Appendix 2022 Bering Sea Pacific Cod September Report

Steven Barbeaux

September 2022

Introduction

For 2022 the Eastern Bering Sea (EBS) Pacific cod stock assessment lead authorship has changed for the first time in ~35 years. Grant Thompson had lead authorship for this stock from the mid 1980's through 2021. The new author, Steve Barbeaux, worked with Grant on the assessment in 2020 and 2021 and worked with Grant through the latest CIE review in 2021. The SSC recommended the new author make minimal changes to the assessment in the 2022 transition year. For the most part the models presented here for 2022 match those accepted for the ensemble for 2021 (Thompson et al. 2021; Table 1), however there are some minor changes explored that were thought to potentially improve the assessment model, or were necessary given software constraints.

Explored changes to the ensemble models:

1) Developing a new script for the seasonally corrected annual weight at length relationship fit outside the model.
2) New algorithm used for constructing the fishery length composition data using a developed R script.
3) Removing the seasonally corrected annual weight at length relationship from the model (NOWL).
4) Alternative aging bias assuming bias in those otoliths aged prior to 2007 and no bias in those aged after 2007 instead of bias assumed in 1994-2007 and 2008+ blocks. (AGE)
5) Alternative input sample size used for the fishery length composition and additional tuning to ensure the Dirichlet multinomial log theta parameter is not fit at or near a bound. (WT)
6) Fitting an additional standard error term on the VAST bottom trawl survey index. (SE)

Data Changes

Seasonally corrected annual weight at length relationship

Since 2015 the EBS Pacific cod stock assessment has used a seasonally corrected annually varying weight at length (WL) relationship derived from a nonlinear regression model developed by (now retired) Grant Thompson in an older (now unsupported) version of Mathcad. As this could not be replicated, the new lead author has developed a generalized additive modeling approach that achieves a similar product.

We started with the same base linear formula across all data for all years 1977-2021

$$log(W) = log(\alpha_1) + \beta_1 log(L)$$

Where W = weight in kg, L is length in cm. A generalized additive model was then fit to take into account annual and week effects:

$$log(W) = Y * log(L) + s(t)\colon \log(L) + s(t)$$

Where *Y* is the factor year and *t* is week of the year. The *s* are cyclic cubic regression splines with basis dimension of K=7 for log length by week and then week (Fig. 1). The basis dimension of 7 was chosen as it best replicated the original model developed by Grant Thompson.

The GAM was then used to predict weight across all years for all 52 weeks and for size bins from 10 to 120 cm at 10 cm increments with the standard bias correction of

$$W = e^{log(\widehat{W})+\sigma^2/2}$$

Where σ is the error term from the GAM.

A linear regression was fit across all predictions for all weeks combined for each year.

$$log(\dot{W}_Y) = log(\alpha_{2Y}) + B_{2Y} log(\dot{L_Y})$$

and the annual deviation in α and β used as annual indices on weight at length (Fig. 2) were calculated as alpha dev = exp(α_1) – exp(α_{2Y}) and Beta dev = exp(β_1)- exp(β_{2Y}). The results show up as annual variability in weight at length in the model (Fig. 3).

Annual length distribution data

The annual fishery length distribution data have been processed differently resulting in a new distribution used in the models (Fig. 4). The change was necessary as the previous author had manually processed the data in Excel, replicating this effort would not be possible, but more tedious than necessary. In developing the script to process the data, the author generalized the code to match that used in the Gulf of Alaska Pacific cod assessment. In prior assessments, for 1977-1990 the raw length measurements were used as the length distributions and for 1991-present fishery length compositions were weighted by catch weight by NMFS area (area), month, and gear and processed in EXCEL. Only areas with registered catch and greater than 30 lengths measured were used in the length composition data. For the current assessment and for all years the annual fishery length distributions were weighted by catch number per haul, vessel, area, month, and gear and processed through a function developed as an R script. For the 2022 models the total number of fish caught were calculated using average weights by area, gear, month, and year strata from the observer data where there were more than 30 fish weighed for each strata. Where there were fewer than 30 fish within a stratum the aggregation level was expanded by the following stratification levels until 30 or more weighed fish were encountered: 1) year, gear, month, 2) year, gear, quarter, 3) year, area, month, and 4) gear and year. An analysis of average weights revealed gear and time of year had a greater impact on average weights than area of capture. Length measurements from year, area, gear, month strata with less than 30 measurements were not included in the distributions. These measurements made up less than 1% of the total length measurements collected.

The overall difference in the distributions when using the new method was small with a slight shift to smaller fish overall (Fig.4) with the greatest impact in the 1977-1989 composition data.

There was a minor change in the survey length distribution produced by RACE for 2021 from those shown in the previous assessment with fewer small fish (< 30 cm) from the distribution, but otherwise remained largely the same (Fig. 5).

Change in assessment results due to data changes

We ran Model 19.12A with both the old and new data sets to examine changes in model results due to changes in the data. In general, the fits remain approximately the same (Table 2), however there was a small increase in estimated recruitment and spawning biomass post-1990 (Fig.6). Natural mortality changed from 0.361 to 0.369, survey catchability changed from 0.92 to 0.87, unfished spawning biomass from 1.30 to 1.31 million tons, the 2022 projected spawning biomass went from 518 kt to 528 kt and $F_{40\%}$ from 0.44 to 0.42 with the 2022 max ABC changing from 175 kt to 179 kt (Table 2 and Fig. 6). The largest change in Age-0 recruitment in 2020 and 2021 was due to a change in the 2021 survey and 2020 and 2021 fishery size composition data.

In review of the 2021 models, we discovered that the addition of the WL relationship resulted in nearly the same or even poorer fit to the length and age composition data (Table 2) in both the new and old configuration. Comparing the 2021 model using the 2021 data with and without the WL relationship shows that not using WL relationship improves the fit (-0.7 log likelihood). In addition, the redevelopment of the WL relationship described above similarly did not improve the model fit (Table 1). The removal of the WL relationship results in an improved retrospective pattern with a Mohn's Rho value closer to 0 across all four models (Table 3). For Model 19.12A for both old and new composition data the removal of the WL relationship results in lower M, a higher survey Q, lower recruitment on average, and lower spawning biomass over the time series (Fig. 7). This results in a lower $F_{40\%}$ and lower recommended ABC for 2022 (Table 2). Model results for all models with and without the WL relationship are provided in Table 4 and Table 5. Only Model 21.2 showed a small improvement (-1.3 LL) with the inclusion of the WL relationship.

