



SCIENTIFIC COMMITTEE
TWENTY-FIRST REGULAR SESSION

Tonga
13-21 August 2025

Data-moderate assessment approaches for Southwest Pacific Ocean (SWPO) striped marlin

WCPFC-SC21-2025/SA-IP-XX
15 June 2025

DRAFT

N. Ducharme-Barth¹

¹NOAA Fisheries, Pacific Islands Fisheries Science Center, 1845 Wasp Boulevard, Bldg. 176, Honolulu HI 96818

Table of contents

1 Executive summary	3
2 Introduction	4
3 Methods	5
3.1 Model Framework	5
3.1.1 Input data	5
3.1.2 Population Dynamics	6
3.1.3 Observation Model	6
3.1.4 Catch Treatment	6
3.2 Prior Development	6
3.2.1 Biological Simulation Framework	6
3.2.2 Maximum Intrinsic Rate of Increase (R_{Max}) Prior	7
3.2.3 Depletion Prior	7
3.2.4 Fishing Mortality Prior	8
3.3 Model Implementation	8
3.4 Model Diagnostics	9
3.4.1 Convergence Diagnostics	9
3.4.2 Data Fits	9
3.4.3 Posterior Predictive Checks	10
3.4.4 Retrospective Analysis	10
3.4.5 Hindcast Cross-Validation	10
3.5 Forecasts	11
4 Results	11
4.1 Model Convergence and Diagnostics	11
4.1.1 Model Fit to Data	11
4.1.2 Model Validation	11
4.2 Population Dynamics and Stock Status	12
4.3 Biological Reference Points	13
4.4 Future Projections	13
5 Discussion	13
5.1 General remarks	13
5.2 Challenges, limitations & key uncertainties	14
5.3 Future stock assessment modeling considerations	14
6 Declaration of Generative AI use	15
7 References	15
8 Tables	18
9 Figures	23
10 Technical Annex	39
10.1 Equilibrium Age-Structured Population Dynamics	39
10.1.1 Survival	39
10.1.2 Fecundity	40
10.1.3 Stock-Recruitment Relationship	40
10.2 Transitional Mean Weight Model	41

1 Executive summary

Previous integrated age-structured assessments of Southwest Pacific Ocean (SWPO) striped marlin (*Kajikia audax*) have encountered significant technical challenges, including poor fits to size composition data, conflicts between data sources, and difficulties in estimating model initial conditions. To address these issues, this study developed a data-moderate assessment approach using Bayesian surplus production models (BSPMs) to provide estimates of stock status and trends for SWPO striped marlin from 1952-2022.

Three BSPM variants were developed using a Fletcher-Schaefer production function framework implemented in Stan: (1) a baseline model with biologically-informed priors, (2) a depletion-constrained model incorporating additional information from New Zealand recreational fishing size data, and (3) an extended model with refined fishing mortality priors. The models were fitted to annual catch data (1952-2022) and a standardized CPUE index (1988-2022), incorporating biological uncertainty through simulation-based priors derived from 500,000 parameter combinations of life history characteristics.

All three models achieved satisfactory convergence and demonstrated good predictive performance. They estimated consistent population trajectories showing decline from unfished conditions in 1952 to minimum levels around the early 1990s, gradual recovery through the 2000s and 2010s, and recent stabilization at moderate depletion levels. Median estimates of depletion and exploitation relative to Maximum Sustainable Yield (MSY) based reference points did not indicate that the stock was overfished or undergoing overfishing in any of the three models. Only the most conservative model variant indicated some recent risk of the stock falling below MSY-based reference points. Projecting the stock forward under status quo catch scenarios (average 2018-2022 removals), ten-year projections (2023-2032) from all models indicated a continued population recovery through 2032.

This data-moderate assessment provides a robust alternative framework for evaluating SWPO striped marlin stock status when data conflicts present challenges for more complex integrated models. While all models suggest the stock is likely not overfished and currently recovering, the substantial uncertainty in absolute population scale emphasizes the importance of continued monitoring and the potential value of developing multiple modeling approaches. Furthermore, surplus production models simplify age-structured population dynamics, which can potentially bias MSY estimates when age-structured processes differ from the aggregated model assumptions. This analysis complements the 2025 revised integrated stock assessment and can provide managers with additional perspective on likely stock status and trajectory.

2 Introduction

Assessments of Southwest Pacific Ocean (SWPO) striped marlin (*Kajikia audax*) have been challenging. 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 appears largely driven by the size composition from multiple fisheries. While estimation of low biomass scale is not inherently problematic, the resulting high ratio of fishing mortality (F) to natural mortality (M) that is maintained over multiple decades in the face of relative stability in catches and catch-per-unit-effort (CPUE) is suspicious and potentially indicative of model mis-specification.

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. 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. Over-simplification 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 Model 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 (Edwards, 2024; Winker et al., 2020). 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 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).

The models begin from an assumed unfished state in 1952 and incorporates process error in population dynamics, observation error in abundance indices, and uncertainty in biological parameters. Several model variants were developed: a baseline model (`0001-2024cpueExPrior`), a depletion-constrained model (`0002-2024cpueDepPrior`) that incorporates additional information about historical stock status from mean size data, and an extension of the depletion-constrained model that assumes a different prior on the variability in fishing mortality (`0003-2024cpueFPrior`). The `0002-2024cpueDepPrior` and `0003-2024cpueFPrior` models include a mean weight-based prior on 1988 depletion, and more detail on the development of this prior can be found in Section 3.2.3. Unless otherwise mentioned, all models used the same prior parameterizations (Table 2) based on the biological simulation filtering described in Section 3.2.

3.1.1 Input data

The BSPMs were fitted to catch and standardized CPUE data for SWPO striped marlin. Annual catch data (in numbers) 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. A single standardized CPUE index (1988-2022) from the diagnostic case of the 2024 assessment (Castillo-Jordan et al., 2024) was used. The index is very noisy (Figure 2). In general, however, there is a slight declining trend over the index period. The decline was most pronounced after 2000, though the trend stabilized somewhat before showing a slight increase in recent years.

3.1.2 Population Dynamics

The population dynamics follow a Fletcher-Schaefer surplus production model with state-space formulation where population depletion x_t (relative to carrying capacity K) evolves according to:

$$x_t = \begin{cases} (x_{t-1} + R_{Max}x_{t-1}(1 - \frac{x_{t-1}}{h}) - C_{t-1}) \times \epsilon_t, & x_{t-1} \leq D_{MSY} \\ (x_{t-1} + x_{t-1}(\gamma \times m)(1 - x_{t-1}^{n-1}) - C_{t-1}) \times \epsilon_t, & x_{t-1} > D_{MSY} \end{cases}$$

where R_{Max} is the maximum intrinsic rate of increase, n is the shape parameter, 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: $D_{MSY} = (1/n)^{1/(n-1)}$, $h = 2D_{MSY}$, $m = R_{Max}h/4$, and $\gamma = n^{n/(n-1)}/(n-1)$.

3.1.3 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.

3.1.4 Catch Treatment

Annual fishing mortality is estimated directly $F_t \sim N^+(0, \sigma_F)$ with catch fitted using a lognormal likelihood with uncertainty $\sigma_C = 0.2$.

Catch is predicted as:

$$C_t = (x_t + \text{surplus production}_t) \times (1 - \exp(-F_t)) \times \epsilon_t \times K$$

3.2 Prior Development

3.2.1 Biological Simulation Framework

Informative priors for key population parameters were developed through biological simulation using Monte Carlo sampling from biologically plausible parameter distributions. 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 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).

3.2.2 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 10.1 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 prior pushforward analysis, similar to (ISC, 2024), was used to further refine this distribution and develop a lognormal prior for R_{Max} . Briefly, random values of R_{Max} along with random values of the other leading parameters of the BSPM (drawn from the prior distributions described in Table 2) were used to drive simulated populations forward subject to direct removal of the observed catch (Figure 1). Based on the results of the prior pushforward (Figure 3), the R_{Max} distribution was further filtered based on the following criteria:

- Depletion in the terminal year $> 2\%$ and $R_{Max} < 1$
- Depletion in the terminal year $> 2\%$, $R_{Max} < 1$, and trend in depletion over the last decade between -5% and 10%

This last filtering step approximates the recent relative trend seen in the index (Figure 2). Fitting a lognormal distribution to the remaining R_{Max} values results in the prior distribution specified in Table 2 and shown in Figure 4. Note that though initial filtering restricted $R_{Max} < 1$, the final lognormal prior still assigns some prior weight to $R_{Max} > 1$.

3.2.3 Depletion Prior

It is not possible to fit to size composition data in a BSPM framework. However, information on fishing mortality and depletion may be contained in the long time series of New Zealand recreational fishing data. This fishery routinely catches the largest individuals and shows a decline in average size since the start of the model period (Figure 5), before which limited large-scale commercial fishing occurred. Assuming constant availability and selectivity, declining mean weight can be an indication of increased mortality. Assuming natural mortality M has remained constant, this indicates that the population may be in a more depleted state than in the initial period. The 0002-2024cpueDepPrior variant was developed to capitalize on this potential additional source of information.

A supplementary analysis was used to estimate relative population depletion in 1988 using New Zealand recreational fishing weight data. A modified Gedamke-Hoenig approach (Gedamke & Hoenig, 2006) for using size composition data to estimate mortality in non-equilibrium situations was applied to estimate change in total mortality from Z_1 (initial period) to Z_2 (final period) at a known change point (1952) when industrial fishing commenced at the start of the model period (more detail in Technical Annex Section 10.2). These two total mortality rates were calculated for each of the biological parameter combinations described in Section 3.2.1 and used to calculate relative spawning stock biomass depletion as the ratio of spawning potential under the two mortality regimes:

$$\text{Relative Depletion} = \frac{SPR_{Z_2}}{SPR_{Z_1}} = \frac{\sum_{a=1}^{A_{Max}} l_{a,Z_2} f_a}{\sum_{a=1}^{A_{Max}} l_{a,Z_1} f_a}$$

where $l_{a,Z}$ is survival to age a under total mortality Z , and f_a is reproductive output at age a (incorporating maturity, weight, sex ratio, and reproductive cycle). Additional technical detail on the population dynamics components of this equation can be found in Section 10.1 of the Technical Annex.

The resulting depletion estimates across successful parameter combinations were used to construct an informative lognormal prior for stock depletion that could be applied in the BSPM in 1988 (Figure 6).

3.2.4 Fishing Mortality Prior

An extension of the prior pushforward analysis described in Section 3.2.2 was applied to develop a prior for the variability in fishing mortality σ_F . While in Section 3.2.2 catch was subtracted directly, random values of $\sigma_F \text{Normal}^+(0, 0.1)$ were used to develop time series of fishing mortality values. Based on the random fishing mortality values, catch was calculated and removed from the population in the same way as described in Section 3.1.4. Based on the results of the prior pushforward, the σ_F distribution was further filtered (Figure 7) based on the same criteria used in Section 3.2.2 with the additional criteria that:

- Depletion in the terminal year $> 2\%$, $R_{Max} < 1$, trend in depletion over the last decade between -5% and 10%, and maximum simulated catch was less than 350,000.

