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

ADVANCING THE SEAPODYM MODELING FRAMEWORK: METHODOLOGY, VALIDATION, AND ANALYSES

WCPFC-SC22-2026-EB-WP01
11 July 2026

Submitted by I. Senina[1], L. Bonnin[1], R. Forestier[1], and S. Nicol[1]

[1] Oceanic Fisheries Programme of the Pacific Community

Contents

1 Executive Summary **4**
- 1.1 Scope of work . . . 4
- 1.2 Key outcomes . . . 4
- 1.3 Report details . . . 4
- 1.4 Main results . . . 4
- 1.5 Acknowledgements . . . 5

2 Introduction **6**
- 2.1 Background and motivation . . . 6
- 2.2 SEAPODYM in the context of tuna stock-assessment models . . . 6
- 2.3 Outline of the paper . . . 7

3 Bridging analysis and environmental links assumptions **8**
- 3.1 Reducing the full model . . . 9
- 3.2 Minimal baselines . . . 10
- 3.3 MLE outcomes . . . 13

4 Fitting to simulated data **17**
- 4.1 Fitting to SPM outputs . . . 17
- 4.2 Fitting to SEAPODYM outputs . . . 19
  - 4.2.1 Fitting to simulated fisheries data . . . 19
  - 4.2.2 Fitting to simulated larval data . . . 20
  - 4.2.3 Fitting to simulated density data . . . 23

5 Likelihood profiling **24**

6 Sensitivity to fixed parameters **27**
- 6.1 Sensitivity to a fixed parameter: σ_m in the albacore model . . . 28

7 Convergence diagnostics and parameter uncertainty **29**
- 7.1 Convergence to the minimum . . . 29
- 7.2 Parameter uncertainty . . . 30
- 7.3 Structural component of uncertainty . . . 31

8 Validation with independent data **32**
- 8.1 Validation framework . . . 32
- 8.2 Statistical metrics by data type . . . 32
  - 8.2.1 Catch and CPUE . . . 32
  - 8.2.2 Length frequency of catch . . . 34
  - 8.2.3 Tagging data . . . 34
  - 8.2.4 Early-life-stage data . . . 35
- 8.3 Aggregation of metrics into global scores . . . 36
- 8.4 Validation of parameter estimates . . . 36
- 8.5 Validation against independent data . . . 37

9 Discussion **38**

10 Declaration of generative AI and AI-assisted technologies in the SC paper preparation process 41

Bibliography 42

1 Executive Summary

1.1 Scope of work

This paper responds to recommendations 676 and 677 of the WCPFC Scientific Committee SC21 meeting, and to the remaining recommendation of the 2019 SEAPODYM review (Dunn and Webber, 2020). It reports on validation of the SEAPODYM reference models for tropical tunas against simulated data, on the standard convergence and goodness-of-fit diagnostics for these models, on sensitivity analyses to fixed parameters and key assumptions, on likelihood profiles, on the sensitivity of model results to the period used in maximum likelihood estimation (MLE), on uncertainty quantification of parameter estimates, and on a bridging analysis in which the spatial and environmental functional relationships of SEAPODYM are progressively re-introduced into a simplified configuration.

1.2 Key outcomes

This paper delivers, for the SEAPODYM reference models of the tropical tunas: (i) the bridging analysis recommended by the 2019 review (Dunn and Webber, 2020), showing that the spatial and environmental structure of SEAPODYM is necessary to quantitatively reproduce tuna population dynamics and to estimate management-relevant quantities; (ii) a demonstration that the current maximum likelihood estimation recovers known parameter values when the model is fitted to data simulated by itself, with and without added observation noise; (iii) likelihood profiles for the stock size; (iv) a method to compute parameter standard errors; and (v) validation of the reference models against out-of-sample data.

1.3 Report details

Section 3 presents the bridging analysis, in which the feeding-habitat taxis, variable diffusion, and spawning-habitat environmental relationships of SEAPODYM are progressively re-introduced into a simplified model, and attributes the resulting change in fit to each mechanism against the tagging and catch data. Section 4 reports the experiments that test whether the estimation framework recovers known parameters. It also details why suggested simulated data (Goethel *et al.*, 2024) could not be used. Section 5 provides likelihood profiles for the stock size. Section 6 examines the sensitivity of the estimates to parameters held fixed. Section 7 documents the convergence diagnostics and the resulting parameter standard errors, and Section 8 evaluates the reference models against independent data inside and outside the estimation window.

1.4 Main results

1. The bridging analysis shows that, starting from a passive-drift configuration of the skipjack reference model, the progressive re-introduction of feeding-habitat taxis and variable diffusion each improves the fit to the tagging and catch data by a quantifiable amount, and that the spawning-habitat environmental relationships are likewise required by the catch data. Each component is recovered by a distinct

data stream, indicating that the SEAPODYM structure is a necessary minimum rather than additional complexity.

2. Fitting to simulated data, in which SEAPODYM is fitted to fisheries data generated from a known parameter set recovers the simulated parameters within their estimated standard errors, confirming that the estimation framework, including the statistical formulation, the model configuration and the fitting workflow, is correct.

3. Fitting to larval data (categorical and continuous), alone and jointly with fisheries data, yields consistent estimates of the spawning-habitat and early-life parameters in case of continuous data.

4. Refitting the model to its own density predictions under a changed biogeochemical forcing shifts the estimated parameters to absorb the forcing bias, giving a direct, controlled measure of the structural uncertainty that the ocean forcing contributes.

5. Likelihood profiles for the skipjack reference model help to identify the minimal stock size compatible with the observed fishing pressure.

6. The convergence diagnostics implemented in the SEAPODYM framework, principally the Hessian-based assessment of the Newton decrement, condition number, positive-definiteness of the Hessian, parameter standard errors and cross-correlations, allow the reported solutions to be confirmed as local minima and parameters estimated with their uncertainties.

7. Systematic validation against independent data outside the estimation period ensures that the reference models remain valid there, so the conclusions do not appear sensitive to the choice of estimation window.

1.5 Acknowledgements

The continued development and application of the SEAPODYM model to the work of the WCPFC Scientific Committee is facilitated through Project 62. The project affiliates the independently funded work on SEAPODYM into the SC's work programme. It is conducted in collaboration with the Oceanic Fisheries Programme (OFP) of the Secretariat of the Pacific Community. This work under Project 62 was supported by a Climate Change and Tuna Project coordinated by the Oceanic Fisheries Programme at SPC, FAO ABNJ2 Tuna Project funded by the Global Environmental Fund, and by the GCF Regional Tuna Programme, Component B, Activity 3.1.

2 Introduction

2.1 Background and motivation

The Spatial Ecosystem and Population Dynamics Model (SEAPODYM Reference Manual, 2022) is a spatially explicit, age-structured model of large-pelagic-fish population dynamics used to provide WCPFC with quantitative advice on the spatial dynamics of tropical tunas under environmental variability and climate change (Lehodey *et al.*, 2013; Bell *et al.*, 2021; Senina *et al.*, 2025).

In 2019, SPC commissioned an independent review of SEAPODYM (Dunn and Webber, 2020). The review found that SEAPODYM is the most advanced high-resolution spatially explicit population-dynamics framework available for tunas and concluded that it is suitable as a complementary tool to the operational stock-assessment platform MULTIFAN-CL (Fournier *et al.*, 1998; Hampton and Fournier, 2001) for management-strategy evaluation, climate-change projection, and ecosystem-based fishery management in the Western and Central Pacific Ocean (WCPO). The review also identified a list of recommendations, of which a response has been provided in SPC OFP (2025). One of the recommendations – a so-called *bridging analysis* or *audit trail*, in which a standard fishery model is first reproduced in SEAPODYM with movement and forage/environment relationships turned off, and the spatial and environmental functional relationships then iteratively re-introduced – was the subject of further discussion at the WCPFC Scientific Committee, leading to recommendations 676 and 677 of the SC21 meeting (SC21, 2025):

> *676. SC21 noted the progress made on the 2019 SEAPODYM Review and recommended that the SSP (Scientific Service Provider) progress validation against other approaches to continue improving this important tool. SC21 agreed that this should include fitting to simulated data, focusing on population dynamics by evaluating variation (e.g., the Indian Ocean yellowfin data produced by SPM, Goethel, et al., 2024).*
>
> *677. SC21 recommended that reporting of the reference models for tropical tunas based on SEAPODYM to SC22 include, as appropriate, standard convergence diagnostics, sensitivity analyses to fixed parameters and key model assumptions, likelihood profiles, sensitivity of model results to data used in the optimization period, and quantification of uncertainty in model estimates.*

This paper addresses both recommendations.

2.2 SEAPODYM in the context of tuna stock-assessment models

SEAPODYM is structurally and methodologically distinct from the classical age-structured stock-assessment frameworks currently used to provide scientific advice for management of tuna stocks. The assessments for WCPO skipjack, yellowfin, bigeye and South Pacific albacore are produced with MULTIFAN-CL (Fournier *et al.*, 1998; Hampton and Fournier, 2001; Castillo Jordán *et al.*, 2022), a length-based, age-structured model with statistical parameter estimation. Stock Synthesis (Methot and Wetzel, 2013) provides an analogous

generalised framework used by many other RFMOs. Both MULTIFAN-CL and Stock Synthesis represent the population on one or a few stationary spatial regions, parameterise movement, where present, as quarterly inter-region transfer coefficients, and treat recruitment, natural mortality and selectivity as time-invariant or as deviations around a mean with no explicit environmental driver. SEAPODYM, by contrast, resolves the spatiotemporal dynamics of a population continuously on a grid with typical resolutions of $2^\circ \times 30$ and $1^\circ \times 30$ days (see e.g., Senina *et al.* (2025)); mechanistically models the movement of tuna density as the combination of physical processes such as passive drift by ocean currents, active directed movement (taxis) along the gradient of a feeding-habitat index, and random non-directional movements as a diffusion process with spatially-varying rate. It uses environmental variables such as temperature, dissolved oxygen, primary production and a forage biomass simulated by the lower- and mid-trophic-level sub-model SEAPODYM–LMTL (Lehodey *et al.*, 2010, 2015) as environmental drivers and mechanistic parameters acting physically on movement or demographic rates or the model state respectively.

These differences carry trade-offs. Environmentally-driven population productivity and movement reduce the dimensionality of the estimation problem and allow consistent projection of the population state under oceanographic scenarios – a property that classical assessment models do not have. On the other hand, a spatially explicit model requires substantially larger input datasets (geo-referenced high-resolution fishery data, ocean forcing fields, micronekton predictions), and has computational costs an order of magnitude higher than that of an integrated assessment. The 2019 review also noted one more cost for the mechanistic approach, namely, the need to make "strong assumptions on the nature and magnitude of the relationship between tuna distribution/movement and the covariates" (Dunn and Webber, 2020, p. 3), and recommended conducting the bridging analysis, in effect, a request to decompose the consequences of these "strong" assumptions step by step. This paper provides a detailed response to why this recommendation is only partially applicable to the SEAPODYM approach.

Since the development of the first operational simulation version of SEAPODYM (Bertignac *et al.*, 1998; Lehodey *et al.*, 2008), the methodological developments of this model have proceeded along two complementary axes: (i) toward more rigorous estimation, including adjoint-based maximum-likelihood inference (Senina *et al.*, 2008; SEAPODYM Reference Manual, 2022), integration of multiple data types (catch and effort, length frequencies, tagging and early-life data) into a single likelihood (Senina *et al.*, 2020a, 2025), and explicit standard-error and likelihood-profile diagnostics, and (ii) toward more flexible, experimentally- and data-based model structure allowing development of fully validated quantitative reference models for four tuna populations explaining the data variability and the impact of environmental and climate change on tuna productivity and distributions (SPC OFP, 2025).

2.3 Outline of the paper

The paper is organised as follows. Section 3 closes the remaining recommendation of the 2019 review by presenting a case for bridging analysis built on the passive-drift-only configuration of the skipjack reference model (Senina *et al.*, 2025). Section 4 reports on the fitting of SEAPODYM to data simulated under three settings: (i) fisheries data, both as

in the original simulation experiment of Senina *et al.* (2008) and in a more recent experiment in which the South Pacific albacore model inputs are generated from its own outputs and the initial parameters are jittered to estimate *a priori* known values; (ii) larval data, both categorical and continuous; and (iii) density data, using the `seapodym_densities` application. Section 5 presents the likelihood profiles for the stock size and discusses the method to determine the population scale within the spatiotemporal model with parameter estimation. Section 6 assesses the sensitivity of model results to fixed parameters (with the oxygen parameter of the yellowfin model as a worked example) and to correlated or unobserved parameters. Section 7 reports the convergence diagnostics for the reference models, estimation of the parameter standard errors, and discusses the structural component of uncertainty in the model estimates. Section 8 details the statistical scores used to evaluate goodness of fit and describes how SEAPODYM reference models are validated against independent data outside the MLE window.

3 Bridging analysis and environmental links assumptions

Dunn and Webber (2020) recommended a bridging analysis in which a standard fishery model is first reproduced in SEAPODYM "where the movement and functional relationships to underlying forage fish and environmental dynamics have been 'turned off' ", and the spatial and environmental relationships are then re-introduced iteratively in order to "confirm that the underlying processes and statistical equations are correct" (Dunn and Webber, 2020, p. 3). In the review, the ADRE of SEAPODYM is viewed as "describing dynamic processes (e.g. spawning, movement, mortality), which are informed by environmental covariates (e.g. temperature, currents, primary production and dissolved oxygen concentration)" (Dunn and Webber, 2020, p. 5); environmental drivers are accordingly treated as *covariates* for fish distribution, and the movement terms and environmental drivers of population dynamics are characterised as "additional complexity" relative to a no-movement, no-environmental baseline. Moreover, the review repeatedly underlines that SEAPODYM rests on "strong assumptions on the nature and magnitude of the relationship between distribution/movement and covariates" (Dunn and Webber, 2020, pp. 3, 7, 13–15, 23). Taken together, these characterisations place SEAPODYM in the paradigm of the Species Distribution Models (SDMs) which are usually built on Generalised Additive Models (GAMs), in which the link between density distribution and environment is phenomenological, a candidate covariate may merely co-vary with the true driver of the dynamics rather than cause it, and the choice of explanatory variables and of their functional form is the critical modelling decision – one that a bridging analysis is the standard tool to justify. We adopt the recommendation but show first that this particular reading of SEAPODYM mis-states the underlying modelling concept, and that the mathematical premise of the proposed reduction, namely that a standard fishery model can be recovered as a reduced case of SEAPODYM with the spatial and environmental terms disabled, does not hold.

3.1 Reducing the full model

First, let us consider the original model structure of SEAPODYM, which numerically solves an age-structured advection-diffusion-reaction equation (ADRE) for the density $N(a,t,\mathbf{x})$ of fish of age a at time t and position $\mathbf{x} = (x,y)$ (SEAPODYM Reference Manual, 2022, §2):

$$\frac{\partial N}{\partial t} + \frac{\partial N}{\partial a} = -\nabla\cdot\big[\big(\mathbf{v}_0 + \chi_a\,\nabla H_a\big)\,N\big] + \nabla\cdot\big[D(H_a)\,\nabla N\big] - \big[M(a,H) + F(a,t,\mathbf{x})\big]\,N, \quad (1)$$

with the spawning being the age boundary condition

$$N(0,t,\mathbf{x}) = H_s(t,\mathbf{x})\,\frac{r\,\bar{N}(t,\mathbf{x})}{1 + b\,\bar{N}(t,\mathbf{x})}, \quad (2)$$

where $\mathbf{v}_0 = \mathbf{v}_0(a)$ is the ocean-current velocity field integrated over the species' vertical layers, age-dependent due to change in vertical layer accessibility across age, χ_a is the age-dependent taxis coefficient, H_a is the age-dependent feeding-habitat index integrating micronekton density across the three pelagic layers according to species' thermal preference and oxygen tolerance, H_s is the spawning-habitat index, $D(H_a)$ is the habitat-dependent diffusion, $M(a,H)$ is the age- and habitat-dependent natural mortality, where H is one of H_s, H_j or H_a depending on the life stage at age a (larval, small juvenile or juvenile and adult), F is fishing mortality, r and b are Beverton–Holt parameters and $\bar{N}(t,\mathbf{x})$ is the local mature-adult biomass. This equation is also complemented by spatial boundary conditions and an initial condition (SEAPODYM Reference Manual, 2022, §2, eqs. 1.2 and 1.4).