Because of the lack of improvement to fit by including it and difficulty in projecting this relationship, I recommend that the seasonally corrected annual weight at length relationship used in the base model be discarded for 2022 and that we explore other options for modeling seasonality and annual changes in growth in 2023.

Model Changes

Alternate aging bias

The 2021 base models fit two periods for aging bias 1994-2007 and 2008-present. The models fit a positive bias (aged older than reality) in the 1994-2007 survey ages and some negative bias (aged younger than reality) in the 2008-present ages (Table 5 and Fig.8). Through isotope analysis Kastelle et al. (2016) validated that the previous aging method was positively biased. This bias is believed to have been corrected in the most recent, 2008-present, aging. The opinion of the Age and Growth Laboratory is that that current methods should no longer be biased (D. Anderl, personal communication). To be in alignment with this opinion we propose fitting models with no assumed bias for 2008-present and only fit aging bias for 1994-2007 data.

Removing the two parameters used to fit the 2008-present aging bias results in an overall degradation in model fit (Table 5). There was a small increase in negative log likelihood in all components, but as would be expected, the largest difference was in the fit to the survey age composition data. Although there is a reduction in model goodness of fit, given the advice of the age and growth laboratory, the authors think the reduced model is a better representation of the actual bias in the age data. If aging in the most recent time period is unbiased, the change in fit may be due to changes in growth in recent years and should be more explicitly explored in future models. Explorations of impacts on fit to the survey and

CPUE data can be found in Table 5 and Table 6. Graphs of changes in fits and differences in parameters over the models explored can be found in Figure 9 and Figure 10. Changes in fits to the survey are provided in Figure 11 and for Model 21.2 to the winter longline CPUE index in Figure 12. Figures 13 through 17 show change in spawning stock biomass and recruitment at age-0 for all of the models and proposed changes. The overall impact of changing the aging bias assumptions on model results varied among models, but in all cases was relatively minor.

In regards to advice from the Age and Growth Laboratory and despite the degradation in model fit I recommend that fitting aging bias for the most recent time period be removed for the 2022 models and that I explore more options for capturing variability in growth in 2023.

Alternate input sample size for fishery and survey length composition

The length composition input sample sizes used the 2021 Pacific cod models were calculated from the number of hauls. For the survey age and length composition the raw number of hauls conducted during the annual survey were used as the input sample size. For the fishery length composition, the input sample sizes were scaled from the number of hauls sampled such that average input sample size for the fishery length composition data equaled the average number of hauls in the bottom trawl survey time series. This reduced the input sample size from 5,625 hauls per year on average to 358. The method of scaling the fishery input sample size to the number of survey hauls was a holdover from the multinomial approach. The 2021 model employed the Dirichlet multinomial and for the length composition data, both survey and fishery, the log theta parameter was fixed at the high bound, as fitting the log theta resulted in high values greater and hampered model conversion (Table 8). A model with high log theta values near the bound may indicate that the input sample sizes are too low or the variance of the other data components such as the indices are too low. In effect the value of using the Dirichlet multinomial as parameterized in Stock Synthesis is that the theta parameter rescales the weighting of the composition data where input sample sizes are too high, however it does not rescale composition data where input samples sizes are too low. Note that the input sample size used in the model is in effect weighting the data within the model in relation to other data and model assumptions. Inappropriately low input sample sizes can down-weight the data in the model.

There are a number of methods currently employed at the AFSC for determining input sample sizes for composition data using multinomial and Dirichlet multinomial distributions. Raw haul numbers are commonly used, as are fixed 'rule of thumb' values, an effective sample size is calculated when VAST is used to estimate age composition data which has been suggested for use (Thorson; personal communication). In addition, a bootstrap approach for calculating effective sample size for the survey size and age composition data has been developed and could be used as input sample size (Hulson et al 2012). A similar bootstrap approach is in development for the fishery size and age composition data.

For 2022 I examined: 1) changing the fishery length composition sample size to the raw number of hauls per year, 2) continue to use the raw number of survey hauls for the age composition data, and 3) scale the survey length composition such that the log theta parameter of the Dirichlet multinomial is not near a bound using an input variance adjustment factor in the Stock Synthesis control file. For the survey length composition this resulted in input sample sizes being increase by a multiple of 5. The Dirichlet multinomial sample size multiplier fit for each model and version are provided in Table 8 and resulting average corrected input sample size for each data type are provided in Table 9. Note that while the theta 'corrected' new sample sizes for the length composition data are increased substantially, the theta ‘corrected’ input sample size for the age composition data drops. The new method for calculating sample sizes results in a substantially higher weight for the length composition data in the objective function going from a ~9,600 LL to ~22,000 LL in all models.

When the length composition input sample sizes were increased the fits to both the survey and age composition data are degraded in all models (Table 5). A reduction in fit to the survey although not wholly unexpected is troubling as it was greater than anticipated and should be further explored when the VAST bottom trawl survey index is updated for this year.

The change in weighting of the length composition data also resulted in an increase in the sigma values for the annually varying selectivity parameters (Table 7). Having increased value of the objective function specifically attributable to the size composition places more emphasis on fitting the length composition data better, the models do so by having the selectivity curves vary more from year to year through increasing these sigmas. With the change in input sample size and increased variability allowed in selectivity, the retrospective pattern across three of the four models is degraded with a substantial increase in the spawning stock biomass Mohn’s Rho values (Table 3).

Parameter estimates also vary more among the models than they had previously. In fitting catchability, the models had ranged between 0.87 and 1.04 for all previous models and versions (Table 6). For the new input sample size method survey catchability fit among the four models ranged between 0.69 and 1.14. Similarly, the range of natural mortality was increased from between 0.33-0.38 in all previous models and versions to 0.31-0.4 in the models with the new input samples sizes. These differences result in larger differences in key model results including reference points and current status among the four ensemble models (Table 7, and Fig. 9 - Fig 17).

The 2021 model’s method of down-weighting the fishery survey sample size to the average number of hauls in the survey has been consistently used in this model for several years, however it is unique to this stock and has little support in the literature. In theory, the parameterization of the Dirichlet multinomial in Stock Synthesis has the ability, through the fitting of the Theta parameter to reduce the input sample size to one consistent with other data in the model and therefore reduction in the initial input sample size would not be required. This is of course assuming that the number of hauls is an adequate proxy for input sample sizes. This method of setting input sample size is commonly used , however it too has mixed quantitative support for use and shown could be an overestimate of sample size in some cases (Pennington and Vølstad 1994).