This last filtering step was done to ensure that observed catch values fell within the 95th percentile of the simulated catch, and with a median simulated catch that approximated the magnitude of observed catches (Figure 8). Fitting a Normal^+ distribution to the remaining σ_F values results in the prior distribution specified in Table 2 and shown in (Figure 9). While all parameters (e.g., R_{Max} and $\log K$) were subject to this new filtering step, only the σ_F distribution showed meaningful change. Therefore, when setting up the 0003-2024cpueFPrior model, the prior distributions for the non- σ_F parameters were assumed to be the same as the other two models.

3.3 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).

3.4 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.

3.4.1 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).

3.4.2 Data Fits

Model fit to the abundance index was 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.

3.4.3 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.

3.4.4 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.

3.4.5 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 3.4.4.

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, $I_{2017} = \check{I}_{2018}$. The absolute scaled error (ASE) of the prediction is:

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

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.

3.5 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 Model Convergence and Diagnostics

All three BSPM model variants achieved satisfactory convergence based on standard Hamiltonian Monte Carlo diagnostics (Table 3). 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.

4.1.1 Model Fit to Data

4.1.1.1 Catch Data Fit

All models fitted the observed catch time series (1952-2022) with a coefficient of variation of 0.2 to account for uncertainty in catch estimates. Since fishing mortality required to generate the catch was directly estimated, model fit to the catch was good (Figure 10). The models consistently predicted lower than observed catch for the large catch event in 1954.

4.1.1.2 CPUE Index Fit

Model fit to the standardized CPUE index (1988-2022) was evaluated using normalized root-mean-squared error (NRMSE). All three models provided reasonable fits to the observed index given the high variability and observation error (Table 3). The `0003-2024cpueFPrior` model showed marginally improved fit with lower RMSE values. All models successfully captured the general declining trend in standardized CPUE from 1988-2022. However, the estimated combination of observation and process error meant models were not tightly constrained to individual observations, resulting in persistent negative residuals in the last decade (Figure 11). Posterior predictive checks confirmed that models replicated key features of the observed CPUE data (Figure 2).

4.1.2 Model Validation

Overall retrospective bias was very low and close to zero for all models (Table 3; Figure 12), indicating lack of persistent bias in terminal estimates as successive years of index data were removed. Estimates of relative exploitation rate (U/U_{MSY}) and relative depletion (D/D_{MSY}) fell within the credible intervals of the full model run 100% of the time (Table 3). For hindcast cross-validation, all models exceeded the performance of the naïve predictor, with the `0003-2024cpueFPrior` model showing the best performance (Table 3; Figure 13). This provides support that models successfully identified and estimated a production function for the stock.

4.2 Population Dynamics and Stock Status

In all models, most leading parameters ($\log(K)$, R_{Max} , σ_P , n , σ_F , and $\sigma_{O,add}$) showed posterior updates from their respective prior distributions (Figure 14), indicating sufficient information in the data to inform parameter estimates. The shape parameter n showed little deviation from the Schaefer-model prior. All models estimated similar levels of σ_P and additional observation error $\sigma_{O,add}$ (Table 4), which is unsurprising given they fitted the same catch and standardized index. Although estimates of σ_P were elevated from the prior, models primarily reconciled index variability by estimating substantial additional observation error $\sigma_{O,add}$, resulting in approximately a 3:1 ratio of total observation error σ_O to process error σ_P . Similar to findings in ISC (2024), there was a trade-off between $\log(K)$ and σ_F where higher population scale was associated with lower fishing mortality and vice versa.

All models showed substantial posterior updates for the R_{Max} parameter, estimating productivity on the lower end of the biological prior. These median estimates of R_{Max} , between 0.2-0.3 (Table 4), are consistent with estimates for 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 (95% credible interval: 0.122-0.215), which falls within the posterior estimates of all three models (Table 4). 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). In the current analysis, the `0003-2024cpueFPrior` model estimated lower productivity R_{Max} with reduced uncertainty (Table 4), consistent with assuming a more depleted stock. Inverse transform sampling to match the R_{Max} posterior distribution from this model reveals that lower productivity is characterized by pronounced shifts toward lower steepness values h and, to a lesser extent, smaller von Bertalanffy k values (Figure 15).

The estimated population depletion time series shows a decline from the assumed unfished state in 1952 to minimum levels around the early 1990s (Figure 16). All three models estimated similar depletion trajectories, with differences primarily related to estimated population scale. The `0003-2024cpueFPrior` model estimated slightly lower depletion levels throughout the time series compared to the other models. From the minimum depletion period, all models estimated gradual recovery through the 2000s and 2010s, with populations stabilizing at moderate depletion levels in recent years. Credible intervals around depletion estimates are relatively wide, particularly in the early period when no abundance index data were available to inform the model, reflecting inherent uncertainty in reconstructing historical population dynamics. The estimated population time series mirrors these depletion patterns. The `0001-2024cpueExPrior` model estimated the highest population scale throughout the time series, while the `0002-2024cpueDepPrior` and `0003-2024cpueFPrior` models estimated lower population levels. While uncertainty in absolute population size is substantial, as reflected in wide credible intervals, relative trends are more precisely estimated.

Posterior estimates of depletion in 1988 differed notably among the three models (Table 4). For the `0002-2024cpueDepPrior` and `0003-2024cpueFPrior` models, these estimates differed from the prior value of $x_{1988} = 0.637$ (Table 2). The baseline model (`0001-2024cpueExPrior`) estimated median depletion of $x_{1988} = 0.87$ (0.59-1.06), while the depletion-constrained model (`0002-2024cpueDepPrior`) estimated $x_{1988} = 0.75$ (0.56-0.94). The `0003-2024cpueFPrior` model, with a more representative prior on fishing mortality variability, estimated the greatest depletion levels in 1988: $x_{1988} = 0.67$ (0.50-0.89).

Process error estimates (Figure 16) showed consistent temporal variability across models, with the largest

deviations occurring during periods of rapid population change in the 1970s and 1980s. This indicates that removals alone could not explain the observed changes in the standardized index. The estimated process error variability σ_P (Table 4), while similar across all models, was lowest for the `0003-2024cpueFPrior` model, indicating greater internal consistency in the model dynamics.

4.3 Biological Reference Points

MSY-based reference points were derived from the Fletcher-Schaefer production function for all three BSPM variants. The relative stock status time series (Figure 16) shows consistent trends across all models, albeit at different relative levels. Median estimates of the D/D_{MSY} ratio indicate the stock was above the MSY reference level for the entire model period. Only the `0003-2024cpueFPrior` model indicates risk of falling below the MSY-based reference point in recent years. Similarly, median estimates of the F/F_{MSY} ratio were consistently below the overfishing reference point, with episodic periods of overfishing risk indicated by the `0003-2024cpueFPrior` model. However, recent trends indicate declining overfishing risk, consistent with estimated stock recovery.

4.4 Future Projections

Ten-year stochastic projections (2023-2032) were conducted for all three BSPM variants under a status quo catch scenario, with projected removals set to the average catch over 2018-2022 (Figure 17). Under status quo catch levels, all models projected continued population recovery through 2032. Depletion trajectories show gradual increases from current levels, with the `0001-2024cpueExPrior` model projecting the most optimistic recovery due to higher estimated population scale. The `0002-2024cpueDepPrior` and `0003-2024cpueFPrior` models showed more modest but consistent increases in depletion over the projection period.

Relative to MSY-based reference points, projections indicate the stock will likely remain above D_{MSY} throughout the projection period for the baseline and depletion-constrained models. The `0003-2024cpueFPrior` model showed greater variability and some risk of declining below D_{MSY} in the near term, though the median trajectory suggests stabilization above the reference point by the end of the projection period. Projected fishing mortality rates remained well below F_{MSY} for all models under the status quo catch scenario, indicating that current harvest levels are likely sustainable and consistent with continued stock recovery. Projection uncertainty increased over time, reflecting both parameter uncertainty and stochastic variability in future population dynamics.

5 Discussion

5.1 General remarks

The three BSPM variants provide a consistent picture of stock trajectory and current status. All models estimated similar depletion patterns: decline from unfished conditions in 1952 to minimum levels around the early 1990s, followed by gradual recovery through recent decades. This trajectory suggests the stock is currently recovering. While median estimates indicate the stock was not overfished or undergoing overfishing during the assessment period, the `0003-2024cpueFPrior` model suggests some recent risk of falling below these thresholds.

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. When these assumptions are uncertain or mis-specified, population scale estimates informed by size composition data can be biased. Production models 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.

5.2 Challenges, limitations & key uncertainties

Several important limitations affect this analysis. The assumption of no age-specific population dynamics may be problematic when mortality or reproductive potential varies significantly by age. This simplification prevents incorporation of biological information beyond production function parameters and excludes explicit use of size composition data.

The aggregated nature of surplus production models can potentially bias MSY estimates when age-specific processes differ from production function assumptions. For species with complex life histories like striped marlin, managers should consider these limitations when applying MSY-based reference points derived from surplus production models.

Data limitations also constrain the analysis. Fitting to a single standardized abundance index limits evaluation of uncertainty in abundance trends. Multiple indices would enable model ensemble approaches to better characterize this uncertainty.

The incorporation of depletion information from New Zealand recreational size data through a 1988 depletion prior, while novel, represents a simplified approach. More sophisticated modeling frameworks like delay-difference or age-structured models could explicitly fit size composition time series in a likelihood-based framework, providing more elegant integration of this information.

Modeling assumptions introduce additional uncertainties. Current modeling assumes the same level of catch uncertainty across all time periods. Future versions could allow time-varying catch uncertainty (e.g., greater uncertainty in historical versus recent observations) or accommodate effort-driven removals similar to catch-errors with effort deviation approaches available in MULTIFAN-CL (Fournier et al., 1998).

Population scale estimates were highly uncertain and sensitive to model configuration choices, while depletion estimates were more precisely estimated. The choice of prior on σ_F was particularly influential. Although this prior was developed using prior pushforward analysis (Kim & Neubauer, 2025), it represents a key uncertainty. Similarly, uncertainty in biological parameters (growth, maturity, natural mortality) due to limited sample sizes contributed to broad initial priors for R_{Max} . While the posterior distribution for R_{Max} was substantially updated from the prior, reducing uncertainty in key biological productivity parameters could help reduce model uncertainty estimates and facilitate development of age-structured models.

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, over-simplification 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.

6 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.

7 References

- 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., Day, J., Teears, T., Davies, N., Hampton, J., McKechnie, S., Magnusson, A., Peatman, T., Vidal, T., Williams, P., & Hamer, P. (2025). *Revised Stock Assessment of Striped Marlin in the Southwest Pacific Ocean: 2025* (WCPFC-SC21-2025/SA-WP-XX).
- 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, 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>
- 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/rssc.12378>
- Gedamke, T., & Hoenig, J. M. (2006). Estimating Mortality from Mean Length Data in Nonequilibrium

- Situations, with Application to the Assessment of Goosefish. *Transactions of the American Fisheries Society*, 135(2), 476–487. <https://doi.org/10.1577/T05-153.1>
- 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>
- Hordyk, A., Huynh, Q., & Carruthers, T. (2024). *openMSE: Easily Install and Load the 'openMSE' Packages* (R package version 1.0.1). <https://openmse.com/>
- 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*, 78(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>
- 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 misspecifications 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>
- 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., 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>
- 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*.
- 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/>
- 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. (2020). Unifying parameterizations between age-structured and surplus production models: An application to atlantic white marlin (*Kajiki albida*) with simulation testing. *Collect. Vol. Sci. Pap. ICCAT*, 76(4), 219–234.