Now let us consider a reduced equation. Setting $\mathbf{v}_0 = 0$, $\chi_a = 0$, $D = 0$, $H_s \equiv 1$ and $H_a \equiv 1$, i.e., the configuration the review specifies, reduces Eqs. 1–2 to

$$\frac{\partial N}{\partial t} + \frac{\partial N}{\partial a} = -\big[M(a) + F(a,t,\mathbf{x})\big]\,N, \qquad N(0,t,\mathbf{x}) = \frac{r\,\bar{N}(t,\mathbf{x})}{1 + b\,\bar{N}(t,\mathbf{x})}. \quad (3)$$

Equation 3 is also known in ecological literature as the Foerster equation – the continuous time-age PDE, which would be solved independently in every spatial cell because all spatial interactions have been removed. It is not a standard fishery model. A standard fishery model (e.g. MULTIFAN-CL, Fournier *et al.*, 1998) contains structural elements that are absent in SEAPODYM and cannot be recovered by zeroing terms in Eqs. 1 and 2, e.g., (i) inter-region transfer coefficients between a small number of stationary regions, (ii) free recruitment deviations per year or quarter in place of the mechanistic H_s×BH link, (iii) time-varying parameterisations of per-fleet selectivity and catchability. The two model families share only an age-structured backbone; neither is a parametric subset of the other. In addition, the Maximum Likelihood Estimation (MLE) method in a standard fishery model relies on a specific set of likelihoods, regional scaling and statistical standardisation of CPUE or effort data, all of which SEAPODYM does not use due to employing a different spatially-explicit mechanistic approach with environmental drivers.

The no-movement reduction is biologically incoherent for highly migratory species such as tropical tunas. With $\mathbf{v}_0 = \chi_a\nabla H_a = D_a = 0$, biomass flux across spatial cells is null, so no migrations between spawning and feeding grounds or between changing

feeding grounds are modelled, hence the catch in any cell can be drawn only from biomass produced locally via local spawning and subsequent survival. A maximum-likelihood fit of Eq. 3 to such data will either fail to converge or, if using the catch-removal method, will converge to physically meaningless parameter combinations in which local fecundity and survival have been inflated until the local catch becomes sustainable everywhere. The no-movement reduction is therefore neither mathematically nor biologically a useful starting point for an audit trail.

3.2 Minimal baselines

Rather than reducing the model to a no-environment, no-movement form, we set up the audit trail from physically plausible configurations of the ADRE and examine separately the contributions of two mechanisms to model fit and predictive skill: 1) the direct movement of fish, and 2) the feeding and predation effects on larval survival and recruitment. The role of direct movement was earlier illustrated by a forward simulation in which the parameters of the skipjack reference model (Senina *et al.*, 2025) are held at their MLE values while only the active (taxis) component of movement is disabled. All likelihood components then deteriorate markedly: with passive movement alone, the model cannot reproduce the spatial distribution that underlies the observed high catches in the western tropical Pacific (Figure 30 of Senina *et al.*, 2025). The simulated distributions differ notably from the reference model, the southern sub-stock becomes more dispersed, while the northern sub-stock shifts farther north and stretches zonally across the Pacific, confirming that the estimated skipjack distribution arises from the interplay of physical transport by currents and fish behaviour (active swimming and habitat selection). This was, however, a forward simulation that held all remaining parameters at the values estimated using the model with active movement, and so cannot prove whether those parameters might re-adjust to compensate for its removal. The bridging analysis presented here lifts that restriction: by re-estimating the model under each reduced configuration, it tests whether ocean currents alone can explain skipjack dynamics as observed through the catch, tag-return and early-larval data.

Second, removing environmental drivers from movement and reproduction simultaneously is impractical because recruitment and movement patterns are inter-related. Namely, H_s determines the spatial pattern of recruitment, and the diffusion and advection parameters that MLE estimates in a sequential build-up are conditional on the spatial distribution of recruits. An $H_s \equiv 1$ spawning would put recruits in the cold edges of the basin where larvae cannot survive in nature; the MLE would then have to compensate for those non-existent larvae either by inflating local mortality there or by distorting adult parameters. Thus, the "improvement" obtained by switching $D_0 \to D(H_a)$ or by adding $\chi_a \nabla H_a$ in the model without mechanistic link between larval survival and environment would not reflect the true movement contribution because it will be confounded with the structural change in the recruitment model. Hence, varying the movement and recruitment processes separately, each against the reference, is needed to avoid conditioning either set of parameters in subsequent estimations.

Thus, we choose two minimal physically plausible baselines rather than the fully reduced Eq. 3:

Baseline 1. Passive advection by ocean currents, isotropic constant diffusion The corresponding equation for baseline B_1 is

$$\frac{\partial N}{\partial t} + \frac{\partial N}{\partial a} = -\nabla\cdot\big(\mathbf{v}_0\, N\big) + D_0\, \Delta N - \big[M(a) + F(a,t,\mathbf{x})\big]\, N, \tag{4}$$

where $\mathbf{v}_0$ is the input vector field of ocean currents in the surface (epipelagic) layer. The baseline B_1 retains the spatial dynamics of the ADRE (so that empirical spatial catches can in principle be supported through transport by currents) and the age structure of the population, but contains none of SEAPODYM's mechanistic habitat features. Note, that this equation describes the spatial dynamics of density of larvae and small juveniles (up to three months for skipjack). For older age classes, such an equation was proposed by Barrier *et al.* (2023) as sufficient to explain the ENSO-induced zonal movements of skipjack tuna.

Baseline 2. Density-dependent recruitment with only temperature-driven recruitment. In the second baseline B_2, only the spawning habitat H_s is modified. The minimal physically plausible recruitment baseline is the thermal-only Gaussian

$$H_s^{(0)}(\mathbf{x},t) = f_1\big(\mathrm{SST}(t,\mathbf{x}); T^*, \sigma\big) = \exp\Big(-\frac{(\mathrm{SST} - T^*)^2}{2\sigma^2}\Big), \tag{5}$$

with two parameters T^* (optimal SST) and σ (thermal tolerance width). It retains the property that larval survival is higher in warm waters and near zero in cold ones, so the MLE under this baseline is biologically plausible and convergent, while it omits the two other factors of the full reference $H_s = f_1 \cdot f_2(P;\alpha) \cdot f_3(F_1;\alpha_F,\beta_F)$, i.e., the prey-availability function (with primary production as a proxy for larval food) and the predation function dependent on the micronekton density in the surface layer.

The bridging analysis builds each baseline up toward the reference model along its own "axis". On the movement axis (full H_s, taxis off) the diffusion is built up from constant to habitat-dependent, and active taxis is then restored to recover the reference; because both $D(H_a)$ and the taxis velocity $\chi_a \nabla H_a$ are driven by the age-dependent feeding habitat H_a, this also introduces the full three-dimensional definition of H_a, integrating the micronekton biomass of the SEAPODYM-LMTL model (Lehodey *et al.*, 2010, 2015) across the three pelagic layers weighted by in-layer temperature- and oxygen-driven accessibility. On the recruitment axis (reduced H_s, reference movement) the spawning habitat is built up from temperature-only to temperature-and-food, and the predation function is then added to recover the reference. The configurations, and the contribution each comparison quantifies, are:

- $\mathbf{B_1}$ ($D = D_0$, $\chi_a = 0$, full H_s): passive transport with constant diffusion – the movement baseline.
- $\mathbf{B_1^+}$ ($D = D(H_a)$, $\chi_a = 0$, full H_s): habitat-dependent diffusion, taxis still off. The comparison $\mathbf{B_1}$ vs $\mathbf{B_1^+}$ quantifies the contribution of variable diffusion, and $\mathbf{B_1^+}$ vs the reference, which differ only in χ_a, quantifies the contribution of active taxis.
- $\mathbf{B_2}$ ($H_s = f_1$, reference movement): SST-only spawning habitat – the recruitment baseline.

$\mathbf{B_2^+}$ ($H_s = f_1 f_2$, reference movement): SST and larval-food (primary-production as a proxy) spawning habitat. The comparison $\mathbf{B_2}$ vs $\mathbf{B_2^+}$ quantifies the contribution of the larval-food function, and $\mathbf{B_2^+}$ vs the reference quantifies the contribution of predation.

In each experiment, the SEAPODYM maximum-likelihood estimation is re-run with the appropriate subset of parameters being released. We report the change in the negative log-likelihood by its data-type components (catch, length frequencies, tagging, early-life data), the change in the estimated parameters that are common to consecutive steps, and the change in the principal management-relevant outputs: temporal and spatial change of tuna biomass. The resulting comparison gives the contribution of each mechanistic feature (habitat-driven diffusion, active taxis on the three-dimensional feeding habitat, and the environmental drivers of recruitment) to the fit and to the model's predictions, exactly as the audit-trail concept of Dunn and Webber (2020) intends.

We adopt the skipjack tuna reference model (same model configuration as in Senina *et al.* (2025) but under ocean forcing ERA5 and at 2° spatial resolution) as the test case for this analysis on the basis of three properties of this species and the reference model. First, skipjack is known to be the most mobile of the four WCPO tropical tunas, so the *a priori* contribution of active movement to model fit is largest and the audit is most informative on the movement axis. Second, skipjack recruitment is highly temperature-sensitive at the larval stage, both empirically and through the hatching-success relationship calibrated from laboratory studies (Fujioka *et al.*, 2024) that is now embedded in the SEAPODYM v5 early-larvae observation model (Senina *et al.*, 2025), so the audit is also informative on the H_s-driven recruitment axis. Third, skipjack is short-lived (~ 5 years), so the response of the population to a change in the mechanistic forcing propagates quickly through the age structure within the MLE window, which makes the audit experiments practical. Note an important difference between the skipjack and default model structures – the age boundary condition does not depend on H_s, which would otherwise imply instantaneous egg mortality – this is to enable the new observation model for skipjack larvae introduced in (Senina *et al.*, 2025). Hence, in the skipjack model configuration, H_s only drives larval survival over time, in other words, H_s deterioration increases the larval natural mortality rate parameterised for the optimal environmental conditions as follows:

$$M(a,H) = m(a)\left(1+\rho(a)+\varepsilon\right)^{1-H}, \qquad (\varepsilon, H) = \begin{cases} (\varepsilon_0,\ H_s) & \text{larvae} \\ (\varepsilon_a,\ H_a) & \text{juveniles and adults} \end{cases} \tag{6}$$

where $\rho(a)$ is the dimensionless age-dependent environmental-vulnerability amplitude, highest (2.7) at age 0, negligible (0.07) by the first autonomous age (one quarter for skipjack) and vanishing (< 0.01) at observed ages (≥ 30 cm fork length), essentially imposing the environmentally-driven mortality for the least resilient but least observed (not directly observed through fisheries data) ages while allowing ε_a to be estimated from the observed ones. This is precisely why the fourth environmental link, the variability of natural mortality as a function of feeding habitat, $M(a,H)$, is not tested here: the reference-model MLE for skipjack estimated the corresponding parameter for juveniles

and adults as $\varepsilon_a = 0$. The audit experiment $M(a, H) \to M(a)$ would therefore by construction return a null change in likelihood. A similar audit could be applied to another species in future work where the variable component of $M(a, H)$ is non-zero in MLE.

3.3 MLE outcomes

Table 1 reports the best negative log-likelihood by data type for both axes and the reference: in each configuration the released parameters were jittered and re-estimated. Figure 1 shows the corresponding change in the spatial distribution of recruits and total biomass, and Figure 2 the same quantities aggregated over the WCPO and EPO regions, where the east–west redistribution across configurations is most apparent.

On the movement axis (B_1), without active taxis (B_1, B_1^+) the population spreads zonally across the basin: recruits form a narrow equatorial band that extends far into the eastern Pacific, and total biomass disperses broadly, over-filling the EPO and the sub-tropics away from the western-Pacific catches (circles on the maps of total biomass). These distributions deteriorate the fit to the catch data, increasing the negative log-likelihood of catch by 50 %.

Replacing constant by habitat-dependent diffusion ($B_1 \to B_1^+$) results in almost indistinguishable distributions, confirming that heterogeneous diffusion alone barely reshapes the field. It improves the total likelihood only marginally ($6.92 \to 6.76$), slightly improving fits for both catch ($0.76 \to 0.72$) and tagging ($T : 3.04 \to 2.91$) data. Besides, the tag fit improvements are attributed to the least-observed Kuroshio region, and the small catch gain reflects only the extra dispersion of biomass in the low-habitat zones, that likely fits the noisy catches there marginally better. Both passive configurations fall short of the reference on *both* discriminating components – the tagging misfit is roughly doubled ($T \approx 2.9$ vs 1.38) and the catch misfit some 40–50 % larger ($C \approx 0.72$–0.76 vs 0.51).

The decisive contribution is *active taxis*: quantified by comparing B_1^+ (full H_s, taxis off) with the reference (full H_s, taxis on), it more than halves the tagging misfit ($T : 2.91 \to 1.38$) *and* removes the catch deterioration ($C : 0.72 \to 0.51$); this single step ($6.76 \to 4.89$) accounts for almost all of the B_1-to-reference improvement ($6.92 \to 4.89$). Directed, habitat-following movement is therefore what *both* the tagging and the catch require: by concentrating biomass in the favourable feeding habitat predicted where the fish are caught, taxis corrects the spatial misplacement that the over-dispersed passive field cannot.

On the recruitment axis (B_2), an SST-only spawning habitat does not affect the fit to the tagging data – movement is already at reference – but largely misfits the catch ($C = 1.11$). Because the estimated solution produces the most extended and dispersed recruit distribution, it reduces the total number of recruits (Figure 2) to reproduce the lower CPUEs of the sub-tropical zones, but compensates for the recruitment decrease in the core skipjack habitat by lowering the predation mortality on juveniles (increasing β_p to steepen the function's age-decline; see Table 1), which move and concentrate in their favourable feeding habitat in the tropical and equatorial zone. One major discrepancy between B_2 and the reference is in the EPO: the model produces very low total biomass there, indicating that the EPO is driven by recruitment rather than by movement from the WCPO.

Table 1: Estimated and fixed parameters of the five bridging configurations. * fixed (not estimated); [/] estimated at the lower / upper boundary; a dash (−) marks a parameter made ineffective by the disabled process. Parameters held fixed at the same value in all five runs are omitted. Negative log-likelihood components: C catch, L length-frequency, T tagging, E early-life.