I recommend that the new weighting of the length composition data be considered for 2022, however acceptance of the new weighting be examined more thoroughly once the new 2022 survey and fishery data are added to the model with further examination of model stability and sensitivity to this change. **In addition, I recommend alternative means for calculating the length and age composition input sample sizes should be explored in 2023 including bootstrap and VAST derived effective sample sizes.**

Fitting additional variance on the VAST survey index

The variance of the VAST survey index is small compared to the previous design based estimates with the design based average survey CV at 0.10 and average VAST based CV at 0.05 (Fig. 18). In addition the VAST estimates have changed as new years have been added to the index. This type of variability is not captured in the variance estimates provided. As is, the low variance estimates for the VAST survey results in the survey index having substantially more influence on results in the current model than the design based survey had in previous models. Fitting an additional variance parameter for survey estimates to account for unknown sources of variability is a common practice (Johnson et al. 2021) and implemented in the latest version of stock synthesis.

The addition of a standard error parameter in all models results in the uncertainty (log SE) of the bottom trawl survey being increased by between 0.15 to 0.19 (Table 6). Although the likelihood for these models is substantially reduced as a function of increased variance around the surveys, the apparent (visually assessed) fit to the survey and CPUE index is substantially worse. Across all ensemble models fitting a higher variance for the survey caters to an improved overall fit to the length composition data as the weighting among model components shifts. With the additional flexibility in the models with the increase in variance for both indices, The retrospective analysis shows a substantial increase in absolute bias in all four ensemble models (Table 3) with the Mohn's Rho across models going from a range of -0.01 to -0.05 to a range of -0.08 to -0.38. Figure 19 includes a graph of the retrospective pattern for Model 19.12A 2021 version spawning stock biomass and Figure 20 includes a graph of the same for Model 19.12A version with the increased index standard errors.

I recommend that fitting additional standard error to the indices not be adopted for this year's set of ensemble models. Additional exploration of proper variance attribution of VAST indices within the assessment model should continue to be explored in 2023.

Additional observations on current ensemble

There is a set of new tools useful for examining stock synthesis model performance described by Carvalho et al. (2021) and provided in the R library ss3diags. All of the ensemble models and versions were analyzed using these tools.

Joint-index residual plots were produced for each data type for all models and versions using the SSplotJABBAres function from the ss3diags R library. This function also produced joint RMSE values for each data type (Table 10). The change in input sample sizes, retuning of the models, and the fitting of additional standard error on the abundance indices resulted in substantial inflation of the RMSE of the abundance indices.

Residual runs tests were performed to examine the distribution of the residuals and whether the residuals were randomly distributed (Table 11). Every model and version, except Model 19.12A Version with no WL (NOWL), 1977-2007 aging bias only (+AGE), new input sample size (+WT), and fitted with additional standard error for the indices (+SE)(Fig. 20), failed in at least one data component. All of the models passed for the mean age residuals, but there were mixed results for all of the other data components. All of the versions with the 2021 length composition input sample sizes failed the runs test for the fishery mean length residuals. Except for all versions of Model 19.12 with annually varying survey catchability and Model 19.12A version NOWL+AGE+WT+SE, the remaining models and versions failed the survey mean length residual runs tests. By version the NOWL+AGE+WT+SE performed the best with 4 failures total across all models and data components, NOWL+AGE+WT next with only 5 failures, the remaining versions had 8 failures each, but in different Models and data components.

The Mean absolute scaled error (MASE) values examine the prediction skill of the models and versions, values greater than 1.0 indicated performance worse than a random walk. Results of the MASE tests are provided in Table 12. For the bottom trawl survey index all models and versions, except Version

NOWL+AGE+WT+SE performed better than the random walk. Only the 2021 version of Model 21.2 performed better than a random walk for predicting the fishery mean length with the NOWL+AGE+WT+SE version performing particularly badly. None of the models or versions predicted survey mean length particularly well with all of the new input sample size versions performing particularly badly. Survey mean age predictions were better than a random walk for all of the 2021 input sample size versions, but worse for all of the models and versions with the new input sample size.

These examinations lead to the conclusion that none of these model or versions are particularly exceptional. Fitting of the fishery length composition data is particularly problematic. Pacific cod grow rather quickly and I believe there are substantial seasonal and spatial influences that are not captured in any of these models.

I recommend that the authors in 2023 re-explore a seasonal model for Bering Sea Pacific cod and in light of the most recent genetic and tagging data (McDermott personal comm.) explore an expanded spatial model that incorporates the western Gulf of Alaska in the model. The genetics and tagging data will be more fully addressed in the complete assessment for November.

References

Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., Kitakado, T., Yemane, D., Piner, K.R. and Maunder, M.N., 2021. A cookbook for using model diagnostics in integrated stock assessments. Fisheries Research, 240, p.105959.

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

Johnson, K.F., A.M. Edwards, A.M. Berger and C.J. Grandin. 2021. Status of the Pacifc Hake (whiting) stock in U.S. and Canadian waters in 2021. Prepared by the Joint Technical Committee of the U.S. and Canada Pacifc Hake/Whiting Agreement, National Marine Fisheries Service and Fisheries and Oceans Canada. 269 p.

Pennington M. and Volstad 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.

Thompson et al. 2021 Assessment of Pacific cod in the Bering Sea. In: Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands regions. North Pacific Fisheries Management Council, 605 W. 4th Avenue Suite 306, Anchorage, AK 99501.

Tables

Table 1. Model features for the ensemble from 2021.

Feature	*M19.12a*	*M19.12*	*M20.1*	*M20.2*
Feature 1: Allow catchability to vary?	No	Yes	No	No
Feature 2: Allow domed survey selectivity?	No	No	Yes	No
Feature 3: Use fishery CPUE?	No	No	No	Yes

Table 2. Comparison of key elements from Model 19.12A for old data and new data with and without (NOWL) the seasonally adjusted annual weight at length relationship.

Label	***Old Data***	***Old Data /NOWL***	***New Data***	***New Data /NOWL***
# Parameters	301	301	301	301
Total Likelihood	10448.3	10447.6	10473.5	10468.0
Survey Likelihood	-7.6	-7.7	-3.7	-4.4
Length comp Likelihood	9602.9	9602.4	9618.7	9616.3
Age comp Likelihood	780.391	780.2	787.0	784.3
Recr. Virgin (n x 10^9)	560.393	534.9	616.9	551.3
M	0.361	0.355	0.369	0.357
BTS Q	0.92	0.94	0.87	0.94
L at Amax	112.1	112.7	110.6	113.3
VonBert K	0.119	0.118	0.122	0.115
Unfished spawning biomass (T x 10^6)	1.300	1.303	1.310	1.321
Bratio_2021	0.39	0.41	0.424	0.41
SPRratio_2020	0.52	0.53	0.52	0.55
$F_{40\%}$	0.36	0.35	0.35	0.33
2022 ABC (t)	174,678	167,833	183,826	161,352

Table 3. Retrospective Mohn's rho from 10-year peal for all models and versions. Version is 2021 = 2021 base models, NOWL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey.

