# ANNEX XX

*26th Meeting of the*
*International Scientific Committee for Tuna*
*and Tuna-Like Species in the North Pacific Ocean*
*Taipei City, Chinese Taipei*
*22-29 June 2026*

# INTERNATIONAL SCIENTIFIC COMMITTEE FOR TUNA AND TUNA-LIKE SPECIES IN THE NORTH PACIFIC OCEAN (ISC) 2024 PACIFIC BLUEFIN TUNA STOCK ASSESSMENT PEER REVIEW [1]

**JUNE 2026**

[1] Prepared for the 26th Meeting of the International Scientific committee on Tuna and Tuna-like Species in the North Pacific Ocean (ISC) held June 22-29, 2026, in Taipei City, Chinese Taipei. Document should not be cited without permission of the authors.

Left Blank for Printing

# International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean (ISC) 2024 Pacific Bluefin Tuna Stock Assessment Peer Review

20–23 March 2026
Hokkaido University, Sapporo, Japan

**Reviewers**

Allan Hicks: allan@fisheriesanalysis.com
Sheng-Ping Wang: wsp@mail.ntou.edu.tw
Tamaki Shimose: shimose@iwate-u.ac.jp

**Chair**

Robert Ahrens robert.ahrens@noaa.gov

**List of acronyms**

ASPM – Age-structured production model
CAAL – Conditional age-at-length
CCSBT – Commission for the Conservation of Southern Bluefin Tuna
CDS – Catch documentation scheme
CKMR – Close kin mark recapture
CPUE –Catch per unit effort
CV – Coefficient of variation
EPO – Eastern Pacific Ocean
GLMM – Generalized linear mixed model
IQ – Individual quotas
ISC – The International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean
LL – Longline
M – Natural mortality
OSA – One-step ahead
PBF – Pacific bluefin tuna
PBFWG – Pacific bluefin tuna working group
PDO – Pacific decadal oscillation
RTM – Real-time monitoring
SRR – Stock-recruitment relationship
SS – Stock Synthesis
SSB – Spawning Stock Biomass
SST – Sea surface temperature
TAWA – Tuna-larvae Abundance in Western-pacific Area
TOR – Terms of reference
VBGF – von Bertalanffy growth function
VDR – Voyage data recorder
VMS – Vessel monitoring systems
WCNPO – Western and Central North Pacific Ocean
WG – Working group
WPO – Western Pacific Ocean

## INTRODUCTION

This document is a consensus report following the review of the International Scientific Committee for tuna and tuna-like species (ISC) stock assessment of Pacific Bluefin tuna in the North Pacific Ocean. This report was prepared by Dr. Allan Hicks, Dr. Sheng-Ping Wang, and Dr. Tamaki Shimose with support from the review chair Dr. Robert Ahrens. The in-person review meeting took place 20-23 March 2026 at the Hokkaido University in Sapporo, Japan. The review meeting was structured according to terms of reference (TOR) provided by the ISC Pacific Bluefin Tuna Working Group (PBFWG). Presentations by PBFWG members were followed by discussions with the review panel members. The PBFWG conducted additional analyses during the meeting when requested, which helped the panel understand the assessment and its input data. Materials provided to the panel included the 2024 assessment base case stock synthesis (SS) files, the 2024 stock assessment report, and background documents describing the development of model inputs and methods for diagnoses. The final day of the review involved further discussion of each of the panel's recommendations to ensure comprehensiveness and clarity. This process naturally led to consensus in the recommendations and as a result the review panel and the chair felt that a single consensus report was the most effective approach to communicate their recommendations to the ISC and PBFWG.

The Review Panel and the review chair (the Panel) thank the members of the ISC PBFWG who attended the review meeting for their hard work, patience with the Panel, and exceptional support. They also thank the Hokkaido University for their generous hospitality during the review meeting and the Japan Fisheries Research and Education Agency (FRA) for hosting an excellent reception.

This document has three sections. The short introduction provides background to outline the motivations and history that led to this review. This is followed by a discussion section which is organized by TOR and provides a narrative of the main topics discussed relevant to the panel's recommendations. The final section provides the panel's specific recommendations and is also organized by TOR. Recommendations have been separated into “short-term” and “long-term”; the former are intended to be addressed for the next stock assessment, and the latter are larger efforts that may take longer to reach resolution.

## 1 BACKGROUND

The 2024 benchmark stock assessment of Pacific Bluefin tuna (*Thunnus orientalis*, PBF) was conducted in the Stock Synthesis modeling platform by the PBFWG using fishery catches, CPUE, and size compositions up to the 2022 fishing year (2022). The previous stock assessment (2022) estimated the 2020 spawning biomass to be between the biomass rebuilding targets of 6.3% $SSB_0$ and 20%$SSB_0$ and projected to be above the 20%$SSB_0$ by the 2029 fishing year. The 2024 stock assessment used the underlying structure of the 2022 stock assessment and addressed many issues including 1) conducting new research on abundance indices, 2) externally estimating variability in the length-at-age, 3) reducing the retrospective pattern in the SSB, and 4) modifying some assumptions to produce a more stable and robust stock assessment.

The terms of reference developed by the PBFWG outline the objectives and scope of an external peer review of the 2024 stock assessment of PBF. They are based upon the peer review process approved by the ISC Plenary in 2024. It is expected that the outcomes and recommendations from this peer review would be incorporated, to the extent possible, in the next PBF assessment, tentatively scheduled for 2027.

# 2 OVERALL CONCLUSIONS OF THE PEER REVIEW

The benchmark stock assessment represents a significant methodological advance over previous assessments, particularly through the adoption of a shortened time series (1983–2022) as the base-case model and improvements to the treatment of key CPUE indices. The shortened time-series starts at a period following relatively moderate catches, except for 1981 and 1982 when the catches were high at a time of estimated low spawning biomass. Catches from 1983 to 1993 were low. The single base case assessment concludes that the stock has rebuilt substantially from historically low levels in 2010, with the 2022 spawning stock biomass (SSB) estimated at 23.2% of $SSB_0$, exceeding the second rebuilding target of 20%$SSB_0$.

The Panel finds that the assessment is generally well-constructed, employs appropriate diagnostic tools with significantly improved retrospective patterns, and provides a defensible characterization of recent stock trends. The hindcasting skill of the model is commendable, and the transparency with which the PBFWG has documented methodological decisions and identified unresolved issues is appreciated.

The Panel agrees on two major topics to improve the assessment of PBF in the future. First, improving the characterization of growth, including variability of length-at-age, would improve fits to the size compositions and the estimation of uncertainty. Second, the removal of recent years of the recruitment index from the model improved the retrospective pattern, but left recent recruitment uninformed. The development of a recruitment index would greatly benefit the assessment of PBF.

# 3 OBJECTIVES

The purpose of the review of the ISC 2024 PBF assessment is to provide information that will improve the analyses used for providing management advice based on the best available science. To this end, the goals and objectives of this review are to:

1. Conduct a peer review of the 2024 PBF stock assessment in accordance with the guidelines described in “Process for the Independent Review of ISC Stock Assessments”.
2. Based on the TOR, provide both short-term and long-term recommendations to improve future assessments scheduled for 2027 and beyond, including but not limited to data inputs, modeling approaches, and addressing uncertainty, etc.
3. In cooperation with PBFWG members, identify options for future research and data collection that will improve future assessments.

# 4 KEY MOTIVATIONS AND DISCUSSION RELATED TO RECOMMENDATIONS

This section of the report contains noteworthy discussion points and key factors that lead to the recommendations related to the subcomponents for each term of reference. Each of the broad topics in the TOR regarding biological assumptions, input data, and model assumptions is concerned with the above-mentioned two major areas of improvement for the current assessment. Therefore, a consistent approach will be necessary for each broad topic in the TOR in order to develop an internally consistent model. For this reason, there may be some duplication of the description in the sub-sections below. In this section, we organize the issues, discussions, and findings under each of the TORs so that the reader has some context for each recommendation.

## 4.1 TOR 1: REVIEW THE BIOLOGICAL ASSUMPTIONS OF THE PBF ASSESSMENT AND PROVIDE PRACTICAL RECOMMENDATIONS TAKING INTO ACCOUNT THE BEST AVAILABLE LIFE HISTORY INFORMATION FOR THIS SPECIES (E.G., LENGTH-WEIGHT RELATIONSHIP, MATURITY, NATURAL MORTALITY, GROWTH, AND STOCK-RECRUITMENT RELATIONSHIP).

### *4.1.1 STOCK STRUCTURE*

Pacific Bluefin tuna in the Pacific Ocean are born in the Western Pacific Ocean (WPO), may migrate to the Eastern Pacific Ocean (EPO) before spawning in the WPO. A small amount of PBF are found in the southern Pacific Ocean. The PBF across this range is considered to be a single stock as spawning is believed to occur only in the WPO. Instead of a spatial model with migration between areas, a single area model using fleets-as-areas was developed for the stock assessment.

The fleets for area approach with time-varying selectivity confounds the model's ability to distinguish abundance from availability. When a fleet's catch-at-age changes, the model has two mechanisms to explain it: a change in stock abundance (e.g. incoming recruitment), or a change in time-varying selectivity. This means that genuine signals of recruitment strength or mortality in the size composition data can be partially or wholly absorbed into selectivity adjustments, reducing the effective information content of those data for estimating population dynamics. This is potentially more challenging with the erosion of CPUE indexes for recruitment and catch rates in northern fishing grounds. The Lee et al. (2017) simulation study showed that a fleets-as-areas model was useful to assess PBF, but was calibrated under historical conditions where SSB was low and distribution contracted. As SSB has now more than doubled since the historical low period, and continues to grow, the spatial distribution of the stock is likely expanding. EPO fleets may be encountering fish of different age compositions than during the calibration period, not because selectivity has changed but because more fish are reaching trans-Pacific migration age. Time-varying selectivity may interpret this as a selectivity shift rather than an abundance increase for those age classes, potentially causing the model to underestimate the rate of recovery for older fish. If this was to occur, this is a directional bias rather than a general unquantified uncertainty. The panel was not able to determine if this is occurring in the stock assessment model, but is a concern.

### *4.1.2 LENGTH-AT-AGE*

The assessment uses a von Bertalanffy growth function (VBGF) with parameters externally fixed from otolith-derived age-length data (Fukuda et al. 2015b): $K = 0.188\ yr^{-1}$, $L_1 = 19.05$ cm (age-0), $L_2 = 118.57$ cm (age-3), implied $L_\infty = 249.9$ cm. The Panel accepts external estimates as appropriate given samples of length for age-0 fish and the well-documented confounding between growth and selectivity in integrated models with spatially mixed fish (Lee et al. 2019).

However, the 2024 assessment has materially changed the role the growth curve plays. With the Japanese longline (LL) CPUE terminating in 2019 and the Japanese troll fishery recruitment index truncated at 2010, size compositions from juvenile-catching fleets (principally Fleets 8, 9, 11, and 12) are now the primary source of information about year-class strength for 2011–2022. The growth curve has therefore shifted from a supporting biological parameter to the primary translation key between observable lengths and the model's internal age structure. The reliability of current stock status and projected spawning biomass is now highly sensitive to growth curve assumptions.

