*J. Agric. Sci., Tokyo Univ. Agric., 61 (1), 17–30 (2016)*
東京農大農学集報，61（1），17–30（2016）

論　文
Articles

# An Evaluation of Initial Weighting Factors for Length Composition Data Used in Integrated Stock Assessment Models

By

Ayumi SHIBANO*†, Minoru KANAIWA**, Yukio ISHIHARA***, Ryosuke UJI****, Tsuyoshi SHIMURA****, Yukio TAKEUCHI***** and Kotaro YOKAWA*****

(Received August 18, 2015/Accepted October 23, 2015)

**Summary**：In stock analysis using integrated age-structured models, weighting factors of each likelihood component (e.g., for abundance indices and size composition data) directly affect the estimation of the model. In a past stock assessment of Pacific bluefin tuna (PBF) conducted in 2012, conflicts were recognized between the length composition dataset of a purse seine fishery operating in the Sea of Japan (PS-SoJ) and the abundance indices derived from other fisheries. After careful consideration, the initial weighting factor used in iterative reweighting method for the length composition dataset of PS-SoJ was reduced. In this paper, we provide a further analysis of procedure for determining the initial weighting factor of the likelihood component for the length composition dataset of PS-SoJ used in the stock assessment model of PBF. We tested five scenarios involving alternative initial weighting factors. Firstly, we estimated an effective sample size as an initial weighting factor considering inherent accuracy and precision of the length composition dataset in the context of cluster sampling. Next, we illustrated the partial likelihood of estimated key parameter (i.e. $R_0$) as indicator of model fit for the tested scenarios to suggest optimal methods. In the results, the model fit was improved in a scenario where the point estimate of the effective sample size except for highly uncertain years was used. The differences between the statistical point estimates of stock dynamics in each scenario would have significant effect on management considering PBF's situation. In conclusion, we discussed appropriate methods of setting initial weighting factors for a variety of data conditions.

***Key words***：stock assessment, weighting factor, effective sample size, length composition, Pacific bluefin tuna

## 1. Introduction

Pacific bluefin tuna (*Thunnus orientalis*, PBF) is one of the most commercially important species in the world because of its global economic importance and intensive international trade[1]. The PBF stock assessment working group (PBFWG) established by the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean (ISC ; http://isc.ac.affrc.go.jp/) provides the details of PBF fisheries in their stock assessment report[2,3]. Japan has a long history of PBF fishing activity going back to pre-historic times. In recent years (1952–2012), Japanese total annual catches have fluctuated between 34,000 MT (in 1956) and 6,000 MT (in 1990 ; Fig. 1)[4]. PBF is caught by various fisheries, and catch data of as many as 25 fisheries in 5 countries are reported to ISC[4]. Among those fisheries, the total catches of all purse seine fisheries in Japan accounts for a large proportion of the worldwide total catch (34.7% ; Fig. 1). Especially, the catch of a Japanese purse seine fishery in

* Department of Bioindustry, Graduate School of Bioindustry, Tokyo University of Agriculture
** Department of Aquatic Bioscience, Faculty of Bioindustry, Tokyo University of Agriculture
*** Tottori Prefectural Fisheries Experimental Station
**** Tottori Prefectural Fisheries Division
***** National Research Institute of Far Seas Fisheries
† Corresponding author (E-mail : ayumi.shibano@gmail.com)

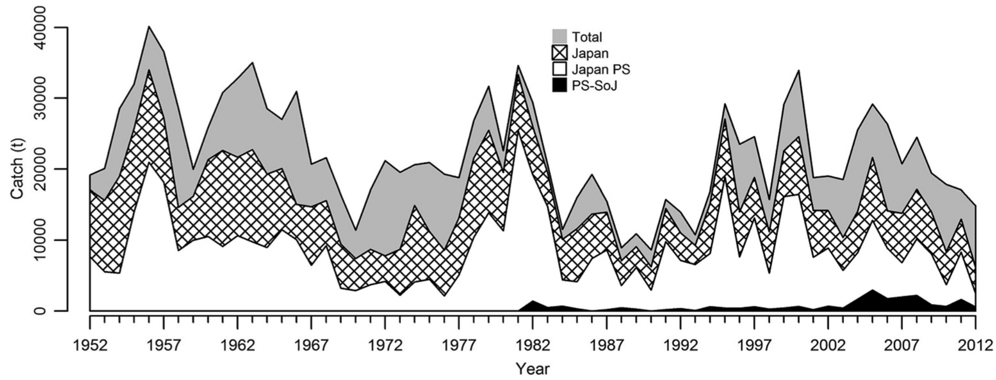


**Fig. 1** Annual catches of Pacific bluefin tuna by the purse seine fishery in the Sea of Japan (PS-SoJ), all purse seine fisheries in Japan (Japan PS), other fisheries in Japan (Japan) and other countries (Total). Partially revised from ISC[4].

the Sea of Japan (PS-SoJ) has been increasing since 2004, and it accounts for the sixth largest catch (7.1%) in the world on the average from 2004 and afterwards[4].

The largest proportion of PBF caught by PS-SoJ is landed at Sakai-minato port. Length composition data for this fishery have been collected at this port by the Tottori Prefectural Fisheries Experiment Station since 1983 (except for 1986 and 1990 when PBF were not caught). The length composition dataset of PS-SoJ is highly suitable as a sample from the entire exploitable stock because it has the largest sample size and high sample coverage, which means high sample coverage to the total catch in the world due to the large catch amount of this fishery (Table 1 and Fig. 1). The PBFWG agreed the length composition dataset of PS-SoJ to be of the highest quality among all fleets exploiting PBF because it contained high representativeness and long time series[2].

In the size sampling of fisheries research, samples are generally drawn from landings which are clusters selected by fishing operations. This type of selection is called cluster sampling[5-7]. Evaluations of the accuracy and precision of the length composition dataset obtained from each fishery should therefore be conducted carefully when used in stock analyses, considering the nature as a cluster sampling fishery.

One measure of accuracy and precision of size composition is effective sample size (ESS), as it shows the amount of information on the population which the sample represents[8]. PENNINGTON *et al.*[6] focused on similarity in samples caught in a single operation and proposed a variance ratio method to estimate ESS based on sampling variance. The method of ESS was applied to the length composition dataset of Northeast Arctic cod (*Gadus mohua*), Northeast Arctic haddock (*Melanogrammus aeglefinus*), and Namibian deepwater hake (*Meluccius paradoxus*) generated by trawl surveys[6]. The result showed that ESS was substantially less than the actual sample size[6] due to the effect of similarity which decreased ESS, thus decreasing accuracy and precision in the length composition of these species that were caught[6,7]. The method presented by PENNINGTON *et al.*[6] was also applied to Pacific Ocean perch (*Sebastes alutus*), and the result in this case suggested as well that a consideration of similarity was important in estimating a reasonable ESS[9,10].

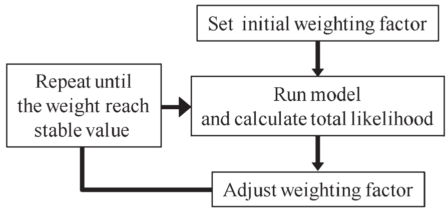


**Fig. 2** Illustration of weighting procedure for length composition data

Stock Synthesis 3 (SS3)[11], an integrated age-structured stock assessment model, was used in the PBF stock assessment of 2012[2]. In the analysis of SS3, each likelihood component (e.g., stock abundance indices and size composition data) is weighted to calculate the total likelihood. Determination of the weighting factors thus affects the stock assessment result directly[12], and METHOT and WETZEL[11] cautioned : "Correct relative weighting among likelihood components is crucial to attain good model performance and good estimates of the variances of the model results."

Several iterative reweighting methods have been devised to realize correct weighting[12,13]. In these methods, weighting is conducted through two stages as shown in Fig. 2. An interpretation of the two stages of data weighting is given in subsection 2.2.

