertainty in growth. Those were obtained by fixing the asymptotic length to the values that had half the likelihood of the MLE, that is 171.922 cm for the G\_high hypothesis and 162.245 cm for the G\_low hypothesis, and estimating all other parameters in the external model fit to otolith and tagging data.

In the assessment models, growth was fixed at the external estimates, except for the coefficient variation at age 1, which was estimate within the assessment models (Figure 3). The variability of size at age includes several sources of variation such as differences in growth rates between fish with the same birthday, differences in birthday within a quarter, when in the quarter a fish was caught, and changes in growth between years. This variability, especially at young ages, may not be adequately captured when estimating growth using daily increment data. It is beneficial to estimate it within the assessment model to draw upon the information contained in the length frequency data, especially for small fish.

#### 3.2. NATURAL MORTALITY

Natural mortality (M) for yellowfin tuna in the EPO was estimated in an external analysis and fixed in the assessment model. M was assumed to vary by age and sex, because (i) natural mortality has been shown to have a consistent pattern of declining with size (Lorenzen, 2022) and (ii) sex ratio data showed preponderance of males at large sizes, while growth is similar for females and males. M was estimated externally by applying a cohort analysis to EPO tagging data obtained by the recent IATTC tagging program (SAC-14-07) and fitting to sex ratio data from the EPO (SAC-16-INF-F). The sex ratio data came from both purse-seine and longline fisheries.

Cohort analysis is used because it is robust to the impact of non-mixing on fishing mortality. However, it instead makes assumptions about the terminal fishing mortality (i.e., no tagged fish are alive after the last recapture). Natural mortality was parameterized using the Lorenzen function to model the decline in M with age and a logistic offset to model an increase in female natural mortality related to maturity. The M assumptions were similar to the 2024 exploratory models but differ slightly from the 2020 benchmark assessment (Figure 4).

Similarly to growth, three hypotheses for natural mortality were implemented in this assessment (Figure 4). The “base” hypothesis was represented by the MLE estimates of natural mortality at age and sex (M) obtained from the cohort analysis. The M\_low and M\_high were obtained by assuming a normal distribution with mean equal to the M MLE and standard deviation equal to the standard error estimated for each value of natural mortality at age and sex from the cohort analysis. The M\_low and M\_high were:

$$M_{low_{a,s}} = \widehat{M}_{a,s} - 1.1759 * SE(\widehat{M}_{a,s})$$

$$M_{high_{a,s}} = \widehat{M}_{a,s} + 1.1759 * SE(\widehat{M}_{a,s})$$

As  $dnorm(x=-1.1759, mean=0, sd=1) / dnorm(x=0, mean=0, sd=1) = 0.5$

### 3.3. REPRODUCTIVE BIOLOGY

New data (Schaffer and Fuller 2023) was used to update the fisheries biology assumptions (SAC-16-INF-F). This includes proportion mature, batch fecundity, and frequency of spawning at length. The estimates of length at 50% maturity were different from the previous estimates. Also, spatial differences are marked. Length at 50% maturity are 77.7 cm and 95 cm for the NE and the SW areas, respectively (Figure 5). Batch fecundity was estimated for both regions combined, as there were few data points for the SW region, and increase faster than weight ( $0.04219w^{1.12444}$ ). The resulting reproductive output differs between the two regions mainly between the sizes of 75 to 125 cm, which covers the range for most adult females.

### 3.4. RECRUITMENT

Recruitment is estimated on a quarterly time step as deviations around a Beverton-Holt stock-recruitment curve (Beverton and Holt, 1957). The Beverton-Holt curve is parameterized so that the relationship between spawning biomass (fecundity in this assessment) and recruitment is determined by the average recruitment produced by an unexploited population (virgin recruitment) and steepness ( $h$ ). Steepness is defined as the proportion of the virgin recruitment that a population produces when reduced to 20% of its virgin state. A steepness of 1.0 implies that the stock may produce recruitments equal to the virgin level, on average, at all levels of spawning biomass, while a steepness of 0.8 indicates that when a stock is at 20% of its virgin spawning biomass, only 80% of the virgin recruitment is produced, on average.

Steepness is a key parameter of a stock assessment, but it is problematic to estimate (Lee et al., 2012). In practice the recruitment of tropical tunas may be more related to the extent of favorable habitat for larvae rather than the spawning biomass (Maunder and Deriso, 2013). The habitat may vary with environmental conditions (SAC-14-06), and decadal changes in productivity may occur. Those factors may be confounded with spawning biomass, making the estimation of steepness challenging. **Three steepness values (1.0, 0.9, and 0.8)** were included to address the uncertainty in the shape of the stock-recruitment relationship (similarly to what is assumed for bigeye tuna in the EPO, SAC-15-02). The three steepness values are weighted based on expert judgement from the risk analysis for the last benchmark assessment (SAC-11 INF-F):  $P(h=1.0) = 0.46$ ,  $P(h=0.9) = 0.32$ ,  $P(h=0.8) = 0.22$ .

Recruitment is assumed to vary lognormally around the stock recruitment relationship with a fixed standard deviation on the logarithm of the recruitment deviations ( $\sigma_R$ ). Ideally the recruitment variability should be estimated using random effects or state-space approaches (Maunder and Thorson, 2019), which can be computationally limiting for large models. As an approximation, integrated models implement a penalized likelihood approach where recruitment variability is constrained by a penalty added to objective function. Recruitments are corrected so that the expected values are unbiased. The bias correction is computed using the method of Methot and Taylor (2011). As recommended in the external reviews, a value of  $\sigma_R (= 1)$  large enough to estimate the individual recruitment without much constraint was used in the model development phase.  $\sigma_R$  was then modified following an iterative process that takes into account both the variability among the estimated recruitments and the uncertainty of each of the estimates (Methot and Taylor, 2011). The bias-correction was applied was computed using the library *r4ss* (Taylor et al., 2021).

### 3.5. MOVEMENT

Tagging data suggests that movement of yellowfin tuna is limited. However, the available data is limited, particularly in release locations, and a comprehensive analysis of the data with respect to the needs of a

spatial stock assessment model has not been conducted. Therefore, it is premature to base movement within a spatial stock assessment on the available tagging data for yellowfin tuna in the EPO.

No movement was assumed in the assessment models, except in the exploratory spatial model. For that model, movement was limited directionally from area 1 to area 2 to area 3 to follow the increase in size in the purse seine fishery associated with dolphins. Movement information is assumed to come from the assumption of asymptotic selectivity in this fishery the fit to its length composition data. Age-specific movement was modelled by fixing the age below which the magnitude of movement was constant and the age above which the magnitude of movement was constant and assuming an exponential increase in movement with age in between those ages. The two levels of movement were estimated. Models with different ages for these levels were conducted. Movement was estimated to be very low and therefore separate assessments for the different regions, as done in the assessment, are consistent with these estimates. It was concluded that a more flexible approach to model movement and more information on movement (e.g., tagging data) is needed to construct a reliable spatial model with explicit movement. None of the models used in the risk analysis explicitly model movement.

### **3.6. FISHERY SELECTIVITY AND DATA WEIGHTING**

The approach used to define fisheries, model selectivities, and weight the length composition data is based on a decision tree, similar to that used for the bigeye ([SAC-15-02](#)) and skipjack ([SAC-15-04](#)) stock assessments. The underlying philosophy of this approach is that the index of abundance and its composition data—standardized using spatiotemporal models to better represent abundance and minimize time trends in selectivity—should provide the primary source of information on population abundance. In contrast, fisheries should be structured to remove the catch at the appropriate length and age and contribute limited information on abundance. This approach assumes that fisheries should exhibit “regular” length composition distributions (i.e., smooth and unimodal), which can be modelled by a “regular” selectivity curve (e.g., double-normal). These assumptions are supported by gear selectivity studies that have shown regular length compositions and selectivities. However, when the index does not reliably inform absolute abundance, fisheries may need to serve as a supplementary source of abundance information.

This decision tree trades off adequately representing selectivity patterns while maintaining model efficiency and stability and avoiding noise in the data being interpreted as signal. While theoretically, all data-rich fishery fleets should employ time-varying selectivity to reduce misspecification and enhance estimation accuracy (Martell and Stewart, 2014; Xu et al., 2019) estimating additional selectivity parameters for each fleet would impact model efficiency and stability.

To implement this philosophy, first the fisheries were defined spatially to construct areas where the length compositions are similar, then a structured framework (“decision tree”) was developed to evaluate whether fisheries exhibit “regular” length compositions and whether a double-normal selectivity curve can adequately predict length composition in the stock assessment. If a fishery does not meet these criteria, further adjustments may be necessary. These include refining fishery definitions, down-weighting or eliminating composition data, and fixing selectivity parameters at appropriate levels.

The approach in [SAC-15-02](#) was slightly modified for yellowfin tuna, and it is summarized in the decision tree in Figure 6. For each fleet, a decision is made based on the magnitude of catches, the quality of the length composition data and whether the data showed unimodal distributions that could be modelled by a double-normal selectivity. Each fleet follows one of three selectivity and data-weighting strategies:

1. Fleets with high catch volumes, rich composition data, and a strong fit to a double-normal selectivity curve should use time-varying selectivity. The time-varying selectivity approach

adopted is to include selectivity blocks every 10 years. Data weighting should apply the Francis weighting method (Francis, 2011). However, if data within a block is poor, then the data for this block should have a zero weight and the selectivity block mirrored to the nearest block.

2. Fleets with low catch volumes, poor composition data, or an inability to fit a double-normal selectivity curve should not have an estimated selectivity in the final models (i.e., it should be fixed to values obtained in preliminary runs or mirrored to other fleets) and the model should not be fit to their composition data.
3. Fleets that do not fit into either of these categories should use constant selectivity and apply 20% of the Francis weight.

Fleets F2, F14, F15 have three selectivity time blocks (splits at years 2000 and 2010), and F3 has two time-blocks (split at year 2010). For F3, the data before the year 2000 had low sample size, was sparse and variable, thus low quality and zero weight. All other fleets for which selectivity is estimated use time-invariant selectivity, with their composition data down-weighted by 80% to minimize their influence on population abundance estimates (Table 3).

The double normal selectivity at length, as parametrized in SS3, has flexible shapes and can be used even to represent asymptotic selectivity curves. In initial runs, all the parameters of the double normal selectivities were estimated. In subsequent runs, some, or all parameters were fixed according to the decision tree and fits to the length composition.

For F15, the most important purse-seine fishery associated with dolphins, both an age-based asymptotic selectivity and a length-based asymptotic selectivity were used to provide more flexibility in the shape of the curve for younger ages, which was needed to adequately model the length composition data.

The largest purse-seine fisheries associated with floating objects (F2, F3, F4) catch mostly smaller fish, but occasionally large fish are also caught. The presence of these large fish in the length composition data may bias the estimation of the selectivity curve. The resulting curve may not adequately model small fish. To minimize this influence, fish larger than 82 cm were given zero weight in the estimation and the selectivity curve was modeled in SS3 as a combination of selectivity, a retention curve (1 for length < 82.5 cm, 0 otherwise), and zero fishing mortality for non-retained sizes. This assumption ensured that most of the catch was accurately removed at small sizes, with a very minor portion of the catch being removed at a smaller incorrect size.

A similar approach was used to model the selectivity for the unassociated fisheries for large sizes (Table 3), but the truncation was done on the left side of the distribution curve, to remove the small sizes.

It is important to carefully consider the assumption of asymptotic selectivity because of the potential influence of this assumption on estimation of the absolute population size and status of the stock, when fitting the models to composition data from those fleets. The fisheries that catch the largest sizes are the ones considered as having asymptotic selectivity (Table 3, fisheries indicated with “A” in the “double-normal” column). These are longline fisheries, the purse-seine fisheries associated with dolphins, or both, depending on the model. It also depends on the area, as those fisheries show spatial variation in length composition. Investigation of historic Japanese longline length composition from the 1960s and 70s (Figure 7b in [SAC-15-03](#)) supported the hypothesis that larger fish were historically found in the core area, which shows mostly fish of intermediate size in the purse-seine fishery associated with dolphin, while large fish are found historically and currently in the offshore and southern coastal areas in both the longline and the purse-seine fishery associated with dolphin. A few vessels from the longline observer program of the Chinese fleets operate in the core area have shown catches of larger sizes than the purse-seine fishery associated with dolphin, indicating that maybe in this area large yellowfin may not be available to the

purse-seine fishery associated with dolphins (unpublished analyses by the staff in collaboration with Chinese scientists).

### **3.7. SELECTIVITY FOR THE INDICES AND WEIGHTING OF THE INDEX COMPOSITION DATA**

Selectivities and composition data weight for the indices of relative abundance do not follow the same decision tree as for the fisheries. The indices and their composition data are assumed to have the most reliable information of absolute abundance and trends in abundance. The selectivity's are assumed to be asymptotic and time invariant, and length composition weighting is based on Francis' method. Unlike the fisheries, for which the composition data are spatially weighted by catch within their respective operational areas, the survey fleet compositions are weighted by fish abundance across the EPO. This allows the index selectivity to be treated as primarily gear-based and approximately constant over time, while fishery selectivity needs to account for temporal changes in availability and fleet distribution.

### **3.8. INITIAL CONDITIONS**

The model is assumed to start from fished state, with the initial recruitment and the initial fishing mortality ( $F_{init}$ ), being estimated, with no penalty associated with initial equilibrium catches. The fishery assumed to correspond to  $F_{init}$  was chosen as a fishery with a wide range of sizes and large catches to best represent the equilibrium fishing mortality at age for the stock (F15 for EPO, NE and NE\_short models and F16 for SW models). Additionally, 16 recruitment (quarter) deviations before the start of the model initial quarter were estimated, so that variability in the initial age structure is accounted for.

### **3.9. MODEL DIMENSIONS**

The model period is 1984-2023, except for the NE\_short models (see section 5), which start in 2006. The start year is the same as the 2020 Benchmark assessment but differs from the previous benchmark stock assessments, which started in 1975, because data from the purse-seine fishery before 1984 with spatial information necessary to standardize the index and length frequencies are limited. Thirty age classes are defined, from 0 quarters to 29+ quarters (7.25 years). The population size structure was defined in 2-cm intervals from 2 to 200+ cm. The model is structured by sex, but only natural mortality differs between females and males. The size compositions are defined using 2-cm intervals, from 20 to 198+ cm, for the fisheries, and 10-cm intervals, from 20 to 170 cm for the purse-seine indices and 60 to 170 cm for the longline index. The models are conditioned on catches and fit to the relative abundance indices and length composition data.

## **4. MODELS**

### **4.1. ANCESTRAL MODEL**

An “ancestral model” was created to provide a foundation for the models used in the stock assessment and risk analysis. All models were derived from the ancestral model which contained 37 fisheries and 7 indices (Table 2). This model is fit to all available data. The ancestral model is not used for management advice or in the risk analysis. Depending on the stock structure hypothesis being represented by an assessment model, different fisheries and indices are turned on and off and the selectivities are set to the values estimated in the ancestral model. The spatial model was implemented by assigning the region of the fishery as indicated Table 1.

### **4.2. REFERENCE MODELS**

A risk analysis approach is used in this benchmark assessment. The first step to apply the risk analysis framework ([SAC-11 INF-F](#)) is to list the unresolved issues and uncertainties that need to be accounted for in the management advice. This includes defining alternative “states of nature” (Hilborn and Mangel,

1997) (i.e. hypotheses) that are considered plausible for describing the population dynamics of yellowfin tuna and address the unresolved issues or represent the uncertainties. Several hypotheses are formulated that represent these different states of nature that are arranged in a hierarchy. The higher-level hypotheses (overarching hypotheses) representing the most important uncertainty (level 1) and lower-level hypotheses nested within the higher level to represent other uncertainties (level 2), and are crossed with the level 3 hypotheses, which encompass parameters for which there is little or no information in the data.

The three levels of hypotheses in the risk analysis for yellowfin tuna are:

**Level 1** - spatial structure;

**Level 2** - uncertainty in biological parameters (growth, natural mortality) and effort creep (1% increase per year in the catchability of the indices of abundance); and

**Level 3** - steepness of the stock-recruitment relationship.

The overarching hypotheses (Level 1) addresses the issue of spatial structure. Although there is some evidence of the existence of northern and southern stocks, the divisions are not clear and mixing between the two potential stocks may be episodic, or the magnitude may vary from year to year ([SAC-14-06](#)). Alternatively, there may be regional dynamics. The delimitation of meaningful regions is challenging and may require the expansion of the tagging effort in the EPO. Delimitation of regions and areas was addressed with cluster analysis of length composition. Stock structure is approximated in two ways, with spatial models and with the areas-as-fisheries approach. The spatial model is considered exploratory at this time and will not be used for management advice as the movement rates were estimated to be very low and there are limitations in the modelling platform to model movement. A NE model starting in 2006 was also developed to address different information on abundance trends between the index of abundance and composition data that indicated a change in the stock or fishery dynamics after the large 1998 El Nino. Although this is not a stock-structure hypothesis, it was included at Level 1 to ensure that the Level 2 and 3 hypotheses were also evaluated for this model.

Level 1 Stock structure:

1. EPO
2. NE
3. NE short
4. SW

The models for the EPO and NE region are fit to indices of abundance based on the purse-seine set associated with dolphin associated and models for the SW region are fit to indices of abundance based from longline.

The level 2 hypotheses were implemented by changing one assumption at a time in the base reference model of each spatial structure. The low and high scenarios for growth (Figure 3) and natural mortality (Figure 4) were based on the uncertainty of the external estimates (values that have approximately half the likelihood as the maximum likelihood estimate). The effort creep scenario of 1% a year of increase in catchability is based on the findings of a recent review ([IOTC-2024-WPPT26DP-16](#))

For the level 3 hypotheses, three values of steepness of the Beverton-Holt stock recruitment relationship ( $h=1.0$ ,  $h=0.9$ ,  $h=0.8$ ) were considered for the third level hypothesis.

One risk analysis was done for each of the four level 1 hypotheses by combining 18 reference models, which resulted in a total of 72 models used. Equal weight was used for all level 2 hypotheses. The weights for three values of steepness (level 3 hypotheses) were based on expert judgement from the risk analysis

done for the 2020 benchmark assessment:  $P(h=1.0) = 0.46$ ,  $P(h=0.9) = 0.32$ ,  $P(h=0.8) = 0.22$ .

## 5. ASSESSMENT RESULTS

### 5.1. MODEL CONVERGENCE AND DIAGNOSTICS

The models were fit by minimizing a penalized negative log-likelihood function (NLL). To ensure that the models obtained the global minima, a series of jitter analyses, which randomly change the initial parameter values to test convergence, were performed until the model passed the jitter test (the initial parameters estimates were ones that produced the lowest NLL among all jittered models). All 72 models passed the jitter analysis, converged (all produces a positive definite Hessian matrix) and are used in the stock assessment and risk analysis (Table 4). The fits were also evaluated using residual analysis and ensuring that selectivity curves were “sensible”. Integrated model diagnostics were used to understand the information content of the data. The diagnostics used were the age-structure production model (ASPM), ASPM with estimated recruitment deviations (ASPM\_dev), ASPM\_dev also fit to the index length composition data, which has an asymptotic selectivity curve (ASPM\_dev+), catch curve analysis (CCA), CCA only fit to the index length composition data (CCA-I), likelihood component profile on the scale parameter ( $\log_{R0}$ ), and retrospective analysis.

#### Residual analysis

Residual analysis provides information on how well the model fits the data and whether they are consistent with the assumptions (e.g., the distributional assumptions and weighting factors used). In general, patterns in residuals are evaluated visually. Many of the weighting factors for the data fits are determined based on tuning methods or assumptions about the influence a data set should have, so there evaluation is less informative.

#### Indices of abundance

All the models visually fit the indices of abundance reasonably well and the fit does not differ among assumptions (Level 2 and Level 3 hypotheses, Figures 7a – 7d). This is also supported by the RMSE of the fit (Figure 7e). However, some patterns can be noted. First, the models cannot match the quarterly variability of the data and show a smoother trend. Also, no model can capture the high increase in the observed indices around year 2000. NE\_short is the model that can capture the overall interannual trend best (Figure 7c). This may be an indication that the new biological assumptions are more consistent with the data from this period. The index for the SW area shows slight incompatibility with the model trends (Figure 7d). This may be due to the uncertainty in the boundaries of the stock/region and potential “contamination” with data from other regions.

#### Length composition

The models fit the length composition well. The fish are taken out at about the right size for the fisheries that have their selectivities fixed at those estimated by the ancestral model (Figure 8a). The EPO model generally fits, on average, the fishery length composition well for the fully weighted data (Figure 8b), except the flat top of the unassociated purse seine (F6) length composition data. The flat top may be a result of the catches coming from a mix of schools of different sized fish as in the other unassociated fisheries, but the average fish sizes for small and large fish being more similar than in other areas (SAC-16-INF-F). Fitting a flat top length composition requires a very steep selectivity curve. The fits to the down weighted data are also reasonable (Figure 8c). The EPO model also fits the index length composition data well, but does not capture the slightly bimodal peak (Figure 8d). The bimodal peak may be due to a mix of sizes from different areas, a concern that was pointed out in the 2024 exploratory models ([SAC-15-03](#)). The bimodality in this model is not as pronounced, most likely due to the improvements in the

standardization methods (e.g. the spatial domain changing with quarters and ENSO condition, which downweights areas on the boundaries that have no catch observations during some time periods).

The NE and NE-shot models also fit the index length composition well, on average, and likewise do not completely capture the peak of the distribution (Figure 8d).

The SW models generally fit the longline fishery (Figure 8c) and index (Figure 8d) length composition data well, but there is an inconsistency with the longline index data measured by the fishers and the observers (Figure 8d). The fishers composition distribution is few centimeters smaller than the observers and does not include the smallest fish. The model fits the fishery data and misfits the whole observer length composition..

Visual inspection of bubble plots allows detection of systematic misfits represented by patterns in the residuals. The patterns in residuals of a particular fishery are generally the same no matter which model is fit to the data (Figures 9a-9h). Most of the fits are typical of stock assessment models fits to length composition data with runs in residuals across lengths for the same period, runs across time for the same length, and diagonal runs across both time and length. Of most concern is residual patterns in the index composition data and the fisheries that assume asymptotic selectivity and have full weighting (Table 3). The core dolphin associated fishery data show some patterns (Figure 9e) which are related to the model not being able to fit the shoulders of the length composition distribution. The EPO, NE, and NE\_short models underpredict the large fish in the second half of the period (Figure 9f), which is consistent with the CCA suggesting the length composition data supports larger biomass in the second half of the time period (Figure 10a). Residual pattern for the offshore floating object fishery (F3) decrease in the SW model in relation to the EPO and NE models, indicating an improvement in the fit (Figure 9b) and inconsistency between the SW and NE areas. The fit improves slightly for main purse-seine associated with dolphins in the core area (F15) in the short model (NE\_short) in relation to the NE and EPO models (Figure 9e). The improvement in fit in the short model is more marked in the data associated with the purse-seine indices (Figure 9h).

### **Selectivity**

Selectivity curves are an important component of the assessment because they have a direct impact on the size of fish caught and the fit to composition data. Fisheries were defined so that regular (i.e., double normal) selectivity curves could well represent the composition data. If not, the composition data was down weighted or not fit at all and the selectivity fixed based on other information (Figure 6). Many of the fisheries had dome-shape selectivities with a long tail for large fish (Figures 10a-10e), including all of those with full weighting that don't assume an asymptotic selectivity, suggesting that even though these fisheries target small or intermediate sized fish they also catch large fish. Some of the fisheries have substantial temporal changes in selectivity at large size.

Of particular interest are the selectivities associated with the fully weighted length composition data (Figures 10d, 10e, 10f). The dolphin associated purse seine fishery in the core area (F15) (Figure 10f) and the dolphin associated index (Figure 10g) combine asymptotic age based and asymptotic length based selectivities to allow flexibility to fit the flat top length composition distribution. Although the age-based selectivity probably does not add much to the resulting age-based selectivity which determines the fish that are removed from the population it may give more flexibility in representing the length composition data.

### **Age-structured production model (ASPM)**

This diagnostic (Maunder and Piner, 2015) may be used to: (i) evaluate model misspecification, (ii) ascertain the influence of composition data on the estimates of absolute abundance and trends in

abundance, and (iii) check whether catch alone can explain the trends in the indices of abundance. The ASPM diagnostic is computed as follows: (i) run the full assessment model; (ii) fix selectivity parameters at the maximum likelihood estimate (MLE), (iii) turn off the estimation of all parameters except the scaling parameters ( $R_0$ ), parameters used to create the initial conditions, and index catchability, and set the recruitment deviates to zero; (iv) fit the model to the indices of abundance only; (v) compare the estimated trajectory to that of the reference model. When the ASPM well fits an index of abundance with contrast, it is likely that the index, in combination with the catches, provides information on absolute abundance (Maunder and Piner 2015). When the catches cannot explain the changes in the indices, the ASPM will fit the index poorly. This can have several causes: (i) the stock is recruitment-driven; (ii) the stock has not yet declined to the point where catch is a major factor influencing abundance, (iii) the full assessment model is misspecified, or (iv) the indices of relative abundance are not proportional to abundance. Checking whether the stock is recruitment-driven involves estimating recruitment deviations when fitting the model (ASPM\_dev). If this is still not able to capture the population trajectory estimated in the integrated model, it can be concluded that the information about scale in the integrated model comes from the length composition data. Large confidence intervals on the abundance estimated by the ASPM also indicate that the index of abundance has little information on absolute abundance.

The ASPM estimates much larger biomass than the full model for all stock structures, indicating that information on recruitment is needed to extract reliable absolute abundance information out of the index of abundance (not shown). The ASPM-dev model estimates much lower biomass than the full model for the EPO, NE, and NE\_short models (Figure 11a) so that the recruitment variation drives trends in abundance to fit the index of abundance (Figure 11b). For these models, the estimated composition data (not fit) is much smaller than the observed due to the higher estimated fishing mortality. For the SW model, the ASPM-dev estimates higher biomass than the full model in the first half of the period. The ASPM-dev+ model with the index composition data estimates absolute biomass and trends in biomass closer to the full model, indicating that information on these quantities is coming from the index composition data, which has an asymptotic selectivity curve. Although, the absolute levels still differ for the NE and NE-short models, so some information on absolute scale must come from other composition data. This diagnostic was also conducted for the low M sensitivity and are generally like the base-reference models.

**Catch-curve analysis (CCA)** is done by fitting the integrated model only to the length composition data, and estimating all parameters except the auxiliary parameters associated with the index (Carvalho et al., 2021, 2017). The decline in the proportion of catch-at-age with age (the catch curve) provides information on fishing mortality (since the natural mortality assumed to be known), and when combined with catch data provides information on abundance. The CCA is used to verify whether the temporal trend implied by the size composition data is consistent with that coming from the index of abundance. If the two trends are similar, then there is more confidence that the estimated abundance trend is accurate. Two variants of the CCA were used, one that is fit to all length composition data (CCA) and other that is fit only to the survey data (CCA-I).

In general, the CCA-I and the CAA show similar estimates of absolute biomass and trends (Figure 11a). CAA estimates an increase in biomass after 2006 while the full assessment model estimates a lower biomass for the EPO, and NE models. This motivated the development of the NE\_short model that starts after this change to account for any changes in the dynamics of the stock or fishery, which may have been caused by the strong El Niño in 1998, followed by a strong La Niña in 1999. In addition, the tagging, otolith and reproductive biology data come from recent years and may best represent the time frame of the NE\_short models. The CAA for the NE-short model highlights some differences in information between the index and composition data about the increase in abundance in the last two decades. The CAA for the SW model

shows very similar abundance levels and trends as the full assessment, indicating the influence of the composition data, except in the final decade where the index composition data supports a rapid decline. This diagnostic was also conducted for the low M sensitivity and are generally like the base-reference models.

### **Likelihood component profile on the global scaling parameter**

A likelihood component profile of the average recruitment in an unfished (virgin) population in logarithm scale,  $\ln R_0$ , is used to determine whether information about absolute biomass scaling is consistent among data sets (e.g. (Francis, 2011; Lee et al., 2014; Wang et al., 2014)). The profile is done by fixing  $\ln R_0$  to a range of values around the maximum likelihood estimate (MLE) and estimating all other parameters, then obtaining the contribution of each data set and penalty components to the likelihood conditioned of the value of  $\ln R_0$ . The profile quantifies how the fit to each data component is degraded by changing the population scale. The data with large amount of information on population scale will show loss of fit (smaller likelihood, or larger negative-log likelihood) as population scale is changed from its best estimate (Lee et al., 2014). If different data components favor different values for  $\ln R_0$ , there is contradictory information among them, conditioned on the model, thus pointing to potential model misspecification.

The main contributions to the likelihood profile, which inform absolute abundance, for all models are the length composition data and the recruitment penalty (Figure 11c). The index of relative abundance supports lower biomass levels but provides much less information than the composition data. The composition data for the different fisheries generally all support similar absolute biomass values. This diagnostic was also conducted for the low M sensitivity and are generally like the base-reference models (Figure 11d).

### **Retrospective analysis**

Retrospective analysis is useful for determining how sensitive model estimates are with the addition of new data. They indicate the reliability of recent estimates of abundance and fishing mortality. The analysis is generally done by consecutively eliminating a year of data from the end of the time series. Inconsistencies in the results of this progressive removal of data are a signal of inadequacies in the assessment models. The assessment model has a quarterly time step, but new data are updated annually (four quarters at once). Thus, the retrospective analysis was done by removing whole years of data at once.

Retrospective analysis was run for each of the EPO models with steepness of 1. The results were similar for all models (level 2 hypothesis). The final biomass estimates had moderate negative bias in some years (Figure 11e).

## **5.2. ESTIMATED TRENDS**

### **5.2.1. RECRUITMENT ESTIMATES**

The recruitment trends show patterns of similarities and differences among spatial structure hypotheses (Figure 12a). All models that compose the ensemble for each spatial structure hypothesis have similar trends in recruitment (Figure 11a). Models for all four spatial structure assumptions estimate two peaks in recruitment, but for the SW model the largest peak occurs in 1998, while in the others it occurs in 1999 (Figure 12b). The second peak is in 2021. The SW models also estimated high recruitment in 2015-2017. The EPO and NE models estimate a regime shift in recruitment to a lower level after this peak, while the SW model does not.

### 5.2.2. SPAWNING BIOMASS

The spawning biomass in the NE is estimated to be about twice the level of that estimated for the SW regardless of the level 2 and 3 hypotheses (Figure 13a). The estimate for the EPO is larger than the sum of the estimates for the two component stocks (Figure 13b). The biomass trends generally follow the recruitment trends. Large spawning biomasses are a result of strong recruitment 2 or 3 years prior. The strong cohorts of 1998 and 1999 in the NE and SW regions show up as large spawning biomasses in 2001 and 2002 in the two regions, respectively. The trends in biomass since 2010 are diametric for the NE and SW regions.

The EPO-wide models, which use the areas-as-fleets approach to model spatial structure, estimates larger and more uncertain spawning biomass levels than the NE and SW combined, indicating that the EPO-wide models have difficulty fitting data with incompatible signals (Figure13b). The NE and NE\_short models estimate very similar spawning biomasses.

### 5.2.3. FISHING MORTALITY

The relative distribution of fishing mortality at age is similar for the EPO, NE and NE\_short models, regardless of the level 2 hypotheses (Figure 14a): the fishing mortality is much higher for the older age classes. The magnitude of the fishing mortality, however, is lower for the EPO models (Figure 14b), which is a consequence of its biomass being estimated higher than the sum of the biomasses for the NE and SW regions. The relative distribution of fishing mortality at age of the SW region follows a different pattern. The fishing mortality on the intermediate aged yellowfin (9-12 quarters of age) is lower since the unassociated catches are lower and the purse-seine fishery associated with dolphins generally catches larger yellowfin. The fishing mortality on the youngest yellowfin (1-4 quarters of age) has steadily increased following the expansion of the floating object fishery in the mid 1990's. After 2015 the fishing mortality of this age group surpasses the 5-8 age class.

The trends in fishing mortality are similar between the NE and the NE\_short models (Figures 14a-14b), indicating that starting the model later does not change the perception of the effects of fishing in recent years. For those two hypotheses, there is a general increase in fishing mortality in all age classes after the year 2006, declining after 2015, with the lowest at the start of the covid19 pandemic, in 2020. After that, the fishing mortality increases, particularly for older yellowfin.

The increase in fishing mortality noticed in the last five years in the NE area is not shared by the EPO model. This may be due to the influence of the SW area, which has stable fishing mortality followed by a sharp decline in 2023. This indicates that using an EPO-wide model may underestimate and mask regional trends in fishing mortality.

### 5.2.4. FISHERIES IMPACT

The impact for each type of fishery on the spawning biomass was estimated by projecting the population without their catches (Wang *et al.* 2009). The increase in spawning biomass relative to the current spawning biomass indicates the impact of those fisheries.

The EPO, NE and NE\_short models estimate similar impacts of the different types of fisheries (Figure 16). The longline fisheries have the smallest impact, while the purse-seine fisheries associated with dolphins have the greatest impact during most of the modelled period. The unassociated fisheries had the second largest impact in the early years, but in the 1990s the impact of the floating-object fisheries started to increase and surpassed that of the unassociated fisheries around 2008.

For the SW models, the impact of the different purse seine set type has changed considerably over time. The longline fishery and the purse-seine associated with dolphins had the largest impact until mid-1990's,

when there was an expansion of the floating object fishery, which steadily increased its impact and became the fishery with the largest impact in this region, larger than all other fisheries combined. The longline fishery has decreased both its effort and its impact on yellowfin in that area. The fishery associated with dolphins has slowly increased its absolute impact in this region, but in proportion it has stayed stable since the year 2000.

