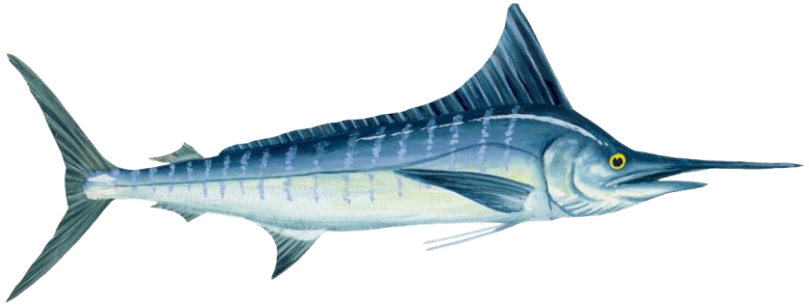
Revised 2024 Stock Assessment of Striped Marlin in the Southwest Pacific Ocean Part 1: Integrated Assessment in Stock Synthesis

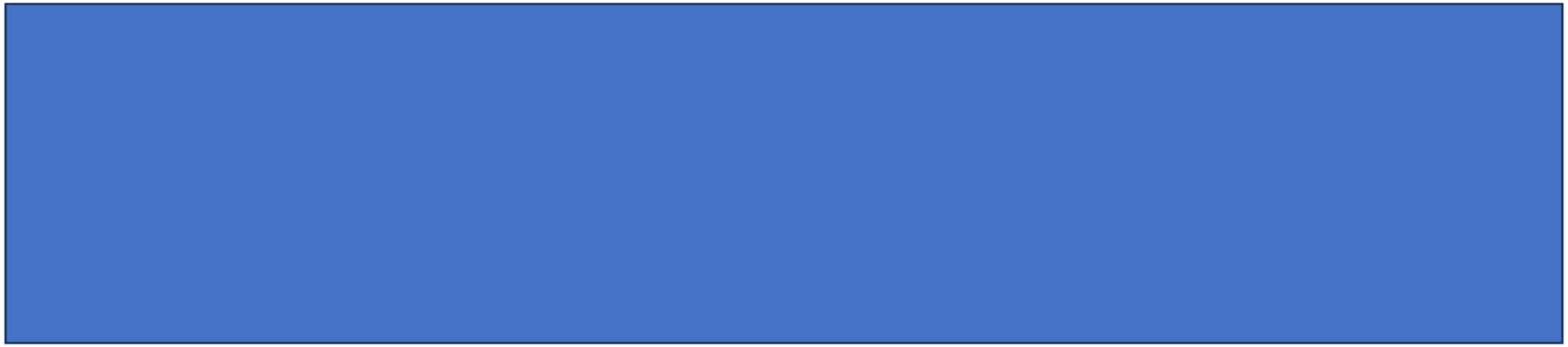


WCPFC-SC21-2025/SA-WP-06
Nuku'alofa, Tonga
13–21 August 2025

C. Castillo-Jordan, N. Ducharme-Barth, J. Day, N. Davies, T. Peatman, K. Kim, P. Hamer

To be continued? …

Recommendations for further work



- Review of the comp.data (length and weight)
- Consider the issue of effective hook effort changes (reductions)
- Priors investigation for ensemble models
- Quarterly time step model to improve resolution
- Use of alternative model platforms

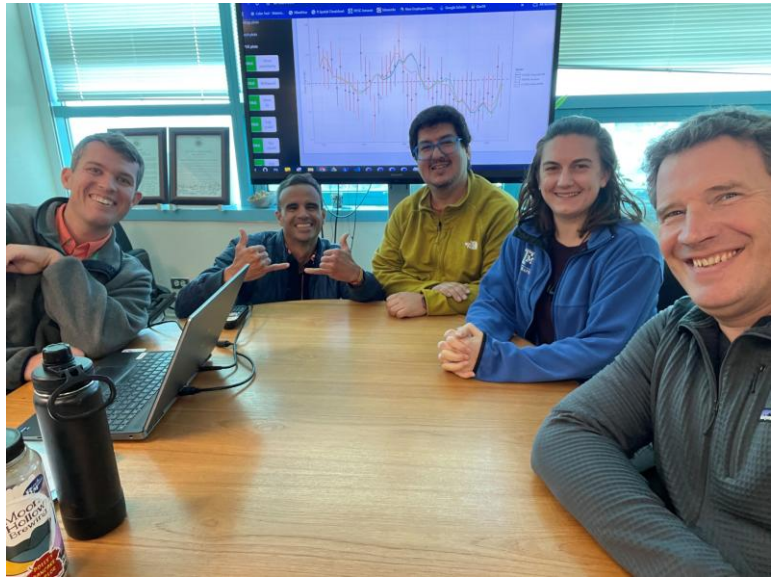
Data Conflicts & Model Limitations

- Conflicts between length and weight frequency data in some fisheries
- Poor fit to CPUE data in recent years
- Fixed initial conditions (1979 start) limiting uncertainty estimation and sensitivity analyses
- Potential overestimation of depletion due to strong scaling effect from size data
- Biomass trend appears more pessimistic than CPUE trends indicate

US Support Initiative

- US delegation offer accepted by SPC and supported by SC20
- In-person workshop held in Hawai‘i (January 2025)
- Collaboration continued through SC21
- Workshop outcomes:

  https://n-ducharmebarth-noaa.github.io/2025-swpomls-meeting/



Assessment Platform Transition
Multifan-CL (MFCL) to Stock Synthesis (SS3)

- Technical Rationale:
  - Annual MFCL model struggled with selectivity resolution given faster early growth
  - Quarterly MFCL improved selectivity but estimated inconsistent quarterly recruitment
  - Size-based selectivity option not available in MFCL at workshop time (now available)
- Platform Decision:
  - Transitioned to SS3 for size-based selectivity capabilities
  - Maintains annual recruitment consistent with species biology
  - Supported by 2025 SPC pre-assessment workshop as efficient pathway for this collaboration

Broader Implications:

SC20 requested 2025 swordfish assessment shift to SS3

2029 striped marlin assessment shift to SS3

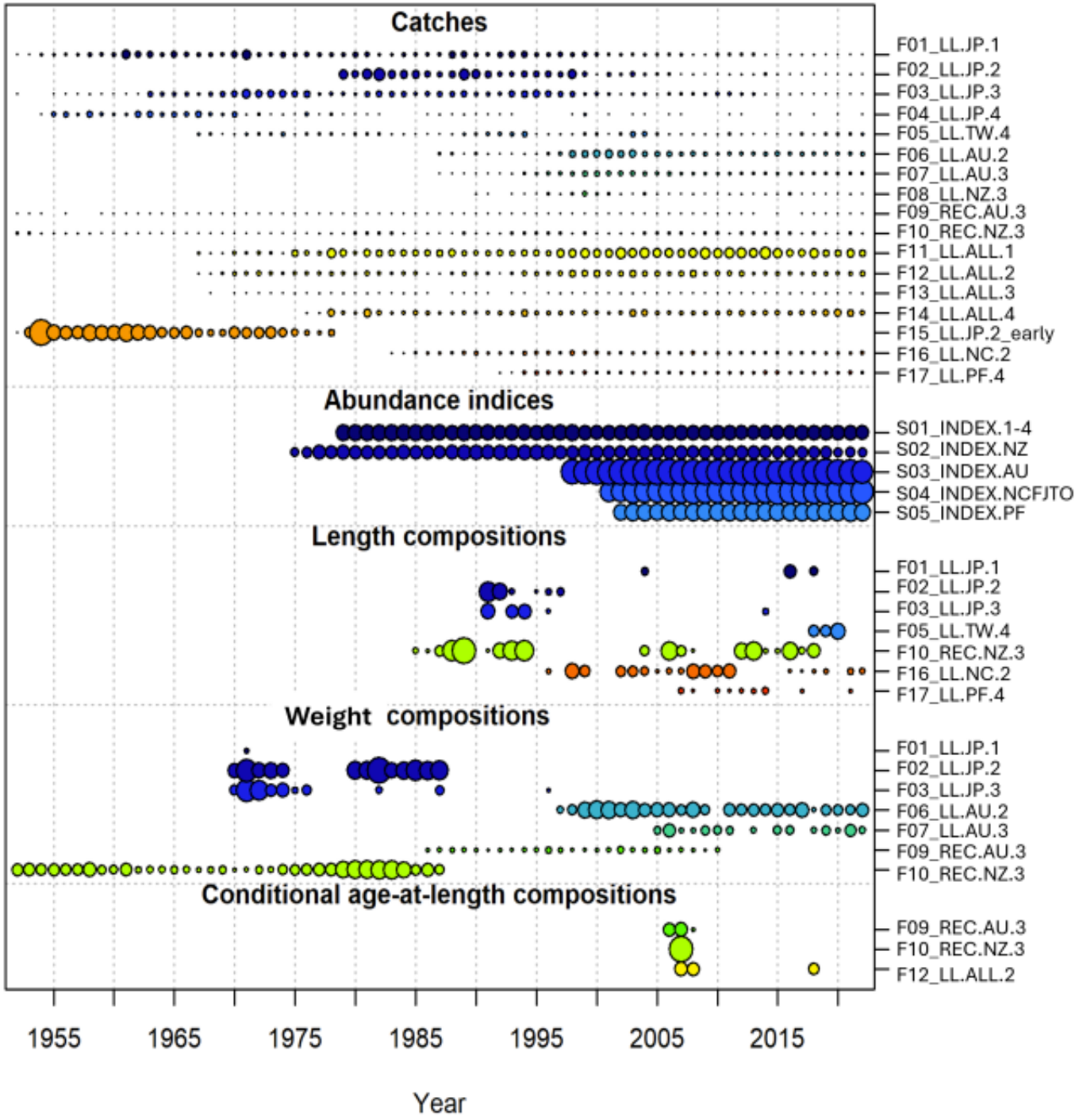
Assessment Platform Transition
Multifan-CL (MFCL) to Stock Synthesis (SS3)

- Integrated SS3 model framework
- Complementary to Bayesian Surplus Production Model (Ducharme-Barth, 2025) – note this came later when the SS3 model failed to resolve concerns around population scale plausibility
- Combined results available for information for SC21 management advice

Key Model Modifications and Improvements

- **Platform & Structure:**
  - Transitioned from MFCL to SS3
  - Extended model start year to 1952 with early catch uncertainty incorporated
- **Selectivity & Data Handling:**
  - Implemented size-based selectivity with modified selectivity functions
  - Major revision and refined filtering of size data inputs (significant data reduction)
  - Separated NC and PF longline fleets from mixed flag fisheries due to different length distributions
  - Applied ad hoc size data weighting to reduce influence of size data