**Table 1** Summary of the Pacific bluefin tuna purse seine fishery in the Sea of Japan and summary statistics for the accuracy and precision of the length composition dataset between 1987 and 2012 (except for 1990 due to no catches of PBF).

| Year | Fishing season | $n$ | $M$ | $m$ | $m/M$ | $m/n$ | $\hat{m}_{eff}$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Median | (95% CI) | Median/$m$ | Median/$n$ |
| 1987 | Aug 10 - Aug 18 | 2 | 1,419 | 791 | 0.557 | 395.5 | 110 | (17 - 195,920) | 0.139 | 54.9 |
| 1988 | Jun 26 - Aug 7 | 10 | 3,539 | 2,006 | 0.567 | 200.6 | 19 | (11 - 74) | 0.010 | 1.9 |
| 1989 | Jun 30 - Aug 5 | 4 | 2,395 | 1,166 | 0.487 | 291.5 | 30 | (7 - 449) | 0.026 | 7.5 |
| 1990 | - | - | - | - | - | - | - | - | - | - |
| 1991 | Jul 19 - Aug 9 | 7 | 2,024 | 1,300 | 0.642 | 185.7 | 9 | (5 - 33) | 0.007 | 1.2 |
| 1992 | Jul 4 - Aug 5 | 6 | 2,913 | 2,220 | 0.762 | 370.0 | 8 | (4 - 178) | 0.004 | 1.4 |
| 1993 | Jul 14 - Jul 23 | 3 | 1,801 | 1,284 | 0.713 | 428.0 | 4 | (2 - 7,209) | 0.003 | 1.2 |
| 1994 | Jun 25 - Jul 4 | 4 | 9,608 | 1,935 | 0.201 | 483.8 | 237 | (77 - 4,598) | 0.123 | 59.3 |
| 1995 | Jul 25 - Jul 31 | 3 | 3,508 | 1,035 | 0.295 | 345.0 | 16 | (7 - 3,915) | 0.015 | 5.3 |
| 1996 | Jun 22 - Jul 25 | 6 | 4,238 | 2,772 | 0.654 | 462.0 | 774 | (253 - 5,311) | 0.279 | 129.0 |
| 1997 | Jul 7 - Aug 29 | 8 | 3,955 | 1,902 | 0.481 | 237.8 | 66 | (31 - 778) | 0.035 | 8.2 |
| 1998 | Jun 1 - Aug 5 | 7 | 4,265 | 2,240 | 0.525 | 320.0 | 6 | (4 - 132) | 0.003 | 0.9 |
| 1999 | Jul 30 - Aug 12 | 8 | 6,129 | 3,333 | 0.544 | 416.6 | 17 | (9 - 533) | 0.005 | 2.1 |
| 2000 | Jul 17 - Aug 9 | 10 | 7,548 | 3,775 | 0.500 | 377.5 | 33 | (14 - 378) | 0.009 | 3.3 |
| 2001 | Jul 10 - Jul 26 | 5 | 2,193 | 1,365 | 0.622 | 273.0 | 164 | (55 - 2,864) | 0.120 | 32.7 |
| 2002 | Jul 6 - Jul 30 | 13 | 5,976 | 3,190 | 0.534 | 245.4 | 31 | (15 - 155) | 0.010 | 2.4 |
| 2003 | Jul 3 - Aug 11 | 14 | 6,649 | 2,895 | 0.435 | 206.8 | 22 | (9 - 75) | 0.008 | 1.6 |
| 2004 | Jun 27 - Aug 12 | 36 | 27,102 | 9,122 | 0.337 | 253.4 | 30 | (19 - 44) | 0.003 | 0.8 |
| 2005 | Jun 17 - Aug 7 | 55 | 47,120 | 15,626 | 0.332 | 284.1 | 102 | (71 - 159) | 0.007 | 1.9 |
| 2006 | Jun 12 - Aug 9 | 50 | 19,418 | 10,814 | 0.557 | 216.3 | 85 | (62 - 128) | 0.008 | 1.7 |
| 2007 | Jun 11 - Aug 2 | 48 | 41,911 | 17,073 | 0.407 | 355.7 | 51 | (32 - 85) | 0.003 | 1.1 |
| 2008 | Jun 3 - Aug 11 | 58 | 44,500 | 19,961 | 0.449 | 344.2 | 74 | (57 - 104) | 0.004 | 1.3 |
| 2009 | June 8 - Aug 21 | 28 | 16,513 | 2,328 | 0.141 | 83.1 | 20 | (9 - 47) | 0.008 | 0.7 |
| 2010 | Jun 1 - Jul 23 | 26 | 17,749 | 3,325 | 0.187 | 127.9 | 45 | (20 - 101) | 0.014 | 1.7 |
| 2011 | May 29 - Aug 3 | 41 | 40,605 | 20,603 | 0.507 | 502.5 | 48 | (29 - 87) | 0.002 | 1.2 |
| 2012 | Jun 2 - Aug 8 | 27 | 8,959 | 8,381 | 0.935 | 310.4 | 57 | (37 - 151) | 0.007 | 2.1 |

$n$: Number of landings
$M$: Number of fish landed
$m$: Actual sample size
$\hat{m}_{eff}$ : Effective sample size assuming the true exploitable stock
CI: Confidence interval

In the stock assessment of PBF conducted in 2012, the PBFWG discussed an iterative reweighting method that used a multinomial likelihood function to determine the weighting factor for size composition data[2, 13, 14]. In this method, the adjusted weighting factor for size composition data is internally estimated[11) . The initial weighting factors were determined externally using the method presented by Pennington *et al.*[6]. In particular, we should note that the PBFWG set an upper limit of initial weighting factor at 51.2 in order to reduce the influence of an unusually high initial weighting factor[2].

Meanwhile, the catch per unit effort (CPUE) time series derived from longline fisheries targeting spawning aggregation and troll fisheries targeting age-0 and age-1 fish, which were different age groups from PS-SoJ that targeted age-2 or older fish, were used as abundance indices in the stock assessment[2]. In this case, each fishery represents the dynamics of different ages. The PBFWG recognized conflicts between these two CPUE series and the length compositions of PS-SoJ, and these conflicts influenced the assessment results.

The PBFWG generally agreed with the importance of CPUE, because it was direct information on stock trend, whereas length composition data was indirect information on it. Consequently the PBFWG reduced the initial weighting factors for the length composition dataset of PS-SoJ in the base-case stock assessment run[2]. Considering the relatively high representativeness of PS-SoJ's length composition dataset, further exploration of weighting methods would contribute the stock analysis.

MAUNDER[15] noted that the maximum likelihood approach[13] ignores the uncertainty of ESS. A key for resolving conflicts between PS-SoJ and abundance indices may exist in the uncertainty of the initial weighting factor of PS-SoJ. Several elements such as number of landings and inconsistency between model assumptions may deteriorate the model fit.

Firstly, with only a few landings there will, by chance, be years in which all landings happen to contain fish of similar size. In these years, estimates of ESS presented by PENNINGTON *et al.* will be much too high and imprecise. It is thus difficult to estimate ESS when the number of landings is small. However, the minimum number of landings that must occur to ensure a low uncertainty in ESS is unknown.

A good example of inconsistency between model assumptions related to length composition is differences of selectivity assumption. If we assume a constant selectivity for a fishery, whereas the fishery actually has an age preference, this increases the uncertainty of ESS. Further description of this point is provided in subsection 2.2.

These factors invalidate the use of ESS presented by PENNINGTON *et al.*[6] as the initial weighting factor of length composition for these years. Instead, giving a zero weight to the dataset with highly uncertain ESS enables the exclusion of the cases in which inconsistency of model assumptions exist. It is therefore meaningful to explore the influence of giving a zero weight on model fit and estimated stock dynamics.

Uncertainty in ESS is important when it is used as a weighting factor in model analyses that use SS3, as SS3 outputs and the good model fit are dependent on interannual variability of initial weighting factors. The uncertainty of ESS should be checked when it is used as an initial weighting factor, and the confidence interval of ESS can be used to measure uncertainty. If the confidence interval is wider than a certain level, the point estimate of ESS is not very useful as an indicator to show the annual trend of accuracy and precision, and it should be treated carefully considering its uncertainty. It is therefore essential to examine the effect of similarity and number of landings on the uncertainty of an estimated ESS.