## 6. BRIDGING ANALYSIS

A traditional “bridging” analysis, where each change is done in sequence from the old to the new configuration, is not possible because of the extensive structural changes in the current assessment compared to the 2020 benchmark assessment (as run in the current SS3 version<sup>3</sup>). Instead, we determine the sensitivity to using the 2020 benchmark growth and natural mortality assumptions in the EPO\_base-1 and NE\_base-1 models of the current assessment (Figures 17a–17b).

The first change is growth. The base growth assumption for the 2020 Benchmark assessment followed a Richards curve with asymptotic length at 182 cm, which is much larger than the about 172 cm assumed in the G\_high hypothesis (and 167 cm or 162 cm assumed in “base” and G\_low) in the current assessment. The former growth curve was a Richard’s curve estimated within an assessment model fit to the daily increments data collected in the early 1980’s and length frequency data (Wild, 1986). In this benchmark assessment, the Richards growth curve is replaced by the growth cessation model fit to recent ageing data from daily increments for younger fish and from tagging data for older fish (SAC-16-INF-F). The new growth curve has a large impact on the spawning biomass ratio and the fishing mortality, as the stock is estimated to be less depleted and subject to lower fishing mortality in the EPO\_base-1 and NE\_base-1 compared to the same models with the former growth curve.

The second change is in natural mortality. In 2020 benchmark assessment, the natural mortality was assumed to vary by age, it was larger for small fish following Hampton (2000), and vary between males and females based on the observed changes in sex ratio at length. In this benchmark assessment, M reduces with length following the Lorenzen curve (Lorenzen et al. 2022) with a logistic offset for females with the parameters estimated fitting to tagging data and sex-ratio data using a cohort analysis (SAC-16-INF-F). Both the intercept, and the shape parameter of the Lorenzen curve were estimated, as well as the logistic offset. The EPO\_base-1 model with the old M assumption shows similar SBR as the EPO\_M\_low-1 or NE\_M\_low-1 models, but higher fishing mortality.

The third change related to reproductive biology. In the 2020 benchmark assessment, the reproductive biology was based on (Schaefer, n.d.). A new study became available (Schaefer and Fuller, 2022b), which also allowed for the estimation of different maturity ogives for the NE and SW regions (SAC-16-INF-F). The models with new reproductive biology assumptions have higher SBR, but similar fishing mortality rates.

Finally, the effect of changing all three aspects of the biology (growth, natural mortality and reproductive biology) simultaneously were assessed. The models with the previous biological assumptions have lower SBR and higher fishing mortality than the current assumptions.

The “base” model for the 2020 benchmark assessment, which was the most conservative scenario in the 2020 risk analysis ([SAC-11-07](#)), is also shown in the figures, for reference. It is noteworthy that the fishing mortalities in that model have higher frequency of variation when compared to the current assessment models. This is the effect of selectivity assumptions. In the 2020 benchmark assessment the selectivities were irregular splines, as opposed to smooth double normal curves in the current assessment.

---

<sup>3</sup> There are no differences in the estimates between the SS3 version used in 2020 and the current version

## 7. RISK ANALYSIS

A risk analysis is implemented in this assessment by quantifying the probability of meeting the target and limit reference points specified in the IATTC harvest control rule from an ensemble of models that represent the uncertainty in the assessment. Resolution [C-16-02](#) defines target and limit reference points (RPs), expressed in terms of spawning biomass ( $S$ ) and fishing mortality ( $F$ ), for the tropical tuna species: bigeye, yellowfin, and skipjack. Those RPs, and the method used to compute them in this document, are described below, as is the harvest control rule (HCR) that implements them.

### 7.1. DEFINITION OF REFERENCE POINTS

#### 7.1.1. LIMIT

The spawning biomass limit reference point ( $S_{\text{Limit}}$ ) is the level of  $S$  that should be avoided as further depletion could endanger the sustainability of the stock. The interim  $S_{\text{Limit}}$  adopted by the IATTC in 2014 is the  $S$  that produces 50% of the virgin recruitment if the stock-recruitment relationship follows the Beverton-Holt function with a steepness of 0.75. This spawning biomass is equal to 0.077 of the equilibrium virgin spawning biomass (Maunder and Deriso 2014). The HCR requires action be taken if the probability ( $p$ ) of the spawning biomass at the beginning of 2020 ( $S_{\text{current}}$ ) being below  $S_{\text{Limit}}$  is greater than 10%. Thus, to provide management advice,  $S_{\text{current}}/S_{\text{Limit}}$  and the probability of this ratio being  $< 1$  (by assuming the probability distribution function for the ratio is normal) are calculated.

The fishing mortality limit reference point ( $F_{\text{Limit}}$ ) is the threshold of fishing mortality that should be avoided because fishing more intensively could endanger the sustainability of the stock. The interim  $F_{\text{Limit}}$  adopted by the IATTC in 2014 is the fishing mortality rate that, under equilibrium conditions, maintains  $S$  at  $S_{\text{Limit}}$ . The HCR requires action to be taken if the probability of the average fishing mortality during 2021-2023 ( $F_{\text{current}}$ ) being above  $F_{\text{Limit}}$  is greater than 10%. Thus, to provide management advice,  $F_{\text{current}}/F_{\text{Limit}}$ , and the probability of this ratio being  $> 1$  (by assuming the probability distribution function for the ratio is normal), are calculated.

#### 7.1.2. TARGET

The spawning biomass target reference point is the level of spawning biomass that should be achieved and maintained. In 2014 the IATTC adopted  $S_{\text{MSY}}$  (the spawning biomass that produces the MSY) as the target reference point defined as the dynamic MSY level ( $S_{\text{MSY}_d}$ ) in the HCR. Here,  $S_{\text{MSY}_d}$  is derived by projecting the population into the future under historical recruitment (bias adjusted) and a fishing mortality rate that produces MSY. The current value of  $S_{\text{MSY}_d}$  used to compute reference points for yellowfin is the last quarter's  $S$  in the projection period. To provide management advice,  $S_{\text{current}}/S_{\text{MSY}_d}$  and the probability that this ratio is  $< 1$  (by assuming the probability distribution function for the ratio is normal with a CV equal to that of  $F_{\text{current}}/F_{\text{MSY}}$ ) are calculated.

The fishing mortality target reference point is the level of fishing mortality that should be achieved and maintained. The IATTC adopted  $F_{\text{MSY}}$  (the fishing mortality rate that produces the MSY) in 2014 as the target reference point. Thus, to provide management advice,  $F_{\text{current}}/F_{\text{MSY}}$  and the probability that this ratio is  $> 1$  (by assuming the probability distribution function for the ratio is normal) are calculated, as is the inverse of  $F_{\text{current}}/F_{\text{MSY}}$  ( $F$  multiplier).

The dynamic MSY ( $\text{MSY}_d$ ) is also derived from the projection for  $S_{\text{MSY}_d}$  and is defined as the total fishery catches in the last four quarters of the projection)

### 7.1.3. PROXY TARGET REFERENCE POINTS

Proxy reference points are cited in the harvest control rules resolution, but no specific proxies were defined. The staff has recommended to use the F30% ([SAC-15-05](#)), which is the fishing mortality that drives the population to 30% of the spawning stock biomass with no fishing (dynamic SSB0 or SSBt, F=0). For illustration purposes, we compute here the Kobe plot using F30% and SSB30% as target reference points.

### 7.2. STOCK STATUS

The spawning biomass ratio for all four spatial-structure hypotheses has been above the limit reference point (Figures 18a -18b) for all the assessment periods, regardless of the level 2 and level 3 hypotheses. The ratio of the spawning biomass to the proxy target reference point  $30\%S_d$  has been above 1 for the most part for the assessment period, with the exceptions of some years for NE and NE\_short (Figures 19a-19b).

With respect to the IATTC interim target and limit reference points, all four spatial-structure hypotheses estimate the same general current stock status (Tables 5a-d and 6). The stock(s) is estimated to be well above the spawning biomass correspond to MSY ( $S_{MSY}$ ) and the staff's proposed MSY proxy  $S_{30\%}$  ([SAC-15-05](#)) with low probability of being below these. The fishing mortality is estimated to be well below the level corresponding to MSY and the MSY proxy  $F_{30\%}$  with low probability of being above these (Figures 21, 22, 23) . The assessment estimates zero probability that the spawning biomass or fishing mortality limit reference points have been breached (Figures 22, 23c, 24c) . The EPO model is the most optimistic.

The most pessimistic models are those with low natural mortality (Figures 24a-c, Table 5a-d). Some of these models estimate that the spawning biomass is below the  $S_{30\%}$  level and the fishing mortality is above the  $F_{30\%}$  level. The high natural mortality levels are generally the most optimistic.

The estimates of the SBR (the ratio of the spawning biomass to the virgin spawning biomass) corresponding to MSY are low (generally below 20%, Table 5) even though the highest fishing mortality is on older yellowfin. The value is higher with lower steepness of the stock-recruitment relationship and lower natural mortality. For example, the SW model with no relationship between stock size and recruitment (steepness equals 1) and high natural mortality has a value of 5% while the NE\_short model with steepness equal to 0.8 and low natural mortality has a value of 32%. The low level of SBR corresponding to MSY might be due to the assumptions about natural mortality declining with age (i.e., high M for juveniles).

#### EPO-wide models

The spawning biomass at the start of 2024 ranged from 1.21 to 2.13 times the spawning biomass at the proxy reference point of  $30\%S0_d$  and ranged from 1.11 to 11.2 times the biomass that produces the dynamic MSY, depending on the model (Table 5a). The fishing mortality of yellowfin in 2021-2023 ranged from 36% to 91% of the proxy fishing mortality reference point and 12% to 91% of the fishing mortality at MSY (Table 5, Figure 20a). The median of the joint probability density function for the ratios of the recent fishing mortality to the proxy and target referent points are 0.559 and 0.397 respectively and there is a 0.2% and 0.4% probability of breaching those reference points (Table 6, Figures 20b-21b). The median of current spawning biomass is 1.73 and 2.38 times the proxy and target biomass reference points and the probability of breaching those reference points is practically zero (Table 6). The overall probability of breaching the limit reference points is also zero for both spawning biomass and fishing mortality (Table 6). The most pessimistic model is the scenario that assumes steepness of 0.8 and low natural mortality (Tables 5a-d). Even for this scenario, the probabilities of breaching the biomass and fishing mortality limit reference points are zero and the spawning biomass has a 5% probability of being below the target RP and a 0.1% probability of being below  $30\%S0_d$ .

### **NE and NE\_short models**

The NE or the NE\_short scenarios provide the same conclusion that the stock has breached either the limit, nor the target (or proxy) reference points, but lead to a slightly less optimistic results than the EPO-wide model (Tables 5b-c). For the NE hypothesis, the spawning biomass at the start of 2024 ranged from 0.95 to 1.69 of 30% $S_{0\_d}$ , and from 0.89 to 7.98 of the  $S_{MSY\_d}$ . The fishing mortality of yellowfin in 2021-2023 ranged from 50% to 108% of the proxy fishing mortality reference point and 16% to 110% of the fishing mortality at MSY (Table 5, Figure 18). The probability of breaching the limit reference points is zero both overall and for the most pessimistic scenario. The probability of breaching the target or proxy reference points is 5.9% for fishing mortality and 4.4% for spawning biomass, respectively. Similar conclusions are drawn for the NE\_short models, which are slightly more optimistic.

### **SW models**

The SW area showed a slightly different stock status, although still optimistic, as no target or reference point was breached (Table 5d). The recent fishing mortality is on average 75.7% of the proxy fishing mortality target reference point and a 16% probability of being higher. The recent fishing mortality is on average half the fishing mortality that produces MSY and a 7.5% probability of being higher. The estimated spawning biomass at the start of 2024 ranged from 0.93 to 2.08 of 30% $S_{0\_d}$  and 0.84 to 13.73 of the  $S_{MSY\_d}$ , and on average was 1.46 of 30% $S_{0\_d}$  and 2.2 of  $S_{MSY\_d}$ . The probability that the spawning biomass was lower than the limit reference point was zero.

## **8. FUTURE DIRECTIONS**

A main uncertainty in the stock assessment of yellowfin tuna in the EPO continues to be the spatial structure. This assessment showed that different areas in the EPO may have different depletion levels. The values used for natural mortality and the reliance on size composition data to inform absolute abundance remain key sources of uncertainty. Growth, especially at older ages, relied on a few high-quality tag returns. All four of these sources of uncertainty could be reduced by a comprehensive tagging program.

### **ACKNOWLEDGEMENTS**

This assessment was only possible due to the diligent work of onboard observers, port samplers, data editors, data specialists, scientists and administrative staff from IATTC and CPCs that provided the data and performed research used in this assessment. The IATTC biology group headed by Dan Fuller was central to the 2025 benchmark assessment as it implemented the growth and reproductive biology studies as well as the tagging studies that were the basis for the updated estimates of growth, natural mortality and reproductive biology of yellowfin tuna in the EPO. Japan, Korea, Chinese Taipei and China collaborated with the staff to improve the longline indices of abundance, Japan and Korea provided the operational-level data used in the final estimates of the longline indices used to fit the SW models. The assessment team is grateful to Rick Methot, Ian Taylor and the SS3 team for constant improvements of the SS3 code. Paulina Llano coordinated the Spanish translation. Cristine Patnode assisted with the edition of the figures.

## REFERENCES

- Beverton, R.J.H., Holt, S.J., 1957. On the dynamics of exploited fish populations, *Gt. Britain Fishery Invest.*
- Carvalho, F., Punt, A.E., Chang, Y.-J., Maunder, M.N., Piner, K.R., 2017. Can diagnostic tests help identify model misspecification in integrated stock assessments? *Fish. Res.* 192, 28–40. <https://doi.org/10.1016/j.fishres.2016.09.018>
- 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., Wetzels, C.R., Doering, K., Johnson, K.F., Methot, R.D., 2021. A cookbook for using model diagnostics in integrated stock assessments. *Fish. Res.* 240, 105959. <https://doi.org/10.1016/j.fishres.2021.105959>
- Francis, R.I.C.C., 2011. Data weighting in statistical fisheries stock assessment models. *Can. J. Fish. Aquat. Sci.* 68, 1124–1138. <https://doi.org/10.1139/f2011-025>
- Hilborn, R., Mangel, M., 1997. *The ecological detective: confronting models with data*, Monographs in population biology. Princeton university press, Princeton.
- Lee, H.-H., Maunder, M.N., Piner, K.R., Methot, R.D., 2012. Can steepness of the stock–recruitment relationship be estimated in fishery stock assessment models? *Fish. Res.* 125–126, 254–261. <https://doi.org/10.1016/j.fishres.2012.03.001>
- Lee, H.-H., Piner, K.R., Methot, R.D., 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. *Fish. Res.* 158, 138–146. <https://doi.org/10.1016/j.fishres.2013.12.017>
- Majumdar, A., Lennert-Cody, C.E., Maunder, M.N., Aires-da-Silva, A., 2023. spatio-temporal modeling for estimation of bigeye tuna catch in the presence of pandemic-related data loss using parametric adjacency structures. *Fish. Res.* 268, 106813. <https://doi.org/10.1016/j.fishres.2023.106813>
- Martell, S., Stewart, I., 2014. Towards defining good practices for modeling time-varying selectivity. *Fish. Res.* 158, 84–95. <https://doi.org/10.1016/j.fishres.2013.11.001>
- Maunder, M.N., Deriso, R.B., 2013. A stock–recruitment model for highly fecund species based on temporal and spatial extent of spawning. *Fish. Res.* 146, 96–101. <https://doi.org/10.1016/j.fishres.2013.03.021>
- Maunder, M.N., Thorson, J.T., 2019. Modeling temporal variation in recruitment in fisheries stock assessment: A review of theory and practice. *Fish. Res.* 217, 71–86. <https://doi.org/10.1016/j.fishres.2018.12.014>
- Methot, R.D., Taylor, I.G., 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. *Can. J. Fish. Aquat. Sci.* 68, 1744–1760. <https://doi.org/10.1139/f2011-092>
- Minami, M., Lennert-Cody, C.E., 2024. Regression Tree and Clustering for Distributions, and Homogeneous Structure of Population Characteristics. *J. Agric. Biol. Environ. Stat.* <https://doi.org/10.1007/s13253-024-00631-z>
- Schaefer, K.M., n.d. Reproductive biology of yellowfin tuna (*Thunnus albacares*) in the eastern Pacific Ocean. *Bull. Inter-Am. Trop. Tuna Comm.* 21, 205–272.
- Schaefer, K.M., Fuller, D.W., 2022a. Horizontal movements, utilization distributions, and mixing rates of yellowfin tuna (*THUNNUS ALBACARES*) tagged and released with archival tags in six discrete areas of the eastern and central Pacific Ocean. *Fish. Oceanogr.* 31, 84–107. <https://doi.org/10.1111/fog.12564>
- Schaefer, K.M., Fuller, D.W., 2022b. Spatiotemporal variability in the reproductive biology of yellowfin tuna (*Thunnus albacares*) in the eastern Pacific Ocean. *Fish. Res.* 248, 106225. <https://doi.org/10.1016/j.fishres.2022.106225>
- Taylor, I.G., Doering, K.L., Johnson, K.F., Wetzels, C.R., Stewart, I.J., 2021. Beyond visualizing catch-at-age models: Lessons learned from the r4ss package about software to support stock assessments. *Fish. Res.* 239, 105924. <https://doi.org/10.1016/j.fishres.2021.105924>

- Wang, S.-P., Maunder, M.N., Piner, K.R., Aires-da-Silva, A., Lee, H.-H., 2014. Evaluation of virgin recruitment profiling as a diagnostic for selectivity curve structure in integrated stock assessment models. *Fish. Res.* 158, 158–164. <https://doi.org/10.1016/j.fishres.2013.12.009>
- Waterhouse, L., Sampson, D.B., Maunder, M., Semmens, B.X., 2014. Using areas-as-fleets selectivity to model spatial fishing: Asymptotic curves are unlikely under equilibrium conditions. *Fish. Res.* 158, 15–25. <https://doi.org/10.1016/j.fishres.2014.01.009>
- Wild, A., 1986. Growth of yellowfin tuna, *Thunnus albacares*, in the eastern Pacific Ocean based on otolith increments. *Bull. Inter-Am. Trop. Tuna Comm.* 18, 423–480.
- Xu, H., Thorson, J.T., Methot, R.D., Taylor, I.G., 2019. A new semi-parametric method for autocorrelated age- and time-varying selectivity in age-structured assessment models. *Can. J. Fish. Aquat. Sci.* 76, 268–285. <https://doi.org/10.1139/cjfas-2017-0446>

**TABLES**

**TABLE 3.** The decisions for selectivity and composition data weighting according to each fishery’s catch amount and composition data quality. The rules on which this decision table is based are illustrated as a flowchart in Figure 6. “Double-normal” indicates whether the length composition data of the fleet can be fit well in the assessment model by using a double-normal selectivity curve, “trc” indicated that a retention curve was used to truncate the curve, “A” indicates that the selectivity was assumed asymptotic. “Data quality” indicates the relative quality of the fleet’s length composition data. “Time blocks” indicates whether and how the selectivity of the fleet is time-varying. “Weighting scaler” indicates how length composition data are weighted in comparison to the Francis weighting method.

#	Type	Fleet name	Catch amount	Double-normal	Data quality	Selectivity	Time blocks	Weighting scaler
1	Fishery	F1_PS_OBJ_North_coastal	low	no	low	Fixed <sup>3</sup>	no	0
2		F2_PS_OBJ_Core	high	yes, trc	high <sup>1</sup>	Estimate	2000; 2010	1
3		F3_PS_OBJ_Offshore	high	yes, trc	high <sup>1</sup>	Estimate	2010 <sup>2</sup>	1
4		F4_PS_OBJ_Galapagos	low	yes, trc	high <sup>1</sup>	Estimate	no	0.2
5		F5_PS_OBJ_South_coastal	low	no	low	Fixed <sup>3</sup>	no	0
6		F6_PS_NOA_North_coastal	high	yes	high	Estimate	no	1
7		F7_PS_NOA_Core_small	high	yes	low	Fixed <sup>3</sup>	no	0
8		F8_PS_NOA_Core_large	high	yes, trc, A	low	Fixed <sup>3</sup>	no	0
9		F9_PS_NOA_Offshore_small	low	yes	low	Fixed <sup>3</sup>	no	0
10		F10_PS_NOA_Offshore_large	low	yes, trc, A	low	Fixed <sup>3</sup>	no	0
11		F11_PS_NOA_Galapagos_small	high	Yes	low	Fixed <sup>3</sup>	no	0
12		F12_PS_NOA_Galapagos_large	high	yes, trc, A	low	Fixed <sup>3</sup>	no	0
13		F13_PS_NOA_South_coastal	low	No	low	Fixed <sup>3</sup>	no	0
14		F14_PS_DEL_North_coastal	high	Yes	high	Estimate	2000; 2010	1
15		F15_PS_DEL_Core	high	yes, A <sup>4</sup>	high	Estimate	2000; 2010	1
16		F16_PS_DEL_Offshore_South	high	no, A	high	Estimate	no	0.2
17		F17_PS_DEL_Galapagos	low	yes, A	high	Estimate	no	0.2
18		F18_PS_DIS_small_North_coastal	low	NA	NA	Fixed	no	NA
19		F19_PS_DIS_small_Core	low	NA	NA	Fixed	no	NA
20		F20_PS_DIS_small_Offshore	low	NA	NA	Fixed	no	NA
21		F21_PS_DIS_small_Galapagos	low	NA	NA	Fixed	no	NA
22		F22_PS_DIS_small_South_coastal		NA	NA	Fixed	no	NA
23		F23_LP	low	yes, A	high	Estimate	no	0.2
24		F24_LL_North_coastal_n	low	yes, A	no data	Mirror F26	no	NA
25		F25_LL_Core_n	low	yes, A	no data	Mirror F26	no	NA
26		F26_LL_Offshore_n	Low / High (SW)	yes, A	mixed	Fixed / Estimate (SW)	no	0 0.2 (SW)
27		F27_LL_Galapagos_n	low	yes, A	no data	Mirror F26	no	NA
28		F28_LL_South_coastal_n	Low / High (SW)	yes, A	mixed	Fixed / Estimate (SW)	no 2010 (SW)	0 0.2 (SW)
29		F29_LL_North_n	low	yes, A	no data	Mirror F26	no	NA
30		F30_LL_Polynesia_n	Low / High (SW)	yes	mixed	Fixed / Estimate (SW)	no	0 0.2 (SW)
31		F31_LL_North_coastal_w		yes, A		Mirror F24		NA
32		F32_LL_Core_w		yes, A	NA	Mirror F25		NA
33		F33_LL_Offshore_w		yes, A		Mirror F26		NA
34		F34_LL_Galapagos_w		yes, A		Mirror F27		NA
35		F35_LL_South_coastal_w		yes, A		Mirror F28		NA

#	Type	Fleet name	Catch amount	Double-normal	Data quality	Selectivity	Time blocks	Weighting scaler	
36		F36_LL_North_w		yes, A		Mirror F29		NA	
37		F37_LL_Polynesia_w				Mirror F30		NA	
38	Survey	S1_EPO	NA	Yes, A <sup>4</sup>		Estimated		0 / 1 (EPO)	
39		S2_PS_NCoastal		Yes, A		Estimated in spatial model		0 / 1 (NC, spatial)	
40		S3_PS_Core		yes, A <sup>4</sup>		Estimated NE, NE_short, spatial model <sup>5</sup>		0 / 1 (NE, NE_short, spatial)	
41		S4_PS_Offshore				Estimated spatial model		0 / 1 (spatial)	
42		S5_LL_Offshore		yes, A		Estimated SW		0 / 1 (SW)	

1 Occasional large sizes, length >100 cm removed

2 Data before 1993 is sparse and was not used

3 Fixed at an early estimate as done in ASPM

4 Age and length selectivity curves were used, increase the flexibility in shapes curves

5 The index for NE and NE\_short model had a spatial domain that comprised the Core and NC area, for the spatial model, the index covered only the core area.

**TABLE 4.** The diagnostics metrics for all models. Gradient is the final gradient of the assessment model, Jitter is the number of jittered starting values for each round until the model passed the jitter test. All models have a positive definite Hessian matrix.

Number	Level 1 Structure	Level 2 Biology/q	Level 3 h	estimated lnR0	parameters	NLL	Gradient	Jitter
1	EPO	base	1	16.1802	249	1531.6	0.00003	20
2		MI	1	12.8659	249	1534.4	0.00011	20 + 10
3		Mh	1	19.5440	249	1531.8	0.00011	20
4		GI	1	16.2418	249	1526.5	0.00009	20 + 10
5		Gh	1	16.1584	249	1533.9	0.00005	20
6		q1	1	16.2594	249	1552.4	0.00005	20 + 10 +10
7	EPO	base	0.9	16.2119	249	1530.4	0.00598	20
8		MI	0.9	12.9094	249	1531.1	0.00006	20 + 10 +10
9		Mh	0.9	19.5642	249	1530.4	0.00094	20
10		GI	0.9	16.2492	249	1530.4	0.00003	20 + 10
11		Gh	0.9	16.2357	249	1532.6	0.00008	20
12		q1	0.9	16.2861	249	1550.0	0.00063	20 + 10 +10
13	EPO	base	0.8	16.2882	249	1524.5	0.00008	21 + 10 +10
14		MI	0.8	12.9762	249	1527.7	0.00006	22 + 10 +10
15		Mh	0.8	19.5918	249	1528.9	0.02541	20
16		GI	0.8	16.2961	249	1523.5	0.00004	20 + 10
17		Gh	0.8	16.2291	249	1531.0	0.00004	20
18		q1	0.8	16.3074	249	1552.7	0.00047	20
19	NE	base	1	15.6582	238	1174.3	0.00004	20
20		MI	1	12.4259	238	1166.2	0.00376	20
21		Mh	1	18.9084	238	1183.6	0.00021	20
22		GI	1	15.6502	238	1177.0	0.00016	20
23		Gh	1	15.6545	238	1174.2	0.00003	20
24		q1	1	15.6991	238	1206.2	0.00007	20
25	NE	base	0.9	15.7032	238	1171.8	0.00015	20
26		MI	0.9	12.4933	238	1162.9	0.00559	20
27		Mh	0.9	18.9405	238	1181.5	0.02244	22 + 10 +10
28		GI	0.9	15.6953	238	1174.4	0.03919	22 + 10 +10
29		Gh	0.9	15.6997	238	1171.7	0.00077	20
30		q1	0.9	15.7401	238	1202.6	0.00002	20 + 10
31	NE	base	0.8	15.7677	238	1169.1	0.00007	20 + 10
32		MI	0.8	12.5921	238	1159.6	0.00066	20
33		Mh	0.8	18.9853	238	1179.2	0.04916	20
34		GI	0.8	15.7602	238	1171.7	0.00057	20
35		Gh	0.8	15.7642	238	1169.0	0.00098	20
36		q1	0.8	15.7979	238	1198.9	0.00008	20
37	NE_short	base	1	15.5175	135	442.8	0.00010	20 + 10
38		MI	1	12.2618	135	438.8	0.00008	20
39		Mh	1	18.8090	135	447.4	0.00006	20 + 10

Number	Level 1 Structure	Level 2 Biology/q	Level 3 h	estimated lnR0	parameters	NLL	Gradient	Jitter
40		Gl	1	15.5079	135	446.8	0.00003	20
41		Gh	1	15.5209	135	441.5	0.00003	20 + 10
42		q1	1	15.5083	135	444.9	0.00007	20 + 10
43	NE_short	base	0.9	15.5607	135	442.6	0.00002	20 + 10
44		MI	0.9	12.3278	135	438.5	0.00008	20
45		Mh	0.9	18.8381	135	447.2	0.00001	20
46		Gl	0.9	15.5514	135	446.6	0.00014	20
47		Gh	0.9	15.5627	135	441.4	0.00005	20 + 10
48		q1	0.9	15.5501	135	444.9	0.00010	20
49	NE_short	base	0.8	15.6213	135	442.4	0.00003	20 + 10
50		MI	0.8	12.4229	135	438.4	0.00113	20
51		Mh	0.8	18.8784	135	447.0	0.00003	20
52		Gl	0.8	15.6126	135	446.4	0.00005	20
53		Gh	0.8	15.6231	135	441.1	0.00010	20
54		q1	0.8	15.6105	135	444.7	0.00069	20 + 10
55	SW	base	1	14.9326	208	924.8	0.00001	20
56		MI	1	11.5782	208	937.4	0.00005	20
57		Mh	1	18.3939	208	915.6	0.00008	20
58		Gl	1	14.9449	208	912.4	0.00009	20
59		Gh	1	14.9338	208	937.8	0.00003	20
60		q1	1	14.9248	208	940.8	0.00001	20
61	SW	base	0.9	14.9610	208	924.8	0.00006	10
62		MI	0.9	11.6272	208	937.5	0.00590	10
63		Mh	0.9	18.4100	208	915.6	0.00009	10
64		Gl	0.9	14.9723	208	912.4	0.00002	10
65		Gh	0.9	14.9624	208	937.8	0.00009	10
66		q1	0.9	14.9533	208	940.9	0.00375	10
67	SW	base	0.8	14.9999	208	925.0	0.00010	10
68		MI	0.8	11.6960	208	937.8	0.00001	10
69		Mh	0.8	18.4318	208	915.7	0.00002	10
70		Gl	0.8	15.0099	208	912.6	0.00003	10
71		Gh	0.8	15.0017	208	938.0	0.00004	10
72		q1	0.8	14.9924	208	941.1	0.00005	10

**TABLE 5a.** Management table for yellowfin tuna in the EPO.  $S_{current}$ ,  $S_0$ ,  $S_{MSY\_d}$ : spawning biomass (metric tons) at the beginning of 2024, in a unfished equilibrium state, and at dynamic MSY, respectively;  $F_{current}$ ,  $F_{MSY}$ : and  $F_{30\%S_0\_d}$  fishing mortality between 2021-2023, at MSY, and at level that takes the population to 30% of dynamic spawning biomass without fishing, respectively;  $S_{LIMIT}$  and  $F_{LIMIT}$ : limit reference points for spawning biomass and fishing mortality, respectively;  $C_{current}$ : total catch of yellowfin in 2023 (metric tons); MSY\_d: dynamic MSY; p(): probability.

	B			Gh			Gl			Mh			MI			Q1		
	1	0.9	0.8	1	0.9	0.8	1	0.9	0.8	1	0.9	0.8	1	0.9	0.8	1	0.9	0.8
<b>EPO</b>																		
MSY (1,000 t)	376	337	325	364	344	307	404	352	330	539	442	399	297	282	275	407	363	332
MSY_d (1,000 t)	416	372	353	409	380	337	448	388	359	578	478	425	341	318	307	431	376	342
$C_{current}/MSY\_d$	0.72	0.80	0.85	0.73	0.79	0.89	0.67	0.77	0.83	0.52	0.62	0.70	0.88	0.94	0.97	0.69	0.79	0.87
$S_{MSY}/S_0$	0.06	0.22	0.27	0.11	0.22	0.28	0.10	0.22	0.27	0.06	0.19	0.25	0.19	0.26	0.31	0.10	0.22	0.28
$S_{current}/S_0$	0.61	0.59	0.59	0.60	0.61	0.56	0.64	0.62	0.60	0.73	0.72	0.70	0.48	0.45	0.43	0.57	0.55	0.53
$S_{current}/S_{LIMIT}$	7.88	7.66	7.63	7.77	7.86	7.23	8.28	8.02	7.83	9.51	9.33	9.11	6.21	5.87	5.55	7.39	7.20	6.92
$p(S_{current} < S_{LIMIT})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$F_{current}/F_{LIMIT}$	0.21	0.25	0.27	0.22	0.25	0.30	0.19	0.23	0.27	0.14	0.16	0.19	0.31	0.37	0.43	0.21	0.24	0.29
$p(F_{current} > F_{LIMIT})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$S_{current}/S_{MSY\_d}$	7.83	2.17	1.78	4.57	2.23	1.62	5.16	2.30	1.82	11.2								
										1	3.26	2.38	1.97	1.35	1.11	4.83	2.23	1.70
$p(S_{current} < S_{MSY\_d})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00
$F_{current}/F_{MSY}$	0.19	0.45	0.54	0.26	0.43	0.61	0.23	0.42	0.53	0.12	0.26	0.36	0.54	0.77	0.93	0.25	0.44	0.59
$p(F_{current} > F_{MSY})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00
$S_{current}/30\%S_0$	1.75	1.70	1.70	1.72	1.75	1.59	1.86	1.78	1.74	2.13	2.09	2.04	1.37	1.29	1.21	1.77	1.71	1.62
$p(S_{current} < 30\%S_0)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$F_{current}/F_{30\%S_0}$	0.55	0.58	0.59	0.56	0.56	0.65	0.49	0.54	0.57	0.36	0.39	0.42	0.80	0.86	0.91	0.54	0.57	0.63
$p(F_{current} > F_{30\%S_0})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00

**TABLE 5b.** Management table for yellowfin tuna in the NE.  $S_{current}$ ,  $S_0$ ,  $S_{MSY\_d}$ : spawning biomass (metric tons) at the beginning of 2024, in a unfished equilibrium state, and at dynamic MSY, respectively;  $F_{current}$ ,  $F_{MSY}$ : and  $F_{30\%S_0\_d}$  fishing mortality between 2021-2023, at MSY, and at level that takes the population to 30% of dynamic spawning biomass without fishing, respectively;  $S_{LIMIT}$  and  $F_{LIMIT}$ : limit reference points for spawning biomass and fishing mortality, respectively;  $C_{current}$ : total catch of yellowfin in 2023 (metric tons); MSY\_d: dynamic MSY; p(): probability.

