SCIENTIFIC COMMITTEE
TWENTY-SECOND REGULAR SESSION
11-19 August 2026
Apia, Samoa (Hybrid)

OPAL: THE OPEN POPULATION ASSESSMENT LIBRARY

WCPFC-SC22-2026-SA-WP02
11 July 2026

Submitted by

N. Ducharme-Barth[1], D. Webber[2], P. Neubauer[3], K. Kim[4] ,A. Magnusson,P. Hamer

[1] NOAA Fisheries, Pacific Islands Fisheries Science Center, 1845 Wasp Boulevard, Bldg. 176, Honolulu HI 96818
[2] Quantifish Limited, Tauranga, New Zealand
[3] Dragonfly Data Science, Wellington, New Zealand
[4] Oceanic Fisheries Programme, Pacific Community, Noumea, New Caledonia

Table of contents

Executive summary

This paper describes `opal` (the **o**pen **p**opulation **a**ssessment **l**ibrary), an open-source, modular R package for fisheries stock assessment, and illustrates its application through two case studies. Development of `opal` is aligned with and supported by *WCPFC Project 123: Scoping the next generation of tuna stock assessment software*, and the current working paper is intended to document a proof of concept for R-based, modular assessment software rather than a finished platform for producing management advice.

`opal` is built on RTMB, which provides automatic differentiation of the objective function, gradient-based optimization, and the Laplace approximation for models with random effects. Writing the model in R keeps the full model definition within a single language, lowering the barrier to inspecting and extending the code. The package builds on the `sbt` package developed for the assessment of southern bluefin tuna, and draws design influences from several existing assessment packages including Stock Synthesis (SS3) and MULTIFAN-CL (MFCL). It estimates age-structured population dynamics with length-based processes; fits CPUE together with length- and weight-composition data under a choice of likelihoods; supports random-effects estimation, catch-based projections, and Bayesian inference via SparseNUTS; and includes a set of built-in diagnostics, comprising estimability and parameter-correlation checks, simulation self-testing, posterior-predictive checks, and one-step-ahead (OSA) residuals.

Two case studies were developed to showcase model features, capabilities and available diagnostics. The first case study applied `opal` to simulated data for Hawai'i 'opakapaka (*Pristipomoides filamentosus*), for which the true dynamics are known from an SS3 operating model. `opal` closely tracked the SS3 spawning-biomass trajectory and the known simulated truth. This case study also illustrated the random-effects, projection, and Bayesian workflows, though the random-effects model was substantially slower than its fixed-effects counterpart. The second case study applied `opal` to real data from the 2020 WCPO bigeye tuna (*Thunnus obesus*) assessment, in a fleets-as-areas configuration, and compared its outputs against the SS3 and MFCL reference models. `opal` closely tracked the spawning-biomass and depletion trajectories of both platforms, though the absolute spawning-biomass scale differed, consistent with differences in selectivity parameterization, composition weighting, and catch-removal calculations across the three models.

Both case studies are illustrative and are not intended to inform management advice. Together they demonstrate that `opal` can estimate population dynamics, accommodate the data volume and dimensionality

of a real tuna assessment, and readily produce diagnostic outputs. They also identify outstanding challenges, including the robustness of optimization from naive starting values, the scaling of random effects to higher-dimensional models, and ongoing code optimization.

We invite SC22 to:

- Note the development of `opal` as a proof of concept for R-based, modular assessment software, and its application to two case studies.
- Comment on the suitability of `opal` for assessments of various species (e.g., tuna, shark and billfish) within the WCPFC.
- Endorse the continued development of `opal` as a core development stream for the next generation of tuna assessment software.

1 Introduction

`opal` — the **o**pen **p**opulation **a**ssessment **l**ibrary — is an open-source, modular R package for fisheries stock assessment. The package is built on RTMB (Kristensen, 2023), an R interface to the Template Model Builder framework (Kristensen et al., 2016), which provides automatic differentiation of the model objective function and supports gradient-based optimization and the Laplace approximation for models that include random effects. Writing the model in R rather than C++ keeps the entire model definition within a single language environment, which lowers the barrier to inspecting, modifying, and extending the underlying code.

`opal` builds directly upon the `sbt` R package developed by Quantifish for the assessment of southern bluefin tuna under the Commission for the Conservation of Southern Bluefin Tuna (CCSBT; Webber et al., 2026), and draws on design influences from a number of existing assessment platforms, including the Stochastic Population over Regional Components model (SPoRC; Cheng et al., 2026), the Woods Hole Assessment Model (Miller & Stock, 2020; WHAM; Stock & Miller, 2021), Stock Synthesis (SS3; Methot & Wetzel, 2013), CASAL2 (Casal2 Development Team, 2026), and MULTIFAN-CL (MFCL; Fournier et al., 1998). The development of `opal` is aligned with and supported by *WCPFC Project 123: Scoping the next generation of tuna stock assessment software*. It provides a proof of concept for the R-based modular software development and takes significant steps towards developing core modules for next generation tuna assessment software.

The intent of `opal` is to provide assessment scientists with a flexible modelling framework that supports the estimation of random effects; facilitates enhanced model diagnostics, including one-step-ahead (OSA) residuals, simulation self-testing, and posterior predictive checks; and integrates directly into a Bayesian estimation framework. These capabilities are described in Section 4.

The package source code, function documentation, and worked examples are publicly available (website, code repository, & documentation repository), and the package is distributed under an open-source license. This paper describes the structure of the model (Section 2), its principal features (Section 4), and two case studies that illustrate its application to a simulated single-species assessment and a real-data tuna assessment (Section 5).

2 Model structure

`opal` estimates age-structured population dynamics and accommodates age- or length-based processes. The population is tracked as numbers-at-age. Biological quantities — weight, maturity, fecundity, natural mortality, and sex ratio — may be supplied on either an age basis or a length basis: a vector defined over age classes is used directly, while a vector defined over length bins is converted to age through a probability-of-length-at-age (PLA) matrix. Selectivity is likewise defined on a length basis and mapped to age through the PLA. This structure allows population demographics to be specified in terms of either age or length, and allows the model to fit multiple data types: relative abundance indices (CPUE), length composition, and weight composition. Where time-varying or hierarchically structured processes are required, parameters may be specified as random effects and integrated out using the Laplace approximation provided by RTMB (Kristensen et al., 2016).

The model is implemented as a single objective function, `opal_model()`, that returns the negative log-likelihood (NLL) for a given set of parameters and data. Rather than encoding the population dynamics and observation model as a single monolithic routine, `opal_model()` orchestrates a sequence of standalone, individually documented module functions, each responsible for one component of the calculation. Population-model components — growth, selectivity, recruitment, and the data likelihoods — are implemented as separate exported functions rather than embedded in a single routine, so that components can be modified, replaced, or extended independently, and so that individual modules can be tested in isolation. The collection of module functions is registered through a single accessor for use when the model is passed to the Bayesian sampler. This modular organization is intended to make the codebase tractable to maintain, human readable and to lower the barrier to community contribution.

The principal modules, in the order they are evaluated, are: mean length-at-age (`get_growth()`); the standard deviation of length-at-age (`get_sd_at_age()`); the PLA matrix (`get_pla()`); weight-, maturity-, and fecundity-at-age (`get_weight_at_length()`, `get_maturity_at_age()`, and a resolver, `resolve_bio_vector()`, that accepts either age- or length-basis inputs); selectivity-at-age (`get_selectivity()`, with the parametric forms `sel_logistic()` and `sel_double_normal()`); initial equilibrium numbers-at-age (`get_initial_numbers()`); the forward population dynamics and harvest (the `do_dynamics()` and `get_harvest_rate()`); recruitment (`get_recruitment()`); the data likelihoods (`get_cpue_like()`, `get_length_like()`, `get_weight_like()`); and the prior and penalty terms (`get_recruitment_prior()`, `evaluate_priors()`).

2.1 Population dynamics

The population is structured by year, season, and age, and may be configured for any number of fisheries.

2.1.1 Growth and the probability of length-at-age

Mean length-at-age, μ_a, is computed from a Schnute parameterization of the von Bertalanffy growth function, anchored by the lengths L_1 and L_2 at two reference ages A_1 and A_2, which matches the SS3 growth pattern with a linearly interpolated coefficient of variation (Methot & Wetzel, 2013):

$$\mu_a = L_1 + (L_2 - L_1)\,\frac{1 - e^{-k(a-A_1)}}{1 - e^{-k(A_2-A_1)}}.$$

Module: `get_growth()`

The coefficient of variation is interpolated linearly with mean length between the two reference ages, and the standard deviation of length-at-age is its product with the mean:

$$\mathrm{CV}_a = \mathrm{CV}_1 + \frac{\mu_a - L_1}{L_2 - L_1}\,(\mathrm{CV}_2 - \mathrm{CV}_1), \qquad \sigma_a = \mu_a\,\mathrm{CV}_a.$$

Module: `get_sd_at_age()`

From μ_a and σ_a, the PLA matrix gives the probability that an individual of age a falls within length bin ℓ, assuming length-at-age is normally distributed. For contiguous bins with edges $\{e_\ell\}$, this is the difference of the normal cumulative distribution function Φ evaluated at the bin edges:

$$P_{\ell a} = \Phi\left(\frac{e_{\ell+1} - \mu_a}{\sigma_a}\right) - \Phi\left(\frac{e_\ell - \mu_a}{\sigma_a}\right), \qquad \sum_\ell P_{\ell a} = 1.$$

Module: `get_pla()`

The PLA is computed once per function evaluation and reused for the conversion of weight, maturity, fecundity, natural mortality, and selectivity between the length and age bases.

2.1.2 Biology and spawning potential

Quantities defined on a length basis are mapped to age through the PLA. For a length-basis vector x_ℓ, the age-basis vector is $x_a = \sum_\ell P_{\ell a}\,x_\ell$ (equivalently $\mathbf{x}_a = \mathbf{P}^\top \mathbf{x}_\ell$); a vector supplied on an age basis is used directly. Weight-at-length follows an allometric length-weight relationship, $w_\ell = \alpha_{\mathrm{LW}}\,m_\ell^{\beta_{\mathrm{LW}}}$, where m_ℓ is the bin midpoint, and is converted to weight-at-age through the PLA. Spawning potential at age is the product of sex ratio (the fraction mature-female-equivalent at age), maturity, and fecundity:

$$\psi_a = r_a\,m_a\,f_a,$$

Modules: `get_weight_at_length()`, `get_maturity_at_age()`, `resolve_bio_vector()`

where r_a, m_a, and f_a are sex ratio, maturity, and fecundity at age, respectively. Alternatively, a pre-computed spawning-potential vector may be supplied directly. All derived biology arrays are passed explicitly to the dynamics routine rather than read from the data object, so that automatic-differentiation gradients propagate correctly if any of these quantities are estimated (for example, growth parameters estimated through the PLA, or natural mortality specified as a function of length).

2.1.3 Initial conditions

The population is initialized at the unfished equilibrium. Unfished survivorship-per-recruit accumulates natural mortality across ages with a plus-group at the oldest age, $\phi^0_a = \phi^0_{a-1}\, e^{-M_{a-1}}$ with ϕ^0_n divided by $1-e^{-M_n}$, and the unfished spawners-per-recruit is $\Phi_0 = \sum_a \psi_a\, \phi^0_a$. Unfished recruitment and the Beverton-Holt parameters follow from unfished spawning biomass B_0 and steepness h:

$$R_0 = \frac{B_0}{\Phi_0}, \qquad a_{\mathrm{BH}} = \frac{4hR_0}{5h-1}, \qquad b_{\mathrm{BH}} = \frac{B_0(1-h)}{5h-1}.$$

Module: `get_initial_numbers()`

When initial fishing mortality is specified, survivorship is recomputed with total mortality $Z_a = M_a + \sum_f F_f^{\mathrm{init}} s_{f,a}$, giving a fished spawners-per-recruit Φ_{eq} and an equilibrium recruitment $R_{\mathrm{eq}} = a_{\mathrm{BH}} - b_{\mathrm{BH}}/\Phi_{\mathrm{eq}}$. Initial numbers-at-age are $N^{\mathrm{init}}_a = R_{\mathrm{eq}}\, \phi_a$, optionally modified by estimated initial age-structure deviations with the lognormal bias correction described below.

2.1.4 Harvest and the forward projection

The population is projected forward over years and seasons. Within each season, survival from natural mortality is $S_a = e^{-M_a/n_{\mathrm{season}}}$. Removals are taken as a harvest fraction rather than a continuous fishing-mortality rate. For each fishery with observed catch, a fully selected harvest rate is obtained by dividing the observed catch by the available biomass (or available numbers, depending on the catch units of the fishery),

$$F_f = \frac{C_f}{\sum_a N_a\, s_{f,a}\, w_{f,a} + \varepsilon}, \qquad h_{f,a} = F_f\, s_{f,a},$$

Module: `get_harvest_rate()`

where $s_{f,a}$ is selectivity-at-age, $w_{f,a}$ is weight-at-age, and ε is a small constant guarding the denominator. The total harvest rate at age is the sum over fisheries, $h_a = \sum_f h_{f,a}$, and predicted catch-at-age is $\hat{C}_{f,a} = h_{f,a} N_a$. To keep the combined harvest fraction feasible, the quantity $1 - \sum_f F_f$ is passed through a differentiable lower-bound function (`posfun`) that returns the argument where it exceeds a small threshold and otherwise smoothly approaches that threshold, adding the resulting penalty to the objective function.

Surviving numbers are advanced by age and season, with recruitment entering the first age class at the start of each year, and the plus-group accumulating survivors from the oldest two age classes. Spawning biomass is $\mathrm{SB}_y = \sum_a N_{y,a}\, \psi_a$. The routine returns numbers-at-age, predicted catch-at-age by fishery, harvest rates, and the spawning-biomass time series for both the exploited and the dynamic unfished populations; their ratio defines dynamic depletion, $\mathrm{SB}_y/\mathrm{SB}^0_y$.

2.1.5 Recruitment

Recruitment follows a Beverton-Holt stock-recruitment relationship with annual log-normal deviations ε_y:

$$R_y = \frac{a_{\mathrm{BH}}\, \mathrm{SB}_y}{b_{\mathrm{BH}} + \mathrm{SB}_y}\ \exp\!\left(\varepsilon_y - \tfrac{1}{2}\, b_y\, \sigma_r^2\right),$$

Module: `get_recruitment()`

where σ_r is the standard deviation of the deviations and b_y is a year-specific bias-adjustment scalar. The lognormal bias-correction term $\frac{1}{2} b_y \sigma_r^2$ optionally can be modulated by a bias-adjustment ramp that reduces the correction in years where recruitment is poorly informed by the data, following the approach used in SS3 (Methot & Wetzel, 2013). Recruitment deviations may be treated as fixed effects (penalized maximum likelihood) or as random effects.

2.2 Likelihood components

The statistical model is written as a joint posterior density for the unknown parameters and latent quantities, conditional on the observed data. In implementation, `opal` evaluates the corresponding negative log posterior as the sum of the data log-likelihood contributions, recruitment-deviation density, parameter prior densities, and a soft harvest-rate constraint. Each data likelihood is controlled by a switch in the data object, so that any combination of CPUE, length composition, and weight composition observations may be included or excluded without altering the model code.

2.2.1 Abundance indices

Relative abundance is represented as an observation model for the logarithm of each standardized index. The CPUE component (`get_cpue_like()`) supports one or more indices, each with its own catchability q_j, additional observation standard deviation τ_j, a power parameter ω_j relating the index to vulnerable abundance, and a cumulative effort-creep term κ_{ij}. For observation i in index j, a latent uncentred index predictor is formed from the vulnerable abundance $\tilde{N}_i$ associated with that observation,

$$\log \hat{I}_{ij} = \log q_j + \omega_j \, \log \tilde{N}_i + \log \kappa_{ij}.$$

Module: `get_cpue_like()`

Within each index, the abundance and effort-creep contribution $\omega_j \, \log \tilde{N}_i + \log \kappa_{ij}$ is mean-centred on the natural scale and $\log q_j$ is then added, so that q_j sets the mean level of the predicted index and remains identifiable rather than being absorbed by the centring. Observation and process variability are combined in quadrature,

$$\sigma_{ij}^2 = \sigma_{\mathrm{obs},ij}^2 + \tau_j^2,$$

where $\sigma_{\mathrm{obs},ij}$ is the input standard error. The resulting observation model is

$$\log I_{ij} \mid \hat{I}_{ij}, \sigma_{ij}^2 \sim \mathrm{Normal}\Big(\log \hat{I}_{ij},\, \sigma_{ij}^2\Big),$$

and the CPUE contribution to the objective is the negative log density of these observations under this normal model.

