

Refinements to Rceattle for eastern Bering Sea pollock

Working paper on 2024 “PM” model refinements

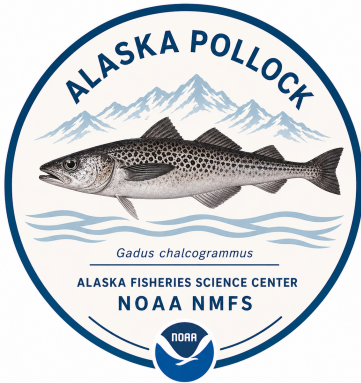
James Ianelli

Grant Adams

August 16, 2026

Abstract

The custom AD Model Builder (ADMB) stock assessment model developed for eastern Bering Sea walleye pollock has a long history of application. For a variety of reasons, including general recommendations from the 2025 Center for Independent Experts (CIE) reviews, the custom ADMB code has been successfully ported to R Template Model Builder (RTMB). That conversion provides access to advanced RTMB features and makes the underlying code easier to understand and modify. Similarly, adopting more generic platforms, such as Rceattle, can provide access to more modern software developments. In this study, we tailored Rceattle code and configurations so that Rceattle can closely mimic the custom ADMB/RTMB assessment code. For example, we added a non-parametric selectivity form as an option within Rceattle. We also introduced a two-dimensional, first-order autoregressive (AR1) selectivity form, which we consider a reasonable alternative for further evaluation.



Rceattle COMPANION REPORT

1 Executive Summary

Building from the ADMB-to-RTMB conversion, this analysis evaluates whether Rceattle can be configured to reproduce the main features of the 2024 eastern Bering Sea pollock “PM” assessment while providing a more general and extensible modeling framework. This work included the ability to introduce dynamic structural equation modeling (DSEM) for evaluating recruitment patterns. This study focuses on comparisons with the custom ADMB/RTMB assessment and next steps for application within the operational stock assessment.

The corrected comparison uses complete ages 1–15 and nominal, unadjusted composition sample sizes in both `pm.tp1` and Rceattle. Applying the corrected ADMB parameters and fishery selectivity in Rceattle reproduces numbers-at-age, spawning biomass, and predicted catch to numerical precision. Mean absolute differences are 0.000113% for spawning biomass and 0.000153% for predicted catch. This fixed-parameter result establishes the population- dynamics bridge.

The independently estimated canonical PM Rceattle fit reaches a small maximum gradient and retains weak Hessian identification in the annual selectivity deviations. Relative to the independently estimated corrected ADMB fit, its mean absolute differences are 9.94% for spawning biomass and 4.76% for recruitment. These differences describe estimation structure and parameterization, while the fixed-parameter calculation isolates software implementation agreement.

The 2D age-by-year AR1 selectivity sensitivity was included as an initial diagnostic attempt. Results appeared broadly similar to the non-parametric constrained selectivity fits. However, because some parameters reached bounds, the model failed to converge. Closer examination of this selectivity form will continue as a research topic.

One-step-ahead residuals and nine terminal-year-consistent retrospective peels have been regenerated from the canonical PM fit. Aggregate composition residual dispersion is lower than the standard-normal reference. Mohn’s rho is 0.261 for spawning biomass, 0.122 for total biomass, 0.132 for recruitment, and -0.211 for fishing mortality. These results focus subsequent work on selectivity identification, retrospective bias, and observation-model sensitivity.

The exploratory Bayesian work remains separate in the [SparseNUTS companion report](#).

2 Purpose and Scope

The September 2026 review focuses on whether the Rceattle configuration has sufficient documentation for Plan Team evaluation. This working paper serves that focused purpose rather than the broader purpose of a full Stock Assessment and Fishery Evaluation (SAFE) chapter.

This working paper therefore records:

- what was run from those development files;
- how the Rceattle outputs compare with the modified ADMB reference;
- which diagnostics are available now;
- which diagnostics remain unfinished; and
- what decisions or requests would help structure the next evaluation pass.

Standard SAFE report sections such as background material, fishery history, ecosystem narrative, reference-point tables, projections, and harvest advice are outside the scope of this document.

3 Initial Rceattle Developments

3.1 Model Inputs

The Rceattle configuration uses the prepared workbook `Data/2024_EBS_pollock_canonical_pm.xlsx`. The data builder retains every age from 1 through 15 and the nominal composition sample size for each fishery, bottom trawl survey (BTS), and acoustic-trawl survey (ATS) observation. Age-1 survey selectivity remains separately parameterized from the age-2–15 curve; the likelihood evaluates one complete 15-age composition. This treatment aligns Rceattle with the corrected `pm.tp1` data and removes the earlier sample-size inflation based on the ages-2–15 subtotal.

The run uses Rceattle version 4.9.1.9001 from upstream commit 7d873ca76b4c, plus the archived `MultinomialPM` patch identified by MD5 586d47343fb386a04dcbfc2904f4f889.

The Rceattle development lane adds `MultinomialPM`, which evaluates the corrected PM stabilized composition criterion

$$-N_y \sum_a o_{y,a} \log \left(\frac{p_{y,a} + c}{o_{y,a} + c} \right),$$

Table 1: Disposition of the initial Rceattle development components.

Item	Initial_development	September_review_result	Plan_Team_interpretation
1	Prepared Rceattle model-input workbook	Used the delivered workbook directly.	Input files are available and reproducible from stored files.
2	Main two-stage fit and 2D AR1 sensitivity	Two-stage NonParametricPM completed; 2D AR1 received a FAIL convergence status.	The NonParametricPM configuration is complete; the relevant Rceattle comparison; 2D AR1 is diagnostic only.
3	Convergence and model diagnostics	Structured convergence checks, OSA diagnostics, and nine two-stage retrospective peels are saved.	All nine final peel fits received OK status; jitters, self-tests, and profiles remain pending.
4	Next-assessment data update workflow	Update script review is complete; execution awaits complete new-year observations.	Useful as a template only after complete new-year observations are inserted.
5	DSEM recruitment example	Script review against the ecosystem and socioeconomic profile (ESP) cohort-table workflow is complete; model execution remains pending.	Placeholder covariates should be replaced with cohort-aligned ESP indicators before fitting.
6	ADMB-to-Rceattle comparison diagnostics	The ADMB-parameter comparison reproduces ADMB to rounding-scale precision.	This is the strongest evidence that the ADMB comparison is correct.

Table 2: Composition-likelihood fits using complete ages 1–15 and nominal, unadjusted sample sizes. Raw objectives differ in included constants and likelihood conventions, so gradients and trajectories provide the meaningful cross-formulation comparisons.

Model	Composition likelihood	Offset	Maximum gradient	Status
Legacy Rceattle AFSC formulation	MultinomialAFSC	0.001	4.40×10^{-5}	WARN
Canonical PM stabilized multinomial	MultinomialPM	0.001	4.67×10^{-5}	WARN
Standard multinomial sensitivity	Multinomial	0.000	7.57×10^{-5}	WARN

where $o_{y,a}$ and $p_{y,a}$ are complete observed and predicted age proportions, N_y is the nominal sample size, and $c = 0.001$ stabilizes the logarithm. The offset changes the score slightly, so this criterion remains a stabilized pseudo-likelihood. A standard multinomial sensitivity uses $-N_y \sum_a o_{y,a} \log(p_{y,a})$ and supplies the statistically conventional reference. The two formulations answer complementary questions: canonical PM measures implementation agreement with corrected `pm.tp1`, while the standard multinomial evaluates sensitivity to the stabilizing convention.

All three fits reach small maximum gradients. Each receives a **WARN** because the Hessian condition number remains near 10^6 and its weakest direction loads on annual selectivity deviations. The likelihood correction therefore improves statistical coherence and preserves a separate selectivity identification issue for continued evaluation. Relative to canonical PM, the standard multinomial sensitivity changes spawning biomass by 10.83%, recruitment by 4.58%, and total biomass by 8.16% on average in absolute percentage terms. This material sensitivity supports reporting both lanes and using the standard multinomial as the principal inferential benchmark.

The next-assessment update script review is complete, and execution awaits complete new-year observations. The script duplicates terminal records and places placeholders for new catch, composition, and survey observations. The script serves as a template until complete new-year observations are available; after population, it creates a complete assessment workbook.

3.2 Selectivity specifications

3.2.1 Non-parametric selectivity

The primary independently fitted comparison is the two-stage NonParametricPM fit. The initial analysis uses an empirical fishery selectivity start because the fishery selectivity likelihood is multimodal. The fit sequence is:

1. turn time-varying selectivity deviations off to establish the selectivity and abundance scale; and
2. enable the deviations and initialize the full fit from the first stage.

This approach completed and produced close agreement with ADMB. The final fit has a **WARN** convergence status because the Hessian condition number is high and the least-identified direction loads almost entirely on `sel_coff_dev`.

The time-varying selectivity penalties use explicit standard deviations on log-selectivity random-walk increments. The fishery and its shared CPUE selectivity block use an SD of 0.500, ATS and AVO use 0.138, and BTS uses 1.000. Thus, the fishery curve has appreciable annual flexibility, whereas the ATS and AVO curves receive stronger temporal smoothing. The fishery specification also retains the ADMB-matched curve-shape penalty weights (12.5, 1/60, and 1) and an average-selectivity centering weight of 10. These values are penalty coefficients rather than SDs.

3.2.2 2D-AR1 selectivity

The 2D AR1 sensitivity was configured for the fishery selectivity block across the full 1964–2023 period supported by fishery age-composition data. The catch-per-unit-effort (CPUE) index uses the same selectivity

Table 3: Estimated 2D AR1 fishery-selectivity correlation and scale parameters.

Quantity	Rceattle parameter	Penalized fixed effects	Final random effects
Year AR1 correlation	Sel_curve_pen1	0.584	0.947
Age AR1 correlation	Sel_curve_pen2	0.120	0.098
Selectivity deviation scale	sel_dev_log_sd	0.500	0.379

Correlations are reported on their -1 to 1 scale, and the selectivity-deviation scale is reported on its positive standard-deviation scale.

through the shared `Selectivity_index`. The 2024 selectivity field is carried forward from 2023, the final year with fishery age-composition data. Other fleet selectivity forms retain their existing configurations. The fishery and CPUE rows were set to `Selectivity = "2DAR1"`, with `Time_varying_sel = "Off"` because the 2D AR1 field provides the age-year variation directly. The original saved sensitivity displayed in the comparison figures covered 12 selectivity bins beginning at age 1 and 61 years (1964-2024).

Core assessment parameters were optimized jointly with the 2D AR1 selectivity field over the full fishery age-composition period. Within the shared fishery/CPUE block, the base selectivity coefficients, year correlation, age correlation, and deviation scale were active fixed-effect parameters. The age-year deviations were first tested as penalized fixed effects (`random_sel = FALSE`) to obtain a non-zero field. The final sensitivity then used `random_sel = TRUE`, so the age-year deviations were treated as random effects and integrated with the Laplace approximation. The estimated correlations and scale are shown in Table 3.

The original Laplace run required about 13 minutes of optimization time. A follow-up run fixed `sel_dev_log_sd` at -0.9702190739, equivalent to an SD of 0.379, and restricted estimation of the field to 1964–2023. Its direct Laplace optimization ran for about 10 minutes before an undefined gradient ended the fit. A penalized-effects initialization followed by Laplace integration also ended with an undefined gradient. Fixing the deviation scale therefore preserved the convergence problem. These timings illustrate a practical cost of integrating the large age–year field even before a successful optimum is reached.

Rceattle performs the optimization using unconstrained internal values. For a correlation, an internal value x is converted using $2\logit^{-1}(x) - 1$; for the deviation scale it is converted using $\exp(x)$. Thus, the final internal values 3.600, 0.197, and -0.969 correspond to the reported values 0.947, 0.098, and 0.379, respectively. The internal values are optimizer coordinates; the transformed values provide the correlations and standard deviations used for interpretation.

The final random-effects run estimated the year correlation at 0.947, the age correlation at 0.098, and the selectivity-deviation scale at 0.379. These values should be interpreted only as diagnostic estimates because the final result failed convergence checks:

- maximum marginal gradient: 17.8, largest on `rec_pars`;
- Hessian inversion failed, leaving standard errors unavailable;
- 56 `log_F` parameters at configured bounds; and
- final convergence status: **FAIL**.

The terminal biomass values remained near ADMB, and the results were similar enough to warrant further evaluation. However, the historical trajectory differences and failed convergence checks mean that closer examination of this selectivity form should continue as a research topic.

The three specifications are compared directly in Figure 1. The ADMB reference and fitted NonParametricPM curves are the relevant bridge comparison; the 2D-AR1 curves are retained only to show the behavior of the failed sensitivity.

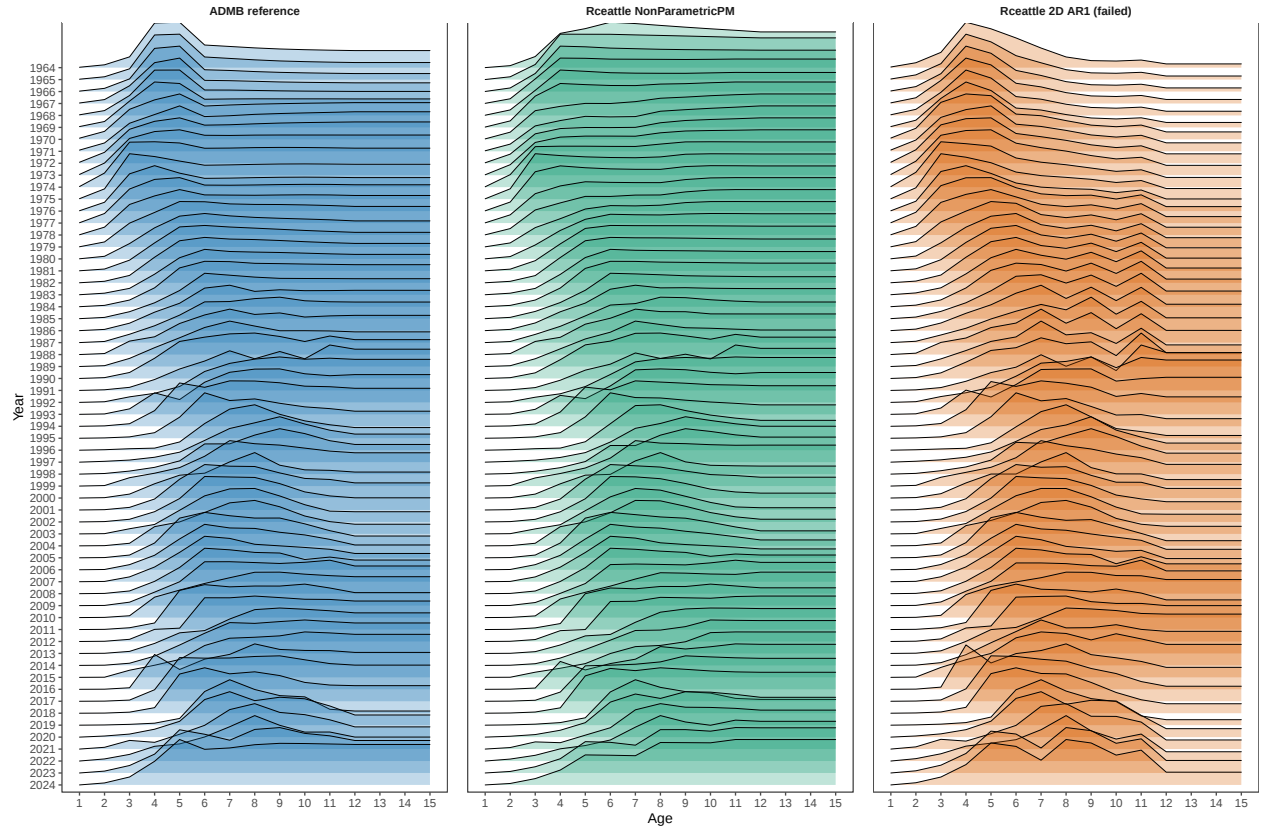


Figure 1: Relative fishery selectivity by year for the modified ADMB reference, fitted Rceattle NonParametricPM model, and Rceattle 2D age-by-year AR1 sensitivity. Each annual curve is normalized to a maximum of one. The 2D-AR1 panel is shown as a diagnostic comparison only because that fit failed convergence checks.