θ	**Description**	B_1	B_1^+	B_2	B_2^+	**Ref.**
	Fit and optimisation					
	$-\ln L$ total ($\times 10^5$)	6.92	6.76	5.54	5.09	4.89
	catch C	0.76	0.72	1.11	0.57	0.51
	length L	2.74	2.75	2.65	2.79	2.64
	tagging T	3.04	2.91	1.42	1.33	1.38
	early-life E	0.38	0.38	0.37	0.39	0.37
	Estimated parameters n	29	37	35	36	36
	Iterations	396	496	419	285	242
	Max. gradient $G_{\max}$	0.2	13.0	83.1	6131.8	558.7
	Reproduction and spawning habitat H_s					
α	shape (curvature) parameter of the normalised Holling-III larval (functional) response to food (f_2 in H_s)	12]	11.9999]	-	0.2673	11.9998]
β_F	predation shape parameter (f_3 in H_s)	2.3115	2.3022	-	-	2.6215
b	slope parameter, Beverton–Holt, nb/km^2	0.1]	0.1]	0.0665	0.1]	0.0842
	Adult feeding habitat H_a					
σ_0	std. dev. of SST Gaussian (age 0), °C	-	[1.0	1.4547	1.6288	1.0658
σ_K	std. dev. of SST Gaussian, oldest cohort, °C	-	[2	[2	[2	[2
T_K	optimal SST, oldest cohort, °C	-	28.5313	27.4936	[15.5	27.2616
b_T	Allometric power in thermal preferences at age	-	[0.5	[0.5	2.9977]	[0.5
$\hat{O}$	dissolved-oxygen threshold, ml/l	-	3.7746	4.0178	3.4921	3.948
eF_1	epipelagic forage contribution	-	1.5]	1.5]	1.4989]	1.5]
eF_6	migrant bathypelagic forage contribution	-	5]	1.4166	1.5928	1.3299
	Movement					
V_m	max. sustainable speed (taxis), BL s^{-1}	-	-	0.6897	0.6979	0.5484
σ	diffusion-rate multiplier	0.4481	0.2591	0.5]	0.4695	0.3924
c	diffusion–habitat coefficient	-	0.5]	0.5]	0.4999]	0.5]
	Natural mortality					
β_p	predation-mortality slope coefficient	[0.1	[0.1	1]	[0.1001	0.1156
β_s	senescence-mortality slope coefficient	0.8721	0.8776	0.8699	0.8226	0.7885
	Fishing mortality					
q_{P1}	Catchability	0.0035	0.0036	0.0032	0.0043	0.0036
s_{P1}	Selectivity SD	4.1966	4.3225	3.935	4.127	3.909
$\hat{l}_{P1}$	Size at maximal selectivity, cm	46.5836	46.8242	45.871	47.3894	46.0444
q_{P2}	Catchability	0.0061	0.0062	0.0058	0.0058	0.0064
q_{P23}	Catchability	0.0091	0.0094	0.0065	0.0087	0.0066
$\hat{l}_{P23}$	Size at maximal selectivity, cm	77.5336	77.9326	72.9407	76.5759	70.8872
q_{P3}	Catchability	0.0026	0.0025	0.002	0.0022	0.0022
s_{P3}	Selectivity SD	7.3142	7.3531	6.8497	7.8504	7.0321
$\hat{l}_{P3}$	Size at maximal selectivity, cm	53.637	53.7915	53.0719	53.9793	53.3622
s_{S5}	Selectivity SD	12.0763	12.1634	9.9322	9.6099	9.156
$\hat{l}_{S5}$	Size at maximal selectivity, cm	61.2215	61.4992	54.9707	55.2235	54.1876
$\hat{l}_{S7}$	Size at maximal selectivity, cm	71.4114	71.2121	71.3571	71.4349	71.1467
q_{L8}	Catchability	1e-04	1e-04	3.3e-5	3.3e-5	3.9e-5
s_{L8}	Selectivity slope	0.1702	0.1741	0.1547	0.1477	0.1627
s_{S10}	Selectivity SD	6.6298	6.6937	5.7239	5.8542	5.4258
$\hat{l}_{S10}$	Size at maximal selectivity, cm	47.6625	47.7481	44.9847	45.6948	44.8722
s_{S11}	Selectivity SD	14.1416	14.142	11.3575	10.1379	11.3928
$\hat{l}_{S11}$	Size at maximal selectivity, cm	72.7501	73.2274	62.6261	65.1786	63.7574
$\hat{l}_{S12}$	Size at maximal selectivity, cm	72.9433	73.4209	68.5224	70.3056	69.6713
q_{P15}	Catchability	0.0229	0.0175	0.0146	0.0101	0.0103

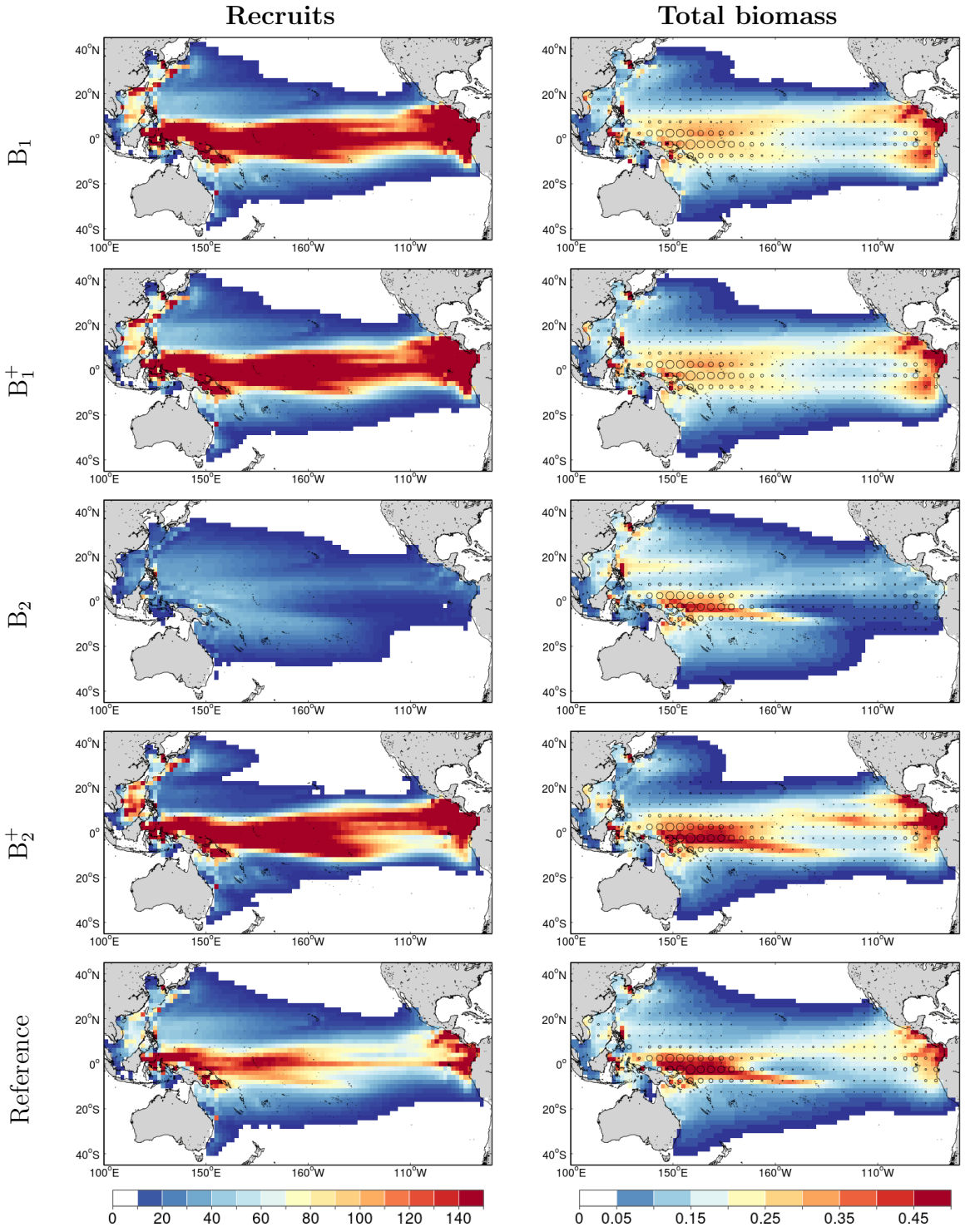

Figure 1: Spatial distribution of recruits (left) and total biomass (right) for the bridging configurations (rows, top to bottom): B_1 (passive, constant diffusion), B_1^+ (passive, habitat-dependent diffusion), B_2 (reference movement, SST-only spawning habitat), B_2^+ (reference movement, SST and larval-food spawning habitat), and the reference model. Circles show observed catches; averages over 2013–2022.

Adding the larval-food function f_2 to H_s (B_2^+) halves the catch misfit ($C : 1.11 \to 0.57$)

and brings the total close to the reference (5.09 vs 4.89). This improvement comes from the higher EPO biomass, which is necessary to sustain the observed local purse-seine catches, although it still does not reproduce the local CPUE distribution as well as the reference.

The remaining predation function f_3 closes the small residual at the reference ($C \to 0.51$). The spatial catch is therefore governed by the environmental drivers of recruitment, and all three drivers are essential to fit the observations: an SST-only spawning habitat misplaces recruits and cannot support the observed catches, whereas accounting for larval food availability and predation on larvae restores them.

The two axes thus provide a clear process-to-data attribution: active taxis is required primarily by the tagging data and, to a lesser extent, by the catch/CPUE; the food (and predation) structure of recruitment is required by the catch. Neither is "additional complexity" over a no-movement, no-environment baseline; each is a necessary component, with movement informed by the tagging data – which recruitment cannot mimic – and recruitment informed by the catch.

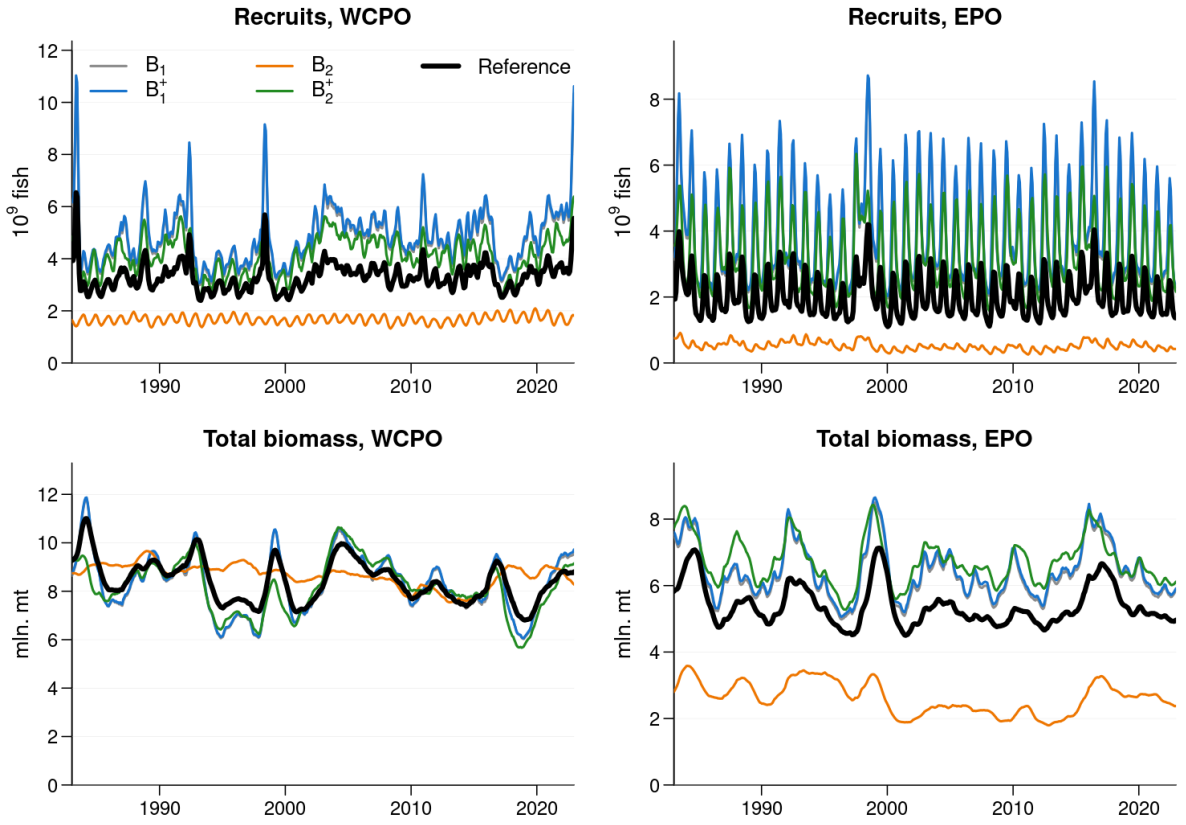

Figure 2: Monthly time series of recruits (top, millions of fish) and total biomass (bottom, millions of metric tons) aggregated over the WCPO (100°E–150°W, left) and EPO (150°W–70°W, right) regions (40°S–45°N), for the five bridging configurations (legend in the top-left panel; labels as in Figure 1), over 1983–2022.

4 Fitting to simulated data

4.1 Fitting to SPM outputs

Following recommendation 676 of the WCPFC Scientific Committee SC21 meeting, the simulated data of Goethel *et al.* (2024) were obtained via the project repository at https://github.com/aaronmberger-nwfsc/Spatial-Assessment-Modeling-Workshop and the Google Drive indicated in the project's Readme. The study provides the estimation-model configurations used by the participating platforms (CASAL2, MFCL, SPASAM, and three Stock Synthesis variants for the YFT case study) and the aggregated diagnostic outputs; the underlying simulated datasets themselves (100 realisations of the Indian Ocean yellowfin operating model, and the spatially disaggregated pseudo-observations of catch, length frequency, tagging and CPUE). Considering how this dataset could be used, we examined whether the methodological pre-conditions for an interpretable comparison between SEAPODYM and the SPM Indian Ocean yellowfin operating model were satisfied. Six structural mismatches between the two modelling frameworks and between the SPM-generated data and the data types that SEAPODYM is designed to assimilate prevent such a comparison from being interpretable. We conclude that SC21 Recommendation 676 (SC21, 2025) is best addressed through the bridging analysis presented here and the within-SEAPODYM identical-twin experiments reported in the three subsections that follow, with the SPM dataset retained as a reference operating model for any future work that addresses one or more of the pre-conditions individually.

1. **Resolution mismatch.** The SPM Indian Ocean yellowfin operating model resolved the basin on a 13×17 grid of $5° \times 5°$ cells at a quarterly time step, with quarterly age classes from age 0.25 to age 7+ years (Goethel *et al.*, 2024). The SEAPODYM tropical-tuna reference models operate at $1° \times 1°$ or $2° \times 2°$ on a monthly time step with monthly age classes (SEAPODYM Reference Manual, 2022, §5)(Senina *et al.*, 2025) which is a $25\times$ spatial, $3\times$ temporal and $3\times$ age refinement relative to SPM. A coherent setup would require either degrading the SEAPODYM resolutions, which is detrimental to the ADI solver and the correctness of the numerical solution, or interpolating the SPM-generated observations (specifically, catches) onto the SEAPODYM grid, which introduces within-cell uniformity artefacts.