2.2.2 Composition data

Length and weight composition observations are represented as draws from probability vectors predicted by the population and observation model. Predicted compositions are obtained by converting predicted catch-at-age $\hat{C}_{f,a}$ to a length basis through the PLA (and, for weight composition, an additional length-to-weight rebinning matrix $\mathbf{G}$):

$$\hat{p}_{f,\ell} \propto \sum_a P_{\ell a}\, \hat{C}_{f,a} \qquad \text{(length)}, \qquad \hat{p}_{f,k} \propto \sum_{\ell,a} G_{k\ell}\, P_{\ell a}\, \hat{C}_{f,a} \qquad \text{(weight)}.$$

Modules: `get_length_like()`, `get_weight_like()`

Predicted proportions are aggregated into the active bin range for the fishery, with the tails optionally accumulated into the minimum and maximum bins, then a small robustness constant is added and the vector is renormalized to sum to one. For observation i of fishery f, let $\hat{\mathbf{p}}_{if}$ denote this predicted composition, n_i the observed or effective sample size, $\mathbf{c}_{if}$ the observed count vector, and $\mathbf{o}_{if}$ the observed proportion vector. The available observation models are

$$\mathbf{c}_{if} \mid \hat{\mathbf{p}}_{if}, n_i \sim \text{Multinomial}(n_i,\, \hat{\mathbf{p}}_{if})\,,$$

for the multinomial option,

$$\mathbf{o}_{if} \mid \hat{\mathbf{p}}_{if}, n_i, \tau_f \sim \text{Dirichlet}(\hat{\mathbf{p}}_{if}\, n_i\, e^{\tau_f})\,,$$

for the Dirichlet option, and

$$\mathbf{c}_{if} \mid \hat{\mathbf{p}}_{if}, n_i, \tau_f \sim \text{Dirichlet-Multinomial}(n_i,\, \hat{\mathbf{p}}_{if}\, e^{\tau_f})\,,$$

for the Dirichlet-multinomial option. The parameter τ_f is a fishery-specific log-scale variance-adjustment parameter that controls the effective sample size or overdispersion. The composition contribution to the objective is the negative log density of the selected observation model. It is evaluated as a batched matrix operation over the observations of each fishery rather than a per-observation loop, which reduces the size of the automatic-differentiation tape.

2.2.3 Priors and penalties

Prior information is incorporated through a flexible prior specification (`evaluate_priors()`) that applies parametric prior densities to named parameters. Recruitment deviations are treated in the same distributional style. By default,

$$\varepsilon_y \mid \sigma_r \sim \text{Normal}(0, \sigma_r^2),$$

with the lognormal recruitment bias adjustment applied in the recruitment equation (Section 2.1.5). Initial

age-structure deviations, when estimated, use the analogous normal density. These terms can be interpreted as penalties in an MLE fit or as prior/random-effect contributions in the joint posterior. This shared formulation allows the same model code to be optimized with `nlminb` or sampled as a Bayesian model (Section 4.1).

3 Development pathway

Software development follows a defined contribution workflow. The package uses a two-branch structure: a `dev` branch for active development and a stable `main` branch for tested, release-ready code. New features are developed on branches forked from `dev` and are expected to include roxygen function documentation and minimal, self-contained `testthat` unit tests. Contributions are submitted as pull requests against `dev`, and are incorporated into `main` only after they pass the package checks — that is, after all tests pass and the package and its vignettes can be built.

Integrated testing operates at several levels. Unit tests verify the behaviour of individual modules; for example, the growth tests confirm that `get_growth()` returns the reference lengths at the reference ages and is monotonic, and that the weight, maturity, and PLA conversions behave as expected. Integration tests verify that the modules combine correctly within the full `opal_model()` objective function and that the objective and its gradient are finite when particular data and likelihood components are active. A numerical-regression baseline (`opal_baseline`), produced and checked through a dedicated vignette, stores a complete reference evaluation of the bigeye case-study model — the objective value, the gradient, the report output, and timing — so that a refactored build can be compared against the reference to detect unintended numerical changes and performance regressions. Continuous-integration checks (`R-CMD-check`) are run automatically on the repository.

4 Features

4.1 Bayesian inference

Prior information can be incorporated either through a penalized maximum likelihood estimation (MLE) approach, in which priors enter the objective function as penalty terms and the model is optimized with `nlminb`, or through direct Bayesian inference. Bayesian estimation is performed with `SparseNUTS` (C. Monnahan, 2026), which implements a sparse-preconditioned variant of the No-U-Turn Sampler that exploits the sparse precision structure of TMB models to improve sampling efficiency for high-dimensional, correlated, and hierarchically structured posteriors (C. C. Monnahan et al., 2026). Since the `opal` objective function and its component modules are constructed in RTMB, the same model definition is used for both the MLE and the Bayesian workflows. Convergence is assessed with the usual Markov chain Monte Carlo (MCMC) diagnostics — split-$\hat{R}$ and bulk and tail effective sample sizes — and the resulting posterior draws are carried directly into the forward projections described below. For the 'opakapaka case study (Section 5.1), a dense-metric sampler run for eight chains mixed well across the full parameter vector (Table 2, Figure 3).

4.2 Projections

`opal` supports catch-based forward projection of the population from an estimated set of parameters. A projection is specified by a number of projection years, a future catch scenario, and a number of uncertainty

iterations, and is run with `project_dynamics()` together with the helper functions `project_rec_devs()` and `project_selectivity()`. Future recruitment deviations are simulated from an ARIMA model fitted to the estimated recruitment-deviation series (or, under the Bayesian workflow, to each posterior draw's own deviation history), and future selectivity is resampled from the most recent estimated years, so that both process components carry forward the variability seen in the historical period.

Projections are conditioned on the specified future catch and propagate parameter uncertainty into the projected quantities. Two approaches to characterizing that uncertainty are available: resampling from a multivariate normal (MVN) approximation to the joint parameter distribution at the mode, and propagation of a posterior sample obtained by MCMC (Section 4.1). Under the MVN approach the scalar parameters are drawn from the Hessian-based covariance while the recruitment and initial-age deviations are set to their conditional expectation given those draws. This preserves the estimated correlation between population scale and the deviations and avoids a Jensen-inequality inflation of mean biomass, so that the uncertainty in the terminal (projection-start) state coincides with the delta-method uncertainty of the historical trajectory. Both approaches yield distributions of projected spawning biomass and depletion under the specified catch scenario (Figure 4).

4.3 Model diagnostics

`opal` provides a set of built-in diagnostics intended to support model evaluation and validation. These include estimability checks and parameter-correlation summaries derived from the fitted model; simulation self-testing, in which data are simulated from the model and re-estimated to confirm that parameters can be recovered; posterior-predictive checking where the observations are compared to the simulated distributions of the predictions given the model and OSA residuals, computed through RTMB's `oneStepPredict()`, which provide residuals that are independent and identically distributed under a correctly specified model and are therefore more readily interpretable than ordinary residuals for non-Gaussian observations (Dunn & Smyth, 1996; Kristensen et al., 2016; Thygesen et al., 2017). The simulation and posterior-predictive capabilities also support the projection and Bayesian workflows described above.

5 Case studies

Two case studies are distributed with the package as worked examples. The first applies `opal` to simulated data for which the underlying population dynamics are known; the second applies it to real data from an existing WCPFC tuna assessment. Detailed configuration, code, and outputs for each are provided in the package vignettes. In both cases the examples are purely illustrative and developed for the purpose of demonstrating the model structure, features, or comparability with existing software platforms. They are not intended to be used as a basis for management advice.

5.1 Hawai'i 'Opakapaka

The first case study uses simulated data for Hawai'i 'opakapaka (*Pristipomoides filamentosus*), a deepwater snapper, for which the true population dynamics are known. The reference model was constructed in SS3 (v3.30.19; Methot & Wetzel, 2013) as part of an associated simulation study (Oshima et al., 2026), which provides both an operating model that generates the data and an estimation model fitted to those data. The `opal` model covered the historical period 1949–2023 with three fleets — a commercial fishery, a

non-commercial fishery, and a research-fishery survey — at an annual time step, and was fitted to two CPUE indices, generated with negligible observation error, together with length composition data from the commercial and survey fleets.

Two versions of the `opal` model were configured in this case study. A baseline fixed effects version was configured to estimate population scale, annual recruitment deviations, initial age structure, and selectivity parameters for the fleets that have length composition data, with the remaining life-history and fishery parameters fixed at the values used to generate the simulated data. Alternatively, a random effects version of the `opal` model was configured identically to the first with the exception that random effects were used to estimate the annual recruitment deviations and the initial age structure, and recruitment variability `log_sigma_r` was estimated. Additionally, because the operating model was built in SS3, the case study also serves as a cross-platform comparison: it provides a basis for evaluating the correspondence between the `opal` and SS3 population-dynamics calculations under identical inputs, and for comparing the spawning-biomass trajectory and parameter estimates from `opal` against the operating-model truth and the SS3 estimation model. As shown in Figure 1, `opal` reproduced the SS3 simulated spawning-biomass trajectory, demonstrating that the internal population dynamics engine is sound.

Both `opal` models converged (Table 1) and completed standard error calculations fairly quickly (~2 seconds for the fixed effects model, ~2 minutes for the random effects model — both of which were substantially faster than the ~6 minutes for the SS3 model) without any issues detected with regards to the gradient or Hessian and all parameters (including `log_sigma_r` in the random effects model) were identifiable according to `check_estimability()`. Using `RTMB`'s `simulate()` function it is possible to quickly generate simulated data from the fitted model. These simulates datasets can either be refit within the model (e.g., simulation self-testing) or used to generate posterior predictive distributions for the observed data. Using the commercial fishery index as an example, Figure 2 shows that the observed CPUE tended to fall within the posterior predictive distribution (Panels A & B); however, the PIT ECDF plot indicated the predictions to be overdispersed relative to the observations. This overdispersion makes sense given that the commercial fishery index was generated with negligible observation error and `RTMB`'s `simulate()` function incorporates both the full uncertainty (observation + process error) from the estimated model.

The fixed effects model was also fitted with the Bayesian workflow (Section 4.1): a dense-metric NUTS sampler was run for eight chains, which were well mixed with split-$\hat{R}$ close to one and large effective sample sizes for the scalar parameters and recruitment deviations (Table 2, Figure 3). The stock was then projected ten years forward under a constant catch held at the terminal year's level, propagating parameter uncertainty through both the MVN and MCMC approaches (Section 4.2). The two approaches agreed closely at the projection-start state, and the projection intervals widened only gradually because 'opakapaka is long-lived and matures slowly, so the existing adult biomass declines nearly deterministically under constant catch while future recruitment and selectivity uncertainty accumulate over the projection years (Figure 4).

5.2 2020 western and central Pacific Ocean bigeye tuna

The second case study used the data from a fleets-as-areas model described in the appendix of the 2020 assessment of bigeye tuna (*Thunnus obesus*) in the western and central Pacific Ocean (WCPO; Ducharme-Barth et al., 2020). The bundled data object contained the biological parameters, catch and standardized CPUE observations, length and weight composition data, configured at a quarterly time step across the

fisheries defined in the fleets-as-areas assessment, with selectivity specified by fishery as either logistic or double-normal forms.

The diagnostic configuration used the quarterly standardized CPUE series as four seasonal indices, fitted length compositions from fisheries 8–14, and fitted weight compositions from fisheries 1, 2, 3, 4, 6, 7, and 15. Sensitivity runs indicated that the size-composition likelihoods pulled the absolute spawning-biomass scale below the SS3 and MFCL reference models, while posterior-predictive composition checks indicated uneven relative weighting among fisheries. The retained fit therefore used fishery-specific variance adjustments to the length-frequency (LF) and weight-frequency (WF) data sets. Weight compositions for fishery 5 remained inactive after diagnostic runs failed to converge cleanly. Recruitment deviations were estimated across the model period, catchability was estimated separately by quarterly index, and selected selectivity parameters were estimated for the fisheries with active composition data. Growth, natural mortality, steepness, recruitment variability, effort creep (assumed to be zero), and the remaining selectivity parameters were fixed at the bundled reference values. The quarterly catch history used by the model is shown by fishery and catch-unit group in Figure 5 and Figure 6.

The final bigeye tuna (BET) model configuration converged according to `nlminb`, with a KKT-style[1] projected gradient close to zero after accounting for active parameter bounds; the larger raw gradient is reported separately in Table 3. Remaining active selectivity bounds are reported in Table 7. Fitting this diagnostic configuration from the retained-cache starting values to the maximum-likelihood estimate took approximately 69 s on the workstation used to build this document; the final optimizer restart recorded 880 iterations, 952 function evaluations, and 881 gradient evaluations. Hessian-based uncertainty was computed with `RTMB::sdreport()` using a supplied fixed-effect Hessian from `optimHess()` with a finite-difference step of 10^{-4}; this Hessian was positive definite, with minimum eigenvalue 7.42×10^{-4}. The additive objective decomposition was still dominated numerically by the downweighted composition likelihoods, especially the weight-composition component, while the CPUE component contributed the abundance trend information and the recruitment-deviation density regularized quarterly departures from the stock-recruitment curve. The estimated unfished spawning biomass was approximately 872,989 mt, close to the MFCL implied scale of 975,288 mt and closer to the SS3 fixed-selectivity scale of 1,076,845 mt than an earlier higher-composition-weight diagnostic fit. The final-year relative spawning biomass was 0.197 when scaled by SB_0 (Table 4). The dynamic depletion ratio was slightly lower (0.182) because it uses the time-varying unfished spawning biomass denominator rather than the static SB_0 denominator.

An additional exploratory BET model was fitted after aggregating the 15 original fisheries to six fisheries using the grouped downweighted LF/WF configuration described above (Table 9). This sensitivity was intended to test whether the many small fisheries were materially affecting the abundance scale or the CPUE fit, rather than to replace the retained diagnostic model. Under the six-fishery aggregation, B_0 was 1,023,476 mt and final-year SB/SB_0 was 0.238, compared with 872,989 mt and 0.197 for the retained 15-fishery model (Table 10). The CPUE log-RMSE was also similar (0.116 for the six-fishery model versus 0.111 for the retained model), and the six-fishery CPUE and biomass trajectories are included in Figure 7 and Figure 21. The six-fishery selectivity curves in Figure 20 should be interpreted as pooled selectivities for the aggregation groups rather than direct replacements for the original fishery-specific curves.

[1] Following the Karush–Kuhn–Tucker (KKT) conditions for bound-constrained optimization, gradient components at an active bound are not treated as evidence of non-convergence provided their sign is consistent with the constraint (e.g., pushing further against, rather than away from, the bound).

Diagnostic plots in the `opal` BET vignette compare observed and predicted CPUE by seasonal index, observed and predicted length and weight compositions, fitted recruitment deviations, and estimated selectivity-at-length and selectivity-at-age curves. The CPUE fit in Figure 7 tracked the broad decline in standardized abundance from the early 1950s through the low-abundance period in the 2000s, while allowing the four quarterly indices to retain their own scale and observation uncertainty; simulated CPUE series from the fitted model showed the amount of variation expected under the fitted lognormal observation model. Additionally, illustrative fits to the same index from MFCL and SS3 models are also shown for reference. The CPUE OSA residuals in Figure 8 were centred close to zero and fell within approximately ±2 residual units, but the Q-Q panels were clearly under-dispersed: standard deviation of normalised residuals (SDNRs) were below one for each index month and for all indices combined (Table 5). For the independent lognormal CPUE likelihood these OSA residuals are equivalent to standardized log residuals, so the flattened Q-Q pattern indicated that the fitted CPUE observation variance was larger than the residual variation, rather than a small number of large outliers (Dunn & Smyth, 1996; Thygesen et al., 2017). The length- and weight-composition diagnostics in Figure 9 and Figure 10 sum variance-adjusted pseudo-counts over all years and independently simulate fitted row-level compositions before pooling those simulated counts within fishery. The corresponding grouped diagnostics for the six-fishery sensitivity are shown together in Figure 11 using the same row-level pooling logic. To show the unaggregated temporal structure, Figure 12 and Figure 13 give year-specific examples for length-composition fishery 12 and weight-composition fishery 15. Example LF and WF composition OSA residual diagnostics are shown for fishery 12 and fishery 15 in Figure 14 and Figure 15; these two panels are examples rather than the full residual set. The by-fishery Q-Q diagnostics in Figure 16 summarise all active composition residuals against a fixed standard-normal reference line, rather than a fitted Q-Q line, so below-one dispersion remains visible; Figure 17 repeats this Q-Q layout for the six-fishery grouped sensitivity using the same `compResidual::resMulti()` residual calculation on the grouped compositions. Following the diagnostic summaries commonly applied in CCSBT assessments of southern bluefin tuna, Table 6 reports finite bin-level OSA SDNRs generated with `compResidual::resMulti` for both the retained 15-fishery model and the matched six-fishery sensitivity. The near-one OSA SDNRs are conditional on the variance-adjusted input sample sizes used in the retained likelihood; the mean input N by retained-model fishery ranges from 0.0124 to 2.14 after downweighting. Because these variance adjustments produce fractional pseudo-counts, including sub-unit effective sample sizes for some rows, the multinomial composition likelihood in this diagnostic fit should be interpreted as a weighted pseudo-likelihood convention rather than a literal integer-count multinomial sampling model. The close Q-Q alignment is therefore a check of the randomized bin-level residual distribution for the weighted pseudo-data likelihood, not evidence that the raw composition data are perfectly fitted.

