The devil is in the details: Investigating sources of bootstrapped bias in the Pacific bluefin tuna assessment and the associated impact on the future projections.

Huihua Lee[a], Hiromu Fukuda[b], Yohei Tsukahara[b], Kevin Piner[a], Mark Maunder[c], and Richard Methot Jr[d].

a: NOAA Fisheries, Southwest Fisheries Science Center, USA

b: Highly Migratory Resources Division, Fisheries Resources Institute, FRA, JAPAN

c: Inter-American Tropical Tuna Commission, USA

d: NOAA Fisheries, Northwest Science Center, USA

Corresponding author: huihua.lee@noaa.gov

April 2021

Working document submitted to the ISC Pacific bluefin tuna Working Group, International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean (ISC), from 20 to 23 and 27 April 2021, Webinar.

1. Introduction

Pacific bluefin tuna are under management measures developed by the Western and Central Pacific Fisheries Commission (WCPFC) and the Inter-American Tropical Tuna Commission (IATTC) to rebuild the population. Both commissions adopted two biomass-based rebuilding targets with specific time periods, 1) the initial rebuilding target (the median spawning stock biomass (SSB_{med}) from point estimates of the assessment between 1952 and 2014) by 2024 with a 98% probability, and 2) the second biomass rebuilding target (20% of $SSB_{F=0}$) by 10 years after reaching the initial rebuilding target or by 2034, whichever is earlier, with a 96% probability. Simulation-based projections were used to calculate the projected SSB and the probabilities of rebuilding to these targets from the terminal year of the assessment model for various harvesting and recruitment scenarios (ISC 2020).

Stock assessment replicates were simulated using the parametric bootstrapping procedure in stock assessment model software (Stock Synthesis (SS), Methot and Wetzel 2013) and then were used in the future projections to account for the uncertainty in the assessment terminal year and recruitment estimates. Because bootstrapped replicates were used to calculate the projected SSB and the probabilities of rebuilding to the management targets, validating the bootstrapped replicates that represent the stock assessment is crucial to ensure correct calculations. However, biases (defined as the difference between the median of bootstrapped replicates and the maximum likelihood estimate (MLE) from the assessment) occurred in the bootstrapping procedure for the latest and previous stock assessments (ISC 2016, ISC 2018, ISC 2020). These bootstrapped replicates showed more smaller SSB (negative biases) in the early modeling years (1952-1974) and more larger SSB (positive biases) in the recent modeling years (1994-2018) than the MLEs of SSB from the assessment model (Fig. 1). Although the source of the biases was not identified, these biased bootstrapped replicates were corrected using an ad-hoc method by adjusting the differences in the median future SSB between bootstrap replicates and the assessment model (Fukuda et al. 2020).

The purposes of this paper are 1) to identify possible sources of these biases in the bootstrapping procedure and 2) to calculate probabilities of achieving the rebuilding targets based on these adjusted bootstrapped replicates. We compare the probabilities of achieving the rebuilding targets among the unadjusted, ad-hoc bias corrected, and adjusted bootstrapped replicates to better understand the impacts. To keep the paper concise, some details in the methodologies were not described in the paper but provided in the stock assessment report (ISC 2020).

2. Review the parametric bootstrapping procedure in the 2020 Pacific Bluefin tuna stock assessment model

The PBF populations and corresponding fisheries data (replicates) were simulated using the same model assumptions, parameter estimates, and observation errors in the 2020 Pacific bluefin tuna stock assessment model. Each annual CPUE-based index of relative abundance observation was sampled from a bias-corrected lognormal error distribution with mean and standard deviation based on the expectation and inputted coefficient of variation from the assessment model. Quarterly size observation for each fleet was sampled from a multinomial error distribution with probability for each size bin and its variance described by the sample size based on the expectation and inputted sample size from the assessment model (Fig. 2). It is noted that due to the minimal sample size at 1 in the bootstrapping procedure, sample sizes < 1 (mostly observations for the super-year) were modified to avoid the runtime error as follows.

Bootstrapping procedure 0: The purpose of this bootstrapping procedure is to generate the bootstrapped replicates from the 2020 assessment model.

i. For the observations with sample size < 1, we replaced these sample sizes with 1.
ii. Generate 1 replicate (turn off the estimation and specified the parameters at the MLE estimates from the 2020 assessment model) in SS.
iii. Put back to the original sample sizes (<1).
iv. Fit this replicate to the assessment model (estimation is on).
v. Repeat steps i-iv 300 times.

3. Identify possible sources of bootstrapped biases in SSB

Simulated data from the bootstrapping procedure 0 were plotted against the assessment expectations. Overall, the medians of the simulated retain catch and of simulated CPUE-based abundance indices conformed with the respective expectations from the assessment model (Fig. 3). The medians of the simulated average sizes (in length or weight) for each fleet (Fig. 4a) generally followed the expected average sizes from the assessment model. The only exception is the simulated average sizes for fleet 10. To quantify biases for a given quantity (q), we used the percent relative errors (RE_i^q) defined by the relative difference between the quantity generated from the bootstrapping procedure ($Boot_i^q$) and the expected quantity (EXP_i^q) from the 2020 Pacific bluefin tuna stock assessment divided by the expected quantity for a given replicate (i). Bias was expressed as the median of relative errors (MRE), and variability was expressed as the standard deviation of the mean of REs (StdRE). We used "unbiased" to describe results with the MRE within -5% to 5%, "moderately unbiased" with the MRE between -10% and -5% or between 5% and 10%, and "biased" with the MRE less than or equal to -10% or greater than or equal to 10%.

Most of the bootstrapped sizes were unbiased (Fig. 4b). Biased sizes were generated for all the years in fleet 10 (MREs > 20%), 2003 in fleet 3 (MRE < -10%), and 2001 in fleet 8 (MRE < -10%). Moderately unbiased sizes were generated for 2017 in fleet 2, 1997-1998 in fleet 5, 1995 and 1997 in fleet 8, 1995 in fleet 9, 1959, 1961, 1971, 1975 and 1976 in fleet 13, 2009, 2010, and 2013 in fleet 14, 2014 in fleet 15, 2003, 2011-2012, 2016, and 2018 in fleet 19, and 2011 in fleet 28. For the fleets other than fleet 10, the moderately unbiased and biased bootstrapped sizes were variable (StdREs > 16) likely due to the very small inputted sample sizes (< 4, Fig. 4b) with the exception of 2010 and 2013 in fleet 14. Based on these results, we assumed that two possible sources could cause bootstrapped biases in SSB. The first source is biased size compositions for fleet 10 and the other source is the small inputted sample size.

3.1. Why simulated size compositions were biased for fleet 10 but unbiased for other fleets?

The added small constant to observed or expected proportions at size to make likelihood calculations more robust (Methot and Wetzel 2013) could be problematic, if it is not well handled, when generating the bootstrapped replicates. In the 2020 stock assessment, a 0.01 constant was added to the observed proportion for each weight bin (fleet 10), whereas a 0.0001 constant was added to the observed proportion for each length bin (fleets 1-9, 12-15, 17-20, and 28) (ISC 2020). These proportions were re-normalized to sum to 1.0 after the constant was added. Unlike the length compositions, the weight compositions for fleet 10 are highly skewed to the right (i.e., more smaller weight data than larger weight data, e.g., Fig. 5a). A larger constant added (e.g., 0.01) to the observed proportion is to fatten the long upper tail of weight distribution while maintaining the mode of the distribution (e.g., Fig. 5a). In the optimization, the use of the added constant helps to fit the composition data efficiently and to reduce the level of misfit of the highly skewed data (i.e., robustification). This resulted that the assessment model estimated the expected weight distribution with a larger average weight (fattened by a 0.01 added constant, e.g., Fig. 5b). The model expectations were then used in the bootstrapping procedure to generate the size composition replicates (e.g., Fig. 5c).

The problem emerged in the bootstrapping procedure when the added constant was used to generate a replicate from the expectation. Then the model fit this replicate with the added constant (robustification). This resulted that, as an example of the 2002 data, the expected average weight became 36.0 kg with a 0.01 added constant, while the original observed average weight (no constant added) was 10.1 kg (Table 1). When the added constant turned back to a very small number (e.g., 0.0001, like length compositions), the expected average weight (11.7 kg) was closer to the observed average weight (no constant added). This unintentional use of the added constant created biased data.

4. Solutions

4.1. Bootstrapping procedure 1: The purpose of this bootstrapping procedure is to create unbiased size compositions for fleet 10.

i. For the observations with sample size < 1, we replaced these sample sizes with 1.

ii. Put the added constant at 0.0001 (instead of 0.01) for fleet 10.

iii. Generate 1 replicate (turn off the estimation and specified the parameters at the MLE estimates from the 2020 assessment model) in SS.

iv. Put back to the original sample sizes (<1).

v. Put the added constant at 0.01 for fleet 10.

vi. Fit this replicate to the assessment model (estimation is on).

vii. Repeat steps i-vi 300 times.

This bootstrapping procedure removed the 0.01 added constant for fleet 10 when generating replicates from the expectations. This created the unbiased sizes for most of the years for fleet 10 (Fig. 6), except for moderately unbiased sizes in 2017 and 2018. Both 2017 and 2018 sample sizes (< 6) were the smallest in fleet 10. Like the bootstrapping procedure 0, biased sizes were generated for 2003 replicate in fleet 3 and 2001 replicate in fleet 8 (MREs < -10%). However, moderately unbiased sizes were generated for 1997-1998 in fleet 5, 1995, 1998, 2005, and 2011 in fleet 8, 1995 in fleet 9, 1954, 1961, 1971, and 1976 in fleet 13, 2009, 2010, and 2013 in fleet 14, 2014 and 2017 in fleet 15, 2003, 2011-2012, 2016, and 2018 in fleet 19, and 2005 in fleet 28. Again, the moderately unbiased and biased bootstrapped sizes were variable (StdREs > 16) likely due to the very small inputted sample sizes (< 4) with the exception of 2010 and 2013 in fleet 14.

The SSBs were negatively biased (MREs between -11% and -17%) from 1952 to 1974. The biases were then reduced after 1975 (-10% < MREs < 10%) with the exceptions that the 2009-2010 and 2017-2018 SSB were positively biased (MREs > 10%) (Fig. 7).

4.2. Bootstrapping procedure 2: The purpose of this bootstrapping procedure is to validate the bootstrapping procedure 1.

i. For the observations with sample size < 1, we replaced these sample sizes with 1.

ii. Put the added constant at 0.0001 (instead of 0.01) for fleet 10.

iii. Generate 1 replicate (turn off the estimation and specified the parameters at the MLE estimates from the 2020 assessment model) in SS.

iv. Put back to the original sample sizes (<1).

v. Put the added constant at 0.01 for fleet 10.

vi. Replace the bootstrapped size compositions with the expected size compositions from the 2020 assessment model.

vii. Fit this replicate to the assessment model (estimation is on).

viii. Repeat steps i-vii 300 times.