2. **Fishing effort is not available.** Catch alone does not inform the spatial distribution of biomass. SEAPODYM uses two catch prediction methods with respective likelihoods (SEAPODYM Reference Manual, 2022, §6.2.1): an effort-based formulation in which predicted catch is the product of catchability, age-based selectivity, fishing effort and modelled biomass-at-age, and a catch-removal formulation in which observed catch is subtracted directly from the modelled local biomass. The catch-removal method accounts for fishing mortality and constrains local biomass *from below*, which enforces $B_{t,i,j} \geq C_{t,i,j}$ via a penalty when local biomass cannot support observed catch, but it does not, on its own, inform the spatial distribution of the population. It can inform only via the cells with largest catches, where SEAPODYM does not predict enough biomass, but all low-catch cells remain unobserved. The dataset of Goethel *et al.* (2024) contains no effort series for any fleet: catch was treated as known without observation error, and fishing mortality entered

the OM as a cell-specific exploitation rate. A SEAPODYM fit to these data would therefore lose its principal fisheries data for inferring biomass spatial structure, the cell-by-cell comparison of predicted and observed CPUE that drives the gradient of the catch likelihood at every grid point. The IO yellowfin fishery is dominated by the purse-seine fleet in the western tropical Indian Ocean and the longline fleet in the central Indian Ocean, so catch-only fitting would constrain biomass density mostly there and leave the rest of the basin essentially data-free.

3. **Tag data resolution mismatch.** Aggregated tag data cannot inform the movement likelihood in SEAPODYM. The tagging likelihood implemented in the skipjack reference model (Senina *et al.*, 2025) is based on the method introduced by Senina *et al.* (2020a): each recapture is treated as a point observation in continuous space and time, smoothed onto the model grid with a bivariate Gaussian kernel; the spatial position and date of release must be exact as those constitute the initial conditions of the ADEs which are solved along with the full population model. In the Goethel *et al.* (2024) dataset, both releases and recaptures had already been aggregated to $5^\circ \times 5^\circ$ cells, to quarterly time steps and to integer quarter-ages, and were supplied as binomial proportions per (cell $\times$ age) with a population-at-risk denominator. From such binning, it is not possible to recover original information.

4. **CPUE indices are not directly comparable.** SEAPODYM does not currently include an abundance-index likelihood as a primary observation model. CPUE enters the SEAPODYM likelihood implicitly, through the Gordon–Schaefer catch equation $C_{f,t,i,j} = q_f \, E_{f,t,i,j} \sum_a s_{f,a} \, N_{a,t,i,j} \, \Delta_i x \Delta_j y$ that predicts catch from effort, catchability, age-based selectivity and modelled biomass-at-age; the predicted CPUE is then $\widehat{CPUE}_{f,t,i,j} = C_{f,t,i,j}/E_{f,t,i,j}$ and is informative only where observed effort is available. The longline CPUE provided by Goethel *et al.* (2024) is a standardised abundance index with no effort field. Fitting to it would require the implementation of a new likelihood component, something like $\widehat{I}_{f,t,i,j} = q_f \sum_a s_{f,a} \, N_{a,t,i,j}$ with an appropriate observation error together with the corresponding adjoint code in the SEAPODYM C++ codebase. This is several weeks of careful development plus convergence testing. It is not justifiable for a comparison whose dimensions are already mismatched.

5. **Habitat-definition mismatch.** The SPM Indian Ocean yellowfin operating model defined habitat preference as a function of sea surface temperature and surface chlorophyll-a concentration (Dunn *et al.*, 2020). SEAPODYM defines the feeding habitat of the adult age classes through the accessibility function that essentially accounts for a three-dimensional environment, integrates the age-dependent thermal preference, an oxygen-tolerance function, and the density of the micronekton functional groups inhabiting three pelagic layers (epipelagic, upper mesopelagic, lower mesopelagic) (SEAPODYM Reference Manual, 2022, §2.2)(Lehodey *et al.*, 2015, 2010). The biological motivation for this difference is well-established: adult yellowfin spend a substantial fraction of their time below the mixed layer chasing migrant mid-trophic prey; their depth use is constrained by ambient oxygen; and yellowfin do not feed on phytoplankton at any life stage. SPM uses surface variables as *covariates* for yellowfin biomass distribution. A discrepancy between

SEAPODYM predictions and the SPM truth in such an exercise would confound (i) SEAPODYM's own structural assumptions, (ii) the SPM surface-only habitat definition, (iii) the differing ocean-forcing fields used to derive the two habitat formulations, and (iv) the data-content mismatches set out in points 1–4 above. The exercise would produce a result, but the result could not be attributed to any single source.

In addition to the structural mismatches, one important difference is worth mentioning. Namely, the critical difference of the number of degrees of freedom to data fitting. The estimation models reported by Goethel *et al.* (2024) estimated of order 10^2 to 10^3 free parameters per replicate. The bulk of the count was in quarterly recruitment deviations, regional recruitment apportionment, multi-region movement matrices with quarterly deviations, per-fishery selectivity parameters and per-index catchability with random-walk deviations (Goethel *et al.*, 2024, Table 2)(Fournier *et al.*, 1998; Methot and Wetzel, 2013). The SEAPODYM tropical-tuna reference models estimate of order a few tens of free parameters (Senina *et al.*, 2025; SEAPODYM Reference Manual, 2022): typically 10–20 biological parameters describing reproduction, mortality, movement and habitat preferences; 1–3 selectivity parameters per fishery; and a single space-invariant catchability per fishery, which per se require a careful structuring of the fisheries data (Forestier *et al.*, 2025). The two regimes correspond to fundamentally different statistical inference – a high-dimensional, weakly-mechanistic fit on the one hand, a low-dimensional, environmentally-constrained fit on the other. With the habitat-definition and forcing mismatch, it is unlikely that the SPM solutions could be recovered. Thus, even if such results could be achieved in the above setup, a SEAPODYM model with a larger percent-relative-error in SSB might still be the better tool for the management questions for which it is the appropriate framework, namely the high-resolution spatial distribution of the population and its response to environmental variability and climate change.

4.2 Fitting to SEAPODYM outputs

In a simulation experiment of Senina *et al.* (2008), catch and length-frequency data were generated from a known parameter set, and the MLE showed that, given the model structure, the function minimiser recovers the model parameters used to produce the simulated outputs within their estimated standard errors. Here we provide three working examples illustrating the model and parameter estimation robustness and revealing the data and ocean forcing requirements to ensure that the methods work.

4.2.1 Fitting to simulated fisheries data

Artificial catch data were constructed from a deterministic forward simulation (no noise added), then used as observations in the optimisation.

The experiment is extended here to the South Pacific albacore configuration, with 16 fisheries and a 2° model grid. Synthetic catch data were generated from a reference forward simulation with a known parameter set, then spatially aggregated from the 2° model grid to the 5° catch data input format. This aggregation step introduces a residual discrepancy between the synthetic observations and the model predictions at the reference parameters. The reference parameter set was perturbed with 10 independent jitter draws;

the MLE was run from each perturbed start with the same free parameters. 8 of 10 runs converged; the 2 non-converged runs were excluded from interpretation. All 8 converged runs returned identical parameter values to 4 significant figures, indicating a unique minimum was found regardless of starting point. Recovered biological parameters lie within 9% of the reference values across all 16 parameters, with the largest discrepancy in nb recruitment (+8.1%) (Table 2). Catch fit at the recovered solution is near-perfect ($r^2 = 1.00$, nRMSE= 0.02), confirming that the optimiser reached a solution consistent with the synthetic data.

Table 2: Biological parameter recovery in the twin experiment. The reference column shows parameter values used to generate the artificial dataset. The recovered column shows the value returned by all 8 converged jitter runs (identical across runs to 4 significant figures). Relative error is $(recovered - reference)/|reference| \times 100$.

Parameter	Reference	Recovered	Rel. error (%)
σ_0	2.384	2.420	+1.5
T_0	26.23	26.62	+1.5
σ_{larvae}	3.005	3.150	+4.8
T^*	24.10	24.46	+1.5
β_F	1.934	1.940	+0.3
σ_m	5.047	4.843	−4.0
T_m	10.03	10.15	+1.2
$\hat{O}$	3.689	3.688	−0.03
E_{11}	2.000	2.103	+5.2
E_{22}	3.795	3.749	−1.2
E_{21}	0.8202	0.8767	+6.9
σ	5.500	5.528	+0.5
V_m	1.057	1.079	+2.1
r	0.007406	0.008007	+8.1
$\hat{t}$	128.5	128.3	−0.2
ϱ_s	1.076	1.077	+0.1

4.2.2 Fitting to simulated larval data

The work of Bonnin and Senina (2024) extended the SEAPODYM likelihood to two new types of early-life-stage observation that are relevant for parameter estimation but that the framework could not previously assimilate: (i) the digitised Nishikawa global larval CPUE atlas (Nishikawa *et al.*, 1985; Buenafe *et al.*, 2022), in which larval counts are reported as *binned categorical* intervals (e.g. 0, 0–0.5, 0.5–1, 1–5, > 5 larvae per 1000 m^3 of water strained), and (ii) geo-referenced female gonad histology samples (Farley *et al.*, 2013, 2014) which yield a continuous, season-aggregated spawning index from the proportion of females in *Actively spawning* or *Spawning capable* reproductive state.

Identical twin experiments. The methodology and the data requirements of the new likelihoods were validated through identical twin experiments using the Pacific yellowfin reference model (Senina *et al.*, 2020a) forced with ERA5–NEMO–PISCES at 1°× monthly

Table 3: Accuracy of the optimisation process to estimate model parameters, depending on quality and quantity of observations. Original parameter values, used to generate pseudo-datasets, are compared to estimated parameter values associated with twin experiments using continuous and categorical pseudo-data with varying numbers of observations. Percentage differences from original values are displayed within parentheses. The range of explorable parameter values is indicated in the columns Min and Max.

				Continuous	Categorical	
				200 obs.	6000 obs.	12000 obs.
Par.	Min	Max	Orig. val.	Est. val.	Est. val.	Est. val.
q_{larvae}	0.0	100	1.0	1 (0%)	0.9 (-10%)	0.93 (-7%)
σ_{larvae}	0.5	10	2.0	2 (0%)	1.81 (-10%)	1.85 (-7%)
T_{larvae}	25.5	35	31.0	31 (0%)	30.52 (-2%)	30.65 (-1%)
α	0.0	10	2.7	2.7 (0%)	2.63 (-3%)	2.75 (+2%)
α_F	0.0	10	1.8	1.8 (0%)	1.74 (-3%)	1.77 (-2%)
β_F	0.0	10	2.3	2.3 (0%)	-	-

resolution. A single original solution (Bonnin and Senina, 2024, Table 2) with known parameter values (q_{larvae}, σ_{larvae}, T_{larvae}, α, α_F, β_F) was used to generate continuous and categorical pseudo-datasets of seasonal larval density (Figs. 5–6 of Bonnin and Senina, 2024), then sub-sampled at varying numbers of grid cells n to characterise the data-quantity requirement (up to 12000 grid cells, a value similar to the 11545 observations of the Nishikawa dataset). Ten jitter runs from randomly perturbed initial values were performed for each twin experiment.

The ability of the optimisation process to estimate model parameters was assessed by computing the percentage difference between original and estimated parameter values.

The original model state, list of free parameters, and explorable range of parameters are described in Table 3.

A small number of continuous observations (200 grid cells) enabled estimation of the known parameter values with high accuracy (Figure 3, Table 3). A much greater number of categorical observations was necessary to reach convergence, with relatively high residual error (7%) for the estimation of q_{larvae} and σ_{larvae} with 12000 observations. The shape parameter in predator-dependent function β_F could not be estimated with categorical data due to correlation with other parameters.

The twin experiments established the threshold of usability of the Nishikawa larval CPUE: its categorical nature, rather than its spatial coverage, is the limiting factor. Re-estimation of all spawning-habitat parameters from the Nishikawa data alone is therefore not attempted in the SEAPODYM reference models; the data are instead used as an independent validation set for the spawning-habitat predictions of the existing models (cf. Section 8.2.4), and – in the configuration adopted in the SEAPODYM v5 skipjack reference model (Senina *et al.*, 2025) – through the early-larvae observation model rather than as a separate likelihood term. As an applied counterpart to the simulation experiments, Bonnin and Senina (2024) fitted the spawning-habitat and spawning-season parameters of the South Pacific albacore reference model to the gonad-derived spawning index (which is

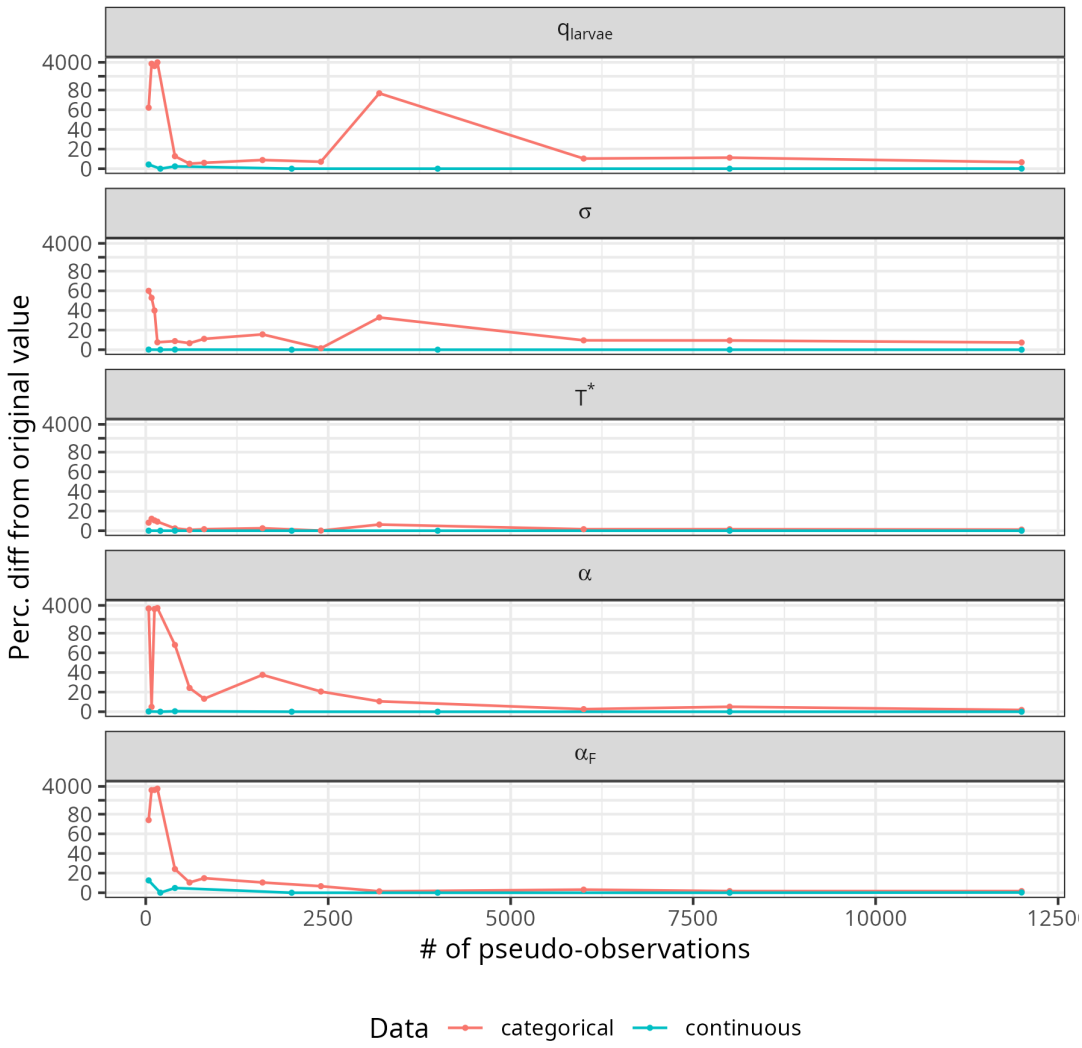

Figure 3: Accuracy of the optimisation process to estimate model parameters, depending on quality and quantity of observations. An original model state was used to generate continuous or categorical larval density index. Seasonal averages were computed from these outputs, with varying quantities of randomly sampled grid cells. These pseudo-datasets were used to estimate model parameters. The ability of the optimisation process to estimate parameter values accurately was evaluated as the percentage difference to the original parameter values.

continuous), with the simultaneous re-estimation of the fisheries catchability and selectivity parameters: the re-fitted model improved the spatial correlation with the data-derived spawning index from $R = 0.31$ to $R = 0.47$ (and from 0.47 to 0.70 in the spawning season alone), shifted the spawning-season peak by ~ 23 days ($\hat{t}$ from 117 to 140; ϱ_s from 1.054 to 1.094, corresponding to a minimum latitude of 16.5° for the spawning migration to occur), and retained an essentially equivalent fit to the catch and length-frequency data (Table 8 and Figs. 16–18 of Bonnin and Senina, 2024).