The CPUE residual pattern supported reducing the fixed CPUE τ (additional observation standard deviation) and/or scaling the CPUE SEs down. The retained BET fits fixed τ at 1e-08 for all four CPUE indices; the only active CPUE parameters were the four catchability terms. Relative to the earlier cached run with $\tau = 0.1$, setting τ near zero reduced the mean CPUE standard deviation from 0.222 to 0.198, but increased the all-index CPUE SDNR only from 0.513 to 0.565. Removing the fixed τ therefore helped only modestly and did not resolve the under-dispersion by itself. Further CPUE upweighting would require scaling the CPUE SEs down and should not be treated as an automatic tuning to SDNR = 1 without checking the consequences to the biomass estimates and composition fits.

The selectivity diagnostics in Table 7 and Figure 19 show where the fitted `opal` selectivity curves differ

from both the MFCL and SS3 reference curves and which selectivity terms remain estimated in the final constrained map. Several weakly identified width terms were fixed at their SS3-initialized values to obtain stable Hessian-based uncertainty; Table 7 lists these fixed terms explicitly. The broader estimated-parameter blocks and their optimizer bounds are summarized in Table 8. The recruitment deviations in Figure 18 show the seasonal pattern of estimated process variation. The biomass and depletion trajectories in Figure 21 show the long-term decline in spawning biomass (SB), static (SB/SB_0), and dynamic depletion ($SB_t/SB_{t,F=0}$) for the retained and six-fishery `opal` models relative to MFCL and SS3 models fitted to the same data set. The retained `opal` uncertainty ribbon is the approximate 95% interval from the positive-definite supplied-Hessian `sdreport()` calculation.

The `opal` spawning-biomass scale in Figure 21 in the retained downweighted fit was much closer to the SS3 and MFCL reference trajectories than initial model runs. However, a review of the document cache and the `opal` BET scripts suggests several relevant differences from the *equivalent* SS3 and MFCL models that could lead to differences in the `opal` estimated biomass scale. First, the double-normal `opal` selectivity curves are length-based (as opposed to age-based cubic splines in MFCL) and are converted through the PLA; the bundled double-normal width starts require conversion from a legacy SS3 scale to the current width scale, and diagnostic runs in the `opal` repository showed that biomass scale is sensitive to this selectivity scaling and map. Second, the fit used a simplified harvest-fraction removal calculation applied at the start of each timestep, no weight composition for fishery 5, and an intentionally downweighted size-composition likelihood. These differences can all change the vulnerable biomass implied by the same catch and CPUE data, so the absolute SB comparison is best viewed as a diagnostic cross-platform check rather than a tuned stock-status result. As with the 'opakapaka example, these results are intended to demonstrate the end-to-end `opal` workflow as a proof-of-concept and do not reflect a carefully considered stock assessment model intended for management advice.

6 Discussion

`opal` provides an age-structured, length-based stock assessment framework implemented in RTMB, with a modular internal structure, support for multiple data types and likelihood forms, random-effects estimation, catch-based projections, Bayesian inference, and a set of built-in diagnostics. The two case studies demonstrate its application to fit length composition, weight composition, and index data in single-species assessments in both a simulated data context, where the truth is known, and with a real data tuna example. Both examples serve to highlight the successful proof-of-concept of the `opal` framework in its ability to adequately estimate population dynamics and easily provide high-quality diagnostic outputs for model evaluation and validation.

The continued development of case-studies is central to the development of `opal`, with case studies serving multiple purposes: they provide a basis for testing and validating the model, they demonstrate the model's capabilities to the broader community, and they provide a starting point for new users to learn how to use the software. Additionally, new modules are added to the code base to meet the needs of the case studies as they arise. The integrated testing approach ensures that new modules do not break existing functionality, while safely providing a pathway to grow model functionality. Use of simulated data in case studies provides an objective basis for evaluating `opal`'s performance, and for cross-platform comparisons with existing software.

One of the primary motivations for the development of `opal` was to enable the calculation of more recently developed statistical diagnostics that are either unavailable or challenging to calculate in existing software

platforms such as SS3 or MFCL. The included case-studies serve to showcase and demonstrate some of these diagnostics that are readily calculated from an `opal` model fit. In both cases, the `check_estimability` function can be used to rapidly identify if the model has converged and if all parameters are identifiable. In the case of the bigeye tuna case study this led to the identification of a number of poorly determined selectivity parameters, which could subsequently be held fixed, and also prompted an investigation of a model with a simplified fisheries structure where all selectivity parameters were estimable. The 'opakapaka case study also demonstrated the use of the `simulate` function to generate posterior predictive distributions for the observed data in order to ascertain whether the model was able to adequately capture the observed data. Additionally, the bigeye tuna case study demonstrated the use of OSA residuals to evaluate the fit of the model to the observed data, and to identify potential issues with the model fit. The OSA residuals can be implicitly calculated as a by-product of the underlying RTMB framework, and are therefore readily available for any model fit.

The other driving motivation for `opal` development was to facilitate the application and implementation of Bayesian inference within integrated assessment models. Bayesian inference provides a more holistic characterization of uncertainty in model estimates through the incorporation of prior information through parameter priors. This can have the additional benefit of stabilizing model estimates, as well as to allow the estimation of parameters that may otherwise be poorly determined. Doing so may allow analysts to move away from grids of alternative fixed parameters model runs to a more simplified model ensemble where uncertainty in key parameters is represented by their prior. Lastly, having a Bayesian stock assessment modelling framework readily available to analysts can formalize the inclusion of information from precursor, external analyses[2] into the stock assessment model.

As mentioned previously, one of the benefits of the case-study driven approach is the ability to conduct cross-platform comparisons with existing and well-established stock assessment software. In the 'opakapaka case study, the `opal` model was able to be closely configured to match the existing SS3 operating and estimation models. Though there are slight differences in the underlying population dynamics calculations[3], the `opal` model was able to closely track the SS3 spawning biomass trajectory which demonstrates the viability of the `opal` population dynamics structure. Given that we had access to the 'known-truth' from simulated data, this provided further confidence in the ability of `opal` to accurately estimate population dynamics. The bigeye tuna case study also demonstrated the ability of `opal` to closely track the spawning biomass and depletion trajectories from both SS3 and MFCL, though there were some differences in the absolute scale of the spawning biomass. These differences are likely due to differences in selectivity parameterization, composition weighting, and catch removal calculations between the three platforms. The case studies also highlighted some of the challenges associated with fitting complex integrated assessment models, particularly when using real data where the underlying truth is unknown. These challenges include issues with parameter identifiability, sensitivity to model configuration, and the need for careful consideration of model diagnostics and outputs.

While the case studies demonstrated `opal`'s ability to both successfully estimate population dynamics, and

[2] For example, the development of movement or population scale priors from external analyses, such as SEAPODYM (Lehodey et al., 2008) or the DTU fine-scale movement model (Magnusson et al., 2025), in a potential future spatially-explicit version of `opal` .

[3] `opal` applies harvest rates at the start of each time step, while SS3 applies a continuous total mortality at the mid-point of each time step. Additionally, all biological calculations in SS3 are based on the population demographics at the mid-point of the time step.

handle the data-volume and high-dimensionality of a real fisheries example[4] they also identified numerous challenges, some still outstanding. Though the 'opakapaka case study demonstrated that `opal` was able to treat and estimate certain parameters as random effects, it resulted in a model that was substantially slower than the fixed effects model. It remains to be seen how this scales to higher dimensionality assessments as preliminary attempts to apply random effects in the bigeye tuna case study did not appear tractable[5]. Similarly, while the `opal` model is now able to fairly rapidly converge to a solution in the highly dimensional bigeye tuna case study, this model begins optimization from fairly plausible initial values. Substantial time and effort went into developing both the plausible initial value set and the associated parameter bounds. Both SS3 and MFCL appear to more robustly converge from naive initial values so further work is warranted to improve the robustness of `opal` 's optimization. However, it is worth noting that both SS3 and MFCL estimate parameters in phases (MFCL has a fairly refined initialization sequence) which may contribute to their better optimization performance, while `opal` simultaneously estimates all parameters without phasing. Additionally, code optimization is still an active phase of development and several areas (e.g., length-weight conversion, selectivity parametrization, and size-composition) of the underlying `opal` model structure were re-factored during the development of the bigeye tuna case-study to improve the model performance. Further code optimization is likely needed to improve the speed and robustness of the model, particularly for high-dimensional assessments.

In addition to the aforementioned issues, the real-world data that formed the basis for the bigeye tuna case study illustrated several further challenges. The initial richer selectivity maps were not fully estimable: several double-normal width and right-tail parameters were weakly identified, some ran to bounds, and Hessian-based uncertainty was not available until the $\log(B_0)$ lower bound was relaxed and selected weakly identified selectivity terms were fixed at their SS3-initialized values. The resulting model is therefore a constrained diagnostic fit rather than a fully tuned assessment. The case study also exposed practical issues that are common in integrated assessment models: the absolute biomass scale is sensitive to selectivity parameterization and composition weighting; the retained diagnostic fit used fishery-specific LF/WF variance adjustments to balance biomass scale and posterior-predictive composition variability; weight-composition fishery 5 was excluded after convergence diagnostics were poor; and the mixed catch units and quarterly time-step encoding require careful checks when comparing against SS3 and MFCL outputs. Finally, the size-composition residual diagnostics are computationally and statistically delicate because they operate on weighted pseudo-data likelihoods, so apparently good OSA residual behaviour must be interpreted alongside effective sample sizes, mean-composition diagnostics, and selectivity plausibility. This reinforces the need for continued development of both simulated and real-data case studies to ensure that the `opal` framework is robust and can be applied to a wide range of fisheries assessment scenarios.

Despite the progress to date, `opal` is far from a finished product. As `opal` transitions into the next phase of development multiple rich avenues for continued work exist across several core areas: documentation, development of the user interface, model testing, code optimization, research, and feature development. This work extends beyond simply coding a technically-sound population dynamics model, but also encompasses

[4] In terms of population dynamics structure the BET case study uses 15 fisheries, 268 time-steps, 40 quarterly age-classes, 95 length-bins, and 200 weight-bins. There are 4 quarterly CPUE indices, and 2 composition types (1,194 WF instances and 568 LF instances). The model is therefore a high-dimensional integrated assessment model that is representative of the complexity of real-world fisheries assessments.

[5] However, these attempts were made prior to certain code optimizations and reparametrizations in the selectivity and size-composition likelihood calculations, so it is possible that random effects estimation may be more tractable in future versions of `opal` .

managing the user experience so that `opal` is an accessible, easy to learn tool that does what is needed. Continued success will depend on both engagement and investment from the broader fisheries assessment community beyond the existing development team. Additionally, a focused and prioritized list of requirements is needed to inform the development pathway and to ensure that the resulting software meets the needs of the community. In the interim, several extensions are planned based on either existing feature lists (Punt et al., 2020) or the development team's experiences with `opal` to date. Areas identified for ongoing work include:

- refinement of existing case studies;
- development of additional case studies (e.g., Indian Ocean yellowfin tuna fleets-as-areas case study based on simulated data);
- code optimization (particularly of the selectivity and size-composition likelihood calculations);
- derivation of biological and fishery reference points;
- additional likelihood components (e.g., age, tag or close-kin mark-recapture data);
- alternative selectivity forms;
- and additional population dimensionality (e.g., spatial structure, sex, or length).

Within the context of the WCPFC, `opal` could be applied to the upcoming assessment of north Pacific Ocean blue shark (*Prionace glauca*), or in a research capacity to the assessment of south Pacific albacore (*Thunnus alalunga*).

The work so far demonstrates the proof of concept of the R-based modular development and has made significant development steps for core model components. We suggest that continued development of `opal` provides an option for the core development stream for the next generation of tuna assessment software and can provide alternative model features and structures that would be suitable for assessments of various species (e.g., tuna, shark and billfish) within the WCPFC or other regional fishery management organizations.

7 Acknowledgements

Development of opal is intended to be an open and collaborative effort, inclusive to the broader fisheries assessment community. To that end we are grateful to the many people that contributed to the project through informal discussions, and participation in periodic update meetings. While not an exhaustive list, this community includes: N. Davies (Te Takina Ltd), S. Hoyle (Hoyle Consulting), M. Maunder (Inter-American Tropical Tuna Commission), R. Bi (IATTC), F. Carvalho (PIFSC), G. Pilling (SPC), J. Hampton (SPC *emeritus*), and many others including participants in the 2025 CAPAM mini workshop on developing the next generation tuna stock assessment model as well as the 2026 SPC Pre-Assessment Workshop. Development of `opal` has received funding support from *WCPFC Project 123: Scoping the next generation of tuna stock assessment software*.

8 Declaration of generative AI use

Generative AI tools were used to assist with code development, refactoring, and documentation, and with drafting and editing text in this document. All analytical choices, model specifications, diagnostic interpretations, and conclusions are those of the authors, who reviewed and verified all AI-assisted output.

9 References

Casal2 Development Team. (2026). *Casal2: Integrated population dynamics model*. https://github.com/Casal2/CASAL2

Cheng, M. Lh., Goethel, D. R., Cunningham, C. J., Hulson, P. F., Ianelli, J. N., & Omori, K. L. (2026). The SPoRC Stock Assessment Package: A Generalized Next-Generation Platform to Assess Spatial, Age and Sex-Structured Populations. *Fish and Fisheries*, *27*(3), 754–770. https://doi.org/10.1111/faf.70082

Ducharme-Barth, N., Vincent, M., Hampton, J., Hamer, P., Williams, P., & Pilling, G. (2020). *Stock assessment of bigeye tuna in the western and central Pacific Ocean (30July) - Rev.03* (SC16-SA-WP-03). https://meetings.wcpfc.int/node/11693

Dunn, P. K., & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, *5*(3), 236–244. https://doi.org/10.1080/10618600.1996.10474708

Fournier, D. A., Hampton, J., & Sibert, J. R. (1998). MULTIFAN-CL: A length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. *Canadian Journal of Fisheries and Aquatic Sciences*, *55*(9), 2105–2116. https://doi.org/10.1139/f98-100

Kristensen, K. (2023). *RTMB: 'R' Bindings for 'TMB'*. https://doi.org/10.32614/CRAN.package.RTMB

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). **TMB** : Automatic Differentiation and Laplace Approximation. *Journal of Statistical Software*, *70*(5). https://doi.org/10.18637/jss.v070.i05

Lehodey, P., Senina, I., & Murtugudde, R. (2008). A spatial ecosystem and populations dynamics model (SEAPODYM) – Modeling of tuna and tuna-like populations. *Progress in Oceanography*, *78*(4), 304–318. https://doi.org/10.1016/j.pocean.2008.06.004

Magnusson, A., Davies, N., Pilling, G., & Hamer, P. (2025). *Scoping the Next Generation of Tuna Stock Assessment Software: Progress Report (Project 123)* (WCPFC-SC21-2025/SA-WP-01).

