North Pacific Fisheries Commission

NPFC-2021-SSC PS07-WP15

Estimation of catch-at-size/age data of Pacific saury using stratified random sampling with proportional allocation

Wen-Bin Huang[1], Yi-Jay Chang[2] and Chih-hao Hsieh[2]

[1]Department of Natural Resources and Environmental Studies, National Dong Hwa University, Hualien 974.
E-mail: bruce@gms.ndhu.edu.tw

[2]Institute of Oceanography, National Taiwan University, Taipei 106.

Summary

Pacific saury is immediately graded into several commercial size categories (by weight) on board in the saury far sea fishery, e.g., the Chinese Taipei saury fishery (CTSF). In this study, a stratified random sampling approach with proportional allocation was employed to refine the estimator of the catch-at-size/age data for the CTSF. We increased the sample size of saury length from ~300-500 per year in 2007-2019 to 2,492 in 2020, amounting to an approximately 5-fold increase, particularly for the abundant categories. Two sample fishing vessels with their own specifications of commercial size categories were employed to involved the potential monthly and/or between-vessel variation. In addition, both age-length keys for saury from Chinese Taipei and Japan were used to estimate the catch-at-age data. The age-0 and age-1 saury accounted for 63.3 % and 36.7 % of the 2020 CTSF catch, respectively. The mean percentage of age-1 fish in the 1° x 1° grids was greater than 91% in June and July, decreased to about 39-47% from August to October, and was less than 29% in November and December.

1. Introduction

To enable the use of age-structured models in Pacific saury stock assessment, the Small Scientific Committee on Pacific Saury (SSCPS) has been encouraging members to share biological data, such as catch-at-size and catch-at-age data, since early 2019 (SSC PS, 2019a). Specifications and a template for submitting data on size composition and catch of Pacific saury were determined in the SSC PS05 meeting in late 2019; the age-length keys for Pacific saury were provided by Japan (SSC PS, 2019b). Submission of documentation detailing the protocols for estimating catch-at-size/age data of Pacific saury is encouraged to facilitate and promote discussion in SSC PS meetings. Sharing of these protocols will improve the development of a robust data set for age-structured stock assessment and will better inform decisions for model construction.

In the Pacific saury far sea fishery, which includes fishing fleets from Chinese Taipei, China, and Vanuatu, saury is graded into several commercial size categories (by weight) on board. In the Chinese Taipei saury fishery (CTSF) six commercial size categories have been defined and used in logbook entries since 2009, including *Extra large* (≤ 6 individuals/kg), *No.1* (7-9 individuals/kg), *No.2* (10-12 individuals/kg), *No.3* (13-15 individuals/kg), *No.4* (16-18 individuals/kg), and *No.5* (≥ 19 individuals/kg). Prior to 2009 only five commercial size categories were used; *No.5* was the last category to be added (Huang, 2007). Immediately after size classification, saury is subsequently frozen and stored on board in 10 kg capacity boxes. The commercial size categories used in the Pacific saury fisheries in China and Vanuatu are similar to the CTSF.

The purpose of sample survey design is to maximize the amount of information obtained, or to minimize the bound on the error of estimation, for a given cost (Scheaffer et al., 2012). If measurements within strata are homogeneous, a stratified random sample is able to produce a smaller bound on the error of estimation than would be produced by a simple random sample of the same size. Proportional allocation is useful if precise estimates are desired for the larger strata in the population (Glasgow, 2005). Proportional allocation sets the sample size in each stratum equal to be proportional to the number of sampling units in that stratum, yielding a self-weighted sample. Therefore, stratified random sampling with proportional allocation can minimize the variance of the estimator if there is considerable variability among the stratum means (Scheaffer et al., 2012). In the practice of sampling for catch estimation, many fishery scientists tend to aim for a minimum level of sampling in all size categories with increased replication in larger and more abundant size classes (Jennings et al., 2001).

