Georges Bank Winter Flounder: Virtual Population Analysis to State-space

Alexander Hansell

**Summary**

Historically, Georges Bank (GB) winter flounder has been assessed using a virtual population analysis (VPA). The VPA has diagnostic issues including major retrospective patterns, which have required adjustments in past assessments. Additionally, the assessment is plagued by poor cohort tracking and a panel recommendation from the last two assessments has been to explore a state-space framework. A research recommendation for the last two management track assessments was to explore moving towards a state-space framework that allows for process error. Here, we explore fitting the 2022 assessment inputs that were used for the last VPA to the Woods Hole Assessment Model (WHAM). We document the benefits of moving to WHAM and we propose a WHAM run for use in the 2025 management track assessment. The proposed run uses logistic- normal distribution on all age composition and uses 2dar1 process on numbers at age. The proposed run has improved diagnostics compared to previous VPA runs and similar reference points and projections.

**GB Winter flounder assessment history:**

In SARC/SAW 28 (2002) the first age based assessment was conducted which used a Virtual Population Assessment (VPA). At SARC 34 (2005) a stock production model incorporating covariates (ASPIC) was used to assess the stock because there was concern over poor age sampling from the commercial fishery. At Groundfish Assessment Review Meeting (GARM III) a VPA was used to assess the stock because age sampling information improved for the commercial fishery (NEFSC 2008). Since 2008 a VPA has been used to assess the stock. The last assessment was the fall 2022 management tack (NEFSC, 2022).

**Data inputs:**

The stock boundary includes statistical areas 522-525, 542, 551-552 and 561-562 (Figure 1). Commercial landings data are available for 1964-2021. Total commercial landings are predominately from the U.S. bottom trawl fleet, but landings from Georges Bank have also been reported in the Canadian groundfish trawl fisheries, since 1964, as bycatch in the haddock and cod fisheries. Ages and lengths are sampled from the US fishery but size information is not available from the Canadian fishery.

GB winter flounder discards are estimated for the large mesh bottom trawl fleet (codend mesh size ≥ 5.5 inches), small mesh groundfish fleet (codend mesh size < 5.5 inches), and the sea scallop dredge fleet. Discards are estimated based on fisheries observer data and the landings data using the combined ratio method described in Wigley et al. (2007). Lengths are available from the observer program and are converted to ages using the northeast fisheries science center (NEFSC) age-length keys.

**Current Assessment:**

The catch at age input to the VPA consisted of combined U.S. and Canadian landings and discards from 1982-2022 for ages 1-6 with a 7+ age group (Figure 2). The VPA was calibrated using abundance at age indices from the NEFSC spring survey (1982-2022, ages 1-7) and fall NEFSC trawl surveys (1981-2021, ages 0-6 lagged forward one year and age) and the Canadian spring bottom trawl surveys (1987-2022, ages 1-7; Figure 3). Stock size was estimated for ages 2-6 in the terminal year+1. The natural mortality rate was assumed as 0.3 per year. Maturity data from the NEFSC spring survey were used to estimate the average proportion mature at age for 1982-2022. A three year moving average was used to estimate maturity using data from the spring NEFSC survey (1982-2022).

Model configurations:

The same data inputs were used as the 2022 management track assessment. No changes were explored regarding natural mortality, stock recruit relationship or environmental covariates; these will be explored in the upcoming research track assessment (2026). Several assumptions were made when moving from VPA to WHAM: 1) logistic selectivity for the commercial fleet and all three indices; 2) input effective sample sizes were assumed to be 200 for all data sources and years. High input effective sample sizes were needed for the self-weighting age composition distributions.

Model exploration was done in an order: 1) age distributions; 2) recruitment assumptions; 3) time varying selectivity; 4) full state-space models. WHAM allows for 10 different age-composition distributions and all were explored for GB winter flounder. Different recruitment assumptions were also explored, including: mean recruitment; deviations from the mean were independent identically distributed (iid) and deviations from the mean were correlated by year with an Ar1 process. Time varying selectivity in the commercial fleet was explored. A series of state-space models were also explored with random effects on numbers at age. A similar method was used

The same diagnostics used during the most recent Atlantic Cod Research Track were used to inform model selection:

1. Convergence: This diagnostic was used to determine if models produced a stable solution. Both first and second order convergence had to occur. First order convergence was determined by the maximum gradient of the negative log-likelihood function being less than e^-4. Second order convergence required the Hessian matrix being invertible.
2. Residuals: Goodness of fit was determined by examining model fit to the different data inputs. One-step ahead (OSA) residuals were used as the primary diagnostic because they help to account for correlations in the data that can be propagated into Pearson residuals (Trijoulet et al 2023). OSA residuals should be normally distributed.
3. Akaike’s Information Criteria (AIC): AIC was used to help determine parsimony and compare models with comparable likelihoods. Smaller AIC scores are better, with an improvement in model being determined if difference in AIC between two models was greater than two.
4. Retrospective patterns: Retrospective error was explored by comparing seven-year Mohn’s rho value for spawning stock biomass and fishing morality.
5. Prediction skill: Prediction skill was estimated by using mean absolute scaled error (MASE). A 20-year peel was used, with prediction error calculated over one- to five-year time periods. First-order model convergence was calculated for each peel. A MASE below one indicates predictive skill that is better than a naïve approach of predicting future observations.
6. Estimation performance: Self-test were used to explore estimation performance. The final candidate model was used to generate 100 datasets with parameters fixed at their estimated values. Simulations were then refit the model to these datasets to investigate relative error in F, SSB, R and convergence rates

**Results:**

WHAM like runs successfully converged and produced similar trends as the VPA for SSB, F, age ones and numbers at age. However, WHAM estimated slightly higher SSB and F since the mid-1990s (Figure 4-5). Selectivity estimates were similar for the commercial fleet and the three indices (Figure 4). Runs exploring the different age-comps supported using a logistic normal age distribution that ignored missing values; the runs using this distribution had some of the best Mohn’s rho values for SSB and F (Table 1). Additionally, this age composition produced decent residual patterns for all data inputs (Figure 7-10). Thus, logistic normal age distribution was used in subsequent model runs.

Incorporating random effects into recruitment did not substantially improve model diagnostics so they were not included (Table 2). Models with random effects on fleet selectivity converged; however, model runs had similar residual patterns in the commercial age-comps and failed to improve Mohn’s rho values (Table 3).

Full state-space models converged and including 2dar1 random effects on numbers at age led to similar AIC and improved Mohn’s rho values for SSB and F (Table 4). Thus, the selected model had similar inputs to the VPA; however, has logistics-normal distribution on all age composition likelihoods and random effects using a 2dar1 process were used on numbers at age.

The selected WHAM run had better retrospective patterns than the VPA and no longer requires a retrospective adjustment (Figure 11-12). MASE scores were above one for each survey and were especially high for the Canadian survey (Table 5; Figure 13-14 The selected model passed a self-test with mean bias below ten percent for R, SSB and F (Table 6; Figure 15).

Biological reference points were similar between the VPA and WHAM (Table 7; Figure 12). Projections were also similar between the two methods; however, WHAM removed an unrealistic decrease in projected biomass from early in the time series (Table 8; Figure 13).