Fleets 8, 9, 10, 11, and 12 — catching predominantly age-0 and age-1 fish across different seasons, gears, and areas — independently show similar directional residual structures in recent years where the observed peak in length frequency near 60-80 cm is often higher and narrower than the expected length frequency near those lengths. This coherence across five independent fleets cannot be attributed to gear-specific or availability effects, which would produce uncorrelated patterns. It is instead the signature of a shared systematic error in how the model allocates fish across the age-0/age-1 size boundary. Fleet 6 (Sea of Japan purse seine, ages 3–6) provides complementary evidence with persistent large negative residuals (overfitting) at 100–150 cm. The model is consistently mischaracterizing observed peaks in the length frequency.

The primary driver is most likely inter-annual variability in juvenile growth rates that is considered collapsed into a single growth curve. A fixed, time-invariant VBGF applied to a population that experiences environmentally driven growth variability will produce this pattern. This mechanism also explains why the pattern has decadal structure that a fixed birthday misspecification cannot produce. The July 1 birthday assumption is therefore not the primary source of misfit, though it remains structurally important for a different reason: by collapsing two sub-cohorts with different growth histories into a single wide-CV distribution, it removes the model's capacity to separately diagnose growth rate variation, cohort mixing, and selectivity misspecification from residual patterns.

Furthermore, residual patterns in length frequency data for the Japanese and Taiwanese LL fleets (Fleets 1 and 3) show consistent underfitting when fish over 270 cm are observed. This is a result of the distribution of length for older fish being too narrow and the model not able to produce fish as large as are observed. This may occur because paired length-age samples did not include large fish, thus the variability was underestimated, or the distribution of length at age is a skewed distribution occasionally seeing very large fish for that age.

Finally, length-at-age samples from the EPO were not available for the external estimation of the growth curve. If these migrating fish of intermediate age are of a different size than the non-migrating fish, there could be a bias for the length-at-age and the variability at those intermediate ages.

The CVs of length-at-age (Tsukahara et al. 2024) were estimated conditional on the same model structure they parameterize. This is a recognized limitation common to integrated assessments: if the fixed VBGF is mis-specifying mean size-at-age in some years, the estimated CV, if large enough, will absorb that discrepancy alongside genuine individual variability, limiting its diagnostic value. The high age-0 CV (0.278) likely reflects both spawning-ground mixing and this residual size-prediction error, but the two contributions are not separately identifiable. Conversely, the small CV for length at older ages is likely underestimated due to the rare occurrence of large fish.

The retrospective instability in terminal recruitment (CV up to 53%) is a symptom of size compositions being translated through a growth curve that does not represent inter-annual growth variability while time-varying selectivity simultaneously adjusts to absorb residual misfit. The direction of any resulting bias in SSB is not obvious a priori and is currently unquantified. The growth curve may be ignoring large fish and their contribution to the spawning biomass.

A number of recommendations are provided to help understand this issue and improve the assessment model. These include ensuring that length-at-age samples from the EPO are including in growth estimation, running sensitivities to alternative growth curves, testing whether temporal structure of length composition residuals correlates with sea surface temperature (SST) anomalies, examining residuals by cohort rather than calendar year, and revisiting the single-birthday assumption.

### *4.1.3 Weight-at-length*

The weight-length relationship ($W = 1.7117 \times 10^{-5} L^{3.0382}$) is fixed and assumed time-invariant, based on Kai (2007). The Panel considers this relationship to be appropriate given available data, but recommends compiling recent data to update the estimates. The current analysis converted “gilled and gutted” fish to whole weight, which comprised approximately 50% of the samples. This may possibly bias the estimates and may be avoided by removing these samples in an updated analysis with additional data.

Additionally, the assumption of time-invariance is worth re-examining in future assessments, given that condition factor can vary with prey availability and that the population has experienced substantial recovery with potentially changed density-dependent effects on body condition.

### *4.1.4 Maturity*

Estimation of maturity-at-age or length was not described in the assessment document, but was briefly presented at the review meeting. The maturity ogive is fixed at: 0% for ages 0–2, 20% at age-3, 50% at age-4, and 100% at age-5+. These estimates are derived from histological studies conducted in the Sea of Japan (Tanaka 2006; Okochi et al. 2016) and Ryukyu Islands (Chen et al. 2006; Ashida et al. 2015), corrected to account for the immature fish in the EPO, and have been carried forward unchanged across multiple assessment cycles. The greatest uncertainty in the maturity ogive is the structural issue related to the EPO component of the stock. A maturity schedule that overestimates the proportion of age-3–5 fish that are mature will systematically overestimate SSB contributions from those age classes, inflating estimated SSB relative to true reproductive output and would affect depletion estimates. The working group should continue to understand the sensitivity of the assessment result to the assumed proportion of ages within the EPO. We recommend that work in this area continue so that a better understanding of the geographic distribution of maturity at age/size is developed.

### 4.1.5 *Recruitment*

#### 4.1.5.1 Steepness and recruitment variability

The Beverton-Holt SRR is used with steepness fixed at h = 0.999, rendering the relationship functionally flat and virtually all recruitment variation is absorbed by estimated recruitment deviations. Iwata (2012) and Iwata et al. (2012b) estimated h near 1.0 from biological data, the ISC working group and the published literature have long recognized the absence of a clear SSB-recruitment relationship for this stock (Sakuramoto 2015; Harford et al. 2017). This implies that the stock may easily recover from low spawning biomass, which has been seen, but also may result in risky MSY-based reference points due to a steep left side of the equilibrium yield curve.

With h = 0.999, sigma-R becomes an important parameter controlling how far recruitment can deviate from a mean, and therefore determining the range of stock trajectories the model can produce. The assessment sets sigma-R = 0.6 through a tuning procedure that matches the empirical spread of the estimated recruitment deviations. A sensitivity run with sigma-R set to 1.0 produced unintuitive results that may have been a result of model configuration such as the sum to zero constraint including uninformed deviates. Even though sigma-R = 0.6 is an empirical estimate and cannot be verified as the true variability of recruitment, the tuning procedure is the best practice resulting in internal model consistency.

#### 4.1.5.2 Projection stationarity and the resampling assumption.

Recruitment deviations are resampled from historical recruitment rather than using the stock-recruit curve because of the absence of a clear SSB-recruitment relationship. Lag-1 autocorrelation was not found in the estimated recruitment deviates, thus the entire period from 1983 to 2020 was resampled for the projections. The published literature (Sakuramoto 2015; Harford et al. 2017) has established that PBF recruitment shows regime-like structure with shifts linked to SST anomalies in the larval habitat and, in some analyses, to PDO phase. Regime shifts in recruitment were detected in 1957, 1972, 1980, 1994, and 2009 (Sakuramoto 2015). Resampling from the entire period ignores the potential for regimes which may have a significant effect on projections. Projections using only recent recruitment deviates, or assuming a distribution centered around a regime-specific mean, may illustrate more realistic near-term projections, or at least indicate the potential if the current assumed regime was to remain throughout the projection period. At a minimum, it may be useful to fully characterize the uncertainty in the projections. It is recognized that the most recent recruitment deviates are not well informed, and the regime may change in the projection period.

### 4.1.6 *Natural mortality*

Natural mortality (M) is fixed as age-specific and time-invariant: M = 1.6 $yr^{-1}$ at age-0 (PBF tagging studies: Takeuchi and Takahashi 2006; Iwata et al. 2012a, 2014), M = 0.386 $yr^{-1}$ at age-1 (adjusted from southern bluefin tuna tagging via life-history scaling: Polacheck et al. 1997), and M = 0.25 $yr^{-1}$ for ages

2+ (median of empirical methods). The Panel accepts the age-2+ value as a reasonable central estimate given available methods, but would like to see more evaluation of uncertainty and a range of plausible values.

# 5 TOR 2: Review PBF fisheries and available data

## 5.1 Fisheries, catch, and size compositions

Pacific bluefin tuna range across the northern Pacific Ocean and occupy different areas with multiple countries capturing this species at different life-stages, with different types of fishing gear, and under different management measures. This, therefore, results in a complex fleet structure needed to capture the different selectivities and data streams. The 2024 stock assessment specified 26 fisheries stratified by country, gear type, season, area, market type, and size of fish caught. The Panel agreed that this is a large number of fleets, and encouraged the Working Group to determine if any efficiencies could be made by combining some fleets that have similar qualities. The Panel greatly appreciated the thorough yet succinct presentations on the fleets given by working group members.

The fleet overview raised three concerns: the extent of size-data coverage, the importance of major data gaps, and whether changes in stock status or fishery management may have altered selectivity in ways that matter for the assessment. The PBFWG responded that targeting practices have changed across fleets and gears over time, and that time-blocked selectivity was implemented to accommodate data gaps and changes in fishing behavior, or fleets were treated separately. These were technically reasonable responses, but they also reinforce the need for clearer documentation showing where changes in selectivity are strongly supported by data or documented fishery changes, and where they may instead be compensating for changing stock dynamics or data limitations.

### *5.1.1 Japanese longline fisheries*

The Japanese longline fishery was classified into two separate fleets (1 and 2) based on the size of fish caught and showed changes in size compositions over time. Some years with low sampling for size compositions were removed from the stock assessment. Fleet 2 contained smaller fish and all size compositions were assigned to this fleet for 2017 and onwards. Catch for these fleets were treated separately from the abundance indices. Continued collection of samples from this fleet is very important.

### *5.1.2 Taiwanese longline fisheries*

The Taiwanese longline fishery was classified into two separate fleets (3 and 4) based on the size of fish caught and represents a well sampled and consistent time series. Catches have increased since 2012, yet 2024 was the first year in which the quota was constrained, resulting in an early closure of the fishery. This may result in a change to the associated CPUE data. Size data represent a large proportion of the fleet.

The Taiwanese longline fleet presentation reinforced a high quality of data, including independent catch documentation scheme (CDS) verification, voyage data recorder (VDR) based effort reconstruction, and very high size sampling and coverage. This infrastructure increases confidence in the transparency and internal consistency of the Taiwanese data inputs, while also highlighting the importance of maintaining careful evaluation of index comparability as the stock rebuilds and quota effects become more relevant. The Panel recommends continued sampling of the Taiwanese fisheries to maintain the data quality.

### *5.1.3 Japanese purse seine fisheries*

The Japanese purse seine fisheries were divided into six fleets (5 through 10) based on area, type of fishery (market or farming), and size of fish caught. Fleet 5 represents a large amount of historical catch from the Pacific Ocean and has had good size sampling since 2014. Size sampling before 2013 was fair and showed smaller fish than in 2014 and later. There is a gap in the size composition between 2006 and 2013 due to reduced reliability of the data, which makes it difficult to determine when the change from smaller to larger fish occurred in this fishery. Fleet 6 represents the market fisheries in the Sea of Japan and has very good sampling for size compositions. The size of the fish has varied over time and may have followed strong cohorts in some time-periods. Fleet 7 represents fisheries capturing fish for farming in the Sea of Japan, which began in 2011. There has been good size sampling since 2017. Fleet 8 is a juvenile fishery for the market in southern Japan with good size sampling showing a single consistent mode in the size composition near 50 cm. Catches were historically larger than recent years, but size compositions are only available from 2002 onwards. Fleet 9 is a juvenile fishery for the market in southern Japan with good size sampling showing consistent modes in the size composition near 45 cm and 75 cm. Catches were historically larger than recent years, but size compositions are only available from 2003 onwards. Fleet 10 is a fishery for farming in southern Japan with a small amount of catch, relative to other fleets, in the last couple of decades. Size sampling is good and the size composition shows a single mode near 50 cm.