4.2.3 Fitting to simulated density data

The different ocean forcing fields used to describe tuna habitat in SPM and SEAPODYM can prevent SEAPODYM from reproducing SPM-generated data. Indeed, a SEAPODYM MLE can be very sensitive to such forcing, with the estimation of model parameters enabling some integration of the biases associated with the ocean model.

To illustrate this sensitivity of a SEAPODYM MLE to its ocean forcing, we show here the recalibration work that was done after a modification of the biogeochemical model that was used to provide ocean forcing fields. The new version of the forcing, hereafter called v2, integrated a modification of the nitrogen fixation mechanism (Aumont *et al.*, 2015) and resulted in a reduction of primary production in the tropical WPO, and thus, in cascade, in a reduction of the euphotic depth and of the layer thickness, and in a rise of layer temperatures (e.g. up to 0.9°C for the epipelagic layer and up to 1.7°C for the upper-mesopelagic layer).

To reduce computing time and memory use, the tuna response to the new environment was recalibrated using a no-fishing framework, using age-structured density predictions from the original MLE as observations. This freed the recalibration from the heavy requirements associated with the integration of fisheries, tagging and early-life data.

The optimisation process resulted in a different MLE, with significant changes to the model parameter values (Table 4). The recalibration mostly affected model parameters associated with early life, with most significant changes to the prey function shape parameter α, the variability of larvae mortality with spawning habitat ϵ_{larvae}, the slope coefficient in predation mortality β_p, and the contributions of the epipelagic, migrant upper-mesopelagic, and highly migrant lower-mesopelagic forage to the feeding habitat, E_{11}, E_{21} and E_{31}.

The modification of the Holling-III larval function response to food, α, is a good example of how SEAPODYM calibration can integrate a bias in the ocean forcing. As indicated previously, the correction of the nitrogen fixation mechanism resulted in a significant decrease of primary production, a proxy for the larval prey density. The model compensated for this decrease of prey by a lower α.

The modification of the variability of larvae mortality with spawning habitat ϵ_{larvae} could be a response to the decrease in epipelagic forage, a proxy for larvae predator density. The corresponding increase of the spawning habitat index may have resulted in higher spawn, which could have been compensated by higher habitat-dependent mortality ϵ_{larvae} to maintain the same level of recruitment. These changes may also have affected the estimation of the slope coefficient in predation mortality β_p, which controls the rate of mortality decrease in the first months of life.

Table 4: Sensitivity of MLE to ocean forcing. Modification of MLE parameter values when changing ocean forcing from v1 to v2.

	MLE v1	MLE v2	% difference
Mortality			
$\bar{m}_p$	0.0549	0.0419	-24%
β_p	0.239	0.118	-51%
$\bar{m}_s$	0.000107	0.0000689	-35%
β_s	1.2	1.27	+6%
ϵ	4.99	4.99	+0%
ϵ_{larvae}	21.4	0.519	-98%
Spawning habitat			
σ_{larvae}	5	5	+0%
T_{larvae}	29.5	30	+2%
α	2.75	0.371	-86%
α_F	2.19	0.0000748	-100%
β_F	3.15	2.79	-11%
Thermal pref.			
σ_0	0.0136	0.01	-26%
σ_m	11.9	10.6	-12%
T_0	29.4	29.1	-1%
T_m	35.9	36	+0%
b_T	2.95	2.74	-7%
Forage scaling			
E_{11}	1.68	3.24	+93%
E_{21}	0.431	0.0501	-88%
E_{31}	0.338	0.05	-85%
Movement			
V_m	0.242	0.153	-37%
σ	0.5	0.475	-5%

5 Likelihood profiling

Likelihood profiling is not embedded in the SEAPODYM workflow in the systematic way it is built into MULTIFAN-CL. The sensitivity of the joint likelihood to the estimated parameters is instead characterised by the variational (adjoint) global sensitivity analysis (Senina *et al.*, 2020a; SEAPODYM Reference Manual, 2022), which returns the exact response of each likelihood component to every parameter across its admissible range (Section 6). The reference manual additionally documents a tool, embedded in the codebase, that projects the likelihood hypersurface onto a pair of parameters, giving a rapid visualisation of a complex likelihood geometry to diagnose pathologies such as non-convergence or spiky gradients. To respond directly to recommendation 677 of the previous review (SC21, 2025), we used the skipjack reference model to profile the principal derived quantity that a classical assessment would profile – the stock size.

The stock size warrants this treatment because, in a spatially explicit model driven by modelled ocean dynamics, local biases in the physical and biogeochemical forcing are inevitable, and the fisheries data themselves are noisy and occasionally carry erroneous effort or catch entries. To reconcile such data with a biased habitat, the model tends to raise the biomass so that fish are present in every cell where catch is observed. SEAPODYM bounds this tendency by imposing a maximum stock size as a penalty term in the joint likelihood. Because the penalty competes directly with the data fit, the choice of its target is not innocuous: too low a ceiling degrades the fit, while too high a ceiling leaves the absolute scale unconstrained. Establishing the stock size therefore requires a series of optimisations and subsequent validations to find the level at which the model reproduces all data types across the full time series without deteriorating the fit in any sub-period.

In the example shown here, the maximum stock was imposed on the WCPO region only, leaving the EPO free, so that the diagnostic isolates the western-Pacific constraint where the bulk of the catch is taken. After estimating the parameters at each candidate stock level, the solution was assessed by the catch-removal method: every fishery was redeclared as catch-removal and run forward, so that the diagnostic measures only whether the predicted biomass field can sustain the observed catches across space, time and age, independently of the effort-based likelihood. The fisheries with the highest catches accumulate most of the local error, so the criterion – fewer than 1% of observed grid cells with relative error exceeding 0.1 – was applied to the two major WCPO purse-seine fleets, fishing in association with floating objects (PS-OBJ) and on free schools (PS-FS). Figure 4 traces this error against the imposed WCPO stock-size ceiling for the estimation period 1995–2012 (black) and the independent validation period 2013–2022 (blue). The PS-OBJ fishery, concentrated in the western equatorial zone and targeting young skipjack, is already sustained at a stock as low as 4000 thousand mt over the estimation period; but this fleet was still developing, and the major expansion of FAD deployment from the 2010s raised its catches enough to require about 6000 thousand mt over the validation period. However, both levels still resulted in high local errors for the free-school fishery over the 2013–2022 validation period, which moves the selected stock size to 7300 thousand mt. Thus, across both fisheries and periods the 1% threshold is crossed at progressively higher stock levels, so that 7300 thousand mt is the smallest ceiling that brings all four error curves below 1%, with the free-school fishery over the validation period as the binding constraint. This ceiling was held fixed across all reference-model fits reported here, including the bridging analysis, so that the population scale is a single validated choice common to every configuration rather than a per-experiment tuning.

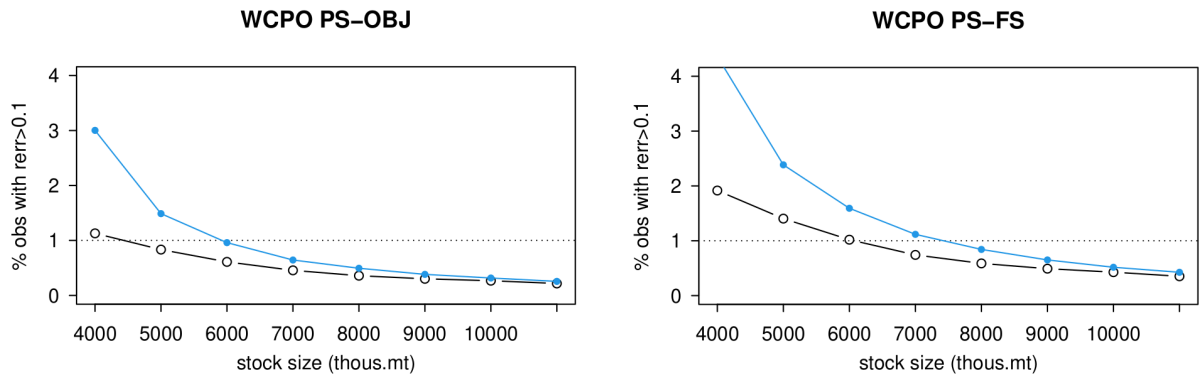

Figure 4: Error profiles against the imposed WCPO maximum stock size for the two major WCPO purse-seine fisheries (left: PS-OBJ, floating objects; right: PS-FS, free schools), predicted by the catch-removal method over the MLE time period 1995–2012 (black line) and the independent validation period 2013–2022 (blue line). The y-axis is the percentage of grid cells with local relative error exceeding 0.1; the x-axis is the imposed maximum stock size (thousand mt).

The most informative feature of Figure 4 is that these error profiles, as well as all likelihood profiles (catch, length, not shown here) do not exhibit a minimum: the error decreases monotonically with stock size and saturates at an asymptote. This is a direct consequence of how absolute scale enters the joint likelihood. The stock-recruitment relationship is not fixed and mortality parameters are re-estimated within their bounds, so demographic parameters are free to produce a higher biomass without penalty, and for the effort-based (catchability) fisheries an increase in biomass is absorbed by a compensating decrease in the estimated catchability, leaving their predicted catches, and hence their likelihood contribution, unchanged. Raising the stock above what the data require therefore costs nothing, while the catch-removal fisheries can only benefit from the additional biomass; the error falls and then flattens rather than taking a parabolic shape. The absolute stock size is thus not identified from above by the MLE. Rather it is only defined from below as a minimal stock that sustains observed fishing pressure given the model spatiotemporal and age distributions.

This reframes the response to recommendation 677. A classical likelihood profile for the stock size, such as in Figures 62–64 of the skipjack stock assessment by Teears *et al.* (2025), cannot be reproduced in SEAPODYM in the sense that it is not obtained by estimation. The scale is penalty-bounded, so the error profiles along with the catch-sustainability criterion presented here constitute an alternative approach for determining the population scale in the spatiotemporal model. Given that this method implies choosing the threshold, it makes the stock size a *validated* choice rather than a likelihood-estimated parameter. Profiling the catch-removal error against the imposed ceiling, applying a cell-fraction threshold to the highest-catch fisheries, and requiring it to hold across the entire validation window is what allows the stock size to be set transparently and reproducibly.

6 Sensitivity to fixed parameters

A reference model is conditioned not only on the data it is fitted to but also on the parameters held fixed during estimation and on the environmental forcing that drives it. The reliability of its conclusions therefore depends on how strongly the solution responds to those fixed choices, and characterising that response is an important part of model development and evaluation. We distinguish four sources of such sensitivity, in increasing order of externality to the fitting procedure: (i) dynamical parameters that the data cannot inform and that are consequently fixed; (ii) parameters that are individually informative but mutually correlated, of which one must be fixed; (iii) input parameters taken from independent studies, each carrying its own uncertainty; and (iv) the ocean forcing itself. We treat each in turn and, where credible alternatives exist, quantify the sensitivity of the fitted solution by re-estimating the remaining parameters under each fixed choice.

Whether a parameter can be estimated at all is established by global sensitivity analysis. SEAPODYM evaluates the sensitivity of each likelihood component to every parameter by the variational (adjoint) method (SEAPODYM Reference Manual, 2022, §4.2.2), which returns the exact gradient of the objective and hence the response of the fit to parameter changes across the admissible range (Senina *et al.*, 2008). A parameter to which the likelihood components are insensitive is not observable: the data carry no information to locate it, its profile is flat, and estimating it would propagate noise rather than signal. Such parameters are fixed at independently justified values. In the skipjack reference model this applies to the stock–recruitment steepness (`a_adults_spawning`), which is observable only through spawning-biomass contrast that the contemporary fishery does not span (Section 7), and to predation mortality at age zero, which the catch and length-frequency data – dominated by older ages – cannot resolve.

A parameter may be individually informative yet inseparable from another. The asymptotic correlations between estimates, read from the off-diagonal of the inverse Hessian at the optimum (Section 7), expose such pairs: as $|\rho|$ approaches one, the two parameters act on the data as a single combined dial and only that combination is determined. Estimation then fixes one member of the pair and retains the other, the choice resting on biological grounds – preferring to constrain the better-known quantity, or an observation nuisance over a dynamical parameter. This choice is not innocuous: a parameter driven to a bound is not a converged estimate but a signal that a correlated fixed parameter has been set wrongly, the optimiser exploiting the remaining freedom to compensate. We illustrate both the diagnosis and its consequence with a one-dimensional sensitivity sweep below.

A third class of fixed parameters is taken directly from independent studies rather than estimated within SEAPODYM. Somatic growth and the maturity ogive are the principal examples: they are derived from dedicated ageing and reproductive studies – for South Pacific albacore, from Farley *et al.* (2013) and Farley *et al.* (2014) – and enter the model as fixed inputs, each with its own observational uncertainty and, occasionally, competing formulations. Where a credible alternative exists, such as a second growth curve or a revised maturity-at-length, we propagate that uncertainty into the reference solution with a dedicated sensitivity run, re-estimating the free parameters under the alternative input and comparing the resulting fits and derived quantities.

Finally, the estimated parameters are sensitive to the ocean forcing itself. Because the optimisation tunes the population dynamics to reproduce the fishery and survey data, it necessarily absorbs the biases of the physical and biogeochemical fields that drive habitat and movement (SEAPODYM Reference Manual, 2022, §4.3); a different forcing relocates the minimum, so each parameter set is conditional on the forcing under which it was obtained (Section 3). Rather than treat any single product as truth, we develop the reference models under two independent forcings – JRA55–NEMO–PISCES and ERA5–NEMO–PISCES – and read the spread between the resulting solutions as a direct, integrated measure of forcing-induced uncertainty.

6.1 Sensitivity to a fixed parameter: σ_m in the albacore model

To assess the sensitivity of the reference-model solution to a fixed, poorly identified parameter, we performed a one-dimensional sweep of σ_m from 1.5 to 7.5 in steps of 0.5 (runs e_01–e_13), with all other parameters re-estimated at each value. Convergence was assessed from the maximum gradient component at the optimum and all thirteen runs (e_01–e_13) converged.

Table 5: Catch, length-frequency, and stock-size likelihood components as a function of fixed σ_m. Total is the sum of the three components shown. Bold values highlight the lowest value in their respective column.

