CCSBT-ESC/2608/20
(ESC Agenda item 9)

The southern bluefin tuna stock assessment in 2026

Eveson J.P.[1], R. M. Hillary[1], A.L. Preece[1], N. Takahashi[2], T. Itoh[2], D.N. Webber[3]

[1] CSIRO, Environment, Hobart, Australia

[2] Fisheries Resources Institute, Japan Fisheries Research Agency, Yokohama, Japan

[3] Quantifish Ltd, Tauranga, New Zealand

Prepared for the 31st Extended Scientific Committee Meeting of the Commission for the Conservation of Southern Bluefin Tuna

24-28 August 2026

Brussels, Belgium

CCSBT-ESC/2608/20

Copyright

© Commonwealth Scientific and Industrial Research Organisation 2026. To the extent permitted by law, all rights are reserved and no part of this publication covered by copyright may be reproduced or copied in any form or by any means except with the written permission of CSIRO.

Important disclaimer

CSIRO advises that the information contained in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law, CSIRO (including its employees and consultants) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

CSIRO is committed to providing web accessible content wherever possible. If you are having difficulties with accessing this document please contact csiro.au/contact.

Acknowledgement of Country

CSIRO acknowledges the Traditional Owners of the lands, seas and waters of the area that we live and work on across Australia and pays its respects to Elders past and present. CSIRO recognises that Aboriginal and Torres Strait Islander peoples have made, and will continue to make, extraordinary contributions to Australian life including in cultural, economic, and scientific domains.

Acknowledgements

Preece, Hillary and Eveson were funded by the Australian Fisheries Management Authority (AFMA) and CSIRO. Webber was funded by the CCSBT.

We acknowledge data collection and analysis by CCSBT members and collation of data through CCSBT's scientific data exchange.

We acknowledge the advisory panel and members' contributions to the development of the new assessment code.

Contents

Abstract 4
1. Introduction 4
2. Input Data 5
3. Model and grid structure options 6
4. Data weighting & fits for the base model 7
   4.1 Data fits 7
   4.2 Data weighting 9
5. Stock status 12
6. Sensitivity tests 15
7. Projections 19
8. Summary 22
References 23
Appendix A: Data fit summary plots for the base model 25
Appendix B: Changes to natural mortality at age 37
Appendix C: Estimated selectivity-at-age by fishery for the base model 38

Abstract

The 2026 stock assessment for southern bluefin tuna provides estimates of current stock status (in 2026). The assessment software has changed to the new R-TMB code that allows for MCMC sampling and estimation of parameter uncertainty within each grid cell. The 2026 assessment models have a similar structure and population dynamics behaviour to the previous models (in ADMB). The 2026 assessment indicates further rebuilding of the stock since the last assessment in 2023. The fits to the data did not indicate any areas of concern. The sensitivity tests were defined to assess impacts on stock status estimates from additional areas of uncertainty. The current assessment indicates that the stock is close to the current rebuilding target (30% TRO, by 2035), and Total Reproductive Output (TRO) at MSY. Current relative TRO is estimated to be 0.28 (0.21-0.37). The projections for the reference set of models indicate that the management procedure rebuilding target will be reached earlier than planned, i.e. by 2028.

1. Introduction

This document describes the first stock assessment of southern bluefin tuna (SBT) using the new assessment model. The previous assessment was implemented in AD Model Builder (ADMB) (Hillary et al., 2023). The new assessment is implemented in R with RTMB (Kristensen, 2024), which provides an R interface to TMB (Template Model Builder; Kristensen et al., 2016). The new SBT model was validated against the ADMB code in 2025 (Anon, 2025a). The assessment code has changed to enable Markov Chain Monte Carlo (MCMC) sampling that accounts for parameter uncertainty within each grid cell and uncertainty across grid cells. The selected MCMC grid of models contains three steepness values (previously 4 values in 2023) and three values of the close kin mark-recapture (CKMR) power parameter for relative reproductive contribution by age, psi (ψ), giving nine cells rather than the previous 108-cell grid (Anon, 2026). The two natural mortality parameters previously included in the grid are now estimated within each MCMC cell. This design represents continuous parameter uncertainty within the reduced grid instead of treating every non-grid parameter only at its maximum-likelihood estimate.

The population dynamics and general model structure remain similar to the previous assessment, but some parameterisations and observation models have changed. These changes are:

- **Selectivity:** a two-dimensional AR(1) process that smooths correlated effects over age and year, where temporal variation is permitted (Anon, 2025b).
- **Natural mortality (*M*):** a length-scaled curve from age 0 through age 25, namely $M_a = M_{10}(L_a/L_{10})^{m_c}$, using mean length-at-age L_a from the first year and season of

the model, followed by a linear increase from M_{25} to M_{30}. Parameters M_{10} and M_{30} are estimated. The exponent m_c is fixed at -1, which makes mortality inversely proportional to length for ages up to 25 (as described by Lorenzen 1996) (see Appendix B for details).

- **CKMR POP likelihood:** when an adult is not directly aged, its capture length is used and the likelihood integrates over its possible ages (Anon, 2024).
- **Length and age compositions:** a wider set of likelihood options, including compound Dirichlet-multinomial and Dirichlet formulations (Anon, 2024).
- **CPUE selectivity:** CPUE-weighted Japanese longline length frequencies estimate the selectivity associated with the Japanese GAM22 CPUE index (Anon, 2025b).
- **CV of CPUE data series:** the year-specific CVs from GAM22 are added to the historical constant CV of 20% (on the log-scale), giving CVs on the linear scale ranging from 20.24% to 30.10%, with mean 21.34% (Itoh et al., 2026).

The previous SBT assessment used several data weighting methods for the abundance, tagging, close-kin, and composition data used in the models (Hillary et al., 2023). This assessment uses one-step-ahead (OSA) randomized quantile residuals (Dunn and Smyth, 1996; Stewart and Monnahan, 2025) as a diagnostic tool for assessing data weighting. OSA residuals are easier to interpret than standardized Pearson residuals for many non-Gaussian observations. The diagnostics inform, but do not determine, the weights selected for this assessment.

2. Input Data

The input data have been updated since the 2023 assessment (Hillary et al., 2023) to include three additional years (Preece et al., 2026). A new data input is a CPUE-weighted Japanese longline length-frequency series used to estimate the selectivity-at-age associated with the CPUE index. The data inputs for the 2026 assessment are:

- **Catch biomass:** total catch biomass for each of six fisheries[1], used as a conditioned input (CCSBT, 2026).
- **Length compositions:** observations for the four longline fisheries (LL_{1-4}).
- **Age compositions:** observations for the Indonesian and Australian surface fisheries.

[1] LL1: Japanese LL areas 4-9 and other longline catches, LL2: Taiwanese albacore LL and gillnet, LL3: Japanese LL in Area 2, LL4: Japanese spawning (Area 1), Indonesian, Australian surface.

- **Longline CPUE:** a standardized relative abundance index in numbers. (Itoh and Takahashi, 2026).
- **CPUE-weighted Japanese length compositions:** Japanese longline size-frequency data weighted by CPUE rather than catch.
- **Aerial survey:** a relative biomass index for ages 2–4 (Eveson and Farley, 2017).
- **Conventional tags:** juvenile releases and subsequent recaptures from the 1990s conventional tagging program.
- **Gene tagging:** number of age 2 release samples, age 3 harvest samples from same cohort, and matches found between them (Preece et al., 2026)
- **CKMR parent–offspring pairs (POPs):** juvenile cohort, Adult Capture Year, Adult Age, Number of POPs and number of comparisons between sampled juvenile and adult fish (Farley et al, 2026).
- **CKMR half-sibling pairs (HSPs):** cohorts, number of comparisons and number of kin from sampled age-3 juveniles (Farley et al, 2026).
- **UAM (unaccounted mortality) estimates**: estimates of the potential scale of non-member (UAM) catches (Edwards and Hoyle, 2026).

3. Model and grid structure options

The population dynamics retain the previous assessment's annual age-structured model with two seasons (season is specified for each of the 6 fisheries). The uncertainty grid crosses three steepness values with three values of the CKMR reproductive parameter, psi (ψ) (Anon, 2026):

Variable	Values	Grid prior	Sampling protocol
Steepness (h)	0.6, 0.7, 0.8	Uniform	Balanced across cells
Psi (ψ)	1.5, 1.75, 2	Uniform	Balanced across cells

The previous assessment constructed a 2,000-member "reference set" of models by resampling the grid elements according to either their prior or posterior distribution. The current assessment samples the posterior within each of the nine grid cells. The final 2,000-draw combined posterior is sampled approximately evenly across the nine cells, which we will refer to as the reference set for consistency with previous terminology. This reference set is used for calculating stock status estimates and for projections. Equal-tailed 95% posterior credible intervals for a given parameter or derived quantity (referred to simply as 95% credible intervals) are calculated as the empirical 0.50, 0.025 and 0.975 quantiles of the 2,000 values from the reference set.

4. Data weighting & fits for the base model

Data-weighting calculations and data fit summaries use the base model, which corresponds to the mid grid cell ($h = 0.7, \psi = 1.75$). Pareto-smoothed importance-sampling leave-one-out information criteria (PSIS-LOO) and the Pareto shape diagnostic are used to assess predictive performance of the base model (Vehtari et al., 2024, 2017). OSA standardized normal residuals and their standard deviation (SDNR) are evaluated at the base model maximum-likelihood estimate. Fits to the data are in Figures in Appendix A.

4.1 Data fits

Fits to the abundance-related data sets:

- *Longline CPUE index*: data points are overwhelmingly contained within the predictive distribution (see Figure A1 in Appendix A) with no obvious trends in the residuals beyond moderate amounts of autocorrelation. There is an apparent under-estimation of the earliest and most recent historically highest CPUE points, very likely driven by the time-dependent weighting of these data based on the relative CV from the standardisation
- *Aerial survey index*: as in previous assessments, barring the historically high 2017 data point, the survey observations are all contained within the predictive interval (see Figure A2) with no obvious long-term residual trends
- *Gene tagging*: the gene tagging matches (Figure A3) are all contained within the predictive interval and fit well and without trend
- *Conventional tagging*: fits to the tag recaptures by release year, release age, and recapture age, pooled across tagger groups, are shown in Figure A4. In general, the fits are good. Recapture estimates are biased for a few release years and ages (e.g. recaptures are underestimated for age-1 releases in 1991) but look very similar to fits from the 2023 assessment, so the change to the natural mortality schedule has not had a negative impact on the conventional tag fits.
- *CKMR POPs*: as in previous assessments we aggregate these data to (i) the juvenile cohort, (ii) the adult capture age, and (iii) the adult capture age level - see Figures A5, A6 and A7, respectively. At all aggregation levels the data are well explained by the predictive distribution, and there are no obvious trends in the fits
- *CKMR HSPs*: the fits to these data at the base level (juvenile cohorts of each comparison group) are summarised in Figure A8 with the vast majority of the points included within the predictive interval. These data are aggregated to the initial cohort level (Figure A9) and all but one of the points are contained within the predictive interval and fitted without apparent trend

- *Overall matches for genetic data sets*: Figure A10 summarises the total (overall years) observed and predicted matches for the genetic data sets (gene tagging and CKMR). We see that for each data set the observed number of matches is very close to the predicted number and well encompassed by the predictive interval

Fits to the catch composition data sets for each fishery:

- LL_1: as in previous assessments these data are fairly well fitted (see Figure A11) using a mix of 3, 4 and 5 year selectivity blocks
- LL_2: Figure A12 shows the summary of the fits to these data which are good overall
- LL_3: this fishery shows more change over time, in both samples taken and shape and location of the size distribution (see Figure A13) than the other longline fisheries. When there are samples, and appreciable levels of catch taken, the fits are fine and really relate to the earliest periods in the fishery
- LL_4: a direct removals approach was investigated for this fishery but caused issues in both optimization and MCMC (e.g., harvest rates pushed too high in some sampled states, resulting in negative numbers of fish for some older age classes), so the choice was made to estimate a single selectivity for this exclusively historical fishery (1964–1990) and the fits are acceptable (see Figure A14)
- *Japanese CPUE fleet length data*: these data have a lot of overlap with LL_1, but are quite different at times. Fits are fairly good (see Figure A15) though note the main difference with LL_1 is the data are cut-off at the expected length at age 4 given this is done to the catch and effort data used to create the standardised CPUE index
- *Indonesian fishery ages*: a mix of 1, 2 and 3-year selectivity blocks are permitted for fitting these data (see Figure A16) to account for strong changes in year to year at times. Note the shift to a stronger and consistent bimodal pattern in the most recent years as area 2 catches are included in the fishery. Fits are similar to previous assessments.
- *Surface fishery ages*: fits are good and similar to previous assessments. Selectivity is permitted to change from year to year from the mid-1990s onwards (see Figure A17).

A plot of the natural mortality curve estimated for the base model is shown in Figure B1 of Appendix B. A sensitivity run was specified to test the effect of the new formulation on the stock status results (see section 6).

Figures showing the estimated selectivity-at-age for each of the fisheries, including the new CPUE fleet, are given in Appendix C. The new AR1 formulation gives less flexibility to the Indonesian selectivity, which has led to an increase in the estimate of natural mortality at age 10 (see Anon. 2026 for a detailed discussion). A sensitivity test ('Indo_sel') was

specified to restore flexibility in the Indonesian selectivity in order to evaluate the impact of this change on stock status results (see section 6).

4.2 Data weighting

Previous assessments deliberately down-weighted CPUE, up-weighted the aerial survey, retained the input composition weights, and investigated the tagging and close-kin weights using data-specific methods (Hillary et al., 2023). The current method for reviewing relative weighting uses OSA residuals to place those diagnostics on a common standardized-normal scale (Figure 1).

For a correctly specified normalized residual, an SDNR near one indicates that the residual spread is broadly consistent with the assumed observation error. That is a diagnostic target, not a rule that every data set must be tuned to one. The accepted SDNR summary shows:

- CPUE has an SDNR near 0.8, consistent with its deliberate conservative weighting.
- The aerial survey, gene tagging, POP, HSP, and conventional-tag SDNRs are near one, with uncertainty intervals that include one.
- Length- and age-composition SDNRs are approximately 0.5–0.75, indicating narrower residuals than the multinomial diagnostic would predict.

Up-weighting the age and length composition data was considered, but it worsened fits to abundance information, particularly HSPs, and did not resolve the narrower residuals (Anon. 2026). Therefore, the base model composition weights were retained. This is consistent with the principle that composition fits should not be improved at the expense of abundance data (Francis, 2011).