DRAFT

8 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	$\rho = -0.3$ with M_{ref}	(Farley et al., 2021)
Reference mortality	M_{ref}	Lognormal	0.36 yr^{-1}	0.44	$\rho = -0.3$ 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	$\rho = -0.5$ with b_w	(Castillo-Jordan et al., 2024)
Weight exponent	b_w	Normal	3.58	0.05	$\rho = -0.5$ 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
Length at first capture	L_c	Lognormal	140 cm	0.125	-	Fishery selectivity

Parameter	Symbol	Distribution	Mean/Mode	CV/Range	Correlation	Source/Notes
Reference ages	age_1, age_2	Fixed	0, 10 years	-	-	Francis parameterization

Table 2: Prior distributions for Bayesian surplus production model parameters. All models assumed the same priors unless explicitly noted.

Parameter	Distribution	Mean	SD	Description
Core Population Parameters				
K	Lognormal	$\log(1,049,036)$	0.46	Carrying capacity
R_{Max}	Lognormal	$\log(0.5099)$	0.46	Maximum intrinsic rate of increase (from biological simulation)
n	Lognormal	$\log(2)$	0.1	Shape parameter (Pella-Tomlinson production function); $n = 2$ is a Schaefer model
Process and Observation Error				
σ_P	Lognormal	$\log(0.0533)$	0.27	Process error standard deviation (Winker et al., 2018)
$\sigma_{O,add}$	Half-Normal	0	0.2	Additional estimated observation error
σ_F	Half-Normal	0	0.025	Fishing mortality variability (0001-2024cpueExPrior & 0002-2024cpueDepPrior)
σ_F	Half-Normal	0	0.047	Fishing mortality variability (0003-2024cpueFPrior)
Depletion Constraint				
(0002-2024cpueDepPrior & 0003-2024cpueFPrior)				
x_{1988}	Lognormal	$\log(0.6374)$	0.2	Initial depletion at $t = 37$ (this analysis)

Table 3: Model summary

Model	Max \hat{R}	Min ESS	Divergent	Tree Depth	BFMI Issues	Index RMSE	Mohn's ρ	Coverage D/D_{MSY}	Coverage U/U_{MSY}
0001	1.009	724	0	0	0	0.28	0.006	0.966	100%
0002	1.009	674	0	0	0	0.27	-0.005	0.863	100%
0003	1.008	728	0	0	0	0.265	-0.026	0.779	100%

Notes:

0001 corresponds to the 0001-2024cpueExPrior model

0002 corresponds to the 0002-2024cpueDepPrior model

0003 corresponds to the 0003-2024cpueFPrior model

- 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
- Index RMSE: Root mean squared error for fit to standardized CPUE index
- 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 4: Posterior estimates with 95% credible intervals for all BSPM parameters

Model	$\log(K)$	R_{Max}	σ_P	Shape (n)	σ_F	$\sigma_{O,add}$	σ_O	x_{1988}
0001	13.8 (13.2-14.6)	0.27 (0.13-0.76)	0.07 (0.04-0.11)	1.97 (1.62-2.37)	0.03 (0.01-0.07)	0.10 (0.04-0.19)	0.27 (0.20-0.35)	0.87 (0.59-1.06)
0002	13.7 (13.2-14.5)	0.23 (0.12-0.56)	0.07 (0.04-0.12)	1.96 (1.62-2.35)	0.04 (0.02-0.08)	0.09 (0.03-0.17)	0.26 (0.20-0.34)	0.75 (0.56-0.94)
0003	13.3 (12.8-14.0)	0.23 (0.13-0.45)	0.06 (0.03-0.10)	1.97 (1.65-2.38)	0.07 (0.03-0.13)	0.09 (0.03-0.16)	0.26 (0.19-0.33)	0.67 (0.50-0.89)

Parameter Definitions:

0001 corresponds to the 0001-2024cpueExPrior model

0002 corresponds to the 0002-2024cpueDepPrior model

0003 corresponds to the 0003-2024cpueFPrior model

- $\log(K)$: Natural logarithm of carrying capacity (total numbers)
- R_{Max} : Maximum intrinsic rate of population increase (per year)
- σ_P : Process error standard deviation
- Shape (n): Pella-Tomlinson production function shape parameter
- σ_F : Fishing mortality variability standard deviation
- $\sigma_{O,add}$: Extra estimated observation error component
- σ_O : Total observation error (input + additional)
- x_{1988} : Population depletion in 1988 (year 37 of model period)

9 Figures

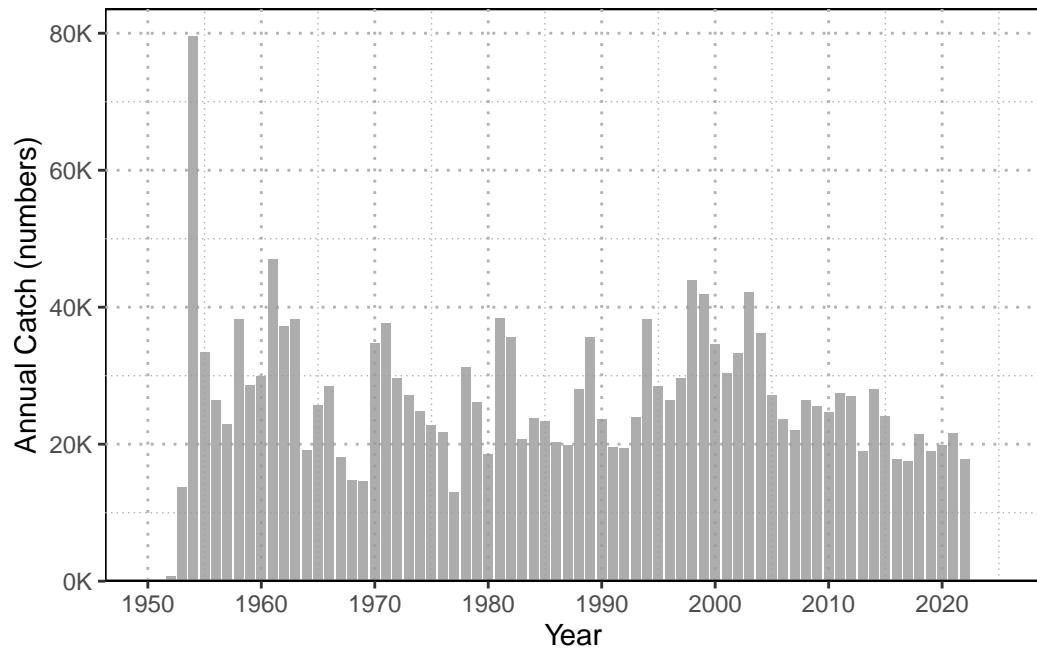


Figure 1: Annual catch of Southwest Pacific Ocean striped marlin (1952-2022). Catch data shows initial low removals in early years, a peak in 1954, followed by relatively stable catches with a slight decline in recent decades.

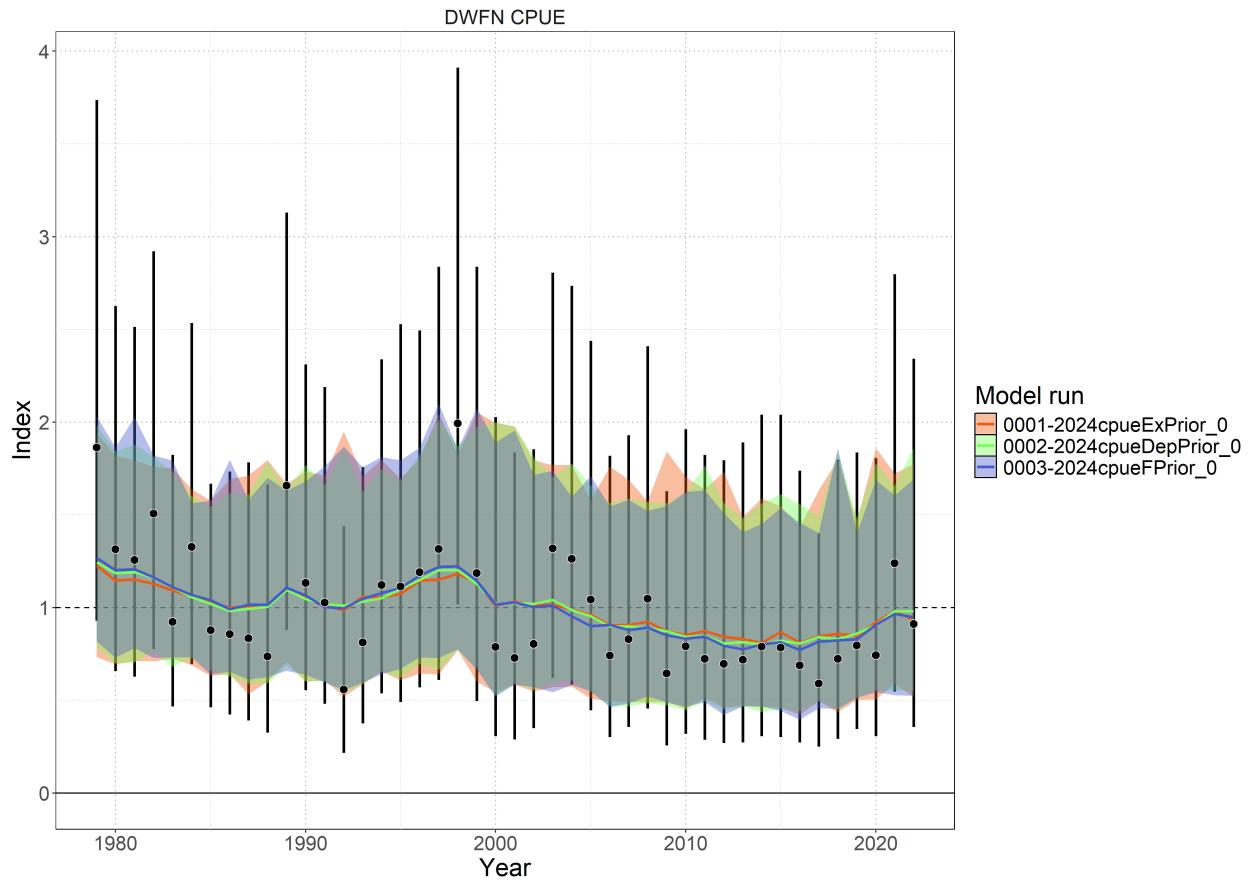


Figure 2: Standardized index (black points), observation error (black bars), and posterior predicted model fits (colored lines) with associated 95% credible interval (colored shading).

