# Development of a Pacific Bluefin Stock Assessment

H.H. Lee[1], K. Piner[1], M.Maunder[2], A. Aires-da-Silva[2]

[1]NOAA Fisheries, SWFSC, 8901 La Jolla Shores Drive, La Jolla, Ca. 92037

[2]IATTC, 8901 La Jolla Shores Drive, La Jolla, Ca. 92037

## Abstract

This paper presents objective criteria by which a stock assessment model was developed for Pacific bluefin tuna. The goal of the work was to create an internally consistent model that follows objective criteria using a series of CAPAM workshops on population modelling as guideline. We assert that agreed data should be considered true. Unacceptable diagnostic for model fit to data or conflict between data series is indicative of model misspecification. Misspecification was addressed using either additional model process in the form of flexible and time-varying selectivity patterns or by adding the un-modelled process to the observation error. To keep the model parsimonious, prioritization criteria were developed to determine which data sources would be addressed by time-varying selectivity and which would be addressed by data weightings.

## Introduction

This paper presents objective criteria by which a stock assessment model was developed for Pacific bluefin tuna (PBF), using the data and life history agreed to by the PBF working group. The objective criteria are an outcome of a series of CAPAM workshops on population modelling. The development of a stock assessment model is comprised of the assessment model processes, data and statistical methods for comparing data to predictions. Once data has been reviewed and accepted, the data should be considered true, as any series considered untrue should have been rejected. The appropriate model process to account for any biases between the population and the data should be included in the dynamic model. Systematic misfit to data or conflict between data within an assessment model should be considered as a diagnostic of model misspecification. Model misspecification can be the result of missing important model processes, incorrect structure of a model process, incorrect specification of a model process parameter, and/or incorrect statistical assumptions about the data. Misspecification occurs for three reasons: 1) random sampling error, 2) misspecification of the observation model (model processes relating dynamics or states to data), and/or 3) misspecification of the system dynamics model (the population dynamics model). Observation model misspecification is a result of the relationship between the data and the underlying system dynamics being incorrect. Systems dynamics model misspecification occurs when the underlying processes governing the population dynamics are incorrectly specified.

Unacceptable model fit (model predictions do not match the data) can be either the magnitude of the residual is larger than implied by the observation error or trends in residual indicating systematic misfit. Data conflicts occur when different data series, given the model structure, provide different information about important aspects of the dynamics. Model misfit and conflict should be interpreted in the context of the random sampling error (treated in the model as observation error). Sampling error occurs because we do not census the entire population. Misfit or conflict can occur because the observation error in the model (data weight) is too small overstating the precision of that data series. In contrast, large observation errors also mean that the data themselves are not informative about model process.

*Dealing with conflict and misfit*

Unacceptable model misfit or conflict between data can be dealt with by either data weightings or model process changes/flexible model parametrization. If the observation error in the model was initially assumed to be too precise, then adding additional observation error is an appropriate solution, such as when residual patterns are random but the standardized residuals are larger than assumed. However when the misspecification is other than incorrect assumption of the data precision, down-weighting does not address the underlying problem (Wang *et al.* 2015). Re-weighting the data only reduces the contribution of misfit to the estimation of the dynamics. When either the system or observation models are misspecified, modelling the correct process is the ideal solution (Lee *et al.* 2015; Lee *et al.* in review). The difficulty in dealing with data conflict arises because the actual misspecified process is often unknown or the process (including variation) is not estimable.

*Prioritize data (Francis 2011)*

Because it is difficult to determine the underlying cause of the model misfit and conflict, we often assume that some data are more reliable than other data for determining particular aspects of the population dynamics (Francis 2011). Models are developed to insure that the estimation of specific aspects of the model is consistent with identified and prioritized data. Other data sources are treated in a manner that diminishes their contribution to these key aspects of the dynamics creating internally consistent models. Important aspects of the model include trends in population, absolute population scale (Kell *et al.* 2014: Lee *et al.* 2014), relative population scale (depletion), and current age structure (important for short-term projections). The types of data available for the PBF assessment are limited to:

*Catch*

Catch provides minimum historical population size and is the direct measures of the anthropogenic disturbance to the system. The effects of this disturbance are what we want to understand. Catch is generally considered known with little error and models should be able to remove the amount assumed.

*Indices of abundance*

Reliable indices of abundance provide us information on trend and in conjunction with catch and an elucidated production function scale provide direct evidence of population scale (Lee *et al.* 2014; Maunder and Piner 2015). Indices of abundance generally are considered the most valuable information on population dynamics and should be adequately fit within the model (Francis 2011).