The Panel observed that several Japanese fleets exhibit bimodal or multimodal size-composition patterns and questioned whether Stock Synthesis can fit such patterns statistically and how those patterns should be interpreted biologically or operationally. The PBFWG indicated that model fits to multimodal data have improved and suggested that the patterns may represent different cohorts recruiting into the fishery, and separation of young ages. The Panel agreed that improved fit is useful, but future reports should explain more clearly what these patterns are believed to represent and whether lack of fit to peak locations, peak heights, and valley depths could result in biases in selectivity, recruitment, or other key parameters.

The Panel further noted that the documentation to date has emphasized model specification more strongly than biological or operational interpretation. The issue is not whether multimodal data can be fitted, but whether the fitted structures are being interpreted clearly enough to support confidence in the model's internal allocation of information among composition, selectivity, and recruitment.

### *5.1.4 Korean offshore purse seine fishery*

The Korean offshore purse seine fishery (fleet 11) typically targets fish less than 30 kg, but has seen an increase in fish over 30 kg since 2019. Purse seiners are the primary source of the catch, but other fisheries contribute to this fleet. Size composition data are available since 2003.

### 5.1.5 *Japanese coastal fisheries*

The Japanese coastal fisheries were separated in the stock assessment by gear type, season, market type, and area. Troll, pole-and-line, and set-net defined seven fleets (12 through 18) and fleet 19 consisted of other gear types including handline and small-scale longline. Three troll fleets represented different seasons and one fishery for farming that began in the 1990s. Three set-net fleets represented different seasons and one fleet was for the northern area. The market troll fleet in season 1 (fleet 13) catches PBF under 50 cm while the market troll fleet in seasons 2-4 (fleet 12) catches fish slightly larger. The size composition data are good for both of these fleets. Size composition was not available for the troll fishery for farming (fleet 14) and is assumed to be composed of age 0 PBF. The pole and line fishery (fleet 15) has limited size data that were not used, thus selectivity was mirrored to fleet 12. The set-net fleets were split by seasons (fleets 16 and 17) and the fishery in the northern area (fleet 18). These size compositions for these fleets differed with fleet 16 catching PBF around 45 cm and 75 cm, fleet 17 catch PBF around 60 cm, 80 cm, and 120 cm, and fleet 18 catching small PBF less than 50 cm. The fleet consisting of other gear types (fleet 19) catches a range of sizes with the highest proportions typically less than 50 cm.

The panel appreciated the detailed descriptions of these fleets and recommends combining some of these fleets where reasonable.

### 5.1.6 *Eastern Pacific Ocean commercial fisheries*

The commercial fisheries in the Eastern Pacific Ocean were divided into two fleets representing U.S. dominant catches (fleet 20) up until 2001 and Mexico dominant catches (fleet 21) from 2001 onward. These fleets are primarily purse-seine fisheries. Catches in the EPO were predominantly from the U.S. fleet until 2001 when the Mexican fleet was dominant. Sampling of PBF was low from 1983 to 2001 when catches were low. There is a large amount of variability in the observed size compositions. Discards are accounted for in these fleets, and is reported to be a small percentage of the catch in the recent Mexican fisheries.

### 5.1.7 *Eastern Pacific Ocean recreational fisheries*

The EPO recreational fisheries were divided into two fisheries depending on when size sampling was conducted. Fleet 22 occurs from 2014 onward and has size observations while fleet 23 occurs from 1983 to 2013 and did not include size observations due to no information on how the program operated prior to 2012. The selectivity for fleet 23 was assumed similar to the estimated selectivity for fleet 22.

The Panel recognizes the important continuous sampling of PBF by the Sportfishing Associate of California, even after the Pacific Bluefin Tuna Port Sampling Program was suspended in 2020 and onwards. The Panel recommends investigating whether these fleets can be combined with a single estimated selectivity.

#### *5.1.8 Discards*

Discards were included as three fleets representing Japanese and Korean purse seine fisheries (fleet 24), Japanese fisheries for farming (fleet 25), and U.S. sport fisheries (fleet 26). The Panel recognizes that discards are challenging to determine and appreciates the modelling approaches they took. The Panel recommends continued monitoring of discards and additional research to determine the best methods to accurately estimate discards.

### 5.2 ABUNDANCE INDICES

The abundance indices in this assessment are the most significant data concern identified by the Panel due to changes in management resulting in the loss of some indices in recent years and other changes over time. The assessment relies on three abundance indices for adults: S1 (Japanese longline, 1993–2019), S2 (Japanese longline, 1983–1992), and S5 (Taiwanese longline south, 2002–2022). A recruitment index S3 (Japanese troll, 1983–2010) is also used. Both the Japanese longline adult index (S1) and the Japanese troll recruitment index (S3) have been truncated due to fishery management changes that likely altered catchability in recent years.

#### *5.2.1 Japanese longline index (S1)*

The introduction of individual quotas (IQ) in Japan in 2020 fundamentally altered fishing behavior, rendering the post-2019 Japanese longline CPUE inconsistent as an abundance index. Therefore, 2019 was the final year of the S1 index. The Panel accepts the decision to omit these years from the CPUE analysis. However, it creates a critical four-year window (2019–2022) in which the only operative adult index is the Taiwanese longline series (S5). This is the period during which the SSB is estimated to have dramatically accelerated its recovery, crossing the 20%$SSB_{f=0}$ second rebuilding target. The entire recent stock recovery narrative rests on a single index series for this terminal period, and that uncertainties in this index's catchability (including the potential for changing availability as the stock rebuilds and distribution shifts) are not fully propagated into the stock status estimates.

Small fish appeared in the catch in 2019 and 2020, causing a very large increase in the S1 CPUE. To include 2019 and 2020 in the CPUE series and maintain consistency across the years, the fish smaller than 60 kg were removed from the CPUE analysis (and 2020 was not included in the stock assessment), and fish smaller than 150 cm were removed from the size compositions. The catch for this fleet was a separate fishery, thus was removed properly from the population. However, the Panel was concerned that the observations were modified instead of modelling potential changes in the population or fisheries. The increased prevalence of small fish may have been a result of a change in selectivity/availability or an incoming recruitment, which may be worth investigating.

#### *5.2.2 Japanese troll recruitment index (S3)*

The exclusion of S3 data beyond 2010 was justified by evidence of catchability changes due to increased troll catch for farming operations and subsequent management interventions (Fukuda 2023). The Panel accepts the technical justification. However, the loss of a recruitment index for 12+ years (2011–2022) creates substantial uncertainty in terminal year recruitment estimates (CV up to 53% for recent recruitments), which the assessment acknowledges. The Panel notes that the period without a reliable recruitment index coincides with the period of greatest management interest, as the stock has been recovering and projections are sensitive to recruitment assumptions.

The loss of a reliable recruitment index following the truncation of the Japanese troll fleet index to 2010 represents one of the most consequential data gaps in the current assessment. The absence of a direct juvenile index for the period 2011–2022 has forced the model to infer year-class strength almost entirely from size compositions of juvenile-catching fleets translated through a fixed growth curve. The Panel considers the reestablishment of a robust, independent juvenile index of abundance to be among the most important data collection priority for the 2027 assessment cycle, of equal or greater urgency than the adult index problem. An ideal juvenile index would be independent of commercial fishing operations and their associated management-induced catchability changes, cover the primary juvenile aggregation areas in the East China Sea and western Pacific, provide annual or seasonal coverage sufficient to resolve inter-annual recruitment variability, and be precise enough to avoid misleading results. Options may include dedicated scientific trawl or purse seine surveys targeting age-0 and age-1 fish in the East China Sea juvenile grounds during the peak aggregation period, acoustic surveys of juvenile aggregations, or a formal expansion of the troll monitoring program into a statistically rigorous index with documented sampling design.

The Panel framed the current situation as a transitional period in which the former recruitment index (S3) has been truncated while the replacement indicators (Japanese larval survey, real-time troll monitoring) are promising but not yet operational within the base-case assessment. The PBFWG acknowledged that the recruitment index through 2010 can still support recruitment estimates for approximately 5–6 subsequent years through cohort propagation. However, by the terminal year (FY 2022), the model is estimating recruitments with no information from this data source. This is not a criticism but simply indicating a limitation in the terminal period.

### *5.2.3 New troll monitoring index (S4)*

The new standardized CPUE index from the Japanese troll monitoring program (Fujioka et al. 2023), covering 2011–2023, is presented as a qualitative indicator of recent recruitment trends but not included in the base case stock assessment. The Panel acknowledges the methodological caution exercised, particularly given potential catchability changes from 2017 onwards. However, the Panel notes that this index shows relatively high recruitment in 2021–2023, which qualitatively contradicts the low recruitment estimated by the base-case model for 2019–2021, and that the PBFWG used this as justification for starting projection resampling from 2021 rather than 2022. Again, the Panel recommends pursuing a consistent juvenile index for recent years to inform incoming recruitment.

### *5.2.4 Taiwanese longline index quality (S5)*

The S5 index, derived from a GLMM standardization of Taiwanese longline CPUE (Yuan et al. 2024), is the only adult index for the terminal period. The Panel notes several potential concerns: (a) the fishing ground is confined to the southern area, and there is evidence of possible changes in fish availability as the stock recovers and potentially expands its distribution; (b) the derivation of fishing effort in days requires multiple estimation steps (VMS/VDR data, logbook records), each of which may introduce additional uncertainty that needs to be carried forward; and (c) Taiwan has not yet approached its catch upper limit, but as SSB grows, changes in fishing strategy or effort distribution could affect catchability in ways not captured by the standardization model.

The Taiwanese longline fishery had not reached its catch upper limit during the assessment period (through fishing year 2022). However, the fishery reached its catch allocation for the first time in 2024, resulting in an early season closure. This indicates that quota-related catchability changes may become increasingly relevant for future assessments as the stock continues to rebuild. The S5 index is derived from standardised CPUE — catch per unit effort measured in fish per fishing day. When a fleet closes early because its quota is exhausted, the fishing days in the final portion of the season are zero, and fish that would otherwise have been encountered are not represented in the effort denominator. Even if a GLMM standardisation is applied, quota-driven truncation of the fishing season introduces a structural catchability change that is not attributable to fish abundance: the index may tend to underestimate abundance in years when the quota binds, and the degree of underestimation may increase as stock abundance grows and the quota constrains effort earlier in the season. This is the same mechanism that rendered the Japanese longline index (S1) unreliable after the introduction of individual quotas in Japan in 2020, necessitating its termination. The Panel considers this an urgent issue.