In this study, we used a stratified random sampling approach with proportional allocation to replace that of the stratified random sampling with disproportional allocation and weighting factor that was reported in the SSC PS05 and SSC PS06 meetings (Huang et al., 2019; 2020). This change was made to refine the estimator of the Pacific saury catch-at-size/age data. Herein, we used the 2020 saury catch and knob length datasets, for the six CTSF commercial size categories, to estimate the catch-at-size/age data of saury.

2. Materials and methods

Two datasets were used in this study:

(1) 2020 landed saury catch data for the six CTSF commercial size categories in the northwestern Pacific. This dataset was compiled from CTSF logbooks, which are collected and managed by the Overseas Fisheries Development Council (OFDC).

(2) 2020 saury knob length data for the six CTSF commercial size categories in the northwestern Pacific. Samples of saury for each commercial size category were obtained from saury catch boxes in October, a month known to have high catch and CPUE. The knob lengths of all individuals in the sample boxes were measured and recorded to the nearest 0.1 mm. Two sample fishing vessels were deployed in 2020. These two sample

fishing vessels operated in similar locations, but differed slightly in their fishing timing; vessel A started fishing in August and ended in December while vessel B started fishing in July and ended in November (Figure 1).

We estimated the catch-at-size/age data of saury caught by the CTSF in 2020 using stratified random sampling with proportional allocation by:

(1) ***Evaluating the sample size of catch boxes for each size category*:** The accumulated number of catch boxes from a sample fishing vessel throughout the 2020 fishing season for the six commercial size categories were assumed as nb_0, nb_1, nb_2, nb_3, nb_4, and nb_5, respectively. The proportion of catch boxes for the six commercial size categories was estimated as P_0, P_1, P_2, P_3, P_4, and P_5, respectively, using a formula: $P_i = 100 * nb_i / \Sigma nb_i$ (i=0, 1~5). The proportions were further simplified to a simplest whole-number ratio, sn_0, sn_1, sn_2, sn_3, sn_4, and sn_5, using a formula: $sn_i = P_i/P_{min}$ (i=0, 1~5; P_{min} = the smallest P_i), as the sample size of catch boxes for each size category. If any P_i was smaller than 7%, its category was excluded from the simplification ratio procedure. This ensured that the total sample size of catch boxes for all commercial size categories from a sample fishing vessel did not exceed 15, which is the current upper limit of workforce and expenditure processing. The sample size of catch boxes assigned to the excluded size category (i.e., less than 7%) was 1.

(2) ***Random sampling of catch boxes for each size category and associated fish measurements:*** The catch boxes of the six commercial size categories from the sample fishing vessel were randomly sampled, after they were landed at the Kaohsiung Fishing Port, using the amount sn_0, sn_1, sn_2, sn_3, sn_4, and sn_5, respectively. Fish in the sample boxes were counted and knob length was measured for each commercial size category. The knob length data of fish samples from the 2 sample vessels was merged by each commercial size category. In addition, the mean fish number within each box for the six commercial size categories were recorded as mfn_0, mfn_1, mfn_2, fmn_3, mfn_4, and mfn_5, respectively.

(3) ***Estimation of saury catch-at-size data*:** From the mean fish number per box (mfn_i) for the six commercial size categories (i = 0~5) and the accumulated numbers of catch boxes (N_i) throughout the fishing season, we can estimate Fn_i (= $mfn_i * N_i$), i.e., individual fish caught for each commercial size category (i) throughout the fishing season. The knob length data for all fish in each size category (i) was simulated through resampling with Fn_i/10,000 individuals from the knob length data in step (2). The saury catch-at-size data was estimated by summarizing the knob length data from the six size categories and then subdividing using 1.0 cm intervals.

(4) ***Estimation of saury catch-at-age data*:** An age-length key for saury was acquired based on percentages of *age-1* and *age-0* fish assessed by otoliths from saury samples for each of the 1.0 cm length intervals. Using the above age-length key and the estimated catch-

at-size data in step (3), the catch-at-age data of saury caught by the CTSF was estimated.