This bootstrapping procedure replaced the bootstrapped size compositions with the expected size compositions from the 2020 assessment model after generating the replicate. Therefore, all the replicates have the same size compositions (e.g., fleet 10 in Fig. 8a). Although this bootstrapping procedure reduced the source of uncertainties in SSB (StdREs are 19-44 in bootstrapping procedure 1 to 9-32 in bootstrapping procedure 2), the MREs of SSB showed a similar pattern to those from the bootstrapping procedure 1 indicating negative biases (MREs between -11% and -21%) from 1952 to 1975 and reduction of biases after 1976 (-10% < MREs < 10%) (Fig. 8b).

4.3. Bootstrapping procedure 3: The purpose of this bootstrapping procedure is to control the dynamics in the early years that showed the negative biases of SSB in the bootstrapping procedure 1.

i. For the observations with sample size < 1, we replaced these sample sizes with 1.

ii. Put the added constant at 0.0001 (instead of 0.01) for fleet 10.

iii. Specify the early recruitment deviations at the assessment estimates.

iv. Generate 1 replicate (turn off the estimation and specified the parameters at the MLE estimates from the 2020 assessment model) in SS.

v. Put back to the original sample sizes (<1).

vi. Put the added constant at 0.01 for fleet 10.

vii. Fit this replicate to the assessment model (estimation is on).

viii. Repeat steps i-vii 300 times.

This bootstrapping procedure controlled the dynamics in the early years by specifying the early recruitment deviations at the assessment estimates. As expected, SSBs were unbiased from 1982 to 1978 (Fig. 9). However, MREs of SSB after 1979 slightly increased compared to those in the bootstrapping procedure 1.

4.4. Bootstrapping procedure 4: The purpose of this bootstrapping procedure is to reduce the variability and possible biases resulting from the small inputted sample sizes when generating the replicates.

i. For all observations, multiply inputted sample sizes by 10 for all fleets.

ii. Put the added constant at 0.0001 (instead of 0.01) for fleet 10.

iii. Generate 1 replicate (turn off the estimation and specified the parameters at the MLE estimates from the 2020 assessment model) in SS.

iv. Put back to the original sample sizes.

v. Put the added constant at 0.01 for fleet 10.

vi. Fit this replicate to the assessment model (estimation is on).

vii. Repeat steps i-vi 300 times.

This bootstrapping procedure increased all inputted sample sizes when generating the replicates. The assessment model then fit these replicates with original sample sizes to maintain the weights among observations. Unlike the bootstrapping procedure 1, unbiased sizes were generated for all years for fleet 10, 2003 in fleet 3, and 2001 replicate in fleet 8 (Fig. 10). However, biased sizes were generated for 2004-2005 replicates in fleet 5 and 2010 replicate in fleet 14, and moderately unbiased sizes were generated for 1995, 1998, 1999, 2001, and 2002 replicates in fleet 5, 1959-1961, 1976, and 2006 replicates in fleet 13, and 2013 replicate in fleet 14. The moderately unbiased and biased bootstrapped sizes were less variable (StdREs < 5) compared to those in bootstrapping procedure 0 or 1 (StdREs > 16). These moderately unbiased and biased bootstrapped sizes resulted from the super year observations not from the small inputted sample sizes.

The SSBs were moderately unbiased for most of the years with the exceptions that the 1955-1956, 1979, 1984, 1989, 1994, 1999, and 2005-2006 SSB were positively biased (MREs > 10%) (Fig. 11). The StdREs of SSB (between 12.5 and 30.8) from this bootstrapping procedure were smaller than those from bootstrapping procedure 1 (between 18.5 and 44.1) (Fig. 11).

5. Associated impact on the future projections

The probabilities of achieving the initial and second rebuilding targets are above the levels prescribed in the WCPFC Harvest Strategy (75% and 60% in 2024 and 10 years after achieving the initial rebuilding target) the current management measure for all the bootstrapping procedures (Table 2). The probabilities of achieving the initial rebuilding target are 100% for most of the bootstrapping procedures, except for bootstrapping procedure 2 (99.7%). The probability of achieving the second rebuilding target from the 2020 base case model without the ad-hoc adjustment was the highest (99.1%), followed by the 2020 base case model with the ad-hoc adjustment (98.9%), bootstrapping procedures 1, 3, and 4 (98.7%), and bootstrapping procedure 2 (98.6%).

The projected SSB were generally similar among the bootstrapping procedures examined (Fig. 13). A notable difference was in the bootstrapping procedure 4, where smaller projected SSB with narrower confidence intervals before 2025 than others.

6. Conclusion

The bootstrapping procedure 1 corrected the undue usage of adding a constant to the compositions in bootstrapping procedure, suggesting that this added constant should be minimal when generating the replicates. When the uncertainties were from data other than compositions (bootstrapping procedure 2), a similar pattern of MREs of SSB validated the processes in bootstrapping procedure 1 (negative biases for early modeling years and moderately positive unbiased for the more recent years) (Fig. 12). When these early negative biases were natively controlled, we saw a similar pattern of MREs of SSB with slightly increased MREs of SSB after 1979 (bootstrapping procedure 3 in Fig. 12). This pattern of MREs of SSB disappeared when generating the replicates with systematically increased (x 10) small inputted sample sizes (bootstrapping procedure 4 in Fig. 12). The bootstrapping procedure 4 improved the overall biases and variabilities of bootstrapped SSB. The reduction of biases and variability was carried into projecting SSB, resulting in smaller uncertainty of projected SSB before 2025. Based on these results, bootstrapping procedure 4 is the best performer among the bootstrapping procedures examined.

The bias due to the bootstrapping procedures used for the projections appeared to have little impact on management interpretations of the probability to achieve the rebuilding targets within the specified time. However, understanding the source of the bias and offering solutions to potentially correct that source of bias is preferable to relying on the ad-hoc approach. We note that our results suggest that the ad-hoc approach appeared to be reasonably successful compared to the approach with no adjustment at all.

Reference

Fukuda, H., Tsukahara, Y., Nishikawa, K., Maunder, M.N. 2020. A review of the issues associated with the bootstrap calculations and future projections in the PBF assessment. ISC Pacific bluefin tuna Working Group, International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean (ISC), April 2020. ISC/20/PBFWG-1/15.

ISC. 2016. The 2016 Pacific bluefin tuna stock assessment. The 16th Meeting of the International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean, July 13-18, 2016. ISC/16/ANNEX/09. http://isc.fra.go.jp/pdf/ISC16/ISC16_Annex_09_2016_Pacific_Bluefin_Tuna_Stock_Assessment.pdf

ISC. 2018. Stock assessment of Pacific bluefin tuna in the Pacific Ocean in 2018. The 18th Meeting of the International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean, July 11-16, 2018. ISC/18/ANNEX/14. http://isc.fra.go.jp/pdf/ISC18/ISC_18_ANNEX_14_Pacific_Bluefin_Tuna_Stock_Assessment_2018_FINAL.pdf

ISC. 2020. Stock assessment of Pacific bluefin tuna in the Pacific Ocean in 2020. The 20th Meeting of the International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean, July 15-20, 2020. ISC/20/ANNEX/11. http://isc.fra.go.jp/pdf/ISC20/ISC20_ANNEX11_Stock_Assessment_Report_for_Pacific_Bluefin_Tuna.pdf

Methot, R.D. and Wetzel, C.R. 2013. Stock Synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research, 142: 86-99. https://doi.org/10.1016/j.fishres.2012.10.012

Table 1. Observed and expected average weight for various constant values added to the observed proportions for fleet 10.

Years	Observed average weight (no constant)	Expected average weight from the assessment (0.01 constant)	Expected average weight from the assessment (0.0001 constant)
1994	4.8	38.6	15.0
1995	10.8	35.0	10.4
1996	17.5	35.8	11.5
1997	29.0	38.8	15.3
1998	26.5	38.9	15.4
1999	17.4	35.8	11.5
2000	23.1	37.3	13.3
2001	16.0	38.5	15.0
2002	10.1	36.0	11.7
2003	22.3	37.7	13.9
2004	51.6	52.7	33.3
2005	32.8	41.2	18.5
2006	31.9	42.7	20.3
2007	35.5	41.8	19.1
2008	23.8	39.8	16.6
2009	16.3	38.2	14.6
2010	39.9	48.6	28.0
2011	19.3	44.8	23.0
2012	16.5	45.2	23.6
2013	24.2	46.5	25.3
2014	17.6	41.0	18.2
2015	24.4	48.3	27.6
2016	39.4	45.5	24.0
2017	27.5	39.0	15.6
2018	59.3	46.5	25.2

Table 2. Probability of achieving the initial and second rebuilding targets from various bootstrapping procedures the current management measure, where the 2020 basecase model (with ad-hoc biased adjustment) is the scenario 1 in the 2020 stock assessment (ISC 2020).

	Probability of achieving the initial rebuilding target at 2024	Probability of achieving the second rebuilding target at 2034
2020 model (without ad-hoc biased adjustment)	100.0%	99.1%
2020 model (with ad-hoc biased adjustment)	100.0%	98.9%
Bootstrapping procedure 1	100.0%	98.7%
Bootstrapping procedure 2	99.7%	98.6%
Bootstrapping procedure 3	100.0%	98.7%
Bootstrapping procedure 4	100.0%	98.7%