The PSIS-LOO diagnostics (Figure 2) quantify predictive contribution and identify influential observations in each data set. They provide insight into which data sets and single observations have a strong influence on the model (Vehtari et al., 2024, 2017). The data sets with the most apparent influence on the model predictions (in a very general sense) are the age and length frequency data – in particular the Indonesian and surface age data, and the length data for longline fleets 1 and 3 and the CPUE “fleet”. The abundance data sets (CKMR, aerial survey, LL CPUE, conventional tags and gene tagging) perform more consistently. Some of the gene tagging observations exert a strong influence on recent estimates in the models, most likely because there are no other data sets that observe the most recent cohorts.

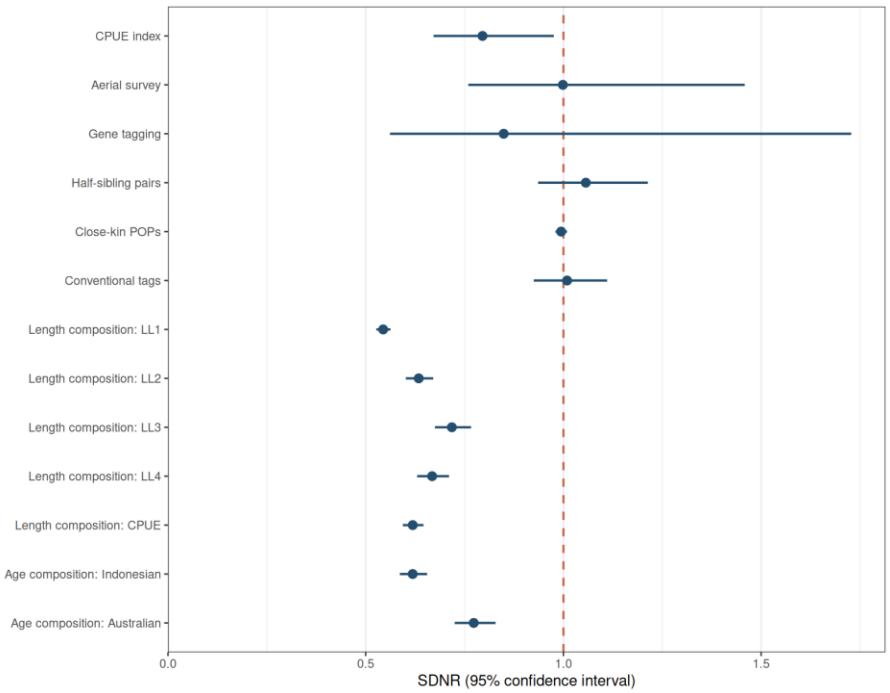


Figure 1. OSA residual SDNR estimates (points) and approximate confidence intervals (whiskers) for assessment data components.

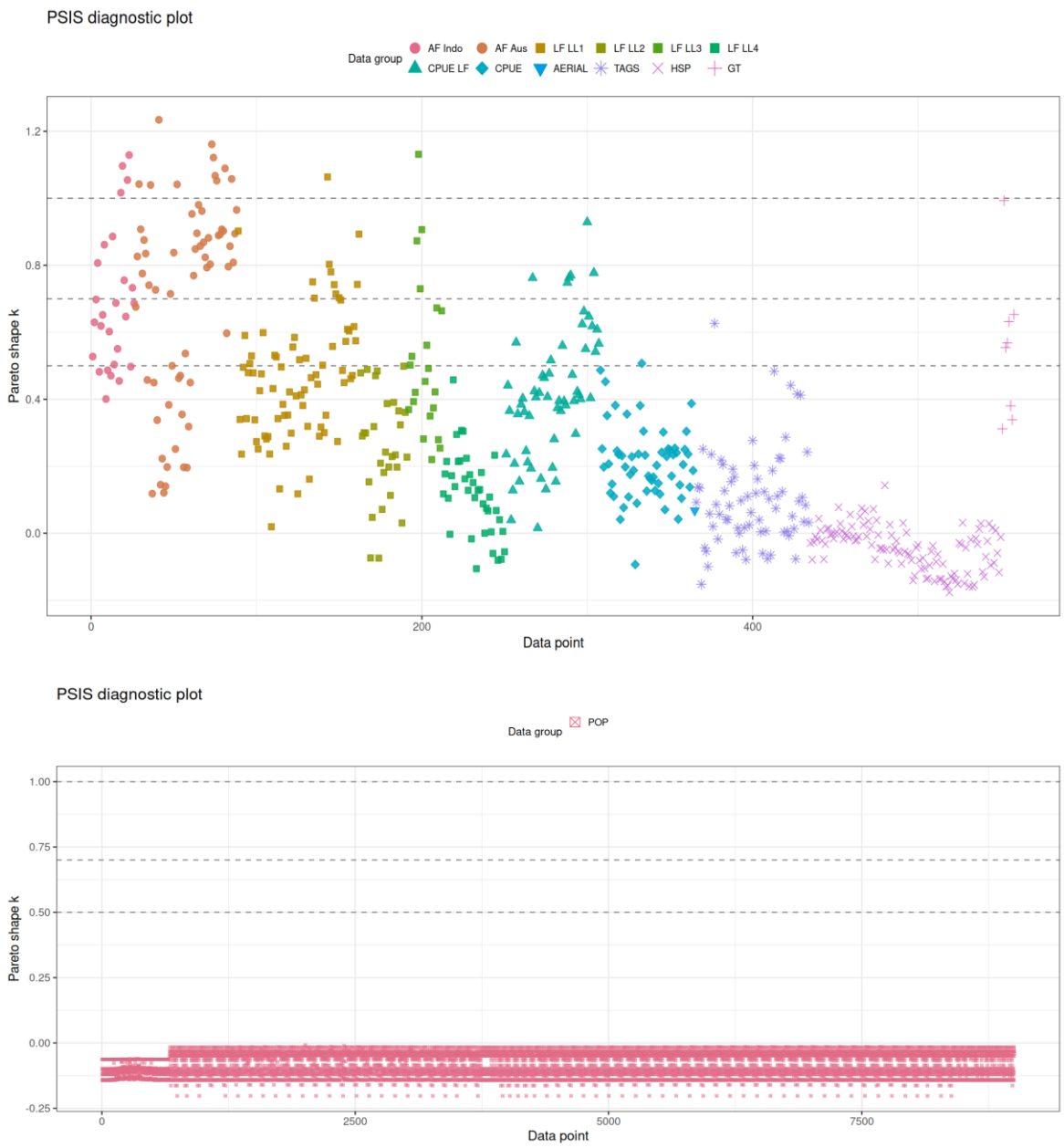


Figure 2. (top) Pareto shape parameter k from the PSIS-LOO cross-validation for all data sets other than POPs, which are shown separately (bottom) because of their much larger number of observations (n=5075). Horizontal reference lines are drawn at k = 0.5, 0.7, and 1.0, with k<0.5 meaning reliable PSIS-LOO estimates, k>0.7 suggesting the estimates may be unreliable, and k>1 meaning the approximation is generally not trustworthy.

5. Stock status

Results from the reference set of models (i.e., the 2,000 draws sampled from the MCMC grid) are used to provide stock status summary statistics (Table 1). Relative total reproductive output (TRO) is TRO divided by unfished TRO for the same draw.

The 2026 estimate of depletion in terms of relative TRO is 0.281 (0.209-0.373) (Table 1). This is an increase from the 2023 estimate obtained from the current reference set as well as from the 2023 assessment. The estimate of depletion of age 10+ biomass in 2026 is 0.231 (0.171-0.313), which is similar to 2023 using the new reference set. Further examination of age structure change in the population from 2023 to 2026, for the base model run, showed a decline in numbers at ages 10–14, increase in numbers at ages 8–9, and an increase in numbers at ages 20–24, which resulted in an increase in TRO but static B10+.

Figure 3 and Figure 4 show the estimated trajectories for relative depletion in TRO and historical recruitment obtained from the reference set of models. The 95% credible interval for recruitment becomes wide in recent years due to lack of data to inform recruitment estimates (the last abundance estimate from gene-tagging is for age-2 fish in 2024, which is the 2022 cohort). Relative TRO has increased since the mid-2000s, and the probability of relative TRO being below 0.2 has decreased to 1.1% in 2026 (Table 1).

MSY in 2026 is estimated to be slightly lower than in 2023 based on median values from the reference set of models; however, it is higher than the median 2023 estimate from the 2023 assessment and has a wider range (Table 1). The fishing mortality (F) and TRO at MSY relative measures have very wide ranges of values, which make them less informative; nevertheless, current F is well below F_{MSY} (median F to F_{MSY} ratio of 0.553). Current TRO is estimated to be at 0.985 of TRO_{MSY}, with the range including values greater than 1.0 (0.629-1.587); this is an increase from the 2023 assessment result.

Table 1. Stock status and equilibrium reference points for 2023 and 2026 derived from the reference set of models (shown are posterior medians and 95% credible intervals). The 2023 estimates from the 2023 stock assessment are included for comparison.

Year (y)	Relative TRO (TRO_y/TRO_0)	P(Relative TRO < 0.2)	Relative B10+	F/F_MSY	TROy/ TRO_MSY	MSY (t)
2023 from previous assessment in 2023	0.23 (0.21-0.29)	-	0.22 (0.19-0.26)	0.46 (0.34-0.65)	0.85 (0.61-1.29)	30,648 (29,152-31,376)
2023 from new assessment	0.255 (0.189-0.338)	0.068	0.231 (0.171-0.310)	0.567 (0.397-0.805)	0.929 (0.590-1.495)	32,984 (26,813-40,131)
2026 from new assessment	0.281 (0.209-0.373)	0.011	0.231 (0.171-0.313)	0.553 (0.383-0.811)	0.985 (0.629-1.587)	32,529 (26,400-39,381)

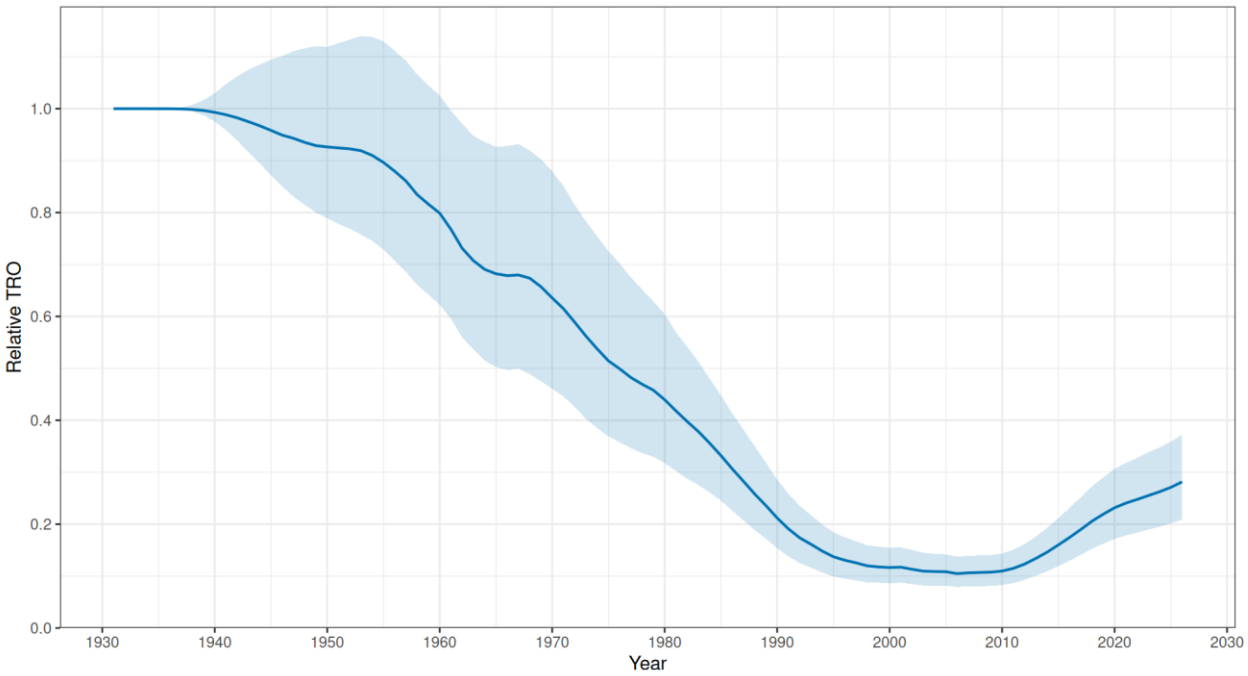


Figure 3. Relative total reproductive output trajectory for the reference set of models. The line shows the posterior median and the shaded ribbon shows the 95% credible interval.

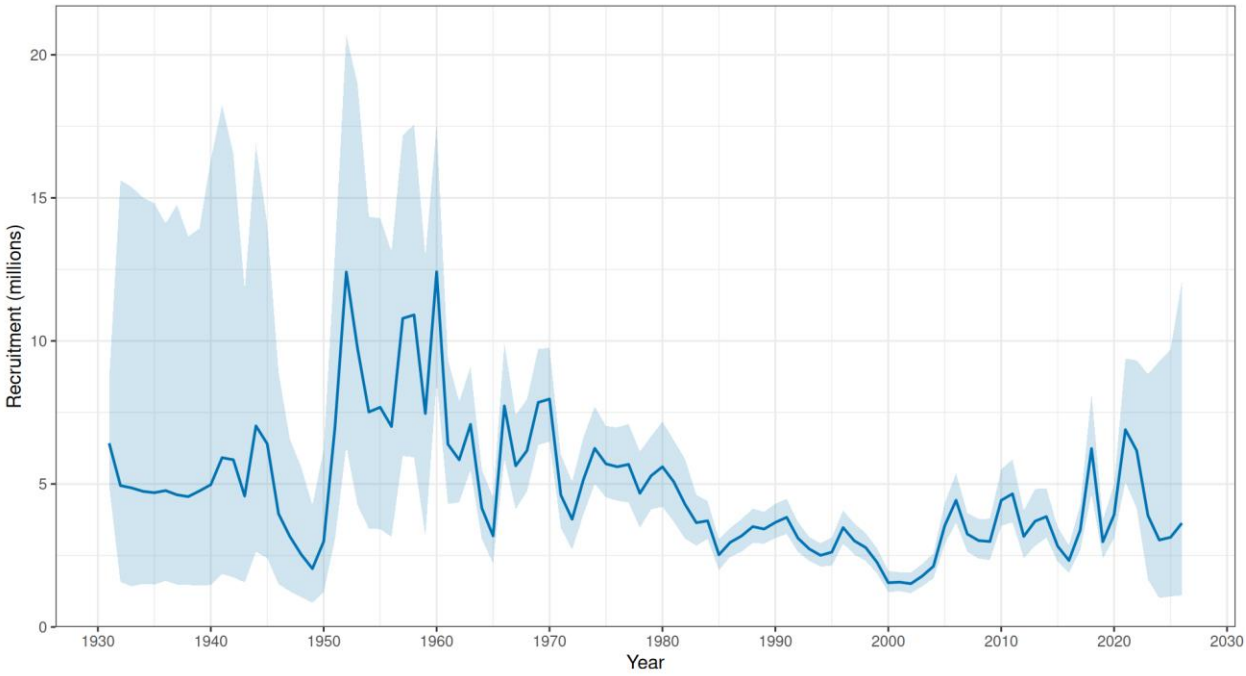


Figure 4. Recruitment trajectory for the reference set of models. The line shows the posterior median and the shaded ribbon shows the 95% credible interval.