Mohn's rho	*Model 19.12*	*Model 19.12A*	*Model 21.1*	*Model 21.2*
2021	-0.05	-0.08	-0.07	0.09
NOWL	-0.08	-0.03	-0.03	0.09
NOWL+AGE	-0.08	-0.02	-0.02	0.07
NOWL+AGE+WT	-0.01	-0.05	-0.02	-0.03
NOWL+AGE+WT+SE	-0.20	-0.26	-0.08	-0.38

Table 4. Aging bias parameter fit for models and versions. Version is 2021 = 2021 base models, NOWL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey.

Label	*Model 19.12*	*Model 19.12A*	*Model 21.1*	*Model 21.2*	*Version*
Age 1 delta 1977	0.342	0.350	0.348	0.349	2021
Age 1 delta 1977	0.343	0.347	0.346	0.349	NOWL
Age 1 delta 1977	0.343	0.347	0.347	0.351	NOWL+AGE
Age 1 delta 1977	0.343	0.344	0.340	0.350	NOWL+AGE+WT
Age 1 delta 1977	0.343	0.344	0.343	0.346	NOWL+AGE+WT+SE
Age 10 delta 1977	1.114	1.005	1.019	0.985	2021
Age 10 delta 1977	1.103	1.040	1.046	0.969	NOWL
Age 10 delta 1977	1.046	0.989	0.990	0.903	NOWL+AGE
Age 10 delta 1977	1.135	1.010	1.253	0.997	NOWL+AGE+WT
Age 10 delta 1977	1.176	1.074	1.298	1.166	NOWL+AGE+WT+SE
Age 1 delta 2008	0.015	0.006	0.006	0.010	2021
Age 1 delta 2008	0.016	0.009	0.009	0.011	NOWL
Age 1 delta 2008					NOWL+AGE
Age 1 delta 2008					NOWL+AGE+WT
Age 1 delta 2008					NOWL+AGE+WT+SE
Age 10 delta 2008	-1.726	-1.488	-1.482	-1.553	2021
Age 10 delta 2008	-1.770	-1.557	-1.552	-1.619	NOWL
Age 10 delta 2008					NOWL+AGE
Age 10 delta 2008					NOWL+AGE+WT
Age 10 delta 2008					NOWL+AGE+WT+SE

Table 5. Comparison of likelihood elements from models with new data. Version is is 2021 = 2021 base models, NOWL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey. Parameters include the annual dev pseudo-parameters.

Label	***Model 19.12***	***Model 19.12A***	***Model 21.1***	***Model 21.2***	***VERSION***
Parameters	342	301	305	302	2021
Parameters	342	301	305	302	NOWL
Parameters	340	299	303	300	NOWL+AGE
Parameters	342	301	305	302	NOWL+AGE+WT
Parameters	343	302	306	304	NOWL+AGE+WT+SE
AIC	21,447	21,549	21,553	21,625	2021
AIC	21,431	21,538	21,546	21,628	NOWL
AIC	21,472	21,584	21,588	21,663	NOWL+AGE
AIC	45,948	46,383	46,202	46,535	NOWL+AGE+WT
AIC	45,914	46,043	45,766	45,777	NOWL+AGE+WT+SE
Total Likelihood	10381.3	10473.5	10471.4	10510.7	2021
Total Likelihood	10373.3	10468.0	10468.2	10512.0	NOWL
Total Likelihood	10395.8	10493.2	10491.2	10531.7	NOWL+AGE
Total Likelihood	22632.1	22890.6	22796.1	22965.7	NOWL+AGE+WT
Total Likelihood	22613.8	22719.4	22577.0	22584.6	NOWL+AGE+WT+SE
Survey Likelihood	-91.3	-3.7	-2.7	-39.6	2021
Survey Likelihood	-92.5	-4.4	-3.5	-40.0	NOWL
Survey Likelihood	-91.7	-3.9	-3.7	-39.6	NOWL+AGE
Survey Likelihood	-83.5	81.2	84.9	177.5	NOWL+AGE+WT
Survey Likelihood	-42.4	-35.8	-40.8	-64.86	NOWL+AGE+WT+SE
Length comp Likelihood	9587.7	9618.7	9617.3	9685.2	2021
Length comp Likelihood	9579.1	9616.3	9616.7	9692.1	NOWL
Length comp Likelihood	9580.6	9618.6	9617.1	9690.1	NOWL+AGE
Length comp Likelihood	21716.1	21854.4	21755.6	21849.7	NOWL+AGE+WT
Length comp Likelihood	21700.7	21801.0	21657.1	21702.6	NOWL+AGE+WT+SE
Age comp Likelihood	775.9	787.1	785.6	786.9	2021
Age comp Likelihood	776.3	784.3	783.5	784.0	NOWL
Age comp Likelihood	796.5	806.9	806.3	805.1	NOWL+AGE
Age comp Likelihood	849.3	844.1	848.3	844.1	NOWL+AGE+WT
Age comp Likelihood	850.3	844.4	857.8	848.4	NOWL+AGE+WT+SE

Table 6: Comparison of key model results from models with new data. Version is 2021 = 2021 base models, NOWL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey.