Since the monthly accumulated numbers of catch boxes by each commercial size category for each 1° x 1° grid can be calculated from the logbook data, the monthly saury catch-at-size/age data for each 1° x 1° grid can be estimated by using the estimation steps described above.

In addition to the age-length keys derived from the CTSF's October samples, two seasonal age-length keys for 2020 were provided by Dr. Satoshi Suyama to adequately differentiate between saury growth rates in the early fishing season (May-July) and the main fishing season (August-December) **(Figure 2)**. The S curves of the age-length keys for the main fishing season in 2020 from Chinese Taipei and Japan were very similar and the length difference is about 0.4 cm. **(Figure 2)**. To estimate the monthly catch-at-size/age data for the 1° x 1° grids throughout the whole fishing season, the age-length key derived from the CTSF and that shared by Dr. Suyama were used for the main fishing season and the early fishing season, respectively.

3. Results

The accumulated numbers (nb_i) and proportions expressed as a percentage (P_i) of catch boxes from the sample fishing vessels for the six commercial size categories in 2020 are listed in Table 1. Based on the pre-determined criteria for proportions larger than 7%, the proportions of the commercial size *No.1, No.2, No.3, No.4*, and *No.5* from sample fishing vessel A were expressed as a simplest whole-number ratio (sn_i) as 1, 2, 3, 4, and 3, and the size *Extra large* with a proportion of less than 7% was assigned a value of 1 (**Table 1**). Herein, the sample size of the catch boxes for the six commercial size categories (nb_i) from sample fishing vessel A were determined (**Table 1**). The sample size of catch boxes for the sample fishing vessel B were also determined in the same way and are listed in **Table 1**.

The knob length frequency distribution for each commercial size category from the 2 sample fishing vessels is shown in the top panel of **Figure 3**. The sample size of the body lengths for the six commercial size categories from the 2 sample fishing vessels were 124, 291, 414, 539, 663, and 461, respectively, and 2,492 fish in total.

The estimated number of individual fish (Fn_i) graded into each commercial size category (i) throughout the whole fishing season for the CTSF is shown in the middle panel of **Figure 3**. This estimate was based on the accumulated numbers of catch boxes (N_i) and the mean fish number per box (mfn_i) in each size category. The re-sample size (Fn_i/10,000) for simulating all fish in each size category based on the knob length data from the box samples is also shown in the middle panel of **Figure 3**. An estimated length-frequency distribution of saury for the CTSF in 2020 was calculated by summarizing the knob length data from the six size categories (bottom panel, **Figure 3**).

Percentages of *age-0* and *age-1* saury within the 2020 saury samples, expressed in 1 cm intervals representing body length, are shown in the top panel of **Figure 4**. Using the age percentage values of the samples (top panel) and the estimated length frequency distribution (middle panel), the age composition of saury caught by the CTSF in 2020 was estimated (bottom panel). The *age-0* and *age-1* saury accounted for 63.3 % and 36.7 % of the 2020 catch, respectively. The fish less than 27 cm were all 0-year-old, while the fish greater than 29 cm in length were all 1-year-old. The cutoff point between the body lengths of these two ages was about 29.2 cm.

The estimated age compositions based on a 1° x 1° grid of fishing months for the entire fishing season in 2020 are shown in **Figure 5**. The mean percentage of *age-1* fish in the 1° x 1° grids was greater than 91% in June and July, decreased to about 39-47% from August to October, and was less than 29% in November and December.

4. Discussions

4.1. Stratified random sampling with proportional allocation approach

Saury catch is graded into several commercial size categories (by weight) on board and then is subsequently frozen and stored by the CTSF. The size categories can be regarded as strata. Since strata are usually more homogeneous than the population as a whole, stratification can lead to large improvements in the precision of estimated parameters (Glasgow, 2005). In addition, due to the saury catch having been categorized/stratified on board, the stratified random sample approach could be regarded as superior in terms of cost and/or administrative convenience for the CTSF. This approach has the potential to reduce the cost per observation in the survey, thereby enabling a larger sample size than that provided by a simple random sample approach (Glasgow, 2005).