Methot, R. D., & Wetzel, C. R. (2013). Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research*, *142*, 86–99. https://doi.org/10.1016/j.fishres.2012.10.012

Miller, T. J., & Stock, B. C. (2020). *The {W}oods {H}ole {A}ssessment {M}odel ({WHAM})*. https://timjmiller.github.io/wham/

Monnahan, C. (2026). *SparseNUTS: Sparse No-U-Turn MCMC Sampling for 'Template Model Builder'*. https://noaa-afsc.github.io/SparseNUTS

Monnahan, C. C., Kristensen, K., Thorson, J. T., & Carpenter, B. (2026). *Leveraging Sparsity to Improve No-U-Turn Sampling Efficiency for Hierarchical Bayesian Models*. arXiv. https://doi.org/10.48550/ARXIV.2603.02437

Oshima, M., Nadon, M., Syslo, J., Hongguang, M., & Carvalho, F. (2026). *Opakapaka Simulation Report*. https://github.com/noaa-pifsc/Opaka_simulation/tree/main

Punt, A. E., Dunn, A., Elvarsson, B. Þ., Hampton, J., Hoyle, S. D., Maunder, M. N., Methot, R. D., & Nielsen, A. (2020). Essential features of the next-generation integrated fisheries stock assessment package: A perspective. *Fisheries Research*, *229*, 105617. https://doi.org/https://doi.org/10.1016/j.fishres.2020.105617

Stock, B. C., & Miller, T. J. (2021). The Woods Hole Assessment Model (WHAM): A general state-space assessment framework that incorporates time- and age-varying processes via random effects and links to environmental covariates. *Fisheries Research*, *240*, 105967. https://doi.org/10.1016/j.fishres.2021.105967

Thygesen, U. H., Albertsen, C. M., Berg, C. W., Kristensen, K., & Nielsen, A. (2017). Validation of ecological state space models using the laplace approximation. *Environmental and Ecological Statistics*, *24*(2),

317–339. https://doi.org/10.1007/s10651-017-0372-4

Webber, D., Parma, A., Ianelli, J., Hillary, R., & Eveson, P. (2026). *Sbt: Southern bluefin tuna stock assessment model*. https://github.com/quantifish/sbt

10 Tables

Table 1: `opal` 'opakapaka model metrics: positive-definite Hessian, maximum absolute gradient (mgc), minimum Hessian eigenvalue, parameter estimability, runtime, number of parameters, and sigma R.

Metric	opal	opal (RE)
Positive-definite Hessian	TRUE	TRUE
Max |gradient| (mgc)	2.50e-03	8.17e-05
Min Hessian eigenvalue	2.47e-01	2.48e-01
Inestimable parameters	0 (all parameters estimable)	0 (all parameters estimable)
Runtime (seconds)	2.23	128.85
N parameters (fixed)	127	9
N parameters (random)	0	119
sigma R	0.52 (fixed)	0.45

Table 2: Posterior summaries and MCMC convergence diagnostics for the 'opakapaka opal model: the MLE estimate, posterior mean and standard deviation, 5th and 95th percentiles, bulk and tail effective sample sizes (ESS), and split-Rhat, for the scalar parameters (unfished spawning biomass and the two CPUE catchabilities) and three representative recruitment deviations.

Parameter	MLE	Mean	SD	q5	q95	ESS bulk	ESS tail	Rhat
log_B0	8.412	8.369	0.069	8.254	8.484	7551.397	4990.793	1.002
log_cpue_q[1]	0.002	0.002	0.012	-0.017	0.021	9055.441	4529.348	1.001
log_cpue_q[2]	-0.004	-0.004	0.038	-0.066	0.058	8109.356	4245.775	1.000
rdev_y[1]	0.134	0.114	0.354	-0.516	0.651	6982.413	4233.302	1.000
rdev_y[38]	-0.690	-0.676	0.227	-1.080	-0.323	7361.246	3667.656	1.002
rdev_y[75]	0.000	-0.003	0.505	-0.841	0.814	8439.138	4562.015	1.002

Table 3: Optimization and objective-component summary for the fitted WCPO bigeye tuna diagnostic model and matched six-fishery aggregation sensitivity. The Max KKT gradient is the maximum projected gradient after accounting for active parameter bounds. Objective components are a switch-based additive audit at the fitted parameter vector: the baseline has CPUE, length-composition (LF), and weight-composition (WF) likelihood switches off, and the CPUE, LF, and WF rows are single-switch differences from that baseline.

Metric	15-fishery BET	6-fishery BET
nlminb convergence	0	0
Objective	4,087.16	2,424.58
Max |gradient|	0.1334	0.0088
Max KKT gradient	0.0018	0.0088
B0	872,989	1,023,476
Baseline contribution	145.26	136.13
CPUE contribution	-149.84	-148.38
LF contribution	1,013.63	717.54
WF contribution	3,078.12	1,719.30
Objective-component total	4,087.16	2,424.58

Table 4: Final-year relative spawning-biomass diagnostics for the fitted WCPO bigeye tuna diagnostic model.

Metric	Value
SB/SB0	0.197
SB/SB_F0	0.182

Table 5: Standard deviation of normalized residuals (SDNR) and median absolute residual (MAR) for CPUE residuals in the fitted WCPO bigeye tuna diagnostic model and matched six-fishery aggregation sensitivity.

Model	Index month	n	SDNR	95% LCI	95% HCI	MAR
15F	All indices	268	0.565	0.521	0.618	0.379
6F	All indices	268	0.575	0.530	0.628	0.397
15F	February	67	0.615	0.526	0.741	0.507
6F	February	67	0.624	0.534	0.752	0.521
15F	May	67	0.492	0.420	0.593	0.354
6F	May	67	0.496	0.424	0.598	0.387
15F	August	67	0.608	0.520	0.733	0.388
6F	August	67	0.623	0.532	0.751	0.445
15F	November	67	0.550	0.470	0.663	0.353
6F	November	67	0.559	0.478	0.674	0.329

Table 6: Composition diagnostics by composition type and fishery for the fitted WCPO bigeye tuna diagnostic model and matched six-fishery aggregation sensitivity. OSA SDNR columns are finite bin-level `compResidual::resMulti()` residual summaries for both model configurations; the six-fishery residuals use grouped observed compositions and fitted grouped predictions from the raw sensitivity cache.

Model	Composition	Fishery	Rows	OSA SDNR	OSA 95% CI	Mean input N	Prop. RMSE	Mean size RMSE	Mean size bias
15F	Length	Fishery 8	122	0.998	0.985-1.011	0.12	0.0184	6.36	-4.27
15F	Length	Fishery 9	88	0.968	0.953-0.983	0.94	0.0231	7.51	-2.65
15F	Length	Fishery 10	92	0.980	0.966-0.995	0.58	0.0274	7.22	-3.26
15F	Length	Fishery 11	89	0.954	0.940-0.969	0.67	0.0215	12.24	4.15
15F	Length	Fishery 12	43	0.968	0.947-0.989	1.18	0.0469	13.82	1.86
15F	Length	Fishery 13	77	0.961	0.946-0.977	1.16	0.0391	15.40	0.51
15F	Length	Fishery 14	57	0.985	0.967-1.004	0.30	0.0448	9.56	1.19
6F	Length	Group 1	122	0.985	0.972-0.998	0.06	0.0179	5.66	-3.13
6F	Length	Group 2	92	0.962	0.948-0.977	1.15	0.0272	6.97	-2.93
6F	Length	Group 3	169	0.965	0.954-0.976	0.71	0.0322	11.07	-1.71
6F	Length	Group 4	89	0.966	0.952-0.981	0.34	0.0214	12.52	4.74
15F	Weight	Fishery 1	238	0.971	0.965-0.977	0.64	0.0068	7.11	1.12
15F	Weight	Fishery 2	112	0.949	0.940-0.958	2.14	0.0066	5.20	-1.88
15F	Weight	Fishery 3	118	0.993	0.984-1.002	0.14	0.0043	7.59	-4.10
15F	Weight	Fishery 4	240	0.979	0.973-0.985	0.54	0.0041	4.53	-1.31
15F	Weight	Fishery 6	158	0.999	0.992-1.007	0.01	0.0079	11.60	-9.25
15F	Weight	Fishery 7	86	0.995	0.984-1.005	0.11	0.0071	4.97	-2.43
15F	Weight	Fishery 15	242	0.953	0.947-0.959	1.37	0.0041	4.34	-1.69
6F	Weight	Group 5	242	0.953	0.947-0.959	1.33	0.0043	4.33	1.02
6F	Weight	Group 6	242	0.967	0.961-0.973	0.69	0.0042	4.90	-2.51

Table 7: Selectivity-parameter diagnostics by fishery for the fitted WCPO bigeye tuna diagnostic model. Estimated terms are estimated in the final constrained map; fixed terms were held at their SS3-initialized values. DN = double-normal. Bound flags identify any remaining estimated selectivity terms that are on parameter bounds.

Fishery	Form	Est. terms	Fixed terms	Bound flag
1	DN	peak	asc. width, desc. width	None
2	DN	peak	asc. width, desc. width	None
3	DN	peak, desc. width	asc. width	desc. width at lower bound
4	DN	peak, asc. width, desc. width	None	desc. width at lower bound
6	DN	peak	asc. width, desc. width	None
7	DN	peak	asc. width, desc. width	None
8	DN	peak	asc. width, desc. width	None
9	DN	peak, desc. width	asc. width	None
10	DN	peak, desc. width	asc. width	None
11	logistic	peak, top	None	peak at upper bound
12	DN	peak, desc. width	asc. width	None
13	DN	peak	asc. width, desc. width	None
14	DN	peak	asc. width, desc. width	None
15	logistic	peak	top	None

Table 8: Estimated parameter blocks and optimizer bounds for the fitted WCPO bigeye tuna diagnostic model. Selectivity rows are summarized as a block because individual fishery terms are listed in the selectivity diagnostics table; recruitment deviations are summarized by their fitted range.

Parameter	n	Estimate	Lower	Upper
log_B0	1	13.680 (B0 = 872,989 mt)	12.000	22.000
log_cpue_q (Feb)	1	0.066	-2.303	2.303
log_cpue_q (May)	1	-0.126	-2.303	2.303
log_cpue_q (Aug)	1	-0.101	-2.303	2.303
log_cpue_q (Nov)	1	0.075	-2.303	2.303
par_sel	21	-6.633 to 1.723	-17.825 to -1.721	-0.020 to 2.205
rdev_y	268	-0.693 to 1.095	-5.000	5.000

Table 9: Six-fishery aggregation used in the exploratory WCPO bigeye tuna BET sensitivity. Original fisheries were pooled only where catch units and selectivity form remained compatible. Catch percentages are calculated separately within metric-ton and thousands-of-fish catch units.

Six-fishery group	Original fisheries	Catch units	Selectivity form	Within-unit catch (%)
1	F8	metric tons	double-normal	63.8
2	F10	metric tons	double-normal	15.4
3	F9, F12, F13, F14	metric tons	double-normal	19.5
4	F11	metric tons	logistic	1.4
5	F1, F2, F3, F4, F5, F6, F7	thousands of fish	double-normal	100.0
6	F15	thousands of fish	logistic	0.0

Table 10: Comparison of the retained 15-fishery WCPO bigeye tuna diagnostic model and the matched six-fishery aggregation sensitivity. Objective values are omitted because the composition data are aggregated differently and therefore are not directly comparable.

Model	Fisheries	Base LF/WF downweight	B0 (mt)	Final SB/SB0	Final SB/SB_F0	CPUE log-RMSE	Max KKT gradient
15F	15	5x	872,989	0.197	0.182	0.111	0.0018
6F	6	10x	1,023,476	0.238	0.244	0.116	0.0088

11 Figures

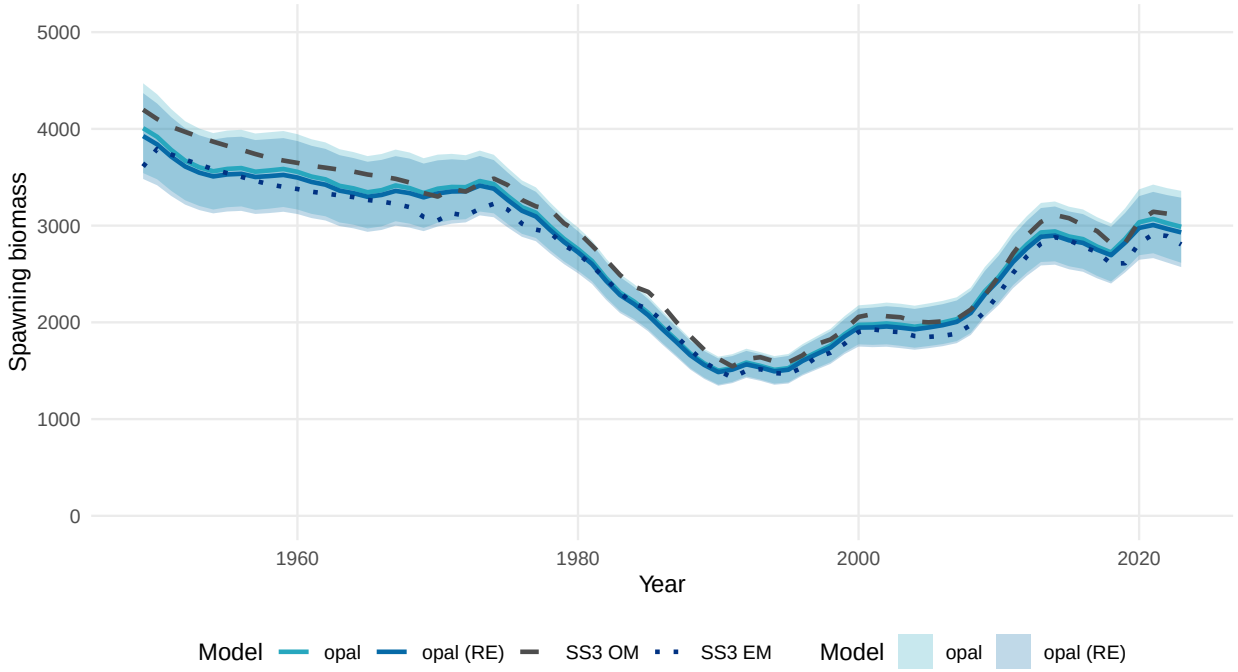

Figure 1: Estimated spawning-biomass trajectory for the 'opakapaka `opal` model estimated using either fixed effects or random effects (RE) compared with the SS3 operating model (OM) and estimation model (EM), 1949–2023.

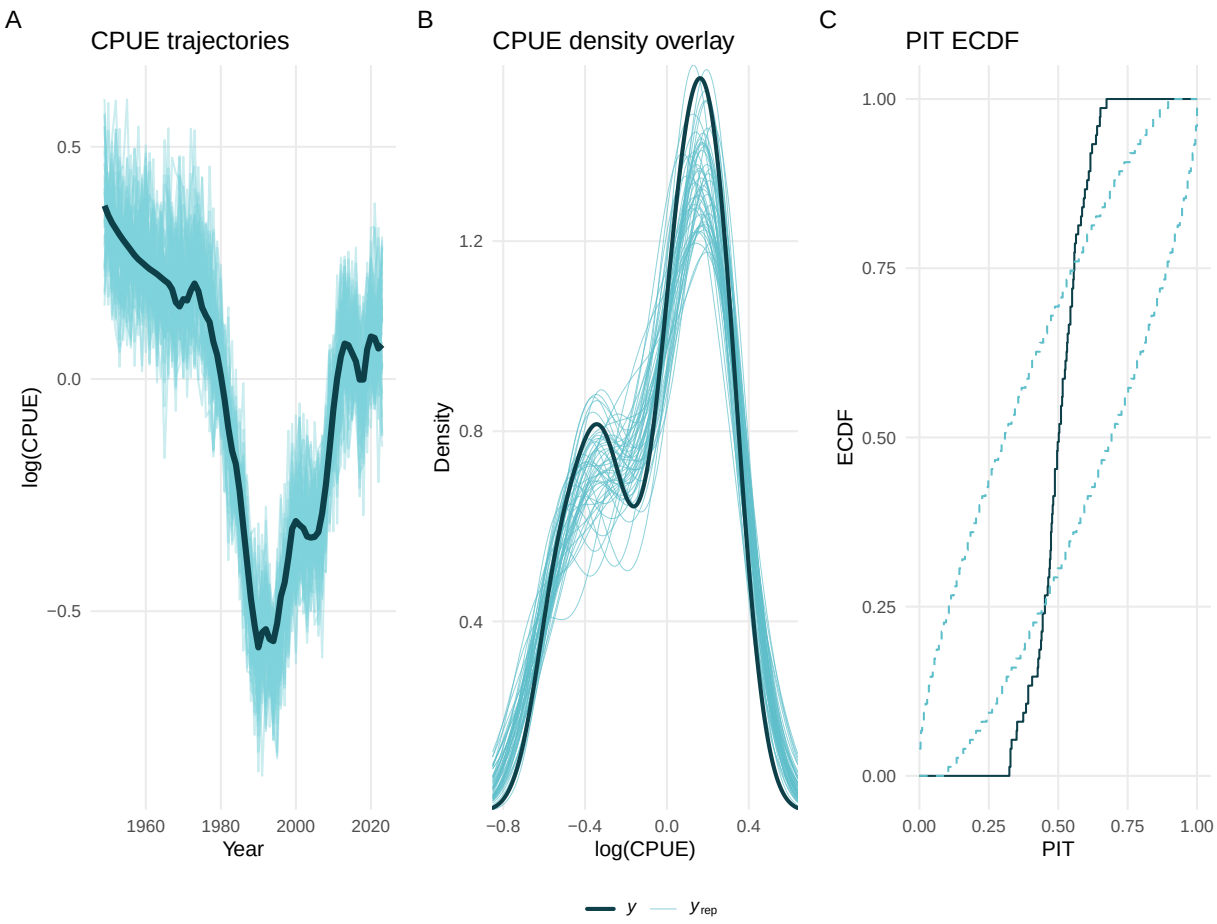

Figure 2: Posterior predictive checks using the `bayesplot` package for the 'opakapaka commercial CPUE index (fishery 1) under the opal model. Simulated replicates in aqua, observations in dark teal. Left to right: replicate trajectories, density overlay, and probability integral transform (PIT) empirical CDF (ECDF) calibration check.

