Western and Central Pacific Fisheries Commission

SCIENTIFIC COMMITTEE
TWENTY-FIRST REGULAR SESSION

Nuku'alofa, Tonga
13-21 August 2025

2025 Stock Assessment of Striped Marlin in the Southwest Pacific Ocean: Part II Data-moderate approach

WCPFC-SC21-2025/SA-WP-07 rev1
01 August 2025

N. Ducharme-Barth[1], **C. Castillo-Jordán**[2], **F. Carvalho**[1], **P. Hamer**[2]

[1]NOAA Fisheries, Pacific Islands Fisheries Science Center, 1845 Wasp Boulevard, Bldg. 176, Honolulu HI 96818
[2]Oceanic Fisheries Programme of the Pacific Community

Table of contents

1 Executive summary

Rev 1: In the initial version, the *median* values listed in Table 8 actually reported the *mean* from the posterior distribution. In the current revision this has been fixed, and the median values are reported.

This working paper presents part II of the 2025 stock assessment of Southwest Pacific Ocean (SWPO) striped marlin (*Kajikia audax*) and describes a data-moderate Bayesian surplus production model (BSPM) approach. This assessment represents a strategic shift from the integrated age-structured models used previously, given challenges with data conflicts, poor fits to size composition data, and difficulties in estimating model initial conditions in the integrated age-structured model. The complexity of integrated age-structured models becomes problematic when fundamental data conflicts exist or when key biological parameters are uncertain, making the simplified yet robust BSPM framework a reasonable alternative. In combination with information from the integrated stock assessment, the BSPM can provide a holistic view on likely stock status and population trajectory.

In order to evaluate stock status, a series of BSPMs spanning 1952-2022 were developed following the Fletcher-Schaefer production model framework, incorporating biological uncertainty through simulation-based priors, process error in population dynamics, and observation error in abundance indices and catch data. Informative priors for maximum intrinsic rate of increase R_{Max} and carrying capacity $\log(K)$ were developed through numerical simulation based on biological and fishery characteristics, with prior distributions refined using pushforward analysis to ensure biological realism. The assessment utilized annual catch data in numbers aggregated into total removals and fitted to standardized CPUE indices, primarily the distant water fishing nation (DWFN) longline index and New Zealand recreational sportfish index. An extensive sensitivity analysis including 38 one-off models was conducted to explore uncertainties in data inputs, model structure, and biological assumptions, with a four-model ensemble ultimately proposed for the provision of management advice.

Given that a surplus production modeling approach was used, conventional WCPFC management metrics related to spawning biomass (SB) were unable to be produced with the surplus production model. Rather than metrics related to SB, stock status was summarized in terms of the following maximum sustainable yield (MSY) and depletion D based reference points: total depletion relative to total depletion at which MSY is produced (D/D_{MSY}), fishing mortality relative to fishing mortality that produces MSY (F/F_{MSY}), and total depletion relative to a generalized limit reference point of 20% depletion from the unfished state ($D/D_{0.2F=0}$). All ratios were calculated with respect to two periods: *recent* and *latest*. For depletion ratios, D_{recent} refers to the average over 2019-2022 and D_{latest} refers to 2022. For fishing mortality ratios, F_{recent} refers to the average over 2018-2021 and F_{latest} refers to 2021. Additionally, note that in this report D is the static depletion of total population numbers relative to the unfished total population numbers, rather than the time-dynamic depletion reported by age-structured models (e.g., MULTIFAN-CL or Stock Synthesis)

Key uncertainties and limitations of the assessment include:

- **Population scale**: Substantial uncertainty in absolute population scale, though minimum scale is well-constrained by the 70-year catch history
- **Stock structure**: Potential stock connectivity issues given genetic evidence of SWPO fish in North Pacific catches

- **Data representativeness**: Uncertainty in how well available abundance indices represent true stock trends and whether catch reporting has been complete over the assessment period
- **Model structure**: The BSPM approach inherently simplifies complex age-structured population dynamics, assuming a single well-mixed population with knife-edged selectivity
- **Biological parameters**: Uncertainty in key biological parameters (e.g., natural mortality, steepness, and growth) propagates through prior distributions and contributes to wide credible intervals around stock status estimates

The general conclusions of this assessment are as follows:

- **Population trajectory**: The SWPO striped marlin stock declined substantially from the assumed unfished state in 1952, reaching minimum levels around the mid-2010s, with strong evidence of recovery since approximately 2015
- **Current stock status**: The stock is estimated to be overfished but not undergoing overfishing, with recent depletion relative to D_{MSY} at 0.77 (95% CI: 0.33 – 2.3) and recent fishing mortality relative to F_{MSY} at 0.77 (95% CI: 0.05 – 1.51)
- **Stock status probabilities**: There is a 74% probability that the stock is below D_{MSY} and only a 22.9% probability that overfishing is occurring
- **Future projections**: Ten-year projections assuming recent average catch levels indicate continued recovery, with median D/D_{MSY} projected to reach 1.32 by 2032 and only a 26.05% chance of remaining overfished

The assessment provides the following recommendations:

- **Stock structure research**: Development of a conceptual model for SWPO striped marlin, potentially in collaboration with the ISC Billfish Working Group, to address stock connectivity given genetic evidence of SWPO fish in North Pacific catches
- **Biological parameter refinement**: Reduction of uncertainty in growth, maturity, natural mortality, and steepness parameters through targeted research, and data collection
- **Future modeling approach**: Progressive development within a Bayesian framework toward full age-structure through delay-difference models, age-structured production models, and fully integrated age-structured models
- **Management strategy evaluation (MSE)**: Development of robust management measures through MSE approaches can provide utility given substantial uncertainties in model inputs

2   Introduction

Stock assessments of Southwest Pacific Ocean (SWPO) striped marlin (*Kajikia audax*) have presented significant technical challenges. As documented in the joint NOAA-SPC modeling meeting report (Ducharme-Barth et al., 2025), key issues identified by WCPFC SC20 included poor fits to size composition data and relative abundance indices, conflicts between different data sources, and difficulties in estimating model initial conditions.

Discussions at the 2025 SPC Pre-Assessment Workshop highlighted additional challenges related to the estimation of total population scale. The low biomass scale indicated by previous versions of the assessment appeared at the time to largely be driven by the size composition from multiple fisheries. While estimation of low biomass scale in these preliminary models (Ducharme-Barth et al., 2025) was not inherently problematic, the resulting high ratio of fishing mortality (F) to natural mortality (M) that was maintained over multiple decades in the face of relative stability in catches and catch-per-unit-effort (CPUE) was noted as suspicious and flagged for further investigation.

The complexity of integrated age-structured models, while providing detailed population dynamics representation, can become problematic when fundamental data conflicts exist or when key biological parameters are uncertain. For SWPO striped marlin, these challenges are compounded by spatiotemporal heterogeneity in fleet coverage, potential non-representativeness in mixed-fleet composition data, and limitations in age data from opportunistic sampling. Changes to productivity assumptions (growth, natural mortality, maturity, and/or steepness) or selectivity patterns will impact biomass scaling through predictions of expected size composition data.

Scaling back model complexity to a data-moderate approach, like Bayesian surplus production models (BSPMs), offers analysts a simplified yet robust alternative for stock assessment when data limitations or conflicts present challenges for more complex models (Kokkalis et al., 2024). These models have proven effective in recent pelagic fish stock assessments (Winker et al., 2018), and are routinely applied for WCPFC shark assessments (ISC, 2024; Neubauer et al., 2019, 2024). They facilitate a more tractable exploration of model assumptions by distilling productivity and fishing assumptions into a restricted parameter subset. The BSPM framework allows for explicit incorporation of parameter uncertainty through informative priors while maintaining computational efficiency and interpretability.

One advantage of the Bayesian approach is the use of prior information to propagate uncertainty in model assumptions. Biological simulation approaches can provide robust methods for parameterizing key population dynamics parameters (ISC, 2024; Pardo et al., 2016). For species with complex life histories like striped marlin, these simulation-based priors can incorporate uncertainty in growth, maturity, natural mortality, and reproductive parameters to derive realistic distributions for maximum intrinsic rate of increase and carrying capacity. Additionally, prior pushforward checks (Kim & Neubauer, 2025; Monnahan, 2024) can be used to further refine parameter distributions by identifying parameter combinations that produce clearly improbable outcomes.

This analysis presents a data-moderate approach for SWPO striped marlin that addresses some of the key technical challenges identified in previous assessments while providing a robust framework for stock status evaluation. The model incorporates biological uncertainty through simulation-based priors and includes approaches to address catch uncertainty and population depletion. It is important to note that this analysis

complements the 2025 revised stock assessment report (Castillo-Jordan et al., 2025). While the relative simplicity of data-moderate approaches offers some advantages relative to more data-intensive approaches like fully-integrated age-structured stock assessment models, particularly when uncertainty in data or key assumptions is high, this comes at the cost of simplifying the model fishery and population dynamics. Oversimplification of key age-structured processes, or lack of spatiotemporal specificity could result in biases in data-moderate approaches. Taking multiple modeling approaches into account can give managers a holistic view on the likely status and trajectory of the stock.

3 Methods

3.1 Modelling framework

To address some of the issues raised in Section 2, a series of Bayesian state-space surplus production models (BSPMs) spanning the period 1952-2022 were developed following the Fletcher-Schaefer production model framework (C. Edwards, 2024; Winker et al., 2019). Development of the BSPM followed the approach of (ISC, 2024; Neubauer et al., 2019) and incorporated recent best practices for surplus production models (Kokkalis et al., 2024) and Bayesian workflows for stock assessment (Monnahan, 2024) as guides for model development, analysis, and presentation.

The models were implemented in the Stan probabilistic programming language (Team, 2024b) to take advantage of enhanced convergence diagnostics, greater efficiency in posterior sampling through the no-U-turn (NUTS) Hamiltonian Monte Carlo (HMC) algorithm (Betancourt & Girolami, 2013), and greater flexibility with model configuration and prior specification. Implementation in R using the *rstan* package (Team, 2024a) provides connection to an ecosystem of R packages for model diagnostics and validation: *bayesplot* (Gabry & Mahr, 2024) for posterior visualization and *loo* (Vehtari et al., 2024).

In their baseline configuration (diagnostic case), the models begin from an assumed unfished state in 1952 and incorporate process error in population dynamics, observation error in the relative abundance index and catch time series, and uncertainty in biological parameters. Population scale or carrying capacity $\log(K)$ is expressed in the same units as the catch data, which represents total numbers of individuals. The models fit to a single time series of relative abundance (e.g., standardized CPUE index) at a time. Rather than subtracting the catch directly, multiplicative survival after fishing mortality was applied. Fishing mortality was estimated internally by fitting to a time series of observed catches. Computationally, this was necessary to avoid placing hard constraints on the posterior parameter space which can cause convergence issues for the HMC sampling algorithm. Additionally, rather than directly estimating the independent fishing mortality values required to produce the catch, fishing mortality was derived from an estimated time-varying catchability trend and a time series of nominal observed longline fishing effort. The primary rationale for this modelling choice being that it is easier to develop a defensible prior for catchability rather than an arbitrary prior for fishing mortality, given that previous work (ISC, 2024) showed a strong correlation between estimated population scale and the fishing mortality prior. Unless otherwise mentioned, all models used the same prior parameterizations (Table 2) based on the biological simulation filtering described in Section 3.5.

Using the diagnostic model[1] as a starting point, a number of one-off sensitivity analyses were conducted to

---

[1]The Stan code for the diagnostic model `bspm_estqsimple_softdep_fullmvprior_x0_sttgamma_flexsigmaC_OPT` is publicly available on GitHub.

explore key model assumptions or data questions. These are described in Section 3.6 and listed in Table 3. Based on the results of the sensitivity analyses a model ensemble was developed for the provision of stock status and management reference points. Full model development code and input data are available online at the following GitHub repository: https://github.com/N-DucharmeBarth-NOAA/2025-swpo-mls-bspm.

3.2 Input data

The BSPMs were fitted to catch and standardized CPUE data for SWPO striped marlin. Effort was used as an input to drive the time series of estimated fishing mortality via the estimated catchability trend.

Annual catch data (in numbers of individuals) spanning 1952-2022 were compiled from the 2024 assessment (Castillo-Jordan et al., 2024) sources and aggregated into total removals by year. The catch series (Figure 1) shows initially low removals in 1952-1953, followed by a high peak in 1954. Overall, however, catches have been relatively stable though catches in recent decades have shown a slight decline.

Annual longline effort (in numbers of hooks) matching the SWPO assessment domain, as specified in Castillo-Jordan et al. (2024), was extracted from publicly available WCPFC data. Longline effort in the assessment domain (Figure 1) increased gradually from zero in the early 1950s to ~100 million hooks per year in the early 1970s where it stabilized between 100-150 million hooks per year through the 1990s. Beginning in 2000 effort increased substantially to ~350 million hooks per year in 2003, and has fluctuated around that level for the last 20 years.

A number of indices were investigated either as one-off sensitivities or included in the structural uncertainty grid (Figure 2). These included the composite distant water fishing nation (DWFN) longline standardized CPUE index (1988-2022) from the diagnostic case of the 2024 assessment (Castillo-Jordan et al., 2024), an Australian longline index (Tremblay-Boyer & Williams, 2024), a New Zealand recreational sportfish index (Holdsworth, 2023), and three different observer data based indices (Neubauer et al., 2025). The Pacific Islands observer program indices contained information from New Caledonia, Fiji, Tonga, and French Polynesia observer programs. Three alternative indices were constructed either by including data from all programs together, including data from just the western programs (New Caledonia, Fiji, and Tonga) or including data just from French Polynesia. The DWFN, Australia and observer without French Polynesia indices show recent increases; the New Zealand and observer with only French Polynesia indices show recent declines; and the all observer program index shows no trend.

3.3 Population dynamics

The population dynamics follow a hybrid Fletcher-Schaefer surplus production model (C. Edwards, 2024; ISC, 2024) with state-space formulation where population depletion x_t (relative to carrying capacity K) evolves according to:

$$x_t = \left(x_{t-1} + \text{surplus production}_{t-1}\right) \times \epsilon_t \times \exp(-F_{t-1})$$

where surplus production is defined as:

$$\text{surplus production}_{t-1} = \begin{cases} R_{Max} \cdot x_{t-1} \left(1 - \frac{x_{t-1}}{h}\right), & x_{t-1} \leq D_{MSY} \\ \gamma \cdot m \cdot x_{t-1} \left(1 - x_{t-1}^{n-1}\right), & x_{t-1} > D_{MSY} \end{cases}$$

Here, R_{Max} is the maximum intrinsic rate of increase, n is the shape parameter, F_{t-1} is the fishing mortality in year $t-1$, and ϵ_t represents multiplicative process error with $\epsilon_t = \exp(\delta_t - \sigma_P^2/2)$ and $\delta_t \sim N(0, \sigma_P)$. The intermediate parameters are: $D_{MSY} = (1/n)^{1/(n-1)}$, $h = 2D_{MSY}$, $m = R_{Max} \cdot h/4$, and $\gamma = n^{n/(n-1)}/(n-1)$. MSY is the maximum sustainable yield.

Fishing mortality F_t is in turn defined as a function of fishing effort and time-varying catchability:

$$F_t = q_{eff} \times \exp(\delta_{q,t} - \sigma_{qdev}^2/2) \times E_t$$

where q_{eff} is the base catchability coefficient, E_t is the fishing effort[2] in year t, and $\delta_{q,t}$ represents temporal variation in catchability effectiveness.

The catchability deviations $\delta_{q,t}$ follow a structured temporal process organized into discrete periods of length n_{step} years. In the diagnostic model $n_{step} = 1$ which allows for annual catchability deviates across 70[3] periods p. Catchability remains constant within periods p, and evolves between periods according to an AR(1) process:

$$\delta_{q,p} = \rho \times \delta_{q,p-1} + \epsilon_{q,p} \times \sigma_{qdev} \times \sqrt{1 - \rho^2}$$

where ρ is the autocorrelation parameter, σ_{qdev} is the standard deviation of catchability innovations, and $\epsilon_{q,p} \sim N(0, 1)$. The initial period is constrained such that $\delta_{q,1} = 0$. The term $\sqrt{1 - \rho^2}$ normalizes the variance of $\delta_{q,p}$ to remain stationary at σ_{qdev}^2 regardless of the value of ρ.

For each year t, the period assignment is calculated as:

$$\text{period}(t) = \min\left(\left\lfloor \frac{t-1}{n_{step}} \right\rfloor + 1, n_{periods}\right)$$

and the catchability deviation for year t is assigned as $\delta_{q,t} = \delta_{q,\text{period}(t)}$. This fishing mortality parametrization is a simplification of the *catch-errors* formulation in MULTIFAN-CL (Fournier et al., 1998) and allows catchability to vary smoothly over time while maintaining temporal structure and avoiding overfitting to annual fluctuations in the effort-catch relationship.

Model initial conditions are defined for $t = 1$ as $x_t = x_0 \times \epsilon_t$ where x_0 is the estimated depletion relative to unfished at the start of the model period. The prior specified for x_0 in the diagnostic case assumes an unfished initial condition with some uncertainty (Table 2).

---

[2]When input in the model the raw input effort, defined as *hundred hooks* in the WCPFC database is divided by `1e6` to re-scale catchability so that it is numerically on the same scale as other estimated parameters. This improves convergence of the HMC sampling of the posterior distribution.

[3]There are 71 total periods in the model period 1952-2022 and fishing mortality is not estimated in the terminal year.

3.4 Observation model

The model fits to a standardized CPUE index using a lognormal likelihood:

$$I_t \sim \text{Lognormal}(\log(q \times x_t) - \frac{\sigma_{O,t}^2}{2}, \sigma_{O,t})$$

where catchability q is analytically derived assuming an uninformative uniform prior, and total observation error $\sigma_{O,t} = \sigma_{O,input} + \sigma_{O,add}$ combines fixed input uncertainty with an estimated additional error component.

3.4.1 Catch treatment

Population removals (catch) were calculated in each time step as:

$$\text{Removals}_{t-1} = \big(x_{t-1} + \text{surplus production}_{t-1}\big) \times \epsilon_t \times (1 - \exp(-F_{t-1})) \times \exp(\log K)$$

As described in Section 3.3, instantaneous fishing mortality F_t was transformed to a discrete proportion of the population that survived fishing, survivorship, through the exponential decay function $\exp(-F_{t-1})$. Removals are simply calculated as the complementary proportion $(1 - \exp(-F_{t-1}))$ of the total production (including surplus production and process error) that are removed due to fishing and converted to absolute units, numbers, through $\exp(\log K)$.

The model fits to the observed total catch in numbers using a Student's t-distribution likelihood:

$$\log(Removals_{obs,t}) \sim \text{Student-t}\left(\nu_c, \log(Removals_t) - \frac{\sigma_{C,t}^2}{2}, \sigma_{C,t}\right)$$

where the degrees of freedom parameter ν_c controls the tail behavior[4], and the observation error $\sigma_{C,t}$ allows for time-varying uncertainty in catch observations. However, in the diagnostic model configuration $\sigma_{C,t} = 0.2$ for all t.

3.5 Prior development

The baseline prior distributions for leading estimated model parameters are specified in Table 2. A description of the derivation of these prior distributions follows.

Informed prior distributions for R_{Max}, and the Pella-Tomlinson shape parameter n were developed using numerical simulation based on biological knowledge. Informed prior distributions for $\log(K)$, and baseline catchability q_{eff} were developed using numerical simulation based on existing fishing data observations (e.g., catch and effort). Priors for other estimated leading model parameters (initial depletion x_0, AR1 correlation in catchability deviates ρ, variability in catchability deviates σ_{qdev}, process error σ_P, index observation error $\sigma_{O,add}$, and degrees of freedom of the catch likelihood ν_{catch}) were based on expert opinion and values observed in the literature. The naïve or input prior distributions were refined using a prior pushforward analysis.

---

[4]As $\nu_c \to \infty$ the likelihood equation converges to the lognormal distribution $\text{Lognormal}\left(\log(Removals_t) - \frac{\sigma_{C,t}^2}{2}, \sigma_{C,t}\right)$

3.5.1 Biological simulation framework

Informative priors for key population parameters were developed through biological simulation using Monte Carlo sampling from biologically plausible parameter distributions[5]. The simulation framework incorporated uncertainty and correlations in life history parameters:

- **Growth parameters**: Francis parameterization with independent sampling
- **Natural mortality**: Reference mortality M_{ref} negatively correlated with maximum age (r = -0.3)
- **Maturity**: Length-based logistic function parameters
- **Reproduction**: Steepness parameter for stock-recruit relationship and random variation around a 50:50 sex-ratio
- **Weight-length relationship**: Correlated allometric parameters (r = -0.5) with biological constraints
- **Fishery selectivity**: Length at first capture for depletion prior

Parameter combinations (Table 1) were filtered to ensure biological realism and included correlations between life history parameters reflecting biological trade-offs (e.g., shorter lived individuals having higher natural mortality rates).

3.5.1.1 Maximum intrinsic rate of increase (R_{Max}) prior

R_{Max} was calculated by numerically solving the Euler-Lotka equation using bounded optimization (nlminb function in R) to find the value of R_{Max} that satisfies the equation within convergence criteria. Starting values and bounds were set to ensure biologically realistic solutions. Formulation of the Euler-Lotka equation followed the approach implemented in openMSE (Hordyk et al., 2024; Stanley et al., 2009):

$\alpha \sum_{a=1}^{A_{Max}} l_a f_a \exp(-R_{Max} \times a) = 1$

where α is the slope at the origin of the stock-recruitment curve, l_a is the probability of survival to age a, and f_a is the fecundity (reproductive output) at age a. Additional technical detail can be found in Section 11.1 of the Technical Annex. Repeating this calculation across the 500,000 parameter combinations resulted in a prior distribution of R_{Max} conditioned on plausible biology and life-history characteristics. The resulting distribution was first filtered to values of $R_{Max} < 1.5$ representative of what the literature (Gravel et al., 2024; Hutchings et al., 2012) indicates is most likely for teleosts. A lognormal distribution was fit to the filtered distribution to derive input prior parameters for the Stan model.

3.5.1.2 Pella-Tomlinson shape parameter (n) prior

Similarly, a prior distribution for the Pella-Tomlinson shape parameter n was derived from the filtered R_{Max} values and associated biological parameter combinations through a two-step process. First, the inflection point of the production function was estimated using the relationship established by Fowler (1988):

$$\phi = 0.633 - 0.187 \times \log(T_G \times R_{Max})$$

where ϕ is the inflection point (representing the population level at which maximum sustainable yield occurs as a fraction of carrying capacity), T_G is the generation time calculated following Grant & Grant (1992) as

---

[5]The R code for the simulation `01-01-define-rmax-prior.r` is publicly available on GitHub.

$T_G = \sum_{a=1}^{A_{Max}} a \times l_a \times f_a / \sum_{a=1}^{A_{Max}} l_a \times f_a$, and R_{Max} is the maximum intrinsic rate of increase.

The Pella-Tomlinson shape parameter n was then derived from the inflection point through the mathematical relationship $\phi = (1/n)^{1/(n-1)}$, solved numerically using bounded optimization in R as done for R_{Max}. This transformation links the biologically-derived inflection point to the production function curvature, where $n = 2$ corresponds to the symmetric Schaefer model (e.g., MSY achieved at depletion $D = 0.5$). Lower values of n ($n < 2$) indicate more productive stocks that achieves MSY at $D < 0.5$ and higher values of n indicate less productive stocks with MSY occurring at $D > 0.5$. A lognormal distribution was fit to the distribution of n to derive input prior parameters for the Stan model. Additionally, linking the shape parameter n to R_{Max} necessitated a multivariate prior structure so the correlation between the two parameters was estimated as well.

3.5.2 Fishing data simulation framework

Prior distributions for carrying capacity $\log(K)$ and initial catchability coefficient q_{eff} were derived using catch and effort data combined with the biologically-informed parameter distributions[6].

3.5.2.1 Carrying capacity prior ($\log(K)$)

The carrying capacity prior was constructed by assuming the maximum observed catch approximates MSY. This provides a conservative assumption for the upper limit of population scale[7] as larger MSY correlates positively with larger population scale. To account for uncertainty in the maximum catch $\sim$ MSY relationship (e.g., true MSY was above or below the maximum observed catch), candidate MSY values were sampled from a lognormal distribution centered on the maximum observed catch with moderate levels of uncertainty ($MSY \sim \text{Lognormal}(\log(\text{max catch}), 0.5)$).

For each combination of MSY, biologically-derived R_{Max} and shape parameter n, carrying capacity was calculated as:

$$K = \frac{MSY}{r \times \phi \times (1 - \phi)^{n-1}}$$

where $\phi = (1/n)^{1/(n-1)}$ represents the inflection point of the production function. For the Fox model special case where $n = 1$, the relationship simplifies to $K = e \times MSY/r$. Non-finite values and extreme outliers ($\log(K) > 20$) were replaced through resampling from the valid parameter space to ensure numerical stability. A lognormal distribution was fit to the filtered distribution to derive input prior parameters for the Stan model.

3.5.2.2 Initial effort catchability prior (q_{eff})

Next, the initial catchability coefficient prior was developed using the relationship between CPUE and stock abundance:

---

[6]This is applied in the R script `03-06-new-logK-prior-pushforward-bspm_estqsimple_softdep_mvprior_x0.r` lines 64-92 which is publicly available on GitHub.

[7]The upper limit of population scale is typically more challenging to estimate than the lower population limit as population scales of infinite sizes can support any observed catches. However, there typically is a hard floor to population scale in that it must be sufficiently large to maintain a population that produces the time series of observed catches.

$$q_{eff} = \frac{CPUE}{K}$$

where CPUE was calculated from early fishery data (first 10 years of catch and effort records) when the stock was assumed to be closer to carrying capacity. The prior distribution was constructed by randomly pairing CPUE values sampled from the early period with carrying capacity values from the $\log(K)$ prior distribution to generate a distribution of q_{eff} values incorporating both the uncertainty in population scale and the variability observed in early fishery CPUE. Again, a lognormal distribution was fit to the filtered distribution to derive input prior parameters for the Stan model.

3.5.3 Other priors

As mentioned previously, initial prior distributions for the remaining leading model parameters (initial depletion x_0, AR1 correlation in catchability deviates ρ, variability in catchability deviates σ_{qdev}, process error σ_P, index observation error $\sigma_{O,add}$, and degrees of freedom of the catch likelihood ν_{catch}) were specified either using expert opinion or literature.

Given that the limited large-scale industrial fishing existed in the SWPO prior to the expansion of the Japanese distant-water longline fleet in 1952, the population was assumed to start at an unfished condition with some uncertainty. Initial depletion x_0 was modeled with a lognormal distribution centered on 1. A lognormal prior was used for the standard deviation of the process error where the parameters were converted from the JABBA default prior for process error (Winker et al., 2018). The degrees of freedom nu_{catch} parameter assumed a gamma prior based on suggestions in the literature that it serves as a sensible default choice (Juárez & Steel, 2010). The prior for index observation error $\sigma_{O,add}$ assumed a half-normal distribution following ISC (2024) which provides sufficient support for a broad range of extra estimated observation error. For the autocorrelation coefficient in catchability deviations (ρ) and the standard deviation of catchability deviations (σ_{qdev}), initial priors were derived from data-filtering simulation results and specified as a bivariate distribution on the transformed scale: $\tanh^{-1}(\rho)$ and $\log(\sigma_{qdev})$.

3.5.4 Prior pushforward

Following Kim & Neubauer (2025), a prior pushforward analysis was conducted to refine the naïve prior distributions derived from the biological and fishing data simulation frameworks. Random parameter combinations drawn from the input prior distributions were passed through the diagnostic Stan model used for inference but with the likelihoods turned off (`fit_to_data = 0L`), generating simulated population trajectories and fishing dynamics under the model structure. This approach ensures that simulated trajectories are consistent with the actual model structure used for inference.

Parameter combinations were filtered through sequential criteria based on population viability and fishery realism (Figure 3):

- **Initial survival filter (`iter_surve`)**: Terminal year depletion $> 1\%$ (`mindep > 0.01`), minimum population size $> 15{,}000$ individuals which matches recent observed catch levels (`minN > 15000`), and biologically reasonable shape parameter (`shape < 20`)
- **Fishing mortality filter (`iter_f`)**: Maximum fishing mortality < 2.5 (`maxF < 2.5`) and average fishing mortality < 0.8 (`avgF < 0.8`) to ensure realistic exploitation levels

- **Catch consistency filter (`iter_catch`):** Simulated catches within realistic bounds relative to observed data: maximum catch between 0.5-5.0× observed maximum, minimum catch > 0.5× observed minimum, and average catch between 0.5-5.0× observed average
- **CPUE consistency filter (`iter_cpue`):** Early period CPUE (first 10 years) between 0.5-5.0× observed early CPUE average to ensure realistic abundance levels during the initial fishery development

Parameter combinations passing all filtering stages (`iter_cpue`) were used to fit updated multivariate prior distributions using maximum likelihood estimation (Figure 4). Given that a number of parameters showed correlations following the prior pushforward, a 7-dimensional multivariate prior was developed for core parameters: $\log(K)$, $\log(R_{Max})$, $\log(n)$, $\log(x_0)$, $\log(q_{eff})$, $\tanh^{-1}(\rho)$ and $\log(\sigma_{qdev})$. Correlations with absolute values < 0.1 were reset to zero to avoid weak spurious correlations, retaining only meaningful parameter relationships. Though $log(x_0)$ was included in the multivariate prior, for convenience, it was treated as an unchanged univariate prior from the initial naïve prior distribution since the pushforward did not provide an update to this parameter.

Prior distributions for process error σ_P, index observation error $\sigma_{O,add}$, and degrees of freedom of the catch likelihood ν_{catch} were unchanged following the prior pushforward analysis and remained as univariate priors. Final prior distributions are listed in Table 2.

3.6 Description of one-off sensitivities

During the course of model development 38 one-off sensitivity models, organized into 12 groups, were investigated to explore alternative assumptions, model structures and sensitivities to data inputs. These models are listed in Table 3 and are described in further detail below.

3.6.1 Index

The diagnostic case model fits only to the DWFN index. One-off model sensitivities were conducted to explore how model estimates changed as a result of alternative indices:

- `0071` : Fits to the Australian longline index
- `0072` : Fits to the New Zealand recreational sportfish index
- `0073` : Fits to the longline observer data combining the programs from New Caledonia, Fiji, Tonga, and French Polynesia
- `0074` : Fits to the longline observer data excluding the French Polynesia program
- `0075` : Fits only to the longline observer data from the French Polynesia program

The French Polynesia observer program data was treated separately as it showed differing trends from the other 3 observer programs (Neubauer et al., 2025). Additionally, given that the shorter indices (Australia and longline observer) provided little information about population scale, several other models were set-up to simultaneously fit them with the two longer indices (DWFN and New Zealand).

- `0129` : Fits to the DWFN & Australia index
- `0130` : Fits to the New Zealand & Australia index
- `0131` : Fits to the DWFN & Observer (New Caledonia, Fiji & Tonga) index
- `0132` : Fits to the New Zealand & Observer (New Caledonia, Fiji & Tonga) index

3.6.2 Catch observation error σ_C

The diagnostic case model assumes a constant catch observation error of $\sigma_C = 0.2$. Three one-off models explored alternative assumptions about catch uncertainty:

- 0076 : Time-varying σ_C following a power curve declining from 0.5 in early years to 0.2 in recent years, assuming improved catch reporting over time
- 0077 : Constant $\sigma_C = 0.1$, representing lower uncertainty in catch observations throughout the time series
- 0078 : Time-varying σ_C following a power curve declining from 0.5 in early years to 0.1 in recent years, combining both temporal improvement and lower overall uncertainty

3.6.3 Early data

Several models explored sensitivity to problematic early data years, particularly 1952 and 1954 which showed unusual catch or effort patterns. In particular the 1952 catch observation was anomalously low especially given the reported effort, and was identified to have high influence on model estimates. This could be a valid observation, as the Japanese longline fleet expanded into the south Pacific in 1952 they fished closer to the equator which tended to have lower striped marlin catch rates than in subsequent years when they fished between Australia and New Caledonia. Alternatively, this outlier point could be the result of erroneous reporting in either catch or effort. In either case, sensitivity to this data point was investigated:

- 0079 : Recalculates the low 1952 catch observation as the expected catch given the effort and average nominal CPUE from 1953-1957 (1952 catch is increased substantially).
- 0080 : Recalculates 1952 effort given the low 1952 catch and average nominal CPUE from 1953-1957 (1952 effort is decreased substantially)

Additionally, the large catch event in 1954 has been heavily scrutinized. This could be a valid observation as the Japanese longline fleet expended considerable effort during the time of year (3rd and 4th quarters of the year) striped marlin form spawning aggregations, in an area between Australia and New Caledonia where these aggregations form. Regardless, sensitivity to this observation was investigated in case it is a reporting error.

- 0081 : The 1954 catch was reduced by 50%
- 0082 : Combined alternative catch values for both 1952 and 1954
- 0083 : Combined alternative 1952 effort and 1954 catch values

These sensitivities address concerns about sensitivity to potentially anomalous observations in the early years of the fishery when reporting systems were less standardized.

3.6.4 Scaling priors

The diagnostic case model estimates a low population scale relative to the prior. Two models tested the impact of forcing the model toward higher population scale estimates:

- 0093 : Modified priors on log(K) and catchability (qeff) to favor higher carrying capacity and lower catchability estimates
- 0096 : Same as 0093 but combined with more conservative shape parameter prior (centered on 2)

These models explored whether alternative prior assumptions about population scale could shift model estimates.

3.6.5   Catch observation model

- `0110` : Replaced the default Student-t likelihood for catch observations with a lognormal error structure to test sensitivity to distributional assumptions about catch observation error

3.6.6   Temporal resolution of catchability deviates

The diagnostic case model estimates catchability deviates in single time steps (annual). Four models explored different temporal resolutions:

- `0111` : Catchability deviates fixed across 2-year periods (n_step = 2)
- `0112` : Catchability deviates fixed across 3-year periods (n_step = 3)

- `0113` : Catchability deviates fixed across 5-year periods (n_step = 5)
- `0114` : Catchability deviates fixed across 10-year periods (n_step = 10)

These sensitivities tested whether the temporal resolution of catchability changes affected model estimates, with longer periods resulting in a more parsimonious model but potentially diminishing the models ability to fit to data.

3.6.7   Catch reporting

Striped marlin are a commercially valuable non-target species, and as such it is assumed that if they are caught they are retained, and recorded in the logbook. However, given that they are a non-target species it is possible that they are not, or have not always been reported in the logbook and that under-reporting of catches or discarding could occur (e.g., given value of striped marlin and hold space required relative to target species). Three models explored potential systematic under-reporting of catches:

- `0115` : Doubled all observed catch values to test sensitivity to potential 50% under-reporting throughout the time series
- `0116` : Applied a linear multiplier increasing from 1x to 2x observed catch over time, representing improving catch reporting
- `0117` : Applied a linear multiplier decreasing from 2x to 1x observed catch over time, representing deteriorating catch reporting in recent years

3.6.8   Process error σ_P

The diagnostic case model estimated non-zero process error in the early model period, before the CPUE index began, and balanced process error with the catchability deviates in order to fit to the catch.