*Composition*

Composition data provide direct observations of catch-at-age and the relative recruitment strength of adjacent cohorts. The primary use of composition data within the integrated model should be to inform the fleet selectivity process for use in removing the correct catch-at-age and estimating recruitment variation. With sufficient knowledge of life history and fishery selectivity, composition data can provide information on population scale but this information is highly sensitive to misspecification (Maunder and Piner 2015). Small differences between observed and predicted composition can impart undue influence on model results and therefore reliable indices of abundance should be prioritized above composition data for model estimates of both scale and trend.

Our goal is to create a dynamic model of all the available data that fits the data well and is internally consistent. Internal consistency implies all data are fit as well as their observational errors and trends in residuals are minimized. Important aspects of the dynamics (scale, trend and relative scale) should be derived from the most trusted data sources.

## Methods and Results

This paper uses the stock assessment software Stock Synthesis (3.24Y). Our modelling approach can be summarized as the following steps:

1. Carefully select the data and estimates of the true sampling error;

2. Create the initial model based on understanding/hypothesis with original sampling error;

3. Determine if indices have information on scale and prioritize data;

4. Apply model diagnostics;

5. Modify or add additional process based on diagnostics and complete steps 4 and 5 again until internally consistent model is developed;

6. Re-weight the data as needed.

*Step 1.*

All data agreed to by the WG in the November 2015 meeting in Taiwan (ISC 2015) were incorporated.

*Step 2.*

The initial model developed used the agreed data sources and life history (ISC 2015). Key elements of the model included separately modeled time-invariant length- and age-based selection patterns for most fisheries taking fish of migratory ages (ages 1-5) (Table 1). For fleets that include a combination of both length- and age- based selection, the length-based selection was modeled as asymptotic while age-based selection was modeled assuming a separate selection parameter for each age. Length-based selection accounts for the contact selectivity of the gear, and age-based selection accounts for age-based availability.

*Step 3.*

Based on model exploration, production model diagnostics and weighting exercises, the Japanese longline CPUE (all 3 series) were deemed the most reliable information for trends, and in conjunction with catch the most reliable for estimating population scale (Figure 1). The Japanese longline CPUE was prioritized for both scale and trend.

Size composition was prioritized relative to other fishery size composition data for fitting (i.e. fit the composition data well to ensure that the fish were removed at the correct size and/or to provide information about abundance or model processes). Fisheries catching a large amount (in weight) of fish were given high priority (F5), fisheries historically taking a large amount of fish were given a medium priority (F13), and fisheries catching a small amount of fish were given low priorities (Table 2). The argument is that composition data should not provide information on abundance and that it is more important to remove fish of the right size for fisheries that catch a large amount of fish, particularly in recent years. High priority fleets were modelled using additional process in the form of time-varying selectivity so that the removals at age were consistent with the data (F4, F5). Fleet size compositions that showed conflict with the prioritized CPUE (F4) were also given high priority for fitting to reduce the misfit causing the conflict. Compositions with a single well-behaved mode (F2, F10, F12, F14) were given low priority because of ease of fitting with a single time-invariant selection. Median and low priority fleets could use data weights to account for misfit (F9, F13, F14). Fleets with CPUE (F1, F6, F12) were not considered in this prioritization because their size composition was both well-behaved and likely not considered for use with a time-varying selection pattern.

*Step 4.*

This initial model was evaluated for data fit using residual analyses (Figure 2) and conflict using a combination of up-weighting analyses (analyses are not shown) and retrospective analyses (Figure 3).

*Step 5.*

Based on the model diagnostics, time-varying selection was added to the initial model for fleets F4 and F5. All other fleets were one-directional down weighted based on the Francis weights (Table 2). This reference model was evaluated for data fit using residual analyses (Figure 4) and conflict using a combination of up-weighting analyses (analyses are not shown) and retrospective analyses (Figure 5).

## Discussion

The reference model provides good fit to all data, including the prioritized CPUE and size data. By separately modeling length- and age-based selection patterns for fisheries taking migratory ages, initial model reduces the misfit to the size data. Modeling process changes and data weightings further improve the model fit. Model conflict was somewhat reduced in the reference model as indicated by retrospective analysis. More work is needed to know if modeling a substitute but incorrect model process is better than eliminating or down-weighting data or accounting for the process error in the estimation of observation error. We note that right-weighting (Francis 2011) resulted in some fleet composition data receiving a higher weight. It is arguable if this is appropriate since the input sample sizes were already down weighted, but given the complexity of the population and the fisheries, we lower the risks of conflict among data sources by only allowing down weighting. We chose only adjusting initial multinomial samples size lower (down-weighting) when these down-weighted fleets produces nearly identical fit to the composition and better fit to the CPUE series. This should be discussed within the WG.