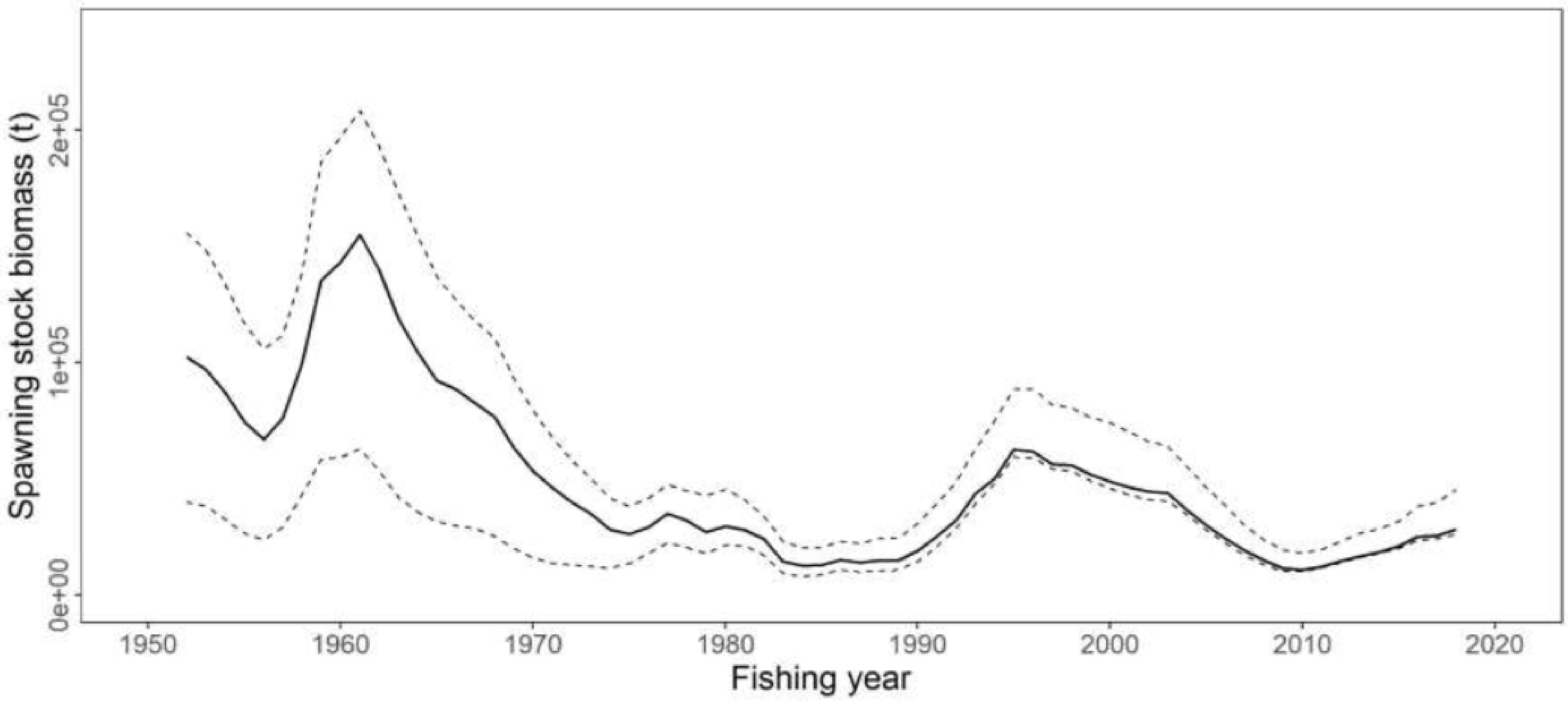


Fig. 1. Maximum likelihood estimate of spawning stock biomass from the 2020 stock assessment (solid line) and the 90% confidence intervals (dash lines) from the bootstrapped procedure in SS.

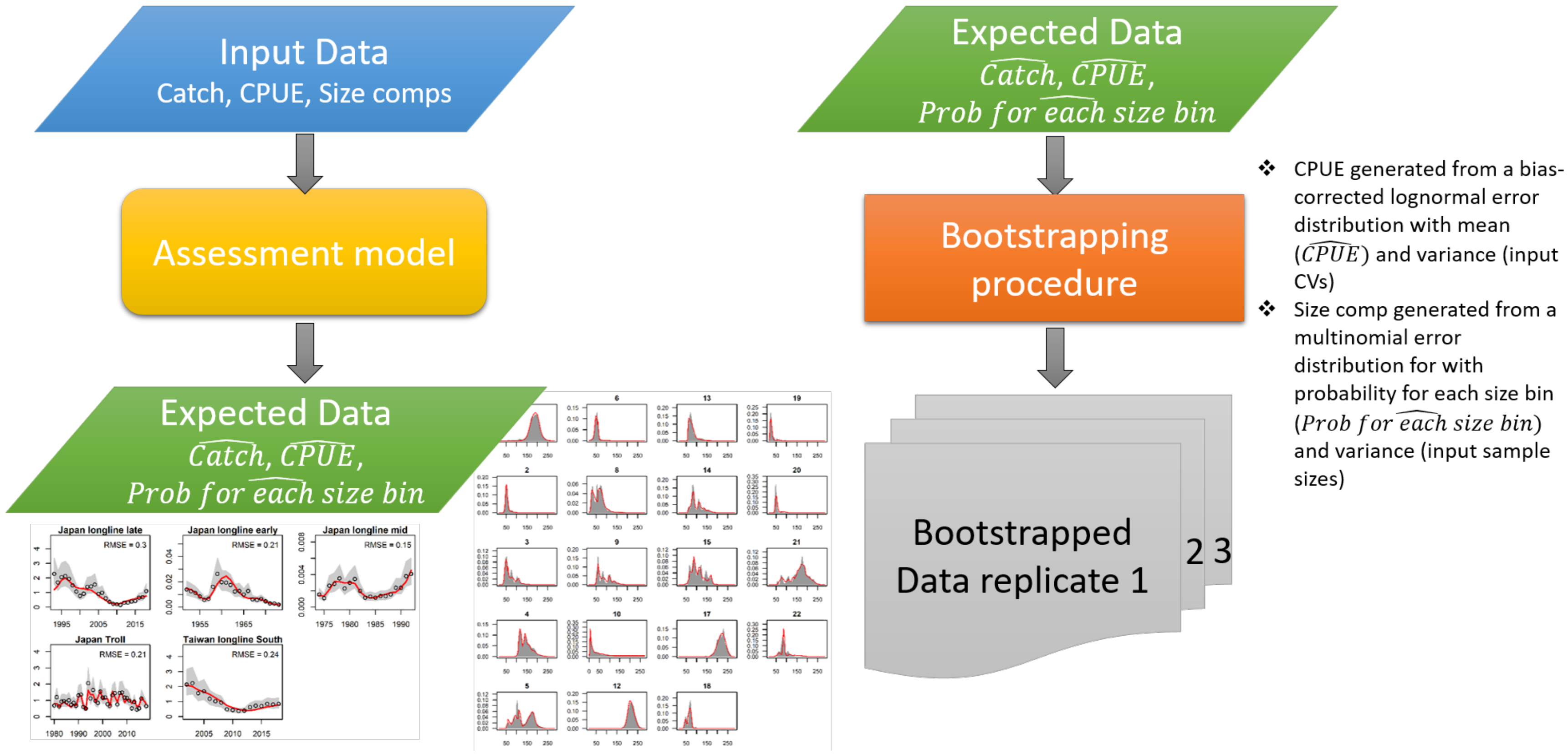
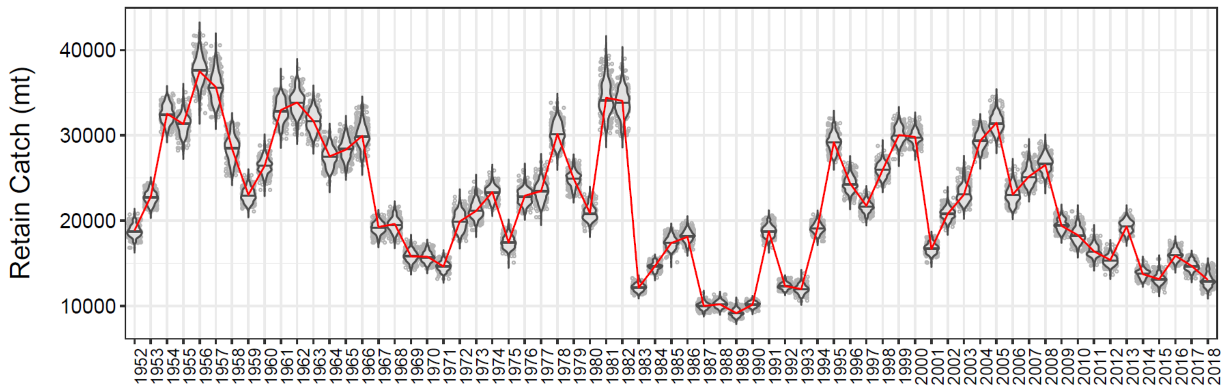


Fig. 2. Diagram of the parametric bootstrapping procedure in the 2020 Pacific Bluefin tuna stock assessment model using stock synthesis.

(a)



(b)