	B			Gh			Gl			Mh			MI			Q1		
	1	0.9	0.8	1	0.9	0.8	1	0.9	0.8	1	0.9	0.8	1	0.9	0.8	1	0.9	0.8
<b>NE</b>																		
MSY (1,000 t)	226	208	200	225	208	200	226	208	200	276	237	220	200	195	196	236	216	205
MSY_d (1,000 t)	238	213	199	237	212	199	238	213	199	284	245	222	209	195	192	235	209	195
$C_{current}/MSY\_d$	0.91	1.01	1.09	0.91	1.02	1.09	0.91	1.01	1.08	0.76	0.88	0.97	1.04	1.11	1.13	0.92	1.03	1.11
$S_{MSY}/S_0$	0.11	0.23	0.28	0.11	0.23	0.28	0.10	0.23	0.28	0.07	0.20	0.26	0.19	0.26	0.31	0.10	0.23	0.28
$S_{current}/S_0$	0.45	0.42	0.40	0.45	0.43	0.40	0.44	0.42	0.39	0.54	0.53	0.50	0.35	0.32	0.29	0.41	0.39	0.37
$S_{current}/S_{LIMIT}$	5.79	5.51	5.15	5.80	5.53	5.16	5.76	5.48	5.11	7.07	6.83	6.52	4.49	4.17	3.76	5.26	5.04	4.75
$p(S_{current} < S_{LIMIT})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$F_{current}/F_{LIMIT}$	0.24	0.29	0.34	0.24	0.28	0.34	0.24	0.29	0.35	0.17	0.21	0.25	0.33	0.39	0.47	0.25	0.30	0.35
$p(F_{current} > F_{LIMIT})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$S_{current}/S_{MSY\_d}$	3.96	1.71	1.32	3.77	1.70	1.31	4.17	1.71	1.32	7.98	2.52	1.83	1.65	1.13	0.89	4.00	1.71	1.32
$p(S_{current} < S_{MSY\_d})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.97	0.00	0.00	0.00
$F_{current}/F_{MSY}$	0.29	0.57	0.75	0.30	0.57	0.75	0.29	0.57	0.75	0.17	0.37	0.51	0.61	0.87	1.10	0.30	0.58	0.75
$p(F_{current} > F_{MSY})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.92	0.00	0.00	0.00
$S_{current}/30\%S_d$	1.41	1.35	1.27	1.42	1.35	1.27	1.41	1.34	1.26	1.69	1.64	1.58	1.12	1.05	0.95	1.36	1.30	1.23
$p(S_{current} < 30\%S_{MSY\_d})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.21	0.83	0.00	0.00	0.00
$F_{current}/F_{30\%S_d}$	0.68	0.73	0.79	0.67	0.73	0.79	0.68	0.73	0.80	0.50	0.53	0.58	0.92	0.99	1.08	0.69	0.74	0.81
$p(F_{current} > F_{30\%S_d})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.41	0.88	0.00	0.00	0.00

**TABLE 5c.** Management table for yellowfin tuna in the NE\_short.  $S_{current}$ ,  $S_0$ ,  $S_{MSY\_d}$ : spawning biomass (metric tons) at the beginning of 2024, in a unfished equilibrium state, and at dynamic MSY, respectively;  $F_{current}$ ,  $F_{MSY}$ : and  $F_{30\%S_0\_d}$  fishing mortality between 2021-2023, at MSY, and at level that takes the population to 30% of dynamic spawning biomass without fishing, respectively;  $S_{LIMIT}$  and  $F_{LIMIT}$ : limit reference points for spawning biomass and fishing mortality, respectively;  $C_{current}$ : total catch of yellowfin in 2023 (metric tons); MSY\_d: dynamic MSY; p(): probability.

	B			Gh			Gl			Mh			MI			Q1		
	1	0.9	0.8	1	0.9	0.8	1	0.9	0.8	1	0.9	0.8	1	0.9	0.8	1	0.9	0.8
<b>NE_short</b>																		
MSY (1,000 t)	195	180	173	195	180	173	195	180	173	251	214	197	168	165	167	193	178	171
MSY_d (1,000 t)	235	216	203	235	217	203	235	216	203	290	260	237	204	193	189	229	210	198
$C_{current}/MSY\_d$	0.92	1.00	1.07	0.92	1.00	1.06	0.92	1.00	1.06	0.75	0.83	0.91	1.06	1.12	1.14	0.94	1.03	1.09
$S_{MSY}/S_0$	0.11	0.24	0.29	0.10	0.23	0.29	0.11	0.24	0.29	0.07	0.20	0.26	0.21	0.28	0.32	0.11	0.24	0.29
$S_{current}/S_0$	0.59	0.56	0.53	0.58	0.56	0.53	0.59	0.57	0.54	0.72	0.70	0.68	0.45	0.42	0.39	0.55	0.53	0.50
$S_{current}/S_{LIMIT}$	7.60	7.30	6.90	7.57	7.27	6.86	7.66	7.36	6.97	9.36	9.12	8.79	5.84	5.49	5.02	7.12	6.84	6.47
$p(S_{current} < S_{LIMIT})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$F_{current}/F_{LIMIT}$	0.21	0.25	0.30	0.22	0.26	0.31	0.21	0.25	0.30	0.15	0.18	0.21	0.31	0.36	0.43	0.23	0.27	0.32
$p(F_{current} > F_{LIMIT})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$S_{current}/S_{MSY\_d}$	4.26	1.81	1.42	4.51	1.82	1.42	4.09	1.82	1.43	9.22	2.76	2.00	1.61	1.16	0.95	4.06	1.74	1.35
$p(S_{current} < S_{MSY\_d})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.84	0.00	0.00	0.00
$F_{current}/F_{MSY}$	0.27	0.52	0.68	0.26	0.52	0.68	0.27	0.51	0.67	0.14	0.31	0.43	0.61	0.85	1.06	0.28	0.54	0.71
$p(F_{current} > F_{MSY})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.83	0.00	0.00	0.00
$S_{current}/30\%S_d$	1.54	1.50	1.43	1.53	1.49	1.43	1.55	1.51	1.45	1.87	1.83	1.78	1.22	1.16	1.08	1.48	1.43	1.37
$p(S_{current} < 30\%S_{MSY\_d})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.10	0.00	0.00	0.00
$F_{current}/F_{30\%S_d}$	0.61	0.65	0.70	0.61	0.65	0.71	0.60	0.64	0.69	0.42	0.45	0.49	0.86	0.92	1.00	0.64	0.68	0.74
$p(F_{current} > F_{30\%S_d})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.47	0.00	0.00	0.00

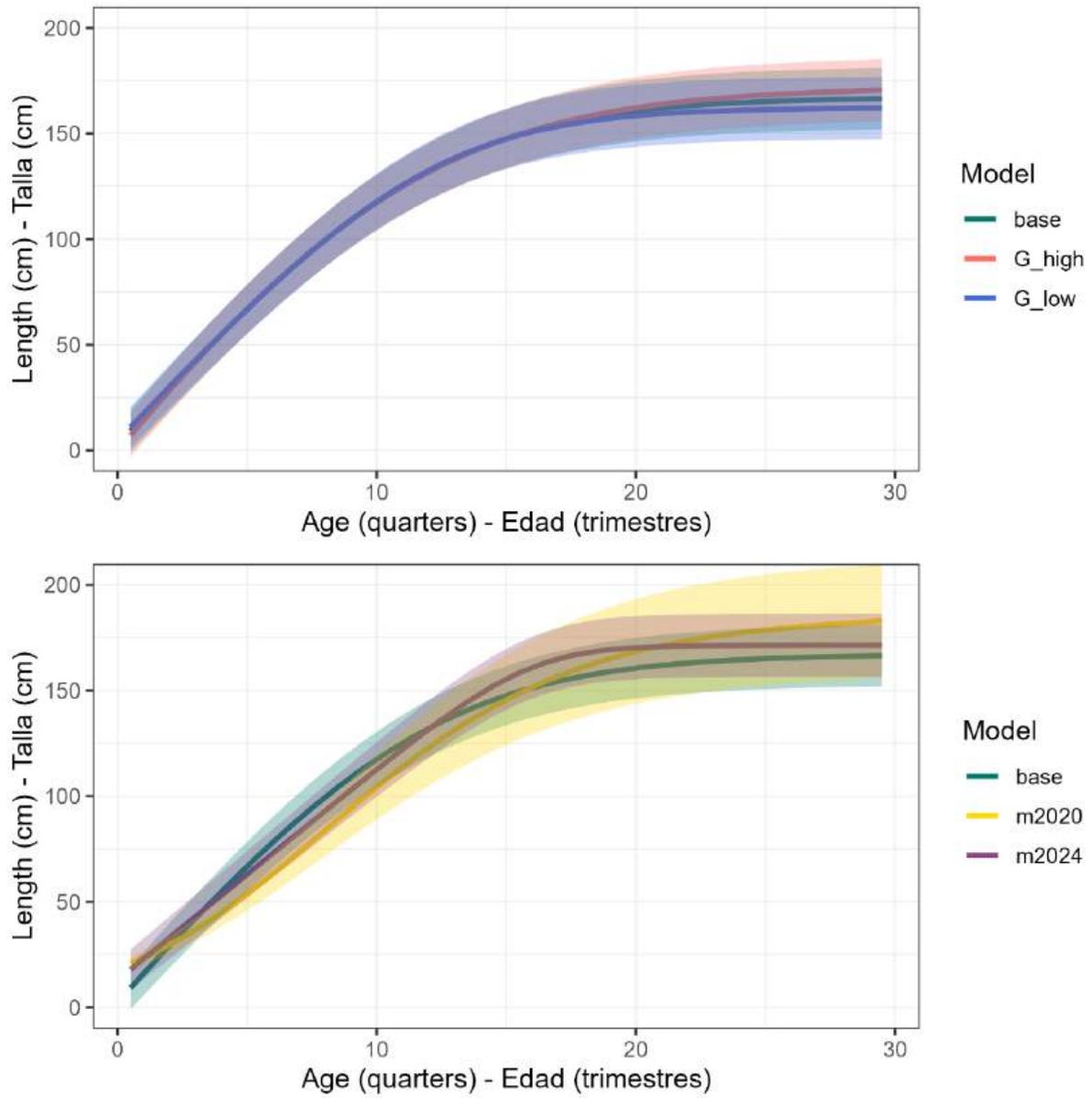
**TABLE 5d.** Management table for yellowfin tuna in the SW.  $S_{current}$ ,  $S_0$ ,  $S_{MSY\_d}$ : spawning biomass (metric tons) at the beginning of 2024, in a unfished equilibrium state, and at dynamic MSY, respectively;  $F_{current}$ ,  $F_{MSY}$ : and  $F_{30\%S_0\_d}$  fishing mortality between 2021-2023, at MSY, and at level that takes the population to 30% of dynamic spawning biomass without fishing, respectively;  $S_{LIMIT}$  and  $F_{LIMIT}$ : limit reference points for spawning biomass and fishing mortality, respectively;  $C_{current}$ : total catch of yellowfin in 2023 (metric tons); MSY\_d: dynamic MSY; p(): probability.

	B			Gh			Gl			Mh			MI			Q1		
	1	0.9	0.8	1	0.9	0.8	1	0.9	0.8	1	0.9	0.8	1	0.9	0.8	1	0.9	0.8
<b>SW</b>																		
MSY (1,000 t)	105	92	85	104	91	84	110	94	87	176	136	121	77	74	72	106	92	85
MSY_d (1,000 t)	110	104	99	104	102	98	111	107	102	168	141	129	93	93	94	104	100	95
$C_{current}/MSY\_d$	0.75	0.79	0.83	0.79	0.80	0.84	0.74	0.77	0.81	0.49	0.58	0.64	0.88	0.88	0.88	0.79	0.82	0.86
$S_{MSY}/S_0$	0.09	0.21	0.27	0.05	0.21	0.27	0.05	0.21	0.27	0.04	0.17	0.24	0.18	0.26	0.31	0.05	0.21	0.27
$S_{current}/S_0$	0.61	0.59	0.57	0.60	0.58	0.56	0.63	0.61	0.59	0.77	0.76	0.74	0.45	0.43	0.40	0.53	0.51	0.49
$S_{current}/S_{LIMIT}$	7.92	7.69	7.39	7.80	7.58	7.28	8.18	7.95	7.64	10.0	9.87	9.65	5.91	5.62	5.25	6.85	6.61	6.31
$p(S_{current} < S_{LIMIT})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$F_{current}/F_{LIMIT}$	0.29	0.34	0.39	0.30	0.34	0.40	0.29	0.33	0.38	0.17	0.20	0.23	0.46	0.53	0.61	0.32	0.37	0.44
$p(F_{current} > F_{LIMIT})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$S_{current}/S_{MSY\_d}$	4.77	2.03	1.50	9.64	1.97	1.46	10.0	2.12	1.56	13.7	3.56	2.45	1.72	1.07	0.84	8.62	1.84	1.34
$p(S_{current} < S_{MSY\_d})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0.97	0.00	0.00	0.01
$F_{current}/F_{MSY}$	0.32	0.57	0.74	0.24	0.58	0.76	0.23	0.54	0.71	0.14	0.28	0.39	0.72	1.05	1.31	0.26	0.63	0.83
$p(F_{current} > F_{MSY})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.70	0.99	0.00	0.00	0.03
$S_{current}/30\%S_d$	1.54	1.48	1.42	1.52	1.47	1.40	1.58	1.53	1.46	2.08	2.04	1.98	1.06	1.00	0.93	1.41	1.35	1.27
$p(S_{current} < 30\%S_{MSY\_d})$	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.30	0.51	0.78	0.00	0.01	0.02
$F_{current}/F_{30\%S_d}$	0.71	0.75	0.81	0.72	0.76	0.82	0.69	0.73	0.79	0.41	0.44	0.47	1.11	1.18	1.27	0.79	0.84	0.91
$p(F_{current} > F_{30\%S_d})$	0.00	0.00	0.02	0.00	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.00	0.84	0.93	0.98	0.01	0.04	0.17

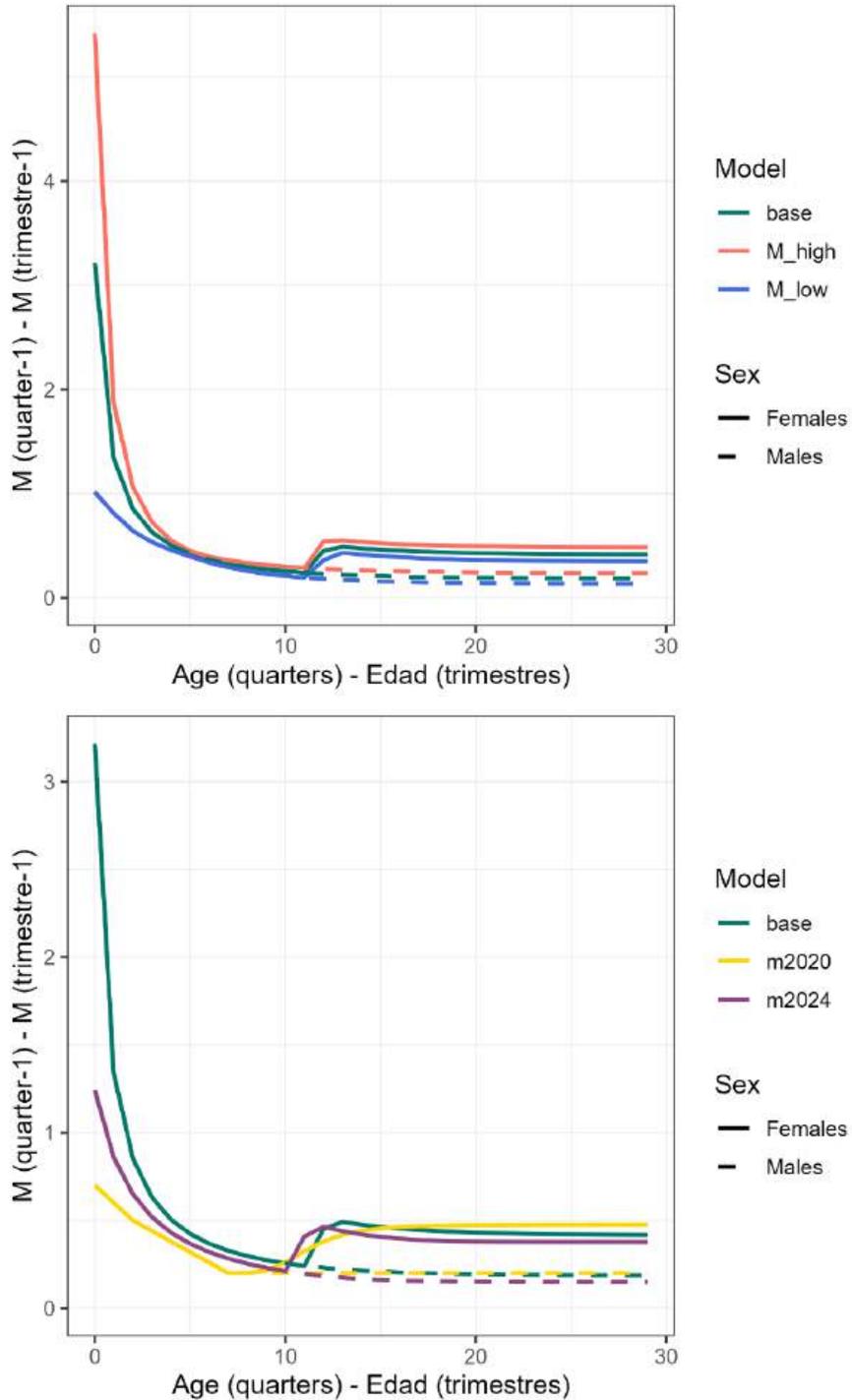
**TABLE 6.** Management quantities for yellowfin tuna in the EPO for each spatial structure hypothesis. The medians (or expected values \*) and probabilities were obtained from the joint probability distributions across models

	<b>EPO</b>	<b>NE</b>	<b>NE_short</b>	<b>SW</b>
$SMSY/SO$ *	0.180	0.189	0.194	0.162
$SMSY_d/SO_d$ *	0.190	0.192	0.201	0.170
$F_{current}/F_{30\%SO_d}$	0.559	0.718	0.643	0.757
$p(F_{current} > F_{30\%SO_d})$	0.002	0.059	0.020	0.161
$F_{current}/F_{MSY}$	0.397	0.532	0.484	0.502
$p(F_{current} > F_{MSY})$	0.004	0.034	0.031	0.075
$F_{current}/F_{LIMIT}$	0.232	0.272	0.243	0.330
$p(F_{current} > F_{LIMIT})$	0.000	0.000	0.000	0.000
$S_{current}/30\%SO_d$	1.73	1.35	1.49	1.46
$p(S_{current} < 30\%SO_d)$	0.0000588	0.044	0.004	0.081
$S_{current}/SMSY_d$	2.38	1.82	1.91	2.22
$p(S_{current} < SMSY_d)$	0.000	0.000	0.000	0.000
$S_{current}/S_{LIMIT}$	7.67	5.43	7.23	7.48
$p(S_{current} < S_{LIMIT})$	0.000	0.000	0.000	0.000

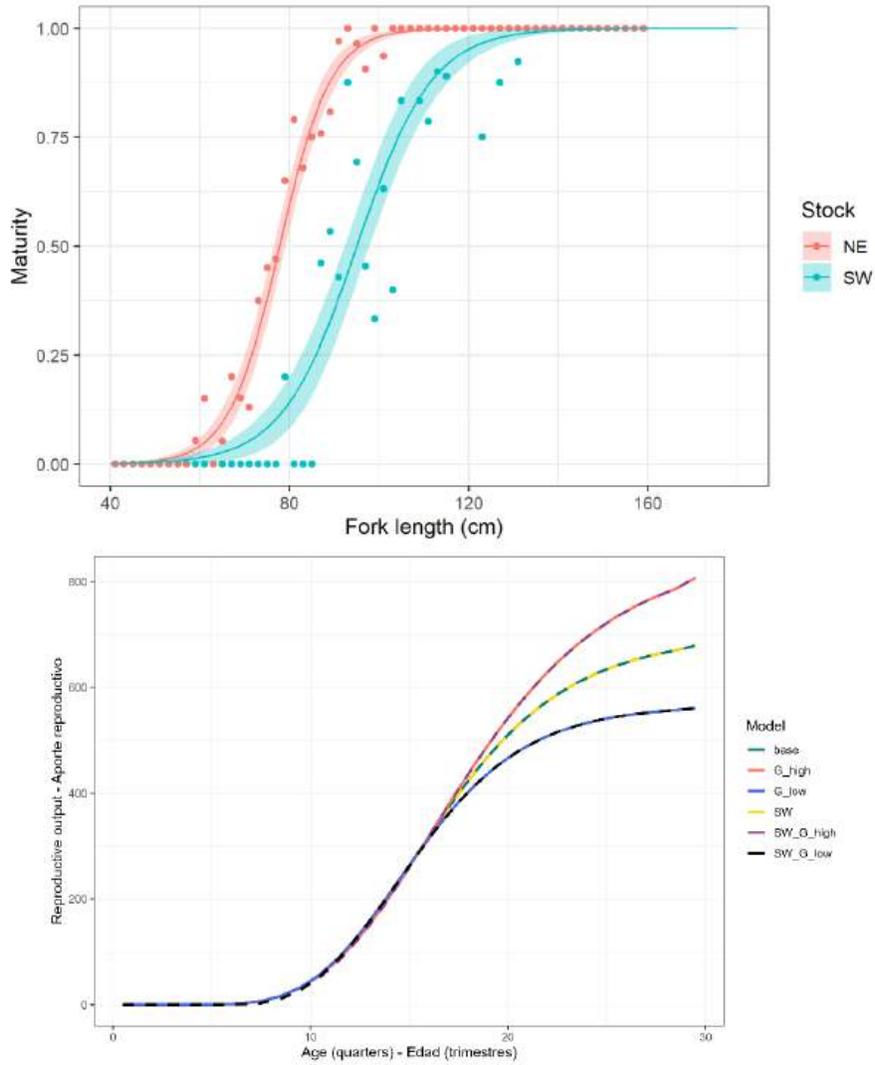
**FIGURES**



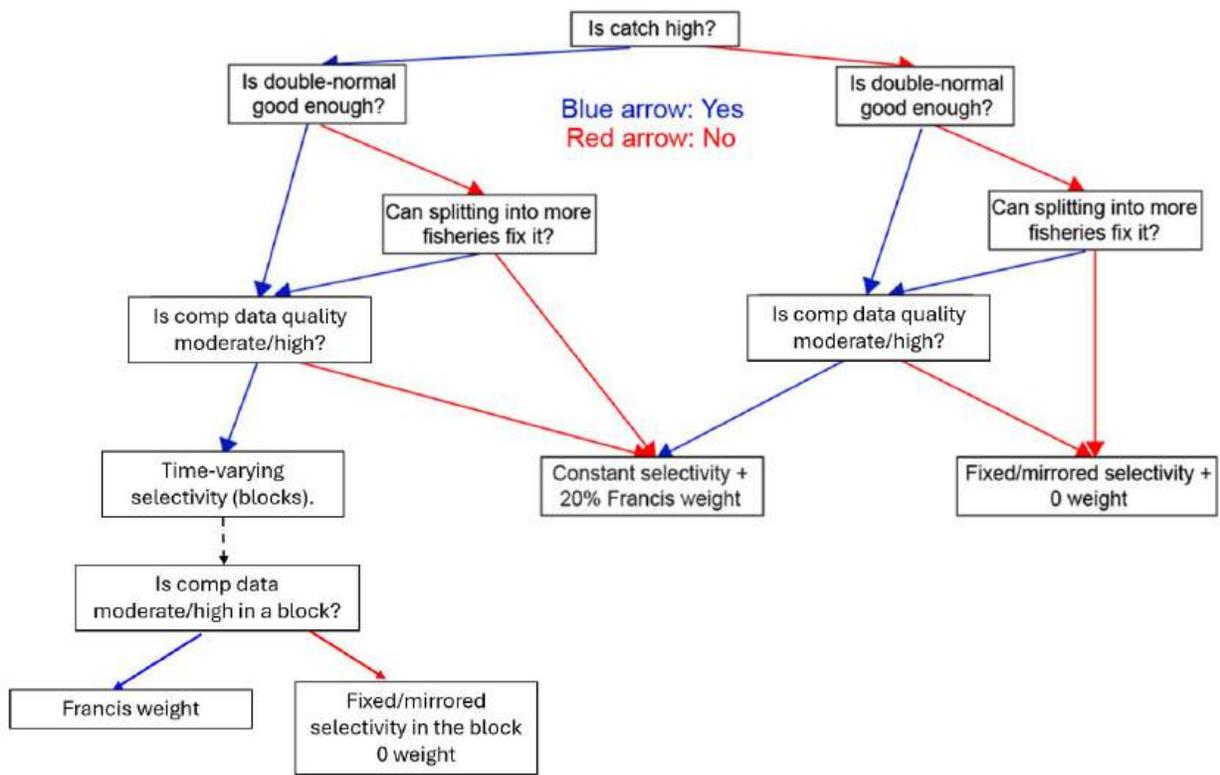
**FIGURE 3.** Growth curves for yellowfin tuna in the EPO. Top: three growth assumptions used in this assessment. Bottom: comparison of the base assumption with the assumptions used in the 2020 benchmark assessment (base) and 2024 exploratory analysis.



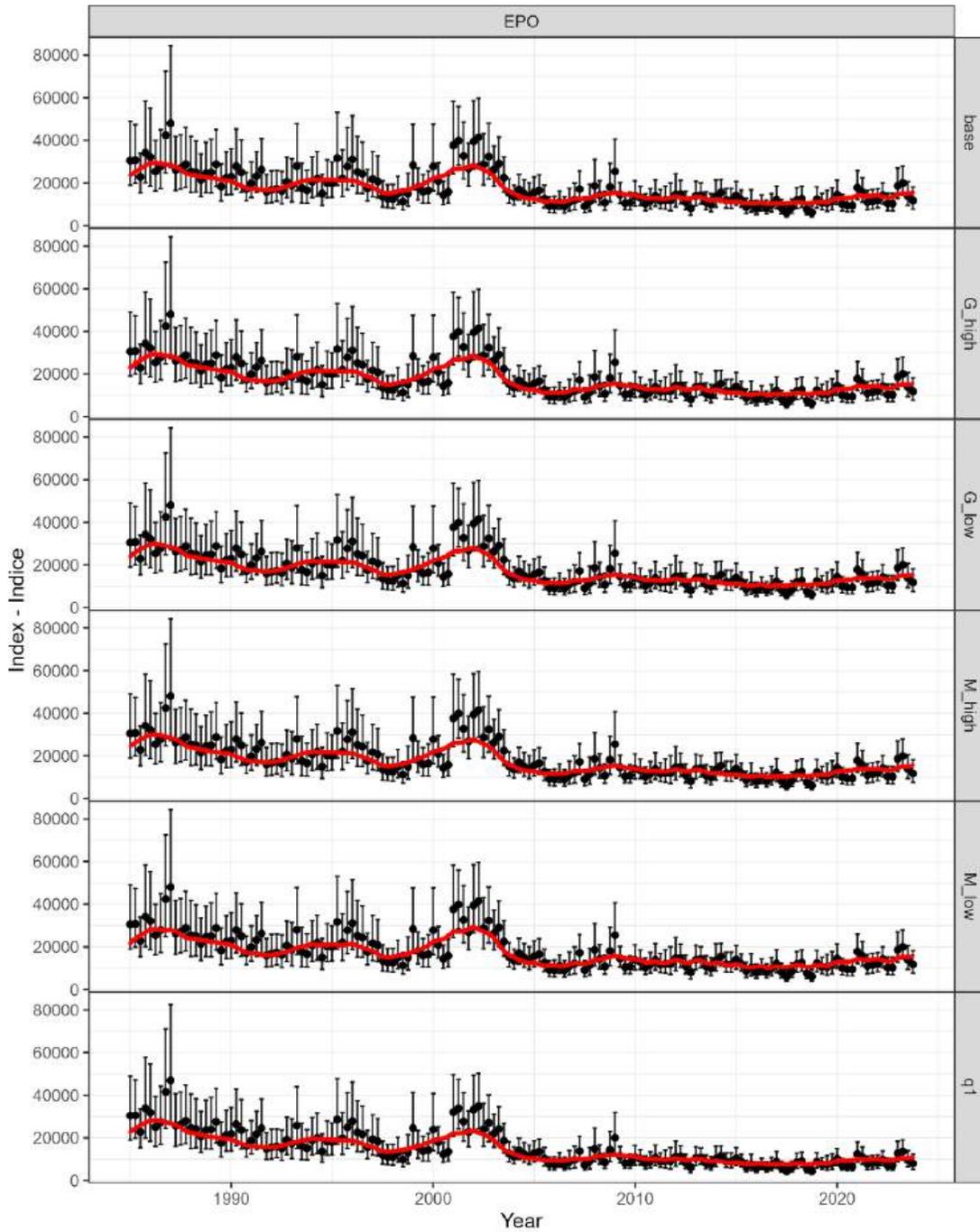
**FIGURE 4.** Natural mortality curves for yellowfin tuna in the EPO. Top: three assumptions used in this assessment. Bottom: comparison of the base assumption with the assumptions used in the 2020 benchmark assessment (base reference model) and 2024 exploratory analysis.



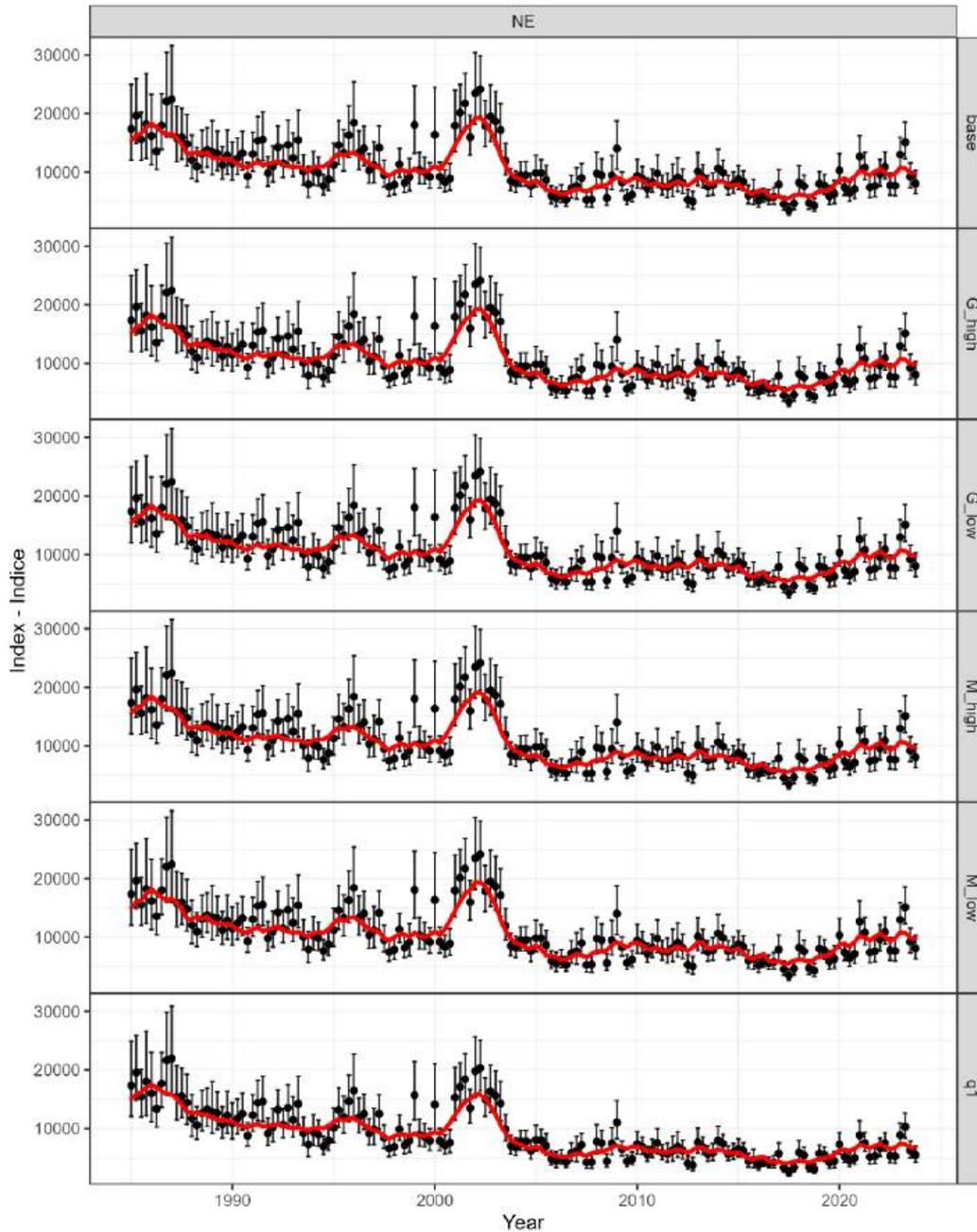
**FIGURE 5.** Top: Maturity at length estimated for yellowfin tuna in the EPO. Bottom:reproductive output at age for yellowfin tuna in the EPO (combination of batch fecundity at length, maturity at length, frequency of spawning at length and length at age).