This reference model could be considered as a starting point for development of a base case model, using the methods discussed. We note that final polishing of these models will be needed to improve performance. One area for consideration is the unreasonably high estimates of initial F on the Japan troll fleet (F6). Because the troll fleet catches only a single age class (age 0), it is probably not the best choice for use as an initial equilibrium fleet on juveniles. We suggest changing the initial equilibrium fleet to a fleet catching a wider range of juveniles (ages) such as fleets F8, F9, or F13. Other minor model changes with this model should be considered as we approach a base case model to insure selection parameters do not end up on bounds due to the large number of parameters in the model.

The assessment model is based on the "The law of conflicting data" which includes the axiom "Data is true", which implies that "Conflicting data implies model misspecification" and therefore the model misspecification should be corrected. However, there are various ways in which data can be unrepresentative of the whole population and correctly modelling the process or observation sub-models is difficult. Some of the data used in the assessment model that shows conflict with the prioritized data may not be representative and be considered for further down weighting if the appropriate processes cannot be modelled.

As highlighted in the November PBF workshop, age-0 bluefin tuna show seasonal patterns in growth. We have not implemented an approach to accommodate this understanding, but simply modelled the F6 fleet affected by the seasonal growth with a very flexible spline selection pattern. Although this effectively fits the size composition and reduced misfit contribution to the total likelihood, it is not the best way to deal with the issue. An appropriate method to deal with growth needs to be addressed in the assessment meeting.

We have modeled the process error ($\sigma_r$) of the spawner-recruit relationship using an approach that decouples estimated recruitment from the predicted to a large degree. This method allows the model to be less sensitive to our assumptions about steepness. Until we better understand the spawner-recruit relationship, this may be a viable approach.

Our model development was contingent on the prioritizing the Japanese longline CPUE series. This is not only because of the good connection between catch and trends in that fisheries CPUE, but also because of the internal consistency of that data with our data sources. Model investigation demonstrated that the Taiwanese longline CPUE was inconsistent with other data including the composition data from that fleet. For these reasons, the Japanese longline data should be given greater priority. If the WG decides to build a model conditioned on the Taiwanese CPUE, the work could be put forward using the same objective criteria proposed in the paper and could be considered as an alternative model.

## References


Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. *Can. J. Fish. Aqua. Sci.* 68: 1124–1138.

ISC. 2015. Report of Pacific bluefin tuna working group intersessional workshop. November 18-25th, Kaohsiung, Taiwan.

Lee, H.H., K.R., Piner, R.D., Methot, Jr., and M.N., Maunder. 2014. Use of likelihood profiling over a global scaling parameter to structure the population dynamics model: an example using blue marlin in the Pacific Ocean. *Fish. Res.* 158: 138-146.

Lee, H.H., Piner, K.R., Maunder, M.N., and R.D. Methot Jr. 2015. Simulation of methods of dealing with age-based movement in PBF stock assessment. ISC/15/PBFWG-2/06

Lee, H.H., L.R., Thomas, K.R. Piner, and M.N., Maunder (in review) Effects of age-based movement on the estimation of growth assuming random-at-age random-at-length data. J. Fish. Bio.

Kell, L.T., P. De Bruyn, M. Maunder, K. Piner, and I.G. Taylor. 2014. Likelihood component profiling as a data exploratory tool for North Atlantic Albacore. Collect. Vol. Sci. Pap. ICCAT. 70(3):1288-1293.

Wang, S.P, Maunder, M.N., Nishida, T., Chen, Y.R . 2015. Influence of model misspecification, temporal changes, and data weighting in stock assessment models: Application to swordfish (*Xiphias gladius*) in the Indian Ocean. *Fish. Res.* 166: 119–128.