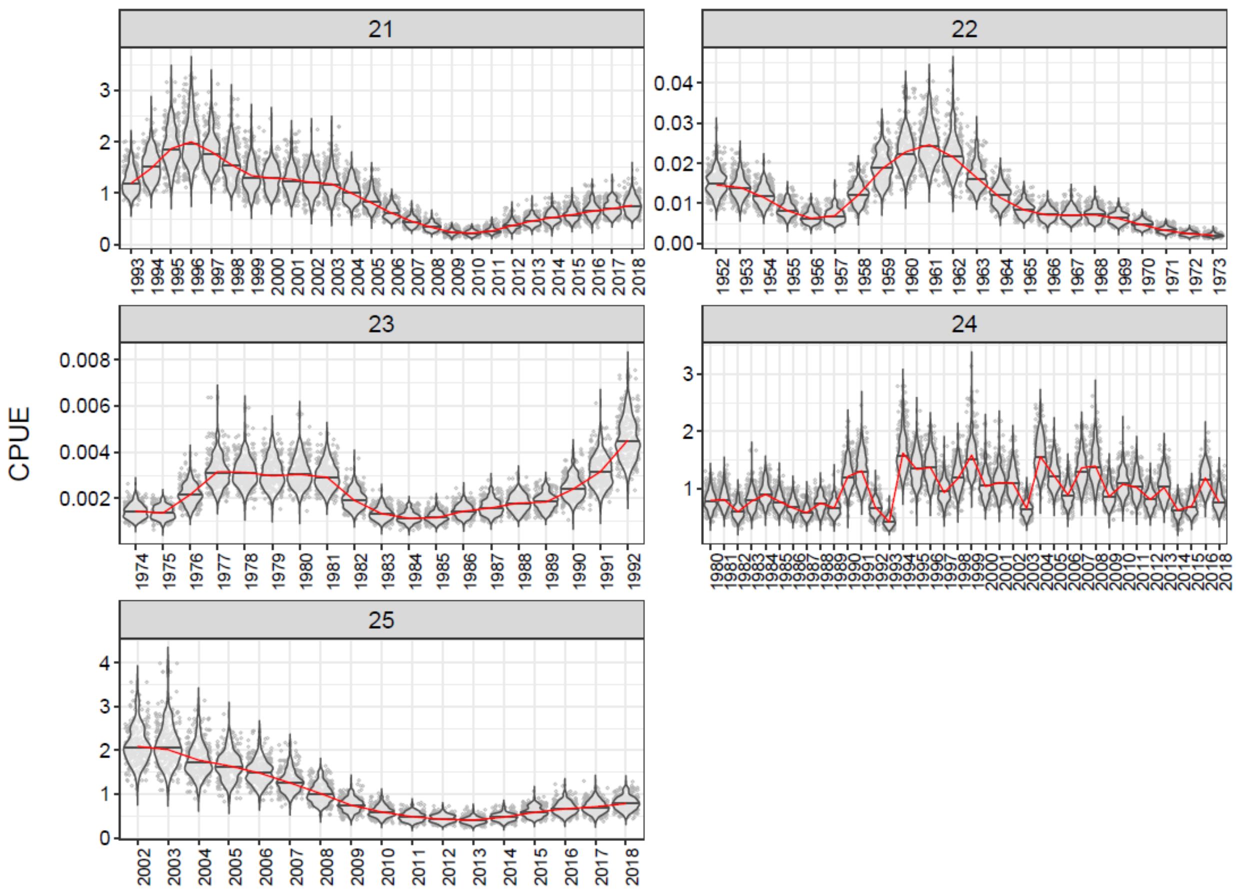
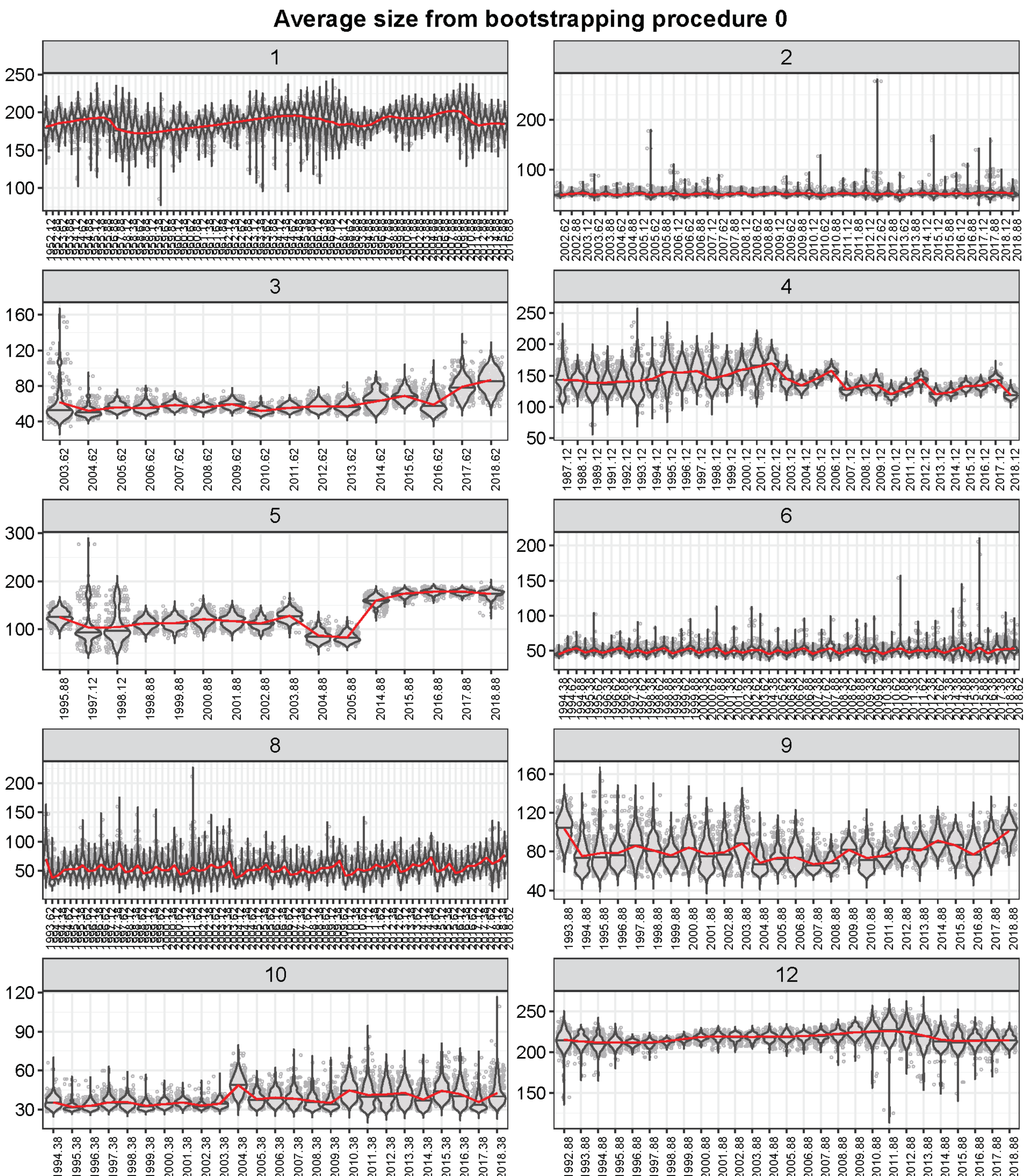


Fig. 3. Retain catch (panel a) and CPUE-based abundance indices (panel b) generated from the bootstrapping procedure 0 (violins), where the horizontal line in each violin indicates the median value and the red solid line indicates the retain catch (panel a) and the expected CPUE (panel b) in the 2020 Pacific Bluefin tuna stock assessment model.


Average size from bootstrapping procedure 0
1
2
3
4
5
6
8
9
10
12

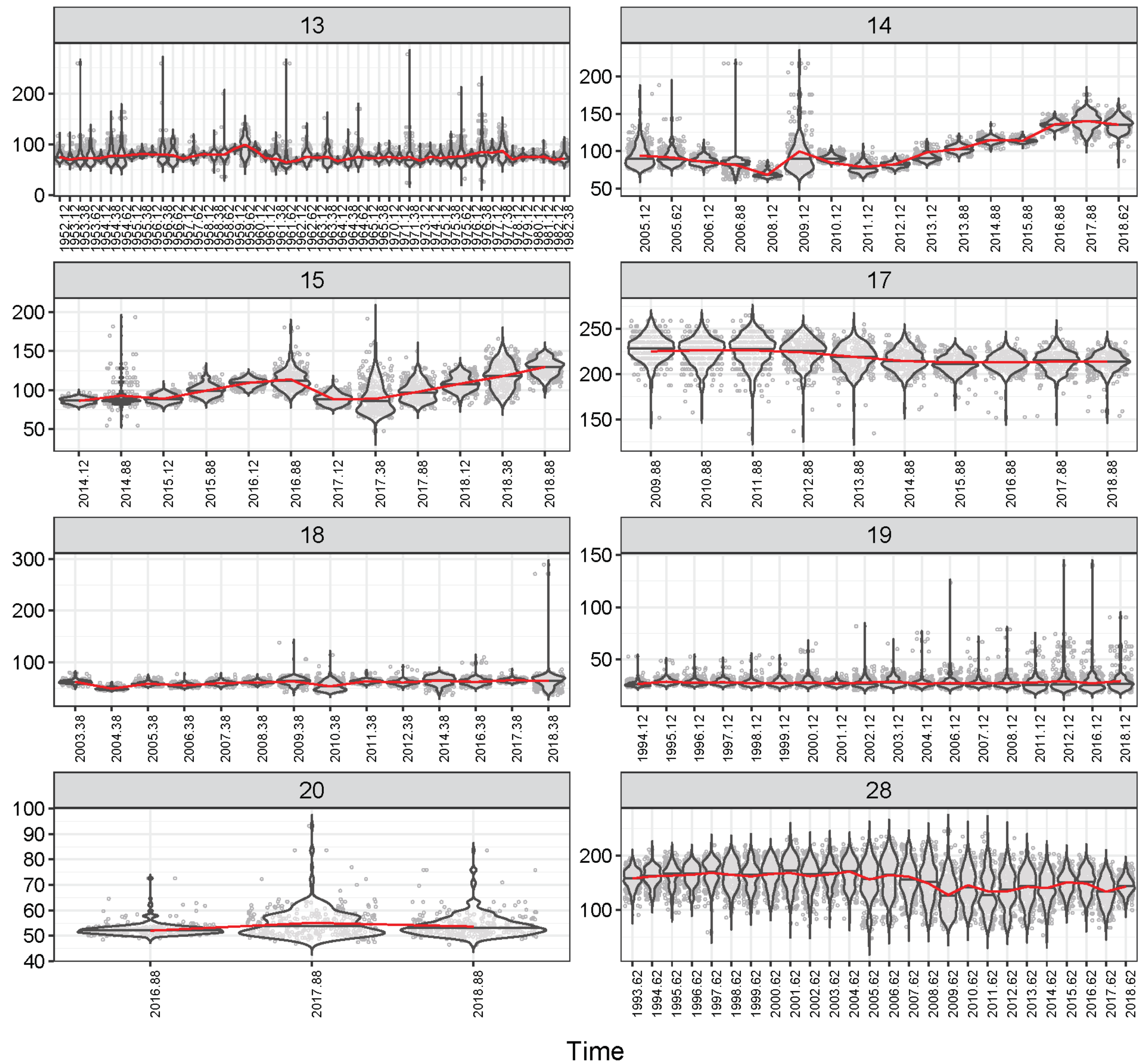
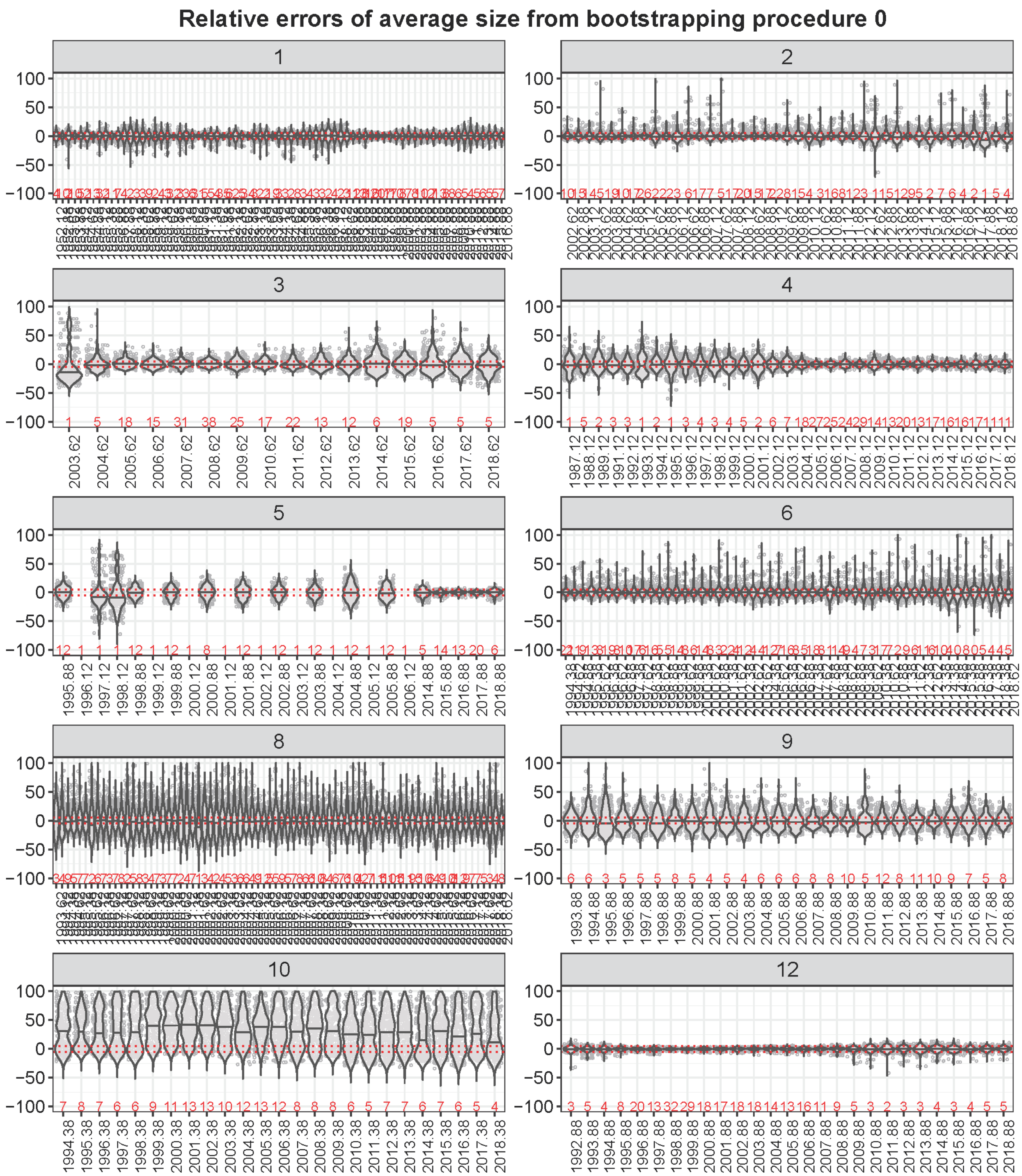


