WCPFC Scientific Committee 22 · SA-WP-06 (Rev.01)

2026 WCPO bigeye assessment

What changed, what the ensemble says—and what remains uncertain

Kyuhan Kim[1] · Thom Teears[1] · John Hampton[2]
Nick Davies[3] · Arni Magnusson[1] · Claudio Castillo Jordán[1]
Tom Peatman[4] · Joe Scutt Phillips[1] · Paul Hamer[1]

[1] Pacific Community (SPC) · [2] Thunnidae Consulting
[3] TeTakina Ltd · [4] Shearwater Analytics

Apia, Samoa · 15 August 2026

The 2026 result in one sentence

STOCK LEVEL

27.7%

Median recent spawning level vs no fishing
80% interval 17.2–39.5%

FISHING PRESSURE

0.84 × FMSY

Median recent F/FMSY
80% interval 0.59–1.21

FAVOURABLE CATEGORY

75.8%

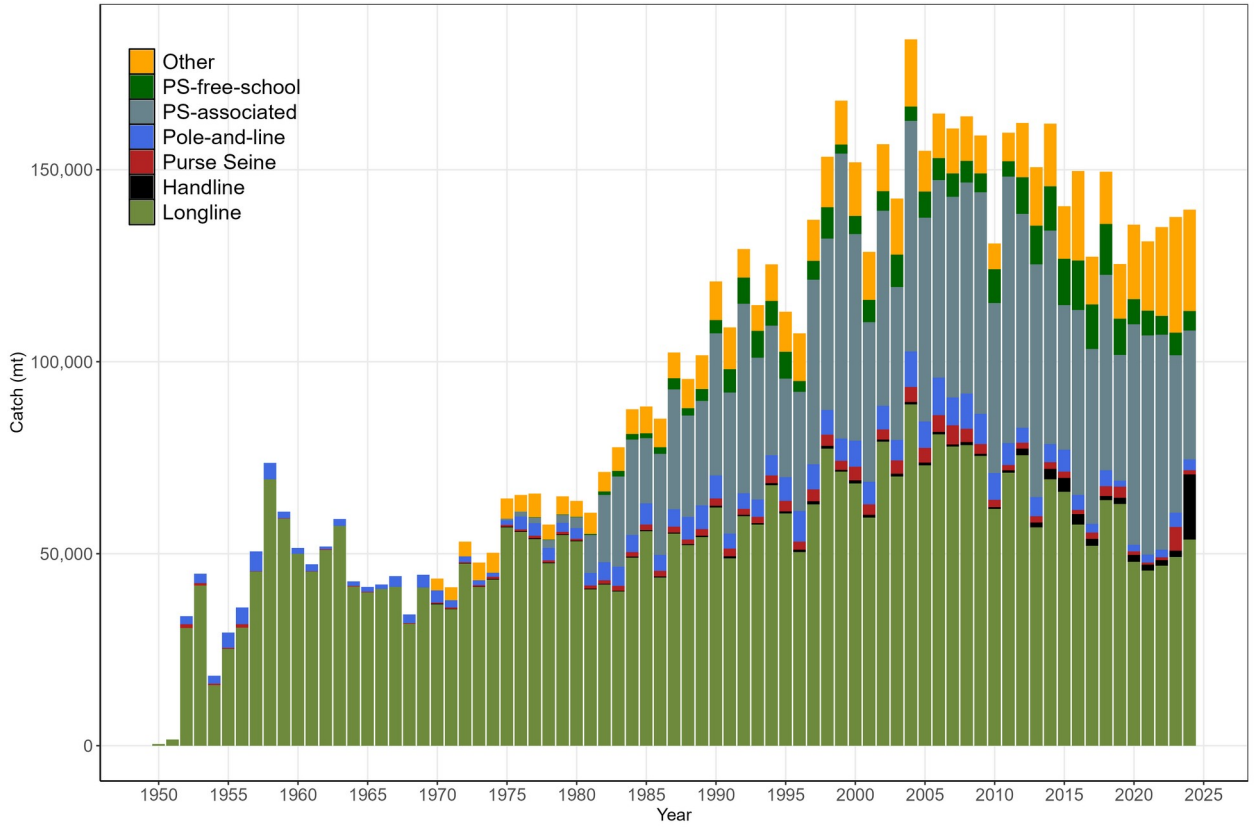
Probability above the limit reference point (LRP) and below FMSY

MOST RESULTS FAVOURABLE
IMPORTANT RISK REMAINS

In the combined results, 18.6% fall below the LRP and 24.1% exceed FMSY. These probabilities overlap and must not be added.

Source: WCPFC-SC22-2026-SA-WP-06, Tables 8 and 10 and Figure 63. Recent depletion averages 2021–2024; recent fishing mortality averages 2020–2023.

A changing fishery, a longer evidence record



1952–2024 · QUARTERLY MODEL

The evidence now includes three more years

- Catch expanded with tropical fisheries, then declined after about 2010.
- Most catch is taken in Regions 2–4.
- In 2015–2024, longline took about half the catch by weight; associated purse seine took about half by number, mainly juveniles.

The apparent 2024 handline increase mainly reflects an Indonesian catch-allocation change—not a demonstrated fleet expansion.

DATA CONTEXT Purse-seine sampling was weaker during 2020–2022, while recent longline data remain subject to revision. These limitations are carried into interpretation—not hidden.

Source: WCPFC-SC22-2026-SA-WP-06, Sections 3.3, 4.3 and 4.5.1; Figure 3. The 2024 Indonesian handline allocation is to be reconstructed historically before the next assessment.

2023 → 2026: uncertainty now crosses both thresholds

	2023 COMBINED		2026 ENSEMBLE	
RECENT DEPLETION	0.35	→	0.277	−21%
	0.30–0.40		0.172–0.395	
F / FMSY	0.59	→	0.838	+42%
	0.46–0.74		0.592–1.206	

Uncertainty now crosses both boundaries · 2023: 0% below LRP and 0% above FMSY; 2026 crosses both.

Comparison does not identify a cause: data years, averaging periods, regions, fisheries, biology, data weighting and how uncertainty was combined all changed.

Sources: 2023 and 2026 assessment reports. Periods: 2023 depletion 2018–21 / F 2017–20; 2026 depletion 2021–24 / F 2020–23. Intervals show the middle 80% of combined results.

2023 → 2026: what changed—and what improved

AREA	2023 ASSESSMENT		2026 ASSESSMENT	WHAT IMPROVED	DETAIL
STRUCTURE	9 regions · 41 fisheries 3,067 parameters	→	5 regions · 33 fisheries 1,997 (−34.9%); fixed natural mortality (M)	Fewer weak contrasts; better numerical stability.	08
EVIDENCE	Earlier CPUE setup Less age-at-length data	→	Regional CPUE + global scaling Weight converted before fitting	Clearer roles; more direct size-and-growth signal.	09
TAGS	τ: tag-count overdispersion Effective τ≈1.02; coarser mixing/reporting	→	Overdispersion + tag handling revised Diagnostic τ=2; ensemble τ=1.2–1.4	Allows extra tag-count variation; tests practical tag influence.	16–19
UNCERTAINTY	Fixed assumption grid Structural choices combined	→	100 planned → 80 retained Equal model weight	Assumption combinations included in status probabilities.	20–23
FIT CHECKS	Instability evident Narrower formal test suite	→	Starting-value tests + self-test Fit profiles; simpler/no-tag checks	Shows which outputs are stable—and which are not.	24–25

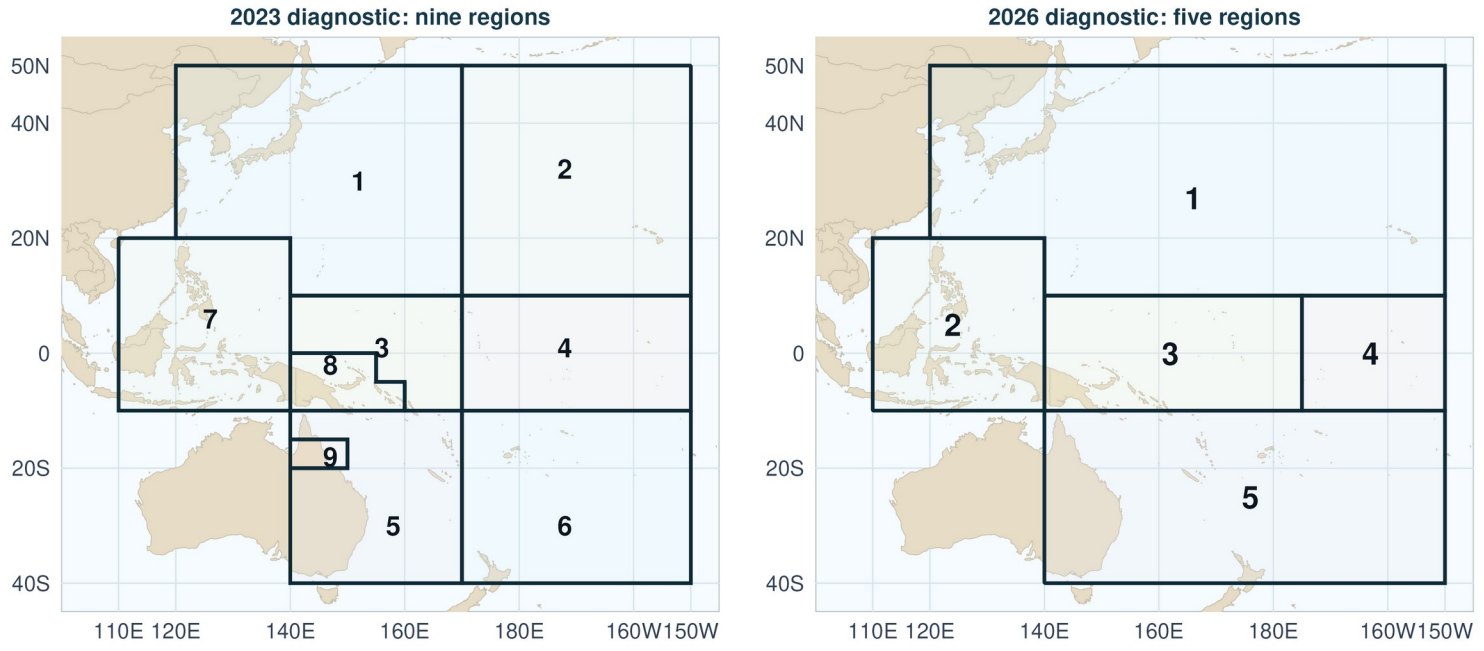
IMPROVED—NOT FINISHED

The rebuild makes the model simpler, improves how evidence is handled, and makes fit checks more transparent. It does not resolve the disagreement over population size or make every parameter reliably estimable.

Source: WCPFC-SC22-2026-SA-WP-06, Executive Summary, Introduction and Sections 4, 6 and 8.1–8.3. "Improved" describes the assessment process and diagnostics; it is not a claim that every change removes bias.

FIGURE 2 · 9 REGIONS → 5

Simpler spatial structure, clearer purpose



WHY FIVE REGIONS?

Keep supported contrasts; remove weak ones

- Removed small regions with little bigeye tag or movement information.
- Combined northern and southern areas where data did not support multiple regions.
- Retained a central-east equatorial region to represent concentrated tag information.

The five-region model has 28 catch fisheries + five regional CPUE indices; fleets remain separate within a region where needed. Simplification reduced instability, but does not ensure every parameter is reliably estimated.

Source: WCPFC-SC22-2026-SA-WP-06, Sections 4.2 and 8.1; Figure 2.

How the evidence base was improved

CPUE

Regional indices + global scaling

Separate regional catch-rate analyses allow regional differences in which fish are caught and how catch rates relate to abundance. A global index links regional population levels, using 1965–1969 in the Diagnostic model.

SIZE COMPOSITION

Weight converted to length

Longline weights were converted before fitting and combined with observed lengths. Stricter filters and a weighting model allowing extra sample variation were applied.

AGE-AT-LENGTH

More direct growth information

More age-at-length observations, including North Pacific samples, were fitted directly and reweighted across regions.

TAG-RECAPTURE

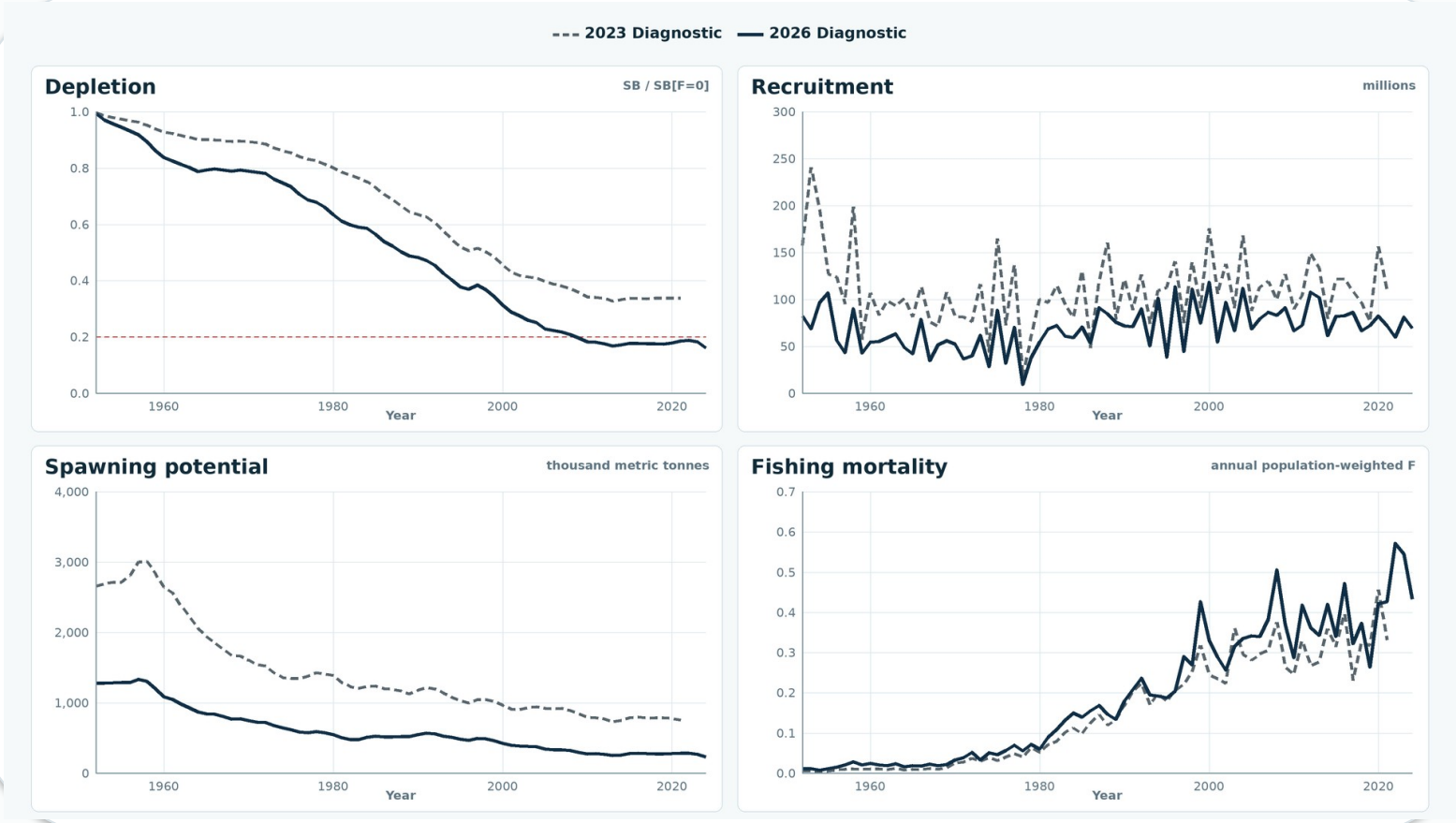
New releases, mixing periods and reporting

Recent Pacific Tuna Tagging Programme (PTTP) data were added; Coral Sea tags were removed; the time allowed for tags to mix was set by release group, and reporting constraints used new tag-seeding evidence.

