REPORT

Review of Four Japanese Stock Assessments in early 2026

By Murdoch McAllister, PhD
The University of British Columbia
Vancouver, B.C., Canada

March 13, 2026

Points of review

A) Determine whether the data used for stock assessment are adequate to understand the stock dynamics of the target species and represent the best scientific information available.

B) Discuss whether the biological parameters used for stock assessment are appropriate.

C) Discuss whether the basic biological information such as distribution, migration pattern, and population are appropriate.

D) Evaluate whether the stock assessment methodology is based on the most appropriate available study and performed analytically.

E) Evaluate whether the data are treated statistically correctly.

F) Evaluate whether the stock assessment result obtained from the input data and methodology used is appropriate.

G) Evaluate the validity of methodology and result used for the future projection.

Stock Assessment #1: Round Herring

A) Determine whether the data used for stock assessment are adequate to understand the stock dynamics of the target species and represent the best scientific information available.

Catch records and abundance indices

The stock assessment uses a long time series of catch records from 1976 to 2023 from landings mostly from Japanese operated vessels in the East China Sea, Western Japan Sea and Northern Japan Sea Areas. At nearly 50 years, this is a sufficiently long time series of catches to provide information about stock dynamics, especially given that round herring is not a very long-lived species, i.e., with a maximum age of two years in catch records. The total recorded annual catch in both biomass and numbers shows quite strong fluctuations, e.g., ranging between about 25,000-45,000 tons up to about the mid-1990s, a sharp decrease to about 10,000 tons in the early 2000s and then increasing and fluctuating between about 20,000-60,000 tons in the most recent decade or so. These wide-ranging fluctuations in catch which roughly parallel fluctuations in the relative abundance indices provided by the egg and larval survey and purse seine abundance indices which run from the mid- 1990s and early 2000s, respectively, to 2023 would appear to be informative about round herring stock dynamics over the past several decades.

There are some potential issues with the completeness of the total annual catch records. It is acknowledged for example that "South Korea's catch in weight was recorded at 14,000 tons from 1976 through 1986 but has not been reported since 1990. … The catch of round herring in China is unknown". No indication is offered for the magnitude of the missing annual catches from South Korea after 1990 or the potential magnitude of round herring catches by China. If the magnitudes of these missing catches are not insignificant, then it is possible that the stock assessment will tend to under-estimate the harvest rates, and also underestimate historical stock abundances. This is because in the stock assessment method used, the larger magnitude of the historical catches, the larger will be the reconstructed numbers at age in the fish stock. Under-estimation of historical stock abundances will tend to lead to under-estimation of abundance based MSY reference points such as spawning stock biomass at MSY and MSY.

The VPA was tuned to two stock abundance indices. One of these was an annual spawning stock biomass index based on an egg and larvae survey starting in the mid-1990s. The other was a stock biomass index based on catch performance reports for large and medium sized purse seine fisheries. The plots of the model predicted index versus the observed index for the egg and purse seine indices both showed moderately good fits to the abundance index data. However, there was one large positive and potentially influential outlier in the fit to the egg index for the very largest observed egg index in 2016. This was preceded by the second largest egg index in 2015, and followed by a further large egg index in 2017 which looks like the third largest index in the time series, and then a strong decrease in the egg index in 2018. One hypothesis is that the egg index suggests a large sudden increase in SSB 2015-2017 which is possible in a short-lived fish like round herring with a maximum observed age of about 2 years.

The large values for the 2016 and 2017 egg indices however were not picked up in the 2015 and 2017 purse seine indices which conformed closely to model predictions of no increase in stock biomass in these years, suggesting that the large egg indices 2015-2017 could potentially represent positive observation errors in the pelagic egg index, and not a large sudden increase in SSB in these years. That the VPA did not fit the large egg index values in 2016 and 2017 but conformed more closely to the purse seine index could reflect the strong influence the commercial fishery data in the VPA. It may still remain plausible that the stock could actually have seen a large short-term burst of spawner biomass in 2015 and 2017 but that the commercial fishery data in terms of catch numbers at age and also the purse seine index used in the VPA and thus also the stock abundance estimates from the VPA failed to pick up this potential large uptick in the stock 2015-2017. There is thus some question as to whether the data used for stock assessment were adequate to understand the stock dynamics of the target species and represent the best scientific information available. The large differences in representation of stock abundance in 2015-2017 between the two indices were glossed over and the alternative hypothesis that there was a short-term large increase in SSB 2015-2017 was ignored, despite the trend coming from the fishery independent abundance index which could potentially be a more reliable indicator of trends in abundance than the fishery dependent index and fishery catch-at-age data. There did not appear to be any discussion to offer potential explanations for the exceptionally large egg index in 2016-2017 and poor fit of the VPA to the egg index for these years. I would thus recommend that the stock assessment document add a few sentences to address the potential issues associated with the large discrepancy in stock trends 2015-2017 between the fishery dependent and fishery independent indices. It is recommended that the VPA be fitted separately to the egg index and the purse seine index to test the sensitivity of VPA model results to the different trends apparent in the two different indices. When the VPA is fitted only to the purse seine index, the value for the b-parameter for the purse seine index could be fixed at the value obtained from the base case VPA when it was fitted to both the egg and purse seine indices.

Commercial length composition data were transformed into age composition data using an age-length key. "From the body length composition stretched by ocean area and month, the catch in number at age was calculated using a length-age key created by month and ocean area based on Ohshimo et al. (2011)." (p. 30). It is possible that growth and length at age could be time varying and using a length-age key based on a study published in 2011 could lead to actual length-age keys that are different from those based on the Ohshimo et al. (2011) study. Application of a length-age key based on a fixed point in time to commercial length composition records when the length-age relationships are time varying could lead to generation of misrepresentations of age-composition records for use in the VPA stock assessment. In addition, as is common with generation of age composition records it is likely that there could be smearing in the age composition records generated from the commercial length composition records. This could result in stronger age classes being under-represented in the generated age-composition data and neighbouring age classes being over-represented. It would be appropriate to develop a simulation-estimation methodology to quantify the expected estimation performance of the current approach to deriving age composition data from length composition data using a fixed length-age key.

B) Discuss whether the biological parameters used for stock assessment are appropriate.

A value for the instantaneous rate of natural mortality was assumed at 0.7 per year based on (Ohshimo 2003, 2009). The Ohshimo method to determine M is not well known outside of Japan. Perhaps the most commonly applied approach to predicting M for a given fish stock is that of Hoenig (1983)[1]. If the Hoenig (1983) method is applied to predict M, based on maximum observed age, and the maximum age was presumed to be 3 years, Hoenig's predicted M would be 1.42 yr^{-1}. A second method, Pauly (1980)[2] predicts M based on estimated regression coefficients for von Bertalanffy K and $L_{infinity}$, and mean water temperature. Using the growth parameters reported in the stock assessment document, and an initial guess of the mean water temperature of 15 C, the Pauly method also gives an M of 1.42 yr^{-1}. In contrast, the maximum observed age would need to be 6 years, for Hoenig's method to predict an M of 0.7 yr^{-1}. If M was 0.7 yr^{-1}, then at age 3, we would expect to see approximately 12% of an unfished cohort surviving. Given that fishing has gone on for several decades, we would expect to see a smaller percentage of age 3 fish but would expect to see at least some age 3 fish in the commercial landings data, if M was 0.7 yr^{-1}. For example, with F at 0.5 and M at 0.7 yr^{-1}, we would still expect to see approximately 2.5% of the cohort at age 3. This would be in contrast, 0.3%, if M was instead 1.42 yr^{-1} and F was on average 0.5 yr^{-1}

Dr. A. Manabe presented at the stock assessment review meeting on 22 January 2026 some study results that indicated that the Ohshimo method tended to produce M estimates on average lower than other commonly applied methods to predict M for several Japanese fish stocks. It was not possible, however, to see whether the estimate of M used in the round herring stock assessment was lower than estimated of M provided by other methods.

---

[1] Hoenig, J. 1983. Empirical use of longevity data to estimate mortality rates. Fishery Bulletin. 81: 989-903.

[2] Pauly, D., 1980. On the interrelationships between natural mortality, growth parameters, and mean environmental temperature in 175 fish stocks. J. Cons. Int. Explor. Mer. 39(2) 175-192.

Sensitivity analysis results in the stock assessment document showed that a higher value for M that was twice the base case value for M would result in larger historical stock biomass estimates, and larger historical SSB estimates with abundance estimates on average being nearly double those given by the base case value for M (Figure 4-2). It also appeared that estimated time trends in stock biomass differed between the different stock assessment scenarios with different assumed values for M. For example, with the higher assumed values for M, the estimated time trends in abundance appeared to be more extreme during periods of stock decline and also periods of stock increase. If the assumed value for M was too small, this could lead to SSB reference points that are also too small relative to true underlying values for the SSB reference points.

C) Discuss whether the basic biological information such as distribution, migration pattern, and population are appropriate.

Figure 2-1 mapped the distribution area of the stock ranging from north to south along the western coast of Honshu and along the western coast of Kyushu (Figure 3-1). The mapped spatial distribution of the stock appears to correspond to the spatial distribution of catches with smallest catches coming from the northern extent, and largest catches coming from the southern extent of the stock's spatial distribution.

D) Evaluate whether the stock assessment methodology is based on the most appropriate available study and performed analytically.

The tuned VPA approach appears to be suitable for the available stock assessment data, i.e., commercial fishery catch-at-age data (derived from annual commercial length composition records and a length-age key) and one fishery independent abundance index and one fishery dependent abundance index. Also, it appears that the fishery harvesting rates have been relatively high which is a requirement for inferences using VPA to be accurate. As the species assessed is very short-lived, i.e., with maximum ages in catches at two years, the methodology could be expected to provide reasonably accurate estimates of stock abundance within a few years of the most recent year of catch data. A maximum age group of 2 years is assumed, and no plus group inclusive of age groups older than age 2 is applied, unlike in the stock assessments for the other species this year. I had asked whether fish older than 2 years had ever been detected in previous samples taken from this fish stock. The answer provided appeared to be, no, no fish older than 2 years had been detected in previous samples taken from catches for this fish stock. Only then, could the assumption of a maximum age group of two years be justifiable. The absence of any fish older than two years, however, would put in question the assumption that the instantaneous rate of natural mortality could be as low as 0.7 yr^{-1}.

The diagnostics for estimation performance coming from the retrospective analysis show no serious retrospective patterns in the key quantities estimated, e.g., SSB, stock biomass, number at age, recruitment and fishing mortality rates. However, as could be expected, the estimates of quantities in the most recent year of each stock assessment show the largest deviations from subsequent updates of the estimated quantities. The effect of tuning to the large surge in egg index estimates for 2015-2017 can be seen in the retrospective analysis. But with additional years of data, the high or low estimates quickly conform to stable estimates with additional years of data added to the assessment.

An important component of the stock assessment is the choice of a stock-recruit function. As with the other three assessed stocks, a so-called "hockey-stick" stock-recruit model had been chosen. One of the justifications for this choice given during the review meeting, was that there was concern that other stock recruit models fitted to the stock-recruit data appeared to lead to extrapolation of spawner abundances that could be associated with upper stock reference points. However, a plot of ln(R/S) versus spawners showed a characteristic linear cloud of data in which a fitted linear model had a significant negative slope (Figure 1). This is a hallmark of Ricker recruitment which implies potential overcompensation at higher stock sizes. Standardized residuals from this fit showed no abnormal residual patterns or outliers. Both models had significant positive lag 1 autocorrelation both at 0.36, both with pvalue = 0.01. When the fit of a Ricker model is compared with the fit of a hockey stick model, the Ricker model was 4.06 AIC units less than that for the hockey stick model.

The hockey stick stock-recruit model which had been considered in some historical stock assessments and in some of Ram Myers' papers a few decades ago has been replaced in most stock assessments globally by the Ricker and Beverton-Holt stock recruit functions. The hockey-stick model fails to accurately represent commonly occurring ecological processes which determine how recruitment of marine fish stocks on average varies with spawner abundance. In Hilborn and Walters (1992[3]) the hockey stock model represents only situations with very strict territoriality and a finite number of spawning nest sites in breeding females. Strict territoriality and a finite number of spawning nest sites is not a known behavioural attribute of marine fishes. The hockey stick model has been identified as having features such as a sharp break between linearly increasing versus constant average recruitment which does not correspond to any known ecological processes commonly shaping fish population dynamics. The hockey stick model also fails on two stringent requirements set for stock recruit model form for fish stocks:

1. The recruitment rate (R/S) should decrease continuously with increases in parental stock size
2. Continuity: The function is a smooth one without any sudden discontinuities or breaks

The Beverton-Holt and Ricker models also have parsimonious mathematical forms. However, both represent commonly occurring ecological processes shaping how recruitment on average varies with spawner abundance and meet all of the conventional requirements for stock-recruit model form. It is recommended that either the Ricker model or Beverton-Holt model be considered as a replacement for the hockey stick model. Based on the much lower AIC in the Ricker model compared to the Hockey stick model, the good fit of the Ricker model to the data (Fig. 1, middle panel), and the lack of realism in the hockey stick model for representing actual ecological processes, a Ricker model would appear to be worth of consideration as a replacement for the hockey-stick model.

E) Evaluate whether the data are treated statistically correctly.