**Discussion:**

Results from the VPA and WHAM produced similar trends; however, WHAM had better diagnostics. Fitting WHAM with a logistics-normal age composition likelihood for all data sources and 2dar1 random effects on numbers at age reduced Mohn’s rho values and no longer required the assessment to have a retrospective adjustment. Because of the improved diagnostics and additional flexibility in WHAM that will allow for future research to improve the assessment, I recommend the selected WHAM run presented here be the new base model for GB winter flounder. Switching to WHAM would allow this platform to be used for the next management track in 2025. Future work could be conducted before the next management track to continue to improve the assessment (e.g., work on time varying selectivity or time-varying q).

The winter flounder research track assessment is scheduled for peer review in 2026, with the group forming this winter. Moving the GB winter flounder stock from VPA to WHAM in the State-space Research Track will allow the winter flounder working group to spend more time on data inputs and the advanced flexibilities of WHAM. For example, the working group could spend more time on fleet structure (US and Canada). Further, the group will have more time to explore stock-recruit relationships and environmental drivers of GB population dynamics.

**References:**

Northeast Fisheries Science Center. 2008. Assessment of 19 Northeast Groundfish Stocks through 2007: Report of the 3rd Groundfish Assessment Review Meeting (GARM III), Northeast Fisheries Science Center, Woods Hole, Massachusetts, August 4-8, 2008. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 08-15; 884 p + xvii. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026.

NEFSC. 2022. Management Track Assessments Fall 2022. US Dept Commer, Northeast Fish Sci Cent Tech Memo. 305; 167p.+xv. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at http://www.nefsc.noaa.gov/nefsc/publications/.

Wigley SE, Rago PJ, Sosebee KA, Palka DL. 2007. The analytic component to the Standardized Bycatch Reporting Methodology Omnibus Amendment: sampling design end estimation of precision and accuracy (2nd edition). NEFSC Ref Doc. 07-09; 156 p.

Trijoulet, V., Albertsen, C. M., Kristensen, K., Legault, C. M., Miller, T. J., & Nielsen, A. (2023). Model validation for compositional data in stock assessment models: Calculating residuals with correct properties. *Fisheries Research*, *257*, 106487.

Table 1: Age composition likelihoods explored.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **age\_comp** | **converged** | **hessian** | **dAIC** | **AIC** | **rho\_R** | **rho\_SSB** | **rho\_Fbar** |
| multinomial | Yes | Yes | 15991.7 | 13643.7 | 1.0373 | 0.3798 | -0.2902 |
| dir-mult | Yes | Yes | 9553.1 | 7205.1 | 0.5262 | 0.1368 | -0.1394 |
| dirichlet-miss0 | Yes | Yes | 91.2 | -2256.8 | 1.0625 | 0.1738 | -0.1668 |
| dirichlet-pool0 | Yes | Yes | 157.4 | -2190.6 | 0.7593 | 0.1575 | -0.1559 |
| logistic-normal-miss0 | Yes | Yes | 359.2 | -1988.8 | 1.2581 | 0.1163 | -0.1239 |
| logistic-normal-ar1-miss0 | Yes | No | 0 | -2348 | 1.2572 | 0.0727 | -0.0906 |
| logistic-normal-pool0 | Yes | Yes | 411.6 | -1936.4 | 0.973 | 0.103 | -0.111 |
| logistic-normal-01-infl | Yes | Yes | 411.6 | -1936.4 | 0.973 | 0.103 | -0.111 |
| logistic-normal-01-infl-2par | Yes | Yes | 411.6 | -1936.4 | 0.973 | 0.103 | -0.111 |
| mvtweedie | Yes | Yes | 411.6 | -1936.4 | 0.973 | 0.103 | -0.111 |

Table 2: Different recruitment assumptions explored.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Run | Random effect | dAIC | AIC | rho\_R | rho\_SSB | rho\_Fbar |
| m1 | None | 0 | -1988.8 | 1.2581 | 0.1163 | -0.1239 |
| m3 | iid | 49 | -1939.8 | 0.7758 | 0.1348 | -0.1357 |
| m2 | Ar1\_y | 91.9 | -1896.9 | 1.5332 | 0.1774 | -0.1732 |

Table 3: Different time varying selectivity assumptions explored for the commercial fleet.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Runs | Model | dAIC | AIC | rho\_R | rho\_SSB | rho\_Fbar |
| m3 | 2dar1 | 0 | -2068.2 | 0.8268 | 0.1998 | -0.2014 |
| m2 | Ar1\_y | 4.9 | -2063.3 | 0.9724 | 0.1925 | -0.2098 |
| m4 | iid | 31.6 | -2036.6 | 0.81 | 0.1502 | -0.1321 |
| m1 | None (logistic) | 80 | -1988.2 | 1.2668 | 0.1264 | -0.1247 |

Table 4: Different state-space models explored.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Runs | Random effects | dAIC | AIC | rho\_R | rho\_SSB | rho\_Fbar |
| m1 | None | 0 | -1988.8 | 1.2581 | 0.1163 | -0.1239 |
| m2 | 2dar1 | 0.9 | -1987.9 | 0.3254 | -0.0342 | 0.0392 |
| m3 | Ar1\_y | 7.3 | -1981.5 | 0.5149 | 0.0423 | -0.0442 |
| m4 | iid | 43.9 | -1944.9 | 0.8835 | 0.0537 | -0.0659 |

Table 5: Average mean absolute scaled error (MASE) estimates over a five-year period.

|  |  |  |
| --- | --- | --- |
| **DFO** | **NEFSC Spring** | **NEFSC Fall** |
| 6.26 | 1.37 | 1.48 |

Table 6: Mean percent bias from self-tests.

|  |  |
| --- | --- |
| VAR | Mean Percent Bias |
| F | 3.71 |
| R | 8.51 |
| SSB | 2.04 |

Table 7: Biological reference points estimated from the VPA and WHAM.

|  |  |  |
| --- | --- | --- |
|  | **VPA** | **WHAM** |
| Fmsy | 0.452 | 0.34 |
| SSBmsy | 7,503  (4,143 - 11,113) | 6,849  (3,862-12,146) |
| MSY | 2,757  (1,811 - 3,918) | 2,445  (1,369 – 4,366) |

Table 8: Projections estimated from the VPA and WHAM.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Year | VPA: Catch | WHAM: Catch | VPA: SSB | WHAM: SSB | VPA: F | WHAM: F |
| 2022 | 278 | 286 | 7,102 | 3,195 | 0.052 | 0.068 |
| 2023 | 2,617 | 1,063 | 5,744 | 2,676 | 0.452 | 0.34 |
| 2024 | 1,646 | 907 | 3,998 | 2,559 | 0.452 | 0.34 |
| 2025 | 1,681 | 916 | 4,589 | 2,684 | 0.452 | 0.34 |