If the pattern of early quota exhaustion continues or intensifies with ongoing stock recovery, the S5 index will become progressively less reliable as a measure of true abundance, potentially at the same time as the stock is approaching or exceeding its rebuilding targets and management decisions are most consequential. This would leave no abundance index for recent years. The Panel makes the following additional recommendations to address this risk: (1) The PBFWG should urgently investigate whether CPUE standardisation methods can account for quota-driven truncation of the Taiwanese fishing season, for example by modelling the relationship between quota exhaustion timing and apparent CPUE, or by developing effort-standardised indices that account for days not fished due to quota closure. (2) The PBFWG should liaise with Taiwanese fisheries authorities to explore whether a supplementary monitoring programme analogous to the Japanese charter monitoring (CM) programme — in which vessels continue to fish under a dedicated scientific permit after their commercial quota is reached — could maintain the continuity of the S5 index independently of quota constraints. (3) Development of alternative adult abundance indices should be treated as the highest-priority data collection need for the 2027 assessment, given that the two historically most important adult indices (S1 and potentially S5) are both vulnerable to management-induced catchability changes as the stock rebuilds. These recommendations may apply to other fisheries that have experienced similar quota limitations (e.g. Japanese longline fisheries).

The Panel recommended that a formal catchability break-point analysis should be conducted before the 2027 assessment, comparing quota-constrained years against earlier years in terms of season timing, effort distribution, vessel participation, and catch rates. The Panel supports a future transition toward VAST or other spatio-temporal models, but recommends that such a transition be based on explicit supporting analysis rather than assumed by default.

# 6 TOR 3: REVIEW THE MODEL CONFIGURATIONS (E.G., FLEET STRUCTURE, SELECTIVITY) AND PROVIDE PRACTICAL RECOMMENDATIONS ON MODELING APPROACHES TO ADDRESS DEFICIENCIES IN INPUT DATA AND EXPLORE ALTERNATIVE MODEL SETTINGS TO IMPROVE FUTURE ASSESSMENTS

The 2024 assessment of PBF presents a single base case model starting in 1983 to estimate current stock status and produce projections under different harvest strategies. The choice of starting in 1983 was justified because of few size compositions before 1994, large catches in 1981 and 1982 when the stock was at one of its lowest depletion levels, convergence issues with a model starting in 1952, and the desire for a more stable and robust model especially to be used in the MSE process. Starting the model in 1983 loses some historical perspective of the past dynamics of PBF, which has been fished for a long time. It also loses the opportunity to understand changes in the stock dynamics due to environmental regime shifts. It is unknown if the estimation of R0 may be impacted by using only recent composition and catch data and if historic R0 values, and thus potential unfished levels, could be obtained in the future. The review panel concurs with the decision to start in 1983, recognizing the paucity of data before then makes it difficult to recreate historic dynamics.

The panel noted that the key interpretive issue is whether the shorter model should be described as a more credible biological representation or more modestly as a conditionally stable framework that reduces strain from historically difficult data. Improved convergence and lower historical data quality provide a practical justification, but do not by themselves demonstrate robustness across reasonable alternative assumptions. The bridging analyses were useful, but additional investigations should show that the main assessment conclusions remain stable when key structural choices are varied.

The following sections highlight some of the model configurations and assumptions discussed by the Panel.

## 6.1 RECRUITMENT TIMING AND THE JULY 1 BIRTHDAY ASSUMPTION: COHORT SLICING

The current base-case model assigns all recruits in a given year to a single July 1 settlement date — a single-birthday assumption. As documented in the technical summary prepared for this review panel (ISC Peer Review Panel, March 2026), this assumption conflates three structurally distinct sources of apparent variance in juvenile length-frequency data: true individual growth variation ($\sigma L$), spawning season spread (fish born in May versus August differ in real age at any sampling date by up to 0.25 years), and cohort-level inter-annual growth variation driven by oceanographic forcing. The current model absorbs all three through an inflated $\sigma L$, producing acceptable length-frequency residuals but misrepresenting the biological growth process. At the fast juvenile growth rates characteristic of age-0 and age-1 PBF, a spawning season spread of May through August generates a mean length difference of approximately 4 cm between early and late sub-cohorts at any given sampling date — comparable to or larger than the biologically correct σL. This misattribution directly distorts selectivity estimation for the juvenile fleets (Fleets 8–12) that are now the primary source of information on recent year-class strength.

Stock Synthesis provides a native mechanism to address this through its multi-settlement framework, which partitions each annual cohort into K sub-cohorts settling at different months within the spawning season. Each sub-cohort is assigned a fraction ($\omega_k$) of total annual recruitment and grows according to the VBGF from its own real age at settlement, rather than from a common July 1 date. The resulting observed length-frequency distribution for an annual cohort is a mixture of K normal distributions, each with the same $\sigma_L$ but with different means because the sub-cohorts are at different real ages. This mechanistically reproduces the observed pattern of high apparent CV at ages 0–1 converging to low CV at ages 3+ without requiring $\sigma_L$ to be inflated beyond a realistic value. The approach operates within the existing annual model structure and requires only changes to the settlement event configuration and the addition of settlement fraction parameters. For PBF, a three-event parameterization centered on months 5.5, 7.0, and 8.5 (reflecting the May–August spawning season) with fractions informed by larval survey data from the TOREDAMA program represents a tractable starting configuration.

The PBFWG should explore the multi-settlement (cohort slicing) framework as an option for the 2027 assessment, with the specific objective of testing whether restoring the variation of the youngest ages to the estimated value from Fukuda et al. (2015b) and Ishihara et al. (2023) improves the coherence of size composition residuals across juvenile fleets without degrading overall model fit. The evaluation should: (1) compare length-frequency residuals, log-likelihood, and selectivity parameter estimates between the inflated-$\sigma_L$ base case and the multi-settlement run, paying particular attention to the cross-fleet coherent residual pattern at 30–60 cm and 60–100 cm identified in Figure 5-7; (2) assess whether the cohort-strength inference from juvenile fleet compositions changes materially when the sub-cohort structure is resolved, given the sensitivity of terminal recruitment estimates to size-to-age translation; and (3) evaluate whether conditional age-at-length (CAAL) data from the Japanese otolith sampling program can be incorporated alongside marginal length-frequency data to provide an independent constraint on the within-cohort length distribution and help separate spawning-season spread from individual growth variance. This analysis would directly address one of the most consequential structural assumptions in the current model configuration and should be treated as a priority analytical task for the next benchmark cycle. This approach will not help mitigate annual variation not accounted for in growth not captured in the growth model. The July 1 start may complicate the implementation of cohort slicing and may need to be assessed.

## 6.2 SELECTIVITY

With 26 fishing fleets and 4 abundance indices, there were many selectivity parameters. This is often expected because of differences between fleets and complex selection patterns due to gear types and area-specific availability. A simulation study (Lee et al. 2017) supports using fleets-as-areas for the PBF assessment. Time-varying selectivity was used, but almost exclusively on age. It may be useful to consider time-varying length selectivity (as with fleets 21 and 22) to account for management, possible length-based movement, and other effects over time. Some fleets used a combination of length- and age-based selectivity to infer a final selectivity pattern. For example, fleet 16 was among a group that invoked zero selectivity at older ages, which could be problematic if catch is in weight and the age-based selectivity is incorrect. The Panel suggests investigating dome-shaped length-based selectivity as an alternative.

### 6.3 DATA WEIGHTING

The effective N approach is reasonable as a first-order correction for known design effects in the sampling — it is better than using nominal sample sizes directly, which would severely overweight strata with large measurement programs. The finding that ESS averages around 16% of nominal N is broadly consistent with design effects reported for other trawl and longline composition programs in the literature.

However, it is common practice in fisheries stock assessment to qualitatively determine a maximum input sample size and weight composition data downwards to balance the influence from each data source. A similar analysis was done on PBF composition data in 2012, but was not referenced in this stock assessment. The Panel recommends updating this analysis and considering the methods described by Hulson & Williams (2024).

### 6.4 BASE MODEL UNCERTAINTY

The 2024 assessment changed the method for estimating confidence intervals from bootstrapping to normal approximation of the Hessian matrix. This change has a direct effect on the reported uncertainty around SSB estimates (e.g., CV of $SSB_{2022}$ = 24%) and affects the probability statements used in management advice.

The Hessian-based approximation assumes asymptotic normality in log-space and a well-behaved, unimodal likelihood surface, but the jitter analysis found evidence of local minima near the base model minimum. Hessian-derived confidence intervals may understate true uncertainty because of normality assumptions and local curvature at the optimum does not fully represent the broader shape of the likelihood. The combination of high parameter dimensionality, local minima, and data-poor terminal years can lead to conditions where the approximation is less reliable. Furthermore, the uncertainty may be best represented by a skewed distribution instead of a normal distribution based on asymptotic confidence intervals derived from the Hessian (Stewart et al 2013).

The Panel is concerned that switching methods between the 2022 and 2024 assessments makes direct comparison of reported uncertainty challenging. Bootstrapped replicates used for the projections may provide insight into the differences between Hessian derived confidence intervals and bootstrapped confidence intervals. Using MCMC to create a posterior distribution would be ideal.

### 6.5 STRUCTURAL UNCERTAINTY

Structural uncertainty was presented through sensitivities (see below). One source of potential structural uncertainty is the decision to start the model in 1983. This decision improved the stability of the model, but it is not known if this is an important source of structural uncertainty. For example, selectivity alternatives were tested in assessments prior to 2024 but not under the current 1983-start framework. Some investigation into the effects of the start period would be useful to investigate.

### 6.6 SENSITIVITIES AND PROFILES

Multiple sensitivities were done to examine the effects of observation processes and system dynamics processes. Changes in the steepness parameter showed an effect on the estimate of depletion, but there is evidence to use a high steepness value. The sensitivity setting sigma-R to 1 showed unexpected results with a very low recruitment deviation in 2021 without information to support it and a higher B0. This may be a result of the model assumptions related to defining the period where recruitment deviations have to sum to 0. The Panel recommends an additional sensitivity (or profile) looking at different values of $L_\infty$ and K, especially its effect on $R_0$, and to use an informative range of values determined from biological plausibility.

## 7 TOR 4: REVIEW MODEL DIAGNOSTICS AND RESULTS

### 7.1 MODEL DIAGNOSTICS

The WG presented many useful diagnostics that helped the Panel understand the strengths and shortcomings of the stock assessment. Convergence was examined in detail and appeared good. Goodness-of-fit to different data sources were presented in useful ways. Sensitivities, profiles, hindcasting, and retrospective analyses provided additional insights into the model.

### 7.2 FITS TO COMPOSITION DATA

There are many methods to diagnose fits to size and age composition data. The WG relied heavily on aggregating observed and fitted compositions over years for each fleet, but also presented residuals and effective sample sizes. The Panel recommended also examining year-specific composition data to determine if trends are present, indicating misspecification. The Panel also recommended adding one-step-ahead (OSA) residuals to the current set of diagnostics to examine fits to composition data, and implementing the guidelines outlined in Stewart and Monnahan (2025).