In this paper, we provide a procedure for determining the initial weighting factor of the likelihood component for the length composition dataset of PS-SoJ used in the stock assessment model of PBF. We tested five scenarios involving alternative initial weighting factors. Firstly, we estimated an ESS as an initial weighting factor considering inherent accuracy and precision of the length composition dataset in the context of cluster sampling. We observed the effect of the initial weighting factors using profiles of partial likelihoods of abundance indices and size composition data. Finally, we discuss appropriate methods of setting initial weighting factors for a variety of data conditions to extend this study to other fisheries and assessments.

## 2. Materials and Methods

### 2.1. Model inputs and settings

We used the same input data and model settings as in the 2012 stock assessment of PBF[2] except for the initial weighting factor for the length composition dataset of PS-SoJ. Abundance indices from 5 fisheries, size composition data from 12 fisheries including PS-SoJ were used to calculate the total likelihood. The summary of PS-SoJ is shown in Table 1, and a description of data collection for this fishery is provided below.

We conducted various simulations of setting different initial weighting factors for the length composition dataset of PS-SoJ, which is one of the size composition datasets used in stock assessments. Length composition data for which the landing date, vessel, and the number of fish in the landing could be identified were used for our analysis. These samples were available from catches landed from 1987 to 2012 (except for 1990 when there were no catches of PBF). Samples were collected at Sakaiminato port mainly by the Tottori Prefectural Fisheries Experiment Station. In recent years, staff from the National Research Institute of Far Seas Fisheries supported these sampling activities. Samples were measured for fork length to the nearest 1 cm.

PS-SoJ captures one school of similar sized fish and lands the school immediately. Fish unloaded by a carrier vessel in a day was defined as a landing. Samples were drawn from each landing, and sample sizes varied by landing. Since the sample size was large enough to expect good coverage across length range in each landing, in this study we considered a landing as a cluster. Length measurement was conducted in almost all landings. When catch amounts were large, carrier vessels unloaded their catches in two consecutive days. In this analysis, the length composition datasets of the same vessel unloading a catch over two consecutive days were combined and treated as a single landing.

Although the fishery season of PS-SoJ targeting PBF is usually from June to July, one exceptional catch of 5.6 MT was observed in October 1997. It is known that the growth of PBF is rapid during summer[16]. From August to October in the Sea of Japan, it is possible that length composition changes due to the growth of fish. It is also possible that fish from other stocks flow into the exploitable stocks of PS-SoJ. For these reasons, the stocks of June to July and October cannot be considered the same, and catches in October were therefore excluded from the dataset in this analysis. In addition, juveniles, called *yokowa*, were excluded from this analysis for three reasons : (1) There was the possibility that juveniles were caught at different fishing grounds from adults and could therefore form different exploitable stocks, (2) since juveniles were not a major target for this observation program, the sampling effort was totally different and varied by landing, and (3) since yokowa are categorized as catches by small purse seiners in the stock assessment, the ESS for yokowa should be calculated separately.

### 2.2. Populations and structure of errors to consider

Correct weighting requires knowledge of the error configuration of our observations[12]. There are four types of error to consider regarding the analysis of proportion at length (Fig. 3). These errors arise because there are four different values : (i) the value in the true population that was accessible to the fishery, $P_{TRUE}$, (ii) the value observed in sample from the true population, $P_{OBSERVED}$, (iii) the value in the model population as a mimic of $P_{TRUE}$ estimated by the population dynamics assumptions, $P_{MODEL}$, and (iv) the value as a mimic of $P_{OBSERVED}$ estimated by the fishery assumptions, $P_{EXPECTED}$.

Firstly, we obtain $P_{OBSERVED}$ from the true population. $P_{OBSERVED}$ does not completely match $P_{TRUE}$ because of observation error, which is the error whose distribution, and likely size, researchers may infer from sampling method (e.g., sample size, number of landings, landing amount and/or representativeness of each landing). This type of error corresponds to precision. Secondly, $P_{MODEL}$ is constructed to mimic $P_{TRUE}$ and estimated from $P_{OBSERVED}$. $P_{MODEL}$ differs from $P_{TRUE}$ because of the simplified assumptions related to population dynamics, and this difference is called process error. This type of error has an influence on accuracy. Thirdly, $P_{EXPECTED}$ is calculated from $P_{MODEL}$ using the simplified selectivity assumptions. This is another source of process error which arises from the difference between the selectivity assumption and the true selectivity. The error we are interested in (i.e., that described in likelihood) is that between $P_{OBSERVED}$ and $P_{EXPECTED}$. This is called total error because it is the sum of observation error and process error.

This characterization of errors gives an obvious interpretation to the two stages of data weighting. As shown in Fig. 2 and 3, initial weighting factors are devised before the model is run. Weights appropriate for observation error are assigned at this stage. Initial weighing factor assumes that the sample represents the population which is accessible to the fishery, and this weighting factor is hereafter referred to as $ESS_{TRUE}$. $ESS_{TRUE}$ reflects the precision of each year, in other words, relative inter-annual variability of the likely size of observation error which is caused by inter-annual variability of the sampling method.

At stage 2, those weighting factors are adjusted to

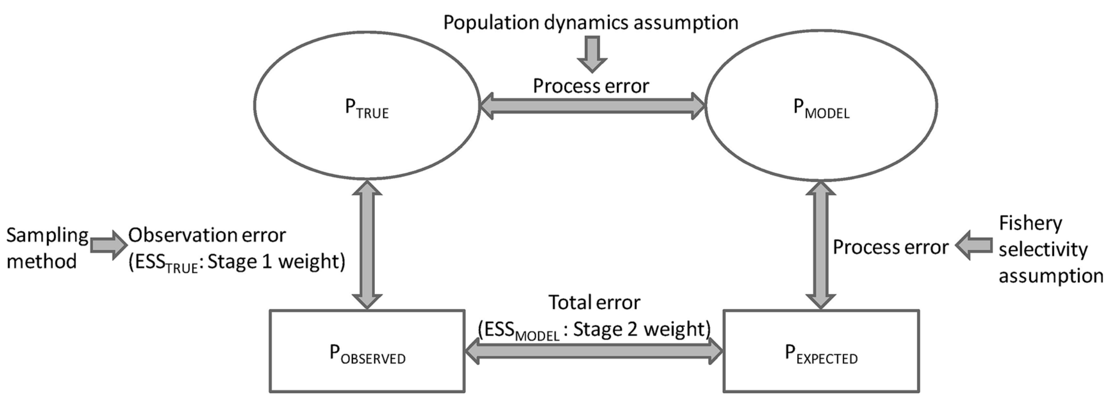


**Fig. 3** Schematic illustration of the four types of error that exist between the four values of proportion at length : (i) the value in the true population, $P_{TRUE}$, (ii) the value observed in sample from the true population, $P_{OBSERVED}$, (iii) the value expected by the population dynamics model as a mimic of $P_{TRUE}$, $P_{MODEL}$, and (iv) the value expected by the fishery model as a mimic of $P_{OBSERVED}$, $P_{EXPECTED}$. Partially revised from Francis[12].

allow for process error and reflect the accuracy in. The weighting adjustment occurs after the model has been run, or sometimes during a model run, and is intended to make the weighting factors more consistent with the model outputs. Thus adjusted weighting factor is hereafter referred to as $ESS_{MODEL}$. $ESS_{MODEL}$ resulting from iterative methods is the indicator of both accuracy and precision. The weighting adjustments usually apply to whole datasets, rather than individual data points. FRANCIS[12] provided an iterative reweighting method that can be used to calculate $ESS_{MODEL}$ based on the concept of $ESS_{TRUE}$ presented by PENNINGTON *et al.*[6].

### 2.3. Evaluation of accuracy and precision

Applying the principle described above, the accuracy and precision of the length composition dataset of PS-SoJ, $P_{OBSERVED}$, is defined as the amount of information on the length composition of the model population that was accessible to this fishery, $P_{EXPECTED}$. The precision of the length composition datasets was evaluated using $ESS_{TRUE}$ which estimates the likely size of observation error produced by cluster sampling. Accuracy of the dataset was evaluated by the weighting adjustment to allow for process error, resulting in $ESS_{MODEL}$ as a reflection of total error and an indicator of both accuracy and precision. Annual estimates of $ESS_{TRUE}$ of the length composition dataset of PS-SoJ were calculated with the assumption that the samples within a year were selected from the same stock.