- `0118` : Restricted process error to be negligible ($\sigma_P \sim 0$) to evaluate the impact of environmental stochasticity on model estimates by forcing the population to follow deterministic surplus production dynamics. This is essentially the same thing as the deterministic recruitment model diagnostic (Kapur et al., 2025; Merino et al., 2022), and can be informative for determining the strength of the estimated production function.

3.6.9 Start year

Given potentially anomalous catch observations in 1952 and 1954, along with uncertainty in the accuracy of early catch observations, four models explored sensitivity to the assumed initial conditions by starting the model in different years:

- 0119 : Model starts in 1955 with assumed initial depletion $x_0 = 0.9$, skipping the early data years

Three models were also tested which started the model in the same year as the CPUE index, and varied the level of initial depletion:

- 0120 : Model starts in 1988 with $x_0 = 0.5$
- 0121 : Model starts in 1988 with $x_0 = 0.7$
- 0122 : Model starts in 1988 with $x_0 = 0.9$

3.6.10 Inclusion of effort deviates

Early model development explored a more complex formulation of fishing mortality F_t where effort deviates were also used in addition to the catchability deviates to link effort to F_t:

$$F_t = q_{eff} \times \exp(\delta_{q,t} - \sigma^2_{qdev}/2) \times E_t \times \exp(\delta_{e,t} - \sigma^2_{edev}/2)$$

This formulation of fishing mortality is closer to the full catch-errors with effort deviation approaches available in MULTIFAN-CL (Fournier et al., 1998). Two models tested whether estimating deviations from the effort time series improved model performance:

- 0124 : Estimated effort deviates with 2-year temporal resolution (n_step = 2)
- 0126 : Estimated effort deviates with 5-year temporal resolution (n_step = 5) for more constrained estimation

3.6.11 Direct estimation of fishing mortality

Rather than estimating fishing mortality indirectly through catchability and effort, two models directly estimated annual fishing mortality rates to see if the simpler formulation for fishing mortality could replicate the results of the diagnostic case model, or if the formulation of fishing mortality significantly impacted model estimates:

- 0127 : Direct F estimation with σ_F calibrated to match the variance in fishing mortality observed in the diagnostic case model
- 0128 : Direct F estimation with σ_F set to 0.25 times the diagnostic case variance, providing more constraint on fishing mortality variability in order to see how estimates of scale are impacted

3.6.12 Alternative R_{Max} prior

The diagnostic case model estimates a fairly productive stock with an R_{Max} that is considerably higher than an analogous species in the Atlantic Ocean, white marlin (*Kajikia albida*). The most recent ICCAT assessment (ICCAT, 2020), implemented using JABBA (Winker et al., 2018), estimated R_{Max} at 0.163 for

white marlin. Sensitivity models assuming this lower level of productivity were developed to see the impact it would have on model estimates, particularly population scale.

- `0133` : Apply lower productivity white marlin Rmax prior
- `0134` : Apply lower productivity white marlin Rmax prior and apply a more conservative shape parameter prior (centered on 2)

3.7 Model Implementation

The models were implemented using the No-U-Turn Sampler (NUTS) in Stan. Each model was run using 5 independent Markov chains with randomly generated starting values to ensure robust exploration of the posterior parameter space. Each chain was sampled for 3,000 iterations, with the first 1,000 iterations discarded as warmup to allow for algorithm adaptation. To reduce posterior sample autocorrelation and storage requirements, every 10th samples was retained from the post-warmup iterations, yielding a final sample size of 1,000 draws from the joint posterior distribution across all chains. The NUTS algorithm was configured with strict adaptation parameters to ensure reliable convergence. The adaptation delta was set to 0.99 to reduce the step size and minimize divergent transitions, while the maximum tree depth was increased to 12 to allow for longer trajectories during the dynamic sampling process.

Model convergence and performance were assessed using multiple diagnostic criteria recommended for Bayesian stock assessment workflows (Monnahan, 2024). The potential scale reduction factor ($\hat{R}$) was required to be less than 1.01 for all parameters, indicating convergence across chains. Effective sample sizes were required to exceed 500 to ensure adequate precision in posterior estimates. Additionally, posterior predictive checks were conducted to evaluate model adequacy through comparison of observed versus predicted index and catch values, while parameter identifiability was assessed through visual examination of posterior distributions and correlation structures among estimated parameters (Gabry et al., 2019).

3.8 Model diagnostics

Model performance was comprehensively evaluated using multiple diagnostic criteria recommended for Bayesian stock assessment workflows (Monnahan, 2024). The diagnostic framework included Stan model convergence criteria, data fits, posterior predictive checks, retrospective analysis, and hindcast cross-validation.

3.8.1 Convergence diagnostics

Conventional Stan model convergence diagnostics and thresholds were employed to identify whether posterior distributions were representative and unbiased (Monnahan, 2024). Models were considered to have 'converged' to a stable, unbiased posterior distribution if they satisfied the following criteria:

- **Potential scale reduction factor** ($\hat{R}$) less than 1.01 for all leading model parameters, indicating convergence across chains
- **Bulk effective sample size** (N_{eff}) greater than 500 for all leading model parameters, ensuring adequate precision in posterior estimates

- **No divergent transitions** during the NUTS sampling process, which can indicate problematic posterior geometry

- **Maximum tree depth** not exceeded during sampling, ensuring adequate exploration of parameter space

Additionally, trace plots and rank plots were examined to visually assess mixing and convergence behavior across chains (Gabry et al., 2019).

3.8.2 Leave-one-out cross-validation diagnostics

Model predictive performance and influential observations were assessed using leave-one-out cross-validation (LOO-CV) implemented through the *loo* package (Vehtari et al., 2017; Vehtari et al., 2024). LOO-CV provides an estimate of the expected log pointwise predictive density (ELPD) while identifying problematic observations that may unduly influence model predictions.

LOO-CV was also used as a factor to compare performance between models. Following the recommendations of Sivula et al. (2020), models were considered to have meaningfully different performance when the ELPD difference exceeded 4 units. For differences larger than 4 units, statistical significance was assessed by comparing the ELPD difference to twice its standard error, with differences greater than $2\times$ the standard error considered statistically significant.

The effective number of parameters (p_loo) was examined as an additional diagnostic measure, with values substantially larger than the nominal number of model parameters indicating potential model misspecification. Catch and index observations were categorized as having low influence (Pareto-k $\leq$ 0.7), high influence (0.7 < Pareto-k $\leq$ 1.0), or very high influence (Pareto-k > 1.0). Observations with high or very-high influence were investigated to identify potential outliers, data entry errors, or systematic model misspecification that could compromise predictive performance.

3.8.3 Data fits

Model fits to the abundance index and catch time series were each quantified using normalized root-mean-squared error (NRMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

$$NRMSE = \frac{RMSE}{\sum_{i=1}^{n} y_i/n}$$

where y_i are the observed values and $\hat{y}_i$ are the model predictions. The average NRMSE across all posterior samples was calculated for each data component. Additionally, Probability Integral Transform (PIT) style residuals were calculated for the fit to the index.

3.8.4 Posterior predictive checks

Posterior predictive checks were conducted to evaluate model adequacy by comparing observed data with data simulated from the fitted model. For each posterior sample, synthetic datasets were generated using the estimated parameters, and agreement between observed and predicted datasets was assessed visually.

3.8.5 Retrospective analysis

Retrospective analysis was conducted for each model by sequentially peeling off a year from the terminal end of the fitted index and re-running the model. Data were removed for up to five years, from 2022 back to 2018. Estimates of x in the terminal year of each retrospective peel were compared to the corresponding estimate of x from the full model run to better understand any potential biases or uncertainty in terminal year estimates. The Mohn's ρ statistic (Mohn, 1999) was calculated and presented. This statistic measures the average relative difference between an estimated quantity from an assessment (e.g., depletion in final year) with a reduced time-series of information and the same quantity estimated from an assessment using the full time-series.

Additionally, following Kokkalis et al. (2024) the proportion of retrospective peels (or coverage) where the relative exploitation rate (U/U_{MSY}) and relative depletion (D/D_{MSY}) were inside the credible intervals of the full model run was calculated. This metric provides an additional perspective on retrospective consistency by evaluating whether the uncertainty estimates from the full model appropriately capture the range of estimates obtained when using reduced datasets.

3.8.6 Hindcast cross-validation

Hindcast cross-validation (Kell et al., 2021) was conducted for each index to determine model performance in predicting the observed CPUE I one-step-ahead into the future relative to a naïve predictor. Briefly, the 'model-free' approach to hindcast cross-validation was used, and used the same set of five retrospective peels described in Section 3.8.5.

The 'model-free' hindcast calculation is described using the model from the last peel $BSPM_{2017}$ as an example. This model fit to index data through 2017 but included catch through 2022. The model estimates of predicted CPUE in 2018 based on $BSPM_{2017}$ is the 'model-free' hindcast for 2018, $\hat{I}_{2018}$. The naïve prediction of CPUE in 2018 is simply the observed CPUE from 2017, $I_{2017} = \ddot{I}_{2018}$. The absolute scaled error (ASE) of the prediction is:

$$ASE_{2018} = \frac{|I_{2018} - \hat{I}_{2018}|}{|I_{2018} - \ddot{I}_{2018}|}$$

Repeating this calculation across all retrospective peels for years 2019-2022 and taking the average across ASE values gives the mean ASE or MASE for the model. An MASE value less than one indicates that the model has greater predictive skill than the naïve predictor.

3.8.7 Model-free hindcasts

Additional model-free hindcasts, similar to those described in Section 3.8.5, where the model was run for the full period 1952-2022 but without the last y years of index data. Hindcasts excluding the last 5, 10, 15, and 20 years of index data were conducted in order to evaluate the models ability to recapture the dynamics of the full model in the terminal year. Good hindcast performance over a long time horizon is an indication of a well-determined production function, and can give an idea of the reliability of the predictions from the full model.

3.9 Forecasts

Simple, status quo catch-based stochastic projections over a 10-year window are provided for each model. For each model, projected catch was taken as the average catch over the terminal 5 years (2018-2022). Process error in the projection period was randomly resampled from the estimated process error in the main model period.

4 Results

4.1 Diagnostic case

4.1.1 Model convergence and diagnostics

The diagnostic case model `0100` achieved satisfactory convergence based on standard Hamiltonian Monte Carlo diagnostics (Table 4). $\hat{R}$ values were below the 1.01 threshold and effective sample sizes exceeding the recommended minimum of 500 (100 samples per chain) for all parameters. No divergent transitions were observed across chains, and maximum tree depth was not exceeded during sampling, indicating successful exploration of the posterior space. Additionally, LOO-CV identified only two observations with Pareto-k > 0.7 indicating the model structure and choice of likelihoods adequately matched the fitted data. Lastly, the effective number of estimated parameters p_loo was well below both the number of observations and the actual estimated parameter count indicating that model is likely to be reasonably well specified.

4.1.2 Model fit to data

4.1.2.1 Catch data fit

The diagnostic case model adequately captured the trend and magnitude of the observed catches (1952-2022) with the obvious exceptions of the 1952 and 1954 observations (Figure 5). Use of a catch likelihood, student-t, with heavier tails made the model more robust to these outlier observations, and accordingly estimated larger and smaller predicted catches, respectively, for these two observations. However, use of the student-t distribution along with an assumed 20% CV on the observation error did mean that catch observations were not fit exactly though the observed catches did tend to fall within the posterior predicted distribution (Figure 6). Given that SWPO striped marlin is a non-target species, some uncertainty in the predicted catches is considered to be appropriate.

4.1.2.2 CPUE index fit

Diagnostic case model fit to the standardized DWFN CPUE index (1988-2022) was evaluated using NRMSE and showed reasonable fits to the observed index given the high variability and observation error (Table 5). The model also successfully captured the general declining trend in standardized CPUE from 1988-2022 (Figure 7). PIT residuals did not show persistent positive or negative patterns, however they did indicate marginally better fit within the last decade of the index relative to the earlier portion(Figure 8). Posterior predictive checks confirmed that models replicated key features of the observed CPUE data (Figure 9).

4.1.3 Model validation

Overall retrospective bias was within the acceptable range proposed by (Hurtado-Ferro et al., 2015), however it did indicate a slight underestimate to the population recovery indicated by the DWFN index (Table 5; Figure 10). This is likely driven by the large observation in 2021 that is used by the full model to drag estimates up from the lower values observed in 2015-2019. Estimates of fishing mortality (U/U_{MSY}) and depletion (D/D_{MSY}) relative to MSY fell within the credible intervals of the full model run 100% of the time (Table 5). For hindcast cross-validation, the diagnostic model substantially exceeded the performance of the naïve predictor (Table 5; Figure 11). This supports the conclusion that models successfully identified and estimated a production function for the stock.

4.1.4 Parameter estimates

Following the guidance from Kim & Neubauer (2025), posterior distributions of parameter estimates were compared to the realized prior in addition to the naïve or input prior. The realized prior represents the effective prior distribution that results from passing naïve parameter combinations through the model structure with likelihoods turned off, revealing how model constraints and structural assumptions modify the input priors before data-based updating occurs (Kim & Neubauer, 2025). In the current analysis, there was little difference between the naïve and realized priors given the approach taken in developing the priors. Based on these comparisons the diagnostic model showed substantial posterior learning across several key leading parameters (Figure 12; Table 6), indicating that the data contained sufficient information to update parameter estimates from their prior distributions.

The carrying capacity $\log(K)$ exhibited substantial posterior learning, with the combination of index and catch data in the diagnostic case model strongly supporting lower population scales than suggested by either the naïve or realized priors. Similarly, the intrinsic rate of increase R_{Max} showed strong posterior updates, with the data favoring lower productivity than the biological simulation framework initially suggested. Inverse transform sampling to match the R_{Max} posterior distribution to the input biological simulation framework prior for R_{Max} indicates that the lower productivity is characterized by pronounced shifts toward lower steepness values h (Figure 13).

The median estimate of R_{Max} from the diagnostic case, 0.396 (Table 6), is considerably higher than estimates for an analogous species, white marlin (*Kajikia albida*), assessed using a JABBA surplus production model (ICCAT, 2020; Winker et al., 2018) in the Atlantic Ocean where $R_{Max} = 0.163$ (95% credible interval: 0.122-0.215). However, it is worth noting that the ICCAT assessment assumed a much more precise prior for R_{Max}, and that the 95% credible intervals of the diagnostic case model almost completely cover the ICCAT posterior distribution. Notably, the Atlantic white marlin JABBA model showed remarkable consistency with stock status and trend estimates from a fully integrated, length-structured model implemented in Stock Synthesis (ICCAT, 2020).

Shape and initial depletion parameters did not demonstrate notable posterior updates (Figure 12). The Pella-Tomlinson shape parameter n showed limited deviation from the prior, suggesting the data provided weak information about production function curvature. Initial depletion x_0 also remained close to the assumed unfished prior, indicating little information in the data to suggest large-scale fishing pressure prior to 1952.

The catchability and temporal variability parameters showed large posterior updates (Figure 12). The base

catchability coefficient q_{eff} showed substantial learning, with strong negative correlations with $\log(K)$. This demonstrates a trade-off between population scale and catchability in order to fit the observed catches. If a more precise prior on q_{eff} were to be developed, this would be expected to set population scale and dramatically reduce uncertainty in model estimates. The autocorrelation parameter ρ and catchability variability σ_{qdev} both showed posterior updates, with the data informing a highly correlated temporal structure of catchability deviations.

Lastly, process error σ_P showed moderate posterior learning, while additional observation error $\sigma_{O,add}$ demonstrated substantial updates from the prior (Figure 12). This indicated that the model estimated additional uncertainty beyond the input index observation error, rather than using process error, to reconcile the observed CPUE variability. This results in a smoother year-on-year population trajectory. The degrees of freedom parameter ν_{catch} for the catch likelihood showed substantial updating (Figure 12), indicating the model preferred a Student-t distribution with very heavy tails in order to adequately account for the catch observations given the level of observation error.

4.1.5 Model-free hindcasts

Model-free hindcasting was conducted to evaluate the model's ability to predict population dynamics when fitted to progressively truncated index datasets. The diagnostic model was compared against retrospective peels excluding the final 5, 10, 15, and 20 years of index data (terminating in 2017, 2012, 2007, and 2002, respectively), with the model providing estimates through the full model period based on based only on the estimated production function and predicted catch.

The model demonstrated remarkably consistent performance across most hindcast scenarios (Figure 14) providing strong evidence of a well-determined production function. In general, the hindcast models successfully captured the general increasing trend in the standardized index after 2015. Model predictions for the DWFN CPUE index showed good coherence between the diagnostic model and all retrospective peels except the 15 year hindcast model, with overlapping 95% credible intervals throughout most of the time series. Estimates from the 15 year hindcast model appear to be strongly influenced by the series of higher observations (2003-2005) which cause it to underestimate the decline in the observed index.

Population dynamics estimates showed reasonable consistency across hindcast scenarios (Figure 15), with the notable exception of the 15 year hindcast model mentioned in the previous paragraph. Depletion trajectories remained largely stable, with all models estimating similar patterns of decline from unfished conditions to minimum levels around the early 1990s, though predictions begin to diverge slightly in the recovery period post 2015. Population scale estimates also showed good agreement, though with increasing uncertainty for earlier time periods in longer hindcasts as expected. The stock status metrics (D/D_{MSY} and F/F_{MSY}) showed fairly good consistency across most hindcast scenarios particularly for F/F_{MSY}. Additionally, given the The estimated trajectories for these key management metrics were nearly identical between the diagnostic model and all retrospective peels, suggesting that management advice derived from this model would have been stable regardless of when the assessment was conducted over the past two decades.

This consistency across multiple extended hindcast periods provides support that the underlying population dynamics are well described, given the available data. This gives confidence that projections made from the full model will likely capture the stock trajectory within the near to medium term, assuming stationarity in future productivity and carrying capacity.

23

4.2 One-off sensitivities

As described in Section 3.6, 38 different one-off sensitivities were conducted to either explore plausible uncertainty in model assumptions and configuration, or to build intuition about the model dynamics and gain confidence in the results.

4.2.1 Index

The first one-off sensitivity group explores alternative abundance indices compared to the diagnostic case model that fits only to the DWFN index. This group includes models testing individual indices (0071-0075) and combinations of indices (0129-0132). In general, model fit to individual indices (0071-0075) was acceptable (Figure 16), though model predictions tended to over-predict the last 5 years of the New Zealand recreational sportfish index (0072). However, observation errors for these observations were fairly large relative to earlier observations meaning that the model was not overly constrained to fit them, and could instead follow the trajectory implied by the production function. Notably, only the two longer indices (DWFN and New Zealand recreational) showed sufficient contrast with the catches to meaningfully move the scale estimate from the prior and models fitting either of those indices estimated similar population scales (Figure 17). The Australia longline index (0071) showed some posterior update but still estimated a larger overall population scale, the observer index with only data from French Polynesia (0075) also showed a noticeable posterior update however had poor convergence diagnostics (Table 4).

For the models fitting to multiple simultaneous indices (0129-0132), estimates of model quantities were largely consistent with the diagnostic case model in terms of scale and trend (Figure 18), even for models fitting to the New Zealand index. This indicated that scale estimates from models fitting to shorter indices may not be reliable. Additionally, fits to the shorter indices were not substantially degraded when also simultaneously fitting to the longer indices.

4.2.2 Catch observation error σ_C

The diagnostic case model assumes a constant catch observation error of $\sigma_C = 0.2$. Three one-off models (0076 - 0078) explored alternative assumptions about catch uncertainty including time-varying error structures and different constant error levels. In general parameter estimates were very consistent regardless of the assumed catch observation error level and temporal pattern (Figure 19). The one exception was with the estimation of ν_{catch}, which showed a preference for a Student-t distribution with lighter tails (e.g., more *Normal* distribution-like) when catch observation error was structured to decline overtime from 0.5 following a power curve (0076 & 0078). This suggests that the earlier observations may warrant a higher level of observation error. Regardless, estimated model quantities, including key management metrics D/D_{MSY} and F/F_{MSY}, were largely unchanged no matter what structure was used with the exceptions that models assuming less catch uncertainty had predictably less estimated uncertainty, and models using a power curve to structure the observation error estimated lower levels of initial nominal CPUE (Figure 20).

4.2.3 Early data

Several models explored sensitivity to problematic early data years, particularly 1952 and 1954 which showed unusual catch or effort patterns. Five one-off models (0079 - 0083) explored alternative assumptions about the catch and or effort in these years. Parameter estimates were very similar across all formulations

(Figure 21). However, estimation of ν_{catch}, showed a preference for a Student-t distribution with lighter tails when either alternative catch or effort was assumed for the 1952 observation. This in combination with the results from Section 4.2.2 suggests that the 1952 observation may be a reporting error and could be modified or downweighted with larger observation error. Fortunately, estimated model quantities, including key management metrics, do not appear sensitive to the treatment of the 1952 observation (Figure 20).

4.2.4   Scaling priors

The diagnostic case model estimates a low population scale relative to the prior. Two models (`0093` & `0096`) tested the impact of forcing the model toward higher population scale estimates through modified priors on carrying capacity and catchability. This was more of an academic sensitivity analysis to see if the model could be moved through the prior information from the low estimate. While the sensitivity was successful in shifting the estimated scale higher (Figure 23), it showed a clear trade-off with the ability to fit the index (Figure 24) and showed an incompatibility between the data and larger estimates of scale. This provides additional evidence that the smaller estimates of population scale are driven by the data.

While not investigated in the current analysis, it would be interesting to see how sensitive estimates of population scale are to just changes in the precision of the catchability prior as suggested in Section 4.1.4.

4.2.5   Catch observation model

A single model (`0110`) replaced the default Student-t likelihood for catch observations with a lognormal error structure which is less statistically robust and more impacted by outlier observations. This tested sensitivity to distributional assumptions about catch observation error. Parameter estimates, particularly those related to population scale, showed some sensitivity to choice of catch observation model (Figure 25), and this did translate into some differences in estimated management quantities (Figure 26). However, as previous sensitivity analyses have identified (Section 4.2.2 & Section 4.2.3), some of these early observations appear less reliable and should potentially be made less influential in the likelihood (e.g., use of a Student-t distribution). This finding, combined with the `0110` model showing substantially more observations with large Pareto-k values (Table 4), did not support further use of the lognormal error structure.

4.2.6   Temporal resolution of catchability deviates

The diagnostic case model estimates catchability deviates in single time steps (annual). Four models explored different temporal resolutions with catchability deviates fixed across 2, 3, 5, and 10-year periods (`0111` - `0114`). Reducing the temporal resolution of the catchability deviates (increasing n_{step}) predictably degraded fit to the catch as n_{step} increased (Table 5; Figure 27). It also changed estimates of σ_P and q_{eff} (Figure 28) which in turn impacted estimates of management quantities particularly depletion and D/D_{MSY} (Figure 29), again particularly for larger values of n_{step}. Reducing the temporal resolution of the catchability deviates did improve model parsimony, and the model with $n_{step} = 2$ (`0111`) was the best performing according to LOO-ELPD. However, the LOO-ELPD for this model was not statistically different to the diagnostic case model (ELPD difference = 0.9074112; Table 4) and model estimates were nearly identical. In general, the increase in process error estimation as n_{step} increased, the reduced ability to fit the catch were undesirable, and the model equivalence according to LOO-ELPD supported maintaining the diagnostic case model with $n_{step} = 1$.

4.2.7 Catch reporting

Striped marlin are a commercially valuable non-target species, and systematic under-reporting of catches or discarding could occur. Three models (`0115` - `0117`) explored potential systematic under-reporting scenarios including constant and time-varying multipliers. As expected, increasing the magnitude of all the catches by $2\times$ (`0115`) shifted the population scale of the population up by ~2 (Figure 30) though relative quantities were unchanged (Figure 31). However, increasing either the terminal (`0116`) or initial (`0117`) catches by $2\times$ resulted in an increase in population scale, change in the depletion trend, and change in the management advice (Figure 31). The `0116` model represented a scenario where potential non-reported catch increased over time, but resulted in a less depleted stock with a removals time series that increased. The `0117` model represented a scenario where non-reported catch decreased over time producing a more depleted stock with a decreasing removals time series.

Though there is no evidence at this stage that this level of systematic under-reporting of catches is occurring for SWPO striped marlin, it is useful to be aware of the impacts that changes to our perception of population removals can have on our understanding of stock status. Additionally, for future analyses that may consider changes to the level of catch reporting, it may be useful to do so on a *per hook* basis to take into account the impact of the change in the number of hooks fished over time on total population removals.

4.2.8 Process error σ_P

The diagnostic case model estimated non-zero process error in the early model period before the CPUE index began. Process error was restricted $\sigma_P \sim 0$ as a one-off sensitivity (`0118`) in order to evaluate if larger initial process error deviates impacted scale estimates or other management quantities. Turning-off the process error did degrade fit to the index slightly (Figure 32), but overall parameter estimates (Figure 33) and management quantities (Figure 34) were largely insensitive to a lack of process error. This confirms that the early process error deviates are not impacting the estimation of key management quantities, and also provides further evidence of a well-determined production function.

4.2.9 Start year

Given potentially anomalous catch observations in 1952 and 1954, four models (`0119` - `0122`) explored sensitivity to the assumed initial conditions. Models started in either 1955 or 1988 with varying levels of initial depletion. In every case, model results (Figure 35) appeared to be fairly sensitive to the choice of prior, however all models except perhaps model `0122` (1988 start with $x_0 = 0.9$) still estimated substantially lower population carrying capacity $\log(K)$ than the prior (Figure 36). However, these models all assumed fairly restrictive priors on x_0, and future analyses could evaluate sensitivity to less informative priors or alternatively develop more specific priors for given years based on a prior pushforward analysis.

4.2.10 Inclusion of effort deviates

Early model development explored a more complex formulation of fishing mortality where effort deviates were used in addition to catchability deviates. Two models (`0124` - `0126`) tested whether returning to a full MULTIFAN-CL style *catch-errors* formulation and estimating effort deviates improved model performance. Estimating effort deviates as expected, did produce an improved fit to the catch (Figure 37) however parameter estimates (Figure 38) and model quantities (Figure 39) showed negligible differences from the diagnostic case

26

model.

4.2.11 Direct estimation of fishing mortality

Rather than estimating fishing mortality indirectly through catchability and effort, two models (`0127` - `0128`) directly estimated annual fishing mortality rates. This tested whether the simpler formulation for fishing mortality could replicate the results of the diagnostic case model, or if the highly parametrized structure of the fishing mortality calculation determined model outcomes. Removing the relationship with effort, and directly estimating fishing mortality allowed the catch to be fit exactly (Figure 40) though this resulted in high levels of catch observations showing large Pareto-k values (Table 4) which is undesirable. Matching the prior variability on fishing mortality σ_F to the level of variability in fishing mortality seen in the diagnostic case model (`0127`) more or less replicated the estimates from the diagnostic case model (Figure 41). Assuming a smaller, more restrictive prior on σ_F (`0128`) resulted in a large increase to population scale to accommodate a time-series of fishing mortality with smaller variability (Figure 41). These results showed that model outcomes were insensitive to how fishing mortality was parametrized, and reaffirmed the sensitivity of model results, particularly population scale, to the priors placed on fishing mortality.

4.2.12 Alternative R_{Max} prior

The diagnostic case model estimated a stock with higher productivity than an analogous species, Atlantic white marlin. Two models (`0133` - `0134`) tested sensitivity to a more conservative productivity assumption based on the white marlin R_{Max} estimates from ICCAT (ICCAT, 2020). As hypothesized, forcing productivity R_{Max} to be lower did result in the model estimating a larger population scale (Figure 42). However, this did not necessarily translate to differences in estimated relative management metrics (depletion, D/D_{MSY}, or F/F_{MSY}), particularly in recent years.

4.3 Model ensemble

Following extensive sensitivity analysis, and iterative rounds of model development, a four model ensemble is proposed to be considered for the provision of management advice:

- `0100` : DWFN index & life history based shape prior
- `0102` : New Zealand index & life history based shape prior
- `0105` : DWFN index & alternative shape prior
- `0107` : New Zealand index & alternative shape prior

Both candidate indices were the only indices of sufficient length and contrast to inform estimates of population scale. Additionally, the shape parameter is very difficult to estimate from data, and as such model estimates are dependent on the choice of prior. While the baseline shape parameter prior (Table 2) may be sensibly derived from striped marlin life history characteristics, this derivation is sensitive to underlying assumptions and relationships (Fowler, 1988; Grant & Grant, 1992). In the absence of other information, a conservative assumption for the shape parameter is to assume it is centered on $n \sim 2$, with uncertainty from Table 2. This reduces the production model to a Schaefer-type production model where $D_{MSY} = 0.5$.

Despite use of a model ensemble to capture uncertainty in stock status and trend, there remains considerable uncertainty not captured within the proposed BSPM ensemble. Important context to the interpretation of ensemble results and conclusions can be found in Table 7.

4.3.1 Ensemble diagnostics and fit

4.3.1.1 Convergence

All ensemble models achieved satisfactory convergence based on standard Hamiltonian Monte Carlo diagnostics (Table 4). All models met convergence criteria with $\hat{R}$ values well below the 1.01 threshold and effective sample sizes exceeding the recommended minimum of 500 (100 samples per chain). No divergent transitions were observed across chains, and maximum tree depth was not exceeded during sampling, indicating successful exploration of the posterior space. The proportion of observations with Pareto-k > 0.7 was very low (<5%) across all models and the effective number of estimated parameters p_loo was well below both the number of observations and the actual estimated parameter count.

4.3.1.2 Data fit

Model fits to the catch data were equivalent between models and captured the trend and magnitude of the catches adequately (Figure 43). For models fitting to the same index (e.g., `0100` & `0105`), the alternative assumption for shape n had no influence on the estimated fit. However, comparing relative fits between models fitting to different indices, the models fitting to the DWFN index did a noticeably better job (lower RMSE) capturing the trend than models fitting to the New Zealand index (Table 5;Figure 44). As mentioned in Section 4.2.1, models fitting to the New Zealand index failed to capture the trend over the last five years.

4.3.1.3 Model validation

Overall retrospective bias was within the acceptable range proposed by (Hurtado-Ferro et al., 2015) for all models, with the models fitting to the New Zealand index showing positive retrospective bias rather than negative for models fitting to the DWFN index (Table 5). Furthermore, all quantities were within the 95% credible interval of the full time-series model for all peels (Table 5; Figure 45).

For hindcast cross-validation, both models fitting to the DWFN index substantially exceeded the performance of the naïve predictor (Table 5; Figure 46). Models fitting to the New Zealand recreational index showed poor performance (MASE > 1), as the models were unable to follow the declining trend over the last five years.

4.3.2 Population dynamics and trends

The estimated depletion[8] time series from the model ensemble shows a decline from the assumed unfished state in 1952 to minimum levels around the mid-2010s (Figure 47). The decline was most pronounced during the 1950s and 1960s, followed by relatively stable depletion levels from the 1970s through the early 2000s which fluctuated around $\sim 0.4 - 0.6$. The stock has been estimated to be recovering from the mid-2010s. Credible intervals around depletion estimates are relatively wide, reflecting inherent uncertainty in the population scale.

The estimated population time series mirrors these depletion patterns. It is important to note that while the uncertainty in absolute population size is substantial, as reflected in wide credible intervals, this uncertainty is asymmetric. The model is fairly confident that the model can not be much smaller than what it currently

---

[8]Note that depletion for the surplus production model is a *static* depletion defined relative to total unfished numbers $D_t = P_t/P_0$; where P_t is the total population level in numbers at time t, and $P_0 = K$ the unfished population numbers or carrying capacity. This is different from the *dynamic* depletion of spawning biomass $SB_t/SB_{t,F=0}$ that is typically reported from integrated age-structured models such as MULTIFAN-CL or Stock Synthesis (Methot Jr. & Wetzel, 2013).

is estimated to be due to the hard constraint of at least being large enough to support the 70 year time-series of catches. As seen in the one-off sensitivities (Section 4.2.4 & Section 4.2.12), uncertainty to the high-side is driven in part by the prior assumptions for $\log(K)$, q_{eff}, and R_{Max}, as well as by model structural choices with regards to treatment of catch uncertainty (Section 4.2.5) and structure of fishing mortality (Section 4.2.10).

Process error estimates showed consistent temporal variability across ensemble models, with the largest deviations occurring in the 1970s and 1980s. This indicates that removals alone could not completely explain the observed changes in the standardized indices in these time periods, suggesting additional factors influenced population dynamics or insufficient observation error was estimated.

Fishing mortality F estimates jump up initially in the mid-1950s with the expansion of industrial fishing into broader areas of the south Pacific. Though F appears highly variable, with substantial uncertainty reflected in wide credible intervals, it generally trends upwards to a peak (~ 0.275) in the early 2000s. Since then, F was generally predicted to decline in recent decades. The recent decline is consistent with the observed recovery in population depletion during the same period.

The catchability deviations show declining trends from elevated levels in the mid-1950s, with the largest declines occurring before 1970 and more gradual decline through the current year. This matches the pattern seen in the estimated nominal CPUE (catch per 1000 hooks) estimates from the model ensemble, where the estimated nominal CPUE is simply the predicted removals divided by the nominal effort. In the earliest years of the fishery nominal CPUE approached 3 fish per 1000 hooks before declining to ~ 0.45 fish per 1000 hooks in 1970, and ~ 0.05 fish per 1000 hooks at the end of the model period.

4.3.3   Stock status

The SWPO striped marlin stock lacks formal, agreed upon reference points. Accordingly, stock status is summarized in terms of the following MSY and depletion based reference points (Table 8): total depletion relative to total depletion at MSY (D/D_{MSY}), fishing mortality relative to fishing mortality at MSY (F/F_{MSY}), and total depletion relative to generalized limit reference point of 20% depletion from the unfished state ($D/D_{0.2F=0}$). All ratios are calculated with respect to two periods: *recent* and *latest*. For depletion ratios, D_{recent} refers to the average over 2019-2022 and D_{latest} refers to 2022. For fishing mortality ratios, F_{recent} refers to the average over 2018-2021 and F_{latest} refers to 2021. Uncertainty in stock status and management reference points is presented in the paper using conventional, equal tailed 95% posterior credible intervals (ETIs). An appendix (Section 12.1) is included which provides the corresponding highest density intervals (HDIs), as these may more appropriately capture uncertainty in skewed distributions (A. M. Edwards & Auger-Méthé, 2025).

On average, the stock is estimated to be in an overfished state according to both D_{recent}/D_{MSY} (0.77 median; $0.33 - 2.3$ 95% credible interval) and D_{latest}/D_{MSY} (0.81 median; $0.32 - 2.36$ 95% credible interval), but is not currently estimated to be undergoing overfishing according to F_{recent}/F_{MSY} (0.77 median; $0.05 - 1.51$ 95% credible interval) or F_{latest}/F_{MSY} (0.69 median; $0.05 - 1.51$ 95% credible interval).

Considering the full range of uncertainty, it is likely that the stock is currently below D_{MSY} (74% for recent period, 70% for latest year) and it is unlikely that the stock is currently undergoing overfishing (-21.9% for recent period, -17.4% for latest year).

Relative to the generalized limit reference point of 20% depletion from the unfished state, the stock shows

a very unlikely probability of being in a critically depleted state. The probability of the stock being below $D_{0.2F=0}$ is 9.2% for the recent period and 8.9% for the latest year, with median ratios of 1.84 (0.73 – 4.72 95% credible interval) and 1.91 (0.72 – 4.81 95% credible interval), respectively.