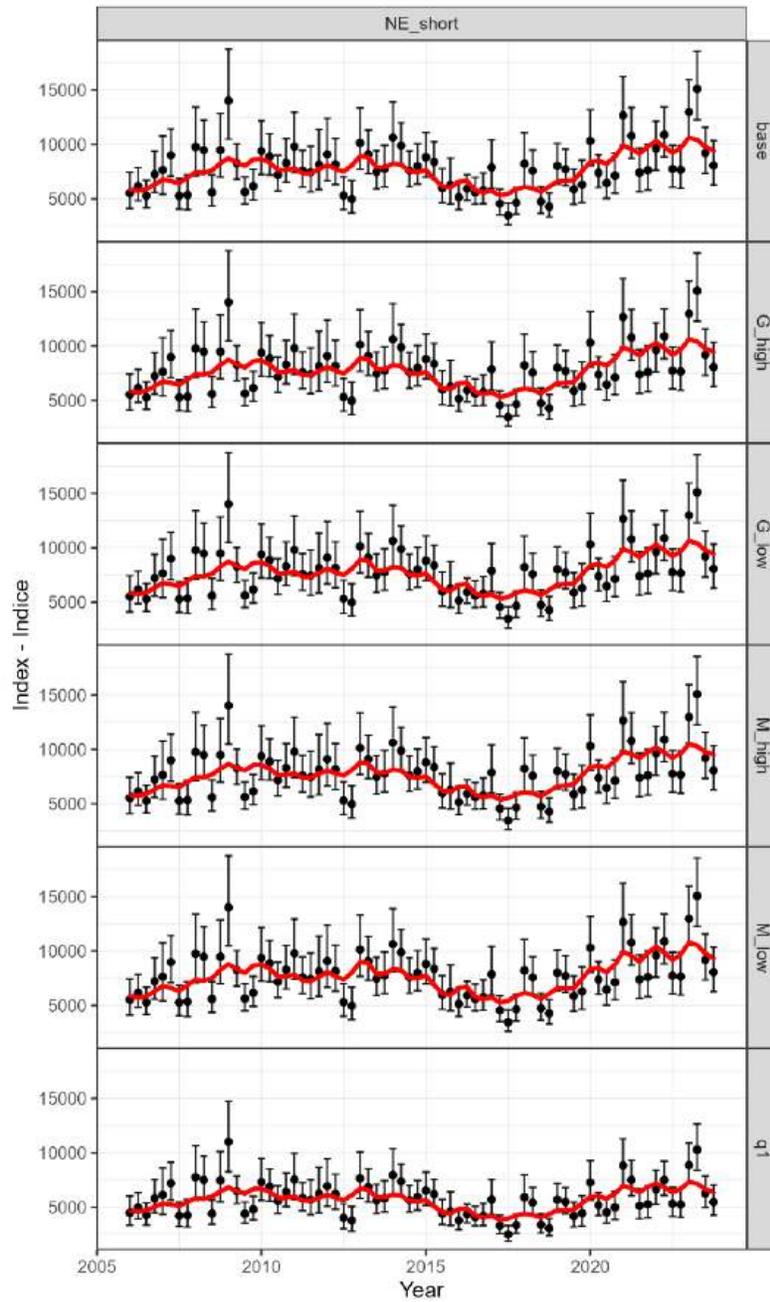
**FIGURE 6.** The decision tree on which the selectivity form and composition data weighting in this benchmark assessment are based on SAC-15-02 with a modification indicated by the dashed arrow.



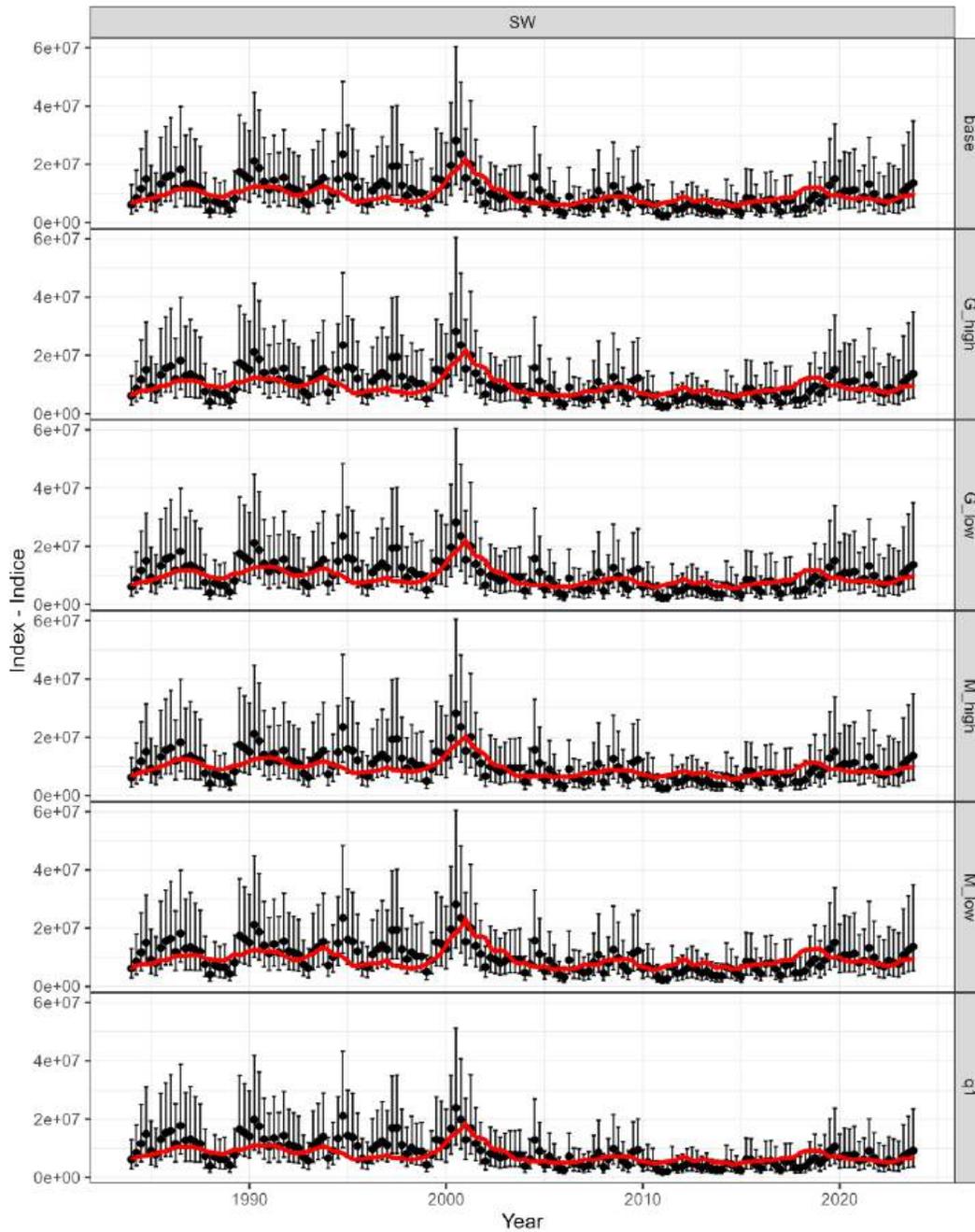
**FIGURE 7a.** Fit to the indices of relative abundance for EPO, steepness =1 . The black dots and error bars represent the observed values and their 95% confidence interval. The solid red lines are predicted values from the stock assessment model.



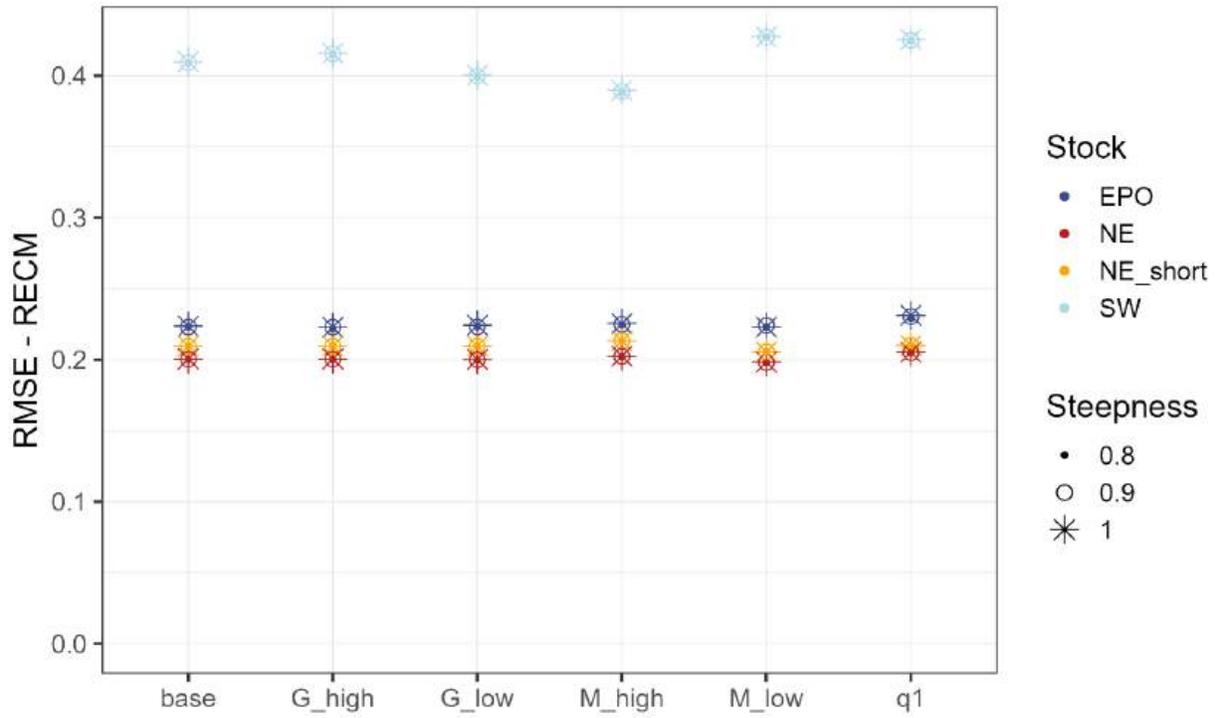
**FIGURE 7b.** Fit to the indices of relative abundance for NE, steepness= 1. The black dots and error bars represent the observed values and their 95% confidence interval. The solid red lines are predicted values from the stock assessment model.



**FIGURE 7c.** Fit to the indices of relative abundance for NE\_short, steepness = 1. The black dots and error bars represent the observed values and their 95% confidence interval. The solid red lines are predicted values from the stock assessment model.

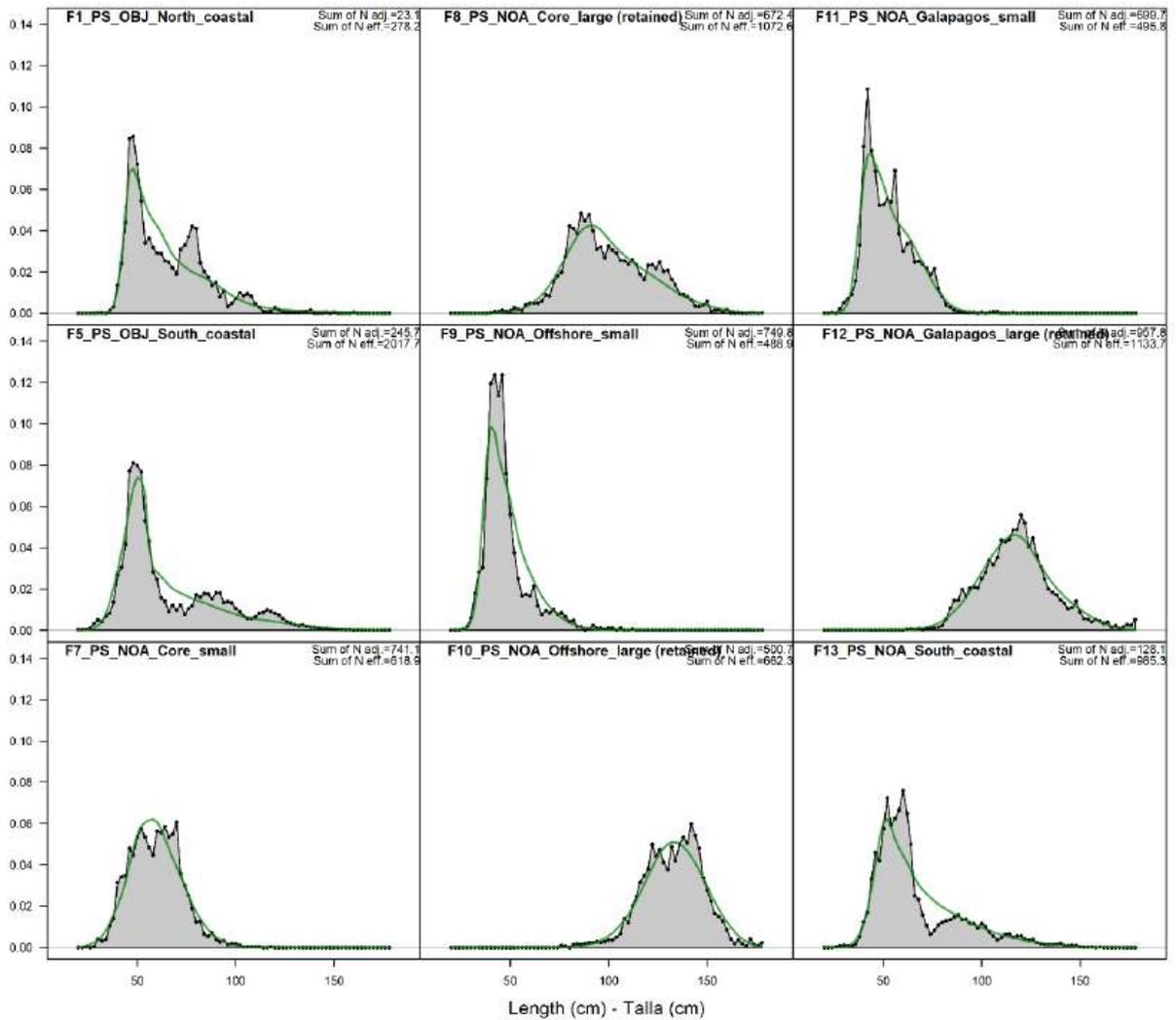


**FIGURE 7d.** Fit to the indices of relative abundance for SW, steepness = 1. The black dots and error bars represent the observed values and their 95% confidence interval. The solid red lines are predicted values from the stock assessment model.



**FIGURE 7e.** Root mean square error (RMSE) for the indices for each model.

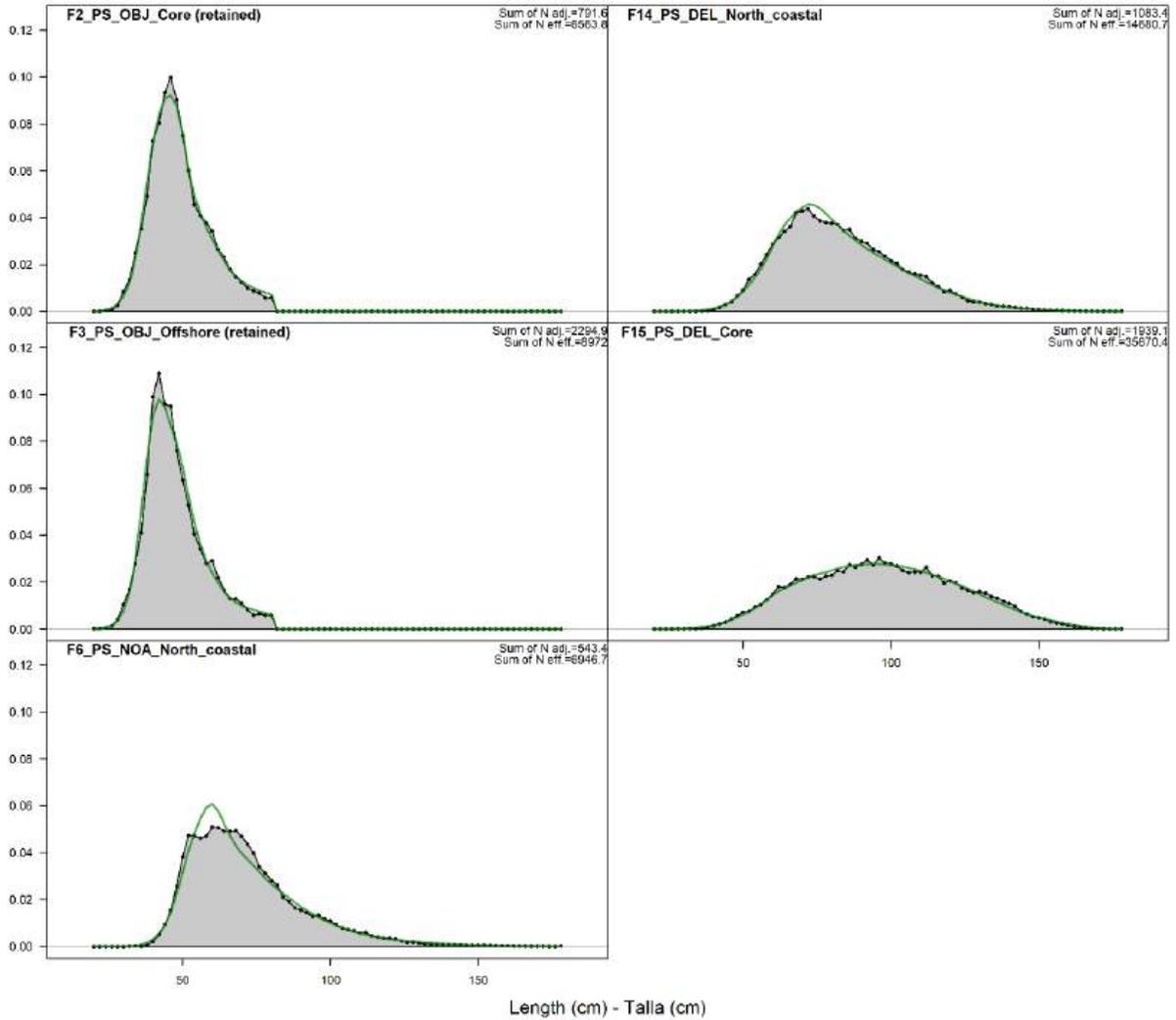
Length comps, aggregated across time by fleet



**FIGURE 8a.** Weighted average observed (shaded area) and predicted length-composition data by the ancestral model for purse-seine fleets that had 0 weight in the models used in the risk analysis and had selectivity curves fixed to the values estimated in the ancestral model

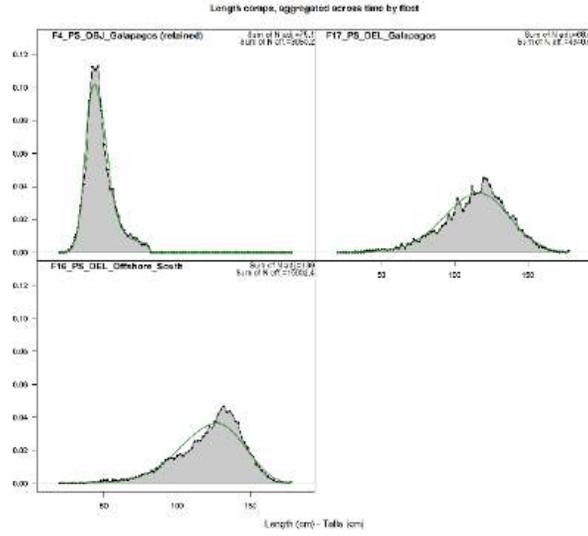
EPO-base-1

Length comps, aggregated across time by fleet

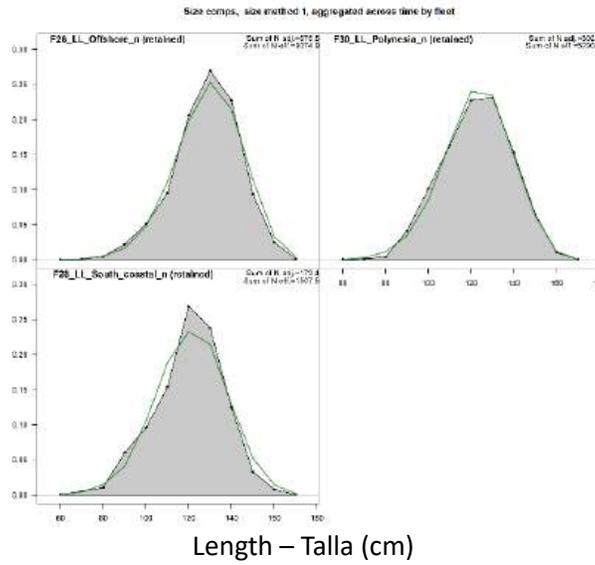


**FIGURE 8b.** Weighted average observed (shaded area) and predicted by the EPO-base-1 (line) length-composition data, by purse-seine fishery, for fleets with weighting scaler = 1.

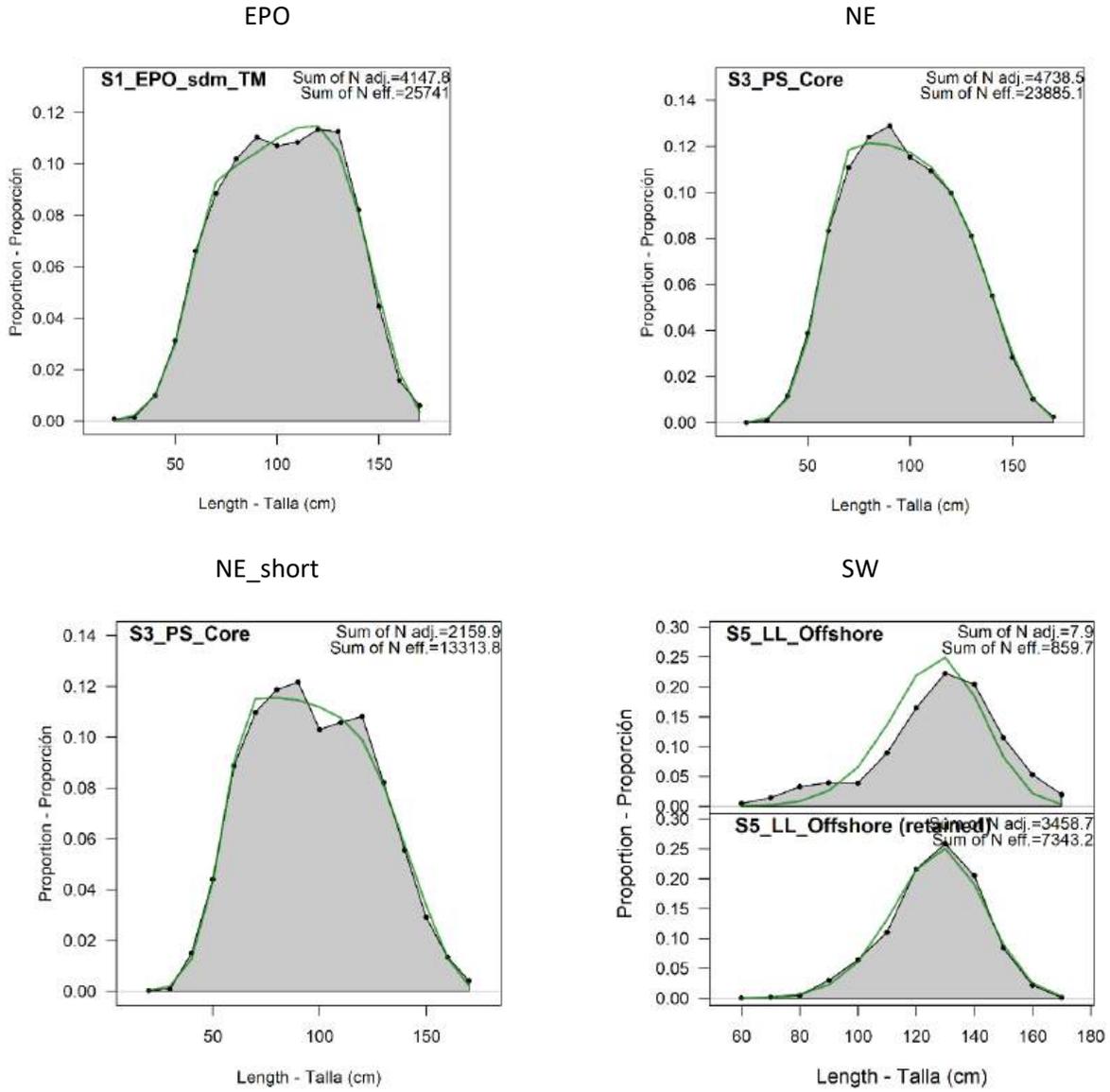
### EPO-base-1



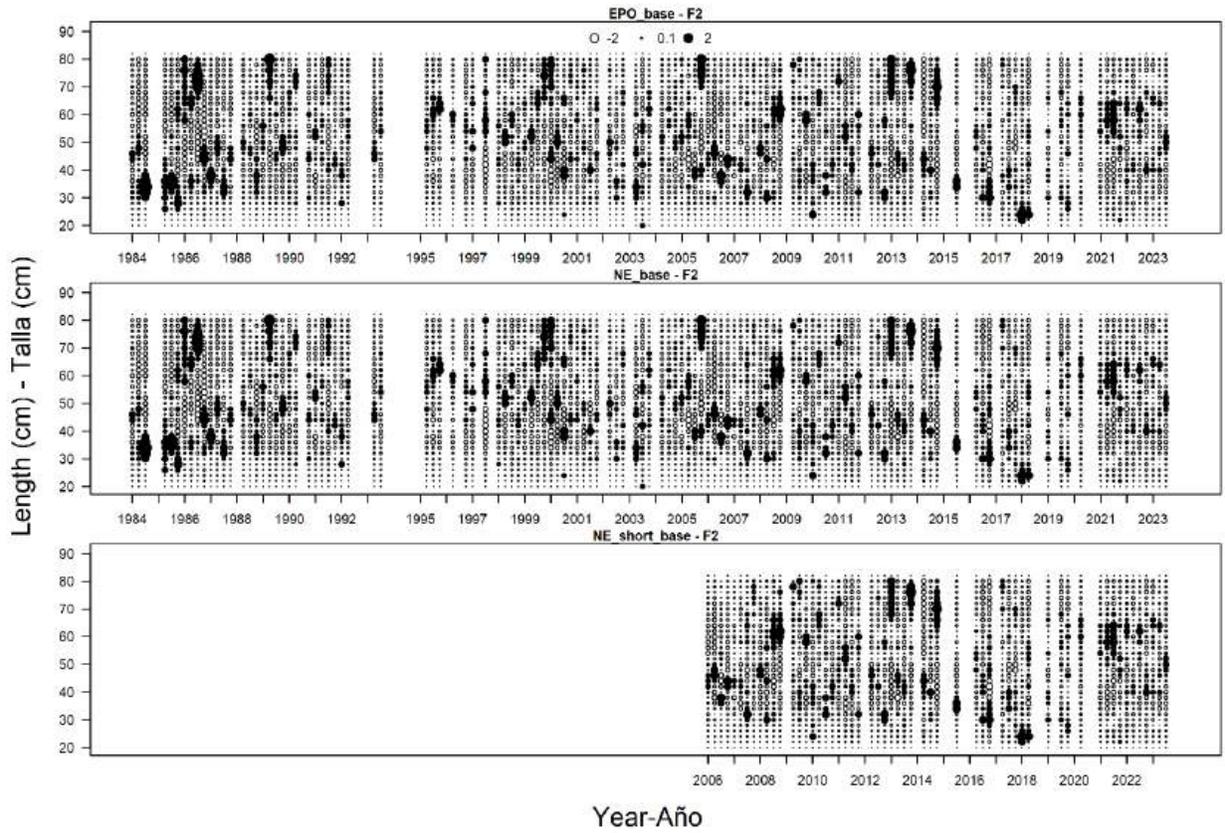
### SW-base-1



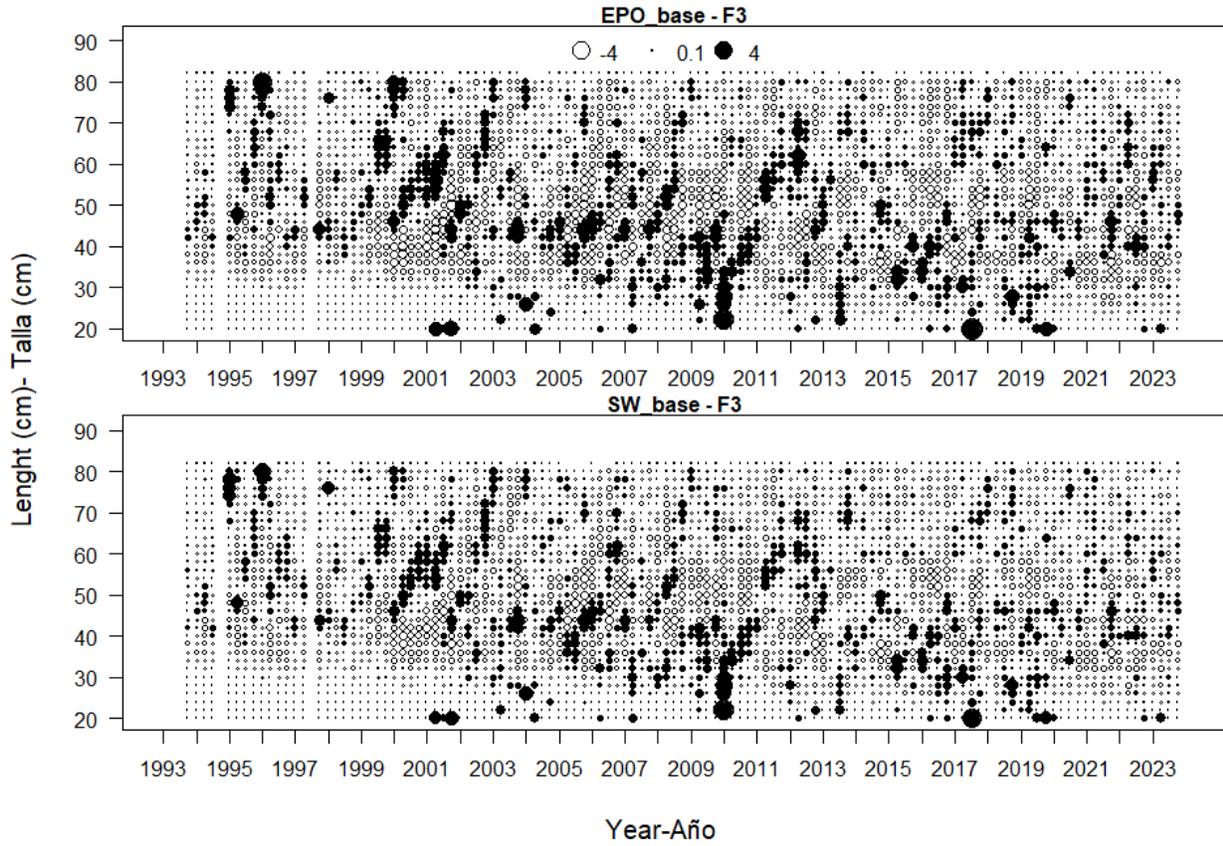
**FIGURE 8c.** Weighted average observed (shaded area) and predicted by the EPO-base-1 (line) length-composition data, by purse-seine fishery, for fleets with weighting scaler = 0.2.