Both abundance indices, i.e., the egg survey index and the commercial purse seine index, were fitted as relative abundance indices with an estimated constant of proportionality, q, which scaled the VPA predicted abundance to the observed annual abundance index. Appropriately, the spawning stock index based on the egg and larvae survey was assumed to be directly proportional to stock size, and the stock biomass index based on the purse seine fishery catch and effort records was assumed to be nonlinearly related to stock biomass with the nonlinearity coefficient estimated at less than 1, indicating hyperstability in the purse seine derived abundance index.

In only one of the four stock assessments were confidence intervals provided for quantities estimated from the VPA. In this stock assessment, although there was tuning to two abundance indices, no confidence intervals were provided for any of the quantities estimated from the VPA.

It is recommended that confidence intervals be provided for these reference point estimates. Google AI summarizes important reasons for this:

Confidence intervals (CIs) are essential alongside parameter estimates (such as means, proportions, or regression coefficients) because they provide critical information about the **precision, reliability, and scientific significance** of a study's findings. Unlike point estimates, which offer only a single "best guess" of a value, CIs provide a range of plausible values that likely contain the true population parameter, thereby accounting for sampling error.

- **Quantifying Precision and Reliability:** A narrow confidence interval indicates high precision and

---

[3] Hilborn, R., & Walters, C. J. (1992). *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty*. Chapman and Hall, New York.

reliability in the estimate, while a wide interval suggests that the estimate is unstable and may vary significantly if the study is repeated.

- **Evaluating Study Robustness:** CIs allow for the comparison of results across different studies to determine consistency. If intervals from different studies overlap significantly, it suggests the results do not differ greatly, regardless of differences in sample size.

- **Assessing Sample Size Adequacy:** The width of the interval is directly related to sample size. A very wide interval often signals that a study was underpowered, or the sample size was too small to draw firm conclusions.

- **Enhancing Communication:** Providing a range of plausible values communicates the inherent uncertainty in statistical analysis more transparently to stakeholders.

It is recommended that biomass estimates could be accompanied by confidence intervals, determined by bootstrapping the tuning index. This could be accomplished at different levels of comprehensiveness, starting with the simplest and most tractable:

1. At the most basic level do bootstrapping accounting for error variability in abundance indices to which the VPA is tuned. This is a minimal level of uncertainty which should be the easiest and most straightforward to implement, and easiest to interpret. Bootstrapping of abundance indices has been addressed in rigorous ways since the early 1990s and there exist numerous publications and implementations of bootstrapping approaches in the context of VPA (e.g., “The "bootstrapping of VPA stock assessment" was a key topic at the Workshop on Risk Evaluation and Biological Reference Points for Fisheries Management, held on November 19-22, 1991, in Halifax, Nova Scotia. https://publications.gc.ca/collections/collection_2016/mpo-dfo/Fs41-31-120-eng.pdf”

2. At the next level, attempt to do bootstrapping of both the abundance index data and catch-at-age data. Thorough understanding of error variability sources in fishery catch-at-age data can be very challenging to establish, especially when catch-at-age data are derived from commercial length composition samples which requires quite strong assumptions to convert sampled length composition to fishery age composition. The length-age key after it is determined, is assumed to be fixed and given with no associated uncertainties in functional form and associated parameter values. It is also assumed that there is stationarity in the length age key over time. Additionally, there exists sampling error in the sampling of lengths from commercial catches in each year. There are thus numerous components of sampling and model fitting error that can contribute to error variability in catch-at-age data derived from length data. A multi-level observation error model would be needed to initiate research into attempts to develop a statistical model that appropriately could represent error variability in catch-at-age data derived from length composition data.

3. Bootstrapping could also allow for uncertainty in parameter values for the instantaneous rate of natural mortality as has been done also in previous bootstrapping implementations of VPA. This would require formulation of a prior distribution for M and then taking random samples from the prior in each bootstrap replicate.

F) Evaluate whether the stock assessment result obtained from the input data and methodology used is appropriate.

The stock assessment results obtained are appropriate and suitable for fisheries management purposes, except for the lack of quantification of uncertainty in most of them. As mentioned above, there exist numerous rationale for why it is appropriate to compute confidence intervals associated with the results obtained out of stock assessment models and in most regions where stock assessments are carried out, it is common practice for confidence intervals or analogous Bayesian probability intervals to be computed and shown for estimated quantities. Confidence intervals provide a standardized representation of uncertainty in estimated quantities to indicate how variable the estimates of quantities of interest could be based on different components of sampling error variability associated with the sampling processes associated with collection and processing of data used in the stock assessment. Implementing the stock assessment modeling in a probabilistic framework would also allow computation of the probability that stock size is increasing or decreasing and the probability that stock size is above pre-defined stock reference points.

The Kobe plot provided appropriately shows the trajectory of stock and fishery status with respect to e.g. the limit and target stock reference points and the F/Fmsy reference point of 1, over the range of years where stock assessment results were obtained. The Kobe plot is shown only for the base case stock assessment model run and there is no representation of uncertainty in the stock status plot. It is recommended that when sensitivity analysis is conducted on different values for M, that two additional Kobe plots are shown in appendices. One of these would be based on the assumption of the largest value for M tried in the sensitivity analysis carried out on M, and the other would be based on the lowest value considered for M (Figure 4-6).

G) Evaluate the validity of methodology and result used for the future projection.

The methodology for the future projection uses the base case model estimates of numbers at age for the starting point for the future projection. No uncertainty is considered in the numbers at age starting point for future projections. This is unrealistic, since it is not possible for there to be complete certainty in the abundance at age in the most recent year of the stock assessment, as this is typically the most uncertain set of abundance at age in a VPA stock assessment, owing to most recent year having the least amount of data being available for abundance estimation. Other VPA stock assessment methodologies applied elsewhere such as at ICCAT, regularly apply bootstrapping of the VPA to generate numerous plausible vectors of numbers at age and other associated quantities such as vulnerability at age in the final year of the stock assessment which then serve as starting points for the projections in which alternative policy options are evaluated. It is recommended that the approach to future projection be updated to incorporate uncertainty in the starting point of the future projections.

Uncertainty, e.g., in recruitment residuals is propagated in the projection of future years. Under a specific harvest policy option fishing mortality rates at age are applied in each future year projected. A single best estimate of fishery selectivity at age is applied based on the average Fs at age based on an average from estimates obtained in recent years: “For selectivity and the average body weight of the catch in the future projections, we continued to use the same values as were used to estimate the various proposed reference points in the “Materials for the Research Institute Meeting” described above (Yoda et al. 2021). Similar to the stock-recruitment relationship, these values are based on the FY 2020 stock assessment, with selectivity being the average values for the period from 2016 to 2018 and the average body weight in catches being the average values for 2017 to 2019. The fishing pressure in 2024 (F2024) was set to the F value that gives the %SPR corresponding to the age-specific fishing pressure from 2021 to 2023 in this year’s assessment under the same selectivity and biological parameters (average body weight, etc.) as those used when calculating the reference points” (p.32). However, Figure 4-7 shows estimates of F for age 0 and age 1-2 fish from 1976 to 2023 that show that the selectivity at age for these age groups can vary considerably between years. It is recommended that uncertainty associated with time varying selectivity be accounted for in the projections and also in the computation of MSY and SBMSY reference points.

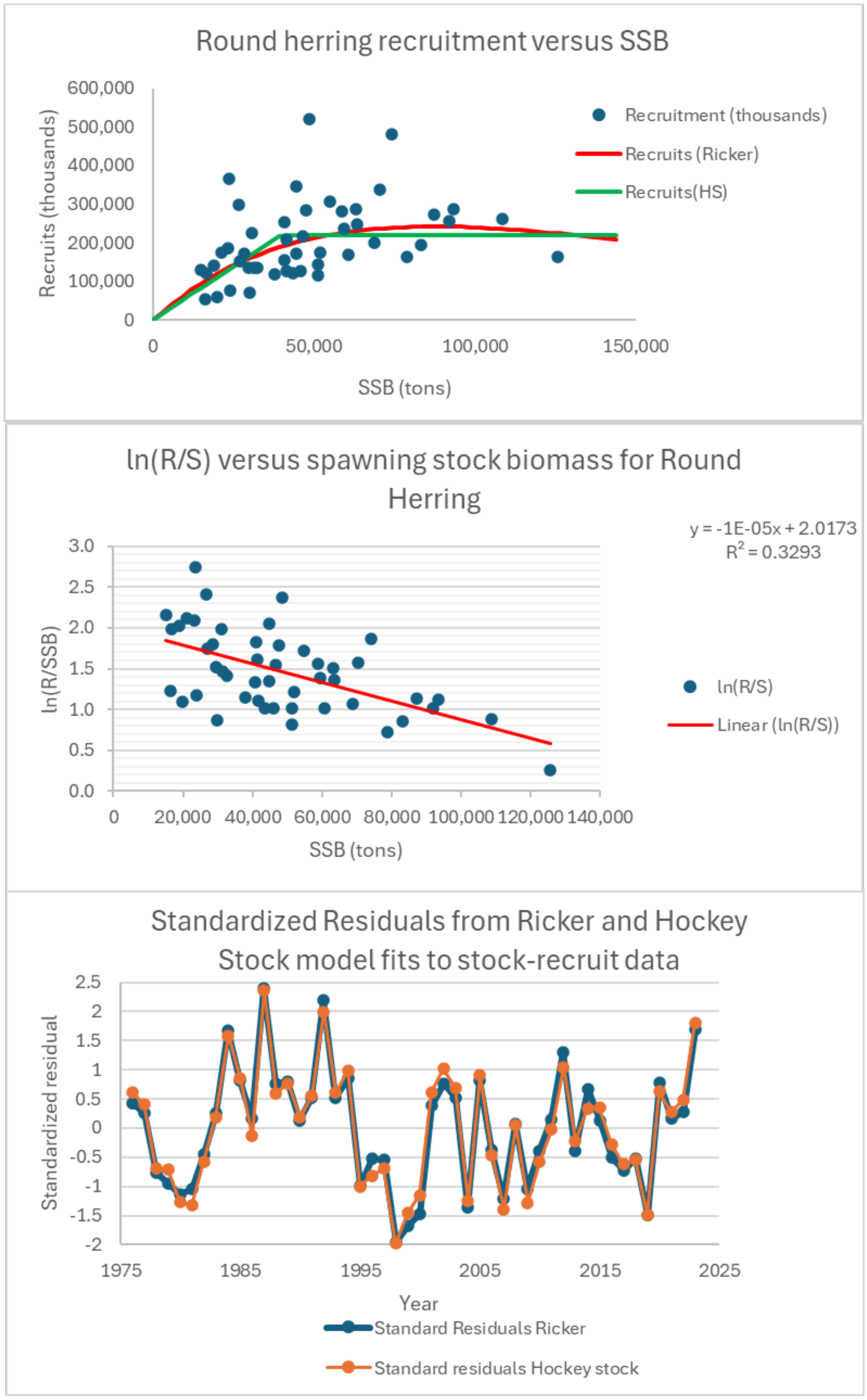


Figure 1. Top panel: Plots of the fits of the Ricker and hockey stick model to the stock-recruit data for round herring using stock-recruit data from the round herring stock assessment document. Middle panel: fit of a linear model to ln(R/S) versus SSB data for round herring. Bottom panel standardized residuals from fit of a Ricker model to the ln(R/S) versus SSB data for round herring. Based on the maximum likelihood fits of the Ricker and hockey stick model, the Ricker model had an AIC 4.06 AIC units smaller than that for the hockey stock model. Both models had significant positive lag 1 autocorrelation both at 0.36, both with pvalue = 0.01.

Stock Assessment #2: Shotted Halibut

A) Determine whether the data used for stock assessment are adequate to understand the stock dynamics of the target species and represent the best scientific information available.

Catch records and abundance indices

The stock assessment uses a time series of reconstructed fishery catch numbers at age from 1993 to 2023 (Figure 3-2) from landings from Japanese operated vessels in the East China Sea. This is despite having a much longer time series of catch biomass going back to 1966 (Figure 3-1). At 30 years, this appears to be a sufficiently long time series of catches to provide information about stock dynamics, given that catch-at-age has a plus group of 4+. The total recorded annual catch in both biomass and numbers shows moderate fluctuations, e.g., ranging between about one to two thousand tons up to about 2010, a pronounced decrease to about 0.3 thousand tons in the last few years of the time series. These fluctuations in catch would appear to be informative about shotted halibut stock dynamics over the past few decades. However, catch biomass was up to as high as about 5 tons in the 1970s, considerably higher than the maximum catch biomass for the years 1993 and after that were included in the VPA (Figure 3-1). An unstandardized stock abundance index was also compiled ranging back also to about 1966 (Figure 4-2). This shows high catch rates from 1966 to the early-1980s fluctuating between about 35 kt to 55 kt, and then a precipitous decline to about 15 kt in the late 1980s with the index fluctuating and gradually increasing up to about 2010 (Fig 4-1, 4-2). After 2010, the unstandardized index drops to an all-time low of about 5.5 kt in 2023. The standardized index available from 1993-2023 shows time trends very similar to the unstandardized index.