To form a stratified random sample, simple random samples are independently drawn from each stratum and the resulting subsamples are weighted and combined (Glasgow, 2005). This approach, i.e., stratified random sampling with disproportional allocation and weighting factor, was applied and reported in the SSC PS05 and SSC PS06 meetings (Huang et al., 2019; 2020). However, this approach resulted in too much effort being allocated to the less abundant groups, such as the *Extra large* size category in the CTSF. The precision of parameter estimates within each stratum is determined by the sample size. Glasgow (2005) suggested that if precise estimates are desired for the more abundant strata in the population, the use of proportional allocation is preferred. In our study, our aim was to improve the precision of the estimator for a population, not each specific commercial size category. It is reasonable therefore to use the proportional allocation approach instead of using fixed samples in strata.

In our study, we increased the size of saury samples from ~ 300-500 per year in 2007-2019 to 2,492 in 2020, amounting to an approximate 5-fold increase. The sample size (n = 124) for the least abundant stratum, *Extra large* size category, in 2020 was about 1.5-2.0 times to previous years while increased about 6-8 times sample sizes were observed in the more abundant strata, e.g., *No.3, No.4*, and *No.5* size categories.

A saury fishing vessel or company has its own commercial specification for size categories, which are determined when setting sail and are not varied throughout the fishing season. Therefore, the estimated length-frequency distribution for a year, month, or a 1^{o} x 1^{o} grid is unbiased due to the use of corresponding catch boxes for each size category with the body length specification of a sample fishing vessel. In addition, we used 2 sample fishing vessels in the study, which contributed to a reduction in uncertainty attributed to potential monthly and/or between-vessel variation.

4.2. Age-length keys

We used two sources of age-length keys for saury: Chinese Taipei and Japan. In the main fishing season, the Chinese Taipei saury samples used for acquiring the age-length key were collected from the high seas, while most saury samples that were used to construct the Japan age-length key were collected in nearshore waters. However, the age-length keys in the main fishing season showed similar S curves in 2020 for both Chinese Taipei and Japan (**Figure 2**). In addition, the age-length key for the early fishing season was shared only from Japan, we encourage other saury fishing fleets to develop the age-length keys, particularly for the early fishing season.

5. References

Glasgow, G. 2005. Stratified sampling types. In K Kempf-Leonard (Ed.), *The Encyclopedia of Social Measurement*, Vol. 3, pp. 683-688. Elsevier ScienceDirect Books. (Available at www.sciencedirect.com/referencework/9780123693983/encyclopedia-of-social-measurement)

Huang, W. B. 2007. Body length, weight, and condition factor of Pacific saury (*Cololabis saira*) from the landed size-classes of Taiwanese catch in comparison with Japanese statistics. *Journal of the Fisheries Society of Taiwan* 34(4): 361-368.

Huang WB, Chang YJ, Hsieh CH. 2019. Enumeration of commercial size category, body length distribution, and age composition of the Pacific saury caught by the Chinese Taipei's saury fishery in 2018. NPFC-2019-SSC PS05-WP03. 6 pp. (Available at www.npfc.int)

Huang WB, Chang YJ, Hsieh CH. 2020. Body length distributions and age compositions of the Pacific saury caught by the Chinese Taipei saury fishery in 2007-2018. NPFC-2020-SSC PS06-WP18. 11 pp. (Available at www.npfc.int)

Jennings S, Kaiser MJ, Reynolds JD. 2001. *Marine Fisheries Ecology*. Blackwell Science, Oxford.

Scheaffer RL, Mendenhall III W, Ott RL, Gerow KG. 2012. *Elementary Survey Sampling*, 7th ed. Brooks/Cole, Boston, Mass.
SSC PS (Small Scientific Committee on Pacific Saury). 2019a. *4th Meeting Report*. NPFC-2019-SSC PS04-Final Report. 48 pp. (Available at www.npfc.int)
SSC PS (Small Scientific Committee on Pacific Saury). 2019b. *5th Meeting Report*. NPFC-2019-SSC PS05-Final Report. 45 pp. (Available at www.npfc.int)

Table 1. Total number of catch boxes, proportion expressed as a percentage, simplification ratio process, and the estimated sample number of catch boxes by commercial size category for the sample saury fishing vessels A and B.

