The SSB difference between base case model and bootstrapping result

Kirara Nishikawa, Yohei Tsukahara and Hiromu Fukuda

Highly Migratory Resources Division, Fisheries Resources Institute, Japan Fisheries Research and Education Agency 2-12-4, Fukuura, Kanazawa-ku, Yokohama, Kanagawa 236-8648, JAPAN

April 2025

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 14 to 18 April 2025, La Jolla, United States of America.

Introduction

During the PBF stock assessment meeting in Feb/March 2024, the PBFWG also conducted the future projections based on the base case (BC) model of the assessment. The uncertainty in the future projection analysis was considered by the combination of the uncertainties came from parameter estimation and from the future recruitment, and these uncertainties were estimated by the bootstrapping method. However, the point estimates of SSB from the BC and the median SSB from the bootstrapping replicates had some discrepancy. This discrepancy was already observed at the time of the 2020 stock assessment (Fukuda et al., 2020). To reduce this kind of discrepancy, the WG conducted ad-hoc bias correction in the 2020 assessment.

In the following assessment in 2022, a detailed analysis was performed and a new bootstrapping methodology to reduce the bias, which was changing the added minimum constant to the composition data bins as well as changing input sample size of the composition data, was developed (Lee et al., 2021) and had reduced the bias to some extent. This additional process to remove the bias was applied in 2024 stock assessment as well. However, bias between the BC point estimates and bootstrapping medians became obvious once again.

In this document, the authors compared the results between BC from stock assessment in 2024 and bootstrapping and investigated the source.

Method

We examined a series of model runs with the different input data (or data files) to investigate the possible reason of the bias. One is the BC from 2024 stock assessment, and another is a model with the expected data distributions in the BC estimation generated by bootstrapping function in the stock synthesis (data_expval.ss), hereafter called expected model. The other is a model with the 300 replicates provided from the bootstrapping function resampled based on numbers of bins for each size composition data in the stock synthesis. To evaluate the contribution of the size composition data for the bootstrapping bias, we also conducted the bootstrapping replicates based on the ASPM-R from 2024 stock assessment diagnostics. For the fully integrated model other than ASPM-R, residuals for size compositions were picked up from each of BC, expected model, and bootstrapping replicate, and the distribution of that by fleet were summarized.

Result

A comparison between the SSB point estimates from the BC and the expected model, and median value from all bootstrapping replicates were shown in **figure 1 for both the fully integrated model and ASPM-R**. For the fully integrated model, the point estimates from BC showed lower values than the point estimates from the expected model and the median values of the bootstrapping replicates throughout the stock assessment period. The difference between the BC point estimates and the median of the bootstrapped replicates were small when the stock size was low, and the difference between them became large when the stock size became high. The point estimates from the expected model was higher than those from BC and medians of the bootstrapping replicates for most of the years.

As for the ASPM-R, bootstrapping replicates and the point estimates from the ASPM-R BC had smaller differences than the differences between those from the fully integrated models. The comparison between ASPM-R base case, ASPM-R expected model and bootstrapped result from ASPM-R was shown in **figure 1-b**. The median value from the bootstrapped replicates of ASPM-R showed similar SSB with the point estimates from ASPM-R BC, but the ASPM-R expected model showed slightly lower values than the rest of the models (**figure 1-b**).

The distributions of SSB from bootstrap replicates (fully integrated model) were shown in **figure 2**. Basically, a unimodal distribution with slightly long tail in higher value was observed in many years. There were some distributions, which have bimodal distributions with biased peak values in low side. The distributions of SSB in 2021 and 2022 from bootstrapping replicates showed longer tails for high side of SSB.

The distributions of the Pearson residual for CPUE were shown in **figure 3**. The residuals of BC and expected model had some differences in all surveys. The residuals from the bootstrapping replicates and expected mode distributed around 0, but the residuals of BC showed some deviation from zero. As for the Japanese Longline (1993-2019), Japanese Troll and Taiwanese Longline CPUE, the trend in residuals in BC showed a changed by 3-5 years scales.

The distributions of the Pearson residuals for size composition by each fleet and by year for BC and the expected model were shown in **figure 4**. The residuals from the expected model had constant distributions over size in each year. However, the residuals from BC had some trend (**figure 4**). **Figure 5** showed summary plots of the Pearson residuals from bootstrapping replicates, BC and the expected model.

Fleet 1-14, 16, 17, 21 showed a similar trend in sum of the residuals between the expected model and median from bootstrapping replicates, but different trends for BC. In this case, residuals from BC fluctuated to be larger or smaller in consecutive some years than the expected model and bootstrapping median. Fleet 18, which was applied the weight composition data, showed completely different trends in residuals among BC, the expected model and median from bootstrapping replicates. About fleet 22, BC, the expected model and median from bootstrapping replicates showed similar trend in general.

Discussion

From the comparison of BC and ASPM-R, differences between the SSB point estimates and bootstrapped replicates in ASPM-R were smaller than those from the fully integrated model. Because the ASPM-R doesn't include the size composition data in the model, thus, the possible source of this bias might came from size composition residuals.

In the 2022 stock assessment, the bias was lightened by increasing the number of resampled size data for all fleets and adding a minimum constant of 0.0001 (instead of 0.01) for weight bin fleets (Lee et al., 2021). For Stock Synthesis version 3.30.14, the number of resampled size data by bootstrapping was based on the "input sample size". Since the stock synthesis version was updated, the method to decide the number of data resampled was changed to the way based on the number of data bins. The bias correcting method by changing the minimum constant was still used in the bootstrapping procedures for the future projection in 2024, but it might show a shortage to correct the bias.

The fleet 5, 6, 9, 10, 18, and 21 had obvious difference in the sum of residuals (> 1.0 in absolute value) between BC and the median values of the bootstrapping replicates. On the other hand, the difference between expected model and bootstrapping are generally small. Since the data replicates of bootstrapping were generated from expected model, so that they should be fitted well basically. The deviation of the BC model from the expected model and bootstrapping median might indicate the difference in error with the random sampling error expected by the SS model. Note that some of these fleets had relatively large input sample size and they were assumed to have the time varying selectivity (prioritized fleet; 5, 6, 9, 21).

On the other hand, the fleet 18, which applied weight bin, had noticeable differences of total residuals between expected model and bootstrap replicates. Accordingly, there are much larger differences among BC, expected model and the median of the bootstrap replicates in that fleet. The recent stock synthesis uses resampling method of the number of size bin. In the PBF stock assessment, the number of data length bins were 65, however, the data weight bins were 29. Thus, the number of weight observations resampled by the current bootstrap procedure was much smaller than it for the length composition fleet. This point was possible sources of the biases in the bootstrapping procedure. The method suggested by Lee et al., 2021, which increased the number of resampled data, might be effective to make smoother weight composition data and to reduce the effect of the minimum constant.

Our observation could be summarized into 2 possible issues regarding the current bootstrapping procedure.

1. The random sampling error was expected for size composition data of the bootstrap replicates, but the base case model might show a kind of trend in error. This trend is more obvious in some of the prioritized fleets, which may have relatively large input sample size. Although the current model reasonably reconciles the size composition data by assuming time varying selectivity, but it might be desirable to further reduce the annual residuals. Also, the input sample size might also related to the bias.
2. As for the weight composition fleet in PBF assessment, our result might suggest that the number of bins, which related to the number of data resampled, is a possible source of the biased error even for the expected model. For the next assessment, it might be necessary to reconsider the weight bin width.

References

- Fukuda H., Y. Tsukahara, K. Nishikawa, M. N. Maunder. 2020. A review of the issues associated with the Bootstrap calculations and future projections in the PBF assessment. ISC/20/PBFWG-1/15
- Lee H.-H., H. Fukuda, Y. Tsukahara, K. Piner, M. Maunder, and R. Methot Jr. 2021. 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. ISC/21/PBFWG-1/07

Figure 1 The distribution of SSB. (a) Top figure is from bootstrapping results based on BC, the median value of the bootstrapping, the point estimated value from BC and the point estimated value from the expected model of BC .(b) Bottom figure is from bootstrapping results based on ASPM-R model, the median value of the bootstrapping, the point estimated value from the ASPM-R model and the point estimated value from the expected value model of ASPM-R.