The stock status indicators suggest that while the median estimates indicate the stock is currently overfished but not undergoing overfishing, there remains uncertainty around these conclusions (Figure 48). The relatively wide credible intervals reflect substantial uncertainty in absolute population scale, however there is still only a 22.92% joint probability that $D_{recent}/D_{MSY} < 1$ and $F_{recent}/F_{MSY} > 1$ so it is unlikely that the stock is simultaneously overfished and undergoing overfishing. The very unlikely risk relative to the generalized limit reference point provides additional confidence that the stock is not in a critically depleted state (Figure 49) and, as described in Section 4.3.2, the population appears to be trending in a positive direction.

4.3.4   Future projections

The 10-year stochastic projections from the ensemble, setting future catch as the recent average catch over the last 5 years (2018-2022), showed the stock expected to continue on the estimated recovery trajectory (Figure 50). Given that the assessment period ends in 2022, projections to the current year 2025 indicate median $D_{latest}/D_{MSY} = 0.99$ (0.24 – 2.42 95% credible interval) with only a 50.78% chance that the stock is overfished according to the MSY based reference point. By next year, 2026, there is a less than 50% chance that the stock is overfished. Looking out to the end of the projection period, 2032, the picture continues to improve with median $D_{latest}/D_{MSY} = 1.32$ (0 – 2.6 95% credible interval) with only a 26.05% chance that the stock is overfished.

Fishing mortality is also predicted to keep declining through 2032 with median $F_{latest}/F_{MSY} = 0.49$ (0.04 – 34.53 95% credible interval) with only a 14.6% chance that overfishing is occurring.

Though circumstances are expected to improve on average, and overfished and overfishing risk levels are expected to decrease throughout the projection period there is still a chance that the population collapses ($D_{2032} < 0.05$) under recent average catch levels by the end of the 10-year projection period. This is estimated to be very unlikely with only a 5.12% chance of occurring. However, risk of stock collapse would be expected to increase if catches rose above recent average levels.

5   Discussion

5.1   General remarks

The BSPM ensemble approach, and sensitivity analyses, present a coherent picture of stock trajectory and current status. The SWPO striped marlin stock was estimated to decline substantially in the 1950s from an unfished condition with minimum levels being experienced within the last decade, and the stock showing signs of recovery. Though median estimates suggest that the stock is overfished and undergoing overfishing at the end of the model period (2022), the stock is expected to no longer be overfished or undergoing overfishing by 2026 assuming recent catches have not increased.

The BSPM approach addresses key limitations encountered in previous integrated assessments while providing robust stock status estimates. The primary advantage lies in condensing complex biological and fishery uncertainties into a manageable parameter set. Integrated assessments can be influenced by size composition

data, which are sensitive to assumptions about growth, mortality, selectivity, and reproduction parameters. Additionally, there is the added burden of ensuring that the input size composition data, along with the indices, are unbiased and representative of the stock dynamics. When these assumptions are uncertain or mis-specified, population scale estimates informed by size composition data can be biased. Production models can help circumvent this issue by aggregating these uncertainties into productivity parameters, allowing more efficient exploration of plausible population dynamics while maintaining biological realism through informed priors. Uncertainty in these processes is propagated both through the prior and through full Bayesian Markov Chain Monte-Carlo sampling of the posterior distribution. Use of a model ensemble approach also captures uncertainties in data or structural components of the model that can not be estimated.

However, BSPMs are an inherent simplification of the stock dynamics. All age-structure is collapsed, a single well mixed population is assumed, and only a single fishery (with implicit knife-edged selectivity) operates. Though previous simulation work (Winker et al., 2019) has shown that correctly specified BSPMs can provide robust estimates of MSY based reference points, this is conditioned on the dynamics of the underlying population not being too different from the assumed BSPM dynamics. If movement dynamics are complex, selectivities are strongly dome-shaped, substantial mortality occurs before reproduction, and/or there is a long lag between recruitment and reproduction this can all introduce bias into the BSPM estimates.

The origin of this analysis was rooted in a skepticism in the depletion and fishing mortality patterns estimated from the integrated assessment model. The BSPM results paint a complimentary picture to the integrated assessment, and demonstrates that estimates of initial depletion and population scale are strongly informed by the magnitude of catches in contrast with the standardized index. Moreover, the predicted recovery is based on a well determined production function as evidenced by the performance of the model-free hindcasts (Section 4.1.5), and should give confidence that projections from this model can capture the stock dynamics.

The estimated trajectory of the stock, including patterns in estimated catchability and nominal CPUE make sense in the context of both Japanese and global longline fishing development (Ahrens, 2010; Honma & Kamimura, 1958). Post World War II MacArthur lines restricted Japanese fishing to the northern hemisphere in the Pacific prior to 1952. Expansion in the south Pacific took place in 1952 in waters close to the equator. Expansion by the Japanese south of $20°S$ did not occur until November 1953 (Honma & Kamimura, 1958) at which point striped marlin was targeted alongside albacore (*Thunnus alalunga*) initially due to high catch rates but eventually to supply the developing Japanese fish ham and sausage industry (Joseph et al., 1974). High catch rates to start the fishery make sense given the targeted nature of the fishery, but also due to fishing on spawning aggregations (Honma & Kamimura, 1958), and early gear configurations (rope or hemp line with few hooks per basket) which kept gear closer to the surface where striped marlin are present. As Japanese (and eventually other DWFNs) longline fishing effort expanded throughout the south Pacific, away from striped marlin hotspots and with more targeted fishing towards albacore and the tropical tunas (bigeye *Thunnus obesus* and yellowfin *Thunnus albacares*) striped marlin catchability and nominal CPUE declined. Furthermore, more modern longline gear configuration (monofilament main lines with large numbers of hooks per basket) places more hooks at deeper depths, which may increase tropical tuna catchability but reduce striped marlin catch rates.

This fisheries context can also provide more insight on the two catch observations (1952 & 1954) that were investigated as being potentially problematic. The initial 1952 low catch observation may be legitimate as effort was not deployed spatially in an area (north of $20°S$) where striped marlin are highly abundant (Honma

& Kamimura, 1958). We can also see that the large catch observation in 1954 may also be legitimate:

- observed nominal CPUE in 1954 (~ 4.22) is comparable to observations in 1953 (~ 3.17) and 1955 (~ 3.94)
- fishing occurred in an area ($15°S - 30°S$) and time period (austral spring and summer) when striped marlin form spawning aggregations (Honma & Kamimura, 1958; Kopf et al., 2012)
- the Japanese considered striped marlin to be a target species in the 1950s and 1960s (Honma & Kamimura, 1958; Joseph et al., 1974)

Of a more technical nature, this assessment applies multiplicative fishing mortality to the population, and estimates fishing mortality from a time-varying relationship with effort by fitting to the observed catch. While such an approach is less conventional for surplus production models[9], and more common in integrated age-structured modelling approaches, it appears to yield important computational benefits particularly when fitting models in a Bayesian context with HMC. Modelling fishing mortality as a multiplicative process ensures that populations stay non-negative, and avoids the hard, implicit constraints that direct subtraction of the catch places on the posterior sampling space. Though applying multiplicative fishing mortality improves model convergence (e.g., 0 or small numbers of divergent transitions), it does mean that fishing mortality, surplus production and process error happen simultaneously rather then sequentially as in the subtractive case. This assumption may not be appropriate for some fisheries which operate in distinct seasons relative to reproduction.

5.2 Challenges, limitations & key uncertainties

While the BSPM ensemble approach provides a robust framework for stock assessment, several important limitations and uncertainties constrain the interpretation of these results and their applicability to management decision-making.

Though not specific to the current SWPO striped marlin stock, assessments are only as good or informative as the underlying data. In this case while the results appear robust, and the model appears to appropriately capture the underlying population dynamics, this is all conditioned on the data being representative of the stock. From the one-off sensitivities it was apparent that introducing alternative trend to the catch via systematic under-reporting (Section 4.2.7) could produce different population trajectories and management advice. Similarly, the population trend is driven by the index. The two primary indices used in the ensemble (DWFN and New Zealand recreational) each have their own separate issues. The DWFN index, while providing a long time series of information, is a composite multi-flag index with limited gear covariates. Changes in fleet composition or fishing location within this index can be incorrectly interpreted by the model as population-level changes. On the other hand the New Zealand recreational index includes more catchability covariates is spatially restricted to index the striped marlin population vulnerable to the East Northland[10] sportfish fishery. While historically this fishery is believed to have interacted with the spawning component of the population during the austral summer as fish approached the southern end of their distribution, there is evidence of increasing striped marlin encounters further south along the western side of New Zealand's South Island (Holdsworth, 2023). If the distribution of striped marlin has shifted or is shifting further south following warmer waters as a result of climate change, then these data may no longer be representative and

---

[9]Most surplus production models subtract observed catch directly (Best & Punt, 2020; Kim & Neubauer, 2025; Millar & Meyer, 2000)

[10]North Island of New Zealand

the decline seen in terminal years of the New Zealand index (Figure 44) may not represent a true decline in the stock.

While R_{MAX} and $\log(K)$ showed meaningful posterior updates, the uncertainty reflected in their prior distributions translated into the large uncertainty evident in model predictions. In the case of R_{MAX}, uncertainty in growth, maturity, natural mortality, and steepness all contributed to the broad prior uncertainty. Reductions in the uncertainty in any one of these biological processes would reduce uncertainty in the prior, and could help reduce uncertainty in model estimated population scale (see Section 4.2.12). Population scale $\log(K)$ estimates also showed particular sensitivity to prior assumptions and model configuration, as well as a strong negative correlation with baseline catchability q_{eff}. Reducing the uncertainty in catchability q_{eff} would presumably directly lower uncertainty in population scale.

In contrast to R_{MAX} and $\log(K)$ which were estimable from the data, the production function shape parameter n was not and posterior estimates did not deviate much from prior assumptions. Uncertainty in shape n translates directly to uncertainty in D_{MSY} and the position of MSY-based reference points. The ensemble approach attempts to capture this uncertainty through alternative shape priors, but the true production function shape remains a key unknown.

While the BSPM approach makes a number of simplifications to the stock population dynamics, assuming a single well-mixed stock may be problematic given the broad spatial distribution and known seasonal migrations (Honma & Kamimura, 1958). Of particular concern is recent genetic work (Martinez, 2021) that identified mixing of SWPO striped marlin and central north Pacific Ocean striped marlin in the catch of the U.S. Hawai'i based longline fishery. This is a fishery that operates exclusively in the north Pacific Ocean around the main Hawaiian islands, away from the typical stock boundary assumed for SWPO striped marlin. Additional genetic work is needed to more properly elucidate the patterns in spatial connectivity. However, at a minimum, evidence of SWPO striped marlin in catch taken in the north Pacific Ocean indicates that the catch levels assumed in the current assessment may be an underestimate of total removals, and that potentially important fishing impacts from north Pacific Ocean fleets on SWPO striped marlin are unaccounted for. Development of a conceptual model for SWPO striped marlin, perhaps in collaboration with the ISC Billfish Working Group, that accounts for this stock structure hypothesis is critically needed so that this important uncertainty can be properly addressed in the next stock assessment.

Though not a limitation unique to the BSPM approach, this assessment assumes stationary productivity and carrying capacity over the 70-year assessment period, which may not reflect reality given environmental changes in the Pacific Ocean over the same time period. Additionally, climate variability affecting prey distribution, ocean temperature, and ecosystem productivity could influence striped marlin population dynamics in ways not captured by the surplus production framework. The large process error deviations estimated in some periods suggest environmental factors may be influencing population dynamics beyond what can be explained by fishing removals alone. More sophisticated modeling approaches incorporating environmental and/or ecosystem drivers are likely needed to appropriately capture the process, however this will likely stretch the limits of what is possible given the data currently available. Absent other sources of information coming online, development of robust management measures (e.g., through a management strategy evaluation; MSE) will likely be more useful than a more sophisticated stock assessment model.

Specifically with regards to management, the substantial uncertainty in absolute population scale, while not preventing relative status determination, constrains the precision of catch-based management advice. Though

uncertainty in population scale is large, it is asymmetrical and the minimum stock size is likely to be well defined which can be useful when developing potential management measures. Additionally, the stock status projections, while suggesting continued recovery under recent catch levels, carry forward all the uncertainties inherent in the assessment model. The projections assume continued stationarity in productivity, catchability, and environmental conditions that may not hold over the 10-year projection period. Furthermore, the status quo catch assumption may not reflect actual future fishing patterns. If management action is to be taken, alternative future projection scenarios (including effort based projection scenarios) should be developed.

Despite these limitations, the BSPM approach provides valuable insights into stock trajectory and status when viewed as part of a broader assessment framework. The consistency of results across sensitivity analyses and the validation through retrospective and hindcast analyses suggest that the qualitative conclusions about stock recovery are robust to many of the identified uncertainties. However, managers should consider these limitations when translating assessment results into specific management actions, particularly those requiring precise abundance estimates or long-term projections.

5.3 Future stock assessment modeling considerations

Future assessment efforts should progressively build toward full age-structure through a systematic approach: delay-difference models, age-structured production models, and then fully integrated age-structured models. This development should occur within a Bayesian framework to inform parameter estimates where prior information exists or can be derived using prior pushforward approaches (Kim & Neubauer, 2025).

Developing Bayesian age-structured modeling approaches can provide more stable estimation of increased parameters due to the use of priors, explicitly account for parameter uncertainty rather than making fixed assumptions, potentially enable informed estimation of difficult-to-observe biological relationships (e.g., stock-recruit relationships), address BSPM limitations by explicitly considering age-structured processes and fishery selectivity, and estimate leading parameters with likelihood components for indices and size composition.

While simplified models offer advantages when data conflicts exist, oversimplification of age-structured processes could result in biases. Future modeling must balance complexity with data quality and conflicts. The transition should be guided by careful evaluation of whether increased model complexity is supported by available data and whether it improves the reliability of management advice. A crucial part of the assessment process is careful scrutiny of the representativeness of the available data, and matching the modeling approach to the information content in the data available. While it is important to also match the modeling approach to our best understanding of stock dynamics, ideally informed by a conceptual model (Minte-Vera et al., 2024), this must be done in a manner that does not overextend the model beyond what can be reliably estimated from the data given parameter uncertainties and structural assumptions.

Lastly, this assessment greatly benefited from collaboration with a broad group of interested parties and stock assessment experts (see Section 6) where feedback was provided in an iterative manner throughout the model development process. Leveraging this diverse pool of experience helped produce a stronger scientific product.

6   Acknowledgements

7   Declaration of generative AI use

A generative artificial intelligence (AI) assistant, Anthropic Claude Sonnet 4.0, was used to parse model code in the preparation of an initial draft of this report.

8   References

Ahrens, R. N. M. (2010). *A global analysis of apparent trends in abundance and recruitment of large tunas and billfishes inferred from Japanese longline catch and effort data.* https://doi.org/10.14288/1.0069329

Best, J. K., & Punt, A. E. (2020). Parameterizations for Bayesian state-space surplus production models. *Fisheries Research*, *222*, 105411. https://doi.org/10.1016/j.fishres.2019.105411

Betancourt, M. J., & Girolami, M. (2013). *Hamiltonian Monte Carlo for Hierarchical Models.* https://doi.org/10.48550/ARXIV.1312.0906

Castillo-Jordan, C., Day, J., Teears, T., Davies, N., Hampton, J., McKechnie, S., Magnusson, A., Peatman, T., Vidal, T., Williams, P., & Hamer, P. (2024). *Stock Assessment of Striped Marlin in the Southwest Pacific Ocean: 2024* (WCPFC-SC20-2024/SA-WP-03 (Ver. 2.01)). https://meetings.wcpfc.int/node/23120

Castillo-Jordan, C., Ducharme-Barth, N. D., Carvalho, F., & Hamer, P. (2025). *Revised 2024 stock assessment of striped marlin in the southwestern Pacific Ocean: Part 1- integrated assessment in Stock Synthesis* (WCPFC-SC21-2025/SA-WP-06).

Ducharme-Barth, N., Castillo-Jordan, C., Sculley, M., Ahrens, R., & Carvalho, J. (2025). *Summary report from the NOAA-SPC assessment model meeting on SWPO striped marlin* (WCPFC-SC21-2025/SA-IP-XX).

Edwards, A. M., & Auger-Méthé, M. (2025). Using highest density intervals can reduce perceived uncertainty in stock assessments. *Fisheries Research*, *288*, 107326. https://doi.org/10.1016/j.fishres.2025.107326

Edwards, C. (2024). *Bdm: Bayesian biomass dynamics model* (R package version 0.0.0.9036).

Farley, J., Eveson, P., Krusic-Golub, K., & Kopf, K. (2021). *Southwest Pacific Striped Marlin Population Biology (Project 99)* (WCPFC-SC-17/SA-IP-11). https://meetings.wcpfc.int/node/12569

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

Fowler, C. W. (1988). Population dynamics as related to rate of increase per generation. *Evolutionary Ecology*, *2*(3), 197–204. https://doi.org/10.1007/BF02214283

Gabry, J., & Mahr, T. (2024). *Bayesplot: Plotting for Bayesian Models* (R package version 1.11.1). https://mc-stan.org/bayesplot/

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian Workflow. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *182*(2), 389–402. https://doi.org/10.1111/rssa.12378

Grant, P. R., & Grant, B. R. (1992). Demography and the Genetically Effective sizes of Two Populations of

Darwin's Finches. *Ecology*, *73*(3), 766–784. https://doi.org/10.2307/1940156

Gravel, S., Bigman, J. S., Pardo, S. A., Wong, S., & Dulvy, N. K. (2024). Metabolism, population growth, and the fast-slow life history continuum of marine fishes. *Fish and Fisheries*, *25*(2), 349–361. https://doi.org/10.1111/faf.12811

Hamel, O. S., & Cope, J. M. (2022). Development and considerations for application of a longevity-based prior for the natural mortality rate. *Fisheries Research*, *256*, 106477. https://doi.org/10.1016/j.fishres.2022.106477

Holdsworth, J. C. (2023). *Striped marlin catch and CPUE in the New Zealand sport fishery, 2019-20 to 2021-22* (New Zealand Fisheries Assessment Report 2023/05).

Honma, M., & Kamimura, T. (1958). *A Population Study on the so-called Makajiki (Striped Marlin) of both Northern and Southern Hemispheres of the Pacific: II. Fishing Conditions in the Southern Hemisphere* (Report of Nankai Regional Fisheries Research Laboratory, No. 8; pp. 12–21).

Hordyk, A., Huynh, Q., & Carruthers, T. (2024). *openMSE: Easily Install and Load the 'openMSE' Packages* (R package version 1.0.1). https://openmse.com/

Hurtado-Ferro, F., Szuwalski, C. S., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., Licandeo, R., McGilliard, C. R., Monnahan, C. C., Muradian, M. L., Ono, K., Vert-Pre, K. A., Whitten, A. R., & Punt, A. E. (2015). Looking in the rear-view mirror: Bias and retrospective patterns in integrated, age-structured stock assessment models. *ICES Journal of Marine Science*, *72*(1), 99–110. https://doi.org/10.1093/icesjms/fsu198

Hutchings, J. A., Myers, R. A., García, V. B., Lucifora, L. O., & Kuparinen, A. (2012). Life-history correlates of extinction risk and recovery potential. *Ecological Applications*, *22*(4), 1061–1067. https://doi.org/10.1890/11-1313.1

ICCAT. (2020). Report of the 2019 ICCAT white marlin stock assessment meeting. *Collect. Vol. Sci. Pap. ICCAT*, *76*(4), 97–181. https://www.iccat.int/Documents/CVSP/CV076_2019/colvol76.html#

ISC. (2024). *Stock Assessment of Shortfin Mako Shark in the North Pacific Ocean through 2022* (WCPFC-SC20-2024/SA-WP-14). https://meetings.wcpfc.int/node/22828

Joseph, J., Klawe, W. L., & Orange, C. J. (1974). *A review of the longline fishery for billfishes in the eastern Pacific Ocean* (R. S. Shomura & F. Williams, Eds.; NOAA Technical Report NMFS SSRF-675; pp. 309–331).

Juárez, M. A., & Steel, M. F. J. (2010). Model-Based Clustering of Non-Gaussian Panel Data Based on Skew-*t* Distributions. *Journal of Business & Economic Statistics*, *28*(1), 52–66. https://doi.org/10.1198/jbes.2009.07145

Kapur, M. S., Ducharme-Barth, N., Oshima, M., & Carvalho, F. (2025). Good practices, trade-offs, and precautions for model diagnostics in integrated stock assessments. *Fisheries Research*, *281*, 107206. https://doi.org/10.1016/j.fishres.2024.107206

Kell, L. T., Sharma, R., Kitakado, T., Winker, H., Mosqueira, I., Cardinale, M., & Fu, D. (2021). Validation of stock assessment methods: Is it me or my model talking? *ICES Journal of Marine Science*, *78*(6), 2244–2255. https://doi.org/10.1093/icesjms/fsab104

Kim, K., & Neubauer, P. (2025). Examining potential biases through prior predictive checks: Prior mis-specifications and their impact on Bayesian stock assessments. *Fisheries Research*, *288*, 107405. https://doi.org/10.1016/j.fishres.2025.107405

Kokkalis, A., Berg, C. W., Kapur, M. S., Winker, H., Jacobsen, N. S., Taylor, M. H., Ichinokawa, M., Miyagawa, M., Medeiros-Leal, W., Nielsen, J. R., & Mildenberger, T. K. (2024). Good practices for surplus

production models. *Fisheries Research*, *275*, 107010. https://doi.org/10.1016/j.fishres.2024.107010

Kopf, R. K., Davie, P. S., Bromhead, D. B., & Young, J. W. (2012). Reproductive biology and spatiotemporal patterns of spawning in striped marlin *Kajikia audax*. *Journal of Fish Biology*, *81*(6), 1834–1858. https://doi.org/10.1111/j.1095-8649.2012.03394.x

Martinez, J. (2021). *Stock Composition Of Striped Marlin (Kajikia Audax) In The Central North Pacific Ocean Inferred By Analyses Of Genome-Wide Molecular Markers*. https://doi.org/10.25773/WN5A-5R23

Merino, G., Urtizberea, A., Fu, D., Winker, H., Cardinale, M., Lauretta, M. V., Murua, H., Kitakado, T., Arrizabalaga, H., Scott, R., Pilling, G., Minte-Vera, C., Xu, H., Laborda, A., Erauskin-Extramiana, M., & Santiago, J. (2022). Investigating trends in process error as a diagnostic for integrated fisheries stock assessments. *Fisheries Research*, *256*, 106478. https://doi.org/10.1016/j.fishres.2022.106478

Methot Jr., 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

Millar, R. B., & Meyer, R. (2000). Non-Linear State Space Modelling of Fisheries Biomass Dynamics by Using Metropolis-Hastings within-Gibbs Sampling. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *49*(3), 327–342. https://doi.org/10.1111/1467-9876.00195

Minte-Vera, C. V., Maunder, M. N., Aires-da-Silva, A., Xu, H., Valero, J. L., Teo, S. L. H., Barría, P., & Ducharme-Barth, N. D. (2024). The use of conceptual models to structure stock assessments: A tool for collaboration and for "modelling what to model." *Fisheries Research*, *279*, 107135. https://doi.org/10.1016/j.fishres.2024.107135

Mohn, R. (1999). The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. *ICES Journal of Marine Science*, *56*(4), 473–488. https://doi.org/10.1006/jmsc.1999.0481

Monnahan, C. C. (2024). Toward good practices for Bayesian data-rich fisheries stock assessments using a modern statistical workflow. *Fisheries Research*, *275*, 107024. https://doi.org/10.1016/j.fishres.2024.107024

Neubauer, P., Castillo-Jordan, C., Day, J., & Hamer, P. (2025). *Exploring the potential for observer CPUE for southwest Pacificswordfish (*Xiphias gladius*) and striped marlin (*Kajikia audax*)* (WCPFC-SC21-2025/SA-IP-13). https://meetings.wcpfc.int/node/26539

Neubauer, P., Kim, K., Large, K., & Brouwer, S. (2024). *Stock Assessment of Silky Shark in the Western and Central Pacific Ocean 2024* (WCPFC-SC20-2024/SA-WP-04-Rev2). https://meetings.wcpfc.int/node/23088

Neubauer, P., Richard, Y., & Tremblay-Boyer, L. (2019). *Alternative assessment methods for oceanic whitetip shark* (WCPFC-SC15-2019/SA-WP-13).

Pardo, S. A., Kindsvater, H. K., Reynolds, J. D., & Dulvy, N. K. (2016). Maximum intrinsic rate of population increase in sharks, rays, and chimaeras: The importance of survival to maturity. *Canadian Journal of Fisheries and Aquatic Sciences*, *73*(8), 1159–1163. https://doi.org/10.1139/cjfas-2016-0069

Sivula, T., Magnusson, M., Matamoros, A. A., & Vehtari, A. (2020). *Uncertainty in Bayesian Leave-One-Out Cross-Validation Based Model Comparison*. arXiv. https://doi.org/10.48550/ARXIV.2008.10296

Stanley, R. D., McAllister, M., Starr, P., & Olsen, N. (2009). *Stock assessment for bocaccio (*Sebastes paucispinis*) in British Columbia waters* (Canadian {Science} {Advisory} {Secretariat} Research Document 2009/055). Fisheries; Oceans Canada. https://www.dfo-mpo.gc.ca/csas-sccs/publications/resdocs-docrech/2009/2009_055-eng.htm

Team, S. D. (2024a). *RStan: The R interface to Stan* (R package version 2.32.6). https://mc-stan.org/

Team, S. D. (2024b). *Stan Modeling Language Users Guide and Reference Manual, v 2.32.6.*

Tremblay-Boyer, L., & Williams, A. (2024). *Standardised cpue indices for the target species in the eastern tuna and billfish fishery–1998 to 2023* (Technical Report Working Paper presented to the 41st meeting of the Tropical Tuna Resource Assessment Group held 16-17 July 2024, Brisbane, CSIRO, Hobart).

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Burkner, P., Paananen, T., & Gelman, A. (2024). *Loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models* (R package version 2.7.0). https://mc-stan.org/loo/

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Winker, H., Carvalho, F., & Kapur, M. (2018). JABBA: Just Another Bayesian Biomass Assessment. *Fisheries Research*, *204*, 275–288. https://doi.org/10.1016/j.fishres.2018.03.010

Winker, H., Mourato, B., & Chang, Y. (2019). Unifying parameterizations between age-structured and surplus production models: An application to atlantic white marlin (*Kajiki albidia*) with simulation testing. *Collect. Vol. Sci. Pap. ICCAT*, *76*(4), 219–234.

9 Tables

Table 1: Biological parameter distributions and sources used in simulation framework

Parameter	Symbol	Distribution	Mean/Mode	CV/Range	Correlation	Source/Notes
Length at age 1	L_1	Lognormal	60 cm	0.2	Independent	(Ducharme-Barth et al., 2025)
Length at age 10	L_2	Lognormal	210 cm	0.2	Independent	(Ducharme-Barth et al., 2025)
Growth coefficient	k	Beta(6.5, 3.5)	~0.65	-	Independent	(Ducharme-Barth et al., 2025)
Length CV	cv_{len}	Uniform	-	[0.05, 0.25]	-	Growth variability
Maximum age	A_{Max}	Lognormal	15 years	0.2	ρ = -0.3 with M_{ref}	(Farley et al., 2021)
Reference mortality	M_{ref}	Lognormal	0.36 yr^{-1}	0.44	ρ = -0.3 with A_{Max}	(Hamel & Cope, 2022) (5.4/A_{Max})
Maturity slope	a_{mat}	Normal	-20	0.2	-	Logistic steepness
Length at 50% maturity	L_{50}	Lognormal	184 cm	0.2	-	(Farley et al., 2021)
Weight coefficient	a_w	Lognormal	5.4×10^{-7}	0.05	ρ = -0.5 with b_w	(Castillo-Jordan et al., 2024)
Weight exponent	b_w	Normal	3.58	0.05	ρ = -0.5 with a_w	(Castillo-Jordan et al., 2024) Constrained: 150-350 kg at 300 cm
Steepness	h	Beta(3, 1.5)	~0.73	-	-	Censored to [0.2, 1.0]
Sex ratio (prop. female)	s	Normal	0.5	0.05	-	Constrained [0.01, 0.99]
Reproductive cycle	-	Fixed	1 year	-	-	Annual spawning
Reference ages	age_1, age_2	Fixed	0, 10 years	-	-	Francis parameterization

Table 2: Prior distributions for diagnostic model leading parameters.

Parameter	Distribution	Mean	SD	Description	Key Correlations
Core Population Parameters					
K	MV Lognormal	$\log(1,110,533)$	0.59	Carrying capacity (N)	$\rho(K, R_{Max}) = -0.22$; $\rho(K, q_{eff}) = -0.241$
R_{Max}	MV Lognormal	$\log(0.6206)$	0.58	Max rate of increase	$\rho(R_{Max}, n) = -0.881$
n	MV Lognormal	$\log(1.16)$	0.54	Pella-Tomlinson shape parameter	
x_0	MV Lognormal	$\log(1)$	0.025	Initial depletion	(Independent)
q_{eff}	MV Lognormal	$\log(0.5019)$	0.95	Initial catchability	$\rho(q_{eff}, \sigma_{qdev}) = 0.559$; $\rho(q_{eff}, \rho) = 0.243$
ρ	MV Transformed Normal	$\tanh^{-1}(0.981)$	0.62	Autocorrelation in catchability deviations	$\rho(\rho, \sigma_{qdev}) = 0.148$
σ_{qdev}	MV Lognormal	$\log(1.619)$	0.3	Catchability variability	
Process and Observation Error					
σ_P	Lognormal	$\log(0.0533)$	0.27	Process error	(Independent)
$\sigma_{O,add}$	Half-Normal	0	0.2	Additional estimated observation error	(Independent)
Catch Likelihood					
ν_{catch}	Gamma	Shape = 2	Rate = 0.1	Student-t degrees of freedom for catch	(Independent)

Notes:

Distributions with MV indicate that the parameter is part of the 7-dimensional multivariate prior.