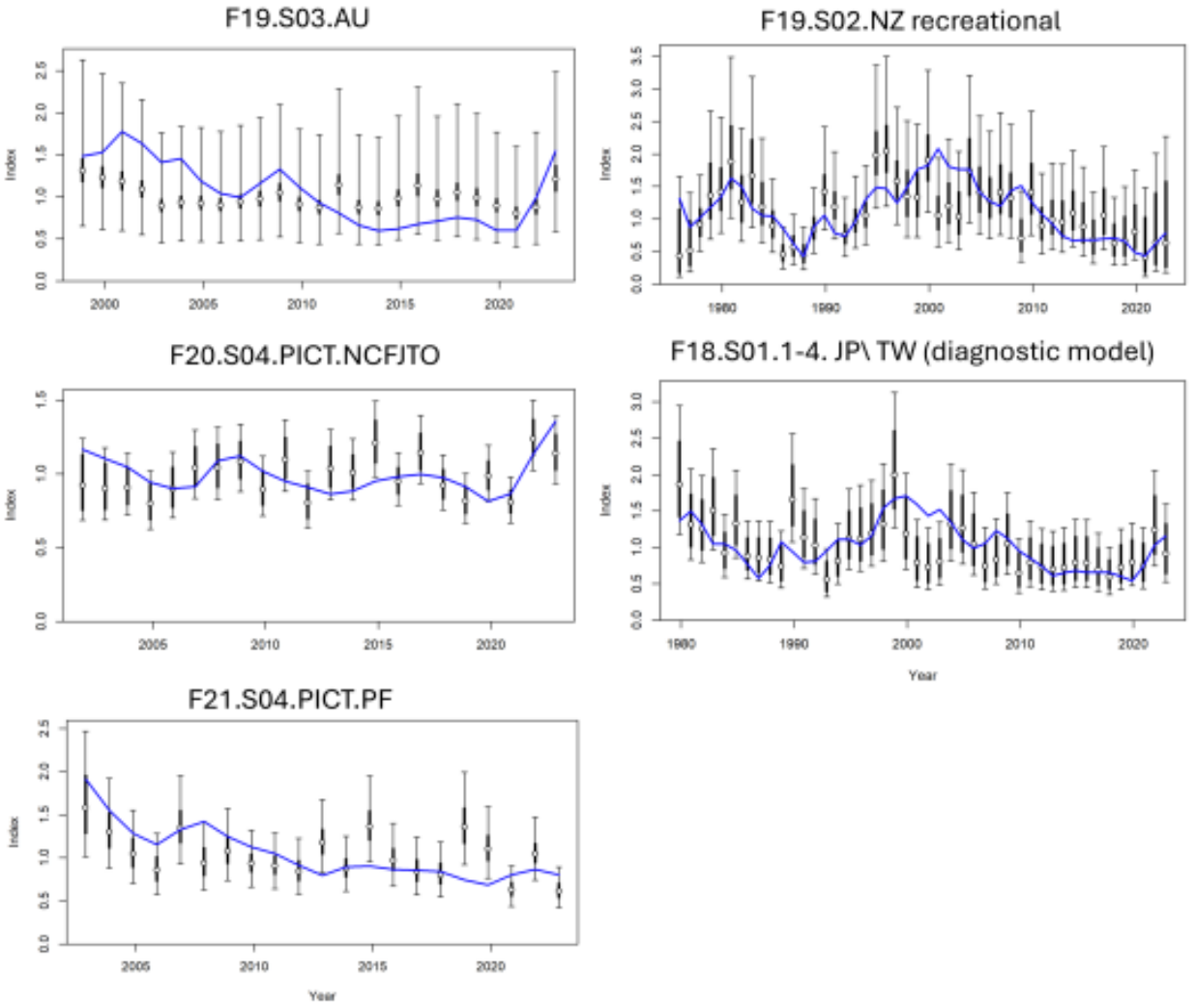
Data



Similar fleet structure 2024
Early JP fleet
NewCal and French P.
Obs CPUE for sensitivities

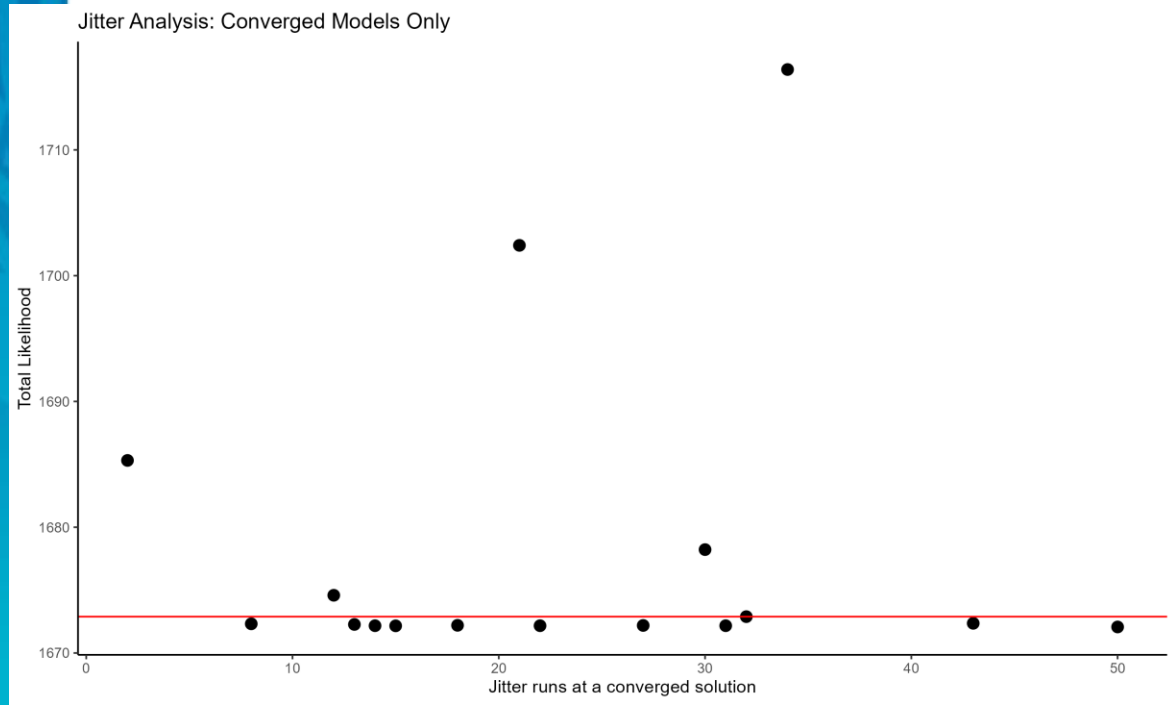
Key Model Modifications and Improvements

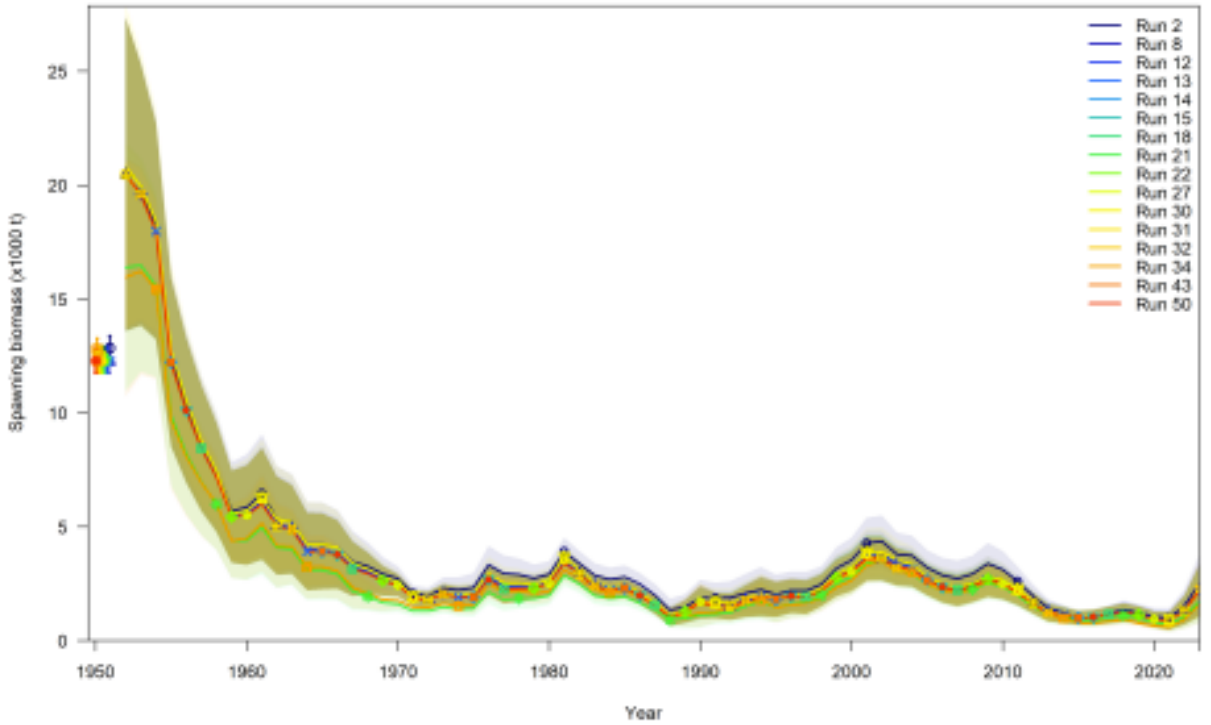
- **Growth & Biological Parameters:**
  - Used CAAL (Catch-at-Age-at-Length) data for internal growth estimation in diagnostic model
- **CPUE Indices Enhancement:**
  - Explored sensitivity to alternative CPUE indices:
    - Australian longline
    - New Zealand recreational
    - Newly developed observer-based longline indices for PICT fleets (Neubauer, 2025, SC21-SA-IP-13)



Results Summary
Diagnostic Model Performance

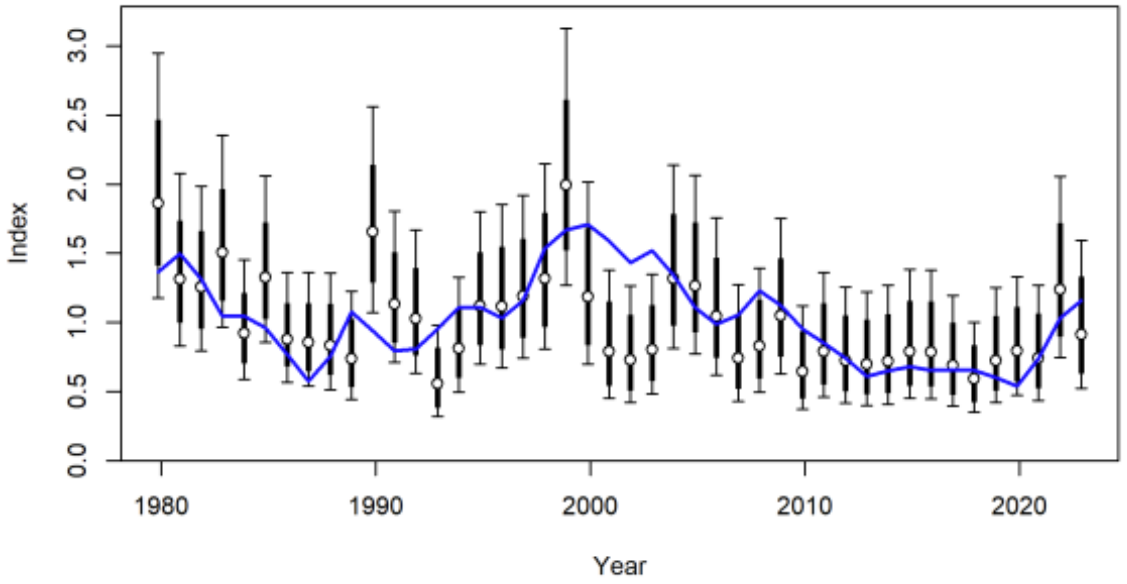
- **Convergence & Model Stability:**
  - Achieved good convergence (max gradient: 4.02×10^{-6}, positive definite Hessian)
  - Only softbound issues with 1952-53 F initialization parameters (not management-relevant)
  - Jitter analysis (50 replicates): 16 converged models found similar likelihood values
  - Model appears close to global minimum, though minimum may not be well-defined

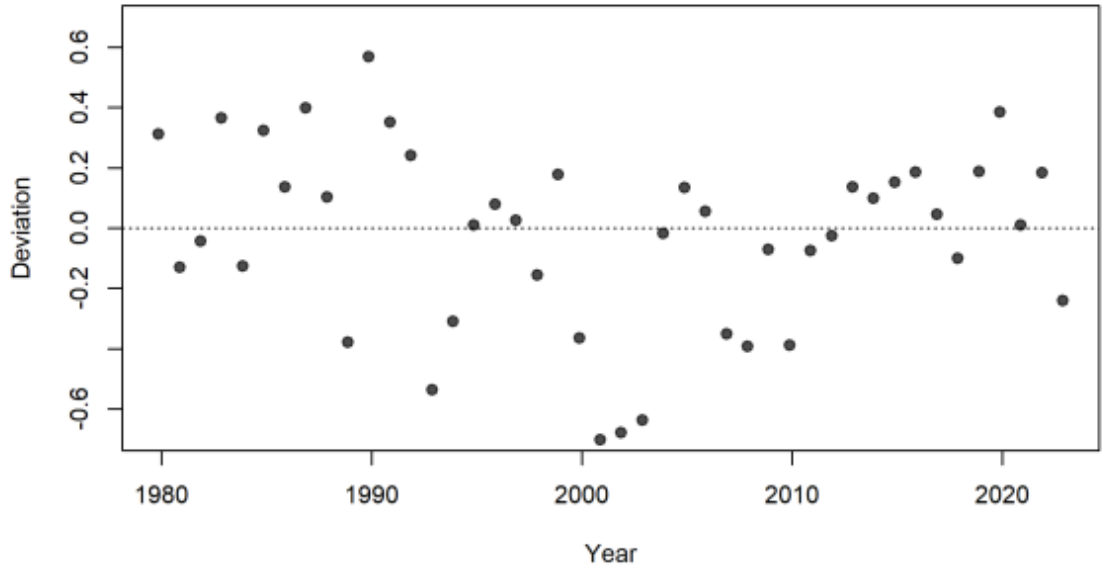




Data Fits - Major Improvements:

- **CPUE:**
  - Substantial improvement over previous 2024 model; captures major trends and recent patterns well, related to data weighting changes and revisions of size data.





