National Marine
Fisheries Service
NOAA

Fishery Bulletin
established in 1881

Spencer F. Baird
First U.S. Commissioner
of Fisheries and founder
of *Fishery Bulletin*

Randomization method for evaluating the reliability of a recruitment index in an integrated assessment model

Huihua Lee (contact author)[1]
Kevin R. Piner (retired)[1]
Mark N. Maunder[2,3]
Hiromu Fukuda[4]

Email address for contact author: huihua.lee@noaa.gov

[1] Fisheries Resources Division
Southwest Fisheries Science Center
National Marine Fisheries Service, NOAA
8901 La Jolla Shores Drive
La Jolla, California 92037-1509

[2] Inter-American Tropical Tuna Commission
8901 La Jolla Shores Drive
La Jolla, California 92037-1508

[3] Center for the Advancement of Population Assessment Methodology
8901 La Jolla Shores Drive
La Jolla, California 92037-1508

[4] Fisheries Resources Institute
Japan Fisheries Research and Education Agency
2-12-4 Fukuura
Yokohama, 236-8648, Kanagawa, Japan

Abstract—The aim of stock assessment modeling is to evaluate current and future stock statuses under various management scenarios. For these assessments, particularly short-term projections, researchers rely heavily on recruitment estimates. However, accurately estimating recruitment levels, especially for the most recent years, is challenging because reliable information is limited. Indices of relative recruitment can provide valuable insights for stock assessment models in such cases. Herein, a method for assessing the reliability of a recruitment index with diagnostic and randomization tests in an age-structured production model (ASPM) is presented. The proposed method can be used to evaluate whether the process variability implied by a recruitment index for the spawner–recruit relationship enhances the connection between observed catches, indices of adult abundance, and the assessment model's production function. We applied this approach to Pacific bluefin tuna (*Thunnus orientalis*) as an illustrative example. Results indicate that the ASPM, with recruitment fluctuations matching the recruitment index for the Sea of Japan (ASPM-R-FIX), improved the fits to recent adult abundance indices over those of the standard ASPM models. The results from statistical analysis further provide strong evidence that extreme recruitment fluctuations in the ASPM-R-FIX enhanced predictions for recent indices for adults. This analysis revealed a strong link between recruitment variability and adult abundance, highlighting the importance of adding a consistent recruitment index to an assessment model to improve management decisions.

Manuscript submitted 1 November 2024.
Manuscript accepted 13 May 2025.
Fish. Bull. 123:179–189 (2025).
Online publication date: 2 June 2025.
doi: 10.7755/FB.123.3.4

The primary purposes of stock assessment models are to evaluate the current status of exploited fish populations and to make predictions about the future status of stocks under different management scenarios (Hilborn and Walters, 1992). The current status of a stock is typically measured by using the abundance in the terminal year relative to a hypothetical unfished state. Both current and historical abundances are affected by past recruitment levels along with natural mortality, individual growth, and catch levels. Recruitment refers to the number of young-of-the-year fish entering the population. The future status of a stock is dependent on future catches and potential future recruitment levels. Forecasting of future recruitments is often informed by the distribution of estimates of historical recruitment over a time period assumed to best reflect the future.

Despite the important role recruitment plays in stock assessment, there remain considerable challenges in estimating the magnitude of each recruitment event (year, quarter, etc.). Integrated assessment models are used to estimate recruitment events by recreating population dynamics (estimate the number of fish by age, sex, etc.) through fitting observed data (e.g., age and size composition of catches and indices of relative abundance) and estimating observation (selectivity, catchability, etc.) and system (recruitment, growth, etc.) processes (Maunder and Punt, 2013). In integrated assessments, recruitment levels can be thought of as the process variability of the spawner–recruit (SR) relationship and are often estimated as deviations from the expectation of that relationship based on an assumed statistical distribution (Maunder and Punt, 2013). When information on recruitment is limited, the model's estimate of recruitment will shrink to the expectation of the SR process.

Complicating the estimation of recruitment levels, observations of a cohort may not begin for several years after hatching, as fish of those ages may not be targeted by fisheries or be observed in surveys. This observation lag becomes especially problematic for the terminal years of the assessment model, for which information on recent cohort strength may be limited or entirely missing because recent cohorts have only partly, or not yet, entered into the fishery. Further complicating the situation is that poorly informed recruitment parameters will often be estimated to account for the misfit to data components of the model due to model misspecification (e.g., temporal variation in selectivity) or sampling variation biasing the recruitment estimates themselves.

Indices of the relative abundance of recruits (i.e., age-0 fish) from fishery-dependent or survey observations can improve assessment results both by enhancing the reliability of recruitment estimates from the assessment model and by providing information on recruitment levels prior to cohorts of recruits entering the fisheries. However, similar to the issues in estimating recruitments as mentioned in the previous paragraph, the relatively free estimation of recruitment levels and their connection to all other data and model processes can make it problematic to determine if the recruitment index is providing reliable information through standard model diagnostics such as residuals (e.g., alternative model run that included an alternative index of recruit abundance in a recent stock assessment for Pacific bluefin tuna; ISC, 2022). If recruitment parameters are poorly informed by other data, the recruitment estimates may allow good model fits to an unrepresentative recruitment index as well as to the catch and composition data and indices of adult abundance included in the likelihood function. This situation could affect the assessment estimates of a stock's status and its future prospects.

Herein, we describe a simple randomization procedure for evaluating if a recruitment index (for age-0 fish) provides reliable information on recruitment levels. The method relies on the relationship between catches, trends in adult abundance (as measured by indices of adult relative abundance), and the stock's production function (i.e., the system processes of growth, natural mortality, and SR relationship). The randomization procedure is used to evaluate if the implied process variability in the SR relationship from a recruitment index improves the production function's prediction of changes in abundance given the observed catches. We provide an illustrative example of the use of this method through application of it to the data and full dynamics model used for Pacific bluefin tuna (*Thunnus orientalis*) because Pacific bluefin tuna have been heavily exploited across all age groups and the majority of the catches of this species, in terms of numbers, consists of age-0 fish (Suppl. Material).