6. Sensitivity tests

A set of 12 sensitivity and projection tests were specified at the 2026 OMMP meeting (Anon. 2026), and three additional sensitivity tests were added after the meetings to give a total of 15 (Table 2).

Table 2. Sensitivity tests defined at the OMMP meeting and three additional tests defined after the meeting (marked as "Additional" in the Priority column).

Test name	Conditioning and projection notes	Grid	Priority
No UAM	Remove non-member catches (unaccounted mortalities, UAM) from conditioning and projections.	Mid-cell	H
CPUE_Drop5	Eliminate the last 5 years of CPUE Series	Mid-cell	H
Omega75	Power function for biomass-CPUE relationship with power = 0.75	Mid-cell	H
Upq2008	Estimate CPUE 2008 change in q for LL1 fleet	Mid-cell	H
LL1_sel	Allow the terminal 3-years to be flexibly estimated to evaluate impact on year-class uncertainty and magnitude	Mid-cell	M
Indo_sel[1]	Restore bi-modality in Indonesian selectivity, more flexibility	Mid-cell	H
NoPOP&HSP	Exclude both close-kin data (Parent-Offspring and Half-Sibling Pairs)	Mid-cell	H
No HSP	Exclude half-sibling-pair close-kin data	Mid-cell	H
GTI[2]	Includes the grid-type trolling index (Itoh 2026) as additional recruitment index. Increase CV of aerial survey to preclude aerial survey dominating the fit, given apparent conflicts in the data	Mid-cell	H
Projections	Alternative allocations	Full grid	H
OldGrid	E.g., 108 (or other) dimension MLE with updated data and values for M0 and M10 (and maybe psi)	Full grid	H
CPUE_fleet	Evaluate impact relative to assuming LL1 selectivity for CPUE predictions (as in previous approach)	Mid-cell	H
CPUE_CV20	Restore the former time-invariant CPUE uncertainty with zero input log-SD and fixed additional log-scale sigma = 0.2	Mid cell	Additional
Old_m	Estimate M0, M4, M10, and M30 using the direct four-anchor natural-mortality parameterization	Mid cell	Additional
Estimate_m_slope	Estimate the length-based natural-mortality exponent while retaining the selected base parameterization	Mid cell	Additional

[1] Achieved by changing the Indonesian separable-AR1 hyperparameters from the base model values (rho_year = 0.98, rho_age = 0.98, sigma = 0.12) to values allowing for greater flexibility (rho_year = 0.5, rho_age = 0.5, sigma = 0.5).

[2] For GTI sensitivity results presented here, the CV of the aerial survey was kept the same as in the base model (tau=0.59).

All sensitivity tests passed the MLE and MCMC acceptance criteria except for the NoPOP&HSP test, which had two divergent transitions but was still considered acceptable for inclusion in comparisons.

The estimates of current stock status (relative TRO) are reasonably consistent across the sensitivity tests (Table 3). Most sensitivity test results are the same as, or slightly more optimistic than, the base model. The largest changes in the 2026 relative TRO from the base model occur for Upq2008 (for which CPUE catchability is split in 2008) and Old_m (which restores the previous direct four-anchor natural-mortality parameterization) (Table 3). The highest current stock status estimates are from the Old_m scenario (relative TRO of 0.339 compared to 0.282 for the base model). The 'Upq2008' and 'No_Pop&HSP' sensitivity tests are most pessimistic relative to the base model but still have median relative TRO values of 0.239 and 0.266, which are greater than the original interim rebuilding target of 0.2.

Table 3. Stock status and equilibrium reference points for the sensitivity tests, with the base model results included for comparison (shown are posterior medians and 95% credible intervals). Relative TRO and relative B10+ are given for 2026; MSY quantities use 2025, the latest fitted catch year. NoPOP&HSP is included under the decision to accept two divergent transitions.

Model	Relative TRO (2026)	Relative B10+ (2026)	F/F_MSY (2025)	TRO/ TRO_MSY (2025)	MSY (t) (2025)
Base	0.282 (0.222-0.353)	0.232 (0.180-0.294)	0.553 (0.466-0.649)	0.987 (0.779-1.237)	32,451 (26,663-39,862)
No UAM	0.290 (0.233-0.362)	0.240 (0.189-0.301)	0.498 (0.424-0.588)	1.014 (0.811-1.265)	31,825 (26,315-38,504)
CPUE_Drop5	0.270 (0.214-0.338)	0.223 (0.176-0.281)	0.592 (0.491-0.710)	0.952 (0.758-1.188)	32,336 (26,708-39,587)
Omega75	0.292 (0.233-0.366)	0.243 (0.191-0.307)	0.521 (0.441-0.615)	1.023 (0.811-1.276)	33,787 (27,334-41,498)
Upq2008	0.239 (0.182-0.305)	0.192 (0.143-0.250)	0.611 (0.511-0.733)	0.831 (0.633-1.062)	32,423 (26,587-39,954)
LL1_sel	0.283 (0.225-0.356)	0.230 (0.181-0.293)	0.570 (0.479-0.673)	0.981 (0.778-1.233)	31,835 (26,114-38,887)
Indo_sel	0.279 (0.224-0.345)	0.230 (0.183-0.288)	0.539 (0.458-0.639)	0.975 (0.783-1.201)	32,246 (26,704-39,756)
NoPOP&HSP	0.266 (0.201-0.349)	0.216 (0.159-0.292)	0.586 (0.476-0.716)	0.931 (0.701-1.229)	32,606 (26,684-39,950)
No HSP	0.287 (0.226-0.363)	0.238 (0.184-0.304)	0.548 (0.459-0.654)	1.007 (0.793-1.279)	32,655 (26,608-39,735)
GTI	0.287 (0.229-0.357)	0.246 (0.194-0.308)	0.629 (0.540-0.735)	1.024 (0.817-1.271)	31,686 (26,251-38,301)
CPUE_fleet	0.294 (0.233-0.363)	0.242 (0.190-0.304)	0.536 (0.450-0.634)	1.027 (0.815-1.271)	32,597 (26,589-39,760)
CPUE_CV20	0.283 (0.224-0.352)	0.232 (0.182-0.292)	0.545 (0.459-0.643)	0.989 (0.781-1.236)	32,557 (26,742-40,189)
Old_m	0.339 (0.266-0.432)	0.258 (0.204-0.331)	0.471 (0.385-0.565)	1.153 (0.911-1.461)	33,435 (27,070-41,387)
Estimate_m_slope	0.287 (0.228-0.362)	0.234 (0.184-0.295)	0.546 (0.460-0.652)	1.004 (0.795-1.268)	32,751 (26,838-40,262)

The OldGrid sensitivity test compares results using the MCMC grid with results using an MLE grid structure similar to the 2023 assessment. For the MLE grid, the "Old_m" parameterisation for natural mortality was used, with parameters M4 and M30 estimated

and M0 and M10 fixed grid elements. The MLE grid included the four steepness values and three psi values from the 2023 assessment (h = 0.55, 0.63, 0.72, 0.8 and psi = 1.5, 1.75, 2), and three M0 and three M10 values informed by the MCMC posterior distributions (M0 = 0.339, 0.368, 0.389 and M10 = 0.099, 0.107, 0.114), giving 108 grid cells in total. As we would expect, relative TRO has greater uncertainty using the MCMC-based reference set than the sampled deterministic MLE grid, since it takes within-cell parameter uncertainty into account (Figure 5). In 2026, the relative TRO from the reference set has a median of 0.281 (95% interval 0.209–0.373), whereas the deterministic MLE grid median is slightly lower and narrower (median 0.280; 95% interval 0.228–0.329). Note that for the sampled MLE grid, 95% intervals are calculated using the 0.025 and 0.975 quantiles of the 2,000 sampled values, which is the same as for the MCMC reference set but the interpretation is different. For the MCMC reference set these are posterior credible intervals; for the deterministic MLE-grid resample they are weighted grid-resampling intervals, not posterior credible or frequentist confidence intervals.

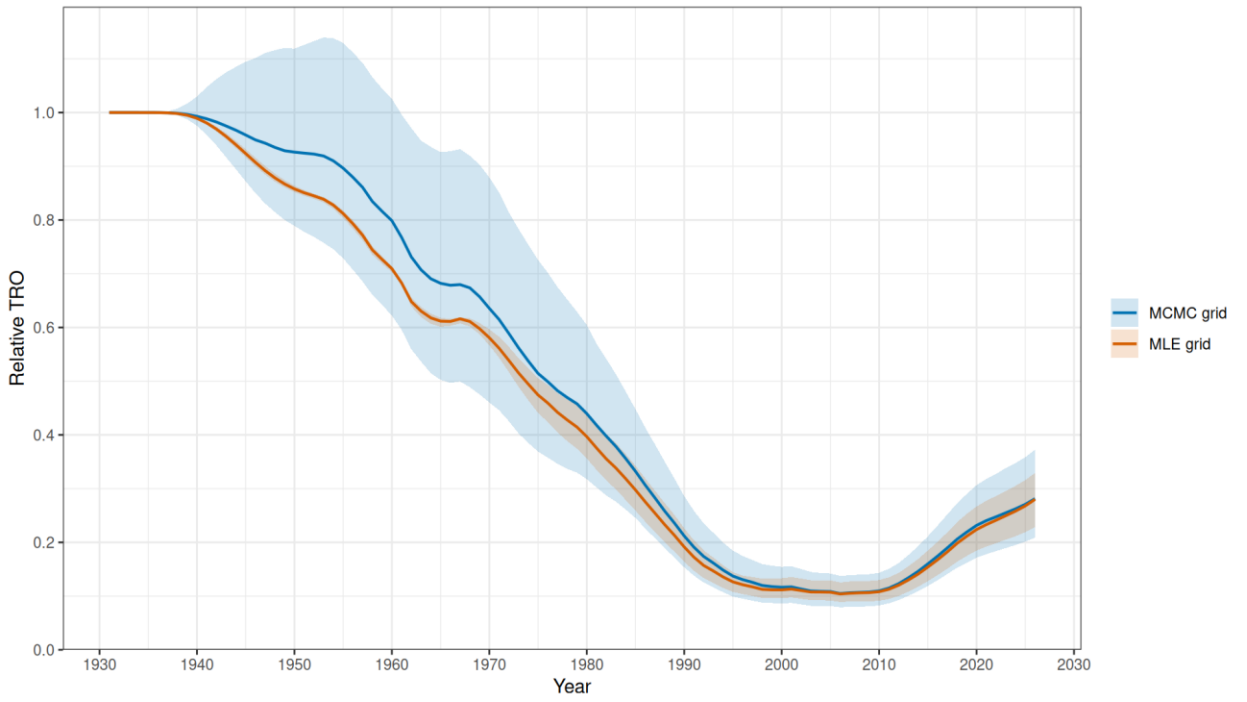


Figure 5. Relative total reproductive output comparison between the MCMC reference set of models and the sampled MLE grid. Lines show medians and shaded areas show 95% intervals (refer to text for explanation of intervals).

7. Projections

Given the major update in both the functionality and the software platform used for the stock assessment, the projection code and current MP (the Cape Town Procedure or CTP) have also been updated. One major change to the mechanics of the projections is the modelling of future recruitment uncertainty. Previously the model used empirical estimates of variance and lag-1 autocorrelation. Now we use an ARIMA (Auto-Regressive Integrated Moving Average) model where the full time-series characteristics of each recruitment history (i.e. per MCMC sample) are estimated and then used in generating future recruitment uncertainty.

The projection results (reference set) show there is a 71% probability of being greater than 0.30 TRO_0 in 2035, which is above the target of 50% (Table 4). There is also a very high probability (99%) of being greater than the original interim target of 0.2 TRO_0 by 2035 (the original interim target was 70%). The estimated depletion in 2035 is 0.339 (0.218-0.542). All of these statistics are an improvement from the 2023 assessment projection results. The mean TAC from 2026 through to 2035 is 25,399 t (25,349-26,249), indicating that current global TAC is approaching upper limits under the current reference set.

Projections were run for two allocation scenarios for the 3000 t increase recommended for 2027-2029 to examine impacts of an alternative scenario: (i) nominal allocations (reference set) and (ii) allocated to Indonesia. Results are almost identical for both allocation scenarios.

Table 4. Projection summary statistics derived using the reference set of models and two options for allocations, and the sampled MLE grid with only nominal allocations. For the reference set results, posterior medians and 95% credible intervals are shown; for the MLE grid, the intervals are the weighted grid-resampling spread, not posterior credible intervals.

Scenario	**P(TRO_{2035} > 0.2 TRO_0)**	**P(TRO_{2035} > 0.3 TRO_0)**	**TRO_{2035}/TRO_0**	**Mean TAC 2026-2035 (t)**
Reference set	0.990	0.712	0.339 (0.218-0.542)	25,399 (25,349-26,249)
Alternative allocation	0.992	0.716	0.341 (0.220-0.544)	25,400 (25,349-26,249)
Sampled MLE grid — nominal allocations	1.000	0.730	0.337 (0.245-0.490)	25,392 (25,349-26,249)

Also included in Table 4 is a comparison of the projection results using the sampled MLE grid to see how much using the MCMC grid as the basis for the reference set has affected

the results. The median values of the summary statistics are similar, but the uncertainty ranges are wider (as expected) using the MCMC grid. Figure 6 shows the projected TRO trajectories, both relative and absolute, out to 2036. For comparison results from the reference set (i.e, the sampled MCMC grid) are shown alongside results from the sampled MLE grid. Several points are worth noting:

- The future uncertainty in the TRO dynamics is notably wider using the MCMC grid relative to the MLE grid. The MCMC approach, as expected, gives a more realistic representation of the model uncertainty as a whole, relative to the MLE grid approach that focused on structural over parametric uncertainty.
- The full MCMC grid appears to show a very slightly higher level of TRO by 2036 and with a flat trend, not the slightly decreasing trend seen in the MLE grid from 2034.
- In both cases (MCMC vs. MLE) the projected relative TRO is above the tuning target level (probability of 0.5 of being at 30% unfished) by 2035 and the median reaches this level by 2028 (Figure 6).
- Even with the additional uncertainty in the MCMC grid there is still a very low probability of being below 20% unfished in the projections, even for the most pessimistic of runs from the MCMC grid.