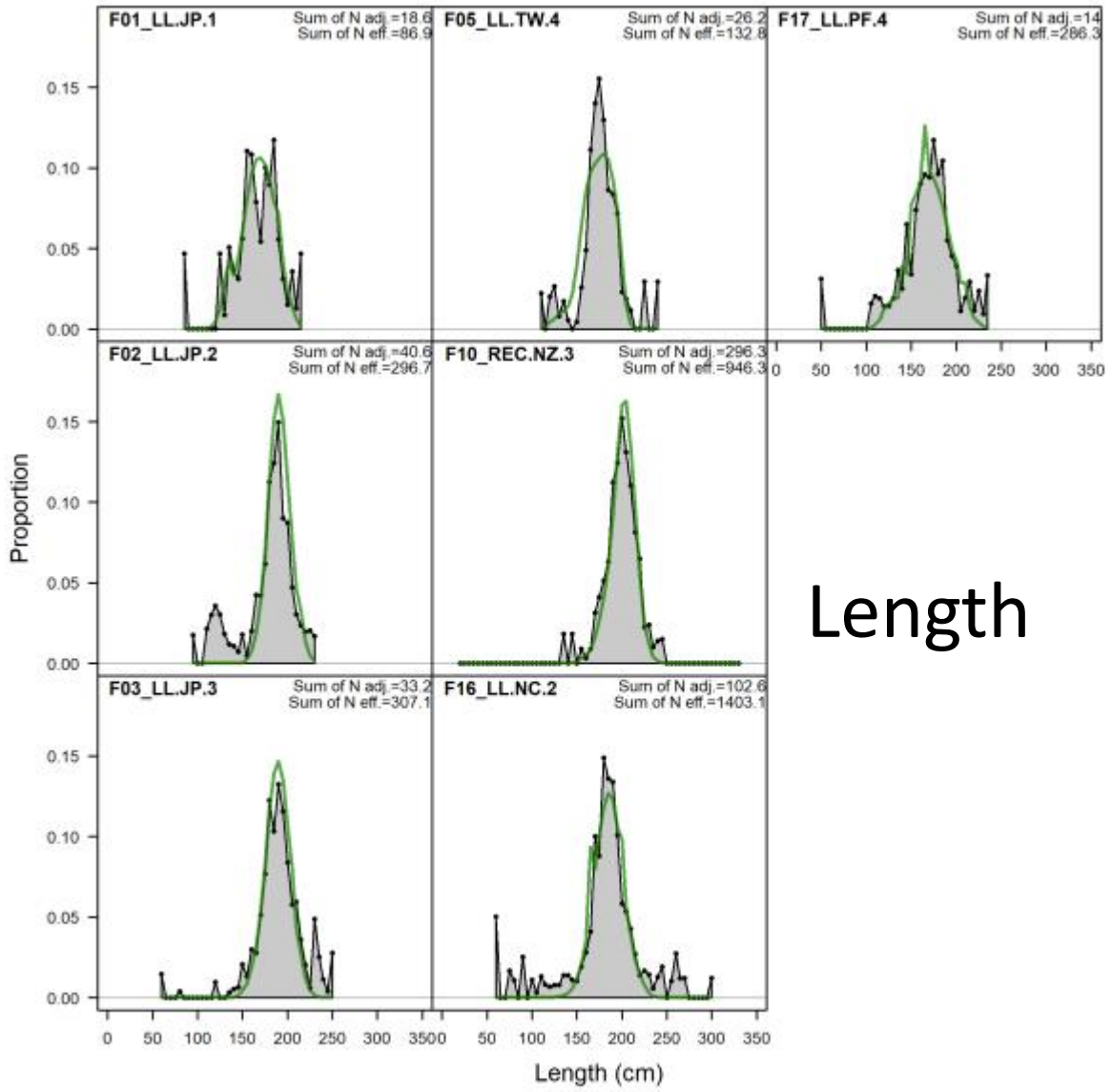
Size Data:

'Length' data remains noisy with spikiness, but main distribution components fit reasonably

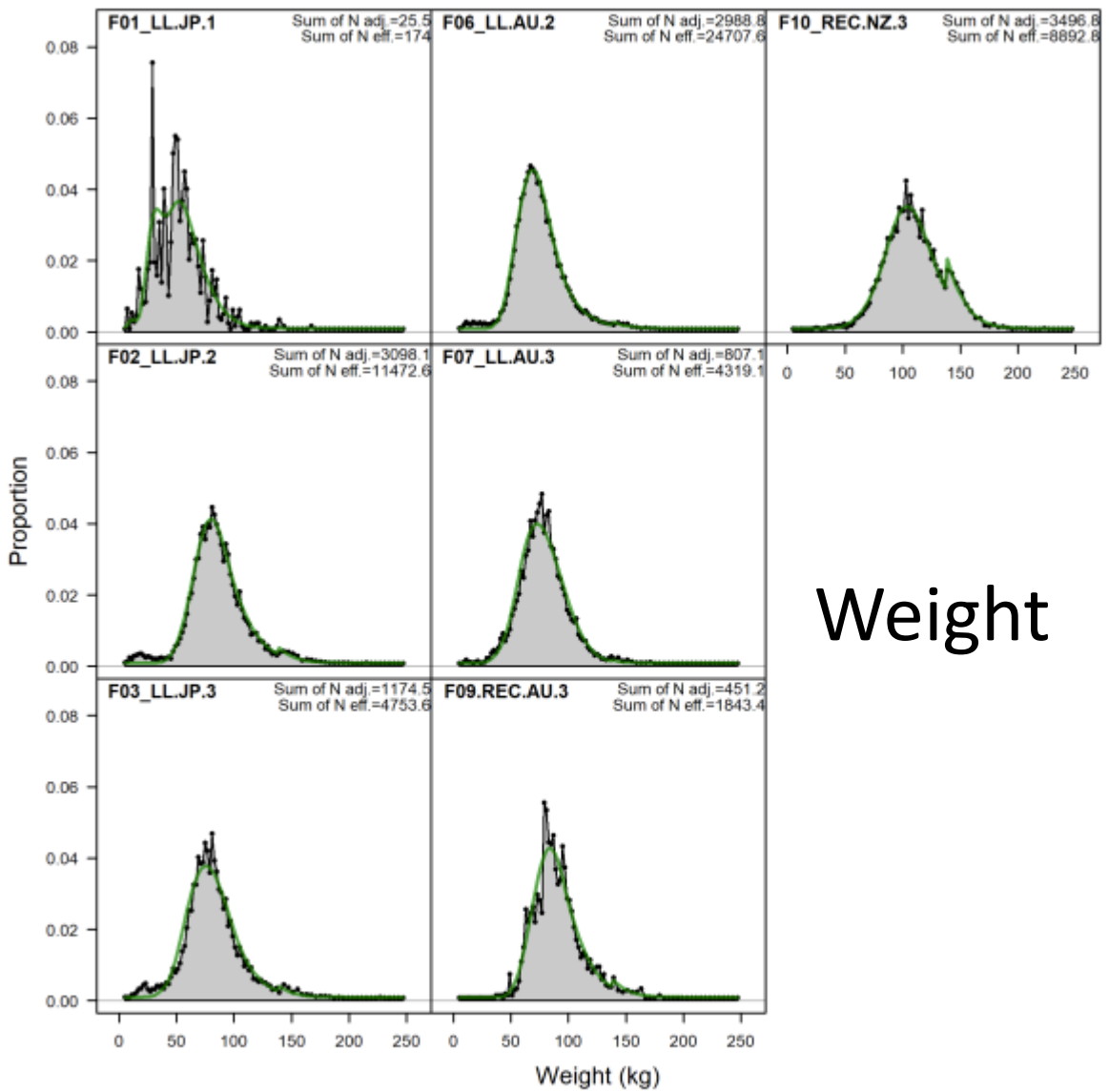
'Weight' data shows excellent aggregate fits (smoother distributions, larger sample sizes)