Table 1. The initial model structure where fleets with size compositions and CPUE number fit to the model are shown. Total 151 parameters.

| Fleet # | Contact selectivity | Age-based availability | Time-varying age-based process | Weighting | # of parameters | CPUE number |
|---|---|---|---|---|---|---|
| 1 | Domed shape (double normal) | None | None | None | 10 (with block) | S1, S2, S3 |
| 2 | Asymptotic (logistic) | Age-specific (ages 1-2) | None | None | 4 | |
| 4 | Asymptotic (logistic) | Age-specific (ages 3-9) | None | None | 9 | |
| 5* | Asymptotic (logistic) | Age-specific (ages 2-10) | None | None | 11 | S21 |
| 6 | Non-parametric (spline) | None | None | None | 9 | |
| 8 | Asymptotic (logistic) | Age-specific (ages 1-4) | None | None | 6 | |
| 9 | Asymptotic (logistic) | Age-specific (ages 1-5) | None | None | 7 | |
| 10 | Asymptotic (logistic) | Age-specific (ages 1-3) | None | None | 5 | |
| 12 | Asymptotic (logistic) | None | None | None | 2 | |
| 13* | Asymptotic (logistic) | Age-specific (ages 1-4) | None | None | 6 | |
| 14 | Asymptotic (logistic) | Age-specific (ages 1-4) | None | None | 6 | |

* Fleets with the majority of catch in weight in average. The fleet 13 is the historical fleet where its size compositions are prior to 1983.

Table 2. The reference model structure where fleets with size compositions and CPUE number fit to the model are shown. Total 295 parameters. Term in parenthesis denotes the priority given to fitting the composition data: N/A = CPUE, H= high priority, M=medium priority, L=low priority based largely on the importance of the fishery catch for estimating dynamics or reducing model conflict.

| Fleet # | Contact selectivity | Age-based availability | Time-varying age-based process | Francis Weighting | # of parameters | CPUE number |
|---|---|---|---|---|---|---|
| 1 (N/A) | Domed shape (double normal) | None | None | 1 | 10 (with block) | S1, S2, S3 |
| 2 (L) | Asymptotic (logistic) | Age-specific (ages 1-2) | None | 1 | 4 | |
| 4 (H) | Asymptotic (logistic) | Age-specific (ages 3-9) | 1987-2014 on ages 5-7 | 1 | 93 | |
| 5* (H) | Asymptotic (logistic) | Age-specific (ages 2-10) | 1995-2014 on ages 2-4 | 0.92 | 71 | S21 |
| 6 (N/A) | Non-parametric (spline) | None | None | 1 | 9 | |
| 8 (L) | Asymptotic (logistic) | Age-specific (ages 1-4) | None | 1 | 6 | |
| 9 (L) | Asymptotic (logistic) | Age-specific (ages 1-5) | None | 0.88 | 7 | |
| 10 (L) | Asymptotic (logistic) | Age-specific (ages 1-3) | None | 1 | 5 | |
| 12 (N/A) | Asymptotic (logistic) | None | None | 1 | 2 | |
| 13* (M) | Asymptotic (logistic) | Age-specific (ages 1-4) | None | 0.94 | 6 | |
| 14 (L) | Asymptotic (logistic) | Age-specific (ages 1-4) | None | 0.63 | 6 | |

* Fleets with the majority of catch in weight in average. The fleet 13 is the historical fleet where its size compositions are prior to 1983.