### 7.3 AGE-STRUCTURED PRODUCTION MODEL

The WG presented results from an age-structured production model (ASPM) using catch data to fit the indices of abundance. This provides a useful insight into whether the catch and index data provide reasonable estimates of population scale given the assumed processes and parameters. A deterministic ASPM was used to diagnose the catch and indices, and an ASPM with recruitment deviations (ASPM-R) used the recruitment deviations estimated from the stock assessment to determine if the indices of abundance are consistent with other data sources, such as the age-0 troll index.

The Panel noted that this diagnostic method could be misinterpreted unless the specific ASPM formulation being used is clearly defined, as the term ASPM has been used in other applications (Punt et al. 1996; Restrepo and Legault 1998). Although the assessment document cited Maunder & Piner (2015), it would be helpful to clearly state what ASPM refers to when presenting this method.

The Panel had a discussion regarding the interpretation of using ASPM and ASPM-R as a diagnostic tool and how independent it is from the base case model. Particularly, estimated recruitment deviations from the base case model are input as fixed in ASPM-R (although the stock assessment document also says "ASPM-R includes the addition of temporal recruitment variation that exactly matches the age-0 troll index"), and thus the Panel questioned how much independent support is provided for the recruitment signal. The improvement in fit to adult longline indices when moving from ASPM to ASPM-R is partly expected since these recruitment estimates originate from the base case model that is fit to these CPUE data.

The Panel recommends a clear description of this diagnostic tool including whether it is an independent validation or a consistency check with base case assumptions, and that it is a test of consistency under current assumptions rather than robustness across alternative structural choices.

### 7.4 Source of Diagnostic Improvement

The Panel asked whether the improvement in diagnostics mainly comes from real structural improvement in the model, or from changes in data usage (shorter start year, truncation of the recruitment index). Better-looking diagnostics do not automatically imply broader robustness, and additional structural investigations may be useful.

### 7.5 Stock assessment results

The results of the base stock assessment were thorough and clearly presented. The Panel especially appreciated the presentations and the opportunity to ask questions. Overall, the results of the stock assessment are useful for the management of PBF and incorporate increasing uncertainty resulting from the reduction or loss in some data sources in recent years.

The fishery impacts plots provided an insight into the cumulative effect of a fishery. These analyses remove the catches from a fishery and recalculate the time-series, assuming that the catch now eliminated from that fleet would not be subsumed by other fleets. It should be noted that these are not impacts in a particular year relative to other fisheries, but a cumulative impact that is an indicator for each fishery.

## 8 TOR 5: Suggest research priorities

The Panel was very impressed with research projects currently underway and greatly appreciated the presentations during the review meeting. This research will help to improve the stock assessment in the future and all of the research projects are important for the future management of PBF.

### 8.1 Taiwanese Longline CPUE Standardization — VAST-at-Age Approach

During the review meeting, the PBFWG indicated that, because the original GLMM and VAST indices show little difference in trend, the Working Group is considering a transition to VAST as further supporting evidence continues to accumulate. The Panel considered that any formal transition in assessment inputs should be based on a clearer statement of what evidence is still needed, particularly with respect to terminal-period stability, catchability interpretation, and implementation practicality, rather than an informal gradual shift.

The VAST-at-age approach is conceptually well-motivated and the spatial structure is well justified. The cohort-resolved findings are biologically plausible and policy-relevant. However, the framework has a significant vulnerability in unaddressed age assignment uncertainty, and the causal interpretation of the observed spatial range expansion needs more rigorous treatment before these indices are used as primary tuning series in the stock assessment. These are addressable issues as the core methodology is sound enough to build on.

The Panel recommended explicitly testing for fleet footprint non-stationarity, for example by comparing spatial coverage of fishing effort across early vs. recent periods, or by running the model with effort as an offset vs. covariate. Additionally, it would be helpful to report convergence diagnostics per age bin, particularly gradient norms and Hessian positive-definiteness for older, data-sparse bins.

### 8.2 Japanese Troll Real-Time Monitoring (RTM) — Juvenile PBF Index of Abundance

The real-time monitoring (RTM) chartered survey is a well-motivated and operationally sensible response to the deficiencies of the sales slip catch series. The fixed-vessel design and VAST standardization are sound choices. However, the proposal is still in an early-stage in several critical aspects: spatial coverage is not documented, the target-species filtering is unfinished, the sampling intensity is modest, and the connection between the local coastal index and population-level recruitment is undemonstrated. The trend agreement with sales slips is reassuring but not sufficient on its own. The program is worth continuing and developing, but it is not yet ready to serve as a primary recruitment index in the stock assessment without further validation work.

During the review meeting, the PBFWG explained that the data source for the recruitment monitoring index remains limited in volume and unstable, and that the time series is still too short for immediate incorporation into the stock assessment. Therefore, the Working Group has not included these data in the model. The Panel agrees, but noted that, given the absence of any direct juvenile abundance index for over a decade, even a developing index with properly characterized uncertainty could provide valuable supplementary information. Any future incorporation would require improvements in data continuity, sample size, time-series length, and standardization methodology.

Some potential improvements include the following:

- Map the spatial coverage of the 14 chartered vessels explicitly, and evaluate whether the combined footprint adequately represents the juvenile distribution domain as understood from tagging and survey data. Consider whether additional vessels in underrepresented areas would be warranted.
- Prioritize completion and validation of the target-species filtering method before incorporating commercial RTM data. Ideally, validate the filter against observer or electronic monitoring data on a subset of trips.
- Examine within-season timing effects — whether fish arrival relative to the sampling window varies across years and whether this creates systematic bias in the annual index. A simple analysis of catch-per-day-at-sea against calendar date across years would reveal whether timing matters.
- Include vessel as a random effect in the VAST standardization if not already present, and report inter-vessel variability in catch rates to assess whether the "fixed vessel" assumption is actually holding catchability stable.
- Clarify the treatment of live releases in the response variable, and if released fish are a meaningful fraction of total catch, use total encounter rate rather than retained catch.
- Plan explicitly for index-to-recruitment validation as the time series lengthens — for example, tracking whether the 2021–2024 RTM index values for the age-0/age-1 cohorts are consistent with subsequent recruitment estimates in the stock assessment once those cohorts reach assessable size.

## 8.3 Larval Index

The Tuna-larvae Abundance in Western-pacific Area (TAWA) index is a conceptually important initiative and the Atlantic bluefin precedent gives it genuine scientific credibility as a long-term goal. The current implementation, however, is in an early exploratory stage with several foundational questions unresolved: the larval transport confounds, the gear transition, the unexplained bimodality, and the post-2020 divergence from the stock biomass trend. The index is not ready for inclusion in the stock assessment in its current form, but the survey program should continue, and the methodological work outlined above would substantially advance its readiness. A clear prioritized work plan with milestones for resolving each of these issues would strengthen the proposal considerably.

The panel considered the larval survey index to be a high-priority development area because fishery-independent information could reduce reliance on indirect model-based inference for juvenile abundance in future terminal periods. Fishery-independent juvenile information is likely to become increasingly important for assessment robustness, given that juvenile information support has been limited after 2010 and that CPUE-based recruitment indices are increasingly vulnerable to management-induced catchability changes. At the present stage, the near-term priority is not immediate model inclusion, but rather improving representativeness, strengthening the quantitative relationship with realized recruitment, and building a time series sufficiently informative for future testing. If successfully developed and incorporated into the stock assessment, this index could provide useful fishery-independent information to support future management of PBF.

Panel recommendations are provided below; in particular:

- Conduct an explicit larval transport/retention analysis using oceanographic model hindcasts to evaluate how much interannual variability in larval density at survey stations is driven by physical advection rather than spawner abundance. This is an important methodological investment for establishing index credibility.
- Validate the 4.55mm size cutoff against gear selectivity curves for both mesh sizes, and if possible, use the overlap period (years with both gear types deployed) to estimate a direct catchability correction rather than relying on the model covariate alone.
- Investigate and resolve the bimodal size distribution before finalizing the response variable. If the two modes correspond to distinct spawning pulses, consider treating them separately or restricting to the larger mode with more resolved age-at-capture.
- Standardize the survey design. Potentially fix a core set of stations that are sampled every year regardless of other survey objectives, so that spatial coverage is consistent over the index period.
- Explicitly test the Nansei area representativeness assumption by comparing years with multi-ground coverage (Sea of Japan + Nansei) against Nansei-only estimates to evaluate whether the single-area index captures the full stock signal.
- Investigate the post-2020 TAWA–SSB divergence before advancing the index to assessment use. This divergence must be explained, not simply noted — it is either a serious methodological problem or a genuinely interesting biological finding, and distinguishing between these has direct implications for how the index should be used and weighted in the stock assessment model.

## 8.4 CKMR

The PBF CKMR program is scientifically well-conceived and is used in the Southern Bluefin Tuna stock assessment. However, PBF faces three compounding challenges that are not yet resolved: insufficient pair counts for reliable standalone estimation, a spawning ground mixing problem that complicates the kinship model structure, and a declining detection rate as the stock recovers that will require either expanded sampling effort or a recalibration of what a CKMR program can realistically deliver. The most productive near-term contribution is likely not standalone biomass estimation but rather using the CKMR results as a consistency check against the stock assessment trajectory while the methodological groundwork for full integration continues to develop. The Panel recommends that CKMR be recognized as a strategic medium-term to long-term research priority for Pacific bluefin tuna.

The panel regarded CKMR as a strategic research investment with high long-term potential and recommended that its practical pathway for use in both assessment and MSE be articulated more clearly. In the near term, CKMR may be more realistic as an external validation tool or supporting research line than as a direct quantitative input to the benchmark assessment.

## 8.5 BIOLOGICAL SAMPLING

Biological samples are extremely informative to the stock assessment due to the high inter-annual variability in recruitment, many fisheries encountering various age classes, and a rapidly increasing stock. For this reason, the Panel strongly encourages each participating country to maintain or improve biological sampling programs to collect basic biological information for the stock assessment. This includes size sampling, the collection of otoliths, determination of maturation state, and observations of weight. Additionally, vertebrae sampling has declined in recent years, which has implications for determination of natal origin and CKMR spatial partitioning support. The Panel is impressed with the current biological research and note this research is dependent on adequate and representative sampling.

The biological sampling program is well-organized, geographically broad, and efficiently serves multiple concurrent research needs. The vertebral natal origin work and the international aging workshop represent genuinely important contributions that address foundational assessment uncertainties. However, the program's connection to the stock assessment model is less explicit. For example, updated parameters exist but are not used, sample size trends are unexplained, and the criteria for parameter revision are absent. For a program that provides the biological foundation for an age-structured assessment model, the pathway from sample collection to parameter update to model input needs to be more formally defined and transparently reported to the assessment working group. This includes conducting sensitivities when parameters are updated.

The Panel considered that not all ongoing biological studies contribute equally to near-term assessment improvement. The greatest value will come from identifying those studies that can be converted into usable information in time for the next assessment. Among growth, maturity, and natural mortality, the Panel considered growth uncertainty to be the most consequential for near-term improvement, given its role as the primary translation key between observed length compositions and internal age structure.