DIFFERENT EVIDENCE ROLES CPUE tracks trends and links regional levels; lengths and ages describe population structure; tags inform population size, movement and survival but depend on mixing and reporting assumptions.

Source: WCPFC-SC22-2026-SA-WP-06, Executive Summary and Sections 4.4–4.7. Full fishery-by-data coverage is in Figure 6.

Stepwise: first compare the two Diagnostics



2023 VS 2026 DIAGNOSTIC

33.80% → 17.39%

The final stock-level ratio is lower

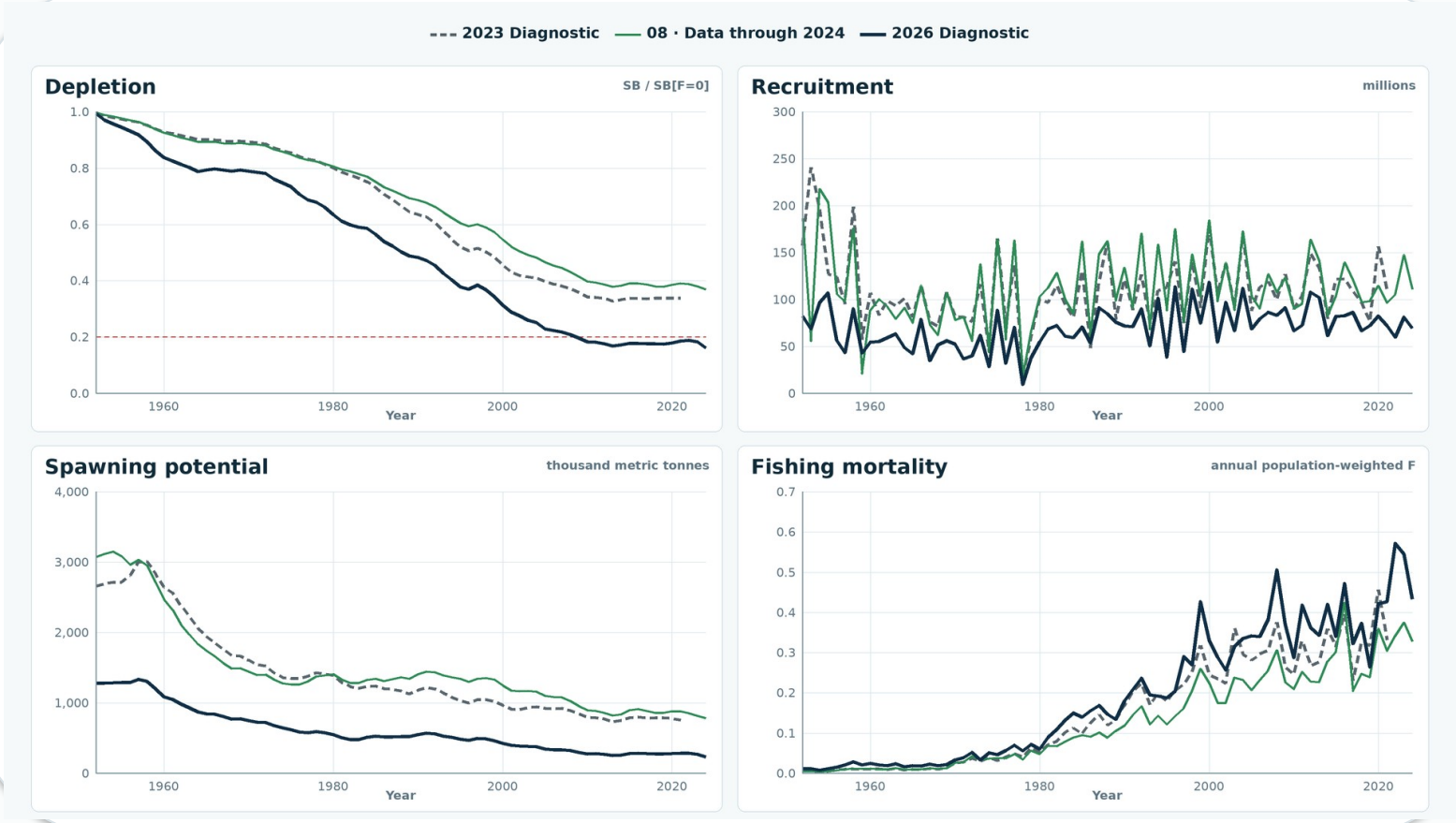
The 2026 Diagnostic also has a smaller spawning population and higher fishing mortality than the 2023 refit.

Recent depletion: 2018–21 vs 2021–24. These are bookends—not a single-change test.

The model-development path is cumulative. Intermediate fits help explain the route; they are not isolated causal tests.

Source: WCPFC-SC22-2026-SA-WP-06, Sections 6.1.1 and 7 and Figure 16.

Stepwise · 08: data through 2024 moved the endpoint



STEP 08 · FIRST NOTABLE CHANGE

2021 → 2024

Data extended through 2024

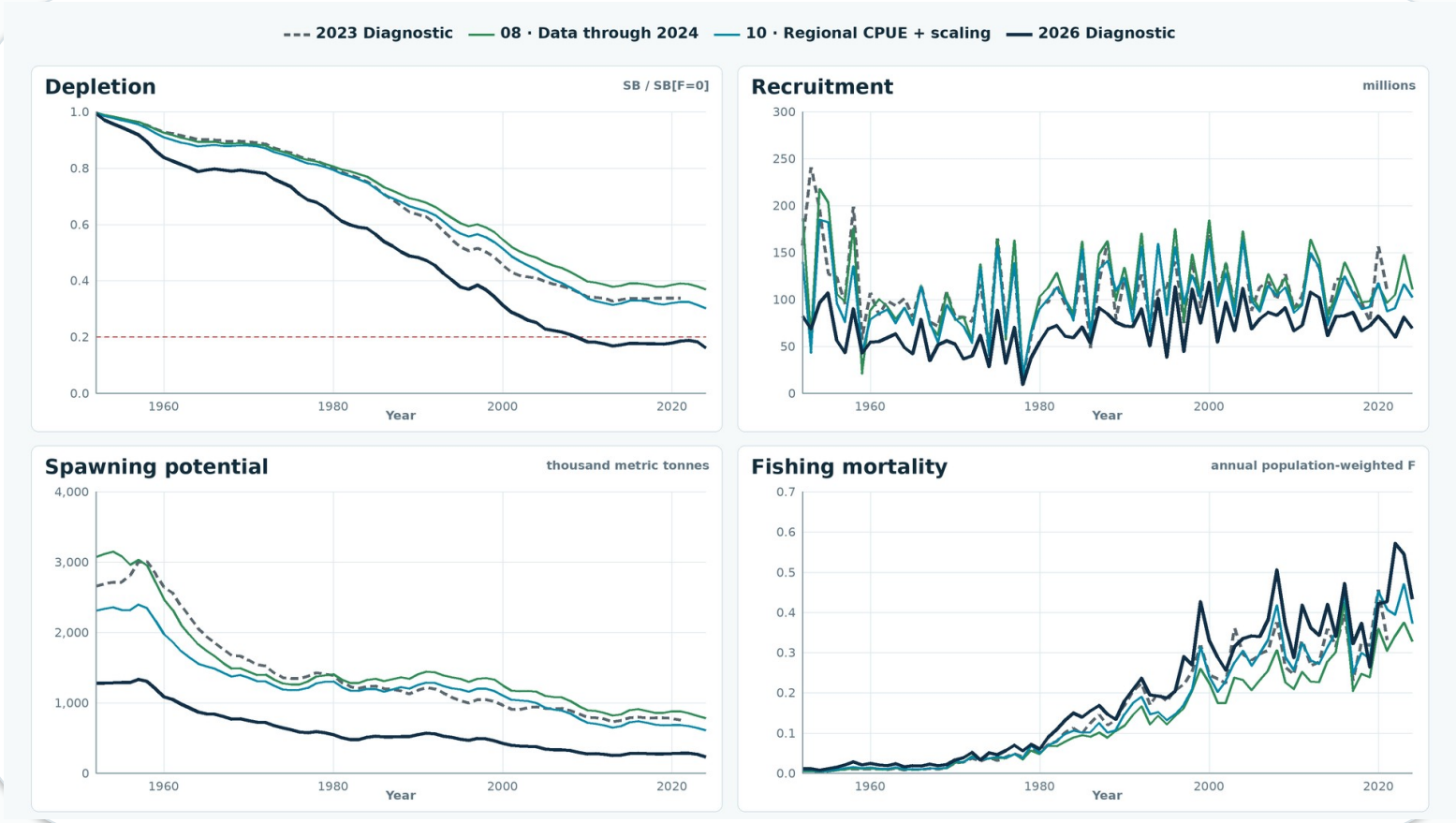
Three additional years move the end of every trajectory. The visible difference combines new observations with a later endpoint.

Not like-for-like: the data window also changed.

The model-development path is cumulative. Step 08 changes the data window as well as the fitted result.

Source: WCPFC-SC22-2026-SA-WP-06, Sections 6.1.1 and 7 and Figure 16.

Stepwise · 10: CPUE + scaling lowered the depletion ratio (more depleted)



STEP 10 · PAW-INFORMED

36.71% → 30.63%

−6.08 percentage points

Regional CPUE + global scaling

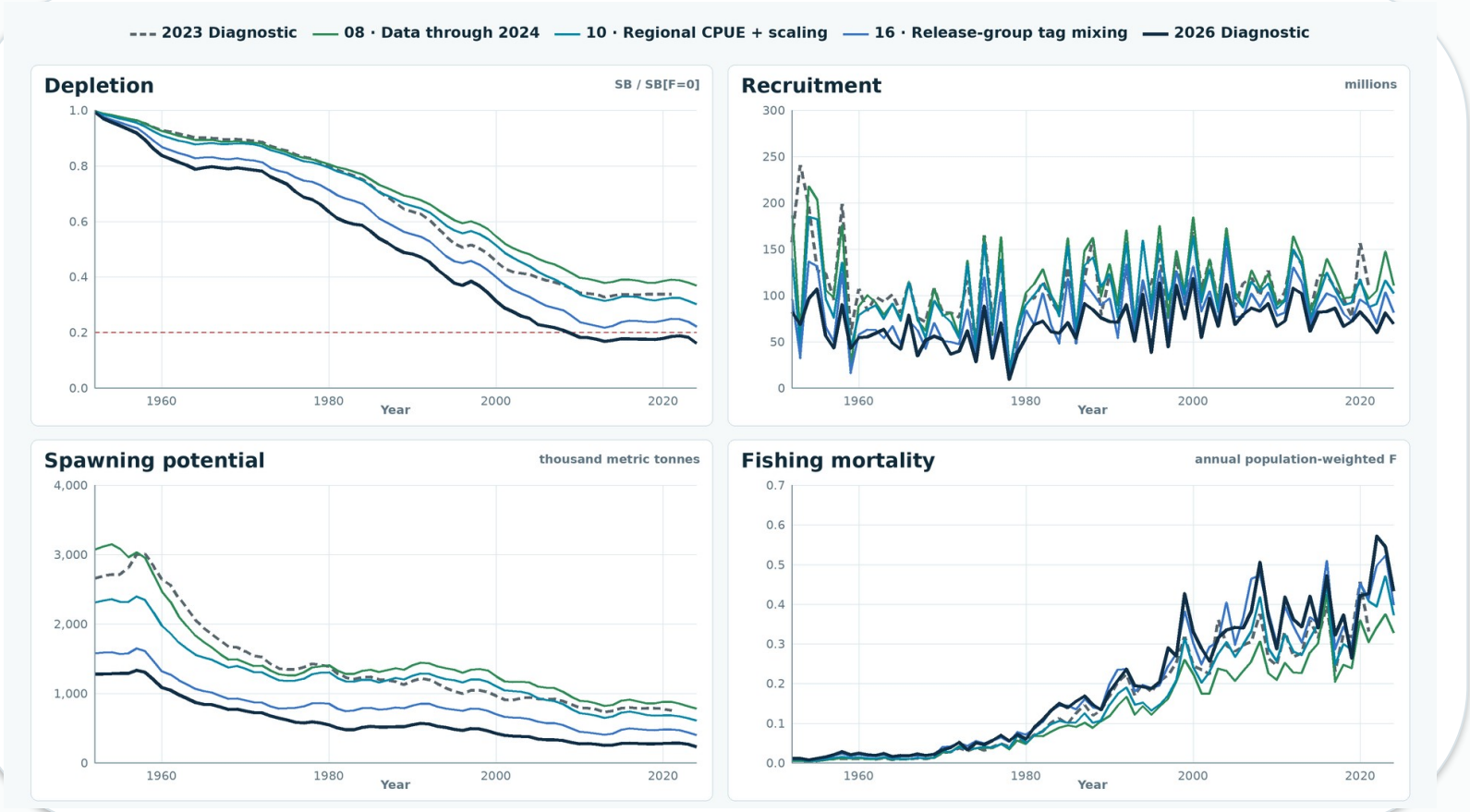
Regional abundance trends and the link among regional population levels were fitted together.