PENNINGTON *et al.*[6] proposed a variance ratio method to estimate $ESS_{TRUE}$, denoted by the symbol $\hat{m}_{eff}$, so that sampling variability based on simple random samples would be the same as the variability based on the cluster sample. The more similar a cluster is, the smaller $\hat{m}_{eff}$ becomes. This method was applied to the length composition dataset of PS-SoJ to evaluate the accuracy and precision of the dataset, since purse seining is essentially a form of cluster sampling.

In the estimation of $ESS_{TRUE}$ assuming cluster sampling, similarity within a cluster is important, as strong similarity decreases the accuracy and precision of the cluster samples for the whole year[5]. PS-SoJ fleets capture one school of fish and land it immediately, indicating that a single landing comprises a single school. The similarity of a landing can be scaled by the length range. When we quantify the length range of a landing, it is appropriate to use measures such as quantile deviation to exclude outliers, since it is possible a few fish that cannot find a school of their own size/age join a school of a different size/age. We observed the quantile deviation of each landing to investigate the effect of similarity on $ESS_{TRUE}$.

Next, we applied the variance ratio method to estimate $\hat{m}_{eff}$ for the length composition dataset of PS-SoJ. Following the notation from PENNINGTON *et al.*[6], a ratio estimator of the mean length of the true exploitable stock, $\hat{S}$, is calculated by

$$\hat{S}=\frac{\sum_{i=1}^{n} M_i \hat{\mu}_i}{\sum_{i=1}^{n} M_i}, \tag{1}$$

where $M_i$ is the number of fish in landing $i$, and $\hat{\mu}_i$ is the mean length of the sample in landing $i$. The estimated variance of the ratio estimator $\hat{S}$ is approximately

$$\mathrm{var}\,(\hat{S})=\sum_{i=1}^{n}\frac{(M_i/\overline{M})^2(\hat{\mu}-\hat{S})^2}{n(n-1)}, \tag{2}$$

where $\overline{M}=\sum_{i=1}^{n} M_i/n$.

The variance of the length composition if the samples are selected randomly, $\hat{\sigma}_x^2$, is estimated as

$$\hat{\sigma}_x^2=\frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i}(M_i/m_i)(x_{i,j}-\hat{S})^2}{M-1}, \tag{3}$$

where $M=\sum_{i=1}^{n} M_i$ is the total number of fish landed in a year, and $x_{i,j}$ is the length of the $j$th fish in landing $i$.

In random sampling, $\hat{\sigma}_x^2/m=\mathrm{var}(\hat{S})$. Then $\hat{m}_{eff}$ is

$$\hat{m}_{eff}=\frac{\hat{\sigma}_x^2}{\mathrm{var}\,(\hat{S})}. \tag{4}$$

### 2.4. Resampling structure using the bootstrap method

To investigate uncertainty of the $ESS_{TRUE}$, we used the bootstrap percentile method[17] to estimate the 95% confidence interval of $ESS_{TRUE}$. Resampling was conducted with consideration given to single-stage cluster sampling[5]. The bootstrap replicates consisted of two primary steps to imitate the structure of actual sampling from the true exploitable stock. This data generation process included (1) selection of the surveyed landing and (2) selection of measured fish within the landing. These steps were then followed by (3) compilation of simulated datasets of landings in each year and (4) estimation of $ESS_{TRUE}$ from the simulated annual dataset that was compiled. The compilation of the simulated annual dataset and $ESS_{TRUE}$ were obtained in each bootstrap trial of length measurements. These steps are illustrated in Fig. 4.

In the first step of resampling, all surveyed landings were resampled with replacements within each year. In the second step, we evaluated the effect that the selection of measured fish and the coverage within a landing had on accuracy and precision. Lengths were resampled with replacements while maintaining the original sample size within each landing.

We compared the width of the $ESS_{TRUE}$ confidence interval and the fluctuation range of the annual medians of $ESS_{TRUE}$. We define wide confidence interval for a year as

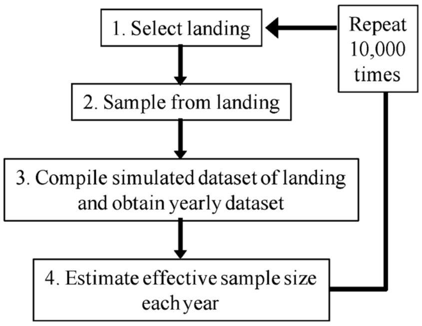


**Fig. 4** Illustration of the resampling methodology using the bootstrap method to investigate the uncertainty of effective sample size.

that wider than the fluctuation range of the medians because such a wide confidence interval would make the annual trend of accuracy and precision uncertain. The bootstrap trial was repeated 10,000 times, and the confidence intervals were estimated using the percentile method through 10,000 runs.

### 2.5. Examining model fit

We tested five scenarios to evaluate the effect of the different initial weighting factors on the estimated stock dynamics and model fit. If the dataset contains internally similar clusters, estimating annual $ESS_{TRUE}$ would require the consideration of two factors : the number of clusters ($n$) and the magnitude of similarity within each cluster. Scenarios 1–3 consider both $n$ and the magnitude of similarity. In scenario 1, the point estimates (without bootstrap) of $ESS_{TRUE}$ multiplied by 0.5 were used as initial values, and an upper limit of 51.2 was set. These are the same values that were used in the 2012 stock assessment[2)]. The method of scenario 1 down-weights datasets with unusually high $ESS_{TRUE}$ but gives the heaviest weight to the dataset via the upper limit. In scenario 2, the point estimates of $ESS_{TRUE}$ were used without any upper limit. This approach is appropriate when the uncertainty of $ESS_{TRUE}$ is sufficiently small in all the years. In scenario 3, the point estimates of $ESS_{TRUE}$ were used in general, and a zero weight was set for a dataset if its $ESS_{TRUE}$ had a wide confidence interval. The criterion of the wide confidence interval follows the description in the preceding subsection 2.4. The approach of scenario 3 is applicable when the size composition datasets contain a dataset for which the accuracy and precision cannot be quantified. In scenario 4, the number of landings ($n$) is used. This approach is appropriate when the magnitude of similarity within each cluster does not vary. In scenario 5, a weight of 1 is set for all the datasets, assuming that there is no difference in ESS across the years. The values in each scenario are shown in Table 2.

**Table 2** Initial weighting factors for the length composition dataset of the purse seine fishery in the Sea of Japan for each initial weighting factor scenario.

| Year | Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 | Scenario 5 |
|---|---|---|---|---|---|
| 1987 | 12.19 | 24.38 | 0 | 2 | 1 |
| 1988 | 8.55 | 17.11 | 17.11 | 10 | 1 |
| 1989 | 12.46 | 24.91 | 24.91 | 4 | 1 |
| 1990 | - | - | - | - | - |
| 1991 | 2.99 | 5.98 | 5.98 | 7 | 1 |
| 1992 | 2.47 | 4.94 | 4.94 | 6 | 1 |
| 1993 | 1.22 | 2.44 | 0 | 3 | 1 |
| 1994 | 51.20 | 137.98 | 0 | 4 | 1 |
| 1995 | 7.26 | 14.52 | 0 | 3 | 1 |
| 1996 | 51.20 | 754.84 | 0 | 6 | 1 |
| 1997 | 23.21 | 46.43 | 46.43 | 8 | 1 |
| 1998 | 2.60 | 5.21 | 5.21 | 7 | 1 |
| 1999 | 7.94 | 15.88 | 15.88 | 8 | 1 |
| 2000 | 15.66 | 31.32 | 31.32 | 10 | 1 |
| 2001 | 51.20 | 127.24 | 0 | 5 | 1 |
| 2002 | 11.42 | 22.84 | 22.84 | 13 | 1 |
| 2003 | 9.78 | 19.56 | 19.56 | 14 | 1 |
| 2004 | 13.58 | 27.16 | 27.16 | 36 | 1 |
| 2005 | 51.20 | 102.41 | 102.41 | 55 | 1 |
| 2006 | 41.13 | 82.26 | 82.26 | 50 | 1 |
| 2007 | 22.87 | 45.74 | 45.74 | 48 | 1 |
| 2008 | 35.68 | 71.36 | 71.36 | 58 | 1 |
| 2009 | 8.95 | 17.89 | 17.89 | 28 | 1 |
| 2010 | 22.64 | 45.28 | 45.28 | 26 | 1 |
| 2011 | 23.77 | 47.55 | 47.55 | 41 | 1 |
| 2012 | 27.64 | 55.28 | 55.28 | 27 | 1 |

The same approach shown in scenario 1 was applied to the commercial purse seine in the eastern Pacific Ocean (PS-EPO). This fishery and PS-SoJ were treated as reference fisheries because both fisheries were considered by the PBFWG to have good sampling program for the size composition data.