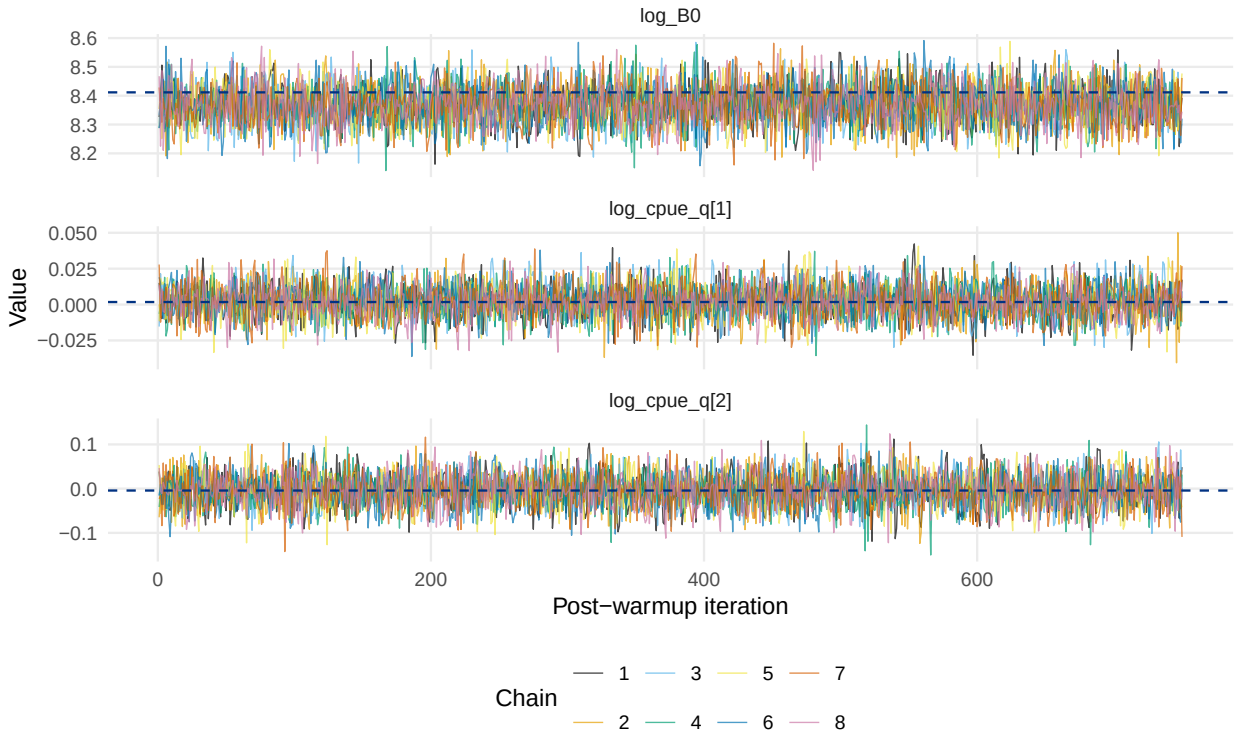

Figure 3: MCMC traces for the scalar parameters of the 'opakapaka opal (with LF) model — unfished spawning biomass and the two CPUE catchabilities, on the log scale — across eight chains (colours). Navy dashed lines mark the MLE. Overlapping, stationary chains indicate good mixing.

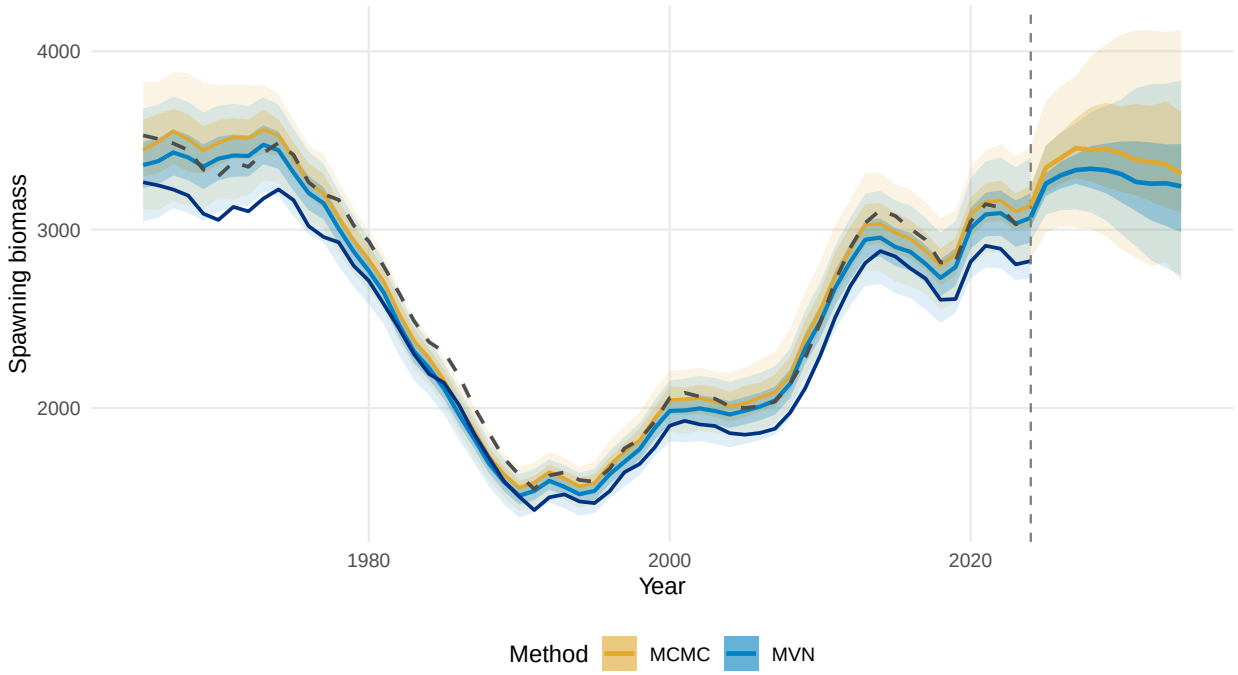

Figure 4: 'Opakapaka spawning biomass: historical uncertainty and a 10-year constant-catch projection (200 draws per method); the last 60 years are shown. Teal solid = estimated MLE; grey dashed = SS3 operating model; navy = SS3 estimation model. Coloured ribbons give the 50% (darker) and 90% (lighter) intervals for the MVN and MCMC approaches, and the amber line is the MCMC historical median. The vertical dashed line marks the end of the historical period (2024).

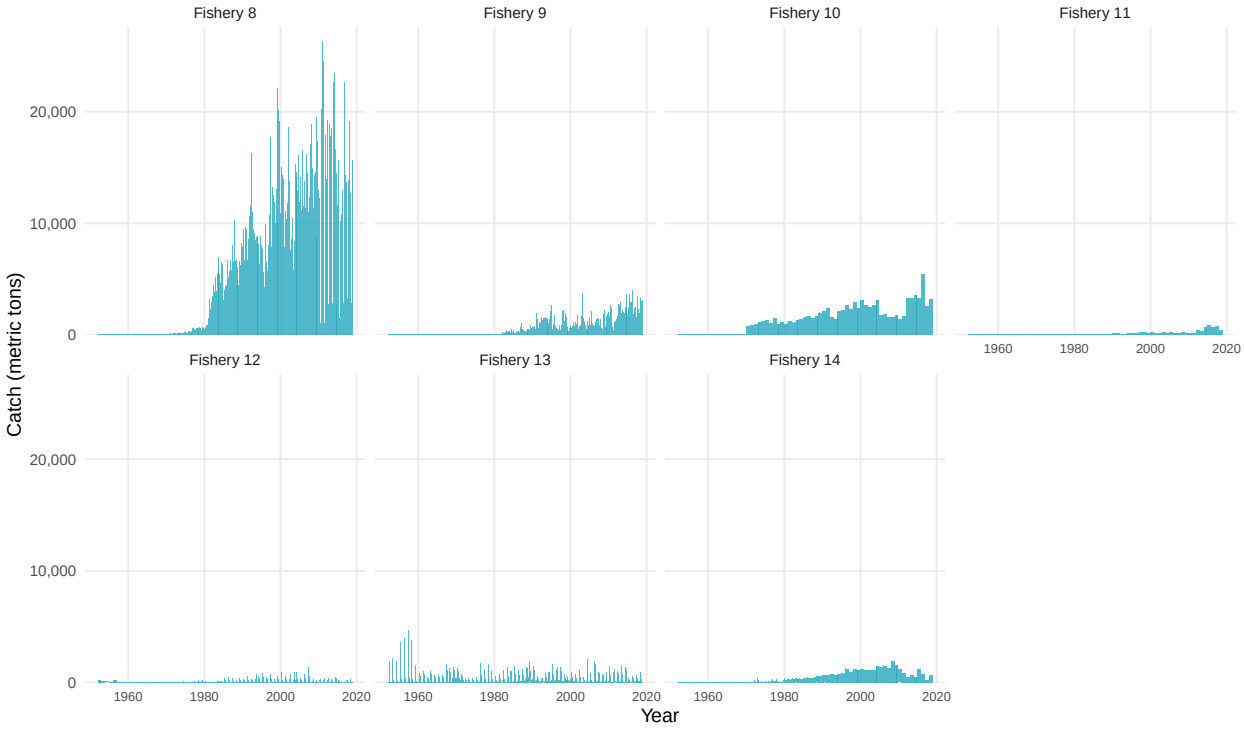

Figure 5: Observed quarterly catch for WCPO bigeye tuna fisheries with catch recorded in metric tons. Facets use a common y-axis scale across the weight-catch fisheries.

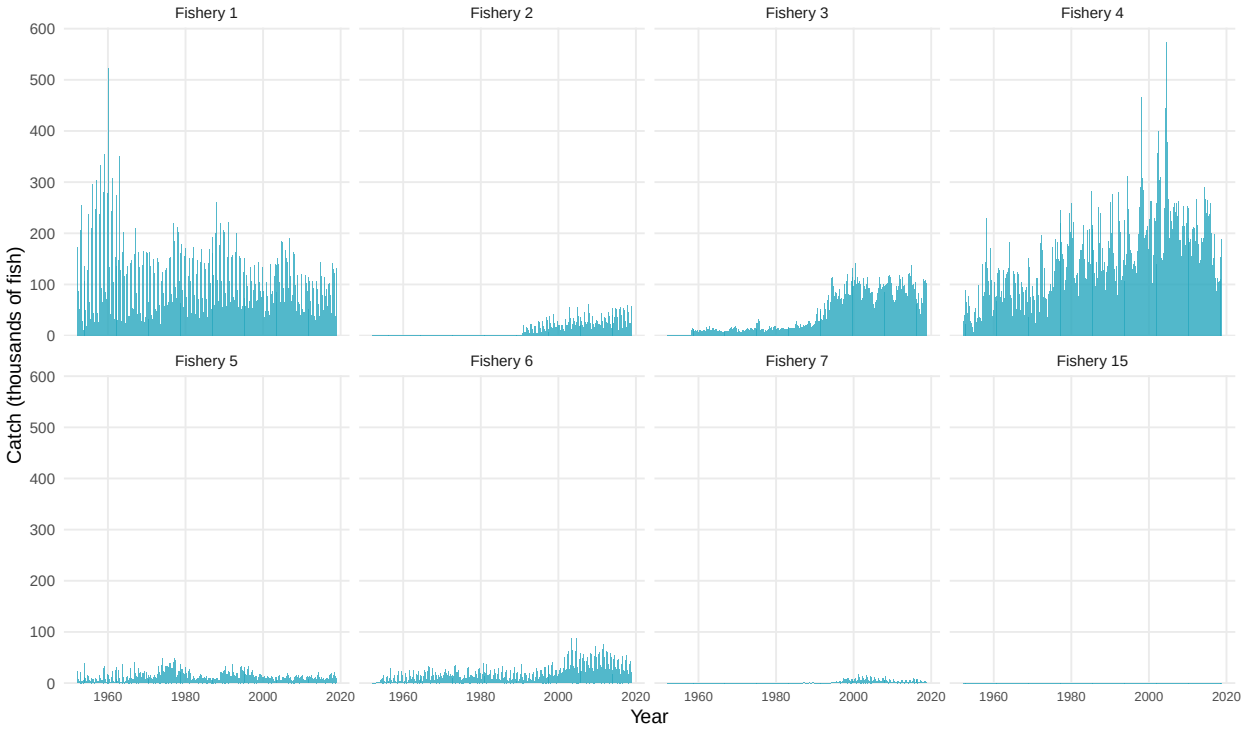

Figure 6: Observed quarterly catch for WCPO bigeye tuna fisheries with catch recorded in thousands of fish. Facets use a common y-axis scale across the numbers-catch fisheries.

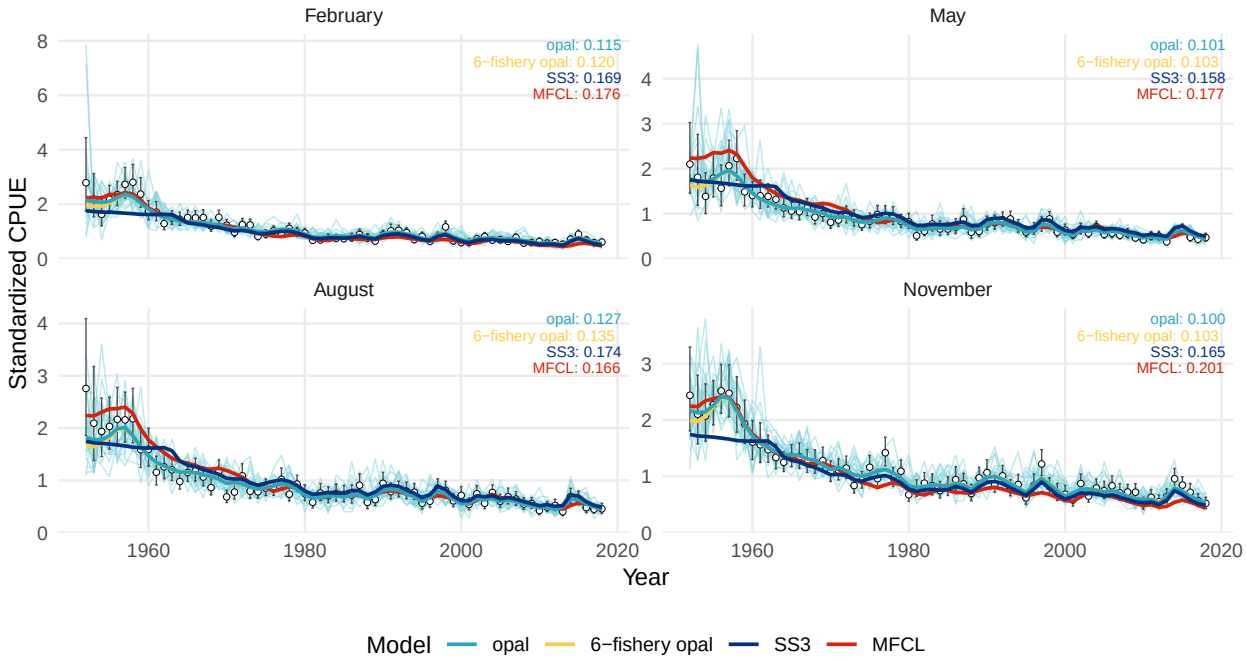

Figure 7: Observed, predicted, and simulated standardized CPUE for the fitted `opal` WCPO bigeye tuna diagnostic model, shown by index month. Vertical error bars show one fitted standard deviation on the log scale; light teal lines show fitted-model simulations from the retained 15-fishery model. The matched six-fishery aggregation sensitivity and illustrative fits by MFCL and SS3 are also shown. The RMSE of the log-residuals is shown for each model and index month.