Materials and methods

Overview

The procedures are outlined in 4 sections. The first section provides a brief description of the study species we used to demonstrate the randomization method, including its biology and the full dynamics model used for it (for details, see Supplementary Material). The next section is about the conversion of the full model into a simplified version used as the basis for the evaluation of the reliability of the recruitment index (an age-structured production model [ASPM]). The third and fourth sections provide explanations of how we used the ASPM for the evaluation and how we quantified the strength of the improvement in model estimation due to inclusion of a recruitment index. The evaluation done with the ASPM is summarized in a flowchart (Fig. 1).

Study model

The current and future statuses for stocks of Pacific bluefin tuna is determined by using an age-structured, single-sex, catch-at-length integrated model fit to data sources compiled quarterly (by calendar season). In this model, the full dynamics of a stock are estimated (e.g., annual recruitment levels) along with all observation model processes (e.g., selectivity and catchability) linking to fleets defined by gear and country of origin. The data components include catches and length composition by fleet, as well as several indices of adult relative abundance. The indices of adult abundance (Fig. 2A) are standardized catch-per-unit-of-effort (CPUE) time series from 3 longline fleets that target large and mature Pacific bluefin tuna: an early time series for the fleet operating out of Japan in 1974–1992, a late time series for the fleet operating out of Japan in 1993–2016, and a time series for the fleet operating out of Taiwan in 2000–2016. The fleets used in this model to remove catch have annual time-varying selectivity, and the fleets associated with CPUE time series have time-invariant selectivity. See Supplementary Material for a further description of this full model developed by using Stock Synthesis, vers. 3.24f (Methot and Wetzel, 2013).

Age-structured production model

The basis of our evaluation method is an ASPM (Maunder and Piner, 2015). This ASPM is an age-structured integrated version of a surplus production model (Schaefer, 1954; Pella and Tomlinson, 1969). In the ASPM, the observation and system processes are separate, as they are in the full model, but predetermined values are specified for most of those processes (e.g., natural mortality, growth, SR relationship, and selectivity). The ASPM is similar to a surplus production model in that only parameters associated with global scaling (unfished abundance by way of unfished recruitment) and relative scaling (depletion levels and catchability) are estimated, by using only catch data and indices of abundance.

The ASPM for the study model was developed by simplifying the full model. In the ASPM, the fleet structure (number of fleets) of the full model is retained, with each fleet defined by country of origin, gear, and areas of operation. We made 3 main changes from the full model. First, we removed size composition data. Therefore, the ASPM likelihood function includes only catch by fleet and the CPUE indices based on the early and late time series for longline fisheries of Japan and on the time series

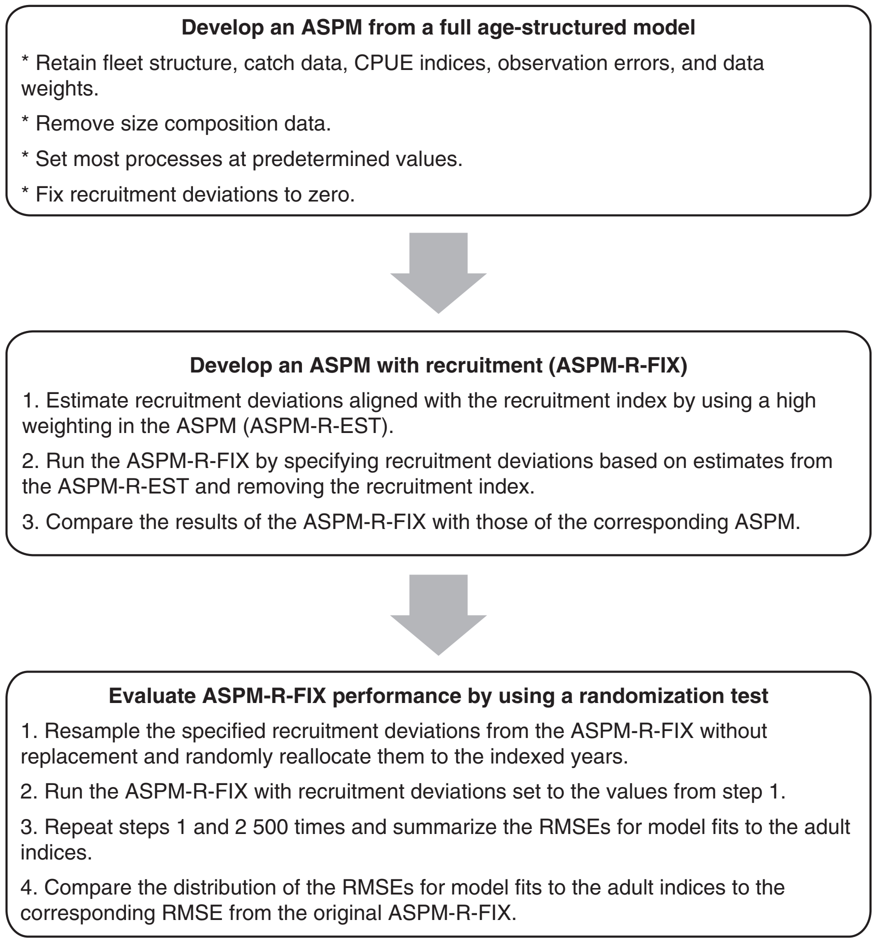

Figure 1

Flowchart summarizing the steps taken to develop an age-structured production model (ASPM) and to evaluate model performance by using a randomization test. The randomization method was used to examine the reliability of a recruitment index and to determine if inclusion of such an index improved model predictions. CPUE=catch per unit of effort; RMSE=root mean square error.

for longline fisheries of Taiwan. The same starting data weights (observation errors) given to each fleet's catch and CPUE index in the full model are retained in the ASPM. For each CPUE index, the observation errors are based on fitting a simple smoother to CPUE observations outside the model and estimating the residual variance for both the full model and the ASPM (for details about the full model, see ISC, 2018). However, unlike for the full model, reweighting (Francis, 2011, 2017) among the types of data was not needed for the ASPM because of the exclusion of composition data.