Ratio change versus the immediately preceding fit.

This is one point on a cumulative path, not the isolated causal effect of changing CPUE alone.

Sources: WCPFC-SC22-2026-SA-WP-06, Sections 6.1.1 and 7 and Figure 16; WCPFC-SC22-2026-SA-IP01.

Stepwise · 16: tag mixing lowered the depletion ratio (more depleted)



STEP 16 · CLEAR, BUT NOT MAJOR · PAW-INFORMED

29.05% → 23.06%

−5.99 percentage points

Mixing set by release group

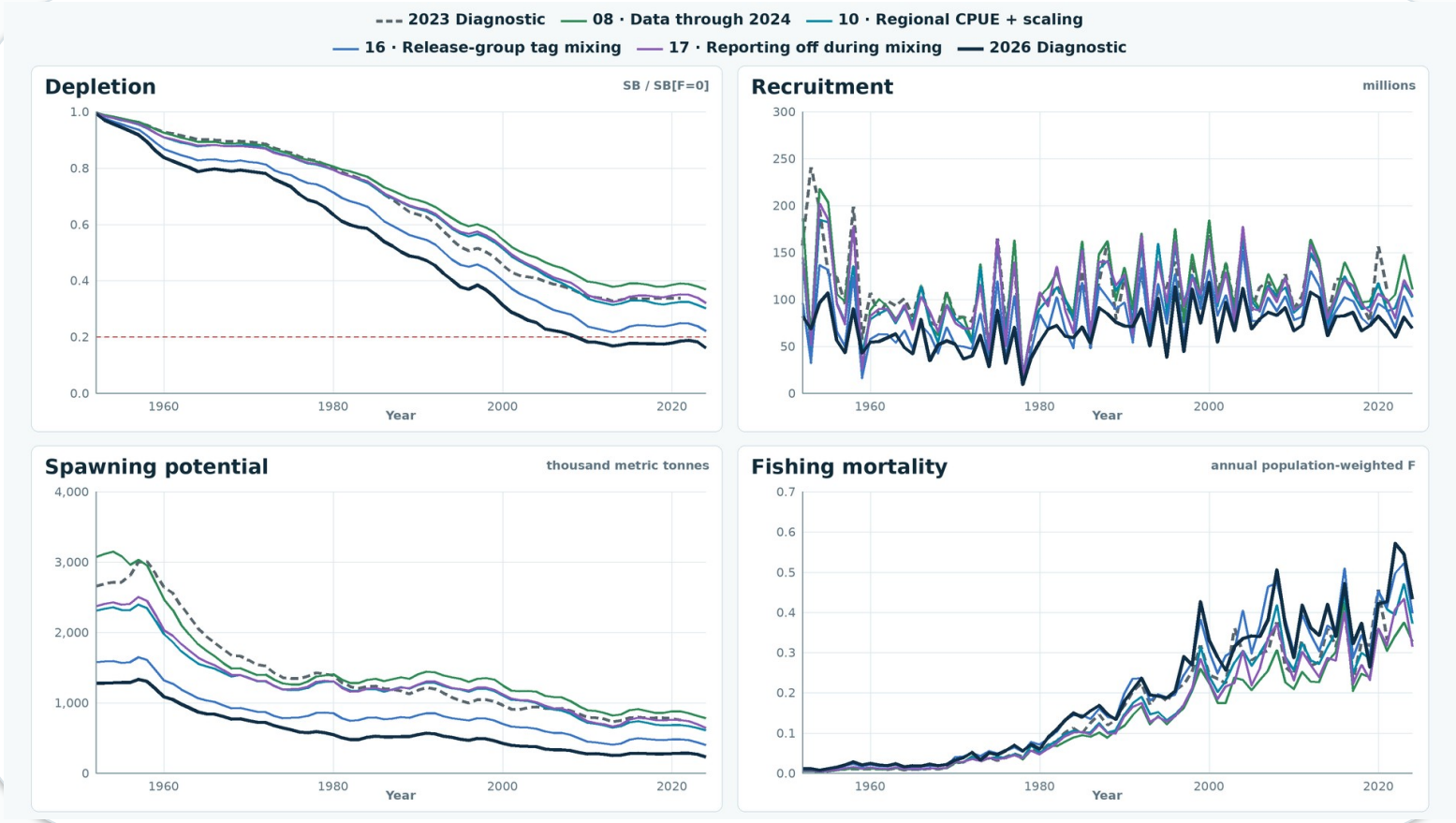
The period before tags inform population scale was tailored to each release group.

Lower inferred scale and higher fishing mortality.

This is one point on a cumulative path, not the isolated causal effect of tag mixing alone.

Sources: WCPFC-SC22-2026-SA-WP-06, Sections 6.1.1 and 7 and Figure 16; WCPFC-SC22-2026-SA-IP01.

Stepwise · 17: tag reporting raised the depletion ratio (less depleted)



STEP 17 · MOST NOTABLE PRE-T SCALING IMPACT

23.06% → 32.85%

+9.79 percentage points

Reporting adjustment off during mixing

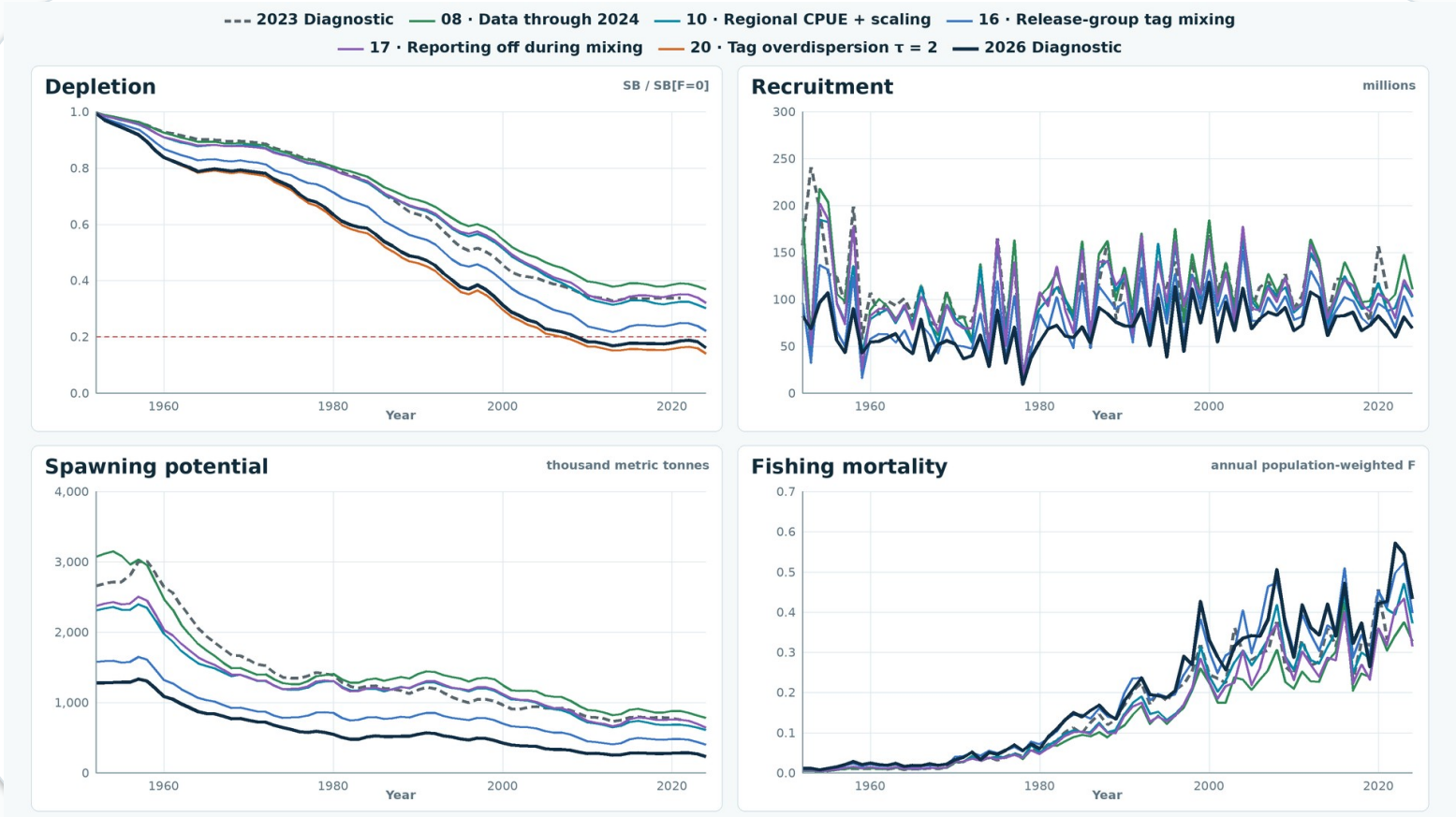
Tag returns during the mixing period were no longer adjusted by the reporting-rate model.

Ratio change versus the immediately preceding fit.

This is one point on a cumulative path, not the isolated causal effect of reporting treatment alone.

Source: WCPFC-SC22-2026-SA-WP-06, Sections 6.1.1 and 7 and Figure 16.

Stepwise · 20: τ=2 produced the largest drop in the depletion ratio



STEP 20 · LARGEST SCALING IMPACT

32.30% → 15.29%

−17.00 percentage points

Tag overdispersion fixed at τ=2

Allowing more extra variation in tag counts reduced their practical influence on population scale.

Later steps lead from this fit to the 2026 Diagnostic.

Each delta is against the immediately preceding fit. The full path is cumulative—not five isolated causal tests.

Source: WCPFC-SC22-2026-SA-WP-06, Sections 6.1.1 and 7 and Figure 16. Full path: Backup 2.

The central issue is estimated population size

“Population scale” means the total population size inferred from the same observed catches

LARGER INFERRED POPULATION

Tag data favour this size

The same catches are a smaller fraction of the population. Estimated fishing impact is lower and stock status is less depleted.

SAME CATCHES

DIFFERENT INFERRED SIZE

SMALLER INFERRED POPULATION

CPUE—and, to a lesser extent, length data—favour this size

The same catches are a larger fraction of the population. Estimated fishing impact is higher and stock status is more depleted.

DISAGREEMENT DOES NOT MEAN BAD DATA

Several fits reproduce similar historical trends and pass key checks, yet imply different total population sizes and status indicators. That disagreement is why the advice combines multiple models.

Source: WCPFC-SC22-2026-SA-WP-06, Executive Summary and Sections 7.2.5 and 8.2. The report does not identify one data source as “wrong.”

τ in one picture: allow extra scatter in field tags

P

POISSON BOUNDARY · τ≈1

A tight count model

After accounting for modelled effects, the model allows only the count variation expected under the Poisson mean–variance rule.

Variance ≈ mean

T

FIELD TAG RECAPTURES

Real-world clustering remains

Release, area, year, fishery, mixing-history and reporting processes remain heterogeneous.

- release groups
- areas
- years
- fisheries
- reporting
- mixing

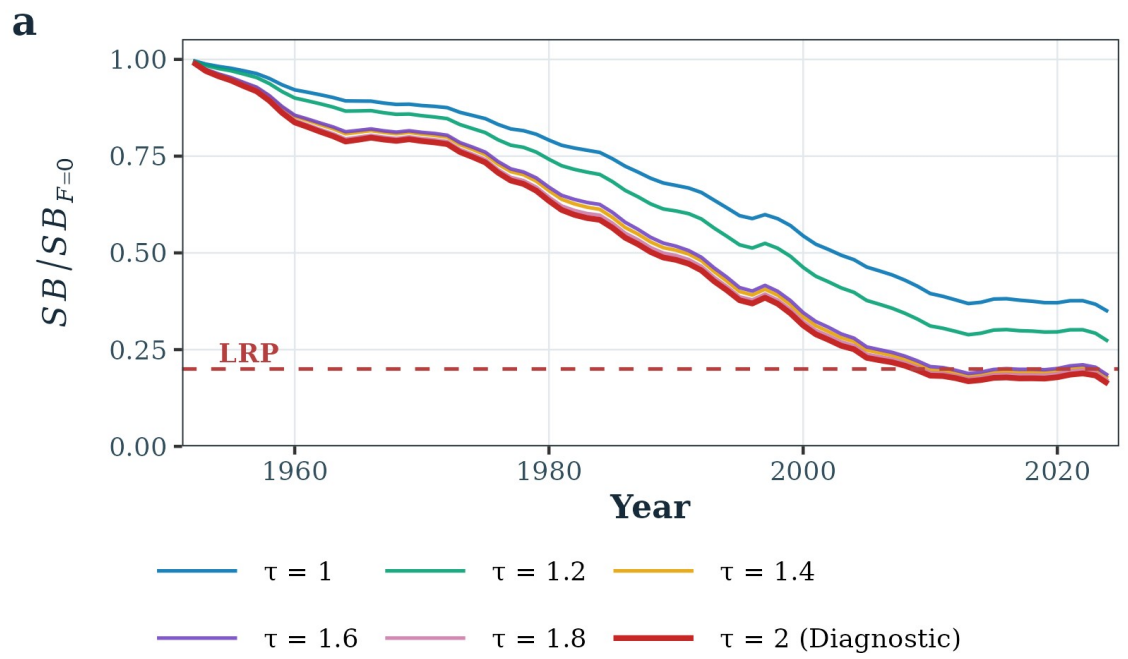
Variance > mean

Moving above τ≈1 allows extra variation, so tags are not treated as more precise—and more influential—than warranted. It changes their practical influence, but does not by itself tell us whether stock status moves up or down.