Fig. 4a. Average size for each fleet generated from the bootstrapping procedure 0 (violins), where the horizontal line in each violin indicated the median value and the red solid lines indicated the expected average size in the 2020 Pacific Bluefin tuna stock assessment model.


Relative errors of average size from bootstrapping procedure 0
1
2
3
4
5
6
8
9
10
12

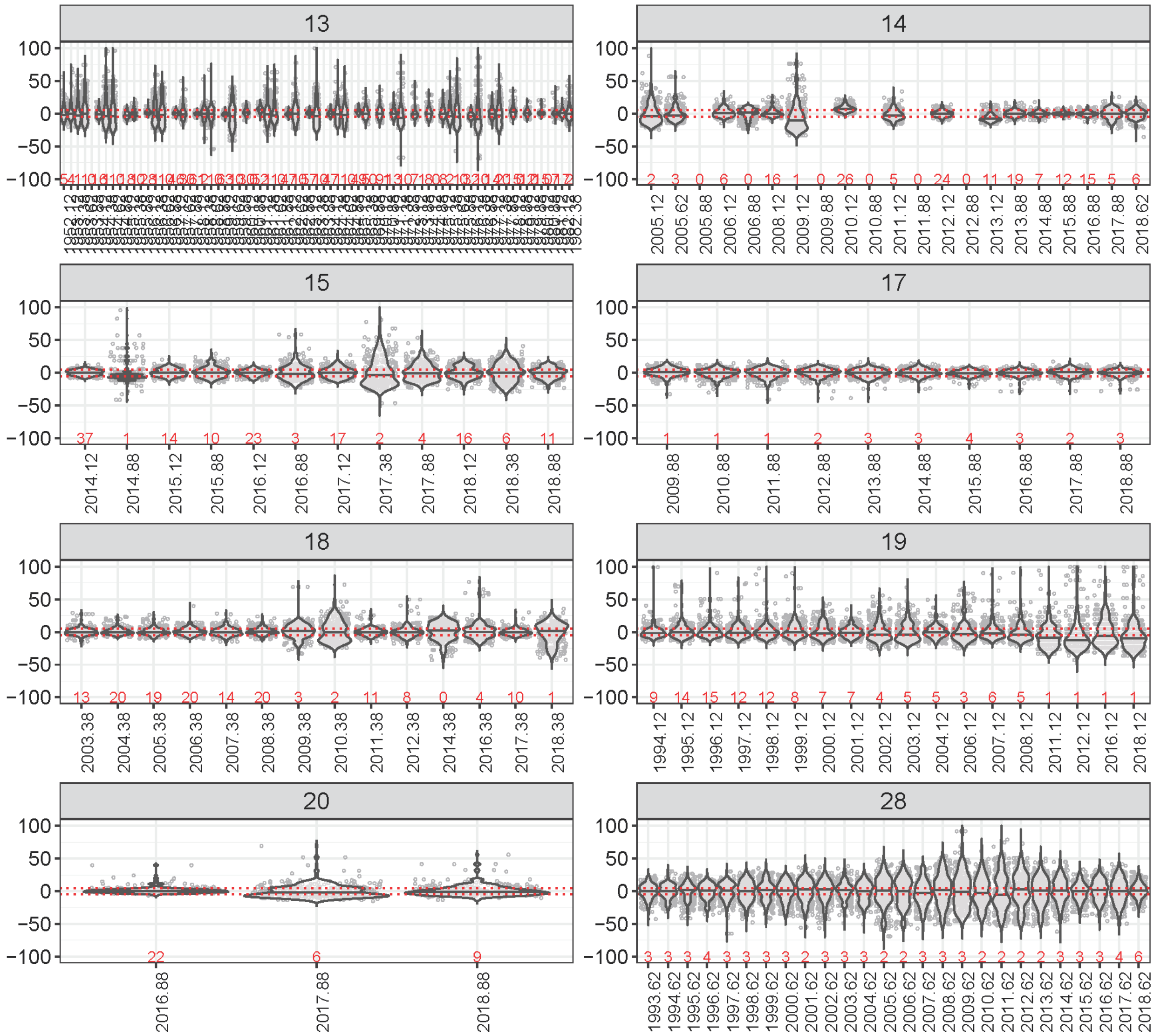


Fig. 4b. Relative errors of the average sizes for each fleet (violins), where the horizontal line in each violin indicated the median value and the red dotted lines indicated the -5% and 5%. Relative errors are defined by the relative difference between the average sizes generated from the bootstrapping procedure 0 and the expected average size in the 2020 Pacific Bluefin tuna stock assessment. The inputted sample sizes are shown in the bottom of each violin.

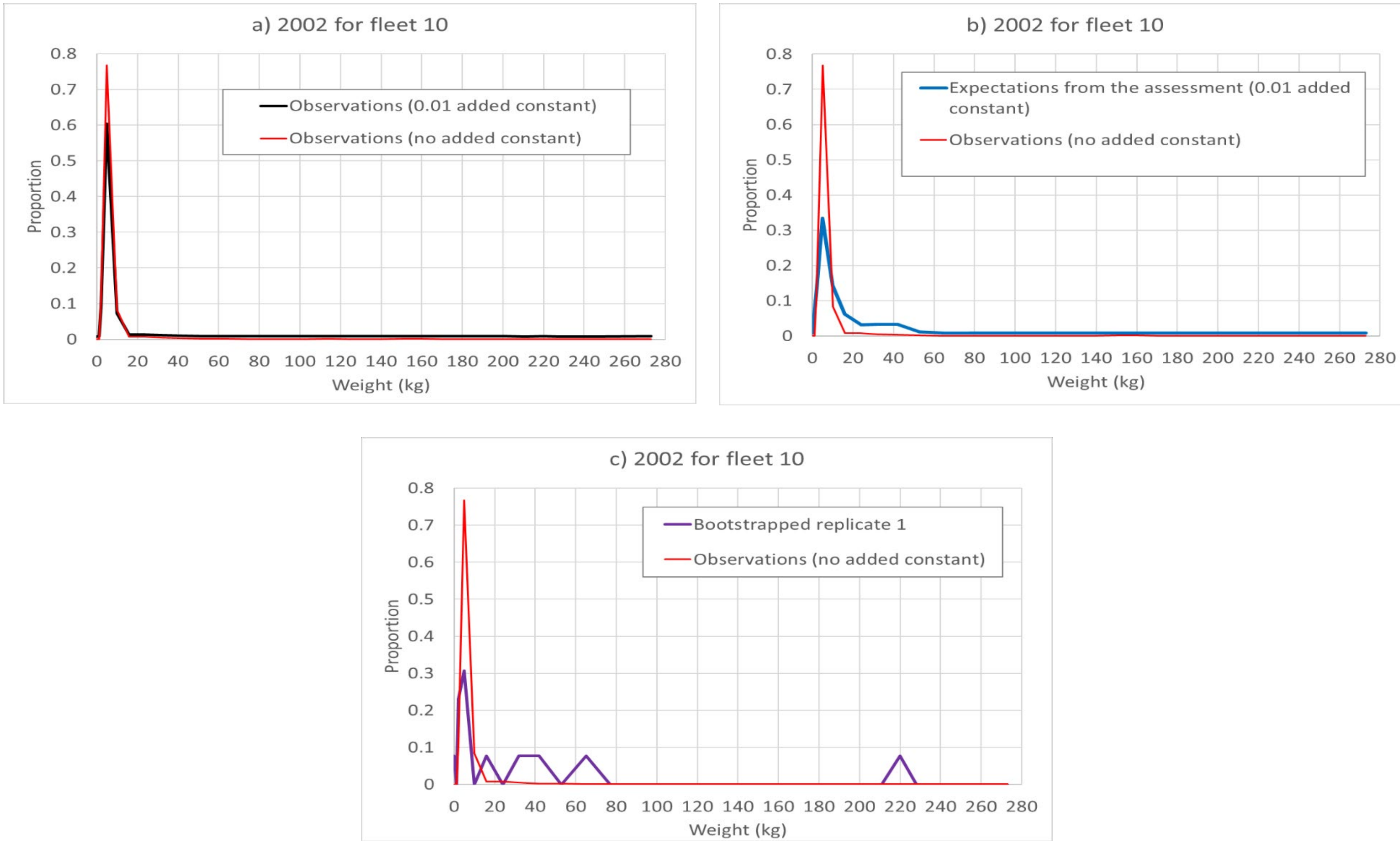


Fig. 5. The 2002 weight distributions for fleet 10, where the red curve indicates the observed proportions, black curve indicates the observed proportions with a 0.01 constant added to each weight bin (panel a), blue curve indicates the expectations from the assessment model with a 0.01 added constant (panel b), and purple curve indicates a bootstrapped replicate that was generated based on the model expectations from the blue curve in panel b (panel c).