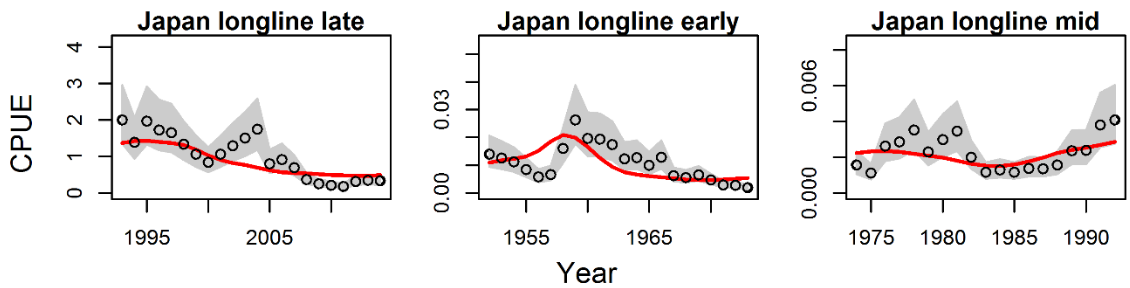
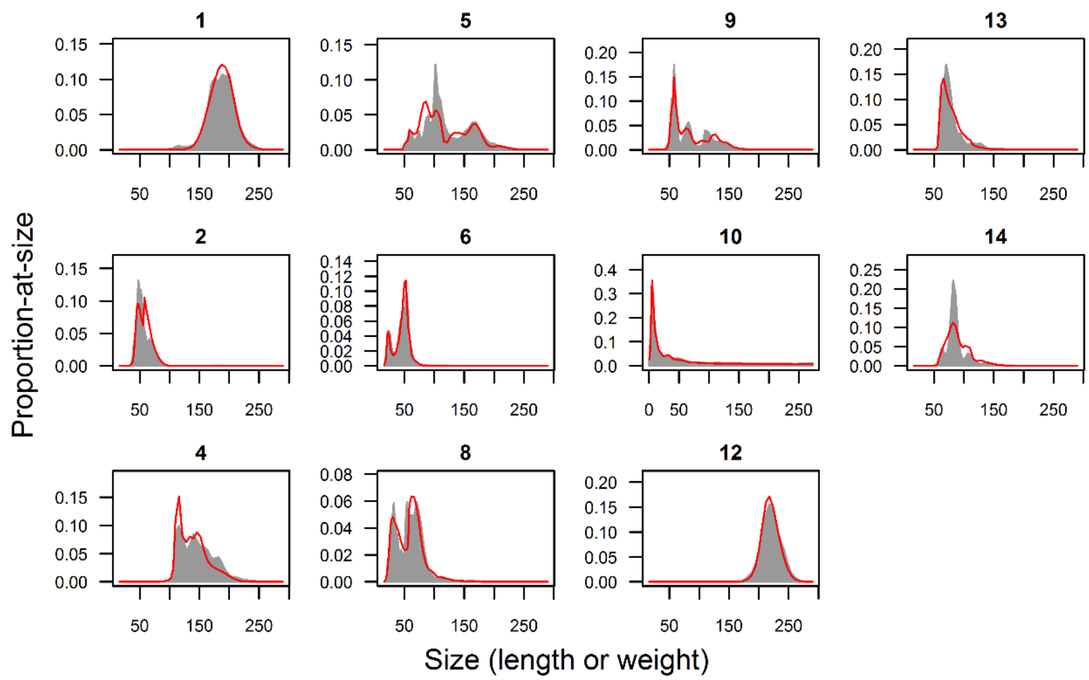



Figure 1. Japanese abundance indices fit using age structured production model as a diagnostic to indicate if there is production function and if catch explain abundance indices.

(a)




(b)

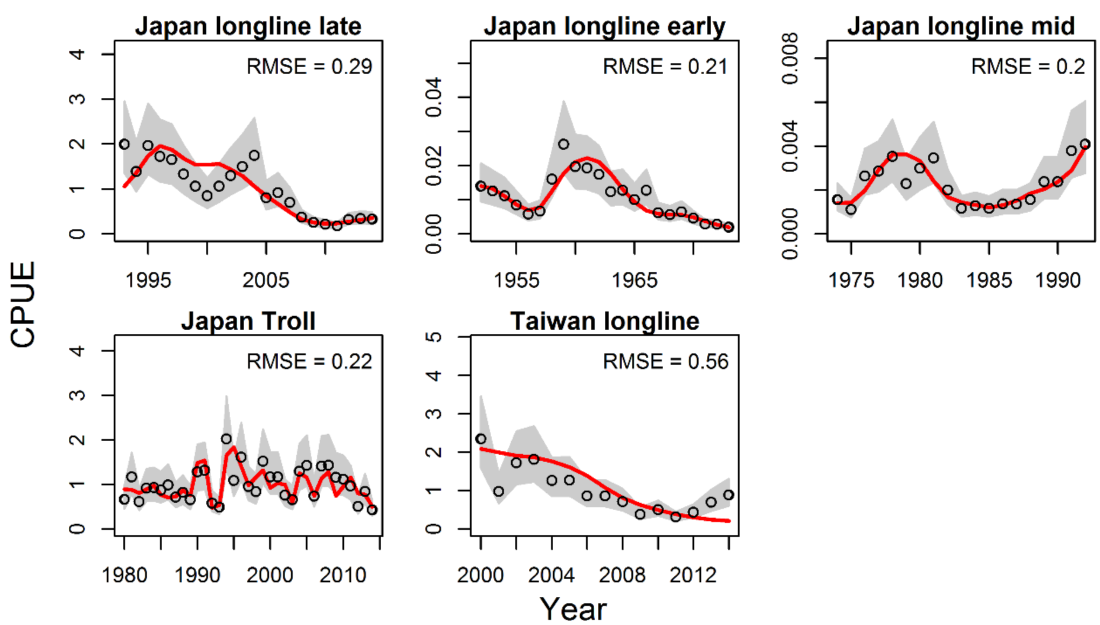



Figure 2. (a) The observed proportion-at-size (gray shaded area) and overall expected fit (red line) for the initial model, where fleets with size compositions fit to the model are shown. (b) The observed CPUE (open circles with gray shaded area) and expected fit (red line) for the initial model, where fleets with CPUE data fit to the model are shown expect for Taiwan longline fishery. The structure for initial model is summarized in Table 1.