# 9 TOR 6: Decision processes are clearly and accurately presented

## 9.1 Stock assessment reporting

The stock assessment was clearly documented and thoroughly presented, with comprehensive descriptions of the data, diagnostic methods, and results. Supplementary materials were provided to address any remaining gaps. The Panel especially appreciated the in-person review meeting and the presentations that were given. This provided an opportunity for discussions to accurately understand the stock assessment and receive insights from researchers familiar with specific data and fisheries. The printed table provided at the review meeting of the qualities and modelling assumptions of the various fleets was very useful.

The Panel emphasized that the role of a review panel extends beyond evaluating technical soundness. Communication quality is a part of assessment quality. For a plenary audience primarily composed of managers, the emphasis should help distinguish between findings that are relatively robust, findings that are conditional on base-case assumptions, and findings that require explicit caution. Otherwise, the presentation may remain overly technical for decision-oriented audiences. The Panel appreciates the technical aspects presented for the review meeting.

An example is to identify main sources of uncertainty that should be considered when making management decisions. For example, describe how much recent recruitment estimates are influenced by observations and how much they are based on model assumptions. This can then be characterized as confidence in projections and inform priorities for data collection.

The Panel noted that the 2024 stock assessment uses data through the end of the fishing year, which for this assessment corresponds to mid-2023. This creates a lag, so management decisions may not be based on the most recent information. With fisheries targeting age-0 PBF, it would be beneficial to use the most up-to-date information when making management decisions. This would be especially useful if a larval or juvenile index were available to indicate the likely strength of the most recent year class. The Panel recognizes that this lag is unavoidable but recommends that the WG consider other methods of communicating recent information to managers when such information cannot be included in the most recent assessment.

# 10 RECOMMENDATIONS

The overall assessment of Pacific bluefin tuna and the presentations given during the review meeting were exceptional and highlighted a highly capable scientific team that collaborated well. The reviewers did not identify any major concerns with the input data, the methodology, or the outputs of the stock assessment. Therefore, the recommendations below reflect potential improvements to the input data and possible areas of improvement in the assessment model.

Two major topics were identified for consideration in the next assessment. First, improvement to the modelling of growth and the variability of length-at-age may improve the fits to length compositions and provide for better estimates of recruitment strength. Second, the loss of a recruitment index in recent years (i.e., CPUE index from the Japanese troll fleet) increases the uncertainty for the terminal year and the projections that inform management. The review panel was pleased to see the innovations and ongoing research with the Japanese troll fleet to develop a new index of juvenile PBF. Accurate and precise estimates of recent recruitment are extremely important to the management of PBF because there are 2 or more years between assessments, PBF grow quickly, and young PBF (less than 3-years old) are encountered by some fisheries. Multi-year gaps between stock assessments and management advice for highly dynamic species and fisheries require accurate information of young fish entering the fishery.

The review panel was impressed with the quality of the work presented at the review and realized that additional staffing would create an enormous boost to the amount of progress that could be made on PBF research and assessment. This would likely influence research and assessments of other fish stocks as well. Therefore,

- the Panel recommends that member countries consider allocating an appropriate level of staffing beyond what is currently provided to support future PBF research and assessments, and stock assessment in general.

The recommendations below are organized by the Terms of Reference provided for the review. Within each TOR, short-term and long-term recommendations are provided, sometimes organized by specific topics.

## 10.1 TOR 1: Review the biological assumptions of the PBF assessment and provide practical recommendations taking into account the best available life history information for this species (e.g., length-weight relationship, maturity, natural mortality, growth, and stock-recruitment relationship).

### *10.1.1 Growth*

Length-at-age was fixed at external estimates, which may introduce bias and other challenges. Improving the length-at-age relationship is one of the highest priority recommendations to improve the PBF stock assessment. Previous PBF assessment results have changed substantially when growth assumptions were revised.

*Short-Term*

- Investigate variability in length-at-age across all ages to improve fits to length compositions. Large variation of age at length in Taiwanese samples may imply differences in growth rates among cohorts (year class) and is recommended to be analyzed.
- Investigate temporal and spatial trends in growth and variability of length-at-age. Variability in growth should also be compared between WPO and EPO for ages 1-6 years.
- Investigate and present sex-specific growth to understand differences between the sexes and potential effects on length compositions.
- Identify the uncertainty in growth parameters for use in sensitivity analyses.

*Long-Term*

- The age-length relationship is different between Japanese and Taiwanese samples. There are two possibilities: 1) ageing bias between Japanese and Taiwanese age readers, or 2) the larger individuals tend to migrate to the southern area during spawning season even in the same age group. The PBFWG should continue to formalize inter-laboratory age-reading comparison, including exchange of age-character materials among Japan, Taiwan, and other relevant participants.
- Develop a program to coordinate and standardize ageing protocols between countries and understand uncertainty (and possibly bias) in the determination of age for PBF. This is motivated by the documented differences in growth trajectories between Japanese, Taiwanese, and U.S. samples, and by the historical sensitivity of PBF assessment results to growth assumptions.

- Test for correlation of growth with environmental variables and density-dependence, which may include using length composition residuals to determine if the temporal structure of the residuals is possibly related to changes in growth.
- Reorganize residuals for each fleet by year class rather than calendar year to test whether residuals for the same cohort are sign-coherent across fleets. This requires no new data and would confirm growth curve misspecification as the shared source.

### *10.1.2 Reproduction*

A considerable amount of research has been conducted on maturity and the working group clearly understands the complications of estimating a stock-wide maturity curve when there is migration of immature fish to the EPO.

*Long-term*

- Continue to monitor maturation rates and investigate the effects of migration and distribution on stock-wide maturity given assumed proportions of ages within the EPO. Maturation rates (0.2 for age-2-yrs, 0.5 for 5yrs, and 1.0 for >=5 yrs) are based on the unverified assumption of distribution and migration ratio between the Sea of Japan, Pacific side of Japan and EPO. The working group will need to continue to understand the sensitivity of these assumptions to the assessment results.

### *10.1.3 Movement and distribution*

The working group clearly presented the assumptions of movement across the stock.

*Long-term*

- Continue investigations of movement and how it affects the distribution of all stages of PBF. Migration rates of the PBF population to EPO and the distribution between the Sea of Japan and the Pacific side of Japan are important to calculations of maturity and growth, as well as understanding the influence of fishing mortality in each area.

### *10.1.4 Recruitment*

Timing of recruitment (i.e., birth date) was discussed during the review.

*Long-Term*

- Conduct investigations to determine ranges of hatch date for PBF in each spawning area. This may be done using length frequencies of age-0 fish or through more complicated analyses using ageing structures (e.g. otoliths). These data are important to characterize variability in length at age 0 and could be used to determine temporal changes in distribution of recruitment between spawning areas.

- Conduct analyses of potential relationships between recruitment strength and environmental covariates. Consider alternative methods of incorporating average recruitment and variability around that average in projections to account for possible environmental regimes.

### *10.1.5 Natural Mortality*

Natural mortality was fixed for ages 0, 1, and 2+ based on analyses by the working group.

*Short-Term*

- Present uncertainty in natural mortality of all ages more explicitly. For example, present a plausible range of M values, compare with other tuna species, develop a prior distribution, determine Lorenzen-type alternatives, or identify assumptions. This may be used to determine sensitivities or to inform estimation of M.

### *10.1.6 Weight-Length*

The relationship between weight-length was last updated in 2007.

Short-Term

- Compile weight-length data that spans the full assessment period (1983-2022), or longer, and update the weight-length analysis using recent data. Determine if observations converted from gilled and gutted fish are necessary for a weight-length curve to use in the stock assessment.

*Long-Term*

- Examine weight-length to determine if important differences occur over space and time.
- Consider if potential high-grading of fatter PBF are selected by fishers under a quota management system, resulting in biases in the length-weight relationship.

## 10.2 TOR 2: Review Pacific bluefin tuna fisheries and available data, especially with regards to catch, size composition, and the abundance indices used in the current assessment, and provide practical recommendations on data curation, analytical methods, how to integrate the data into the assessment, and new data collection methods.

Members of the working group presented a wide range of data and analysed those data using modern methods with proper justification for the methods used to develop the input data. The development of alternative recruitment indices (e.g. real-time troll monitoring, larval survey) is considered the highest research priority by the reviewers because the current recruitment index ending in 2010 informs recruitment estimates for that period and there is little information to inform recruitment in recent years, thus relying on model assumptions.

*Short-Term*

- Consider sources of data or the development of collection programs that would inform recruitment in recent years (also see research section below) to provide accurate information for the period between stock assessments.
- Examine the effects of management on Taiwanese LL CPUE and best approaches for monitoring, analysis, and modelling. The Taiwanese LL CPUE is one of the most important sources of data for the assessment of PBF following the discontinuation of the Japanese longline index. Recommendations include:
  - Develop a plan for evaluating whether catchability has been or will be compromised by quota constraints.
  - Ensure that future reports describe directly how quota-constrained years are screened or adjusted when evaluating index continuity.
  - Consider using VAST or similar tools to identify the most representative data source and model formulation. Any future replacement of the current South-region GLMM should be based on explicit comparative analysis and clearly documented supporting evidence, rather than assumed in advance.
- Consider if there are potential changes in catchability and selectivity for the Japanese LL CPUE index in the future due to recent management actions (e.g. discards).
- Inventory age data and determine if it may be included in the stock assessment to inform the estimation of growth parameters, recruitment, and other processes.
- Monitor discards and improve on the analysis and reporting of discard mortality.

*Long-Term*

- Standardize CPUE analyses using modern approaches such as spatiotemporal modelling for indices and length compositions (e.g. using software packages such as VAST and sdmTMB).
- Develop programs to expand the collection and determine ages from samples stock-wide.

## 10.3 TOR 3: REVIEW THE MODEL CONFIGURATIONS (E.G., FLEET STRUCTURE, SELECTIVITY) AND PROVIDE PRACTICAL RECOMMENDATIONS ON MODELING APPROACHES TO ADDRESS DEFICIENCIES IN INPUT DATA AND EXPLORE ALTERNATIVE MODEL SETTINGS TO IMPROVE FUTURE ASSESSMENTS.

Due to a dynamic wide-ranging stock spanning multiple countries, there are many intricacies leading to a complex model configuration. The working group provided clear presentations of model configurations and assumptions made. The Panel appreciates the availability of a printed sheet describing the many fleets. The working group was very responsive to suggestions and alternative model configurations to examine during the review meeting. This was greatly appreciated by the review panel.

*Short-Term*

- Examine fleet structure and ways to optimize the number of fleets and surveys. This may include using time-varying selectivity and/or catchability to model changes within a fleet.
- Adjust recruitment periods in SS to avoid recruitment deviations that depart from zero without data to inform them.