Table 4: ADMB-parameter comparison summary.

Metric	Value
N ratio minimum	0.999995
N ratio maximum	1.000007
SSB ratio minimum	0.999997
SSB ratio maximum	1.000003
SSB mean absolute percent difference	0.000113
Catch mean absolute percent difference	0.000153

4 ADMB Comparison Results

The comparison analysis separates implementation agreement from estimation agreement.

The first comparison applied the ADMB parameter estimates and fishery selectivity-at-age in Rceattle and compared the resulting output with the ADMB pollock assessment model. This checks whether the same estimates produce the same derived quantities. The agreement is effectively exact.

The two-stage NonParametricPM fit is an optimization and model-behavior test. It estimates the Rceattle model independently and asks whether the optimizer reaches a scientifically similar solution.

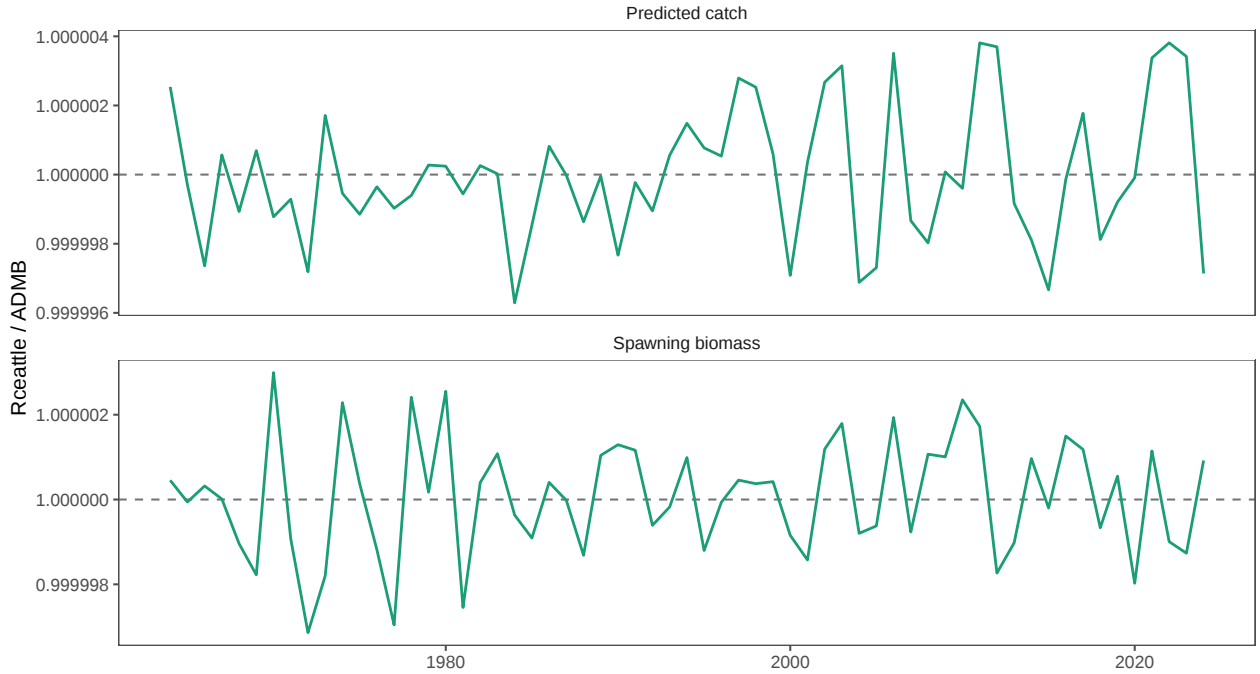


Figure 2: Fixed-parameter Rceattle-to-ADMB ratios for spawning biomass and predicted catch.

The independently estimated canonical PM configuration differs most during the early period and in the terminal recruitment estimate. The fitted comparison table quantifies those estimation differences, while the fixed-parameter ratios remain the regression gate for implementation equivalence. The standard multinomial sensitivity in Table 2 provides the observation-model check.

Table 5: Agreement between fitted Rceattle trajectories and the modified ADMB reference.

Quantity	Correlation	Mean % difference	Maximum % difference	2024 % difference
Historical 2D AR1 sensitivity				
Recruitment	1.00	7.40	21.66	-16.92
Spawning biomass	0.92	38.54	308.54	5.65
Total biomass	0.87	29.17	254.57	2.44
Canonical MultinomialPM				
Recruitment	1.00	4.76	19.96	-19.96
Spawning biomass	0.98	9.94	89.61	0.95
Total biomass	0.98	7.15	72.47	-2.13
Standard multinomial sensitivity				
Recruitment	1.00	3.29	15.84	-15.84
Spawning biomass	0.98	14.68	113.20	8.70
Total biomass	0.98	10.93	93.54	4.43

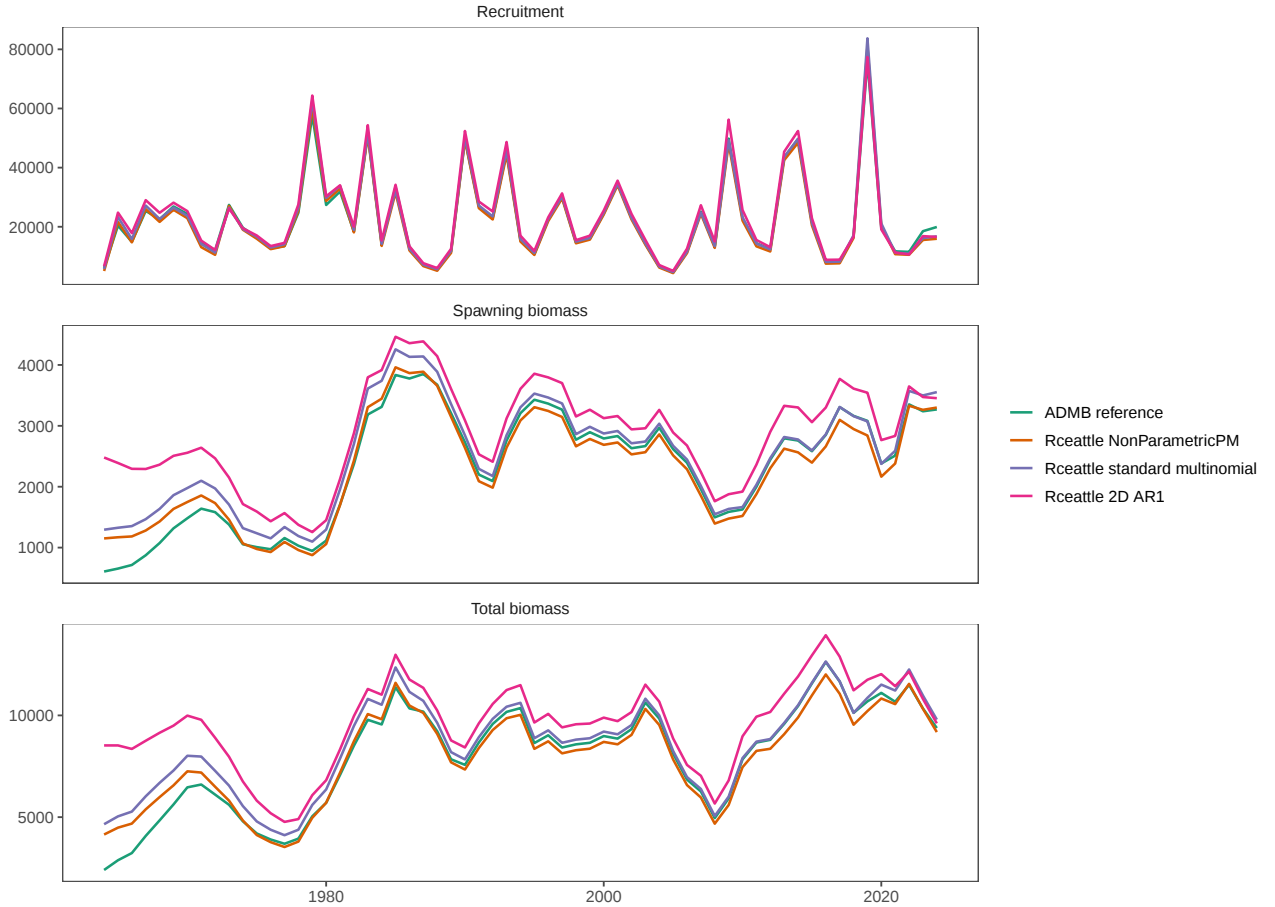


Figure 3: Corrected ADMB reference and independently estimated Rceattle trajectories for the canonical PM, standard multinomial, and historical 2D AR1 sensitivity configurations.

Table 6: Stored Rceattle convergence checks for the evaluated configurations.

Model	Status	Objective	Max gradient	Worst parameter	Positive de
Stage 1 scale fit	OK	1,976.562634	0.000215	sel_coff	TRUE
Canonical PM two-stage fit	WARN	722.487452	0.000047	rec_pars	TRUE
Standard multinomial sensitivity	WARN	2,840.124006	0.000076	sel_coff_dev	TRUE
2D AR1 selectivity sensitivity	FAIL	1,270.801558	17.771505	rec_pars	NA

5 Diagnostics

The base Rceattle configuration has acceptable gradients and a positive definite Hessian, but the Hessian-conditioning warning is substantive. It says the remaining weak direction is almost entirely in the time-varying fishery selectivity coefficients. That is a diagnostic issue to resolve or clearly bound before this fit is promoted from comparison evidence to an operational assessment candidate.

5.1 MCMC integrations

The Markov chain Monte Carlo (MCMC) analysis uses the canonical PM two-stage NonParametricPM base model. Recruitment deviations are represented as latent random effects when constructing the sampling object, which exposes the joint sparse precision structure while preserving the base model likelihood and fitted starting point. SparseNUTS version 1.0.2 at upstream commit 2f3f1626 was used; the installed commit matched the current repository head when the run began.

The production analysis used 8 independent chains scheduled across 2 parallel workers, 2,000 iterations per chain, and 1,000 warmup iterations. The call used `skip_optimization = TRUE`, preserving the canonical two-stage fit as the sampling reference and bypassing SparseNUTS’s internal marginal re-optimization. The sampler explicitly used `metric = "sparse"`, with an adaptation target of 0.95 and a maximum tree depth of 12. The saved joint precision matrix contains 99.51% nonzero entries. Thus, the call exercises SparseNUTS’s sparse-matrix path, while the fitted Rceattle parameterization produces an effectively dense precision structure. Figure 4 makes that structure explicit and shows why sparsity yields limited computational benefit for this model.

Sparse-precision construction required 679.3 seconds. The sampler itself averaged 31.4 minutes per chain, with a range of 10.4–53.8 minutes. Mean warmup and retained-sampling times were 26.3 and 5.2 minutes per chain, respectively. Summed across chains, sampling used 251.4 chain-minutes. Scheduling 2 chains at a time required approximately 137.3 minutes of elapsed computing time, excluding report generation and small file-processing overheads.

The 8,000 retained draws had divergence and maximum-tree-depth rates of 0.00% and 0.00%, respectively. Maximum split R-hat was 1.299 for `sel_inf_dev[7]`. Minimum bulk ESS was 20 for `sel_inf_dev[7]`, and minimum tail ESS was 24. The sampler-event diagnostics are clean, while the R-hat and ESS results identify a small set of selectivity-related directions that would benefit from longer chains or reparameterization.

The pairs plot focuses on `sel_inf_dev[7]`, `log_sel_slp_dev[7]`, `log_sel_slp_dev[6]`, `sel_inf_dev[8]`, `index_log_q[3]`, `log_F[2]`. These parameters had the smallest bulk ESS values in the rerun and therefore provide the most direct visual check for chain separation, strong posterior correlation, or narrow joint geometry.

The convergence statistics, trace behavior, and sampler-event counts provide the basic screening results for this integration. Parameters with elevated R-hat or low effective sample size identify directions that require additional sampling or model reparameterization before the posterior is used for assessment inference. The saved starting point has a maximum marginal gradient of 2.608 for the Laplace-marginal object used to construct the precision matrix. That warning means the sparse preconditioner inherits substantial local uncertainty from the weakly identified selectivity direction in the base fit, so the sampler diagnostics deserve primary weight in judging this run.

Table 7: Diagnostic status for the September 2026 review.

Diagnostic	Current_status	Use_for_September_2026_review
ADMB-parameter comparison	Completed and saved in fixed-parameter comparison outputs.	Primary numerical agreement check.
Two-stage NonParametricPM fit	Completed with WARN status from Hessian conditioning.	Primary independently fitted Rceattle comparison.
2D AR1 selectivity sensitivity	Received a FAIL convergence status, an unsuccessful Hessian inversion, and parameter-bound warnings.	Requires further research before it can be considered as an alternative selectivity form.
OSA residuals	Completed for 1,864 residuals and saved in RDS outputs.	Useful, but analytical and exactly reproduced index groups need interpretation.
Rceattle retrospectives	Nine two-stage peels completed for the NonParametricPM fit with OK final-fit status.	The revised peels preserve data availability and constrain terminal selectivity parameters to the preceding year.
MCMC integrations	Eight SparseNUTS chains sample the NonParametricPM base model with a sparse metric.	Provides posterior-integration diagnostics for the canonical base configuration.
Legacy ADMB retrospectives	Existing ADMB retro.rep files are present.	Provides historical context distinct from the Rceattle retrospective peels.
Jitters, self-tests, and likelihood profile	Pending.	Needed to complete a full diagnostics package.

Table 8: SparseNUTS sampling configuration and basic diagnostics for the canonical PM two-stage base model. Divergence and tree-depth percentages refer to retained sampling iterations after warmup.