Source: WCPFC-SC22-2026-SA-WP-06, Executive Summary and Section 5.5.3. τ is part of the negative-binomial model for tag counts; it is not simply a manual weight.

FIGURE 56 · CHANGE ONE SETTING AT A TIME

Most of the τ response occurs below about 1.4



ONE-SETTING SENSITIVITY TESTS AROUND THE DIAGNOSTIC MODEL

EARLIER

Documentation stated τ=2; MFCL applied effective τ≈1.02.

NEAR 1 → 1.4

Large change in estimated population size and depletion.

ABOVE 1.4

Much smaller change; 1.4 already reduces tag influence substantially.

ENSEMBLE

τ=1.2, 1.3 and 1.4 cover the main transition.

1.2–1.4 are alternative assumptions—not an estimated confidence interval. τ=1.3 was not tested in Figure 56.

WHAT IS SHOWN

Fitted values only

No within-model parameter uncertainty.

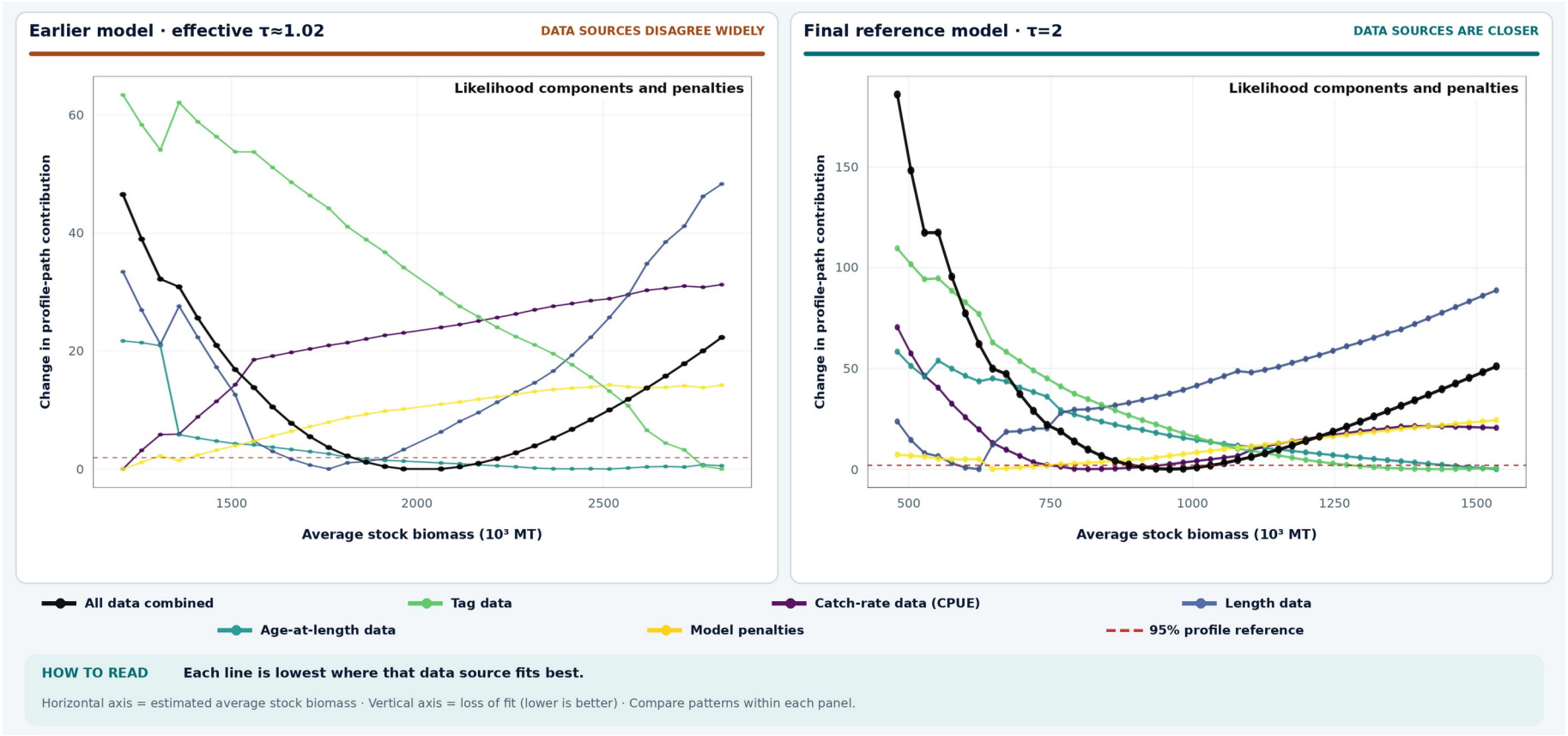
2023 BOUNDARY

Correction—not invalidation

The τ correction alone does not invalidate the 2023 assessment.

Source: WCPFC-SC22-2026-SA-WP-06, Sections 5.5.3, 6.2.2 and 7.4.8, and Figure 56. These are one-off sensitivity tests. τ=1.3 was an intermediate ensemble value, not a Figure 56 run.

Earlier near-Poisson vs final τ=2: the scale conflict narrows



WHAT CHANGED The final setup reduces—but does not remove—the disagreement among data sources. Other model settings also differ, so this is a full-setup comparison—not a controlled τ-only test.

Sources: archived Kflow profile outputs and WCPFC-SC22-2026-SA-WP-06, Figure 33. Biomass x-axis ranges differ and each curve is reset within its panel; compare shapes, not raw heights or x positions.

Why advice combines 80 plausible models

EVIDENCE

SAME DATA

One evidence base, but real disagreement over estimated population size.

PLAUSIBLE ASSUMPTIONS

80 MODELS

Alternative biological assumptions and choices about how observations are modelled.

COMBINED ADVICE

COMBINED RESULT

A quality-screened, equal-weight result from all retained models—not one chosen fit.

DIAGNOSTIC MODEL

Reference fit for development

Used to show fit, checks and sensitivity tests; it lies near the more-depleted edge of the ensemble.

UNCERTAINTY ENSEMBLE

Combined result for advice

Keeps plausible population-size assumptions instead of choosing one as true.

INTERPRETATION

Probabilities apply to this model set

Equal weights describe uncertainty across the models and assumptions included here.

ONE DIAGNOSTIC FIT IS NOT THE ADVICE

No one fit resolves the population-size disagreement. Advice therefore uses the combined results from all 80 retained models.

Source: WCPFC-SC22-2026-SA-WP-06, Sections 6.2.2, 7.5.1 and 8.2.

RANGES INFORMED BY SINGLE-CHANGE SENSITIVITY TESTS

The ensemble varies six model assumptions

h

RECRUITMENT RESILIENCE

Steepness (h): how well recruitment is maintained when spawning stock is low.

M

NATURAL MORTALITY

Deaths other than fishing; fixed in each fit and varied across models after unstable joint M–growth estimation in 2023.

K

TAG MIXING RULE

K=0.05–0.35; controls when recaptures are treated as sufficiently mixed.

τ

EXTRA TAG-COUNT VARIATION

Overdispersion (τ) fixed at 1.2, 1.3 or 1.4 to cover plausible tag influence.

R

REPORTING BEFORE MIXING

Two alternatives for adjusting reporting rates before tags are treated as mixed.

E

FISHING-EFFICIENCY CHANGE

Five effort-creep combinations for the primary and secondary CPUE series.

FROM TESTS TO ENSEMBLE

Ranges → 100 random combinations

Single-change sensitivities informed the ranges; random combinations represented interactions.

BOUNDARY

Structured—not exhaustive

The global-CPUE scaling period mattered in a separate test, but was not varied here.

Source: WCPFC-SC22-2026-SA-WP-06, Section 6.2.2. Full design and retention details are in backup.

80 models passed the quality screen; each has equal weight

PLANNED

100

Random configurations across the six assumptions.

RETAINED

80

All met the fitting threshold.
10 high-gradient and 10 incomplete fits were excluded.

PARAMETER UNCERTAINTY

62 + 18

62 include parameter draws.
18 contribute their fitted result only.

EQUAL TOTAL WEIGHT PER RETAINED MODEL

The two tag-reporting options were planned 50/50, but retention was 34/46 because one option caused more numerical failures. This reflects numerical stability—not stronger scientific support.

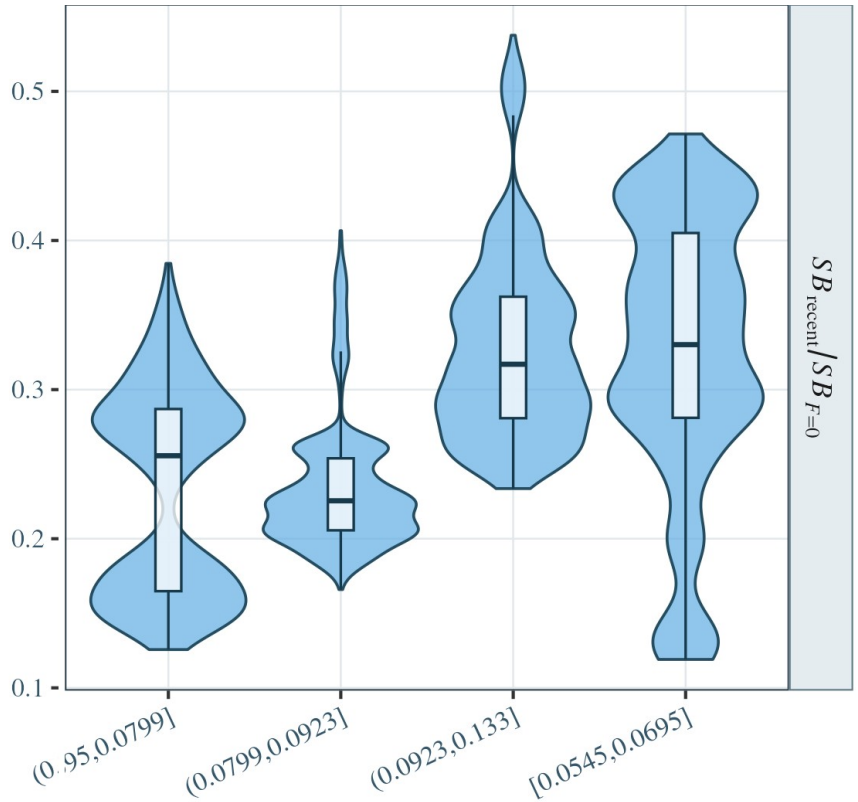
WITHIN-MODEL PARAMETER UNCERTAINTY

It is included for 62 models and unavailable for the other 18. Those 18 are not failed fits; all met the fitting threshold.

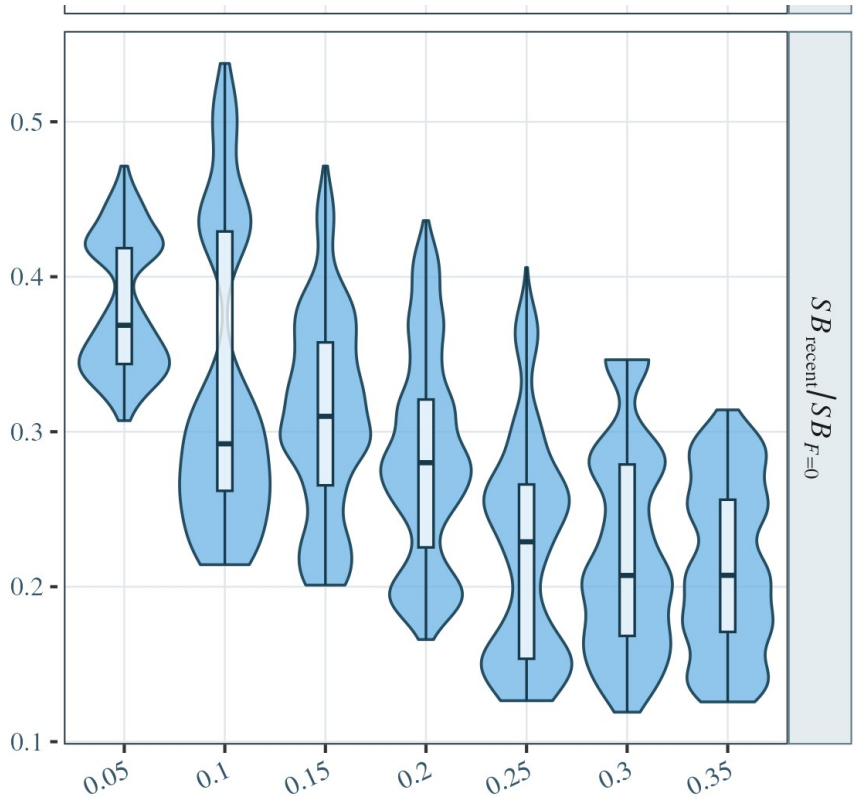
Source: WCPFC-SC22-2026-SA-WP-06, Section 7.5.1 and Tables 8 and 12. Every retained model receives equal total model weight.

Natural mortality and tag mixing show the clearest grouped differences

NATURAL MORTALITY (M)



TAG-MIXING RULE (K)

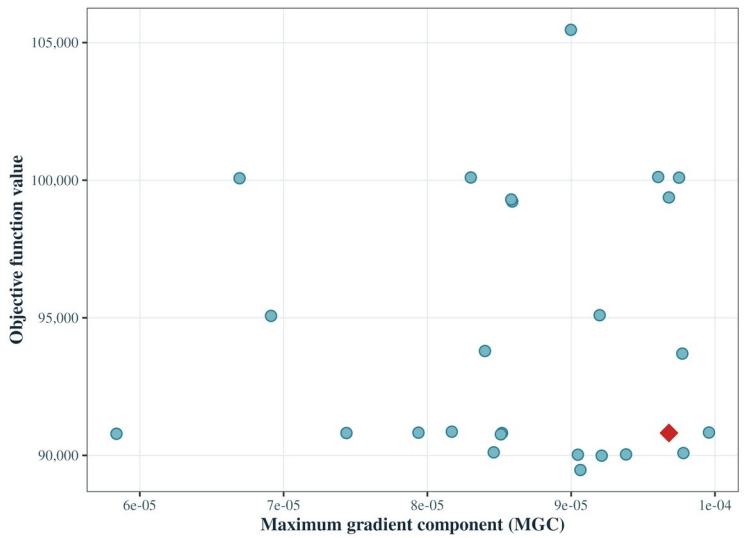


LARGEST GROUPED DIFFERENCES τ matters in one-change tests; M and K separate the ensemble most clearly. All six assumptions vary together, so the τ groups overlap even though τ strongly affected the sensitivity below about 1.4.