- Scrutinize estimated selectivity parameters to identify those that have no influence on the outcomes and can be fixed or eliminated to create a simpler parameterization. For example, dome-shaped parameters that are outside of the range of observed data.
- Examine alternative approaches to modelling the variability of length-at-age, including
    - Enhance variability at the youngest ages, possibly by using cohort slicing to explain additional variability due to time of birth;
    - Expand variability at the oldest ages especially since some length compositions showing occasional fish larger than 270 cm which is outside of the range of lengths possible for a 20 year old fish. A lognormal distribution for length-at-age may be useful.
- Examine the history of a fleet and potential management changes that may support the model to incorporate a change in selectivity. Include a fleet-by-fleet summary table or appendix linking each major selectivity block to the documented fishery change or management action that motivated it (also noted below).
- Use current best practices for data weighting which may include estimating an additive standard deviation for CPUE indices, the Francis (2011) approach for length and age compositions, and possibly other methods. Clearly describe the method used, the outcomes, and any assumptions.
- Evaluate whether data-quality differentials across fleets (ranging from “fair” to “very good,” with two fleets having no size data and sampling coverage from 3% to 83%) are adequately reflected in the effective sample sizes or data weights used in the model. A summary table linking fleet data quality, sampling coverage, and model weight would improve transparency.
- The PBFWG should explore the multi-settlement (cohort slicing) framework as an option for the 2027 assessment, with the specific objective of testing whether restoring the variation of the youngest ages to the estimated value from Fukuda et al. (2015b) and Ishihara et al. (2023) improves the coherence of size composition residuals across juvenile fleets without degrading overall model fit. This analysis would directly address one of the most consequential structural assumptions in the current model configuration and should be treated as a priority analytical task for the next benchmark cycle. The July 1 start may complicate the implementation of cohort slicing and may need to be assessed.

*Long-Term*

- Examine alternative stock assessment software platforms for future stock assessments due to the reduction of continuing support of stock synthesis.

### 10.4 TOR 4: REVIEW THE MODEL DIAGNOSTICS AND RESULTS, WITH PARTICULAR ATTENTION TO THE TREATMENT OF THE UNCERTAINTY AND PROVIDE PRACTICAL RECOMMENDATIONS.

Many diagnostics were provided by the working group for the base model and sensitivities. Uncertainty was large in the terminal year, mostly due to uninformed recent recruitment. The review panel appreciated the willingness of the working group to discuss multiple diagnostics and reasons for large uncertainty in recent years.

*Short-Term*

- Report one-step ahead (OSA) residuals and associated diagnostics along with Pearson residuals (or similar outputs showing under- and over-fitting). See Stewart and Monnahan (2025).
- Reduce the reliance on fits to composition data aggregated across years and examine fits to length compositions within and across individual years.
- Examine sensitivities for reasonable and justifiable ranges of important fixed parameters to determine if it represents an important source of parameter uncertainty that is not accounted for in the base assessment. Reasonable and justifiable should be determined from statistical distributions representing the likely range of each parameter.
- Conduct a formal sensitivity analysis testing alternative growth parameterisations bracketing plausible ranges of K and $L_2$ (or $L_\infty$), reporting effects on terminal recruitment, SSB, and projection probabilities.
- Examine sensitivities to alternative model structures, such as the start date, to determine important sources of structural uncertainty that are not accounted for in the base assessment results.
- ASPM and ASPM-R are useful diagnostic tools especially when used in combination with other diagnostic tools (e.g. examination of residuals). A wide range of methods should be used to diagnose the stock assessment for PBF, some of which are mentioned above.
- Future assessments should explore combined structural sensitivity scenarios (e.g., simpler selectivity combined with alternative recruitment or M assumptions) in addition to single-parameter perturbations. This will better characterize the full range of structural uncertainty including the effects of multiple changes.
- Provide a concise summary distinguishing between sensitivities that materially affect terminal status (decision-relevant) and those that mainly affect fit without changing the main conclusions (diagnostic-only). Uncertainty treatment should be interpreted in terms of decision relevance, not presented only as a list of runs or scenarios.
- Conduct a sensitivity to assumptions of dome-shaped selectivity to examine the consequences of a simpler parameterization (i.e. asymptotic) for important fleets to determine whether terminal status and key conclusions are materially affected by more complex parameterizations.
- Present a comparison of Hessian-derived and bootstrap-derived confidence intervals for key quantities (SSB, depletion ratio) to validate that the Hessian approximation is adequate. If bootstrapping remains computationally feasible, maintain it as a check on Hessian-based intervals in future benchmark assessments.

*Long-Term*

- Consider an ensemble model that includes only important sources of parameter and structural uncertainty that are not included or represented in the base assessment.

## 10.5 TOR 5: SUGGEST RESEARCH PRIORITIES TO IMPROVE CURRENT KNOWLEDGE OF ESSENTIAL POPULATION AND FISHERY DYNAMICS, WITH THE IDENTIFICATION OF PRIORITIES TO IMPROVE NEAR-TERM ASSESSMENTS.

The review panel was very impressed with research projects currently underway and greatly appreciated the presentations during the review meeting. This research will help to improve the stock assessment in the future. The recommendations below relate to each research item, but some long-term general research recommendations include

- Continue to develop standardized fishery-independent surveys to inform abundance, size/age compositions, and biological relationships. This includes, but is not limited to, the Japanese larval survey and real-time troll monitoring.
- Examine alternative spatiotemporal modelling platforms for the standardization of indices to determine if there are useful features specific to the data of interest.

### 10.5.1 VAST Taiwanese LL index

*Short-Term*

- The Panel recommends that the PBFWG conduct a formal comparative evaluation of the VAST-based and GLMM-based Taiwanese longline indices before any transition from the current South-region GLMM to VAST. This evaluation needs to include a comparison of trend consistency, terminal signal, retrospective stability, spatial coverage, and sensitivity to catchability assumptions. The evidence for any transition must be fully documented.
- Propagate age assignment uncertainty into the VAST-at-age indices for the Taiwanese LL fishery, either through bootstrapped age-length keys or by treating age-bin membership probabilistically within the model.
- Report diagnostics for each age-bin in the VAST-at-age indices especially for older, data-sparse bins.

*Long-Term*

- The PBFWG is encouraged to evaluate the feasibility and potential benefits of developing age-specific CPUE indices from the VAST framework.
- Along with evaluation of the VAST index as a potential index for use in the assessment, it would be useful to examine methods to incorporate the challenges imposed by quota restrictions in recent years.

### 10.5.2 Troll Monitoring

*Short-Term*

- The Panel recommends continuing to invest in the real-time troll monitoring program with the objective of achieving a viable time series and a documented standardization protocol before the next benchmark assessment.
- At this time, the Review Panel agrees that this index is not yet ready for formal inclusion and considers it appropriate to continue treating it as a developmental research product. However, given that the assessment has operated without any direct juvenile abundance index for over a decade, even a developing index with properly characterized uncertainty could provide valuable supplementary information. Future work needs to identify the main milestones required for

eventual assessment use, including improvements in data stability, standardization methodology, time-series length, and model integration.

- Examine differences between traditional and chartered trips in the RTM to determine if fisher behavior is different between trip types, and whether they both are representative of juvenile distribution.
- Map the distribution of chartered and traditional trips in the RTM to determine if they cover the past and possible future juvenile distribution domain, and determine if additional vessels in other areas would be warranted.
- Examine whether fish arrival relative to the sampling window varies across years and whether this creates a potential bias in the annual index.
- Include vessel as a random effect in the VAST standardization, if not already present.
- Clarify the treatment of live releases in the response variable.
- Consider how the chartered RTM may be expanded or improved, especially if the traditional RTM continues to decrease.
- Provide regular reports on the real-time troll monitoring development at each subsequent working group meeting, because the weakening of post-2010 juvenile support remains an important structural issue for future assessments.

*Long-Term*

- Plan for index validation as the time series lengthens. For example, tracking whether the 2021–2024 RTM index values for the age-0/age-1 cohorts are consistent with subsequent recruitment estimates in the stock assessment once those cohorts are informed by other data sources.
- Identify possible concerns with monitoring juvenile abundance with the RTM if the juvenile area inhabited expands with an increasing population, or a third spawning area becomes important.

### *10.5.3 Larval Survey*

*Short-Term*

- The PBFWG is encouraged to prioritize the continued development of the Japanese larval survey index as a fishery-independent data source for the stock assessment.
- Determine if the larval survey is best suited as an index of spawning abundance or an index of age-0 recruits. This may be done by comparing the index with different stock assessment outputs.
- Before considering this index for formal assessment use, further work is necessary in three specific areas.
    1. The quantitative relationship between larval and subsequent realized recruitment requires strengthening, if used as a recruitment index.
    2. Researchers must ensure that spatial and temporal coverage is representative of the main spawning and nursery areas.
    3. The PBFWG is advised to evaluate whether the larval survey and the real-time troll monitoring program could be used in combination to provide complementary recruitment information that neither source alone can deliver.

- Conduct a larval transport/retention analysis using oceanographic model hindcasts to evaluate how much interannual variability in larval density at survey stations is driven by physical advection rather than spawner abundance. This is important to establish index credibility.
- Investigate and resolve the bimodal size distribution and examine the consequences of the 4.55 mm size cutoff. Consider alternative survey designs to reduce any effects, or alternative modelling methods to account for two possible birth dates.
- Compare the Nansei area representativeness assumption by comparing estimates from years with data from both spawning areas.
- Investigate and explain the post-2020 TAWA and assessment SSB divergence before advancing the index to assessment use.

*Long-Term*

- Investigate the behavior of different sizes of larvae and their response to the net. This will help to inform catchability.
- Clearly define the design of the survey, especially the spatial coverage and if it is representing the entire area possibly occupied by PBF larvae. Define a core area that is consistently sampled so that spatial coverage is consistent over time and may expand if necessary.
- Assess whether this index can become a formal candidate for testing within the assessment framework, and is expected to report on its development status at each subsequent working group meeting.

### *10.5.4 CKMR*

*Short-Term*

- The PBFWG is advised to develop a clear statement of the intended role and timeline for CKMR within the PBF assessment and MSE framework, including practical short-, medium-, and long-term goals.
- Report CKMR needs to distinguish clearly between long-term scientific potential and realistic near-term operational application.
- Define the potential roles of CKMR which include serving as an external check on absolute abundance estimates, providing independent information on adult population size, or supplying stock-structure information for MSE input.
- Report results of a formal power analysis specifying the required number of POPs and HSPs to achieve a useful precision under a range of plausible stock sizes and compare this against current and projected sampling rates.
- Present the assumptions in the CKMR analysis and consider the consequences of violating these assumptions. For example, equal sex ratios and no skip spawning.

*Long-Term*

- Coordinate sample collection with existing port-sampling and observer programs to maximize efficiency.

- Prioritize international collaboration to expand sample sizes and introduce additional expertise (such as from the CCSBT CKMR program).
- Commission expert code review, or utilize existing reviewed tools, before utilizing the CKMR data for assessment conclusions.
- Evaluate the costs and trade-offs of CKMR with other potential fishery-independent survey opportunities.