Label	***Model 19.12***	***Model 19.12A***	***Model 21.1***	***Model 21.2***	***VERSION***
M	0.337	0.369	0.364	0.363	2021
M	0.342	0.357	0.354	0.353	NOWL
M	0.328	0.348	0.345	0.350	NOWL+AGE
M	0.381	0.339	0.312	0.396	NOWL+AGE+WT
M	0.401	0.364	0.288	0.411	NOWL+AGE+WT+SE
BTS Q	1.04	0.87	0.90	0.867	2021
BTS Q	1.01	0.94	0.95	0.908	NOWL
BTS Q	1.04	0.94	0.95	0.877	NOWL+AGE
BTS Q	0.79	1.00	1.14	0.685	NOWL+AGE+WT
BTS Q	0.72	0.94	1.23	0.678	NOWL+AGE+WT+SE
BTS SE+	0.15	0.19	0.16	0.18	NOWL+AGE+WT+SE
CPUE Q				0.0003	2021
CPUE Q				0.0003	NOWL
CPUE Q				0.0003	NOWL+AGE
CPUE Q				0.0004	NOWL+AGE+WT
CPUE Q				0.0003	NOWL+AGE+WT+SE
CPUE SE+				0.17	
L at Amax	118.686	110.612	112.958	115.160	2021
L at Amax	115.876	113.26	114.202	116.899	NOWL
L at Amax	115.999	112.928	113.566	114.013	NOWL+AGE
L at Amax	111.537	115.060	105.588	111.207	NOWL+AGE+WT
L at Amax	111.285	115.211	103.655	111.903	NOWL+AGE+WT+SE
VonBert K	0.101	0.122	0.115	0.104	2021
VonBert K	0.109	0.115	0.112	0.099	NOWL
VonBert K	0.107	0.113	0.111	0.105	NOWL+AGE
VonBert K	0.126	0.112	0.152	0.125	NOWL+AGE+WT
VonBert K	0.126	0.110	0.154	0.124	NOWL+AGE+WT+SE
Unfished spawning biomass (T x 10^6)	1.339	1.312	1.313	1.310	2021
Unfished spawning biomass (T x 10^6)	1.350	1.321	1.330	1.325	NOWL
Unfished spawning biomass (T x 10^6)	1.391	1.353	1.361	1.357	NOWL+AGE
Unfished spawning biomass (T x 10^6)	1.427	1.393	1.698	1.488	NOWL+AGE+WT
Unfished spawning biomass (T x 10^6)	1.411	1.301	1.737	1.431	NOWL+AGE+WT+SE
Recr. Virgin (n x 10^9)	460.768	616.934	587.841	594.063	2021
Recr. Virgin (n x 10^9)	487.184	551.436	538.143	547.416	NOWL
Recr. Virgin (n x 10^9)	448.448	530.263	520.981	560.104	NOWL+AGE
Recr. Virgin (n x 10^9)	719.073	483.368	451.887	849.215	NOWL+AGE+WT
Recr. Virgin (n x 10^9)	850.824	571.669	373.608	928.794	NOWL+AGE+WT+SE
2022 ABC (t)	114,901	183,826	172,691	151,208	2021
2022 ABC (t)	132,621	161,532	154,662	134,968	NOWL
2022 ABC (t)	114,434	155,920	150,405	145,152	NOWL+AGE
2022 ABC (t)	204,251	137,028	142,616	233,444	NOWL+AGE+WT
2022 ABC (t)	167,343	62,215	138,340	195,555	NOWL+AGE+WT+SE
F40%	0.29	0.35	0.35	0.36	2021
F40%	0.31	0.33	0.33	0.34	NOWL
F40%	0.30	0.32	0.32	0.33	NOWL+AGE
F40%	0.34	0.30	0.27	0.35	NOWL+AGE+WT
F40%	0.36	0.33	0.25	0.37	NOWL+AGE+WT+SE
Bratio_2021	0.352	0.424	0.408	0.367	2021
Bratio_2021	0.382	0.407	0.398	0.357	NOWL
Bratio_2021	0.360	0.400	0.393	0.367	NOWL+AGE
Bratio_2021	0.479	0.377	0.380	0.473	NOWL+AGE+WT
Bratio_2021	0.424	0.284	0.405	0.447	NOWL+AGE+WT+SE
SPRratio_2020	0.60	0.52	0.54	0.56	2021
SPRratio_2020	0.58	0.55	0.56	0.58	NOWL
SPRratio_2020	0.60	0.55	0.56	0.57	NOWL+AGE
SPRratio_2020	0.48	0.58	0.56	0.45	NOWL+AGE+WT
SPRratio_2020	0.49	0.63	0.58	0.46	NOWL+AGE+WT+SE

Table 7: Tuned sigma values for annually varying parameters. Version is 2021 = 2021 base models, NOWL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey. Note that the NOWL and NOWL+AGE versions were not retuned from the 2021 values.

Model	***Model 19.12***	***Model 19.12A***	***Model 21.1***	***Model 21.2***	***Version***
ln sigma R	0.6637	0.6651	0.6663	0.6453	2021,NOWL,NOWL+AGE
ln sigma R	0.6719	0.6604	0.7170	0.6132	NOWL+AGE+WT
ln sigma R	0.7037	0.7235	0.6280	0.6623	NOWL+AGE+WT+SE
L min	0.1752	0.1757	0.1730	0.1749	2021,NOWL,NOWL+AGE
L min	0.2965	0.2077	0.1518	0.2012	NOWL+AGE+WT
L min	0.2067	0.2021	0.1978	0.1978	NOWL+AGE+WT+SE
Ascend_se (fishery)	0.1595	0.1634	0.1819	0.1903	2021,NOWL,NOWL+AGE
Ascend_se (fishery)	0.2525	0.2481	0.2710	0.2657	NOWL+AGE+WT
Ascend_se (fishery)	0.2509	0.2442	0.2795	0.2521	NOWL+AGE+WT+SE
End_logit (fishery)	0.7610	0.8870	0.6760	1.3919	2021,NOWL,NOWL+AGE
End_logit (fishery)	1.4967	1.2715	1.3599	1.8832	NOWL+AGE+WT
End_logit (fishery)	1.5607	1.3512	1.3937	1.5919	NOWL+AGE+WT+SE
Ascend_se (survey)	0.8394	0.8342	0.7610	0.7428	2021,NOWL,NOWL+AGE
Ascend_se (survey)	1.3657	1.2910	1.4924	1.4711	NOWL+AGE+WT
Ascend_se (survey)	1.3777	1.3255	1.4270	1.5538	NOWL+AGE+WT+SE
Peak (survey)	0.2255	0.2194	0.2071	0.2033	2021,NOWL,NOWL+AGE
Peak (survey)	0.3462	0.3199	0.3758	0.3697	NOWL+AGE+WT
Peak (survey)	0.3508	0.3328	0.3445	0.3909	NOWL+AGE+WT+SE

Table 8. Dirichlet multinomial sample size multiplier. Grey values were fixed near the upper bound.