(A) Sample saury fishing vessel A

Size category	*Extra_Large*	*No_1*	*No_2*	*No_3*	*No_4*	*No_5*	Total
Total number of catch boxes	154	2,789	6,130	9,573	12,184	8,393	39,223
Percentage (%)	0.4%	7.1%	15.6%	24.4%	31.1%	21.4%	100.0%
Simplification	NA (< 7%)	7.1/7.1 = 1.0	15.6/7.1 = 2.2	24.4/7.1 = 3.4	31.1/7.1 = 4.4	21.4/7.1 = 3.0	NA
Sample size of catch boxes	1	1	2	3	4	3	14

(B) Sample saury fishing vessel B

Size category	*Extra_Large*	*No_1*	*No_2*	*No_3*	*No_4*	*No_5*	Total
Total number of catch boxes	1,125	17,935	18,512	19,619	9,101	5,934	72,226
Percentage (%)	1.6%	24.8%	25.6%	27.2%	12.6%	8.2%	100.0%
Simplification	NA (< 7%)	3.0	3.1	3.3	1.5	1.0	NA
Sample size of catch boxes	1	3	3	3	2	1	13

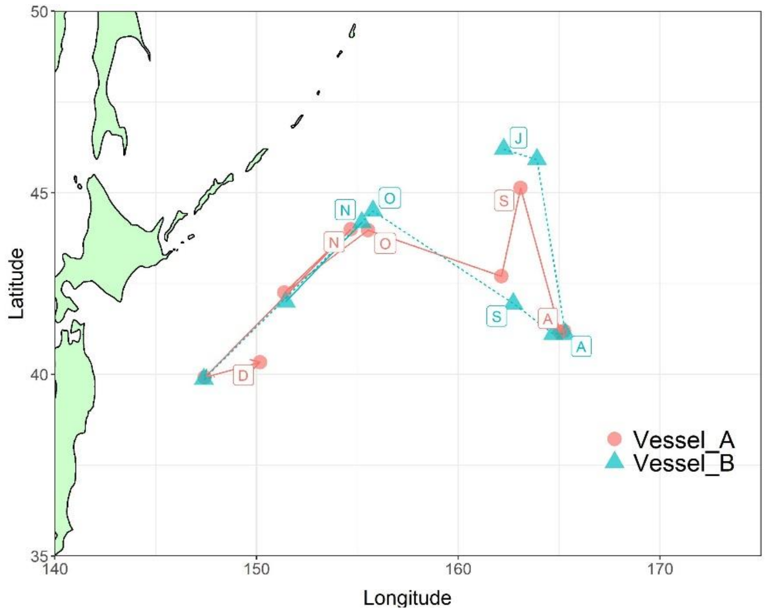

Figure 1. Monthly locations of the sample saury fishing vessels in 2020. J = July; A = August; S = September; O = October; N = November; D = December.