Quantity	Value
SparseNUTS version	1.0.2
Upstream commit	2f3f1626
Chains	8
Iterations per chain	2,000
Warmup per chain	1,000
Retained draws	8,000
Metric	sparse
Adaptation target (adapt_delta)	0.95
Maximum tree depth	12
SparseNUTS internal optimization	Skipped
Precision-matrix nonzero entries	99.51%
Maximum marginal gradient used for Q	2.608 (warning)
Post-warmup divergences	0.00%
Maximum-tree-depth transitions	0.00%
Maximum R-hat	1.299
Minimum bulk ESS	20
Minimum tail ESS	24
Precision construction	679.3 seconds
Mean sampler time per chain	31.4 minutes
Approximate parallel elapsed time	137.3 minutes

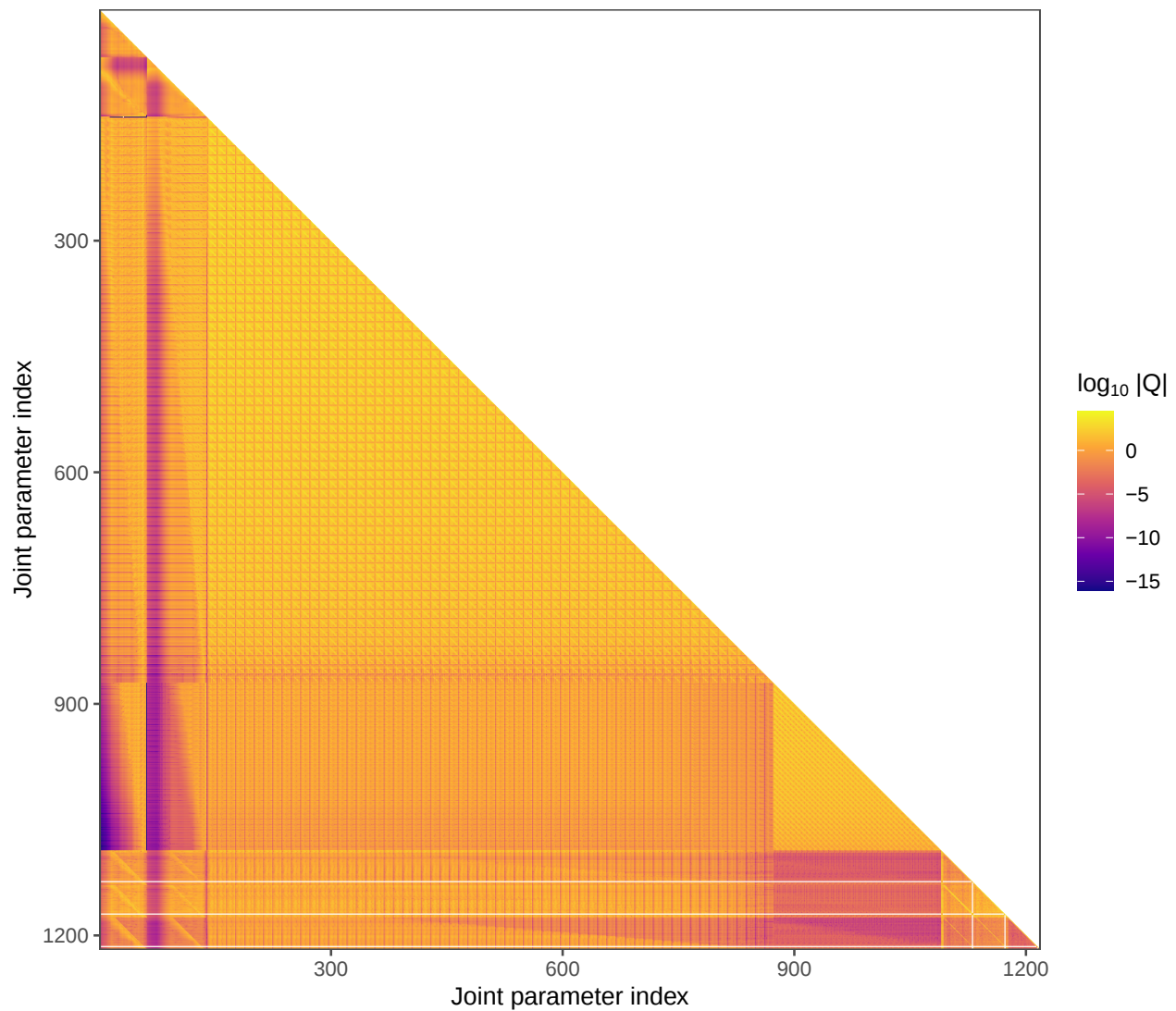


Figure 4: Stored triangular view of the joint precision matrix used by SparseNUTS for the canonical PM Rceattle fit. Axes follow the joint parameter-vector order; color gives the base-10 logarithm of absolute precision. The broad filled structure shows that the selected sparse-matrix path receives an effectively dense precision matrix.

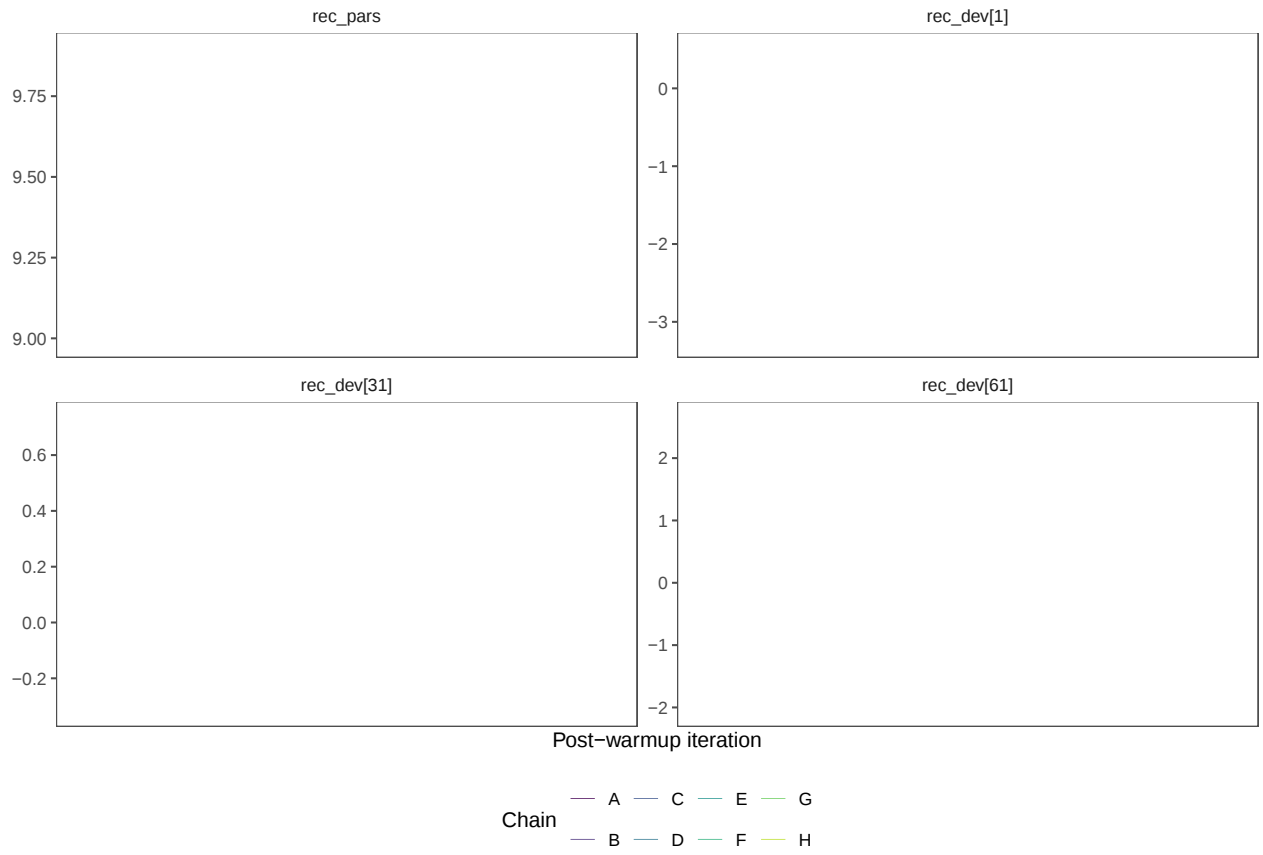


Figure 5: Post-warmup SparseNUTS traces for the recruitment scale and representative early, middle, and late recruitment deviations from the NonParametricPM base model. Each color identifies one chain.

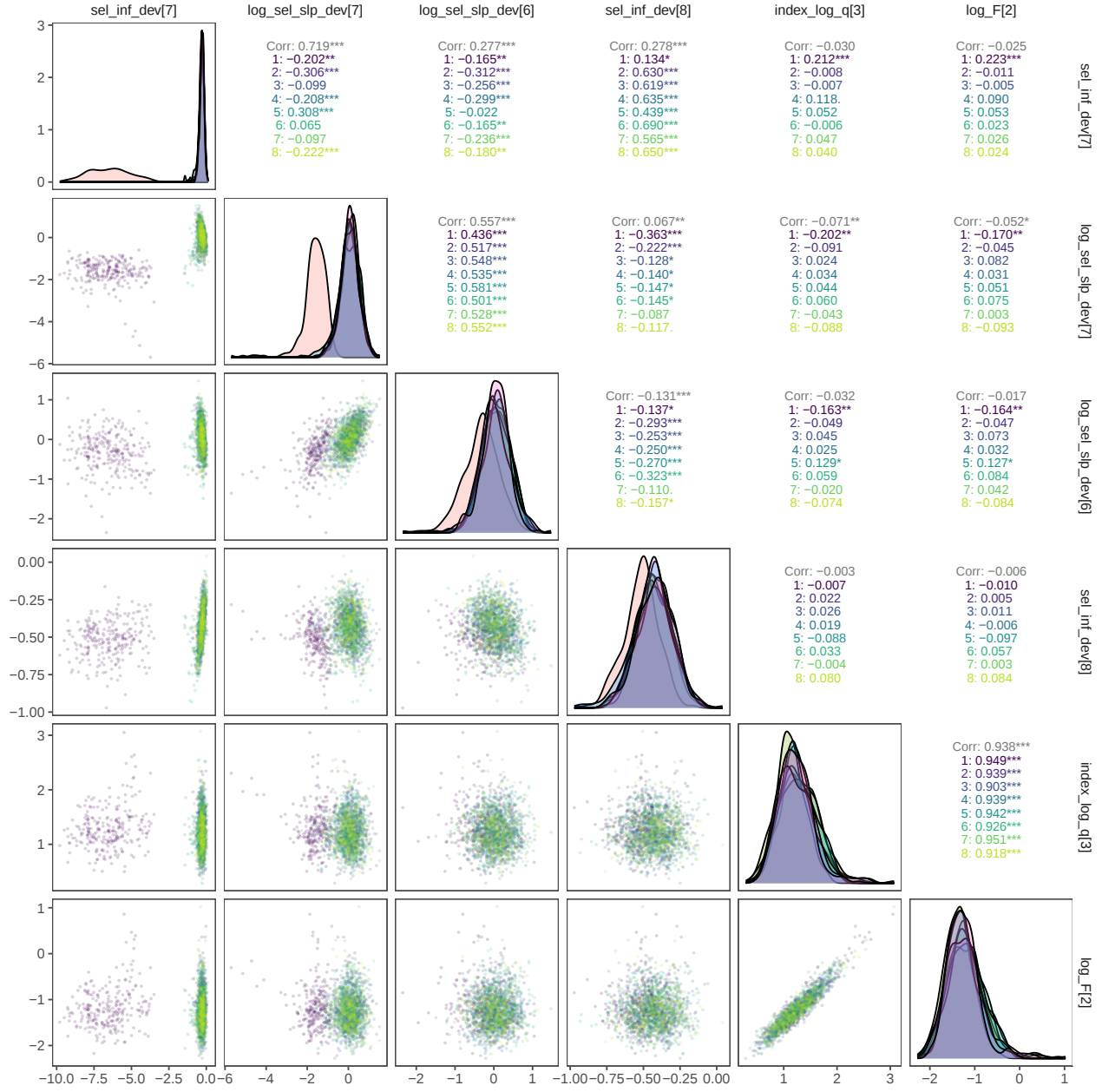


Figure 6: Pairs plot for the six slowest-mixing SparseNUTS parameters, selected by bulk effective sample size in the eight-chain canonical PM run. Colors identify chains, diagonal panels show marginal densities, lower panels show joint draws, and upper panels show correlations. Systematic separation of chain colors indicates incomplete mixing.

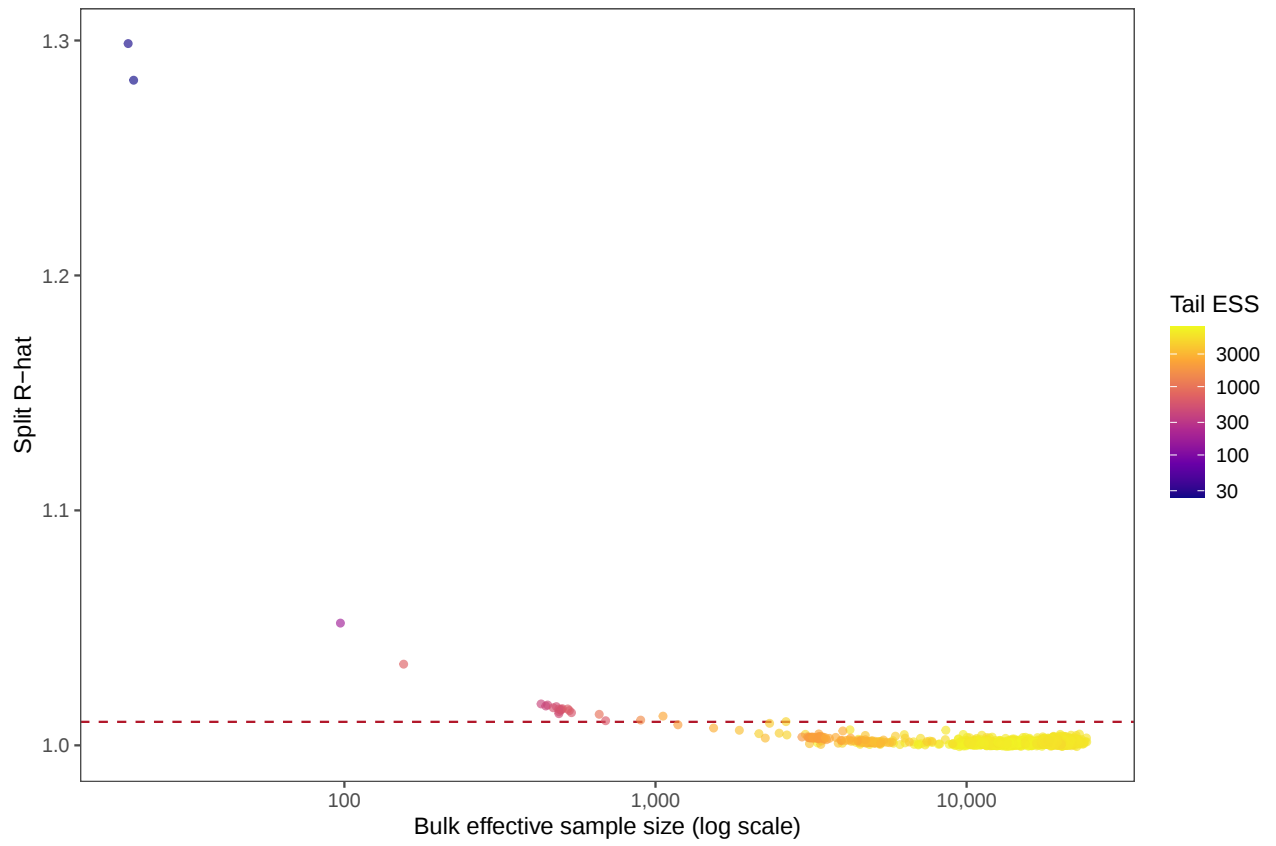


Figure 7: Effective sample size and split R-hat for SparseNUTS parameters in the NonParametricPM base model. The horizontal line marks R-hat 1.01, and point color shows tail effective sample size.