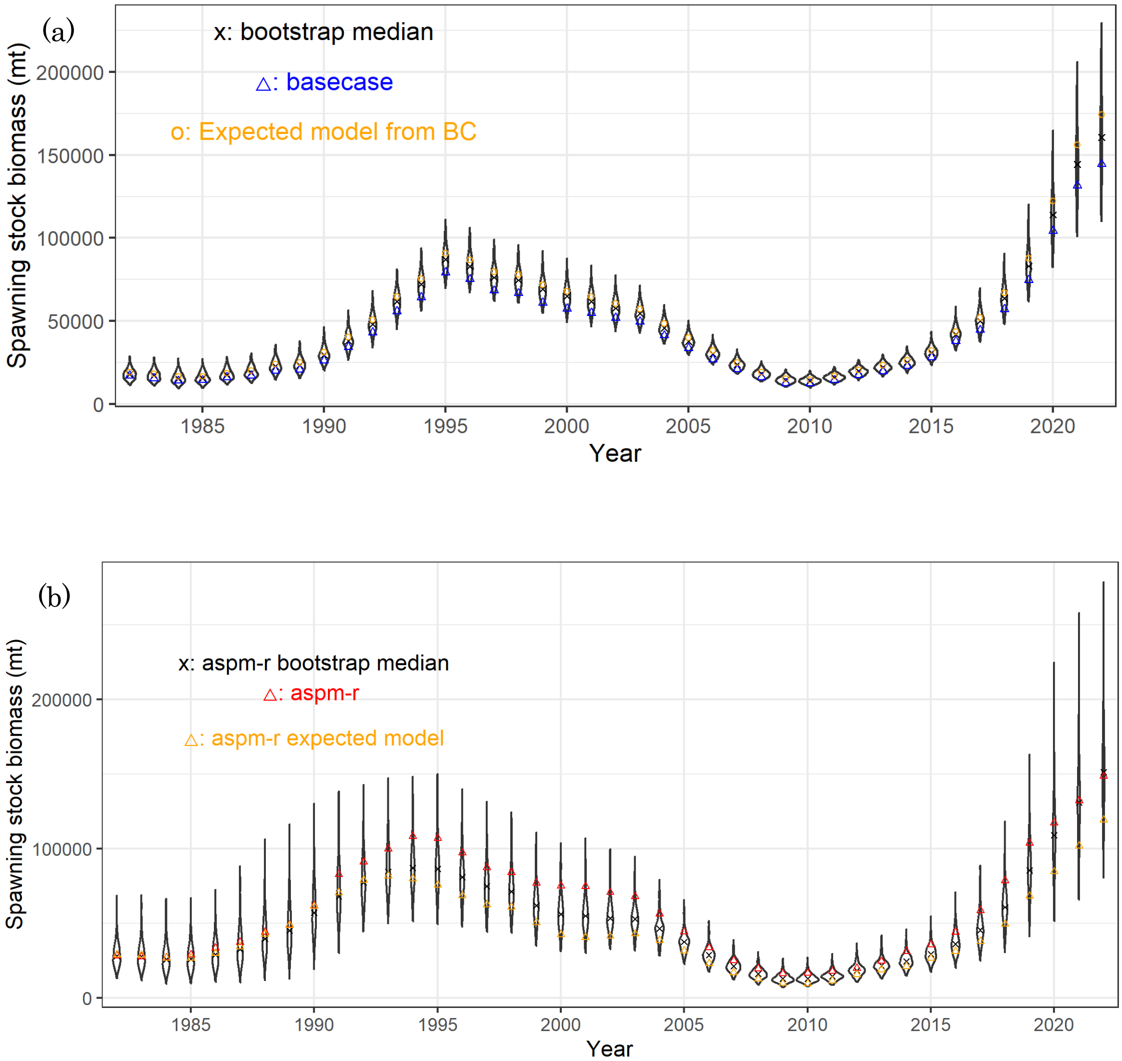

Figure 2 The distribution of SSB estimated by the bootstrapping and the values of median value of bootstrapping (black), point estimated from BC (blue) and the expected model (orange). The binwidth for all years were 1000t.

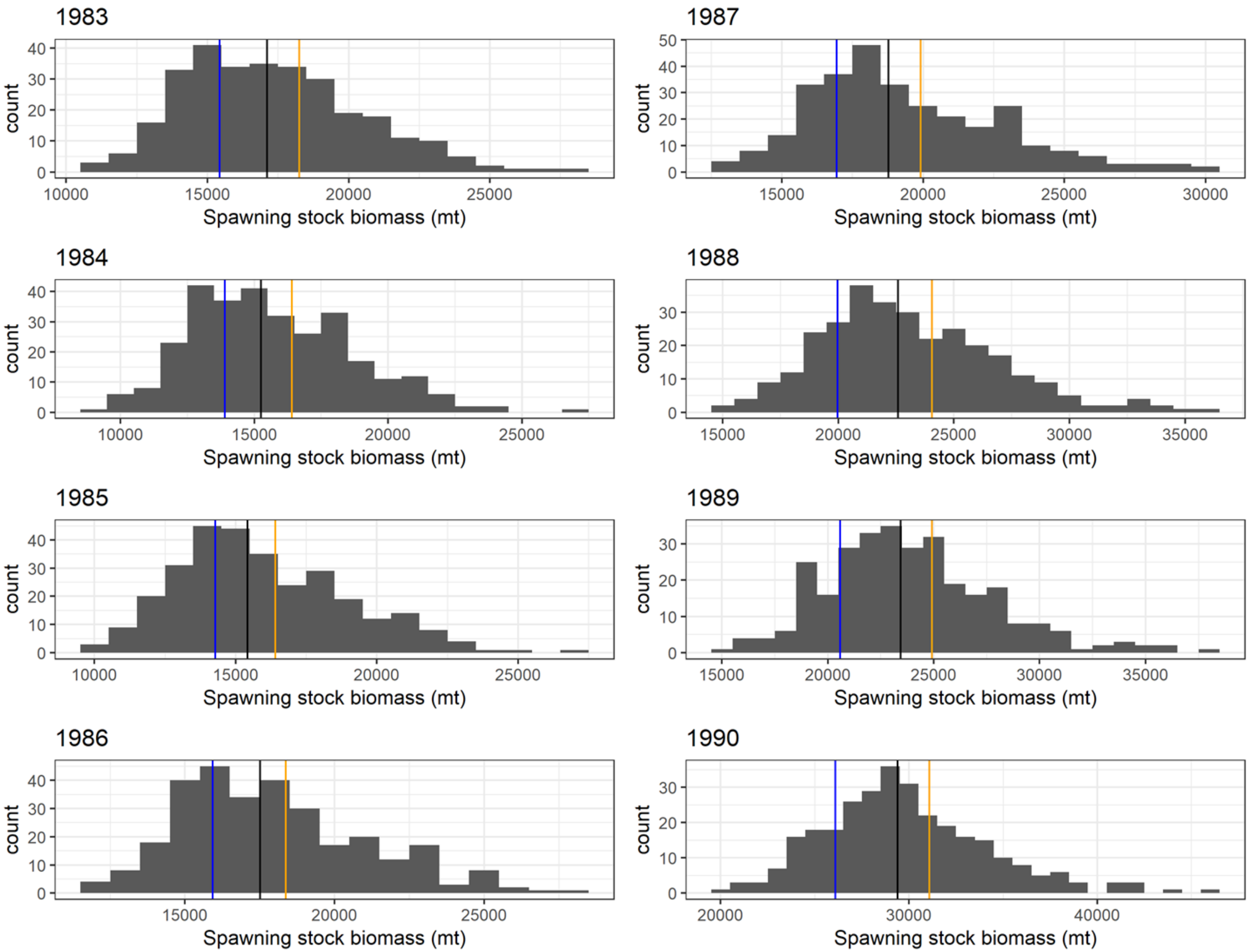

Figure 2 Continue.

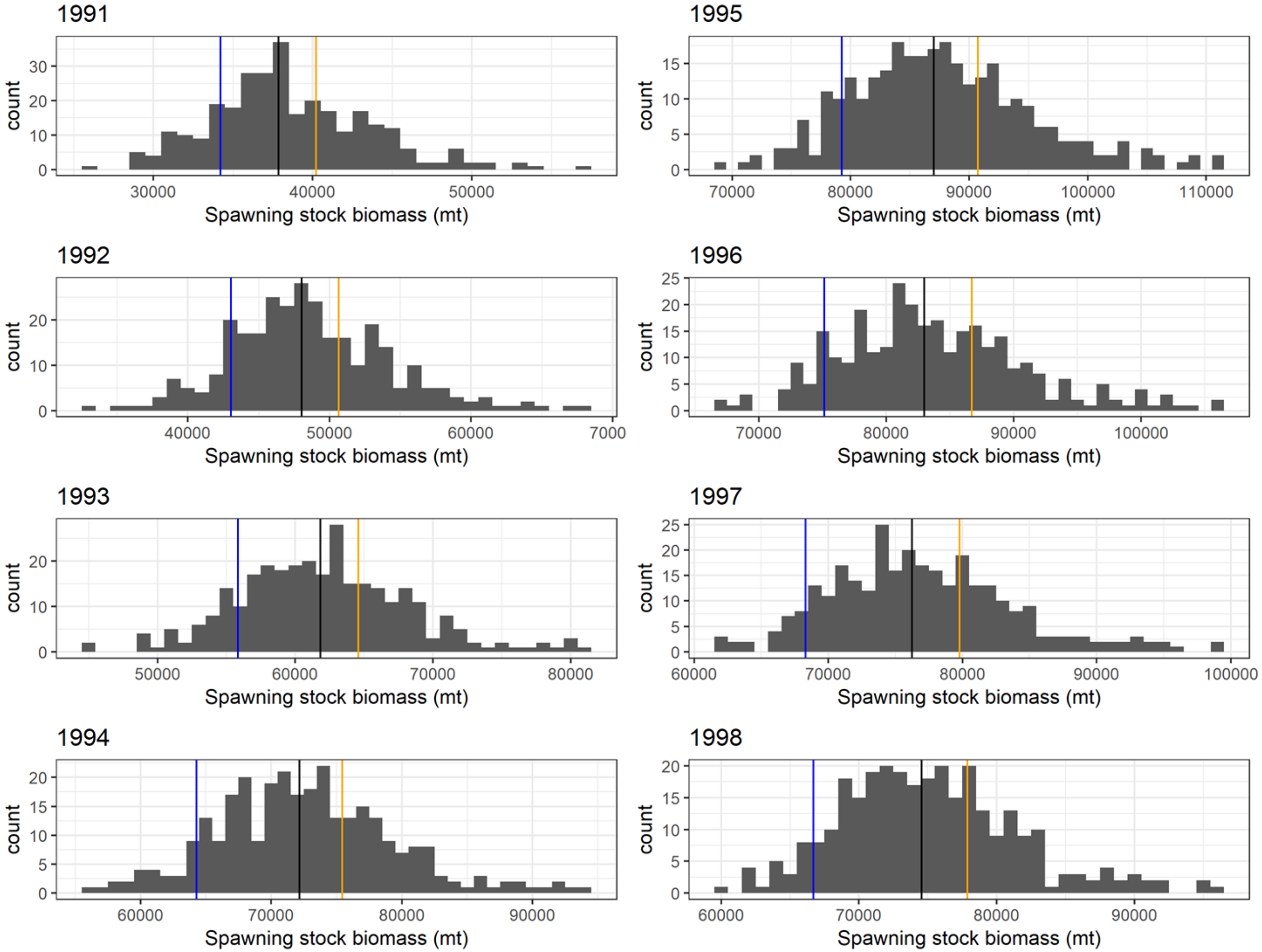

Figure 2 Continue.