Consistent underestimation of smaller fish frequencies across multiple fisheries

Aggregate fits



Length



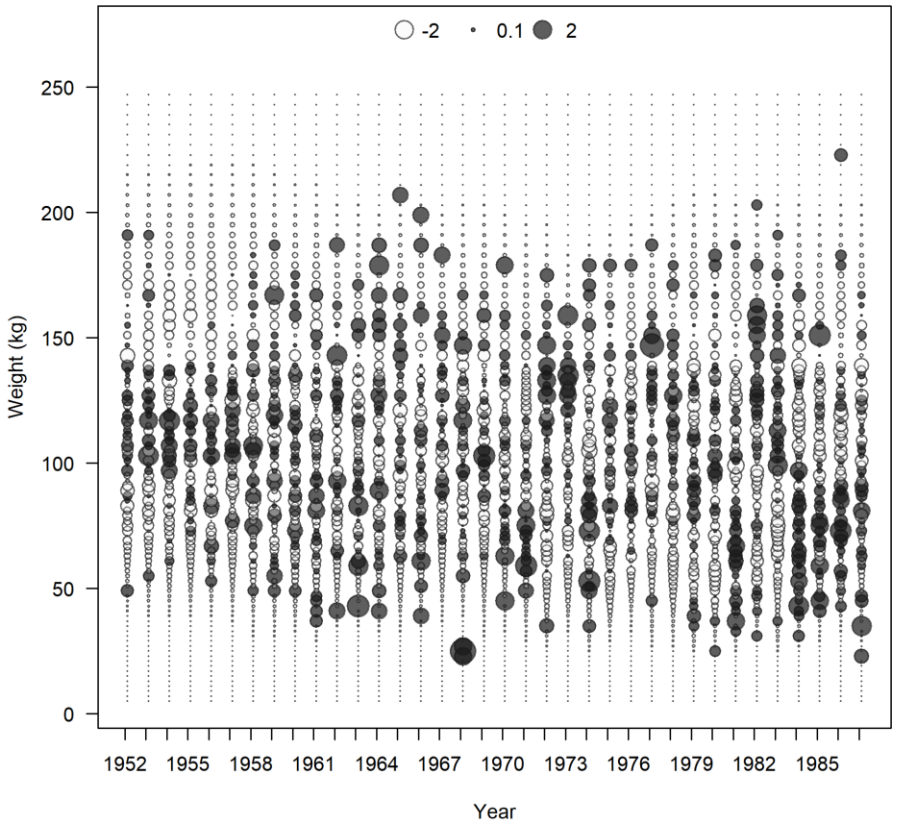
Weight

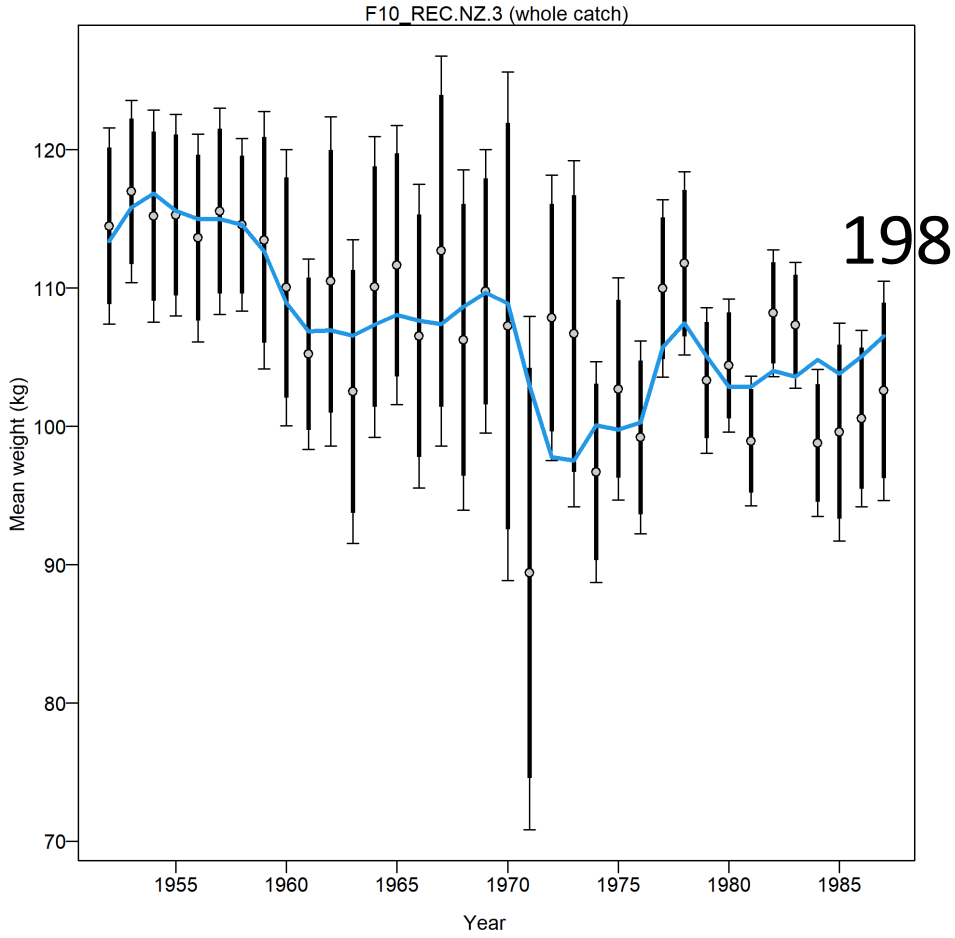
Figure 16: Composite (all time periods combined) observed (grey) and predicted (green line) length frequency for fisheries with length frequency data.

Figure 18: Composite (all time periods combined) observed (grey) and predicted (green line) weight frequency for fisheries with weight frequency data.