The initial weighting factors of the other fisheries were calculated in two steps following the stock assessment conducted in 2012[2)]. Firstly, actual sample size was used in general, and the maximum initial weighting factor was set to 200 (i.e., the initial weighting factor was 200 if the actual sample size was larger than 200). Secondly, the actual sample size of each fishery was scaled as the relative weights between years whose average equivalent to the average ESS of the reference fisheries (i.e., PS-SoJ and PS-EPO).

We estimated ESS assuming the model exploitable stock, $ESS_{MODEL}$, using one of the iterative reweighting methods of FRANCIS[12] to apply the weight adjustment factor $w$ for each scenario (see equation TA 1.8 of FRANCIS[12]). Reweighting was repeated until the estimate of $w$ converged to 1. The optimal values of virgin recruitment ($R_0$) in the tested scenarios were estimated using $ESS_{MODEL}$.

When we decide which method of calculating ESS is appropriate for use in actual stock assessment, it is necessary to confirm that the likelihood is maximized around the optimal $R_0$[18] and to compare the likelihoods between the scenarios of different ESSs, because the best model is the one with the highest likelihood[19].

The likelihood profile was calculated by varying $R_0$ as a fixed value and re-estimating the other parameters for each scenario. The parameter $R_0$ was changed from 9.1 to 10.1 in steps of 0.2, because the optimal $R_0$ was 9.6 for all scenarios. We chose the partial likelihood profile of $R_0$ as an indicator of model fit, because $R_0$ is a key parameter that can scale the biomass size which is generally of interest in stock assessments.

The stock assessment duration was from 1952 to 2012. We compared the estimates of SSB among the scenarios to evaluate the influence of $ESS_{MODEL}$ on the stock assessment results.

## 3. Results

The quantile deviations of fish lengths within each landing each year as a measure of the similarity are shown as histograms (Fig. 5). The quantile deviations of fish lengths from most landings were below 10 cm, inferring that 50% of fish in a landing generally comprised a single class or were of very close year classes (Fig. 5).

A summary of PS-SoJ and statistics for the ESS assuming the true exploitable stock, $ESS_{TRUE}$, are presented in Table 1. Throughout the sampling history, samplers kept the same sample coverage (i.e., $m$ divided by $M$) which was 0.495 on average. They also retained the actual sample size per landing (i.e., $m$ divided by $n$) which was around 300 on average. From 2004, the number of vessels, the number of landings $n$, and the number of fish landed $M$ increased. The number of landings $n$ increased from 6.9 to 41.0 on average, and the annual actual sample size $m$ increased proportionally to $n$, from 2,075.6 to 11,914.8 on average. The estimates of $ESS_{TRUE}$ were substantially less than $m$ throughout the sampling history.

The 95% confidence intervals of $ESS_{TRUE}$ for each year, estimated using the bootstrap method, are shown in Table 1 and Fig. 6. The widths of confidence intervals for $ESS_{TRUE}$ were sufficiently narrow in most years and clearly showed annual fluctuation (Table 1 and Fig. 6). The confidence intervals of $ESS_{TRUE}$ in years with small $n$ (1987, 1993, 1994, 1995, 1996, and 2001) were wider than the annual fluctuation range of medians (Table 1 and Fig. 6). In these years, the values of $n$ were smaller than 6. Confidence intervals in 1989 and 1992 were not wide, although the numbers of landings were small (Table 1

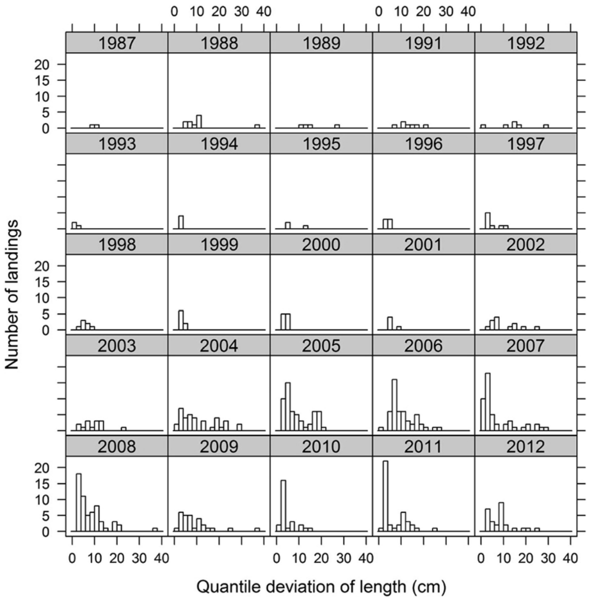


**Fig. 5** Histograms of the quantile deviations of the lengths for each year.

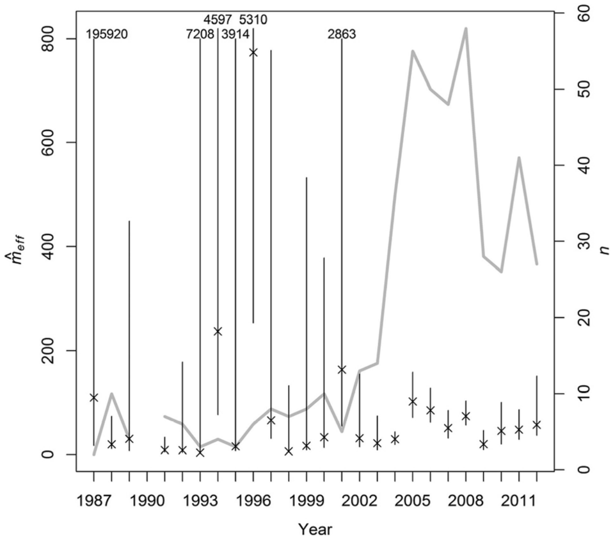


**Fig. 6** Estimated effective sample size ($\hat{m}_{eff}$) for each year. The points (×) are medians and the vertical lines are 95% confidence intervals. The numbers at the upper side are the upper limits for each year. The gray line is the number of landings ($n$) for each year.

and Fig. 6).

For each scenario, estimates of the weighting adjustment factor $w$, and the ESS assuming the model exploitable stock, $ESS_{MODEL}$, are shown in Table 3. $ESS_{MODEL}$ was smaller than initial weighting factors across all scenarios except for scenario 5.

Profiles of scaled negative log likelihoods (NLLs) based on each data component for virgin recruitment parameter $R_0$ are shown in Fig. 7. NLLs were re-scaled by subtracting the minimum values and dividing the maximum values of the NLL of each component as calculated under a different fixed $R_0$. Abundance indices from 5 fisheries were used to calculate the total likelihood in this study following the 2012 PBF stock assessment[2]. Since decomposed partial scaled NLLs of each fishery's abundance indices exhibited similar patterns in all scenarios, the likelihood profiles for abundance indices are shown as the total of 5 fisheries for each scenario (Fig. 7a).