Second, because composition data were not included in the ASPM, selectivity parameters for each fleet were specified (set and not estimated) at values taken from the full model. We developed the ASPM using 2 different selectivity assumptions based on how selectivity is estimated in the full model: 1) the selectivity of fleets varies over time (referred to as *time varying*), and 2) the selectivity of fleets is constant over time (referred to as *time invariant*). Option 1 (time-varying selectivity) is the parameterization of the applied assessment used for management because it better approximates catch at age by considering changing fishery behaviors and fish migrations. Option 2 (time-invariant selectivity) more fully eliminates the influence of the quarterly observations of composition data on the full dynamics through selectivity. However, for brevity, we present figures and more in-depth analyses only for option 1, and results from the use of option 2 are included to analyze model sensitivity in relation to that base case.

Third, by setting recruitment deviations to zero, we forced recruitment to be the deterministic expectation of the SR relationship defined by Beverton and Holt (1957).

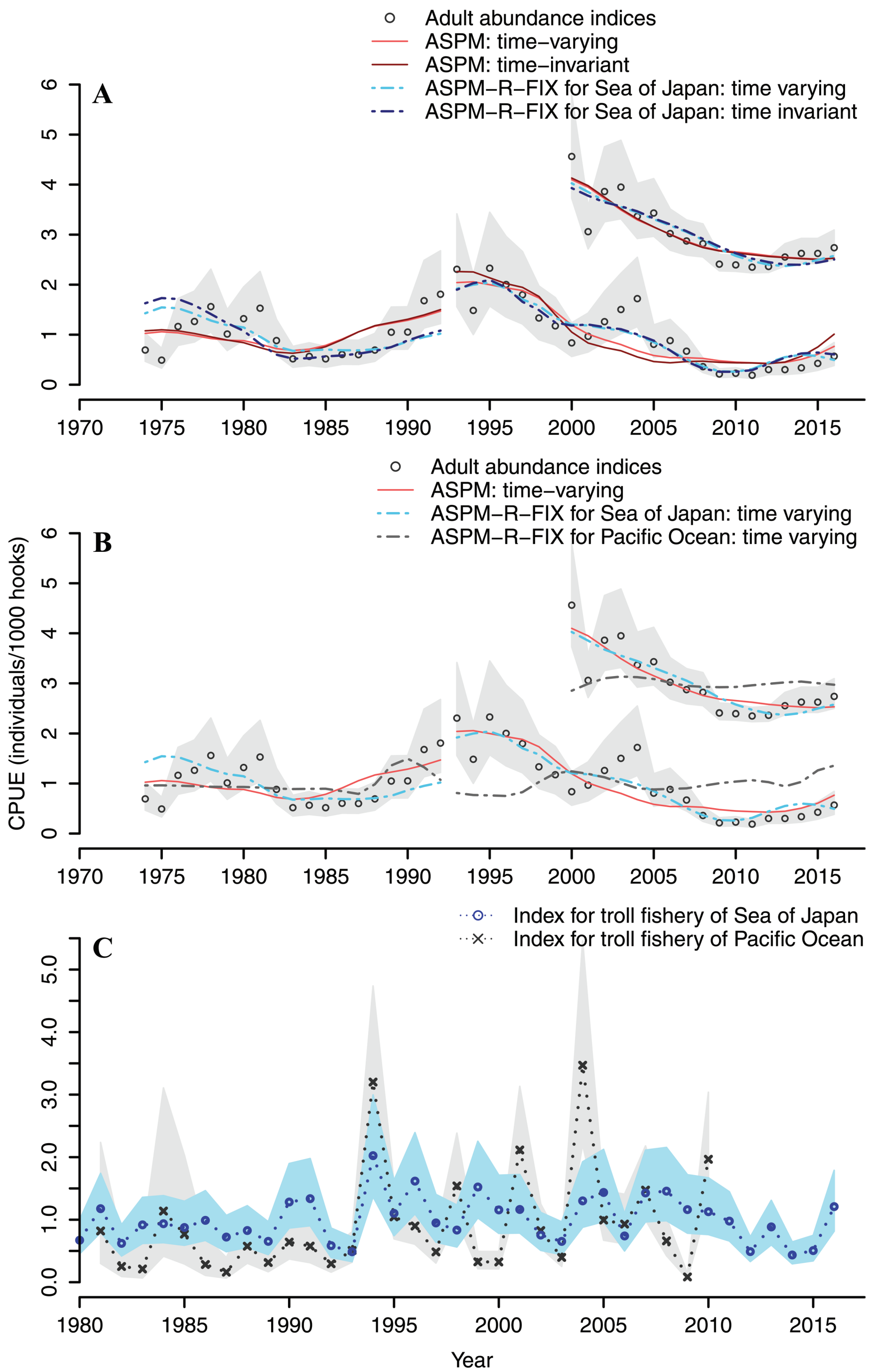

Figure 2

(**A** and **B**) Comparison of observed (open circles) and predicted (lines) indices of abundance of adult Pacific bluefin tuna (*Thunnus orientalis*) based on catch-per-unit-of-effort (CPUE) data from longline fleets in the Sea of Japan or the Pacific Ocean from 1974 through 2016 and (**C**) observed recruitment indices of age-0 Pacific bluefin tuna based on CPUE data from troll fisheries from 1980 through 2016. A randomization method, in versions of the standard age-structured production model (ASPM) and of the ASPM with temporal recruitment fluctuations fixed to match a recruitment index (ASPM-R-FIX), was used to evaluate the reliability of a recruitment index and to determine if inclusion of such an index improved model predictions. In the models, fleet selectivity was assumed to vary or remain constant over time. In panel A, lines represent estimates from either time-varying or time-invariant models for only the Sea of Japan, and in panel B, lines indicate estimates from only time-varying models for either the Sea of Japan or the Pacific Ocean. The gray areas in all panels (and the blue area in panel C) represent the 95% confidence intervals around observed values.

In the ASPM, the estimated parameters are unfished recruitment, catchability, and those representing the initial conditions (recruitment offset for the first year of the model and 2 parameters of equilibrium fishing mortality). The unfished recruitment parameter in the ASPM corresponds to the carrying capacity in a surplus production model. Estimating parameters representing the initial conditions allows the model to begin with the dynamic period at a level of depletion consistent with the catch and CPUE data.