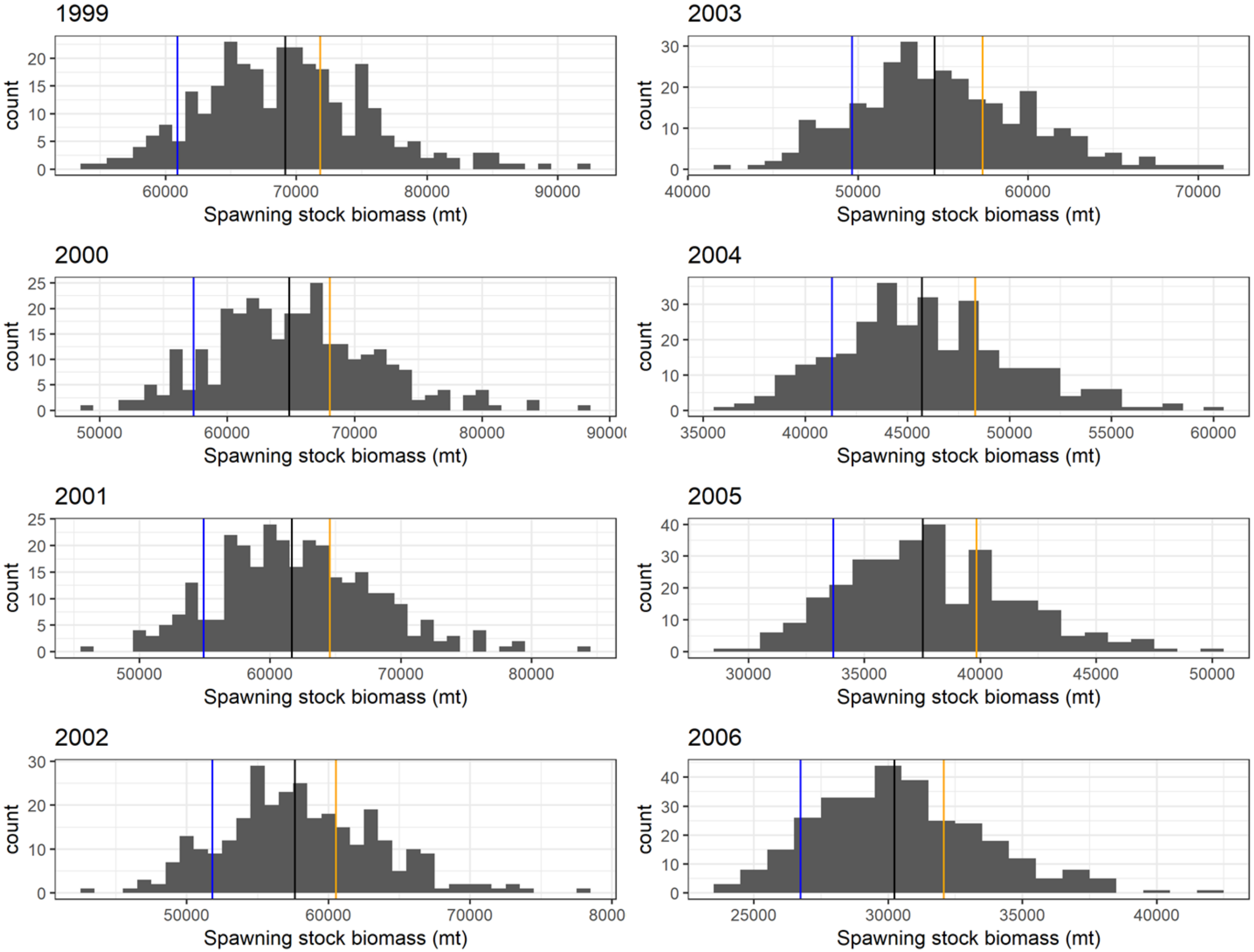

Figure 2 Continue.

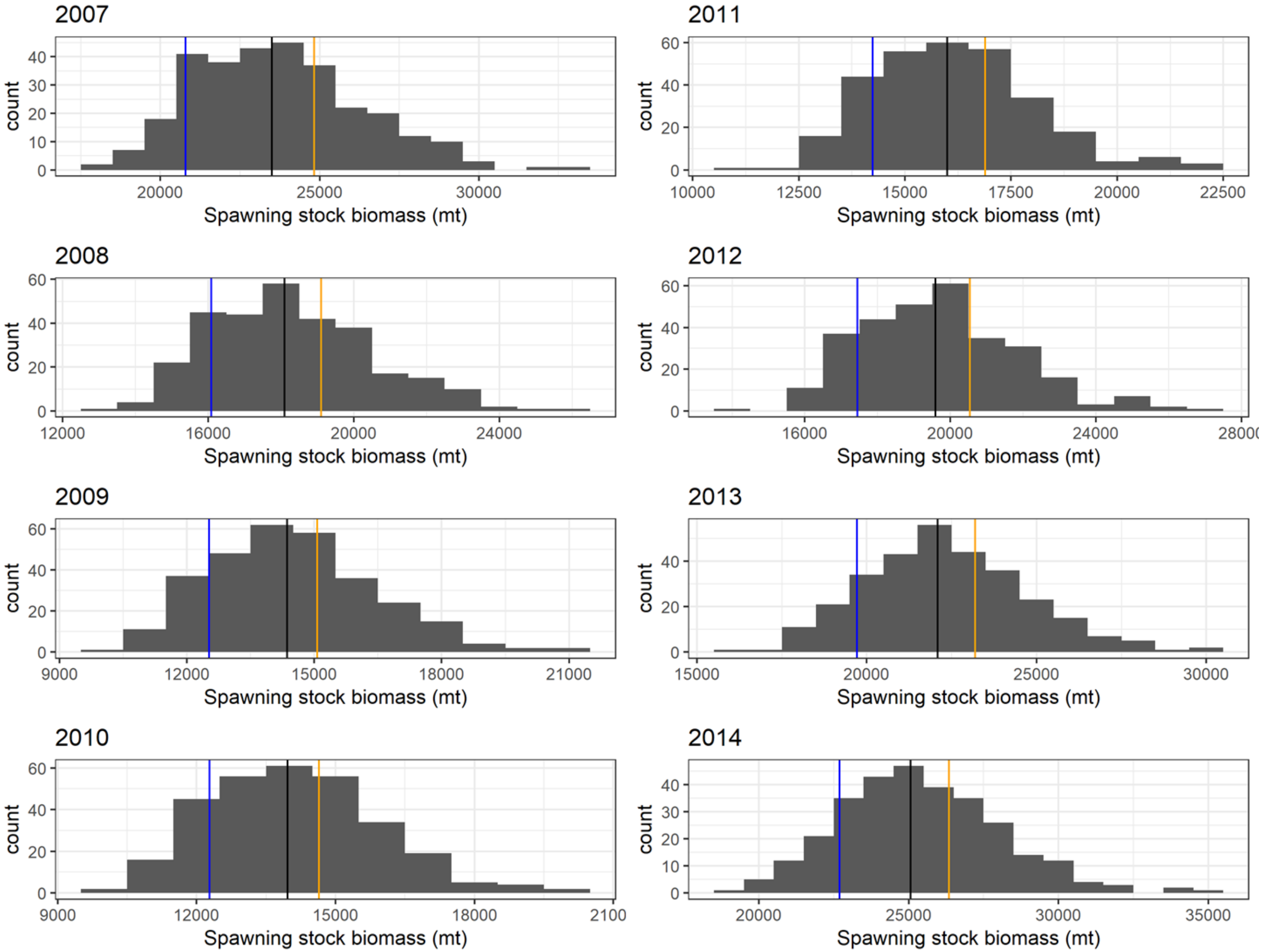

Figure 2 Continue.