Changing the initial weighting factor had little effect on the fit of abundance indices (Fig. 7a). The likelihood profiles for the length composition dataset of PS-SoJ exhibited remarkable differences in patterns among scenarios (Fig. 7b). Scaled NLLs in scenarios 1, 2, and 5 resulted in similar patterns, and they slightly increased around the optimal $R_0$, whereas scaled NLLs in scenarios 3 and 4 came close to the minimum. The likelihood pro-

**Table 3** Estimates of ESS assuming the model exploitable stock and estimates of adjustment factor $w$ for the length composition dataset of the purse seine fishery in the Sea of Japan for each initial weighting factor scenario.

| Year | Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 | Scenario 5 |
|---|---|---|---|---|---|
| 1987 | 2.01 | 1.41 | 0 | 0.57 | 3.43 |
| 1988 | 1.41 | 0.99 | 1.83 | 2.84 | 3.43 |
| 1989 | 2.05 | 1.44 | 2.67 | 1.14 | 3.43 |
| 1990 | - | - | - | - | - |
| 1991 | 0.49 | 0.35 | 0.64 | 1.99 | 3.43 |
| 1992 | 0.41 | 0.29 | 0.53 | 1.71 | 3.43 |
| 1993 | 0.20 | 0.14 | 0 | 0.85 | 3.43 |
| 1994 | 8.44 | 7.99 | 0 | 1.14 | 3.43 |
| 1995 | 1.20 | 0.84 | 0 | 0.85 | 3.43 |
| 1996 | 8.44 | 43.71 | 0 | 1.71 | 3.43 |
| 1997 | 3.83 | 2.69 | 4.98 | 2.28 | 3.43 |
| 1998 | 0.43 | 0.30 | 0.56 | 1.99 | 3.43 |
| 1999 | 1.31 | 0.92 | 1.70 | 2.28 | 3.43 |
| 2000 | 2.58 | 1.81 | 3.36 | 2.84 | 3.43 |
| 2001 | 8.44 | 7.37 | 0 | 1.42 | 3.43 |
| 2002 | 1.88 | 1.32 | 2.45 | 3.70 | 3.43 |
| 2003 | 1.61 | 1.13 | 2.10 | 3.98 | 3.43 |
| 2004 | 2.24 | 1.57 | 2.91 | 10.24 | 3.43 |
| 2005 | 8.44 | 5.93 | 10.98 | 15.65 | 3.43 |
| 2006 | 6.78 | 4.76 | 8.82 | 14.22 | 3.43 |
| 2007 | 3.77 | 2.65 | 4.91 | 13.66 | 3.43 |
| 2008 | 5.88 | 4.13 | 7.65 | 16.50 | 3.43 |
| 2009 | 1.47 | 1.04 | 1.92 | 7.97 | 3.43 |
| 2010 | 3.73 | 2.62 | 4.86 | 7.40 | 3.43 |
| 2011 | 3.92 | 2.75 | 5.10 | 11.66 | 3.43 |
| 2012 | 4.56 | 3.20 | 5.93 | 7.68 | 3.43 |
| $w$ | 0.16 | 0.06 | 0.11 | 0.28 | 3.43 |

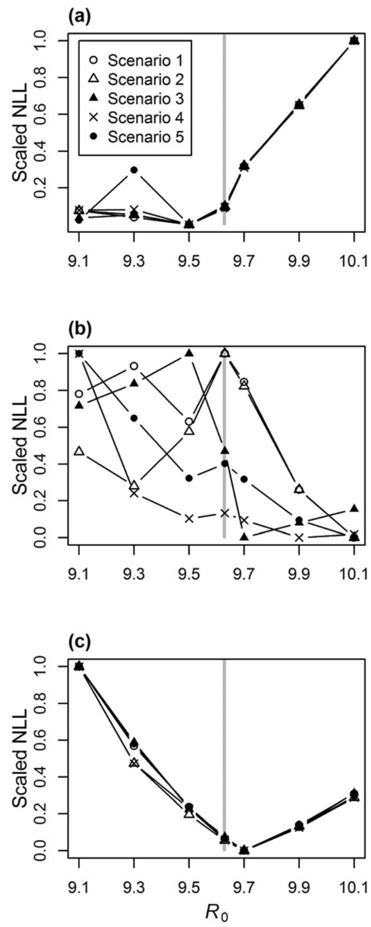


**Fig. 7** Profiles of scaled negative log likelihoods (NLLs) based on each data component for virgin recruitment parameter $R_0$. The profiles of scaled NLLs for all abundance indices used in the stock assessment conducted in 2012 (a), the length composition dataset of the purse seine fishery in the Sea of Japan (b), and all size composition datasets used in the stock assessment conducted in 2012 other than the purse seine fishery in the Sea of Japan (c), are shown. The gray line represents the position of optimal $R_0$ (9.6) for all initial weighting factor scenarios.

**Table 4** Negative log likelihoods (NLLs) of the abundance indices component and the size composition component. NLLs are calculated in different model runs based on the different weighting factor scenarios for the length composition dataset of the purse seine fishery in the Sea of Japan.

| Scenario number | Description | Indices | Size compositions | | | Total |
|---|---|---|---|---|---|---|
| | | Total | Total | PS-SoJ | Others | |
| Scenario 1 | Point estimates of $ESS_{TRUE}$ with an upper limit of 51.2. The same values as in the stock assessment of 2012. | -81 | 2,303 | 43 | 2,324 | 2,217 |
| Scenario 2 | Point estimates of $ESS_{TRUE}$ without upper limit. | -80 | 2,303 | 44 | 2,323 | 2,219 |
| Scenario 3 | Point estimates of $ESS_{TRUE}$ with the exceptions of 0 when confidence intervals of $ESS_{TRUE}$ are wide (1987, 1993, 1994, 1995, 1996 and 2001). | -84 | 2,293 | 30 | 2,327 | 2,204 |
| Scenario 4 | Number of landings ($n$). | -82 | 2,322 | 61 | 2,325 | 2,235 |
| Scenario 5 | 1 for all datasets. | -81 | 2,308 | 47 | 2,324 | 2,221 |

file for size composition datasets of other fisheries showed similar patterns in all scenarios (Fig. 7c).

NLLs in absolute terms at optimal $R_0$ are presented in Table 4. The NLL of length composition dataset of PS-SoJ in scenario 3 decreased to 30 compared with the value of 43 in scenario 1. On the other hand, the NLL in scenario 4 was 61, the largest among all scenarios.

The time series of SSB among all scenarios generally indicated different patterns, especially after the 1970s. Compared with scenario 1, SSB in scenarios 2 and 4 showed maximum differences of approximately 1,614 MT (1.7%), and SSB in scenario 3 showed a maximum difference of roughly 2,290 MT (2.3%).

## 4. Discussion

### 4.1. Methods of setting the initial weighting factor

We have shown several methods of determining the initial weighting factor considering the uncertainty of the ESS that could be used in stock assessments of PBF. Although this analysis can be conducted with virtually generated data, we used the data actually used in the real stock assessment. Our general policy is that we should have multiple options of determining the weighting factors so that we can make flexible decisions based on data conditions.

If the sample is obtained from statistically well-designed sampling, samples from each cluster will closely represent the true exploitable stock. Accuracy and precision will increase whether the sample size per cluster or the number of clusters ($n$) is increased. In this case, the actual sample size can be set as the ESS, assuming an equal weight per sample. We did not use this assumption here, however, because our data contained internally similar clusters.

Methods based on the ESS assuming the true exploitable stock, $ESS_{TRUE}$, presented by PENNINGTON *et al.*[6] were used in scenarios 1–3. These methods will be discussed in the following subsection, considering the results about the uncertainty of $ESS_{TRUE}$.

The number of landings is also a measure of the inter-annual variability of accuracy and precision, and it is a known value with certainty. We used the number of landings in scenario 4. If the magnitude of similarity varies from landing to landing, even though the number of landings is the same, weighting only by the number of landings may not be appropriate. In addition, the number of landings alone is not satisfactory to represent accuracy and precision of length composition, because relative increase of the number of landings does not always contribute to the accuracy and precision for the whole year, due to similarity of size in each landing. If each landing closely represents the true exploitable stock, the dataset should not be rejected regardless of the number of landings, in order to attain good model fit.

It is not realistic to set the ESS equally across years for PS-SoJ as in scenario 5, as inter-annual variability of accuracy and precision usually exists due to unpredictable environmental conditions and/or the presence of operational patterns of this fishery.