5.2 OSA Residuals

OSA residuals run on the pinned Rceattle `MultinomialPM` development lane. All panels in Figure 8 apply to the final two-stage `NonParametricPM` Rceattle fit: the independently fitted non-parametric selectivity configuration used for the modified ADMB comparison. They represent the fitted canonical PM configuration exclusively; the failed 2D-AR1 sensitivity and fixed-parameter bridge calculation use separate results. The composition diagnostics use `afscOSA` version 0.0.1 and pass Rceattle’s internally calculated residuals to `afscOSA::run_osa()`. The residual builder uses the nominal, unadjusted composition sample sizes and complete ages 1–15. This follows the current [afscOSA interpretation guidance](#) while retaining Rceattle’s residual calculation.

Following that guidance, Figure 8 should be read from top to bottom. The aggregate age-composition fits are the first check for systematic misfit, including patterns that could indicate an unsuitable selectivity form. The red vertical ranges contain 95% of data simulated from the fitted multinomial model conditional on the aggregate sample size and fixed parameter estimates. They represent conditional simulation variability. The quantile–quantile plots then assess whether the OSA residuals resemble a standard normal distribution. The standard deviation of normalized residuals (SDNR) should be near one, and the displayed tail statistics provide a separate check on extreme residual behavior. Finally, the OSA and Pearson bubble plots are used to look for age, year, or cohort patterns that may suggest the source of misfit. Each statistic and nominal interval contributes to an integrated diagnostic interpretation rather than an automatic model-rejection rule.

The aggregate observed and fitted compositions are generally close, although the survey panels contain some age-specific departures. The OSA distributions are narrower than the standard-normal reference for all three composition sources: SDNR is 0.71 for the fishery, 0.84 for BTS, and 0.70 for ATS. Their tail diagnostics also show shorter-than-expected tails. Together these results indicate underdispersion relative to the fitted multinomial observation model. Model selection requires this information alongside the full suite of diagnostics. The bubble plots should instead be used to identify persistent age, year, or cohort patterns and to guide focused checks of selectivity, composition weighting, and observation-model assumptions. One Pearson residual (8.86) is capped at 6 in the figure, following the package’s plotting convention.

5.3 Retrospective patterns

Nine retrospective peels were run from the final two-stage `NonParametricPM` fit. The unpeeled model ends in 2024, and the peeled fits end in 2023 through 2015. Each peel preserves the data-availability pattern of the 2024 model. Thus, fishery age compositions end one year before the peel terminal year, survey series retain their historical sampling schedule and terminal lags, and earlier discontinued series retain their original ending years.

Each peel recalculates the empirical fishery-selectivity start and then applies the established two-stage sequence: Stage 1 fits the base curve with annual deviations disabled, and Stage 2 estimates the annual deviations from the Stage-1 solution. The terminal-year fishery and mirrored catch-per-unit-effort selectivity parameters share the preceding-year parameter values. All nine final Stage-2 fits received OK convergence status, with maximum gradients from $2.9\text{e-}05$ to $8\text{e-}05$.

The `NonParametricPM` fishery selectivity is parameterized at length. Consequently, the terminal selectivity-at-length curve exactly equals the preceding-year curve. Rceattle converts that curve to selectivity at age using a year-specific growth and age–length transition matrix, so the reported selectivity-at-age curve can change between those years. This distinction preserves the intended selectivity constraint while retaining each year’s biological size-at-age information. These Rceattle fits are separate from the legacy AD Model Builder (ADMB) retrospective files.

The estimated Mohn’s rho values are 0.261 for spawning biomass, 0.122 for total biomass, 0.132 for recruitment, and -0.211 for fishing mortality. The positive biomass and recruitment values indicate that the peeled models generally estimate higher quantities at their terminal years than the corresponding unpeeled estimates. The negative fishing-mortality value indicates lower terminal fishing mortality in the peels. Spawning biomass shows the largest average relative difference. The consistent OK statuses and small gradients support interpretation of these values as properties of the revised retrospective specification, while the magnitude of

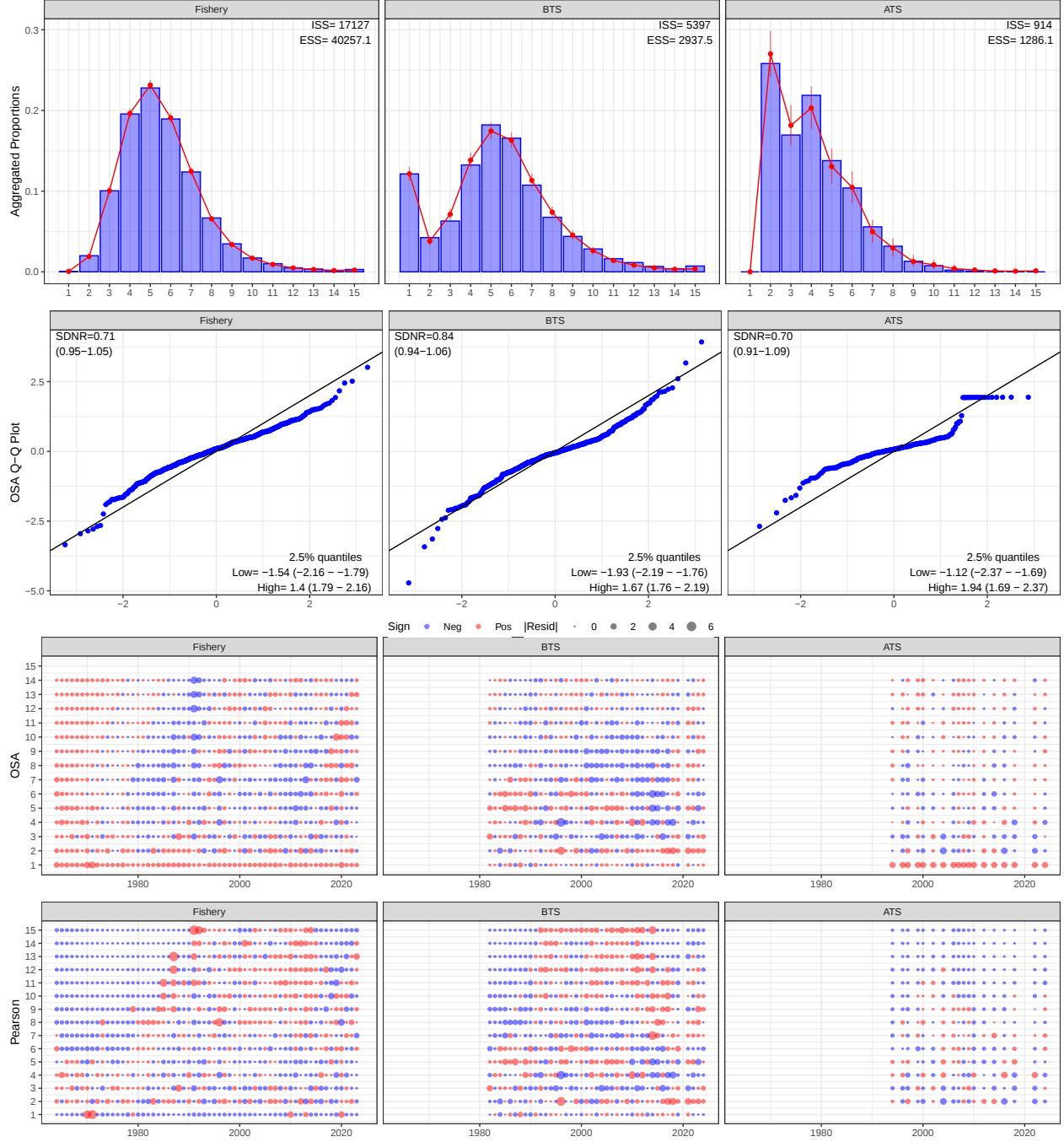


Figure 8: Age-composition diagnostics for the final canonical PM two-stage Rceattle fit. The fit uses complete ages 1–15 and nominal, unadjusted sample sizes. Diagnostics were produced with afscOSA for the fishery, bottom trawl survey (BTS), and acoustic-trawl survey (ATS). The top row compares aggregate observed proportions (blue bars) with fitted proportions (red points and lines); red vertical ranges contain 95% of data simulated conditionally from the fitted multinomial model, aggregate sample size, and fixed parameter estimates. The second row shows quantile–quantile plots of Rceattle’s internal one-step-ahead (OSA) residuals with the standard deviation of normalized residuals (SDNR), its expected 95% interval, and lower and upper tail diagnostics. The lower rows show OSA and Pearson residuals by age and year; red and blue indicate positive and negative residuals, respectively, and bubble area represents magnitude. Absolute residuals greater than 6 are capped at 6 for display. The fitted model retains ATS 2020; its sample size of one yields zero rounded counts in the afscOSA plotting conversion, so the plotting layer omits that row.

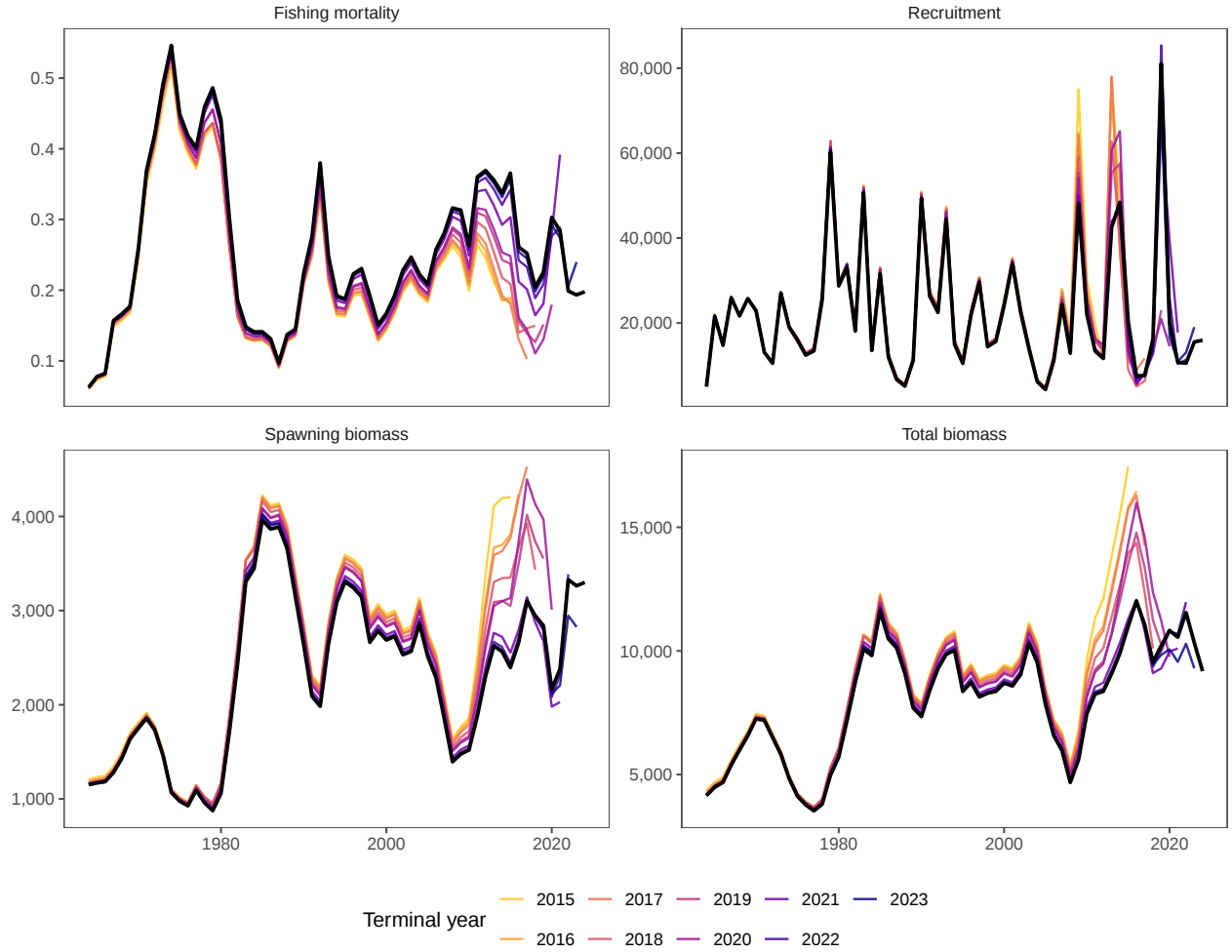


Figure 9: Nine two-stage retrospective peels for the fitted Rceattle NonParametricPM configuration. The black line is the unpeeled model ending in 2024; colored lines are successive peels ending in 2023 through 2015 and are plotted through each peel’s terminal year. Each peel preserves the original data-availability lags, recalculates the empirical selectivity start, and fixes terminal-year selectivity-at-length parameters to their preceding-year values. All nine final Stage-2 fits received OK convergence status.

Table 9: Mohn’s rho for the nine two-stage Rceattle NonParametricPM retrospective peels. Values are the mean relative difference between a peeled estimate in its terminal year and the corresponding estimate from the unpeeled model. Each peel preserves the 2024 data-availability pattern and fixes terminal-year selectivity-at-length parameters to their preceding-year values.

Quantity	Peels	Mohn’s rho
Total biomass	9	0.122
Spawning biomass	9	0.261
Recruitment	9	0.132
Fishing mortality	9	-0.211

Table 10: Projection files generated from the saved Rceattle fit for the spmR/SPM workflow.

spmR Projection Files

file	role	exists	md5
spm.dat	SPM setup	TRUE	8b4f5e91196e976769547c1c86d1e93d
pm.prj	Rceattle-derived species input	TRUE	7bcc2c2e9be26ab9a68d3f335350eefe
tacpar.dat	TAC parameters	TRUE	3b02746ed92fe8984e83896029935fba
spm	SPM executable	TRUE	6f52316e14f001e397a765ebfd71f022
age_schedules.csv	Rceattle-derived age schedules	TRUE	a66020aa97f44441bdb33a164557e499

spawning-biomass rho motivates continued evaluation of selectivity, survey influence, and recent recruitment estimates.

6 Projections

The saved canonical PM two-stage Rceattle fit supplies every population and fishery input for the Standard Projection Model workflow implemented in **spmR** (Ianelli 2026). The run uses **spmR** version 0.3.0 from pinned main-branch commit `e86fc6aa`. The adapter in `R/write_spmr_projection_inputs.R` writes terminal abundance, natural mortality, maturity, spawning and fishery weights, terminal fishery selectivity, recruitment, and spawning-biomass histories to `pm.prj`. The reproducible runner `R/run_spmr_tier3_projection.R` then applies those inputs to all seven Tier 3 alternatives. The report reads the saved projection products and therefore reads one auditable Rceattle lineage throughout the section.