Run	σ_m	Catch	LF	Stock	Total
e_01	1.5	1097800	511410	47938	1657148
e_02	2.0	1094000	514870	33874	1642744
e_03	2.5	1083800	**505980**	**30198**	1619978
e_04	3.0	1070800	506660	30774	1608234
e_05	3.5	1064400	506360	31846	1602606
e_06	4.0	1059900	506030	32386	1598316
e_07	4.5	1057400	506180	32682	**1596262**
e_08	5.0	1058100	507540	33828	1599468
e_09	5.5	1060400	509230	34482	1604112
e_10	6.0	1063100	510950	34933	1608983
e_11	6.5	1047300	523560	36559	1607419
e_12	7.0	1045700	529590	36341	1611631
e_13	7.5	**1044400**	535760	35897	1616057

Total likelihood decreases monotonically from e_01 to e_07 ($\sigma_m = 1.5$ to 4.5), reaching a minimum at e_07, before increasing again towards e_13. e_06 (4.0) and e_08 (5.0) sit close to this minimum, suggesting a broad, shallow optimum between roughly 3.5 and 5.0. Catch likelihood contributes most of the variation whereas length-frequency and stock-size likelihood are comparatively flat, with a small dip in likelihood in the e_04–e_08 range.

Based on this profile, $\sigma_m \approx 4.5$ (e_07) appears preferable, sitting at the minimum of the total likelihood; however, the difference between e_05, e_06, e_07, and e_08 is small relative to the spread of the full range, indicating this parameter is weakly identified by the available data and the choice within $[3.5, 5.0]$ has limited impact on overall fit.

7 Convergence diagnostics and parameter uncertainty

We assess the reliability of each reference-model fit at two levels – the attainment of a genuine minimum and the uncertainty of the estimated parameters – both read from a single Hessian evaluated at the converged parameter vector. Note, the correctness of the function gradient that drives the optimisation, which SEAPODYM evaluates by the hand-written adjoint (reverse-mode) of the forward model, exact to machine precision, is established with the help of the Taylor test by comparing it against finite differences every time the forward code is modified or new functionalities are added (SEAPODYM Reference Manual, 2022).

7.1 Convergence to the minimum

Although in theory the convergence to a minimum requires zero gradient, in practice we deliberately avoid relying on the raw gradient norm or the maximal term of the gradient vector, $\max_i |g_i| < \delta$, as a convergence criterion, because the gradient is not scale-invariant. As it depends on the arbitrary scale of the objective function, a fixed threshold confounds this scale with proximity to the minimum. Rather, at the stopping point, which may be either dictated by the quasi-Newton function minimiser (FM) not making progress, or deliberately chosen when likelihood has stabilised after a number of unsuccessful iterations, we compute the Hessian H of the objective function and summarise it by two scale-aware quantities:

- positive-definite Hessian matrix, so the eigenvalues of H, let us denote them μ_i, must all be positive at a local minimum;
- the Newton decrement $\lambda^2 = g^\top H^{-1} g$, whose half $\frac{1}{2}\lambda^2$ is the decrease in the objective predicted by the local quadratic model on taking a full Newton step to its minimum. The ratio $\frac{1}{2}\lambda^2/L$, with L the objective value, is the relative improvement still attainable. It is robust to the finite-difference step and a scale-invariant diagnostic, when it is small of order $10^{-15} - 10^{-6}$, we can safely declare convergence.

Note, the Hessian is computed separately, using the parfile with parameter estimates. It is approximated by right-side finite differences (left-side near the upper boundary) with Richardson step correction of the exact gradient, which requires two function evaluations per parameter. This makes this computation rather costly, so it is usually performed for the solutions that pass primary validation of all data fits and parameters (see section 8).

An example fit with the tagging likelihood alone illustrates how the diagnostics work. It estimates five parameters governing movement, thermal habitat and growth, and at the stopping point the largest gradient component is $\max_i |g_i| = 5.7$ – well above widely used $\delta < 0.001$, which would report the run as unconverged. The Hessian, however, is positive definite (eigenvalues from 1.4×10^3 to 4.9×10^5), and the Newton decrement gives a remaining objective improvement of $\frac{1}{2}\lambda^2 = 4.4 \times 10^{-5}$ against $L = 102061$ – a relative improvement of 4×10^{-10}, with a largest relative Newton step of 8×10^{-4}. The fit is therefore at the minimum to numerical precision; the large gradient component reflects a stiff, well-determined direction (V_m, see below), not a failure to converge, and it would

change by orders of magnitude under a different likelihood scaling whereas $\frac{1}{2}\lambda^2/L$ would not.

As a direct check that this is the minimum rather than a premature halt, we continued the optimisation until $\max_i |g_i| < 0.1$, reducing the gradient norm from 5.7 to 0.07. The parameter estimates then changed by at most 0.02% and their standard errors by at most 0.13%; the likelihood did not improve (stayed 102061), the Hessian remained positive definite and the Newton decrement fell further to 3×10^{-13}. The additional iterations thus moved the gradient norm, not the solution or its uncertainty, confirming empirically what the decrement anticipated, and that it is the positive-definite Hessian, not the gradient threshold, that defines convergence here.

Beyond this single fit, the convergence to the minimum can be checked by re-optimising the function from jittered initial parameters. If the restarts recover the same solution, i.e., reference estimates, they are thus robust to the choice of starting values. At this stage, other local minima, if inferior to the reference, may be rejected.

7.2 Parameter uncertainty

The inverse Hessian H^{-1} is the asymptotic covariance of the estimates: its diagonal gives the standard errors of Table 6 and its off-diagonals the parameter correlations. For the example fit, the directed-movement scale (V_m) and the thermal-tolerance function slope (b_T) are tightly determined (coefficients of variation of 0.2% and 0.1%), whereas the two diffusion parameters – the base diffusion (σ) and the habitat-dependent "seek" diffusion (c) – are far looser (CV of 25% and 12%) and strongly anti-correlated ($\rho = -0.89$). Consistently, the dominant eigenvector of the covariance matrix, i.e. the least-identified parameter combination, lies almost entirely in the σ–c plane (weights +0.88 on c and −0.47 on σ): the tag data constrain the total diffusivity well but separate its base and seek components only weakly. All other directions are well identified.

Table 6: Estimated parameters, standard errors and coefficients of variation (CV) for the example skipjack reference fit, derived from the inverse Hessian at the converged solution. The fit is at a minimum (positive-definite Hessian, relative remaining improvement 4×10^{-10}).

θ	Parameter description	Estimate	Std. error	CV
σ_0	std. dev. of SST Gaussian (age 0), °C	2.000	0.0170	0.85%
b_T	Allometric power in thermal preferences at age	2.285	0.0025	0.11%
σ	diffusion-rate multiplier	0.0542	0.0136	25.2%
V_m	max. sustainable speed (taxis), BL s^{-1}	1.176	0.0018	0.16%
c	diffusion–habitat coefficient	0.191	0.0238	12.5%

Note one more important diagnostic, which is derived from Hessian's eigenvalues, called condition number $\kappa = |\mu_{\max}|/|\mu_{\min}|$, which, when small ($< 10^3$), indicates a well-conditioned optimisation problem, values $10^3 - 10^6$ indicate a moderate ill-conditioning and larger values – severe ill-conditioning. This number is a robust diagnostic, e.g., in the example above, with both large and small $\max_i |g_i|$, the condition number barely moved, being $\kappa = 353.126$ and $\kappa = 353.122$ respectively. For the ill-conditioned problems, this

number remains meaningful, serving as an indicator of the presence of unobservable (non-identifiable) and correlated parameters.

The σ–c looseness above is one instance of the two forms of weak identification that these diagnostics distinguish. A *trade-off* constrains a combination of parameters while leaving its components individually loose, read from the dominant eigenvector of the covariance; *saturation* instead leaves a single parameter on a flat region of its own functional form, where the fit no longer responds to it. The full joint skipjack fit, which combines tagging, catch and larval data over 36 estimated parameters (Table 1, column 'Ref.'), is far more ill-conditioned than the tag-only fit above ($\kappa = 3.1 \times 10^7$) yet remains firmly at its minimum (relative remaining improvement 4×10^{-6}). The movement and habitat parameters are well-determined and carry none of this ill-conditioning: σ and c parameters were identified with a small $CV = 2\%$ and not cross-correlated. For the full model, its ill-conditioning is caused by a few poorly observable and cross-correlated parameters. Thus, the least-curved (flattest) direction is essentially one *saturated* parameter – the prey (net-primary-production as a proxy) function coefficient of the spawning habitat (α) which is driven by the catch likelihood to high values, at which it reaches a plateau of the likelihood. The dominant correlation direction is a *trade-off* (variance inflation 4.5) among the recruitment scale (σ_0), the larval catchability (q_l) and two mortality parameters. It is therefore an identifiability problem, localised by the diagnostics. The remedies – holding these parameters fixed or constraining them from external information – would not compromise convergence, while likely speeding it up, reducing the standard errors and simplifying the overall diagnostics.

7.3 Structural component of uncertainty

The standard errors above are *conditional*: the inverse Hessian quantifies how well the data determine the parameters within a given minimum, which is obtained with a fixed model structure, a fixed data structure, a fixed partition into estimated and held-fixed parameters, fixed likelihoods and weights, and a fixed ocean forcing. It does not measure the uncertainty in those choices themselves, so the reported standard errors are a lower bound on the total uncertainty. The *structural* component is not read from a single covariance matrix but from an ensemble over structures, and the diagnostics of this section already show it at work: the base and seek diffusion parameters that are loose and strongly anti-correlated in the tag-only fit (σ–c CV 25%/12%, $\rho = -0.89$) become tight and uncorrelated ($CV = 2\%$) in the full joint fit. The same parameters, estimated under a richer structure and larger data set, carry a different uncertainty; the choice of structure therefore acts directly on the parameter uncertainties. The bridging analysis of Section 3 illustrates how the model fit and parameter identifiability depend on the model structure.

The partition into estimated and fixed parameters is structural: holding a parameter fixed removes its variance from H^{-1} but transfers uncertainty onto the parameters correlated with it, so the sensitivity analysis of Section 6 – with the oxygen parameter of the yellowfin model as a worked example – is the necessary complement to the Hessian standard errors. The choice of forcing is likewise structural: the skipjack estimates change only modestly between the ERA5 and JRA55 reanalyses (Senina *et al.*, 2026), evidence that they are structurally robust. This robustness is expected, however,

because ERA5 and JRA55 differ only in their atmospheric forcings, which are both reanalyses, whereas they share the same numerical and parametric setup of the coupled NEMO–PISCES models. It is not guaranteed for more distant forcing choices, which is well illustrated by the density-fitting experiment of Section 4.2.3: after a change in the PISCES biogeochemistry – a modified nitrogen-fixation scheme that lowered tropical primary production in the WCPO – refitting the model to its own density predictions shifted the primary-production-dependent parameters substantially, a direct, controlled measure of the structural uncertainty a change of biogeochemical forcing introduces. Consistently, the skipjack model of Senina *et al.* (2025), which used the JRA55–NEMO–PISCES.v1 forcing, diverged substantially from the models forced with PISCES.v2 (presented in Section 3 and Senina *et al.* (2026), Appendix A) in both its estimated parameters and the modelled tropical distribution of skipjack across the Pacific Ocean.

Together, the conditional standard errors of this section and the results of Sections 3, 4.2.3, and 6 complete the answer to the SC21 request to address the uncertainty in the model estimates (SC21, 2025).

8 Validation with independent data

8.1 Validation framework

Validation of a SEAPODYM solution proceeds in three steps as described in the Reference Manual (SEAPODYM Reference Manual, 2022, §6.3): (i) evaluation of the improvement in fit to the data brought by maximum-likelihood estimation, by comparison with the prior (non-optimised) model; (ii) analysis of the model dynamics under the estimated parameter values, to verify that the resulting spatiotemporal predictions are biologically and oceanographically plausible; and (iii) measurement of the goodness of fit of the MLE solution against an *independent* dataset – defined as any subset of observations that was not used in the likelihood.

Because SEAPODYM is a spatiotemporal model with its dimensions observed through multi-data likelihood, validation is performed at multiple hierarchical levels to ensure consistency across space, time, demographic strata, and fleets. The model is fitted to up to four data types, each one mapped to the underlying age- and space-resolved density field $N(a, t, \mathbf{x})$ through its own observational model: catch/CPUE, length frequency of catch, conventional tag recaptures, and early-life-stage observations (Senina *et al.* 2025; PAW workshop slides, PAW_SEAPODYM-Validation 2026). The validation diagnostics described below are computed for each of these data types and aggregated either as scalar global scores or as spatial maps of grid-cell residuals.

8.2 Statistical metrics by data type

8.2.1 Catch and CPUE

For catch data, validation is performed by fishery, by region and by grid cell, each one aggregated as the time-series or as the spatial map. The fit of the total catch (or the catch by fishery) time series is examined for the absence of systematic bias (mean of residuals

close to zero), of temporal trend (slope of a linear regression of residuals on time close to zero) and of seasonal pattern in the residuals (usually checked at the EEZ level).

Scalar global scores are typically displayed on a Taylor diagram (Taylor, 2001). The spatiotemporal (3D) catch field is collapsed to a single (1D) vector of observations and model prediction. For an observation–prediction dataset (x_i, y_i) of length n, with sample means $\bar{x}, \bar{y}$ and standard deviations σ_x, σ_y, the following base metrics are computed. The squared Pearson correlation

$$r^2 = \frac{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right)^2}{\sigma_x^2\,\sigma_y^2} \tag{7}$$

measures the fraction of variability of the data that is captured by the model; the variance ratio

$$VR = (\sigma_y/\sigma_x)^2 \tag{8}$$

measures whether the model under- or over-disperses the data; and the centred normalised root-mean-square difference (Taylor, 2001)

$$err = \frac{1}{\sigma_x}\left(\frac{1}{n}\sum_{i=1}^{n}\left((x_i - \bar{x}) - (y_i - \bar{y})\right)^2\right)^{1/2} \tag{9}$$

measures how well the model fluctuations around the mean match those of the data. The triple (r^2, VR, err) is the standard set of summaries plotted on a Taylor diagram (Taylor, 2001), where the best possible score is $r^2 = 1$, $VR = 1$, $err = 0$. The Reference Manual (SEAPODYM Reference Manual, 2022, §6.3) provides the historical example in which these three metrics were used to validate the skipjack reference model fitted with catch and length frequencies alone (CL) against the same model fitted with catch, length frequencies and tag recaptures (CLT); the addition of tag data was shown to improve the fit on every metric and every data type.

A solution is considered to pass validation when $r^2 > 0.7$, $0.8 < VR < 1.2$ and $err < 0.5$. The upper bound on VR is rarely binding, because the catch-likelihood term tends to favour diffusion-driven solutions, which reduce the modelled variance and so keep VR below 1. For fisheries with heavily skewed or low-quality data, values of err up to ~ 0.8 are accepted. Likewise, for the fisheries whose catch is derived from effort, $err < 0.8$ may be accepted when the reasons for the misfit are external to the model and cannot be overcome.

At the grid-cell level, three complementary spatial metrics are mapped to detect localised misfit:

$$R^2_{\mathrm{SGF}} = 1 - \frac{\sum_i (x_i - y_i)^2}{\sum_i (x_i - \bar{x})^2}, \qquad MARE = \frac{\sum_i |x_i - y_i|}{\sum_i x_i}, \tag{10}$$

together with a grid-cell Pearson r^2. Tuna catch data are extremely skewed: most of the catch is concentrated in a small "core area" (typically containing $\sim$ 80% of the total catch), and a global mean absolute error is dominated by zero-inflated peripheral cells. Normalisation of the mean absolute error by the data range (NMAE) is therefore used to assess whether the model successfully discriminates "hotspots" from periphery

(good performance: NMAE $\ll 0.1$). Inside the core area, robust performance is defined as $R^2_{\mathrm{SGF}} > 0.5$, $r^2 > 0.7$ and $MARE < 0.3$.