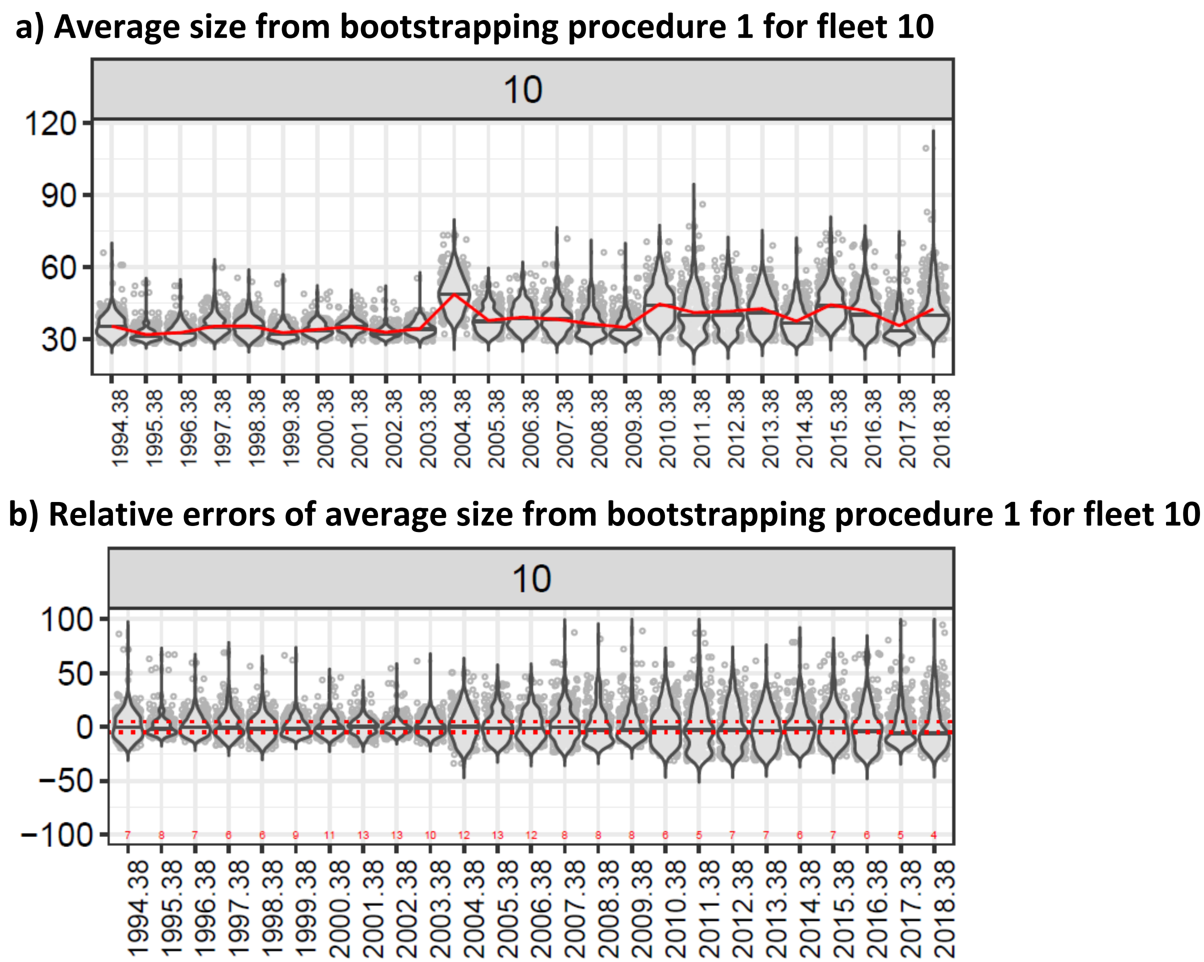


Fig. 6. (a) Average size for fleet 10 generated from the bootstrapping procedure 1 (violins in panel a) and (b) Relative errors of the average sizes (violins in panel b), where the horizontal line in each violin indicated the median value, the red solid line indicated the expected average size in the 2020 Pacific Bluefin tuna stock assessment model, and the red dotted lines indicated the -5% and 5% relative errors. Relative errors are defined by the relative difference between the average sizes generated from the bootstrapping procedure and the expected average size in the 2020 Pacific Bluefin tuna stock assessment. The inputted sample sizes are shown in the bottom of each violin in panel b.

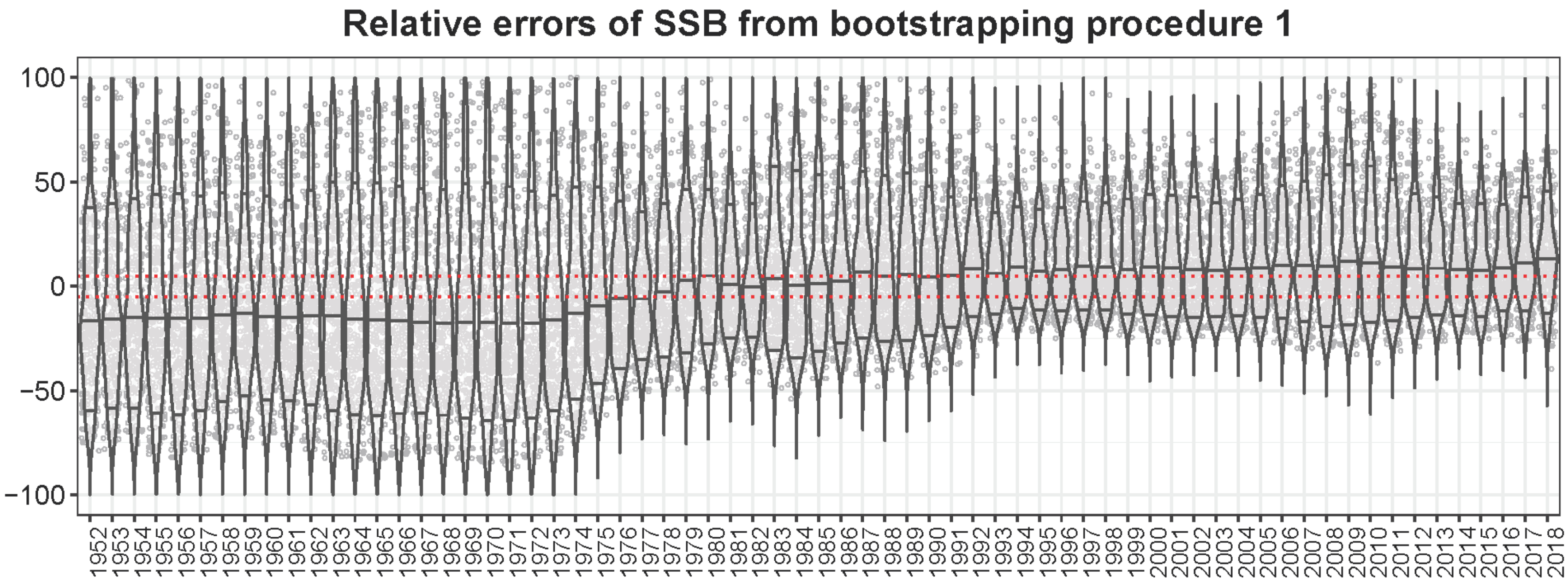


Fig. 7. Relative errors of spawning stock biomass from the bootstrapping procedure 1 (violins), where the horizontal lines in each violin indicated the 0.1 percentile, median, and 0.9 percentile. The red dotted lines indicated the -5% and 5% relative errors.

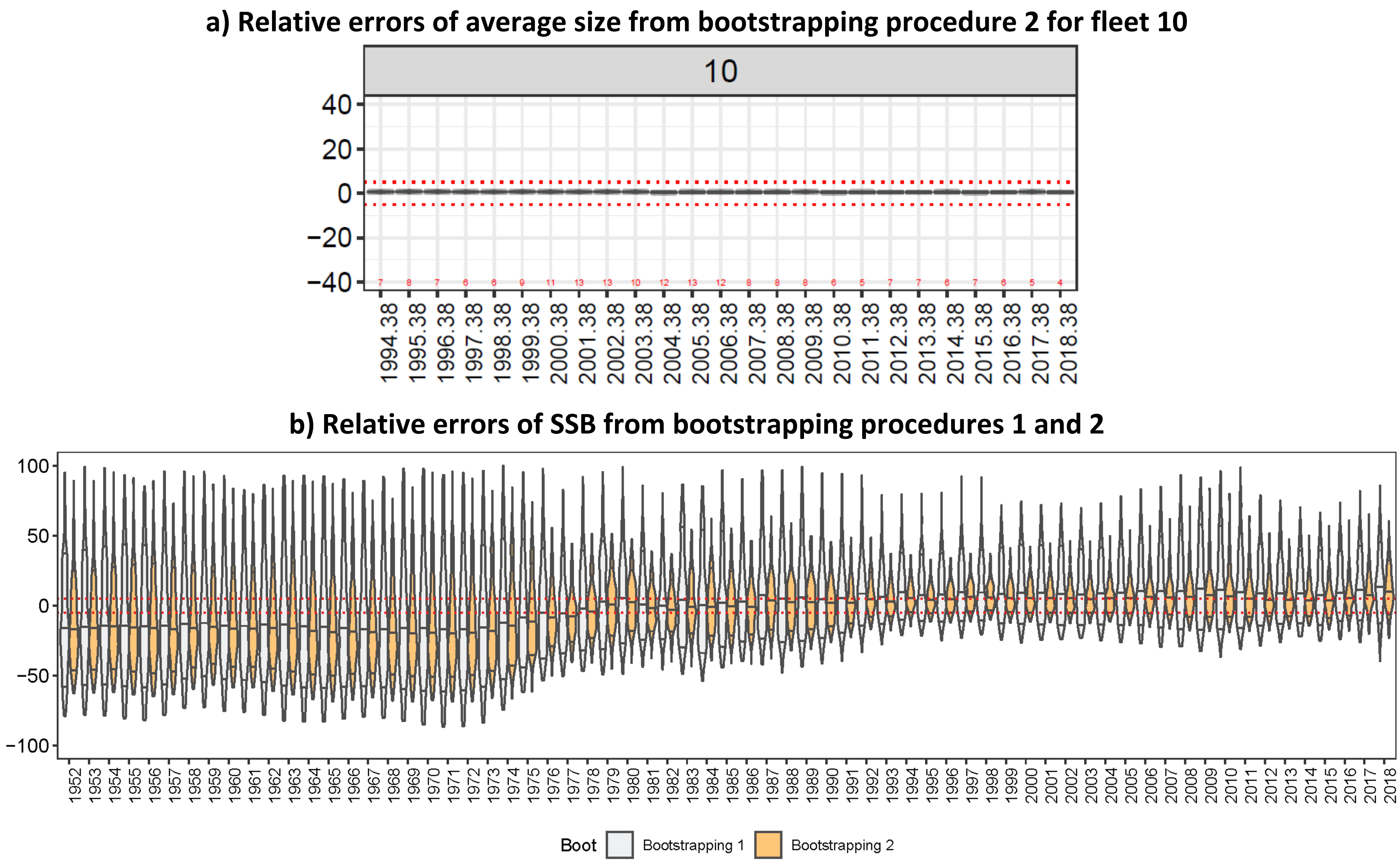


Fig. 8. Relative errors of (a) average size for fleet 10 generated from the bootstrapping procedure 1 and (b) of spawning stock biomass from the bootstrapping procedure 1 (grey violins) and 2 (orange violins), where the horizontal lines in each violin indicated the 0.1 percentile, median, and 0.9 percentile. The red dotted lines indicated the -5% and 5% relative errors.