Comparison of age-structured production model variants

Traditionally, the ASPM does not include estimates of temporal variability in recruitment. However, for short-lived species or environmentally sensitive species, simultaneous increases in abundance, which is reflected in the index of adult abundance and in catch data due to high recruitment, cannot be represented in the ASPM. Consequently, with the ASPM, estimated abundance is often much higher than that estimated with the full model, but the ASPM still does not fit well to the index of relative adult abundance (Minte-Vera et al., 2017). The difference in the population scale between the ASPM and the full model can be somewhat reduced by including the specification and estimation of recruitment variation in the ASPM (e.g., Minte-Vera et al., 2017). Therefore, the ASPM in which recruitment levels are estimated (ASPM-R-EST) can have a good fit to the index of relative abundance. However, when there is little information in the recruitment index, the results will be less precise with recruitment estimated in the ASPM-R-EST, and the recruitment variation may account for bias caused by model misspecification (Minte-Vera et al., 2017; Merino et al., 2022). Conversely, when fluctuations in recruitment depicted in the recruitment index are reliable, we hypothesize that adding estimation of these fluctuations (process variation) to the ASPM should improve the predictions (e.g., goodness of fit) indicated in the adult abundance indices.

To evaluate whether including a recruitment index contributes reliable information on recruitment levels, we used our study model. Historically, the majority of catch of Pacific bluefin tuna, in terms of numbers, has consisted of age-0 fish (Suppl. Material), primarily taken by troll fishing operations in the Sea of Japan, with some taken from the Pacific Ocean. Age-0 fish caught in the Sea of Japan originate from spawning grounds in the western North Pacific Ocean, between Taiwan and Japan, and in the Sea of Japan itself (Uematsu et al., 2018). In contrast, age-0 fish caught in the Pacific Ocean are spawned exclusively in the western North Pacific Ocean. Two CPUE indices for troll fisheries, which can be used as recruitment indices, were developed by using generalized linear models (Ichinokawa et al.[1]; Fig. 2C). The index of recruitment for age-0 fish from the Sea of Japan was used in an earlier stock assessment (ISC, 2018), and the recruitment index for age-0 fish from the Pacific Ocean was used to monitor recruitment in that region. Both indices for age-0 fish are used separately in the analyses described in the next few paragraphs.

To derive the temporal recruitment levels that exactly matched the recruitment index, we constrained the ASPM to estimate recruitment deviations that aligned with the variability of the recruitment index, with no error (ASPM-R-EST), using a weighting 10 times higher than that used for the adult abundance indices. In all other years without a recruitment index, the deviations were maintained at the expectation of the SR process. We re-ran the ASPM-R-EST with these estimates of recruitment deviations specified (fixed), without the recruitment index and the additional weighting (ASPM-R-FIX). In this context, recruitment deviations are the parameters in the model, and recruitment levels are the derived quantities of interest, calculated by combining the deviations with the mean recruitment.

The results from the ASPM-R-FIX were then compared to results from the original ASPM (for a comparison of data types used and parameters estimated among the full model, ASPM, and ASPM-R-FIX, see Supplementary Table). A lower root mean square error (RMSE) of the fit to each adult abundance index in ASPM-R-FIX, relative to that of the original ASPM, indicates a better fit and greater consistency between the recruitment index and indices of adult abundance given the production function. This procedure was completed separately for the index for age-0 fish from the Sea of Japan and the index for age-0 fish from the Pacific Ocean. Furthermore, it was conducted separately for the time-invariant and time-varying versions of the ASPM-R-FIX with the index for age-0 fish from the Sea of Japan.

Randomization tests: performance of model with fixed recruitment deviations

We evaluated if the ASPM-R-FIX improved fits to the adult abundance indices relative to the associated fits of the ASPM, using a randomization procedure for determination of the probability that any improvement in predictions from the ASPM-R-FIX could have been due to random chance. We resampled the fixed recruitment deviations from the ASPM-R-FIX without replacement and randomly reallocated them to the indexed years. The ASPM-R-FIX was re-run with recruitment deviations fixed at these values, and the RMSE of the fit to the indices of adult abundance was calculated. Five hundred iterations of this procedure created a distribution of RMSE for each of the 3 indices of adult abundance. The RMSE from the original ASPM-R-FIX was compared to the RMSE distribution for the randomized version of the ASPM-R-FIX fit to each adult abundance index. If the RMSE of the original ASPM-R-FIX was smaller (better fit) than 95% of the randomized ASPM-R-FIX, we would have strong evidence that model improvement resulted from including recruitment consistent with the recruitment index.

[1] Ichinokawa, M., K. Oshima, and Y. Takeuchi. 2012. Abundance indices of young Pacific bluefin tuna, derived from catch-and-effort data of troll fisheries in various regions of Japan. Int. Sci. Comm. Tuna Tuna-like Species North Pac. Ocean, Work. Group Rep. ISC/12-1/PBFWG/11, 35 p. [Available from website.]

We conducted 3 versions of the resampling experiment: 1) resampling all recruitment levels, 2) resampling only extreme recruitment levels, and 3) resampling only moderate recruitment levels. Extreme recruitment deviations were defined as those outside the 25th or 75th percentiles. Moderate recruitment deviations were those inside the 25th and 75th percentiles. In the analysis with resampling of only extreme recruitment, moderate recruitment deviations were fixed at year-specific values. In the experiment in which only moderate recruitment levels were resampled, extreme recruitment deviations were fixed at year-specific values. The analysis with resampling of all recruitment levels was used to judge if improvement for the recruitment index was nonrandom. The experiment with resampling of only extreme recruitment was used to determine if improvement was due to the recruitment CPUE only picking up extreme recruitment levels. The analysis in which only moderate recruitment levels were resampled was used to judge if the recruitment index was reliable for even small recruitment fluctuations. For brevity, only the analysis with all recruitment levels was done for the time-invariant model.

Results

Performance of production model

The long-term trends in abundance as indicated in the indices for adults include decadal periods of increasing and decreasing abundance (e.g., a 2-way trip; Fig. 2A). With both versions of the ASPM (time invariant and time varying), we were able to predict the inflection point of abundance increasing in the mid-1980s because of lower catches and the subsequent shift to decreasing abundance that began in the mid-1990s. The time-invariant and time-varying versions of the model had similar fits to the adult abundance indices, with the exception of the time-varying model having a better fit to the index based on the late time series for longline fisheries of Japan (1993–2016) (Table 1).