Figure 7 shows the projected TAC out to 2035 under the action of the CTP. Catches are already set out to 2029, so the first TAC projected decision is implemented in 2030 and then in 2033. In 2030, the projected estimated catch increase is 3,000 t for all runs (Figure 7). However, in 2033 we see a 50% chance of no change in TAC, zero probability of it going down, and a range of possible increases up to and including the maximum level of 3,000 t. Based on median values, the TAC will not change from 2030 onwards - this is a specific design feature in that the CTP was designed to become less reactive as the target TRO rebuilding level is approached. This demonstrates that the CKMR part of the CTP is detecting the rebuilding target being approached around 2030 and changes the features of the CTP (from being reactive to a more passive phase) accordingly.

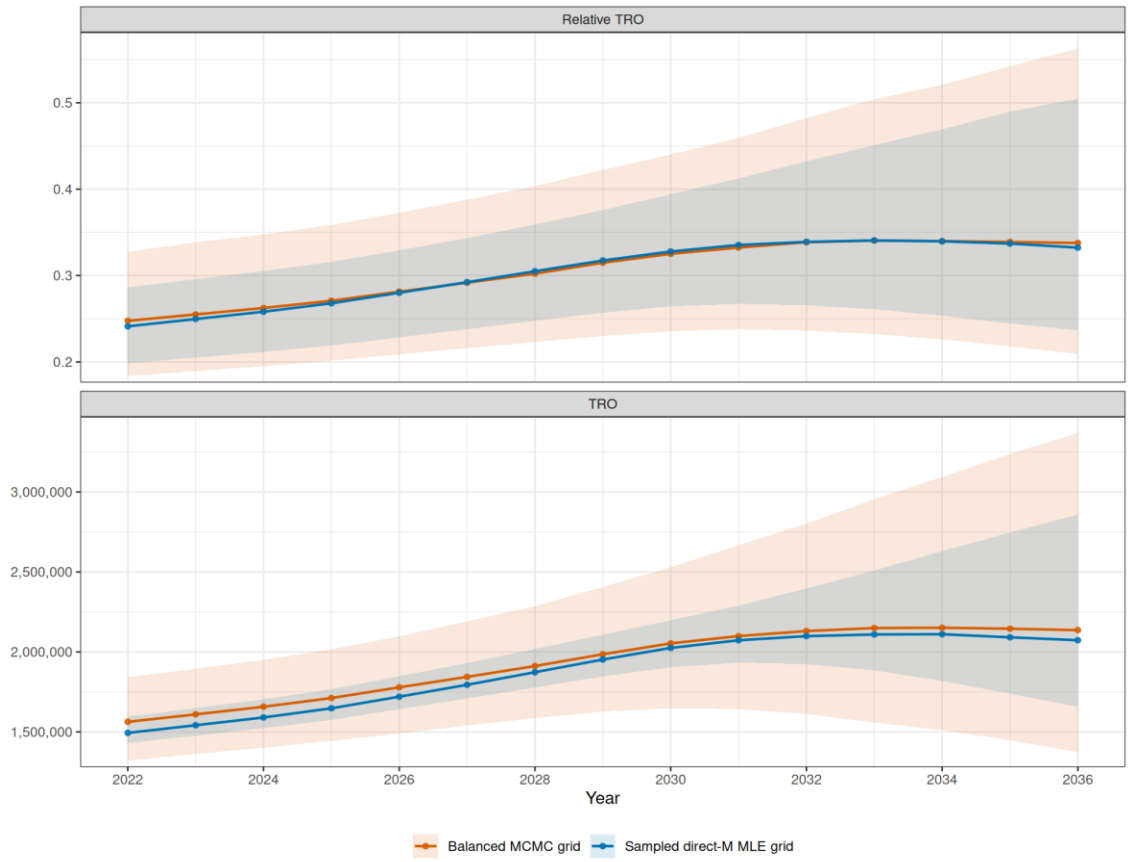


Figure 6. Total reproductive output (TRO) both relative (top) and absolute (bottom) derived using the MCMC reference set and the sampled MLE grid. Lines show medians and shaded areas show 95% intervals.

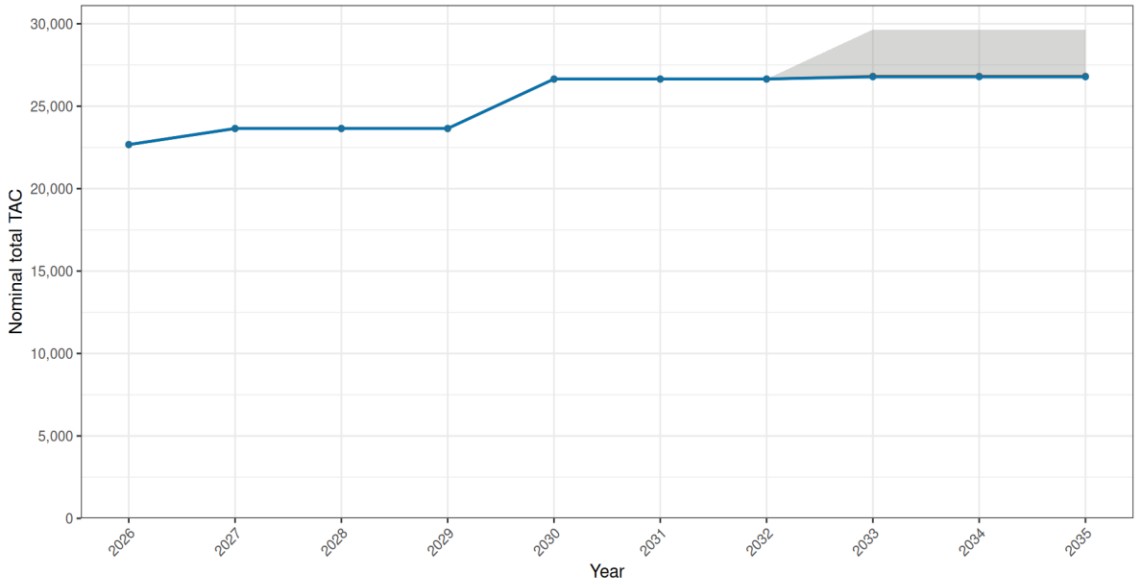


Figure 7. Nominal total TAC for the reference set projections under the nominal fleet allocation. The line shows the median and the shaded region shows the 95% interval.

8. Summary

The 2026 stock assessment for southern bluefin tuna provides estimates of current stock status (in 2026) using the new assessment software in R-TMB that allows for MCMC sampling and estimation of parameter uncertainty within each grid cell. As expected, uncertainty ranges are wider using the MCMC grid for the reference set results.

The 2026 assessment indicates that there has been further rebuilding of the stock since the last assessment in 2023. Current relative TRO is 0.281.

The fits to the data did not indicate any areas of concern. Sensitivity tests were defined to assess impacts on stock status estimates from additional areas of uncertainty and did not identify any areas of concern.

The current assessment indicates that the stock is close to the current rebuilding target, and TRO at MSY. The projections for the reference set of models indicate that the management procedure rebuilding target (30% TRO, by 2035), will be reached earlier than planned, i.e. by 2028 (median of the projections).

References

Anon, 2024. Report of the Fourteenth Operating Model and Management Procedure Technical Group, Commission for the Conservation of Southern Bluefin Tuna, 24-28 June 2026, Seattle, USA.

Anon, 2025a. Report of the Thirtieth Meeting of the Scientific Committee, Commission for the Conservation of Southern Bluefin Tuna, August 2025, Online.

Anon, 2025b. Report of the Fifteenth Operating Model and Management Procedure Technical Group, Commission for the Conservation of Southern Bluefin Tuna, 3-4 July 2025, Online.

Anon, 2026. Report of the Sixteenth Operating Model and Management Procedure Technical Group. Commission for the Conservation of Southern Bluefin Tuna, 22-26 June, Seattle, USA.

CCSBT, 2026. Secretariat Review of Catches. CCSBT-ESC/2608/04.

Dunn, P.K., Smyth, G.K., 1996. Randomized quantile residuals. Journal of Computational and Graphical Statistics 5, 236–244. https://doi.org/10.1080/10618600.1996.10474708

Edwards and Hoyle, 2026. Estimates of unreported SBT catch by CCSBT non-cooperating non-member states between 2007 and 2023. CCSBT-OMMP/2606/04, CCSBT-ESC/2608/BGD-01.

Eveson, P., Farley, J., 2017. The aerial survey index of abundance: 2017 updated results (CCSBT-ESC/1708/06). Commission for the Conservation of Southern Bluefin Tuna.

Farley J, Eveson P, Gunasekera R. 2026. Update on the SBT close-kin tissue sampling, processing and kin-finding 2026. CCSBT-ESC/2608/11.

Francis, R.I.C.C., 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences 68, 1124–1138. https://doi.org/10.1139/F2011-025

Hillary, R.M., Preece, A.L., Takahashi, N., Davies, C.R., Itoh, T., 2023. The southern bluefin tuna stock assessment in 2023 (CCSBT-ESC/2308/16). Commission for the Conservation of Southern Bluefin Tuna.

Itoh, T., 2026. Trolling indices for age-1 southern bluefin tuna: update of the grid type trolling index in 2026 (CCSBT-ESC/2608/25). Commission for the Conservation of Southern Bluefin Tuna.

Itoh T. and N. Takahashi. 2026. Update of CPUE abundance index using GAM for southern bluefin tuna in CCSBT (GAM22) up to the 2025 data. CCSBT-ESC/2608/BGD 03, CCSBT-OMMP/2606/06.

Itoh, T., D. Webber, A. Parma (2026) Calculation of the coefficient of variation of GAM22 CPUE index for the SBT assessment in 2026. CCSBT-ESC/2608/38.

Kristensen, K., 2024. RTMB: R interface to TMB for fully symbolic AD models.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., Bell, B.M., 2016. TMB: Automatic differentiation and laplace approximation. Journal of Statistical Software 70, 1–21. https://doi.org/10.18637/jss.v070.i05

Lorenzen, K. 1996. The Relationship Between Body Weight and Natural Mortality in Juvenile and Adult Fish: A Comparison of Natural Ecosystems and Aquaculture. Journal of Fish Biology 49(4), 627–642.

Preece, A., Eveson, P., Hillary R., 2026. Review of data inputs and preliminary work for the 2026 SBT stock assessment (CCSBT-OMMP/2606/09). Commission for the Conservation of Southern Bluefin Tuna.

Preece AL, Eveson JP, Hartog JR, Dell J, Aulich J, Potter N, Cooper S, Grewe P, Lansdell M. 2026. Update on the SBT gene-tagging recruitment monitoring program 2026. CSIRO, Australia. CCSBT-ESC/2608/10.

Stewart, I.J., Monnahan, C.C., 2025. Diagnosing common sources of lack of fit to composition data in fisheries stock assessment models using one-step-ahead (OSA) residuals. Canadian Journal of Fisheries and Aquatic Sciences 82, 1–13. https://doi.org/10.1139/cjfas-2025-0158

Vehtari, A., Gelman, A., Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing 27, 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Vehtari, A., Simpson, D., Gelman, A., Yao, Y., Gabry, J., 2024. Pareto smoothed importance sampling. Journal of Machine Learning Research 25, 1–58.

Appendix A: Data fit summary plots for the base model

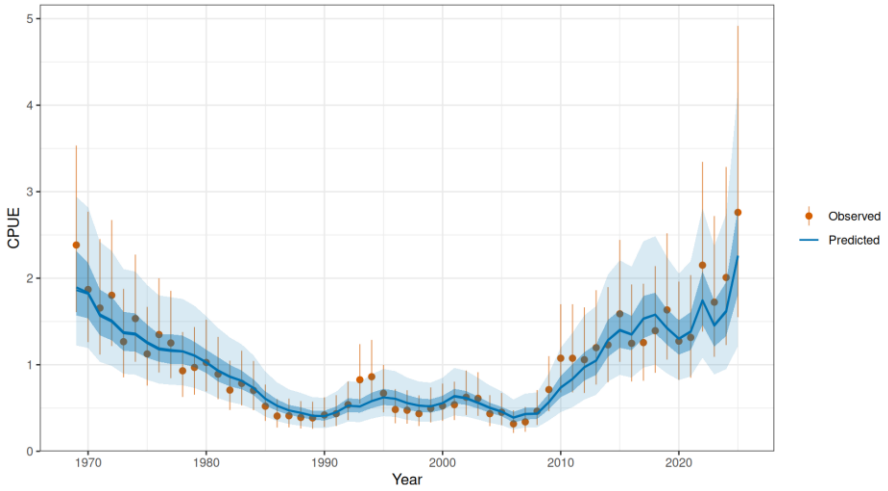


Figure A1. Fit to the Japanese longline CPUE relative abundance index. Observations and their approximate 95% intervals are orange. The posterior predictive median is the dark blue line; darker and lighter shading show the interquartile and 95% credible intervals.

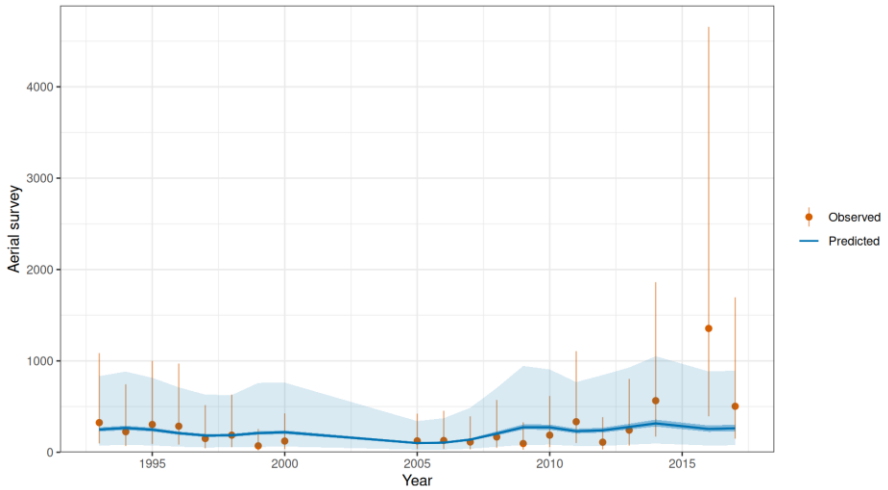


Figure A2. Fit to the aerial-survey relative biomass index. Observations and their approximate 95% intervals are orange. The posterior predictive median is the dark blue line; darker and lighter shading show the interquartile and 95% credible intervals.