Applying a stock assessment method that could incorporate the catch biomass and stock abundance index data stretching back to 1966 could potentially be considerably more informative about stock dynamics than applying an untuned VPA starting in 1993. Age-structured stock reduction analysis (SRA) is a more flexible stock assessment methodology that unlike VPA does not require age composition data to be available for the full time series of application; SRA instead requires an uninterrupted long (i.e., multi-decadal) time series of total catch biomass (see for example McAllister and Ianelli 1997, Can. J. Fish. Aquat. Sci.[4]). SRA models can also be fitted to the available catch-at-age data that can start well after the start of the catch biomass series, and require at least one relative index of abundance that spans at least the final two decades of the fishery. This alternative stock assessment approach which could incorporate catch biomass and abundance index data all the way back to 1966 could provide considerably more information about the form and parameters of the stock-recruit function for this stock than the current VPA approach. SRAs can also be fitted to either reconstructed catch-numbers-at-age or fishery length compositions (see, for example, Wor et al. 2018, Fisheries Research[5]; Licandeo et al. 2020, CJFAS[6]).

It was observed that "The number of young fish caught has remained few, and the proportion of age 3 and age 4 fish in the catch is gradually increasing" since the early 2000s (p. 5). On p. 9 it is stated "it is important to note that the proportion of small bottom trawls in the catch from the stock is increasing…. The assessment of this stock is largely based on information from the two-boat offshore bottom trawls (west of Hamada), and all age-specific catch calculations and standardized CPUE are based on the catch information and biological measurement results of the two-boat offshore bottom trawls (west of Hamada)". The apparent increase in the proportion of age 3 and 4 fish in the catch in recent years could reflect either a systematic decrease in

---

[4] McAllister, M.K., and Ianelli, J.N. 1997. "Bayesian stock assessment using catch-age data and the sampling/ importance resampling algorithm" Can. J. Fish. Aquat. Sci. 54, 284-300.

[5] Wor, C., van Poorten, B., Licandeo, R., Walters, C.J. 2018. Stock reduction analysis using catch-at-length data: Length-SRA. Fisheries Research 208: 124-132.

[6] Licandeo, R., Duplisea, D. E., Senay, C., Marentette, J. R. and McAllister, M.K. 2020. Management strategies for spasmodic stocks: a Canadian Atlantic redfish fishery case study. Can. J. Fish. Aquat. Sci. 77(4): 684-702.

recruitment rates or a gradual shift in fishery selectivity to favouring ages 3 and 4 over ages 1 and 2 (or both). That the available abundance indices (Figs. 4-1 and 4-2) both continue to decrease in recent years even with large decreases in catch biomass over the past two decades (Fig. 3-1) suggests that dwindling recruitment rates may at least be partly to blame for the long-term declines in the abundance indices and decreasing proportion of age 1 and 2 fish in the commercial fishery age composition data.

It is stated on page 6 of the assessment document "A tuning using standardized CPUE for the two-boat offshore bottom trawls in the VPA was not performed because retrospective bias was observed that overestimated stock biomass and SSB and underestimated the fishing coefficient". Despite apparent retrospective bias patterns when the VPA is tuned to the abundance index, this may not necessarily be a good reason to redo the VPA without tuning to an abundance index. There may be other reasons than tuning to the abundance index for retrospective patterns to occur. For example, the M applied in the VPA may be incorrect, or there may be issues with computing catch-numbers-at age from commercial length composition data using the length-age key. By not tuning the VPA to the abundance index which the VPA was tuned to in the previous stock assessment of shotted halibut, important information on fishing mortality rate and abundance trends may be left out when the VPA is run without tuning to an abundance index. A VPA that is tuned to an abundance index rather than an untuned VPA could help to more decisively distinguish between the alternative hypotheses that recruitment rates over recent years have been decreasing and fishery selectivity has been stationary versus recruitment rates have not been systematically decreasing but fishery selectivity has been shifting away from younger ages 1 and 2 to increased selectivity for ages 3 and 4.

Commercial length composition data were transformed into age composition data using an age-length key. "The age-length key used to calculate the age-specific catch of this stock is based on the measurements in the 1990s." It is possible that growth and length at age could be time varying and using a length-age key based on 1990 could lead to actual length-age keys that are different from those based on 1990 information. Application of a length-age key based on a fixed point in time to commercial length composition records when the length-age relationships are time varying could lead to generation of misrepresentations of age-composition records for use in the VPA stock assessment.

On page 9 it is also stated that "These measurements lack the large individuals measurements…. [T]he lifespan of this species is about 7 years, and the proportion of 3 and 4+ ages in the catch is high, so the fishing coefficient may change even after 4+ years." It is thus possible that information about age composition in the catch, fishing mortality at age, and time varying selectivity may be lost by the use of a plus group of 4+ years when it may have been possible to implement a plus group at an older age, e.g. 6 or even 7 years. It is thus recommended that scientists investigate whether it may be appropriate to increase the plus group to older than 4+ years.

In addition, as is common with generation of age composition records it is likely that there could be smearing in the age composition records generated from the commercial length composition records. This could result in stronger age classes being under-represented in the generated age-composition data and neighbouring age classes being over-represented. It may be appropriate to develop a simulation-estimation methodology to quantify the expected estimation performance of the current approach to deriving age composition data from length composition data using a fixed length-age key and also considering plus groups older than ages 4+.

B) Discuss whether the biological parameters used for stock assessment are appropriate.

A value for the instantaneous rate of natural mortality was assumed at 0.35 per year based on (Tanaka 1960). The Tanaka method to determine M is not well known outside of Japan. Perhaps the most commonly applied approach to predicting M for a given fish stock is that of Hoenig (1983). If the Hoenig (1983) method is applied to predict M, based on maximum observed age, and the maximum age was presumed to be 7 years, Hoenig's predicted M would be 0.6 yr^{-1}. In contrast, the maximum observed age would need to be 12 years, for Hoenig's method to predict an M of 0.35 yr^{-1}. If M was 0.35 yr^{-1}, then at age 7, we would expect to see approximately 8.6% of an unfished cohort surviving. Given that fishing has gone on for several decades, we would expect to see a smaller percentage of age 7 fish but would expect to see at least some age 7 fish in the commercial landings data, if M was 0.35 yr^{-1}. For example, with F at 0.5 and M at 0.35 yr^{-1}, we would still expect to see approximately 0.3% of the cohort at age 7. This would be in contrast, 0.02%, if M was instead 0.7 yr^{-1} and F was on average 0.5 yr^{-1}.

Dr. A. Manabe presented at the stock assessment review meeting on 22 January 2026 some study results for several Japanese fish stocks that indicated that methods applied in Japanese stock assessments for predicting M tended to produce M estimates on average lower than methods commonly applied in other regions to predict M. It was not possible, however, to see whether the estimate of M used in the shotted halibut stock assessment was lower than estimated of M provided by other methods.

Sensitivity analysis results for the year 2023 in the stock assessment document showed that a higher value for M that was 0.1 yr^{-1} higher than the base case value for M would result in larger recruitment, stock biomass, and SSB estimates for 2023 (Figure 4-5). It is recommended that further attention is given to identifying an appropriate base case value for M to apply in this stock assessment, especially given Dr. Manabe's findings in his study about comparing M estimates from different methodologies for Japanese fish stocks.

On p. 4, the fraction maturing at age is given for males and females. For SSB calculations the proportion maturing at age should be referenced to females only. However, it appears that the proportion maturing at age for SSB calculations was incorrectly referenced to fraction maturating-at-age in males.

C) Discuss whether the basic biological information such as distribution, migration pattern, and population are appropriate.

Figure 2-1 shows the "Shotted halibut stock distribution in the southwestern part of the Sea of Japan". The two main spawning areas mapped fall within the mapped spatial distribution of the catches. As the mapped area of the stock falls within a very small portion of the total mapped ocean area, it is recommended that future maps of fishing and spawning areas for shotted halibut zoom in to show more spatial resolution, so that the mapped area is cropped to show with higher resolution and detail the mapped spawning and fishery areas for the stock (e.g., between 120-140 degrees East to West and between 30-40 degrees North).

D) Evaluate whether the stock assessment methodology is based on the most appropriate available study and performed analytically.

The untuned VPA approach appears to be too simple for the available stock assessment data, i.e., commercial fishery catch-at-age data (derived from annual commercial length composition records and a length-age key) and leaves out the one available standardized fishery dependent abundance index 1993-2023. If the VPA was instead tuned to the standardized index, and an appropriate value for M was applied based on further research such as that by Dr. Manabe, greater confidence could be reached in the stock abundance and fishing mortality rate estimates obtained from the VPA.

The diagnostics for estimation performance coming from the retrospective analysis show no serious retrospective patterns in the key quantities estimated, e.g., SSB, stock biomass, number at age, recruitment and fishing mortality rates (Supplementary Figure 2-1).

An important component of the stock assessment is the choice of a stock-recruit function. As with the other three assessed stocks, a so-called "hockey-stick" stock-recruit model had been chosen. One of the justifications for this choice given during the review meeting, was that there was concern that other stock recruit models fitted to the stock-recruit data appeared to lead to extrapolation of spawner abundances that could be associated with upper stock reference points. A plot of ln(R/S) versus spawners showed an increasing trend in ln(R/S) as spawner biomass increased (Figure 2). This suggests that as stock size has decreased, stock productivity has also decreased and suggests that over the range of observed estimated values for spawning stock biomass there is a lack of compensation in stock productivity. This is a very concerning pattern in population dynamics, because for a normally behaving stock we would expect to see ln(R/S) increasing as spawning stock decreases. Observing decreasing stock productivity as the stock decreases suggests an unstable situation in which even with reductions in total annual catches and fishing mortality rates, productivity will continue to decrease. Possible hypotheses for decreasing stock productivity as stock decreases could be increased predation or an increase in unreported landings taken from the stock. Because the slope of the fit of a linear model to ln(R/S) versus SSB data is positive, a Ricker model will fail to represent the stock-recruit data, since the slope in the ln(R/S) versus S observations must be negative for a Ricker model to fit. A hockey stock model appears to fit the data (Figure 2) and it is unlikely that a Beverton-Holt model would fit the data any better, since the Beverton-Holt model also would predict compensatory recruitment dynamics. An alternative, more nuanced approach to representing stock-recruit dynamics could be to consider a stock-recruit model with time varying stock productivity. This could be accommodated better if a stock assessment method such as SRA (e.g., Licandeo et al. 2020) had been applied that could be fitted to the much longer time series of catch and stock biomass data going back to 1966, rather than only back to 1993. This is because a model fitted to a much longer time series of catch and abundance data when the abundance and catch data were much higher in earlier years could be more informative about stock-recruit model parameters and hypotheses for time varying stock-productivity.

One anomalous artifact associated with the use of the hockey stick model is the prediction of quasi-triangular average yield-at-age versus spawning stock biomass curves, as was initially pointed out by Professor Yamakawa, the other reviewer, at the stock assessment review meeting (Supplementary Figure 3-2). In fisheries, plots of average yield-at-age versus spawning stock abundance could never be expected to conform to a semi-triangular distribution as seen in Supplementary Figure 3-2. Yield versus SSB curves under a given harvest policy option could only be expected to have relatively smooth curvatures without a sharp angular peak as seen in Supplementary Figure 3-2 where for ages 2 and older there is a sharp peak to the yield curve for each age. Although a hockey stock model would appear to provide a reasonable fit to the stock-recruit data for this stock, it is recommended that a replacement stock-recruit function be adopted, so as to avoid the making of predictions of yield curves and population dynamics that fail to conform to credible population dynamics behaviours in the real world.

E) Evaluate whether the data are treated statistically correctly.

In only one of the four stock assessments were confidence intervals provided for quantities estimated from the VPA. In this stock assessment, no confidence intervals were provided for any of the quantities estimated from the VPA. As I did for the other three stocks, it is recommended that confidence intervals be computed and provided for abundance, fishing mortality rate, and management reference points. Such an approach could be more readily accommodated by applying a VPA tuned to an available abundance index.

F) Evaluate whether the stock assessment result obtained from the input data and methodology used is appropriate.

The stock assessment results obtained are appropriate and suitable for fisheries management purposes, except for the lack of quantification of uncertainty in most of them. As mentioned above, there exist numerous rationale for why it is appropriate to compute confidence intervals associated with the results obtained out of stock assessment models and in most regions where stock assessments are carried out, it is common practice for confidence intervals or analogous Bayesian probability intervals to be computed and shown for estimated quantities. Confidence intervals provide a standardized representation of uncertainty in estimated quantities to indicate how variable the estimates of quantities of interest could be based on different components of sampling error variability associated with the sampling processes associated with collection and processing of data used in the stock assessment. Implementing the stock assessment modeling in a probabilistic framework would also allow computation of the probability that stock size is increasing or decreasing and the probability that stock size is above pre-defined stock reference points.

The Kobe plot provided appropriately shows the trajectory of stock and fishery status with respect to e.g. the limit and target stock reference points and the F/Fmsy reference point of 1, over the range of years where stock assessment results were obtained. The Kobe plot is shown only for the base case stock assessment model run and there is no representation of uncertainty in the stock status plot. It is recommended that when sensitivity analysis is conducted on different values for M, that two additional Kobe plots are shown in appendices. One of these would be based on the assumption of the largest value for M tried in the sensitivity analysis carried out on M, and the other would be based on the lowest value for M considered in the sensitivity analysis for M.

G) Evaluate the validity of methodology and result used for the future projection.