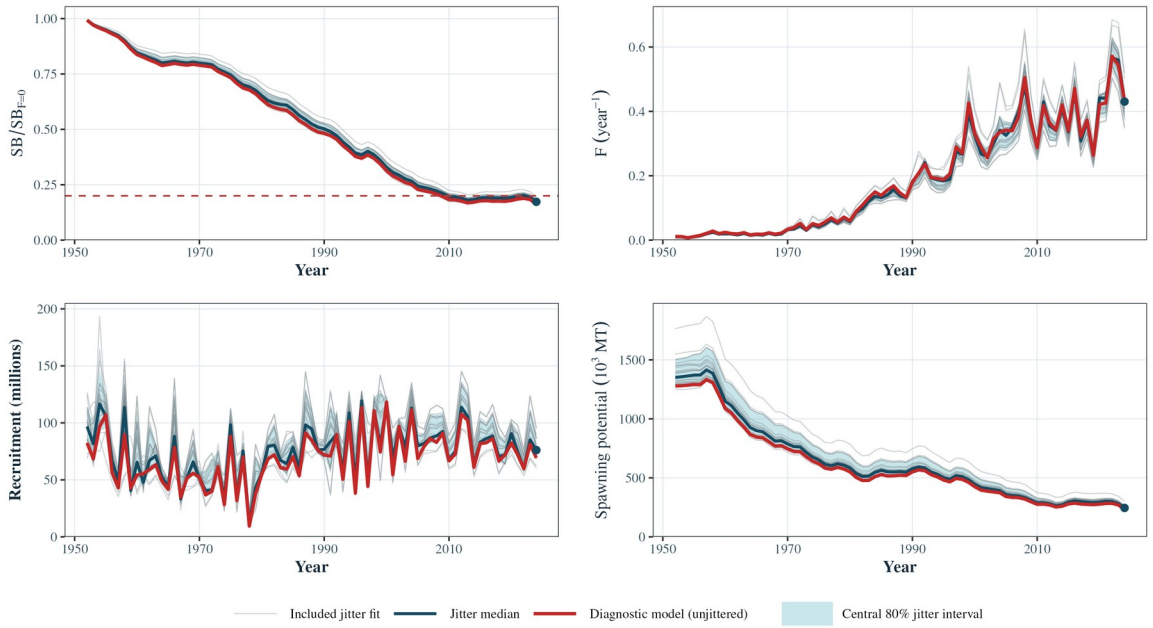
Source: WCPFC-SC22-2026-SA-WP-06, Sections 7.4 and 7.5.1, Figures 58–60. Grouped summaries while all other assumptions vary—not controlled cause-and-effect tests.

Starting-value tests found different fits—but status varied less

FIT SCORE · FIGURE 27



STATUS TIME SERIES · FIGURE 28



30 starts → 25 retained · 10 reached a lower objective value

Status trajectories varied less

Spread is smaller than in fit score and some parameters

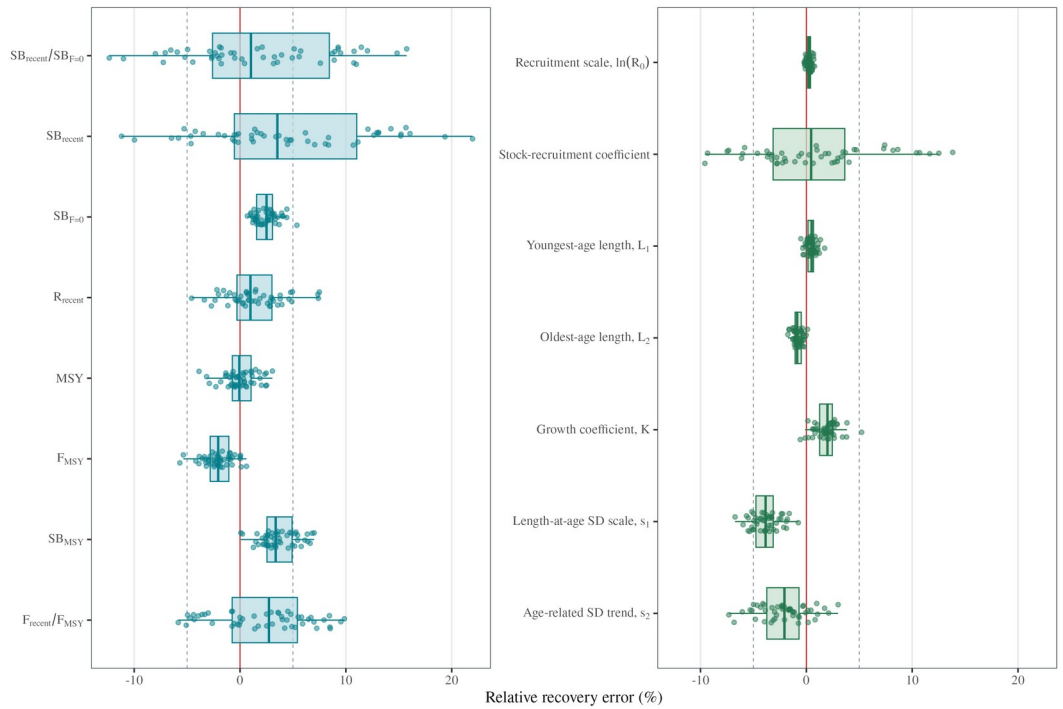
WHAT THE TEST SHOWS

This was the first phase-aware jitter in this assessment series: starting values were perturbed through the full fitting path. It reveals local fit structure, but does not show that every ensemble model—or the combined advice—is unstable to the same degree.

Source: WCPFC-SC22-2026-SA-WP-06, Sections 7.2.1–7.2.2 and Figures 27–29. Figures 27–28 are shown without redrawing; the Diagnostic passed the gradient and curvature checks (MGC 9.68×10−5; positive-definite Hessian).

Self-test: the model largely recovers known simulated values

FIGURE 30 · RECOVERY ERROR FROM 50 SIMULATED-DATA REFITS



REFITS

50 / 50 completed

MEDIAN RECOVERY ERROR

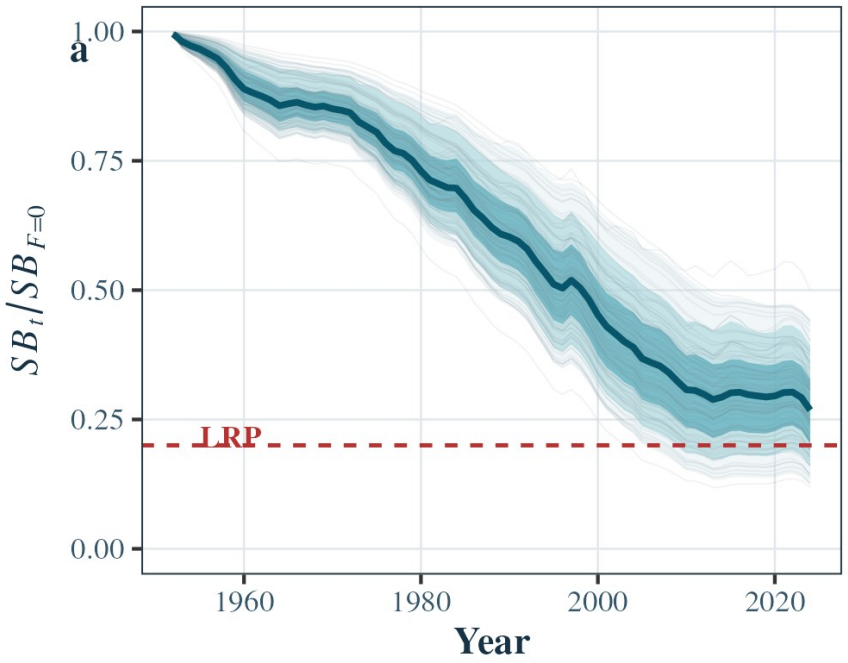
−2.1% to +3.5%

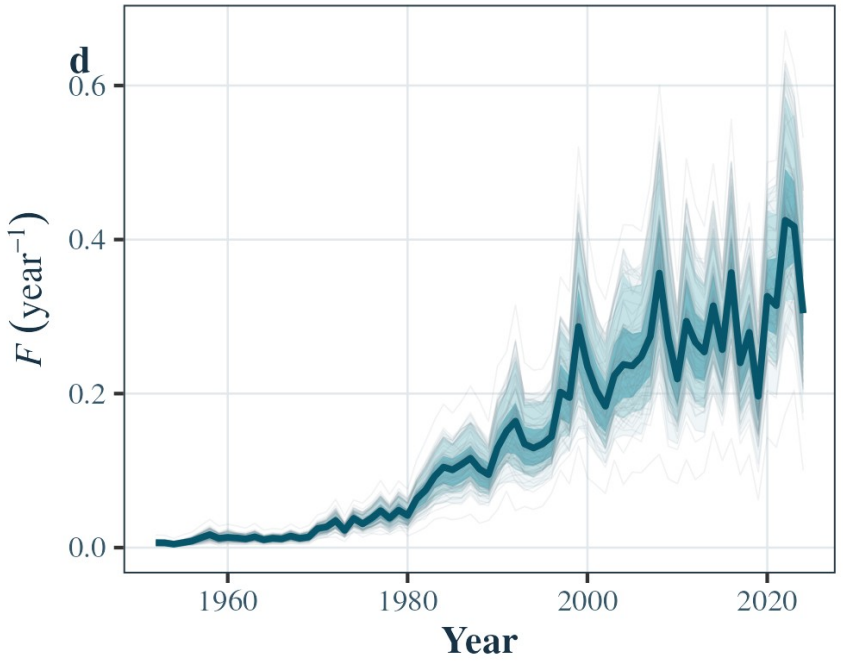
WHAT THIS SUPPORTS

This supports using the Diagnostic model as the starting point for sensitivity tests and ensemble models. It asks whether the model recovers values under its own assumptions—not whether those assumptions are fully realistic or one population size is uniquely correct.

Source: WCPFC-SC22-2026-SA-WP-06, Section 7.2.3 and Figures 30–31. The self-test evaluates internal recovery under known simulated values; it is not an external validation of the assessment assumptions.

The stock became more depleted until about 2010, then stabilised





50% interval 80% interval 95% interval

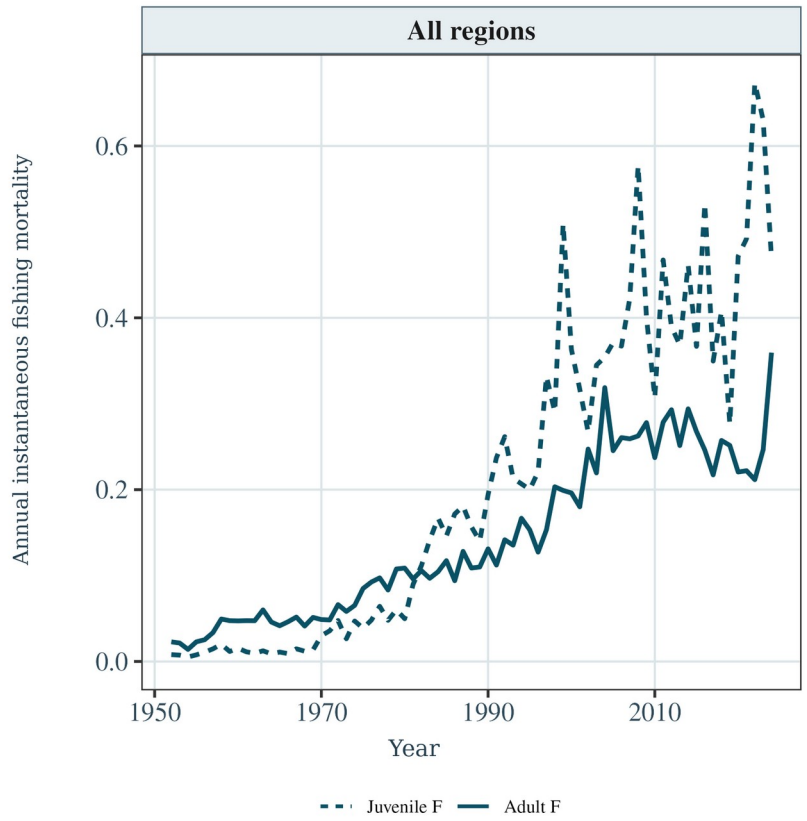
Median

READ THE HISTORY Across the 80 models, the stock-level ratio declined to about 2010 and was comparatively stable afterwards. Fishing pressure rose until around 2000, then remained variable without a clear long-term increase.

Source: WCPFC-SC22-2026-SA-WP-06, Figure 62 and Sections 7.5.1, 8.1 and 8.3. Shading combines differences among retained models and available within-model uncertainty.

FIGURE 48 · ALL REGIONS

Recent fishing mortality remains higher for juveniles



LONG-TERM PATTERN

The gap opens around 1980

Juvenile F rose above adult F as associated purse-seine fisheries expanded and has remained higher stock-wide.

RECENT YEARS

High and variable juvenile pressure

Juvenile F peaked in 2022–2023. The terminal adult-F rise may partly reflect revised Indonesian catch allocation.