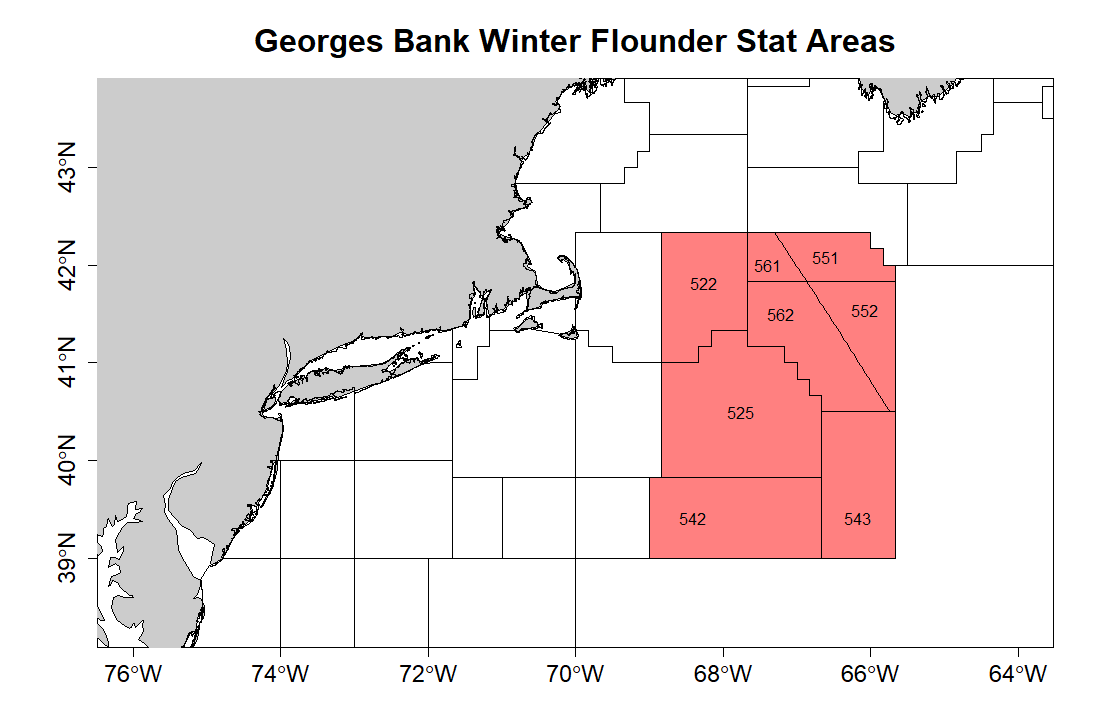


Figure1: GB winter flounder stat areas.

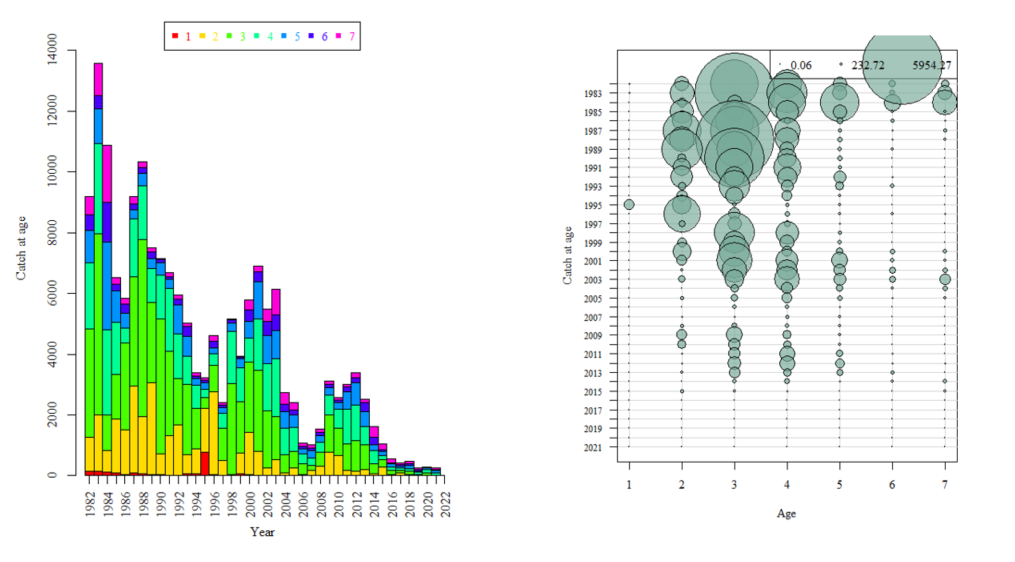


Figure 2: Catch at age for GB winter flounder in the VPA.

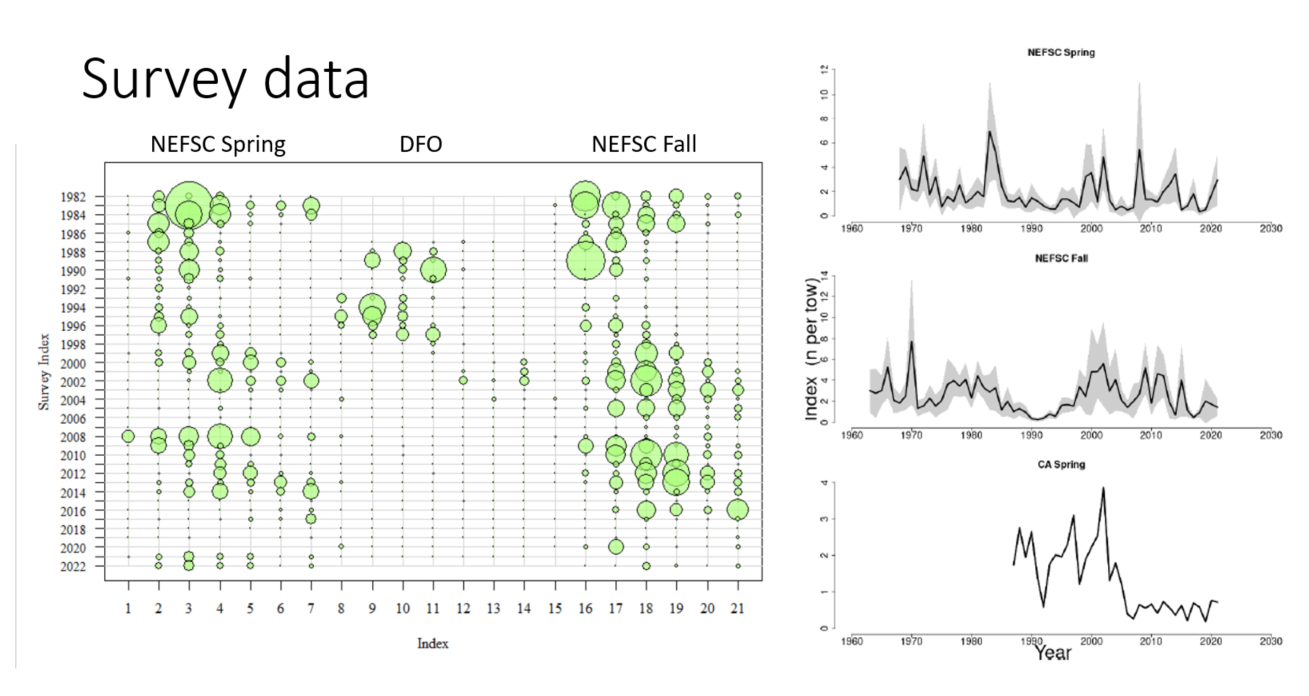


Figure 3: Survey data for GB winter flounder.