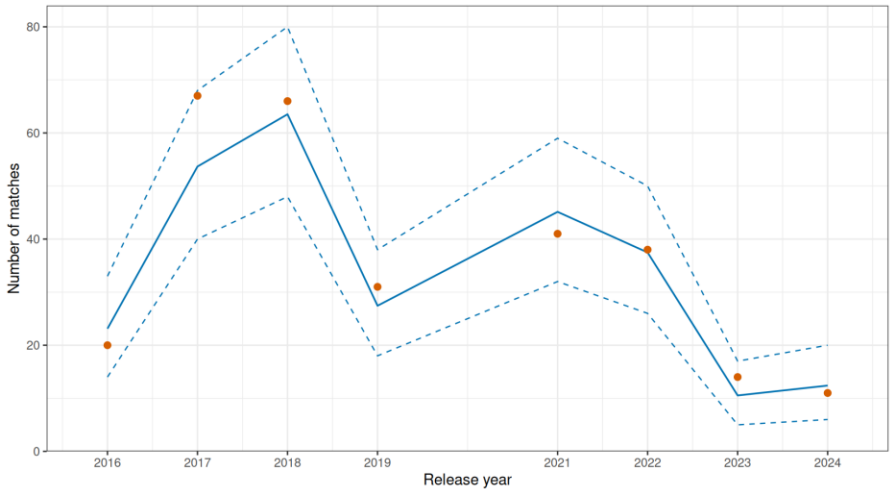


Figure A3. Fits to the gene tagging data. Orange points are the observations, the solid blue line is the posterior median and the dotted lines encompass the approximate 95% posterior predictive credible interval.



Figure A4. Observed and fitted conventional-tag recaptures by release year, release age, and recapture age, pooled across tagger groups. Orange bars are observed counts; blue points and lines are fitted MLE expected counts, and vertical blue bars are equal-tailed 95% posterior credible intervals for the pooled expected counts.

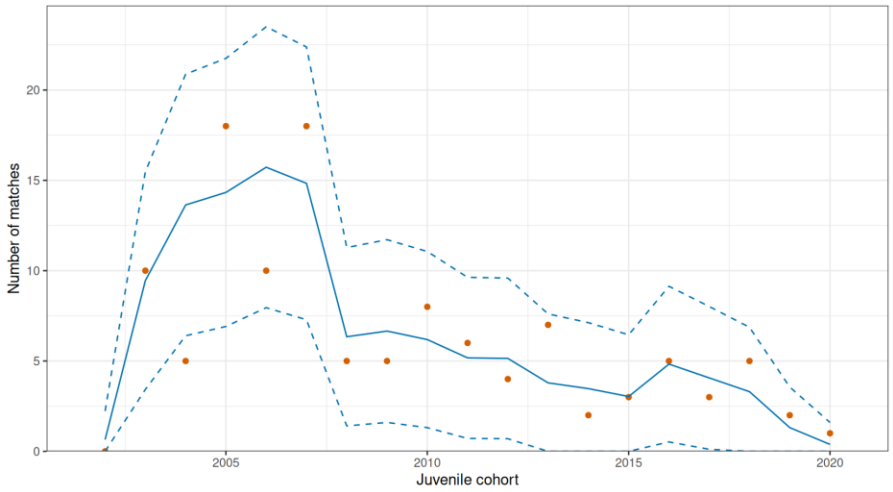


Figure A5. Fits to the CKMR POP data aggregated to the juvenile cohort level. Orange points are the observations, the solid blue line is the posterior median and the dotted lines encompass the approximate 95% posterior predictive credible interval.

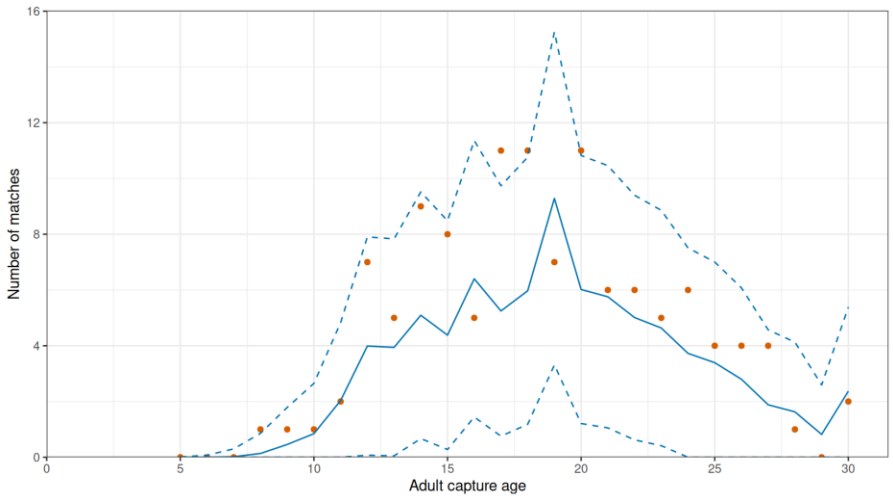


Figure A6. Fits to the CKMR POP data aggregated to the adult capture age level. Orange points are the observations, the solid blue line is the posterior median and the dotted lines encompass the approximate 95% posterior predictive credible interval.

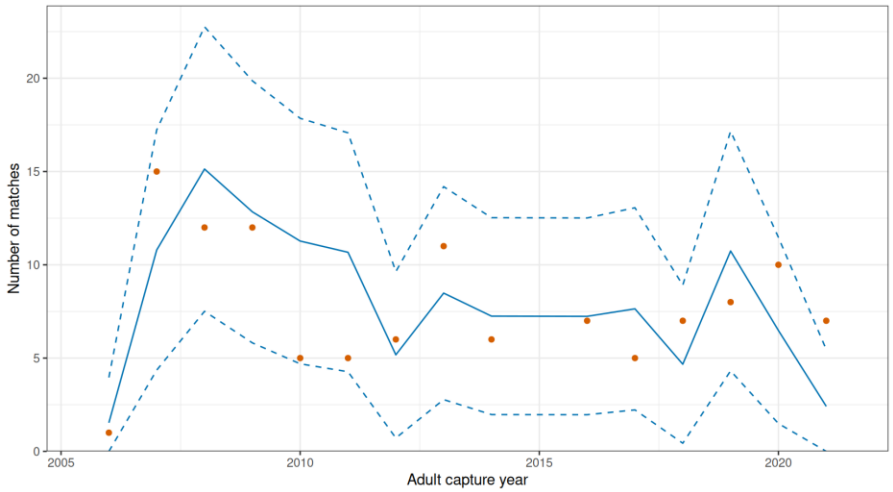


Figure A7. Fits to the CKMR POP data aggregated to the adult capture year level. Orange points are the observations, the solid blue line is the posterior median and the dotted lines encompass the approximate 95% posterior predictive credible interval.

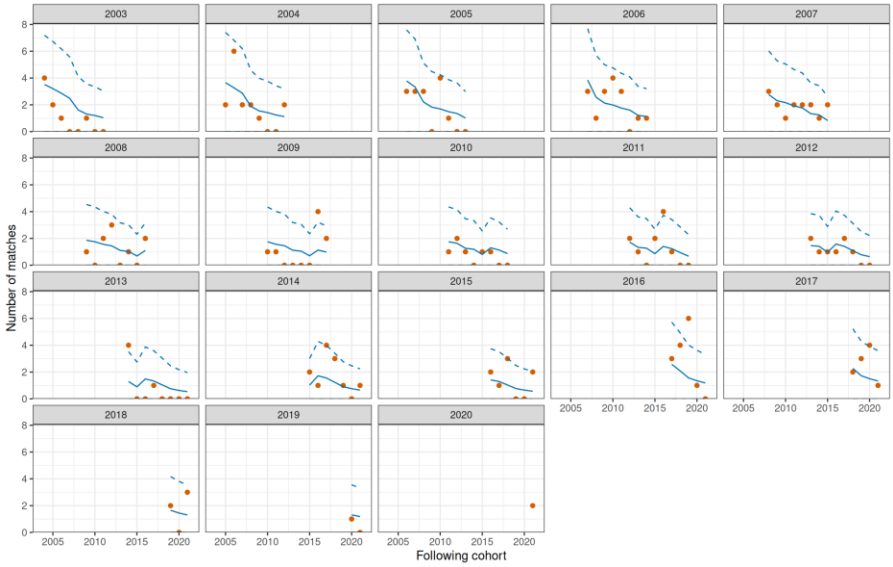


Figure A8. Fits to the CKMR HSP data at the base level. Orange points are the observations, the solid blue line is the posterior median and the dotted lines encompass the approximate 95% posterior predictive credible interval.

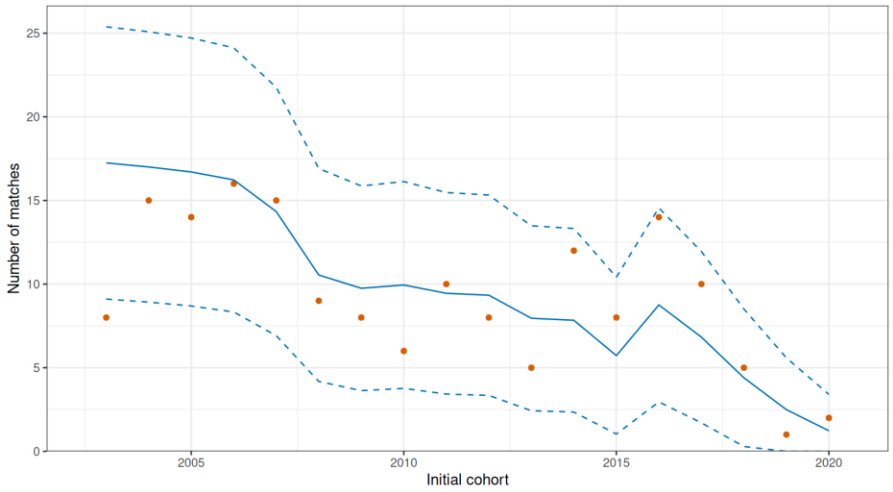


Figure A9. Fits to the CKMR HSP data aggregated to the initial cohort level. Orange points are the observations, the solid blue line is the posterior median and the dotted lines encompass the approximate 95% posterior predictive credible interval.

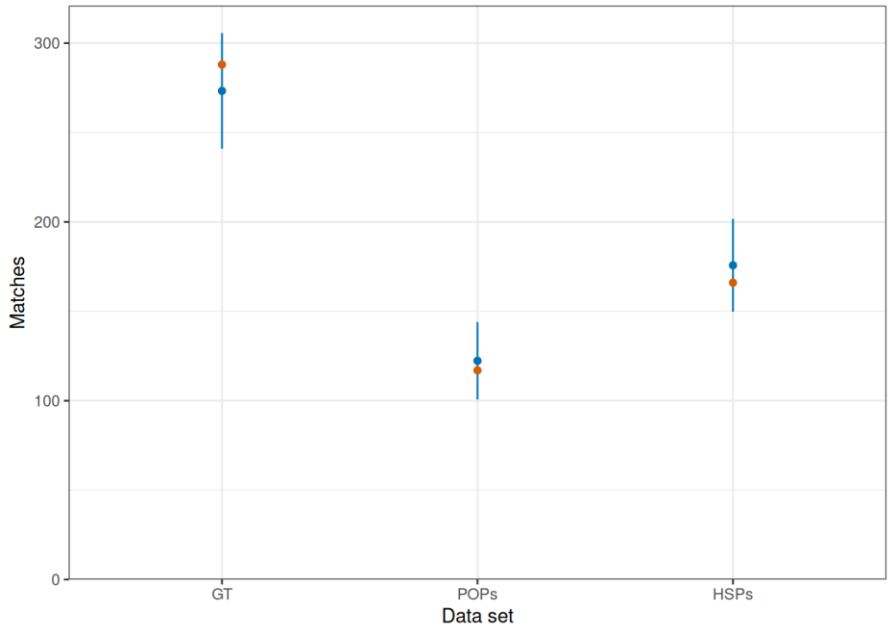


Figure A10. Fits to the overall number of matches for the gene tagging (left), POPs (middle), and HSPs (right). Orange points are the observations, the blue point is the posterior median and the whiskers encompass the approximate 95% posterior predictive credible interval.

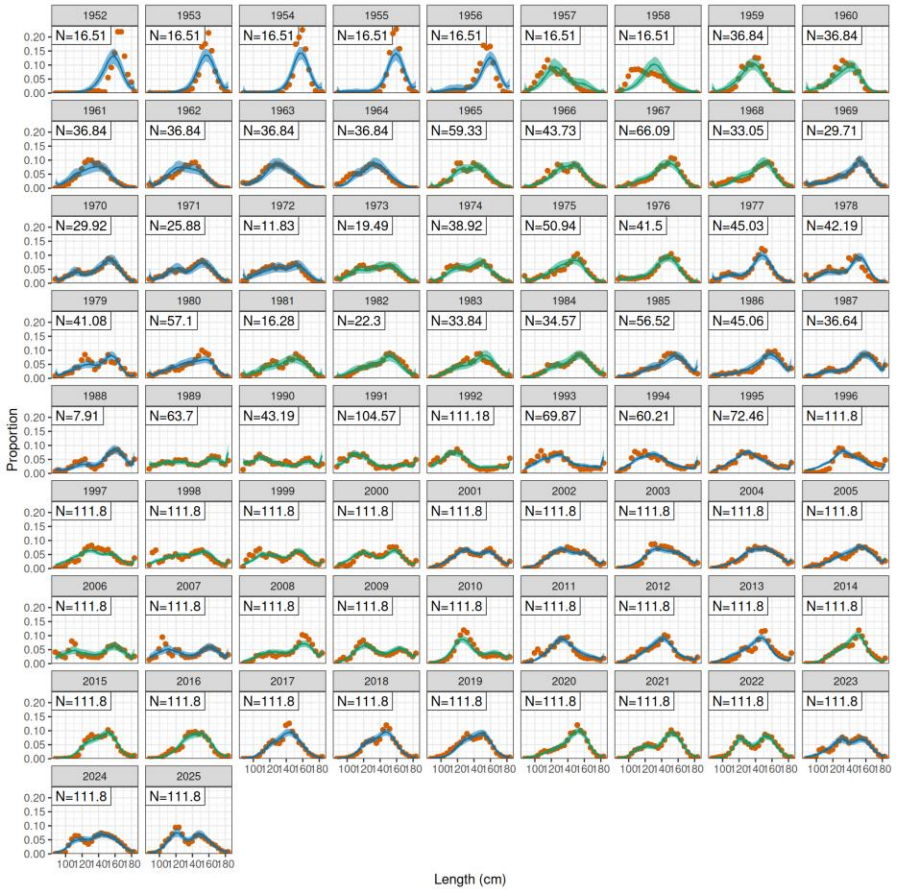


Figure A11. Fit to the LL1 length-frequency data. Orange points are observations; the line and shading show the posterior median and approximate 95% posterior predictive interval. Colour changes mark years in which selectivity may change.