Performance of model with fixed recruitment deviations: goodness of fit

Both the time-varying and time-invariant ASPM-R-FIX that included recruitment fluctuations that exactly matched the recruitment index for the Sea of Japan (Fig. 2C) captured the long-term trends in abundance along with some shorter-term fluctuations as indicated in the indices of adult abundance (Fig. 2A). The time-varying ASPM-R-FIX had improved fits to both the indices based on the late time series for longline fisheries of Japan and the time series for longline fisheries of Taiwan over the fits of the ASPM to those time series (Fig. 2A, Table 1), but the time-invariant ASPM-R-FIX had an improved fit to only the late time series for Japan. Overall, only the improvement in the fit to that index was substantial. With both versions of the ASPM-R-FIX, we estimated a similar magnitude of the global scale with less uncertainty (i.e., standard deviation) than that of both versions of the ASPM (Table 1).

In contrast, with the time-varying ASPM-R-FIX that included recruitment fluctuations that exactly matched the recruitment index for the Pacific Ocean (Fig. 2C), unfished biomass was estimated at over 2 million metric tons, a level 2.9 times higher than that estimated with the time-varying ASPM. This version of the ASPM-R-FIX had

Table 1

Comparison of model gradients, estimates of unfished spawning biomass, and the root mean square errors (RMSEs) for the fits to 3 indices of adult abundance for versions of the standard age-structured production model (ASPM) and of the ASPM with temporal recruitment fluctuations fixed to match the recruitment indices based on data from the Sea of Japan or the Pacific Ocean (ASPM-R-FIX). The models were fitted to time series of catch per unit of effort of Pacific bluefin tuna (*Thunnus orientalis*) from 3 longline fleets: an early and a late time series for the fleets operating out of Japan in 1974–1992 and 1993–2016 and a time series for the fleet operating out of Taiwan in 2000–2016. Model versions include those in which fleet selectivity was assumed to remain constant over time or to vary over time. Biomass values are given in metric tons, and standard deviations of those estimates are provided.

			ASPM-R-FIX		
	ASPM		Sea of Japan		Pacific Ocean
Model aspect compared	Time invariant	Time varying	Time invariant	Time varying	Time varying
Gradient	8.95×10^{-6}	4.14×10^{-6}	4.28×10^{-6}	4.15×10^{-6}	4.02×10^{-9}
Population scale					
Unfished spawning biomass	807,486	814,201	665,813	674,555	2,350,090
Standard deviation	(11,274)	(10,723)	(1689)	(1345)	(67,228)
Trend (RMSE)					
Japan 1974–1992	0.40	0.39	0.45	0.41	0.39
Japan 1993–2016	0.51	0.43	0.29	0.29	0.87
Taiwan 2000–2016	0.31	0.32	0.33	0.30	0.58

a degraded fit to the 2 recent adult abundance indices and did not capture the long-term trends in abundance (Fig. 2B, Table 1).

Performance of model with fixed recruitment deviations: randomization tests

With all recruitments randomized, we found no statistical evidence of improved predictions from the time-varying ASPM-R-FIX that included recruitment fluctuations that exactly matched the recruitment index for the Pacific Ocean for the 3 adult abundance indices based on the early time series for the fleet operating out of Japan (P=0.19), the recent time series for the fleet operating out of Japan (P=0.62), and the recent time series for the fleet operating out of Taiwan (P=0.41) (Fig. 3). In contrast, results from use of the time-varying ASPM-R-FIX that included recruitment fluctuations that matched the recruitment index for the Sea of Japan provide strong statistical evidence of improved predictions for the 2 recent indices of adult abundance (P<0.05) (Fig. 3). For the time-invariant ASPM-R-FIX including the recruitment index for the Sea of Japan, we found strong statistical evidence of improved predictions for both the recent adult abundance indices (P<0.05) (for comparisons, see Supplementary Figure). However, no improvement is evident in the fit to the early index of adult abundance in fisheries of Japan (time-invariant model: P=0.48; time-varying model: P=0.22; for comparisons, see Supplementary Figure).

When considering only the extreme recruitment fluctuations in the time-varying ASPM-R-FIX fit to data from the Sea of Japan, we found strong statistical evidence for improved predictions for both recent adult abundance indices (P<0.05; Fig. 4, B and C). However, with only the extreme fluctuations used in analysis, there is no evidence of improvement for the early adult abundance index for the fleet from Japan (P=0.39; Fig. 4A). In the experiment in which only moderate recruitment fluctuations were used, evidence of improved predictions was strong for the recent index of adult abundance in fisheries of Japan (P<0.05; Fig. 4B). In contrast, there was no evidence of improvement for the early index from Japan (P=0.11) and the recent index from Taiwan (P=0.28) in that experiment (Fig. 4, A and C).

Discussion

This work supports the notion that there is a nonrandom and informative connection between the variability in recruitment as depicted by the recruitment index for the Sea of Japan and subsequent changes in adult abundance indices given the observed catches. We assume that this connection exists because the interannual variability in recruitment indicated in the abundance index for age-0 fish from the Sea of Japan is close to the variability of the true population. This improvement in population dynamics as given by improved fit to indices of adult abundance is strongest for the recent indices and weakest for the earliest index.

This lack of connection to the early index for adult abundance in longline fisheries of Japan can be explained by the time series having relatively fewer observations of cohorts born after 1980, which is the start of the time series on which the recruitment index is based, in comparison to the number of observations for other indices of adult abundance in longline fisheries. Longline fleets target Pacific bluefin tuna at ages older than 6 or 7 years, and the longline fleet operating out of Taiwan generally targets even older fish. Consequently, only a few observations of cohorts born after 1980 could be included in the early longline time series, and we should not expect the recruitment index to have much connection to the early data. In contrast, Pacific bluefin tuna can live longer than 20 years, and the recent longline CPUE indices may include more than 15 years of observations on the strength of many cohorts represented in the recruitment index. Additionally, because the longline fleet operating out of Japan targets somewhat younger adults than the longline fleet of Taiwan, more cohort information may be in the time series for Japan over the same period. This difference in the number of observations could explain why the randomization scenario that evaluated only moderate fluctuations was significant only for the recent index based on longline data for fisheries of Japan. Additionally, the Pacific bluefin tuna population was highly depleted after 2000, and the adult population consisted of relatively few cohorts. Therefore, annual recruitment levels likely had a pronounced effect on the recent adult abundance.