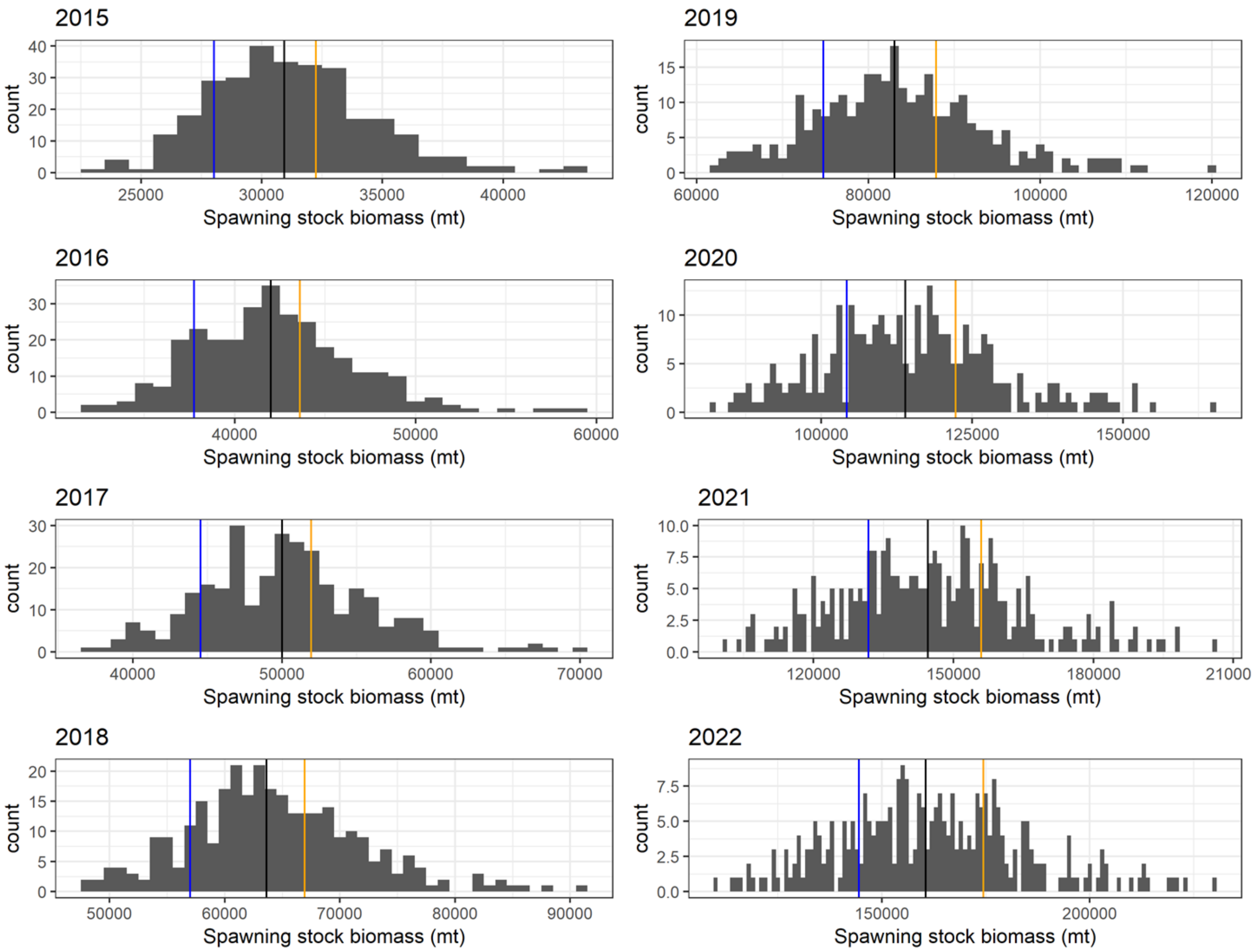

Figure 3 The comparison of the residuals (Observed value-Expected value) from CPUE. Black point indicates the median of the residuals from bootstrapping results, blue points indicate the residuals from BC and orange points indicate the residuals from the expected model.

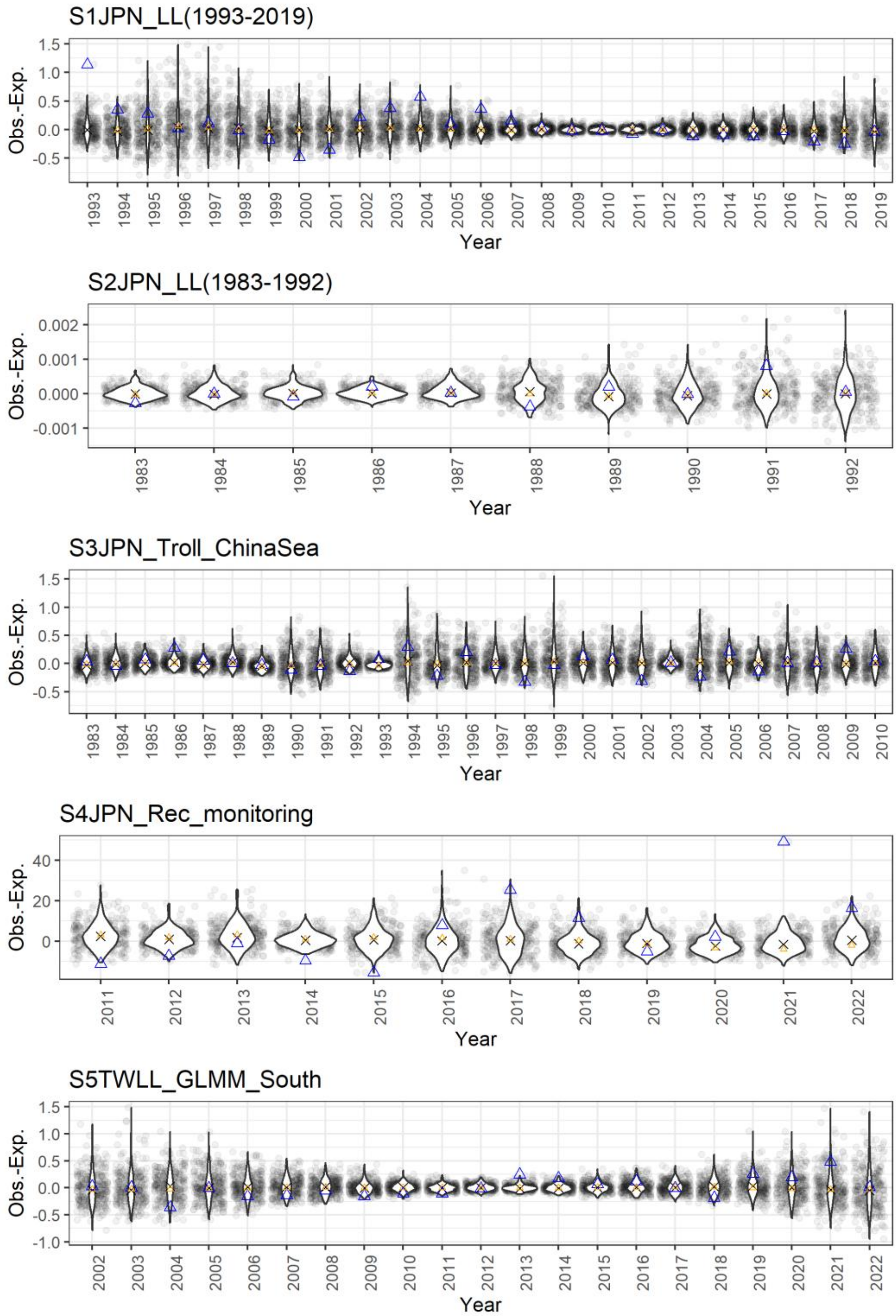

Figure 4 Bubble plot of the Pearson residuals from BC (blue points) and the expected model (orange points). Open circle indicates more than 0, and filled circle indicates less than 0.