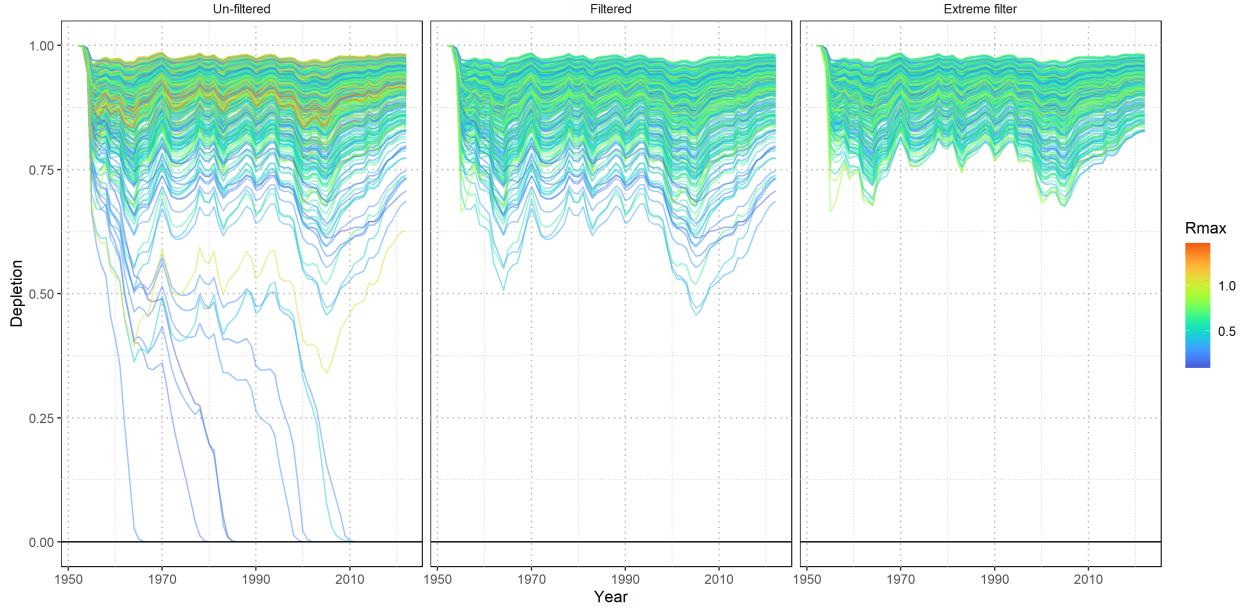


Figure 3: Prior pushforward check showing simulated population depletion trajectories under different filtering scenarios. Lines represent population trajectories colored by maximum intrinsic rate of increase (R_{Max}). Panels show progressively stricter filtering criteria from left to right.

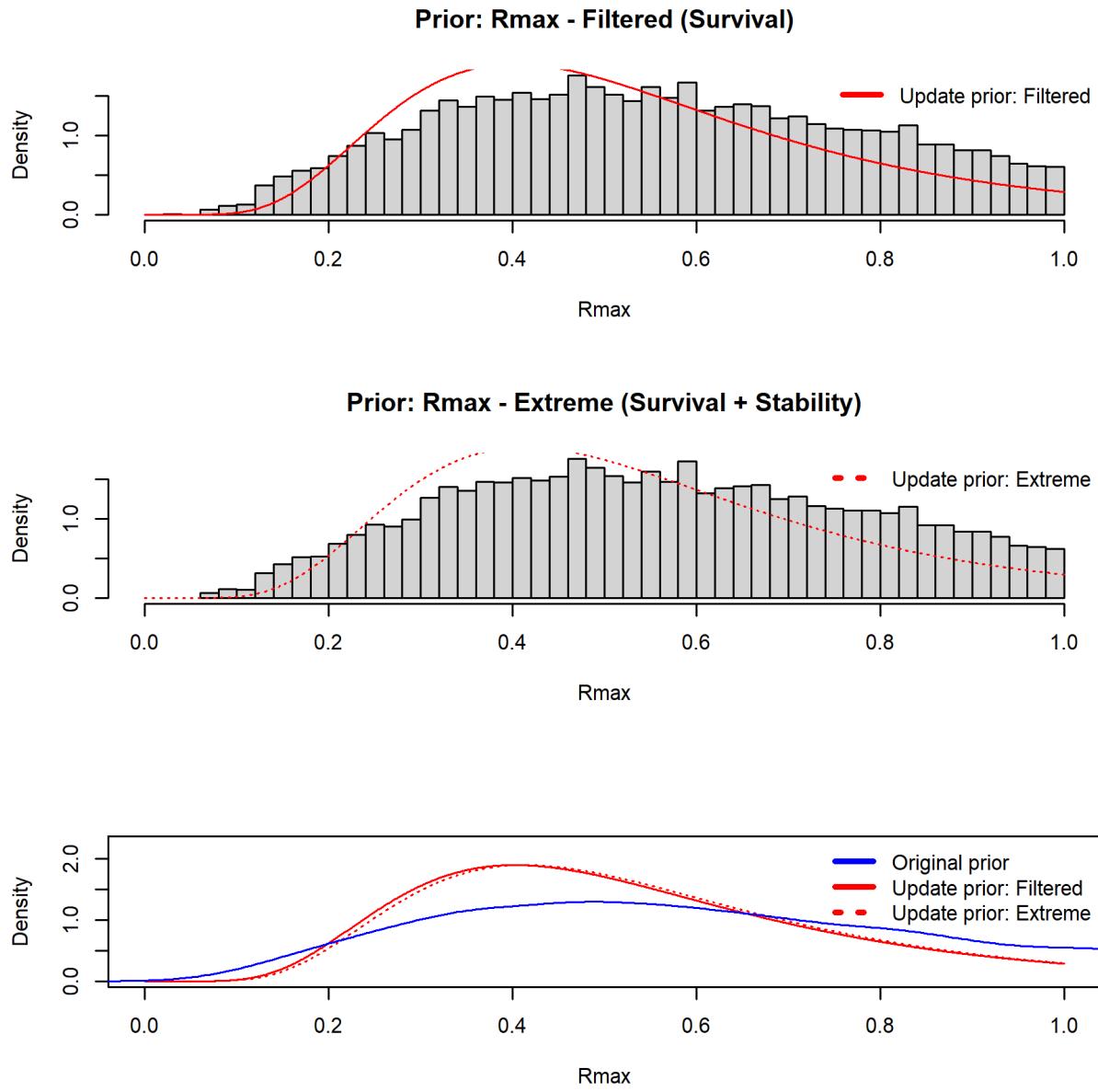


Figure 4: Prior distributions for maximum intrinsic rate of population increase (R_{Max}). Upper panel shows baseline filtered values (gray histogram) with fitted lognormal distribution (red line). Middle panel shows extreme filtered values with fitted distribution (dotted red line). Lower panel compares original simulation values (gray), viable populations (blue), and final lognormal priors (red).

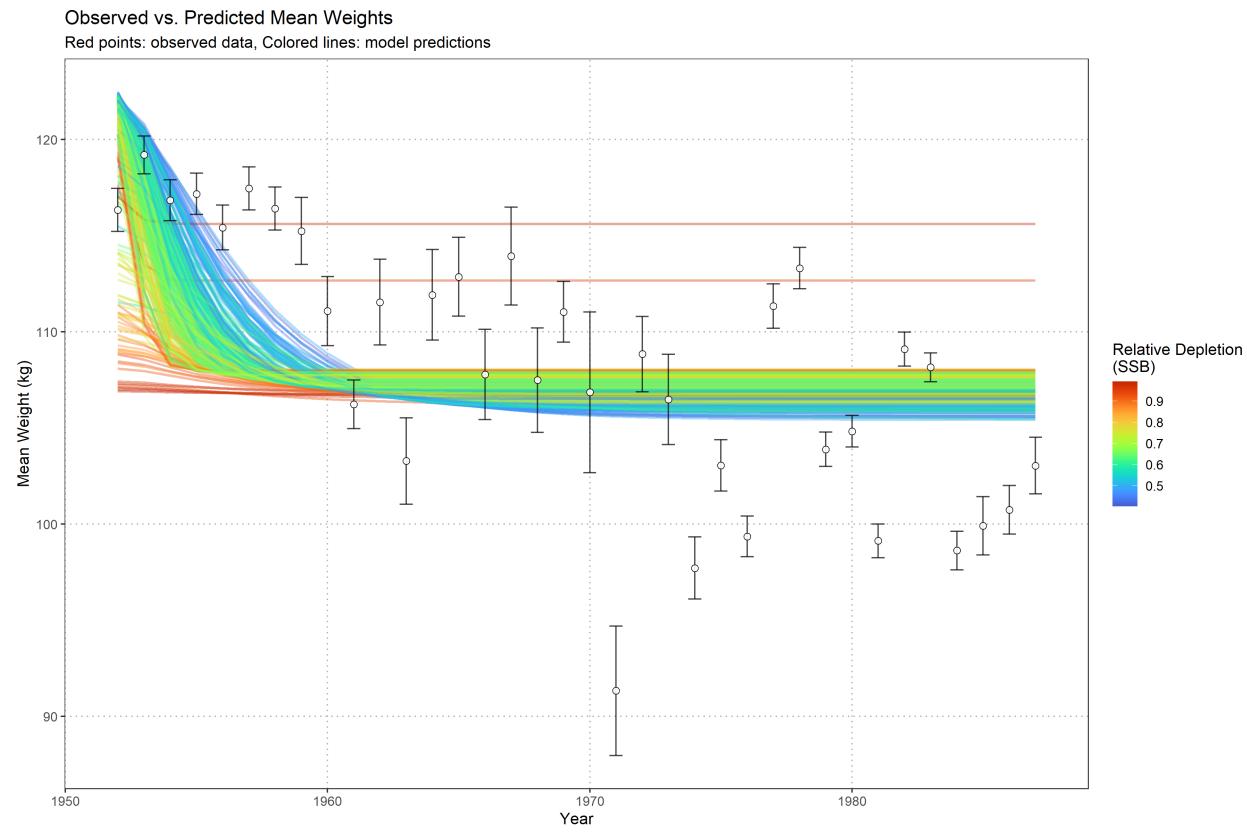


Figure 5: Observed versus predicted mean weights from New Zealand recreational fishery data. Points show observed mean weights with error bars, colored lines show model predictions from the transitional mean weight model colored by relative spawning stock biomass (SSB) depletion.