Validation of CPUE is less straightforward than that of catch. Because the SEAPODYM likelihoods fit catch in each grid cell, the fit is implicitly weighted by effort; but observed effort drops sharply at the edges of a fishery's range, so a trivial absolute error on catch can manifest as a very large error on CPUE in those edge habitats. CPUE validation is therefore carried out only at small EEZ scales located inside the species core area, where data coverage is more uniform in time and space and the observed CPUE can be aggregated reliably (PAW_SEAPODYM-Validation 2026). The relevant metrics at this scale are the root-mean-square error and the Pearson r on the seasonal cycle.

Individual fisheries contribute unequally to the global score. Where some fisheries provide higher-quality data than others (e.g. longline catch reporting versus pole-and-line at the basin scale), fishery-specific likelihood weights can be adjusted to improve the overall fit without compromising consistency. In the South Pacific albacore reference model, for example, a custom up-weighting of the longline-fishery likelihoods reduced the relative spatial error of total catch from 65.7% to 58.9% and improved every aggregate score: Pearson r^2 from 0.59 to 0.72, *VR* from 1.1 to 0.82, *err* from 0.7 to 0.53, and R^2_{SGF} from 0.85 to 0.90 (PAW_SEAPODYM-Validation 2026).

8.2.2 Length frequency of catch

The age-structured density predictions are converted into length frequencies of catch via the length-at-age relationship and the fishery-specific age-based selectivity (SEAPODYM Reference Manual, 2022, §6.2). Validation is performed by fishery and at several temporal strata (seasonal, quarterly aggregated, and annual quarter-of-year averages). Two metrics quantify the divergence between observed and predicted length-frequency distributions:

$$NRMSE_{\bar{x}} = \frac{1}{\bar{x}}\sqrt{\frac{1}{n}\sum(x_i - y_i)^2}, \qquad SMD = \frac{\bar{x} - \bar{y}}{\sqrt{\sigma_x^2 + \sigma_y^2}}, \tag{11}$$

where $NRMSE_{\bar{x}}$ (normalised by the mean of observations) measures the divergence in shape between the two distributions across size bins and *SMD* is the standardised bias of the mean lengths. The conventional thresholds for accepting a fishery's length-frequency fit are $NRMSE_{\bar{x}} < 0.5$ and $|SMD| < 0.5$, the latter ensuring that the mean bias remains smaller than half the pooled standard deviation (PAW_SEAPODYM-Validation, 2026). Per-fishery absolute errors on the mean, median and mode of the length-frequency distributions complement the two scalar metrics.

In the SEAPODYM v5 skipjack reference model (Senina *et al.*, 2025), length-frequency fits for the principal fisheries pass these thresholds (e.g. $NRMSE_{\bar{x}} = 0.18$–0.43 for the WCPO pole-and-line fishery and the major purse-seine fleets, both for the 1995–2012 MLE window and the 1979–2022 extended-validation window).

8.2.3 Tagging data

For tag recapture data, validation operates on three views of the predicted distribution: 2D recapture maps, 1D longitudinal and latitudinal profiles aggregated over latitude

bands and longitude bands respectively, and totals by tag recapture group. The metrics used are the variance-normalised root-mean-square error

$$NRMSE_{\sigma_x} = \frac{1}{\sigma_x}\sqrt{\frac{1}{n}\sum_i (x_i - y_i)^2}, \tag{12}$$

the pattern correlation

$$PC = \frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_i (y_i - \bar{y})^2 \, \sum_i (x_i - \bar{x})^2}}, \tag{13}$$

the Kling–Gupta efficiency (Gupta *et al.*, 2009)

$$KGE = 1 - \sqrt{(r-1)^2 + (bias-1)^2 + (vratio-1)^2}, \quad bias = \bar{y}/\bar{x}, \quad vratio = \frac{\sigma_y/\bar{y}}{\sigma_x/\bar{x}}, \tag{14}$$

and the global triple (r^2, VR, err) defined in Eqs. 7–9.

A useful property of the SEAPODYM tagging likelihood is that it allows clean cross-validation by tag-recapture group: scores are evaluated on the recapture groups used in the MLE (typically $\approx 8,000$ recaptured tags of the WCPO PTTP and the EPO tagging programmes that satisfy the time-at-liberty and core-area criteria of Senina *et al.* 2025) and separately evaluated on an equal number of *independent* recapture groups withheld from the MLE. For the skipjack reference model (Senina *et al.*, 2025), the in-sample triple is (r^2, VR, err) = (0.65, 1.0, 0.62) and the out-of-sample triple is (0.72, 0.7, 0.53), with the out-of-sample fit equal to or better than the in-sample fit across all three metrics – evidence that the estimated movement parameters generalise beyond the recapture groups used in the optimisation.

8.2.4 Early-life-stage data

Two types of early-life-stage observation are available for the tropical tunas: (i) categorical larval CPUE from historical Japanese plankton-net surveys (Senina *et al.*, 2025, in the convention of the SEAPODYM v5 early-larvae observation model), and (ii) continuous probabilities of larval presence predicted from environmental covariates by Boosted Regression Tree (BRT) models. Both are compared to the model-predicted early-larvae density field via $NRMSE_{\sigma_x}$ (Eq. 12), the Pearson r on the modelled seasonality, and the larval-data component of the negative log-likelihood. Strata for assessment are spatial (regional aggregates and grid-cell residuals), temporal (seasonality), and environmental (larval counts at SST, dissolved oxygen, primary production and micronekton density). Because early-life data are sparse and not used to define an independent validation window, metrics are evaluated for the MLE data only.

In the SEAPODYM v5 skipjack reference model (Senina *et al.*, 2025), the early-larvae predictions satisfy the conventional robustness threshold of $NRMSE_{\sigma_x} < 0.5$ when projected onto each of four environmental covariates (SST: 0.32, O_2: 0.27, primary production: 0.40, micronekton density: 0.40); the same scores for a candidate solution *without* the early-larvae observation model degrade to 0.48, 0.73, 0.74 and 0.65 respectively, confirming that the early-life-stage data are informative for the spawning-habitat parameters and that the v5 early-larvae model captures their environmental dependence (PAW_SEAPODYM-Validation, 2026).

8.3 Aggregation of metrics into global scores

Most of the metrics in Sections 8.2.2–8.2.4 are computed within a stratum – a fishery for the length-frequency data, a tag-recapture group for the tagging data, or an environmental-factor bin for the early-life-stage data. To summarise the fit of a data type by a single global score, the per-stratum values $\{M_j\}$ are combined with weights $\{w_j\}$ that reflect the importance of each stratum.

The standardised mean bias *SMD* (Eq. 11) and the Kling–Gupta efficiency *KGE* (Eq. 14) are aggregated as a weighted arithmetic mean, which preserves the sign of the metric:

$$\overline{M} = \frac{\sum_j w_j M_j}{\sum_j w_j}. \tag{15}$$

Note that for *SMD*, Eq. 15 gives the *net bias*.

By contrast, the error metrics, i.e., the normalised root-mean-square errors $NRMSE_{\bar{x}}$ (Eq. 11) and $NRMSE_{\sigma_x}$ (Eq. 12), are aggregated as a weighted quadratic mean, consistent with the additivity of squared errors:

$$\overline{M} = \left(\frac{\sum_j w_j M_j^2}{\sum_j w_j} \right)^{1/2}. \tag{16}$$

For the pattern correlation, *PC* (Eq. 13), which is bounded on $[-1, 1]$ and whose sampling distribution is strongly skewed near the bounds, Eq. 15 is applied to the Fisher z-transformed values and the resulting weighted mean is back-transformed:

$$\overline{PC} = \tanh\left(\frac{\sum_j w_j z_j}{\sum_j w_j} \right), \; z_j = \operatorname{arctanh}(PC_j), \tag{17}$$

which removes the downward bias incurred when correlation coefficients are averaged directly (Silver and Dunlap, 1987).

The weights used in Eqs. 16–17 differ between data types. For length-frequency data, w_j is the sampled catch of fishery j, so that minor or poorly sampled fisheries contribute proportionally less to the global score. Note that the sampled catch must be expressed in the same unit across all fisheries for the weighting to be consistent. For tagging data, w_j is the number of recaptures in group j. The early-life-stage metrics are aggregated over their environmental factors with equal weight ($w_j \equiv 1$), for which Eq. 16 reduces to the unweighted quadratic mean.

8.4 Validation of parameter estimates

The statistical scores summarised above evaluate whether the optimised model reproduces the data. A complementary, equally important step is to verify that the estimated parameter values are biologically and physically realistic and consistent with independent studies. The Reference Manual (SEAPODYM Reference Manual, 2022, §6) formalises this as the recommendation to perform jitter optimisations to verify convergence to a common minimum (Section 7) and to confront parameter estimates with independent observations whenever such observations exist.

In practice two situations are recurrent. First, parameter estimates may converge at a boundary of the admissible interval. For example, for the yellowfin tuna reference model, early MLE runs estimated the thermal preference of adults at the upper bound (36°C) with large temperature tolerance (8°C), accompanied by a poor fit of the purse-seine catch in the EPO. Inspection of the forcing fields showed that the layer-1 depth used by SEAPODYM exceeds the mixed-layer depth, and in particular the $2\,\mathrm{ml\,l^{-1}}$ oxycline in the OMZ; correcting the layer-1 temperature towards the value of the surface layer that yellowfin occupy in such conditions produced a realistic thermal estimate ($T_m = 26.8$°C, $\sigma_m = 4.5$°C, $b_T = 2.7$) and a substantial improvement of the EPO purse-seine catch fit (e.g. Pearson $r = 0.82 \rightarrow 0.91$ for fishery S7) (PAW_SEAPODYM-Validation, 2026). Second, parameter estimates can be confronted with independent biological datasets. For example, parameters describing movement and habitat preference can be validated against archival tagging data, which provide individual-fish records of depth and ambient temperature. Preliminary analyses of WCPO archival data (Scutt Phillips *et al.*, 2022, to be extended with EPO archival data) support the SEAPODYM hypothesis that the depth of yellowfin diel migration decreases as the vertical thermal gradient strengthens, and allow the MLE age-dependence of thermal habitat to be compared with the in-situ ambient temperatures recorded on tagged fish.

8.5 Validation against independent data

Two notions of *independent* data are used in SEAPODYM validation. The first – and the one most directly relevant to recommendation 677 of the SC21 Summary Report (SC21, 2025) – is temporal: the MLE is run on a subset of the available time series, and the goodness of fit is evaluated separately on the data inside the MLE window and on the data outside it. This decomposition isolates the sensitivity of model conclusions to the choice of estimation period and tests whether the reference models retain predictive skill before and after that window. For the SEAPODYM v5 skipjack reference model (Senina *et al.*, 2025), the MLE window is 1995–2012; the goodness-of-fit metrics for catch and length-frequency data over the extended window 1979–2022 are within the same robustness thresholds as those obtained on the MLE window (e.g. length-frequency $NRMSE_{\bar{x}} = 0.18$–0.48 over 1979–2022 versus 0.19–0.48 over 1995–2012; tagging out-of-MLE-group metrics reported in Section 8.2.3). The full set of out-of-window diagnostics for skipjack, yellowfin and bigeye is reported in Senina *et al.* (2026).

The second notion of independent data is spatial: a SEAPODYM solution obtained for one ocean basin can be applied to simulate the dynamics of the same species in a different basin and compared, on the second basin, to a baseline MLE solution fitted to the second basin's own data. The Reference Manual (SEAPODYM Reference Manual, 2022, §6.3) describes this strategy on the South Pacific albacore reference model of Senina *et al.* (2020b), whose MLE solution was used to simulate the Atlantic albacore stock and shown to reproduce the Atlantic dynamics with statistical metrics comparable to those of the Atlantic baseline MLE. Cross-basin validation is most informative for species for which several ocean-basin reference models exist.

9 Discussion

Addressing the request made by the previous Scientific Committee review, this paper has described the methodology behind the SEAPODYM reference models for tropical tunas and the analyses that support their credibility. Five results are worth discussing: i) the bridging analysis and what it implies for how the model integrates the impact of the environment on tuna dynamics; ii) the robustness and reproducibility of the adjoint-based estimation; iii) the sensitivity of the reference solutions to the choice of estimation period; iv) the caveats in the estimation of the population scale with the mechanistic spatiotemporal model; and v) the role of ocean forcings in developing reference models.

Bridging analysis and environmental drivers. The bridging analysis was set up not as a reduction to a no-movement, no-environment form – which is neither a parametric sub-case of a standard fishery model nor biologically coherent for a highly migratory species – but as two build-up sequences from physically plausible baselines, each removing one group of mechanisms from the reference. Along the movement axis, replacing constant with habitat-dependent diffusion barely reshaped the field, whereas restoring active taxis more than halved the tagging misfit and removed the catch deterioration of the over-dispersed passive solution; along the recruitment axis, an SST-only spawning habitat could not support the observed catch, which the larval-food and then the predation term progressively restored. The two axes give a clean process-to-data attribution: active taxis is required primarily by the tagging data (and, to a lesser extent, by the catch), while the environmentally driven structure of recruitment is required by the catch. Each mechanism is therefore a necessary component rather than additional complexity, and each is informed by a distinct dataset – tagging informs movement, and catch informs recruitment. These results reframe the review's concern over "strong assumptions": the shape and magnitude of the relationships between population processes and the environment are estimated, not assumed – indeed a driver can be switched off by the estimation itself, as in the null food-dependence of skipjack adult and juvenile mortality ($\varepsilon_a = 0$) or the near-vanishing prey coefficient of the bigeye spawning habitat.

We stress that the present reference structure is not a final, universal model, but a stage in a framework that evolves as new data are integrated. For example, if archival-tagging records of vertical habitat use are brought into the estimation – preliminary WCPO analyses have shown that the depth of yellowfin diel migration is linked to the vertical thermal gradient (Scutt Phillips *et al.* 2022; Section 8.4) – additional environmental variables are likely to enter the habitat description. The mixed-layer depth is a natural candidate, acting as a physical barrier to the diel vertical migration of many pelagic organisms and thereby shaping the vertical, and hence the accessible horizontal, distribution of habitat.

Robustness and reproducibility of the estimation. The reliability of the reference models rests on an estimation method that is both robust and reproducible. The objective gradient is supplied by a hand-coded adjoint of the forward model and verified against finite differences by a Taylor test whenever the code changes. That the method recovers truth is shown directly by identical-twin experiments: fitting SEAPODYM to data simulated from a known parameter set recovered those values within their estimated standard

errors in the cases examined to date; for the South Pacific albacore configuration, eight of ten jittered restarts converged to an identical solution – to four significant figures – reproducing every biological parameter to within 9% and fitting the synthetic catch almost exactly.

Convergence of the reference fits is established by scale-invariant diagnostics – the Newton decrement, the condition number and a positive-definite Hessian. These diagnostics confirm that a solution is at a minimum even where the raw gradient norm is misleading. The inverse Hessian additionally yields the parameter standard errors and correlations. Implemented numerical tools turn parameter estimation into a controlled workflow: global sensitivity analysis establishes which parameters the data can inform, jittered restarts help avoid local minima, the inverse-Hessian correlations expose parameters that act as a single combined pair, weakly unobservable or degenerate parameters can be fixed at independently justified values, and a parameter driven to a bound is considered not as a converged estimate but as a signal that a correlated fixed parameter was set wrongly. The few weakly identified quantities in the skipjack model, the base and habitat-dependent diffusion, which degenerate under tagging alone but are resolved by the joint likelihood, and the saturated prey coefficient of the spawning habitat, are handled without compromising the fit.