The methodology for the future projection uses the base case model estimates of numbers at age for the starting point for the future projection. No uncertainty is considered in the numbers at age starting point for future projections. This is unrealistic, since it is not possible for there to be complete certainty in the abundance at age in the most recent year of the stock assessment, as this is typically the most uncertain set of abundance at age in a VPA stock assessment, due to most recent year having the least amount of data being available for abundance estimation. Other VPA stock assessment methodologies that have been applied elsewhere such as at ICCAT, regularly apply bootstrapping of the VPA to generate numerous plausible vectors of numbers at age and other associated quantities such as vulnerability at age in the final year of the stock assessment which then serve as starting points for the projections in which alternative policy options are evaluated. It is recommended that the approach to future projection be updated to incorporate uncertainty in the abundance at age in the starting year for future projections.

Uncertainty, e.g., in recruitment residuals is propagated in the projection of future years. Under a specific harvest policy option fishing mortality rates at age are applied in each future year projected. A single best estimate of fishery selectivity at age is applied based on the average Fs at age based on an average from estimates obtained in recent years. This appears to be a reasonable assumption based on Figure 4-6 which suggests that the ratios of F at age for ages 1 to 4+ appear to be relatively stable from 2013 to 2023.

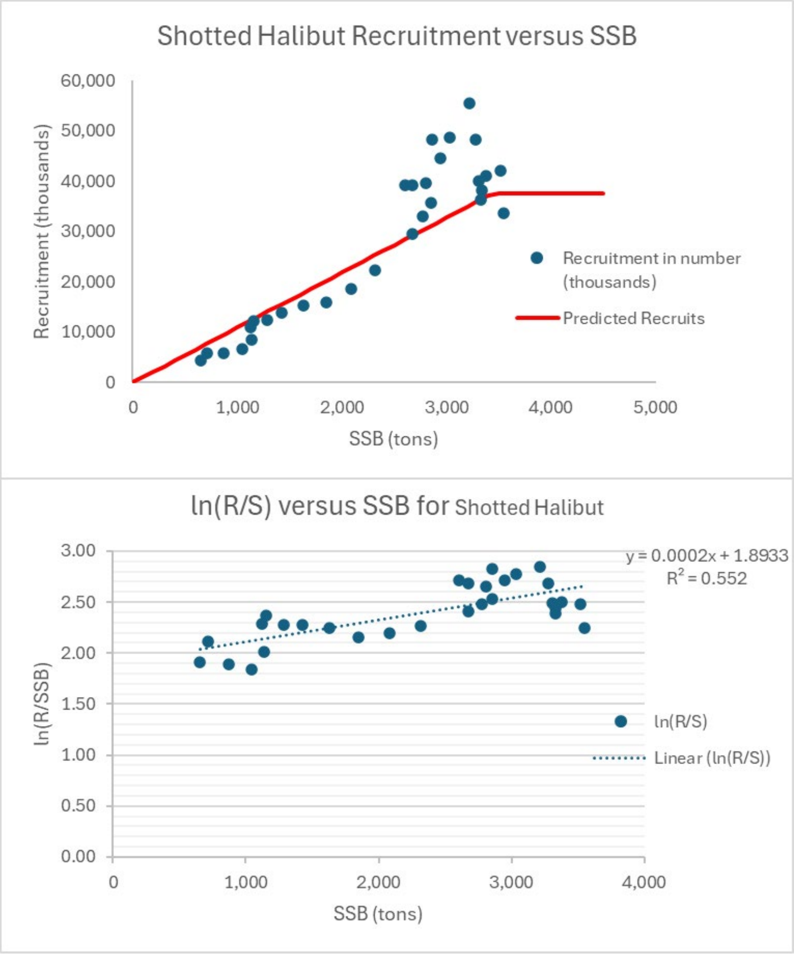


Figure 2. Top panel: Plots of the fits of the hockey stick model to the stock-recruit data for shotted halibut using stock-recruit data from Table 4-1 in the shotted stock assessment document. Bottom panel: fit of a linear model to ln(R/S) versus SSB data for shotted halibut.

Stock Assessment #3: Splendid Alfonsino

A) Determine whether the data used for stock assessment are adequate to understand the stock dynamics of the target species and represent the best scientific information available.

Catch records and abundance indices

The stock assessment uses a time series of reconstructed catch-at-age records from 1998 to 2023 from landings from Japanese operated vessels in coastal and offshore waters south of Honshu and southwest of Kyushu. At 26 years, this is a relatively short time series of catches to provide information about stock dynamics, especially given that splendid alfonsino is relatively long-lived species, i.e., with a maximum age of 15+ years in catch records. The total recorded annual catch in numbers shows relatively small fluctuations, e.g., ranging between about 600-900 (x10,000 fish) per year between 1998 and 2011 and smaller fluctuations between about 500-600 (x10,000 fish) between 2012-2023. These small-fluctuations in total catch to the present since 1998 would not appear to be very informative about splendid alfonsino stock dynamics over the past few decades, especially because the eleven time series of cpue (standardized and non-standardized) available from eleven different fishing areas for splendid alfonsino (Figure 4-1) do not appear to show any apparent consistent responses to the fluctuations in total catches since 1998. However, the plotted time trend in catch weight starts in 1976 (Figure 4-1) and shows total catch biomass increasing from a little over 2000 tons in 1976 up to about 10,500 tons in about 1984. Catches subsequently decrease to about 6500 tons in 1996. Applying a different stock assessment modeling approach such as an age-structured stock reduction analysis that was fitted to either the available fishery length composition or age composition data (McAllister and Ianelli 1997, cjfas; Licandeo et al. 2020, cjfas) could help to provide time series of abundance estimates with a much wider range of abundances and thereby an improved basis for understanding long-term stock dynamics.

There appears to be relatively little information about cohort strength in the commercial fishery catch-at-age data for splendid alfonsino (Figure 3-2). The time series of catch-at-age data with numerous age groups in Figure 3-2 surprisingly show very little interannual variation in cohort strength over the 26-year time series of catch-at-age data. Quite commonly in commercially fished stocks, a large cohort will be visible in a large diagonal band that progress through the catch-at-age time series in catch-at-age plots or tables, with a large cohort first emerging in the youngest age category in a given year and then progressing into each older age category in each successive year. Small cohorts will be visible as very thin and progressively thinning diagonal bands in fishery catch at age plots or tables. No such bands demarking smaller or larger cohorts are visible in the catch-at-age plot shown in Figure 3-2.

The estimated time series of recruitment seen in Figure 4-3 is extraordinarily smooth, as was pointed out by the other reviewer of this and the other three stock assessments, Professor Yamakawa. The time series of estimated recruitment in Figure 4.3 shows some gradual systematic undulations with runs of recruitment trending up and down over series of three to seven years for each leg of increasing or decreasing recruitment. This is in contrast to strong patterns of interannual variability seen in estimated time series of recruitment for other assessed fish stocks which have more intensive sampling programs for the age composition of annual catches. While it may be possible for there to be very little interannual variability in cohort strength over this 26-year period, we would expect to see more interannual variability in cohort strength than is visibly present. Possible reasons for why actual interannual variability in recruitment in this fish stock is not detected with the stock assessment methodology that has been applied include the following. Firstly, the fishery catches sampled might be in locations where there is differential mixing of cohorts such that fish from larger cohorts tend to be sampled more thinly than fish from smaller cohorts. This could result from territoriality or more pronounced density dependent dispersal of fish from large cohorts. Secondly, the incorrect age-assignment in the transformation of fishery length sample compositions into fishery catch-number-at-age using an age-length key could lead to smearing of fish abundance in larger cohorts into adjacent cohorts that have lower abundance. The abundance of large cohorts in catch-at-age records would thus be under-estimated and the abundance of adjacent smaller cohorts would be over-estimated. These two alternative processes could generate overly smooth time series of estimates of recruitment.

The VPA was tuned to nine stock abundance indices, one from each of nine different Japanese fishing grounds for splendid alfonsino (Supplementary Figure 2-2). It was assumed that each index provided an independent sample of recruited fish stock abundance. However, the nine abundance index time series mostly show quite strongly differing time trends in cpue on the different fishing grounds. The plots of the VPA model predicted index versus the observed index for the nine cpue time series shows mostly quite poor fits to these abundance index data, with time series 4, Kozushima, showing the least poor fit (Supplementary Figures 2-2 and 2-3). But even for the least poorly fitting cpue series, there appear to still be apparent time trends in model fit residuals (Figure 2-3). The generally poor fits of the VPA to all nine abundance indices leaves open the question about what is the actual underlying trend in total recruited stock biomass. Should we believe for example that the Kozushima index provides the best indicator of total stock abundance since the model fits this time series the least-worst? Or is this simply the outcome of tuning the VPA to all nine time series, with the Koshuma index corresponding to the least squares VPA model fit to all nine indices even if it is not necessarily the most accurate representation of stock trends? The fitting of a VPA model to nine abundance indices that all show different time trends thus leaves considerable uncertainty over what the actual stock trend could be.

An alternative approach that could be considered would be to treat each fishing ground as a separate stratum and to compute a swept-area abundance estimate for each fishing ground area. This would be facilitated if the same type of fishing gear was used on each fishing ground over the years and catch per unit of effort by year was then computed from these same gear types on each of the fishing grounds. The average catch per unit of fishing effort (cpue) per fishing ground per year could then be used as a representation of average fish density on the fishing ground per year. The swept area estimate of abundance for each year could then be obtained by taking the product of average cpue and the total area of the fishing ground. An index of total abundance per fishing ground per year could then be obtained. The total swept area estimate of recruited stock biomass in a given year would then be the sum of the swept area abundance estimate per fishing ground in a given year. This would then provide a single time series of abundance for the entire recruited stock biomass based on all nine fishing grounds. The swept area biomass contributed by each fishing ground would be in proportion to the product of annual average catch rate and the total stock area of the fishing ground. The total area of each fishing ground could be computed using habitat suitability indices such as bottom depth and bottom type that are known to be consistent determinants of splendid alfonsino presence. A paper that documents this fishery dependent swept area biomass estimation approach is Kirchner and McAllister (1999, Fish. Res.[7]). The VPA could then be tuned to the resulting single total swept area abundance index which would provide a representation of the total stock biomass and how it has varied over the years where the total cpue index is available. A more credible fit of the VPA to a single abundance index that was derived from cpue data from all nine fishing grounds could then be obtained than from fitting the VPA to nine different abundance indices from the nine fishing grounds.

Commercial length composition data were transformed into age composition data using two age-length keys. The biologists "created two age-length keys, one each for the coastal and offshore fishing grounds, and applied them to each landing port and fishing method in order to calculate the catch in number at age, which consists of fish age 2 to 14 and fish age 15 and older" (p.32). This appears to be an appropriate approach given the widely differing age compositions of fish sampled on near shore and offshore fishing grounds. It also appears that it is assumed that the age-length keys are stationary and not time varying, although the inputs to the keys are updated each year as more data are gathered: "The relationships between age and body length and age and grade (body weight) use data compiled over multiple years, which are updated every year with additional age assessment information. The addition of such information also updates the average body weight at age and age-length key every year" (p.32). While this approach allows for the inputs to the keys to be updated as more data are gathered, it still assumes that once the key is formed based on all available data, the keys themselves when applied to all years of fishery length compositions are time invariant. In contrast, it is possible that growth and length at age could be time varying and using two fixed length-age keys (one each for coast and offshore fishing grounds) based on data from all years of available data could lead to actual length-age keys

---

[7] Kirchner, C.H., and McAllister, M.K. 2001. "A new improved method to compute swept area estimates of biomass from commercial catch rate data: application to Namibian orange roughy (Hoplostethus atlanticus)." Fish. Res. 56: 69-88.

that vary over time and are different from those based on the study. Application of length-age keys that are each assumed to be stationary over years to commercial length composition records when the length-age relationships are time varying could lead to generation of misrepresentations of age-composition records for use in the VPA stock assessment. If available data permit, it may be of interest to test whether the null hypothesis that growth and age length keys are not time varying. If this hypothesis were found to be inconsistent with available data, it may then be appropriate to formulate an approach that allowed for the length-age keys to be time varying. In addition, as is common with generation of age composition records it is likely that there could be smearing in the age composition records generated from the commercial length composition records. This could result in stronger age classes being under-represented in the generated age-composition data and neighbouring age classes being over-represented. It would be appropriate to develop a simulation-estimation methodology to quantify the expected estimation performance of the current approach to deriving age composition data from length composition data using a fixed length-age key.

B) Discuss whether the biological parameters used for stock assessment are appropriate.