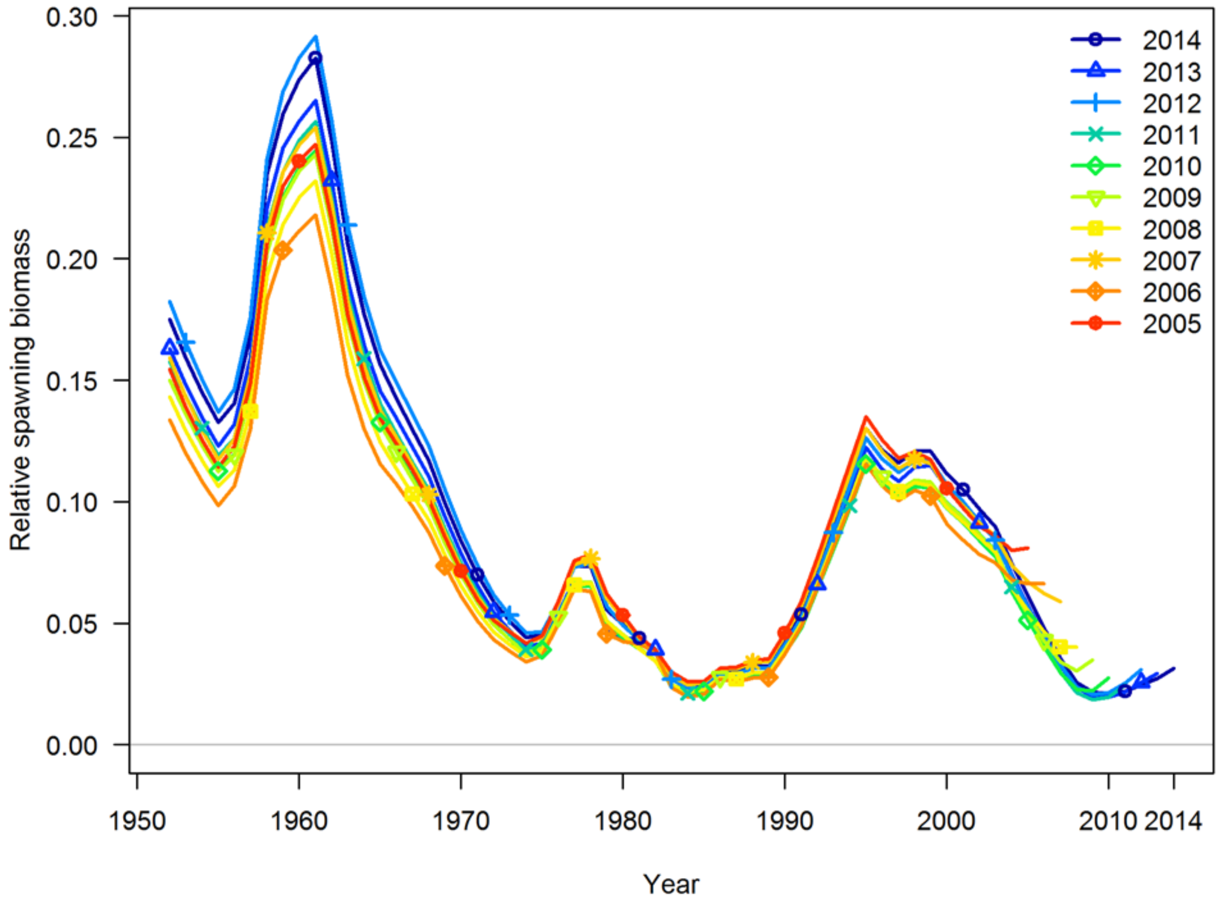
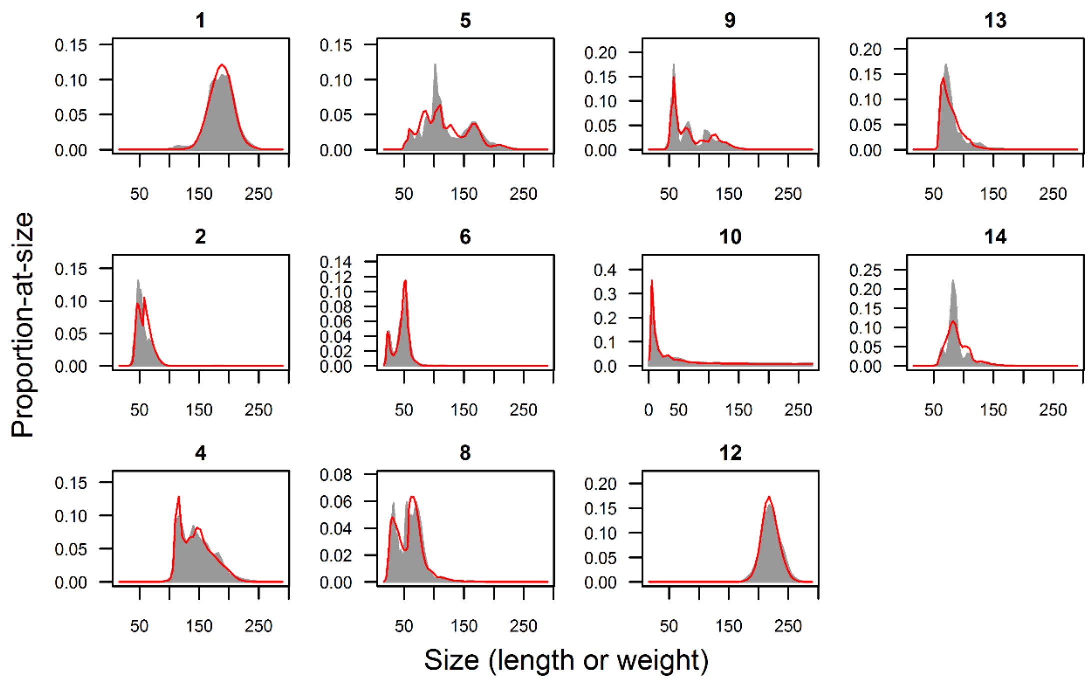