### *10.5.5 Biological sampling*

*Short-Term*

- The PBFWG is strongly encouraged to explicitly prioritize growth-related biological sampling and analysis as the most consequential biological parameter issue for near-term assessment improvement. This prioritization is based on three main factors.
    1. The growth curve now serves as the primary translation key between observed length compositions and internal age structure.
    2. There is documented sensitivity of PBF assessment conclusions to growth assumptions.
    3. Unresolved inter-laboratory differences in age reading exist between Japanese, Taiwanese, and U.S. samples.
- Among ongoing biological studies, distinguish between biological sampling priorities capable of delivering assessment-ready inputs before the 2027 cycle and those remaining longer-term research directions.
- Biological sampling priorities need to link directly to their expected contribution to reducing uncertainty in the assessment. Studies with the highest near-term impact potentially include the collaborative aging workshop, temporal and spatial growth variability analysis, and evaluation of environmental correlates of growth.
- Evaluate the updated maturity parameters from Ashida et al. (2023) for inclusion in the assessment, or provide an explicit justification for why the current maturity ogive is preferred.
- Document the size and age distribution of gonad samples and evaluate whether the current sampling design adequately covers the full maturity transition range, particularly for age 3–6 fish. If not, targeted sampling of smaller fish at ports near spawning grounds during the spawning season should be considered.
- Report the aging workshop results — even preliminary inter-reader precision statistics — as soon as available. The ISC panel should be aware of the magnitude of any between-lab aging differences, as these have direct implications for interpreting growth model uncertainty in the current assessment.
- Characterize the seasonal distribution of port sampling for each biological sample type, and evaluate whether sampling coverage is adequate for capturing biologically important periods.
- Evaluate whether any ongoing studies related to size-composition interpretation can help distinguish more clearly among cohort signals, fishery heterogeneity, and selectivity effects in the multimodal patterns observed in several fleets.

*Long-Term*

- Prioritize collaboration between countries for sharing biological information and understanding spatial differences.

## 10.6 TOR 6: COMMENT ON WHETHER THE STOCK ASSESSMENT METHODS, RESULTS, AND ASSESSMENT DECISION PROCESS ARE CLEARLY AND ACCURATELY PRESENTED IN THE MEETING REPORTS AND THE STOCK ASSESSMENT REPORT.

The working group provided a complete and thorough stock assessment document for the review, as well as background information useful to understand the PBF assessment in more detail. Discussions during the review meeting provided detailed information requested by reviewers. Recommendations below are suggestions for topics where additional explanation in the assessment document may be useful to expand upon and highlight their importance.

*Short-Term*

- Clearly state that the Japanese troll CPUE is no longer used after 2010 and that data informing recruitment becomes progressively weaker after that point. This should be communicated as increasing uncertainty in both stock status determination and near-term projections. Describe that the Japanese larval survey and real-time troll monitoring are promising replacement indices but are not yet fully operational in the assessment.
- Clearly state in the assessment report that adult abundance in the terminal period is supported by a single abundance index (Taiwanese longline, S5) following the discontinuation of the Japanese longline index. Recognize that this represents a narrower evidence base than was available in previous assessments, and reflects an increase in uncertainty presenting the stock status.
- Address whether the model's improvement mainly comes from better biological and fishery interpretation, or mainly from more flexible technical settings.
- For selectivity, provide clearer justification for major time-varying changes, especially where size-composition structures are complex or multimodal. Include a fleet-by-fleet summary table or appendix linking each major selectivity block to the documented fishery change or management action that motivated it.
- Explain the basis for data weighting more transparently, including whether weights reflect statistical diagnostics, historical practice, or expert judgment, and how the weighting interacts with the increased importance of the Taiwanese index in the terminal period. Data weighting should reflect the reliability and information content of different data sources, not only goodness of fit.
- Note clearly in the documentation where important structural uncertainties have been addressed outside the current year's core uncertainty design (e.g., selectivity tested in prior assessment cycles but not re-tested under the 1983-start framework) so that reviewers can assess whether the sensitivity is still adequate.

- The assessment report, particularly sections intended for plenary audiences, should explicitly distinguish between: (a) robust findings (e.g., directional rebuilding trend since approximately 2010), (b) conditional findings (e.g., precise terminal SSB, probability of reaching 20%SSBF=0), and (c) findings requiring explicit caveats (e.g., near-term projections dependent on terminal recruitment assumptions with CVs up to 53%).
- Results not consistent across a wide range of alternative structural assumptions should use language closer to “usable with conditions” rather than “fully robust”.
- Where reporting is intended for decision-oriented audiences, reduce purely technical presentation that may obscure the practical interpretation needed by managers. Help managers distinguish between findings they can use with confidence and findings that still require caution.
- In future reports, link uncertainty in fixed parameters (e.g. growth, maturity, and natural mortality) assumptions more directly to spawning biomass interpretation and rebuilding conclusions, rather than presenting them only as background parameter descriptions.

*Long-Term*

- The Japanese larval survey index and real-time troll monitoring should be described as promising replacement pathways for recruitment information going forward, while recognizing that they are not yet fully operational assessment inputs. Future assessment documents should describe the validation status, intended role, and integration pathway of these emerging indicators.

## 11 References


Ashida H, Ishihara Y, Watanabe S, Ohshimo S, Tanaka Y. 2023. Revisiting the spawning fraction of Pacific bluefin tuna (*Thunnus orientalis*) caught in the Sea of Japan based on diurnal changes in spawning markers. Fisheries Research. 266. https://doi.org/10.1016/j.fishres.2023.106788

Ashida H, Suzuki N, Tanabe T. Suzuki N. Aonuma Y. 2015. Reproductive condition, batch fecundity, and spawning fraction of large Pacific bluefin tuna *Thunnus orientalis* landed at Ishigaki Island, Okinawa, Japan. Environ Biol Fish. 98, pp 1173–1183. https://doi.org/10.1007/s10641-014-0350-8

Chen KS. Crone P. Hsu CC. 2006. Reproductive biology of female Pacific bluefin tuna *Thunnus orientalis* from south-western North Pacific Ocean. Fish Sci. 72. pp 985–994. https://doi.org/10.1111/j.1444-2906.2006.01247.x

Franics RICC. 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68. pp 1124–1138. https://doi.org/10.1139/F2011-025

Fukuda H. 2023. Comprehensive model diagnostics to investigate the cause of a systematic retrospective pattern of SSB in Pacific Bluefin tuna Stock Synthesis model used for the 2022 assessment. ISC/23/PBFWG 02/05

Fukuda H, Yamasaki I, Takeuchi Y, Kitakado T, Shimose T, Ishihara T, Ota T, Watai M, Lu HB. and Shiao JC. 2015. Estimates of growth function from length-at-age data based on otolith annual rings and daily rings for Pacific Bluefin tuna. ISC/15/PBFWG-2/11.

Harford WJ, Karnauskas M, Walter JF, Liu H 2017. Non-parametric modeling reveals environmental effects on bluefin tuna recruitment in Atlantic, Pacific, and Southern Oceans. Fisheries Oceanography, 26(4), 396-412. https://doi.org/10.1111/fog.12205

Hulson PF, Williams BC. 2024. Inclusion of ageing error and growth variability using a bootstrap estimation of age composition and conditional age-at-length input sample size for fisheries stock assessment models. Fisheries Research. Vol 270. https://doi.org/10.1016/j.fishres.2023.106894

Ishihara T, Tanaka H, Fukuda H, Tawa A, Ashida H, Tanaka Y. 2023. Estimation of confidence intervals for the von Bertalanffy growth function parameters using the bootstrap method with a data set of direct age estimates from otoliths. Working paper submitted to the ISC PBF Working Group Meeting, 21-24 March 2023, Shinagawa, Japan. ISC/23/PBFWG-1/10: 8p.

Iwata, S. 2012. Estimate the frequency distribution of steepness for PBF. Working paper submitted to the ISC PBF Working Group Meeting, 10-17 November 2012, Honolulu, Hawaii, USA. ISC/12/PBFWG-3/01: 13p. Available at: http://isc.fra.go.jp/pdf/PBF/ISC12_PBF_3/ISC12_PBFWG-3_01_Iwata.pdf

Iwata S, Fujioka K, Fukuda H, and Takeuchi Y. 2012. Reconsideration of natural mortality of age -0 PBF and its variability relative to fish size. Working paper submitted to the ISC PBF Working Group Meeting, 31 January-7 February 2012, La Jolla, California, USA. ISC/12/PBFWG-1/13.

Iwata, S., Fukuda, H., Abe, O., and Takeuchi, Y. 2012b. Estimation of steepness of PBFT -By using biological feature. Working paper submitted to the ISC PBF Working Group Meeting, 31 January-7 February 2012, La Jolla, California, USA. ISC/12/PBFWG-1/15: 9p.

Kai M. 2007. Weight-length relationship of North Western PBF. Working paper submitted to the ISC PBF Working Group Meeting, 11-18 December 2007, Shimizu, Japan. ISC/07/PBFWG-3/07.

Lee HH, Piner KR, Maunder MN, Taylor IG, Methot RD. 2017. Evaluation of alternative modelling approaches to account for spatial effects due to age-based movement. Can. J. Fish. Aquat. Sci. 74(11). pp 1832-1844. https://doi.org/10.1139/cjfas-2016-0294

Lee HH, Piner KR, Taylor IG, Kitakado T. 2019. On the use of conditional age at length data as a likelihood component in integrated population dynamics models. Fish. Res. 216:204-211.

Okochi Y, Abe O, Tanaka S, Ishihara Y, Akio Shimizu. 2016. Reproductive biology of female Pacific bluefin tuna, *Thunnus orientalis*, in the Sea of Japan. Fisheries Research. 174. pp 30-39. https://doi.org/10.1016/j.fishres.2015.08.020

Polacheck T, Hearn WS, Miller C, Whitelaw W, and Stanley C. 1997. Updated estimates of mortality rates for juvenile SBT from multi-year tagging of cohorts. CCSBTSC/9707/26.

Punt AE. 1996. A multi-component variant of the ASPM approach for use in the assessment of north Atlantic albacore (*Thunnus alalunga*). ICCAT Col. Vol. Sci. Pap.43: 347-360.

Restrepo VR, Legault CM. A stochastic implementation of an age-structured production model. University of Alaska Sea Grant College Program, Rep. 1998(98-01):435-50.

Sakuramoto K, 2015. A stock-recruitment relationship applicable to Pacific bluefin tuna and the Pacific stock of Japanese sardine. American Journal of Climate Change, 4:446-460. http://dx.doi.org/10.4236/ajcc.2015.45036

Stewart IJ, Hicks AC, Taylor IG, Thorson JT, Wetzel C, Kupschus S. 2013. A comparison of stock assessment uncertainty estimates using maximum likelihood and Bayesian methods implemented with the same model framework. 142. pp 37-46. https://doi.org/10.1016/j.fishres.2012.07.003

Stewart IJ, Monnahan CC. 2025. Diagnosing common sources of lack of fit to composition data in fisheries stock assessment models using one-step-ahead (OSA) residuals. Can. J. Fish. Aquat. Sci. 82. pp 1-13. https://doi.org/10.1139/cjfas-2025-0158

Takeuchi Y, and Takahashi M. 2006. Estimation of natural morality of age-0 PBF from conventional tagging data. Working paper submitted to the ISC PBF Working Group Meeting, 31 January-7 February 2006, Shimizu, Japan. ISC/06/PBFWG/07.

Tanaka S. 2006. Maturation of Bluefin Tuna in the Sea of Japan, ISC/06/PBF-WG/09.

Tsukahara Y, Ishihara T, Chang SK, Shiao JC, Lee HH and Fukuda H. 2024. Re-evaluation of coefficient of variance (CV) in growth curve using the latest otolith data. ISC/24/PBFWG-1/01