Time series size data fits

F10_REC_NZ_3
WF

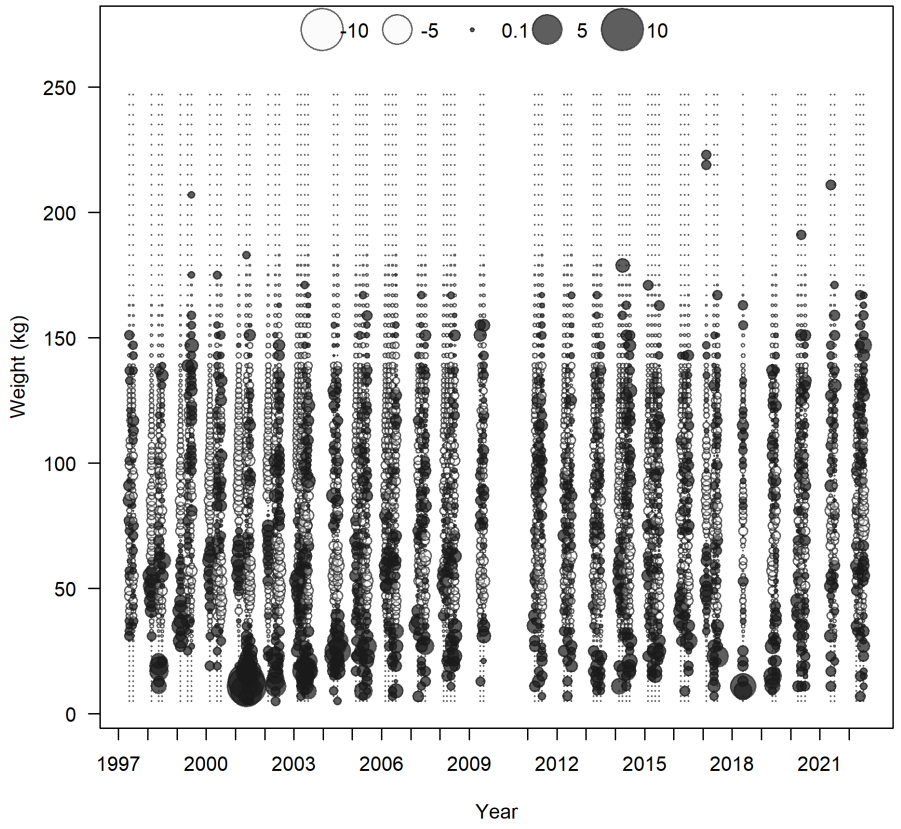


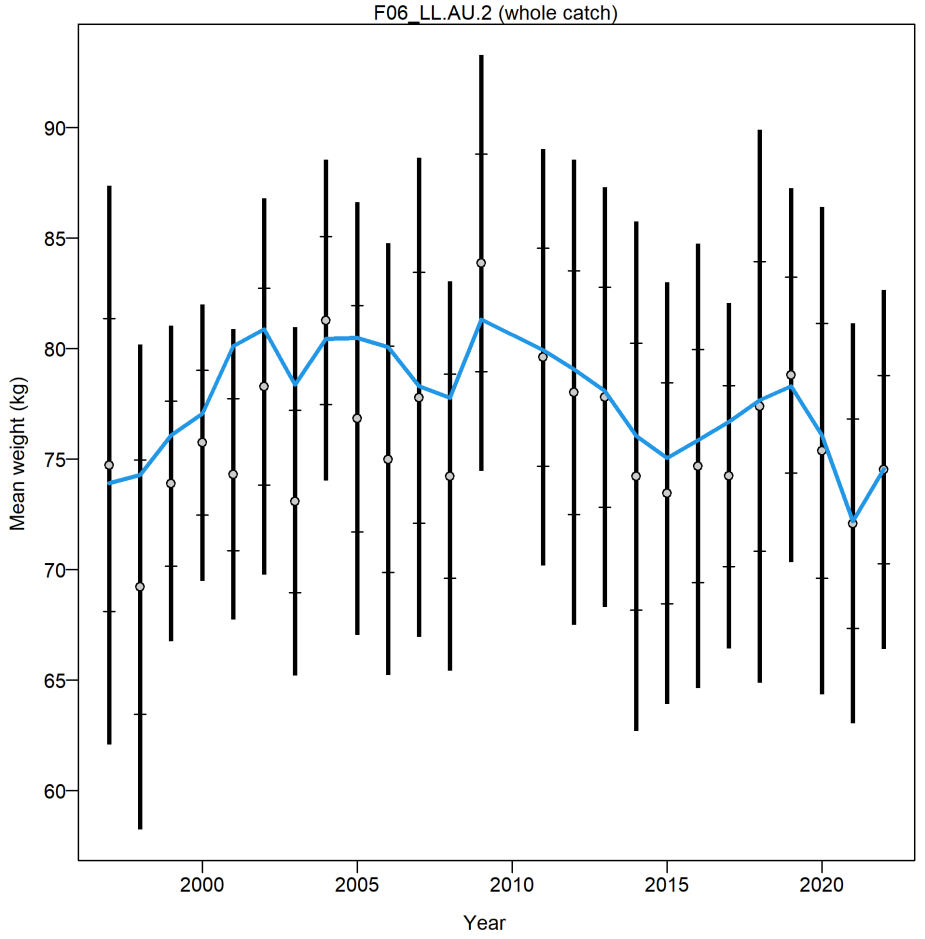


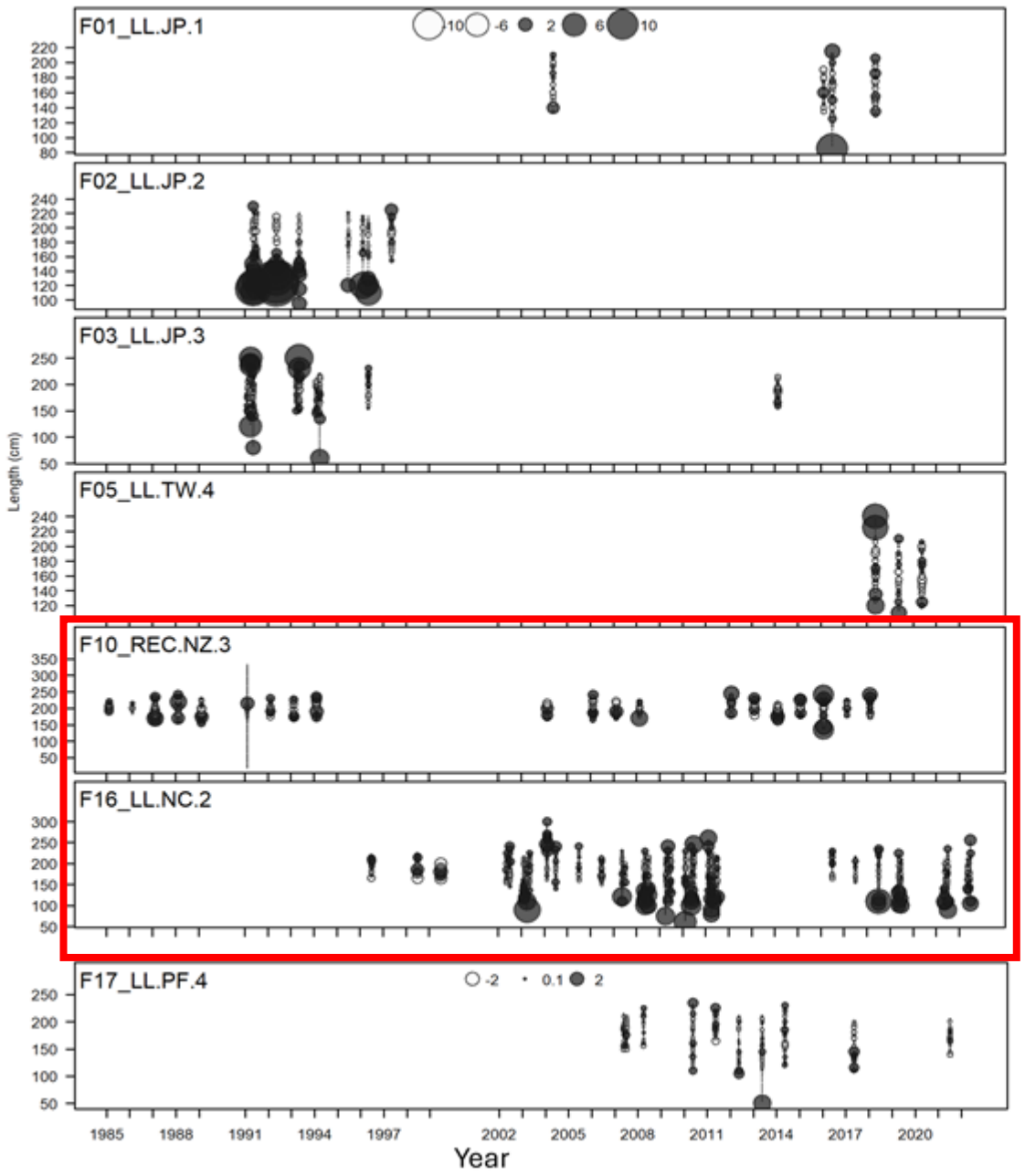
1988

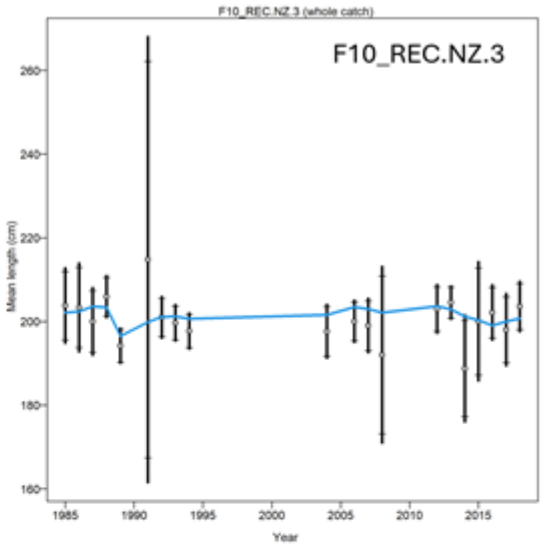
Time series size data fits

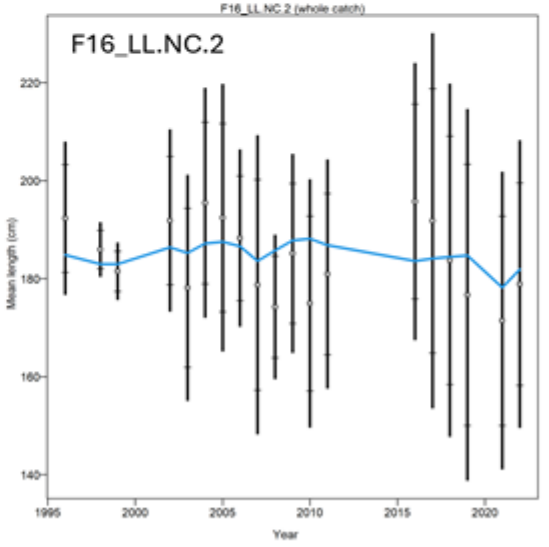
F06_LL_AU_2
WF











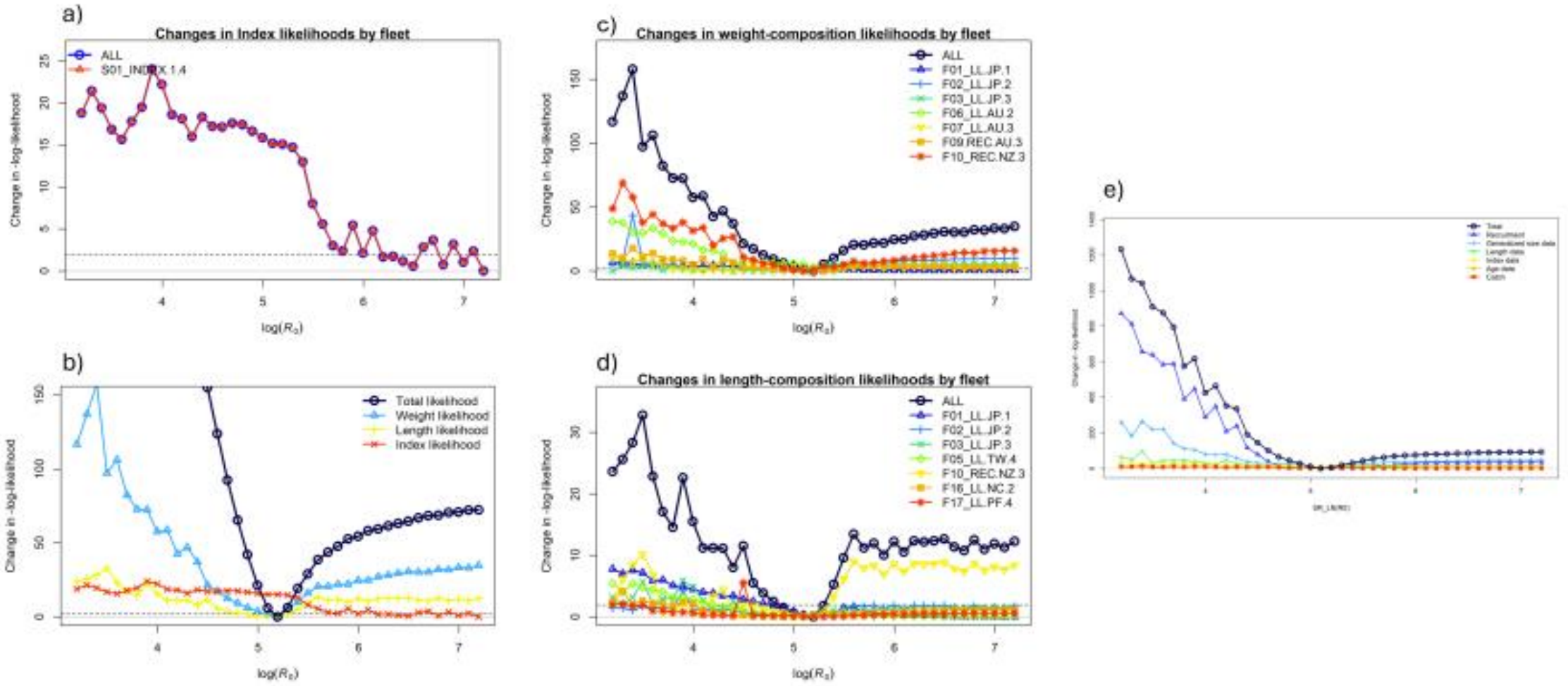
LF data

Increase filtering of the LF data
No much influence in the model
No trends

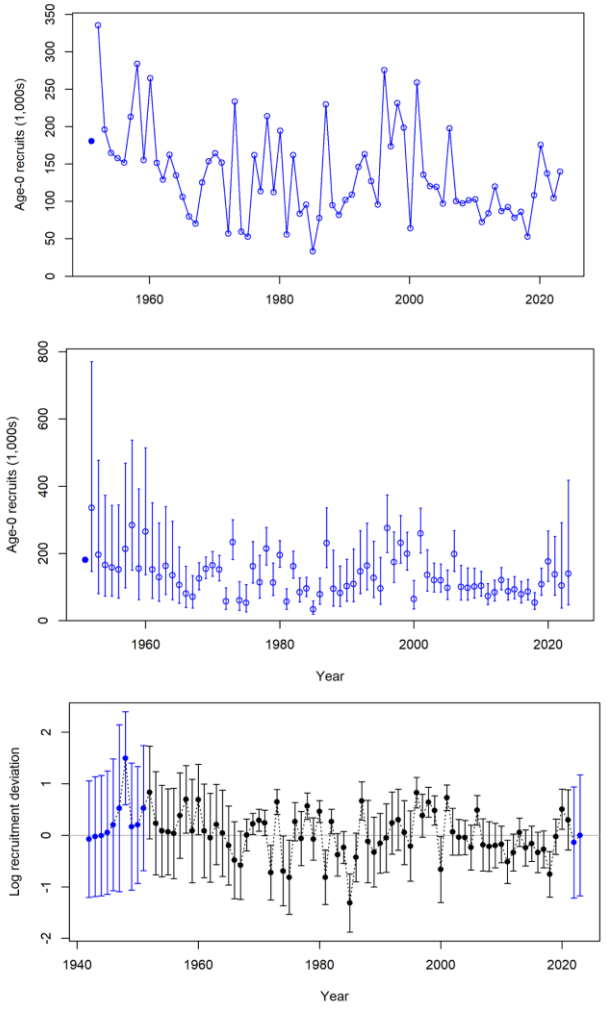
Data Fits - Improvement:

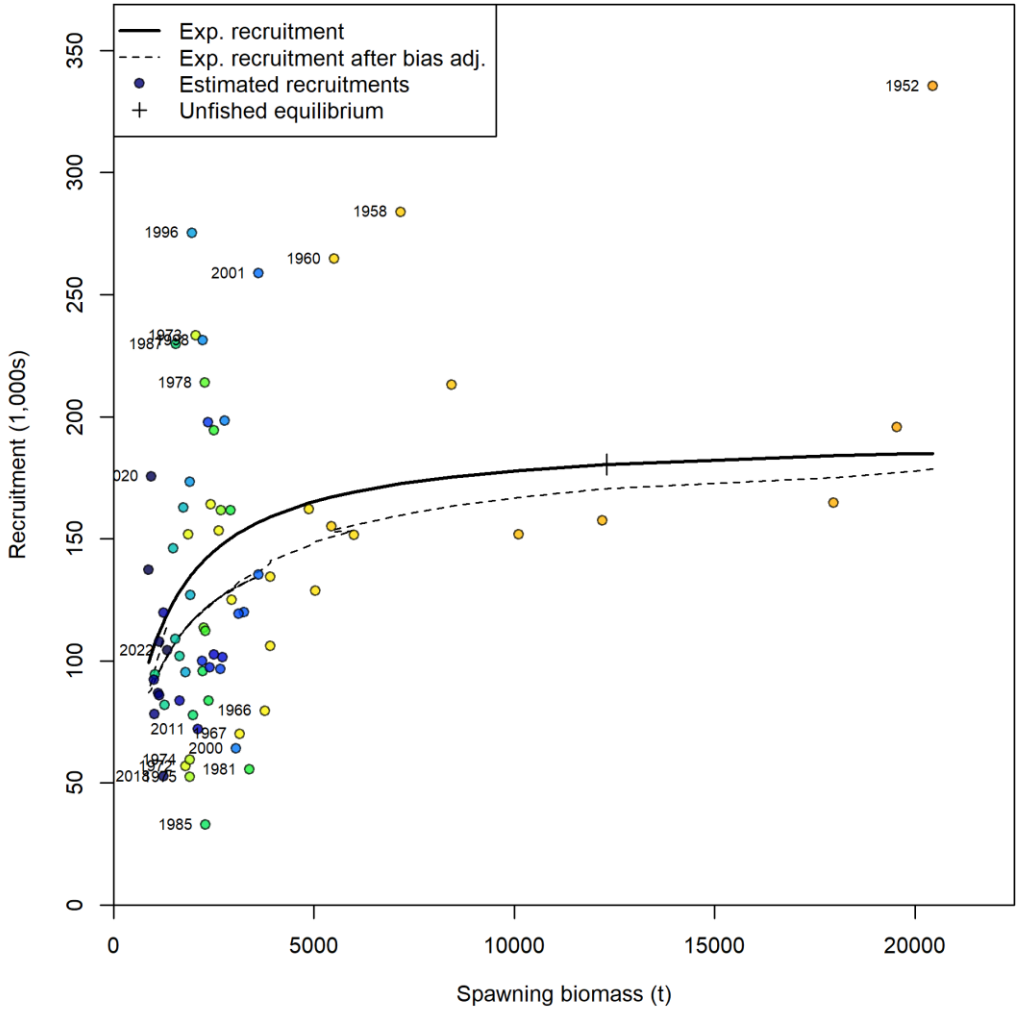
- **Critical Findings - Likelihood Profiles:**
  - Recruitment penalty (SigmaR) dominates other data sources - this is a concern
  - Suggests potential stock productivity mis-specification requiring larger recruitment deviations to fit data
  - Of the data components, the weight data (AU.LL.2, Rec.NZ.3) is most influential on population scale (R0), followed by length data and CPUE (similar)
  - Data conflict moderated but not eliminated – CPUE, although less influential, still prefers higher R0 than size data
  - Length and weight data no longer in conflict with each other (large revision of size data and removal of problematic data)