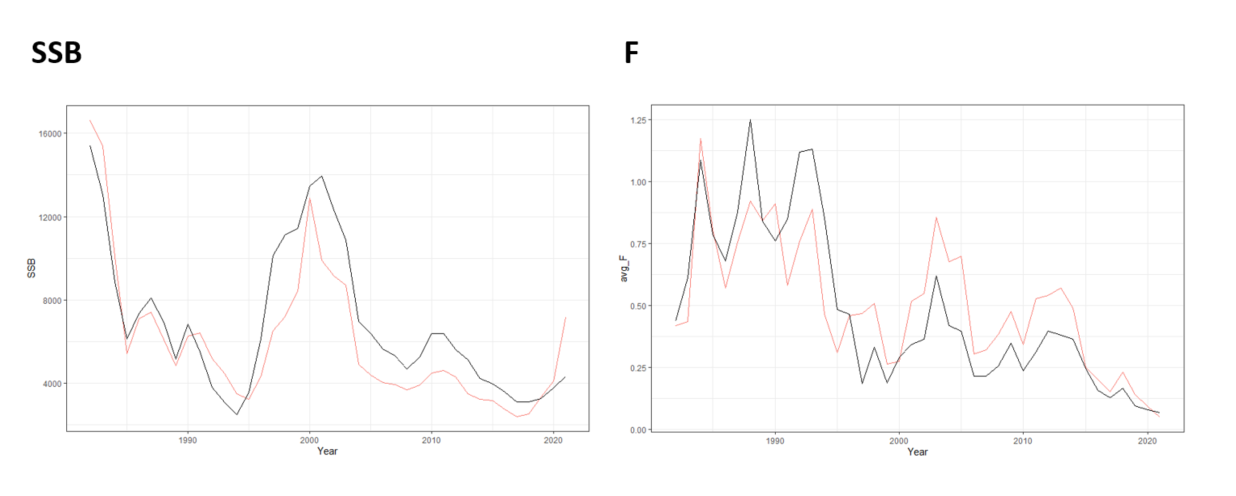


Figure 4: Comparison of VPA (red) and WHAM (black).

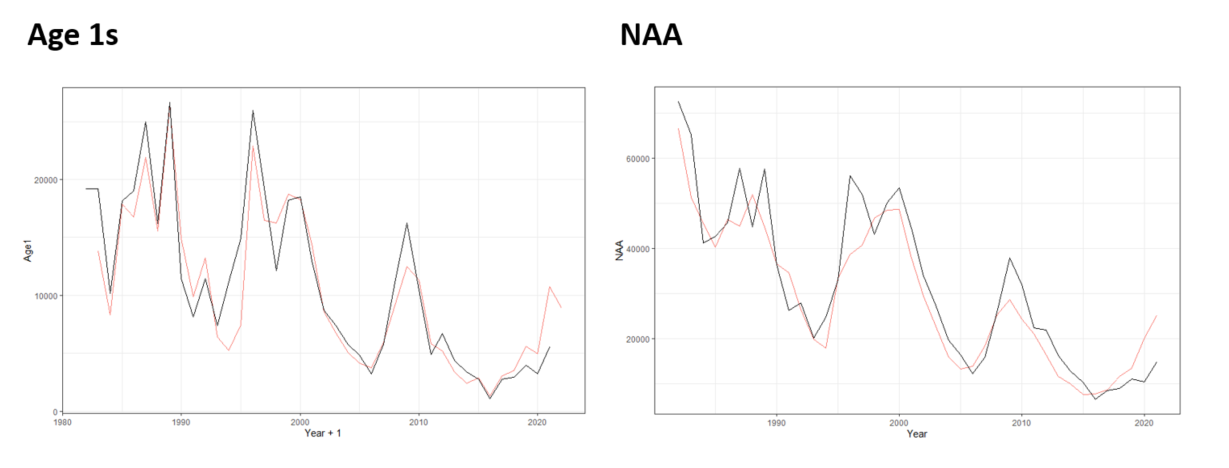


Figure 5: Comparison of VPA (red) and WHAM (black).

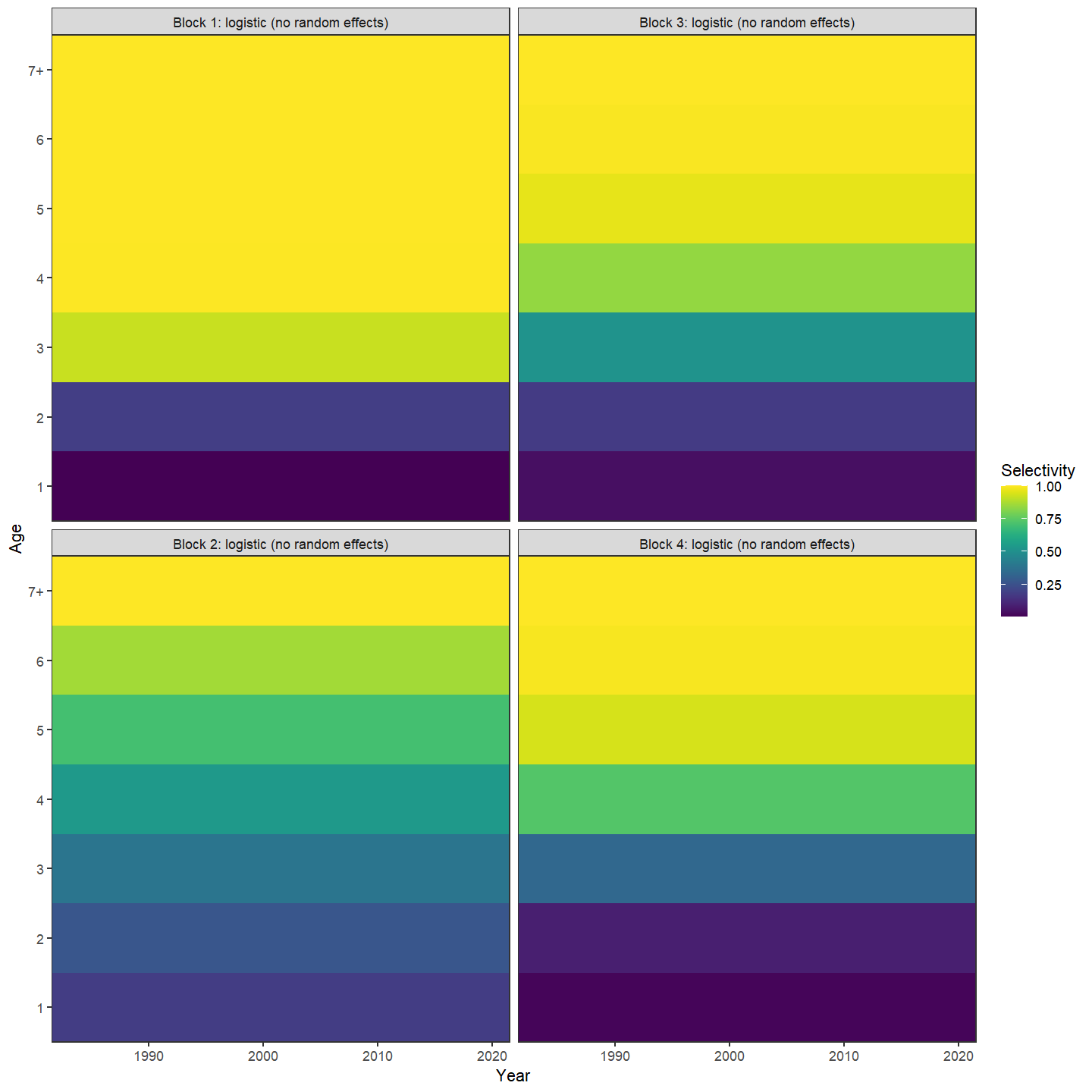


Figure 6: Selectivity: Block 1 = commercial; Block 2 = Spring nefsc; Block 3 = Fall nefsc; and Block 4 = DFO.