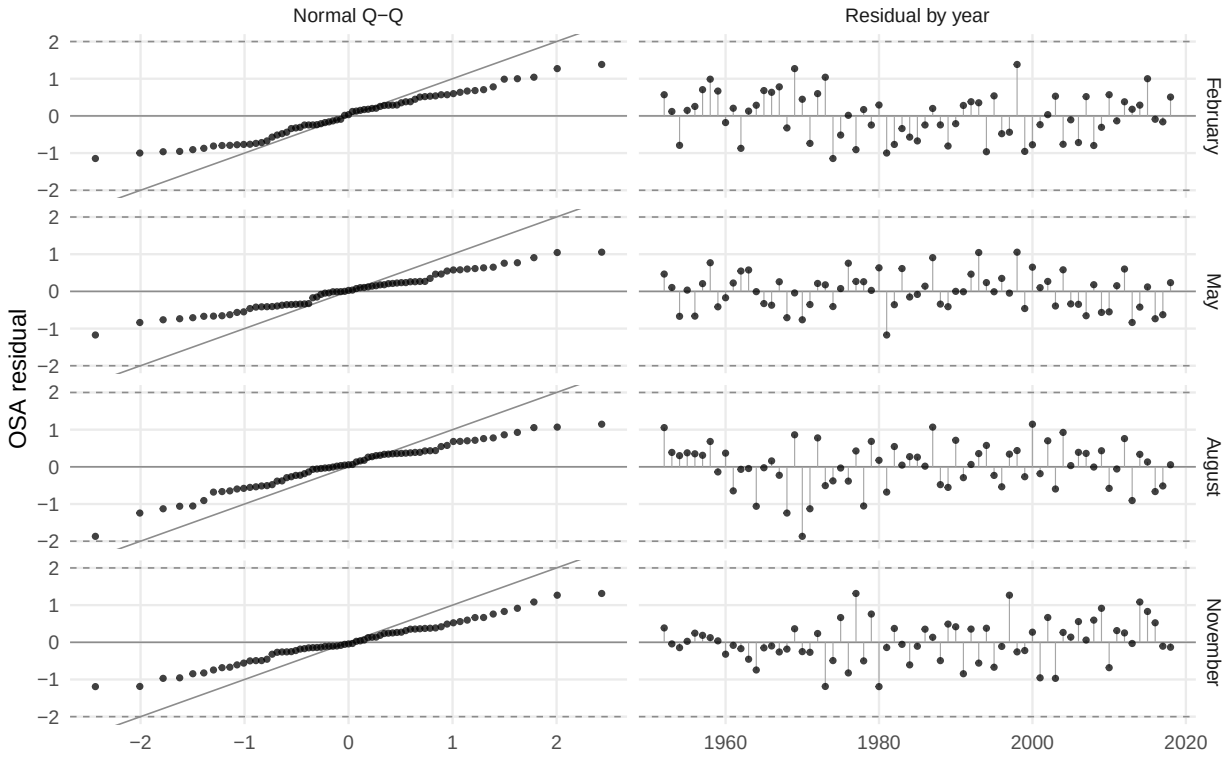

Figure 8: OSA residual diagnostics for the CPUE component of the fitted WCPO bigeye tuna diagnostic model, shown by index month. For this independent lognormal CPUE likelihood, the OSA residuals are equivalent to standardized log residuals; the flattened Q-Q pattern shows under-dispersion relative to the fitted CPUE observation variance.

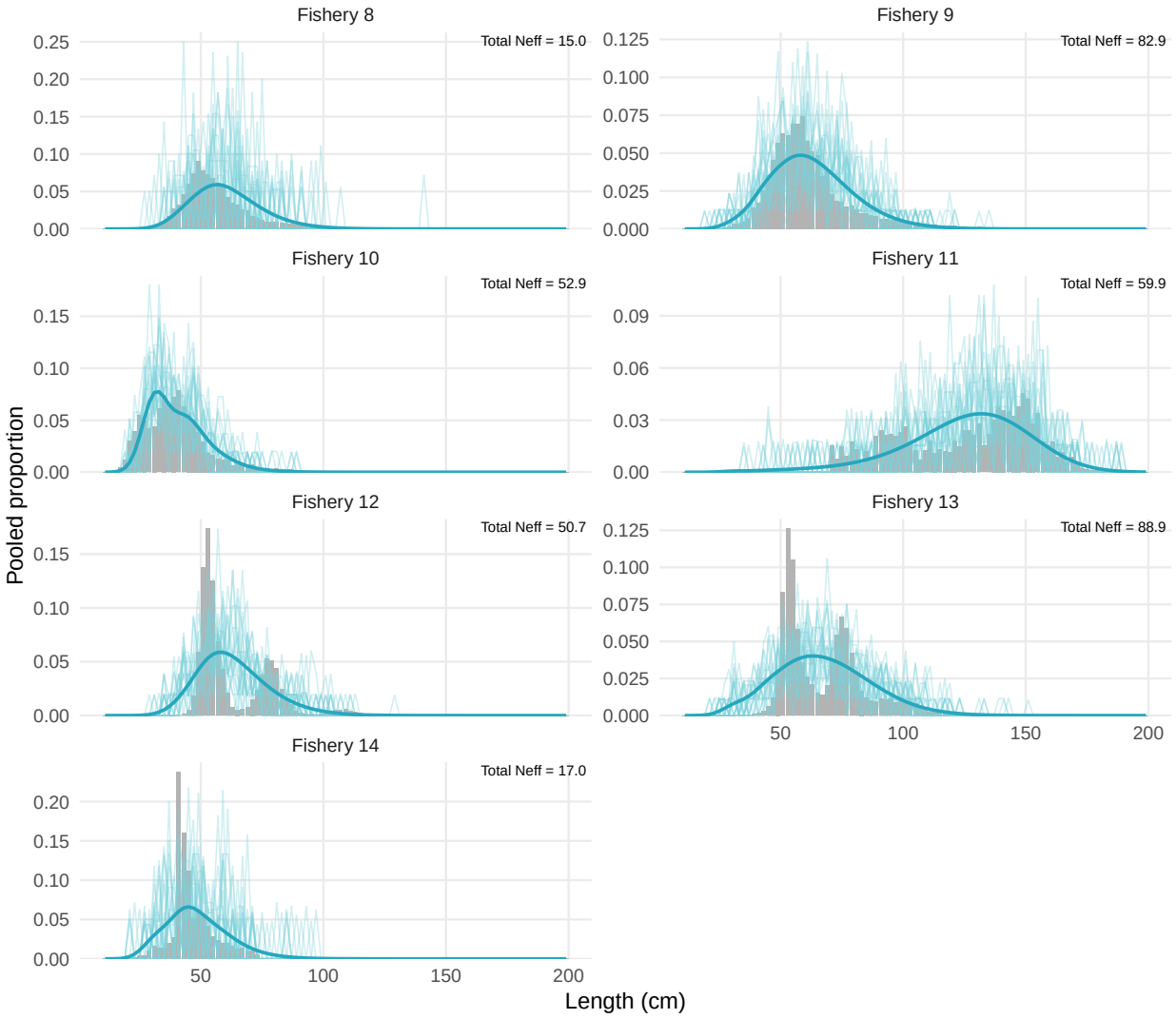

Figure 9: Observed, predicted, and simulated length-composition proportions for WCPO bigeye tuna fisheries 8-14, summed over all years using the variance-adjusted effective sample sizes. Each simulated line independently simulates the fitted row-level compositions and then pools those simulated counts over all years within fishery.

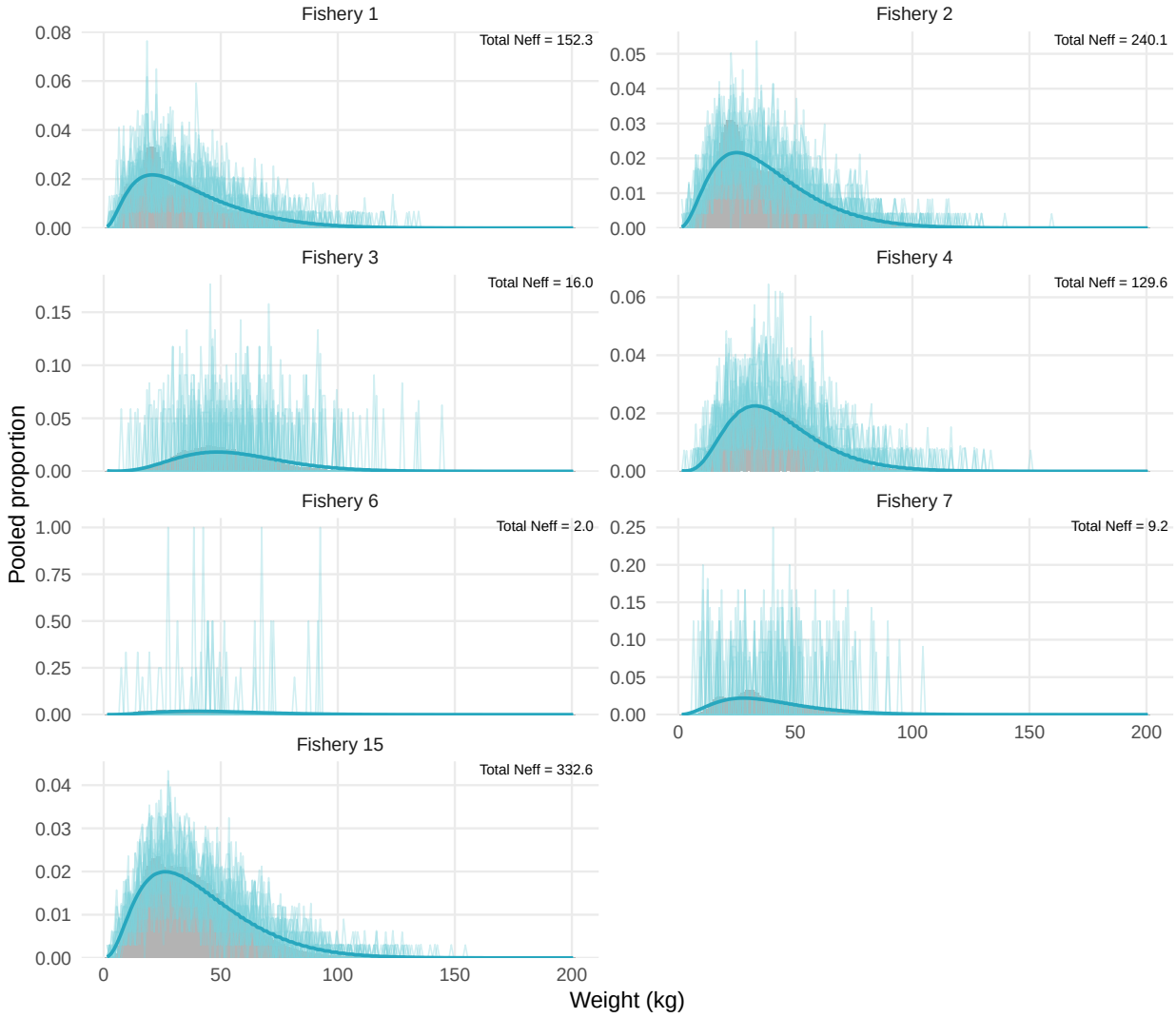

Figure 10: Observed, predicted, and simulated weight-composition proportions for active WCPO bigeye tuna weight-composition fisheries, summed over all years using the variance-adjusted effective sample sizes. Each simulated line independently simulates the fitted row-level compositions and then pools those simulated counts over all years within fishery.

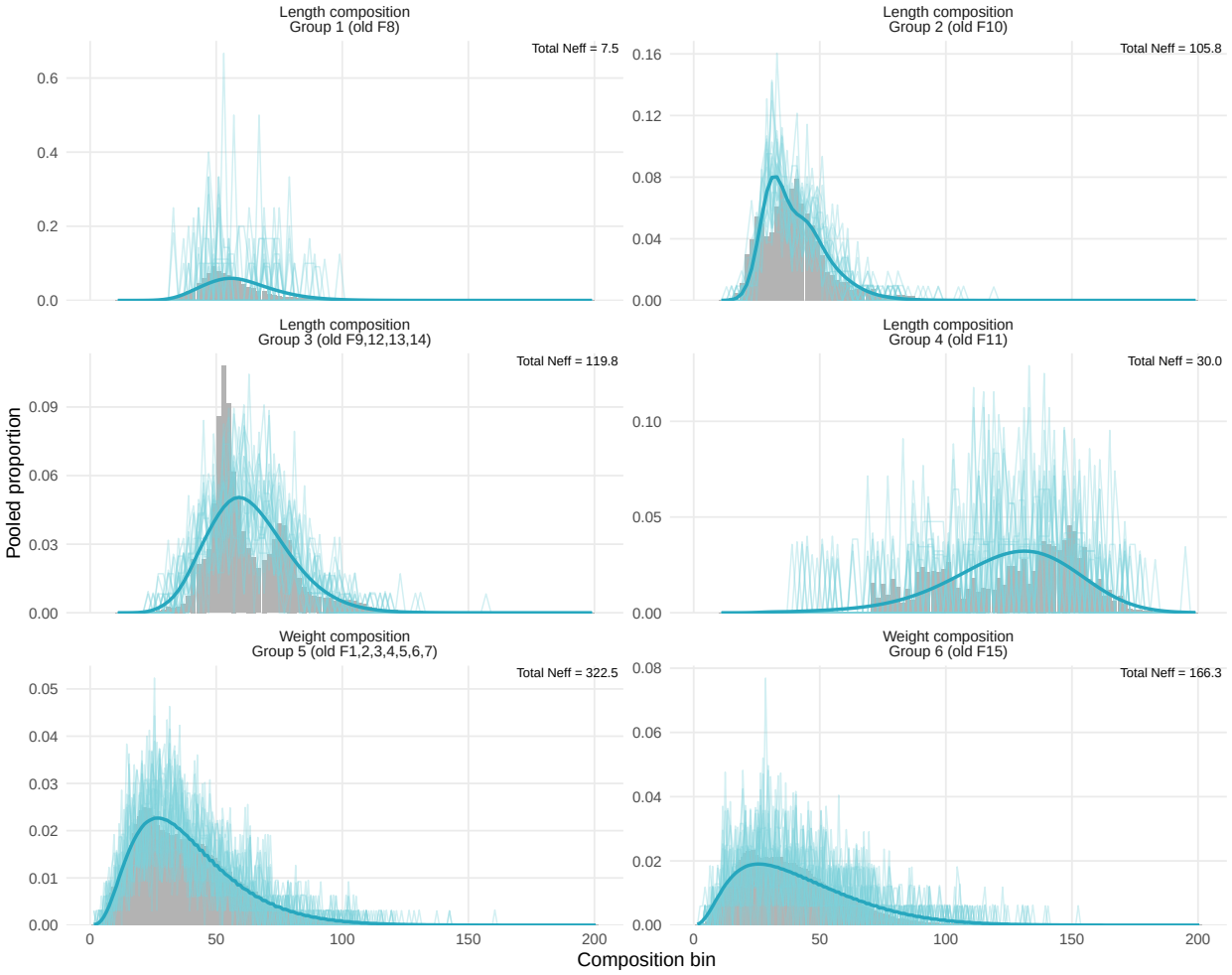

Figure 11: Observed, predicted, and simulated composition proportions for the matched six-fishery WCPO bigeye tuna aggregation sensitivity, summed over all years using the variance-adjusted effective sample sizes. Length-composition groups are shown for groups 1–4 and weight-composition groups for groups 5–6; facet labels identify the original BET fisheries pooled into each group. Each simulated line independently simulates the fitted row-level compositions and then pools those simulated counts over all years within group.