Table 3: Model descriptions

Model	Group	Description
0100	Final ensemble	Diagnostic case model: DWFN index & baseline priors
0102	Final ensemble	New Zealand index & baseline priors
0105	Final ensemble	DWFN index & alternative shape prior
0107	Final ensemble	New Zealand index & alternative shape prior
0071	One-off: Index	Australia index
0072	One-off: Index	New Zealand index
0073	One-off: Index	Observer (all) index
0074	One-off: Index	Observer (New Caledonia, Fiji & Tonga) index
0075	One-off: Index	Observer (French Polynesia) index
0129	One-off: Index	DWFN & Australia index
0130	One-off: Index	New Zealand & Australia index
0131	One-off: Index	DWFN & Observer (New Caledonia, Fiji & Tonga) index
0132	One-off: Index	New Zealand & Observer (New Caledonia, Fiji & Tonga) index
0076	One-off: sigma_c	Time-varying sigma_c: Power curve from 0.5-0.2
0077	One-off: sigma_c	Constant sigma_c: 0.1
0078	One-off: sigma_c	Time-varying sigma_c: Power curve from 0.5-0.1
0079	One-off: Early data	Alternative 1952 catch
0080	One-off: Early data	Alternative 1952 effort
0081	One-off: Early data	Alternative 1954 catch
0082	One-off: Early data	Alternative 1952 & 1954 catch
0083	One-off: Early data	Alternative 1952 effort & 1954 catch
0093	One-off: Scaling priors	Force higher scale (logK & qeff)
0096	One-off: Scaling priors	Force higher scale (logK & qeff), alternative shape prior
0110	One-off: Catch obs model	Lognormal error structure
0111	One-off: qdev steps	n_step = 2
0112	One-off: qdev steps	n_step = 3
0113	One-off: qdev steps	n_step = 5
0114	One-off: qdev steps	n_step = 10
0115	One-off: Catch reporting	2x observed catch
0116	One-off: Catch reporting	Linear increase to 2x observed catch
0117	One-off: Catch reporting	Linear decrease from 2x observed catch
0118	One-off: Process error	Process error restricted to 0
0119	One-off: Start year	Start in 1955: x0 = 0.9
0120	One-off: Start year	Start in 1988: x0 = 0.5
0121	One-off: Start year	Start in 1988: x0 = 0.7
0122	One-off: Start year	Start in 1988: x0 = 0.9
0124	One-off: Est. effort devs	Estimate effort deviates & n_step = 2
0126	One-off: Est. effort devs	Estimate effort deviates & n_step = 5
0127	One-off: Direct F estimation	sigmaF to match diagnostic case
0128	One-off: Direct F estimation	0.25x diagnostic case sigmaF

Table 3: Model descriptions

Model	Group	Description
0133	One-off: Rmax prior	Apply lower productivity Rmax prior
0134	One-off: Rmax prior	Apply lower productivity Rmax prior and alternative shape prior

Table 4: Model diagnostics and fit statistics

Model	n par	n obs	Max $\hat{R}$	Min ESS	Divergent	Tree Depth	BFMI Issues	LOOIC	p LOO	ELPD LOO	Prop Pareto k
0100	150	114	1.008	788	0	0	0	45.4	32.7	-22.7	0.982
0102	150	118	1.008	688	0	0	0	64.9	32.9	-32.4	0.983
0105	150	114	1.008	677	0	0	0	46.1	33.0	-23.0	0.982
0107	150	118	1.007	523	0	0	0	66.1	33.3	-33.1	0.966
0071	150	95	1.007	792	0	0	0	-1.6	35.7	0.8	0.937
0072	150	118	1.014	610	0	0	0	65.3	33.7	-32.7	0.983
0073	150	92	1.008	738	0	0	0	-0.7	27.5	0.4	1.000
0074	150	92	1.008	760	0	0	0	30.2	28.9	-15.1	0.978
0075	150	90	1.036	230	0	0	0	58.3	29.7	-29.2	0.989
0129	150	139	1.014	666	0	0	0	24.7	36.5	-12.3	0.971
0130	150	143	1.012	760	0	0	0	56.2	39.3	-28.1	0.986
0131	150	136	1.005	710	0	0	0	35.9	35.5	-17.9	0.985
0132	150	140	1.011	783	0	0	0	61.7	36.7	-30.8	0.979
0076	150	114	1.008	644	0	0	0	49.8	19.2	-24.9	0.991
0077	150	114	1.013	693	0	0	0	32.0	57.3	-16.0	0.991
0078	150	114	1.010	779	0	0	0	28.6	26.9	-14.3	0.991
0079	150	114	1.010	754	0	0	0	19.5	33.6	-9.7	0.939
0080	150	114	1.010	773	0	0	0	20.9	34.3	-10.5	0.930
0081	150	114	1.006	748	0	0	0	45.8	32.1	-22.9	0.982
0082	150	114	1.009	748	0	0	0	21.3	33.4	-10.7	0.974
0083	150	114	1.008	537	0	0	0	21.5	33.3	-10.7	0.956
0093	150	114	1.006	749	0	0	0	52.4	35.3	-26.2	0.982
0096	150	114	1.007	746	1	0	0	57.0	35.1	-28.5	0.974
0110	149	114	1.008	729	0	0	0	1457.1	49.6	-728.5	0.711
0111	115	114	1.009	747	0	0	0	43.6	30.3	-21.8	0.974
0112	104	114	1.007	764	0	0	0	55.1	28.3	-27.5	0.974
0113	94	114	1.007	815	0	0	0	43.6	26.9	-21.8	0.982

Table 4: Model diagnostics and fit statistics

Model	n par	n obs	Max $\hat{R}$	Min ESS	Divergent	Tree Depth	BFMI Issues	LOOIC	p LOO	ELPD LOO	Prop Pareto k
0114	87	114	1.007	745	0	0	0	53.8	27.9	-26.9	0.991
0115	150	114	1.010	764	0	0	0	44.8	32.2	-22.4	0.991
0116	150	114	1.010	730	0	0	0	45.8	32.6	-22.9	0.982
0117	150	114	1.011	688	0	0	0	42.3	31.4	-21.2	0.982
0118	150	114	1.009	581	0	0	0	52.0	29.3	-26.0	0.991
0119	147	111	1.007	701	0	0	0	5.3	7.5	-2.6	0.977
0120	114	69	1.007	771	1	0	0	3.9	5.8	-2.0	0.971
0121	114	69	1.008	612	0	0	0	5.1	6.5	-2.6	0.971
0122	114	69	1.008	737	0	0	0	6.3	6.2	-3.1	0.971
0124	185	114	1.007	658	0	0	0	65.2	53.4	-32.6	0.860
0126	164	114	1.006	771	0	0	0	61.0	48.6	-30.5	0.904
0127	149	114	1.006	700	0	0	0	60.8	60.2	-30.4	0.535
0128	149	114	1.011	702	0	0	0	62.8	57.1	-31.4	0.544
0133	150	114	1.007	742	0	0	0	45.5	31.7	-22.8	0.982
0134	150	114	1.015	762	0	0	0	44.7	31.5	-22.4	0.982

Notes:

- n par: Number of parameters in the model

- n obs: Number of observations used in model fitting

- Max $\hat{R}$: Maximum potential scale reduction factor across all parameters (should be < 1.01)

- Min ESS: Minimum effective sample size across all parameters (should be > 500)

- Divergent: Number of divergent transitions during sampling (should be 0)

- Tree Depth: Whether maximum tree depth was exceeded during sampling

- BFMI Issues: Whether low Bayesian Fraction of Missing Information was detected

- LOOIC: Leave-one-out information criterion (lower is better)

- p LOO: Effective number of parameters estimated by LOO (should be < n par & n obs)

- ELPD LOO: Expected log pointwise predictive density from LOO (higher is better)

- Prop Pareto k: Proportion of Pareto k values <= 0.7 (should be high for reliable LOO estimates)

Table 5: Model performance and validation statistics

Model	Index RMSE	Catch RMSE	Mohn's ρ	MASE	Coverage D/D_{MSY}	Coverage U/U_{MSY}
0100	0.257	0.250	-0.087	0.587	100%	100%
0102	0.317	0.251	0.069	2.577	100%	100%
0105	0.258	0.250	-0.099	0.566	100%	100%
0107	0.317	0.256	0.087	2.579	100%	100%
0071	0.086	0.246	NA	NA	NA%	NA%
0072	0.316	0.250	NA	NA	NA%	NA%
0073	0.064	0.245	NA	NA	NA%	NA%
0074	0.188	0.244	NA	NA	NA%	NA%
0075	0.390	0.251	NA	NA	NA%	NA%
0129	0.189	0.248	NA	NA	NA%	NA%
0130	0.231	0.247	NA	NA	NA%	NA%
0131	0.230	0.251	NA	NA	NA%	NA%
0132	0.271	0.249	NA	NA	NA%	NA%
0076	0.261	0.335	NA	NA	NA%	NA%
0077	0.253	0.210	NA	NA	NA%	NA%
0078	0.256	0.307	NA	NA	NA%	NA%
0079	0.256	0.222	NA	NA	NA%	NA%
0080	0.256	0.223	NA	NA	NA%	NA%
0081	0.259	0.240	NA	NA	NA%	NA%
0082	0.256	0.221	NA	NA	NA%	NA%
0083	0.256	0.221	NA	NA	NA%	NA%
0093	0.275	0.247	NA	NA	NA%	NA%
0096	0.298	0.249	NA	NA	NA%	NA%
0110	0.250	0.201	NA	NA	NA%	NA%
0111	0.255	0.256	NA	NA	NA%	NA%
0112	0.260	0.297	NA	NA	NA%	NA%
0113	0.249	0.273	NA	NA	NA%	NA%
0114	0.239	0.316	NA	NA	NA%	NA%
0115	0.256	0.254	NA	NA	NA%	NA%
0116	0.258	0.228	NA	NA	NA%	NA%
0117	0.251	0.282	NA	NA	NA%	NA%
0118	0.281	0.258	NA	NA	NA%	NA%
0119	0.255	0.218	NA	NA	NA%	NA%
0120	0.272	0.170	NA	NA	NA%	NA%
0121	0.272	0.170	NA	NA	NA%	NA%
0122	0.279	0.171	NA	NA	NA%	NA%
0124	0.249	0.234	NA	NA	NA%	NA%
0126	0.245	0.238	NA	NA	NA%	NA%
0127	0.247	0.221	NA	NA	NA%	NA%

Table 5: Model performance and validation statistics

Model	Index RMSE	Catch RMSE	Mohn's ρ	MASE	Coverage D/D_{MSY}	Coverage U/U_{MSY}
0128	0.268	0.221	NA	NA	NA%	NA%
0133	0.260	0.252	NA	NA	NA%	NA%
0134	0.259	0.250	NA	NA	NA%	NA%

Notes:

- Index RMSE: Root mean squared error for fit to standardized CPUE index (lower is better)

- Catch RMSE: Root mean squared error for fit to catch data (lower is better)

- Mohn's ρ: Retrospective bias statistic (values close to 0 indicate low bias)

- MASE: Mean Absolute Scaled Error from hindcast cross-validation (< 1 indicates model outperforms naïve predictor)

- Coverage D/D_{MSY}: Percentage of retrospective peels where relative depletion estimates fall within full model credible intervals

- Coverage U/U_{MSY}: Percentage of retrospective peels where relative exploitation rate estimates fall within full model credible intervals

Table 6: Posterior estimates (median and 95% credible intervals) of the leading model parameters for models in the final ensemble.

Model	$\log(K)$	R_{Max}	n	x_0	q_{eff}	ρ	σ_{qdev}	σ_P	$\sigma_{O,add}$	ν_{catch}	σ_F
0100	12.786	0.396	1.353	1.000	2.227	0.993	1.495	0.058	0.084	3.797	—
	(12.323,	(0.155,	(0.500,	(0.948,	(0.602,	(0.980,	(1.013,	(0.034,	(0.029,	(2.119,	
	14.046)	0.866)	3.544)	1.053)	8.505)	0.997)	2.211)	0.094)	0.154)	7.667)	
0102	12.631	0.543	1.090	1.000	2.502	0.993	1.478	0.069	0.184	3.664	—
	(12.142,	(0.185,	(0.446,	(0.952,	(0.513,	(0.979,	(1.016,	(0.038,	(0.100,	(1.967,	
	14.378)	1.168)	3.042)	1.051)	9.714)	0.997)	2.210)	0.115)	0.291)	7.697)	
0105	12.751	0.322	2.913	1.002	2.356	0.993	1.547	0.059	0.083	3.747	—
	(12.198,	(0.121,	(1.172,	(0.950,	(0.596,	(0.979,	(1.041,	(0.036,	(0.029,	(2.000,	
	14.209)	0.722)	8.168)	1.045)	8.655)	0.997)	2.191)	0.101)	0.160)	7.582)	
0107	12.571	0.438	2.351	1.000	2.673	0.993	1.503	0.067	0.184	3.746	—
	(12.015,	(0.153,	(0.899,	(0.954,	(0.565,	(0.980,	(1.006,	(0.038,	(0.103,	(1.988,	
	14.428)	1.008)	6.335)	1.052)	9.271)	0.997)	2.147)	0.112)	0.291)	7.349)	

Table 7: Key sources of uncertainty

Type	Rationale	Uncertainty	Impact	Confidence
Data				
CPUE	Best available long-term indices (DWFN & New Zealand recreational)	DWFN index is a composite, multi-flag index with limited gear covariates. New Zealand index is spatially restricted	Changes in the fleet composition and or fishing location in the DWFN index can be interpreted by the model as population level changes. For the New Zealand index, if the stock distribution has shifted this may impact representativeness	Medium
Catch	Best available information	Situationally targeted species so catch reporting may be inconsistent	Reported catches are highly influential to estimated population scale	High
Model				
Bayesian surplus production model	Parsimonious and robust model	Over simplifies population and spatial dynamics	Unknown	Medium
Spatial assumptions				
No spatial structure	Little tagging data to understand structure	Unclear	Potentially important not quantified impact unknown	Low

Type	Rationale	Uncertainty	Impact	Confidence
Key parameter uncertainty				
Productivity (R_{MAX})	Estimated from informative prior	Uncertainty in key biological processes	Wide prior contributes to high uncertainty within model runs	High
Population scale ($\log(K)$ & q_{eff})	Estimated from informative prior	Scale prior dependent on maximum observed catch being representative of MSY	Broad prior contributes to high uncertainty within model runs	High
Shape n	Estimated from informative prior	Alternative priors used to capture shape of the production function	Uncertainty in where MSY occurs as a fraction of the unfished condition	Medium
Structural uncertainties				
Catch observation error	Fixed $\sigma_C = 0.2$	Choice of σ_C and observation model impacts estimated removals	Removal estimates may not exactly match observations	Medium
Estimation uncertainty				
MCMC	Full Bayesian estimation integrating uncertainty over key parameters	Estimated	Basis for model ensemble	High
Other sources of uncertainty				

Type	Rationale	Uncertainty	Impact	Confidence
Stock structure	Genetic sampling of catch in the north Pacific indicates presence of SWPO fish	Not considered	Actual population removals may be under-counted impacting scale and stock status	Low

Table 8: SWPO Striped Marlin stock status summary

Summary

Year of assessment: 2025	**Depletion (D_{recent})**	Likely (74%) to be below D_{MSY}		Stock is overfished
Last year of data: 2022	**Fishing mortality (F_{recent})**	Unlikely (23%) to be above F_{MSY}		Overfishing is not occurring
	Projection	F about as likely as not (33-66%) to decline further by 2027		The stock is unlikely (<33%) to be undergoing overfishing in the near term under recent average catch levels.
		D likely (>66%) to decline further by 2027		The stock is about as likely as not (33-66%) to be overfished in the near term under recent average catch levels.

Reference points	Metric	Median [2.5%-97.5% CI]	Likelihood	Recent trend / projection
Depletion	LRP (D_{MSY})	0.48 [0.26 − 0.7]		
Abundance	LRP (P_{MSY})	155,183 n [63,037 − 808,861]		
Abundance	LRP ($0.2 \times \log(K)$)	65,041 n [36,198 − 327,844]		
Catch	MSY	29,962 n [25,828 − 184,069]		
Fishing mortality	F_{MSY}	0.23 [0.08 − 0.69]		

Estimates	Metric	Median [2.5%-97.5% CI]	Likelihood	Recent trend / projection
Depletion	D_{latest}	0.38 [0.14 − 0.96]		Increasing
Depletion	D_{recent}	0.37 [0.15 − 0.94]		Increasing

Summary

Abundance	P_{latest}	121,943 n [34,067 – 1,479,253]		Increasing
Abundance	P_{recent}	117,967 n [34,199 – 1,442,511]		Increasing
Catch	C_{latest}	17,488 n [11,545 – 25,988]		Stable, decreasing
Catch	C_{recent}	20,570 n [13,357 – 28,058]		Stable, decreasing
Fishing mortality	F_{latest}	0.15 [0.01 – 0.47]		Decreasing
Fishing mortality	F_{recent}	0.17 [0.01 – 0.49]		Decreasing

Status	**Metric**	**Median [2.5%-97.5% CI]**	**Likelihood**	**Recent trend / projection**
Depletion	D_{latest}/D_{MSY}	0.81 [0.32 – 2.36]	Likely (>66%) to be below D_{MSY}	
Depletion	D_{recent}/D_{MSY}	0.77 [0.33 – 2.3]	Likely (>66%) to be below D_{MSY}	
Fishing mortality	F_{latest}/F_{MSY}	0.69 [0.05 – 1.51]	Unlikely (<33%) to be above F_{MSY}	
Fishing mortality	F_{recent}/F_{MSY}	0.77 [0.05 – 1.51]	Unlikely (<33%) to be above F_{MSY}	

Projections	**Metric**	**Median [2.5%-97.5% CI]**	**Likelihood**	**Recent trend / projection**
Depletion	D_{proj}^{2027}/D_{MSY}	1.1 [0.13 – 2.47]	About as Likely as Not (33-66%) to be below D_{MSY}	D_{proj} increasing
Fishing mortality	F_{proj}^{2027}/F_{MSY}	0.6 [0.04 – 19.3]	Unlikely (<33%) to be above F_{MSY}	F_{proj} decreasing

Notes: IPCC likelihood categories: >99% = Virtually Certain, >90% = Very Likely, >66% = Likely, 33-66% = About as Likely as Not, <33% =

Unlikely, $<10\%$ = Very Unlikely, $<1\%$ = Exceptionally Unlikely. In the absence of formal reference points, overfished and overfishing designations are based on MSY reference points. For depletion ratios, D_{recent} refers to the average over 2019-2022 and D_{latest} refers to 2022. For fishing mortality ratios, F_{recent} refers to the average over 2018-2021 and F_{latest} refers to 2021. Projections are based on recent average catch over the last 5 years: 2018-2022.