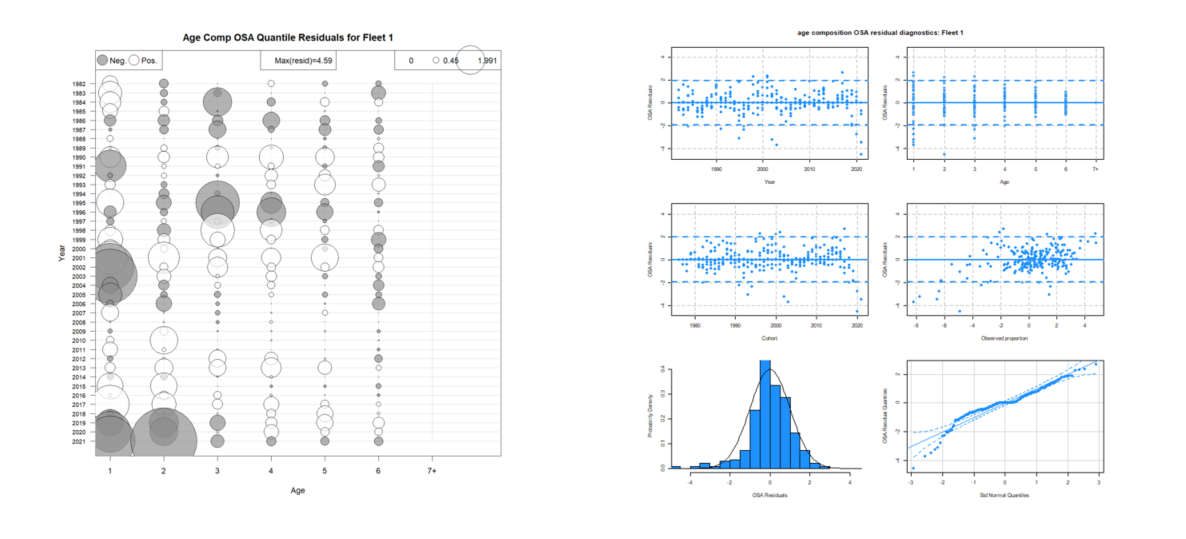


Figure 7: Commercial fleet OSA residuals using logistic-normal age composition likelihood.

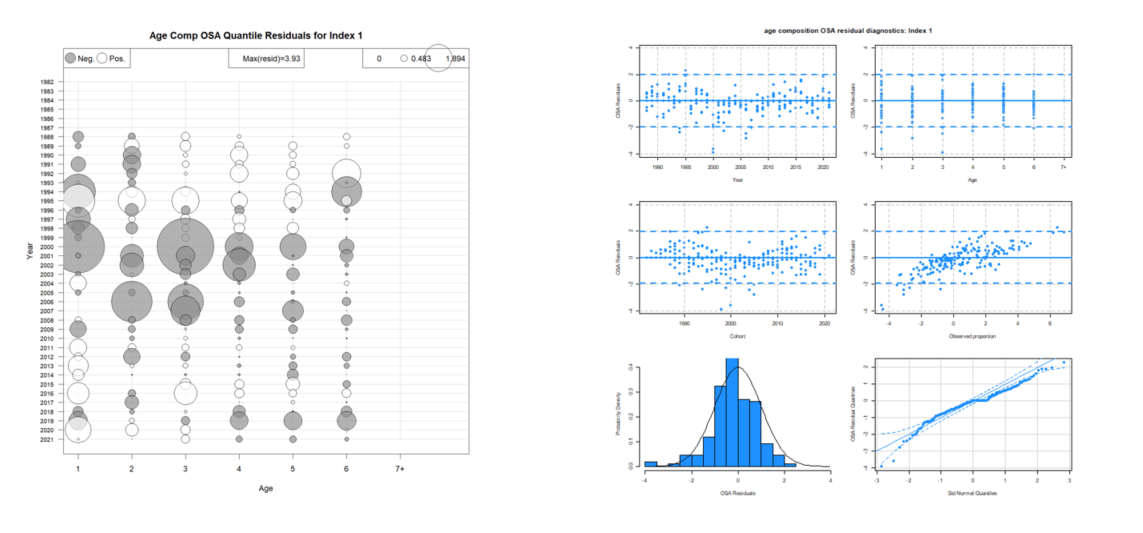


Figure 8: DFO survey OSA residuals using logistic-normal age composition likelihood.

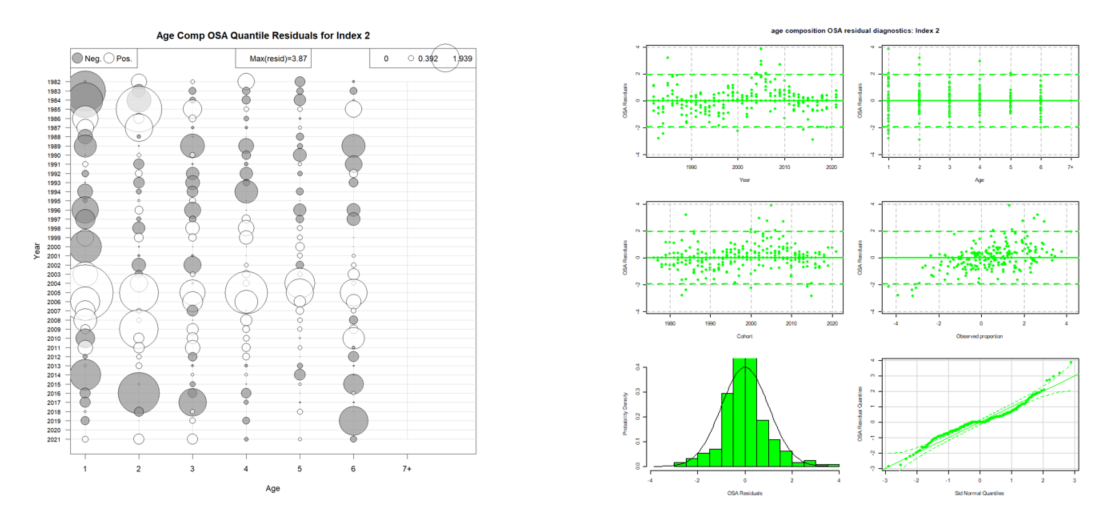


Figure 9: NEFSC spring survey OSA residuals using logistic-normal age composition likelihood.