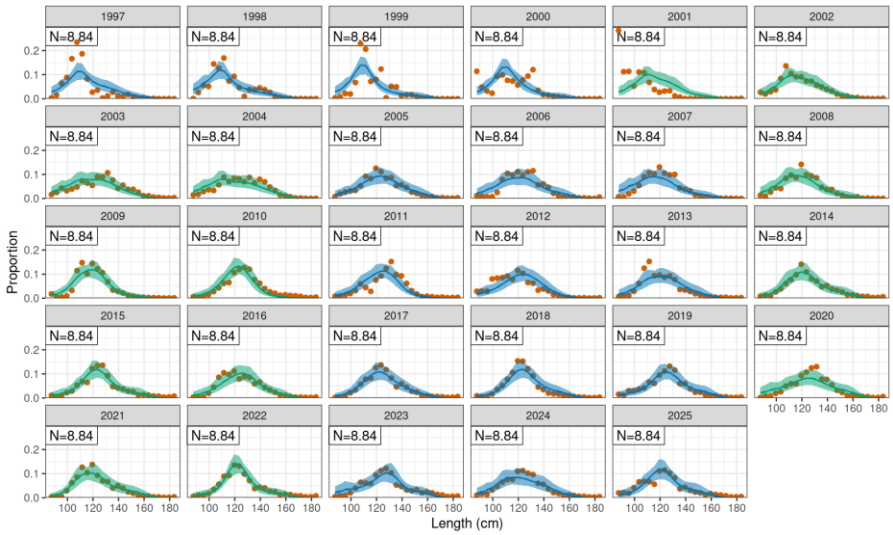


Figure A12. Fit to the LL2 length-frequency data. Orange points are observations; the line and shading show the posterior median and approximate 95% posterior predictive interval. Colour changes mark years in which selectivity may change.

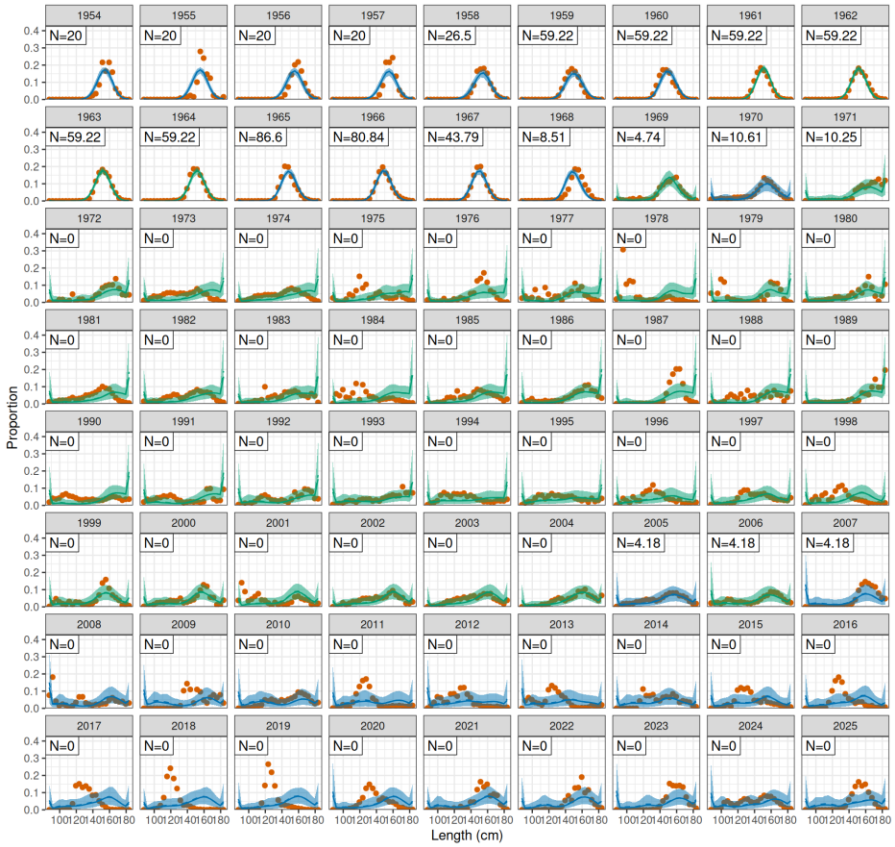


Figure A13. Fit to the LL3 length-frequency data. Orange points are observations; the line and shading show the posterior median and approximate 95% posterior predictive interval. Colour changes mark years in which selectivity may change. Note: 1972-2004 and 2008-2025 are not fitted.

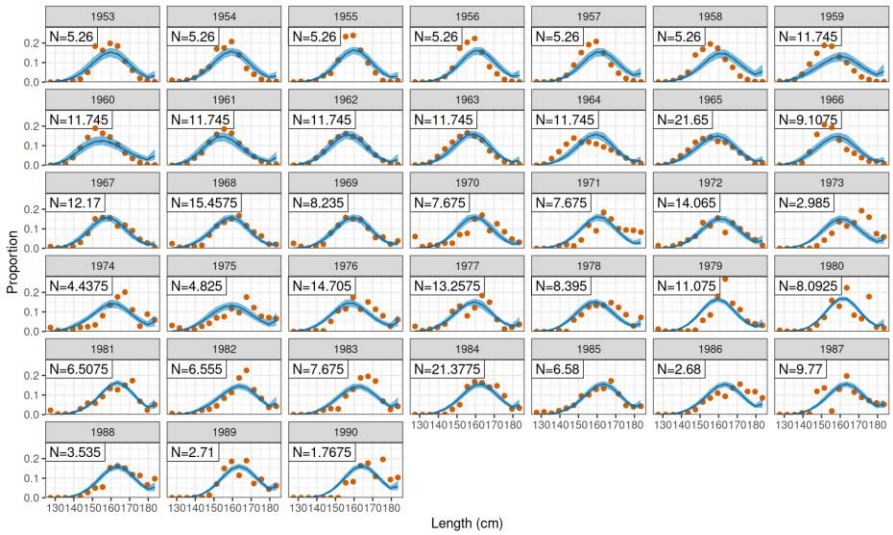


Figure A14. Fit to the LL4 length-frequency data. Orange points are observations; the line and shading show the posterior median and approximate 95% posterior predictive interval.

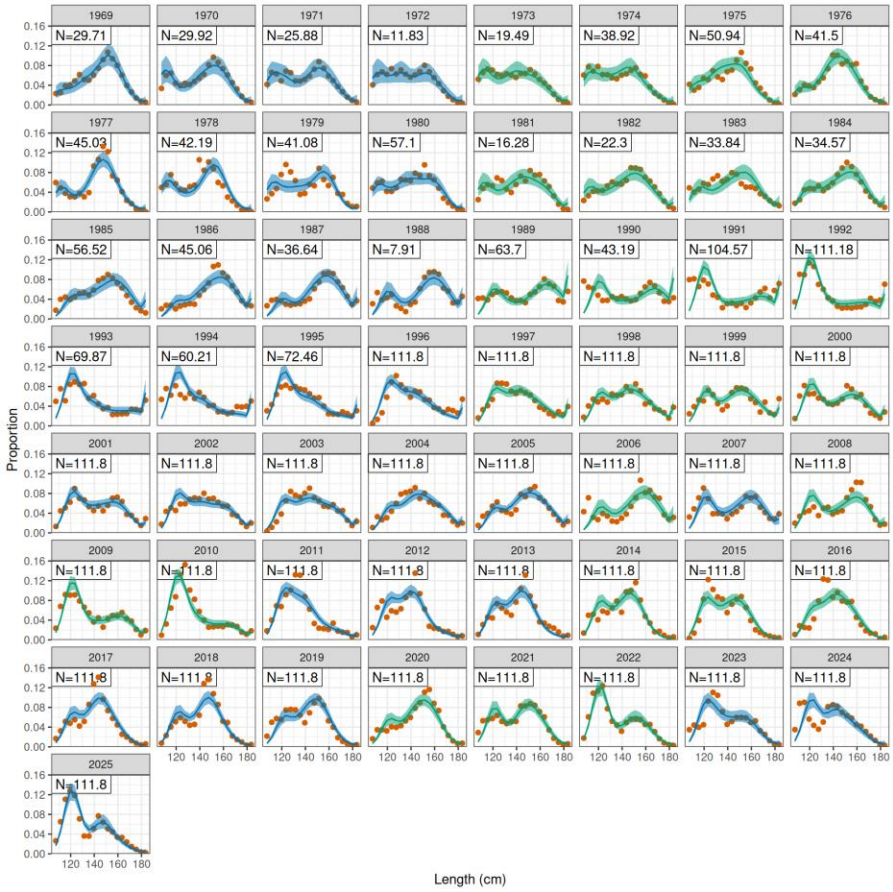


Figure A15. Fit to the CPUE-weighted Japanese longline length-frequency data. Orange points are observations; the line and shading show the posterior median and approximate 95% posterior predictive interval. Colour changes mark years in which selectivity may change.

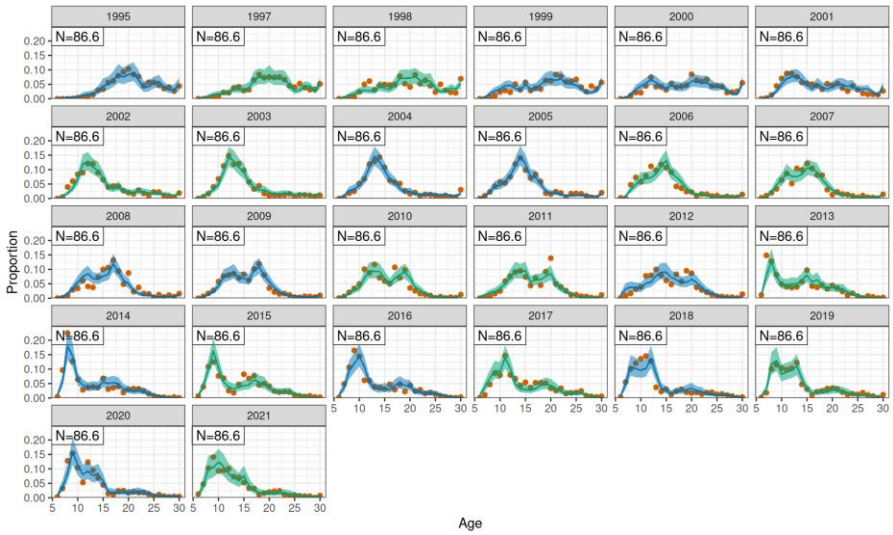


Figure A16. Fit to the Indonesian age-frequency data. Orange points are observations; the line and shading show the posterior median and approximate 95% posterior predictive interval. Colour changes mark years in which selectivity may change.

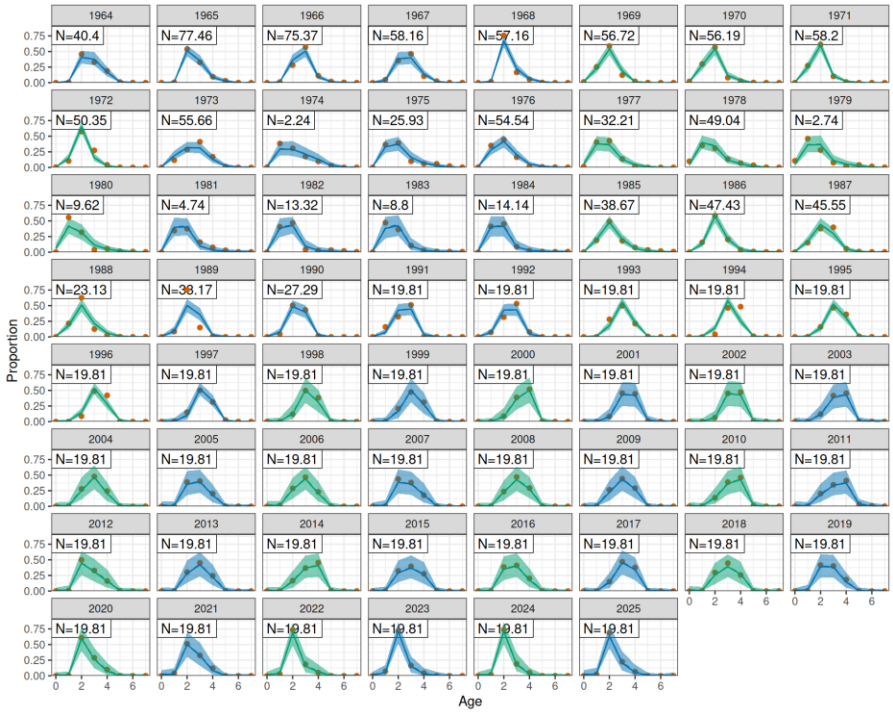


Figure A17. Fit to the Australian surface fishery age-frequency data. Orange points are observations; the line and shading show the posterior median and approximate 95% posterior predictive interval. Colour changes mark years in which selectivity may change.

Appendix B: Changes to natural mortality at age

The new length-based natural-mortality model is specified as $M_a = M_{10}(L_a/L_{10})^{m_c}$ for ages 0 through 25, followed by a linear increase to age 30 (M_{30}). Length at age (L_a) is taken to be the mean length-at-age from the first year and season of the model. Parameters M_{10} and M_{30} are estimated, and the exponent m_c is fixed at -1, which makes mortality inversely proportional to length for ages up to 25 (as described by Lorenzen 1996).

This new model provides a smoother, more parsimonious and biologically interpretable curve. It retains the broad shape of the previous assessment (Figure B1) while avoiding estimating four mortality parameters (M_0, M_4, M_{10}, M_{30}), which caused issues within the new MCMC model. In particular, M_{10} was estimated to be higher than M_4, which was not considered biologically probable (Figure B1).