Validation and sensitivity to the estimation window. Recommendation 677 asked specifically whether model conclusions depend on the period used for estimation. Decomposing the goodness of fit into the data inside and outside the MLE window answers this directly: for the skipjack reference model, estimated over 1995–2012, the catch and length-frequency metrics over the extended 1979–2022 window remained within the same robustness thresholds as those on the estimation window (length-frequency $NRMSE_{\bar{x}}$ of 0.18–0.48 versus 0.19–0.48), and the tagging fit evaluated on recapture groups withheld from the estimation was equal to or better than the in-sample fit on every metric (out-of-sample $(r^2, VR, err) = (0.72, 0.7, 0.53)$ against in-sample $(0.65, 1.0, 0.62)$). The reference solutions therefore retain predictive skill before and after the estimation window, and their movement parameters generalise beyond the data used to fit them.

Estimation of the stock size. Although we can readily evaluate management-relevant quantities, estimating absolute population size is not the primary focus of our quantitative spatiotemporal modelling. The absolute population scale is highly dependent on the model configuration, in particular the ocean forcings and spatial resolution. Because SEAPODYM is a spatiotemporal model fitted to catch data at the grid-cell level, the simulated biomass must be sufficient to support the recorded catch in those specific cells. Consequently, local biases in the forcing data and the limited spatial heterogeneity of a coarse grid can artificially inflate the overall stock estimate. Therefore, the selected scale must be re-evaluated whenever the forcing and/or spatial resolution is refined.

Sensitivity to the ocean forcing. A dependence of model dynamics on oceanographic drivers implies that the model parameters must be re-estimated every time the ocean variables change. Fitting to the observed datasets, the optimisation process absorbs absolute biases inherent in the physical and biogeochemical forcings. In the yellowfin model, for example, a reduction in modelled primary production in a different forcing is compensated

by a lower prey-function coefficient, leaving recruitment essentially unchanged. Because each parameter set is strictly conditional on its specific forcing, we do not treat any single model output as absolute truth. Instead, as the computational performance and estimation framework of SEAPODYM continue to improve, we aim to develop reference models under several independent forcings (e.g. ERA5–NEMO–PISCES, JRA55–NEMO–PISCES) and under alternative assumptions for the most uncertain parameters, and to use the spread between these solutions as an integrated measure of structural uncertainty.

Taken together, methods and analyses presented in this paper support the tropical-tuna reference models as validated, uncertainty-quantified and stable with respect to the estimation period, while making explicit that the model structure itself remains open to refinements as new ocean forcings and new observational datasets become available.

10 Declaration of generative AI and AI-assisted technologies in the SC paper preparation process

The generative AI tool Claude was used to assist with drafting, editing and language proofreading of the text in this document, in particular converting the R code of the model validation tools into LaTeX mathematical notation, searching for relevant literature on validation metrics, and typesetting and formatting the tables and equations. All analytical choices, the design of the presented analyses, model specifications, diagnostic interpretations, and conclusions are those of the authors, who reviewed and verified all AI-assisted output. The methodology, model implementation, and scientific conclusions were developed by, and are the sole responsibility of, the authors. The SEAPODYM model and its codebase were developed independently, without generative-AI assistance.

Bibliography

Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., and M. Gehlen. 2015. PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies. *Geoscientific Model Development*, 8, 2465–2513. https://doi.org/10.5194/gmd-8-2465-2015

Barrier, N., Lengaigne, M., Rault, J., Person, R., Ethe, C., Aumont, O., and O. Maury. 2023. Mechanisms underlying the epipelagic ecosystem response to ENSO in the tropical Pacific. Progress in Oceanography, 208, 103002. https://doi.org/10.1016/j.pocean.2023.103002

Bell, J., Senina, I., Adams, T., ... Tsamenyi, M., Williams, P. 2021. Pathways to sustaining tuna-dependent Pacific Island economies during climate change. Nature Sustainability, 2021, 4(10), pp. 900–910.

Bertignac, M., Lehodey, P., and J. Hampton. 1998. A spatial population dynamics simulation model of tropical tunas using a habitat index based on environmental parameters. *Fisheries Oceanography* 7(3–4): 326–334.

Bonnin, L. and I. Senina. 2024. SEAPODYM: Integration of new data to inform modelling of tuna's early life stages dynamics. In The 20st Regular Session of the Scientific Committee of the Western Central Pacific Fisheries Commission, 14–21 August 2024, Manila, Philippines. Volume WCPFC-SC20-2024/EB-IP-33.

Buenafe, K.C.V., Everett, J.D., Dunn, D.C. et al. 2022. A global, historical database of tuna, billfish, and saury larval distributions. Sci Data 9, 423. https://doi.org/10.1038/s41597-022-01528-7.

Castillo Jordán, C., Teears, T., Hampton, J., Davies, N., Scutt Phillips, J., McKechnie, S., Peatman, T., MacDonald, J., Day, J., Magnusson, A., Scott, R., Scott, F., Pilling, G., and P. Hamer. 2022. Stock assessment of skipjack tuna in the western and central Pacific Ocean. In *The 18th Regular Session of the Scientific Committee of the Western and Central Pacific Fisheries Commission*, 10–18 August 2022, Online, Volume SC18-SA-WP-01 - Rev.05.

Dunn, A., and D. N. Webber. 2020. Review of SEAPODYM, including recent developments and as an ecosystem model for tropical tunas and important bycatch species in the Western Pacific Ocean. Ocean Environmental Ltd, Wellington, New Zealand.

Dunn, A., Rasmussen, S., and S. Mormede. 2018. Spatial Population Model User Manual, SPM v1.1-2018-05-31 (rev. 1291). NIWA Technical Report 138. NIWA, Wellington, New Zealand, 210 p.

Dunn, A., Hoyle, S. D., and S. Datta. 2020. Development of spatially explicit operating models for yellowfin tuna populations in the Indian Ocean. In *The 22nd Session of the IOTC Working Party on Tropical Tunas (Assessment Meeting)*, Victoria, Seychelles, IOTC-2020-WPTT22(AS)-19.

Farley, Jessica H., Ashley J. Williams, et al. 2013. Reproductive Dynamics and Potential Annual Fecundity of South Pacific Albacore Tuna (Thunnus alalunga). PLOS One 8.4. https://doi.org/10.1371/journal.pone.0060577.

Farley, Jessica H., Simon D. Hoyle, et al. 2014. Maturity Ogives for South Pacific Albacore Tuna (Thunnus alalunga) That Account for Spatial and Seasonal Variation in the Distributions of Mature and Immature Fish. en. In: PLOS One 9.1. https://doi.org/10.1371/journal.pone.0083017.

Forestier, R., Bonnin, L., Senina, I., Vidal, T., Ghergariu, M., and S. Nicol. 2025. A comprehensive method to integrate unbiased fisheries data in spatially-explicit population dynamics models. In *The 21st Regular Session of the Scientific Committee of the Western and Central Pacific Fisheries Commission*, 13–21 August 2025, Nuku'alofa, Tonga. Volume WCPFC-SC21-2025/SA-IP-21.

Fournier, D. A., Hampton, J., and J. R. Sibert. 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.

Fujioka, K., Aoki, Y., Tsuda, Y., Okamoto, K., Tsuchida, H., Sasaki, A., and H. Kiyofuji. 2024. Influence of temperature on hatching success of skipjack tuna (Katsuwonus pelamis): Implications for spawning availability of warm habitats. *Journal of Fish Biology* DOI: 10.1111/jfb.15767

Goethel, D. R., Berger, A. M., Hoyle, S. D., McKechnie, S., Vincent, M. T., Davies, N., Hampton, J., Castillo Jordán, C., Kolody, D. S., and A. E. Punt. 2024. Driving with your eyes closed: results from an international, blinded simulation experiment to evaluate spatial stock assessments. DOI: 10.1111/faf.12819.

Gupta, H. V., Kling, H., Yilmaz, K. K., and G. F. Martinez. 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology* 377(1–2): 80–91.

Hampton, J., and D. A. Fournier. 2001. A spatially disaggregated, length-based, age-structured population model of yellowfin tuna (*Thunnus albacares*) in the western and central Pacific Ocean. *Marine and Freshwater Research* 52(7): 937–963.

Lehodey, P., Senina, I., and R. Murtugudde. 2008. A spatial ecosystem and populations dynamics model (SEAPODYM) – modeling of tuna and tuna-like populations. *Progress in Oceanography* 78(4): 304–318.

Lehodey, P., Murtugudde, R., and I. Senina. 2010. Bridging the gap from ocean models to population dynamics of large marine predators: a model of mid-trophic functional groups. *Progress in Oceanography* 84(1–2): 69–84.

Lehodey, P., Senina, I., Calmettes, B., Hampton, J., and S. Nicol. 2013. Modelling the impact of climate change on Pacific skipjack tuna population and fisheries. *Climatic Change* 119(1): 95–109.

Lehodey, P., Conchon, A., Senina, I., Domokos, R., Calmettes, B., Jouanno, J., Hernandez, O., and R. Kloser. 2015. Optimization of a micronekton model with acoustic data. *ICES Journal of Marine Science* 72(5): 1399–1412.

Methot, R. D., and C. R. Wetzel. 2013. Stock Synthesis: a biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research* 142: 86–99.

Nishikawa, Y., Honma, M., Ueyanagi, S., and Kikawa, S. 1985. Average Distribution of Larvae of Oceanic Species of Scombroid Fishes, 1956–1981. (Far Seas Fisheries Research Laboratory, 1985).

Senina, I., Bonnin, L., and R. Forestier. 2026. Validation of SEAPODYM models. Presentation to the SPC Pre-Assessment Workshop, 24–27 March 2026, Noumea, New Caledonia, https://github.com/PacificCommunity/ofp-SAM-SPC-PAW

Western and Central Pacific Fisheries Commission. 2025. Summary report of the 21st Regular Session of the Scientific Committee. WCPFC, Pohnpei, Federated States of Micronesia, posted 3 November 2025.

Scutt Phillips, J., Lehodey, P., Hampton, J., Hamer, P., Senina, I., and S. Nicol. 2022. Quantifying rates of mixing in tagged WCPO skipjack tuna. In *The 18th Regular Session of the Scientific Committee of the Western and Central Pacific Fisheries Commission*, 10–18 August 2022, Online, Volume SC18-SA-WP-04.

SEAPODYM: spatial ecosystem and population dynamics model for migratory age-structured stocks [version 4.0]. Reference Manual, Pacific Community, 186pp., https://purl.org/spc/digilib/doc/vxsz3

Senina, I., Sibert, J., and P. Lehodey. 2008. Parameter estimation for basin-scale ecosystem-linked population models of large pelagic predators: application to skipjack tuna. *Progress in Oceanography* 78(4): 319–335.

Senina, I., Lehodey P., Sibert J., Hampton J., 2020. Improving predictions of a spatially explicit fish population dynamics model using tagging data. Can. J. Fish. Aquat. Sci., 18 pp, dx.doi.org/10.1139/cjfas-2018-0470.

Senina, I., Lehodey P., Hampton J., Sibert J. 2020. Quantitative modelling of the spatial dynamics of South Pacific and Atlantic albacore tuna populations. Deep Sea Research II, doi.org/10.1016/j.dsr2.2019.104667.

Senina, I., Bonnin, L., Lengaigne, M., Kiyofuji, H., Buenafe, K., Fuller, D., Hampton, J., and S. Nicol. 2025. Reference model of skipjack tuna in the Pacific Ocean using SEAPODYM integrating catch, length, tagging, and early-life history data. In *The 21st Regular Session of the Scientific Committee of the Western and Central Pacific Fisheries Commission*, 13–21 August 2025, Nuku'alofa, Tonga. Volume WCPFC-SC21-2025/SA-IP-20.

I. Senina, L. Bonnin, R. Forestier, M. Lengaigne, K. Guihou, J. Temple-Bouyer, and S. Nicol. 2026. SEAPODYM applications to Pacific tuna species In *The 22nd Regular Session of the Scientific Committee of the Western and Central Pacific Fisheries Commission*, 11–19 August 2026, Apia, Samoa. Volume WCPFC-SC22-2026/EB-IP-03.

Silver, N. C. and Dunlap, W. P. 1987. Averaging correlation coefficients: Should Fisher's z transformation be used? Journal of Applied Psychology, 72(1), 146–148. https://doi.org/10.1037/0021-9010.72.1.146

SPC OFP. 2025. Updates on the 2019 SEAPODYM Review. In *The 21st Regular Session of the Scientific Committee of the Western and Central Pacific Fisheries Commission*, 13–21 August 2025, Nuku'alofa, Tonga. Volume WCPFC-SC21-2025/EB-WP-06.

Teears, T., Davies, N., Hamer, P., Magnusson, A., Peatman, T., Scutt Phillips, J., Pilling, G., Scott, R., Vidal, T., and J. Hampton. 2025. Stock assessment of skipjack tuna in the western and central Pacific Ocean. In *The 21st Regular Session of the Scientific Committee of the Western and Central Pacific Fisheries Commission*, 13–21 August 2025, Nuku'alofa, Tonga. Volume WCPFC-SC21-2025/SA-WP-02.

Taylor, K. E. 2001. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres* 106(D7): 7183–7192.

List of Tables

1 Estimated and fixed parameters of the five bridging configurations. * fixed (not estimated); [/] estimated at the lower / upper boundary; a dash (−) marks a parameter made ineffective by the disabled process. Parameters held fixed at the same value in all five runs are omitted. Negative log-likelihood components: C catch, L length-frequency, T tagging, E early-life. 14

2 Biological parameter recovery in the twin experiment. The reference column shows parameter values used to generate the artificial dataset. The recovered column shows the value returned by all 8 converged jitter runs (identical across runs to 4 significant figures). Relative error is $(recovered - reference)/|reference| \times 100$. 20

3 Accuracy of the optimisation process to estimate model parameters, depending on quality and quantity of observations 21

4 Sensitivity of MLE to ocean forcing 24

5 Catch, length-frequency, and stock-size likelihood components as a function of fixed σ_m. Total is the sum of the three components shown. Bold values highlight the lowest value in their respective column. 28

6 Estimated parameters, standard errors and coefficients of variation (CV) for the example skipjack reference fit, derived from the inverse Hessian at the converged solution. The fit is at a minimum (positive-definite Hessian, relative remaining improvement 4×10^{-10}). 30

List of Figures

1 Spatial distribution of recruits (left) and total biomass (right) for the bridging configurations (rows, top to bottom): B_1 (passive, constant diffusion), B_1^+ (passive, habitat-dependent diffusion), B_2 (reference movement, SST-only spawning habitat), B_2^+ (reference movement, SST and larval-food spawning habitat), and the reference model. Circles show observed catches; averages over 2013–2022. 15

2 Monthly time series of recruits (top, millions of fish) and total biomass (bottom, millions of metric tons) aggregated over the WCPO (100°E–150°W, left) and EPO (150°W–70°W, right) regions (40°S–45°N), for the five bridging configurations (legend in the top-left panel; labels as in Figure 1), over 1983–2022. 16

3 Accuracy of the optimisation process to estimate model parameters, depending on quality and quantity of observations 22

4 Error profiles against the imposed WCPO maximum stock size for the two major WCPO purse-seine fisheries (left: PS-OBJ, floating objects; right: PS-FS, free schools), predicted by the catch-removal method over the MLE time period 1995–2012 (black line) and the independent validation period 2013–2022 (blue line). The y-axis is the percentage of grid cells with local relative error exceeding 0.1; the x-axis is the imposed maximum stock size (thousand mt). 26