Prior Distribution: Relative Depletion (SSB)

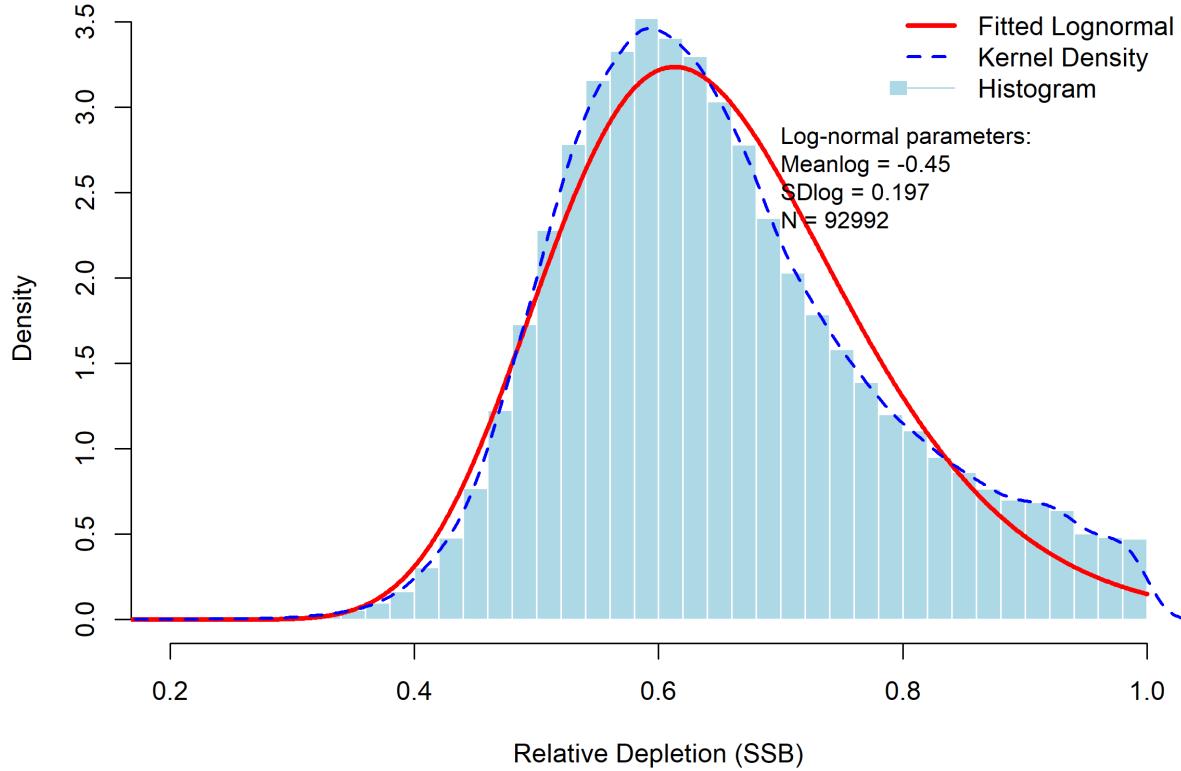


Figure 6: Prior distribution for relative depletion in 1988 derived from New Zealand recreational weight data. Light blue histogram shows depletion estimates from biological parameter combinations that successfully fit the transitional mean weight model. Blue dashed line shows kernel density estimate, red line shows fitted lognormal distribution used as prior in the depletion-constrained BSPM.

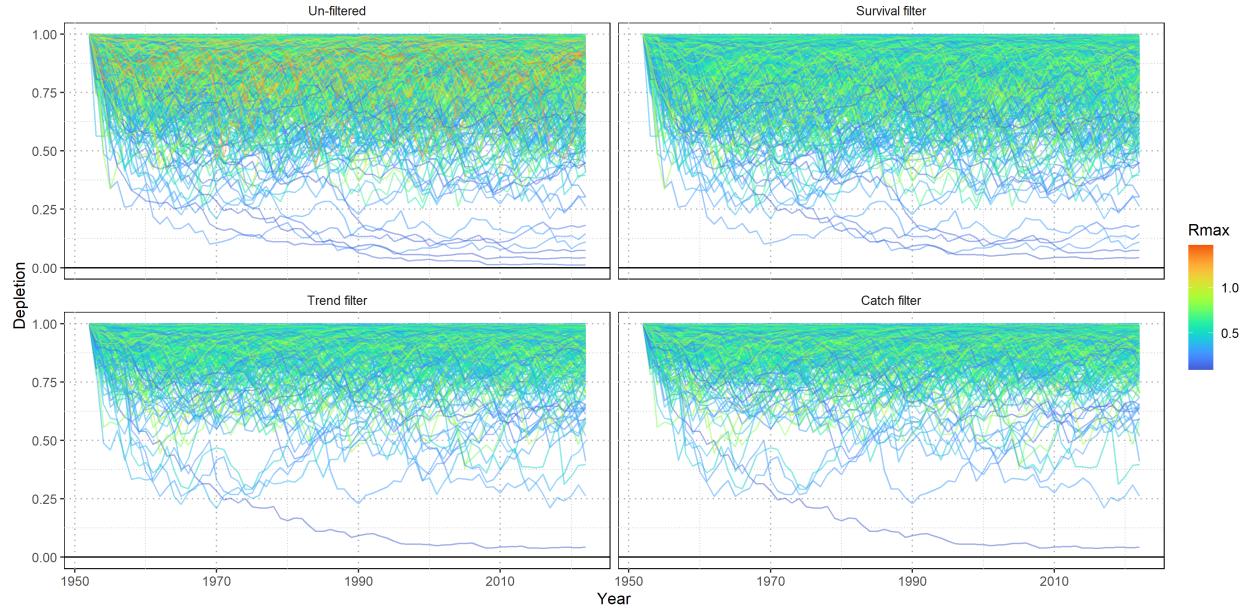


Figure 7: Prior pushforward check for population depletion trajectories under different filtering scenarios. Each line represents a simulated population trajectory colored by fishing mortality variability (σ_F).

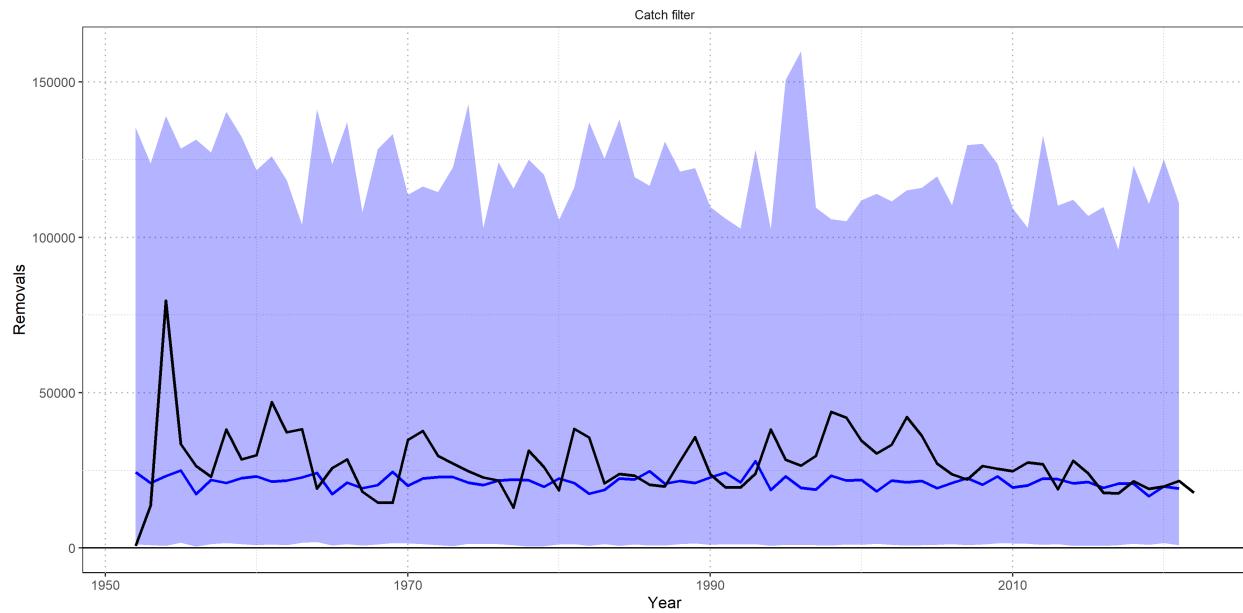


Figure 8: Simulated removals from the fishing mortality variability (σ_F) prior pushforward analysis. Solid blue line indicates the median simulated removals with 95th percentile (shaded region). Observed removals are shown by the solid black line.

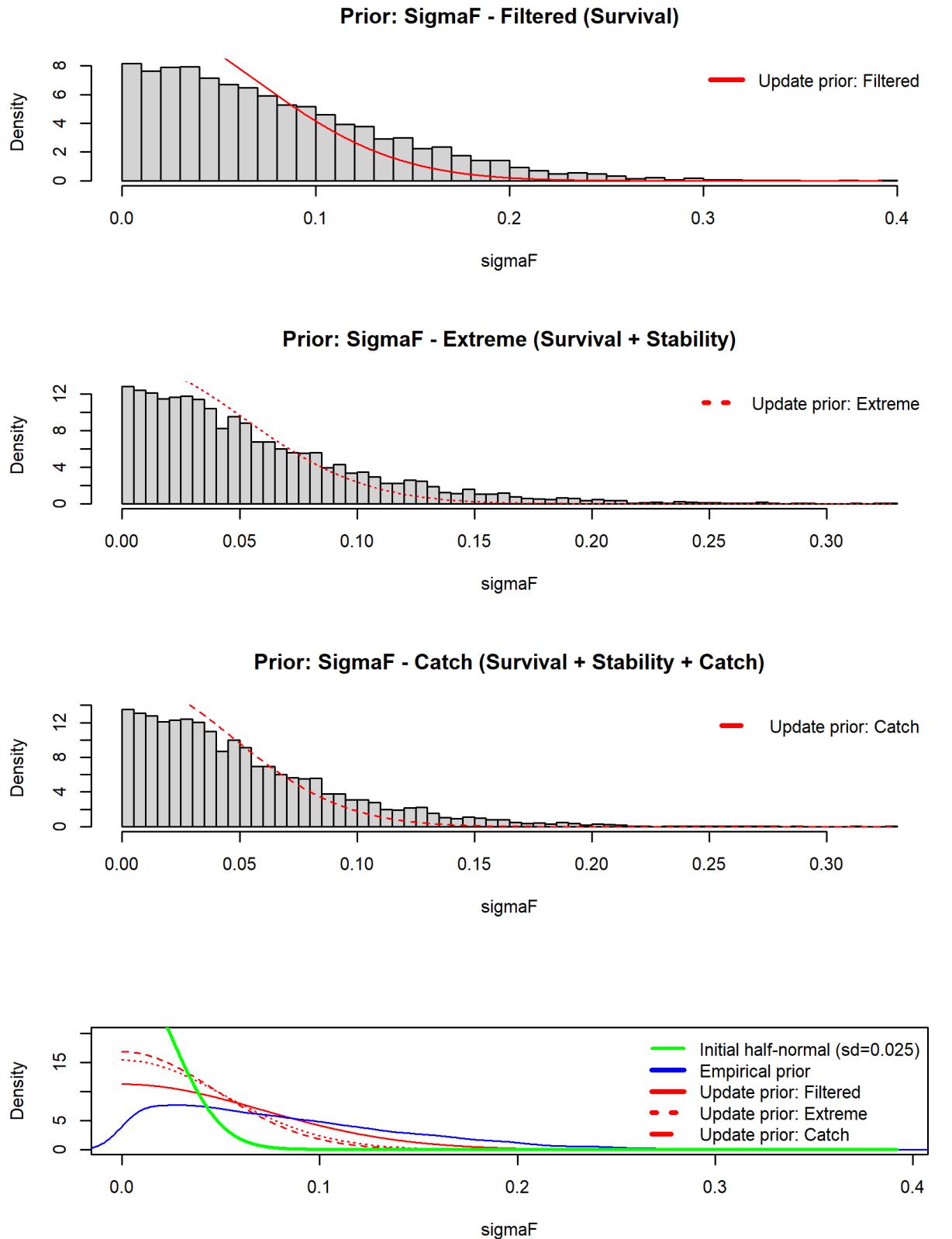


Figure 9: Prior distributions for fishing mortality variability (σ_F) under different filtering scenarios. Top three panels show progressively stricter filtering (baseline, extreme, catch-based) with fitted distributions (red lines). Bottom panel compares all distributions: original simulation (gray), viable populations (blue), final priors (red), and prior used in models 0001 and 0002 (green).