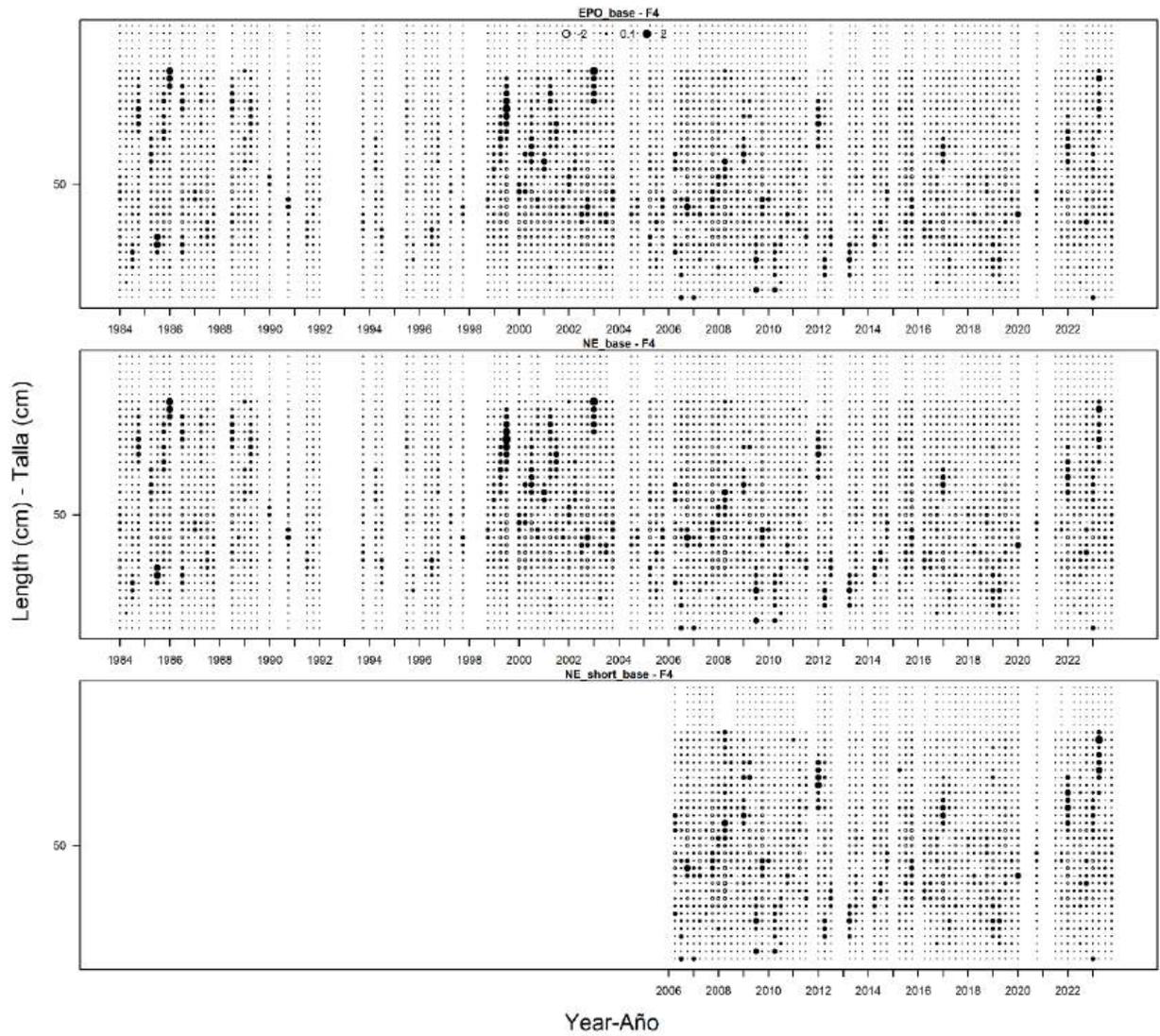
**FIGURE 8d.** Weighted average observed (shaded area) and predicted (line) length-composition data associated with the index of abundance for each stock structure hypothesis (base-1 models)



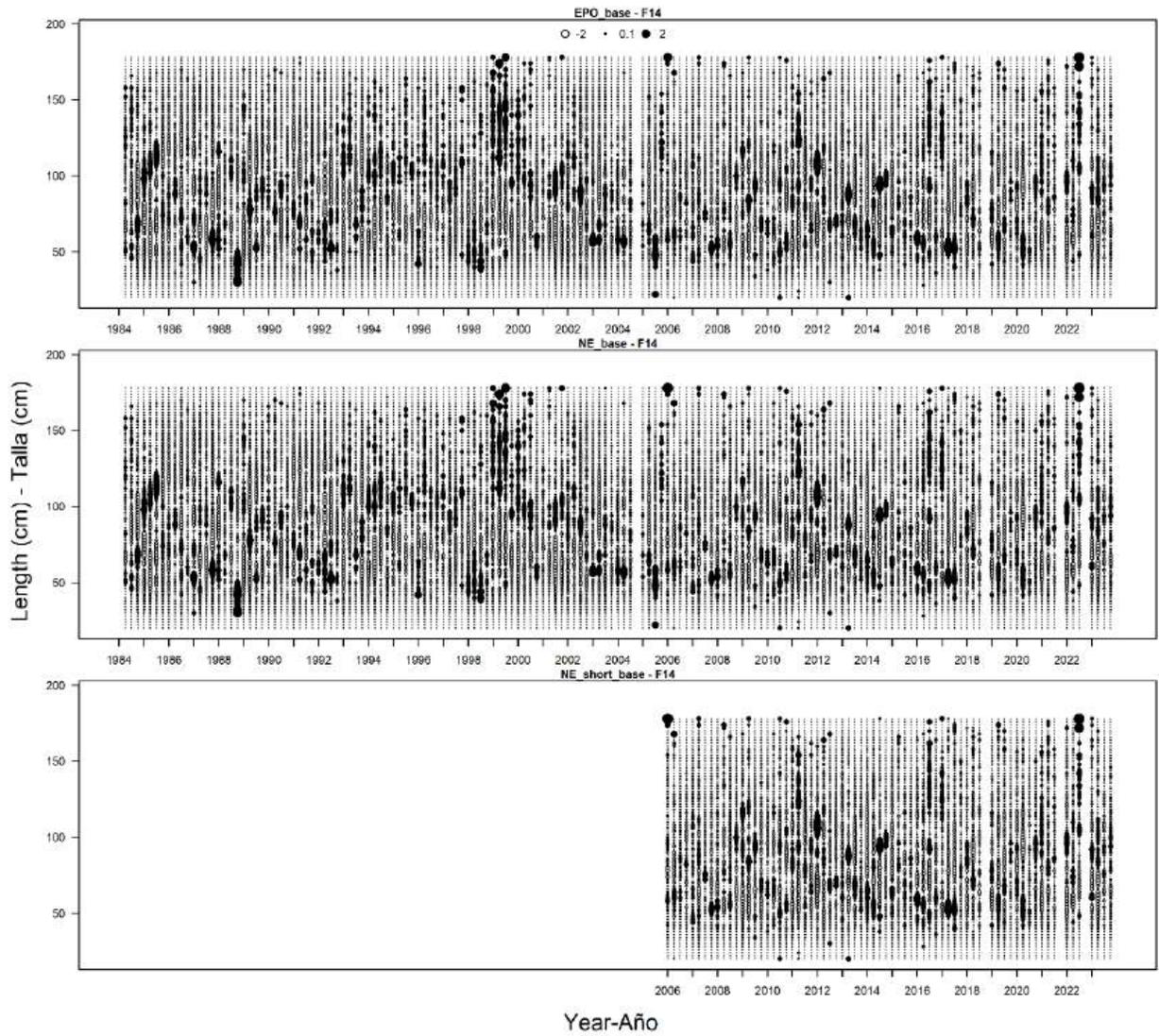
**FIGURE 9a.** Bubble plots: Pearson residuals for the length composition of the F2 OBJ fishery in the different models (with  $h=1$ ).



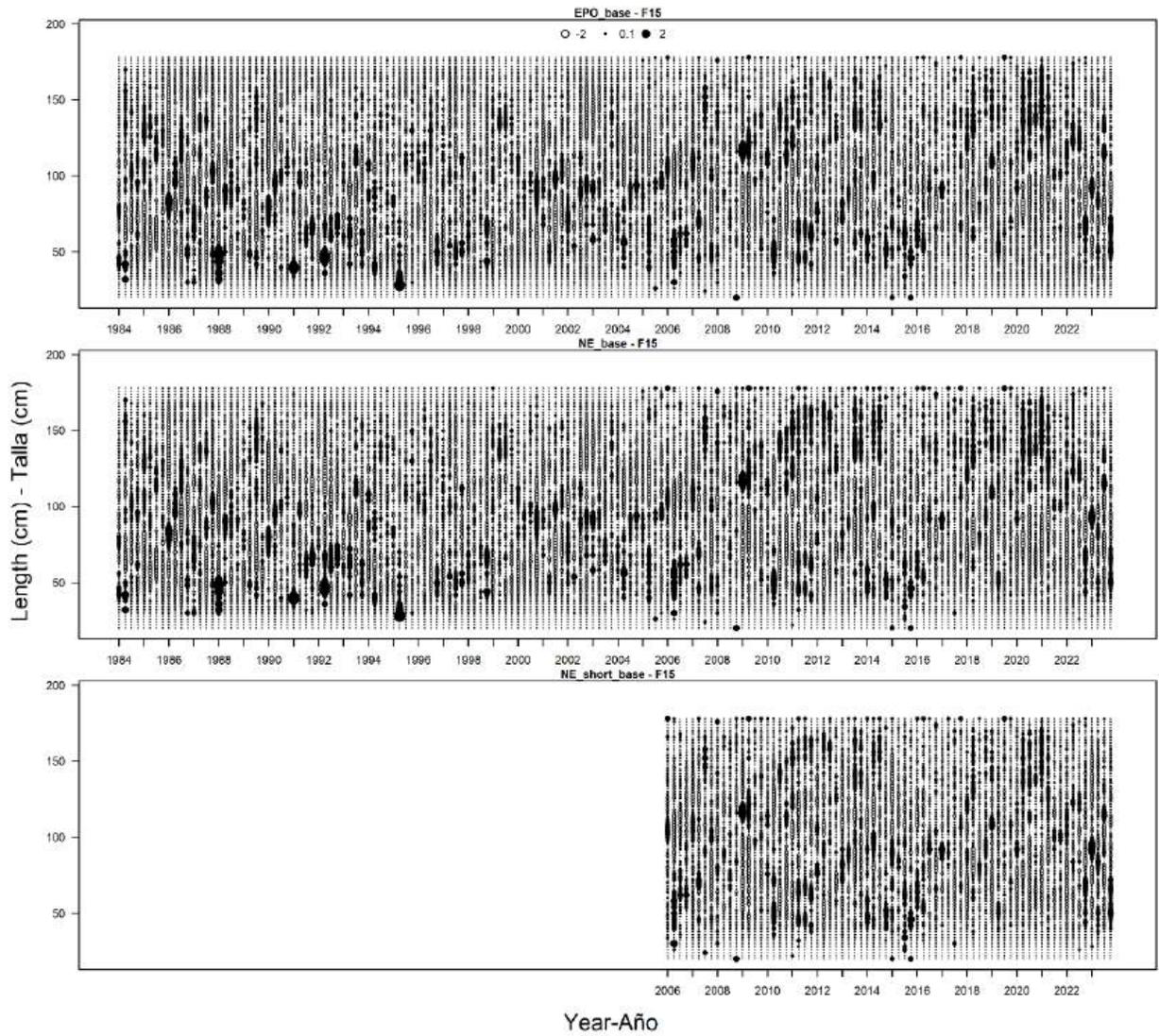
**FIGURE 9b.** Bubble plot: Pearson residuals for the length composition of the F3 OBJ fishery in the different models (with  $h=1$ ).



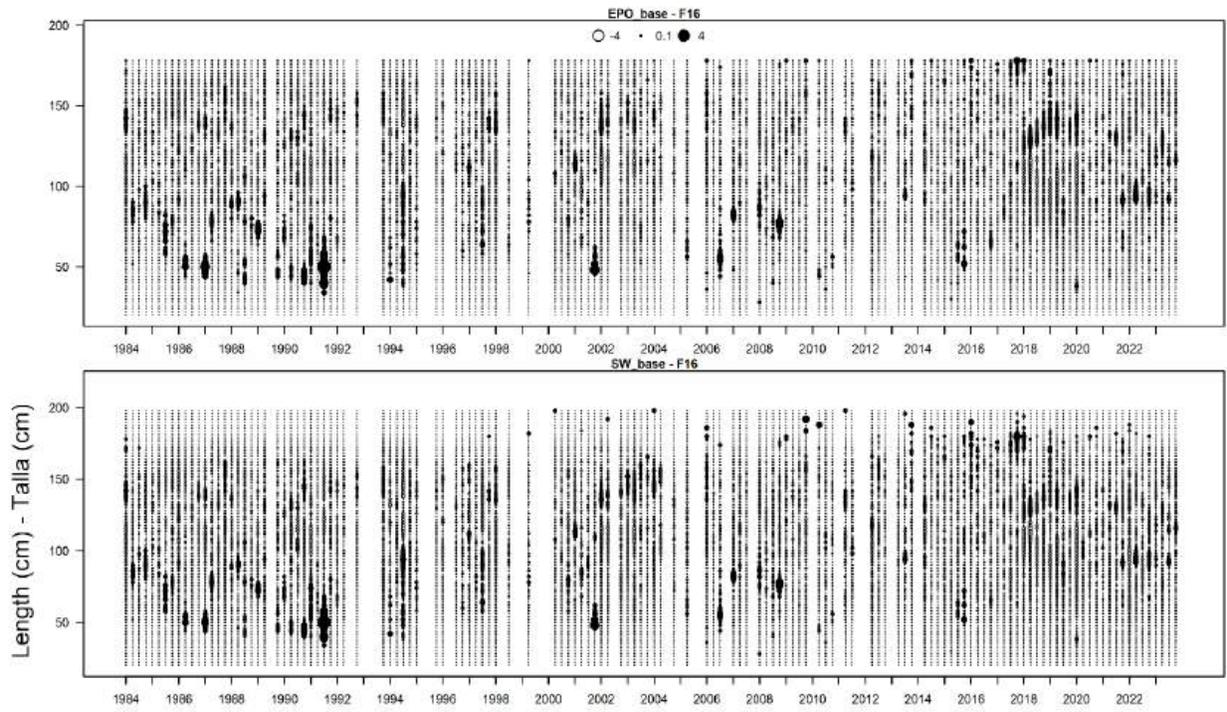
**FIGURE 9c.** Bubble plot: Pearson residuals for the length composition of the F4 OBJ fishery in the different models (with  $h=1$ ).



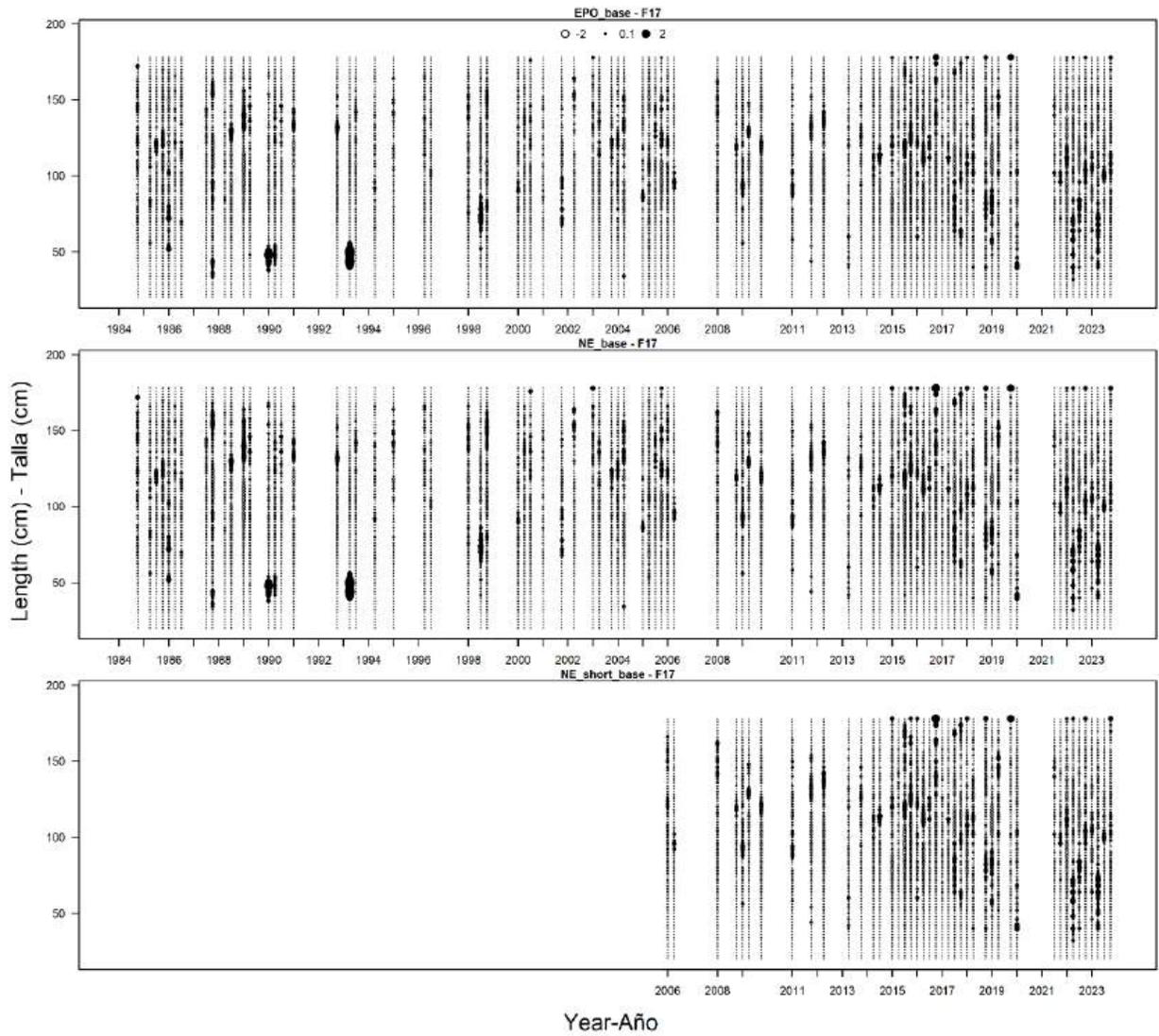
**FIGURE 9d.** Bubble plot: Pearson residuals for the length composition of the F14 DEL fishery in the different models (with  $h=1$ ).



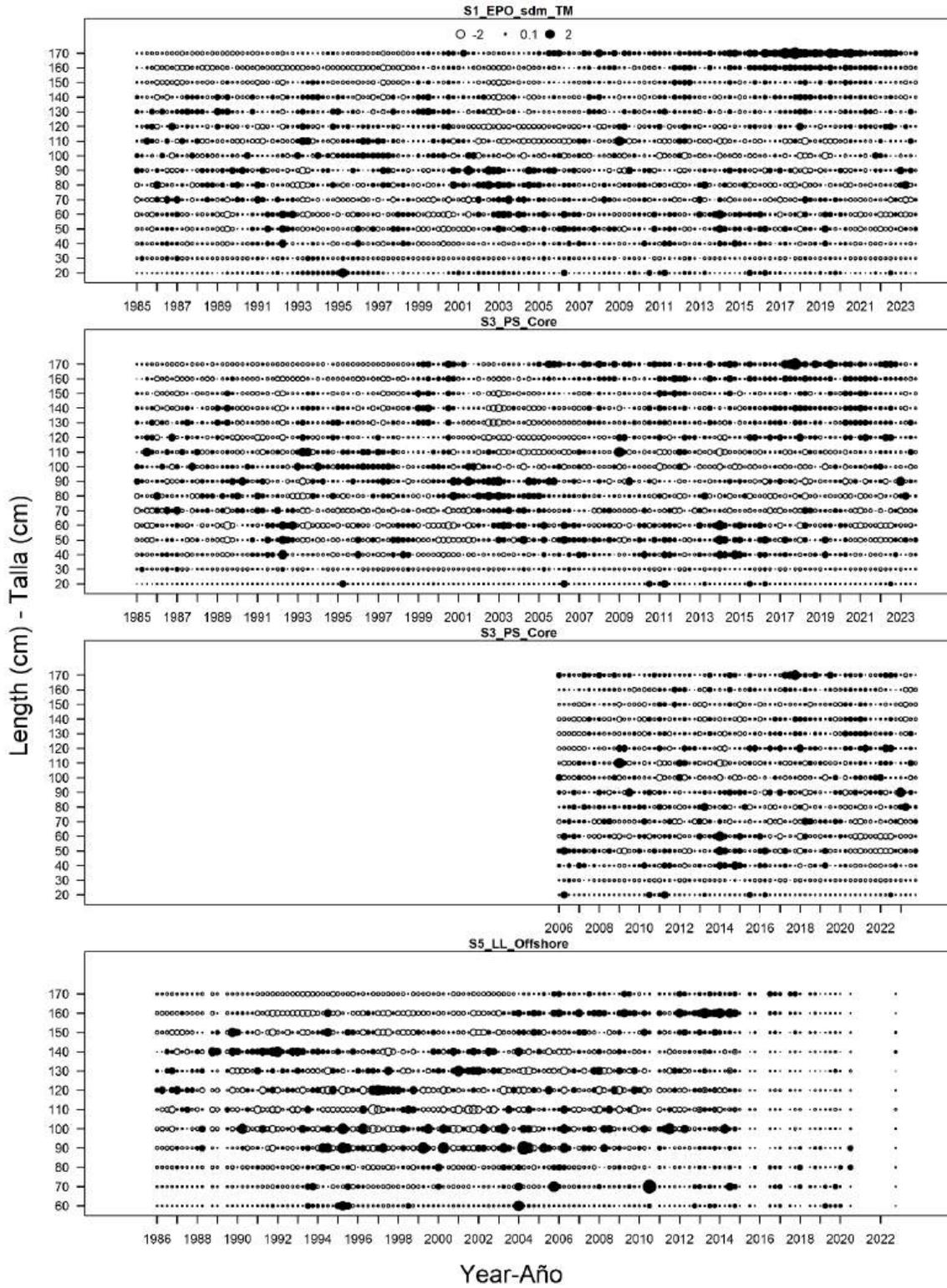
**FIGURE 9e.** Bubble plot: Pearson residuals for the length composition of the F15 DEL fishery in the different models (with  $h=1$ ).



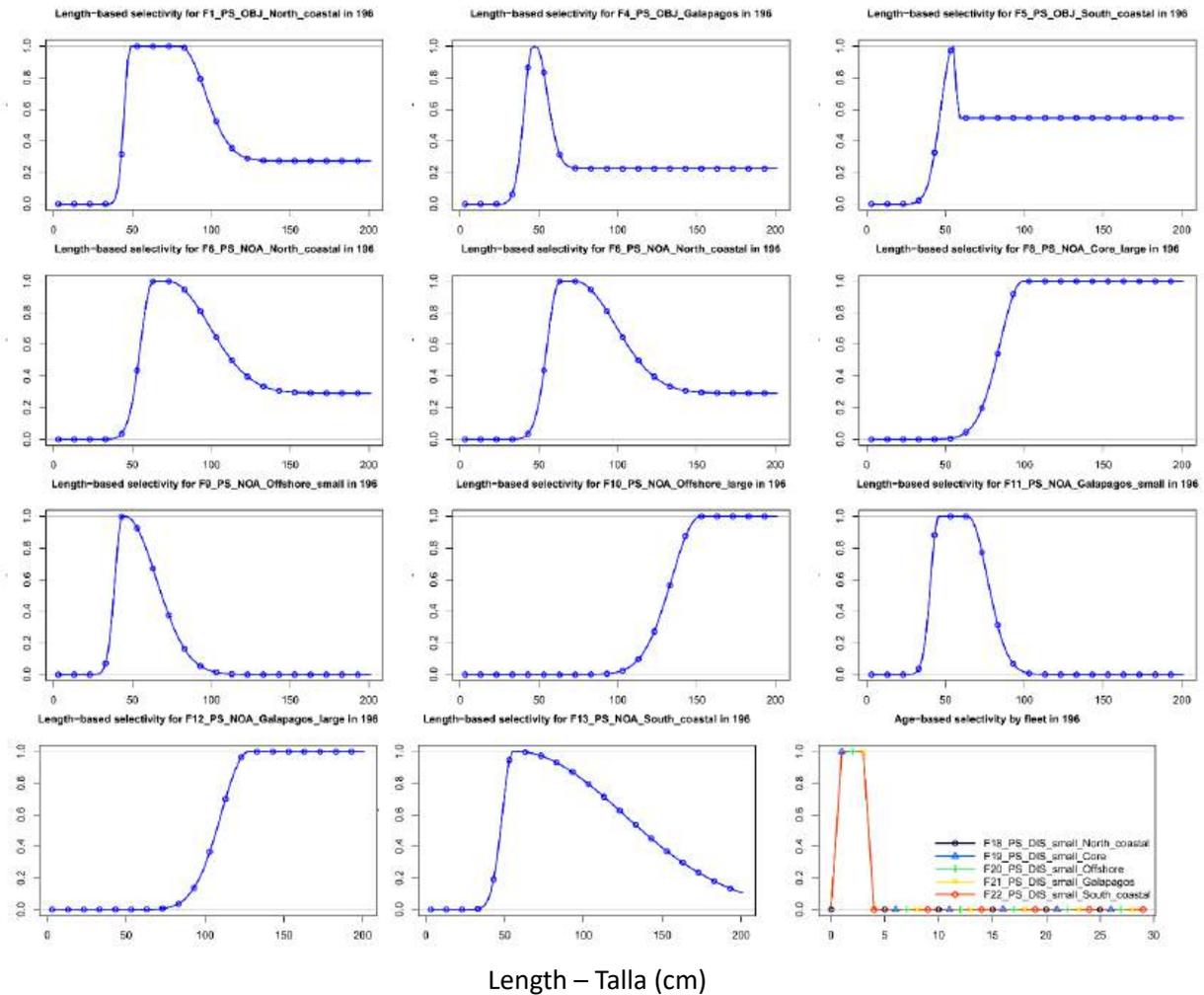
**FIGURE 9f.** Bubble plot: Pearson residuals for the length composition of the F16 DEL fishery in the different models (with  $h=1$ ).



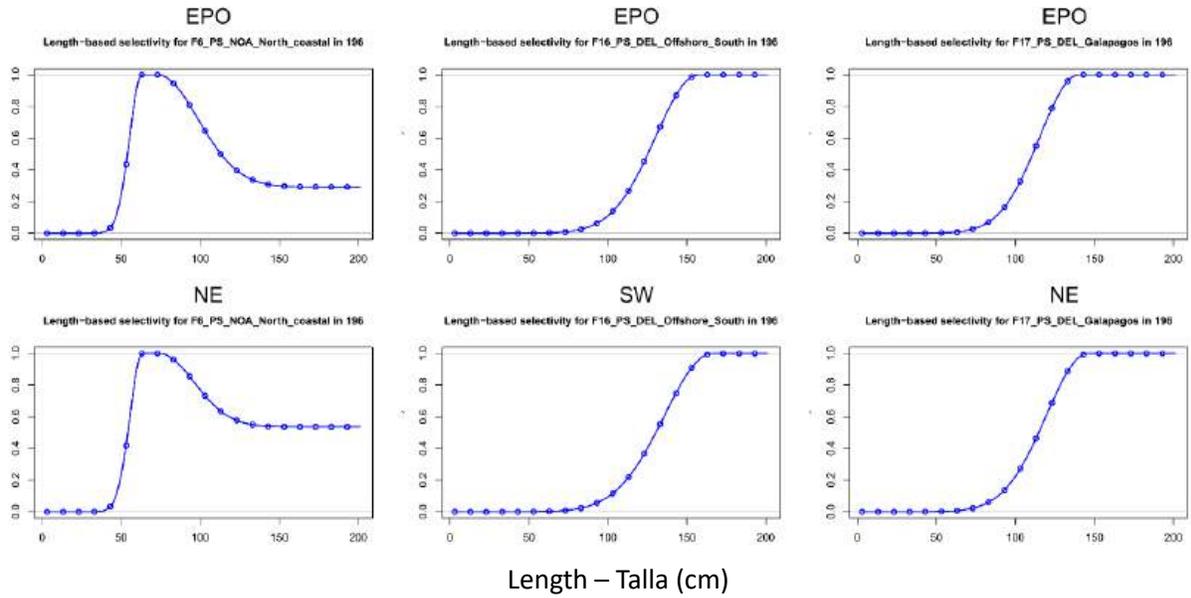
**FIGURE 9g.** Bubble plot: Pearson residuals for the length composition of the F17 DEL fishery in the different models (with  $h=1$ ).



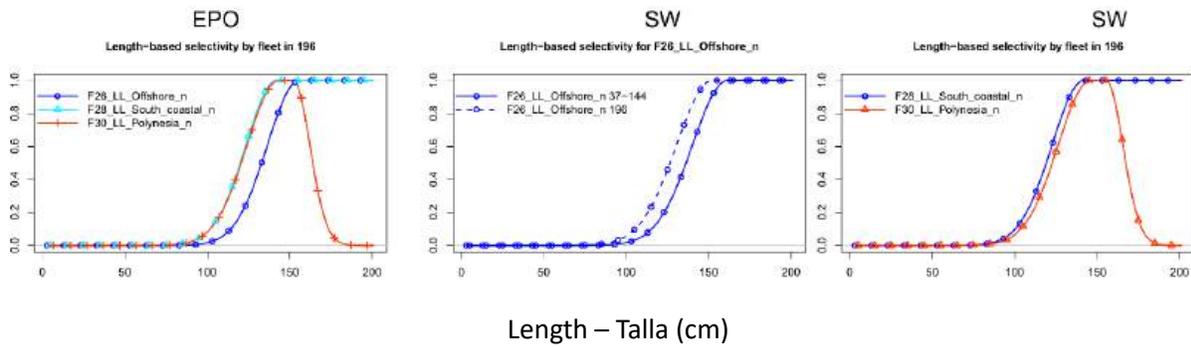
**FIGURE 9h.** Bubble plot: Pearson residuals for the length composition of the indices of abundance in the different models.



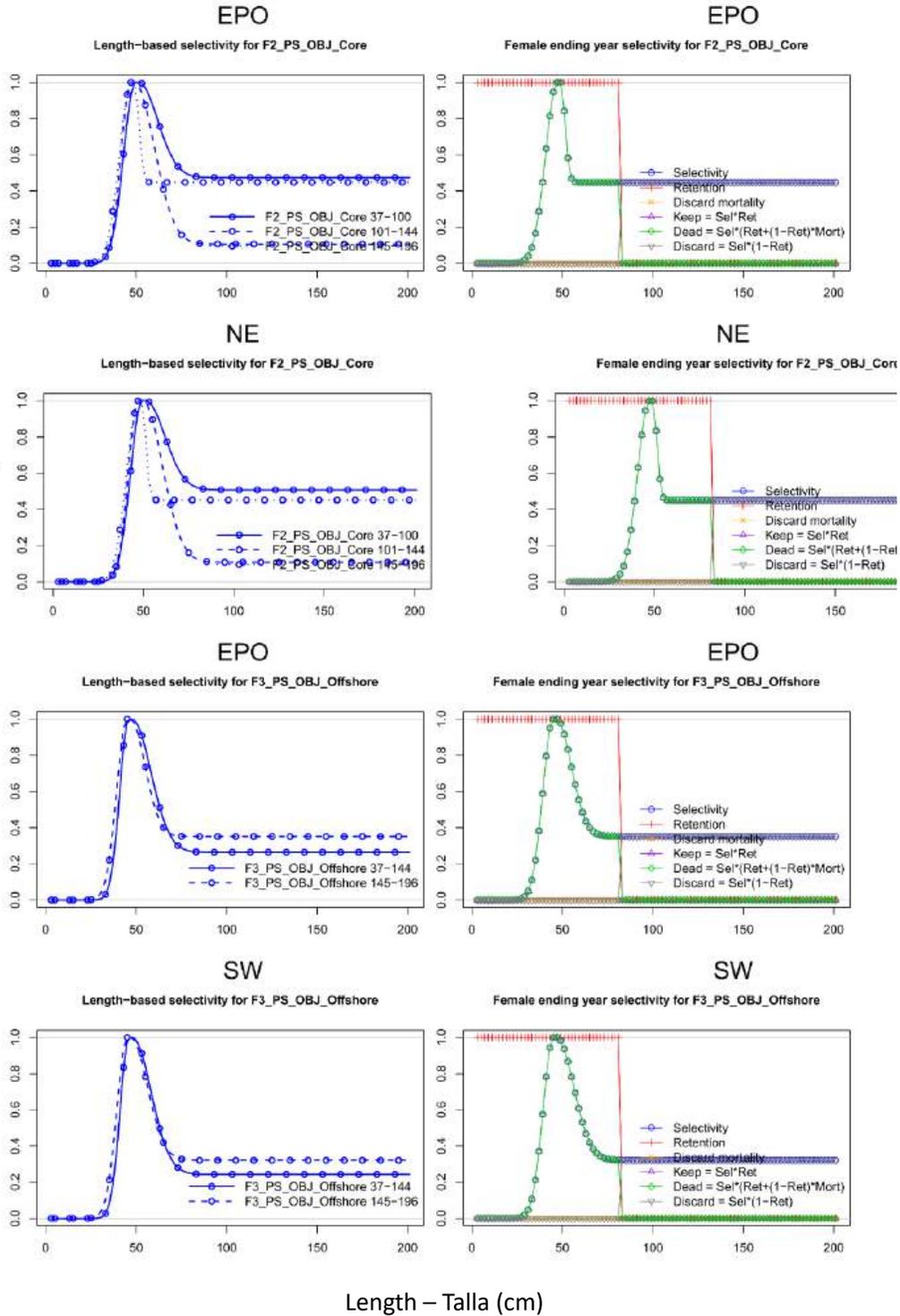
**FIGURE 10a.** Assumptions for fixed selectivity curves. The selectivities were estimated in the ancestral model and fixed in the other models, except for sorting discard fisheries (F18-F22) which were fixed in all models.



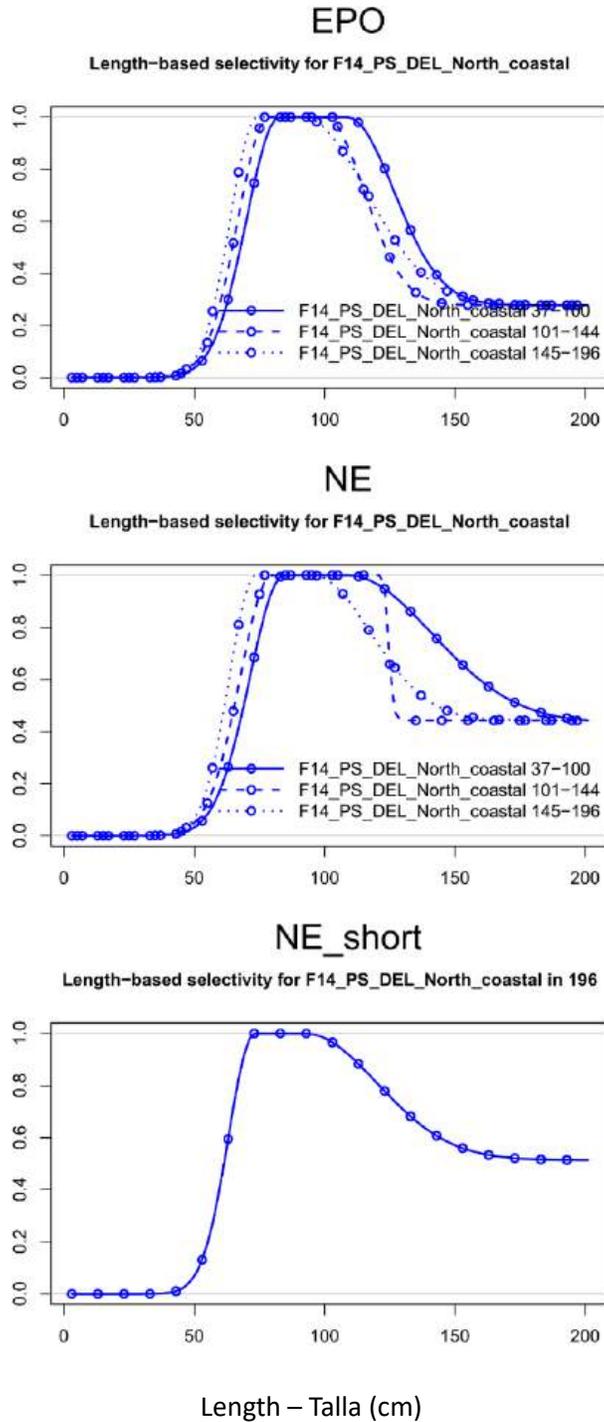
**FIGURE 10b.** Assumptions for selectivity curves for fleets whose length composition data had 0.2 weighting scalar in the risk analysis models.