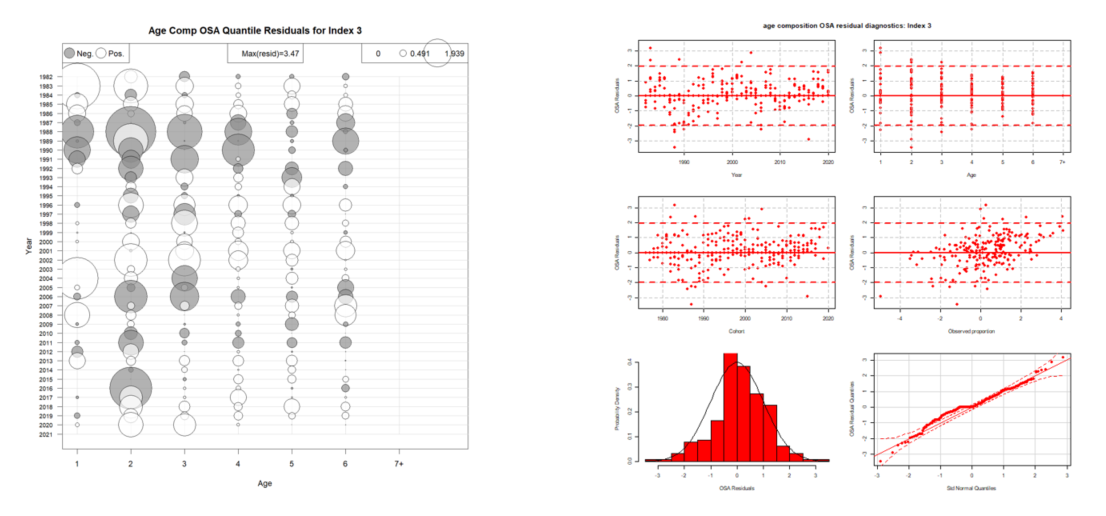


Figure 10: NEFSC fall survey OSA residuals using logistic-normal age composition likelihood.

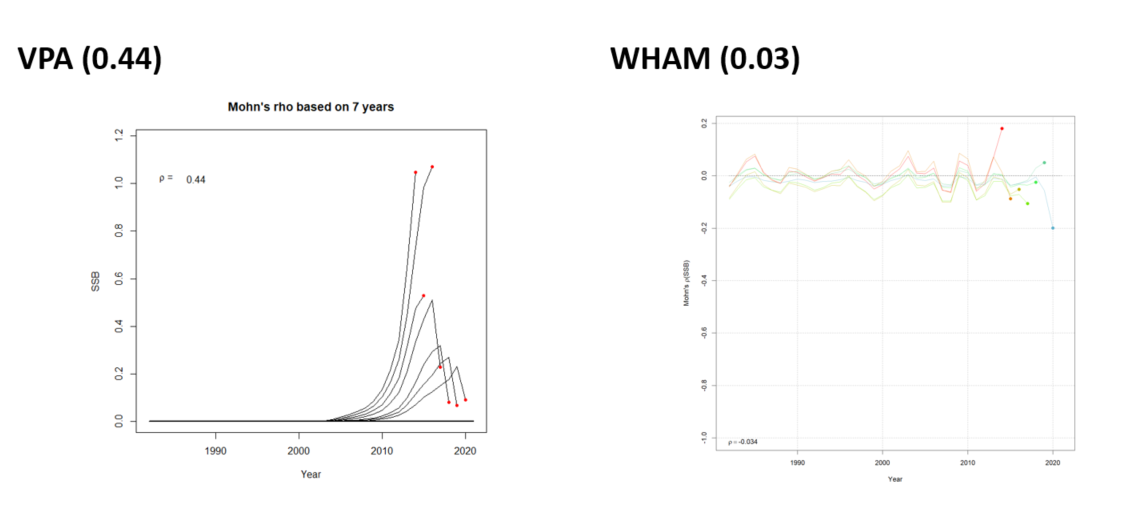


Figure 11: Comparison of retrospective patterns for VPA and WHAM.

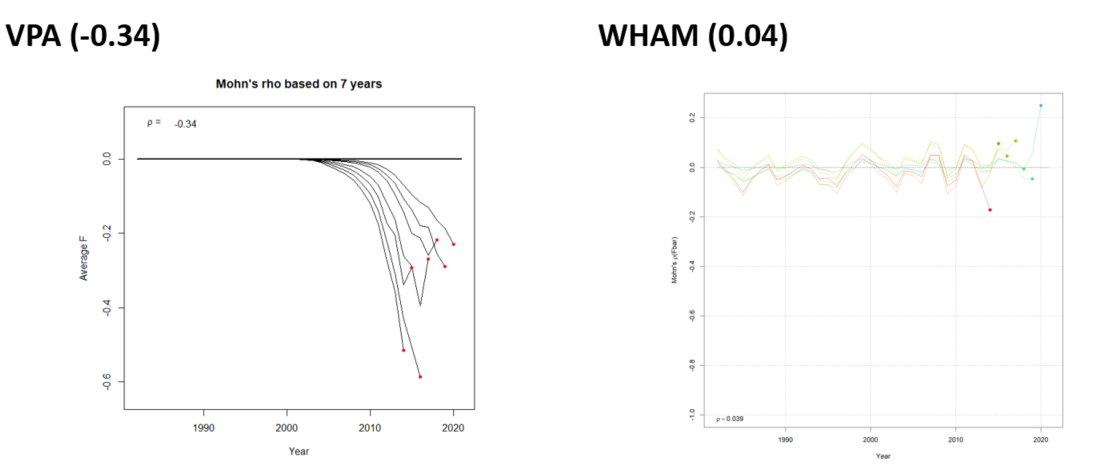


Figure 12: Comparison of retrospective patterns for VPA and WHAM.