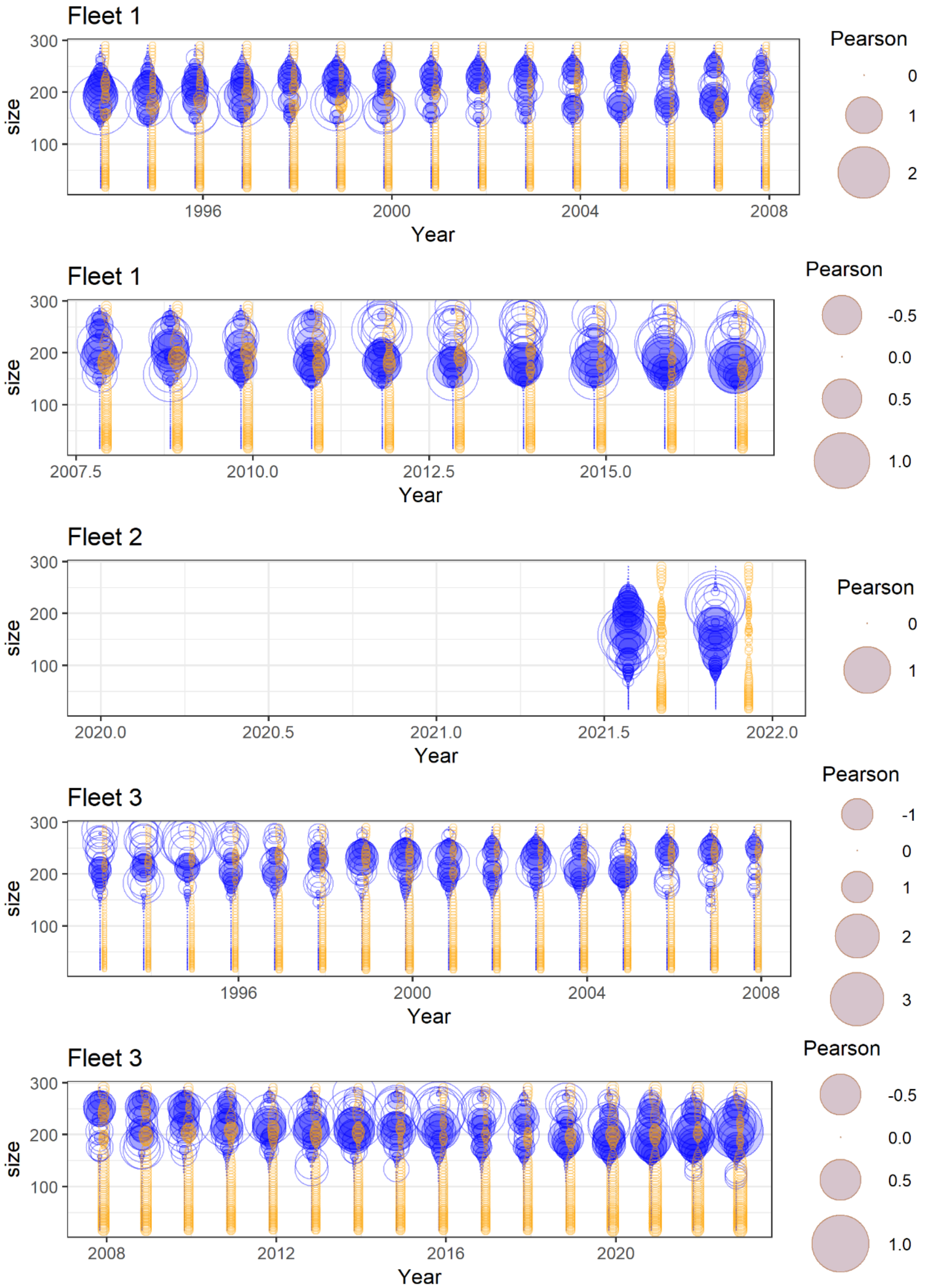

Figure 4 Continue.

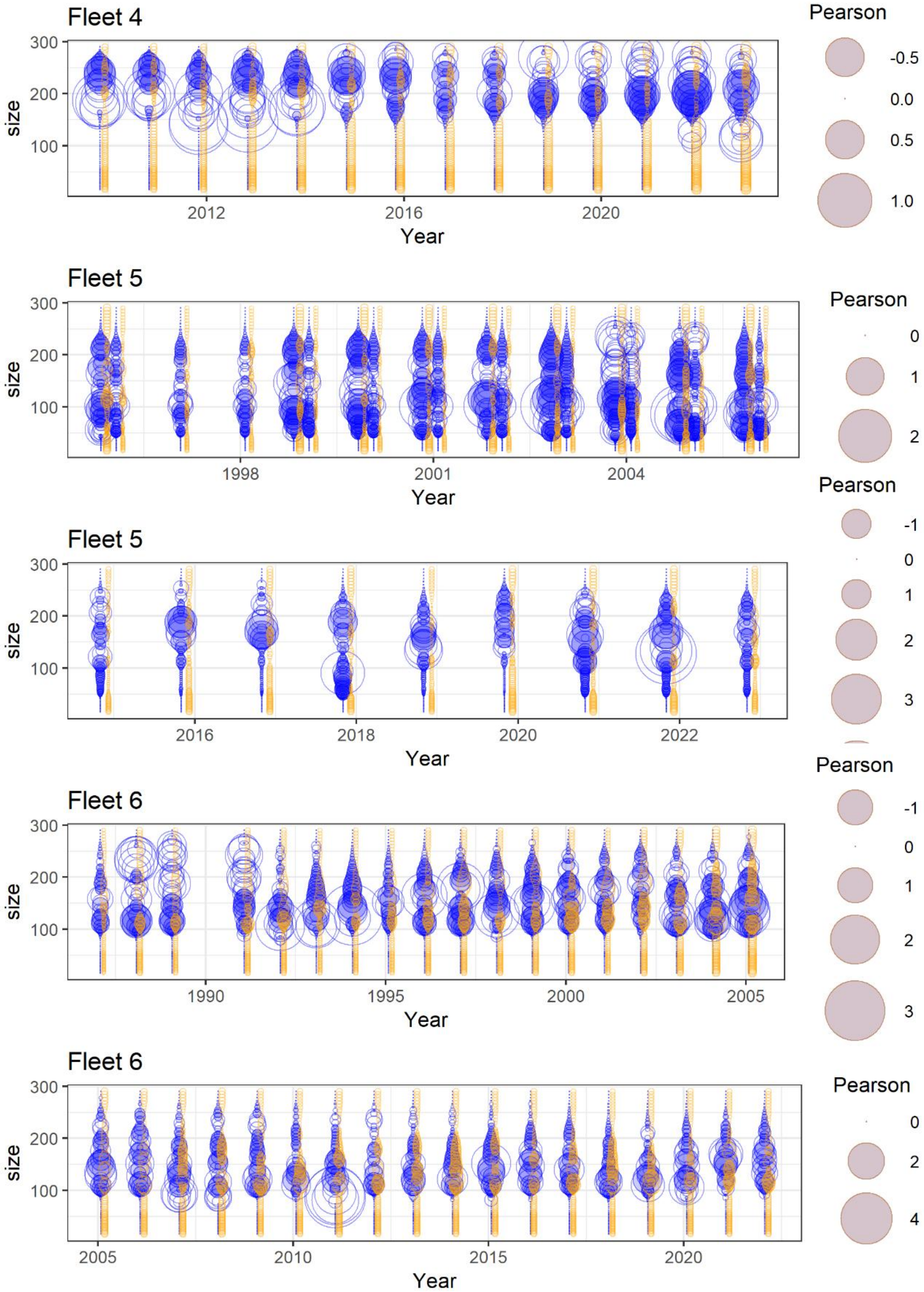

Figure 4 Continue.

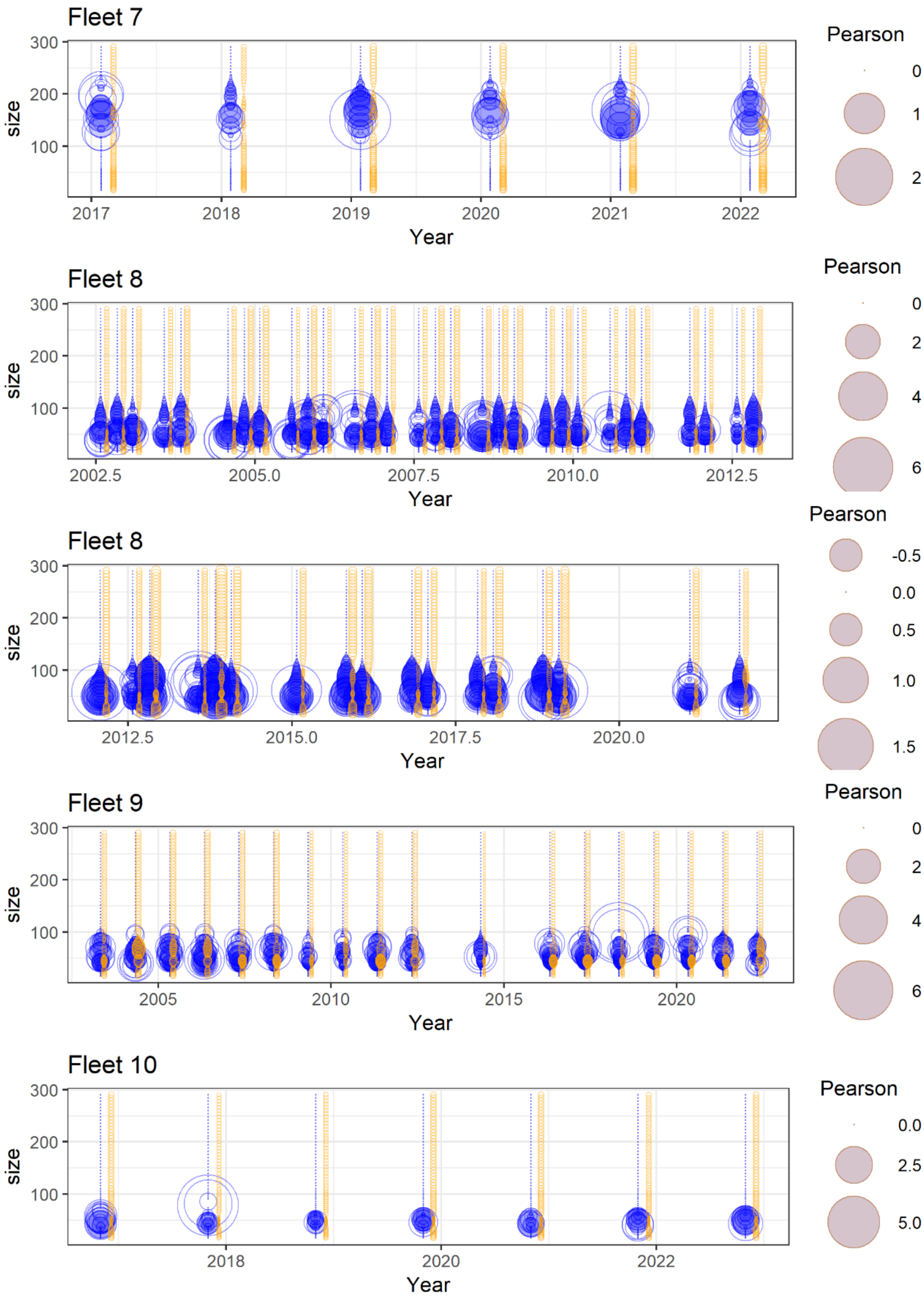

Figure 4 Continue.