**FIGURE 10c.** Assumptions for selectivity curves for fleets whose length composition data had 0.2 weighting scalar in the SW models and 0 in the other models (longline fisheries).



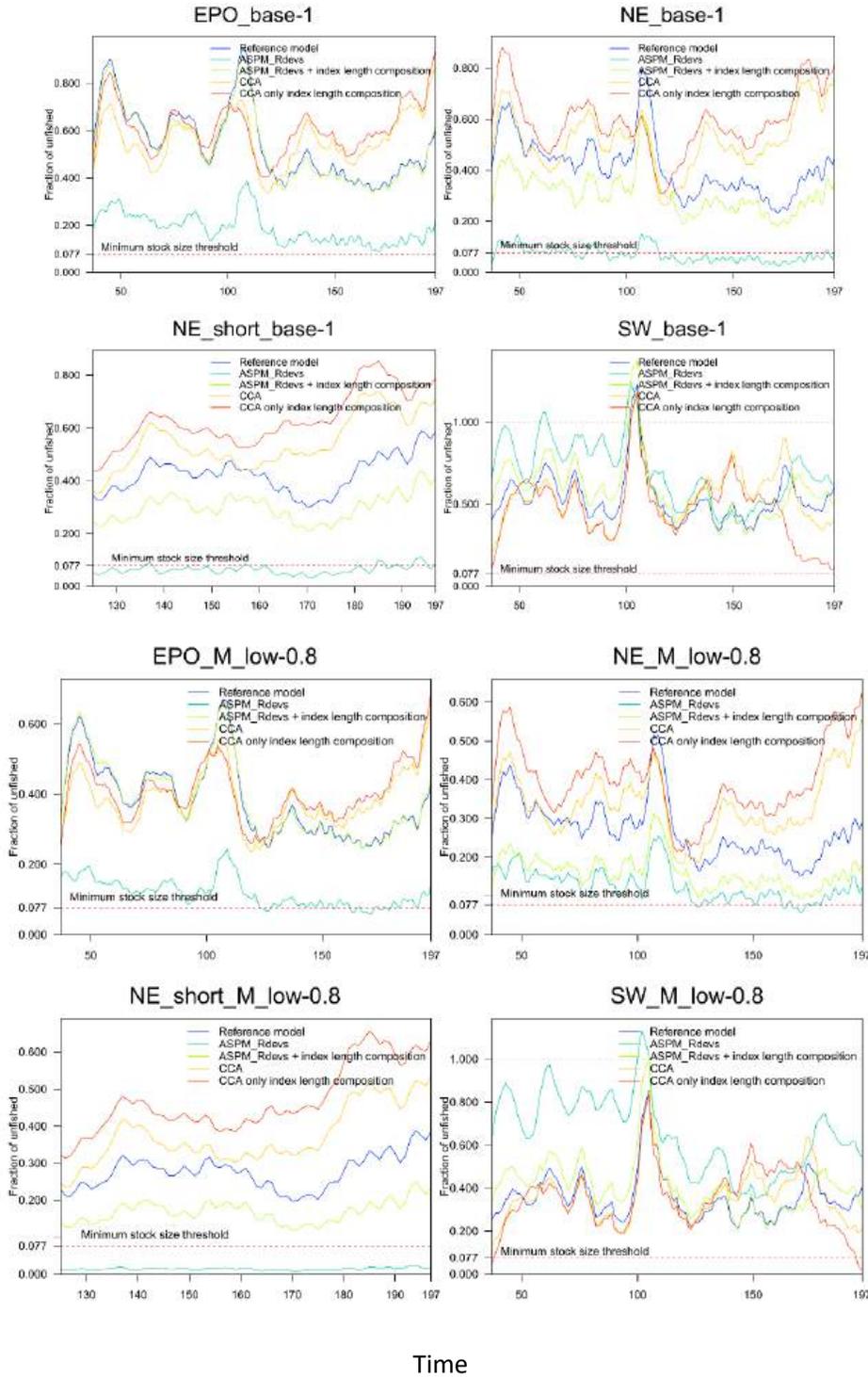
**FIGURE 10d.** Assumptions for selectivity and retention curves for purse-seine fisheries on floating object for fleets whose length composition data had 1 weighting scalar.



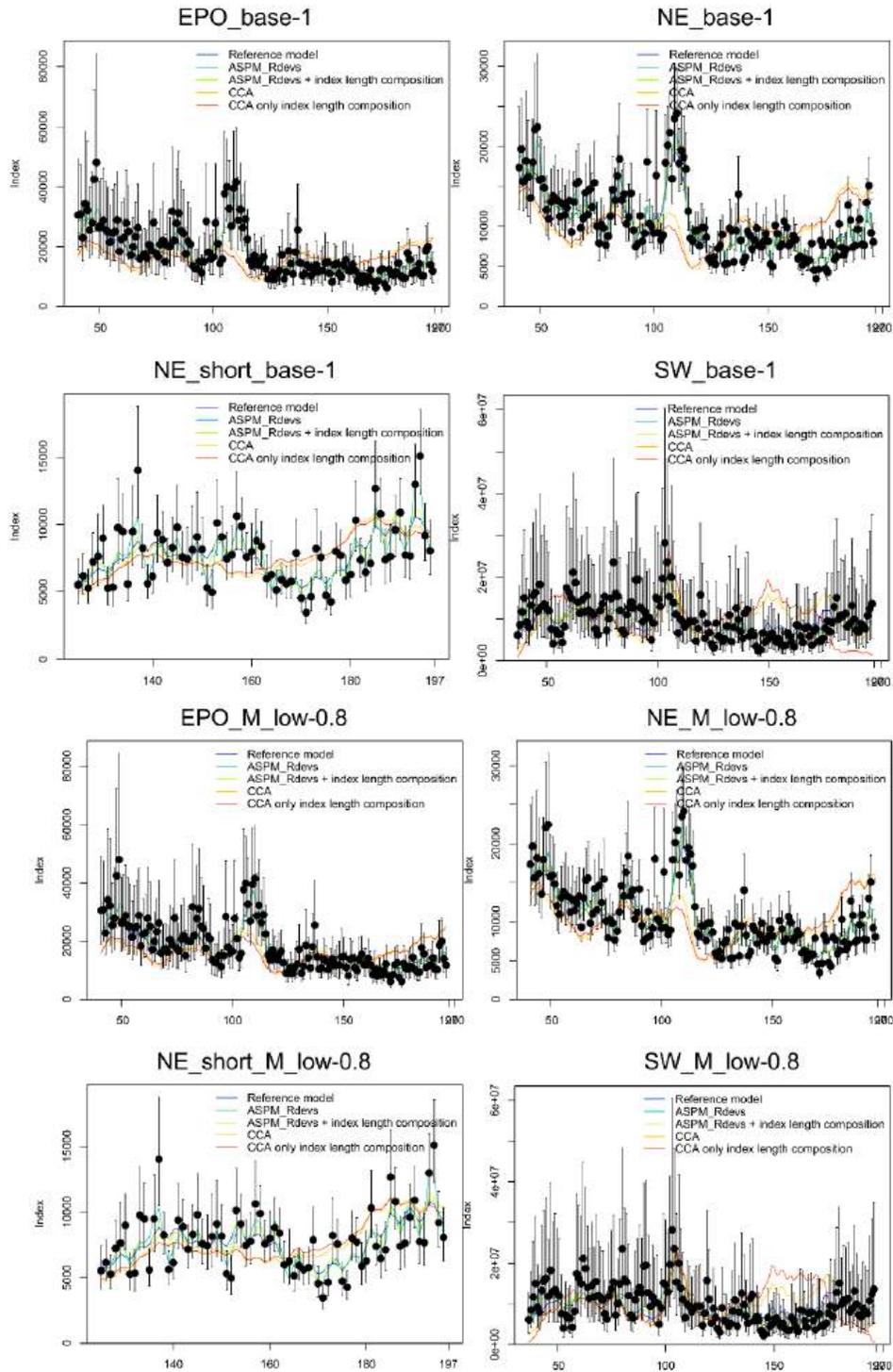
**FIGURE 10e.** Assumptions for selectivity curves for the purse-seine fishery associated with dolphins in the North coastal are F14 (weighting scalar = 1).





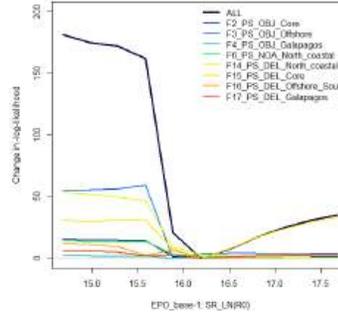
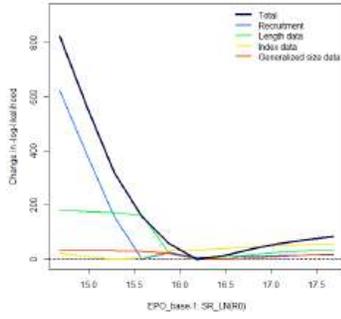


**FIGURE 11a.** Integrated model diagnostics: spawning biomass ratios of yellowfin tuna in the EPO estimated reference models and the corresponding diagnostic models (ASPM-dev, CCA). The lines represent the maximum likelihood estimate (MLE). The red dotted line at 0.077 indicates  $S=S_{LIMIT}$ .

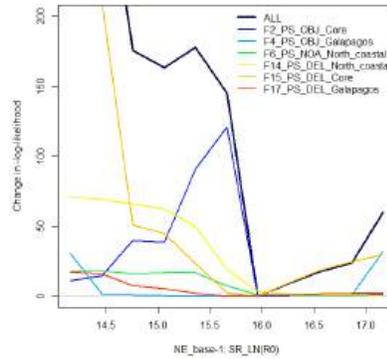
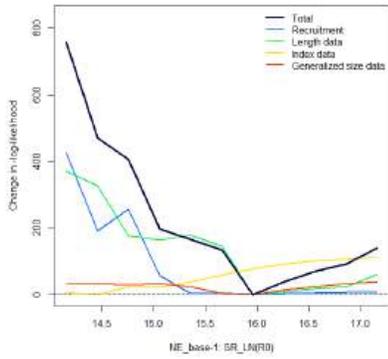


**FIGURE 11b.** Integrated model diagnostics: fits to indices of abundance estimated by each reference model and the corresponding diagnostic models. The lines represent the maximum likelihood estimates (MLE), the dots are the observed values.

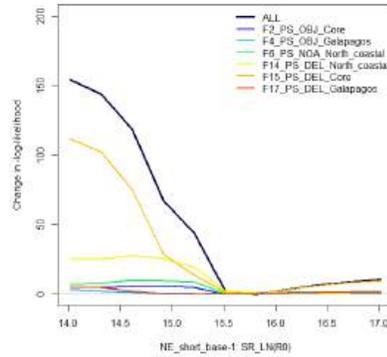
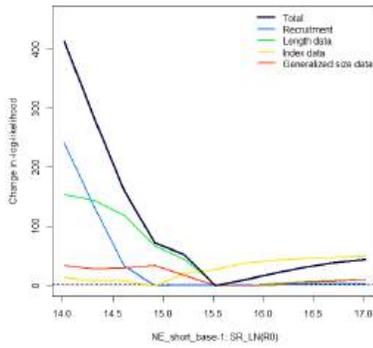
### EPO\_base-1



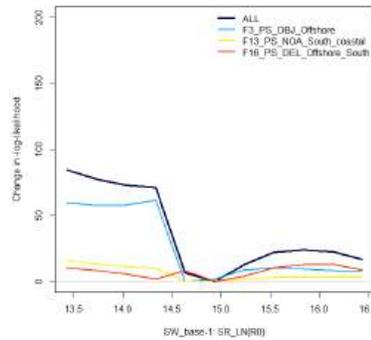
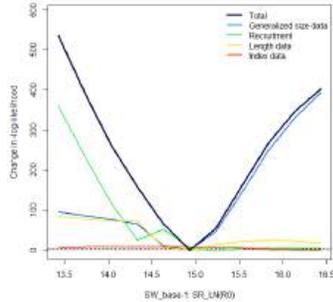
### NE\_base-1



### NE\_short\_base-1

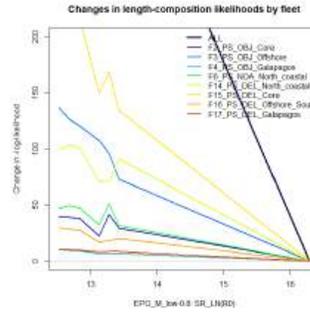
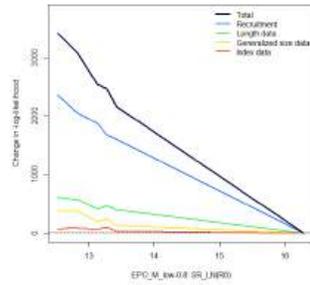


### SW\_base-1

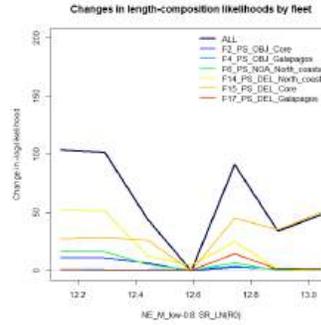
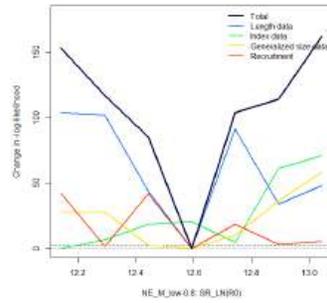


**FIGURE 11c.** R0 profile: likelihood profile for  $\ln R_0$  (scaling parameter) for base-1 reference models for yellowfin tuna in the EPO.

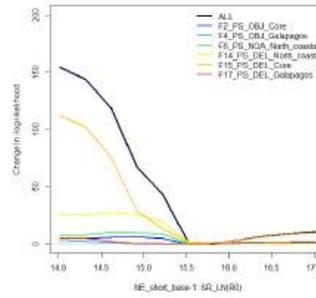
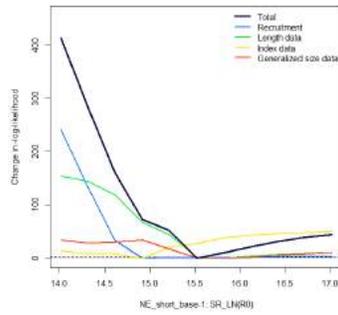
EPO\_M\_low-0.8



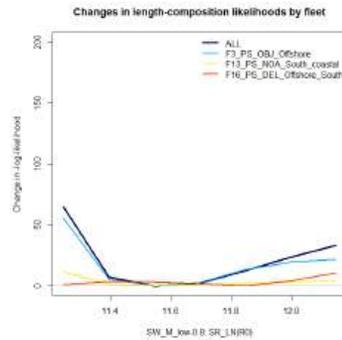
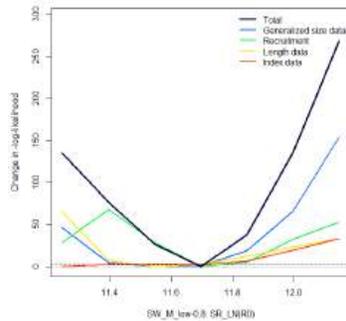
NE\_M\_low-0.8



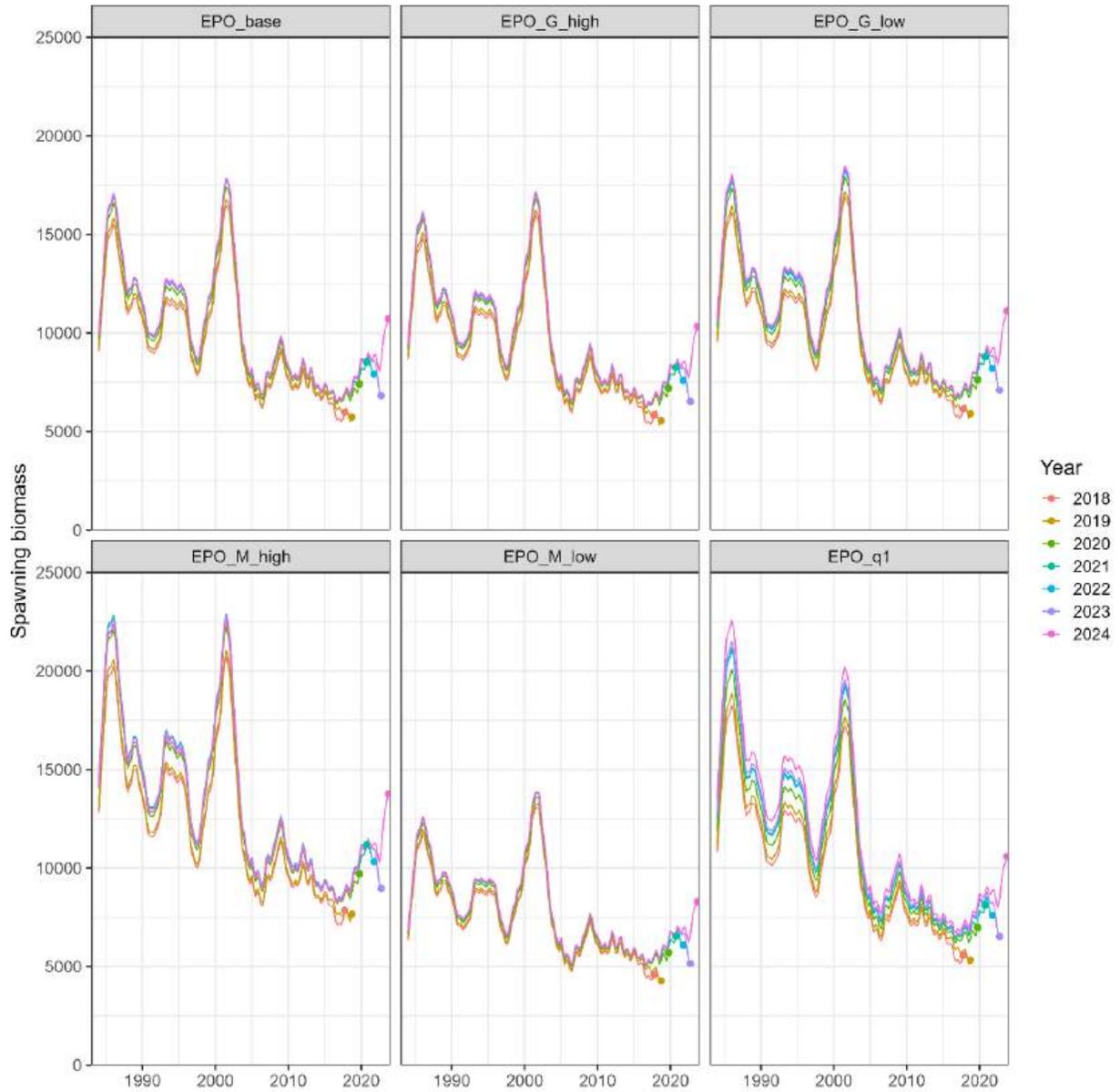
NE\_short\_M\_low-0.8



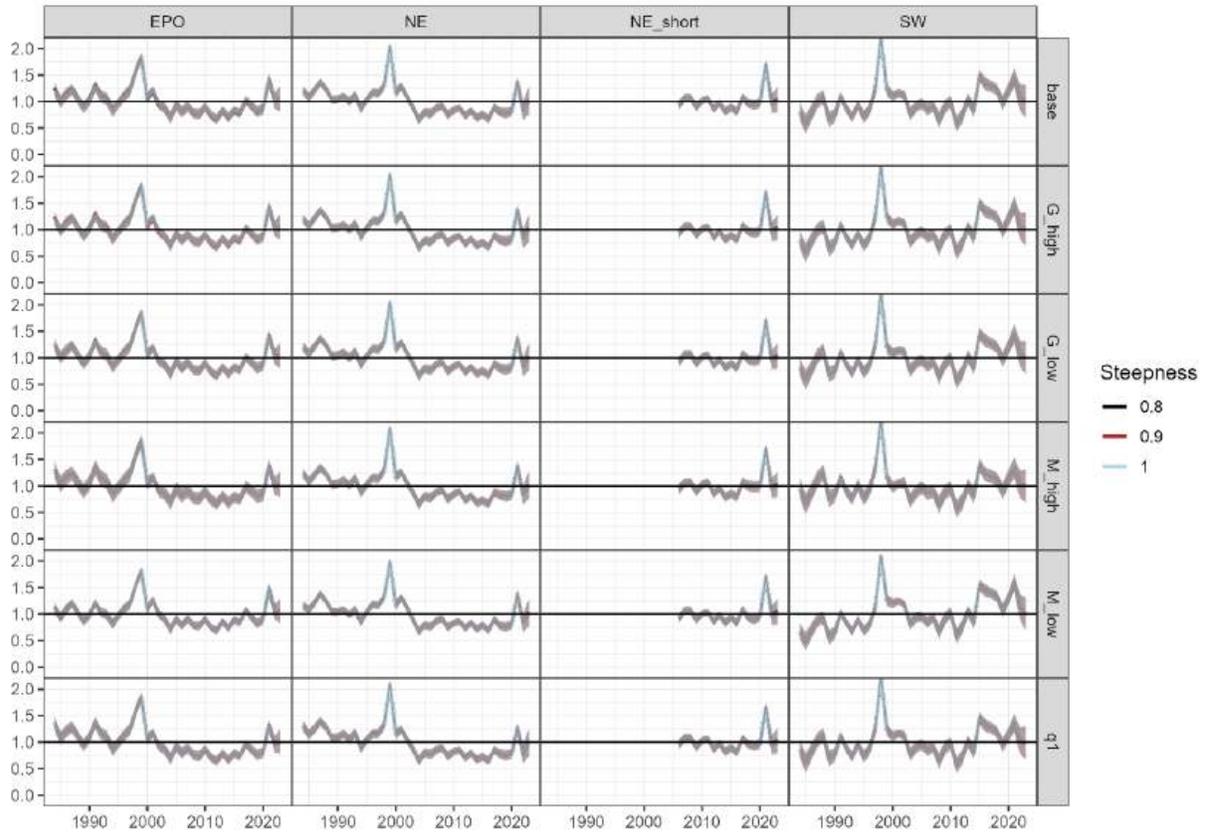
SW\_M\_low-0.8



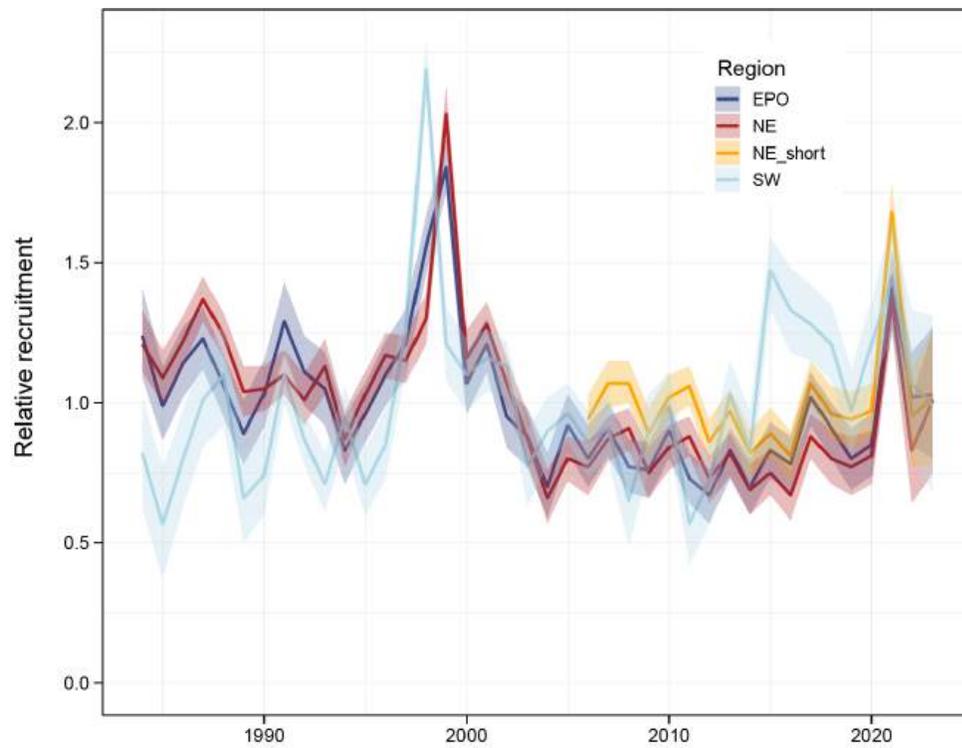
**FIGURE 11d.** R0 profile: likelihood profile for  $\ln R_0$  (scaling parameter) for M\_low-0.8. reference models for yellowfin tuna in the EPO.



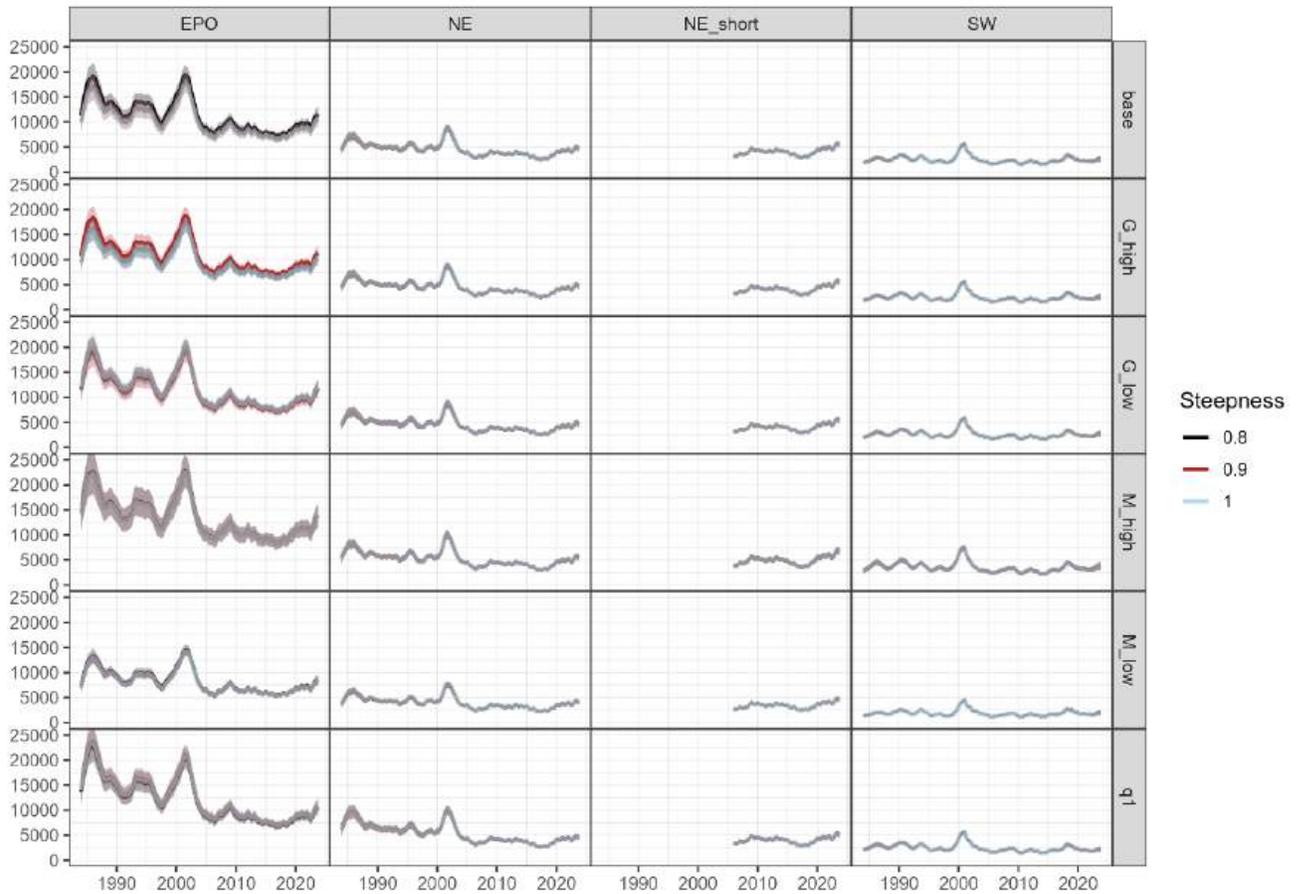
**FIGURE 11e.** Retrospective analysis for EPO models with  $h=1$ .



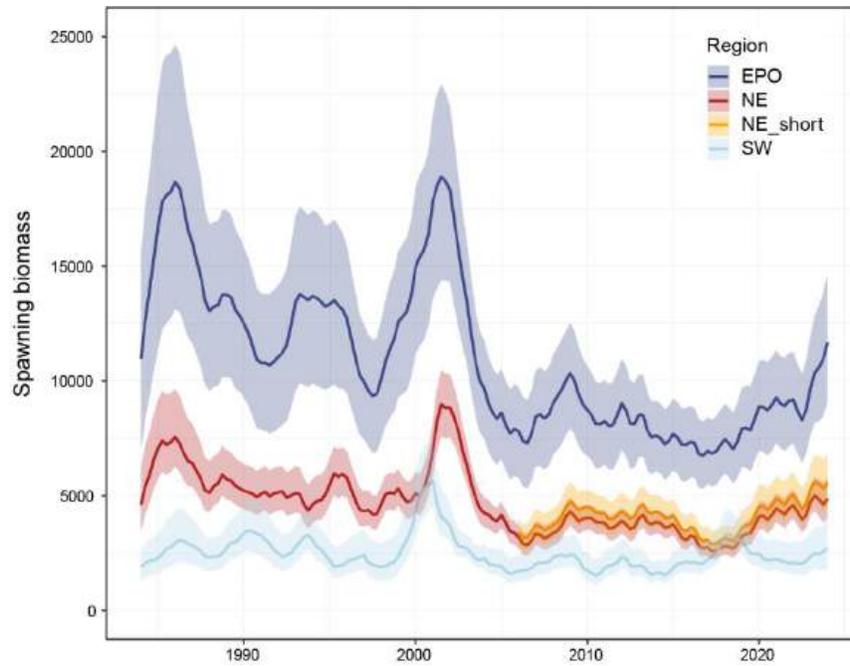
**FIGURE 12a.** Comparison of estimated relative annual recruitment of yellowfin tuna for each hypothesis of stock structure and level 2 hypothesis with 80% confidence intervals. The panels have the results for each steepness values (level 3).



**FIGURE 12b.** Comparison of multi-model estimates of relative annual recruitment of yellowfin tuna for each hypothesis of stock structure



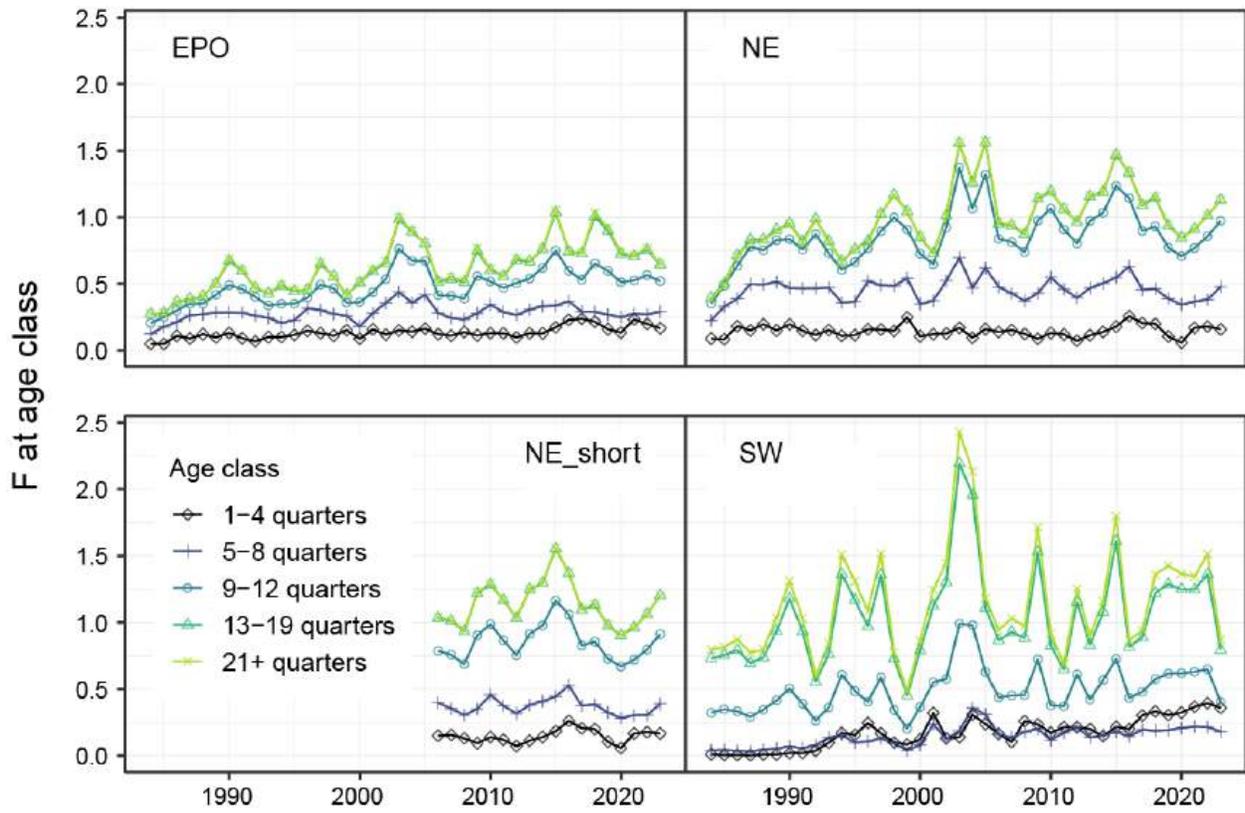
**FIGURE 13a.** Comparison of estimated spawning biomass of yellowfin tuna for each hypothesis of stock structure and level 2 hypothesis with 80% confidence intervals. The panels have the results for each steepness values (level 3).