INTERPRETATION

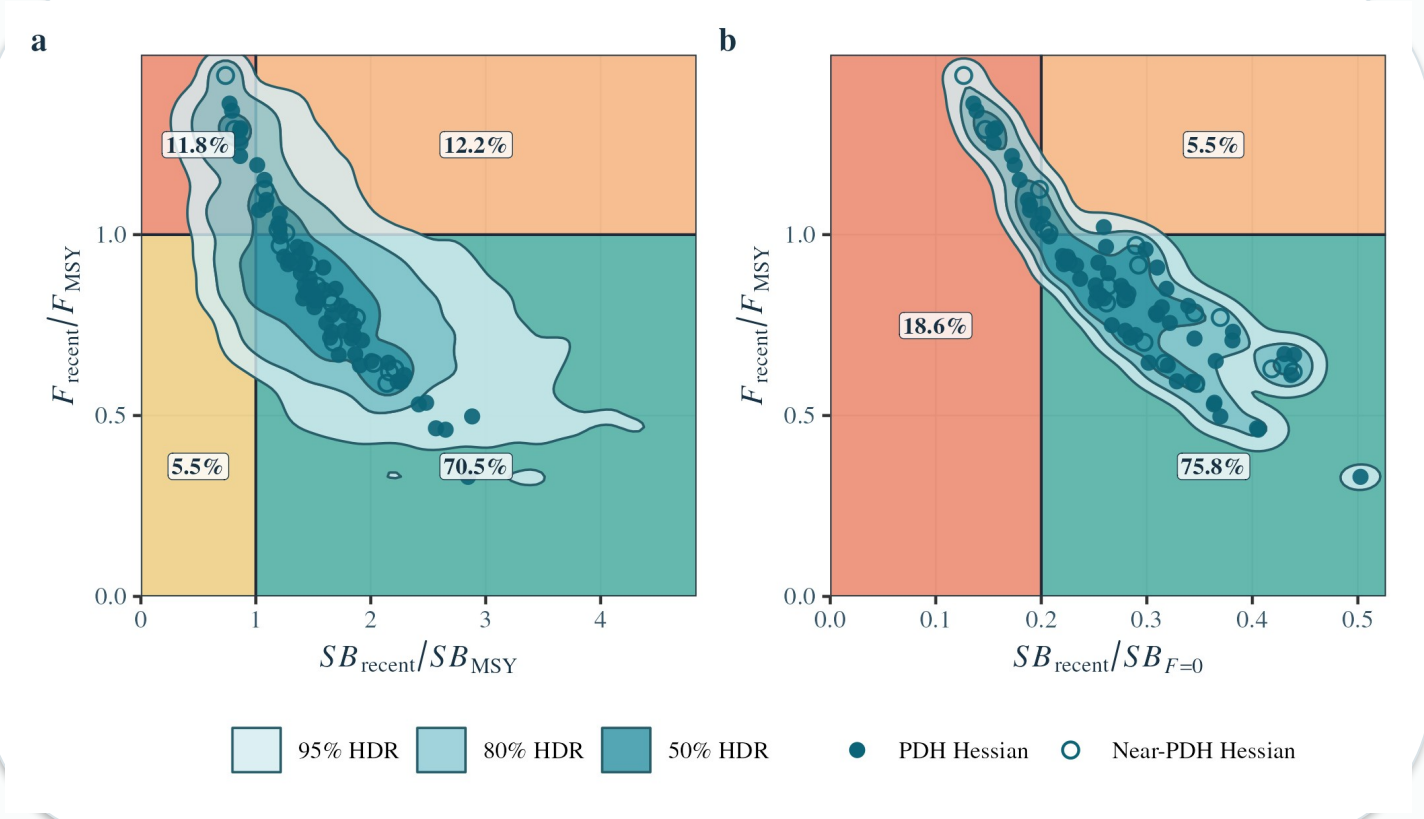
One reference-model history—not ensemble uncertainty

Figure 48 shows the Diagnostic-model history. It does not quantify each fishery's contribution or show ensemble uncertainty.

KEY PATTERN The age pattern—not every regional spike—is the message: recent fishing mortality remains considerably higher for juveniles than adults.

Source: WCPFC-SC22-2026-SA-WP-06, Section 7.3.9 and final Figure 48. The displayed panel is the all-regions crop from the Rev.1 report figure and extends through 2024.

Current status: most results are favourable, but uncertainty crosses both thresholds



READ THE MAJURO PANEL FIRST

Status relative to the LRP and FMSY

x-axis: recent depletion; LRP at 0.20
y-axis: recent F/FMSY; threshold at 1
dots: one fitted result per retained model
contours: where combined results concentrate

75.8% favourable

18.6% below LRP

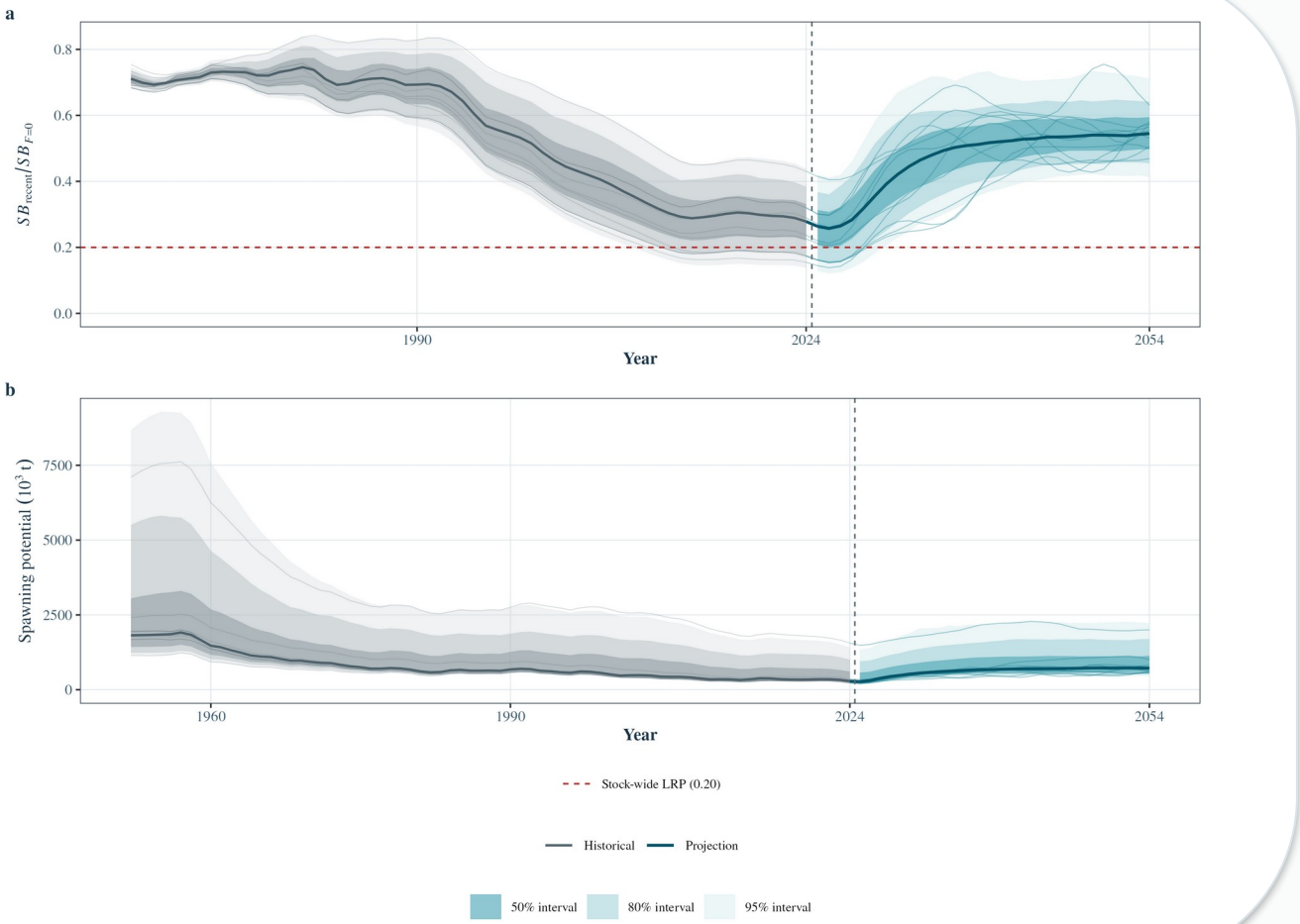
24.1% above FMSY

The last two overlap. Do not add them.

The median is above the LRP and below FMSY, but combined results cross both thresholds. 'Most results favourable' does not mean zero risk.

Source: WCPFC-SC22-2026-SA-WP-06, Figure 63 and Tables 8 and 10. Kobe probability below SBMSY: 17.3%.

Slightly below the CMM objective; conditional projections rise



CMM 2025-02 OBJECTIVE

Recent level is slightly below 2012–2015

2012–15		2021–24	CMM ratio
28.8%	→	27.8%	96.7%

Average across one fitted result from each model; parameter uncertainty is not included. This differs from the 27.7% combined median.

YEAR · MEDIAN DEPLETION METRIC · BELOW LRP

2030	0.355	5.2%
2040	0.517	0 of 800
2054	0.545	0 of 800

SCENARIO—NOT FORECAST

Catches stay fixed by fishery and quarter at 2022–2024 averages; recruitment is resampled.
Includes 80 model assumptions; excludes within-model parameter uncertainty, catches departing from the scenario and changing fishing patterns.
Zero of 800 runs is not zero real-world risk.

Source: WCPFC-SC22-2026-SA-WP-06, Sections 5.8.5 and 7.5.6, Tables 9 and 13 and Figure 68. Projection = 80 models × 10 recruitment sequences.

What the 2026 assessment concludes

LONG-TERM HISTORY

Decline, then comparative stability

Spawning potential declined until about 2010 and has since been more stable. Fishing pressure rose until around 2000, then broadly stabilised with strong year-to-year variation; recent juvenile F remains higher than adult F.

CURRENT STATUS

Most results favourable—important risk remains

Median depletion is 0.277 and median F/FMSY is 0.838. The distribution assigns 18.6% below the LRP and 24.1% above FMSY; the CMM ratio is 96.7%.

COMPARED WITH 2023

Less favourable status—not a one-change result

The 2026 results are less comfortable. The Diagnostic model lies near the more-depleted edge of the retained ensemble, and the shift cannot be attributed to τ—or any other single update—alone.

HOW TO INTERPRET IT

Use the combined 80-model result

Absolute population size remains uncertain: no single data source, tag-release group or τ identifies one size as uniquely correct. Status indicators therefore use the equal-weight ensemble, not one Diagnostic fit.

BOTTOM LINE **Less comfortable than 2023, but the median remains on the favourable side of the LRP and FMSY thresholds. Both the uncertainty and the population-scale conflict matter for interpretation.**

Source: WCPFC-SC22-2026-SA-WP-06, Executive Summary and Sections 8.1–8.3.

FOUR UNRANKED WORKSTREAMS

Priorities before the next benchmark

1

TAG EVIDENCE

Understand field-tag variability

Continue external tag analysis (Project 123) and a focused review.
Build one consistent tag-recapture model across programmes and test it by simulation.

2

WHAT CAN BE ESTIMATED

Test what the model can support

Test sensitivity to starting values, recruitment and movement.
Expand self-tests and compare alternative movement and recruitment structures.

3

OBSERVATION MODELS + DATA

Strengthen population-scale evidence

Test regional recruitment and size-data coverage.
Revisit size weighting, CPUE scaling and historical Indonesian catch allocation.

4

SOFTWARE + PROCESS

Make critical settings reproducible

Automate checks of critical model settings and their interactions.
Allow more time for data preparation, development and research.

WHAT THIS MEANS NOW Use the ensemble for current advice; improve the evidence for absolute population size before the next full benchmark assessment. These workstreams strengthen the next assessment—they do not replace the present one.

Source: WCPFC-SC22-2026-SA-WP-06, Executive Summary recommendations and Section 8.3. Actions are grouped for presentation; numbering does not imply a ranking.

MANY PEOPLE MADE THIS ASSESSMENT POSSIBLE

Acknowledgements

FISHERY KNOWLEDGE AND DATA

Across the WCPO

WCPFC members, territories and data providers; observers, port samplers, tagging teams and fishers.

ASSESSMENT COLLABORATION

From data to review

2026 PAW participants; Pacific Community data, assessment and computing teams; the SPC sclerochronology laboratory; and Japan FRA for age–length data.

PROGRAMME SUPPORT

Enabling the work

WCPFC Scientific Services and Additional Resourcing to SPC, and EU PEUMP Phase 2 support for the CPUE work.

A PERSONAL THANK YOU

To my family—for their patience and support throughout this assessment.

Thank you

Questions · technical backup follows

Source: WCPFC-SC22-2026-SA-WP-06, Section 9. Personal closing line added by the presenter.