Likelihood profiles



Recruitment – diagnostic model





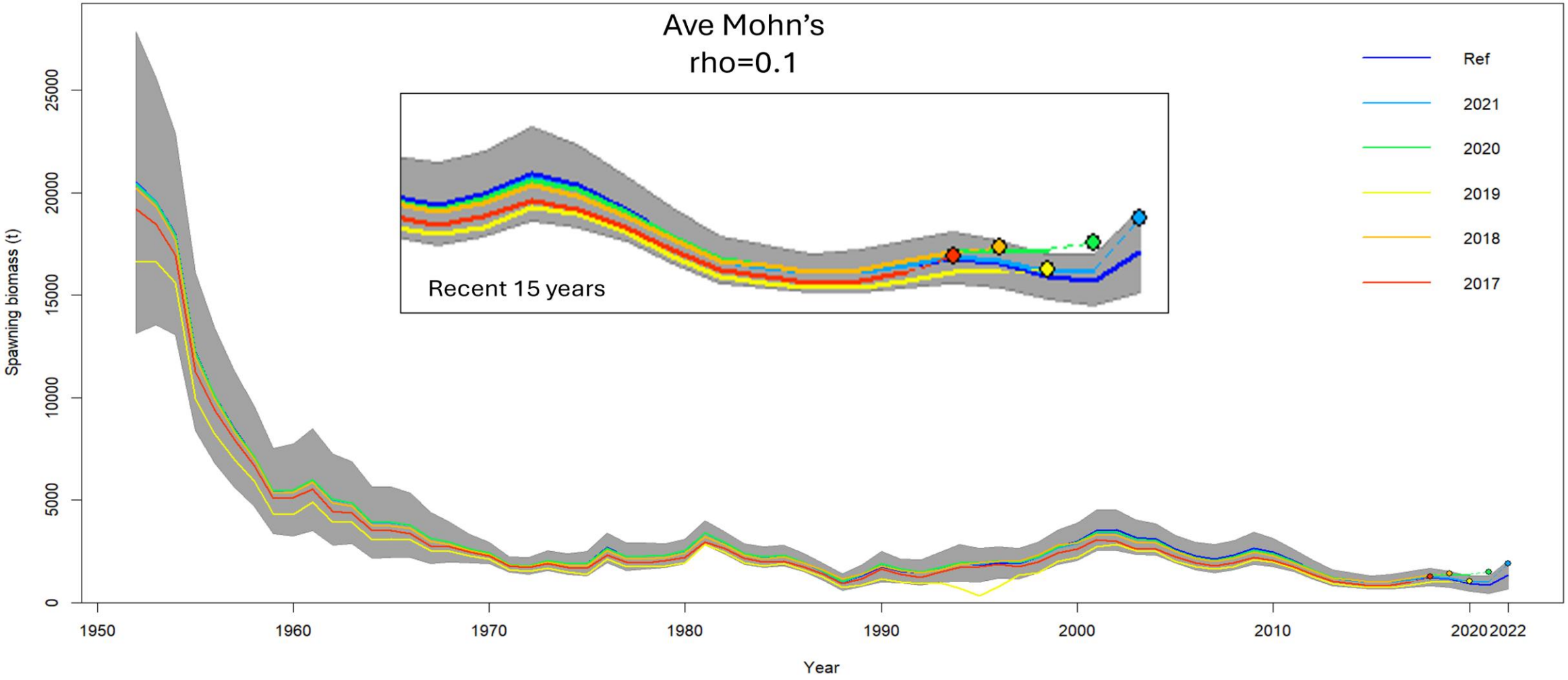
Additional Diagnostic Results

- **Retrospective Analysis:**
  - Spawning Biomass: Positive bias across peels (Mohn's ρ = 0.102) - not significant
  - Suggests potential underestimation of current stock size or overestimation of recent fishing impact
  - All peels within 95% confidence intervals except 2020 peel (ρ = 0.458) - high sensitivity to 2020-2021 data (AU fleet big catches)

- **Fishing Mortality:**
  - Minor negative bias (Mohn's ρ = -0.165)
  - 2020 peel shows strongest effect for both metrics

- **Hindcast Performance:**
  - Good prediction accuracy (MASE = 0.79)

Retrospective analysis (spawning biomass)

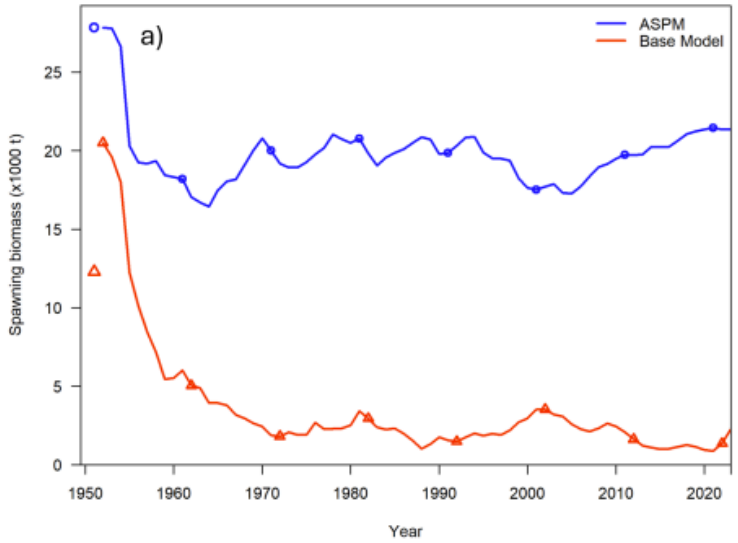


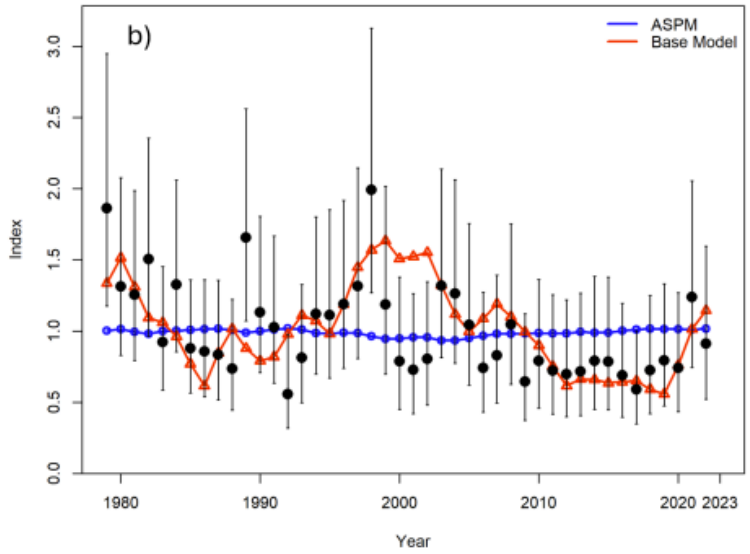
Additional Diagnostic Results

- **Age-Structured Production Model (ASPM) Analysis**
  - ASPM without recruitment deviations could not fit CPUE time series or produce sensible biomass trajectory
  - Suggests recruitment penalty is overwhelming CPUE signal in likelihood
  - When recruitment deviations freed (SigmaR increased to 1.0), ASPM estimated small population scale and fit CPUE
  - Conclusion: CPUE index being overwhelmed by other likelihood components

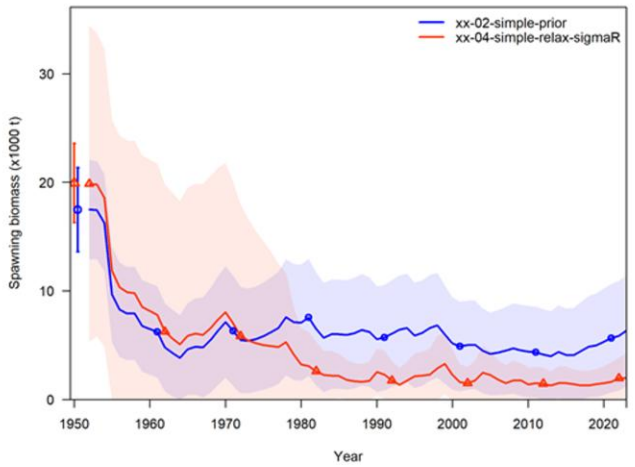
- **Model Specification Issues:**
  - Production function in SS3 model appears poorly determined
  - May indicate model mis-specification and/or over-constrained
  - Cannot produce well-determined production function