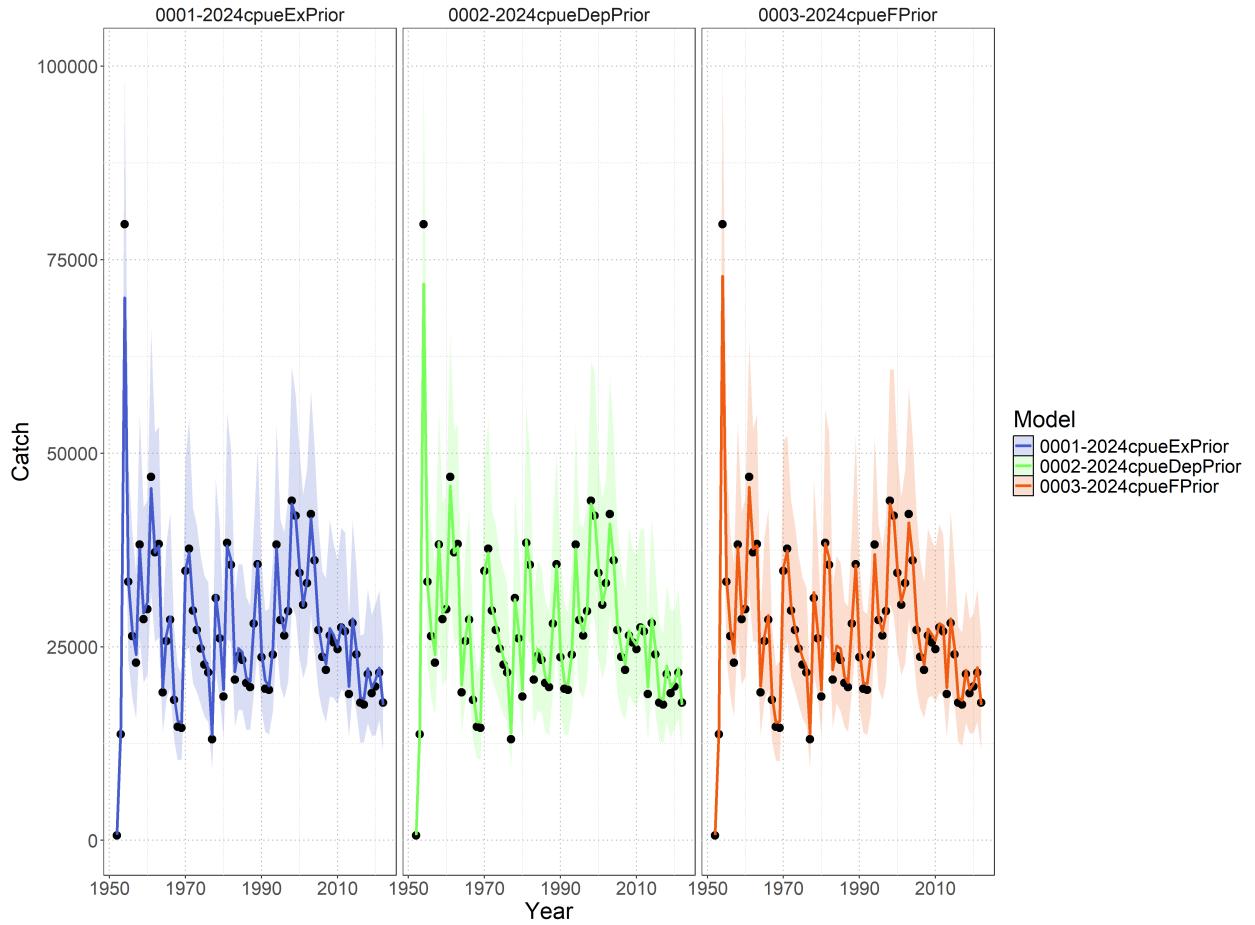


Figure 10: Time series of observed catch (black dots with uncertainty bounds) and model-predicted catch for three Bayesian surplus production models fitted to SWPO striped marlin data from 1950-2022. The left panel shows model 0001-2024cpueExPrior (baseline model with expert priors), the central panel shows model 0002-2024cpueDepPrior (depletion-constrained model incorporating 1988 depletion prior from New Zealand recreational weight data), the right panel shows model 0003-2024cpueFPrior which assumes a less restrictive prior on fishing mortality variability. Colored lines represent posterior median predictions with shaded uncertainty bands showing 95% credible intervals.

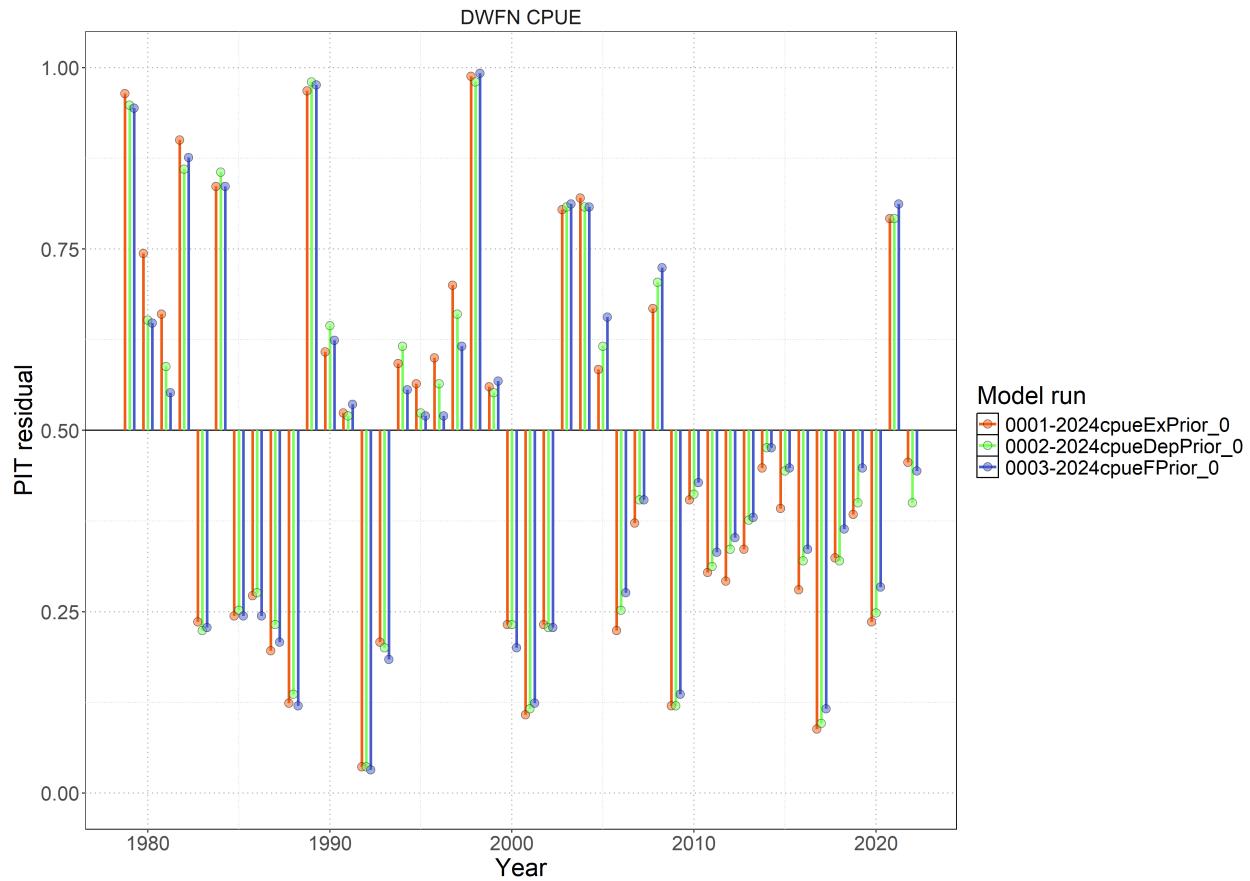


Figure 11: Probability-integral transform (PIT) style residuals from each model (colored points) fit to the standardized index.

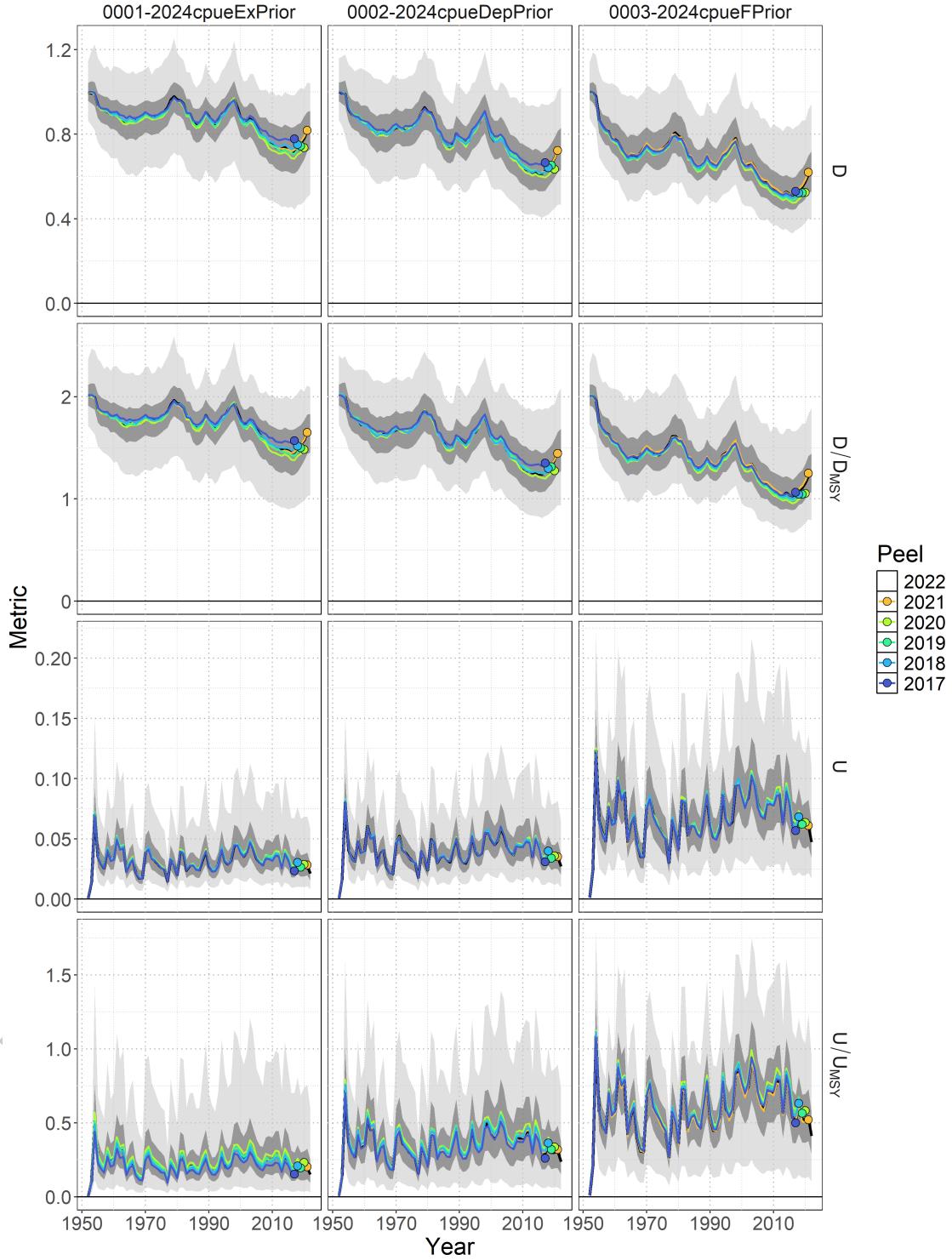


Figure 12: Retrospective analysis for all models with respect to time series of depletion D_t , depletion relative to depletion at MSY D_t/D_{MSY} , exploitation rate U_t , and exploitation rate relative to the rate of exploitation that produces MSY U_t/U_{MSY} . The base model with data included through 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.