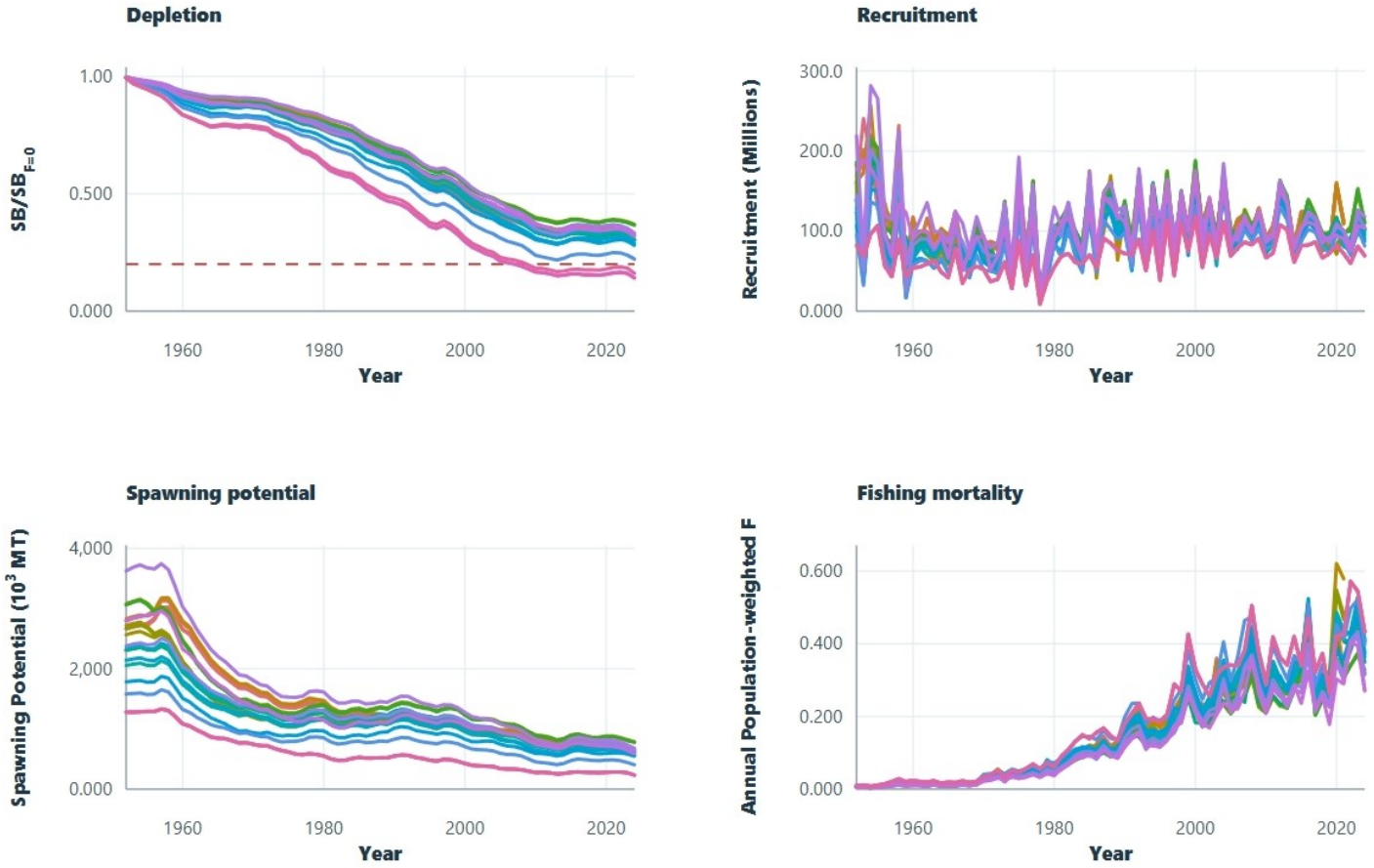
Exact periods, intervals and thresholds

MEASURE	EXACT COMPARISON	2026 RESULT	INTERVAL OR THRESHOLD
Recent depletion **SBrecent / SBF=0**	Mean 2021–2024 spawning potential relative to the 2014–2023 no-fishing level	**0.277**	50%: 0.217–0.336 80%: 0.172–0.395 95%: 0.136–0.442 LRP 0.20 · 18.6% below
Stock relative to MSY **SBrecent / SBMSY**	Mean 2021–2024 spawning potential relative to the equilibrium MSY level in each model	**1.532**	80%: 0.822–2.572 Higher than 1 indicates above SBMSY
Fishing pressure **Frecent / FMSY**	Mean 2020–2023 fishing mortality relative to equilibrium FMSY in each model	**0.838**	50%: 0.702–0.983 80%: 0.592–1.206 95%: 0.460–1.351 24.1% above 1
CMM 2025-02 objective **D2021–24 / D2012–15**	Mean depletion in 2021–2024 relative to 2012–2015, using one fitted value from each model	**0.278 / 0.288 = 0.967**	Central fits only; separate from the combined 0.277 median

The 18.6% below-LRP and 24.1% above-FMSY probabilities overlap and must not be added. Intervals are conditional on the equal-weight ensemble.

Source: WCPFC-SC22-2026-SA-WP-06, Tables 8–10 and Figure 63. Table 8 intervals are equal-tailed.

All 23 fitted stepwise configurations



HOW TO USE THE FULL PATH

Later fits inherit earlier changes, except for the two Step-14 alternatives. Intermediate fits were not fully tuned and diagnosed. Use the path to trace development—not to assign isolated causal effects.

23 fitted configurations
22 numbered steps

Source: WCPFC-SC22-2026-SA-WP-06, Sections 6.1.1 and 7, Figures 16–17. Interactive viewer: pacificcommunity.github.io/ofp-sam-bet-2026-stepwise/interactive-model-viewer.html

PDH-only and draw-count checks changed the summaries little

SUMMARY	ALL RETAINED MODELS	PDH-ONLY MODELS	LARGEST 50–100 DRAW DIFFERENCE
Depletion	0.277 (0.172–0.395)	0.275 (0.173–0.393)	0.0005
SB / SBMSY	1.532 (0.822–2.572)	1.522 (0.828–2.752)	0.0396
F / FMSY	0.838 (0.592–1.206)	0.838 (0.577–1.201)	0.0034

PDH-ONLY COMPARISON

Similar medians and 80% intervals

Restricting the calculation to the better-conditioned fits changed the three summaries little.

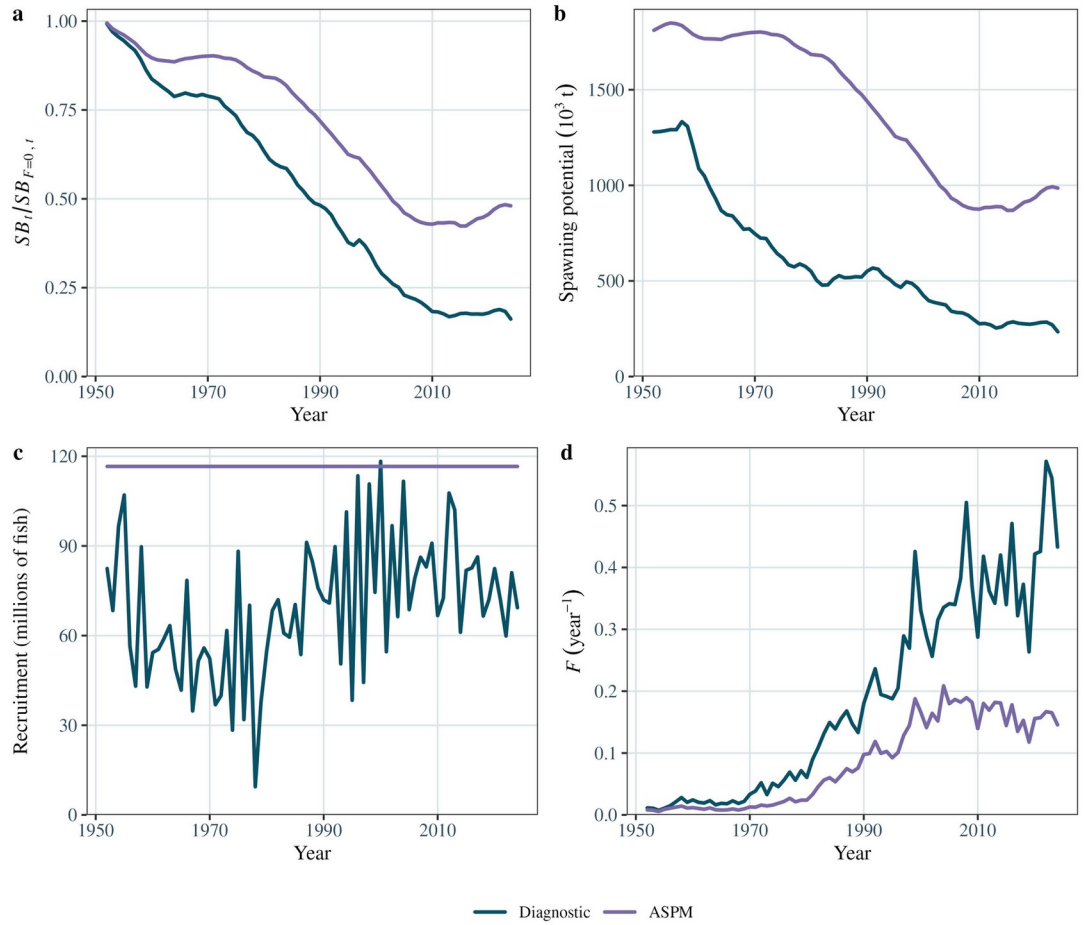
IMPORTANT BOUNDARY

This is not a filtering-impact test

These checks do not quantify the effect of the 20 planned models excluded by the quality screen.

Source: WCPFC-SC22-2026-SA-WP-06, Section 7.5.1 and Table 12. Parentheses are 80% intervals.

ASPM: CPUE implies a larger stock scale



2024 DEPLETION

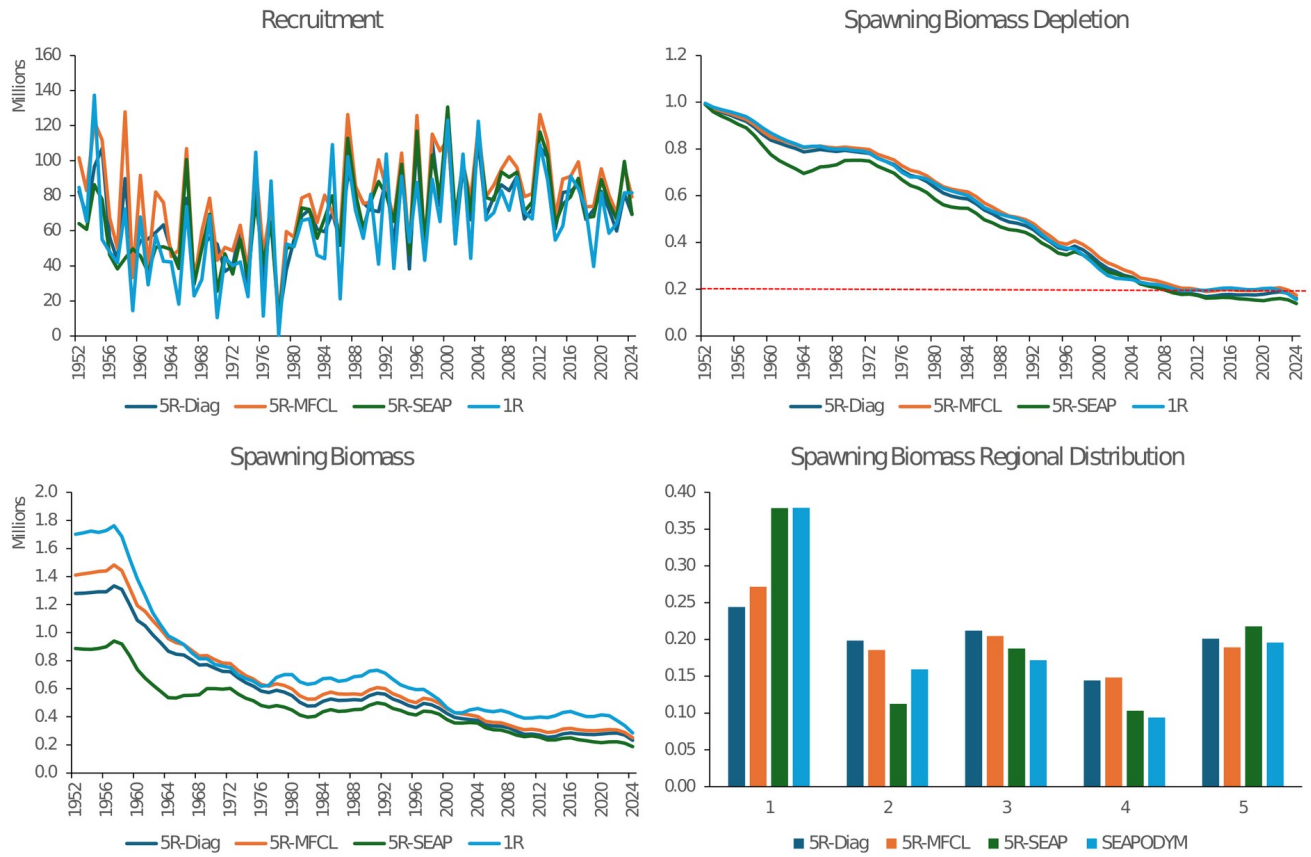
0.480 ASPM
0.162 integrated model

The CPUE-only, constant-recruitment model supports the population-scale tension seen in the likelihood profiles.

It is a diagnostic comparison—not an alternative assessment result.

Source: WCPFC-SC22-2026-SA-WP-06, Section 7.2.6 and Figure 38. The models make different structural and data assumptions.