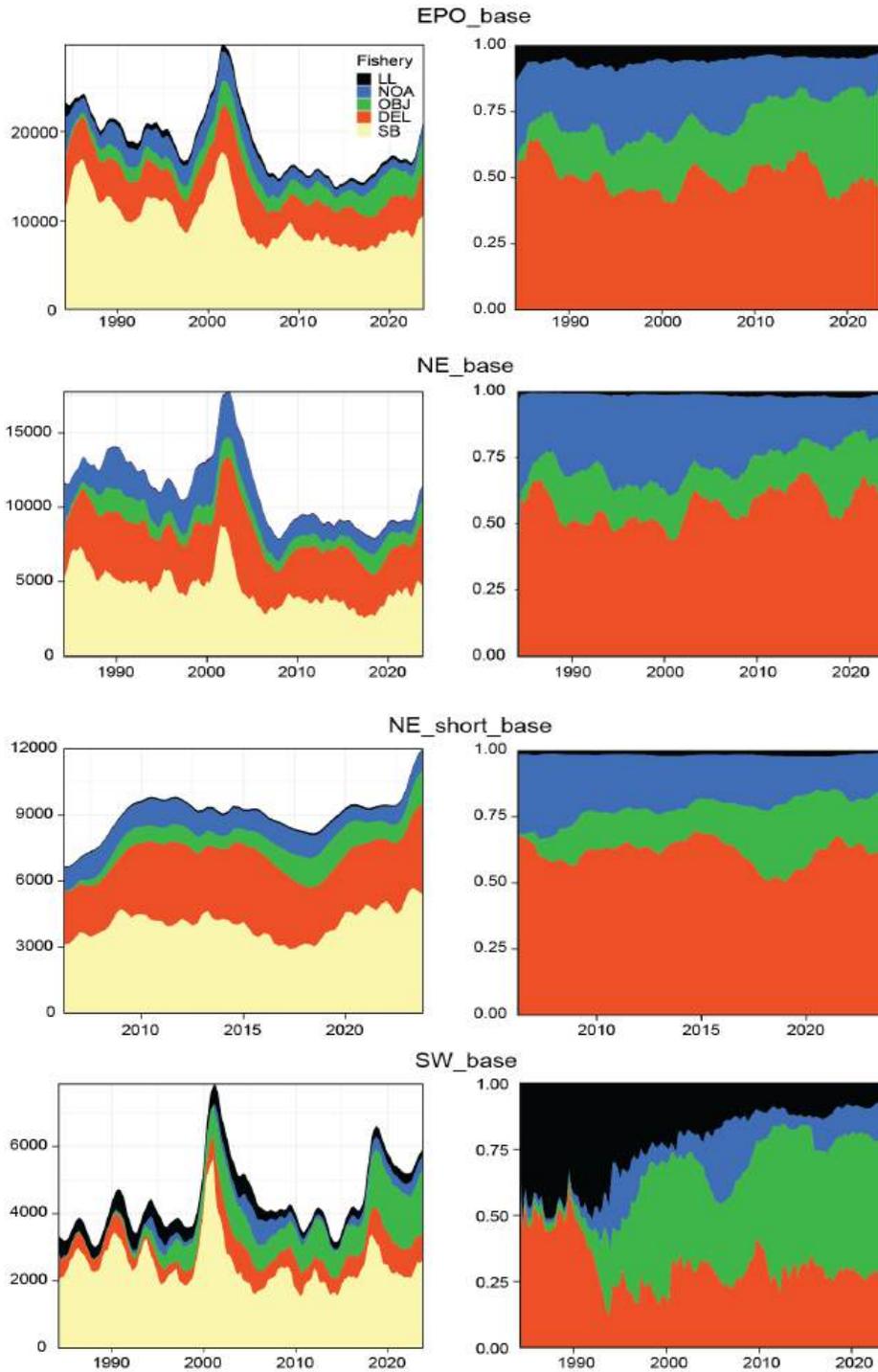
**FIGURE 13b.** Comparison of multi-model estimated spawning biomass of yellowfin tuna for each hypothesis of stock structure with 80% confidence intervals.



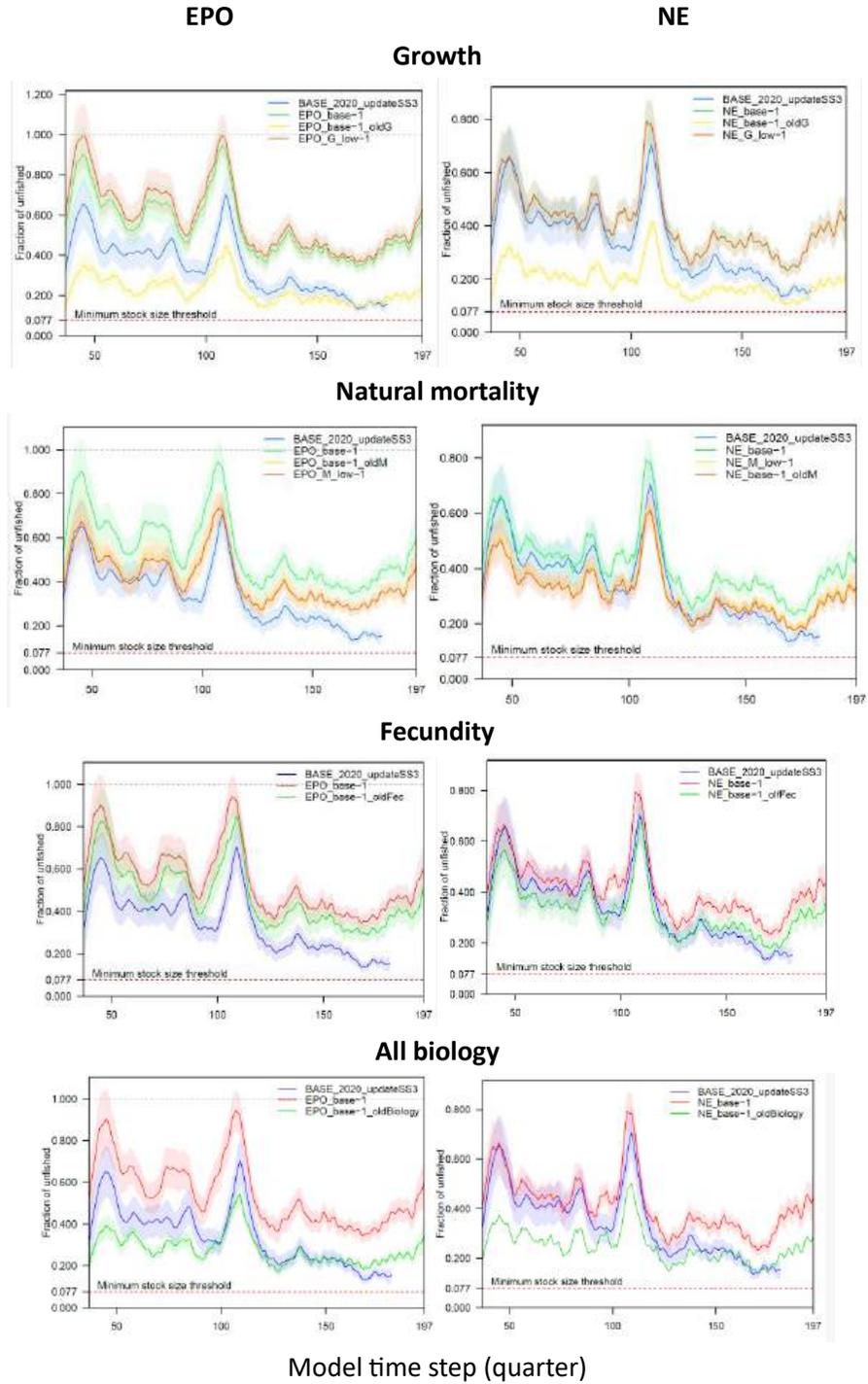
**FIGURE 14a.** Comparison of average annual fishing mortality of yellowfin by age group for each stock structure hypothesis (level 1) and level 2 hypothesis. The values for each model and age group are weighted across the steepness values (level 3).



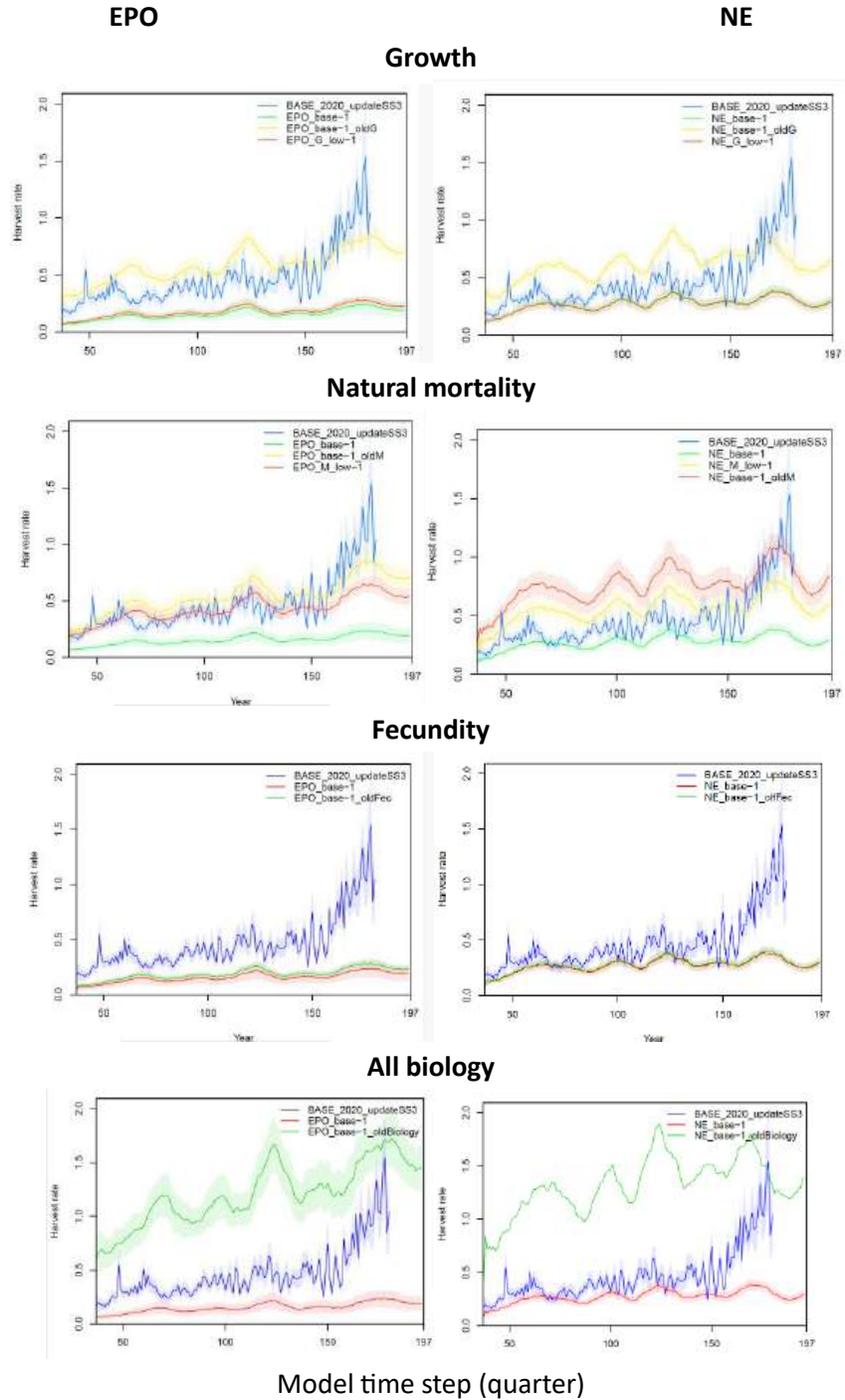
**FIGURE 15b.** Annual fishing mortality at age (sum of the four quarterly estimates within a year) of yellowfin by age group for each hypothesis of spatial structure (level 1). The values for each age group are weighted across level 2 and level 3 hypotheses.



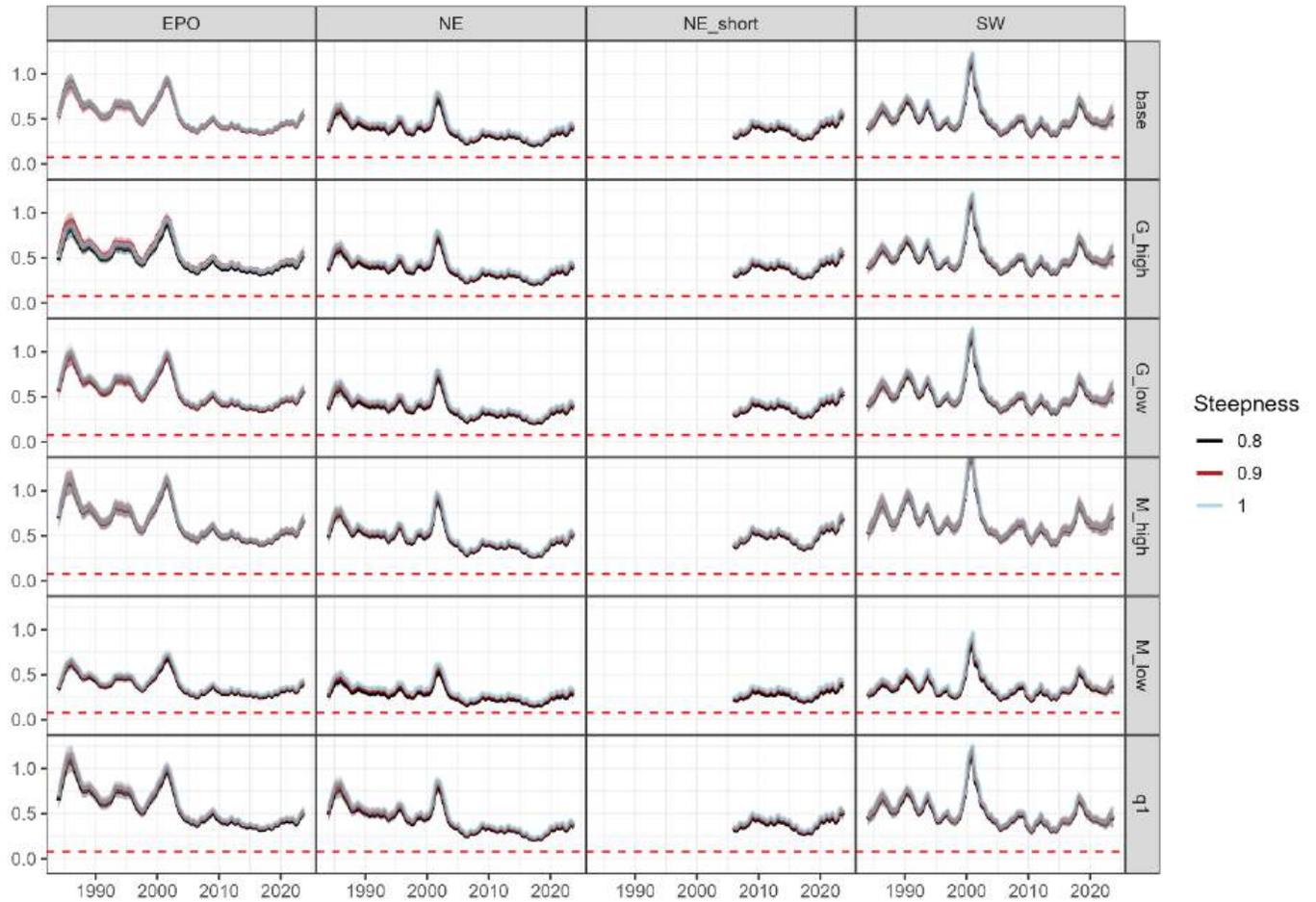
**FIGURE 16.** Impact of the different fishing methods on the spawning biomass. Left panels: comparison of spawning biomass trajectory of a simulated population of yellowfin tuna that was never exploited (colored area) and that predicted by the stock assessment model (SB, yellow shaded area), and the impact of each fishing method (purse-seine on floating objects OBJ, also includes sorting discards and pole and line, purse-seine associated with dolphins DEL, purse-seine unassociated NOA and longline LL fisheries) for each stock structure hypothesis calculated from the base reference models with steepness of 1. Right panels: Proportional impacts.



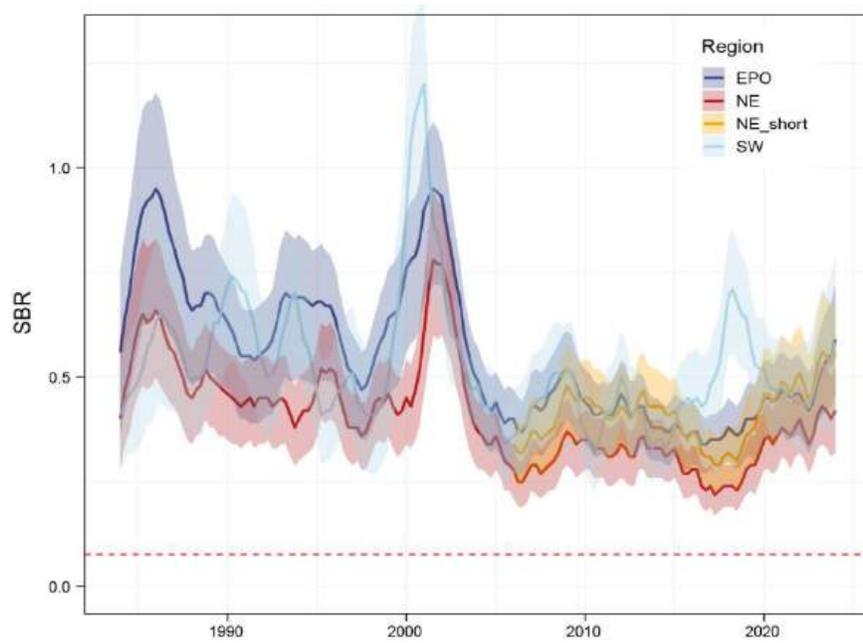
**FIGURE 17a.** Bridging: comparison of estimates of spawning biomass ratio when using assumptions from the 2020 benchmark assessment.



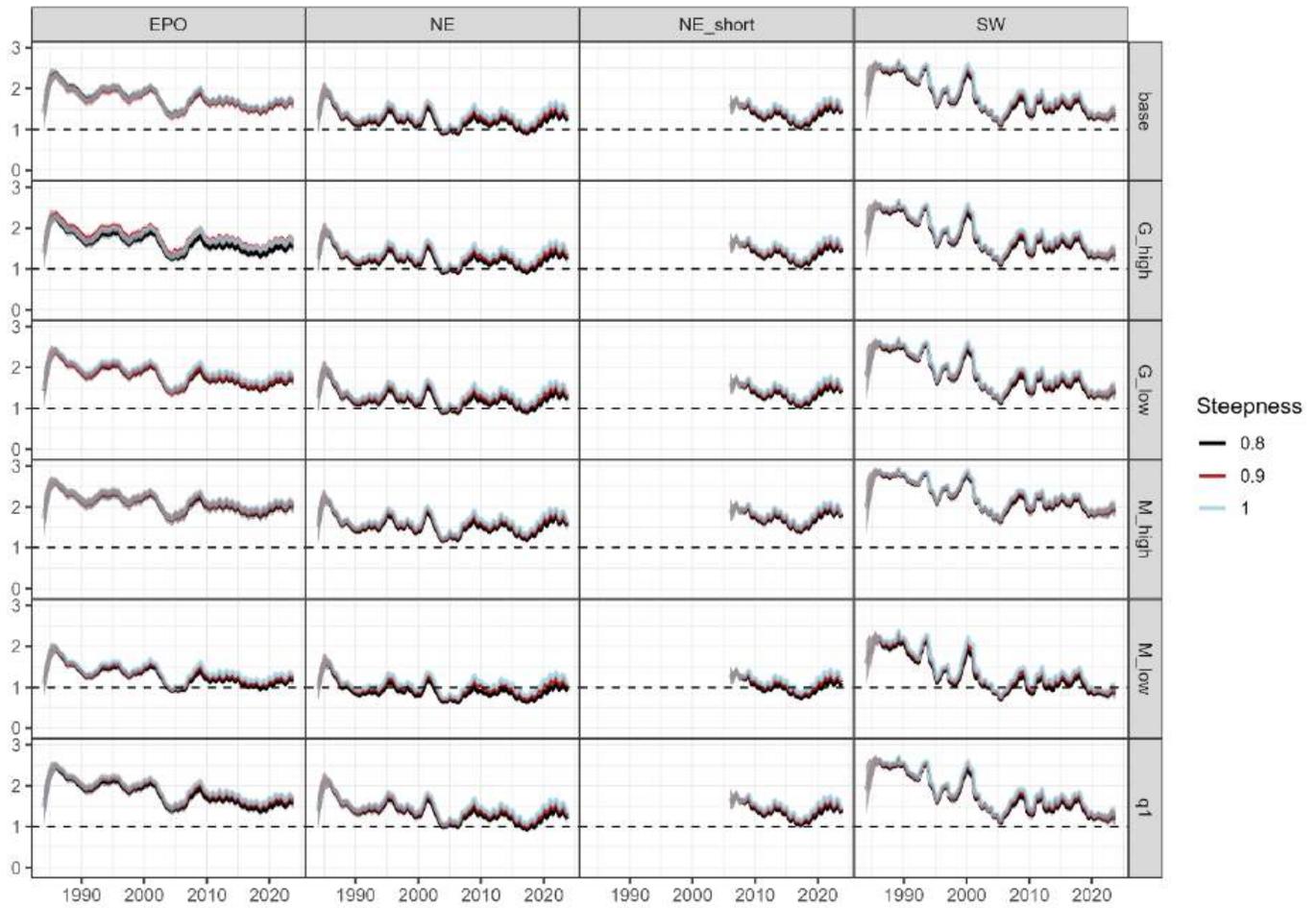
**FIGURE 17b.** Bridging: comparison of estimates of proxy for fishing mortality when using assumptions from the 2020 benchmark assessment.



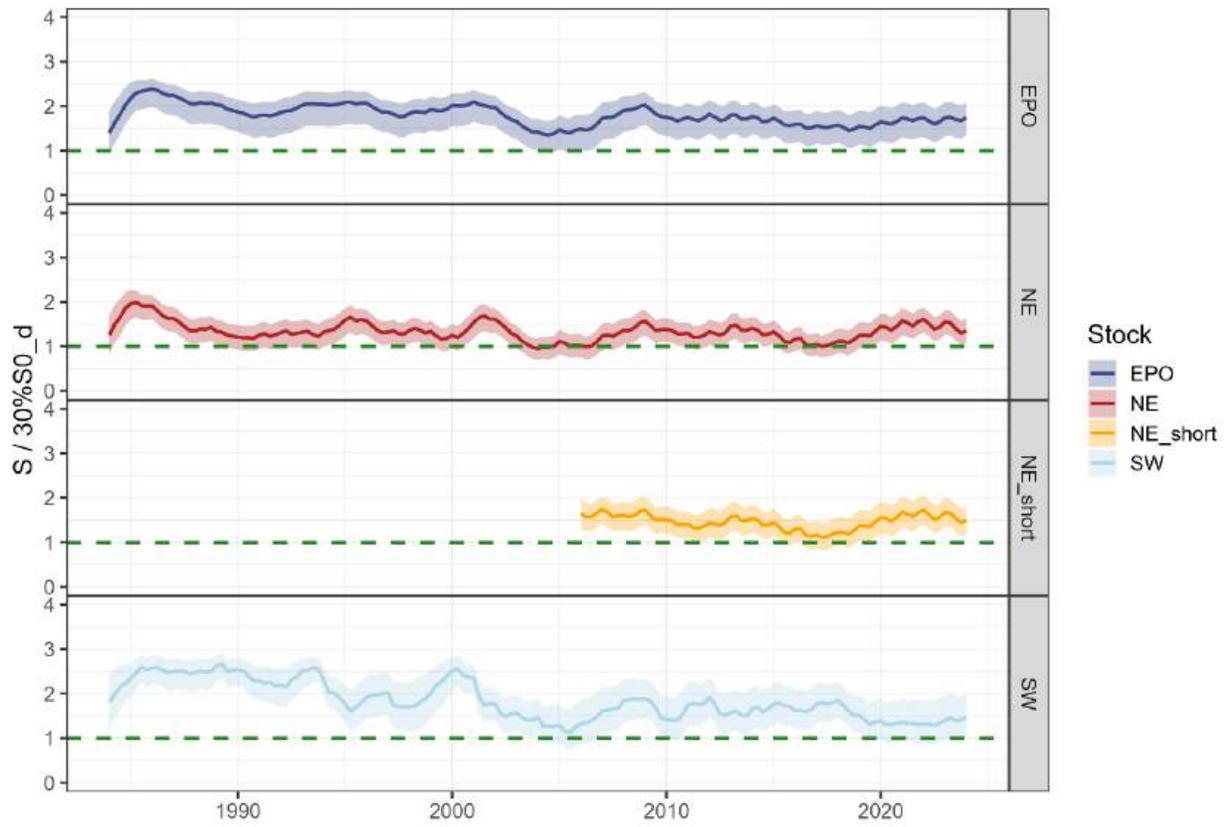
**FIGURE 18a.** Comparison of estimated spawning biomass ratio of yellowfin tuna for each hypothesis of stock structure and level 2 hypothesis with 80% confidence intervals. The panels have the results for each steepness values (level 3). The dashed line indicates the spawning biomass limit reference point of 0.077.



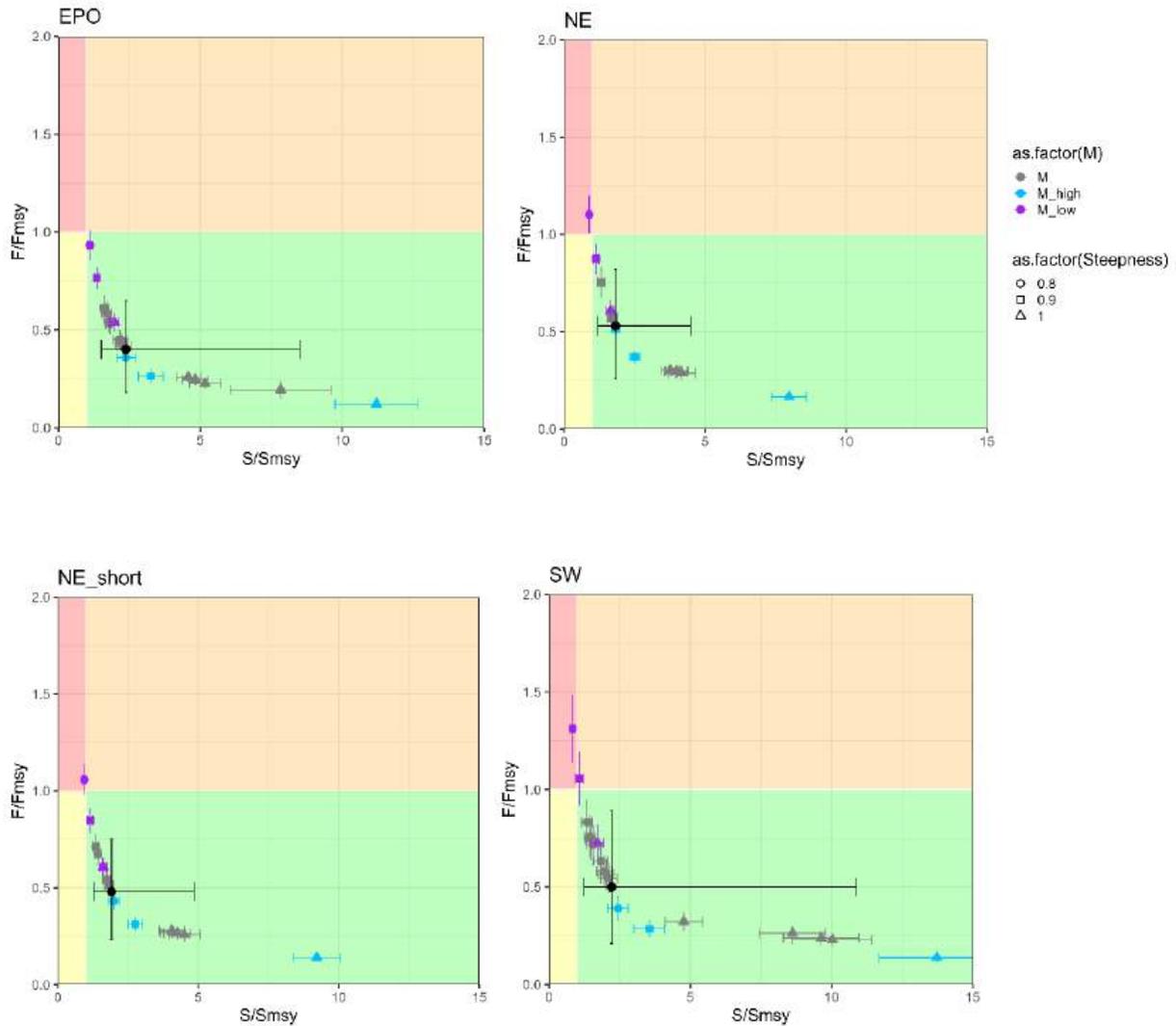
**FIGURE 18b.** Comparison of multi-model estimated spawning biomass ratio (spawning biomass over equilibrium virgin spawning biomass) of yellowfin tuna for each hypothesis of spatial structure with 80% confidence intervals. The red dashed line (at 0.077) indicates the SBR at the limit reference point  $S_{LIMIT}$ .



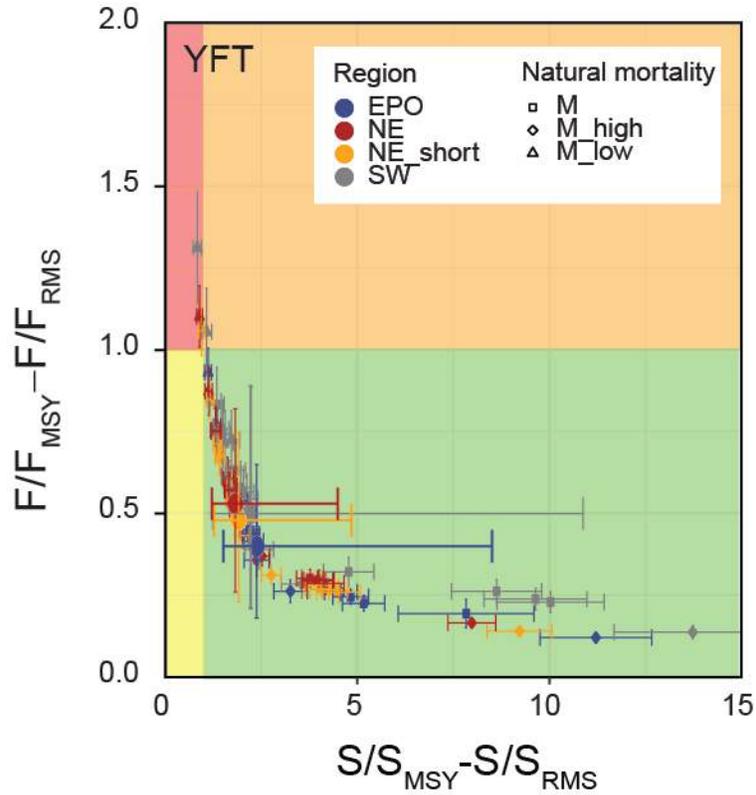
**FIGURE 19a.** Comparison of estimated ratio of  $S$  to  $30\%S_d$  of yellowfin tuna for each hypothesis of stock structure and level 2 hypothesis with 80% confidence intervals. The panels have the results for each steepness values (level 3).



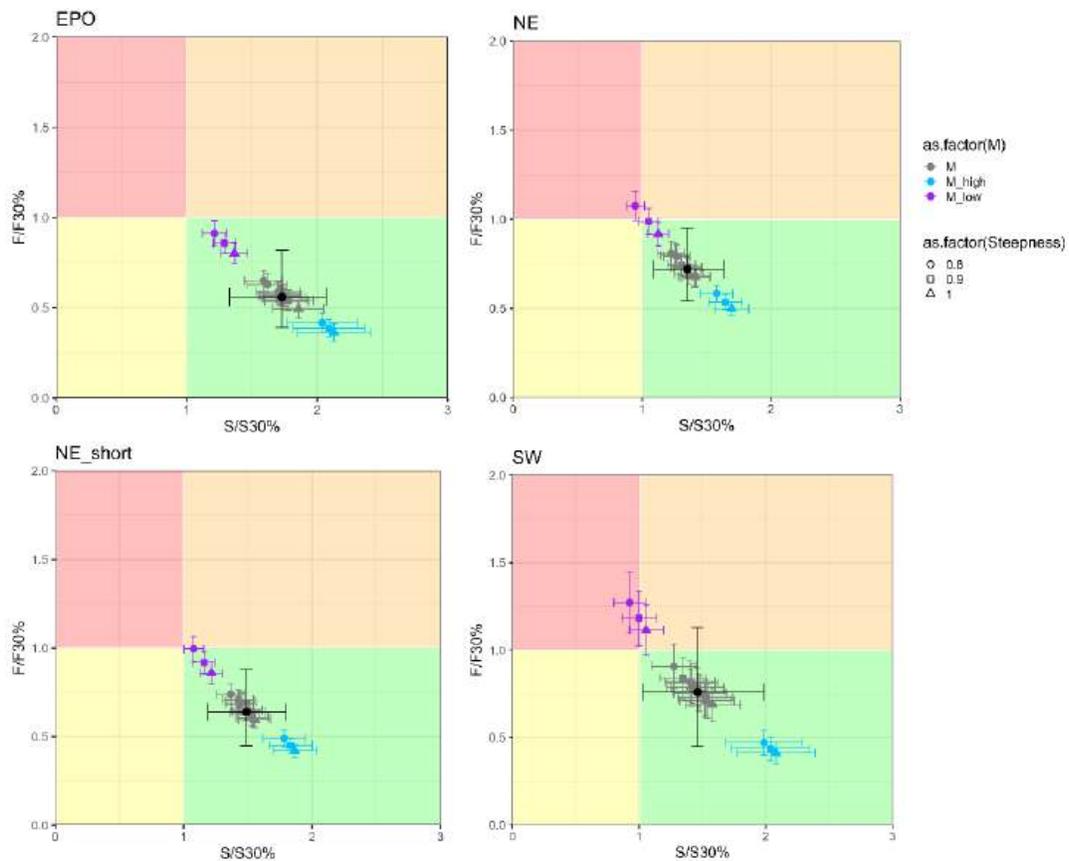
**FIGURE 19b.** Comparison of multi-model estimates of the ratio of  $S$  to  $30\%S_d$  of yellowfin tuna for each hypothesis of stock structure.