A value for the instantaneous rate of natural mortality was assumed at 0.1 per year based on (Tanaka 1960). The Tanaka method to determine M is not well known outside of Japan. Perhaps the most commonly applied approach to predicting M for a given fish stock is that of Hoenig (1983). If the Hoenig (1983) method were to be applied to predict M, based on maximum observed age, and the maximum age was presumed to be 26 years, Hoenig's predicted M would be 0.16 yr^{-1}. In contrast, the maximum observed age would need to be 40 years, for Hoenig's method to predict an M of 0.1 yr^{-1}. If M was 0.1 yr^{-1}, for age 15, we would expect to see approximately 22% of an unfished cohort surviving. For age 26 we would expect to see 7.4% of a cohort surviving. Given that fishing has gone on for several decades, we would expect to see a much smaller percentage of age 15 fish, if M was 0.1 yr^{-1}. For example, with F at 0.15 yr^{-1} (approximate average 1998-2023 in Fig. 4-6) and M at 0.1 yr^{-1}, we would still expect to see approximately 2.35% of the cohort at age 15. The percentage remaining in a cohort would be in contrast, 1%, if M was instead 0.16 yr^{-1} and F was on average 0.15 yr^{-1}. Thus, it appears that the value for M assumed could potentially be too low, when considering an alternative method that has more commonly applied method to predict M for fish stocks. Using a value for M that is too low compared to the actual value, would be expected to generated abundance estimates that are too low on average. Predictions of the effects of using values for M either too low or too high are shown in Figure. 4-4 which confirms that applying a value for M 1.5 times higher than the base case value would generate estimates of spawning stock biomass, stock biomass and recruitment roughly 30-50% larger than under the base case assumed value for M.

Dr. A. Manabe presented at the stock assessment review meeting on 22 January 2026 some study results that indicated that for several Japanese fish stocks some M-estimation methods used in Japanese stock assessment tended to produce M estimates on average lower than other commonly applied methods to predict M. It was not possible, however, to see whether the estimate of M used in the splendid alfonsino stock assessment was lower than estimated of M provided by other methods. If the assumed value for M was too small, this could lead to SSB reference points that are also too small relative to true underlying values for the SSB reference points.

C) Discuss whether the basic biological information such as distribution, migration pattern, and population are appropriate.

Figure 2-1 mapped distribution of fishing grounds for the stock ranging from the southern central coast of Honshu to several hundreds of kilometers offshore and extending southwest offshore from the western coast of Kyushu (Figure 2-1). The mapped spatial distribution of the stock appears to correspond to the spatial distribution of catches with smallest catches coming from the northern extent, and largest catches coming from the southern extent of the stock's spatial distribution. There appears to be a very large area in the north Pacific Ocean where splendid alfonsino are caught. Are there any studies to assess stock-structure throughout the North Pacific Ocean? Can this reasonably be assumed to behave as a single breeding stock?

Or could there be more than one separate breeding stock of S. Alfonsino in the North Pacific Ocean?

D) Evaluate whether the stock assessment methodology is based on the most appropriate available study and performed analytically.

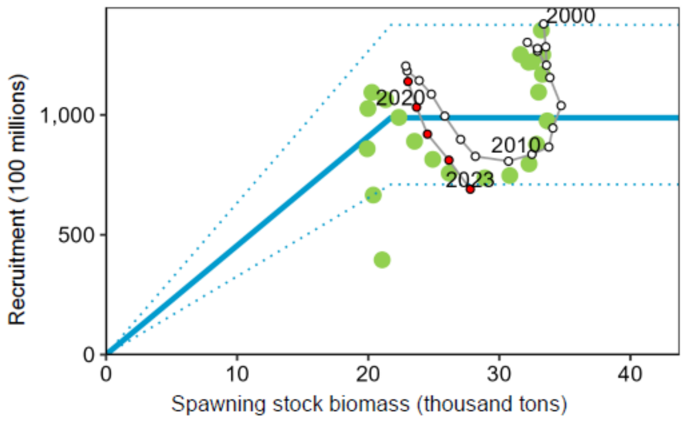
The tuned VPA approach appears to be suitable for the available stock assessment data, i.e., commercial fishery catch-at-age data (derived from annual commercial length composition records and length-age keys). However, as pointed out above, the ability to provide credible estimates of the stock trends may be compromised by fitting the VPA to nine different commercial fishery abundance index time series from nine different fishing grounds for the stock. This resulted in relatively poor fits of the VPA to all 9 abundance index time series. As mentioned above, a more credible result could have been obtained if an effort was made to derive a single abundance index for the stock which was derived from the catch, effort, and stock area information from the nine different fishing grounds. It would have added credibility to the stock assessment estimation of abundance time series if the VPA could also have been tuned to at least one fishery independent abundance index. Also, it appears that the fishery harvesting rates have been moderate to light and have changed relatively little between 1998 and 2023. Selectivity and fishing mortality rates for the youngest age classes, i.e., 2-4 years have been very low for the full time series (Figure 4-5). This could potentially limit the ability of the stock assessment to estimate cohort strength. It is thus likely that there exists considerable uncertainty in estimates of relative cohort strength as a result.

Due to the apparent sparse information in the stock assessment data that were used on cohort strength, it appears that there could be considerable value to the stock assessment in developing a new annually run fishery independent survey that could provide new data to help to improve estimates of cohort strength. This survey could be a fishery independent bottom long-line survey that was located in a stock area that was known to have high influxes and relatively high densities of younger age groups of splendid alfonsino, e.g., ages 2-5 years.

The diagnostics for estimation performance coming from the retrospective analysis show no serious retrospective patterns in SSB, stock biomass, number at age for older fish, and fishing mortality rates for older fish (Supplementary Figure 2-5). However, for numbers of age 2 and 3 and recruits there appears to be some pronounced retrospective patterns. This shows a potential to either considerably over-estimate or under-estimate numbers at age 2 and 3 and recruitment strength for a few years in a row. The retrospective patterns indicate that estimates of age 2 and age 3 and recruitment in a VPA assessment of this stock can be highly uncertain. This makes sense due to the relatively limited amount of information in the fishery data about cohort strength, especially given the very low relative selectivity for ages 2-4 fish.

For retrospective patterns in biomass, fish number, and fishing mortality rate at age are very difficult to inspect due to cluttering the time series of different years and age groups together. As there are age groups running from 2 to 14+ years old, plotting time series for all fourteen of these retrospective time series on the same panel makes it impossible to closely inspect for retrospective patterns per age group. To make the patterns easier to inspect, it is recommended that the range of values for biomass in the biomass plot are narrowed to a minimum of 30000 and maximum of 50000, the fish numbers at age and fishing mortality at age retrospective plots be separated into separate plots containing smaller groups of ages (as was done in Figure 4-5).

An important component of the stock assessment is the choice of a stock-recruit function. As with the other three assessed stocks, a so-called “hockey-stick” stock-recruit model had been chosen. One of the justifications for this choice given during the review meeting, was that there was concern that other stock recruit models fitted to the stock-recruit data appeared to lead to extrapolation of spawner abundances that could be associated with upper stock reference points. I compared the stock-recruit data shown in Figure 4-9 with those listed in Table 4-1 and found that the units of recruitment shown in Figure 4-9 and Table 4-1 differ by four orders of magnitude. On the vertical axis of Figure 4-9, the maximum recruitment values plotted are at approximately 140,000 million alfonsino recruits (copied below). In contrast in Table 4-1, maximum recruitment in Table 4-1 (top three lines of Table 4-1 copied below) is at 14 million recruits. It would appear that Figure 9-1 and Table 4-1 cannot both be correct. It is recommended that the stock assessment scientists check to make sure which set of recruit abundances, i.e., those in Figure 4-9 or those in Table 4-1 are correct and then also to check whether the estimated stock-recruit parameters were estimated using the correct recruitment abundances, and if not, then consider redoing the stock-recruit model fitting and parameter estimation.



Function type: HS Autocorrelation: 0, Optimization method L2, AICc: -0.38

Fig. 4-9. Relationship between spawning stock biomass and recruitment (stock-recruitment relationship)

Table 4-1. Results of stock analysis of the Splendid Alfonsino - Pacific Japan in the fishing grounds, off the Kanto coast to those around the Izu Islands and the southern offshore seamount area of Shikoku

Year	Catch (Thousand tons)	Stock biomass (Thousand tons)	Spawning stock biomass (Thousand tons)	Exploitation rate (%)	Number of recruits at age 2 (millions)	%SPR	F/Fmsy
1998	5.5	44.2	33.4	13	14	25.8	0.85
1999	6.0	45.2	33.5	13	13	23.8	0.94
2000	7.1	45.7	33.6	16	12	21.3	1.06

On page 6 of the report it is stated "Larger splendid alfonsino may also prey on juvenile splendid alfonsino (Ikeda 1980)." The presence of cannibalism in a fish stock suggests that a Ricker model with overcompensation may be more appropriate as the stock-recruit model than other stock recruit functions such as Beverton-Holt or hockey stick (Hilborn and Walters 1992[8]; Skoglund et al. 2022[9]).

Furthermore, a plot of ln(R/S) versus spawners showed an array of data in which a fitted linear model had a significant negative slope (Figure 3, middle panel, pvalue for slope coefficient =0.008). This is a characteristic of Ricker recruitment which implies compensation in recruitment rates at low stock sizes and potential overcompensation at higher stock sizes. Standardized residuals from this fit showed no outliers (Figure 3 bottom panel). However, there was a very strong positive autocorrelation in standardized residuals, with the autocorrelation coefficient estimated at 0.83 and a pvalue of 3E-08. This strong residual pattern when a simple Ricker model is fit, suggests potentially strong external forcing of recruitment with individual forcing events persisting for a number of years in a row. The strong residual pattern could alternatively be a byproduct of time series errors in estimates of recruitment strength. The strong time series pattern in recruitment anomalies indicates further that there could be considerable benefits to improving the empirical basis for estimating cohort strength, e.g., with the implementation of a new fishery independent survey targeting younger age

---

[8] Hilborn, R., & Walters, C. J. (1992). *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty*. Chapman and Hall, New York.

[9] Skoglund, S., Whitlock, R., Petersson, E., Palm, S., and Leonardsson, K. 2022. From spawner habitat selection to stock-recruitment: Implications for assessment. Ecology and Evolution 12(12):e9679. doi: 10.1002/ece3.9679.

classes of splendid alfonsino. Despite there being very large positive autocorrelation from a Ricker model fit, and it being likely that a similar strong autocorrelation would be found in a hockey stick model fit, it does not appear that the stock assessment tried to estimate or account for autocorrelation in stock-recruit residuals. Supplementary Table 6-1 indicates "No" under autocorrelation and it is said on page 12: "Recruitment residual autocorrelation was not accounted for in the analysis". Lag 1 recruitment residual autocorrelation, however, can be tested for in the stock-recruit model residuals. This can be done by regressing standardized residuals in year y with residuals in year y-1. The estimated slope coefficient gives the lag 1 autocorrelation coefficient. I tried to fit a hockey stick model to the splendid alfonsino stock-recruit data shown in Table 4-1 but found that the estimation was unstable and there appeared to be insufficient information in the data to fit a Hockey stick model. To facilitate model fitting, I fixed the slope coefficient for the increasing leg of the hockey stick model to a value equivalent to that shown in Supplementary Table 6-1 and estimated the maximum average recruitment value on the right side of this leg. Based on the HS model fit shown in Figure 3 below, I computed the standardized residuals for the HS model fit. Based on these standardized residuals, I found very large positive recruitment residual autocorrelation, i.e., at 0.85, pvalue =1E-07 (Figure 3 below). The assertion in the report, e.g., Supplementary Table 6-1 that there was no autocorrelation in recruitment residuals with the hockey stick model is incorrect and the estimation of lag 1 autocorrelation in hockey stick model residuals needs to be redone and the results obtained would need to be used in stock projections to evaluate alternative harvest management options.

One anomalous artifact associated with the use of the hockey stick model is the prediction of quasi-triangular average yield-at-age versus spawning stock biomass curves (Supplementary Figure 3-2). In actual fisheries, plots of average yield-at-age versus spawning stock abundance could never be expected to conform to a semi-triangular distribution as seen in Supp. Figure 3-2 where for ages 4 and older there is a sharp angled peak to the yield curve for each age. Yield versus SSB curves under a given harvest policy option could only be expected to have relatively smooth curvatures without a sharp angular peak as seen in Supp. Figure 3-2. Although a hockey stock model would appear to provide a reasonable fit to the stock-recruit data for this stock, it is recommended that a replacement stock-recruit function be adopted, so as to avoid the making of predictions of yield curves and population dynamics that fail to conform to credible population dynamics behaviours in the real world.

E) Evaluate whether the data are treated statistically correctly.

Both indices were fitted as relative abundance indices with an estimated constant of proportionality, q, which scaled the VPA predicted abundance to the observed annual abundance indices. Appropriately, the nine commercial catch per unit index time series that were based on the fishery catch and effort records were assumed to be linearly related to stock biomass.

In this stock assessment 95% confidence intervals were provided for quantities estimated from the VPA, i.e., for spawning stock biomass, stock biomass and recruitment (as shown in Supplementary Figure 2-4). A simple bootstrapping approach based on the tuning indices was appropriately applied.

F) Evaluate whether the stock assessment result obtained from the input data and methodology used is appropriate.

That the MSY (47,000 tons) reported in the main results table on page 3 of the report is estimated to be larger than the SSB (24,300) that on average produces MSY does not appear to be credible. This implies a harvest rate larger than 100% which is not credible. This appears to be a typo, as the recent commercial catches are over an order of magnitude smaller than this amount.