No-tag check: the broad pattern remains

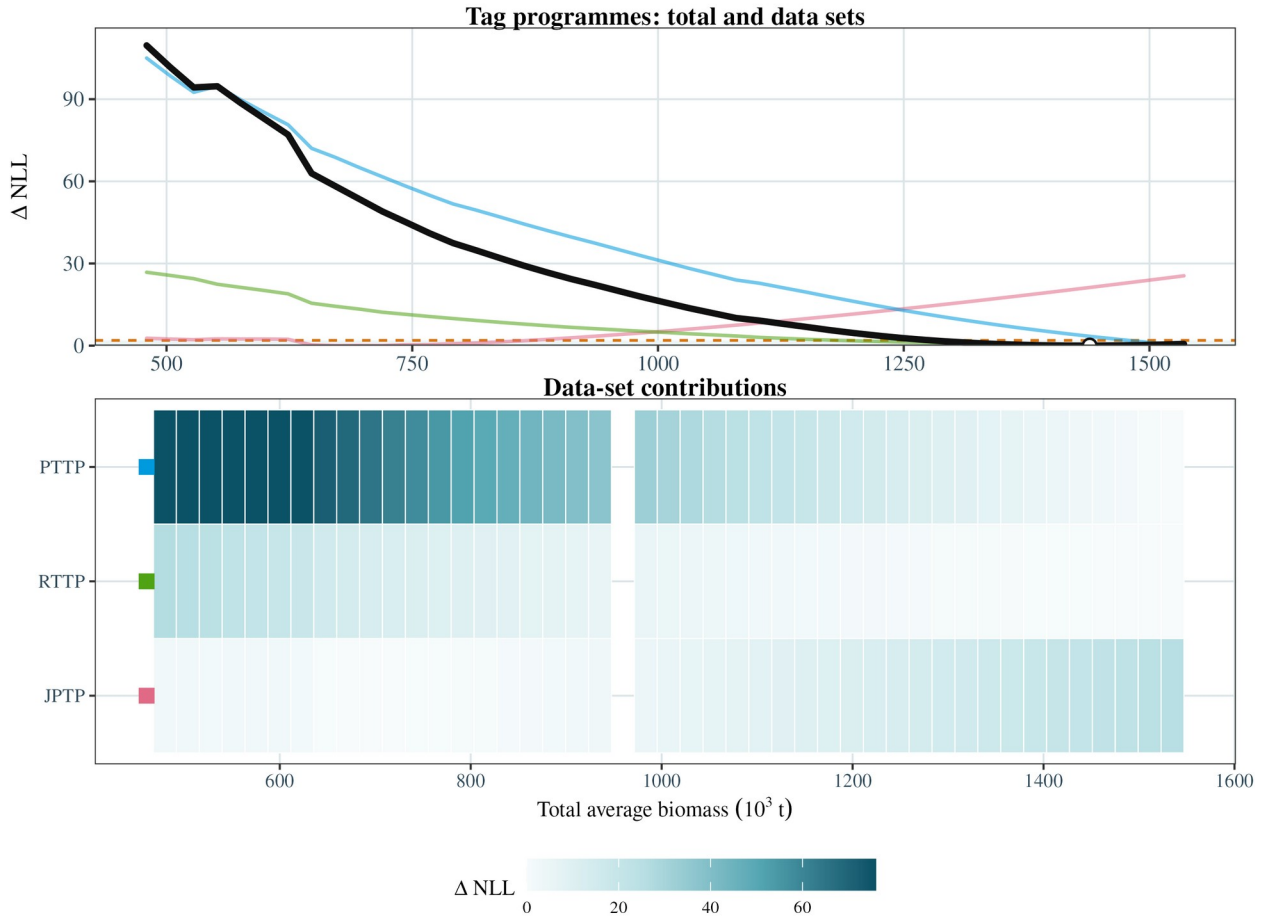


CLOSEST COMPARISON

Removing tags gives about +14% recruitment, +9% spawning biomass and +9% depletion. This supports the broad signal; it does not show that tags are irrelevant.

Source: WCPFC-SC22-2026-SA-WP-06, Section 7.2.7 and Appendix A.1, Figure A3. Movement is fixed for this comparison.

Which tagging programmes favour a larger population?



PROGRAMME SIGNAL

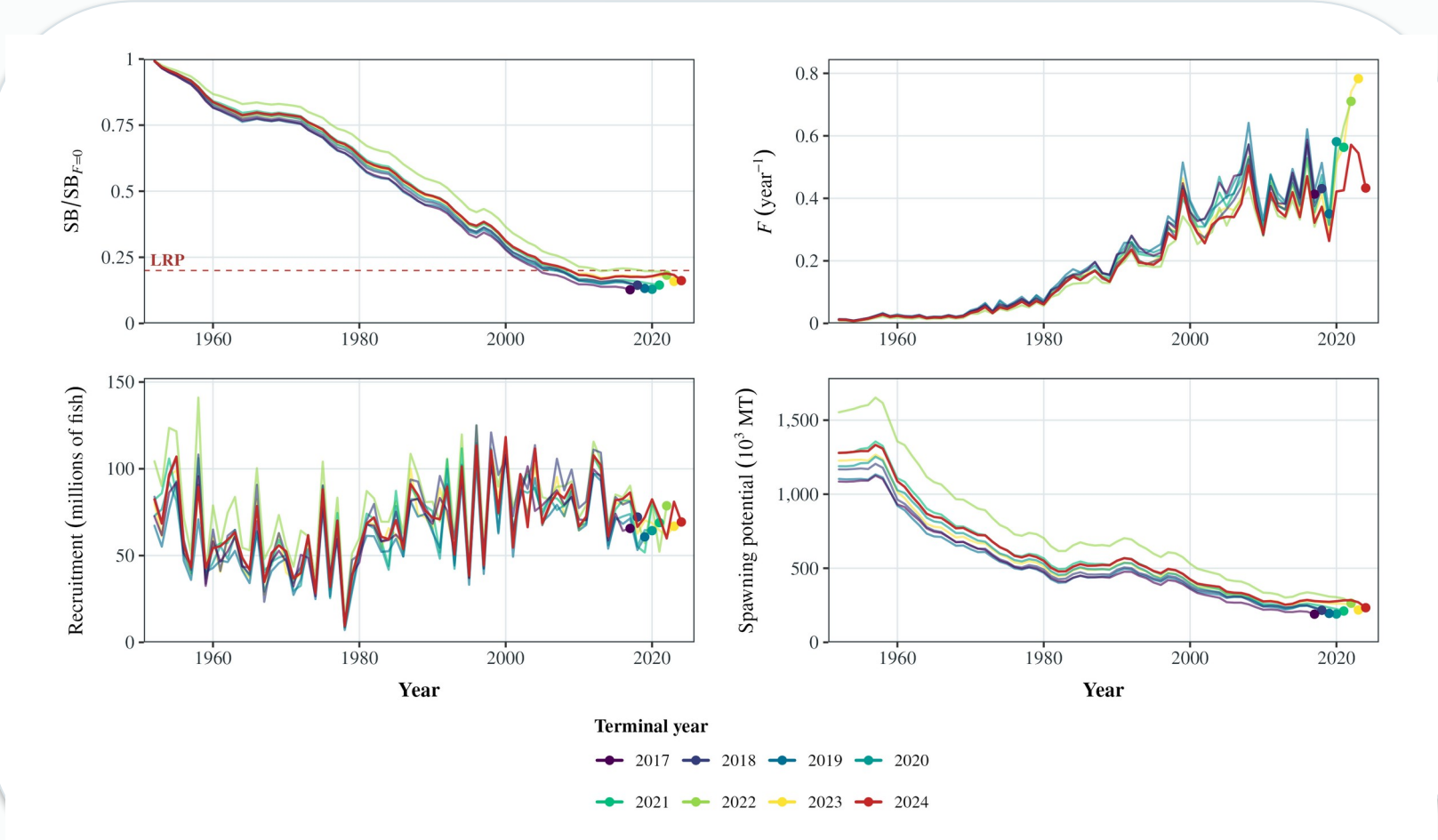
The total tag likelihood favours a larger population

PTTP	strongest preference for a larger population
RTTP	weaker high-scale preference
JPTP	favours a lower scale

Compare curve shapes: each component is reset at its own best fit. This profiles population size—not τ—and is not an uncertainty interval.

Source: WCPFC-SC22-2026-SA-WP-06, Section 7.2.5 and Figure 37.

Final Diagnostic retrospective: seven peels



MOHN'S ρ

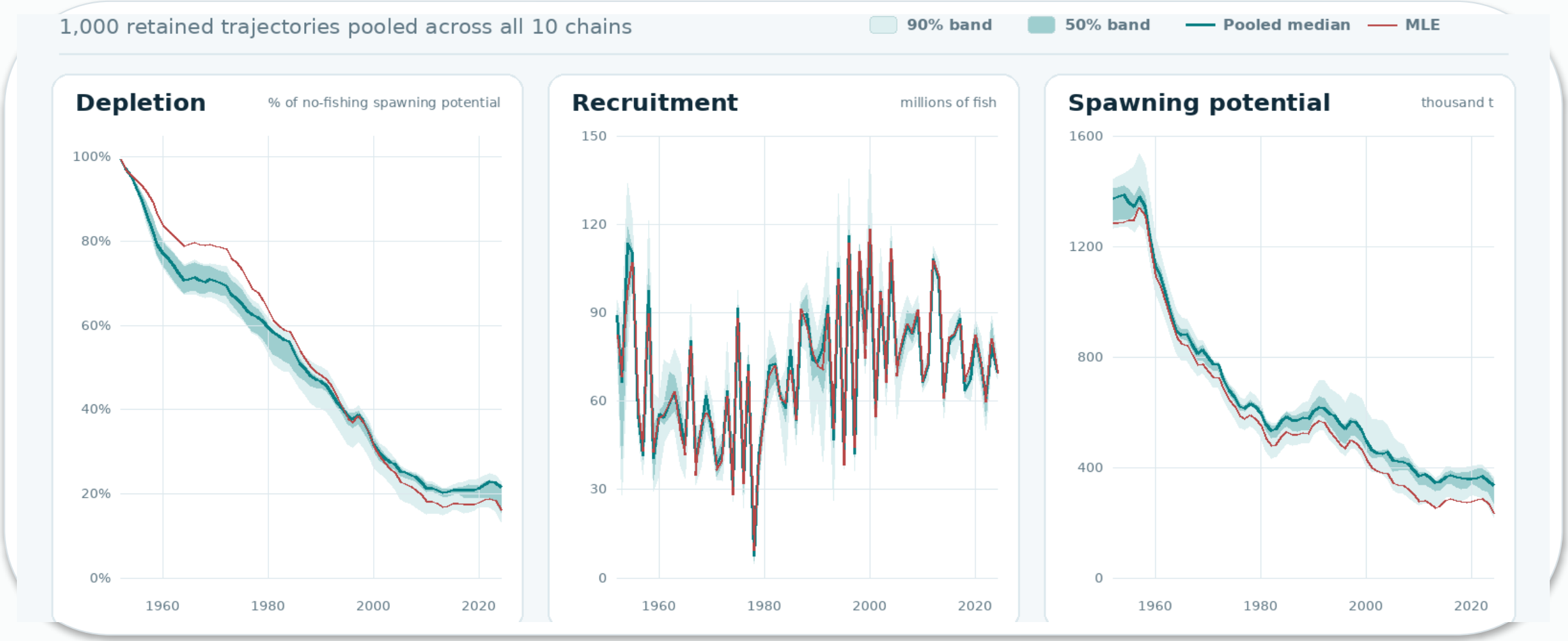
Depletion	**−0.195**
Spawning potential	**−0.235**
Recruitment	**−0.062**
Annual F	**+0.308**

Broad history is retained; endpoint bias is clearer for depletion, spawning potential and F.

Retrospective consistency is not accuracy. This is one Diagnostic model—not the 80-model ensemble.

Source: WCPFC-SC22-2026-SA-WP-06, Section 7.2.4 and Figure 32.

MCMC uncertainty compared with the MLE



Poor chain mixing means these are not reliable posterior estimates. It may reflect strong parameter correlation, alternative/local modes and a difficult likelihood surface.

Source: BET 2026 mfclrtmb MCMC feasibility run; mfclrtmb 0.3.3 and SparseNUTS 1.0.2. Annual aggregation follows the Diagnostic figure definitions; ten chains pooled; all started at the 2026 Diagnostic MLE. Experimental result—not used for assessment advice.

Two legends. One strong line of continuity.

John Hampton

Nick Davies

WHAT STAYED CONSISTENT

The assessment changed. Their advice, review and decades of experience carry the science forward.

Photos supplied by the presenter. Recognition reflects their continuing contribution to WCPO tuna assessment science and software.

mfclrtmb: keep the contract, change the engine

01

Keep the MFCL contract

Same inputs and outputs → transform, compare and reproduce.

02

Use RTMB inference

AD · random effects/Laplace · SparseNUTS / tmbstan MCMC.

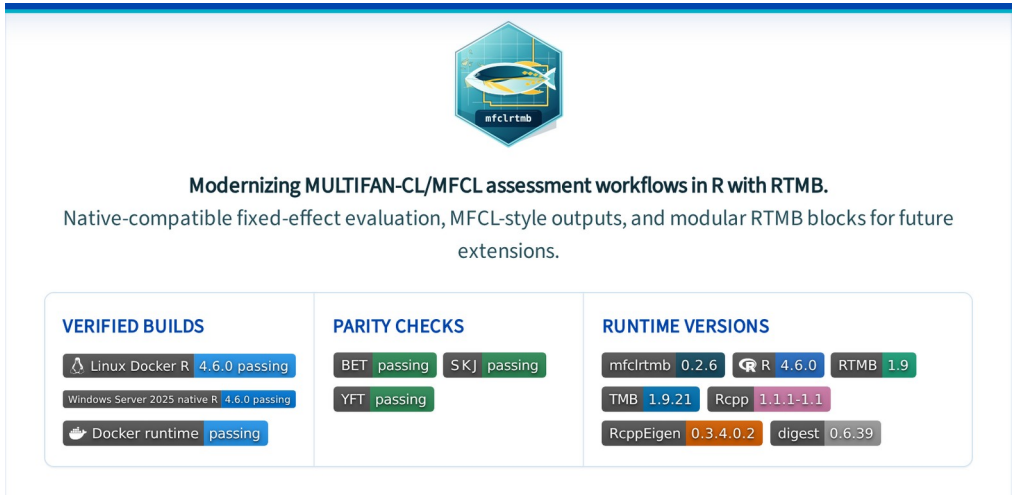
03

Transformable by design

R-native modules can simplify, replace or extend the model—even beyond MFCL.

CONTINUOUS INTEGRATION

MFCL parity checks guard consistency · BET 2026 / SKJ / YFT



pacificcommunity.github.io/ofp-sam-mfclrtmb

SC22 information paper · SA-IP02 · Section 4.2

Current developer: Kyuhan Kim · MFCL parity checks: Nick Davies · improved through John Hampton's feedback

TRANSITION BRIDGE

promising path · not a blanket replacement yet

Sources: WCPFC-SC22-2026-SA-IP02, Section 4.2 (p. 16); mfclrtmb documentation and development repository (accessed 15 August 2026). Website image captured 15 August 2026.

Resources for independent checking

2026 assessment
VIEW CODE

Full report, figures and tables

diagnostics
VIEW CODE

Fits, residuals and diagnostics

80-model viewer
VIEW CODE

Filter inputs and inspect status

Jitter results
VIEW CODE

Alternative starts and stability

Stepwise viewer
VIEW CODE

Trace model development

Self-test results
VIEW CODE

Recovery under configured model

One-off sensitivity viewer
VIEW CODE

Biological and observation-process tests

Retrospective results
VIEW CODE

Peels and terminal-year stability

Profiles
VIEW CODE

Inspect data conflict on scale

2023 assessment
VIEW

Source for assessment comparison

All links are public SPC/WCPFC assessment products or source comparisons referenced by WCPFC-SC22-2026-SA-WP-06.