It should be noted that this evaluation of the reliability of the recruitment index is weak for the last few cohorts in both recent time series, as these cohorts would not yet have made it to being caught by the longline fleets. Still, for applied work, it may be reasonable to conclude that if the levels of recruitment for the other years were strongly connected to adult abundance, the levels of recruitment for the most recent years should be as well. However, this assumption may not hold if catchability of the index is altered by changes in data collection, environmental conditions, or spatial patterns of fish distribution or by future management actions. Therefore, these factors should be reevaluated in the continued development and application of this recruitment index as data for more years are added.

It should also be noted that the convenience of setting steepness at 1.0 in developing the ASPM-R-FIX may complicate comparisons between the results of the ASPM and ASPM-R-FIX for stocks with strong maternal effects (for which steepness is much lower than 1.0). However, in cases involving stocks with a strong maternal effect, it may be advisable to conduct the randomization experiment by using deviations of the recruitment index around the expected SR function estimated in the ASPM-R-FIX to represent independent and identically distributed random variables. More work in this area is warranted.

In this analysis, the population of Pacific bluefin tuna appears to be one for which fishing is a major contributor to changes in adult abundance. However, this randomization evaluation should work, possibly better, in a system where recruitment fluctuations are the primary driver of

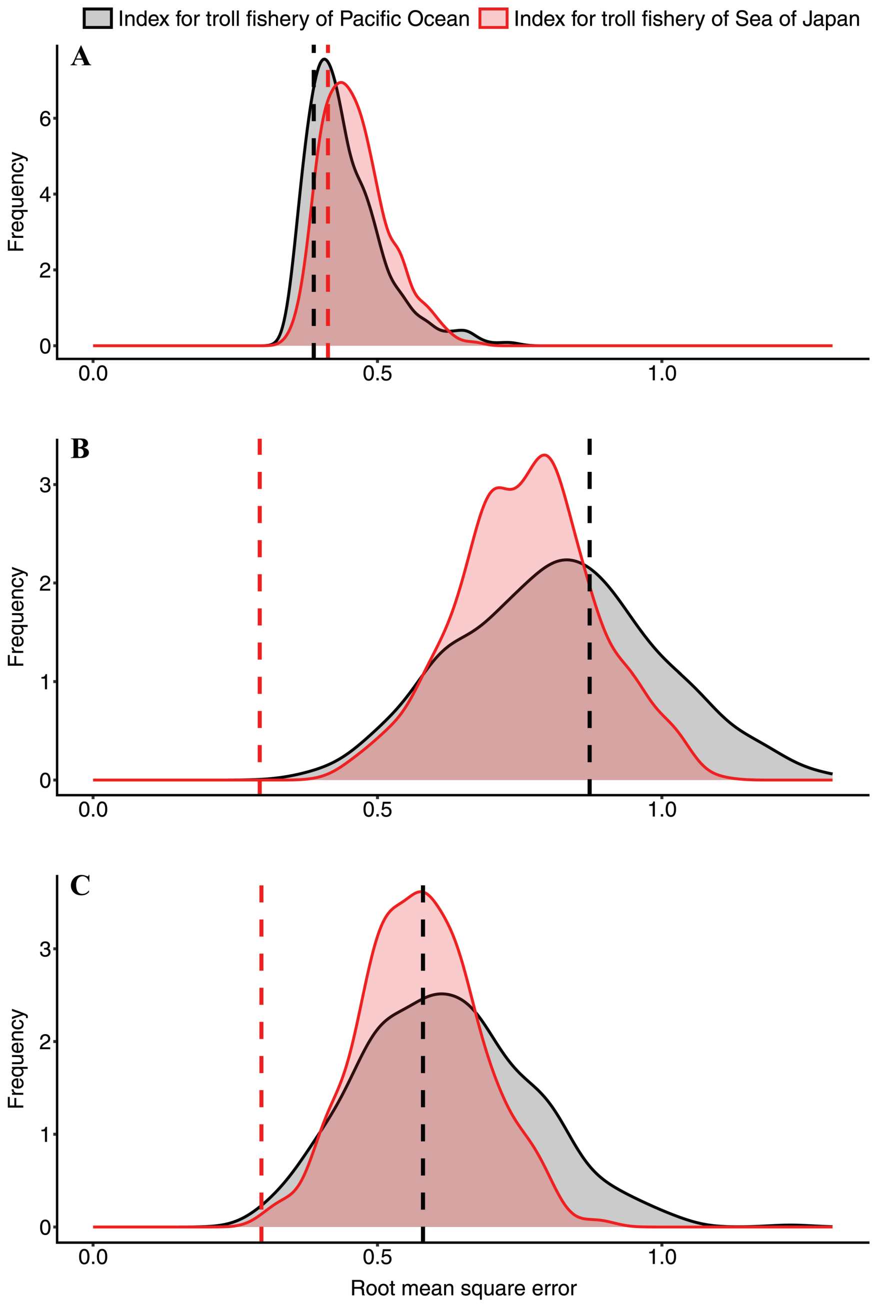

Figure 3

Distributions of the root mean square errors (RMSEs) from 500 iterations of a randomization method used to evaluate the fit of the time-varying versions of the age-structured production model with fixed recruitment deviations (ASPM-R-FIX) to each of the 3 indices of abundance of adult Pacific bluefin tuna (*Thunnus orientalis*) based on catch-per-unit-of-effort (CPUE) time series from longline fleets: (**A**) an early and (**B**) a late time series for the fleets operating out of Japan in 1974–1992 and 1993–2016 and (**C**) a time series for the fleet operating out of Taiwan in 2000–2016. For comparison, RMSEs are provided for the fit of the time-varying versions of the ASPM-R-FIX with recruitment fluctuations fixed to match recruitment indices (CPUE data for age-0 fish) from the troll fishery in the Sea of Japan (vertical red dashed lines) or the Pacific Ocean (vertical black dashed lines) to the 3 abundance indices. All recruitment levels were resampled without replacement and reallocated to the respective year, and the model was run again with recruitment deviations fixed at these values.

