



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
05 June 2025

N. Ducharme-Barth¹

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

Table of contents

1	Introduction	3
2	Methods	4
2.1	Model Framework	4
2.1.1	Population Dynamics	4
2.1.2	Observation Model	4
2.1.3	Catch Treatment	4
2.2	Prior Development	4
2.2.1	Biological Simulation Framework	4
2.2.2	Maximum Intrinsic Rate of Increase (R_{max}) Calculation	5
2.2.3	Depletion Prior	6
2.3	Model Implementation	7
2.3.1	Computational Setup	7
2.3.2	Model Configurations	7
2.3.3	Model Diagnostics	8
2.4	Uncertainty Characterization	8
3	Declaration of Generative AI use	8
4	References	8
5	Tables	10

1 Introduction

Assessment of Southwest Pacific Ocean (SWPO) striped marlin (*Kajikia audax*) has 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 (Castillo-Jordan et al., 2024).

Discussions at the 2025 SPC Pre-Assessment Workshop also highlighted challenges related to the estimation of total population scale, where 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 additional 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. In the current case, changes to productivity (growth, natural mortality, maturity, and/or steepness) or selectivity assumptions 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 also 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 of the advantages of the Bayesian approach is the through the use of prior information to propagate uncertainty in model assumptions, and biological simulation approaches can provide a 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 can be used to further refine parameter distributions by identifying parameter combinations that produce clearly improbable outcomes.

This analysis presents a BSPM approach for SWPO striped marlin that addresses 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.

2 Methods

2.1 Model Framework

A Bayesian state-space surplus production model was developed following the Fletcher-Schaefer production function framework (Edwards, 2024; Winker et al., 2020). The model was implemented in the Stan probabilistic programming language (Team, 2024) using Hamiltonian Monte Carlo sampling for posterior inference. The model spans the period 1952-2022, beginning from an assumed unfished state.

2.1.1 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 $\gamma = n^{n/(n-1)}/(n-1)$.

2.1.2 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.

2.1.3 Catch Treatment

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

Catch is predicted as:

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

2.2 Prior Development

2.2.1 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, with weight-length relationships constrained to produce 150-350 kg individuals at 300 cm length. Correlations between life history parameters reflect biological trade-offs, with faster-growing species typically having higher natural mortality rates and shorter lifespans.

2.2.2 Maximum Intrinsic Rate of Increase (R_{max}) Calculation

For each of the 500,000 parameter combinations, 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 .

2.2.2.1 Equilibrium Age-Structured Population Dynamics

Survival schedule was calculated assuming constant natural mortality:

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

2.2.2.2 Fecundity Calculation

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:

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

2.2.2.3 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: pushforward]. The resulting distribution was fitted to a lognormal prior for use in the BSPM [FIGURE: Rmax prior].

2.2.3 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: NZ weight data and depletion estimation], prior to which limited large-scale commercial fishing occurred. A supplementary analysis estimated initial 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. The method assumed a two-period mortality model where total mortality increased 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.

2.2.3.1 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.

2.2.3.2 Relative Depletion Calculation

For each successful parameter combination, relative spawning stock biomass depletion was calculated 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).

2.2.3.3 Prior Construction

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. This approach incorporates both parameter uncertainty from the biological simulations and observation uncertainty from the recreational weight data into the stock assessment prior specification.

2.3 Model Implementation

2.3.1 Computational Setup

Models were implemented using the No-U-Turn Sampler (NUTS) in Stan with the following configuration: - 5 chains with random starting points - 3,000 iterations per chain (1,000 warmup) - Thinning: every 10th sample retained - Adaptation parameters: `adapt_delta = 0.99` , `max_treedepth = 15` - Final sample size: 1,000 posterior samples

2.3.2 Model Configurations

Two primary model configurations were developed:

1. **Base BSPM** (`bspm_estF.stan`): Estimated fishing mortality without depletion prior
2. **Depletion-constrained BSPM** (`bspm_estF_softdep.stan`): Includes 1988 depletion prior

Both models used the same prior parameterizations (Table 2) based on the biological simulation filtering.

2.3.3 Model Diagnostics

Model performance was assessed using: - Convergence diagnostics: $\hat{R} < 1.01$, effective sample size > 500 - Posterior predictive checks for index and catch fits - Parameter identifiability through posterior examination

[TABLE: model convergence diagnostics]

2.4 Uncertainty Characterization

The analysis explicitly incorporated multiple sources of uncertainty:

1. **Biological uncertainty:** Through comprehensive life history parameter simulation
2. **Catch uncertainty:** Via lognormal catch likelihood with $\sigma_C = 0.2$
3. **Observation uncertainty:** Through time-varying index uncertainty plus additional error
4. **Process uncertainty:** Through estimated recruitment/population variability

Results are presented as posterior distributions reflecting the full uncertainty structure, with management quantities calculated from the joint posterior distribution of all model parameters.

[FIGURE: posterior parameter distributions and correlations]

3 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.

4 References

- Castillo-Jordan, C., Day, J., Tears, 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>
- 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>
- 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>
- 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/>
- 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>

- 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. (2024). *Stan Modeling Language Users Guide and Reference Manual*, v 2.32.6.
- 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 albidia*) with simulation testing. *Collect. Vol. Sci. Pap. ICCAT*, 76(4), 219–234.

5 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	r = -0.3 with M_{ref}	(Farley et al., 2021)
Reference mortality	M_{ref}	Lognormal	0.36 yr^{-1}	0.44	r = -0.3 with A_{max}	(Hamel & Cope, 2022)
Maturity slope	a_{mat}	Normal	-20	0.2	-	(5.4/ A_{max}) 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	r = -0.5 with b_w	(Castillo-Jordan et al., 2024)
Weight exponent	b_w	Normal	3.58	0.05	r = -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

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 observation error
σ_F	Half-Normal	0	0.025	Fishing mortality variability
Depletion Constraint (Constrained Model Only)				
x_{1988}	Lognormal	$\log(0.6374)$	0.2	Initial depletion at $t = 37$ (this analysis)