Figure 3. Nine-year retrospective analysis of relative spawning biomass to its unfished spawning biomass from the initial model. Each line represents a model fit with sequentially one less year of data.

(a)




(b)

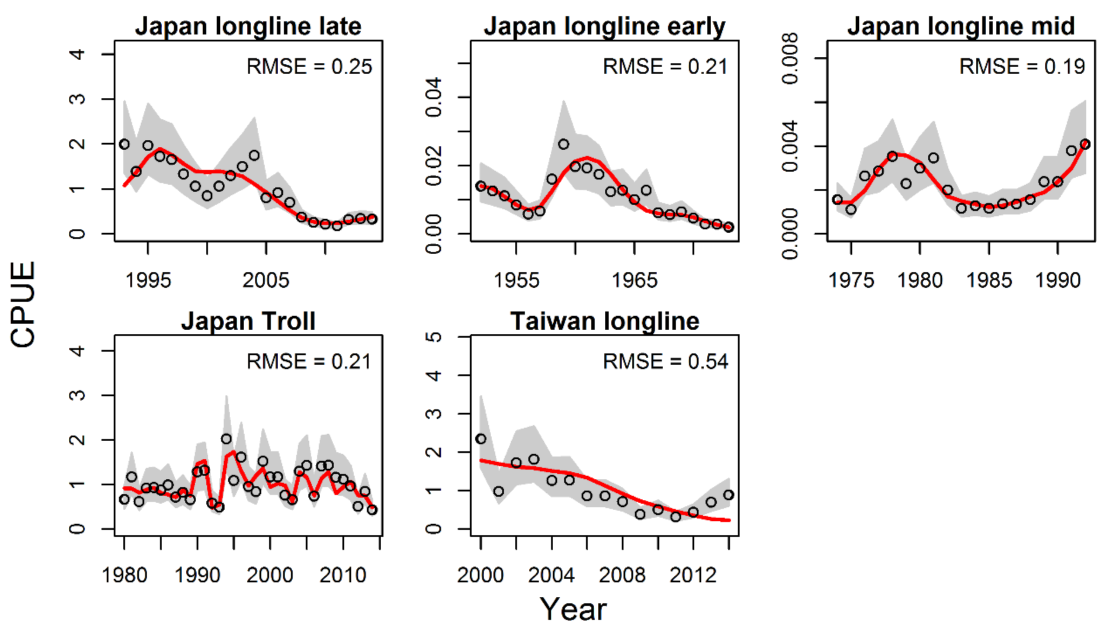



Figure 4. (a) The observed proportion-at-size (gray shaded area) and overall expected fit (red line) for the reference model, where fleets with size compositions fit to the model are shown. (b) The observed CPUE (open circles with gray shaded area) and expected fit (red line) for the reference model, where fleets with CPUE data fit to the model are shown expect for Taiwan longline fishery. The structure for initial model is summarized in Table 1.