6.1 Seven Tier 3 Scenario Results

Catches for 2025 and 2026 are fixed at 1,350 thousand t in every alternative; the comparison therefore begins with 2027. Values in Table 13 are simulation means from 1,000 SPM simulations.

6.2 Alternative 2 Fixed-Catch Projection

This additional run applies Alternative 2 and fixes catch at 1,300 thousand t for 2025–2032. The projection horizon includes every fixed-catch year.

6.3 Projection Figures

The figures summarize the 1,000 simulations for each alternative. Lines show simulation medians and ribbons show central 90% intervals.

7 ESP-linked Recruitment Research

This work is an exploratory recruitment-model extension, separate from the assessment-model fit diagnostics. It connects Rceattle with biological hypotheses and cohort-aligned indicators developed in the ecosystem and socioeconomic profile (ESP). The dynamic structural equation modeling (DSEM) proof of concept used Rceattle 5.7.0 from the `dsem-v5-integration` branch and `dsem` version 3.0.0 in a project-specific R library. This arrangement preserves the established Rceattle 4.9.1 installation. The DSEM configuration changes recruitment deviations from penalized fixed effects to random effects (`random_rec = TRUE`) because the dynamic structural equation model treats recruitment deviations as a stochastic process.

Before introducing environmental paths, the experimental branch reproduced the established two-stage NonParametricPM objective exactly (713.6765). The compatibility fit had a positive-definite Hessian and a

Table 11: Key settings and source lineage for the Rceattle-derived SPM projection.

Generated SPM Setup and Lineage

Item	Value
Rceattle source fit	nonparametric_pm
Source checksum	650473177656b2c1e4122a9b3f0552be
Composition likelihood	MultinomialPM
Composition sample sizes	nominal, unadjusted
spmR version	0.3.0
spmR commit	e86fc6aa
Terminal assessment year	2024
Spawning timing	0.25 year from January 1 (start of April)
Rceattle encoding	spawn_month = 3 elapsed months; $3/12 = 0.25$
SPM encoding	Spawnmo = 4 as a one-based calendar-month code; $(4 - 1)/12 = 0.25$
Timing compatibility check	Matched: both calculations apply survival through 0.25 of the year
Spawning-biomass timing convention	Mature spawning biomass after survival through 0.25 of the year:
Projection begin year	2025
Projection years	14
Simulations	1000
Alternatives	1, 2, 3, 4, 5, 6, 7

Table 12: SPM projection alternatives included in the generated setup file.

SPM Projection Alternatives

Alternative	Name	Description
1	Maximum permissible ABC	Maximum permissible Tier 3 ABC harvest rate.
2	Author-specified ABC	Author-specified adjustment to the Tier 3 ABC harvest rate
3	Average recent F	Recent average fishing mortality from the Rceattle-derived i
4	Alternative SPR rate	User-specified SPR rate of 0.60.
5	No fishing	Zero fishing mortality after the fixed-catch years.
6	OFL threshold determination	OFL harvest rate for threshold and status-determination cal
7	Status-determination ramp	Maximum permissible ABC for three years followed by the 0

Table 13: Tier 3 projection results for all seven FMP alternatives using inputs generated from the Rceattle EBS pollock model. Catch and biomass quantities are in thousand t; $B/B_{35\%}$ is spawning biomass relative to the Tier 3 $B_{35\%}$ proxy.

Seven Tier 3 Projection Alternatives

Alt	Scenario	2027					B_1
		Catch_2027	ABC_2027	OFL_2027	SSB_2027	F_2027	
1	Maximum permissible ABC	1,691	1,691	2,042	2,391	0.350	
2	Author-specified ABC	1,691	1,691	2,042	2,391	0.350	
3	Average recent F	1,200	1,200	2,042	2,459	0.240	
4	Alternative SPR rate	786	786	2,042	2,513	0.150	
5	No fishing	0	0	2,042	2,608	0.000	
6	OFL threshold determination	1,605	1,605	1,605	2,074	0.390	
7	Status-determination ramp	1,514	1,806	1,806	2,241	0.330	

Table 14: Age-specific schedules generated from the terminal-year Rceattle fit and applied in SPM.

Projection Age-Specific Schedules

Age	Spawning_weight	Fishery_weight	Maturity	Natural_mortality	Selectivity
1	0.085	0.025	0.000	0.900	0.000
2	0.196	0.186	0.008	0.450	0.048
3	0.322	0.401	0.289	0.300	0.204
4	0.534	0.667	0.641	0.300	0.480
5	0.621	0.663	0.842	0.300	0.968
6	0.707	0.720	0.901	0.300	0.957
7	0.822	0.791	0.947	0.300	0.942
8	0.946	0.905	0.963	0.300	1.367
9	1.051	1.009	0.970	0.300	1.363
10	1.113	1.061	1.000	0.300	1.358
11	1.178	1.138	1.000	0.300	1.464
12	1.290	1.257	1.000	0.300	1.462
13	1.332	1.267	1.000	0.300	1.462
14	1.367	1.290	1.000	0.300	1.462
15	1.415	1.368	1.000	0.300	1.462

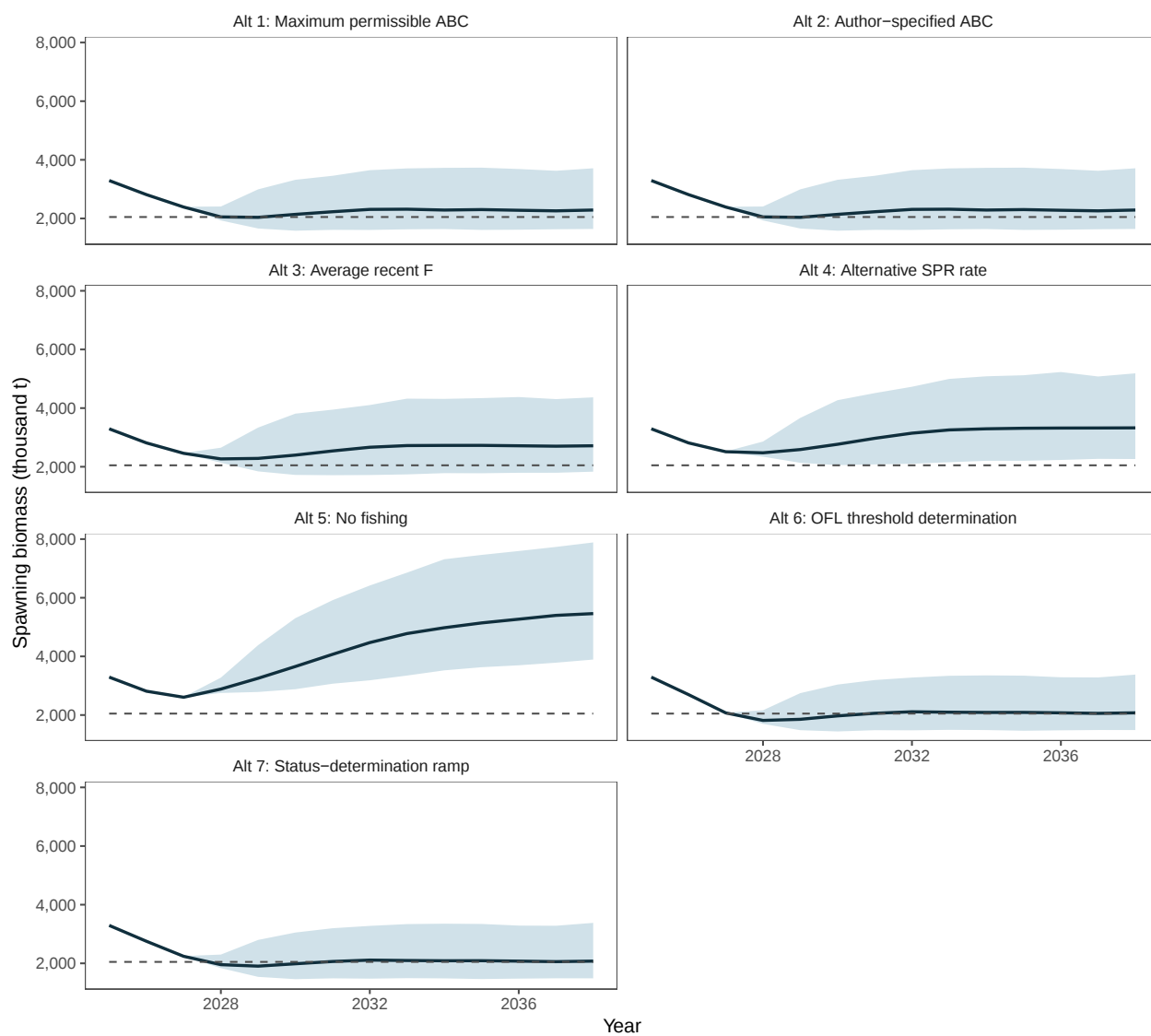


Figure 10: Projected spawning biomass under all seven Tier 3 alternatives. Lines are simulation medians, ribbons span the 5th to 95th percentiles, and dashed lines mark B35%.

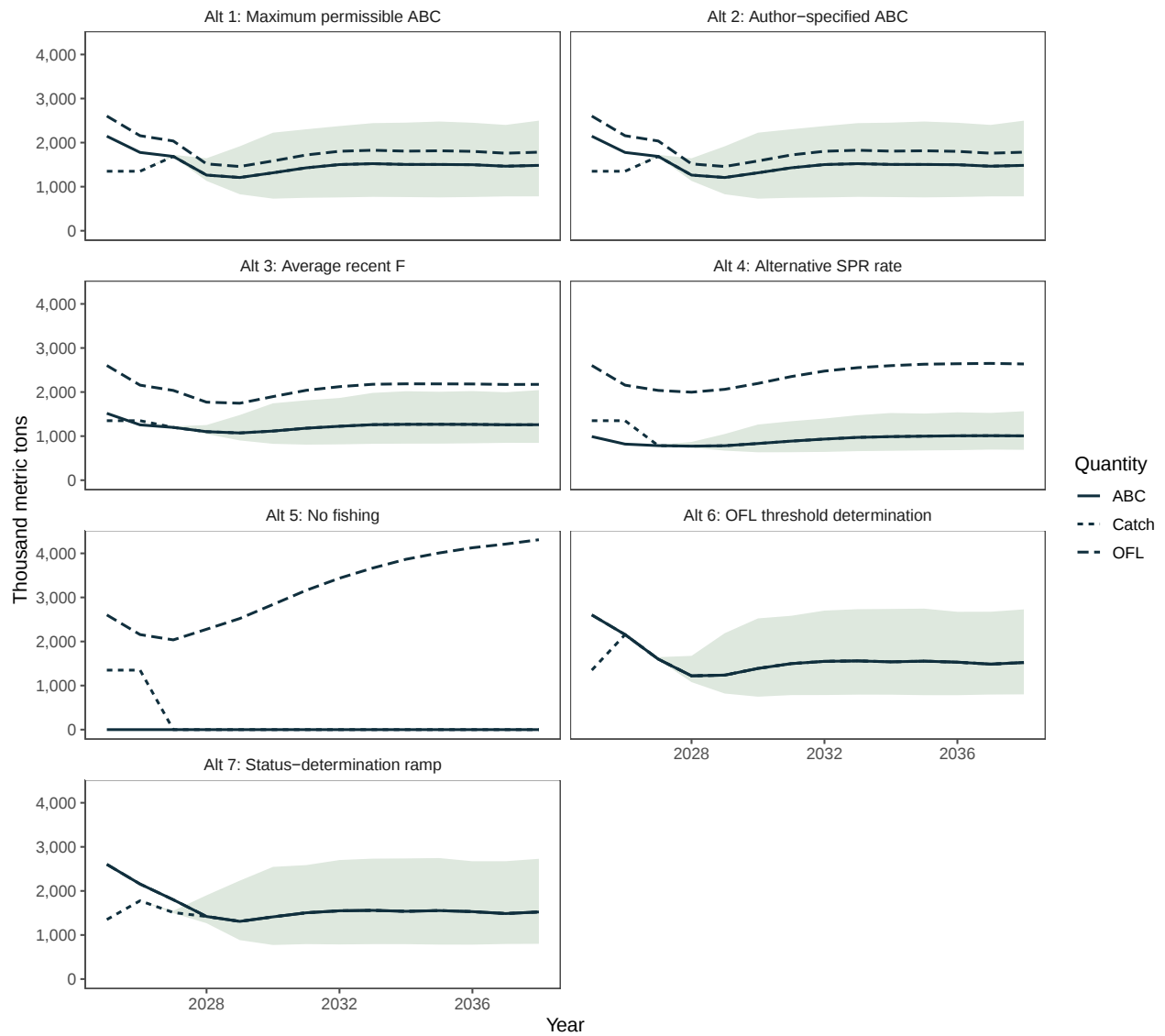


Figure 11: Projected catch, ABC, and OFL under all seven Tier 3 alternatives. Lines are simulation medians and catch ribbons span the 5th to 95th percentiles.

Table 15: Candidate ABC projection given assumed future catches. Catch is fixed at 1,300 thousand t; ABC is the maximum permissible ABC calculated by SPM. $B/B_{35\%}$ is mean spawning biomass relative to the Tier 3 proxy.