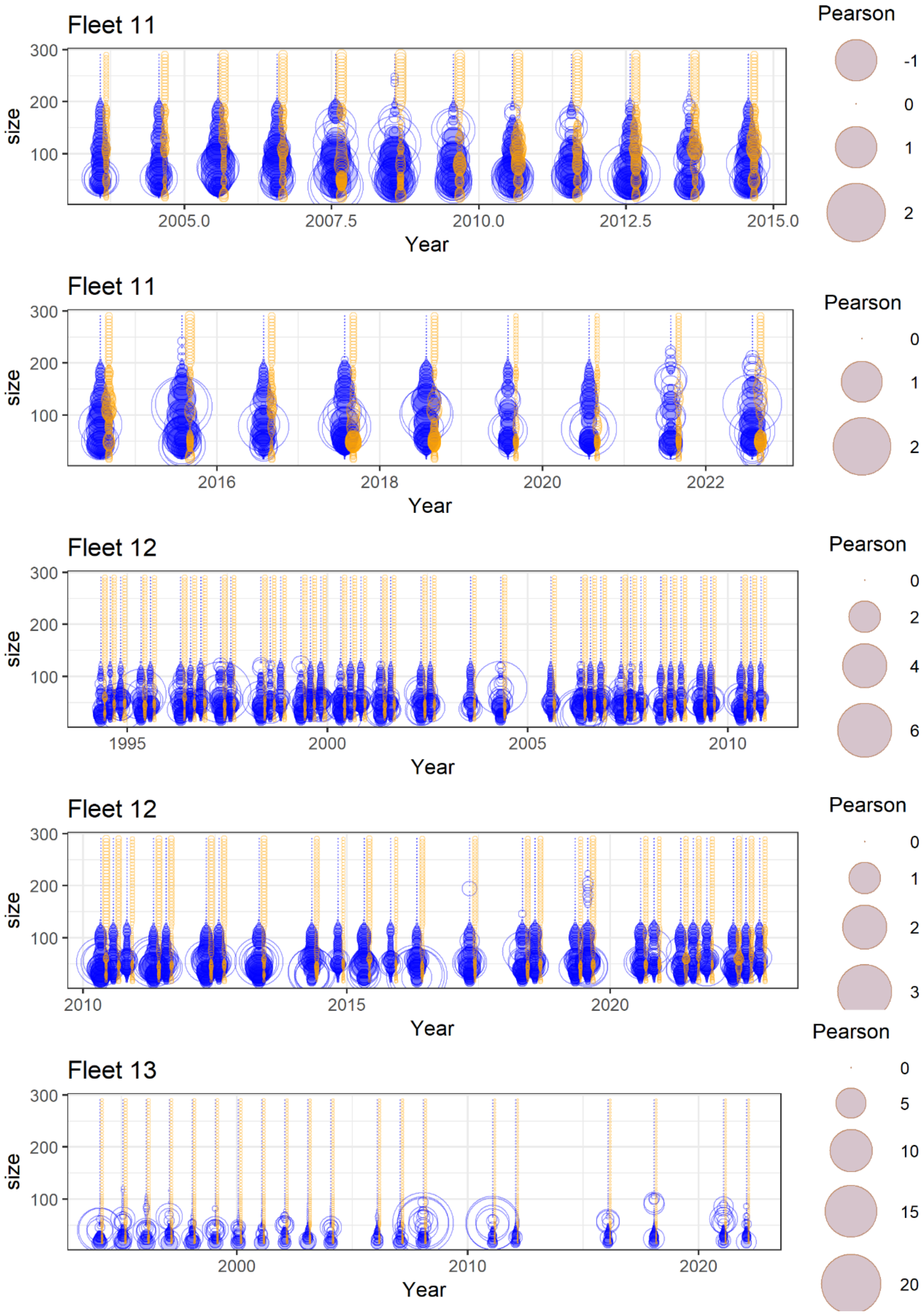

Figure 4 Continue.

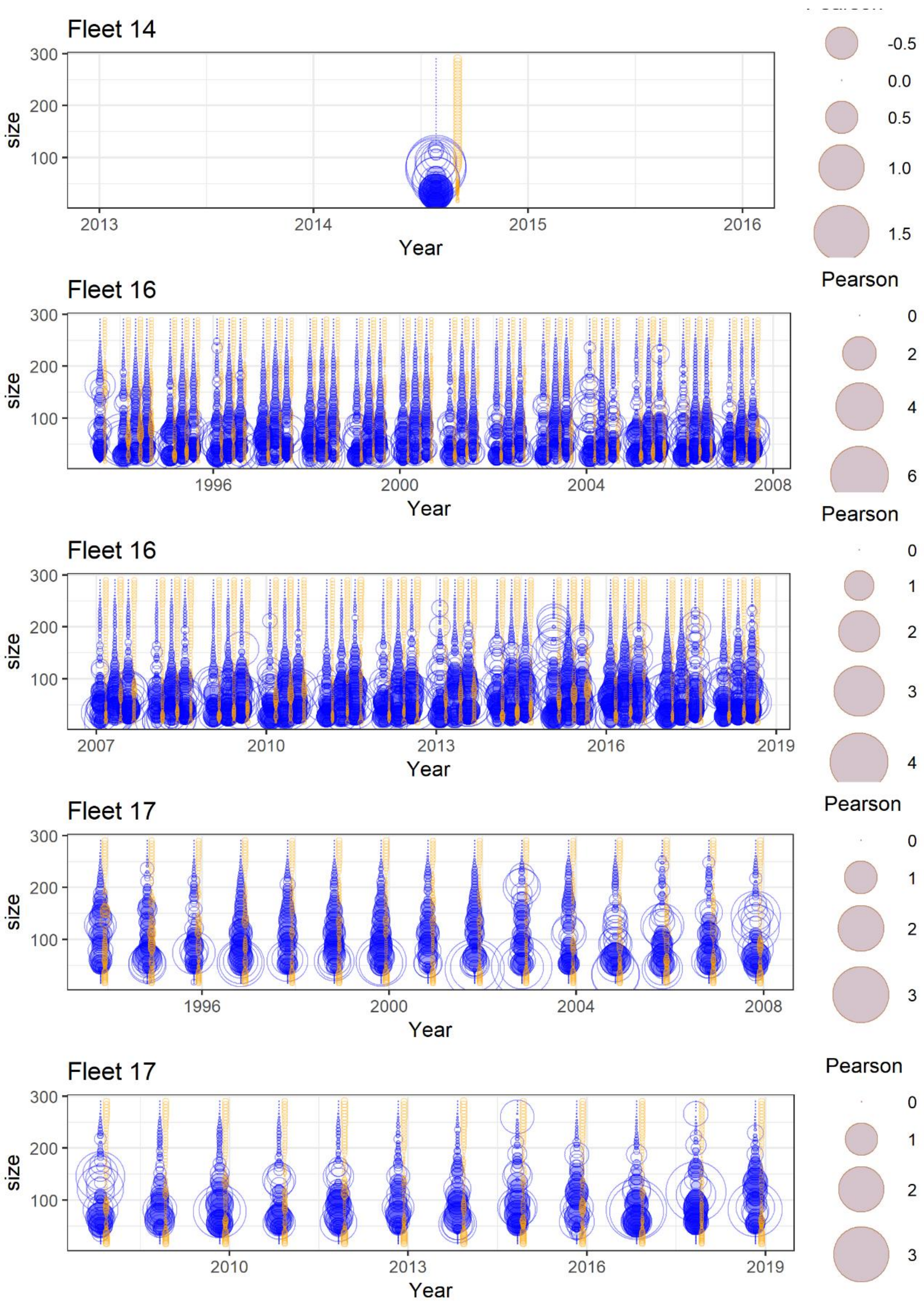

Figure 4 Continue.