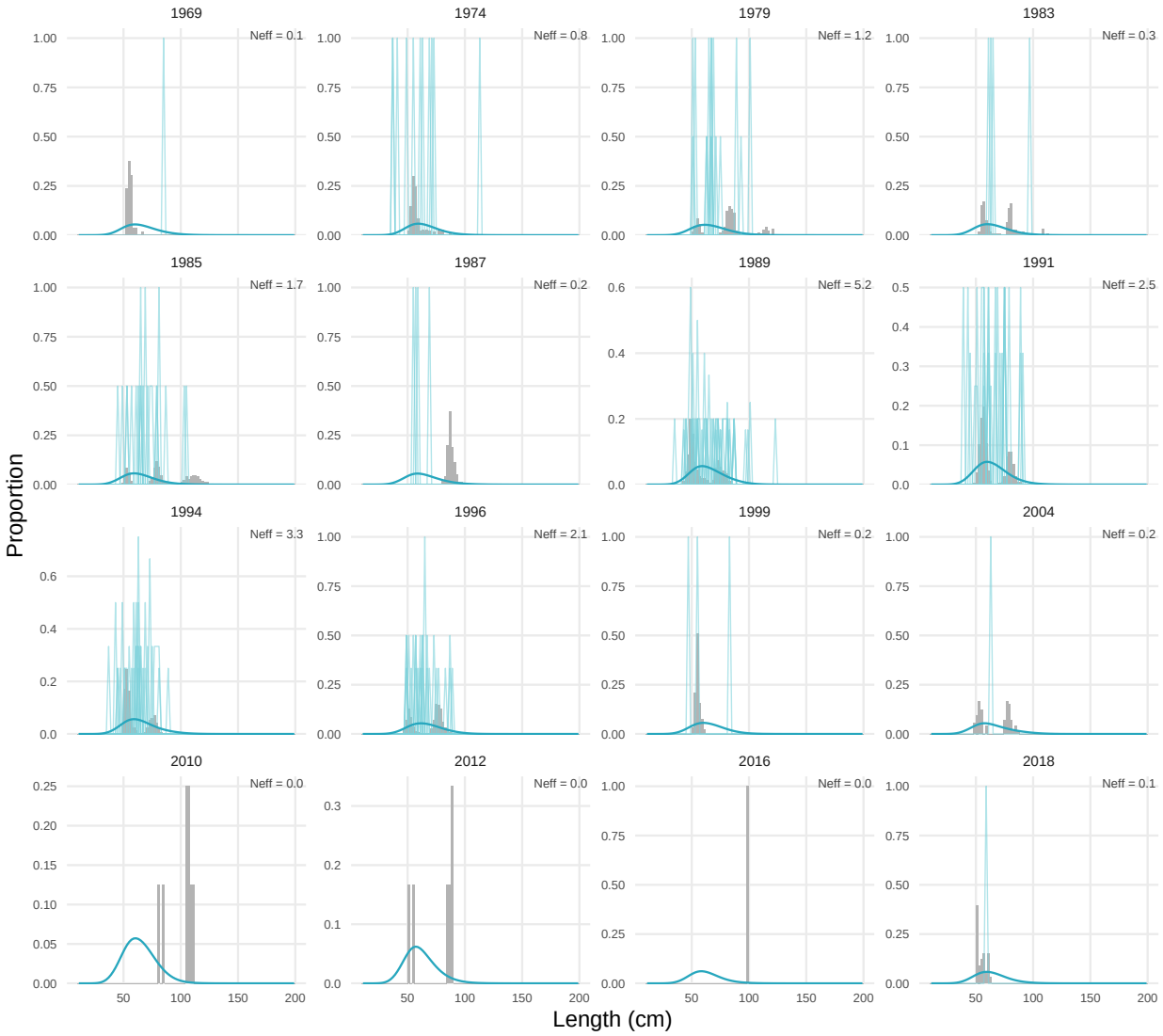

Figure 12: Year-specific observed, predicted, and simulated length-composition proportions for selected years of WCPO bigeye tuna fishery 12. Unlike the pooled length-composition diagnostics in Figure 9, this plot does not aggregate over all years; each panel has its own fitted and simulated composition. Panel text gives the variance-adjusted effective N used by the likelihood for that year.

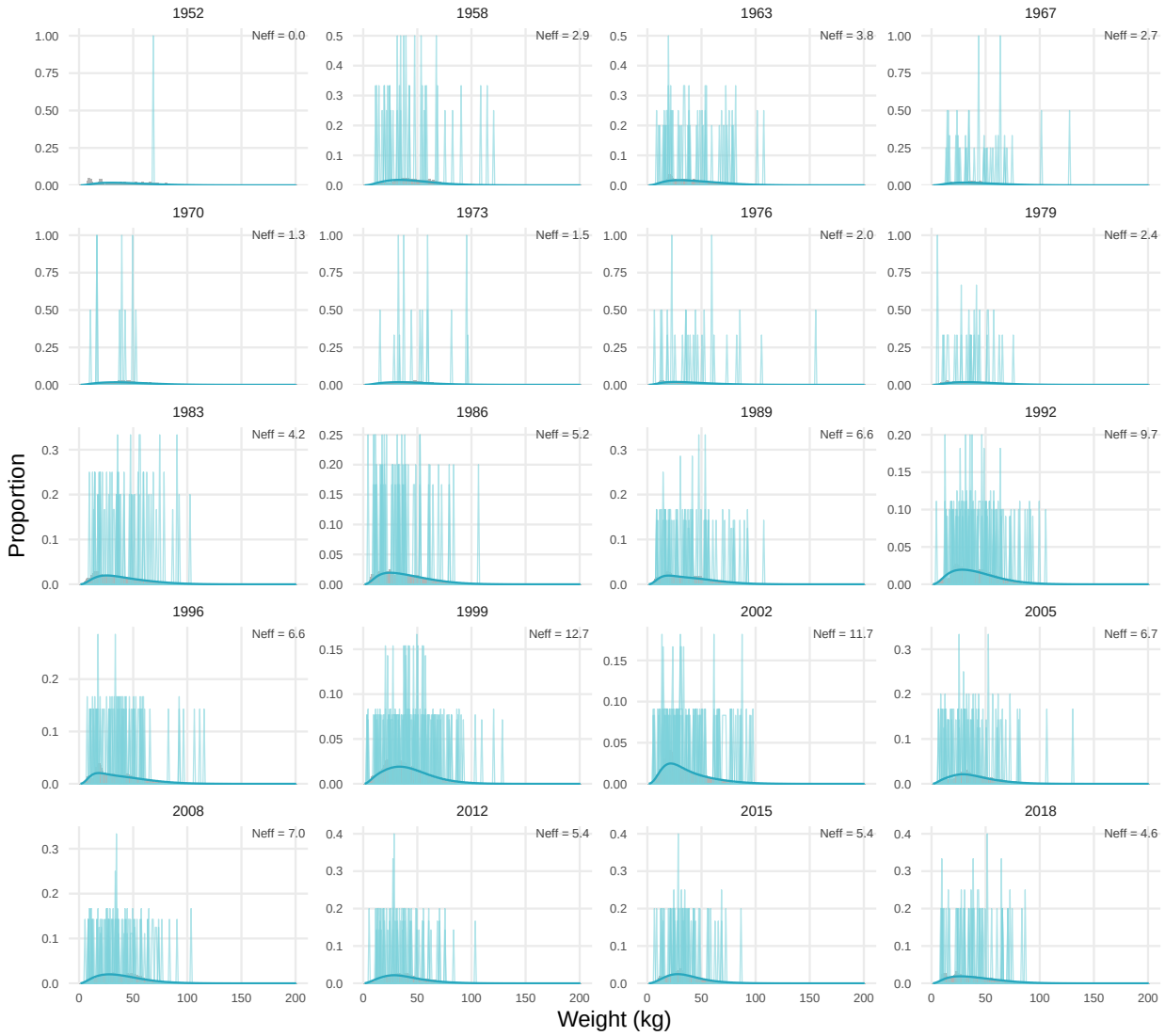

Figure 13: Year-specific observed, predicted, and simulated weight-composition proportions for selected years of WCPO bigeye tuna fishery 15. Unlike the pooled weight-composition diagnostics in Figure 10, this plot does not aggregate over all years; each panel has its own fitted and simulated composition. Panel text gives the variance-adjusted effective N used by the likelihood for that year.

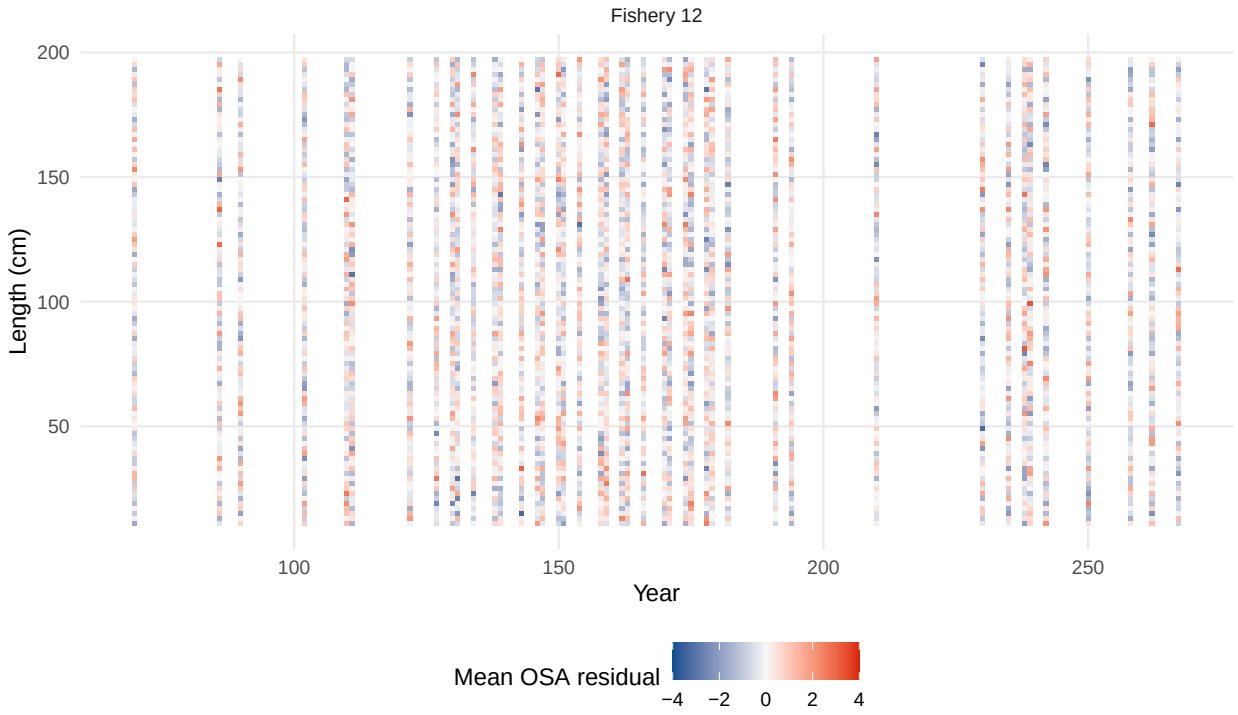

Figure 14: Length-composition OSA residual diagnostics for fishery 12 in the fitted WCPO bigeye tuna diagnostic model. Colours show the mean OSA residual in each year-length cell, with blue indicating negative residuals and red indicating positive residuals.

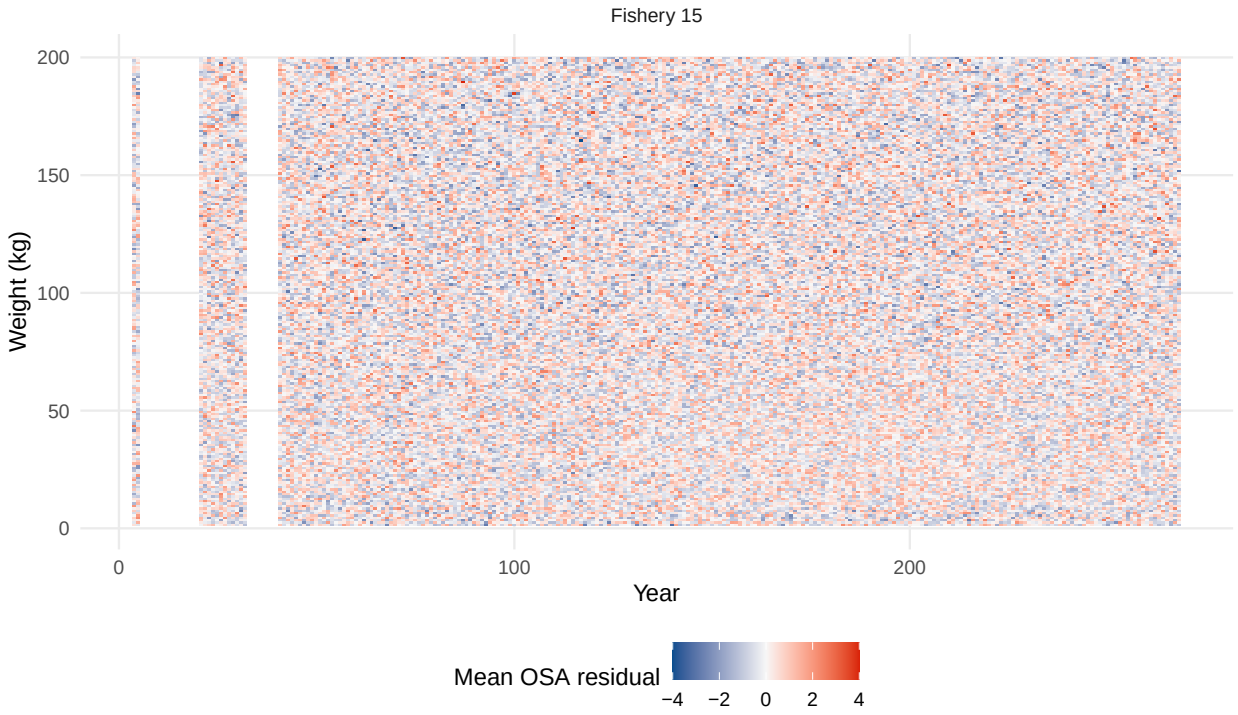

Figure 15: Weight-composition OSA residual diagnostics for fishery 15 in the fitted WCPO bigeye tuna diagnostic model. Colours show the mean OSA residual in each year-weight cell, with blue indicating negative residuals and red indicating positive residuals.

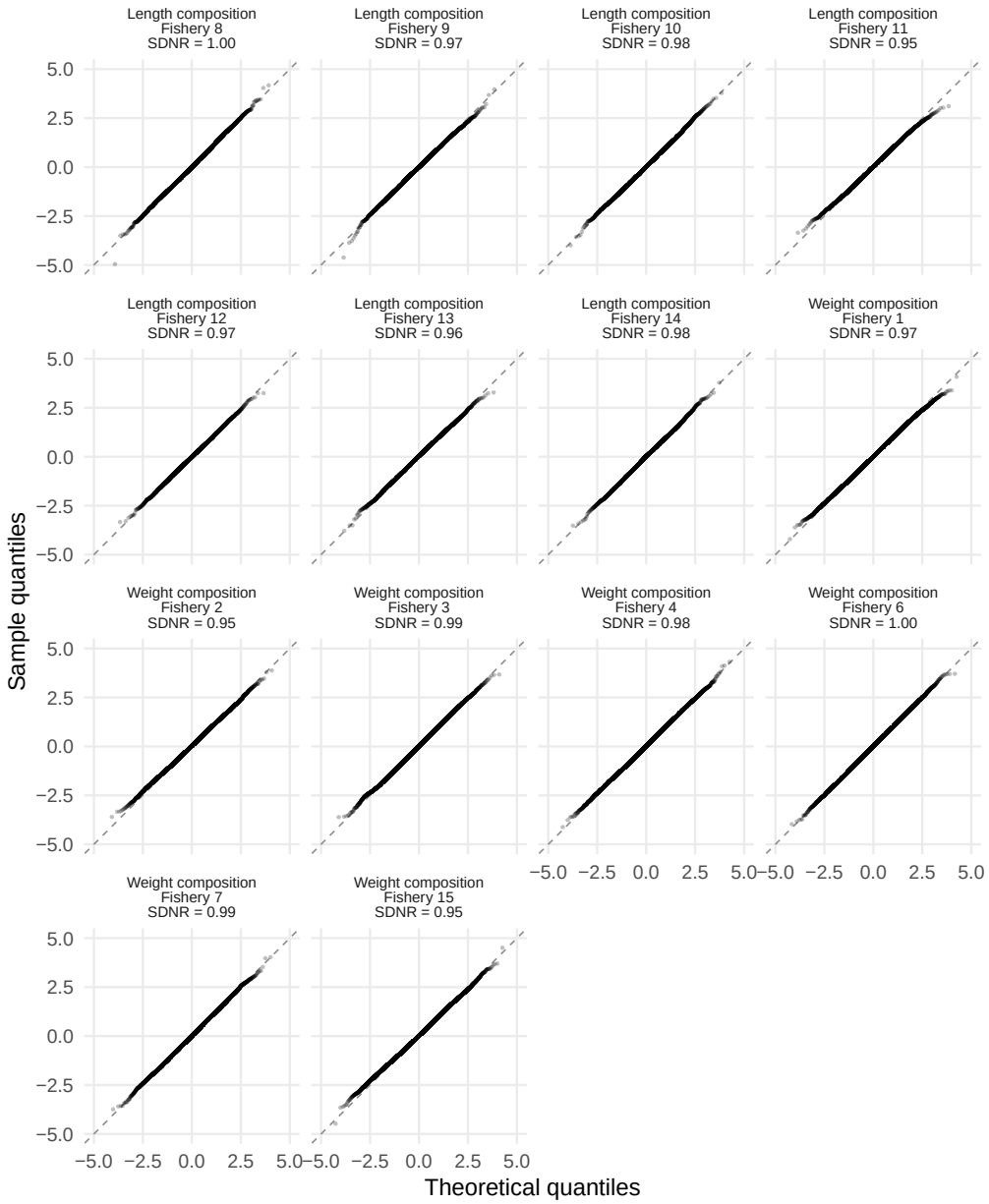

Figure 16: Normal Q-Q diagnostics for length- and weight-composition OSA residuals in the fitted WCPO bigeye tuna diagnostic model, shown separately by composition type and fishery. The grey line is the fixed standard-normal one-to-one reference; it is not fitted separately within panels.