Label	*Model 19.12*	*Model 19.12A*	*Model 21.1*	*Model 21.2*	*Version*
Fishery Length	1	1	1	1	2021
Fishery Length	1	1	1	1	NOWL
Fishery Length	1	1	1	1	NOWL+AGE
Fishery Length	0.643	0.607	0.658	0.633	NOWL+AGE+WT
Fishery Length	0.644	0.609	0.675	0.647	NOWL+AGE+WT+SE
Survey Length	1	1	1	1	2021
Survey Length	1	1	1	1	NOWL
Survey Length	1	1	1	1	NOWL+AGE
Survey Length	0.589	0.622	0.578	0.547	NOWL+AGE+WT
Survey Length	0.595	0.640	0.602	0.587	NOWL+AGE+WT+SE
Survey Age	0.496	0.419	0.434	0.384	2021
Survey Age	0.470	0.441	0.448	0.393	NOWL
Survey Age	0.394	0.366	0.371	0.324	NOWL+AGE
Survey Age	0.249	0.290	0.250	0.235	NOWL+AGE+WT
Survey Age	0.245	0.284	0.228	0.247	NOWL+AGE+WT+SE

Table 9. Resulting average input sample size after Dirichlet multinomial sample size multiplier applied.

Label	*Model 19.12*	*Model 19.12A*	*Model 21.1*	*Model 21.2*	*Version*
Fishery Length	358	358	358	358	2021
Fishery Length	358	358	358	358	NOWL
Fishery Length	358	358	358	358	NOWL+AGE
Fishery Length	3616	3416	3701	3560	NOWL+AGE+WT
Fishery Length	3625	3424	3795	3640	NOWL+AGE+WT+SE
Survey Length	358	358	358	358	2021
Survey Length	358	358	358	358	NOWL
Survey Length	358	358	358	358	NOWL+AGE
Survey Length	1054	1111	1033	979	NOWL+AGE+WT
Survey Length	1063	1144	1076	1050	NOWL+AGE+WT+SE
Survey Age	177	150	155	137	2021
Survey Age	168	158	160	140	NOWL
Survey Age	141	131	133	116	NOWL+AGE
Survey Age	89	104	89	84	NOWL+AGE+WT
Survey Age	87	102	82	88	NOWL+AGE+WT+SE

Table 10. Joint RMSE values (Carvalho et al. 2021) for all models and versions. Version is 2021 = 2021 base models, NOWL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey.

Label	*Model 19.12*	*Model 19.12A*	*Model 21.1*	*Model 21.2*	*Version*
Indices	5.9	13	13.1	11.7	*2021*
Indices	5.8	13	13.1	11.8	*NOWL*
Indices	5.9	13	13	11.8	*NOWL+AGE*
Indices	7.3	17.2	17.2	16.7	*NOWL+AGE+WT*
Indices	*25.8*	*25.7*	*28.9*	24.1	*NOWL+AGE+WT+SE*
Length Comp	*3.8*	*3.9*	*3.9*	4	*2021*
Length Comp	*3.7*	*3.9*	*3.9*	*4*	*NOWL*
Length Comp	*3.8*	*4*	*3.9*	*4.1*	*NOWL+AGE*
Length Comp	*3*	*3.4*	*2.6*	*3.3*	*NOWL+AGE+WT*
Length Comp	*2.7*	*3.4*	*3*	*3.1*	*NOWL+AGE+WT+SE*
Age Comp	*4.9*	5.5	5.5	*6.4*	*2021*
Age Comp	5	5.4	5.4	6.3	*NOWL*
Age Comp	5.5	6.1	6.1	*6.7*	*NOWL+AGE*
Age Comp	*6.6*	*6.2*	*6.8*	*6.4*	*NOWL+AGE+WT*
Age Comp	*6.7*	*6.9*	*7.2*	*7.2*	*NOWL+AGE+WT+SE*

Table 11. Residual runs test (Carvalho et al. 2021) p-Values for fit to survey and fishery CPUE indices for all models and versions. The p-value is a test of whether the observed residual distribution is further than three standard deviations away from the expected residual process average of 0. Red values are significantly different at α = 0.05. Version is 2021 = 2021 base models, NOWL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey.

Version	*Model 19.12*	*Model 19.12A*	*Model 21.1*	*Model 21.2*	*Label*
2021	0.315	0.315	0.566	0.008	BT Survey index
NOWL	0.315	0.147	0.147	0.008	BT Survey index
NOWL+AGE	0.315	0.315	0.315	0.008	BT Survey index
NOWL+AGE+WT	0.135	0.013	0.135	0.147	BT Survey index
NOWL+AGE+WT+SE	0.021	0.58	0.008	0.129	BT Survey index
2021				0.120	Fishery Index
NOWL				0.120	Fishery Index
NOWL+AGE				0.120	Fishery Index
NOWL+AGE+WT				0.024	Fishery Index
NOWL+AGE+WT+SE				0.000	Fishery Index
2021	0.019	0.002	0.012	0.000	Fishery Length
NOWL	0.002	0.012	0.002	0.000	Fishery Length
NOWL+AGE	0.002	0.003	0.002	0.000	Fishery Length
NOWL+AGE+WT	0.049	0.099	0.087	0.024	Fishery Length
NOWL+AGE+WT+SE	0.000	0.209	0.155	0.091	Fishery Length
2021	0.129	0.001	0.001	0.000	Survey Length
NOWL	0.129	0.001	0.001	0.000	Survey Length
NOWL+AGE	0.326	0.001	0.001	0.000	Survey Length
NOWL+AGE+WT	0.039	0.348	0.533	0.111	Survey Length
NOWL+AGE+WT+SE	0.081	0.326	0.081	0.199	Survey Length
2021	0.512	0.512	0.512	0.08	Survey Age
NOWL	0.512	0.512	0.512	0.08	Survey Age
NOWL+AGE	0.704	0.057	0.057	0.219	Survey Age
NOWL+AGE+WT	0.355	0.355	0.448	0.541	Survey Age
NOWL+AGE+WT+SE	0.704	0.355	0.355	0.355	Survey Age

Table 12. Mean absolute scaled error (MASE) values for model data components for all models and versions. Version is 2021 = 2021 base models, NOWL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey. Red values indicate predictions skills worse than a random walk (Carvalho et al. 2021).