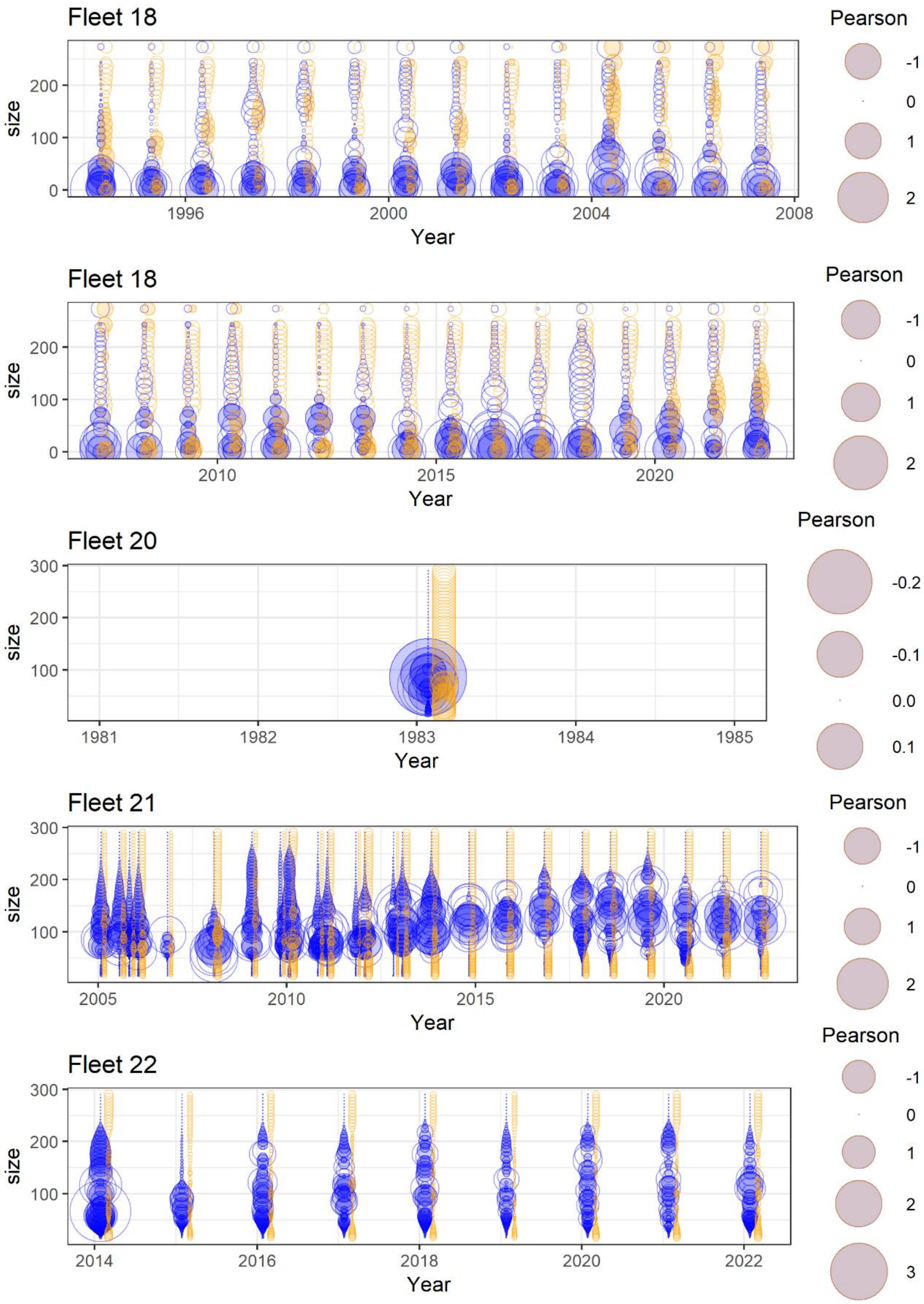

Figure 5 Summary plot of the Pearson residuals from boot strapping results(black points and violin plot), BC(blue points) and the expected model (orange points).

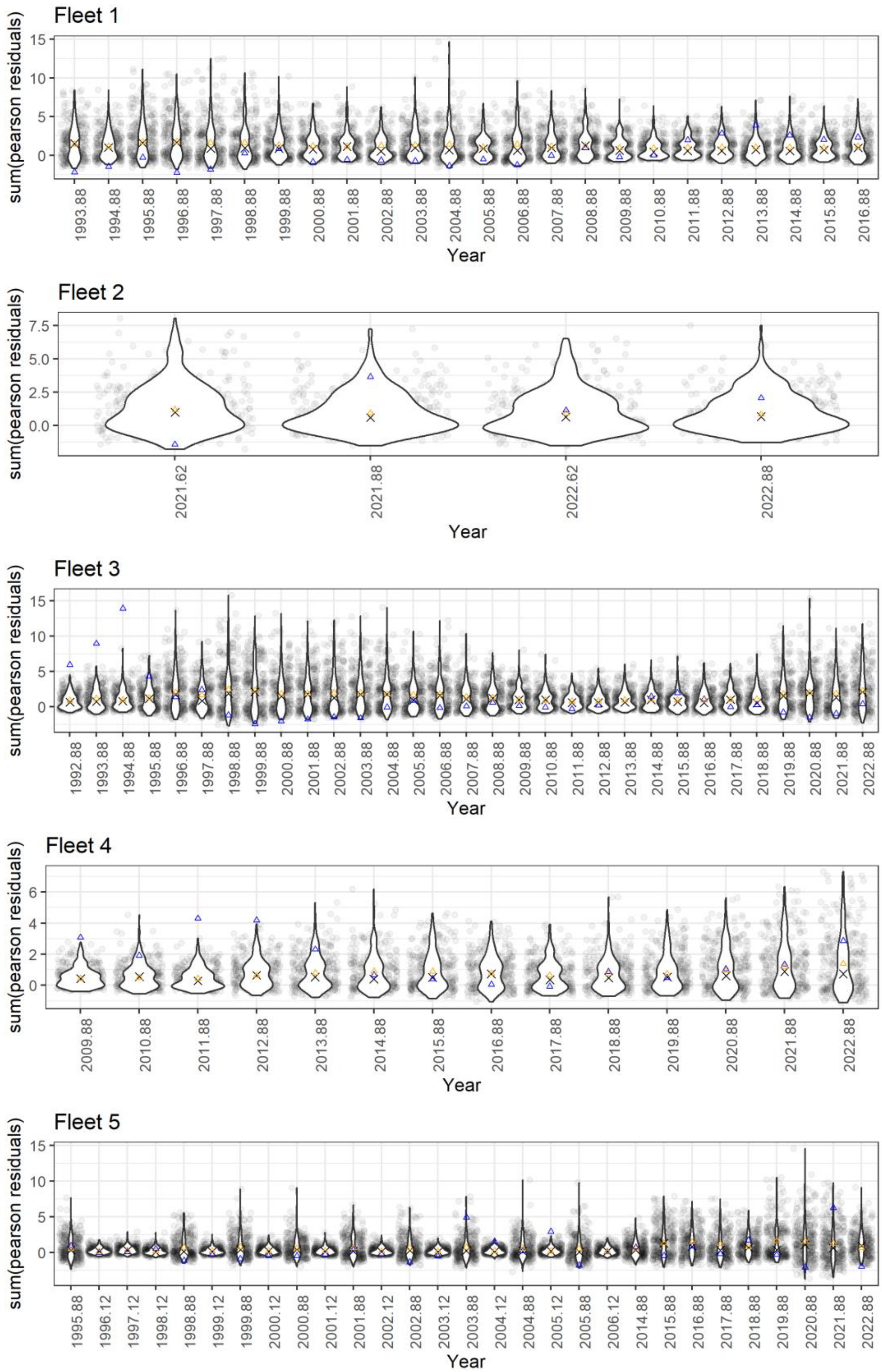

Figure 5 Continue.