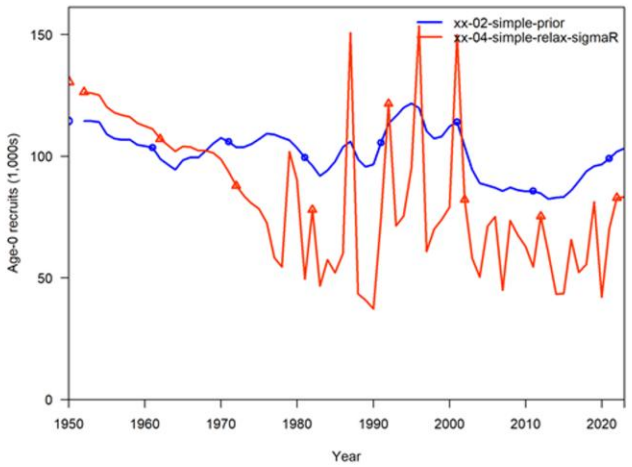
ASPM

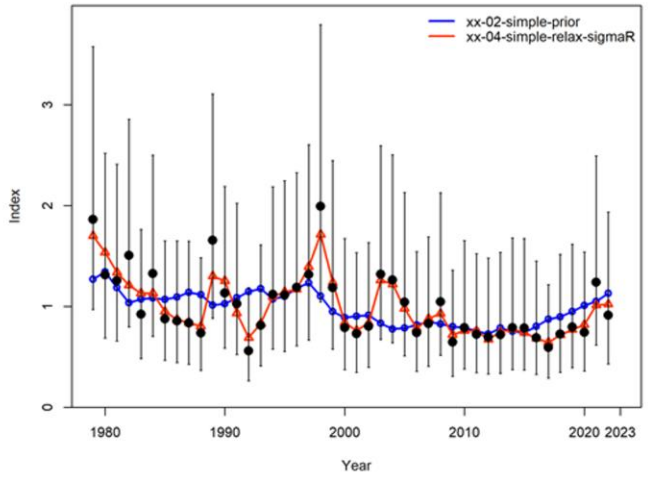




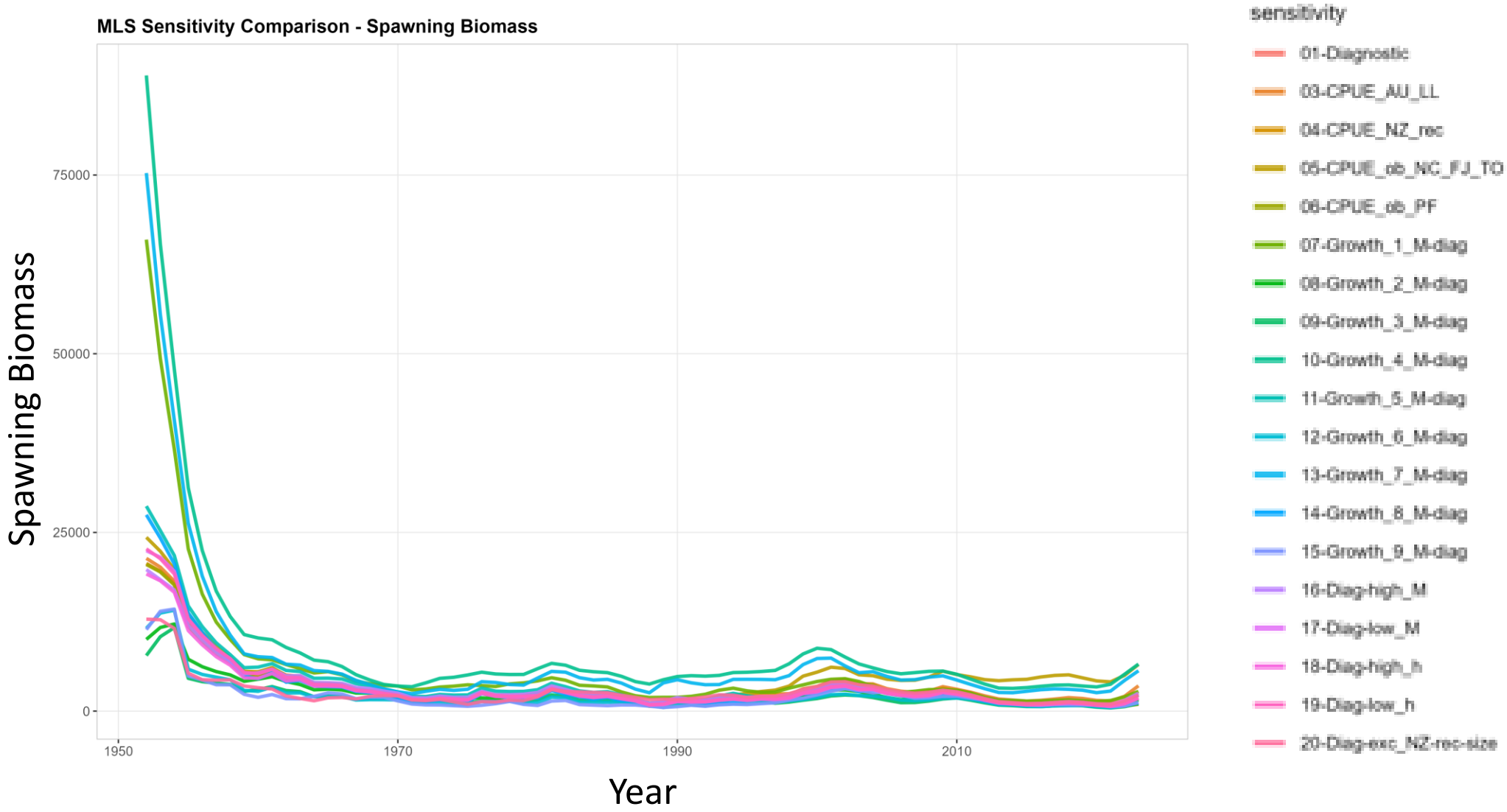
ASPM with recruitment



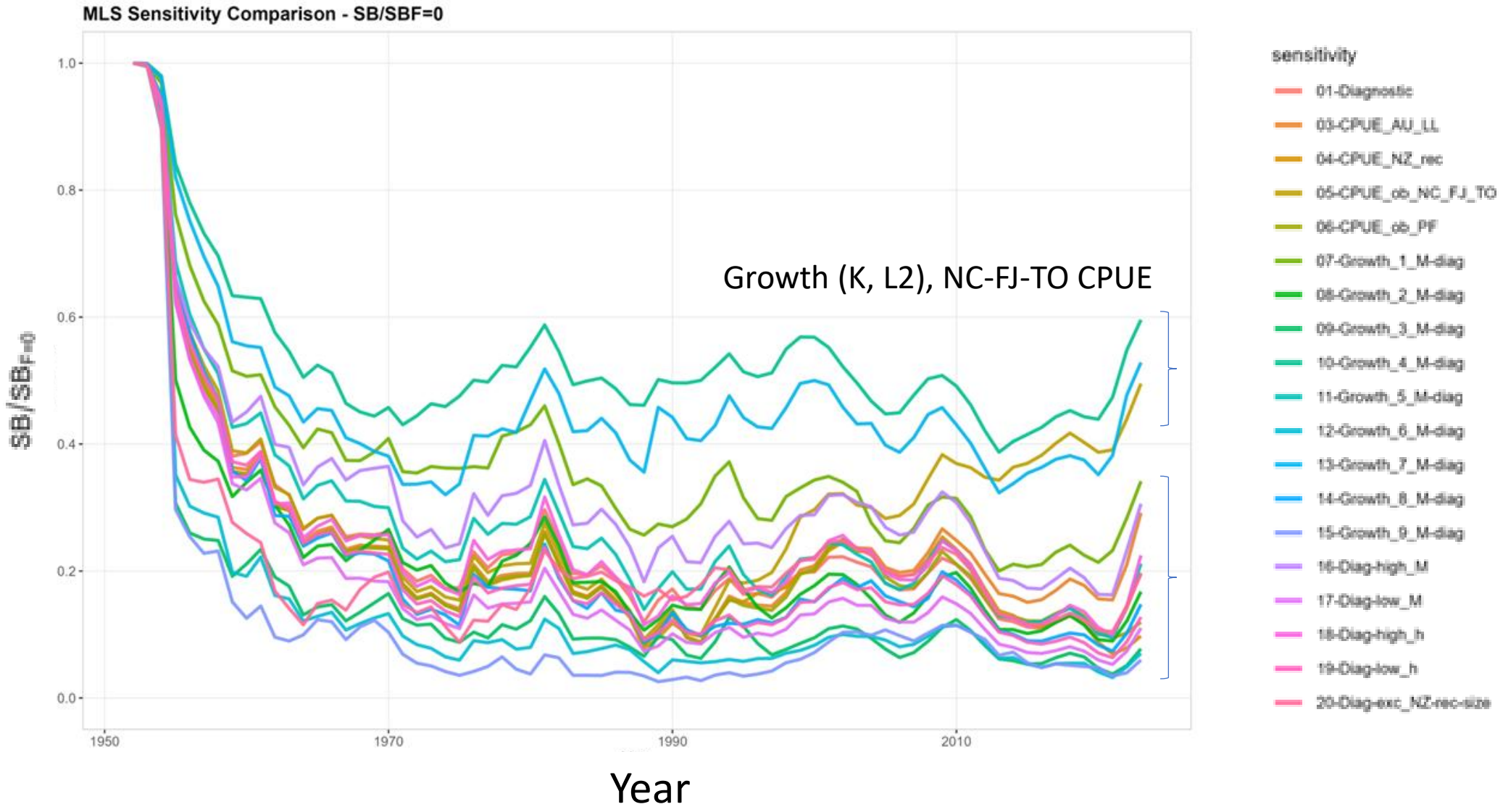




Sensitivity plots – spawning biomass



Sensitivity plots – depletion



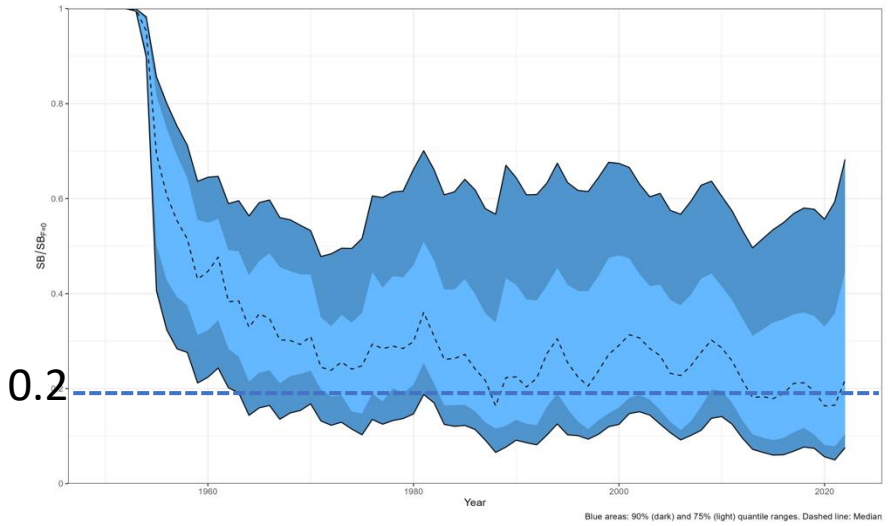
Biology exploratory grid

- Biology – explore sensitivity to alternative biology/productivity parameter – population scale sensitivity

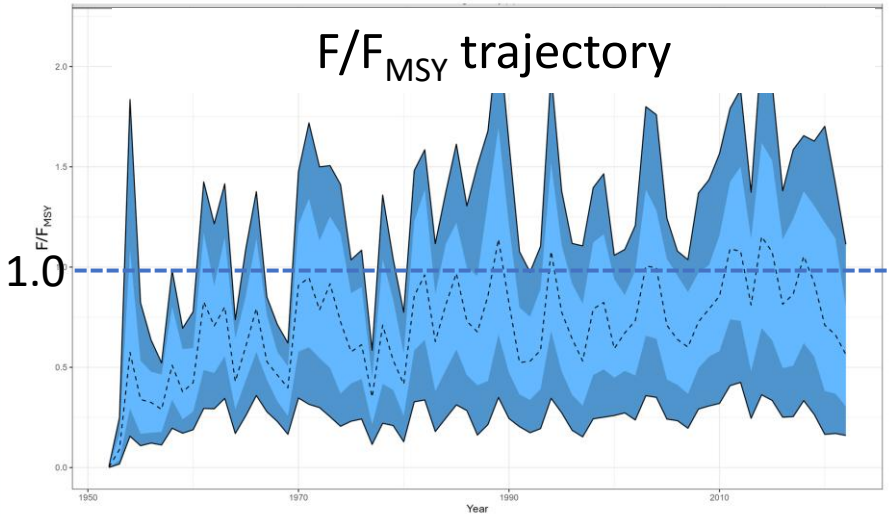
Axis	Option 1	Option 2	Option 3
Growth K	0.2	0.5	0.8
Growth L2	195	210	215
M-at-age	low (diag x 0.65)	diag	high (diag x 1.35)
Steepness	0.75	80	0.95

Results – biology grid x 41 models

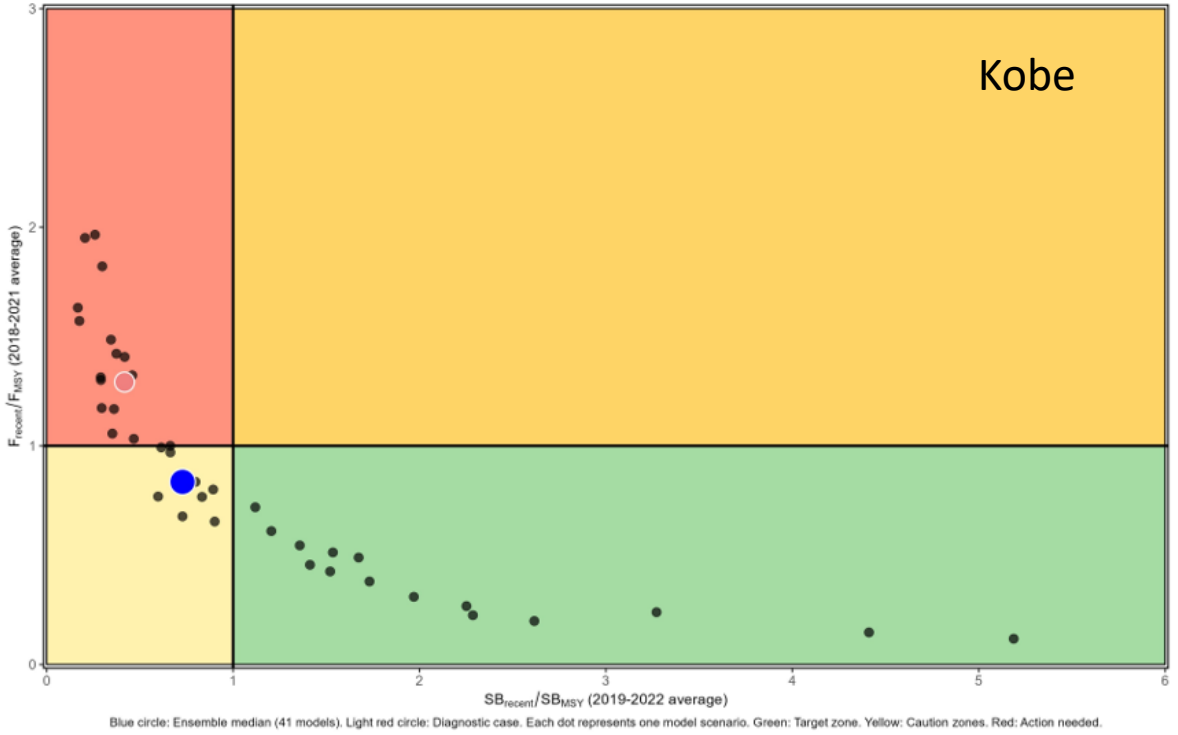
$SB/SB_{F=0}$



F/F_{MSY} trajectory



Plots capturing stock status uncertainty based on an exploratory model grid of biological parameters: growth, natural mortality, and steepness



Kobe scatter plot for 'recent' years (i.e. last 4 year avg) – 41 models from exploratory biological uncertainty grid
No estimation uncertainty, large spread