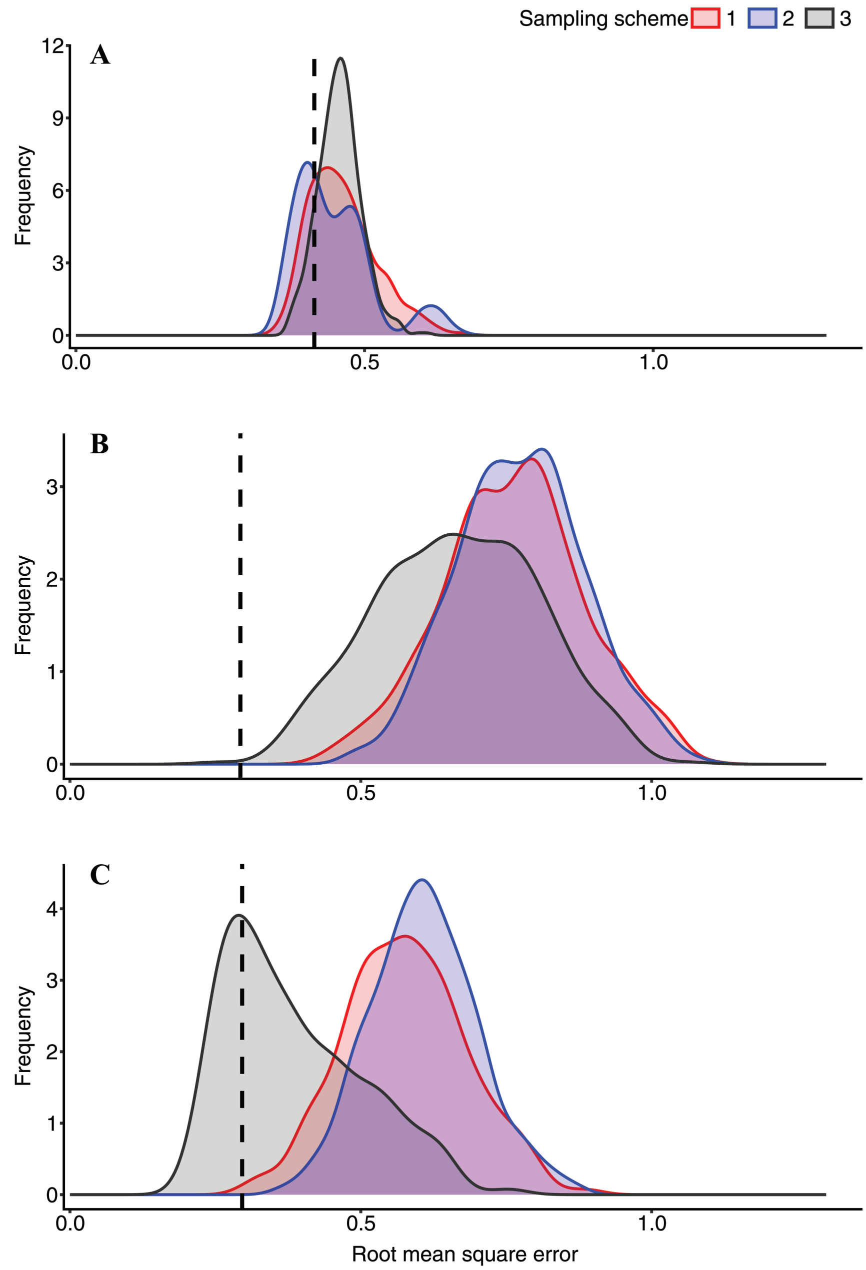

Figure 4

Distributions of the root mean square errors (RMSEs) from 500 iterations of a randomization method used to evaluate the fit of the time-varying versions of the age-structured production model with fixed recruitment deviations (ASPM-R-FIX) to each of the 3 indices of abundance of adult Pacific bluefin tuna (*Thunnus orientalis*) based on catch-per-unit-of-effort (CPUE) time series from longline fleets: (**A**) an early and (**B**) a late time series for the fleets operating out of Japan in 1974–1992 and 1993–2016 and (**C**) a time series for the fleet operating out of Taiwan in 2000–2016. Three sampling schemes were used in the randomization method: 1) all recruitment levels, 2) only the extreme recruitment levels (outside the 25th or 75th percentile), or 3) only the moderate recruitment levels (within the 25th and 75th percentiles) were resampled without replacement and reallocated to the respective year, and the model was run again with recruitment deviations fixed at these values. For comparison, the vertical dashed line in each panel indicates the RMSE for the fit of the time-varying versions of the ASPM-R-FIX, with recruitment fluctuations fixed to match the recruitment index (CPUE data for age-0 fish) from the troll fishery in the Sea of Japan, to the 3 abundance indices.

adult abundance, given that adding reliable process variability to the SR function should still improve ASPM-R-FIX predictions. In contrast, the precision with which the recruitment index reflects actual recruitment fluctuations will affect the evaluation method, as it does not include this observation error. In the case of Pacific bluefin tuna, the troll fishery used to create the recruitment index is the largest fishery numerically in the assessment, and the data from the Sea of Japan accounts for most catches in the troll fishery, likely eliminating many sources of potential uncertainty. For other systems in which various forms of sampling error are larger, the results may not be as clear.

An important consideration in all analyses done with models like the ASPM is elimination of the challenges associated with fitting composition data but still retaining the age structure of the model. Age structure is important for work on Pacific bluefin tuna because fisheries target different segments of the age structure of populations, and these segments can vary somewhat from year to year. Therefore, we used only the time-varying ASPM-R-FIX to conduct the detailed randomizations, the scenarios in which only extreme recruitment levels or only moderate recruitment levels were resampled. The randomization results from both the time-varying and time-invariant ASPM-R-FIX were nearly identical, giving greater confidence in the potential for improvement in model estimation that including a recruitment index in the actual stock assessment can bring.

The concepts behind this work are not new, and the burden of creating the ASPM is not substantial because software packages like Stock Synthesis offer controls that make it relatively easy to eliminate data influence and adjust parameterization to control model complexity. It is relatively clear that adding process variability in the recruitment index based on data from the troll fishery in the Sea of Japan improved model predictions of the effects of fishery catches and that such an index should therefore be included in the applied assessment model. Inclusion of a reliable recruitment index will improve model predictions and understanding of the consequences of past removals. A better understanding of the historical drivers of population dynamics will lead to more accurate predictions of the consequences of future management decisions.