Presuming that the MSY estimate reported on page 3 of the report is a typo, the stock assessment results obtained are appropriate and suitable for fisheries management purposes. For this stock assessment, confidence intervals were appropriately computed and shown for the time series of stock biomass, SSB and recruitment (Supplementary Figure 2-4). This shows a moderate amount of uncertainty in stock biomass, SSB and recruitment estimates starting in about 2025, which could be expected since this is a relatively long-lived species and the catch-at-age data go up to age 15+ which means that cohort reconstruction will be less certain for the past decade of the time series where abundance information by cohort will be less complete. As mentioned above, there exist numerous rationale for why it is appropriate to compute confidence intervals associated with the results obtained out of stock assessment models and in most regions where stock assessments are carried out, it is common practice for confidence intervals or analogous Bayesian probability intervals to be computed and shown for estimated quantities. Confidence intervals provide a standardized representation of uncertainty in estimated quantities to indicate how variable the estimates of quantities of interest could be based on different components of sampling error variability associated with the sampling processes associated with collection and processing of data used in the stock assessment. Implementing the stock assessment modeling in a probabilistic framework would also allow computation of the probability that stock size is increasing or decreasing and the probability that stock size is above pre-defined stock reference points.

The Kobe plot provided appropriately shows the trajectory of stock and fishery status with respect to e.g. the limit and target stock reference points and the F/Fmsy reference point of 1, over the range of years where stock assessment results were obtained. The Kobe plot is shown only for the base case stock assessment model run and there is no representation of uncertainty in the stock status plot. It is recommended that when sensitivity analysis is conducted on different values for M, that two additional Kobe plots are shown in appendices. One of these would be based on the assumption of the largest value for M tried in the sensitivity analysis carried out on M, and the other would be based on the lowest value considered for M (Figure 4-6).

G) Evaluate the validity of methodology and result used for the future projection.

The methodology for the future projection uses the base case model estimates of numbers at age for the starting point for the future projection. No uncertainty is considered in the numbers at age starting point for future projections. This is unrealistic, since it is not possible for there to be complete certainty in the abundance at age in the most recent year of the stock assessment, as this is typically the most uncertain set of abundance at age in a VPA stock assessment, due to most recent year having the least amount of data being available for abundance estimation. Other VPA stock assessment methodologies applied elsewhere such as at ICCAT, regularly apply bootstrapping of the VPA to generate numerous plausible vectors of numbers at age and other associated quantities such as vulnerability at age in the final year of the stock assessment which then serve as starting points for the projections in which alternative policy options are evaluated. It is recommended that the approach to future projection be updated to incorporate uncertainty in the starting point of the future projections.

Uncertainty, e.g., in recruitment residuals is propagated in the projection of future years. Under a specific harvest policy option fishing mortality rates at age are applied in each future year projected. A single best estimate of fishery selectivity at age is applied based on the average Fs at age from F estimates obtained in recent years: "The current fishing pressure is set to the F value that gives the %SPR corresponding to fishing pressure from 2021 to 2023 in this assessment, using the same selectivity and biological parameters (average body weight, etc.) as those used in the calculation of proposed reference points" (p. 51). Figure 4-5 shows estimates of F for age 2-15+ that show that the selectivity at age, i.e., relative fishing mortality rate at age between age groups can vary to some extent between years. This is apparent for ages 5-8, and ages 9-15+, since the relative fishing mortality rate at age changes over years with the fishing mortality rate for some ages climbing above and below that for neighbouring age groups within the 1998-2023 time series. It is thus recommended that uncertainty associated with time varying selectivity be accounted for in the projections and also in the computation of MSY and SBMSY reference points.

The report did not test for autocorrelation in its fitted hockey stick model, as was mentioned on page 12 of the report. However, I fitted a simple hockey stick model and found very large positive recruitment residual autocorrelation, i.e., at 0.85, pvalue =1E-07 (Figure 3). If the autocorrelation in stock-recruit model residuals is as large as this, and autocorrelation in recruitment residuals were simulated in stock projections using this value, results of harvest control policy options could be quite different compared to when zero autocorrelation in recruitment residuals is assumed, as in the current stock assessment for splendid alfonsino. It is thus recommended that the stock assessment scientists test for autocorrelation in hockey stick recruitment model residuals and apply the estimate of the autocorrelation in stock projections done to evaluate outcomes of alternative harvest management options.

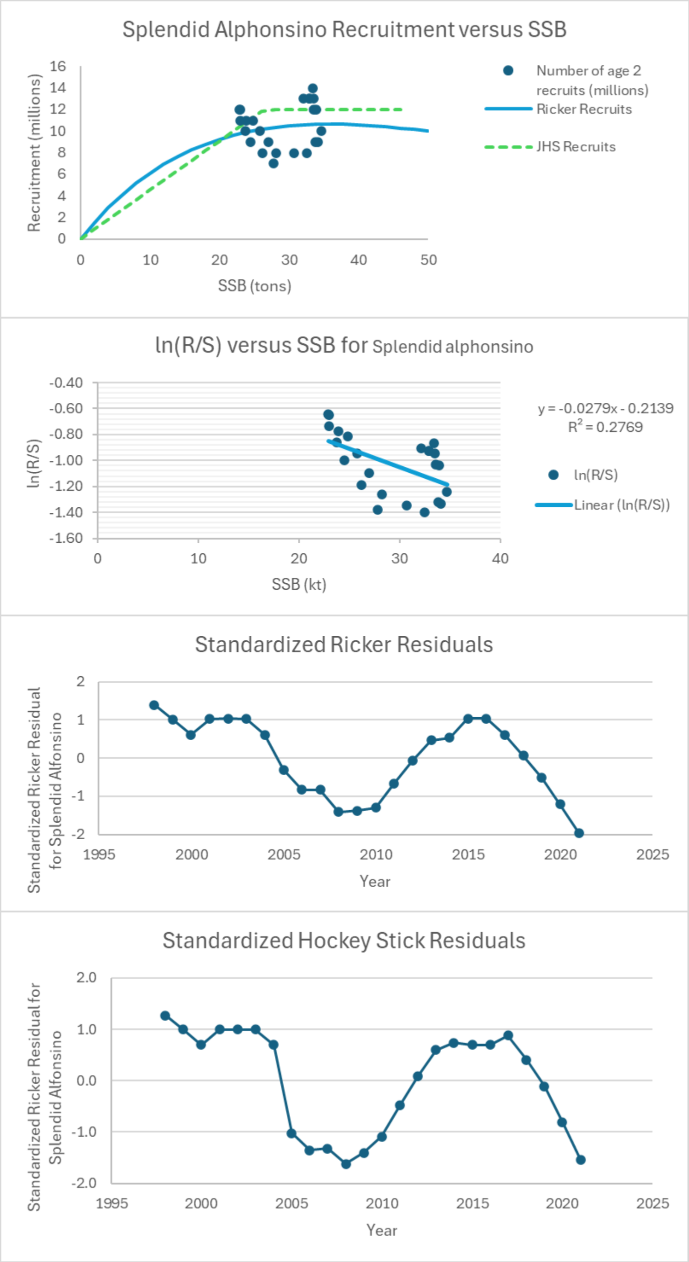


Figure 3. Top panel: Plots of the fits of the Ricker and hockey stick model to the stock-recruit data for splendid alfonsino using stock-recruit data from the splendid alfonsino stock assessment document. Middle panel: fit of a linear Ricker model of ln(R/S) versus SSB data for splendid alfonsino. Bottom two panels: The time series of standardized residuals from fit Ricker and hockey stock models for splendid alfonsino. The auto-correlation coefficient at lag 1 for the Ricker model was estimated at 0.84, with pvalue = 3E-08. That for the hockey stock model was 0.85 with a pvalue of 1E-07.

Stock Assessment #4: Japanese Anchovy – Seto Inland Sea

A) Determine whether the data used for stock assessment are adequate to understand the stock dynamics of the target species and represent the best scientific information available.

Catch records and abundance indices

A time series of fishery catch and pelagic egg survey records over four decades long is available for the stock assessment and could be expected to provide sufficient empirical information for understanding long-term stock dynamics, especially for such a short-lived species as anchovy which has an average "lifespan of approximately two years" (page 6) a maximum age of four years (Fadeev 2005[10]). On p. 7, it is stated: "The catch in weight of Japanese anchovy (the catch in weight of fish presumed to be "Japanese anchovy" according to commercial size category in the Statistics of Fishery and Aquaculture Production, without the inclusion of Whitebait (fish that are 1 to 2 months old))" is used in the stock assessment. The time series of catch biomass of "individuals in post-juvenile stage (aged 3 months and older)" goes back to 1981 and varies from about 65,000-95,000 tons 1981-1986, drops to about 17,000 tons in 1998, and then increases to fluctuate between about 25,000-40,000 tons up to 2023. This large variation in catch biomass over a long time series together with a long time series of egg production survey biomass estimates from 1981-2023 could be expected to be informative about stock dynamics. The reduction in catch biomass in the 1990s is accompanied by a long gradual increasing trend in the egg production index starting in the late 1990s.

The catch biomass of whitebait (fish aged 1-2 months) from 1981-2023 is shown in supplementary Figure 9-1 and Table 9-1. This shows large catch weights ranging from 16,000 tons in 1981 to 53,000 tons in the mid-1980s and values fluctuating between about 20,000 to 39,000 tons subsequently. The long-term average whitebait catch has been 30,679 tons, whereas the long-term average post-juvenile catch biomass has been 38,466 tons. The percentage of whitebait catch biomass to the total of whitebait and post-juvenile catch biomass has ranged between 20-63% with the average being 46% (Table 1 below). As the catch biomass of white bait is nearly as large as the catch biomass of post-juvenile anchovy over the time series, a large proportion of the stock is being removed by the fishery without this being accounted for in the stock assessment.

Typically, fishery stock assessments aim to assess the potential impacts on the fish stock of total fishery removals from the stock (Hilborn and Walters 1992). This stock assessment evaluated instead the impact of removals only from the fishery on post-juvenile stage anchovy and ignored the potential impacts on the stock of the fishery for whitebait. The stock assessment at present can thus only assess the potential fishing mortality rates coming from the fishery for post-juveniles but does not quantify the fishing mortality rates coming from the fishery for whitebait. With the relatively large catch biomasses of whitebait removed per year and the catch biomass of whitebait being nearly as large in magnitude as catch biomass for post-juveniles, i.e., on average 46% of the combined whitebait and post-juvenile catch biomass, a considerable source of fishing mortality on the stock is thus not being assessed. Also, given that the body masses of individual whitebait fish are much less than those for post-juvenile anchovy, the number of whitebait fish caught per ton will be much larger than that caught per ton of post-juvenile fish. Therefore, while the total biomass of the whitebait catch has been on average lower than that of the post-juvenile fish, the total number of whitebait caught per year could actually on average be closer to and possibly exceed in many years the total number of post-juvenile fish caught. Catch numbers of age 1 and 2-month-old fish in the post-juvenile fishery in several years even exceeds catches of ages 3-25 month old fish as can be seen in Supplementary Figure 2-1. However, the catch numbers shown in Figure 2-1 exclude the total catches of whitebait caught in the whitebait fishery (Table 9-1) and correspond to the catches in the fishery targeting post-juvenile fish as represented in Figure 3-2. The natural mortality rate, however, of whitebait could be expected to be much larger than that for post-juvenile fish, because rates of natural mortality have been found to be on average higher for younger smaller fish in a

---

[10] Fadeev, N.S., 2005. Guide to biology and fisheries of fishes of the North Pacific Ocean. Vladivostok, TINRO-Center. 366 p

cohort (Lorenzen et al. 2022[11]). A much larger natural mortality rate for whitebait than post-juvenile anchovy could tend to dampen the magnitude of estimates of fishing mortality rates for whitebait, if whitebait were to be included as a component of the stock assessment for Japanese anchovy.

In fisheries stock assessment, recruits are typically defined as those age-groups of fish that are the youngest of all age groups to be harvested (Hilborn and Walters 1992). As summarized by Google AI: "In fisheries science, recruits are young fish that have survived the initial, high-mortality stages of life (egg, larval, and juvenile) to reach a specific age, size, or developmental stage. They represent the replenishment of a fish population and are the individuals that become available for fishing." In the Japanese anchovy population, the youngest age groups to be harvested are one-month-old anchovies which are called whitebait and, technically, recruits in the Japanese anchovy population could thus be defined as one-month-old whitebait. However, as the catches used in the stock assessment consisted of post-juvenile fish as young as 3 months old, recruits were instead defined as 3-month-old post-juvenile fish. The VPA that was applied could be expected to accurately reconstruct the numbers at age of post-juvenile anchovy providing that it had accurate catch-at-age and natural mortality rate at age inputs, among other things. However, abundance estimates of age 3-month post juveniles from the VPA could be impacted by variable fishing mortality rates on whitebait. Interannual variability in fishing mortality rates on whitebait could be a significant component of variability impacting VPA estimates age 3-month abundance. In contrast, an assessment that included the whitebait fishery could be expected estimate the time series of recruitment to the whitebait fishery that was untainted by interannual variability in fishing mortality rates, and could also represent the availability of age 3-month post-juvenile fish to the post juvenile fishery associated with alternative harvest rates in the whitebait fishery. A stock-recruit analysis that assessed the relationship between age one-month recruits and spawning stock biomass could be expected to more accurately and precisely estimate the relationship between recruitment and spawning stock size, since it would not be impacted by variable fishing mortality rates over years on whitebait. It is thus recommended that the stock assessment scientists consider including annual catches of whitebait in their VPA stock assessment and redefining recruits as age one-month fish, rather than age three-month fish. For VPA to include whitebait catches, this would require an estimate of the instantaneous rate of natural mortality for age one to two-month whitebait and compilation of the time series of whitebait catch numbers from 1981-2023. This would then allow estimation of fishing mortality rate on whitebait in each historical year, and the pre-whitebait fishery and post-whitebait fishery population abundance of whitebait in each historical year.