Results: management quantities

Metric	Mean	Median	Min	10%ile	90%ile	Max
SB_{latest}	2527	1304	469	670	6611	7571
SB_{recent}	1964	1070	488	627	4809	5780
$TotalBiomass_{latest}$	11258	8702	2600	3936	20544	33243
$TotalBiomass_{recent}$	9891	7393	2266	3280	16989	30188
F_{latest}	0.16	0.11	0.03	0.05	0.37	0.60
F_{recent}	0.26	0.15	0.05	0.05	0.68	0.99
SB_{MSY}	2071	2081	758	946	2940	4463
F_{MSY}	0.30	0.25	0.05	0.09	0.55	0.86
F_{recent}/F_{MSY}	0.90	0.84	0.12	0.24	1.57	1.97
F_{latest}/F_{MSY}	0.61	0.56	0.08	0.16	1.11	1.53
SB_{recent}/SB_{MSY}	1.13	0.73	0.17	0.29	2.29	5.19
SB_{latest}/SB_{MSY}	1.45	0.90	0.16	0.29	3.10	6.80
$SB_{recent}/SB_{F=0}$	0.25	0.18	0.03	0.06	0.65	0.75
$SB_{latest}/SB_{F=0}$	0.32	0.22	0.03	0.07	0.89	0.99

The critical uncertainty: population scale

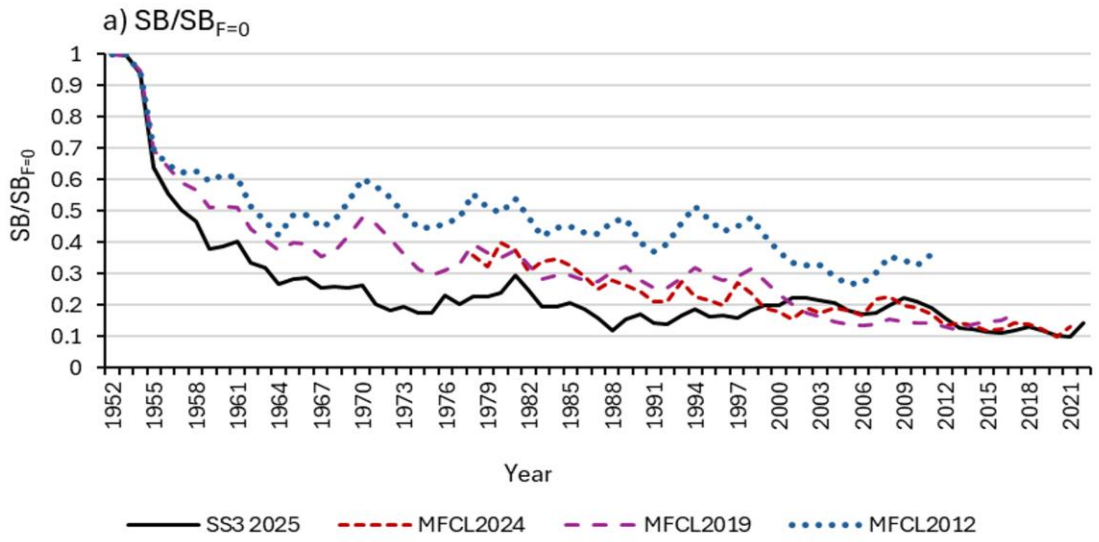
- There are outstanding issues with the integrated assessment that could not be resolved in terms of the model's ability to estimate a reliable stock productivity function and isolate the key drivers of the low population scale estimation.

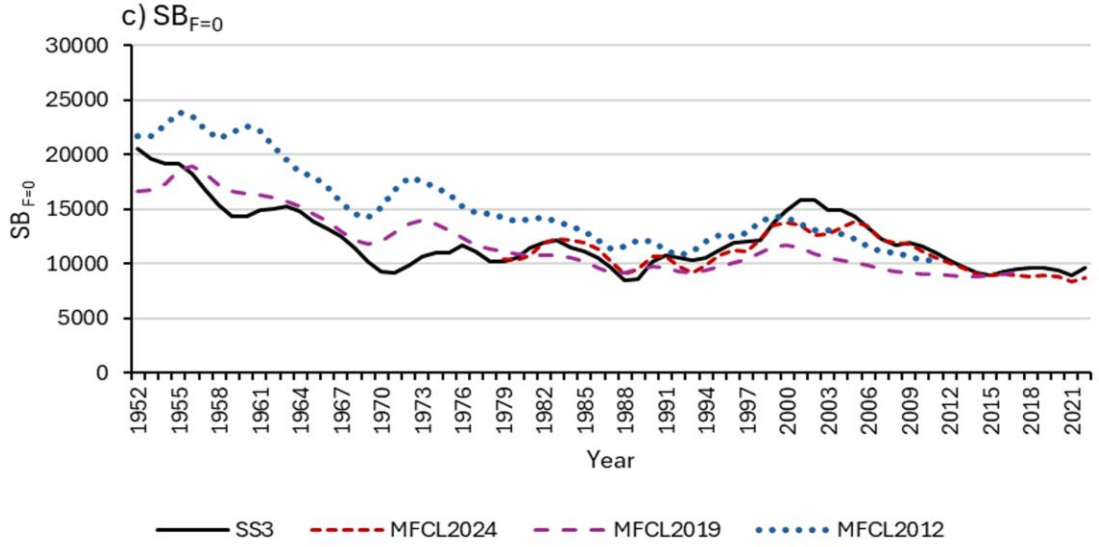
- **A "Low Population Size"** : The revised assessment, like its predecessors, consistently estimates a very small but highly productive stock, with dynamics driven by recruitment variability. This likely derives from model mis-specification involving fixed productivity assumptions. This issue is addressed further by the other population model that was developed (BSPM) (next presentation).

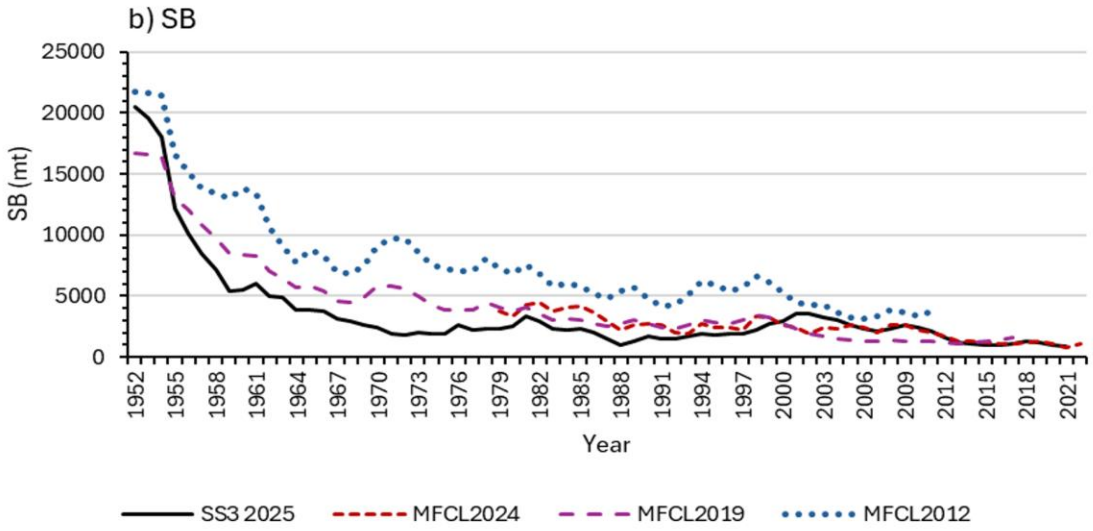
- **This seems highly improbable**: This small population size is difficult to reconcile with the observed catches, CPUE and the vast spatial scale of the SWPO. Simple calculations suggest the estimated biomass is too low to sustain the catches and catch rates observed in the fishery.

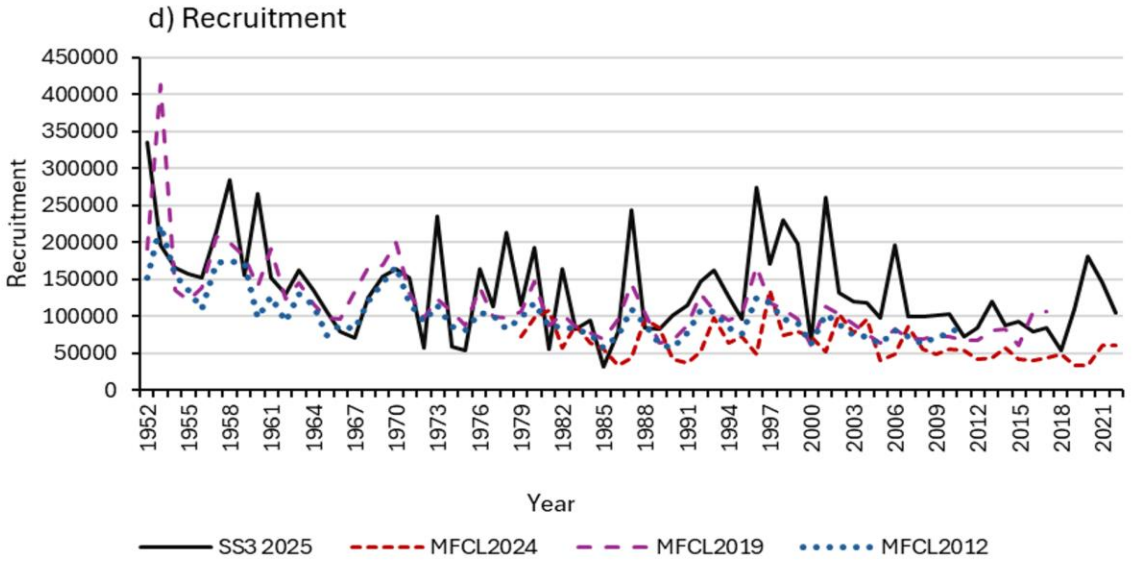
- **Why It Matters:** The model's reference points (e.g., $SB/SB_{MSY,}$ F/F_{MSY} and $SB/SB_{F=0}$) are scaled by this low population size. If the true population is larger, these reference points are likely biased low, and the stock status as presented would be overly pessimistic.

Comparisons of SWPO striped marlin assessments: MFCL - 2012, 2019, 2024 and SS3 - 2025









Worst-case scenario?

- The current assessment outcomes need to considered with the strong caveats around the population scale estimates and could be viewed as a **worst-case scenario**. The model fits the data with a small population, but a larger population size is more plausible.

- Uncertainty in the absolute population scale translates directly into **uncertainty in the management reference points**. This means the model's conclusions about overfished and overfishing status are likely unreliable, likely over pessimistic.

- Without a reliable, independent estimate of absolute population size, the model's outputs for **management advice are highly questionable**.

- Further- note the advice in the paper: " **… given the concerns around the biomass scale uncertainty, the model grid was not specifically designed as grid (or ensemble) for management advice."** If we had high confidence in the assessment – uncertainty characterisation would have received greater attention.

Management advice considerations

- When interpreting the results, the Scientific Committee (SC21) must seriously consider the **uncertainty in the population scale**.
- The estimated reference points are **likely overly pessimistic.** Assuming a higher true population scale would lead to more optimistic reference points and a better stock status.
- The assessment has empirical data, such as several CPUE indices (excluding the French Polynesia and New Zealand indices) that show a **recent increase in abundance**, suggesting some recent strong recruitment.
- The need for a reliable population scale highlights the urgency of implementing new methods like Close-Kin Mark-Recapture (CKMR) to provide a direct estimate of abundance.
- Given the concerns (unresolved issues) with the revised integrated assessment, consider the **alternative production model (SA-WP-07)** a more defensible source of management advice.

Recommendations

- **Improve Biological Data**: Invest in collecting better data on growth, movement, and population structure. Specifically, explore epigenetic aging to get more accurate age data for both striped marlin and swordfish.
- **Implement CKMR**: Use the proven Close-Kin Mark-Recapture (CKMR) method to get a more reliable estimate of the striped marlin population size. This will help resolve key uncertainties in the assessment.
- **Modelling enhancements:**
  - Explore a Bayesian approach for future integrated stock assessments (this can better capture the uncertainty in population scale).
  - Continue to develop and use observer-based CPUE data, as it offers a robust way to track abundance trends, with more operational covariates available for CPUE analysis.
- **Conduct a workshop** to discuss assessment challenges and share best practices for bycatch species like billfish and sharks.

To be continued? …