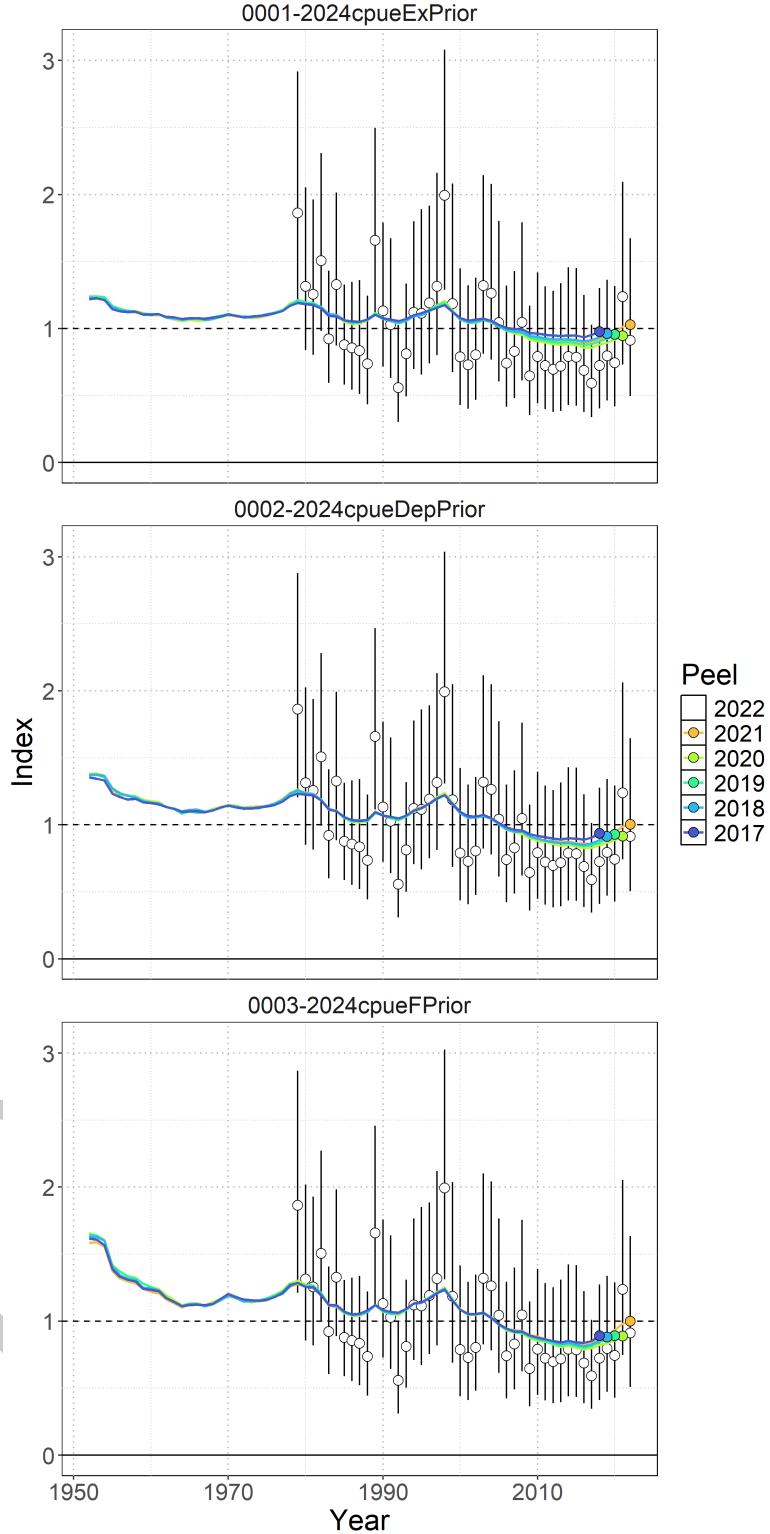


Figure 13: Standardized indices of relative abundance used in the Bayesian Surplus Production Model. Open circles show observed values (standardized to mean of 1; black horizontal line) and the vertical bars indicate the observation error (95% confidence interval). One year ahead ‘model-free’ hindcast 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.

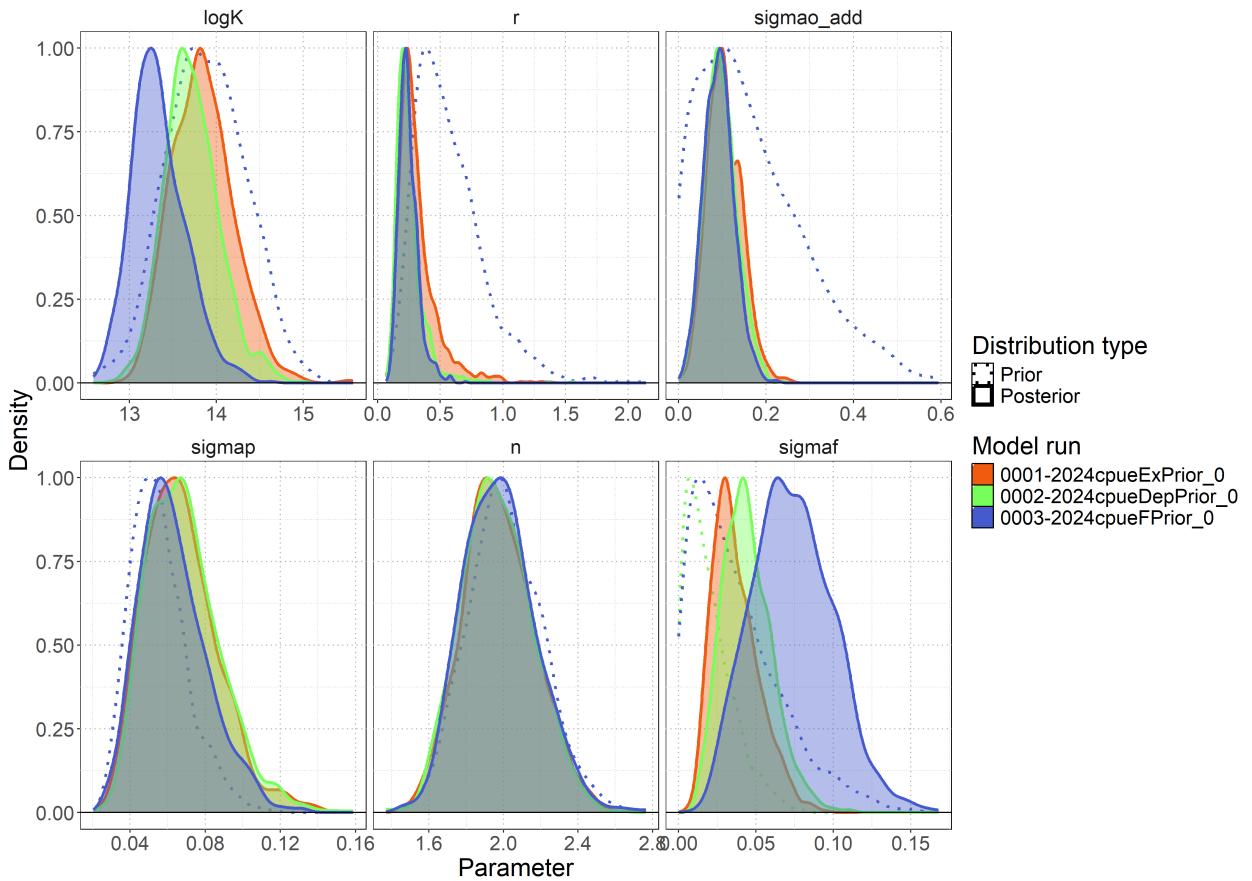


Figure 14: Posterior parameter distributions (solid lines) for leading parameters ($\log(K)$, R_{Max} , σ_P , shape (n), σ_F , $\sigma_{O,add}$, and $\sigma_{O,sc}$), relative to their assumed prior distributions (dotted lines) for all models.