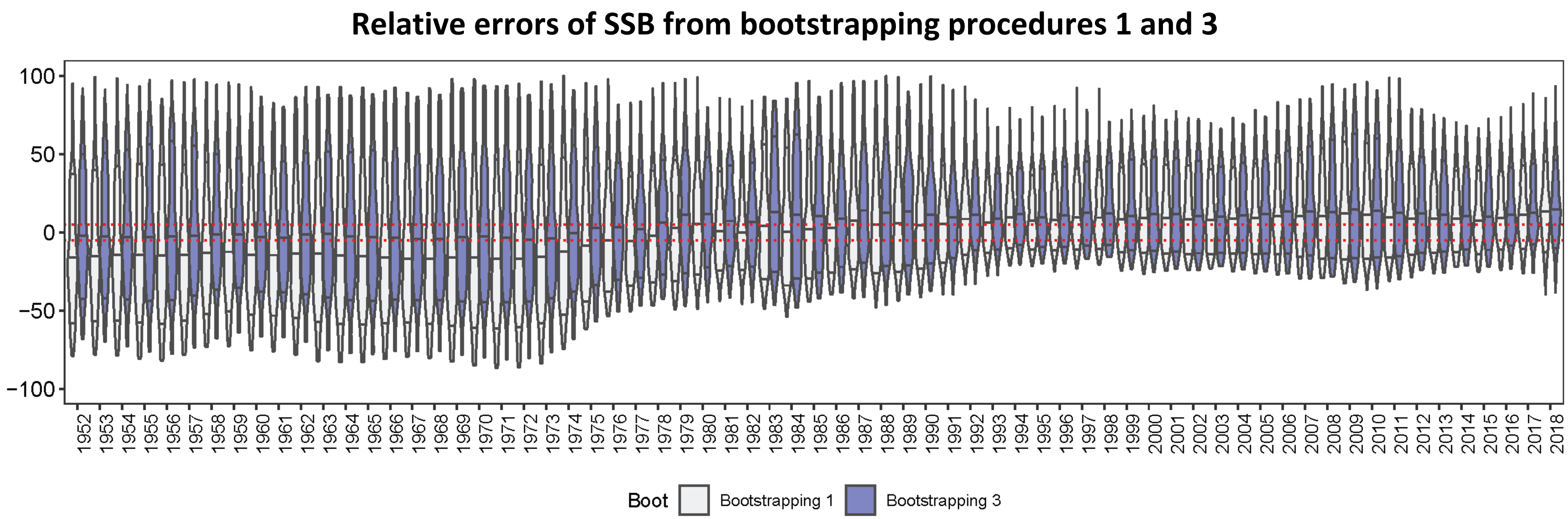
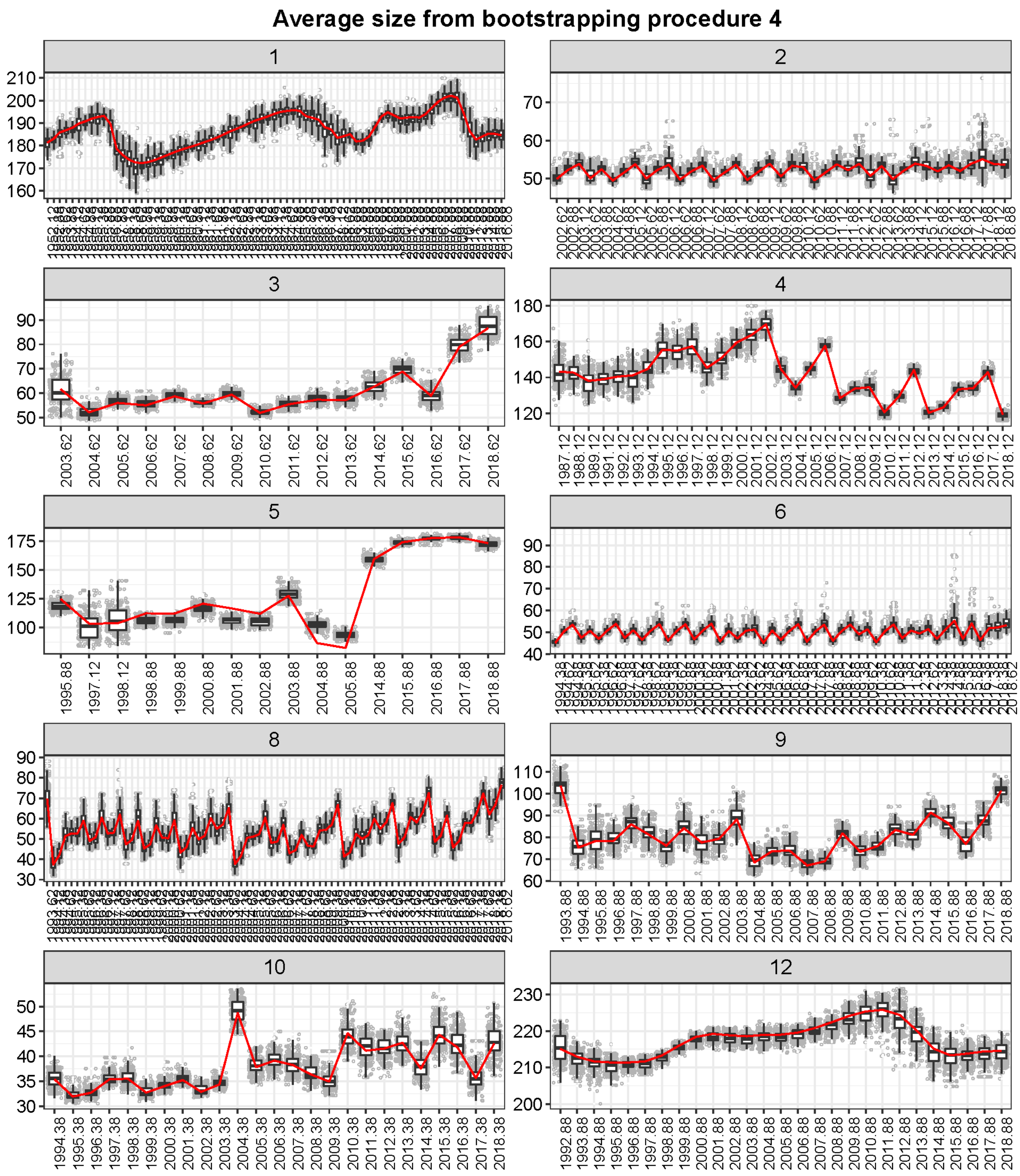


Fig. 9. Relative errors of spawning stock biomass from the bootstrapping procedure 1 (grey violins) and 3 (blue violins), where the horizontal lines in each violin indicated the 0.1 percentile, median, and 0.9 percentile. The red dotted lines indicated the -5% and 5% relative errors.


Average size from bootstrapping procedure 4
1
2
3
4
5
6
8
9
10
12

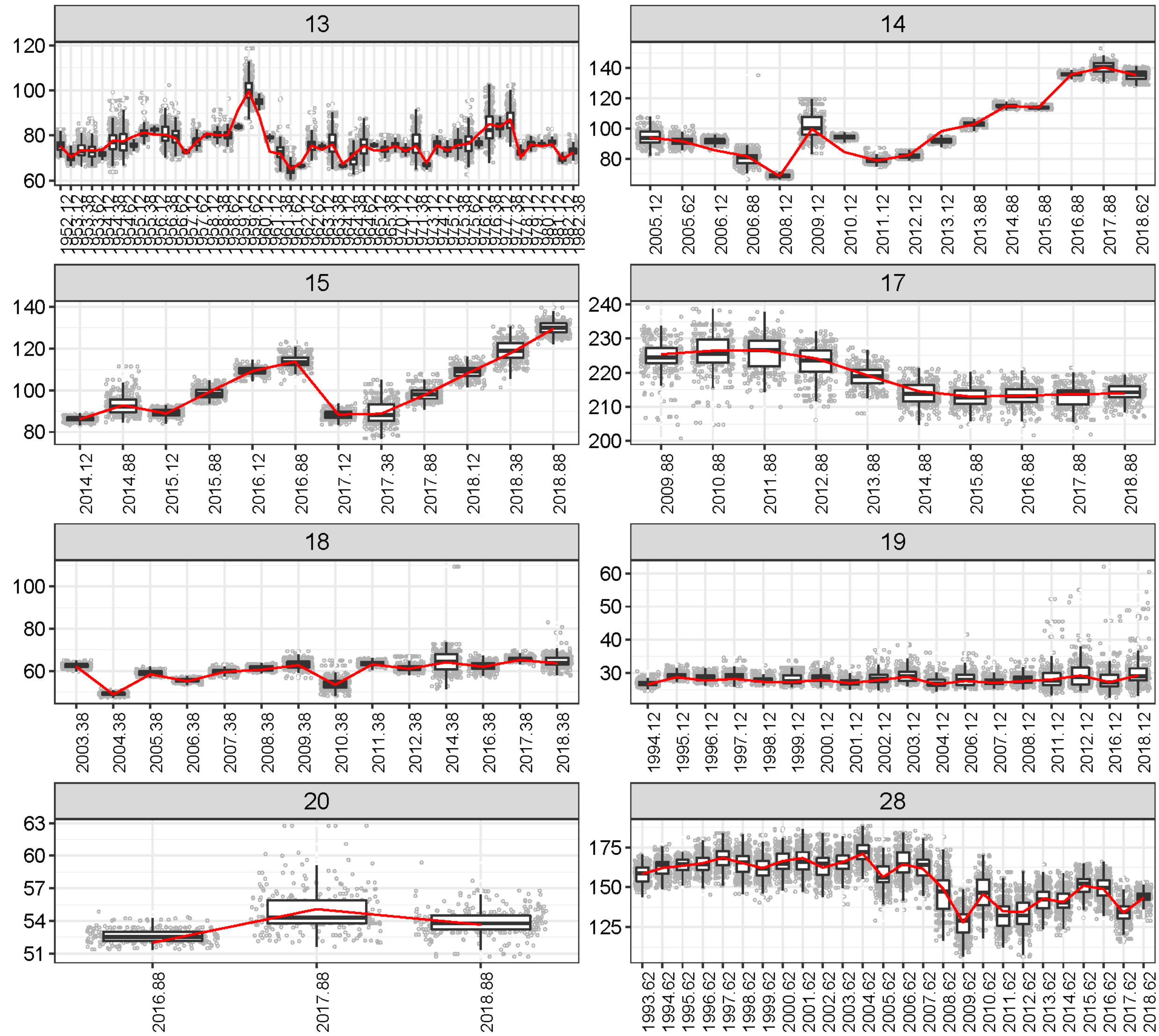


Fig. 10. Average sizes for each fleet generated from the bootstrapping procedure 4 (violins), where the horizontal line in each violin indicated the median value and the red solid lines indicated the expected average size in the 2020 Pacific Bluefin tuna stock assessment model.