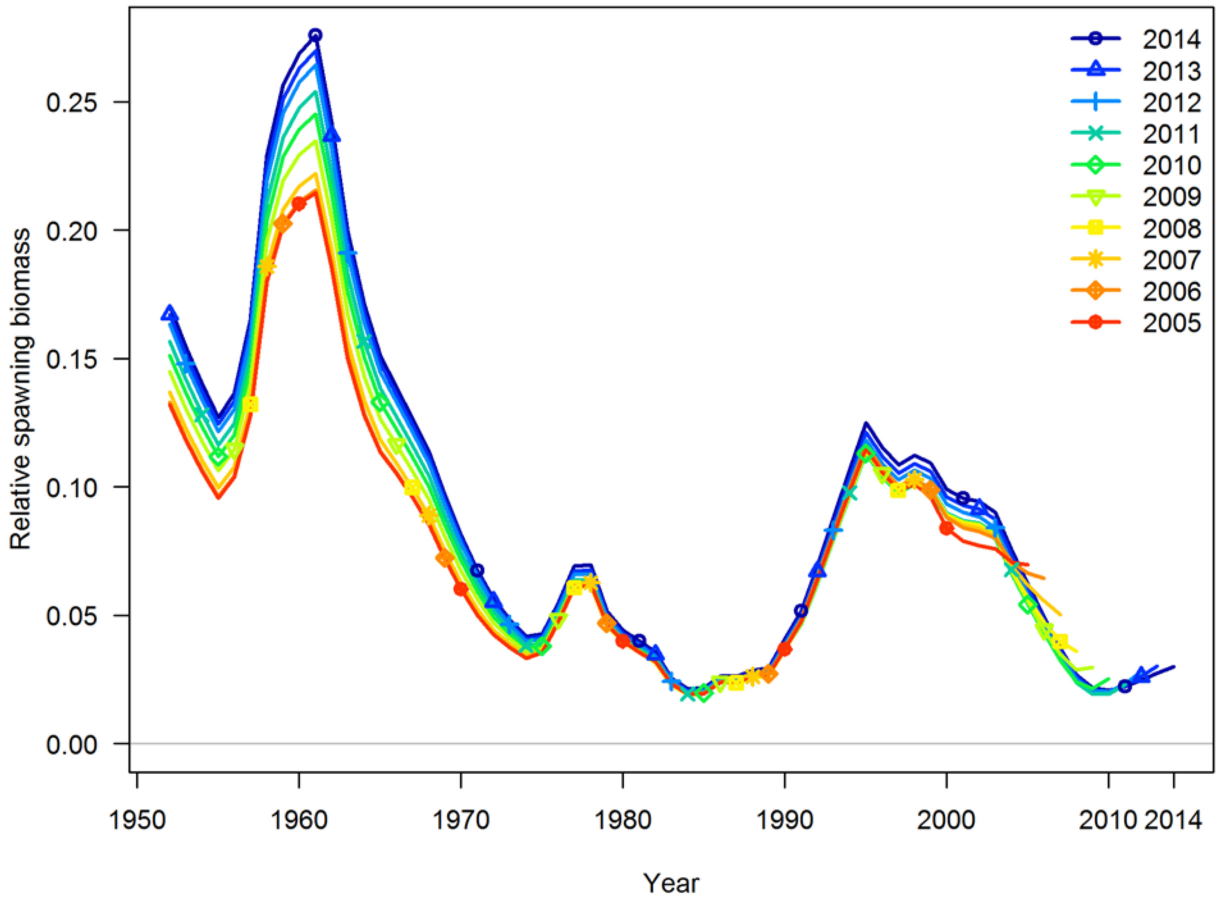



Figure 5. Nine-year retrospective analysis of relative spawning biomass to its unfished spawning biomass from the reference model. Each line represents a model fit with sequentially one less year of data.