Version	*Model 19.12*	*Model 19.12A*	*Model 21.1*	*Model 21.2*	*Label*
2021	0.19	0.36	0.36	0.51	BT Survey Index
NOWL	0.17	0.35	0.35	0.50	BT Survey Index
NOWL+AGE	0.18	0.35	0.34	0.50	BT Survey Index
NOWL+AGE+WT	0.26	0.48	0.47	0.68	BT Survey Index
NOWL+AGE+WT+SE	*1.03*	*1.14*	*1.10*	0.99	BT Survey Index
2021				0.55	CPUE Index
NOWL				0.53	CPUE Index
NOWL+AGE				0.47	CPUE Index
NOWL+AGE+WT				*1.04*	CPUE Index
NOWL+AGE+WT+SE				*2.46*	CPUE Index
2021	*0.33*	*0.31*	*0.33*	*0.38*	Fishery Mean Length
NOWL	*0.29*	*0.31*	*0.31*	*0.38*	Fishery Mean Length
NOWL+AGE	*0.28*	*0.30*	*0.31*	*0.37*	Fishery Mean Length
NOWL+AGE+WT	*0.42*	*0.29*	*0.37*	*0.43*	Fishery Mean Length
NOWL+AGE+WT+SE	*0.61*	*0.45*	*0.50*	*0.50*	Fishery Mean Length
2021	*1.00*	0.93	0.92	*1.00*	Survey Mean Length
NOWL	0.93	0.92	0.91	0.99	Survey Mean Length
NOWL+AGE	0.96	0.92	0.91	*1.00*	Survey Mean Length
NOWL+AGE+WT	*1.43*	*1.28*	*1.30*	*1.37*	Survey Mean Length
NOWL+AGE+WT+SE	*1.51*	*1.80*	*1.77*	*1.75*	Survey Mean Length
2021	0.83	0.76	0.77	0.78	Survey Mean Age
NOWL	0.77	0.74	0.74	0.79	Survey Mean Age
NOWL+AGE	0.87	0.89	0.87	0.89	Survey Mean Age
NOWL+AGE+WT	*1.35*	*1.09*	*1.10*	*1.21*	Survey Mean Age
NOWL+AGE+WT+SE	*1.34*	*1.58*	*1.58*	*1.59*	Survey Mean Age

Figures

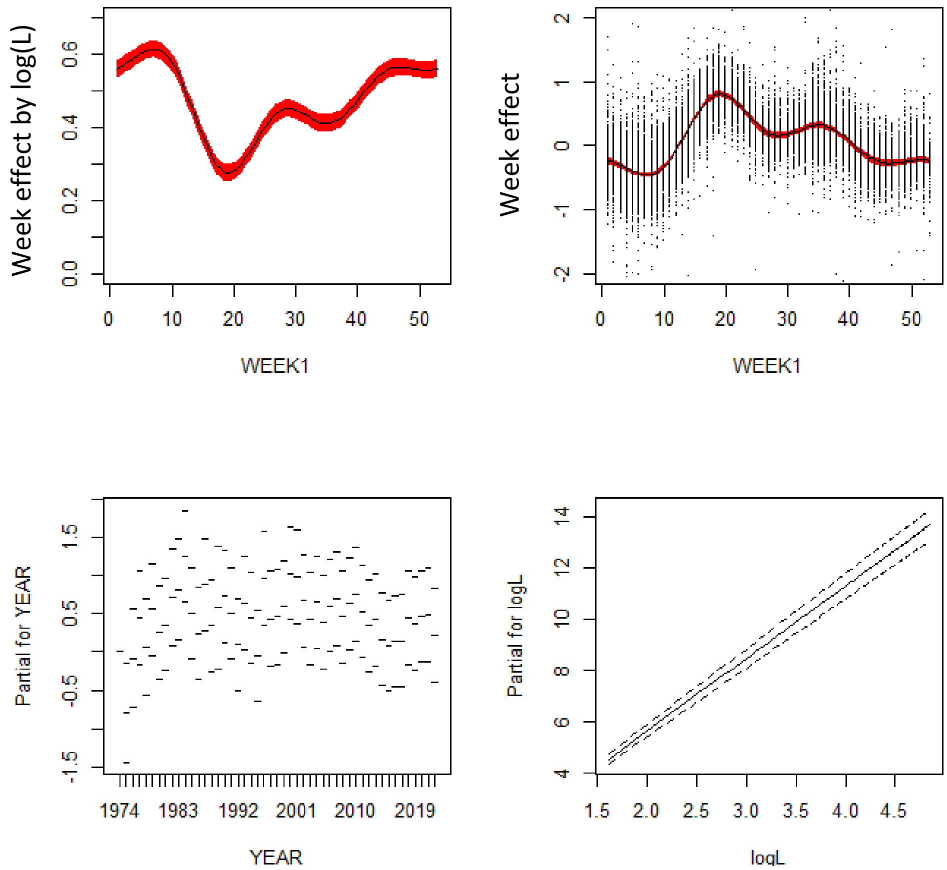


Figure 1. GAM model of weekly and annual effects on weight at length.

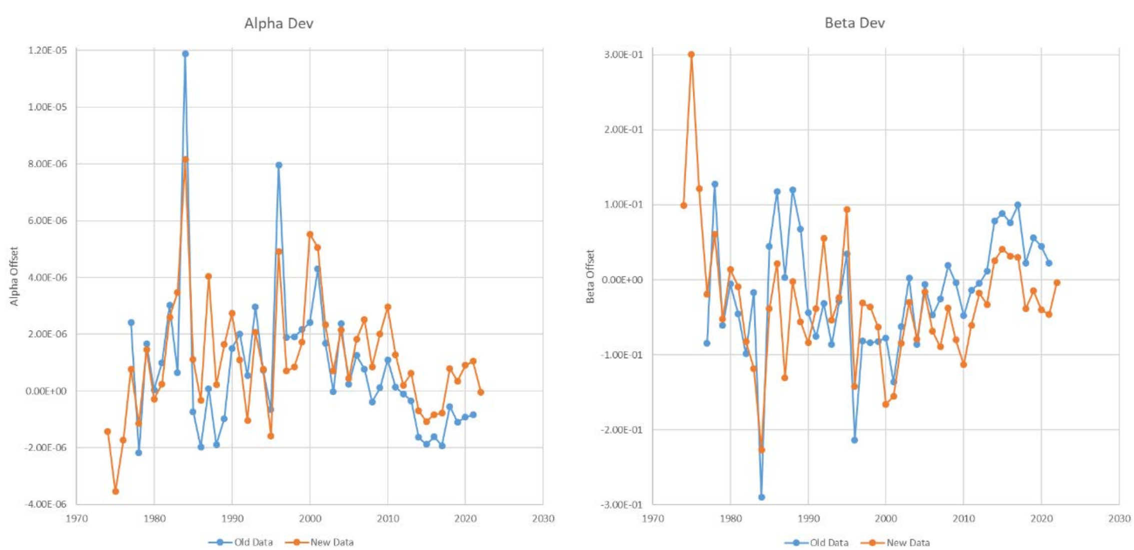


Figure 2. Annual deviation indices for Alpha and Beta for the weight at length relationship used in the assessment models for 1974-2022 for old and new method.