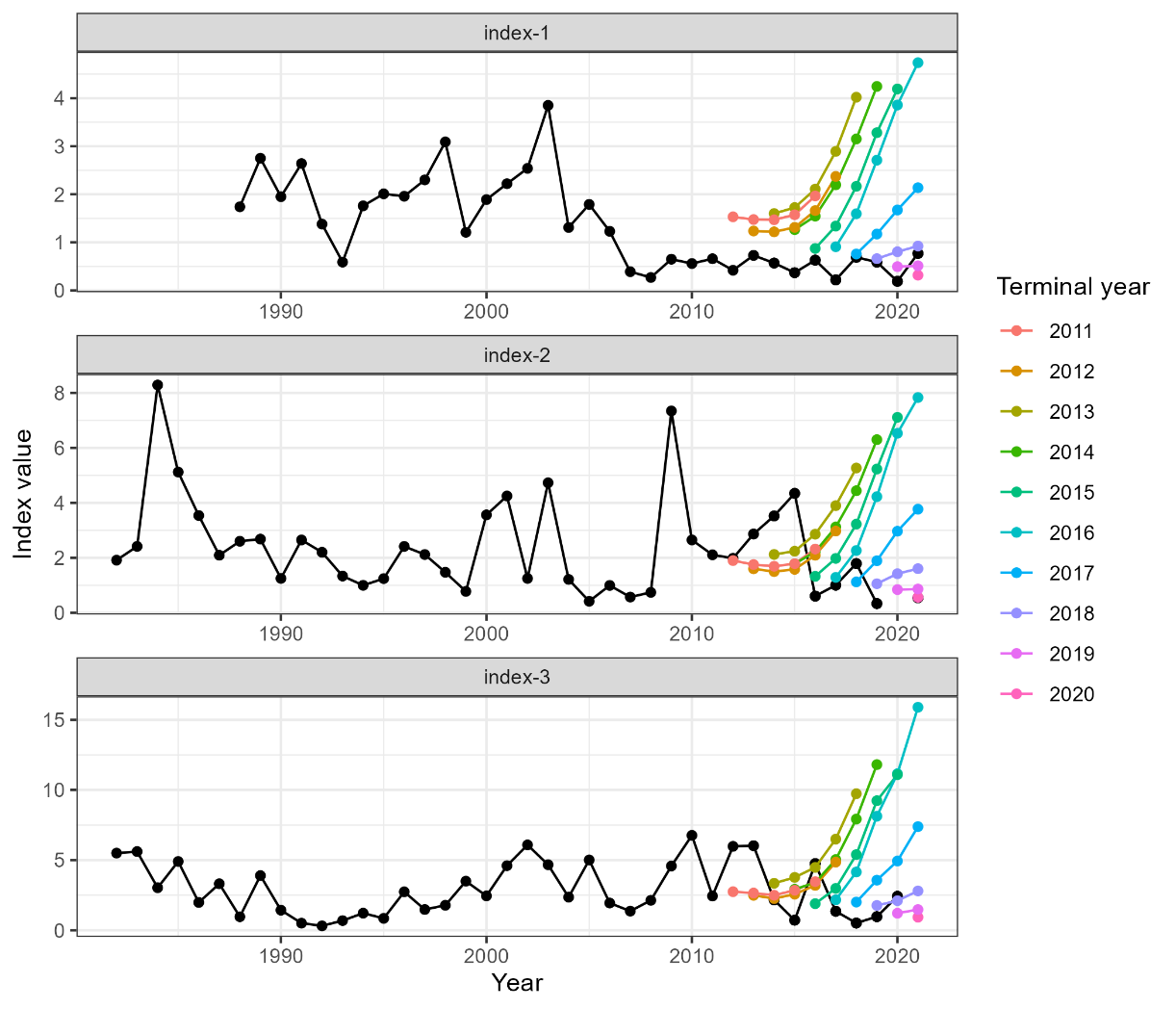


Figure 13: MASE for each index.

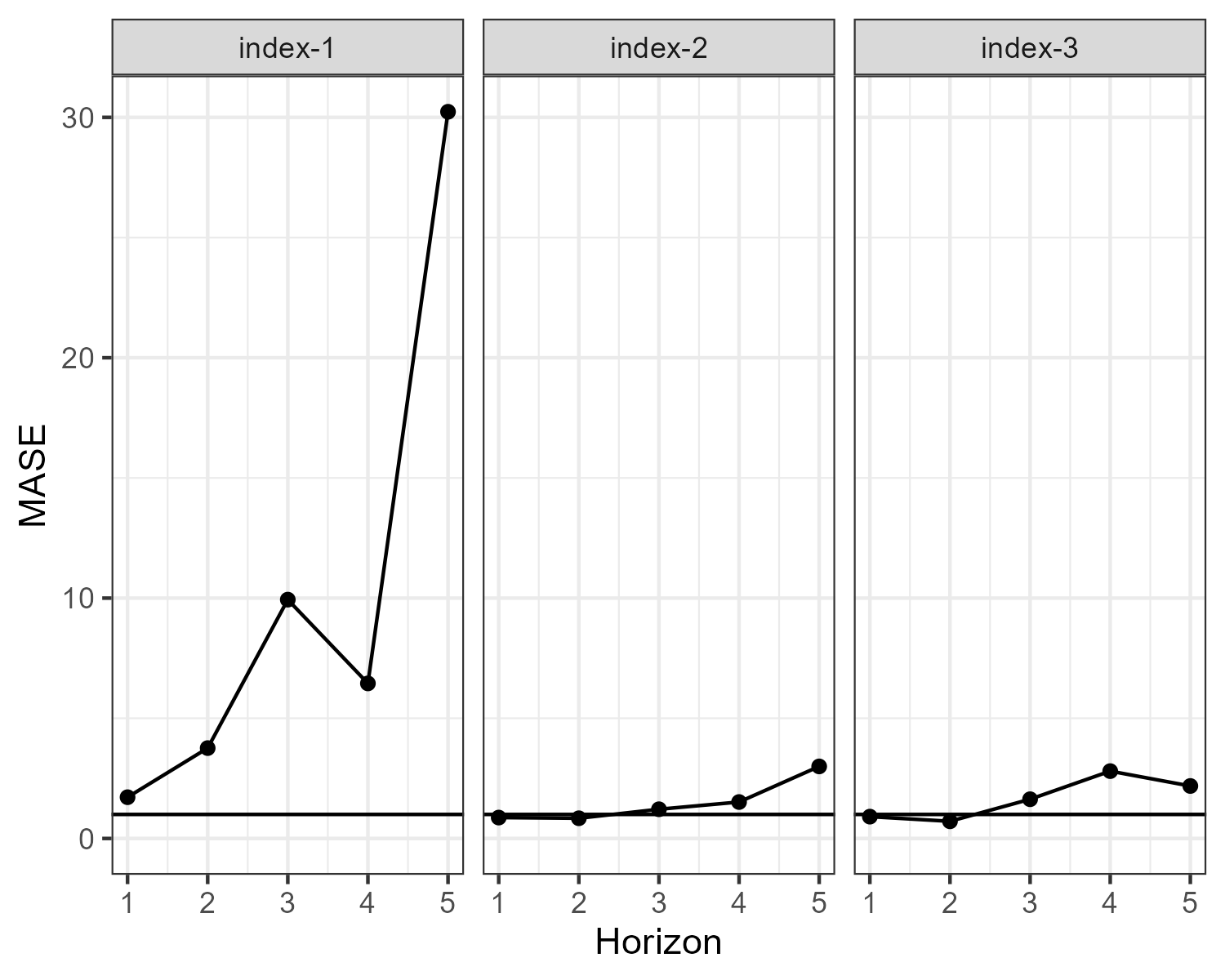


Figure 14: MASE for each index.