### 4.2. Uncertainty of $ESS_{TRUE}$

We suggested here one criterion of uncertainty of $ESS_{TRUE}$ based on confidence interval. The confidence intervals of $ESS_{TRUE}$ in six years (1987, 1993, 1994, 1995, 1996, and 2001) were wider than that in other years, an outcome that is attributable to the strong similarity in each landing and the small number of landings (Table 1 and Fig. 6).

On the other hand, the confidence interval range of $ESS_{TRUE}$ was small when the length range was relatively wide, despite the small number of landings. The differences in confidence interval ranges show that uncertainty is affected by both the similarity and the number of landings. This becomes clear when we compare the results of different years in which the numbers of landings were the same. In 1989 and 1994, for example, the number of landings was 4 (Table 1), and the length ranges in 1989 were relatively wide, whereas the ranges in 1994 were very narrow (Fig. 5). The difference in similarity resulted in different confidence interval ranges for $ESS_{TRUE}$ in each year (Table 1 and Fig. 6). Each landing in 1989 thus relatively closely represented the true exploitable stock, and the uncertainty of $ESS_{TRUE}$ was low under the assumption of constant selectivity.

Considering the uncertainty, it is obvious that the high values of $ESS_{TRUE}$ observed in 1994, 1996, and 2001 do not constitute evidence that length composition datasets were unusually accurate and precise, because the confidence intervals of $ESS_{TRUE}$ in these years were extremely wide (Table 1 and Fig. 6).

In all years other than the above six, the confidence intervals of $ESS_{TRUE}$ were sufficiently narrow and annual fluctuations of $ESS_{TRUE}$ were clear (Table 1 and Fig. 6), suggesting that the uncertainty of $ESS_{TRUE}$ was low.

In the case of 1997, there could be different opinions on whether the confidence interval was wide or narrow. Nevertheless, the important points are that the estimate of the ESS can be uncertain, and the uncertainty is dependent on the number of landings and the similarity of each landing. When the clusters are internally similar (i.e., the sample from each cluster contains only a little information on the true exploitable stock), it is important to evaluate the uncertainty of $ESS_{TRUE}$ outside the stock assessment model.

In the definition of wide confidence interval used in this study, the threshold of confidence interval could be larger and laxer if sampling process changes every year, or if age/size structure of the true exploitable stock remarkably fluctuates. In this study, annual fluctuation of ESS in PS-SoJ does not seem to be large, and the criterion of uncertainty may appropriately be applied. In other fisheries, however, there could be cases of large annual fluctuation of ESS, and examination of various criteria could be necessary.

Considering the results related to the uncertainty of $ESS_{TRUE}$, setting an upper limit as in scenario 1 is not appropriate for PS-SoJ, because the heaviest weight is given to datasets with highly uncertain $ESS_{TRUE}$. Scenario 2 is not appropriate either, because the datasets of PS-SoJ contained several years with highly uncertain $ESS_{TRUE}$. The method of scenario 3 is thus the best choice, because it allows for the uncertainty of $ESS_{TRUE}$.

For many of the years where the ESS is set to zero in scenario 3, the calculated $ESS_{TRUE}$ is far higher than it should reasonably be. These high values could have been caused by the change of selectivity for this fishery. One of the simplest ways to address this problem is to remove these length composition data from input of the model, as demonstrated in scenario 3. However, simply omitting them may be unsatisfactory. The most reasonable method would be combination of scenario 3 and 4, which means using $ESS_{TRUE}$ of Pennington *et al.*[6] in general, applying the number of landings as a correction term. This would address the case of unusually high ESS when the number of landings was small. We did not examine this method because, related to the criterion whether the confidence interval of $ESS_{TRUE}$ was wide or narrow, there would be many different approaches to combine $ESS_{TRUE}$ and the number of landings. As a common factor throughout all approaches, we should examine the uncertainty and the reasonability of ESS considering the characteristics of the species and fishery.

### 4.3. Effect on model fit

As mentioned in the Introduction, there were some conflicts of model fit between the length composition dataset of PS-SoJ and abundance indices in the 2012 stock assessment[2]. These conflicts can also be seen in the results of this study, as scaled NLLs in scenario 1 slightly increased around the optimal $R_0$ (Fig. 7b).

Although the changes of the NLL of the length composition dataset of PS-SoJ was small, the results suggest that there are two positive effects of rejecting datasets with highly uncertain $ESS_{TRUE}$ (scenario 3 in Fig. 7b and Table 4) on the fit of PS-SoJ to other datasets. Firstly, the pattern of the likelihood profile of PS-SoJ was improved compared with the 2012 stock assessment (scenario 1 in Fig. 7b). This result shows the conflict between abundance indices and length composition of this fishery was reduced. Secondly, in absolute terms, the NLL of PS-SoJ became smaller than that of the 2012 stock assessment (Table 4). These results indicate that extremely wide confidence intervals invalidate the use of $ESS_{TRUE}$ as an initial weighting factor in SS3. If we consider the median or other central tendency, the range of the confidence interval divided by the median could be used as a measure of uncertainty. When that is done, only 1987, 1993 and 1995 have high values (over 200). However, for 2004 and afterwards, this measure is less than 2 in all years, while only 1988, 1991, 1996, 2002 and 2003 have values much less than 10 prior to 2004. Assuming $ESS_{TRUE}$ of over-200 was uncertain and rejecting the years which correspond to this criterion would produce NLL between scenario 2 and 3, which means better model fit than scenario 1.

Furthermore, weighting the length composition dataset of PS-SoJ with the number of landings had an unfavorable influence on the fit of the dataset (Table 4). Given these results, the accuracy and precision of cluster samples should be evaluated while considering the similarity or the representativeness of each landing for the correct estimation of $ESS_{TRUE}$. Although there could be better methods to take the length composition of each landing into consideration, we tested ESS of PENNINGTON *et al.*[6] as one candidate and demonstrated the improvement in model fit.

The length composition dataset is another source of information on stock abundance[20]. We focused on this point and changed the weighting of the length composition dataset of PS-SoJ, which is just one of the size composition datasets, for the following reasons : (1) The catch amount of this fishery has been increasing in recent years (Fig. 1), (2) abundance indices derived from this fishery were not used in the 2012 stock assessment[2], highlighting the importance of the length composition data of PS-SoJ as one of the few sources of information on stock abundance for this fishery, (3) the abundance indices used in the stock assessment were derived from other fisheries targeting age groups that PS-SoJ did not target[2], and (4) the uncertainty of abundance indices derived from longline fisheries was a concern[21].

The ESS is one of indicators that show the amount of information on the population which the sample represents[8]. If we estimate $ESS_{TRUE}$ for each fishery using the method in this study as an independent factor of stock assessment model assumptions, the length composition dataset that accurately and precisely describes stock abundance should be given an adequate weight in the model analysis. In the case that conflicts among data sources is identified, this approach should be better able to enhance the reliability of stock assessment results compared with reducing all initial weighting factors for length composition datasets. On the other hand, when ad-hoc approach leads to reasonable results, the method shown here would provide the same results, and there is little necessity of using this method.

### 4.4. Importance in stock assessment

The ISC PBFWG has expressed deep concern for the stock status of PBF[3]. Recognizing the importance of developing reference points for the conservation and management of PBF, the WCPFC[22] formally adopted regulations on the catch of young PBF. In this study, differences of approximately 2,290 MT (2.3%) in SSB were observed between the results in scenario 3 and the 2012 stock assessment. We consider it is very difficult to have certainty for the difference within a range of 2.3% in stock assessments. However, in the situation of PBF with its high market value and international attention to its stock status, difference between point estimates of SSB would have socio-economically important impact on decision of management. The stock assessment should therefore be conducted carefully while minimizing all uncertainties, and the method described in this study would be helpful in achieving reliable stock assessment results.

As seen in the case of PS-SoJ, commercial fishery data has biases derived from a variety of sources such as the operational pattern or market forces, and usually these biases are not fully adjusted. When all input data is obtained from commercial fisheries, and fishery-independent and statistically well-designed indices are not available, appropriate weighting of each input data would be important in conducting reliable stock assessments.