Candidate ABC Projection Given Assumed Future Catches

Year	Catch	ABC	OFL	Mean B	$B/B_{35\%}$
2025	1,300	2,146	2,603	3,300	161%
2026	1,300	1,789	2,171	2,840	139%
2027	1,300	1,716	2,076	2,484	121%
2028	1,300	1,504	1,805	2,282	111%
2029	1,300	1,407	1,703	2,278	111%
2030	1,300	1,440	1,748	2,350	115%
2031	1,300	1,486	1,814	2,441	119%
2032	1,300	1,531	1,876	2,530	123%

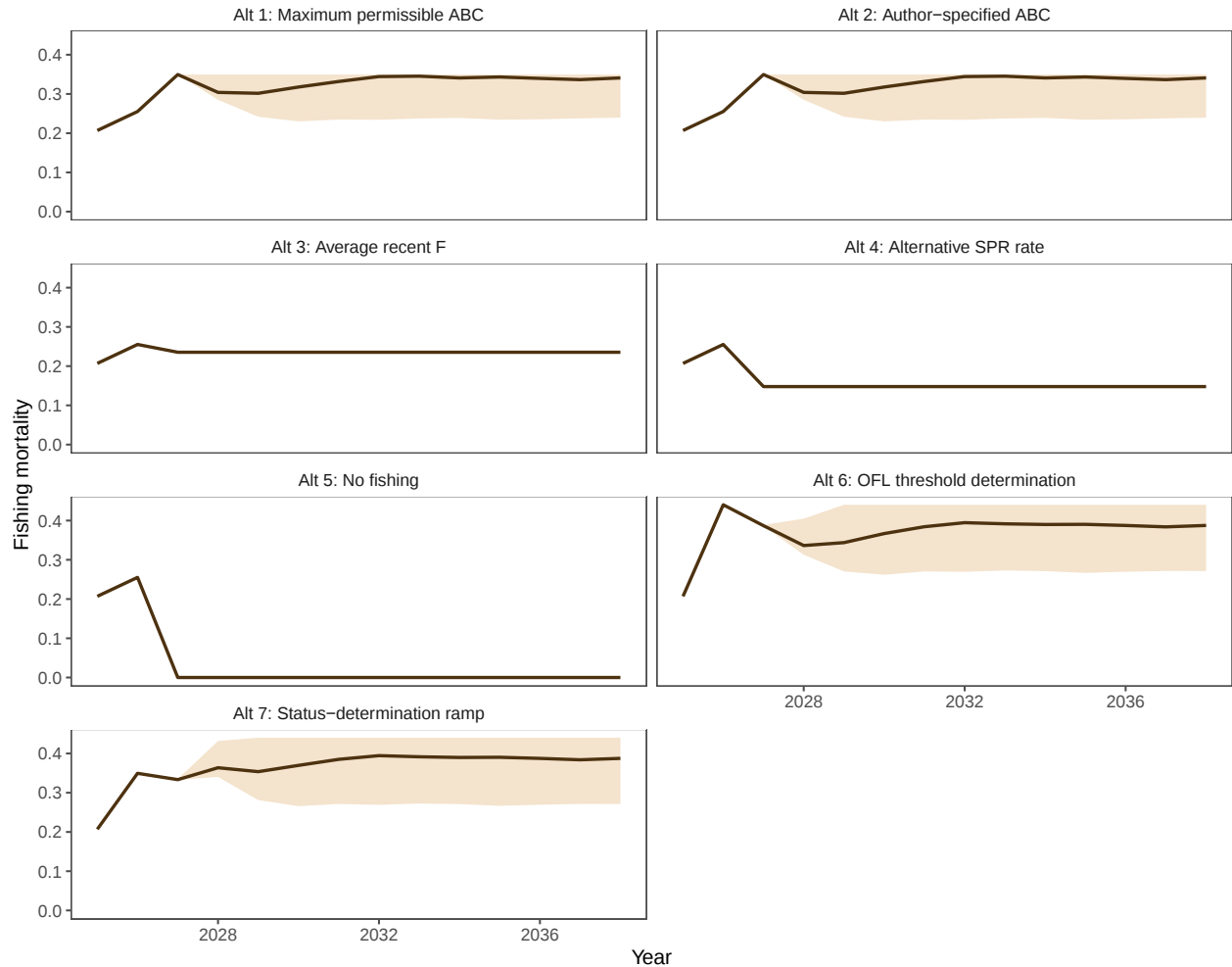


Figure 12: Projected fishing mortality under all seven Tier 3 alternatives. Lines are simulation medians and ribbons span the 5th to 95th percentiles.

Table 16: Mean SPM results by alternative and projection year. Biomass, catch, ABC, and OFL are in thousand t.

SPM Projection Means

Alt	Year	B	Catch	ABC	OFL	F	$B/B_{35\%}$
1	2025	3,293	1,350	2,146	2,603	0.207	161%
1	2026	2,813	1,350	1,777	2,156	0.255	137%
1	2027	2,391	1,691	1,691	2,042	0.349	117%
1	2028	2,099	1,311	1,311	1,576	0.309	102%
1	2029	2,143	1,281	1,281	1,552	0.303	105%
1	2030	2,261	1,381	1,381	1,676	0.308	110%
1	2031	2,349	1,455	1,455	1,767	0.314	115%
1	2032	2,416	1,498	1,498	1,820	0.317	118%
1	2033	2,453	1,532	1,532	1,861	0.319	120%
1	2034	2,447	1,534	1,534	1,862	0.318	119%
1	2035	2,439	1,526	1,526	1,853	0.319	119%
1	2036	2,427	1,520	1,520	1,844	0.318	118%
1	2037	2,423	1,515	1,515	1,839	0.318	118%
1	2038	2,438	1,520	1,520	1,846	0.317	119%
2	2025	3,293	1,350	2,146	2,603	0.207	161%
2	2026	2,813	1,350	1,777	2,156	0.255	137%
2	2027	2,391	1,691	1,691	2,042	0.349	117%
2	2028	2,099	1,311	1,311	1,576	0.309	102%
2	2029	2,143	1,281	1,281	1,552	0.303	105%
2	2030	2,261	1,381	1,381	1,676	0.308	110%
2	2031	2,349	1,455	1,455	1,767	0.314	115%
2	2032	2,416	1,498	1,498	1,820	0.317	118%
2	2033	2,453	1,532	1,532	1,861	0.319	120%
2	2034	2,447	1,534	1,534	1,862	0.318	119%
2	2035	2,439	1,526	1,526	1,853	0.319	119%
2	2036	2,427	1,520	1,520	1,844	0.318	118%
2	2037	2,423	1,515	1,515	1,839	0.318	118%
2	2038	2,438	1,520	1,520	1,846	0.317	119%
3	2025	3,293	1,350	1,517	2,603	0.207	161%
3	2026	2,813	1,350	1,256	2,156	0.255	137%
3	2027	2,459	1,200	1,200	2,042	0.235	120%
3	2028	2,317	1,120	1,120	1,819	0.235	113%
3	2029	2,397	1,117	1,117	1,811	0.235	117%
3	2030	2,540	1,183	1,183	1,937	0.235	124%
3	2031	2,662	1,236	1,236	2,046	0.235	130%
3	2032	2,766	1,271	1,271	2,118	0.235	135%
3	2033	2,836	1,310	1,310	2,190	0.235	138%
3	2034	2,856	1,326	1,326	2,220	0.235	139%
3	2035	2,863	1,327	1,327	2,225	0.235	140%
3	2036	2,859	1,329	1,329	2,228	0.235	140%
3	2037	2,857	1,330	1,330	2,229	0.235	139%
3	2038	2,873	1,333	1,333	2,234	0.235	140%
4	2025	3,293	1,350	²⁵ 990	2,603	0.207	161%
4	2026	2,813	1,350	819	2,156	0.255	137%
4	2027	2,513	786	786	2,042	0.148	123%
4	2028	2,525	784	784	2,020	0.148	123%

Table 17: Exploratory DSEM model comparison using the common cold-pool and next-year spawning-biomass proxy data set. Delta AIC is measured from the lowest AIC. All fits had positive-definite Hessians and maximum gradients below 0.0001. The WARN status reflects Hessian conditioning along the pre-existing selectivity-deviation direction.

Model configuration	AIC	Delta AIC	Maximum gradient	Positive-definite Hessian
IID	4,115.6	24.0	8.60×10^{-5}	TRUE
Cold pool	4,117.4	25.9	6.27×10^{-5}	TRUE
Next-year SSB proxy	4,091.6	0.0	4.15×10^{-5}	TRUE
Cold pool + next-year SSB proxy	4,093.2	1.6	5.28×10^{-5}	TRUE

maximum gradient of 4.4×10^{-5} . Its convergence summary carried the same Hessian-conditioning warning associated with a selectivity-deviation direction. This checkpoint supports use of the branch for the exploratory DSEM comparison while retaining the established fit as the assessment baseline.

The implementation should instead use the [ecosystem and socioeconomic profile \(ESP\) early-survival pathway synthesis](#) as the covariate-selection and data-assembly framework. That synthesis identifies two primary prerecruit filters for eastern Bering Sea pollock: an age-0 energy and condition filter before winter, and an age-1 to age-2 predation/overlap filter before age-3 recruitment. Transport and nursery-delivery processes remain useful supporting hypotheses where spawning-location, larval-transport, prey-match, and age-0 abundance products are available.

The first integrated test used two available age-1-year indicators: cold-pool extent and next-year spawning biomass as a provisional proxy for adult pollock biomass. Each series was standardized over its observed years and aligned with the assessment’s age-1 recruitment year. Both series occurred in every candidate model as latent first-order autoregressive processes with an observation-error standard deviation fixed at 0.1. The four configurations differed only in their paths to recruitment: IID recruitment, cold pool, adult biomass, and both paths together. This common-data design makes their Akaike information criterion (AIC) values comparable and follows the Structural Causal Enhanced Stock Assessment Modelling framework’s use of a common time-series data set across candidate structures ([Champagnat et al. 2026](#)).

The proxy distinction is important. The ESP cohort-table builder currently defines `adult_pollock_biomass_t1` as spawning biomass in the following year, `lead(SSB, 1)`. It therefore represents an assessment-derived spawning-biomass series rather than an independently calculated age-3+ biomass or a direct cannibalism index. The internal scenario name retains `adult_biomass` for reproducibility, while the tables and figures identify the quantity as the next-year spawning-biomass proxy.

The basic testing directed acyclic graph (DAG) in Figure 13 translates those pathways into a small set of cohort-aligned hypotheses. Arrows state the assumed direction of influence to be tested. Causal interpretation and the choice of simultaneously fitted paths require supporting analysis.

Figure 14 shows why the two persistence estimates differ. Cold-pool extent varies sharply among adjacent observations, whereas the spawning-biomass proxy remains above or below its mean for extended periods. The latter pattern is consistent with its estimated AR1 coefficient near 0.99 and motivates the predictive safeguards described below.

The next-year spawning-biomass-proxy model had the lowest AIC (Table 17; Figure 15). Its standardized path coefficient to recruitment was -0.358 (standard error 0.072; approximate 95% interval -0.500 to -0.217). The combined model estimated a similar proxy coefficient of -0.367 (standard error 0.074), while its cold-pool coefficient was -0.056 (standard error 0.095). The cold-pool-only coefficient was 0.040 (standard error 0.107). The spawning-biomass-proxy configurations reduced estimated unexplained recruitment variance by about 25% relative to IID recruitment; the cold-pool-only configuration preserved essentially the IID variance.

Figure 16 distinguishes estimated environmental persistence from recruitment effects. The negative spawning-biomass-proxy effect is consistent across its single-path and combined configurations. Both cold-pool

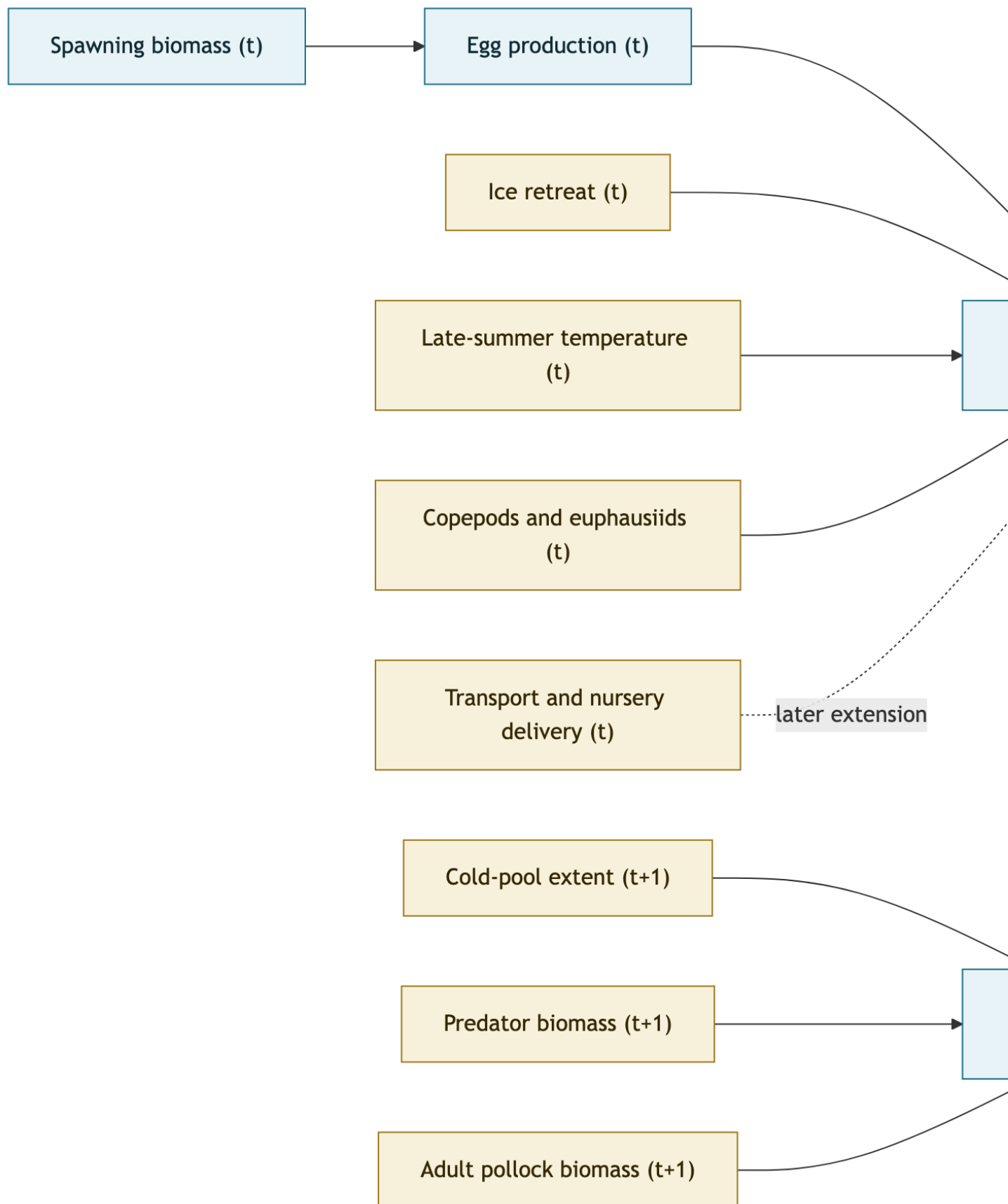
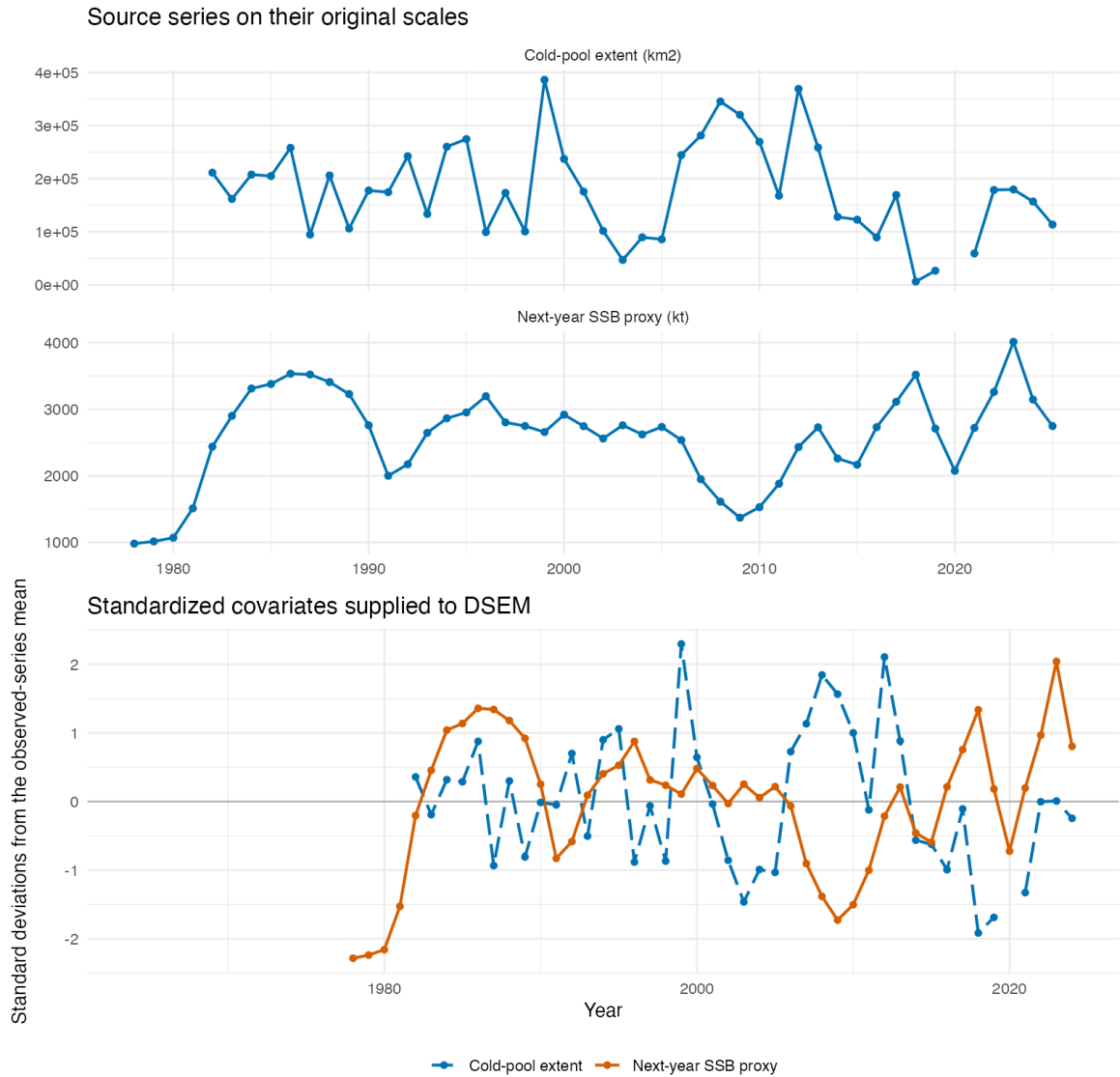