There appears to be a moderate amount of information about cohort strength in the commercial fishery catch-at-age data for Japanese anchovy (Figure 3-2, 3-3). The time series of catch-at-age data with only two visible age groups in Figure 3-3 shows a moderate amount of interannual variation in cohort strength in age 0 and age 1 fish. Between some years, the catch number at age increases or decreases by more than double and less than half. This is a plausible indication of moderate interannual variability in cohort strength.

The estimated time series of recruitment seen in Figure 4-3 and Table 4-1 shows a relatively limited amount of interannual variability in recruitment overall with the average absolute annual change in recruitment being about 35%. There are also sets of years where there is relatively little change in recruitment, e.g., between 1993 and 1999, but then some larger increases or decreases between years that correspond to the large increases or decreases in catch number at age for age zero, especially. The relatively small amount of interannual variation in cohort strength for up to seven years in a row, e.g., starting in 1993 where interannual changes in recruitment are no larger than about 25% and an average absolute change in recruitment between years of 10%, is curious. This is in contrast to strong patterns of interannual variability seen in estimated time series of recruitment for other assessed fish stocks which have more intensive sampling programs for the age composition of annual catches (see Fig. 4 below). While it may be possible for there to be relatively little interannual variability in cohort strength over sequences of several years, we would expect to see more pronounced interannual variability in cohort strength in a small pelagic forage fish species such as anchovy whose annual recruitment rates could be impacted more by interannual changes in oceanographic and ecological conditions than demersal species at higher tropic levels (Schwartzlose et al. 1999[12]; Pikitch et al.

---

[11] Lorenzen, K., Camp, E.V., Garlock, T.M. 2022. Natural mortality and body size in fish populations. Fisheries Research 252: https://doi.org/10.1016/j.fishres.2022.106327.

[12] R.A. Schwartzlose, J. Alheit, A. Bakun, T.R. Baumgartner, R. Cloete, R.J.M. Crawford, W.J. Fletcher, Y. Green-

2012[13]).

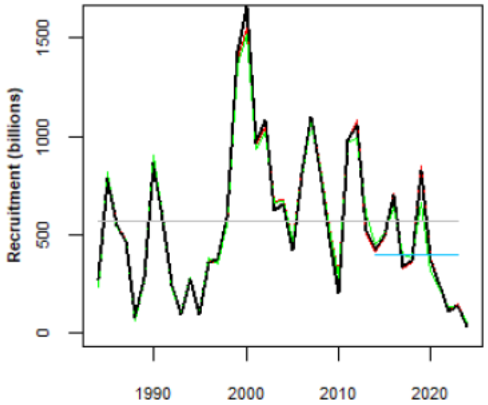


Figure 4. The time series of November recruitments for South African anchovy (de Moor 2025[14]). Note that this is just from one other anchovy population. It would be of interest to investigate the patterns of interannual variability and other properties of time series of recruitment for additional anchovy stocks where stock assessments have provided these.

Possible reasons for why actual interannual variability in recruitment in this fish stock is not detected with the stock assessment methodology that has been applied include the following. Firstly, the whitebait fisheries could be larger in years where whitebait abundance is larger, effectively targeting and fishing down larger cohorts before they recruit to the post-juvenile fishery. Secondly, incorrect age-assignment in the transformation of fishery length sample compositions into fishery catch-number-at-age using an age-length key could lead to smearing of fish abundance in larger cohorts into adjacent cohorts that have lower abundance. The abundance of large cohorts could thus be under-estimated and the abundance of adjacent smaller cohorts would be over-estimated. These two alternative processes could generate overly smooth time series of estimates of recruitment.

Commercial length composition data were transformed into age composition data using fishery length composition data for post-juvenile sized fish. "The catch in number at monthly age and by month followed the calculation method used in the stock assessments of this stock until FY 2021 (Kono and Takahashi 2022). In addition to the catch of "Japanese anchovy" according to commercial size category in the Statistics of Fishery and Aquaculture Production and the catch in weight by sea area in the Seto Inland Sea, the catch in weight by month, the body length composition by month, and body length-weight equations of the major fisheries cooperatives in each sea area were used in the calculation. The catch in number at monthly age and by month in each area was aggregated to obtain the total catch in number at monthly age and by month for this species in the entire Seto Inland Sea. The sea area was divided into five areas: eastern Seto Inland Sea, Hiuchi-nada, Aki-nada, Iyo-nada, and Suo-nada" (p.28-29). This approach assumes that the length-age and body length-weight relations (equations 1-4) are time invariant. In contrast, it is possible that growth and length at age could be time varying and using the fixed equations 1-4 could lead to actual length-age keys that vary over time and are different from those based on the study. Application of length-age keys that are each assumed to be stationary over years to commercial length composition records when the length-age relationships are time varying could

---

Ruiz, E. Hagen, T. Kawasaki, D. Lluch-Belda, S.E. Lluch-Cota, A.D. MacCall, Y. Matsuura, M.O. Nevarez-Martinez, R.H. Parrish, C. Roy, R. Serra, K.V. Shust, M.N. Ward, J.Z. Zuzunaga. 1999. Worldwide large-scale fluctuations of sardine and anchovy populations. S. Afr. J. Mar. Sci., 21 (1999), pp. 289-347, [10.2989/025776199784125962](https://doi.org/10.2989/025776199784125962)

[13] E. Pikitch, P.D. Boersma, I. Boyd, D. Conover, P. Cury, T. Essington, S. Heppell, E. Houde, M. Mangel, D. Pauly. 2012. Little Fish, Big Impact: Managing a Crucial Link in Ocean Food Webs. Lenfest Ocean Program, Washington, DC (2012), p. 108

[14] De Moor, C.L. Stock recruitment relationships for South African anchovy. IWS/2025/Sardine/2025/P2

lead to generation of misrepresentations of age-composition records for use in the VPA stock assessment. If available data permit, it may be of interest to test whether the null hypothesis that growth and age length keys are not time varying. If this hypothesis were found to be inconsistent with available data, it may then be appropriate to formulate an approach that allowed for the length-age keys to be time varying. In addition, as is common with generation of age composition records it is likely that there could be smearing in the age composition records generated from the commercial length composition records. This could result in stronger age classes being under-represented in the generated age-composition data and neighbouring age classes being over-represented. It would be appropriate to develop a simulation-estimation methodology to quantify the expected estimation performance of the current approach to deriving age composition data from length composition data using a fixed length-age key.

B) Discuss whether the biological parameters used for stock assessment are appropriate.

The instantaneous rate of natural mortality was assumed to be 2.1 yr^{-1} for fish age 0 to 1 and 2.0 yr^{-1} for fish age 2 and older per year. If the Hoenig (1983) method is applied to predict M, based on maximum observed age, and the maximum age was presumed to be 4 years, Hoenig's predicted M would be 1.06 yr^{-1}. In contrast, the maximum observed age would need to be 2 years, for Hoenig's method to predict an M of 2.0 yr^{-1}. If M was 2.0 yr^{-1}, for age 2, we would expect to see approximately 1.8% of an unfished cohort surviving. Given the level of fishing that has gone on since the 1980s, we would expect to see a smaller percentage of age 2 fish, if M was 2 yr^{-1}. For example, with F at 1.0 yr^{-1} and M at 2.0 yr^{-1}, we would still expect to see approximately 0.25% of the cohort at age 2 years old. This latter percentage of fish remaining at age 2 years under a total Z of 3.0, is so small that it might not be visible in Fig. 3-3. However, very small contributions of age 2+ fish are visible in some years for catch in weight at age (Fig. 3-2). Thus, it appears that Hoenig-based predictions of M would not suggest considerably different values than those values for M assumed in this stock assessment.

C) Discuss whether the basic biological information such as distribution, migration pattern, and population are appropriate.

Figure 2-1 mapped the distribution of fishing and spawning grounds for the stock with the spawning ground for the stock ranging from the southern coast of Honshu west of Osaka to several hundreds of kilometers offshore and extending southwest to the southern coast of Kyushu (Figure 2-1). Curiously, the mapped survey points for the egg and larvae surveys are mainly in the straits between Shikoku and Honshu and Kyushu and Honshu (Figure 7-1) and overlap considerably with the fishing grounds of the Seto inland sea. The egg and larval survey area correspond only to a relatively small fraction of the total spawning area for the stock. It is presumed that though the egg and larvae surveys correspond to a relatively small fraction of the total spawning ground for Japanese anchovy that they would still provide an unbiased estimate of the time series of spawning stock biomass for Japanese anchovy. If the proportion of the stock that spawns in the area surveyed for eggs and larvae changes markedly between years or changes systematically over years, then the index could either be an imprecise index of stock abundance or may give misleading information on trends in total spawning stock abundance.

D) Evaluate whether the stock assessment methodology is based on the most appropriate available study and performed analytically.

The tuned VPA approach appears to be suitable for the available stock assessment data, i.e., commercial fishery catch-at-age data (derived from annual commercial length composition records and length-age keys) and a stock index based on an egg and larval survey. However, as pointed out above, a large component of harvest from this fish population, i.e., the whitebait fishery is not accounted for in the VPA model applied. As mentioned above, a more credible result could have been obtained if the VPA was extended to include the annual removals of age 1 and 2 month old whitebait in the whitebait fishery. This could potentially have improved estimates of cohort strength and provided a time series of age 1 month old recruitment that was not impacted by interannual variation in fishing mortality rates on age 1 and 2 month old fish caught in the whitebait fishery.

The time step in the VPA is one year per age group, such that all fish in their first year are lumped together as age zero fish, all fish in their second year are lumped together as age 1 fish, and all fish of ages 2 and older are lumped together as age 2+ fish. However, the determination of catch-age composition firstly compiles numbers of fish at age by month in each month of the year. It also uses a matrix of natural mortality by age in months old (Supplementary Table 2-4). Following this, total catch at age by age in years for ages 0, 1 and 2+ years old fish are compiled. The estimates of natural mortality rate at age increase as age in months decreases to the youngest age considered. No estimates of natural mortality rate at age, however, are provided for ages 1 and 2-month-old fish. Given that the starting point in formulation of catch-at age is by month, and natural mortality rates increase as age gets younger in months, it would appear that information on fishing mortality rate and abundance may be lost by wrapping the monthly age composition records into annual time steps. One possible alternative VPA approach that might retain information in the monthly catch-age composition data more precisely would be to do a VPA using the M-at-age and age compositions records at monthly time steps, but still tuning the VPA model to the total abundance of mature fish by year. If this approach was too detailed, other slightly more aggregated VPA approaches could be considered. For example, the VPA could be structured based on two-month time steps in term of time steps and age steps, so that the whitebait fishery catch at age could also be incorporated into the VPA model.

The diagnostics for estimation performance coming from the retrospective analysis show some retrospective patterns in SSB, stock biomass, and fishing mortality rates (Supplementary Figure 2-3). This shows a potential to either considerably over-estimate or under-estimate biomass, SSB, and recruitment strength and under-estimate fishing mortality rate prior to all catch at age data becoming complete for a given cohort. The retrospective patterns indicate that estimates of stock biomass, SSB, recruitment and fishing mortality rate in a VPA assessment of this stock for up to about the most recent two years of estimates can be highly uncertain. The retrospective patterns could potentially arise if the egg and larval index is incorrectly tracking trends in stock abundance. This could potentially occur if there exist time trends in the proportion of spawning stock that spawn within the area surveyed by the egg and larval survey. The retrospective patterns such as those observed could also occur if there exist recent time trends in fishery selectivity at size or age that are not accounted for in the VPA, e.g., a time trend in increasing selectivity for smaller-sized fish in the past decade.

The fit of the VPA model to the egg and larvae survey index shows considerable imprecision in the residuals from the model fit (Supplementary Figure 2-4). Time trends in residuals are also apparent in Supplementary Figures 2-5 and 2-6. Although the authors estimated a small positive lag 1 autocorrelation coefficient in the residuals which was not statistically significant (p.32) the undulating pattern in residuals remains quite pronounced (Figure 2-6). In the 1980s, the residuals are mostly negative. From 1992 to about 2003, the residuals are mostly positive, from about 2009-2015, mostly negative, and then from 2018-2023, mostly positive. The time series pattern in residuals of the VPA fit to the egg and larval index suggests that the egg and larval index could be deviating systematically from stock biomass trends over years. As mentioned above, this could result from the proportion of the spawning stock inside the egg and larval survey area varying systematically over years. It could also come from systematic variation in growth and length-weight parameters for the stock over years that are not accounted for in the formulation of catch-age data from fishery length composition information.

An important component of the stock assessment is the choice of a stock-recruit function. As with the other three assessed stocks, a so-called “hockey-stick” stock-recruit model had been chosen. One of the justifications for this choice given during the review meeting, was that there was concern that other stock recruit models fitted to the stock-recruit data appeared to lead to extrapolation of spawner abundances that could be associated with upper stock reference points.