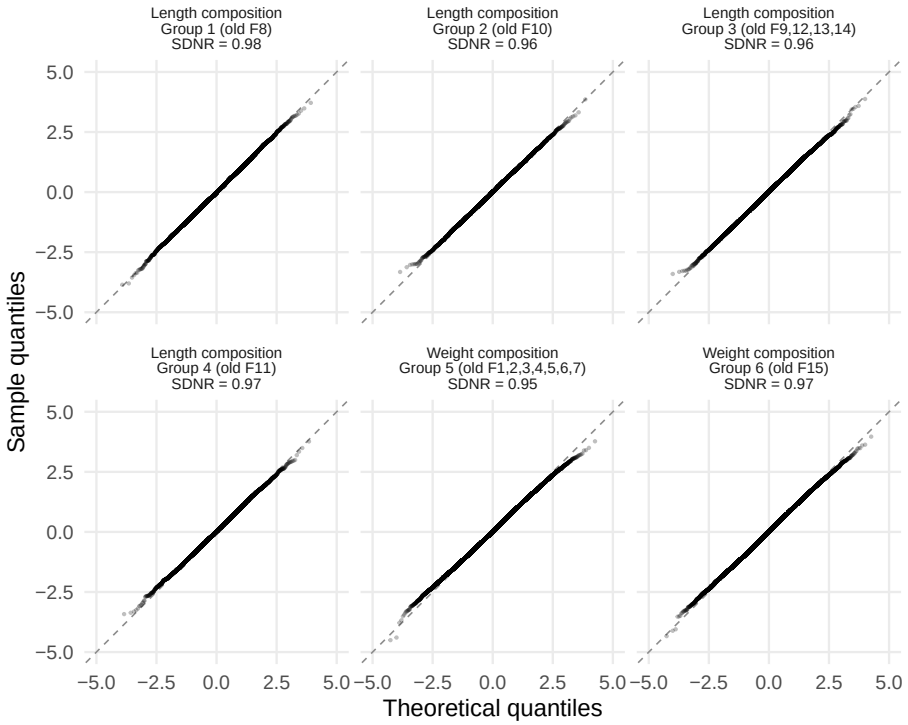

Figure 17: Normal Q-Q diagnostics for grouped length- and weight-composition OSA residuals in the matched six-fishery WCPO bigeye tuna aggregation sensitivity. These panels repeat the layout of Figure 16 using finite bin-level `compResidual::resMulti()` residuals from the grouped observed and predicted compositions; the grey line is the fixed standard-normal one-to-one reference.

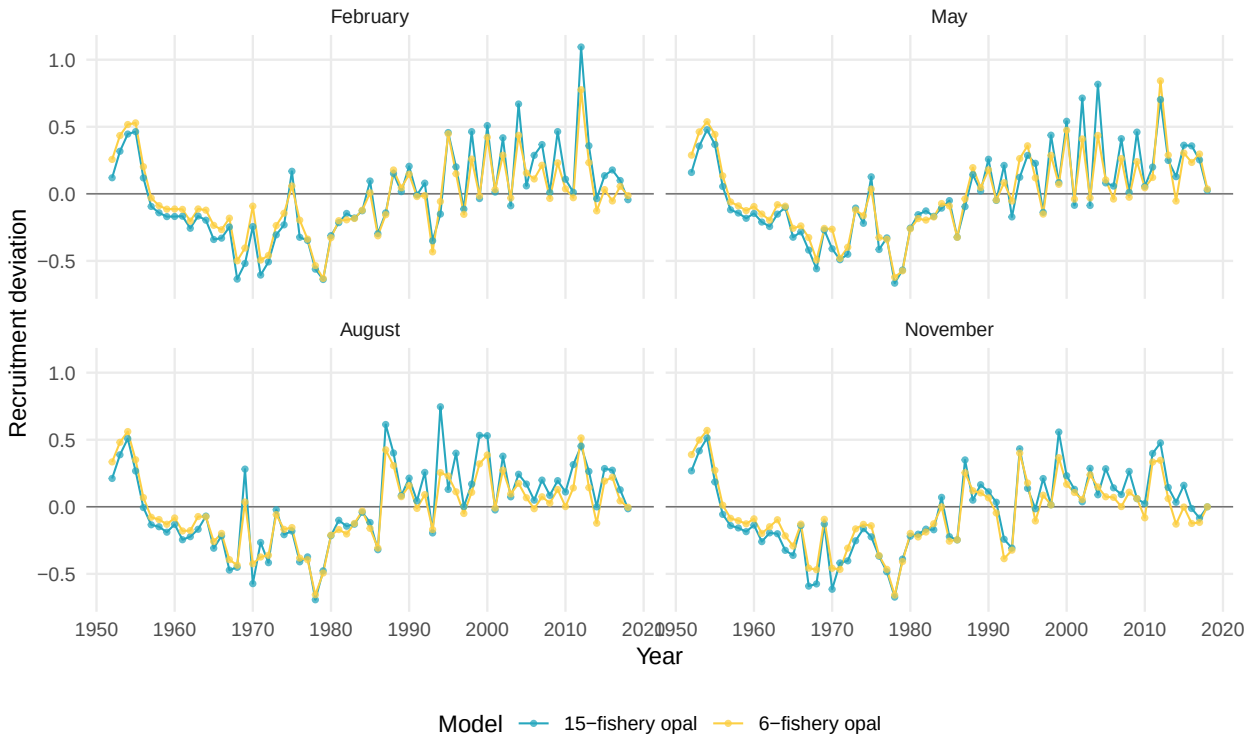

Figure 18: Estimated recruitment deviations for the fitted WCPO bigeye tuna diagnostic model and matched six-fishery aggregation sensitivity, shown by seasonal index month.

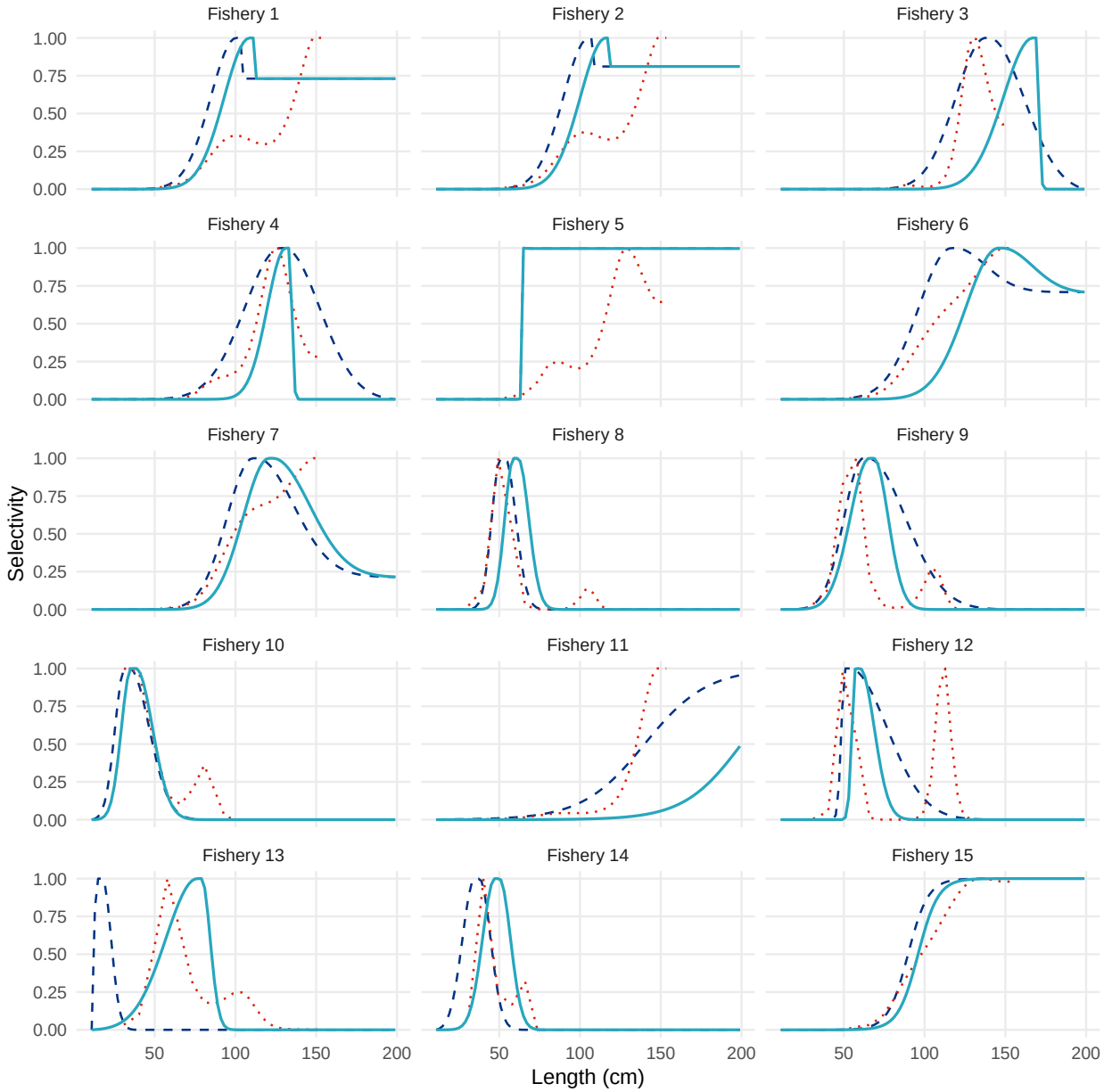

Figure 19: Selectivity-at-length curves for WCPO bigeye tuna fisheries. Teal lines show the fitted opal selectivity curves; dashed navy lines show the SS3 reference selectivity curves used to initialise the opal model; dotted coral lines show the MFCL reference selectivity curves.

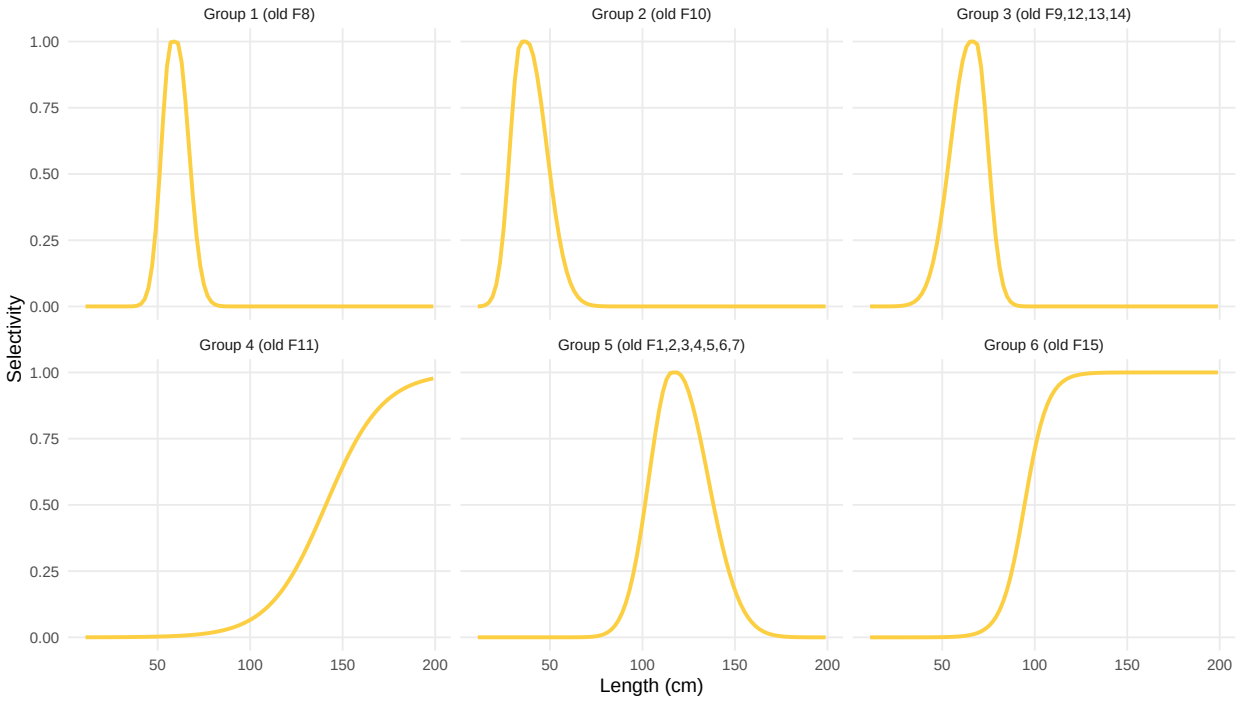

Figure 20: Estimated selectivity-at-length curves for the six-fishery WCPO bigeye tuna aggregation sensitivity. Facet labels identify the original BET fisheries pooled into each aggregation group.

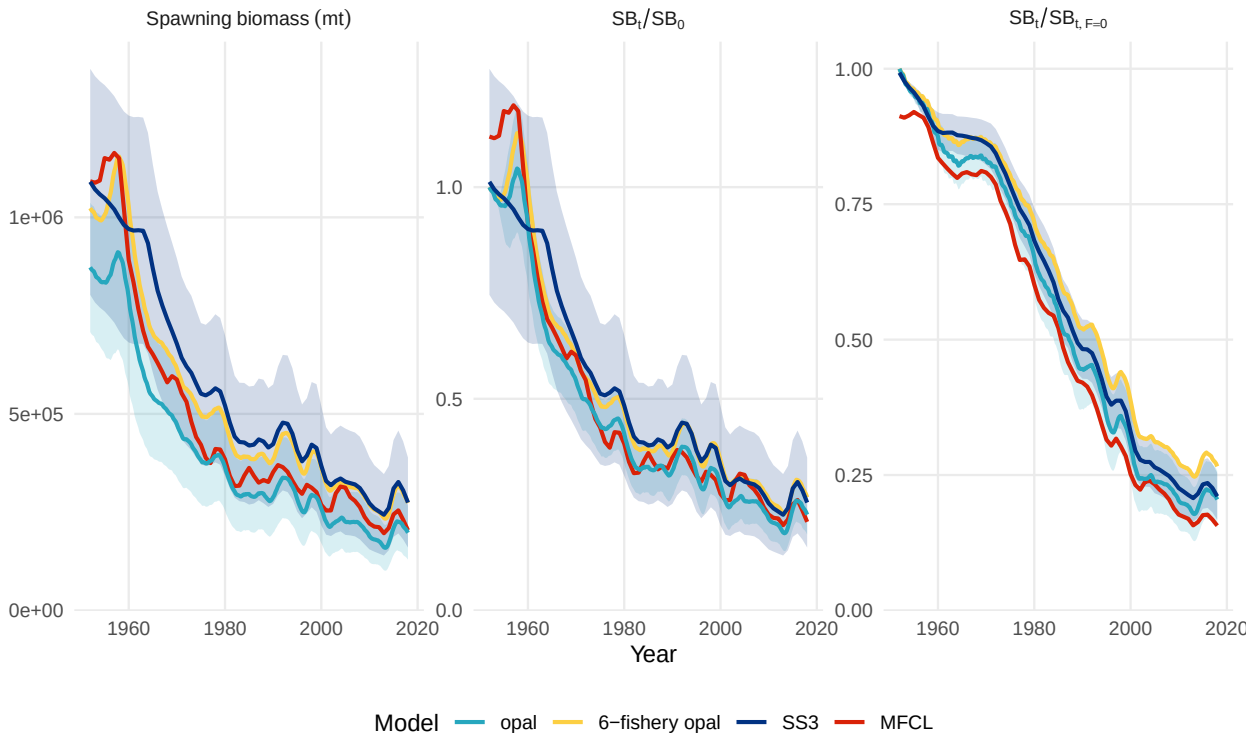

Figure 21: Estimated spawning biomass and relative biomass trajectories for the fitted WCPO bigeye tuna diagnostic model and the matched six-fishery aggregation sensitivity. Shaded ribbons show approximate 95% intervals from model-based standard errors; the retained `opal` ribbon uses `RTMB::sdreport()` with a supplied positive-definite fixed-effect Hessian. The `opal` models are compared with the SS3 and MFCL reference models; external-model SB/SB0 uses the same median SSB/depletion reference scale reported in the text.