Figure 13: Basic cohort-aligned DAG for initial Rceattle DSEM recruitment tests. Blue boxes are stock states or biological processes, yellow boxes are candidate explanatory indicators (including physical conditions, prey, and predator or adult biomass), and the green box is the recruitment outcome. Subscripts identify the cohort's age-0 year (t), following age-1 year ($t+1$), and age-3 recruitment year ($t+3$). Solid pathways define the two initial prerecruit hypotheses; the dashed transport pathway is retained as a later extension because



Cold-pool observations: 42; next-year SSB-proxy observations: 47.

Figure 14: Covariates used in the exploratory DSEM tests. The upper panels show cold-pool extent and the next-year spawning-biomass proxy on their original scales. The lower panel shows the standardized values supplied to DSEM; zero is each observed series' mean and one unit is one observed-series standard deviation. Breaks mark missing annual observations. The model period contains 42 cold-pool and 47 spawning-biomass-proxy observations.

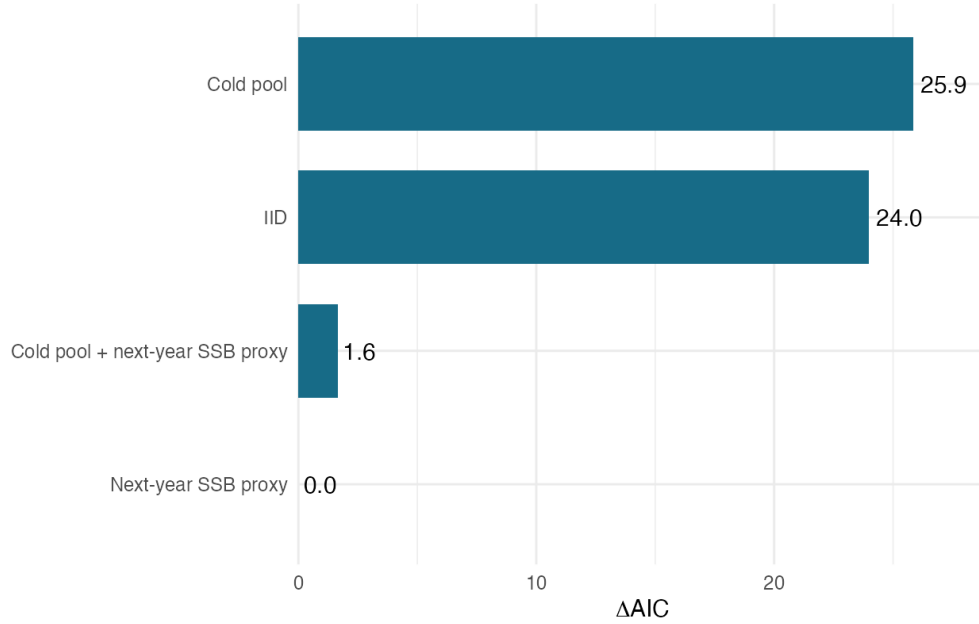


Figure 15: Relative AIC for the four exploratory DSEM recruitment configurations. Every model contains the same standardized cold-pool and next-year spawning-biomass proxy series; bars therefore measure support for the specified paths to recruitment under this common data treatment.

recruitment intervals span zero. The proxy’s autoregressive coefficient is approximately 0.99, meaning adjacent annual values are nearly unchanged relative to the full variation in the series. Such a persistent series can explain historical recruitment variation by tracking stock size, time, or another shared process. Leave-future-out tests provide a stronger predictive check: a model is fitted through a given year and evaluated on recruitment in a later withheld year using only previously available information. Improved withheld-year predictions would establish practical predictive value. They would still provide limited evidence for cannibalism because causal interpretation additionally requires an appropriate independent measure of adult biomass or predation, correct biological timing, and adequate control of shared assessment inputs and confounding. This separation among historical fit, predictive skill, and causal interpretation follows the evaluation sequence emphasized by Champagnat et al. (2026).

7.1 Unit-normal randomization check

A fixed-seed randomization check replaced both observed covariate series with independent draws from a standard normal distribution while preserving each series’ missing-year pattern. The check used seed 20260812 and repeated the same four DSEM structures, population model, estimation settings, and NonParametricPM selectivity configuration. This design asks whether unrelated annual covariates can receive apparently strong recruitment effects through the fitting procedure alone.

The randomized cold-pool substitute had a recruitment coefficient of -0.08 (95% interval -0.25 to 0.08), and the randomized spawning-biomass substitute had a coefficient of -0.11 (-0.31 to 0.08). Their combined-model intervals also included zero. Every random-covariate configuration had a higher AIC than randomized IID recruitment, with differences of 0.7–1.4 units. Estimated unexplained recruitment variance declined by only 2–6%. Together, these results show a much weaker pattern than the strong negative coefficient and 24-unit AIC improvement found for the observed spawning-biomass proxy.

This single fixed-seed realization is a procedural check. Repeated randomizations would form an empirical null distribution for the largest coefficient, AIC improvement, and variance reduction expected from chance covariates.

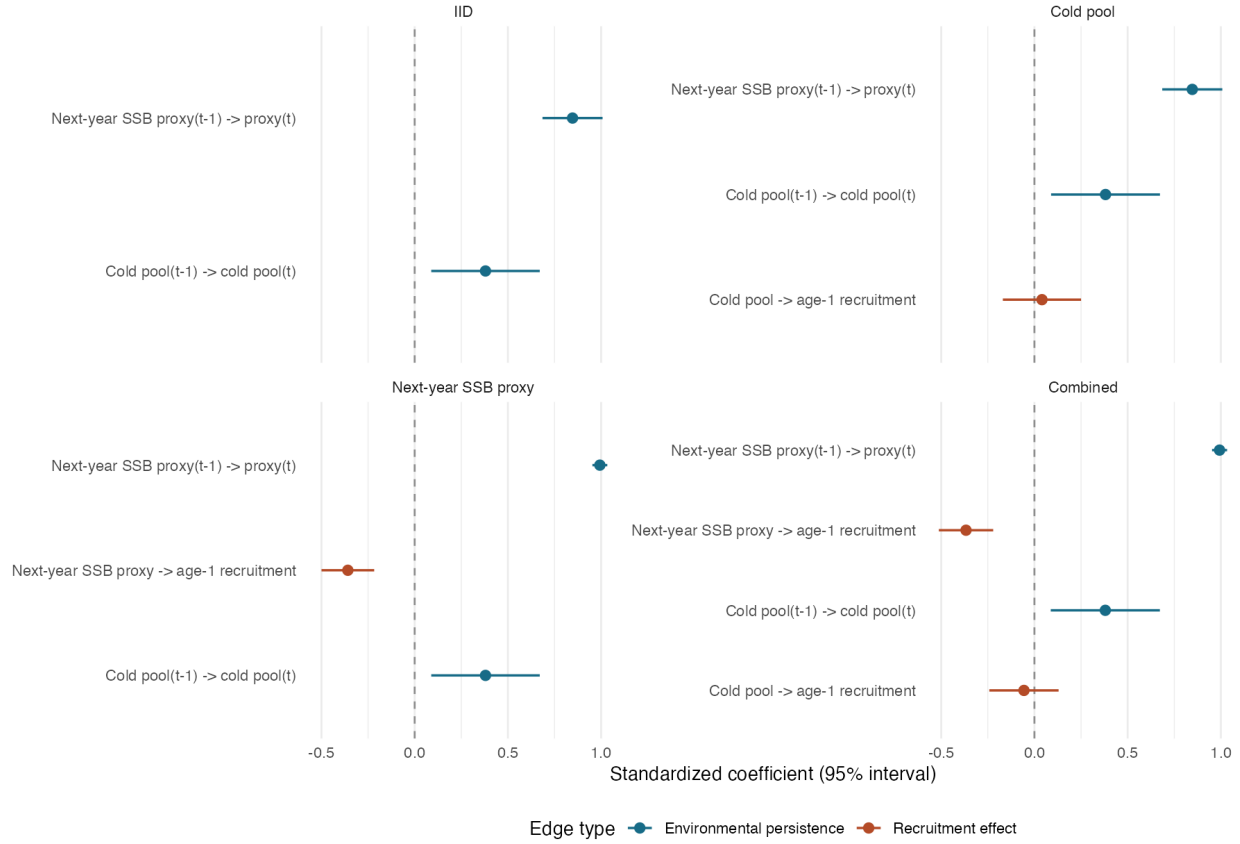


Figure 16: Standardized coefficients for every directed edge estimated in the four exploratory DSEM configurations. Points are maximum-likelihood estimates and horizontal lines are approximate 95% Wald intervals. Blue edges describe first-order environmental persistence; orange edges describe contemporaneous effects on age-1 recruitment. Variance and observation-scale parameters are excluded because they are scales rather than directed effects.

Table 18: DSEM randomization check using two independent unit-normal covariate series. Delta AIC is measured from the randomized IID model, so positive values indicate weaker support than IID. Coefficient intervals are approximate 95% Wald intervals.

Model configuration	AIC	Delta AIC from IID	Recruitment coefficient (95% interval)
IID	4,184.7	0.0	—
Random cold-pool substitute	4,185.7	1.0	cold-pool substitute: -0.08 (-0.25, 0.08)
Random SSB-proxy substitute	4,185.4	0.7	SSB-proxy substitute: -0.11 (-0.31, 0.08)
Both random substitutes	4,186.1	1.4	cold-pool substitute: -0.10 (-0.26, 0.07); SSB-proxy substitute: -0.11 (-0.31, 0.08)

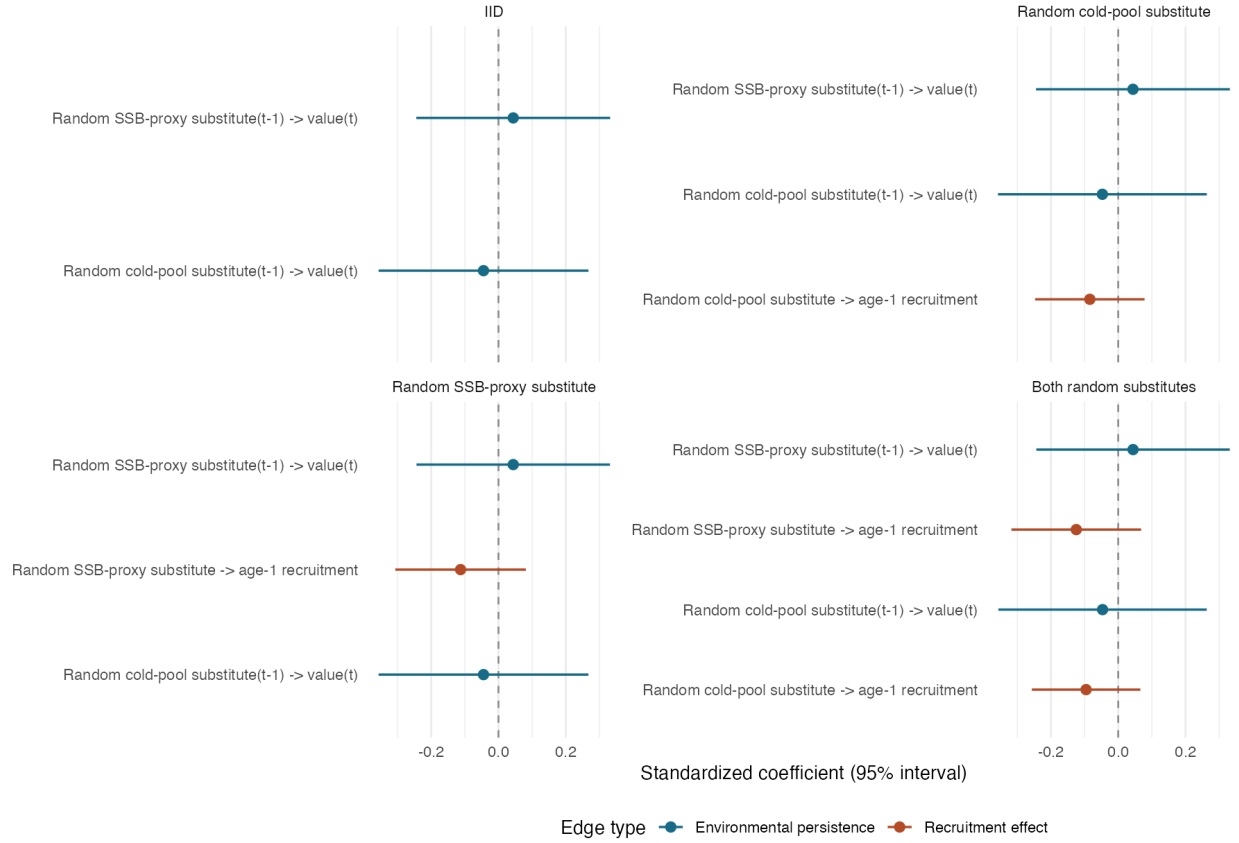


Figure 17: Standardized coefficients from the unit-normal DSEM randomization check. The two candidate covariate series are independent standard-normal draws that preserve the observed missing-year patterns. Points are maximum-likelihood estimates and horizontal lines are approximate 95% Wald intervals. Both randomized recruitment-effect intervals include zero.