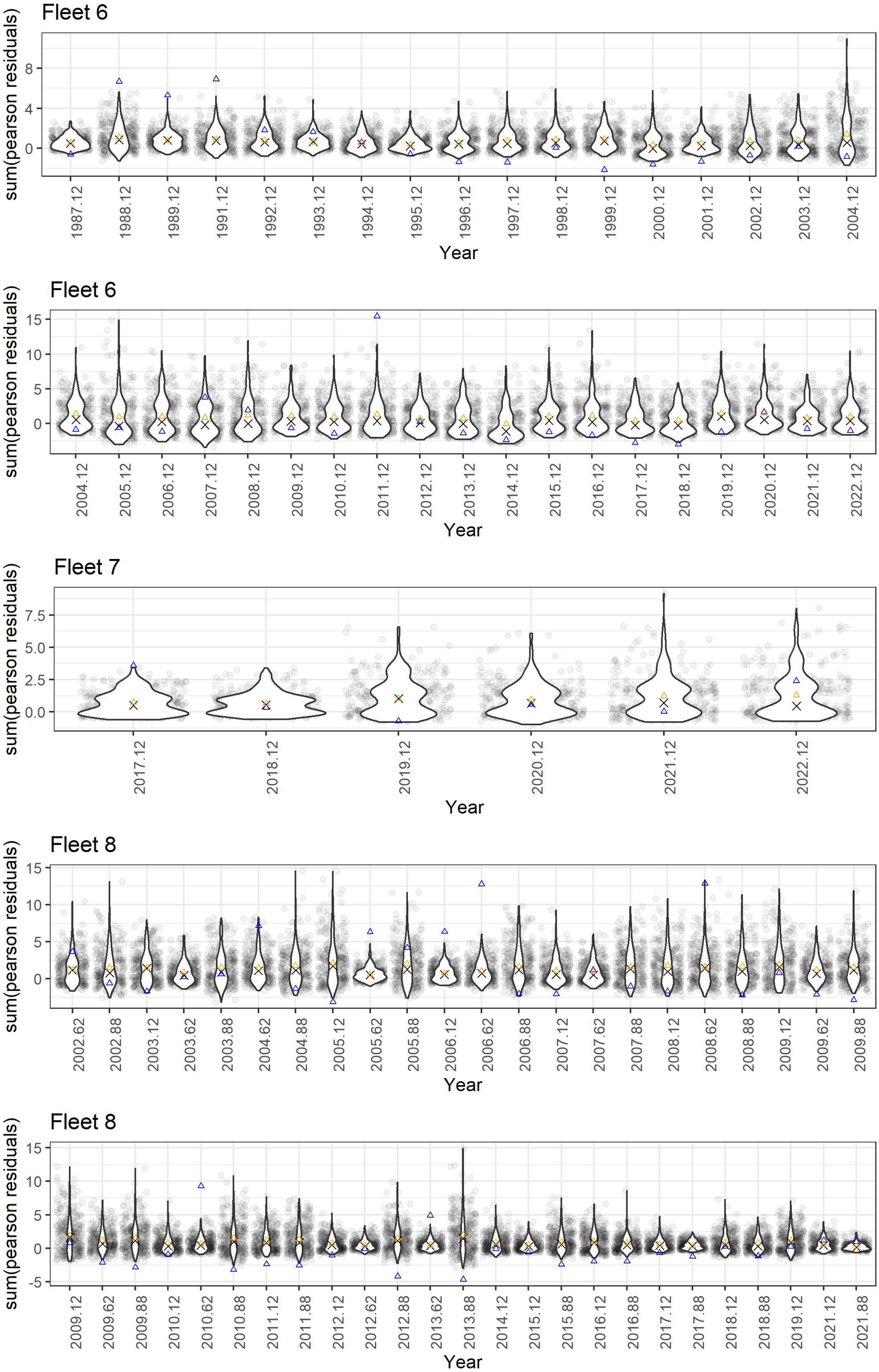

Figure 5 Continue.

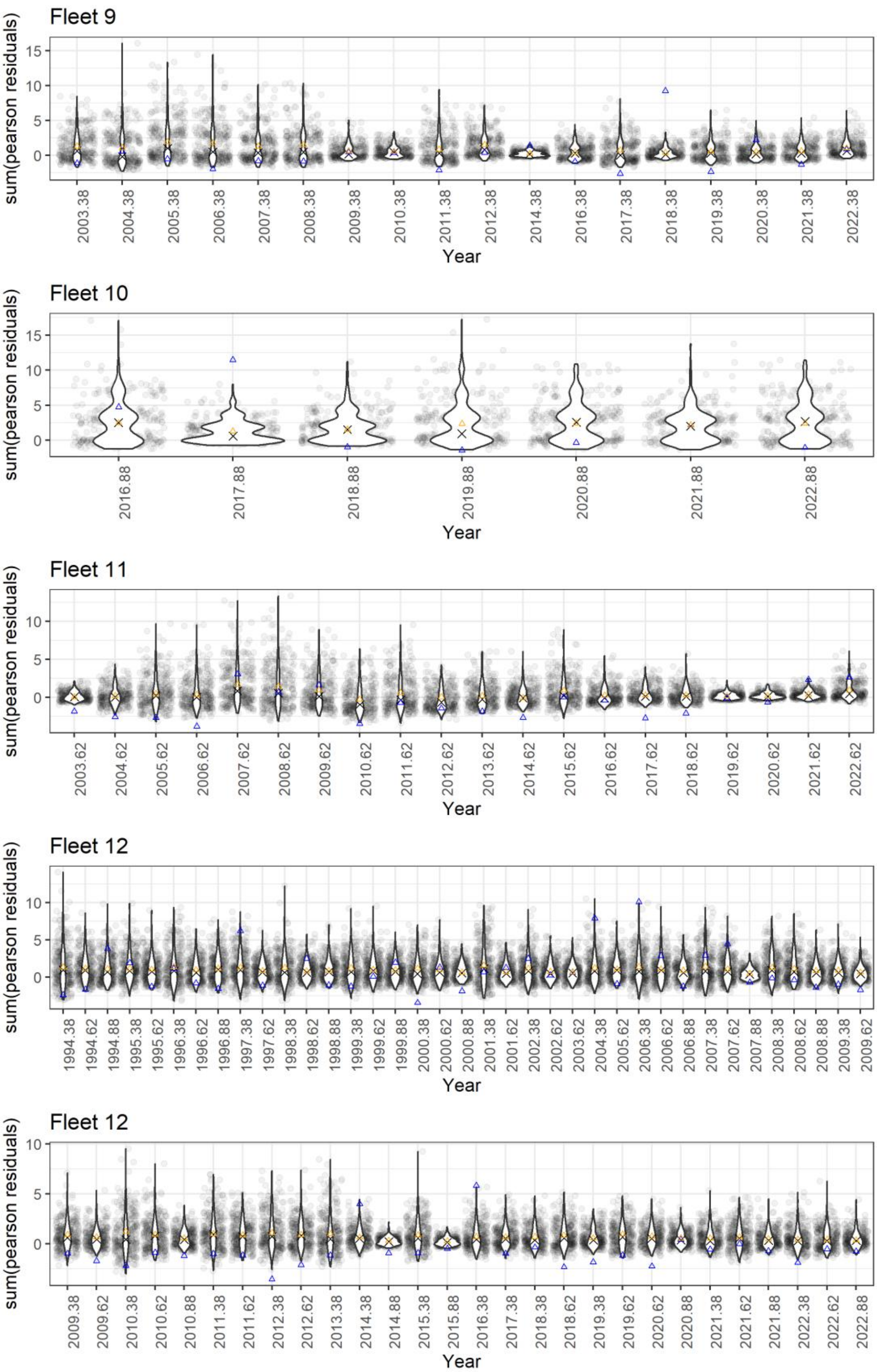

Figure 5 Continue.

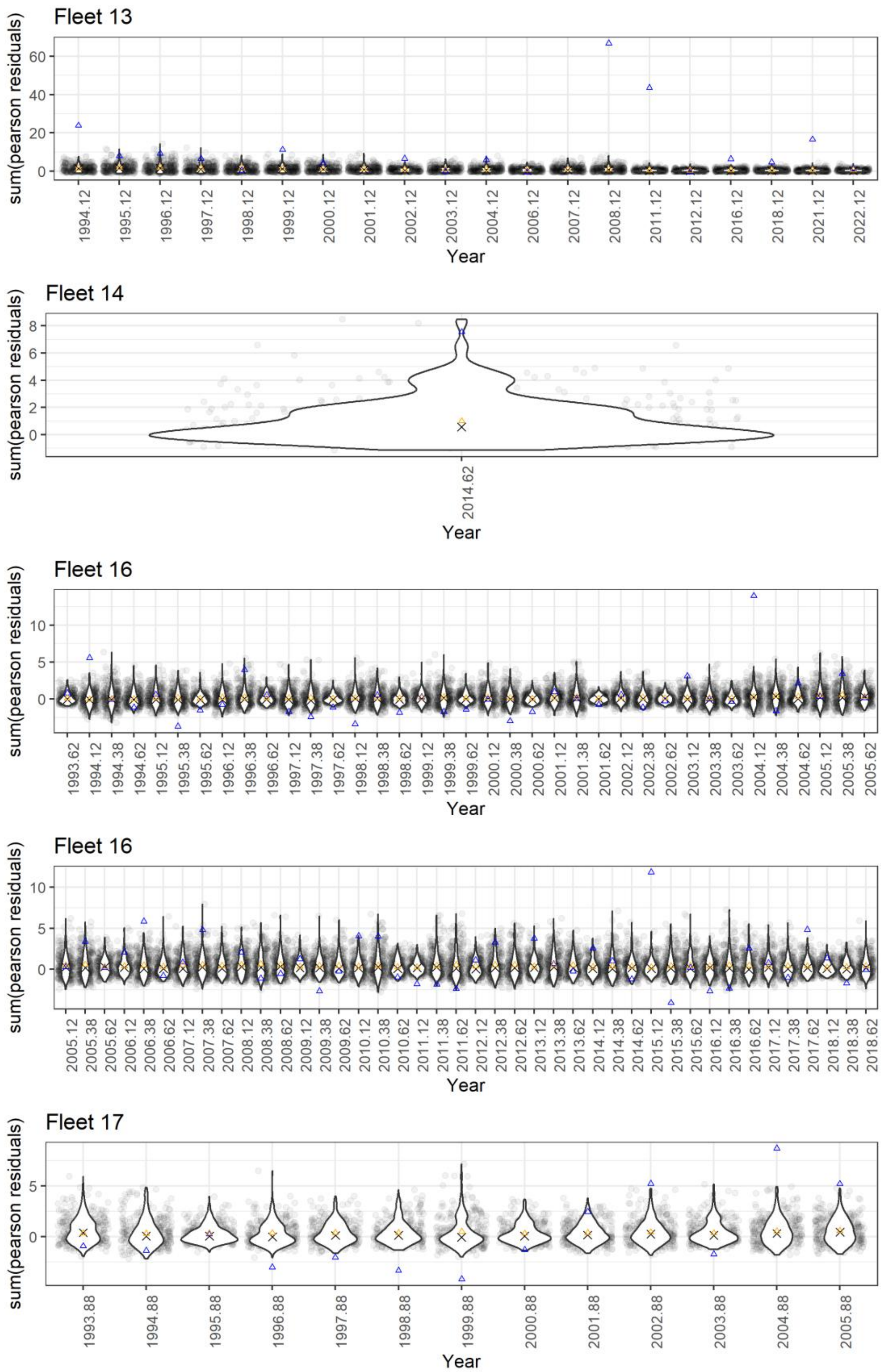

Figure 5 Continue.