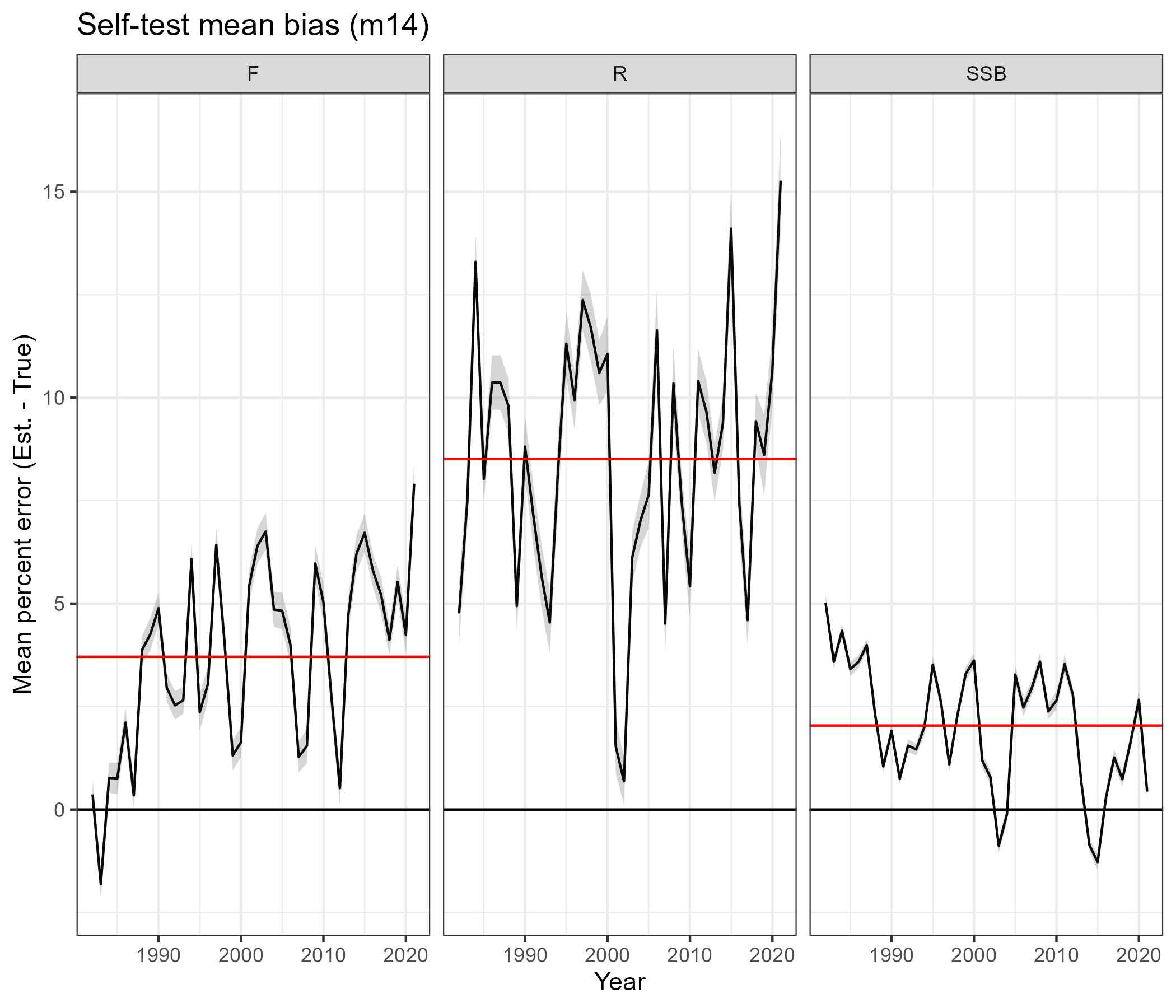


Figure 15: Mean percent bias for key variables produced from self-tests.

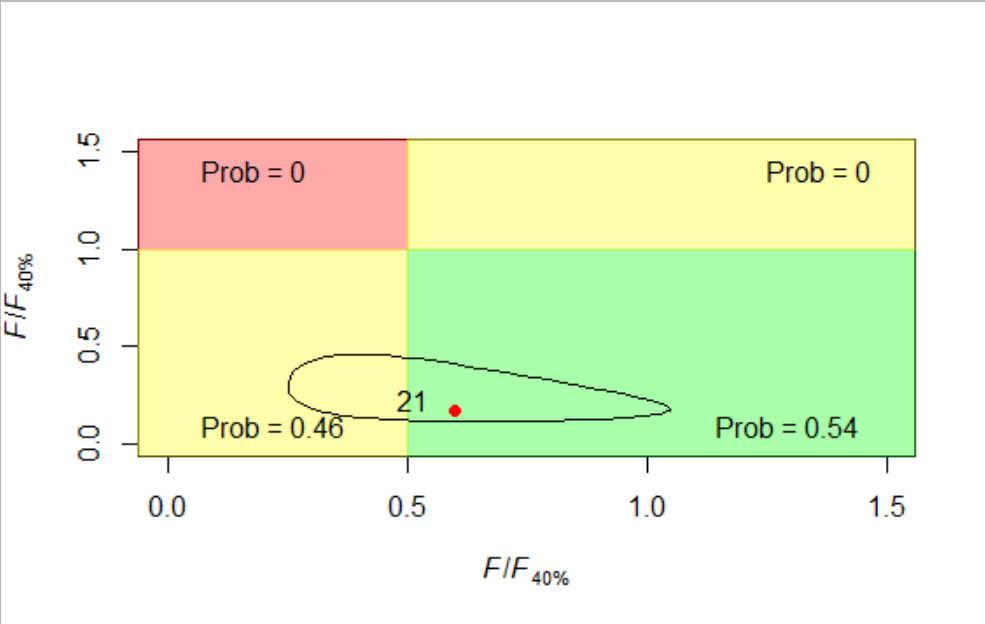


Figure 16: Comparison of stock status in terminal year from VPA and WHAM. Adjusted retrospective value from the VPA is the red circle.

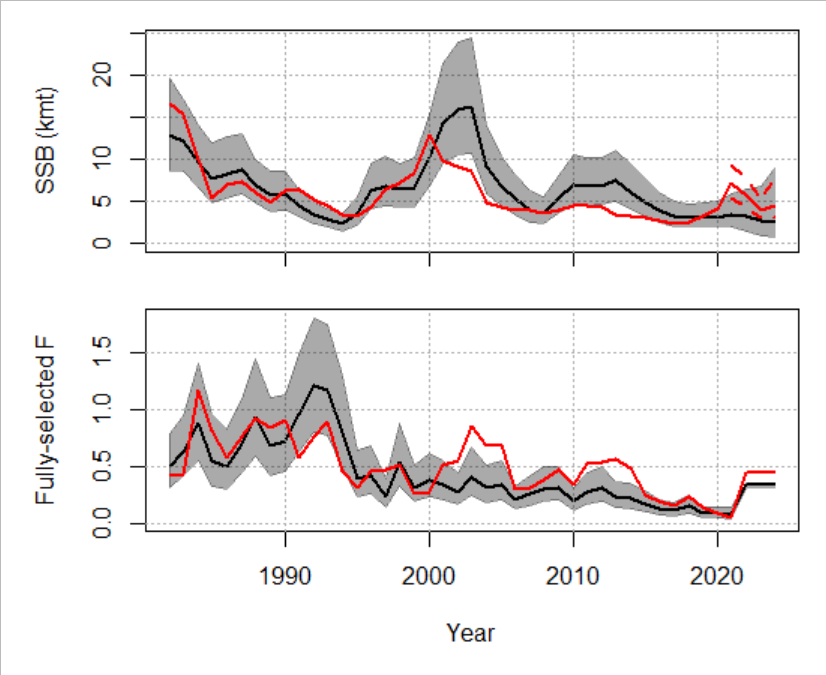


Figure 17: Comparison of VPA and WHAM stock trends/projections.