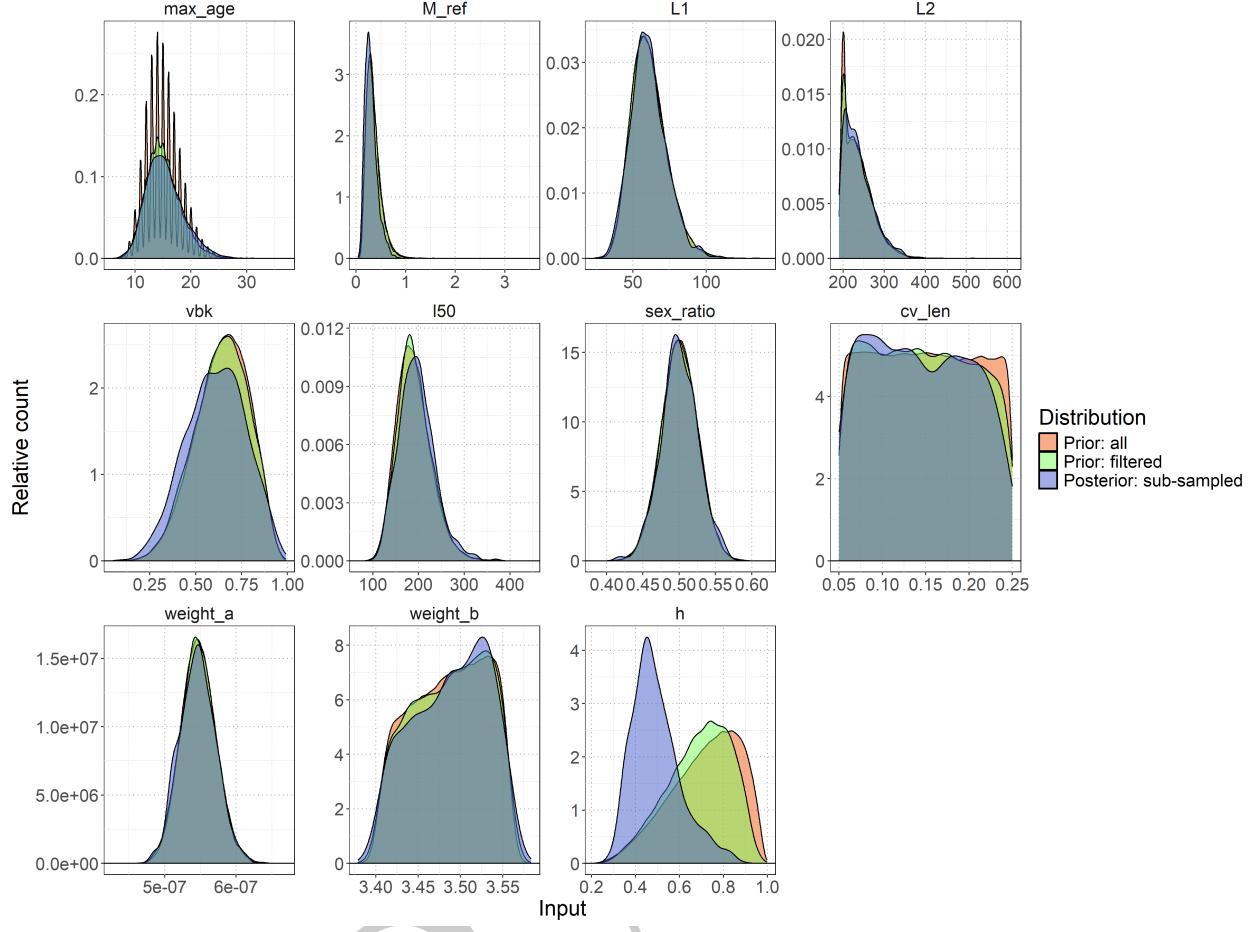


Figure 15: Distributions of biological parameters (maximum age max_age , natural mortality reference M_ref , length at birth $L1$, asymptotic length $L2$, von Bertalanffy growth coefficient vbk , length at 50% maturity $l50$, sex ratio at birth sex_ratio , coefficient of variation in length cv_len , weight-length relationship parameters $weight_a$ and $weight_b$, and steepness h) used in the rmax sub-sampling process. Original biological prior distribution (orange shading), distribution based on extreme filtering criteria following the prior pushforward analysis (green shading), and final sub-sampled distribution selected to match the posterior R_{Max} distribution (blue shading).

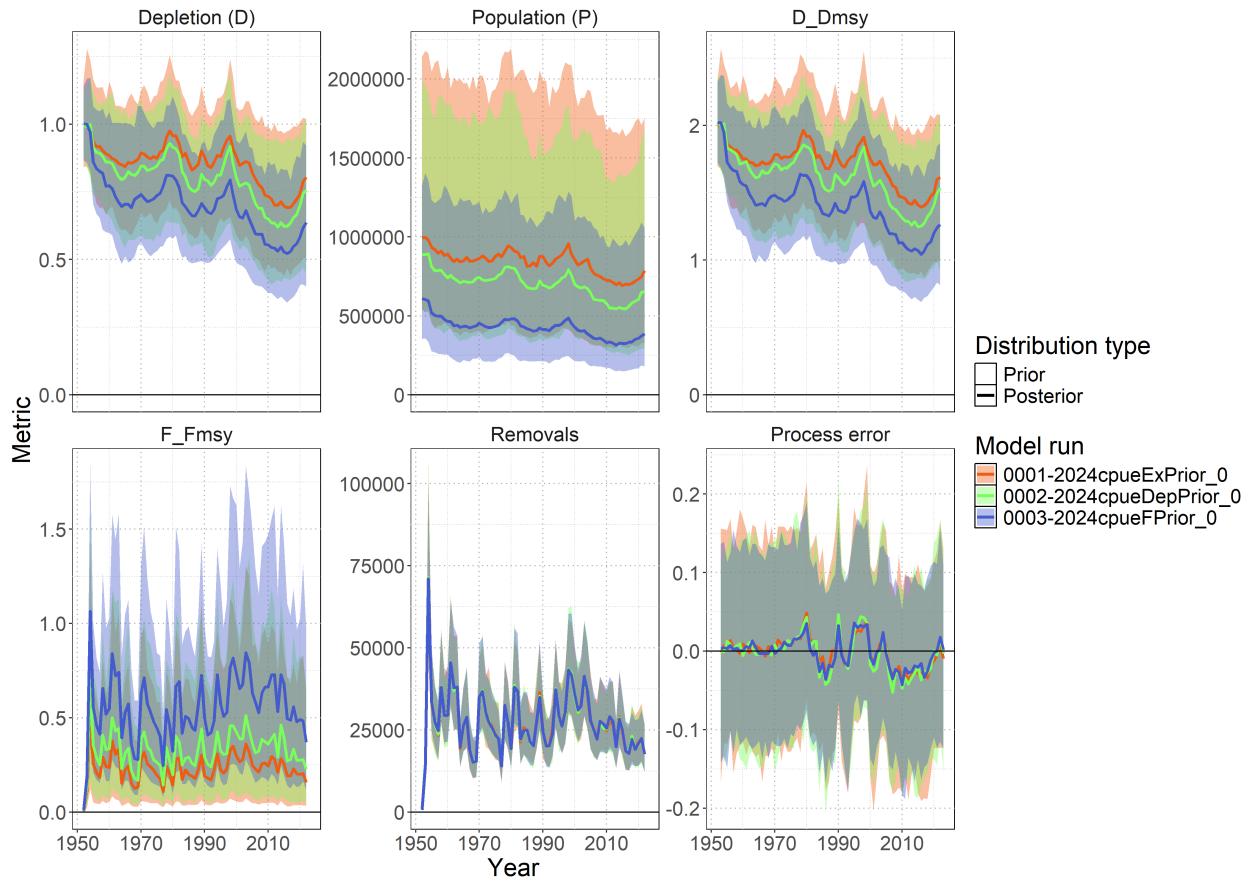


Figure 16: Posterior time series estimates (1952-2022) for key population metrics across all three BSPM models. Solid lines show median estimates with 95% credible intervals (shading). Metrics include depletion (D), population (P), relative depletion (D/D_{MSY}), relative fishing mortality (F/F_{MSY}), removals, and process error.

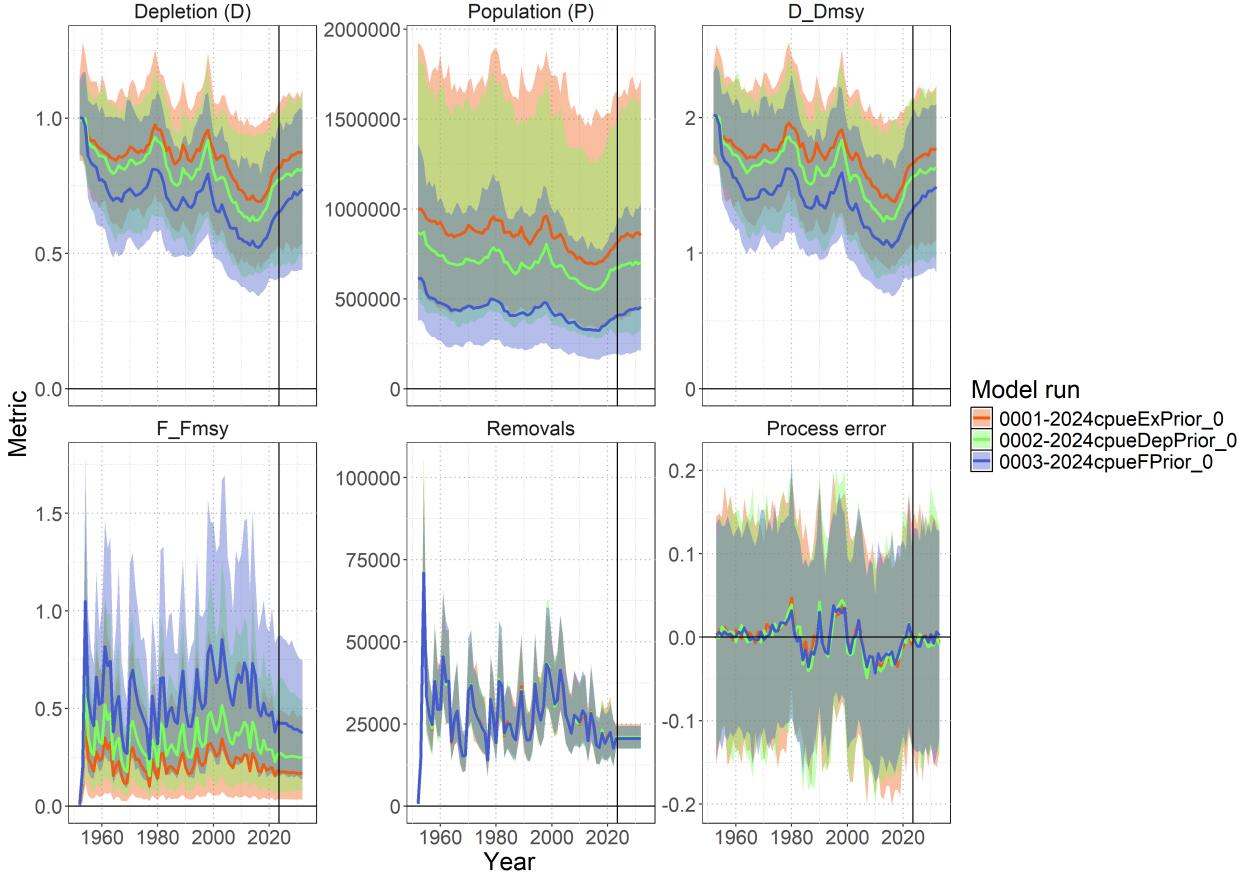


Figure 17: Ten-year stochastic projections (2023-2032) under status quo catch scenario. Colored lines show median projections with 95% credible intervals for all three models. Projections assume average 2018-2022 catch levels with resampled process error. Black vertical line indicates start of projection period. Metrics include depletion (D), population (P), relative depletion (D/D_{MSY}), relative fishing mortality (F/F_{MSY}), removals, and process error.

10 Technical Annex

Additional technical detail of population dynamics components used to solve the Euler-Lotka equation for R_{Max} and implement the Gedamke-Hoenig approach (Gedamke & Hoenig, 2006) for using size composition data to estimate mortality in non-equilibrium situations.

10.1 Equilibrium Age-Structured Population Dynamics

10.1.1 Survival

Survival schedule was calculated assuming constant natural mortality:

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

10.1.2 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:

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

10.1.3 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 (Figure 3). The resulting distribution was fitted to a lognormal prior for use in the BSPM (Figure 4).

10.2 Transitional Mean Weight Model

For each biological parameter combination, total mortality rates Z_1 and Z_2 were estimated by fitting a transitional mean weight model to observed temporal changes in recreational catch weights. The expected mean weight during the transition follows:

$$\bar{W}(d) = W_\infty - \frac{Z_1 Z_2 (W_\infty - W_c) [Z_1 + k + (Z_2 - Z_1) e^{-(Z_2+k)d}]}{(Z_1 + k)(Z_2 + k) [Z_1 + (Z_2 - Z_1) e^{-Z_2 d}]}$$

where: - d = time since mortality change (years after 1952) - W_∞ = asymptotic weight from growth parameters
- W_c = weight at first capture (corresponding to L_c) - k = von Bertalanffy growth coefficient

The mortality parameters were estimated in a likelihood framework by maximizing the likelihood of the observed mean weights given a time series of predicted mean weights $\bar{W}(d)$. Optimization was constrained such that $Z_1, Z_2 \geq M_{ref}$ to ensure biologically realistic mortality estimates.