7.2 Interpretation relative to recruitment per spawner

The negative proxy coefficient is consistent with compensatory density dependence. A plausible biological mechanism is cannibalism, matching the adult-pollock-biomass to juvenile-predator-overlap pathway in Figure 13. The current analysis supports the broader density-dependent pattern and treats cannibalism as a mechanism for further testing.

This result connects closely with the assessment’s recruitment-per-spawner diagnostic. That diagnostic is

$$\log(R_t/SSB_t) = \log(R_t) - \log(SSB_t).$$

Its decline with spawning biomass shows lower recruitment per spawner at higher stock size, a familiar signature of compensation. The DSEM models recruitment deviations directly, so it avoids placing spawning biomass in the response denominator. Its predictor, however, is next-year spawning biomass and remains closely related to parental spawning biomass through stock persistence. In the available ESP cohort table, parental spawning biomass and the next-year proxy have a correlation of 0.84. Logged recruits per spawner correlate -0.70 with parental spawning biomass and -0.72 with the next-year proxy. These similar relationships indicate that the DSEM coefficient and the recruitment-per-spawner figure largely describe the same compensatory stock-recruit signal.

The ratio diagnostic also contains mathematical coupling because spawning biomass occurs on both the horizontal axis and in the response denominator. Consequently, its negative slope alone provides incomplete evidence for a biological mechanism. The DSEM result strengthens evidence for compensation by using recruitment deviations as the response, while the shared assessment origin and temporal persistence of the predictor still limit independent causal interpretation.

A stronger cannibalism test would replace the proxy with independently calculated age-3+ biomass, vulnerable adult biomass, or a predation index that combines predator abundance, diet, ration, and juvenile overlap. A targeted comparison should include parental spawning biomass alone, adult biomass alone, both predictors, and adult biomass mediated through juvenile-adult overlap. That comparison can distinguish the general stock-recruit relationship from an added cannibalism pathway.

Each integrated fit required approximately 2.5–2.7 minutes for the Laplace-integrated hindcast on one processor. All four fits produced positive-definite Hessians and maximum gradients below 0.0001. The near-unit spawning-biomass-proxy autoregressive coefficient (approximately 0.993), the shared trends in the proxy and recruitment, and the exploratory proxy status warrant predictive checks before biological interpretation. Leave-future-out prediction and sensitivities to the adult-biomass definition provide the next evaluation step.

The Gulf of Alaska pollock application in Champagnat et al. (2026) provides the closest assessment-integrated precedent. That study compared IID, autoregressive, single-covariate, and causal recruitment structures using fit, parsimony, unexplained recruitment variance, and one-year-ahead forecast performance. The present eastern Bering Sea results represent an initial structural and numerical test; predictive validation remains the next checkpoint.

For a first DSEM evaluation, the two branches in Figure 13 should be fitted separately before considering the reduced integrated-gate model. This makes it possible to distinguish whether any improvement comes from the age-0 condition hypothesis, the age-1/2 overlap hypothesis, or an unsupported combination of correlated indicators.

The practical implementation started from a cohort-aligned input table with one row per year class. The ESP workflow uses suffixes to make timing explicit: `_t` for the cohort’s age-0/spawning year, `_t1` for the following age-1 year, and `_t3` for age-3 recruitment. Candidate fields include spawning biomass, ice retreat, late-summer SST, wind mixing, copepod/euphausiid biomass, age-0 energy, cold-pool extent, age-1 abundance, adult pollock biomass, arrowtooth biomass, juvenile-predator overlap, and age-3 recruitment. For the Rceattle script, the age-1-year columns replaced the earlier placeholder `env_data` values, and the modeled response is the assessment’s age-1 recruitment deviation. This timing choice follows the integrated assessment process while preserving the ESP table’s explicit cohort alignment.

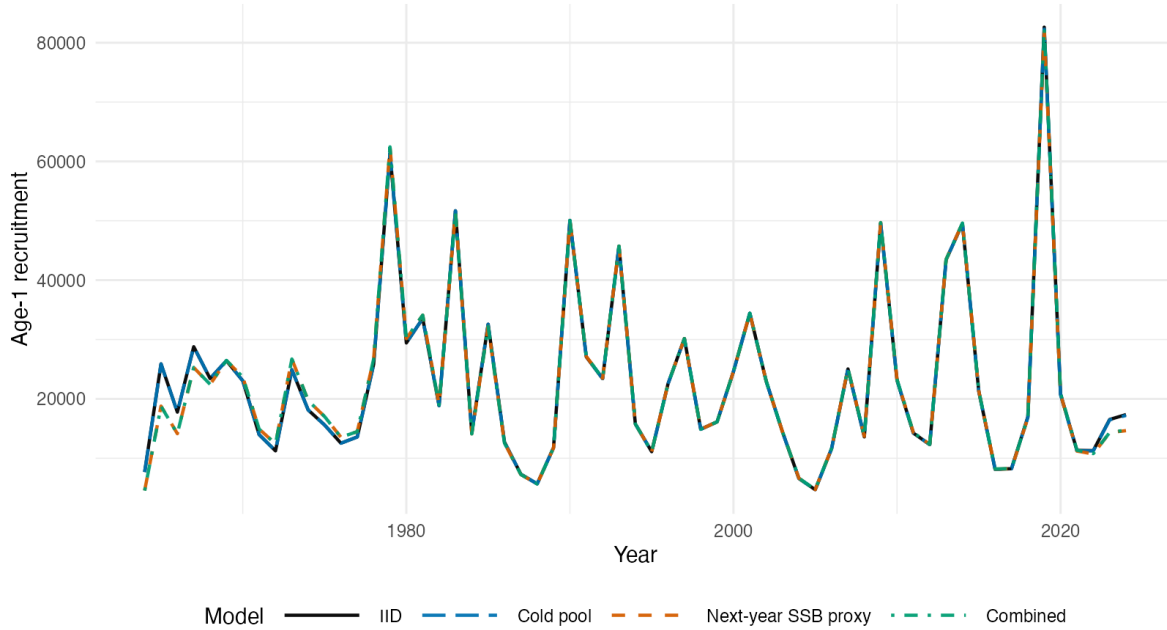


Figure 18: Estimated age-1 recruitment trajectories from the four exploratory DSEM configurations. Models use the same assessment observations, population dynamics, NonParametricPM selectivity, cold-pool series, and next-year spawning-biomass proxy; differences arise from the specified recruitment process.

The proof of concept completed the IID, cold-pool, adult-biomass, and combined comparisons. The next hypothesis-specific sequence is:

1. evaluate leave-future-out recruitment predictions for the four completed models;
2. fit an age-0 condition model using the best-supported energy or condition indicator;
3. extend the predator hypothesis with predator biomass and a juvenile-predator overlap index;
4. fit a reduced integrated-gate model with one representative age-0 condition indicator and one representative predation/overlap indicator; and
5. reserve the transport/nursery-delivery model until the required spawning-location, particle-tracking, prey-match, and age-0 abundance products are available.

Each DSEM result should be compared with the IID baseline using the Akaike information criterion (AIC), standardized path coefficients, convergence status, and sensitivity to alternative proxies and influential year classes. This keeps the DSEM work tied to the ESP biological hypotheses while avoiding an unrestricted environmental screen with weakly identified annual series.

8 Summary

The corrected Rceattle lane improves statistical coherence by retaining complete ages 1–15, using nominal composition sample sizes, and separating the canonical PM stabilized criterion from a standard multinomial sensitivity. The fixed-parameter bridge reproduces corrected `pm.tp1` spawning biomass and predicted catch at numerical precision, which validates the tested population dynamics and parameter mapping. Independently estimated trajectories show larger differences and therefore describe model estimation structure rather than a software-translation error.

The diagnostic results identify a focused development path for the September 2026 Plan Team evaluation. The canonical PM fit reaches a small gradient, while Hessian conditioning isolates weak identification in annual selectivity deviations. Composition OSA residuals remain underdispersed. Nine terminal-year-consistent peels converge successfully and produce positive retrospective bias in spawning biomass, total biomass, and

recruitment, together with negative bias in fishing mortality. The historical 2D AR1 sensitivity retains its research status. The sparse-metric posterior run has zero divergences, while its maximum R-hat of 1.299 and minimum bulk effective sample size of 20 show incomplete mixing in survey-selectivity parameters. The posterior therefore remains exploratory and fails the convergence gate for assessment inference.

Immediate work should retain the corrected ADMB bridge as a regression gate, evaluate selectivity and survey influence on the retrospective pattern, compare the canonical PM and standard multinomial trajectories, and refine the annual-selectivity parameterization. The next-assessment workbook should be populated after complete new-year observations arrive. Recruitment research should continue with documented, cohort-aligned ecosystem and socioeconomic profile covariates and explicit out-of-sample evaluation.

9 Reproducibility

R version 4.6.1 (2026-06-24)

Platform: aarch64-apple-darwin23

Running under: macOS Tahoe 26.6.1

Matrix products: default

BLAS: /Library/Frameworks/R.framework/Versions/4.6/Resources/lib/libRblas.0.dylib

LAPACK: /Library/Frameworks/R.framework/Versions/4.6/Resources/lib/libRlapack.dylib; LAPACK version 3.

locale:

[1] C.UTF-8/C.UTF-8/C.UTF-8/C/C.UTF-8/C.UTF-8

time zone: America/Los_Angeles

tzcode source: internal

attached base packages:

[1] stats graphics grDevices utils datasets methods base

other attached packages:

[1] tidyr_1.3.2	tibble_3.3.1	scales_1.4.0
[4] readr_2.2.0	Rceattle_4.9.1.9001	posterior_1.7.0
[7] Matrix_1.7-5	knitr_1.51	gt_1.3.0
[10] ggthemes_5.2.0	ggplot2_4.0.3	ggridges_0.5.7
[13] dplyr_1.2.1	asar_2.5.0.9000	afscOSA_0.0.1

loaded via namespace only:

[1] gtable_0.3.6	tensorA_0.36.2.1	TMB_1.9.21
[4] xfun_0.60	GGally_2.4.0	lattice_0.22-9
[7] tzdb_0.5.0	vctrs_0.7.3	tools_4.6.1
[10] generics_0.1.4	parallel_4.6.1	pkgconfig_2.0.3
[13] checkmate_2.3.4	RColorBrewer_1.1-3	S7_0.2.2
[16] distributional_0.8.1	lifecycle_1.0.5	compiler_4.6.1
[19] farver_2.1.2	stringr_1.6.0	tinytex_0.60
[22] litedown_0.9	htmltools_0.5.9	yaml_2.3.12
[25] pillar_1.11.1	crayon_1.5.3	abind_1.4-8
[28] commonmark_2.0.0	ggstats_0.13.0	tidyselect_1.2.1
[31] digest_0.6.39	stringi_1.8.7	reshape2_1.4.5
[34] purrr_1.2.2	labeling_0.4.3	cowplot_1.2.0
[37] fastmap_1.2.0	grid_4.6.1	cli_3.6.6
[40] magrittr_2.0.5	withr_3.0.3	backports_1.5.1
[43] bit64_4.8.2	rmarkdown_2.31	bit_4.6.0
[46] otel_0.2.0	hms_1.1.4	evaluate_1.0.5

Table 19: Files used or referenced by this working paper.

Role	Location	Available	Size
SparseNUTS companion report	Local HyperText Markup Language (HTML) file	TRUE	8610.8 KB
Prepared model-input workbook	Local workbook	TRUE	128.6 KB
BTS covariance input	Local data file	TRUE	12.7 KB
Saved fitted Rceattle objects	Local R serialized-data (RDS) file	TRUE	87955.2 KB
OSA residuals	Local RDS file	TRUE	89.8 KB
OSA diagnostics	Local RDS file	TRUE	1.0 KB
Nine-peel Rceattle retrospective	Local RDS file	TRUE	368274.4 KB
Nine-peel Mohn's rho summary	Local CSV file	TRUE	0.2 KB
SparseNUTS fit	Local RDS file	TRUE	156286.8 KB
SparseNUTS sampler diagnostics	Local CSV file	TRUE	0.1 KB
SparseNUTS parameter summary	Local CSV file	TRUE	213.6 KB
SparseNUTS run metadata	Local RDS file	TRUE	1.2 KB
Nine-peel two-stage retrospective script	Local R script	TRUE	8.0 KB
Fixed-parameter comparison output	Local RDS file	TRUE	646.1 KB
Fixed-parameter comparison summary	Local comma-separated-values (CSV) file	TRUE	0.3 KB
Two-stage fit script	Local R script	TRUE	16.3 KB
Diagnostics script	Local R script	TRUE	4.9 KB
Next-assessment update script	Local R script	TRUE	2.6 KB
DSEM example script	Local R script	TRUE	15.8 KB
ADMB comparison script	Local R script	TRUE	10.0 KB
Modified ADMB retrospective file	Local ADMB output	TRUE	0.0 KB
Legacy ADMB retrospective file	Local ADMB output	TRUE	0.0 KB

[49]	viridisLite_0.4.3	markdown_2.0	rlang_1.3.0
[52]	Rcpp_1.1.2	glue_1.8.1	xml2_1.6.0
[55]	vroom_1.7.1	jsonlite_2.0.0	R6_2.6.1
[58]	plyr_1.8.9	fs_2.1.0	

References

Champagnat, Juliette, Cole C. Monnahan, Jane Y. Sullivan, et al. 2026. “Causal Models as a Scientific Framework for Next-Generation Ecosystem and Climate-Linked Stock Assessments.” *Fish and Fisheries*, ahead of print. <https://doi.org/10.1111/faf.70092>.

Ianelli, Jim. 2026. *spmR: Standard Projection Model*. <http://afsc-assessments.github.io/spmR/>.

Appendix

Acronyms

This appendix follows the [Alaska Stock Assessment Report \(ASAR\) package](#) glossary convention. The project glossary uses ASAR’s `\newacronym` format and adds terms specific to this Rceattle comparison. The quoted label “PM” is a model name.

Table 20: Acronyms used in this working paper.

Acronym	Definition
ADMB	AD Model Builder
AIC	Akaike information criterion
AR1	first-order autoregressive
ATS	acoustic-trawl survey
AVO	acoustic vessels of opportunity
BTS	bottom trawl survey
CIE	Center for Independent Experts
CPUE	catch per unit effort
CSV	comma-separated values
DAG	directed acyclic graph
DSEM	dynamic structural equation modeling
ESP	ecosystem and socioeconomic profile
HTML	HyperText Markup Language
IID	independent and identically distributed
MCMC	Markov chain Monte Carlo
MVN	multivariate normal
OSA	one-step-ahead
RDS	R serialized data
RTMB	R Template Model Builder
SAFE	Stock Assessment and Fishery Evaluation
SDNR	standard deviation of normalized residuals
SEM	structural equation model
SPM	Standard Projection Model
SST	sea surface temperature