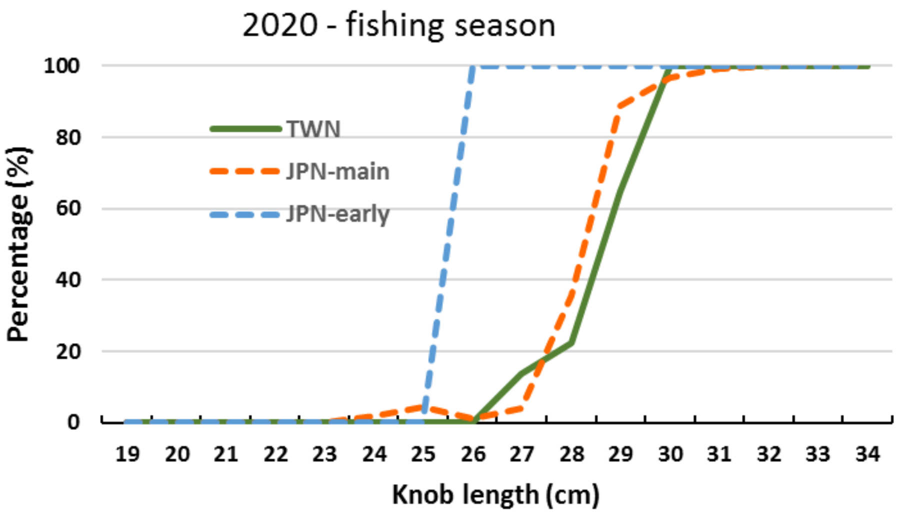

Figure 2. Comparison of the age-length keys for saury in the main fishing season, August to November, between Chinese Taipei and Japan in 2020. Green solid lines and orange dotted lines represent percentages of the *age-1* fish estimated by Chinese Taipei and Japan, respectively. The number in the brackets indicates otolith sample size used for age determination. The age-length keys for saury in the early fishing season, June to July, from Japan in 2020 are shown with the blue dotted line.

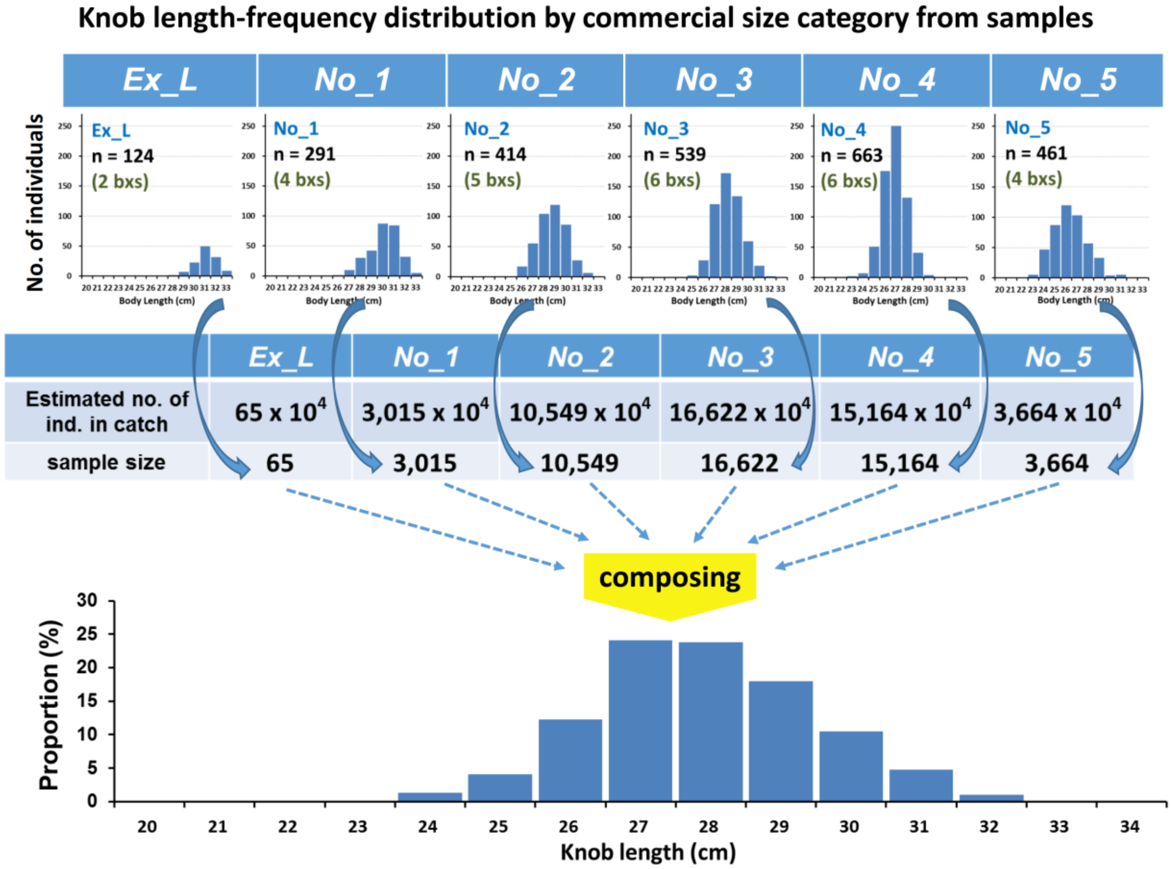

Figure 3. A schematic diagram for the estimation of the saury length-frequency distribution in 2020.