10   Figures
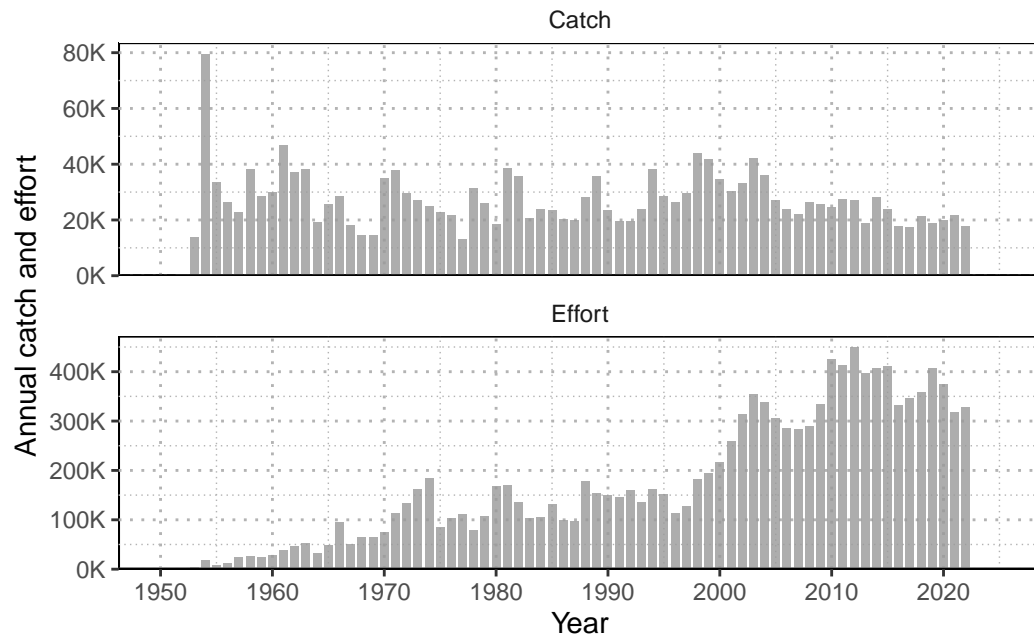
Figure 1: Annual catch (numbers; individuals) of striped marlin and nominal longline effort (hooks fished; thousands) in the Southwest Pacific Ocean (1952-2022).
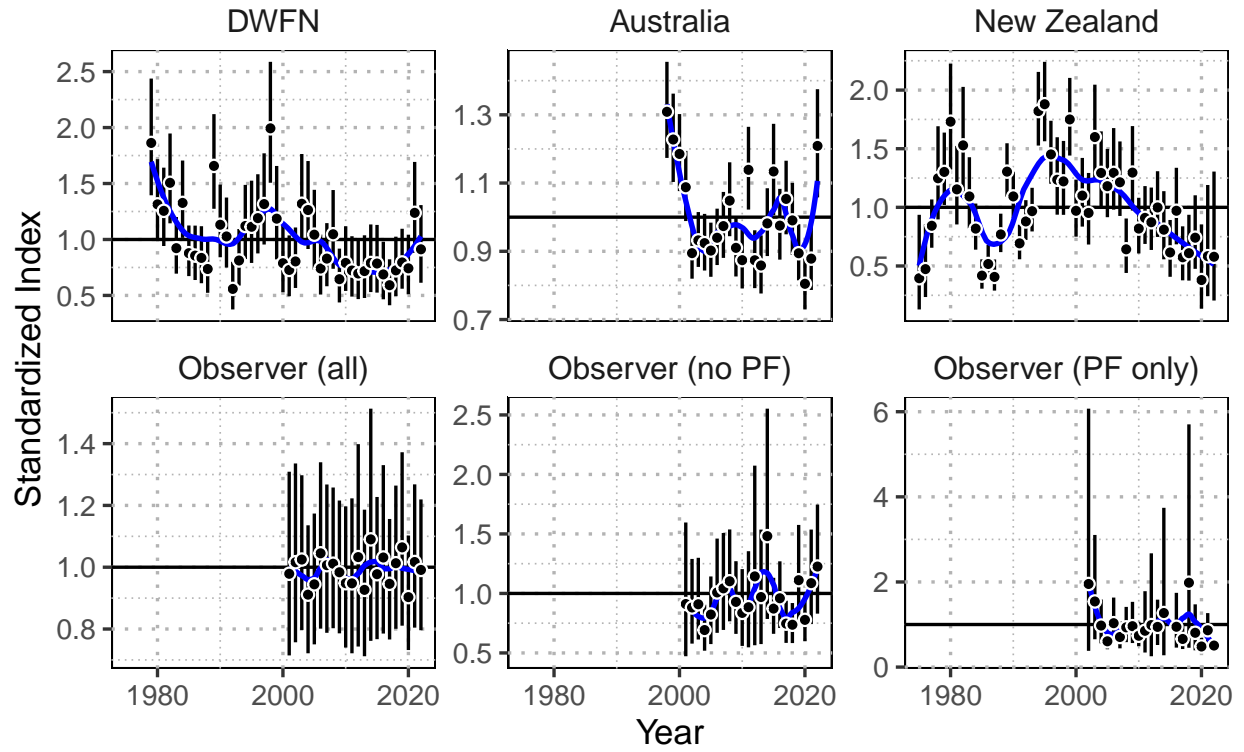
Figure 2: Standardized index (black points), and observation error (black bars) for alternative model indices. The trend in the indices is shown in blue using a loess-smooth.
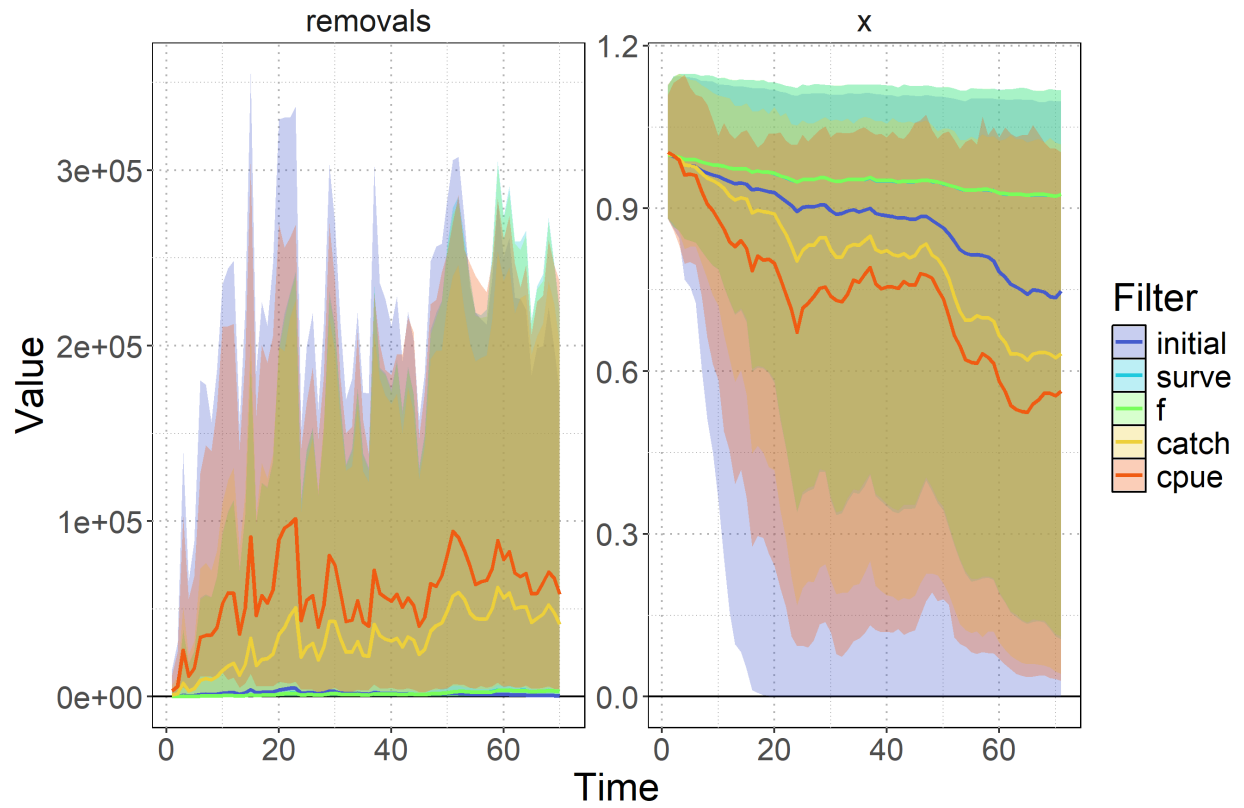
Figure 3: Prior pushforward check showing simulated removals and population depletion trajectories under different filtering scenarios (median: colored lines) with 95% quantile of trajectories (shaded polygon).
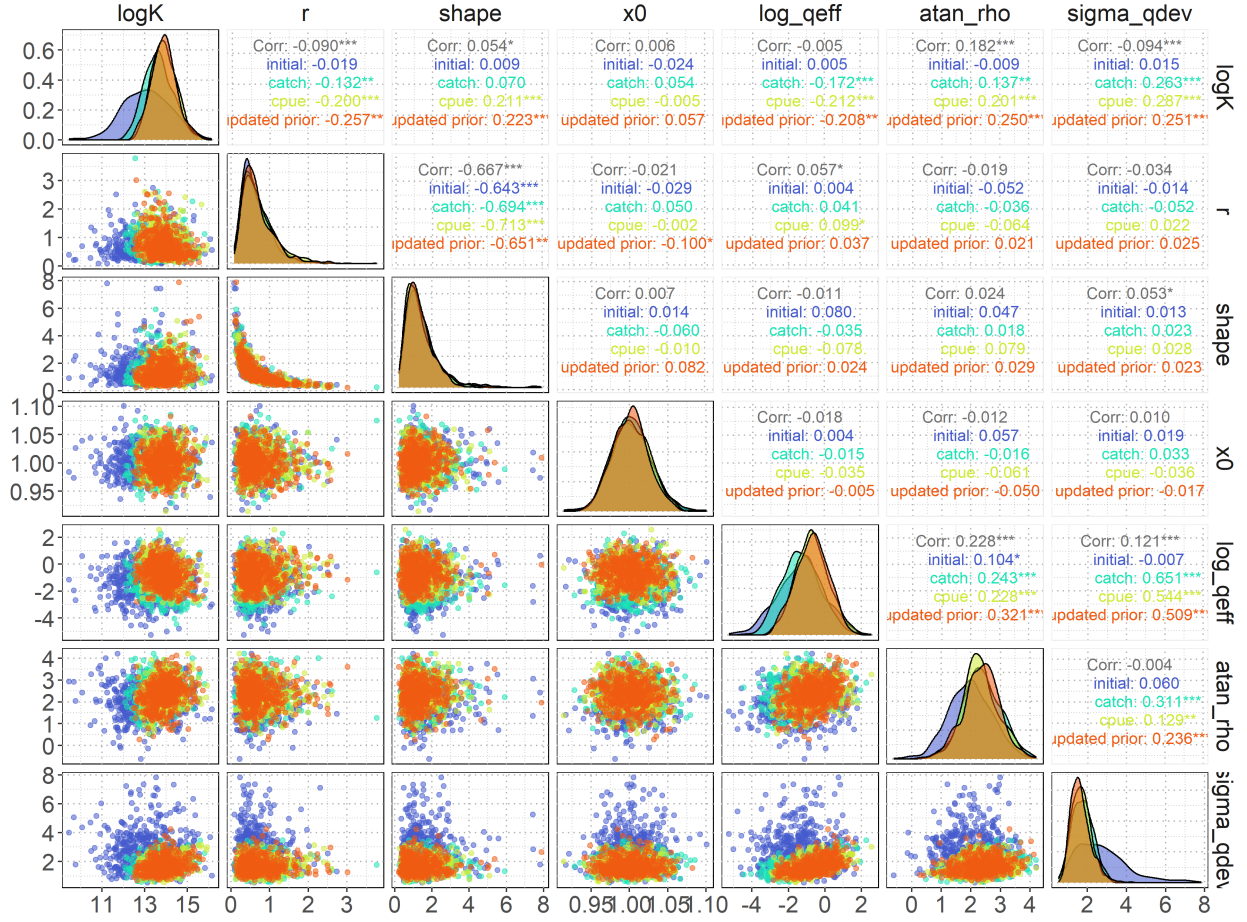
Figure 4: Prior pushforward diagnostic pairs plot showing parameter correlations and marginal distributions for different filtering scenarios. The plot compares initial priors (blue), catch-filtered priors (cyan), cpue-filtered priors (lime green), and the final updated priors following fitting to the cpue-filtered distribution (orange) across key model parameters (logK, r, shape, x0, log_qeff, atan_rho, sigma_qdev). Diagonal panels show marginal density distributions with overlaid curves for each filtering scenario. Off-diagonal scatter plots display pairwise parameter relationships colored by filtering type. Correlation coefficients in the upper triangle quantify linear relationships between adjacent distributions, with statistical significance indicated by asterisks.
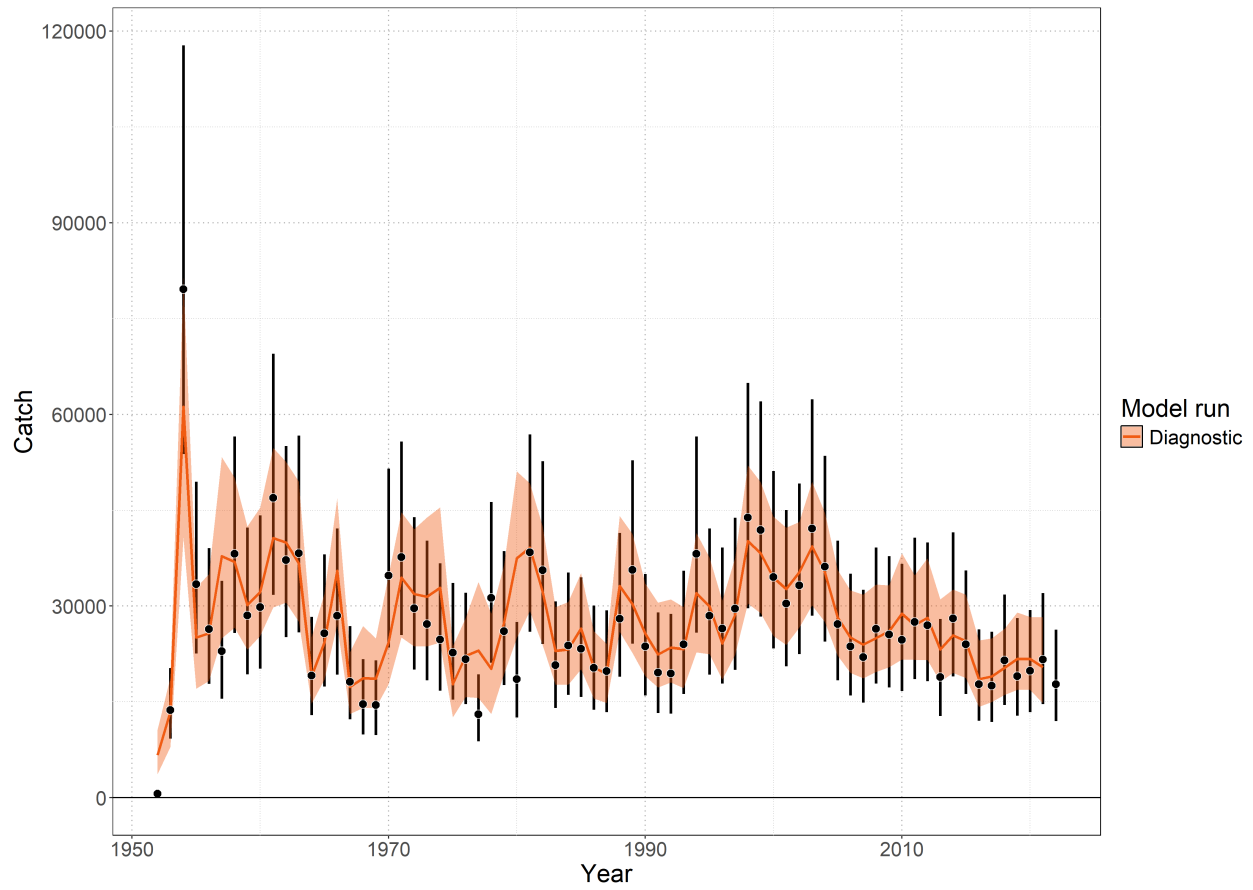
Figure 5: Model fit to catch data showing observed catch (black points with 95% error bars) and model-predicted catch (colored lines and shaded ribbons representing 95% credible intervals).
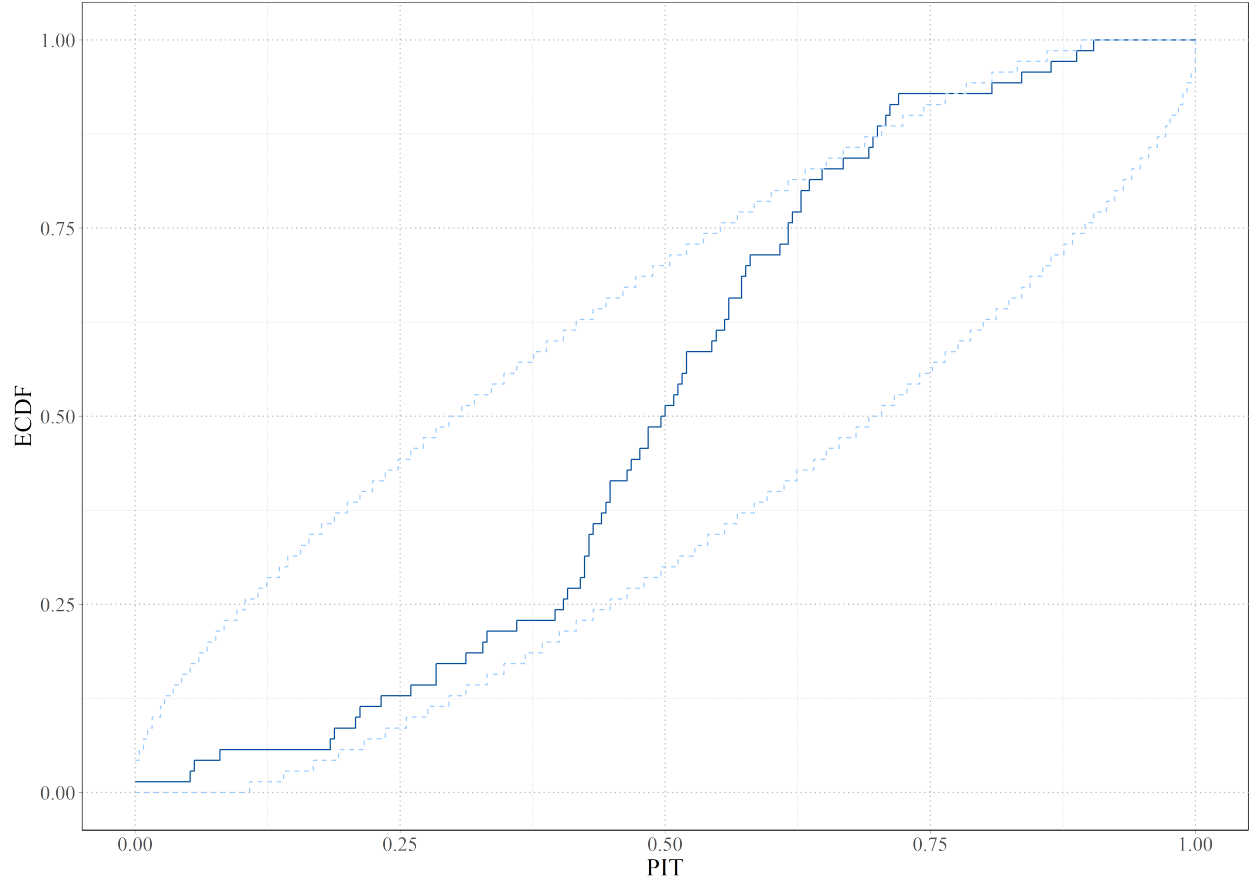
Figure 6: Posterior predictive check for catch data using probability integral transform empirical cumulative distribution function (PIT-ECDF). The solid blue line shows the empirical CDF of PIT values calculated from observed catch data relative to posterior predictive distributions. Light blue dashed lines indicate 99% simultaneous confidence bands under the assumption of correct model specification. The solid blue line should be contained within the 99% simultaneous confidence bands.
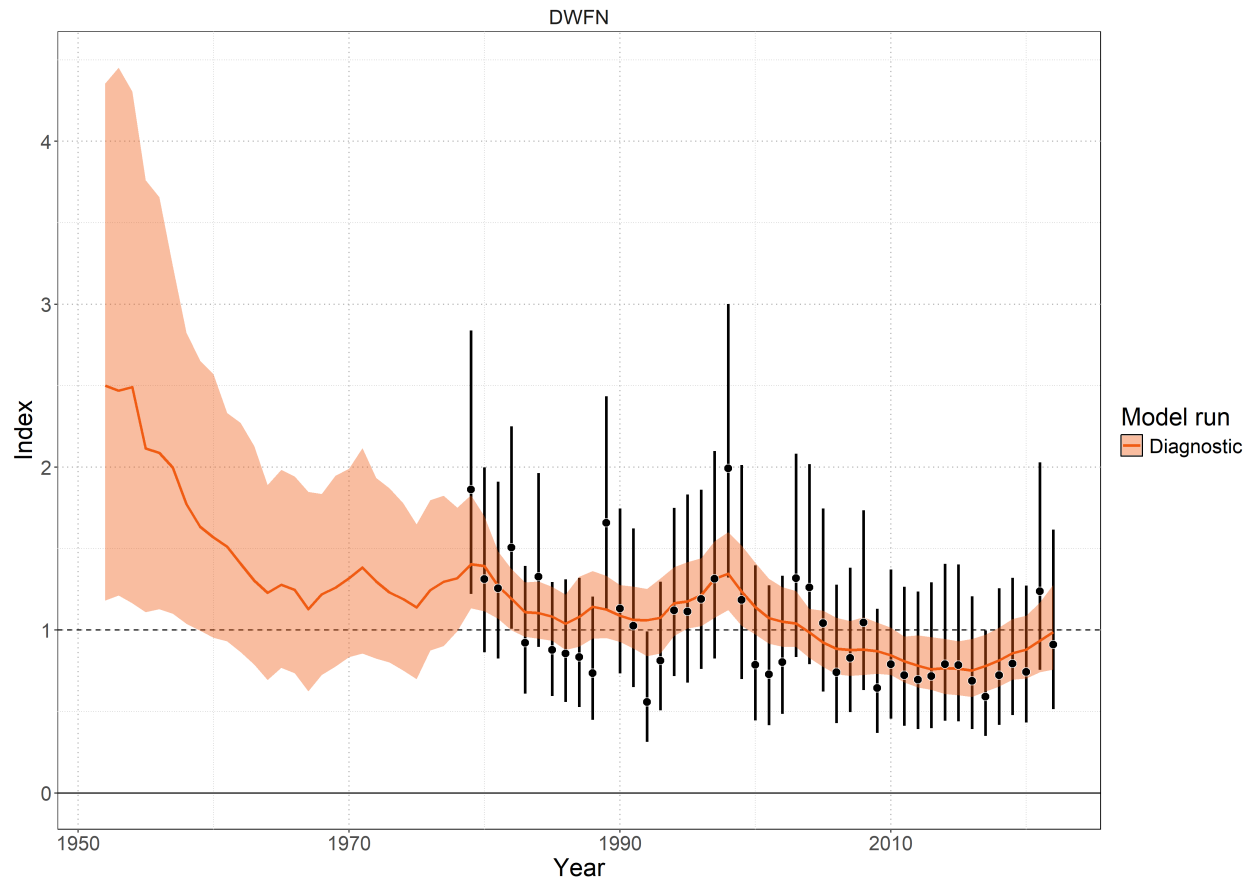
Figure 7: Model fit to the standardized index data showing observations (black points with 95% error bars) and model-predictions (colored lines and shaded ribbons representing 95% credible intervals).
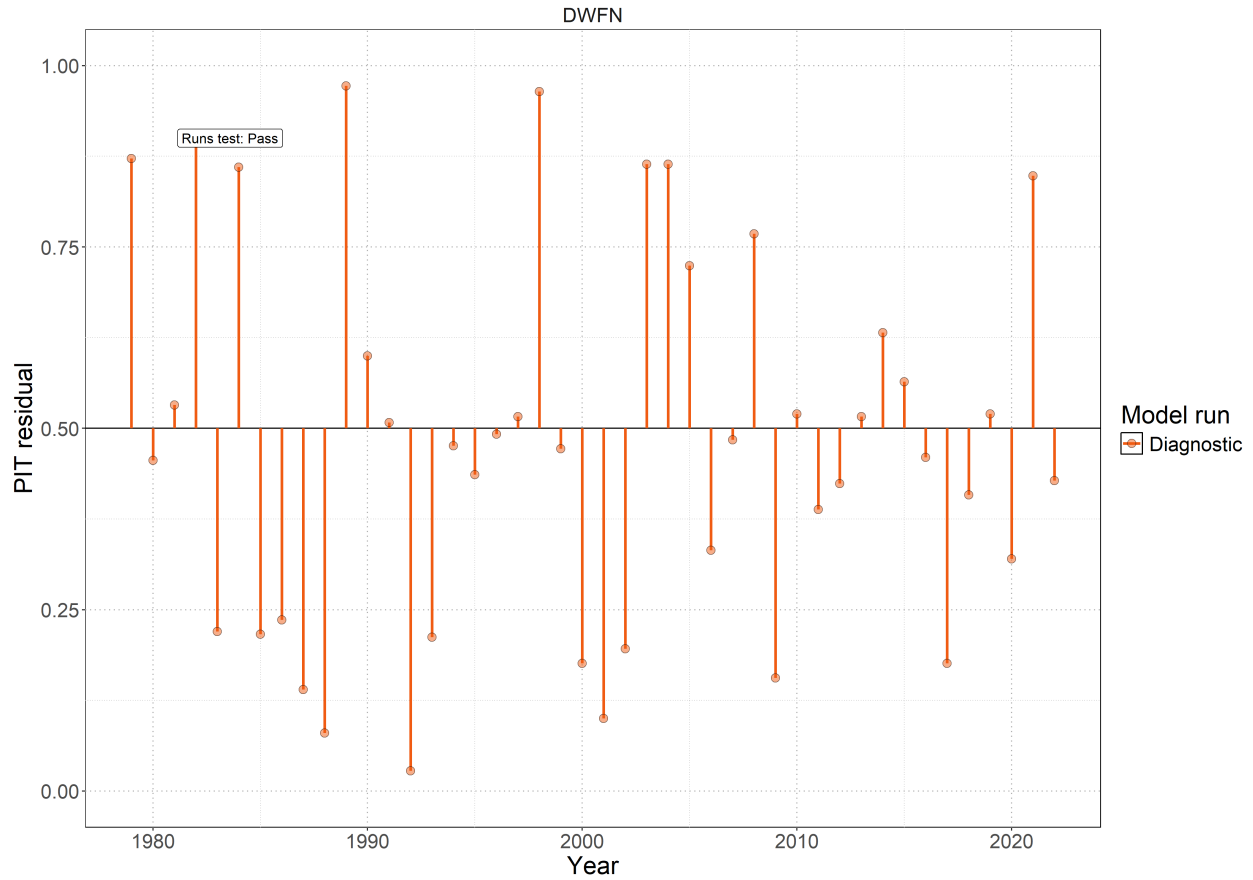
Figure 8: Probability Integral Transform (PIT) residuals for catch data from the diagnostic model over time. Orange points with vertical lines show PIT residual values for each year, where values should be uniformly distributed between 0 and 1 under correct model specification. The 'Runs test: Pass' annotation indicates that the runs test for temporal autocorrelation in the residuals passed at the 5% significance level, suggesting no systematic patterns in the residuals over time.
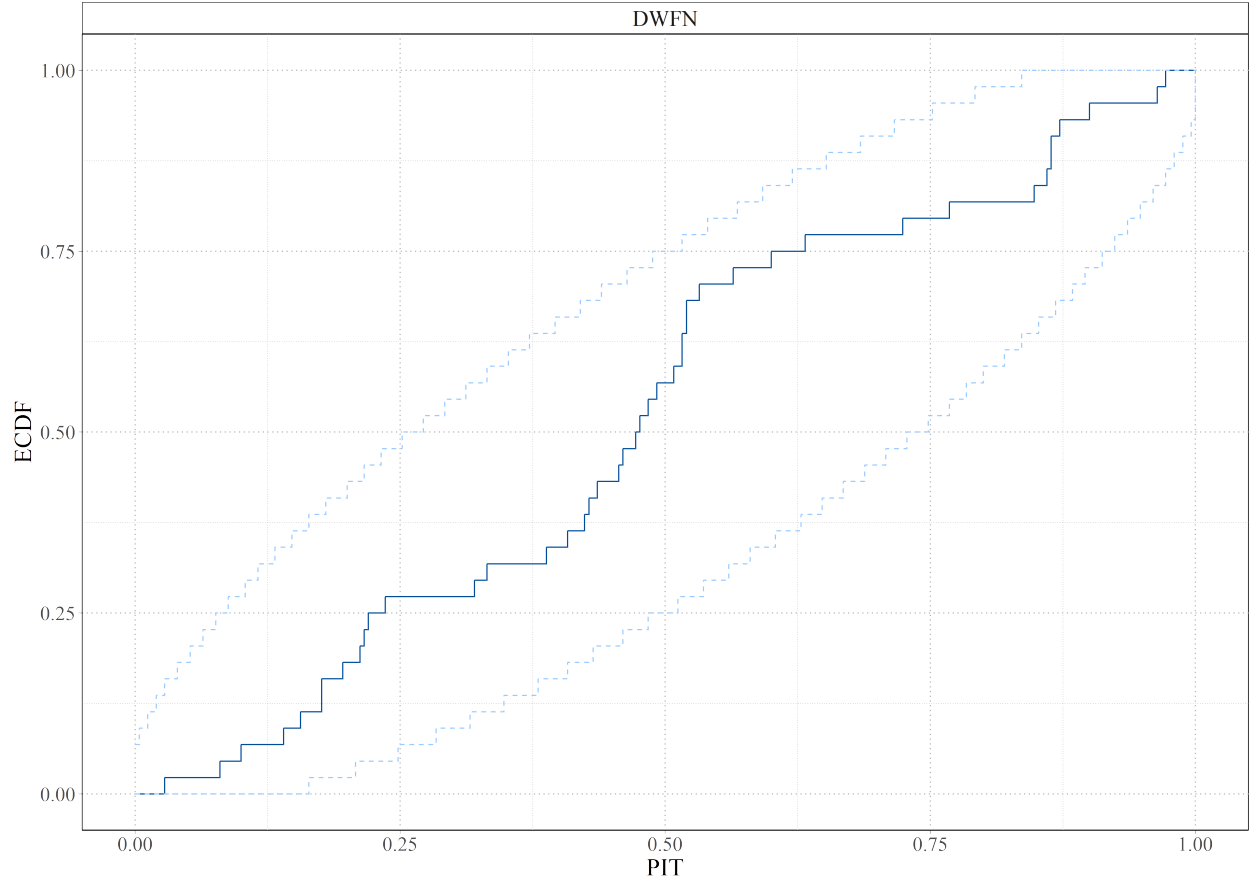
Figure 9: Posterior predictive check for index data using probability integral transform empirical cumulative distribution function (PIT-ECDF). The solid blue line shows the empirical CDF of PIT values calculated from observed catch data relative to posterior predictive distributions. Light blue dashed lines indicate 99% simultaneous confidence bands under the assumption of correct model specification. The solid blue line should be contained within the 99% simultaneous confidence bands.
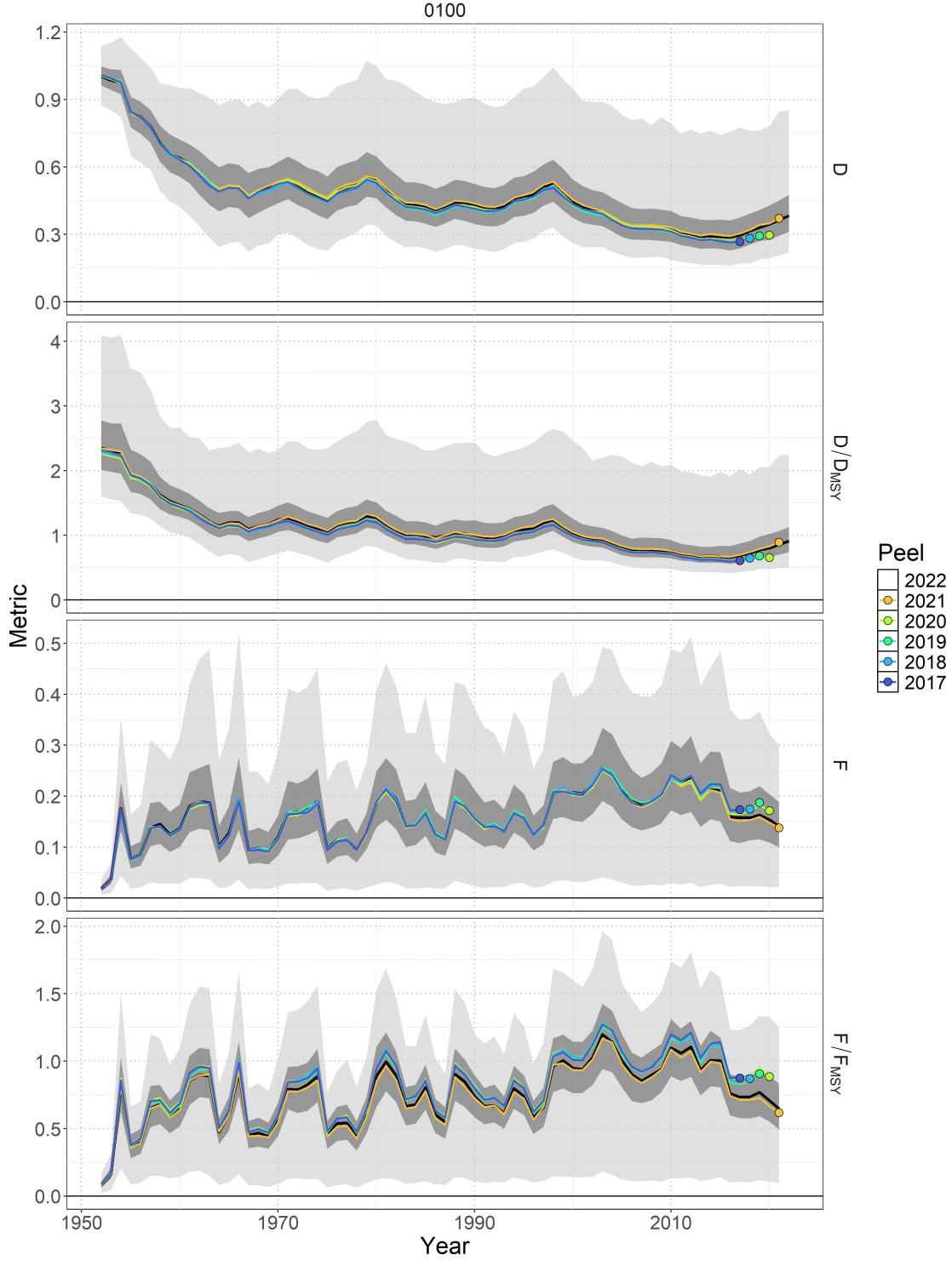
Figure 10: Retrospective analysis for the diagnostic case model (`0100`) with respect to time series of depletion D_t, depletion relative to depletion at MSY D_t/D_{MSY}, fishing mortality F_t, and fishing mortality relative to the rate of fishing mortality that produces MSY F_t/F_{MSY}. The base model with data included through 2022 (black line – median; dark shading – 50% credible interval; light shading – 95% credible interval) is shown relative to the retrospective models. Colored lines correspond to the last year of index data and the colored point indicates the estimate in the last year of the retrospective peel.
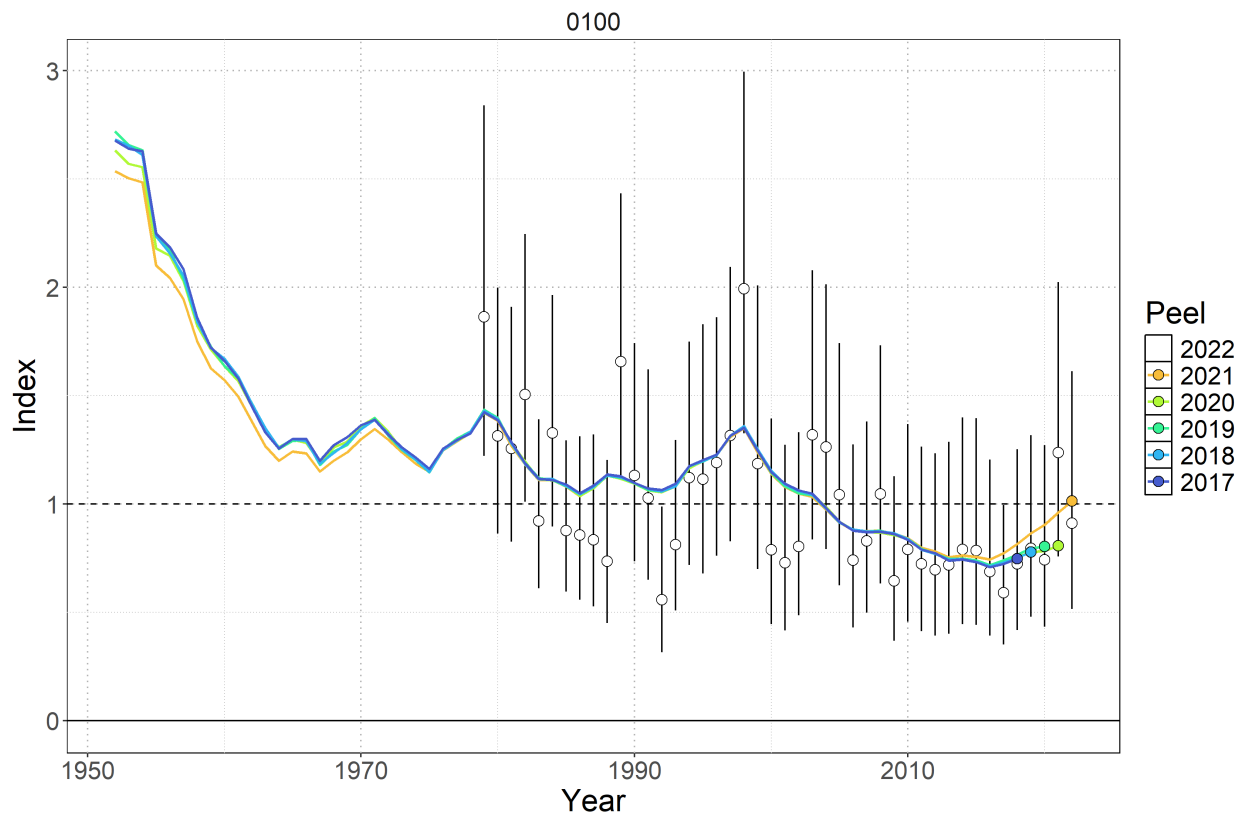
Figure 11: Standardized indices of relative abundance used in the diagnostic case model (`0100`). Open circles show observed values (standardized to mean of 1; black horizontal line) and the vertical bars indicate the observation error (95% confidence interval). One year ahead 'model-free' hindcast predictions are shown by the colored lines, where the color indicates the last year of index data seen by the model. The predicted value is shown one year-ahead with the colored point.
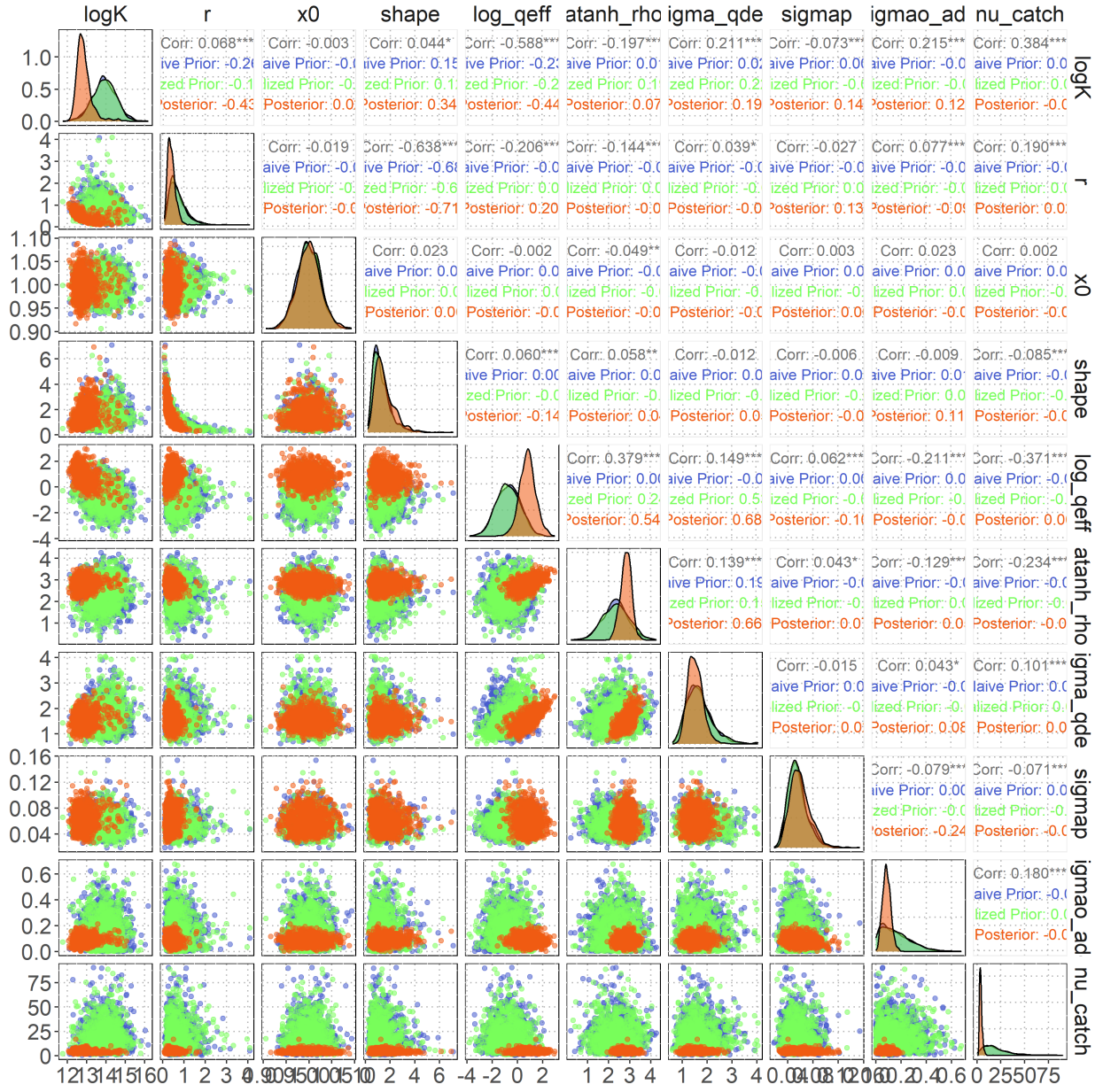
Figure 12: Prior-posterior comparison pairs plot for the diagnostic case model showing naïve prior (blue), realized prior (green), and posterior (orange) distributions across leading model parameters. Diagonal panels display marginal density distributions for each parameter. Off-diagonal scatter plots show pairwise parameter relationships colored by distribution type. Upper triangle correlation coefficients quantify linear relationships between distributions, with statistical significance indicated by asterisks. The comparison illustrates sequential parameter refinement from initial biological/expert knowledge (naïve prior) through model structure filtering (realized prior) to data-based learning (posterior).
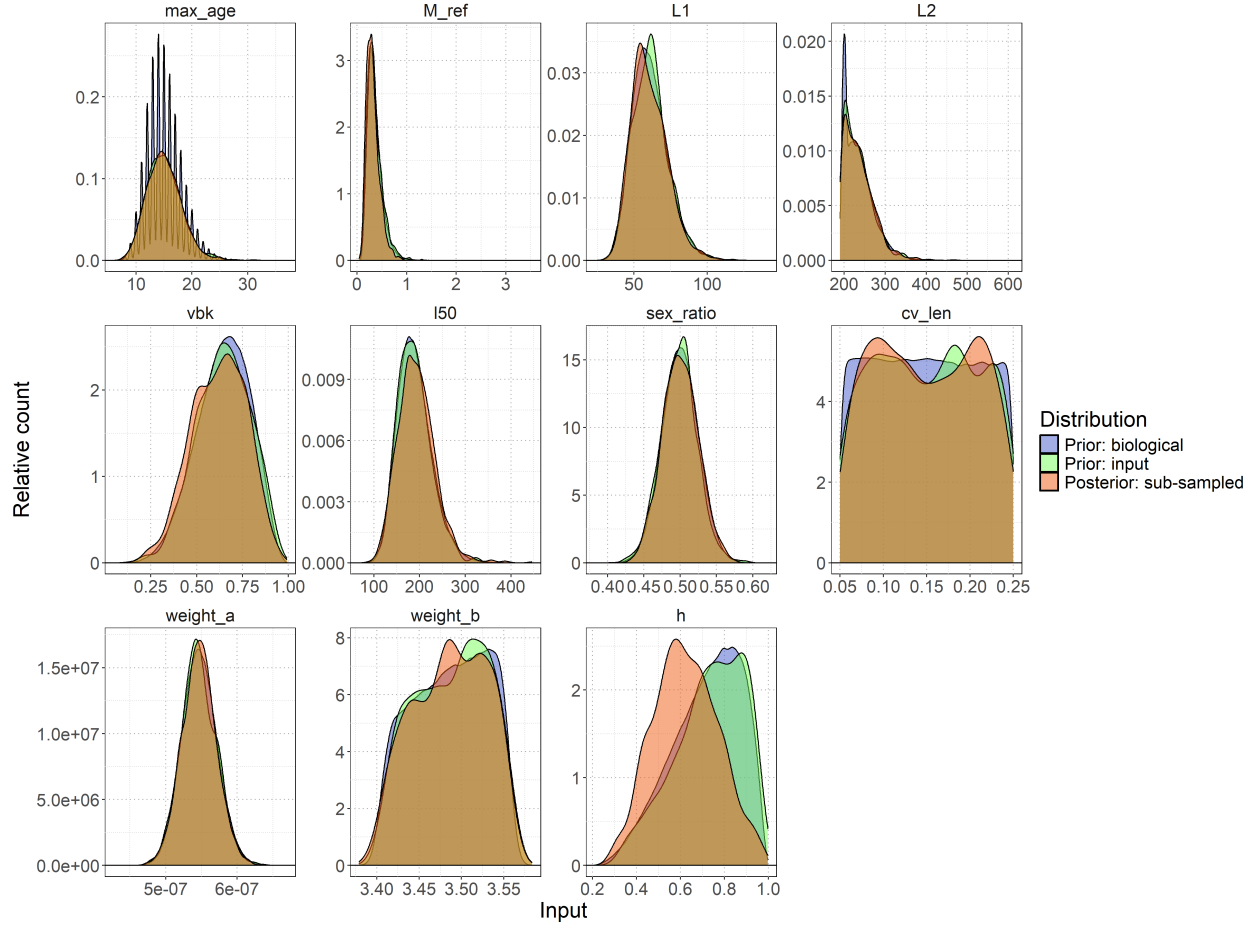
Figure 13: Distributions of biological parameters (maximum age *max_age*, natural mortality reference *M_ref*, length at birth *L1*, asymptotic length *L2*, von Bertalanffy growth coefficient *vbk*, length at 50% maturity *l50*, sex ratio at birth *sex_ratio*, coefficient of variation in length *cv_len*, weight-length relationship parameters *weight_a* and *weight_b*, and steepness *h*) used in the rmax sub-sampling process. Original biological prior distribution (blue shading), input prior distribution following the prior pushforward analysis (green shading), and final sub-sampled distribution selected to match the posterior R_{Max} distribution (orange shading).
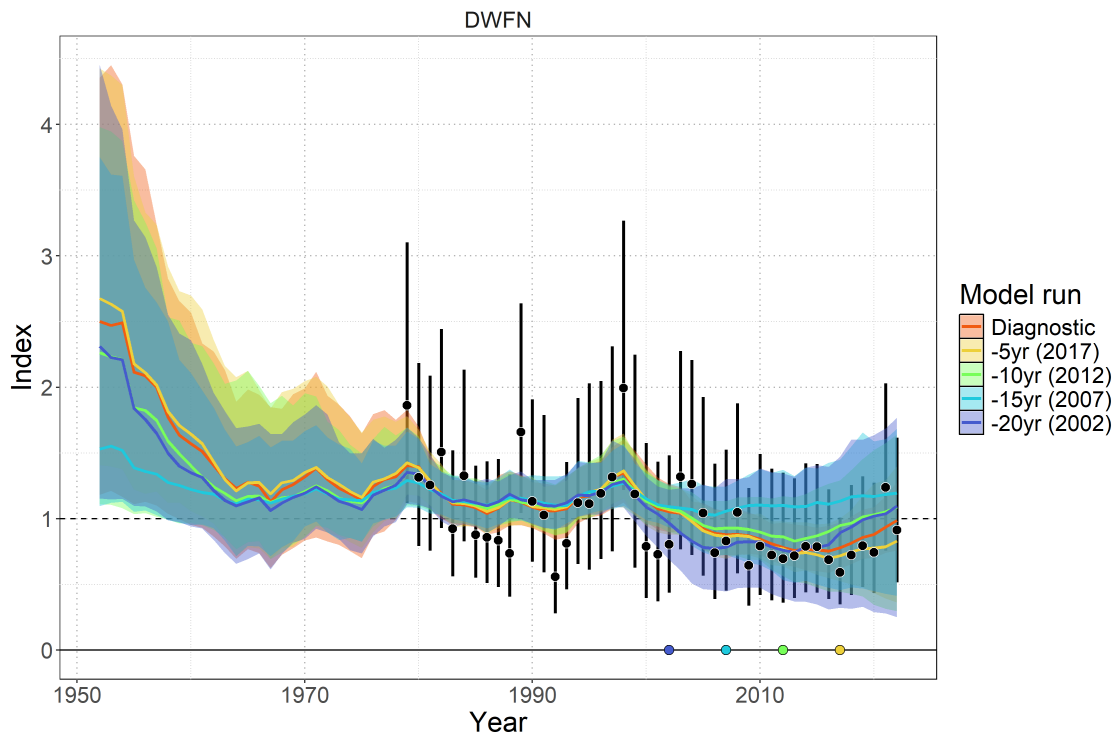
Figure 14: Model-free hindcast performance for the DWFN CPUE index. The diagnostic model fitted to the full dataset (orange line and shading) is compared against retrospective models excluding the final 5, 10, 15, and 20 years of index data. Colored lines show model predictions with 95% credible intervals (shading). Observed index values are shown as black points with error bars. Colored points at the bottom indicate the terminal year of each retrospective peel: 2017 (yellow), 2012 (green), 2007 (cyan), and 2002 (blue).

Figure 15: Model-free hindcast performance for key population dynamics parameters. Time series show posterior distributions (median and 95% credible intervals) for depletion (D), population (P), fishing mortality (F), relative depletion (D/D_{MSY}), relative fishing mortality (F/F_{MSY}), removals, process error, nominal CPUE, and catchability deviates. The diagnostic model (orange) is compared against retrospective models terminating in 2017 (yellow), 2012 (green), 2007 (cyan), and 2002 (blue). Colored points indicate the terminal year of each retrospective peel.

Figure 16: One-off sensitivity group: `Index`.Model fit to standardized index data showing observations (black points with 95% error bars) and model-predictions (colored lines and shaded ribbons representing 95% credible intervals). The diagnostic model (`0100`) is shown in orange.
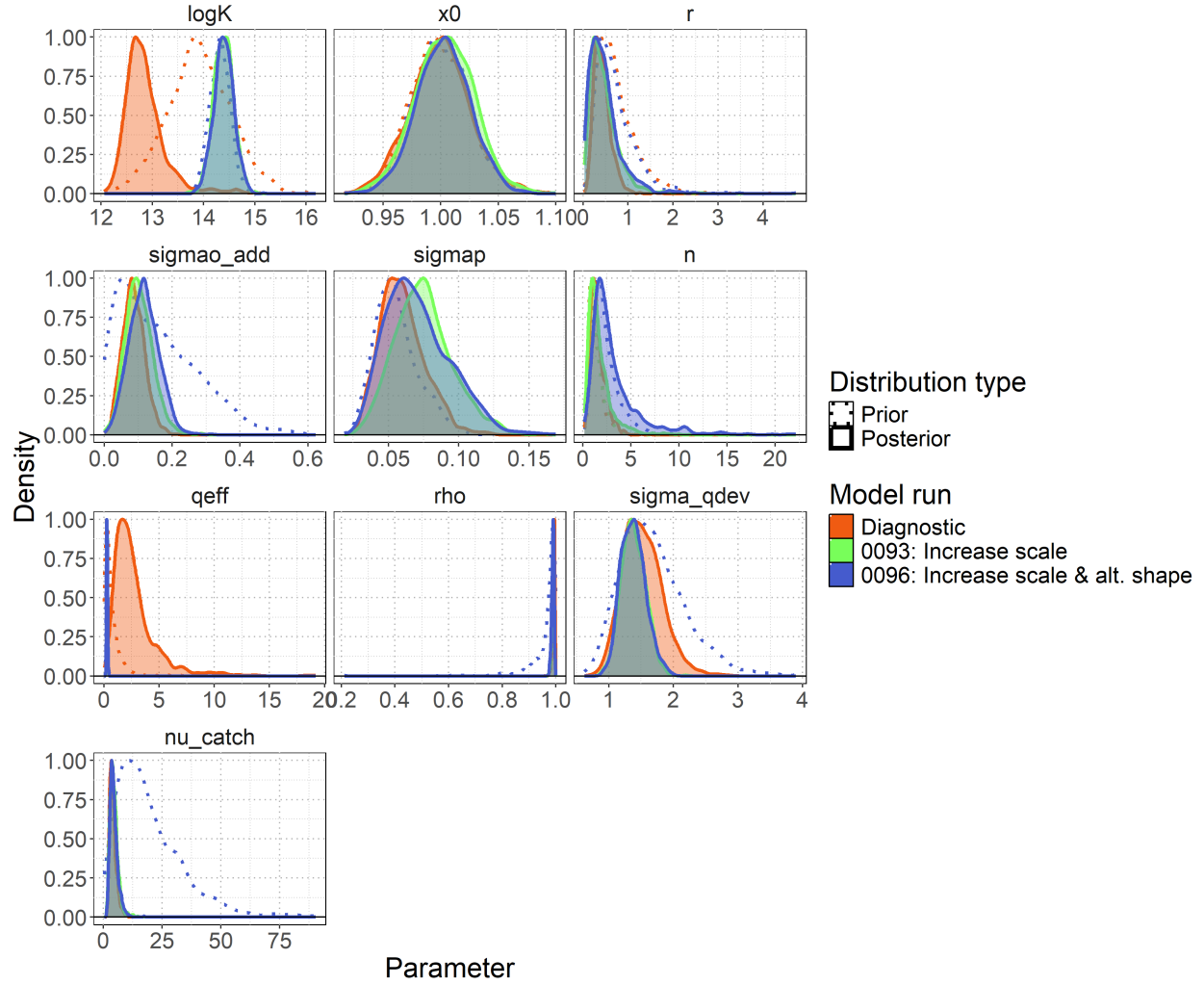
Figure 17: One-off sensitivity group: `Index` .Prior and posterior parameter distributions showing the evolution from prior (dotted line) to posterior (shaded) distributions for leading model parameters: carrying capacity (logK), intrinsic growth rate (r), observation error (sigmao_add), process error (sigmap), shape parameter, effective catchability (qeff), autocorrelation in catchability (rho), catchability deviation variability (sigma_qdev), initial depletion (x0), and catch likelihood degrees of freedom (nu_catch). The diagnostic model (`0100`) is shown in orange.