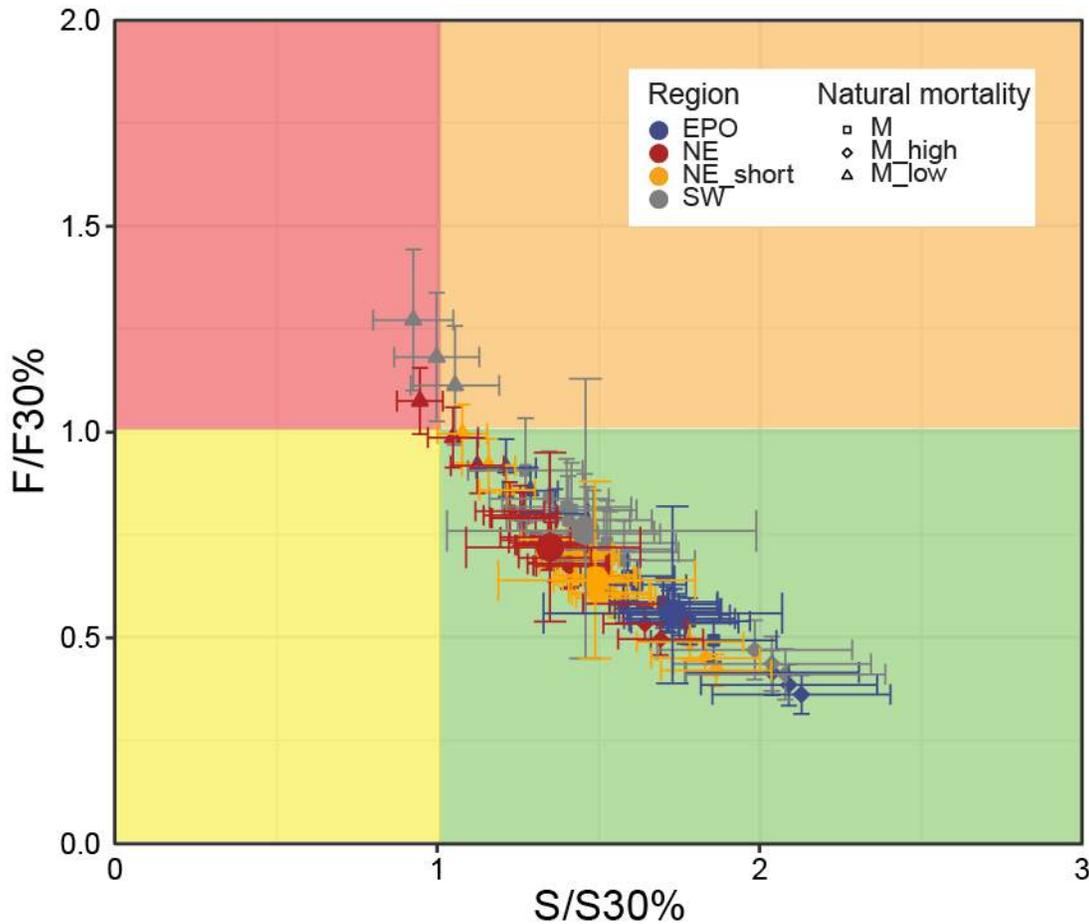
**FIGURE 20a.** Kobe plot of the most recent estimates of spawning biomass ( $S$ ) and fishing mortality ( $F$ ) relative to their MSY reference points ( $S_{MSY,d}$  and  $F_{MSY}$ ) for each stock structure. Each dot is based on the average  $F$  over the most recent three years, 2021-2023, and the  $S$  for the first quarter of 2024 and the error bars represent the 80% confidence interval of model estimates. The black dot and error bars represent the medium and 80% confidence interval of combined values, respectively Estructura.



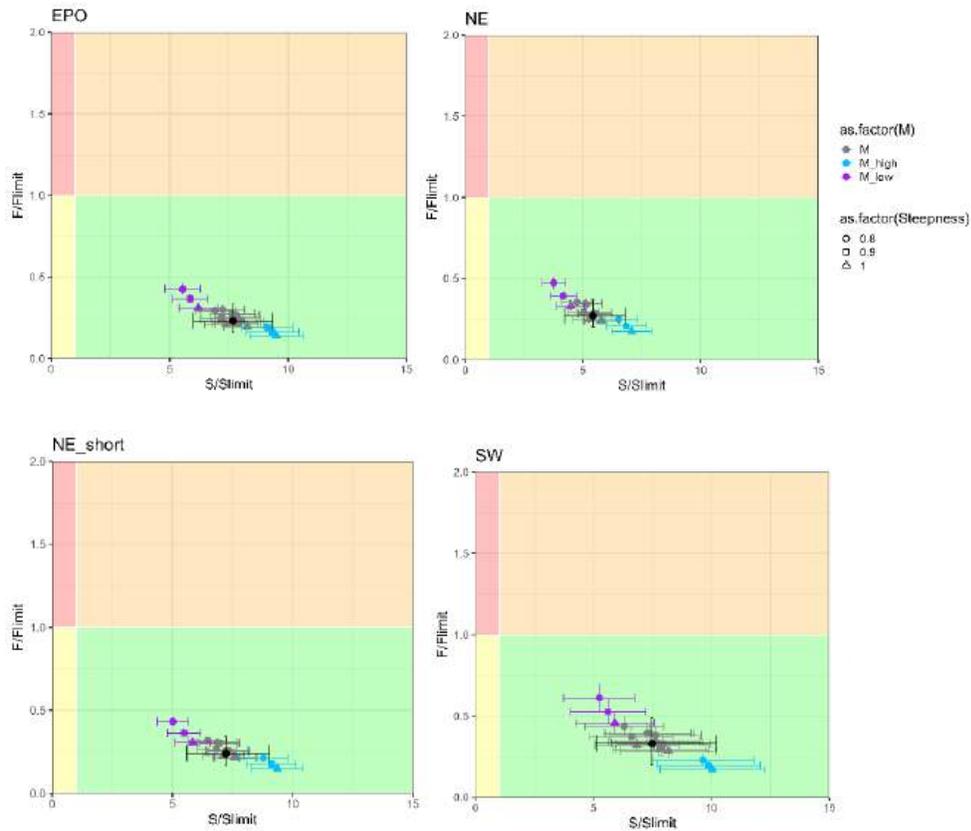
**FIGURE 20b.** Kobe plot of the most recent estimates of spawning biomass ( $S$ ) and fishing mortality ( $F$ ) relative to their target reference points ( $S_{MSY_d}$  and  $F_{MSY}$ ) for each hypothesis of spatial structure. Each dot is based on the average  $F$  over the most recent three years, 2021-2023, and the  $S$  for the first quarter of 2024 and the error bars represent the 80% confidence interval of model estimates. The larger dots represent the combined result for each spatial structure hypothesis.



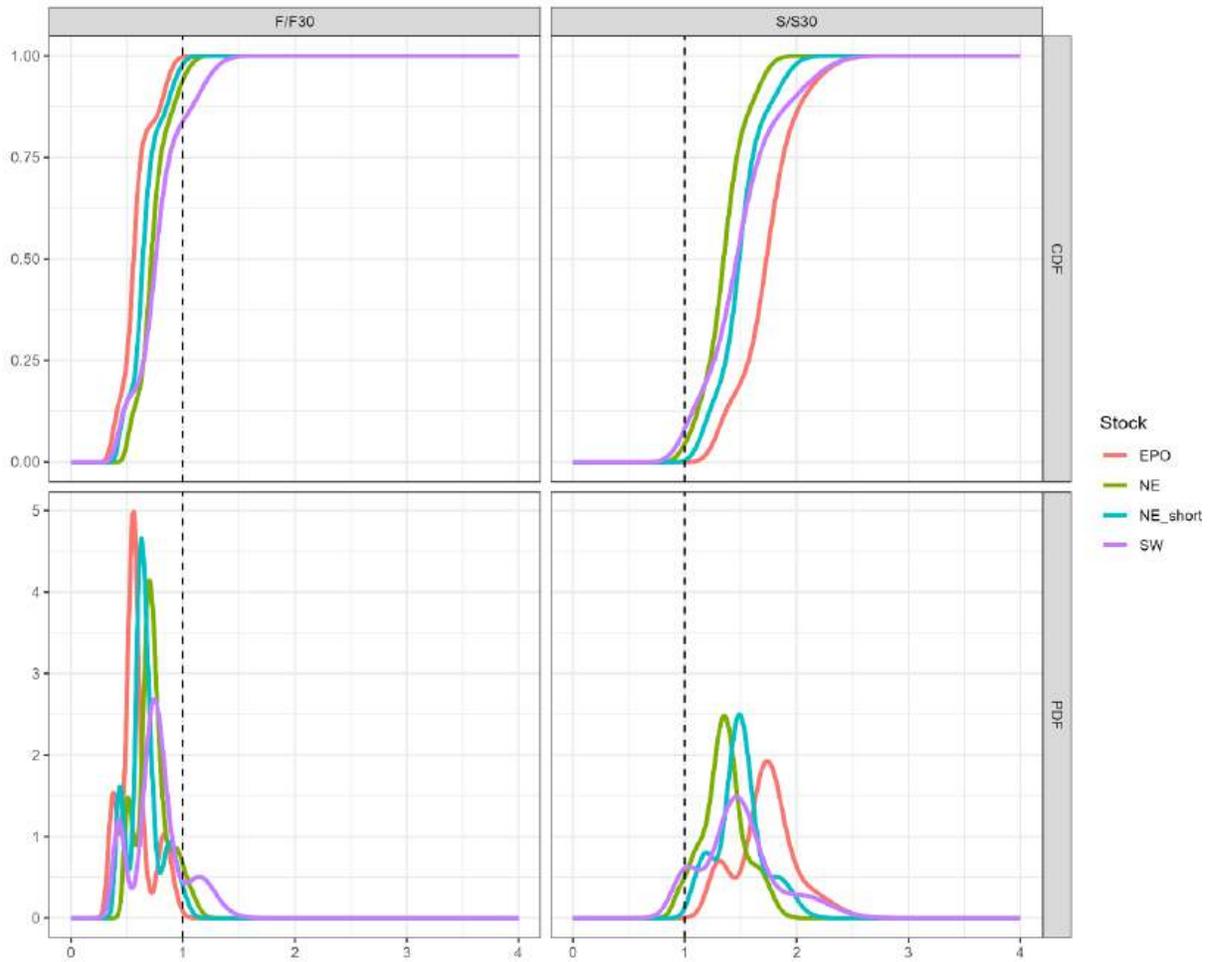
**FIGURE 221a.** Kobe plot of the most recent estimates of spawning biomass ( $S$ ) and fishing mortality ( $F$ ) relative to their proxy target reference points ( $30\%S_d$  and  $F_{30\%S_d}$ ) for each stock structure. Each dot is based on the average  $F$  over the most recent three years, 2021-2023, and the  $S$  for the first quarter of 2024 and the error bars represent the 80% confidence interval of model estimates. The larger dots and error bars represent the median and 80% confidence interval of combined values.



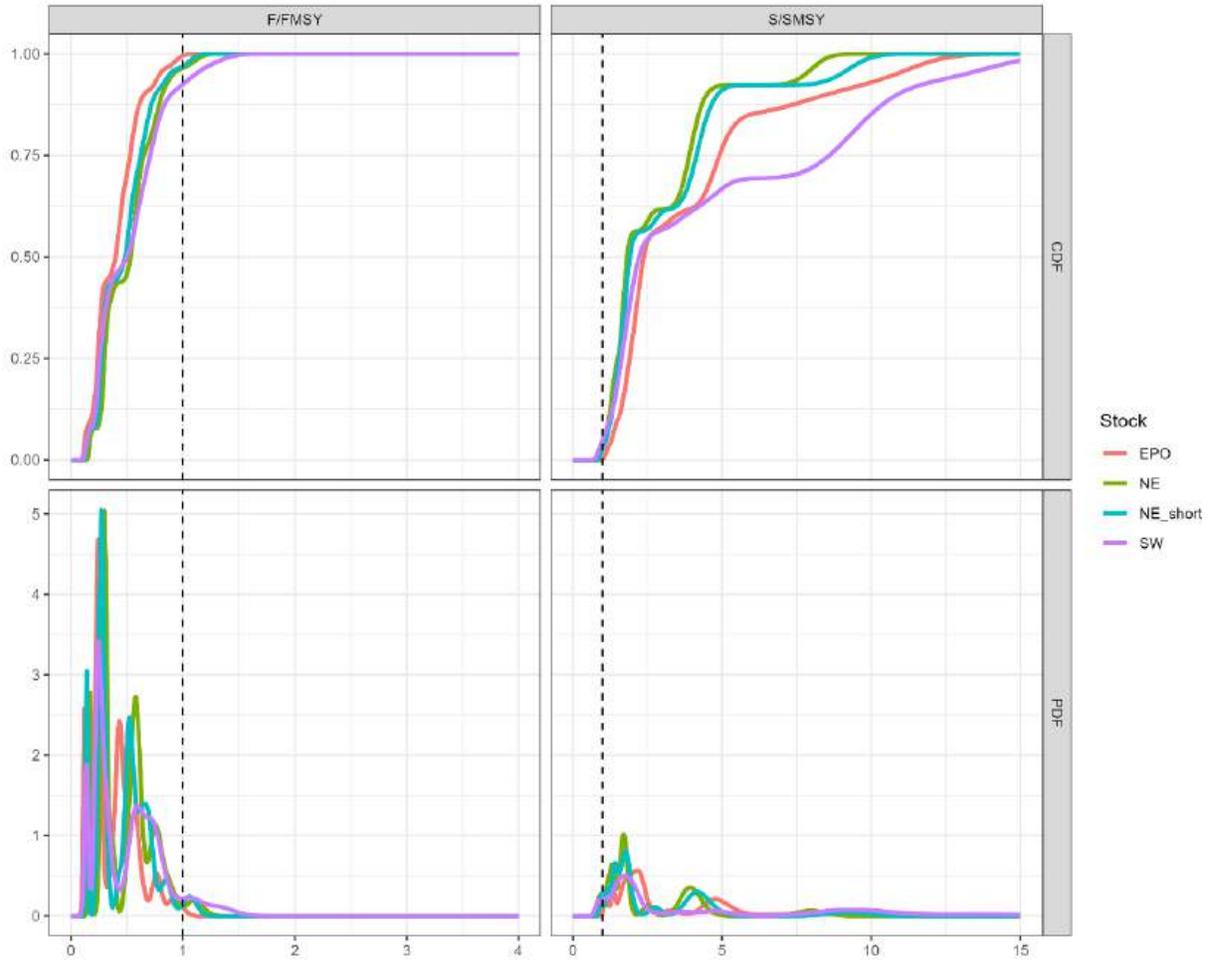
**FIGURE 21b.** Kobe plot of the most recent estimates of spawning biomass ( $S$ ) and fishing mortality ( $F$ ) relative to their proxy target reference points ( $30\%S_d$  and  $F_{30\%S_d}$ ) for each hypothesis of spatial structure. Each dot is based on the average  $F$  over the most recent three years, 2021-2023, and the  $S$  for the first quarter of 2024 and the error bars represent the 80% confidence interval of model estimates. The larger dots represent the combined result for each spatial structure hypothesis.



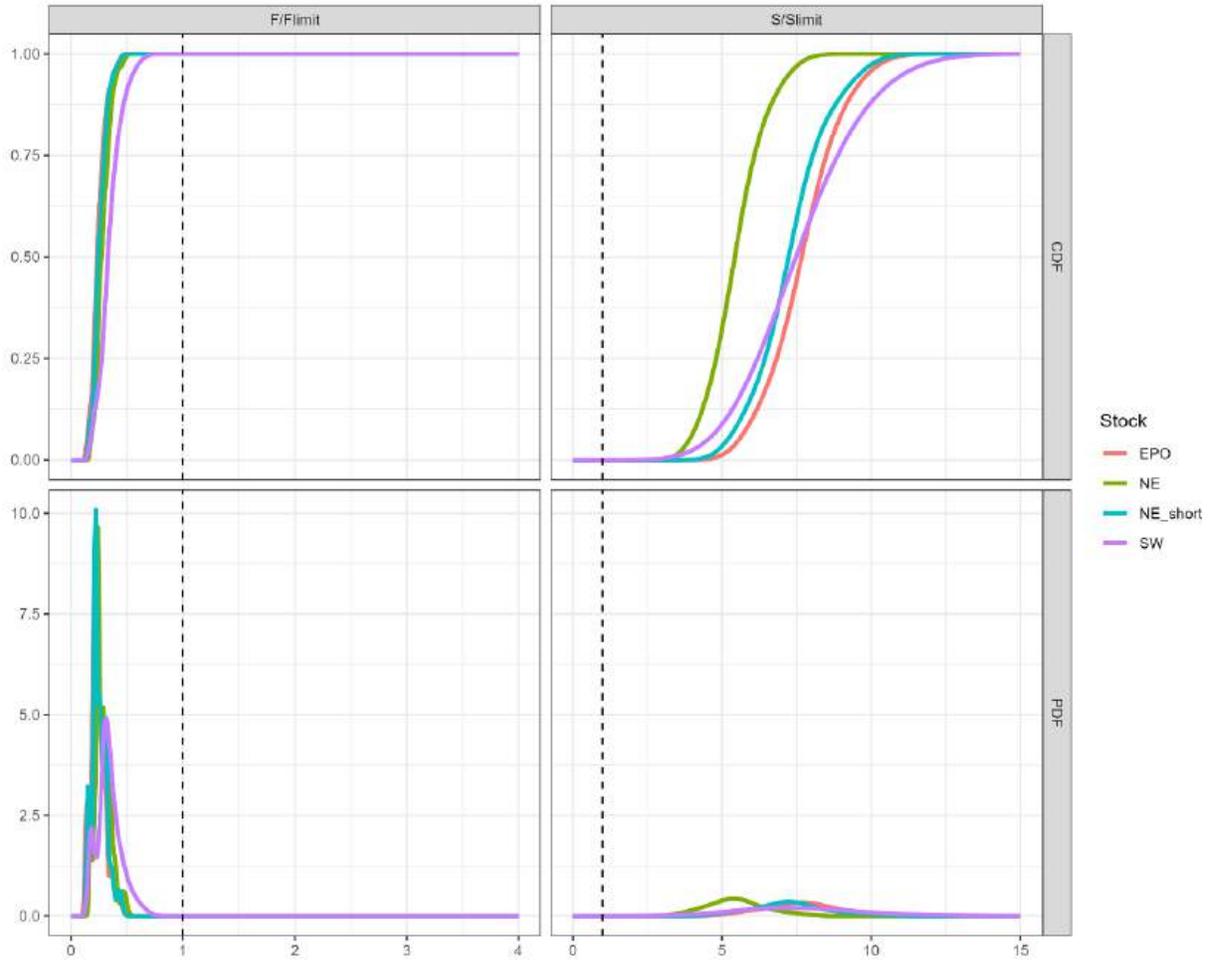
**FIGURE 22.** Kobe plot of the most recent estimates of spawning biomass ( $S$ ) and fishing mortality ( $F$ ) relative to their limit reference points ( $S_{Limit}$  and  $F_{Limit}$ ) for each stock structure. Each dot is based on the average  $F$  over the most recent three years, 2021-2023, and the  $S$  for the first quarter of 2024 and the error bars represent the 80% confidence interval of model estimates. The black dot and error bars represent the medium and 80% confidence interval of combined values, respectively.



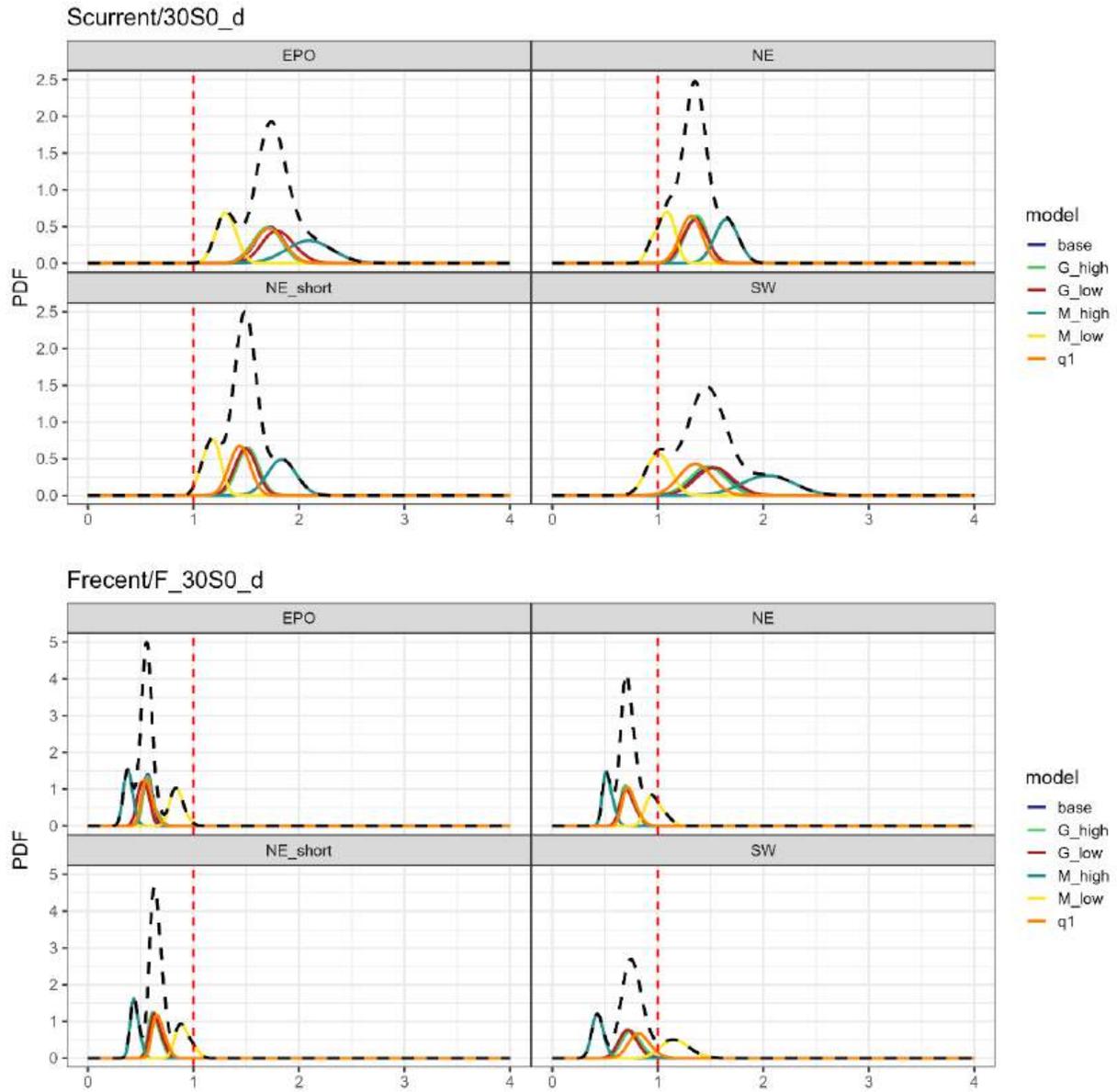
**FIGURE 23a.** The joint probability and cumulative distribution functions for spawning biomass ( $S$ ) in the first quarter of 2024 and fishing mortality ( $F$ ) in 2021-2023 relative to their proxy target reference points ( $30\%S_d$  and  $F_{30\%S_d}$ ) for each stock structure hypothesis.



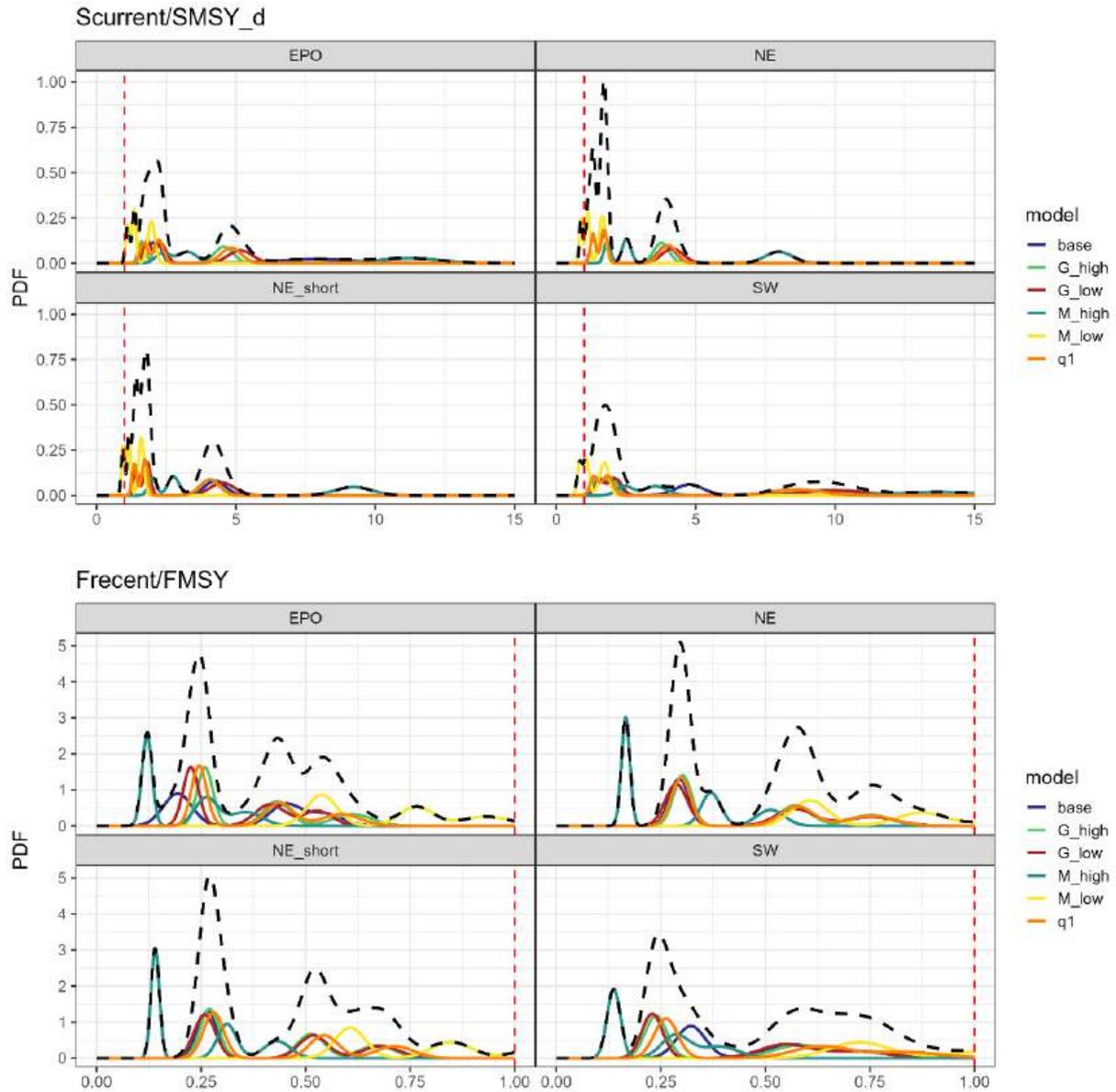
**FIGURE 23b.** The joint probability and cumulative distribution functions for spawning biomass ( $S$ ) in the first quarter of 2024 and fishing mortality ( $F$ ) in 2021-2023 relative to their MSY reference points ( $S_{MSY_d}$  and  $F_{MSY}$ ) for each stock structure hypothesis.



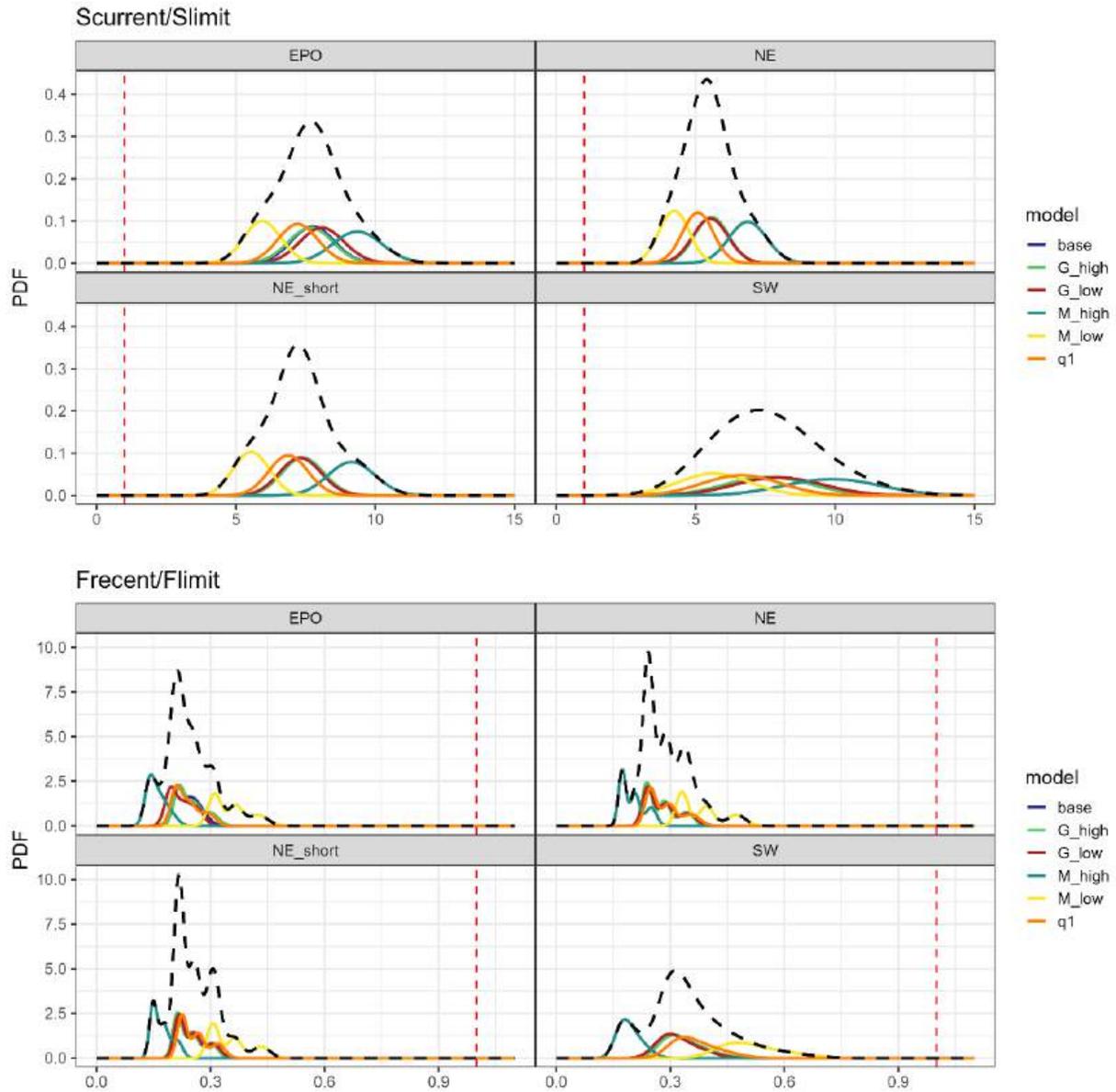
**FIGURE 23c.** The joint probability and cumulative distribution functions for spawning biomass ( $S$ ) in the first quarter of 2024 and fishing mortality ( $F$ ) in 2021-2023 relative to their limit reference points ( $S_{Limit}$  and  $F_{Limit}$ ) for each stock structure hypothesis.



**FIGURE 24a.** The joint probability distribution functions broken down into different components (level 2 hypotheses) of the four stock structure hypotheses for spawning biomass ( $S$ ) in the first quarter of 2024 and fishing mortality ( $F$ ) in 2021-2023 relative to their proxy target reference points ( $30\%S_d$  and  $F_{30\%S_d}$ ) for each stock structure hypothesis. The level 3 hypotheses (steepness values) were integrated out.



**FIGURE 24b.** The joint probability distribution functions broken down into different components (level 2 hypotheses) of the four stock structure hypotheses for spawning biomass ( $S$ ) in the first quarter of 2024 and fishing mortality ( $F$ ) in 2021-2023 relative to their MSY reference points ( $S_{MSY_d}$  and  $F_{MSY}$ ) for each stock structure hypothesis. The level 3 hypotheses (steepness values) were integrated out.



**FIGURE 24c.** The joint probability distribution functions broken down into different components (level 2 hypotheses) of the four stock structure hypotheses for spawning biomass ( $S$ ) in the first quarter of 2024 and fishing mortality ( $F$ ) in 2021-2023 relative to their limit reference points ( $S_{Limit}$  and  $F_{Limit}$ ) for each stock structure hypothesis. The level 3 hypotheses (steepness values) were integrated out.