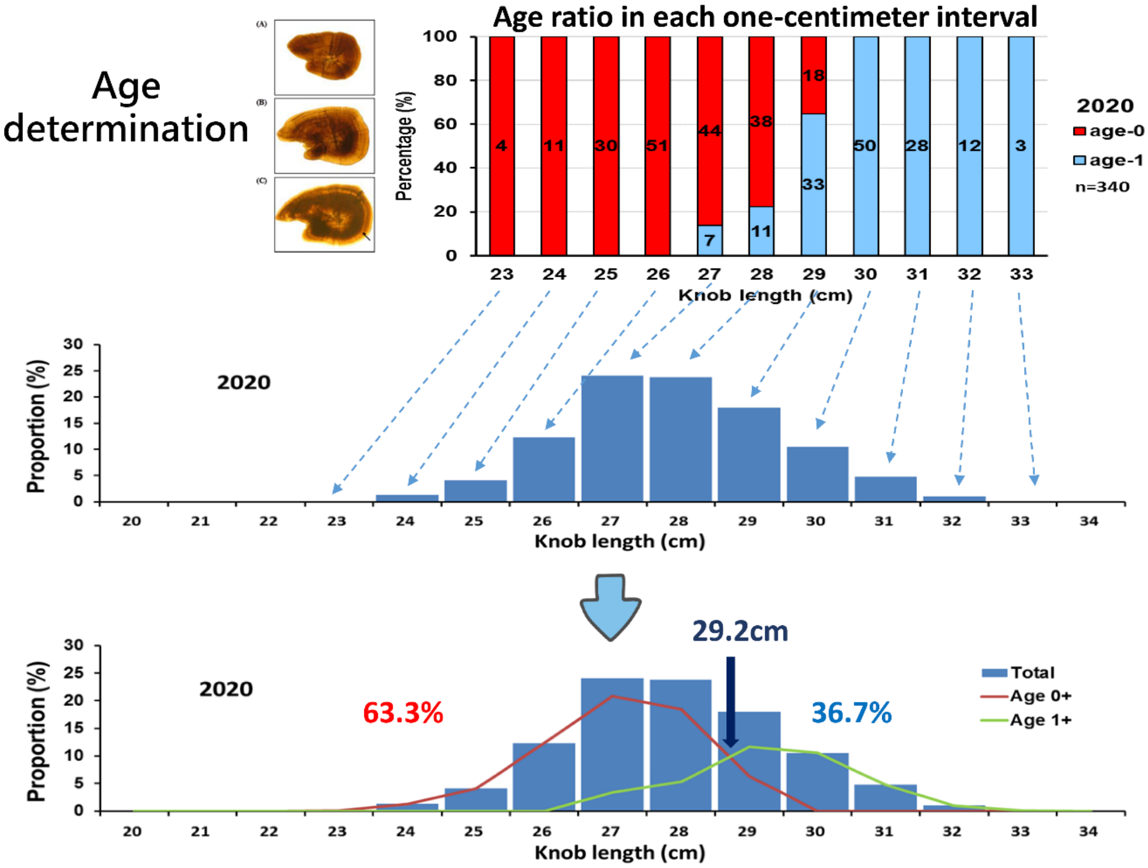

Figure 4. A schematic diagram for estimation of the age composition of saury in 2020. Red and green lines represent proportions of the catch for *age-1* and *age-0* saury, respectively.