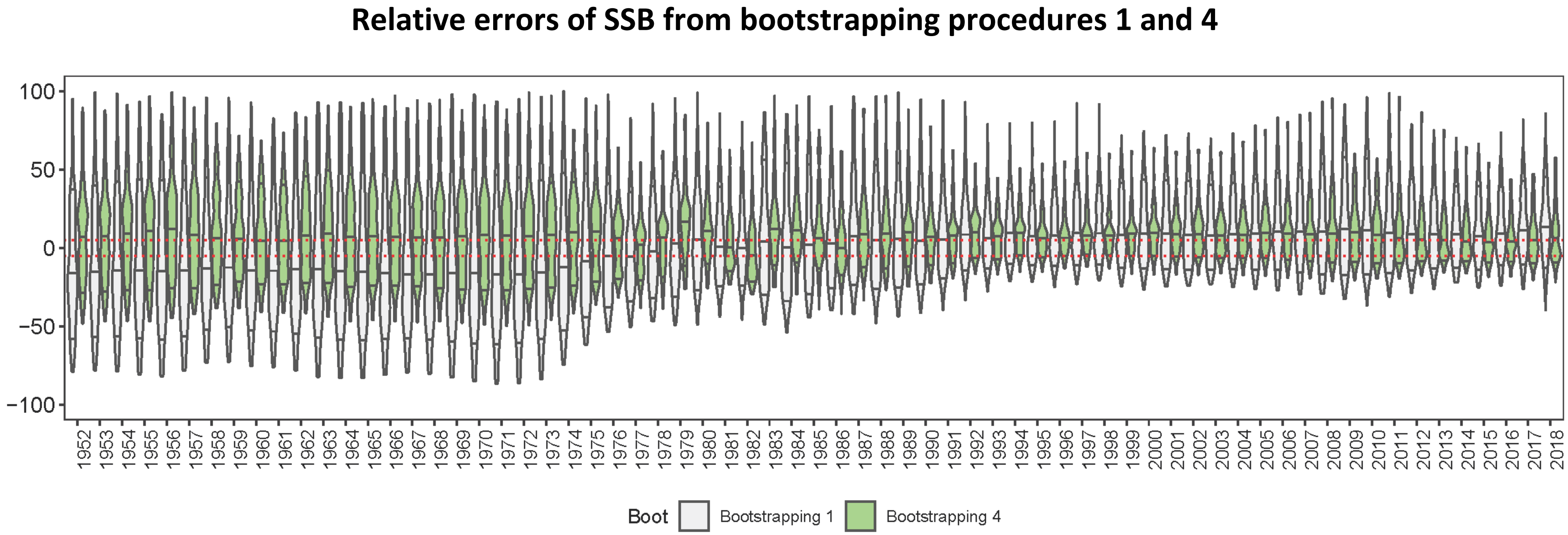


Fig. 11. Relative errors of spawning stock biomass from the bootstrapping procedure 1 (grey violins) and 4 (green violins), where the horizontal lines in each violin indicated the 0.1 percentile, median, and 0.9 percentile. The red dotted lines indicated the -5% and 5% relative errors.

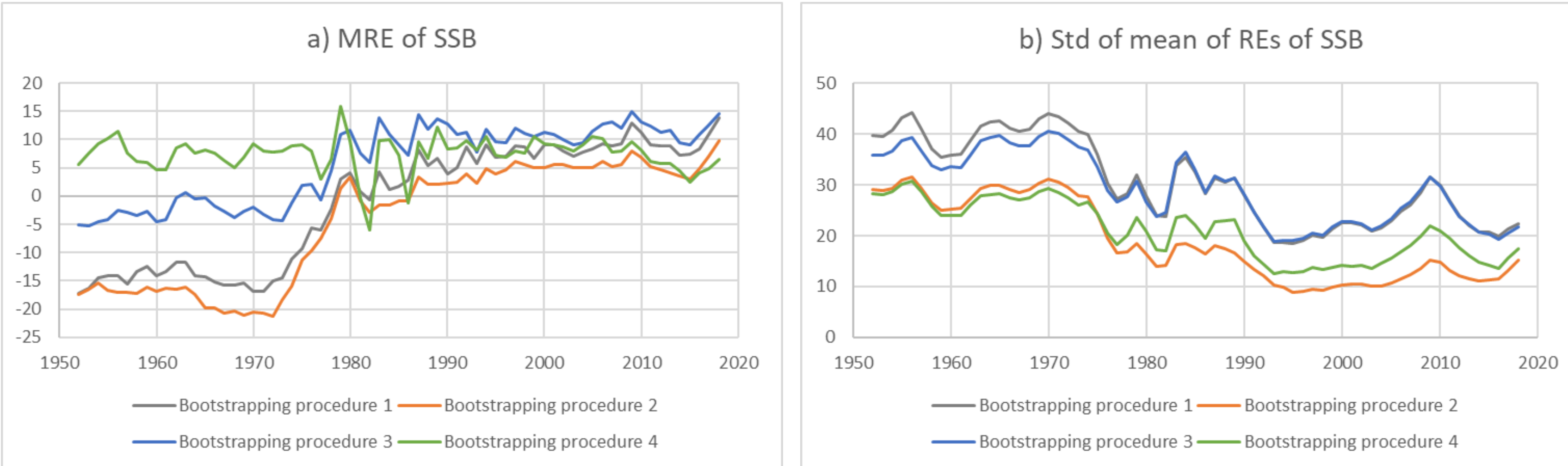


Fig. 12. Median (panel a) and standard deviation of mean (panel b) of relative errors of spawning stock biomass from the bootstrapping procedures 1 (grey violins), 2 (orange violins), 3 (blue violins), and 4 (green violins).

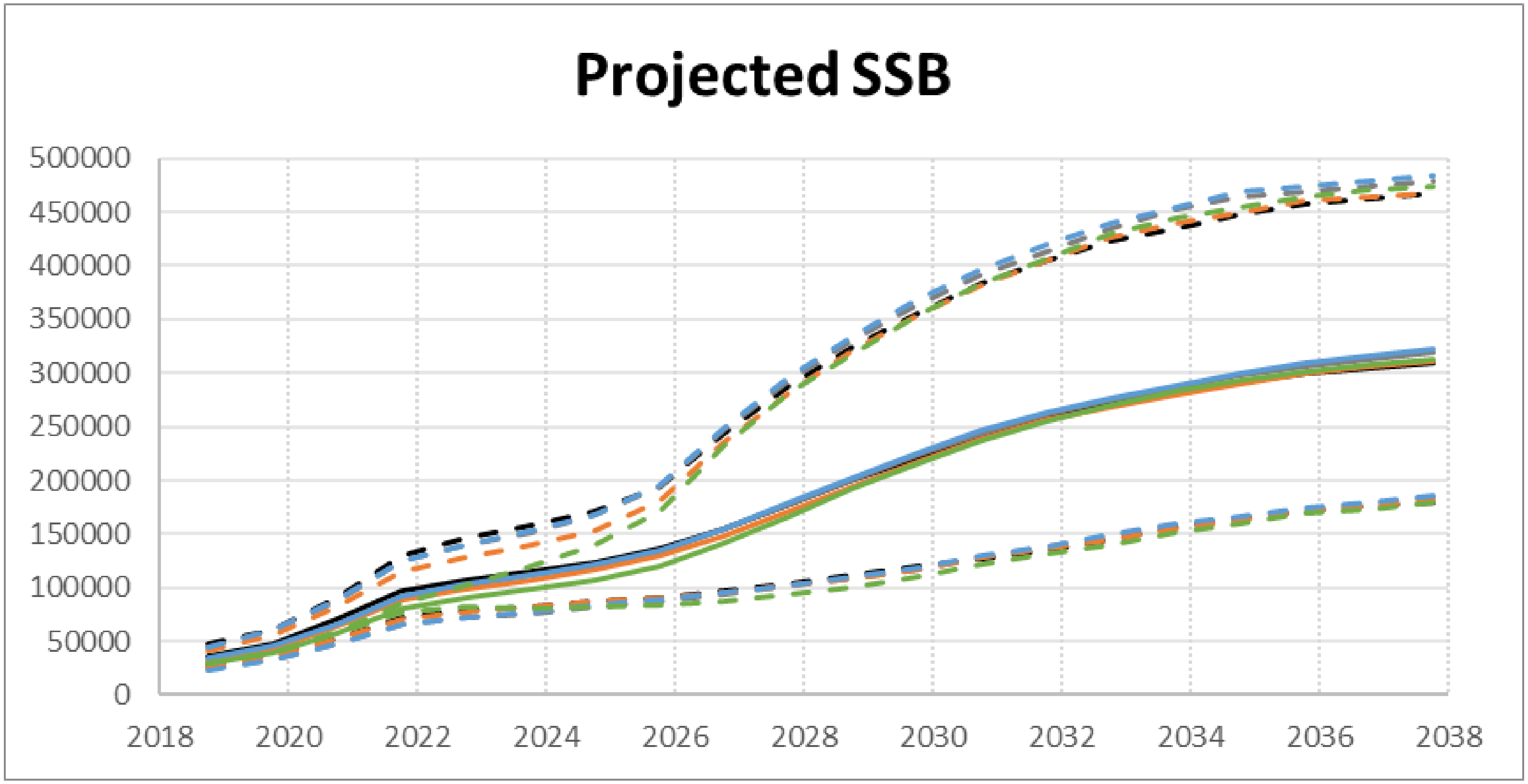


Fig. 13. Comparisons of projected SSB from the 2020 basecase model (no ad-hoc biased adjustment) (black lines), bootstrapping procedures 1 (grey lines), 2 (orange lines), 3 (blue lines), and 4 (green lines), where solid lines indicate median projected SSB and dash lines indicate associated 90% confidence intervals.