Conclusions

The results of this study indicate that the diagnostic test in the ASPM is a useful tool for assessing drivers of both long- and short-term dynamics of populations of Pacific bluefin tuna and for evaluating potential model improvements. Using a well-defined system with informative contrast, we proved that one effective approach to developing a more complex model is to consider the additional data and associated model processes that improve ASPM performance. Notably, the time-varying ASPM-R-FIX, in which recruitment fluctuations derived from data for the troll fishery in the Sea of Japan are explained, was found to have improved predictive performance for adult abundance trends over the standard ASPM, particularly in evaluations with the late index for fisheries of Japan and the index for fisheries of Taiwan. The results of this analysis reveal a strong connection between recruitment variability and subsequent adult abundance, underscoring the importance of integrating a consistent recruitment index into stock assessments. Although the results showcase the effectiveness of the ASPM-R-FIX in accounting for process variability within the production function, they also highlight the need to carefully consider factors such as changes in fishery practices, environmental conditions, and data collection methods. Incorporating a reliable recruitment index strengthens the model's predictive capabilities, enabling more accurate management decisions. As more observations become available, ongoing evaluation and refinement of the recruitment index will be essential to maintain the robustness of these models.

Resumen

El objetivo de la modelación en evaluación de poblaciones es evaluar el estado actual y futuro de las poblaciones en distintos escenarios de manejo. Para estas evaluaciones, especialmente las proyecciones a corto plazo, los investigadores se basan en gran medida en las estimaciones del reclutamiento. Sin embargo, estimar con precisión los niveles de reclutamiento, especialmente para los años más recientes, es un reto porque la información fiable es limitada. En tales casos, los índices de reclutamiento relativo pueden proporcionar información valiosa para los modelos de evaluación de poblaciones. Aquí se presenta un método para evaluar la fiabilidad de un índice de reclutamiento con pruebas de diagnóstico y aleatorización en un modelo de producción estructurado por edades (ASPM). El método propuesto puede utilizarse para evaluar si la variabilidad de proceso que implica un índice de reclutamiento para la relación desovantes-reclutamiento mejora la conexión entre las capturas observadas, los índices de abundancia de adultos y la función de producción del modelo de evaluación. Aplicamos este enfoque al atún aleta azul del Pacífico (*Thunnus orientalis*) como ejemplo ilustrativo. Los resultados indican que el ASPM, con fluctuaciones de reclutamiento que coinciden con el índice de reclutamiento para el Mar de Japón (ASPM-R-FIX), mejoró los ajustes a los recientes índices de abundancia de adultos respecto a los modelos ASPM estándar. Los resultados del análisis estadístico proporcionan además evidencia sólida de que las fluctuaciones extremas del reclutamiento en el ASPM-R-FIX mejoraron las predicciones de los índices recientes de adultos. El presente análisis reveló un fuerte vínculo entre la variabilidad del reclutamiento y la abundancia de adultos, destacando la importancia de añadir un índice consistente de reclutamiento a un modelo de evaluación para mejorar las decisiones de manejo.

Acknowledgments

We thank all reviewers both known and anonymous for their helpful critiques. A special thank you goes to A. Punt, who provided constructive comments on an earlier draft, and to B. Muhling. We recognize the Pacific Bluefin Tuna Working Group of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean for prompting this work.

Literature cited

Beverton, R. J. H., and S. J. Holt.
1957. On the dynamics of exploited fish populations. G. B. Minist. Agric. Fish. Food, Fish. Investig. Ser. 2, vol. 19, 533 p. HMSO, London.
Francis, R. I. C. C.
2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68:1124–1138. Crossref
2017. Revisiting data weighting in fisheries stock assessment models. Fish. Res. 192:5–15. Crossref
Hilborn, R., and C. J. Walters.
1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty, 570 p. Routledge, Chapman and Hall Inc., London.
ISC (International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean).
2018. Stock assessment of Pacific bluefin tuna (*Thunnus orientalis*) in the Pacific Ocean in 2018. Report of the Pacific Bluefin Tuna Working Group. Annex 14. 18th meeting of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean, Yeosu, Republic of Korea, July 11–16, 2018. Int. Sci. Comm. Tuna Tuna-like Species North Pac. Ocean ISC/18/ANNEX/14, 147 p. [Available from website.]
2022. Stock assessment of Pacific bluefin tuna (*Thunnus orientalis*) in the Pacific Ocean in 2022. Annex 13. 22nd meeting of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean, Kona, Hawai'i, U.S.A., July 12–18, 2022. Int. Sci. Comm. Tuna Tuna-like Species North Pac. Ocean ISC/22/ANNEX/13, 150 p. [Available from website.]
Maunder, M. N., and A. E. Punt.
2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. 142:61–74. Crossref
Maunder, M. N., and K. P. Piner.
2015. Contemporary fisheries stock assessment: many issues still remain. ICES J. Mar. Sci. 72:7–18. Crossref
Merino, G., A. Urtizberea, D. Fu, H. Winker, M. Cardinale, M. V. Lauretta, H. Murua, T. Kitakado, H. Arrizabalaga, R. Scott, et al.
2022. Investigating trends in process error as a diagnostic for integrated fisheries stock assessments. Fish. Res. 256:106478. Crossref
Methot, R. D., Jr., and C. R. Wetzel.
2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fish. Res. 142:86–99. Crossref
Minte-Vera, C. V., M. N. Maunder, A. M. Aires-da-Silva, K. Satoh, and K. Uosaki.
2017. Get the biology right, or use size-composition data at your own risk. Fish. Res. 192:114–125. Crossref
Pella, J. J., and P. K. Tomlinson.
1969. A generalized stock production model. IATTC Bull. 13(3):421–496. [Available from website.]
Schaefer, M. B.
1954. Some aspects of the dynamics of population important to the management of the commercial marine fisheries. IATTC Bull. 1(2):27–56. [Available from website.]
Uematsu, Y., T. Ishihara, Y. Hiraoka, T. Shimose, and S. Ohshimo.
2018. Natal origin identification of Pacific bluefin tuna (*Thunnus orientalis*) by vertebral first annulus. Fish. Res. 199:26–31. Crossref