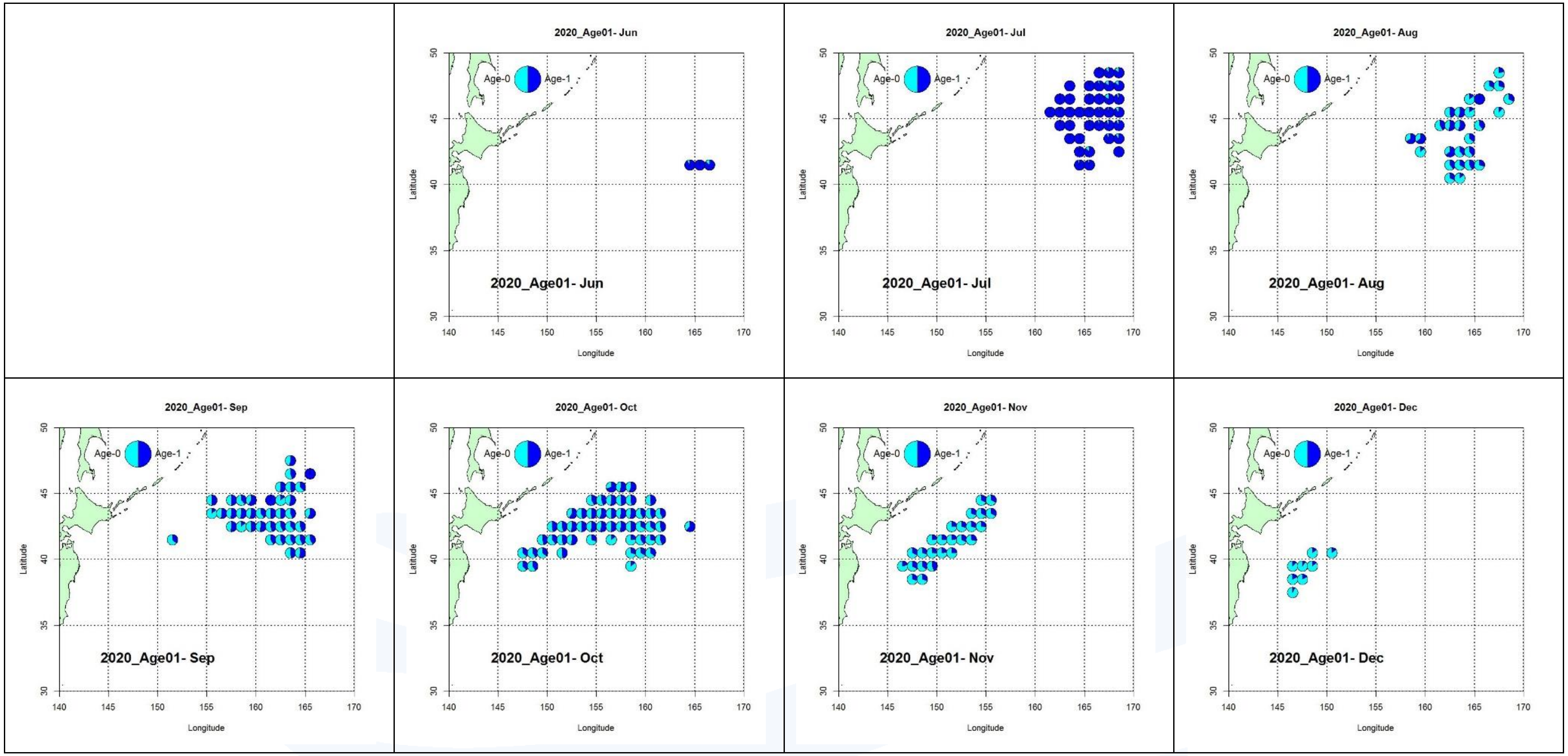

Figure 5. Pie plots of the estimated monthly age compositions within 1° x 1° grids for the *age-0* and *age-1* saury caught by the CTSF in 2020. Estimations in the early fishing season, June and July, used the Japanese age-length key; estimations in the main fishing season, from August to December, used the Chinese Taipei age-length key.