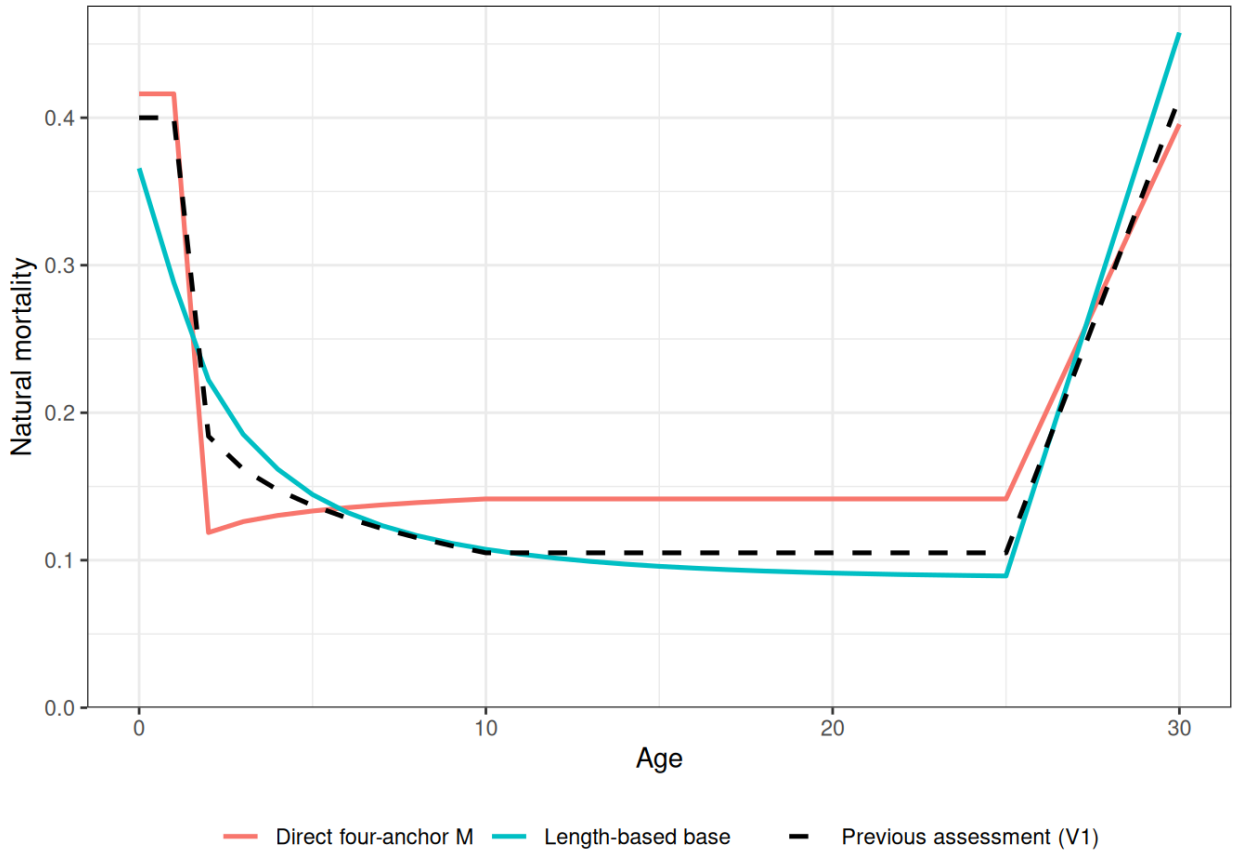


Figure B1: Natural mortality at age estimated for the base model (which uses the new length-based formulation) and for the four-parameter sensitivity model (Old_m). Also included for comparison is the natural mortality curve obtained using the previous-assessment model (i.e., the ADMB model used for the 2023 assessment, recoded in RTMB).

Appendix C: Estimated selectivity-at-age by fishery for the base model

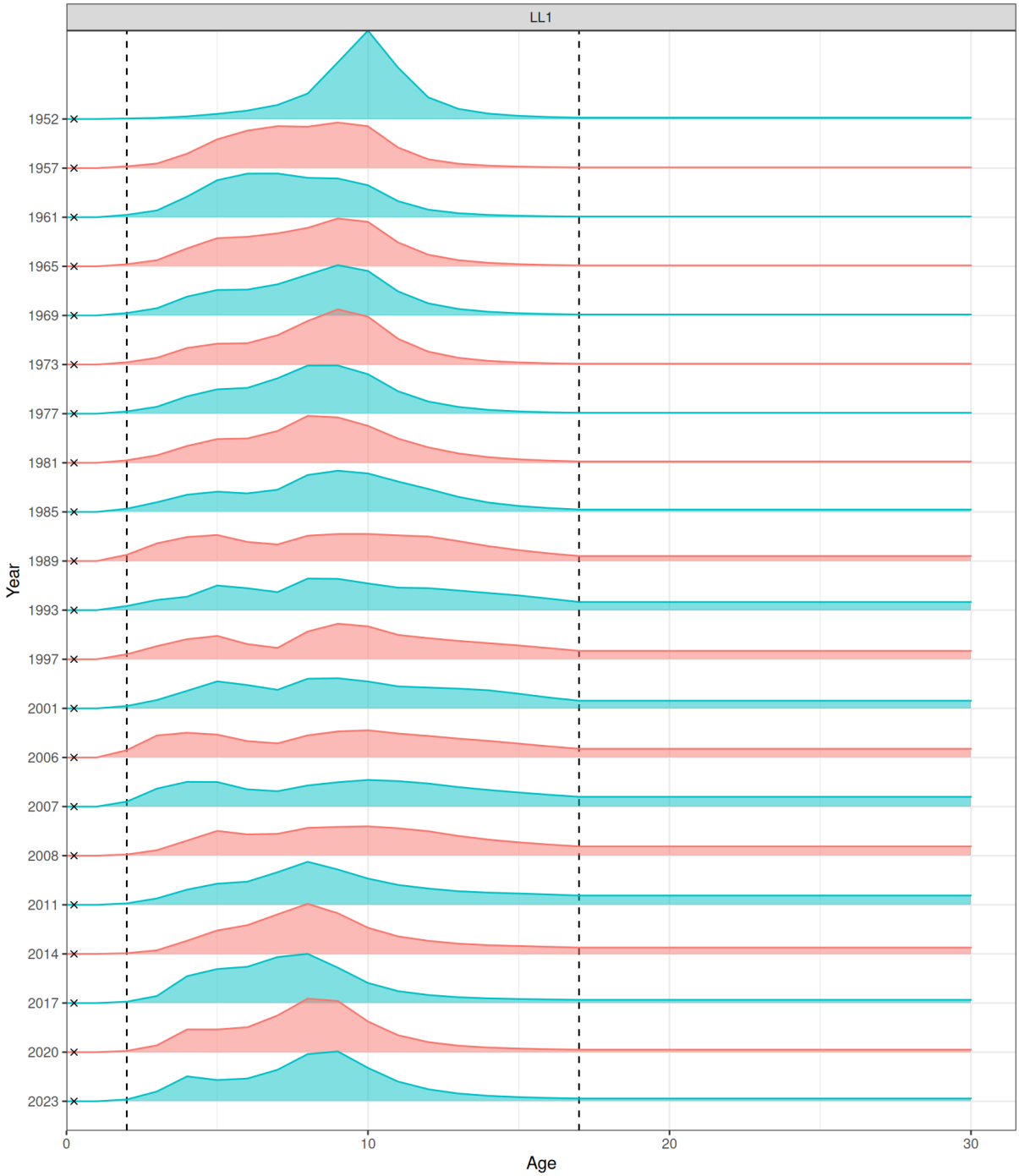


Figure C1: Estimated selectivity-at-age for LL1 from the base model. Dashed vertical lines mark the minimum and maximum ages over which selectivity is estimated; crosses mark years with composition observations.

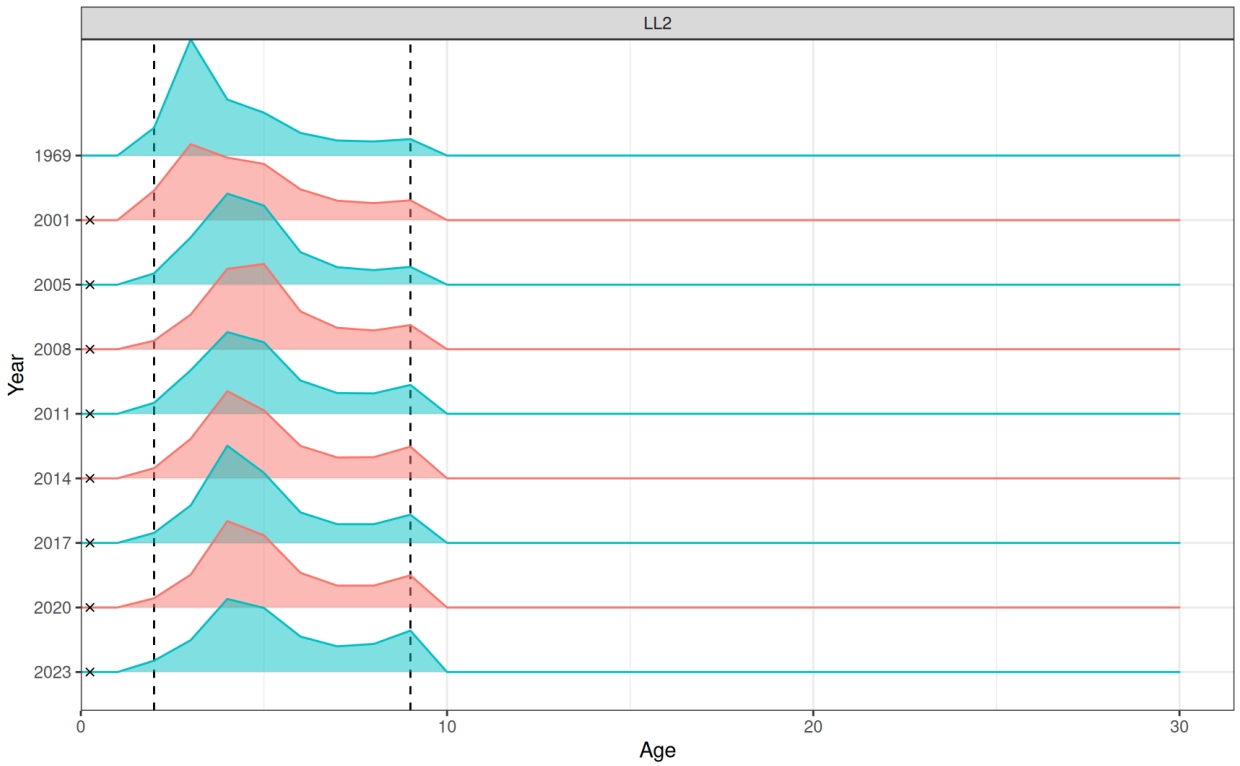


Figure C2: Estimated selectivity-at-age for LL2 from the base model. Dashed vertical lines mark the minimum and maximum ages over which selectivity is estimated; crosses mark years with composition observations.

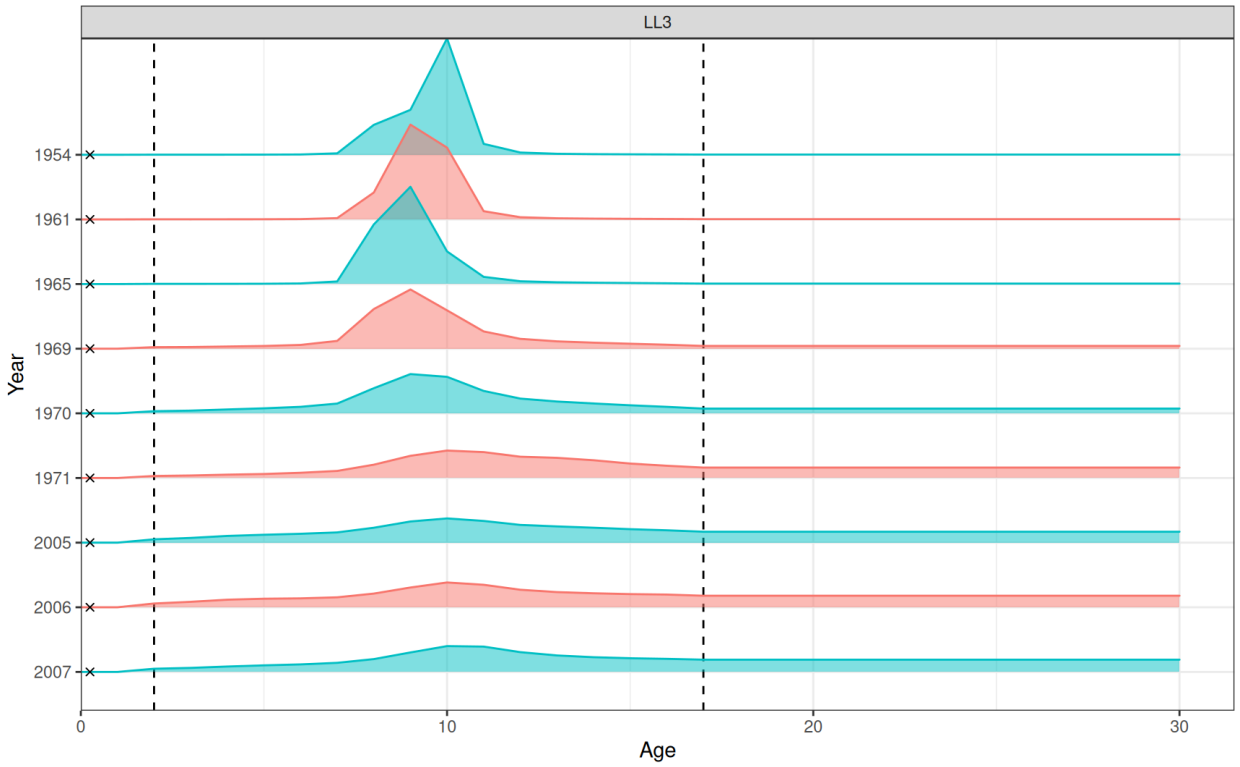


Figure C3: Estimated selectivity-at-age for LL3 from the base model. Dashed vertical lines mark the minimum and maximum ages over which selectivity is estimated; crosses mark years with composition observations.