Furthermore, a plot of ln(R/S) versus spawners showed an array of data in which a fitted linear model had a significant negative slope (Figure 5, middle panel, pvalue for slope coefficient =5E-05). This is a characteristic of Ricker recruitment which implies compensation in recruitment rates at low stock sizes and potential overcompensation at higher stock sizes. Standardized residuals from this fit showed no outliers

(Figure 5 bottom panel). There no significant autocorrelation in standardized residuals for the fitted Ricker and Hockey stock models, e.g., for the Ricker model, the autocorrelation coefficient estimated at 0.09 and a pvalue of 0.55. For the hockey stick model this was 0.07, pvalue = 0.65.

Although a hockey stock model would appear to provide a reasonable fit to the stock-recruit data for this stock (Figure 4-9, and Figure 4 below), it is recommended that a more realistic stock-recruit function be adopted such as the Ricker model.

E) Evaluate whether the data are treated statistically correctly.

The fishery independent egg and larval survey index of abundance was fitted as a relative abundance index with an estimated constant of proportionality, q, which scaled the VPA predicted abundance to the observed annual abundance indices.

In this stock assessment no confidence intervals were provided for quantities estimated from the VPA. As noted for the round herring and shotted halibut stock assessments, it is recommended that confidence intervals be computed for spawning stock biomass, stock biomass and recruitment. A simple bootstrapping approach could be applied based on the tuning index.

F) Evaluate whether the stock assessment result obtained from the input data and methodology used is appropriate.

As noted above, the VPA stock assessment left out a substantial amount of fishery extractions from the assessed exploited fish stock. Although a substantial whitebait fishery has been in place historically catching on average about 46% of the total stock biomass removed annually based on records of whitebait and post-juvenile fishery biomass caught from 1981-2023 (Table 4-1, Supplementary Table 9-1), the current stock assessment considered only the fishery for post juvenile anchovy which were 3 months old and older. As mentioned above, the proper technical definition of a recruit from the fish population assessed should be an age 1 fish, not an age 3 fish, since ages 1 and 2 month old fish could be expected to experience significant fishing mortality rates in the whitebait fishery and the records suggest that there could be substantial interannual variability in whitebait fishery removals (Supplementary Table 9-1). The stock assessment's estimated time series of recruitment this is not correct, because it used an incorrect definition of recruits, i.e., age 3-month fish, and not age 1-month fish. Also, in computing the maximum sustainable yield associated with the Japanese anchovy population, it considered only fishery removals of post-juvenile fish when in fact, fishery removals also occur for younger age 1 and 2-month-old fish in the whitebait fishery. The maximum sustainable yield computation would thus need to find optimums considering a target fishing mortality rate on the whitebait component of the population in addition to a fishing mortality rate for the post-juvenile component of the population. Instead, the current stock assessment ignored the whitebait fishery in computing an MSY and MSY reference points for the post-juvenile component of the Japanese anchovy population.

The stock assessment results obtained are appropriate and suitable for fisheries management purposes only for the post-juvenile component of the fishery. For this stock assessment, however, no confidence intervals were computed and shown as they should have been for the time series of stock biomass, SSB, fishing mortality rate, and recruitment. As mentioned above, there exist numerous rationale for why it is appropriate to compute confidence intervals associated with the results obtained out of stock assessment models and in most regions where stock assessments are carried out, it is common practice for confidence intervals or analogous Bayesian probability intervals to be computed and shown for estimated quantities. Confidence intervals provide a standardized representation of uncertainty in estimated quantities to indicate how variable the estimates of quantities of interest could be based on different components of sampling error variability associated with the sampling processes associated with collection and processing of data used in the stock assessment. Implementing the stock assessment modeling in a probabilistic framework would also allow computation of the probability that stock size is increasing or decreasing and the probability that stock size is above pre-defined stock reference points.

The Kobe plot provided appropriately shows the trajectory of stock and fishery status with respect to e.g. the limit and target stock reference points and the F/Fmsy reference point of 1, over the range of years where stock assessment results were obtained. The Kobe plot is shown only for the base case stock assessment model run and there is no representation of uncertainty in the stock status plot. It is recommended that when sensitivity analysis is conducted on different values for M, that two additional Kobe plots are shown in appendices. One of these would be based on the assumption of the largest value for M tried in the sensitivity analysis carried out on M, and the other would be based on the lowest value considered for M (Figure 4-6).

G) Evaluate the validity of methodology and result used for the future projection.

The methodology for the future projection uses the base case model estimates of numbers at age for the starting point for the future projection. No uncertainty is considered in the numbers at age starting point for future projections. This is unrealistic, since it is not possible for there to be complete certainty in the abundance at age in the most recent year of the stock assessment, as this is typically the most uncertain set of abundance at age in a VPA stock assessment, due to most recent year having the least amount of data being available for abundance estimation. Other VPA stock assessment methodologies applied elsewhere such as at ICCAT, regularly apply bootstrapping of the VPA to generate numerous plausible vectors of numbers at age and other associated quantities such as vulnerability at age in the final year of the stock assessment which then serve as starting points for the projections in which alternative policy options are evaluated. It is recommended that the approach to future projection be updated to incorporate uncertainty in the starting point of the future projections.

Uncertainty, e.g., in recruitment residuals is propagated in the projection of future years. Under a specific harvest policy option fishing mortality rates at age are applied in each future year projected. A single best estimate of fishery selectivity at age is applied based on the average Fs at age from F estimates obtained in recent years: “F is assumed to correspond to an F value that results in a 37,000-ton catch, based on the average of F from 2019 to 2023, with selectivity the same as the average F value from 2018 to 2022 and biological parameters under the same conditions as those used at the “Research Institute Meeting”.” (p. 27). Figure 4-5 shows estimates of F for ages 0 and 1 and 2+ that show that the selectivity at age, i.e., relative fishing mortality rate at age between age groups can vary markedly between years, with the relative magnitudes of fishing mortality rates switching considerably over years such that in some years age 0 has higher F and in other years ages 1 and 2+ has the highest estimated F. It is thus recommended that uncertainty associated with time varying selectivity be accounted for in the projections and also in the computation of MSY and SBMSY reference points.

Table 1. Catch biomass of whitebait (Supplementary Table 9-1), and post-juvenile anchovy (Table 3-1).

Year	Whitebait	Catch of Japanese Anchovy[15]	Catch considered in stock assessment	%whitebait[16]
1981	16,319	67,532	64,815	20%
1982	26,264	76,088	71,150	27%
1983	45,012	80,945	76,297	37%
1984	33,443	75,012	66,617	33%
1985	50,224	99,729	93,295	35%
1986	53,385	92,887	83,932	39%
1987	38,042	36,004	33,762	53%
1988	46,157	57,595	52,729	47%
1989	45,071	40,279	28,085	62%
1990	34,426	31,548	23,400	60%
1991	36,223	42,281	41,084	47%
1992	27,700	27,287	24,125	53%
1993	27,327	24,929	22,474	55%
1994	24,577	20,511	18,331	57%
1995	24,983	22,648	18,283	58%
1996	25,557	19,802	19,246	57%
1997	22,615	20,042	15,625	59%
1998	21,446	16,480	15,176	59%
1999	37,123	22,696	22,139	63%
2000	34,780	36,523	34,831	50%
2001	26,413	35,790	31,388	46%
2002	35,299	43,066	36,634	49%
2003	37,813	33,827	27,823	58%
2004	26,239	32,917	25,705	51%
2005	20,598	35,895	33,099	38%
2006	18,410	41,883	35,298	34%
2007	26,340	36,555	32,707	45%
2008	27,861	39,316	35,059	44%
2009	27,186	44,667	39,749	41%
2010	36,555	36,678	34,116	52%
2011	19,782	37,993	37,221	35%
2012	33,048	39,956	37,336	47%
2013	29,932	43,564	41,974	42%
2014	26,291	43,740	41,736	39%
2015	26,715	41,217	39,450	40%

---

[15] he catch in weight of fish presumed to be "Japanese anchovy" according to commercial size category in the Statistics of Fishery and Aquaculture Production, without inclusion of Whitebait (fish that are 1 to 2 months old). The values for 2023 are provisional.

[16] Percentage white bait computed as the percentage biomass of whitebait to the total of whitebait and post-juvenile whitebait used in the stock assessment.

2016	31,324	42,727	40,162	44%
2017	25,222	41,210	40,557	38%
2018	21,975	37,963	35,476	38%
2019	29,224	31,501	25,821	53%
2020	27,459	40,760	39,491	41%
2021	39,005	35,955	32,937	54%
2022	25,453	45,100	42,469	37%
2023	30,388	44,806	42,423	42%
Average	30,679	42,277	38,466	46%
Minimum	16,319	16,480	15,176	20%
Maximum	53,385	99,729	93,295	63%

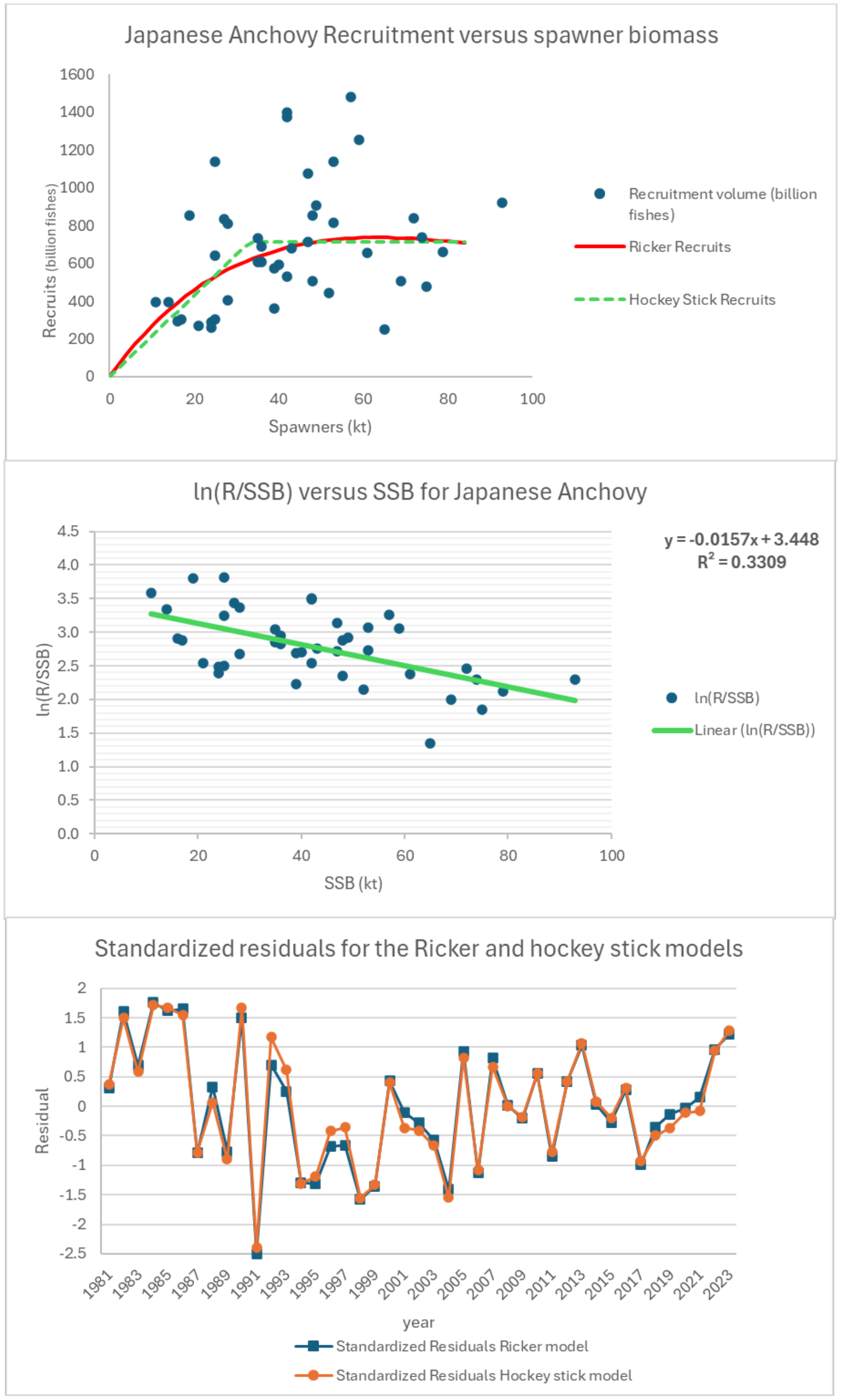


Figure 5. Top panel: Plots of the fits of the Ricker and hockey stick model to the stock-recruit data for Japanese anchovy using stock-recruit data from the stock assessment document (Table 4-1). Middle panel: fit of a linear Ricker model of ln(R/S) versus SSB data for Japanese anchovy. Bottom two panels: The time series of standardized residuals from fit Ricker and hockey stock models for Japanese anchovy. The auto-correlation coefficient at lag 1 for both models was small and not statistically significant. There no significant autocorrelation in standardized residuals for the fitted Ricker and Hockey stock models, e.g., for the Ricker model, the autocorrelation coefficient estimated at 0.09 and a pvalue of 0.55. For the hockey stick model this was 0.07, pvalue = 0.65.

Other General Recommendations

1. On business cards please ask scientists, managers and administrators to include a small headshot so the face of the person can be seen and it will be easier to remember people’s names.
2. In stock assessment review meetings, please have participants wear a name badge, in addition to having their name plate put in front of their seat at the table, also to make it easier to learn people’s names.