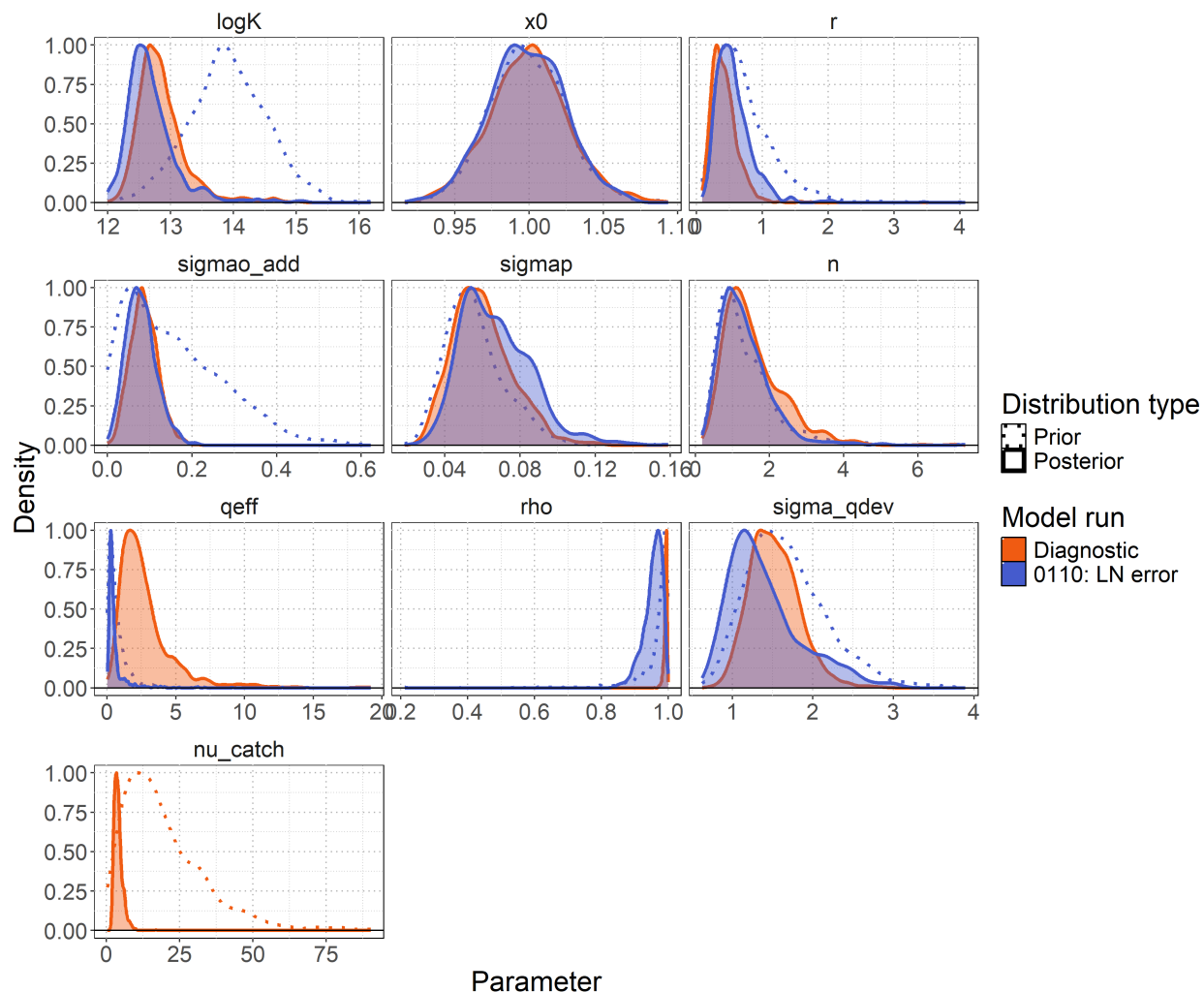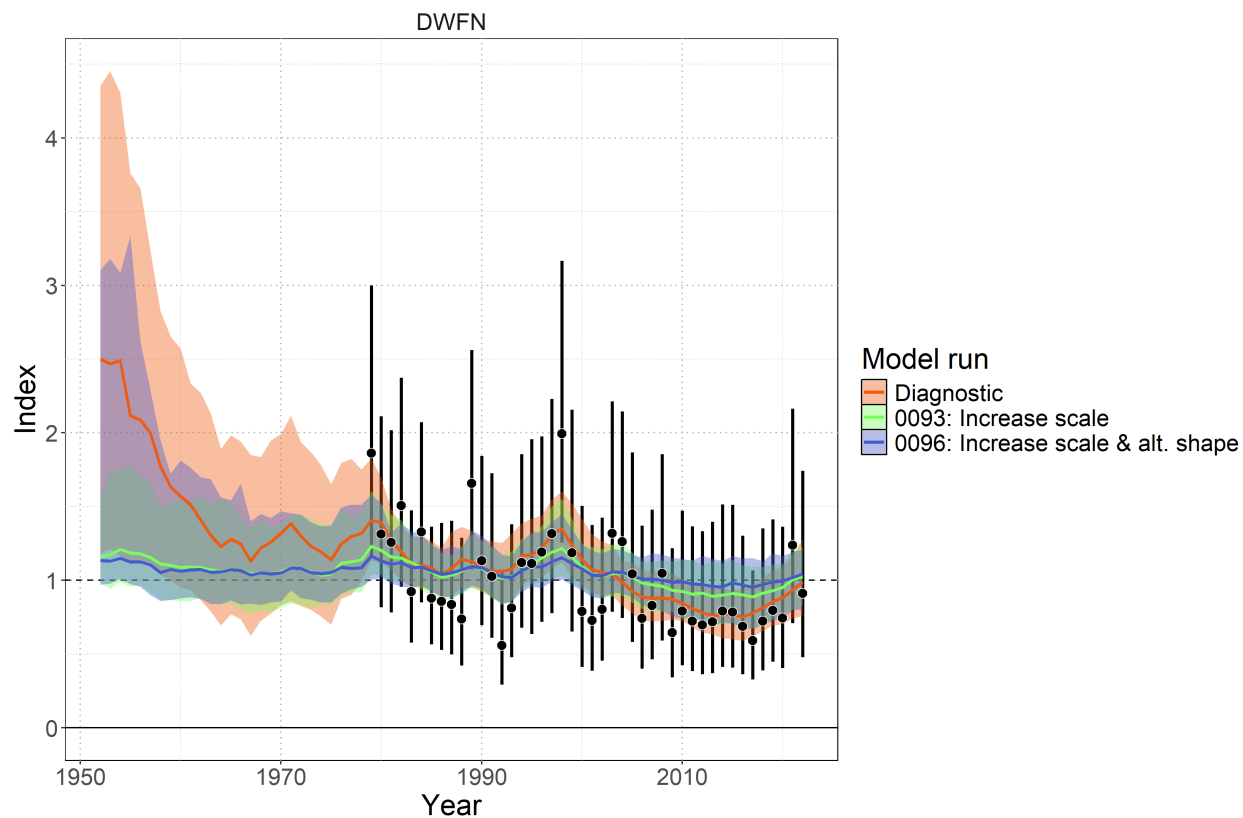Figure 18: One-off sensitivity group: `Index` . Posterior time series distributions for key derived quantities over time (line = median, shading = 95% credible interval): depletion (D), absolute population size in numbers (P), fishing mortality (F), stock status relative to MSY reference points (D/D_{MSY}, F/F_{MSY}), total removals in numbers, process error, nominal CPUE (numbers caught per 1000 hooks), and catchability deviates. The diagnostic model (`0100`) is shown in orange.

Figure 19: One-off sensitivity group: `Catch observation error` . Prior and posterior parameter distributions showing the evolution from prior (dotted line) to posterior (shaded) distributions for leading model parameters: carrying capacity (logK), intrinsic growth rate (r), observation error (sigmao_add), process error (sigmap), shape parameter, effective catchability (qeff), autocorrelation in catchability (rho), catchability deviation variability (sigma_qdev), initial depletion (x0), and catch likelihood degrees of freedom (nu_catch). The diagnostic model (`0100`) is shown in orange.
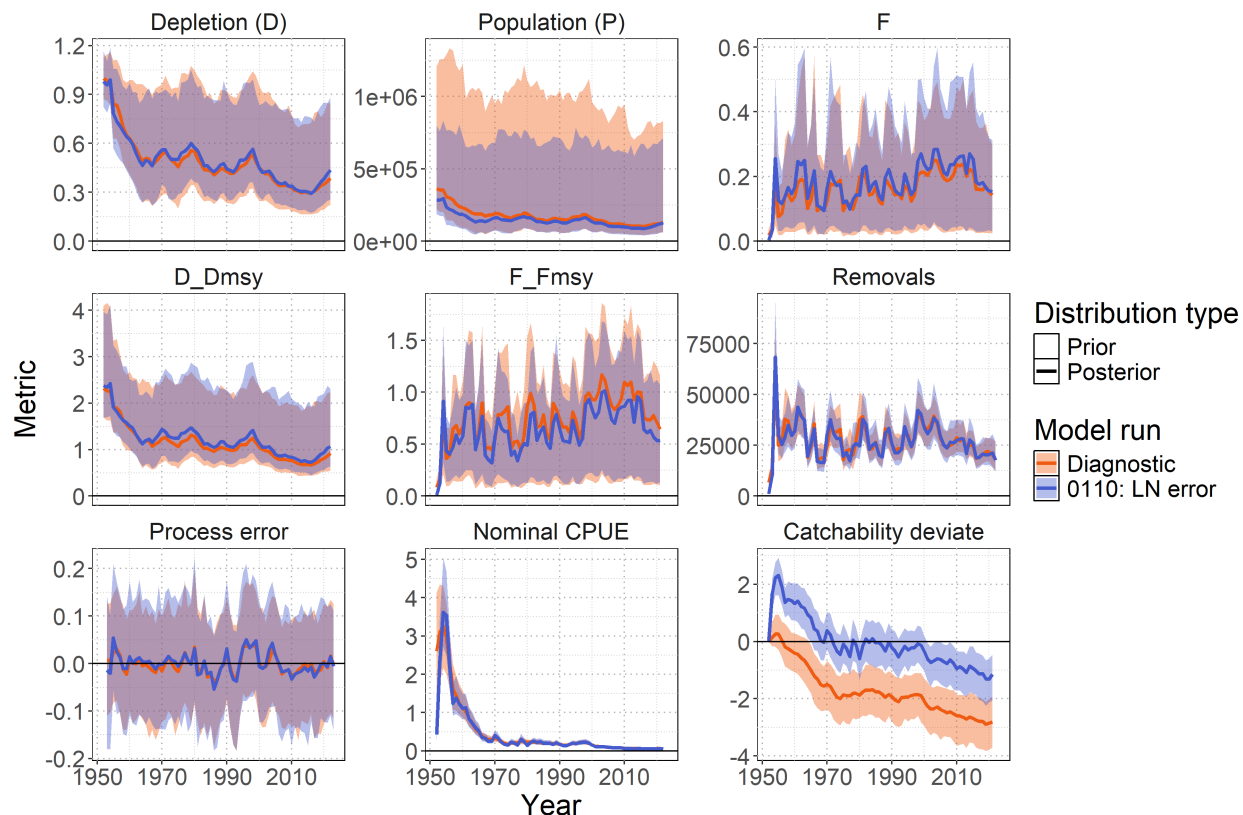
Figure 20: One-off sensitivity group: `Catch observation error`. Posterior time series distributions for key derived quantities over time (line = median, shading = 95% credible interval): depletion (D), absolute population size in numbers (P), fishing mortality (F), stock status relative to MSY reference points (D/D_{MSY}, F/F_{MSY}), total removals in numbers, process error, nominal CPUE (numbers caught per 1000 hooks), and catchability deviates. The diagnostic model (`0100`) is shown in orange.
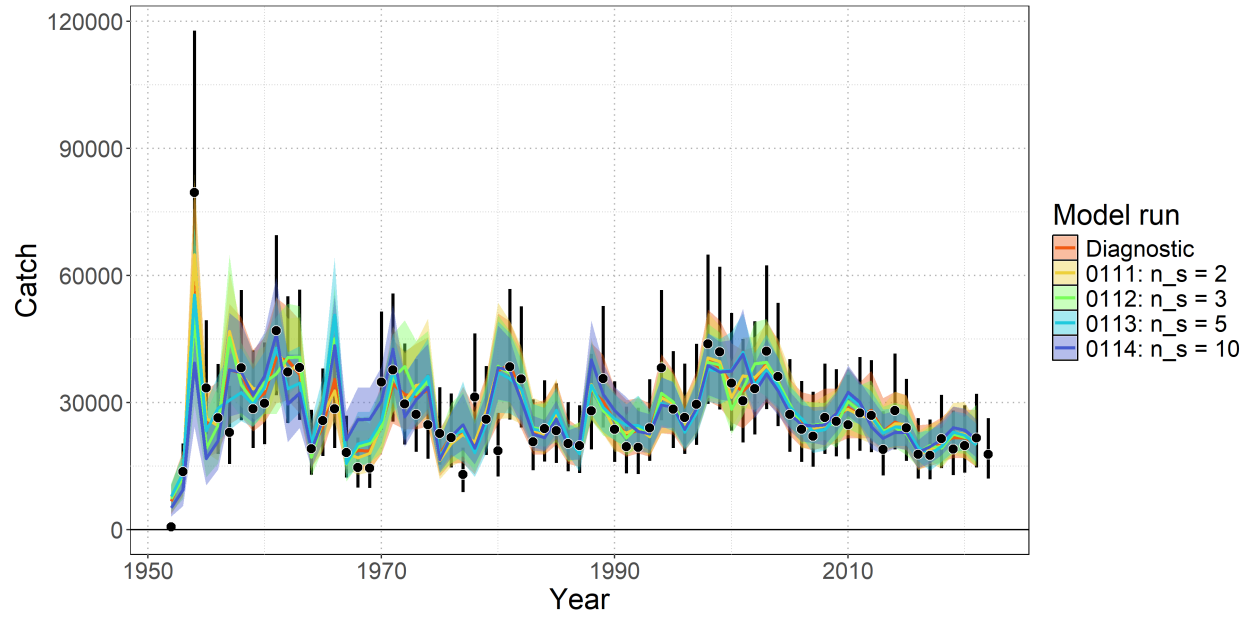
Figure 21: One-off sensitivity group: `Early data`. Prior and posterior parameter distributions showing the evolution from prior (dotted line) to posterior (shaded) distributions for leading model parameters: carrying capacity (logK), intrinsic growth rate (r), observation error (sigmao_add), process error (sigmap), shape parameter, effective catchability (qeff), autocorrelation in catchability (rho), catchability deviation variability (sigma_qdev), initial depletion (x0), and catch likelihood degrees of freedom (nu_catch). The diagnostic model (`0100`) is shown in orange.

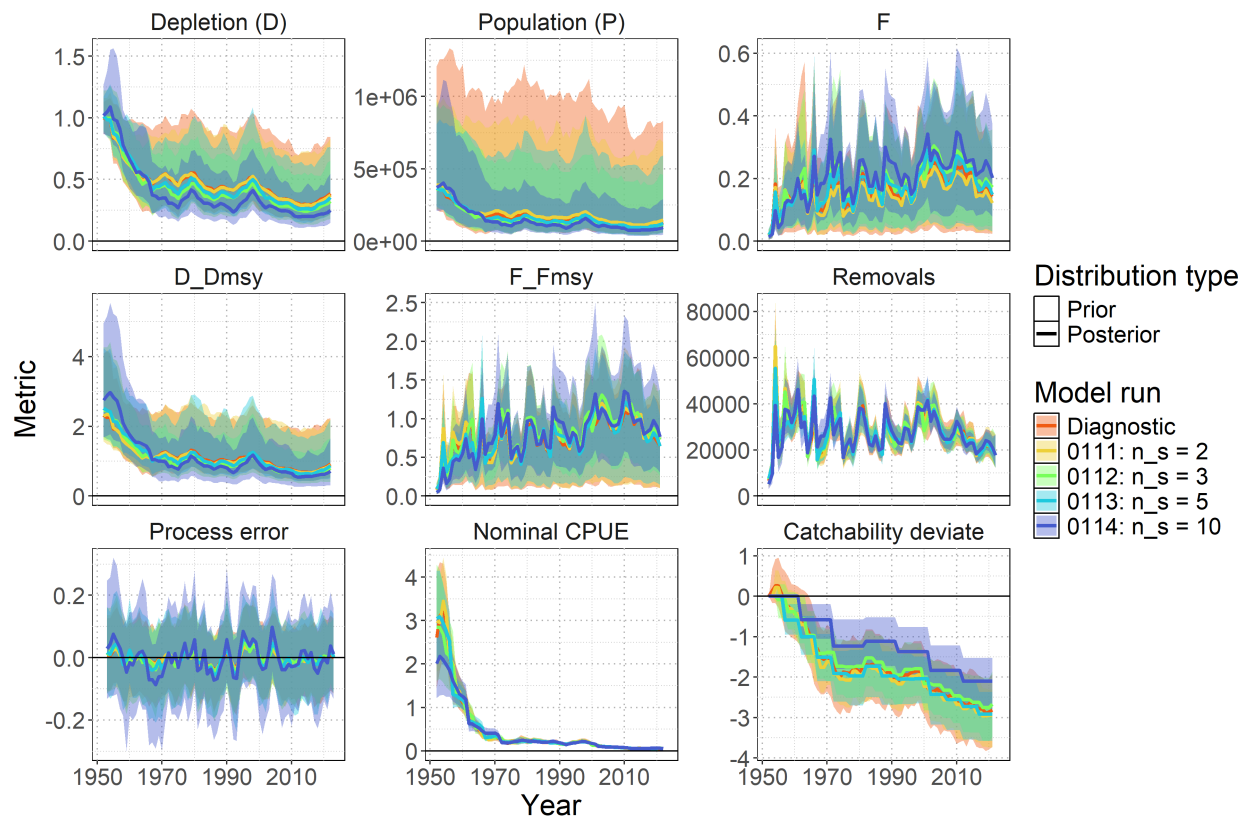Figure 22: One-off sensitivity group: `Early data`. Posterior time series distributions for key derived quantities over time (line = median, shading = 95% credible interval): depletion (D), absolute population size in numbers (P), fishing mortality (F), stock status relative to MSY reference points (D/D_{MSY}, F/F_{MSY}), total removals in numbers, process error, nominal CPUE (numbers caught per 1000 hooks), and catchability deviates. The diagnostic model (`0100`) is shown in orange.

Figure 23: One-off sensitivity group: `Scaling priors`. Prior and posterior parameter distributions showing the evolution from prior (dotted line) to posterior (shaded) distributions for leading model parameters: carrying capacity (logK), intrinsic growth rate (r), observation error (sigmao_add), process error (sigmap), shape parameter, effective catchability (qeff), autocorrelation in catchability (rho), catchability deviation variability (sigma_qdev), initial depletion (x0), and catch likelihood degrees of freedom (nu_catch). The diagnostic model (`0100`) is shown in orange.

Figure 24: One-off sensitivity group: `Scaling priors`. Model fit to standardized index data showing observations (black points with 95% error bars) and model-predictions (colored lines and shaded ribbons representing 95% credible intervals). The diagnostic model (`0100`) is shown in orange.
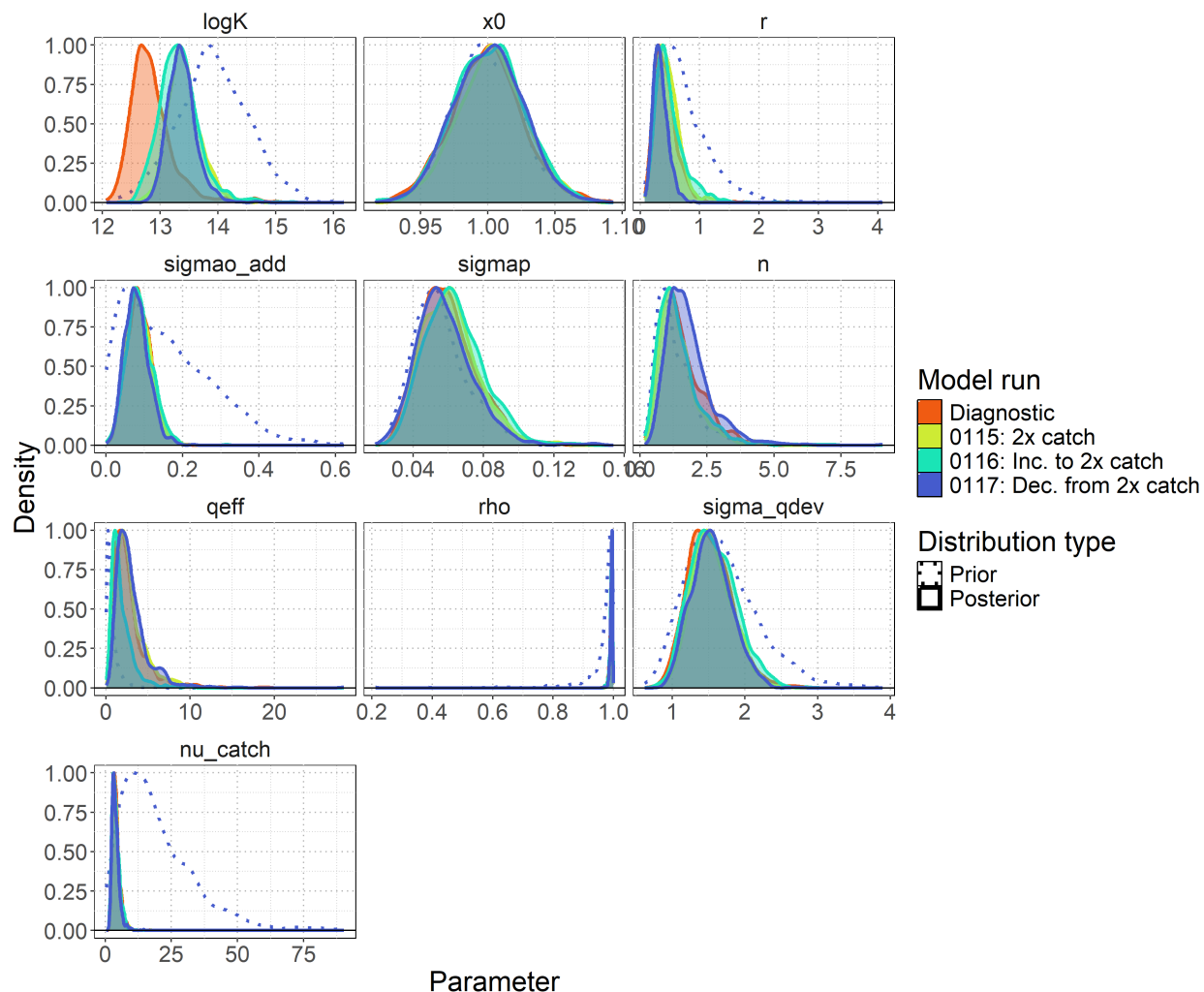
Figure 25: One-off sensitivity group: `Catch observation model` . Prior and posterior parameter distributions showing the evolution from prior (dotted line) to posterior (shaded) distributions for leading model parameters: carrying capacity (logK), intrinsic growth rate (r), observation error (sigmao_add), process error (sigmap), shape parameter, effective catchability (qeff), autocorrelation in catchability (rho), catchability deviation variability (sigma_qdev), initial depletion (x0), and catch likelihood degrees of freedom (nu_catch). The diagnostic model (`0100`) is shown in orange.
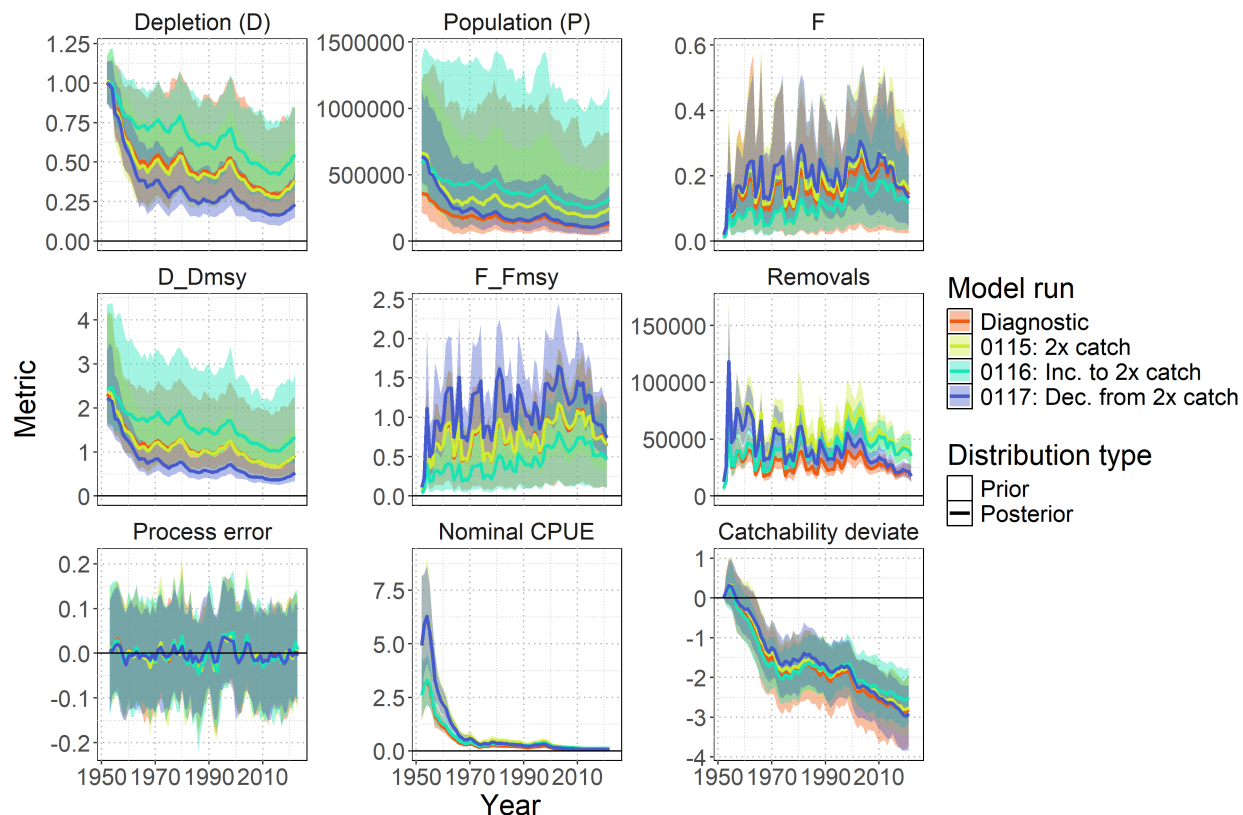
Figure 26: One-off sensitivity group: `Catch observation model`. Posterior time series distributions for key derived quantities over time (line = median, shading = 95% credible interval): depletion (D), absolute population size in numbers (P), fishing mortality (F), stock status relative to MSY reference points (D/D_{MSY}, F/F_{MSY}), total removals in numbers, process error, nominal CPUE (numbers caught per 1000 hooks), and catchability deviates. The diagnostic model (`0100`) is shown in orange.

Figure 27: One-off sensitivity group: `Temporal resolution of catchability deviates`. Model fit to catch data showing observed catch (black points with 95% error bars) and model-predicted catch (colored lines and shaded ribbons representing 95% credible intervals). The diagnostic model (`0100`) is shown in orange.
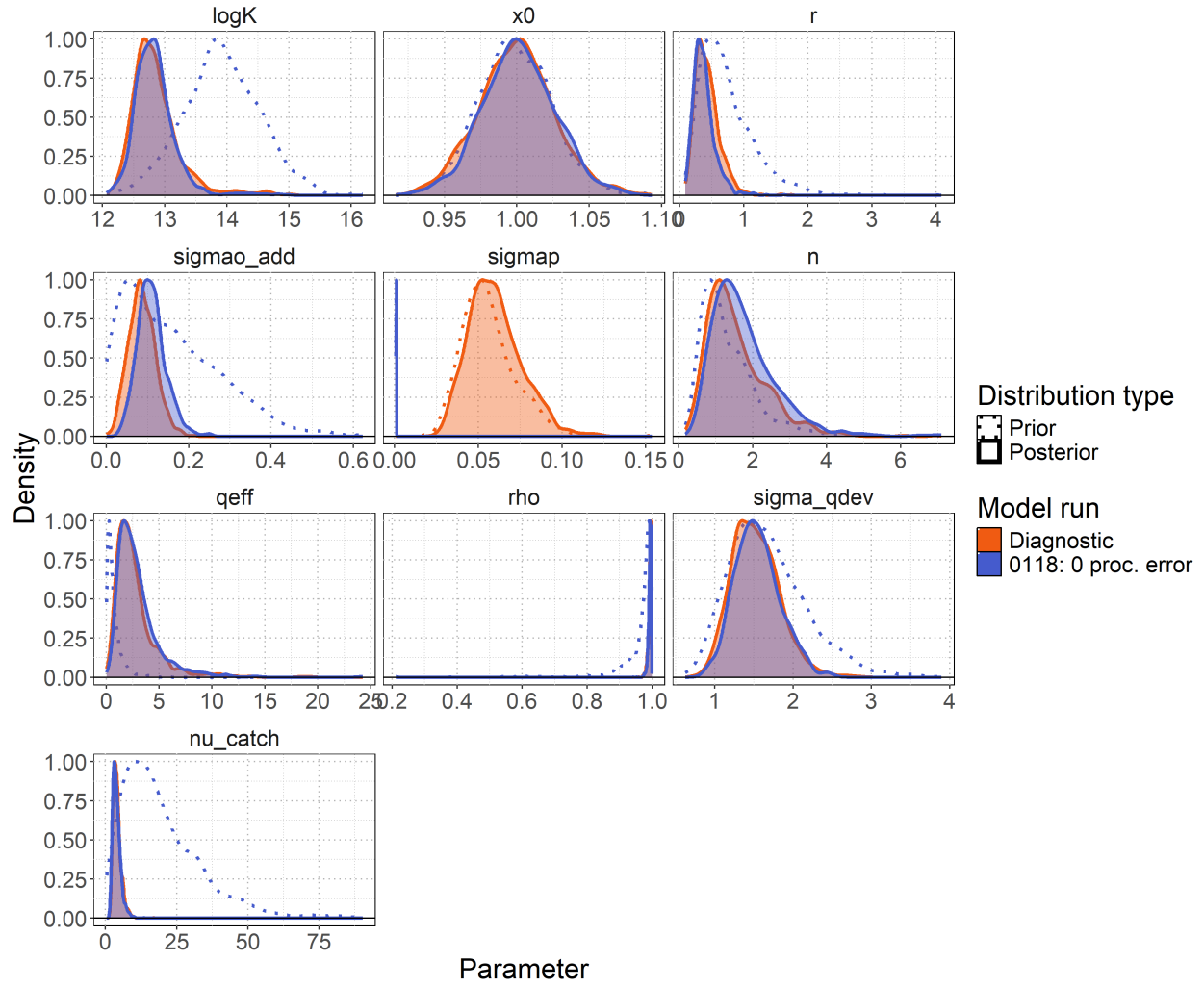
Figure 28: One-off sensitivity group: `Temporal resolution of catchability deviates` . Prior and posterior parameter distributions showing the evolution from prior (dotted line) to posterior (shaded) distributions for leading model parameters: carrying capacity (logK), intrinsic growth rate (r), observation error (sigmao_add), process error (sigmap), shape parameter, effective catchability (qeff), autocorrelation in catchability (rho), catchability deviation variability (sigma_qdev), initial depletion (x0), and catch likelihood degrees of freedom (nu_catch). The diagnostic model (`0100`) is shown in orange.
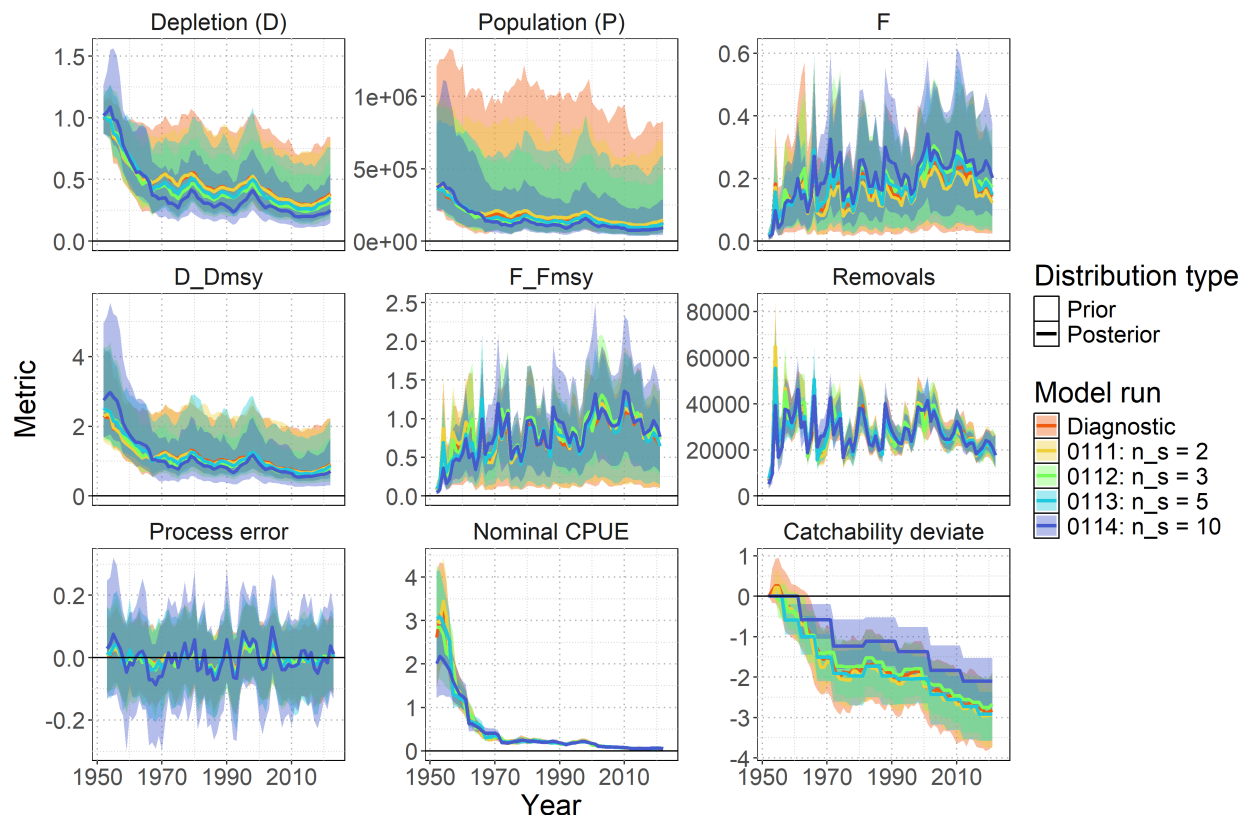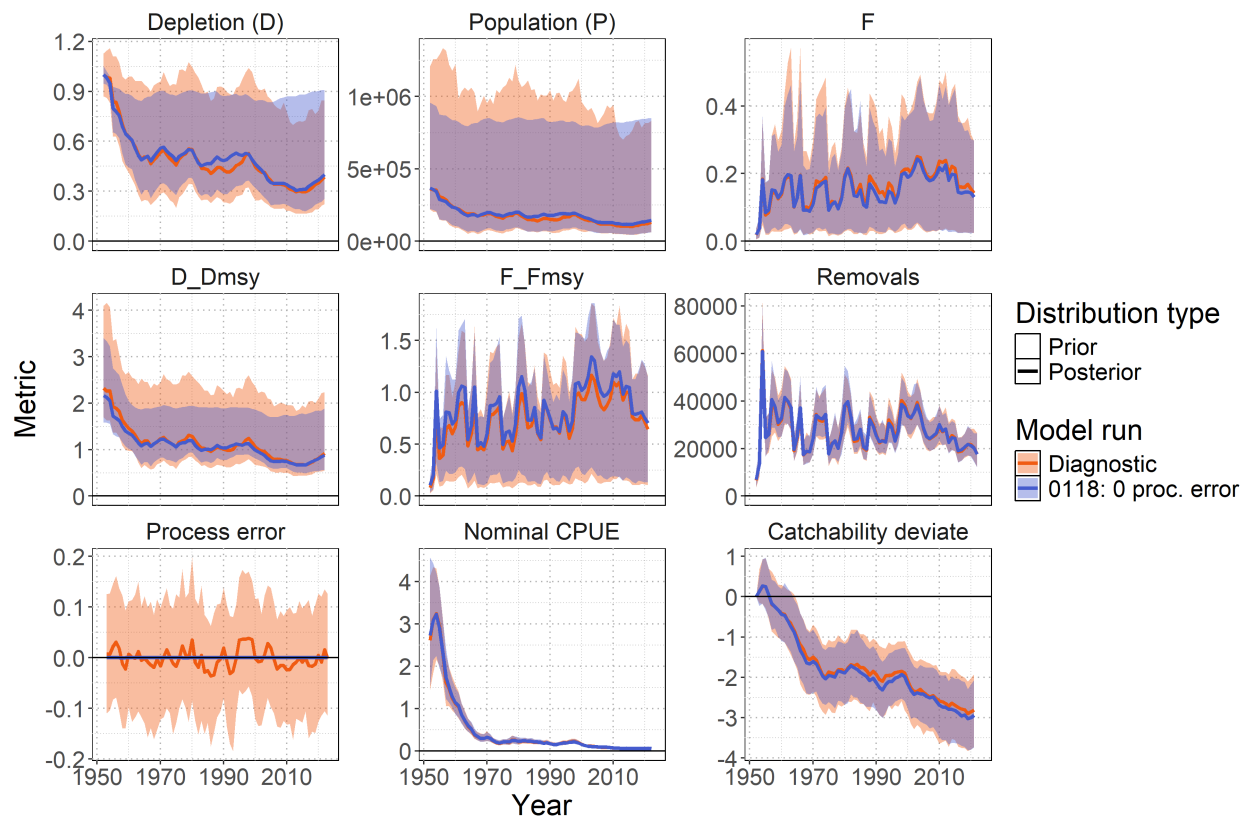
Figure 29: One-off sensitivity group: `Temporal resolution of catchability deviates`. Posterior time series distributions for key derived quantities over time (line = median, shading = 95% credible interval): depletion (D), absolute population size in numbers (P), fishing mortality (F), stock status relative to MSY reference points (D/D_{MSY}, F/F_{MSY}), total removals in numbers, process error, nominal CPUE (numbers caught per 1000 hooks), and catchability deviates. The diagnostic model (`0100`) is shown in orange.
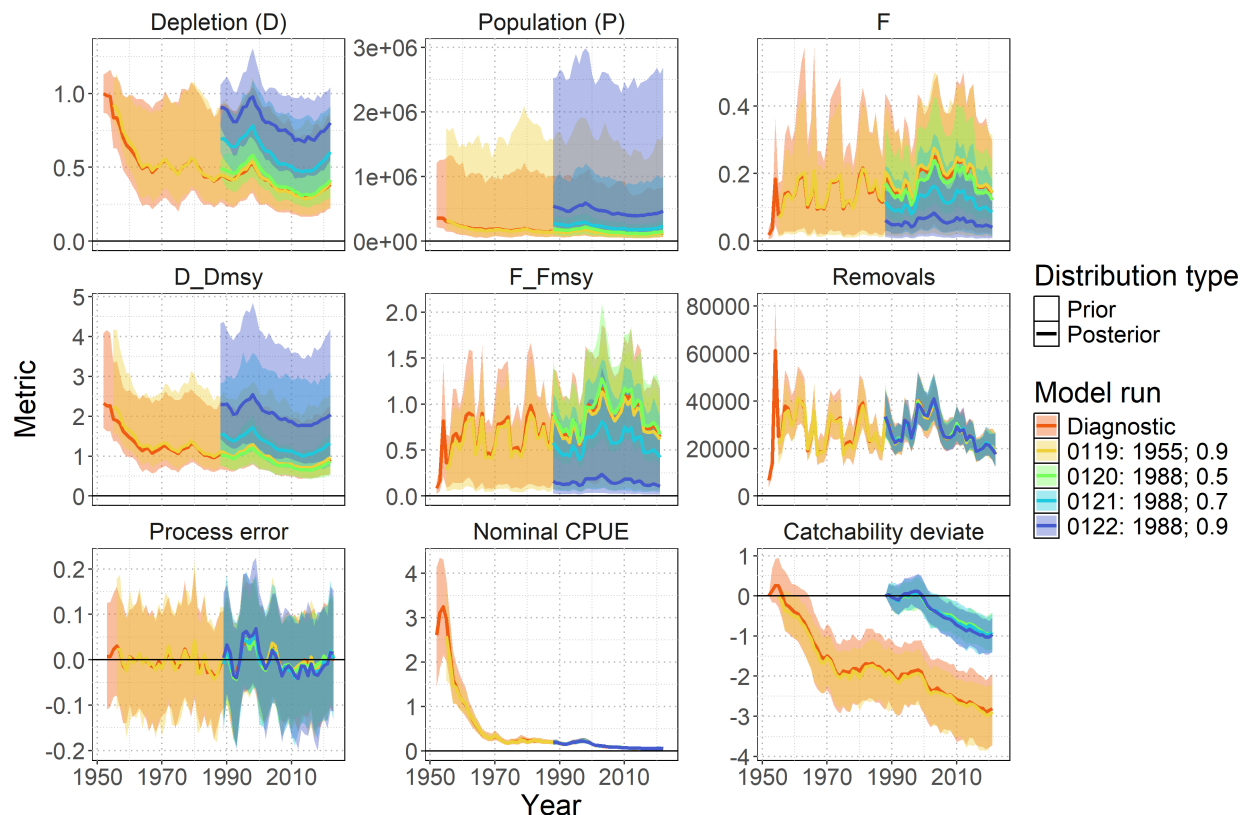
Figure 30: One-off sensitivity group: `Catch reporting`. Prior and posterior parameter distributions showing the evolution from prior (dotted line) to posterior (shaded) distributions for leading model parameters: carrying capacity (logK), intrinsic growth rate (r), observation error (sigmao_add), process error (sigmap), shape parameter, effective catchability (qeff), autocorrelation in catchability (rho), catchability deviation variability (sigma_qdev), initial depletion (x0), and catch likelihood degrees of freedom (nu_catch). The diagnostic model (`0100`) is shown in orange.