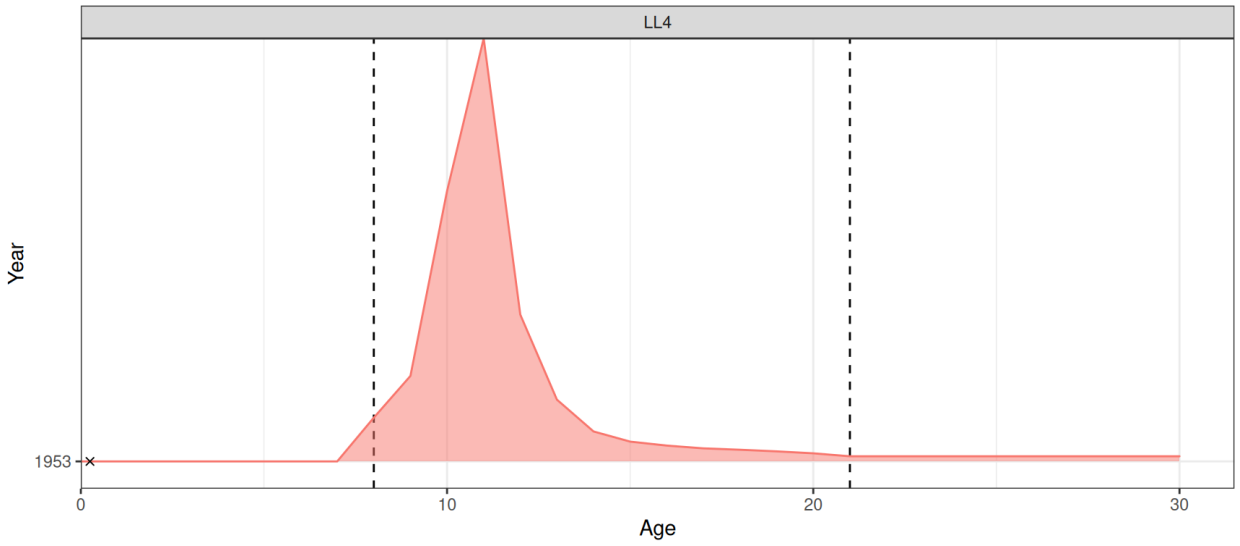


Figure C4: Reviewed MLE selectivity-at-age for the single LL4 block beginning in 1953 from the base model. Its 14 age effects use LL3's fixed rho-year 0.5, rho-age 0.5, and sigma 0.75 hyperparameters.

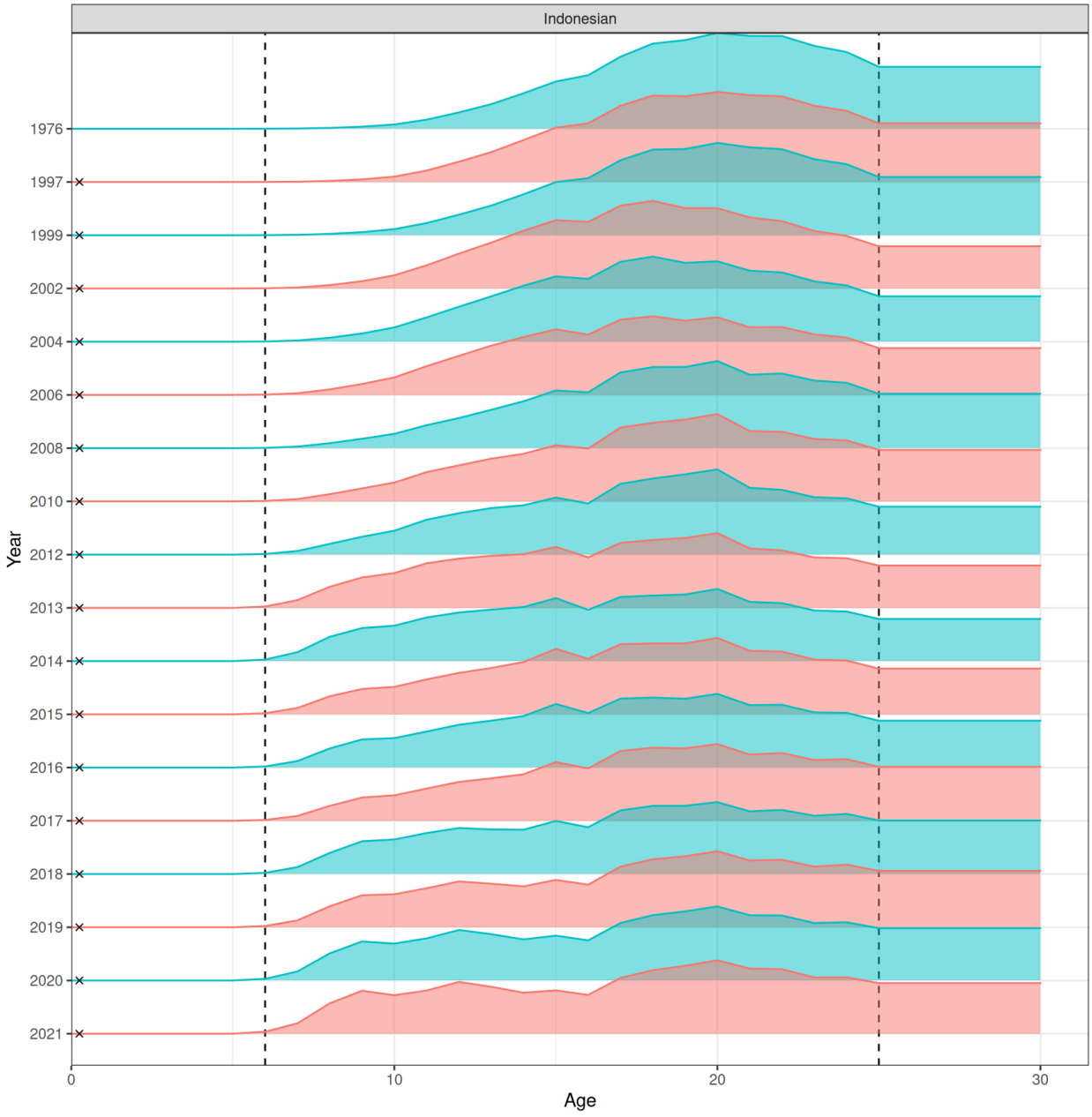


Figure C5: Estimated selectivity-at-age for the Indonesian fishery from the base model. Dashed vertical lines mark the minimum and maximum ages over which selectivity is estimated; crosses mark years with composition observations.

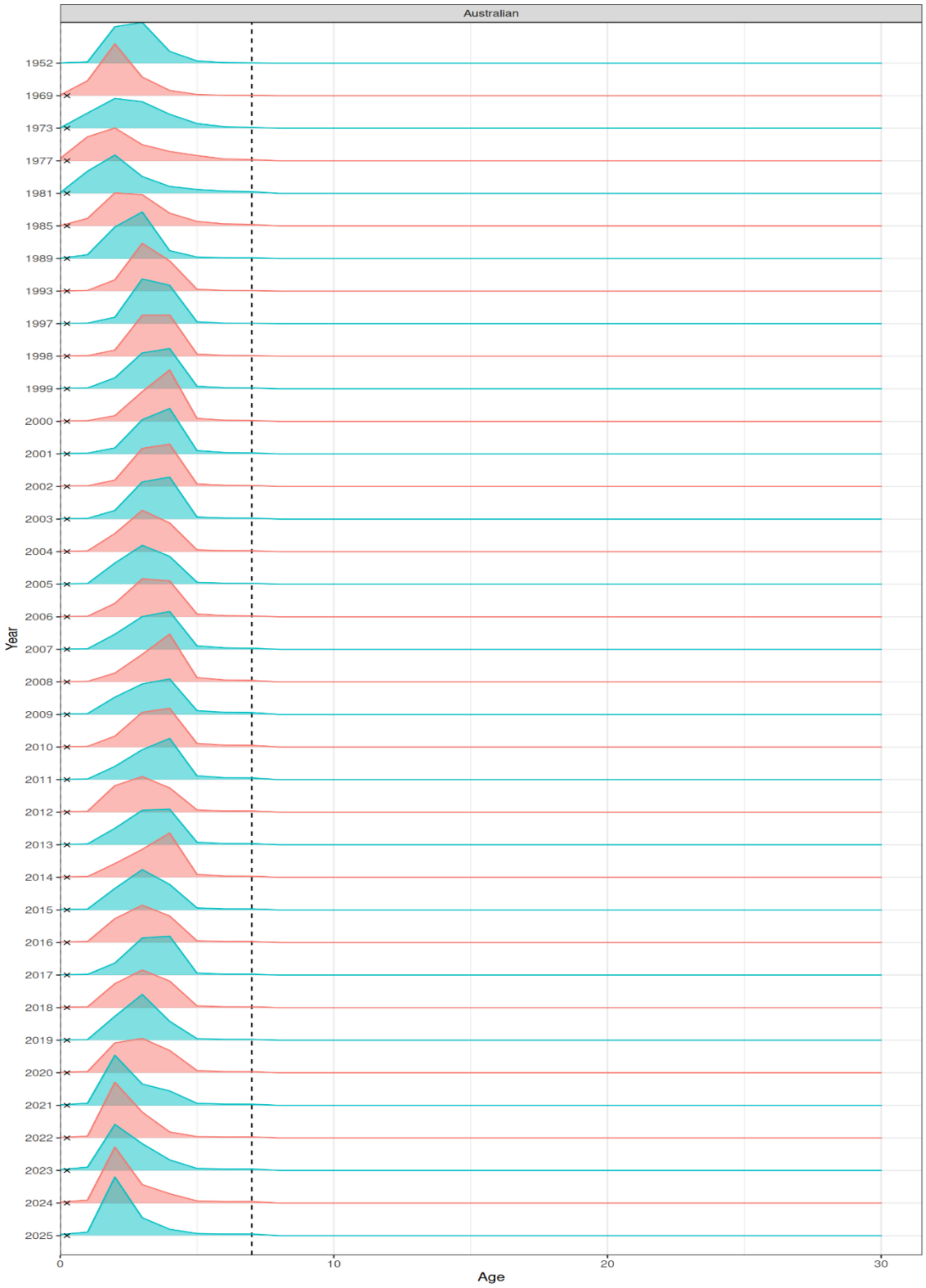


Figure C6: Estimated selectivity-at-age for the Australian surface fishery from the base model. Dashed vertical lines mark the minimum and maximum ages over which selectivity is estimated; crosses mark years with composition observations.

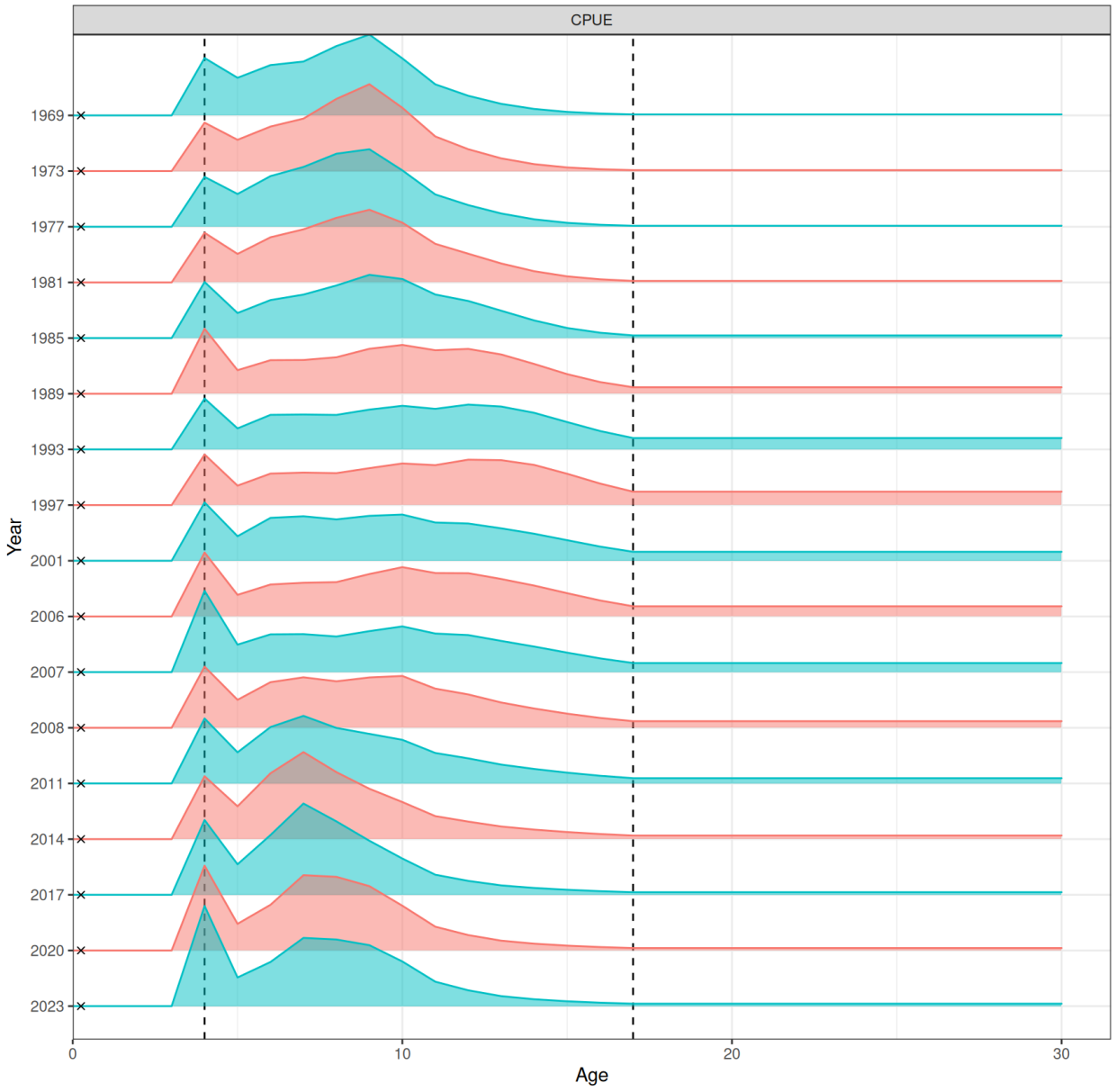


Figure C7: Estimated selectivity-at-age for the CPUE fleet from the base model. Dashed vertical lines mark the minimum and maximum ages over which selectivity is estimated; crosses mark years with composition observations. The high value at the first displayed age is an edge effect from the lower bound of the CPUE selectivity window rather than evidence for an additional younger-age mode.