We have presented here methods of setting initial weighting factors for length composition datasets based on $ESS_{TRUE}$ and identified the importance of estimating $ESS_{TRUE}$ as an objective indicator of accuracy and precision, while considering the number of landings and the similarity within each landing. We demonstrated that giving a zero weight to datasets with highly uncertain $ESS_{TRUE}$ improved the model fit of PS-SoJ, and observed important differences in SSB. The improvement of the model fit should give scientists greater confidence in stock assessment results when they provide advice on the management of PBF.

Although we examined five scenarios here, there would be many other approaches. Iterative calculation such as the bootstrap method used in this study would be helpful to select an optimal scenario. We believe that an approach similar to scenario 3 would be optimal even if we test a wider range of candidates. For other fisheries, however, different scenarios may be optimal, and a wider range of scenarios should be considered.

Extending the study to other fisheries and assessments would be helpful in setting more reasonable weighting factors for each length composition dataset compared with using weighting factors estimated internally in the assessment model. Moreover, testing some of these and a wider range of scenarios across fisheries and assessments would give more general and useful results.

## Acknowledgement

The authors thank the members of the ISC Pacific bluefin tuna Working Group for important contributions to this study. The authors also thank Jon BRODZIAK

(NMFS) and Cleridy Lennert-Cody (IATTC) and Shelley Clarke (Sasama) for valuable comments on an earlier draft of this manuscript and Kazuki Murai (Tokyo University of Agriculture) for assistance in data analysis. Finally, the authors wish to acknowledge RICC Francis (NIWA) for helping to improve the manuscript.

**References**


1) Majkowski J (2007) Global fishery resources of tuna and tuna-like species. <ftp://ftp.fao.org/docrep/fao/010/a1291e/a1291e. pdf> FAO Fisheries Technical Paper, 66pp. (Accessed August 11, 2015)

2) ISC Pacific bluefin tuna Working Group (2012) Stock Assessment of Pacific Bluefin Tuna in 2012, 10-17 November 2012, Honolulu, Hawaii, USA. <http://isc.ac.affrc.go.jp/pdf/Stock_assessment/ PBF_stock_assessment_report_2012.pdf> (Accessed August 11, 2015)

3) ISC Pacific bluefin tuna Working Group (2014) Stock assessment of Bluefin Tuna in the Pacific Ocean in 2014. <http:// isc.ac.affrc.go.jp/pdf/2014_Intercessional/Annex4_Pacific%20Bluefin%20Assmt%20Report%20 2014-%20June1-Final-Posting.pdf> (Accessed August 11, 2015)

4) ISC (2012) The reported total annual landings. <http://isc.ac.affrc.go.jp/fisheries_statistics/> (Accessed August 11, 2015)

5) Cochran W G (1977) Sampling Techniques. New York : John Wiley and Sons, 428pp.

6) Pennington M, Burmeister L-M, Hjellvik V (2002) Assessing the precision of frequency distributions estimated from trawl-survey samples. *Fish. Bull.* **100** : 74-80.

7) Nelson G A (2014) Cluster Sampling : A pervasive, yet little recognized survey design in fisheries research. *Trans. Am. Fish. Soc.* **143** : 926-938.

8) Pennington M, Vølstad J H (1994) Assessing the effect of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. *Biometrics* **50** : 725-732.

9) Hulson P J F, Hanselman D H, Quinn T J (2011) Effects of process and observation errors on effective sample size of fishery and survey age and length composition using variance ratio and likelihood methods. *ICES J. Mar. Sci.* **68** : 1548-1557.

10) Hulson P J F, Hanselman D H, Quinn T J (2012) Determining effective sample size in integrated age-structured assessment models. *ICES J. Mar. Sci.* **69** : 281-292.

11) Methot R D, Wetzel C R (2013) Stock Synthesis : a biological and statistical framework for fish stock assessment and fishery management. *Fish. Res.* **142** : 86-99.

12) Francis R I C C (2011) Data weighting in statistical fisheries stock assessment models. *Can. J. Fish. Aquat. Sci.* **68** : 1124-1138.

13) McAllister M K, 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.

14) Kai M (2008) Sensitivity analysis of stock assessment for Pacific bluefin tuna using Stock Synthesis 2, 28 May-4 June 2008, Shimizu, Japan. ISC, 19pp.

15) Maunder M N (2011) Review and evaluation of likelihood functions for composition data in stock-assessment models : Estimating the effective sample size. *Fish. Res.* **109** : 311-319.

16) Bayliff W H (1993) Growth and age composition of northern Bluefin tuna, *Thunnus thynnus*, caught in the eastern Pacific Ocean, as estimated from length-frequency data, with comments on trans-Pacific migrations. *IATTC Bulletin* **20** : 501-540.

17) Davison A C, Hinkley D V (1997) Bootstrap Methods and their Application. Cambridge : Cambridge University Press, 582pp.

18) Venzon D J, Moolgavkar S H (1988) A method for computing profile-likelihood-based confidence-intervals. *Appl. Stat. J. Roy. Statist. Soc. Ser.* C **37** : 87-94.

19) Hilborn R, Mangel M (1997) The Ecological Detective : Confronting Models with Data. Princeton University Press, Princeton, NJ, USA.

20) Folmer O, Pennington M (2000) A statistical evaluation of the design and precision of the shrimp trawl survey off West Greenland. *Fish. Res.* **49** : 165-178.

21) Oshima K, Mizuno A, Ichinokawa M, Takeuchi Y, Nakano H, Uozumi Y (2012) Shift of fishing efforts for Pacific bluefin tuna and target shift occurred in Japanese coastal longliners in recent years, 10-17 November 2012, Honolulu, Hawaii, USA. <http://isc.ac.affrc.go.jp/pdf/PBF/ISC12_PBF_3/ISC12_PBFWG-3_05_Oshima.pdf> (Accessed August 11, 2015)

22) WCPFC (2014) Conservation and management measure to establish a multi-annual rebuilding plan for Pacific bluefin tuna. <https://www.wcpfc.int/system/files/CMM%20 2014-04%20Conservation%20and%20Management%20 Measure%20to%20establish%20a%20multi-annual%20 rebuilding%20plan%20for%20Pacific%20Bluefin.pdf> (Accessed August 11, 2015)

# 統合資源評価モデルに用いる 体長組成データの重みづけ初期値評価

芝野あゆみ*†・金岩　稔**・石原幸雄***・氏　良介****・
志村　健****・竹内幸夫*****・余川浩太郎*****

（平成 27 年 8 月 18 日受付/平成 27 年 10 月 23 日受理）

**要約**：統合モデルを用いた資源解析における最重要課題の一つとして，各尤度コンポーネント（資源量指数，サイズ組成データ等）の重みづけが解析結果に直接的に影響することが挙げられる。2012 年に行われた太平洋クロマグロの資源評価では，この課題に対処するべく慎重な議論が交わされたが，主要な漁業である日本海まき網のサイズ組成データと他漁業の資源量指数の間で尤度コンポーネントの最適値に矛盾が生じ，妥当な解決策には至らなかった。そこで本研究では体長組成データの妥当な重みづけ初期値を決定する手法について，有効サンプル数とその信頼範囲を用いて決定する方法のパフォーマンスを検討した。初めに，体長組成データそのものの精度を評価するため，クラスターサンプリングを考慮した有効サンプル数を推定した。次に，最適な重みづけ手法を探るためシミュレーションを行い，各手法における尤度への影響を調べた。その結果，本研究の手法により日本海のまき網漁業のサイズ組成データと他漁業の資源量指数間に見られる矛盾は軽減され，尤度も改善された一方で，過去の資源評価と本研究における親魚量推定値の差は大きくはないことが明らかになった。しかし，世界的に注目が集まる本種の場合，僅かな差であっても資源管理に大きな影響を及ぼすと考えられる。また，重みづけ値をどのように決定するかという問題はあらゆる魚種の資源評価に共通するものであり，様々な漁業におけるデータの重みづけ初期値について考察する。

**キーワード**：資源評価，重みづけ値，有効サンプル数，体長組成，太平洋クロマグロ

* 東京農業大学生物産業学研究科生物産業学専攻
** 東京農業大学生物産業学部アクアバイオ学科
*** 鳥取県水産試験場
**** 鳥取県水産課
***** 水産総合研究センター国際水産資源研究所
† Corresponding author (E-mail : ayumi.shibano@gmail.com)