Figure 31: One-off sensitivity group: `Catch reporting`. Posterior time series distributions for key derived quantities over time (line = median, shading = 95% credible interval): depletion (D), absolute population size in numbers (P), fishing mortality (F), stock status relative to MSY reference points (D/D_{MSY}, F/F_{MSY}), total removals in numbers, process error, nominal CPUE (numbers caught per 1000 hooks), and catchability deviates. The diagnostic model (`0100`) is shown in orange.
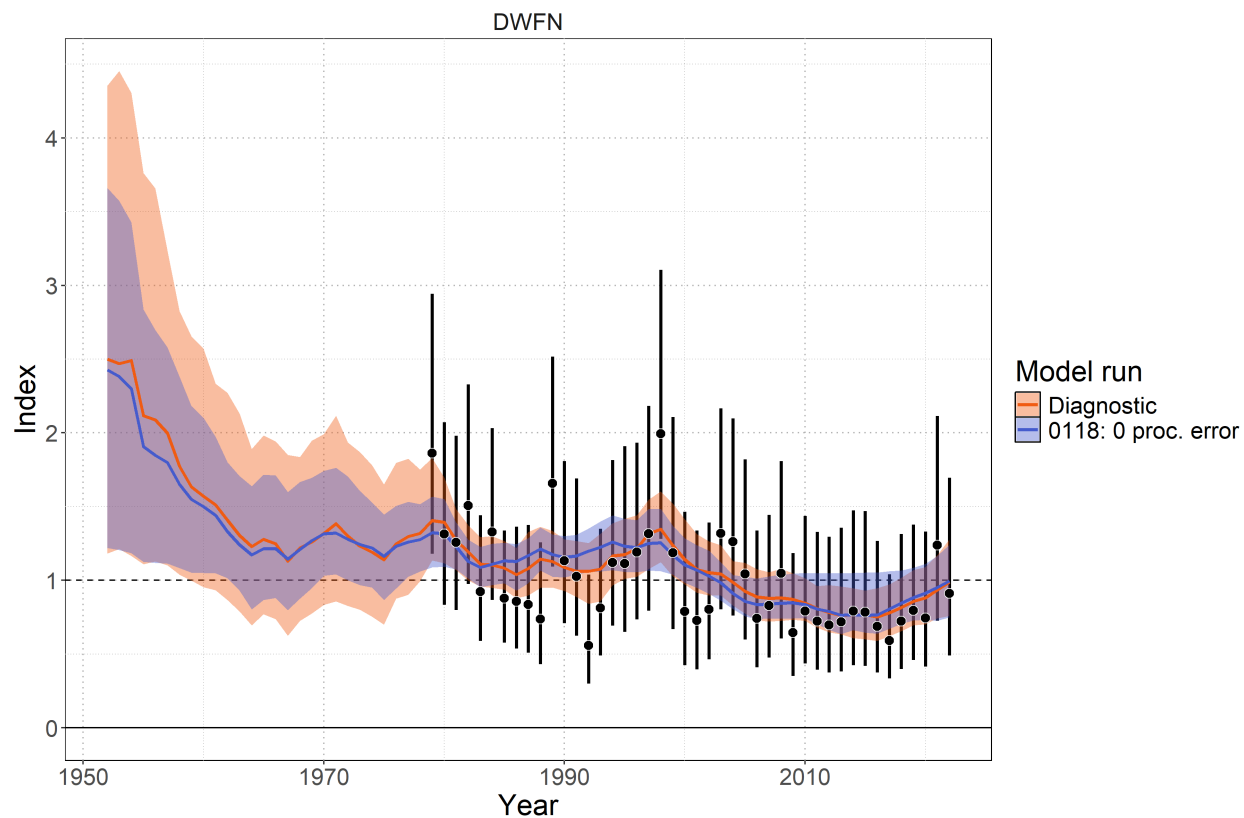
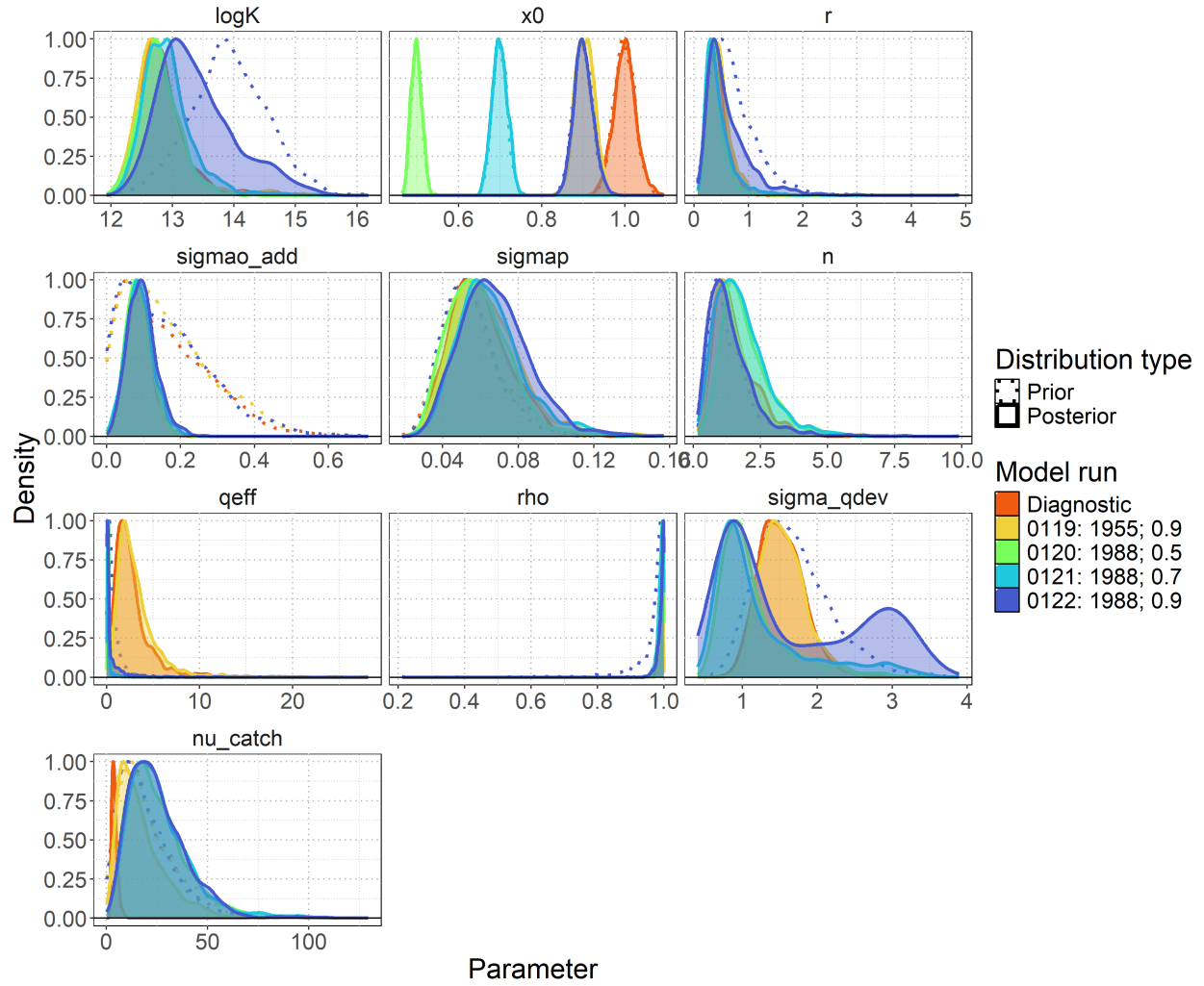Figure 32: One-off sensitivity group: `Process error`. Model fit to standardized index data showing observations (black points with 95% error bars) and model-predictions (colored lines and shaded ribbons representing 95% credible intervals). The diagnostic model (`0100`) is shown in orange.

Figure 33: One-off sensitivity group: `Process error`. Prior and posterior parameter distributions showing the evolution from prior (dotted line) to posterior (shaded) distributions for leading model parameters: carrying capacity (logK), intrinsic growth rate (r), observation error (sigmao_add), process error (sigmap), shape parameter, effective catchability (qeff), autocorrelation in catchability (rho), catchability deviation variability (sigma_qdev), initial depletion (x0), and catch likelihood degrees of freedom (nu_catch). The diagnostic model (`0100`) is shown in orange.
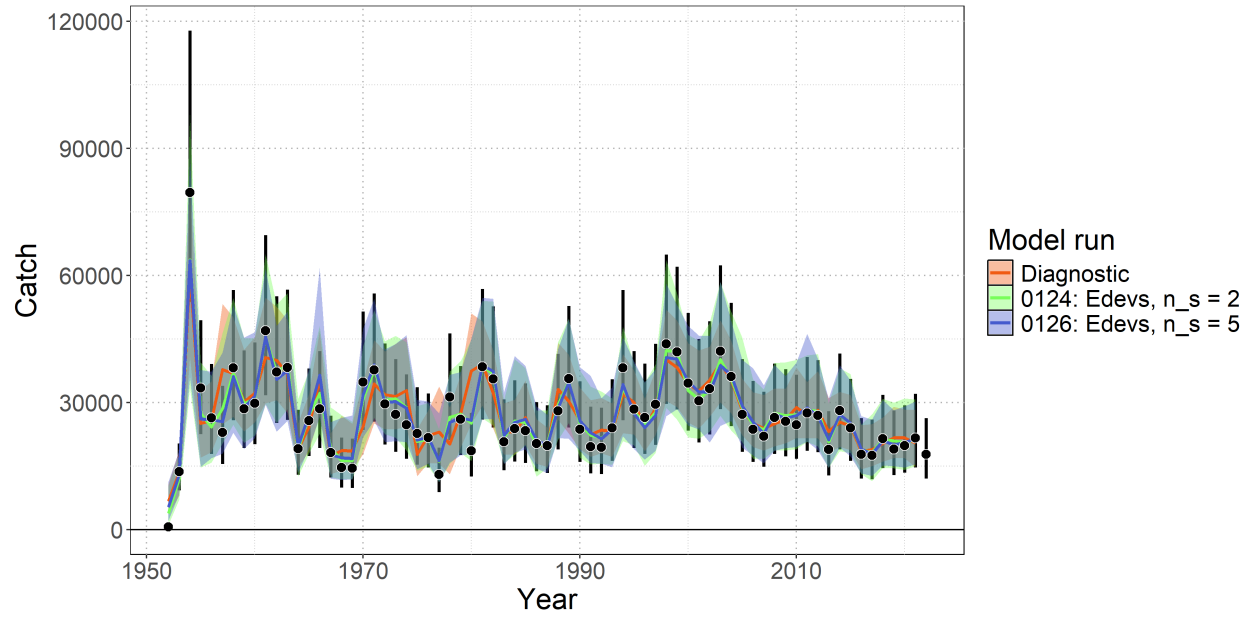
Figure 34: One-off sensitivity group: `Process error`. Posterior time series distributions for key derived quantities over time (line = median, shading = 95% credible interval): depletion (D), absolute population size in numbers (P), fishing mortality (F), stock status relative to MSY reference points (D/D_{MSY}, F/F_{MSY}), total removals in numbers, process error, nominal CPUE (numbers caught per 1000 hooks), and catchability deviates. The diagnostic model (`0100`) is shown in orange.

Figure 35: One-off sensitivity group: `Start year`. Posterior time series distributions for key derived quantities over time (line = median, shading = 95% credible interval): depletion (D), absolute population size in numbers (P), fishing mortality (F), stock status relative to MSY reference points (D/D_{MSY}, F/F_{MSY}), total removals in numbers, process error, nominal CPUE (numbers caught per 1000 hooks), and catchability deviates. The diagnostic model (`0100`) is shown in orange.

Figure 36: One-off sensitivity group: `Start year`. Prior and posterior parameter distributions showing the evolution from prior (dotted line) to posterior (shaded) distributions for leading model parameters: carrying capacity (logK), intrinsic growth rate (r), observation error (sigmao_add), process error (sigmap), shape parameter, effective catchability (qeff), autocorrelation in catchability (rho), catchability deviation variability (sigma_qdev), initial depletion (x0), and catch likelihood degrees of freedom (nu_catch). The diagnostic model (`0100`) is shown in orange.

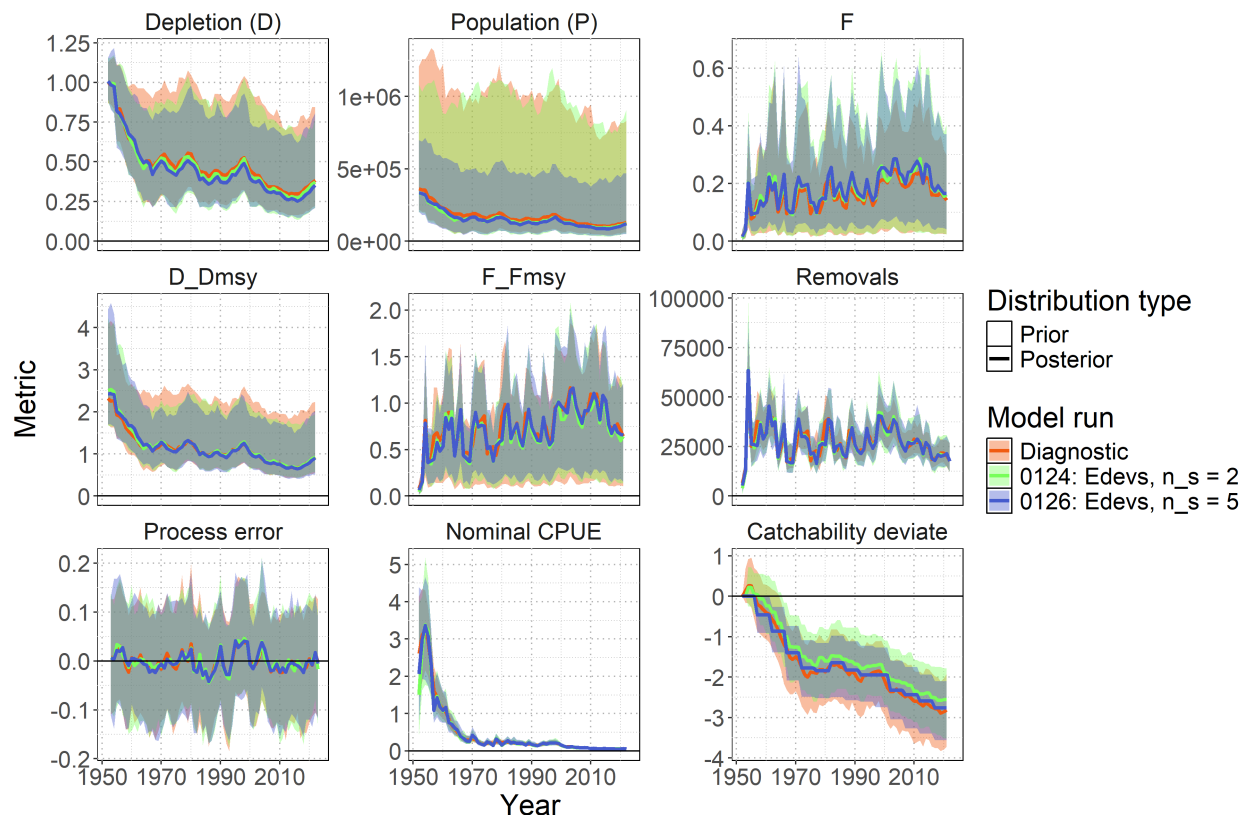Figure 37: One-off sensitivity group: `Inclusion of effort deviates`. Model fit to catch data showing observed catch (black points with 95% error bars) and model-predicted catch (colored lines and shaded ribbons representing 95% credible intervals). The diagnostic model (`0100`) is shown in orange.
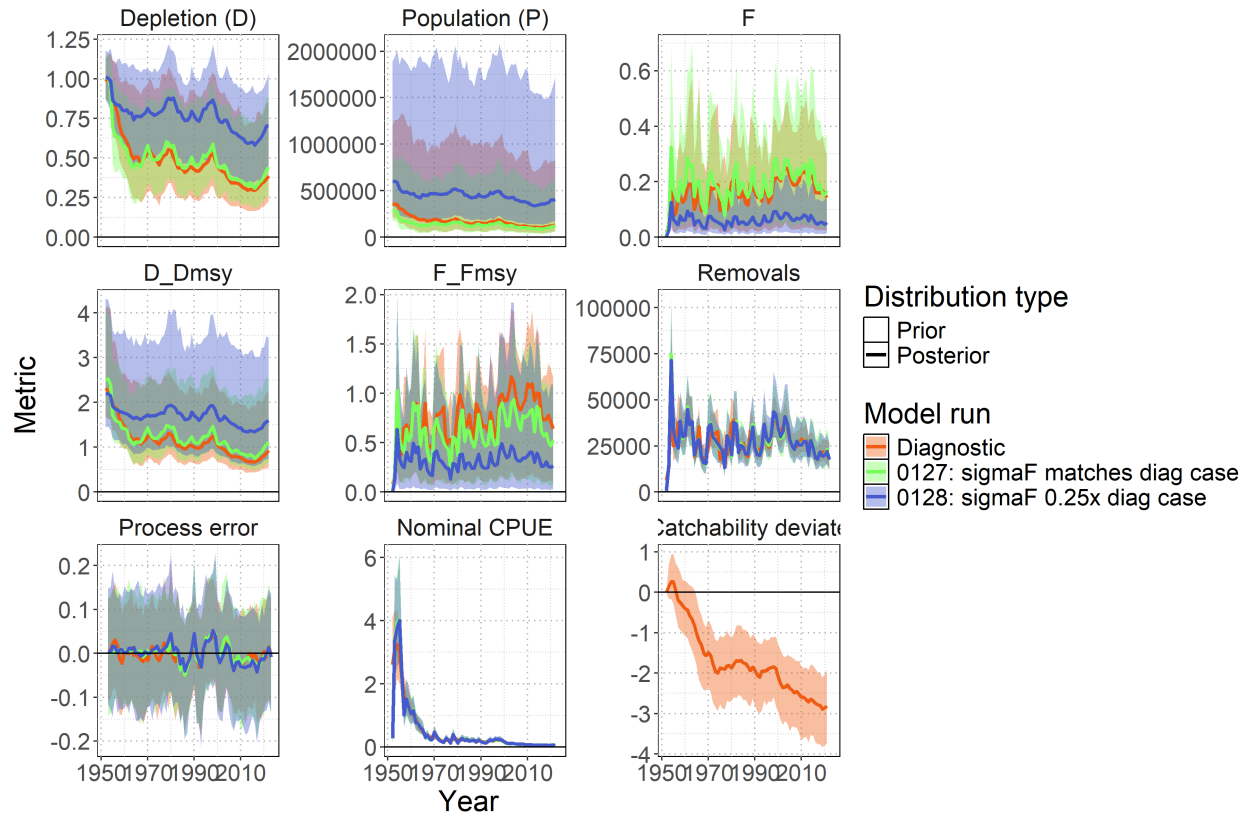
Figure 38: One-off sensitivity group: `Inclusion of effort deviates`. Prior and posterior parameter distributions showing the evolution from prior (dotted line) to posterior (shaded) distributions for leading model parameters: carrying capacity (logK), intrinsic growth rate (r), observation error (sigmao_add), process error (sigmap), shape parameter, effective catchability (qeff), autocorrelation in catchability (rho), catchability deviation variability (sigma_qdev), initial depletion (x0), and catch likelihood degrees of freedom (nu_catch). The diagnostic model (`0100`) is shown in orange.

Figure 39: One-off sensitivity group: `Inclusion of effort deviates`. Posterior time series distributions for key derived quantities over time (line = median, shading = 95% credible interval): depletion (D), absolute population size in numbers (P), fishing mortality (F), stock status relative to MSY reference points (D/D_{MSY}, F/F_{MSY}), total removals in numbers, process error, nominal CPUE (numbers caught per 1000 hooks), and catchability deviates. The diagnostic model (`0100`) is shown in orange.
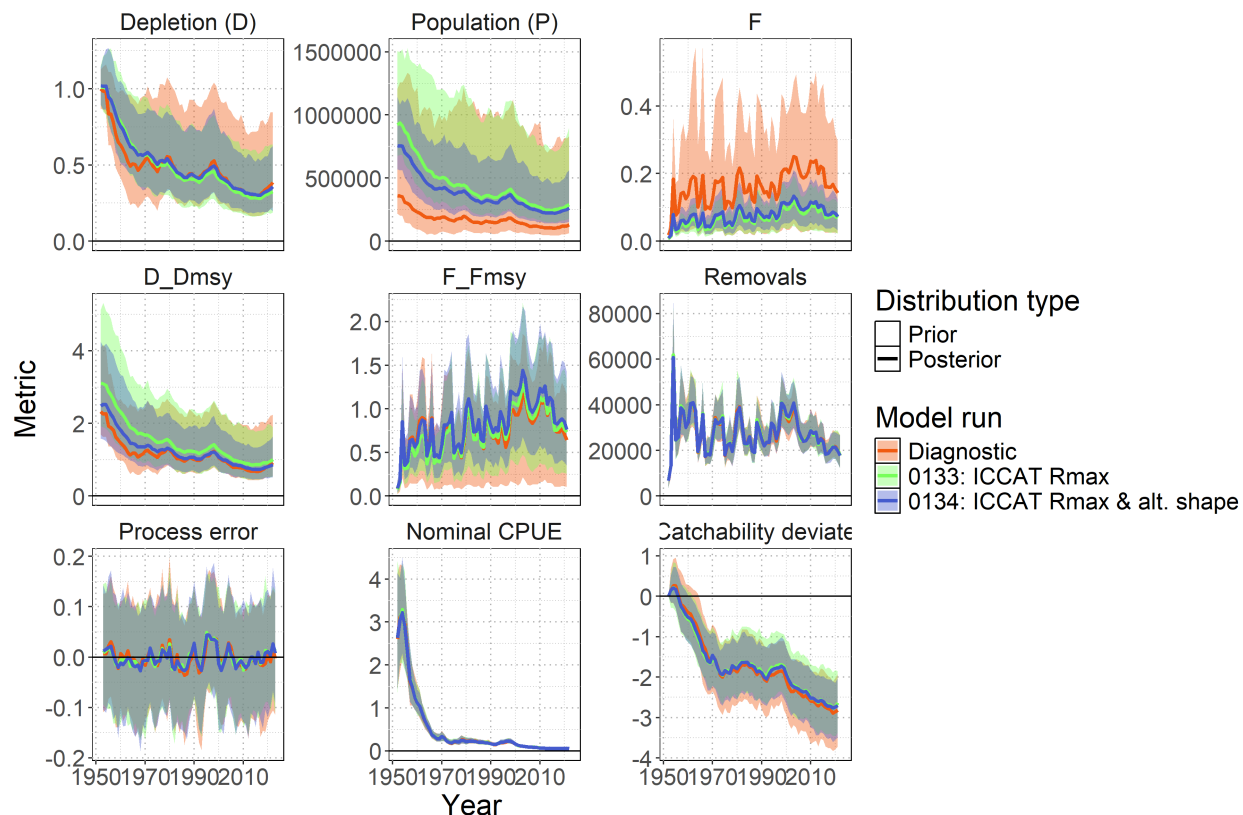
Figure 40: One-off sensitivity group: `Direct estimation of fishing mortality`. Model fit to catch data showing observed catch (black points with 95% error bars) and model-predicted catch (colored lines and shaded ribbons representing 95% credible intervals). The diagnostic model (`0100`) is shown in orange.

Figure 41: One-off sensitivity group: `Direct estimation of fishing mortality`. Posterior time series distributions for key derived quantities over time (line = median, shading = 95% credible interval): depletion (D), absolute population size in numbers (P), fishing mortality (F), stock status relative to MSY reference points (D/D_{MSY}, F/F_{MSY}), total removals in numbers, process error, nominal CPUE (numbers caught per 1000 hooks), and catchability deviates. The diagnostic model (`0100`) is shown in orange.
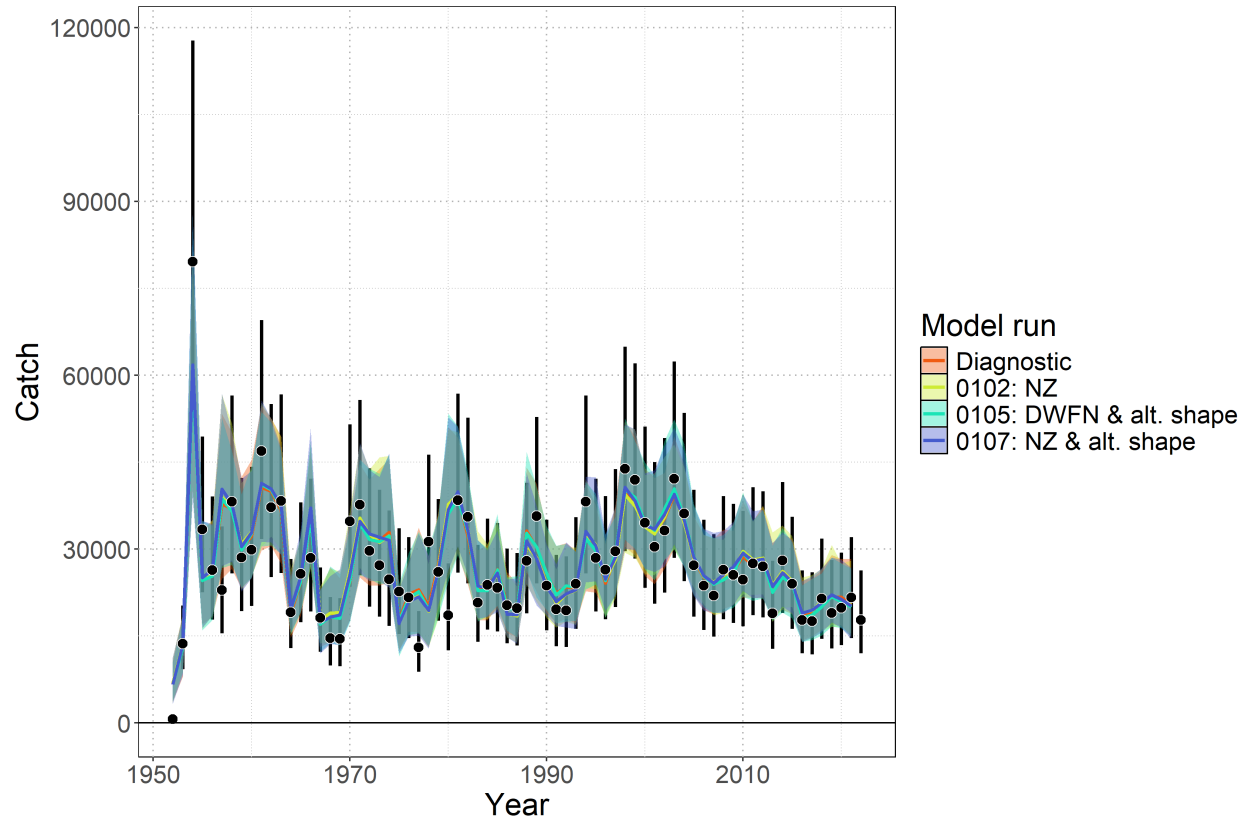
Figure 42: One-off sensitivity group: `Alternative RMax prior`. Posterior time series distributions for key derived quantities over time (line = median, shading = 95% credible interval): depletion (D), absolute population size in numbers (P), fishing mortality (F), stock status relative to MSY reference points (D/D_{MSY}, F/F_{MSY}), total removals in numbers, process error, nominal CPUE (numbers caught per 1000 hooks), and catchability deviates. The diagnostic model (`0100`) is shown in orange.

Figure 43: **Model ensemble**. Model fit to catch data showing observed catch (black points with 95% error bars) and model-predicted catch (colored lines and shaded ribbons representing 95% credible intervals). The diagnostic model (`0100`) is shown in orange.
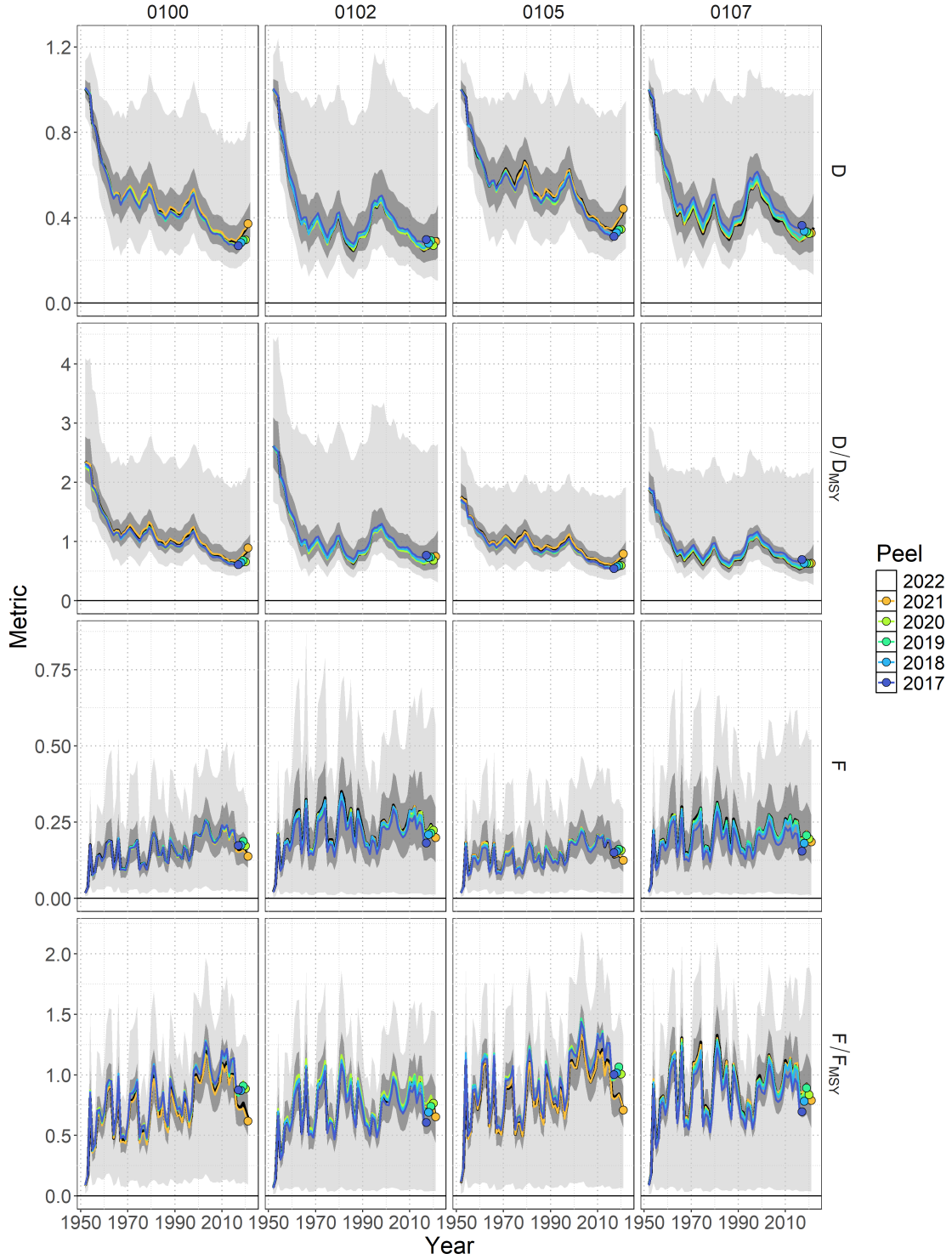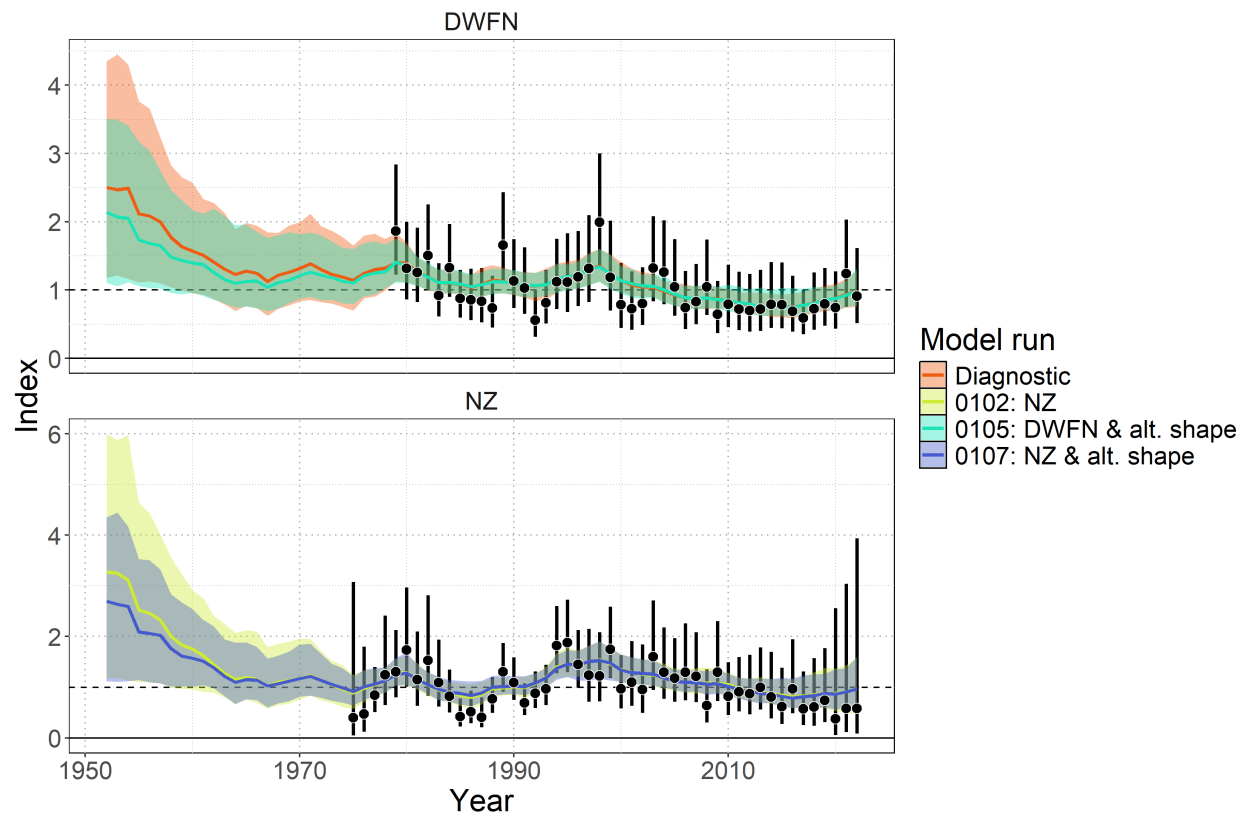
Figure 44: **Model ensemble.**. Model fit to standardized index data showing observations (black points with 95% error bars) and model-predictions (colored lines and shaded ribbons representing 95% credible intervals). The diagnostic model (`0100`) is shown in orange.

Figure 45: Retrospective analysis for the four ensemble models with respect to time series of depletion D_t, depletion relative to depletion at MSY D_t/D_{MSY}, fishing mortality F_t, and fishing mortality relative to the rate of fishing mortality that produces MSY F_t/F_{MSY}. The base model with data included through 2022 (black line – median; dark shading – 50% credible interval; light shading – 95% credible interval) is shown relative to the retrospective models. Colored lines correspond to the last year of index data and the colored point indicates the estimate in the last year of the retrospective peel.
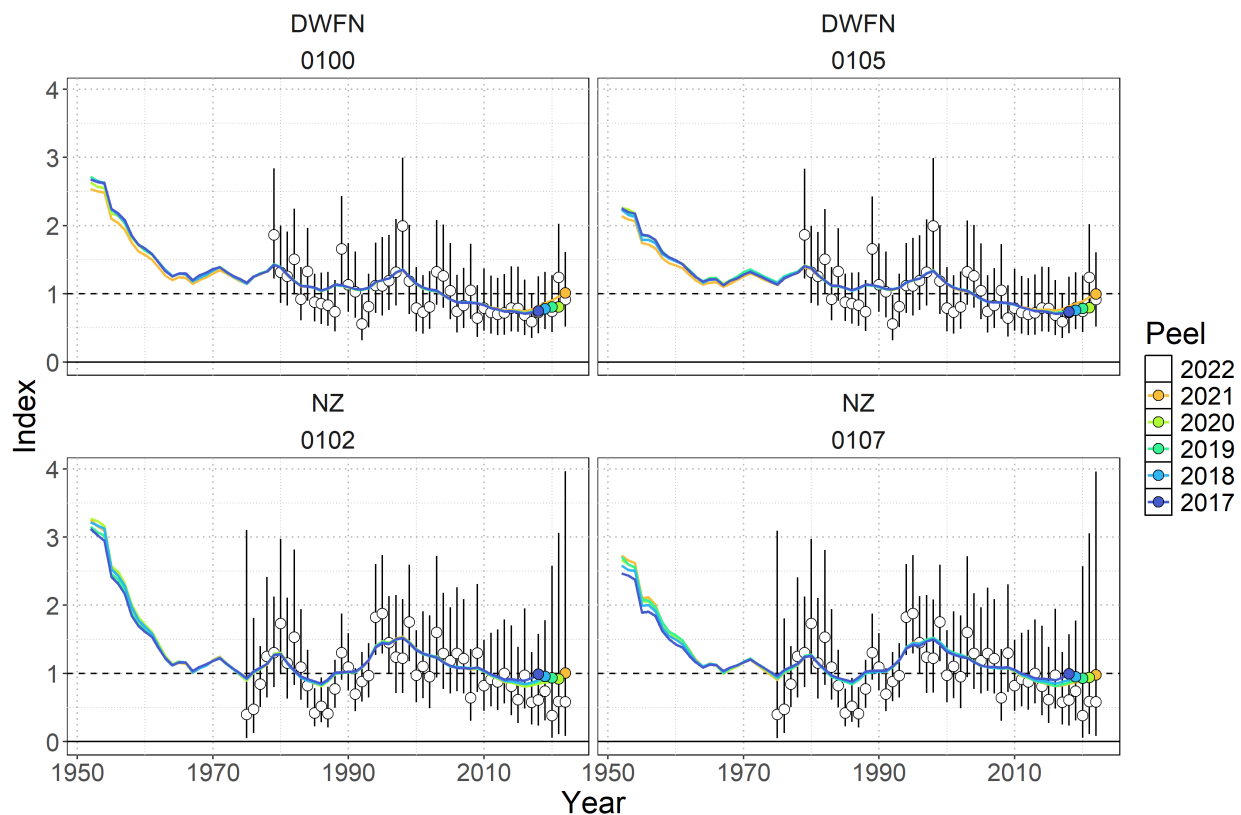
Figure 46: Standardized indices of relative abundance used in the model ensemble. Open circles show observed values (standardized to mean of 1; black horizontal line) and the vertical bars indicate the observation error (95% confidence interval). One year ahead 'model-free' hindcast predictions are shown by the colored lines, where the color indicates the last year of index data seen by the model. The predicted value is shown one year-ahead with the colored point.
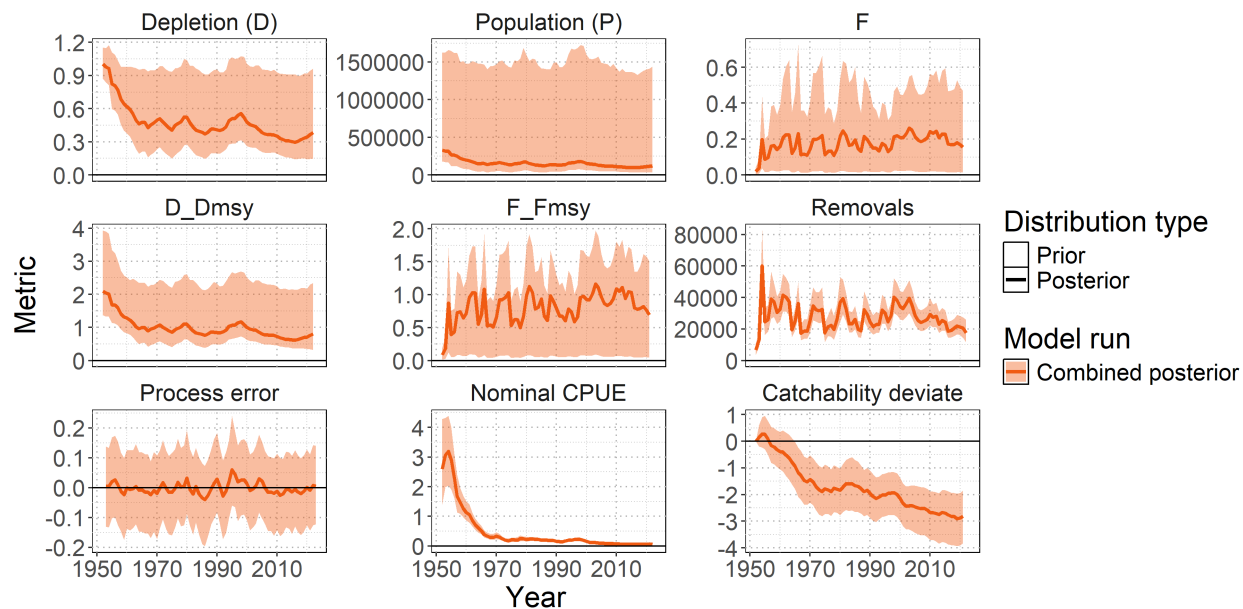
Figure 47: **Model ensemble**. Posterior time series distributions for key derived quantities over time (line = median, shading = 95% credible interval): depletion (D), absolute population size in numbers (P), fishing mortality (F), stock status relative to MSY reference points (D/D_{MSY}, F/F_{MSY}), total removals in numbers, process error, nominal CPUE (numbers caught per 1000 hooks), and catchability deviates. The model ensemble is shown in orange.
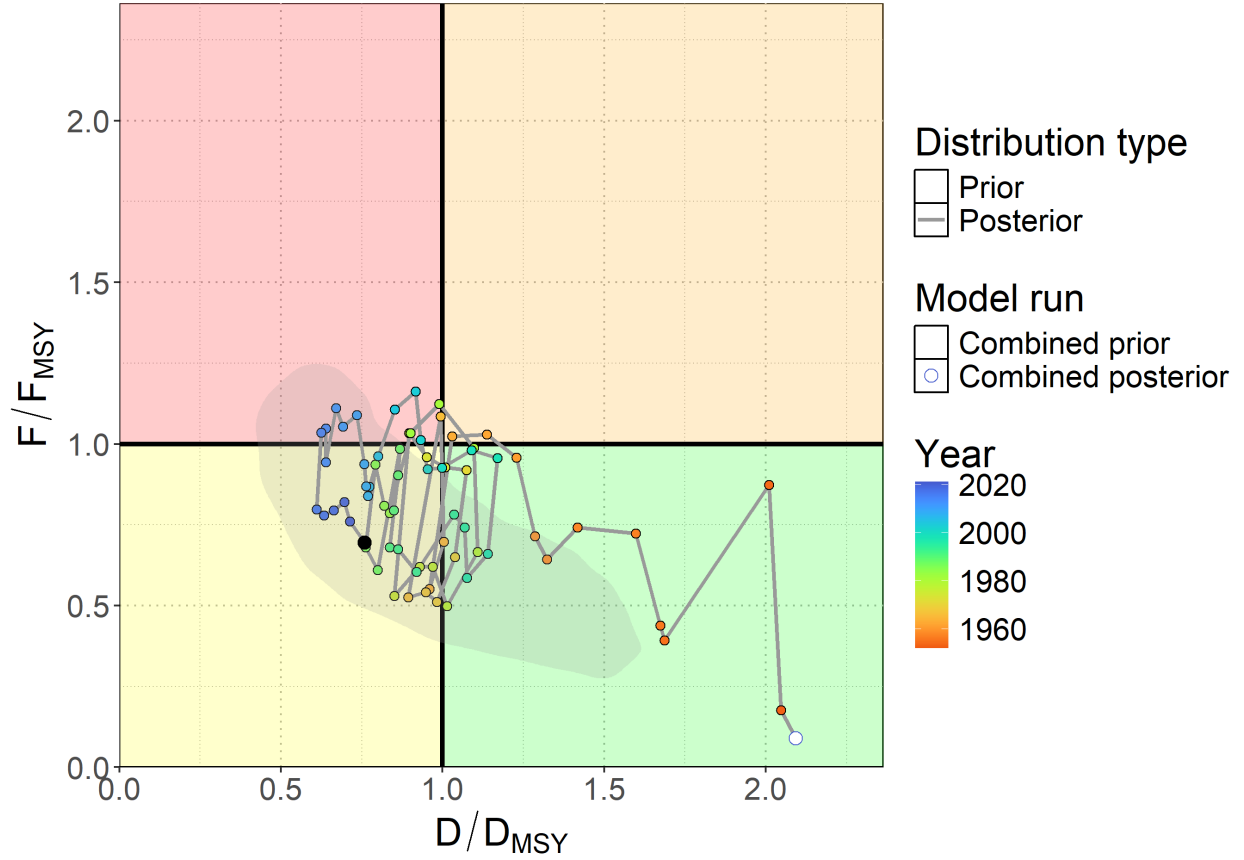
Figure 48: **Model ensemble**. Kobe plot showing the median posterior estimate of *latest* stock depletion relative to depletion at MSY (D_{latest}/D_{MSY}) and fishing mortality relative to MSY (F_{latest}/F_{MSY}). Points represent model estimates colored by year (blue $\sim$ 2020, green $\sim$ 1985, orange $\sim$ 1950), with connecting line showing the trajectory over time. The gray shaded contour is the bi-variate 95% credible distribution around the terminal year estimate D_{2022}/D_{MSY} and F_{2021}/F_{MSY}.
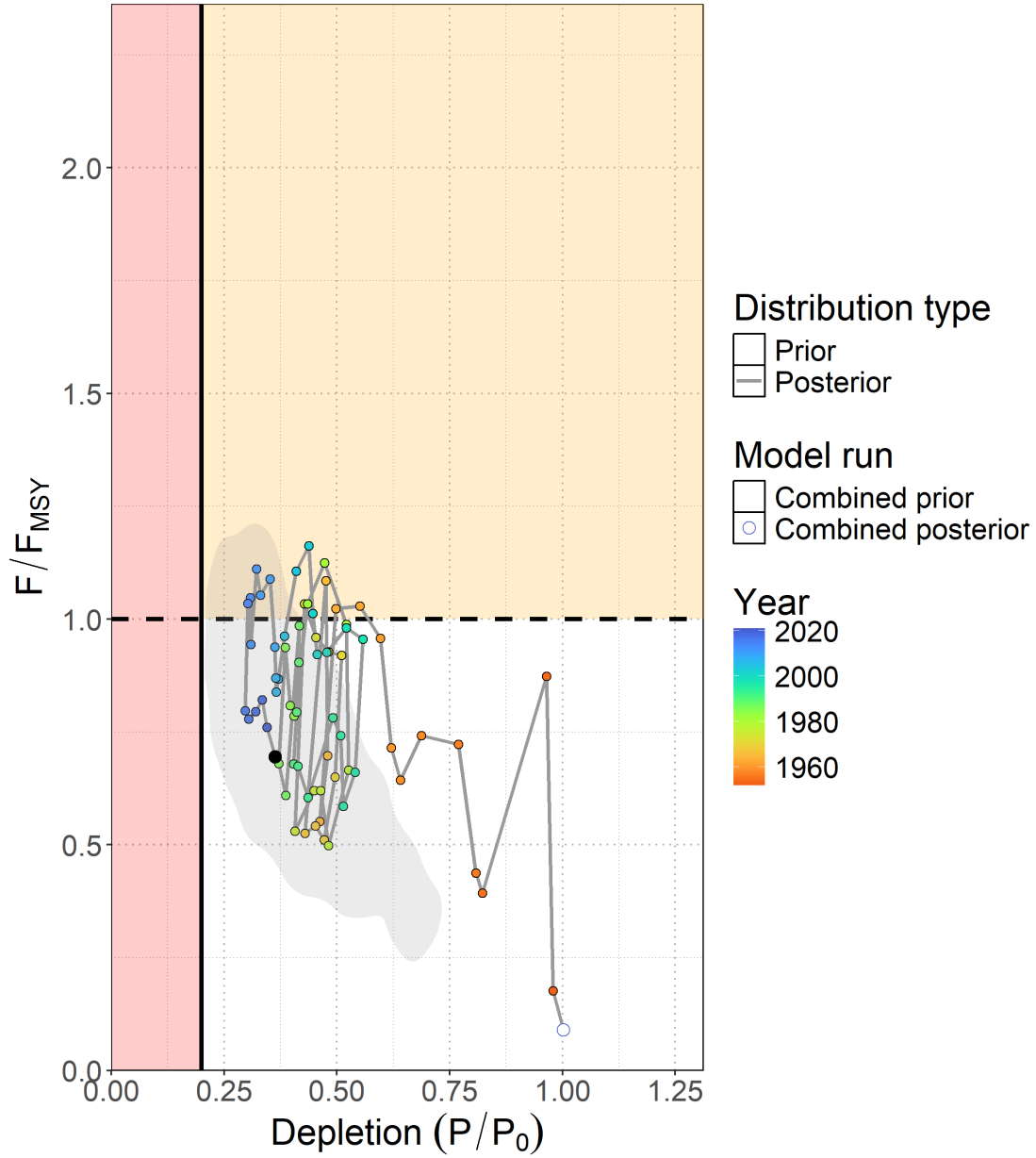
Figure 49: **Model ensemble**. Majuro plot showing the median posterior estimate of *latest* stock depletion ($D_{latest} = P_{latest}/P_0$) and fishing mortality relative to MSY (F_{latest}/F_{MSY}). Points represent model estimates colored by year (blue $\sim$ 2020, green $\sim$ 1985, orange $\sim$ 1950), with connecting line showing the trajectory over time. The gray shaded contour is the bi-variate 95% credible distribution around the terminal year estimate D_{2022} and F_{2021}/F_{MSY}.
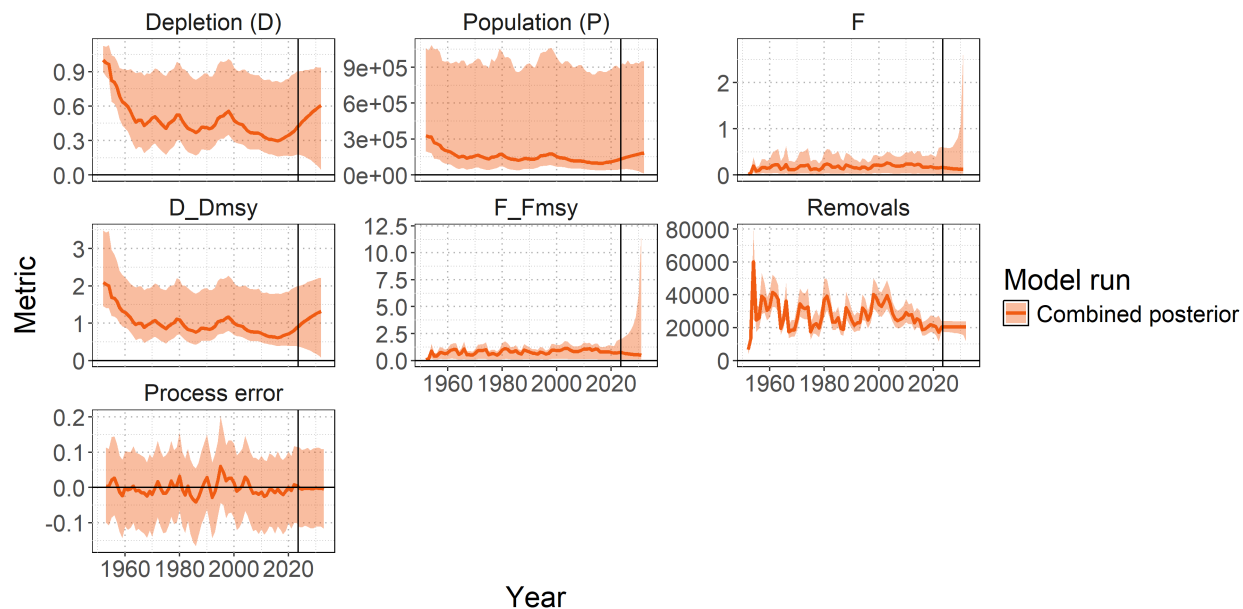
Figure 50: **Model ensemble**. Posterior time series distributions for key derived quantities over time during the forecast period 2023-2032 (line = median, shading = 90% credible interval): depletion (D), absolute population size in numbers (P), fishing mortality (F), stock status relative to MSY reference points (D/D_{MSY}, F/F_{MSY}), total removals in numbers, process error, nominal CPUE (numbers caught per 1000 hooks), and catchability deviates. The model ensemble is shown in orange. A 90% credible interval is shown to restrict the y-axis of the fishing mortality F panel which shows high values of F in the projection period as a very small percentage of populations are estimated to go to zero under recent average catch levels.

11 Technical Annex

Additional technical detail of population dynamics components used to solve the Euler-Lotka equation for R_{Max}.

11.1 Equilibrium Age-Structured Population Dynamics

11.1.1 Survival

Survival schedule was calculated assuming constant natural mortality:

$$l_a = \begin{cases} 1 & \text{if } a = 1 \\ l_{a-1} \times \exp(-M_{ref}) & \text{if } 1 < a < A_{Max} \\ \frac{l_{A_{Max}-1}}{1-\exp(-M_{ref})} & \text{if } a = A_{Max} \text{ (plus group)} \end{cases}$$

11.1.2 Fecundity

Reproductive output incorporated biological constraints:

$$f_a = \frac{\psi_a \times W_a \times s}{\rho}$$

where: - ψ_a = proportion mature at age a - W_a = weight at age a (proxy for reproductive capacity) - s = sex ratio (proportion female) - ρ = reproductive cycle (years between spawning events)

Maturity-at-age was derived from length-based logistic maturity:

$$\psi_L = \frac{\exp(a_{mat} + b_{mat} \times L)}{1 + \exp(a_{mat} + b_{mat} \times L)}$$

where $b_{mat} = -a_{mat}/L_{50}$, then integrated over the length-at-age distribution to obtain ψ_a.

Length-at-age followed the Francis parameterization:

$$L_a = L_1 + (L_2 - L_1) \times \frac{1 - \exp(-k(a - age_1))}{1 - \exp(-k(age_2 - age_1))}$$

with length variability modeled using probability density functions linking length bins to ages.

Weight-at-age used the allometric relationship:

$$W_a = a_w \times L_a^{b_w}$$

11.1.3 Stock-Recruitment Relationship

The slope at the origin of the stock-recruitment curve was:

$$\alpha = \frac{4h}{(1 - h)\phi_0}$$

The calculation incorporated steepness (h) through the Beverton-Holt stock-recruitment relationship. Spawners-per-recruit in the unfished state was:

$$\phi_0 = \sum_{a=1}^{A_{Max}} l_a f_a$$

This α parameter links the intrinsic rate of increase to recruitment compensation, ensuring that R_{Max} estimates reflect realistic population productivity under the assumed stock-recruitment dynamics.

Successful simulations (those yielding $R_{Max} > 0$ and $R_{Max} < 1.5$) were filtered to create a plausible prior distribution. The resulting distribution was fitted to a lognormal prior for use in the BSPM.

12 Appendix

12.1 Highest Density Intervals vs. Equal-Tailed Intervals

Highest density intervals (HDIs) provide an alternative to conventional equal-tailed intervals (ETIs) for characterizing uncertainty in Bayesian stock assessments. While both represent credible intervals containing a specified probability mass (e.g., 95%), they differ fundamentally in how this probability is distributed.

Equal-Tailed Intervals (ETIs) place equal probability (2.5%) in each tail of the distribution, resulting in the familiar percentile-based intervals (2.5th to 97.5th percentiles for 95% intervals). In contrast, Highest Density Intervals (HDIs) contain the 95% most probable values, defined as the shortest possible interval containing the specified probability mass. Critically, all values within an HDI are more credible (have higher probability density) than any values outside it, whereas ETIs can exclude highly probable values while including less probable ones. This is more clearly demonstrated in the following figure (Figure 51), where the ETI and HDI are presented for D_{recent}/D_{MSY}. As seen in panel A of Figure 51, the lower tail of the ETI inappropriately excludes high probability observations. In panel B of Figure 51, we see that all low probability observations are appropriately excluded. This shrinks the range of the credible interval, and shifts it left to exclude lower probability values in the right tail.

HDIs offer several advantages over ETIs: they are always the shortest possible credible interval (reducing perceived uncertainty), provide more intuitive interpretation by including only the most probable values, and are better suited for the right-skewed distributions common in stock assessments. However, HDIs may require explanation to stakeholders familiar with conventional intervals, and unlike ETIs, change under nonlinear transformations of the parameter.

As applied to stock assessments, HDIs typically result in narrower uncertainty intervals and can shift interval bounds toward lower values in right-skewed distributions (A. M. Edwards & Auger-Méthé, 2025). This can affect perceptions of stock status, particularly when intervals approach management reference points. The following table (Table 9) presents the median, ETI and HDI of all key stock status and management reference points for consideration by the WCPFC Scientific Committee.

Table 9: Comparison of equal-tailed intervals (ETI) and highest density intervals (HDI) for the posterior distributions of key stock status and management reference points.

| Category	Metric	Median	95% ETI	95% HDI
Reference points				
Depletion	LRP (D_{MSY})	0.48	[0.26 − 0.7]	[0.26 − 0.7]
Abundance	LRP (P_{MSY})	155,183 n	[63,037 − 808,861]	[37,772 − 556,497]
Abundance	LRP	65,041 n	[36,198 − 327,844]	[28,895 − 221,199]
	($0.2 \times \log(K)$)			
Catch	MSY	29,962 n	[25,828 − 184,069]	[23,040 − 103,619]
Fishing mortality	F_{MSY}	0.23	[0.08 − 0.69]	[0.04 − 0.58]
Estimates				
Depletion	D_{latest}	0.38	[0.14 − 0.96]	[0.1 − 0.9]
Depletion	D_{recent}	0.37	[0.15 − 0.94]	[0.11 − 0.89]
Abundance	P_{latest}	121,943 n	[34,067 − 1,479,253]	[8,612 − 925,445]
Abundance	P_{recent}	117,967 n	[34,199 − 1,442,511]	[13,983 − 911,307]
Catch	C_{latest}	17,488 n	[11,545 − 25,988]	[10,947 − 25,108]
Catch	C_{recent}	20,460 n	[16,891 − 24,770]	[16,962 − 24,792]
Fishing mortality	F_{latest}	0.15	[0.01 − 0.47]	[0 − 0.4]
Fishing mortality	F_{recent}	0.17	[0.01 − 0.49]	[0 − 0.42]
Status				
Depletion	D_{latest}/D_{MSY}	0.81	[0.32 − 2.36]	[0.24 − 2.02]
Depletion	D_{recent}/D_{MSY}	0.77	[0.33 − 2.3]	[0.24 − 1.95]
Fishing mortality	F_{latest}/F_{MSY}	0.69	[0.05 − 1.51]	[0.01 − 1.35]
Fishing mortality	F_{recent}/F_{MSY}	0.77	[0.05 − 1.51]	[0.02 − 1.34]
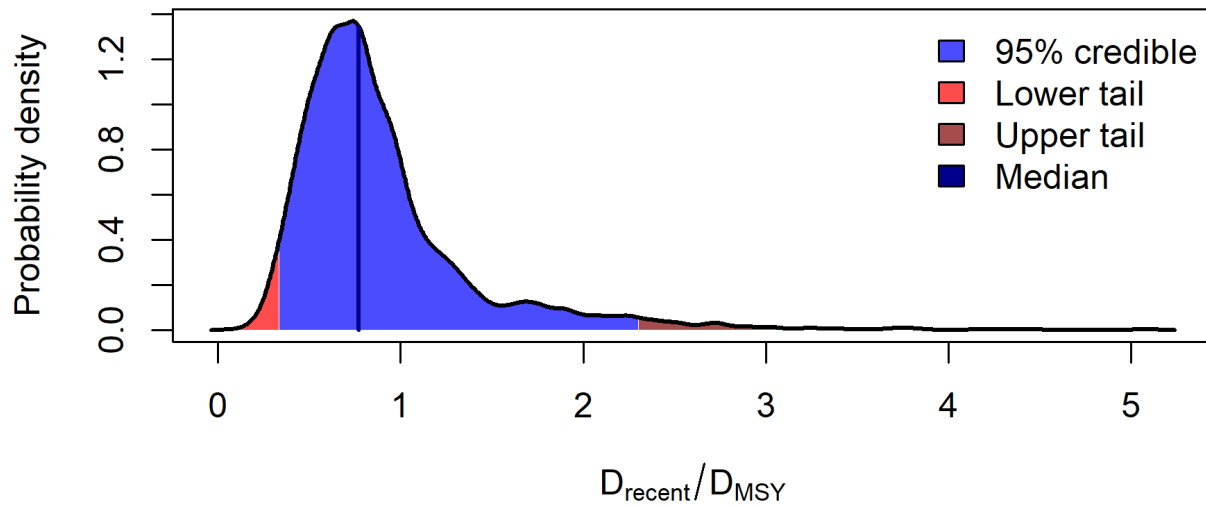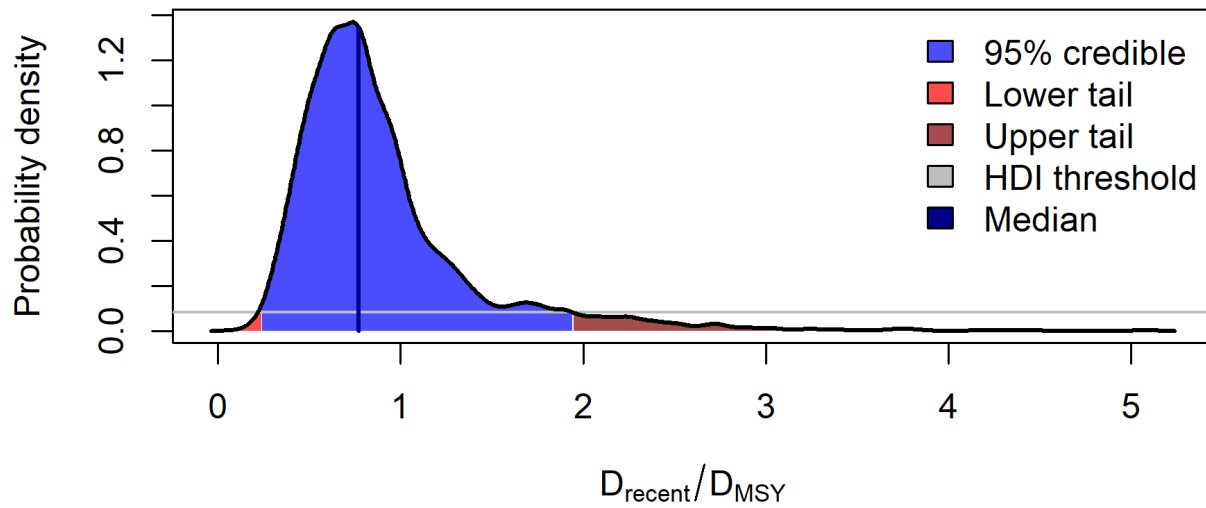
Figure 51: Comparison of equal-tailed intervals (ETI) and highest density intervals (HDI) for the posterior distribution of recent depletion relative to MSY (D_{recent}/D_{MSY}). Panel A shows the 95% ETI with equal probability (2.5%) in each tail. Panel B shows the 95% HDI, which contains the most probable values and is bounded by the horizontal threshold line (gray) where all values above have higher probability density than values below.