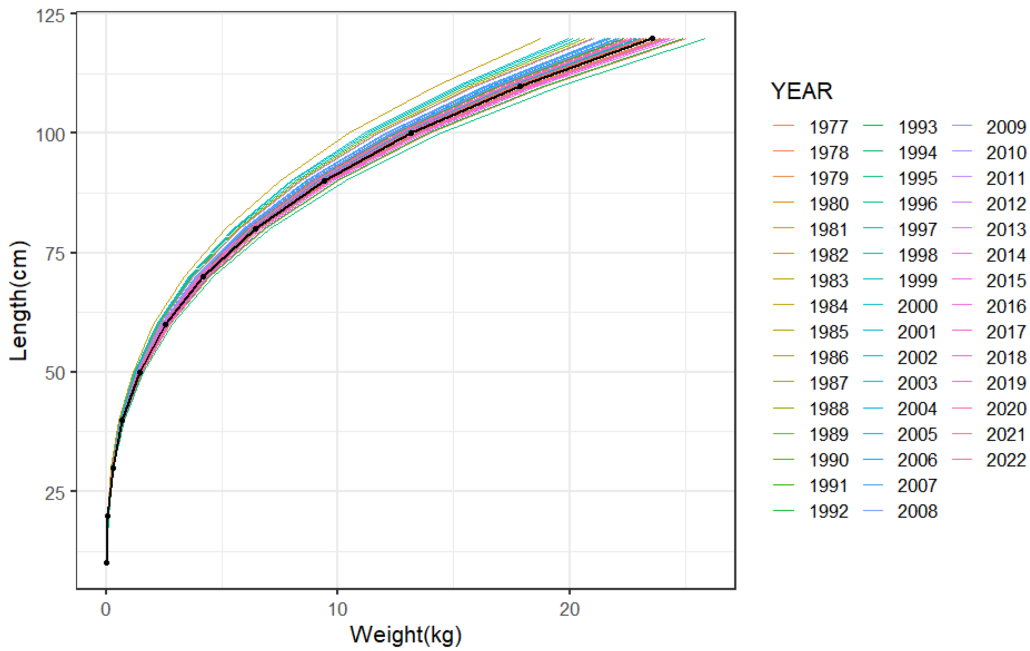


Figure 3. Variability in weight at length for BS Pacific cod 1977-2022 for new method. The black line is the overall weight at length relationship for all data.

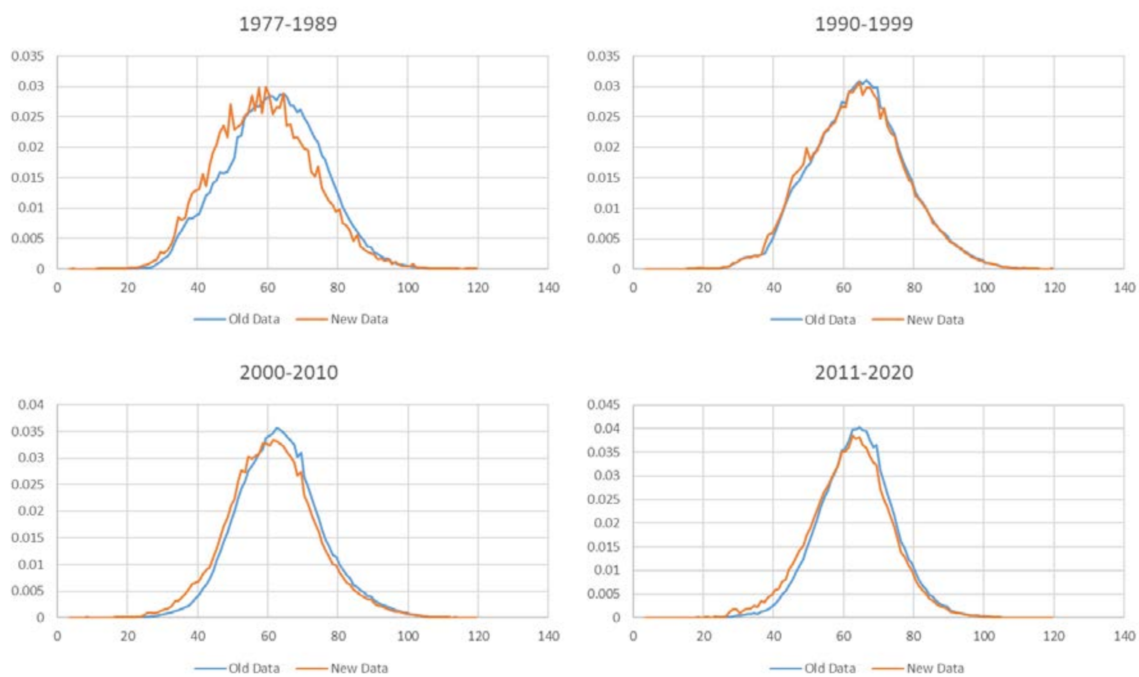


Figure 4. Overall fishery length distributions summed by each decade from the method used by the previous lead author (Old Data) and the new lead author (New Data).

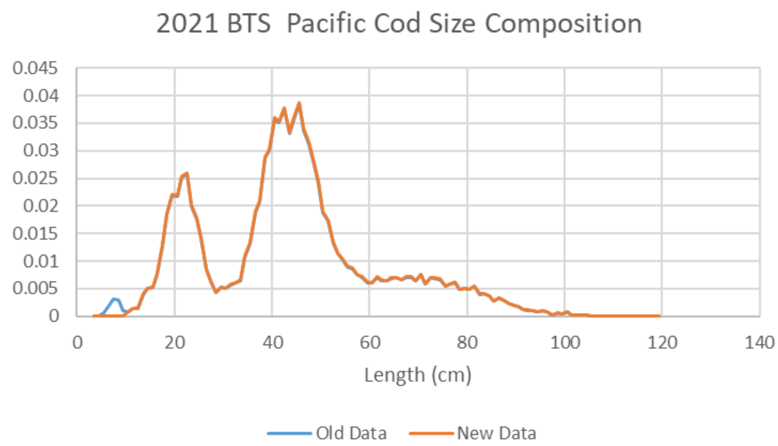


Figure 5. 2021 Bottom trawl survey Pacific cod size composition data for old data and new data.

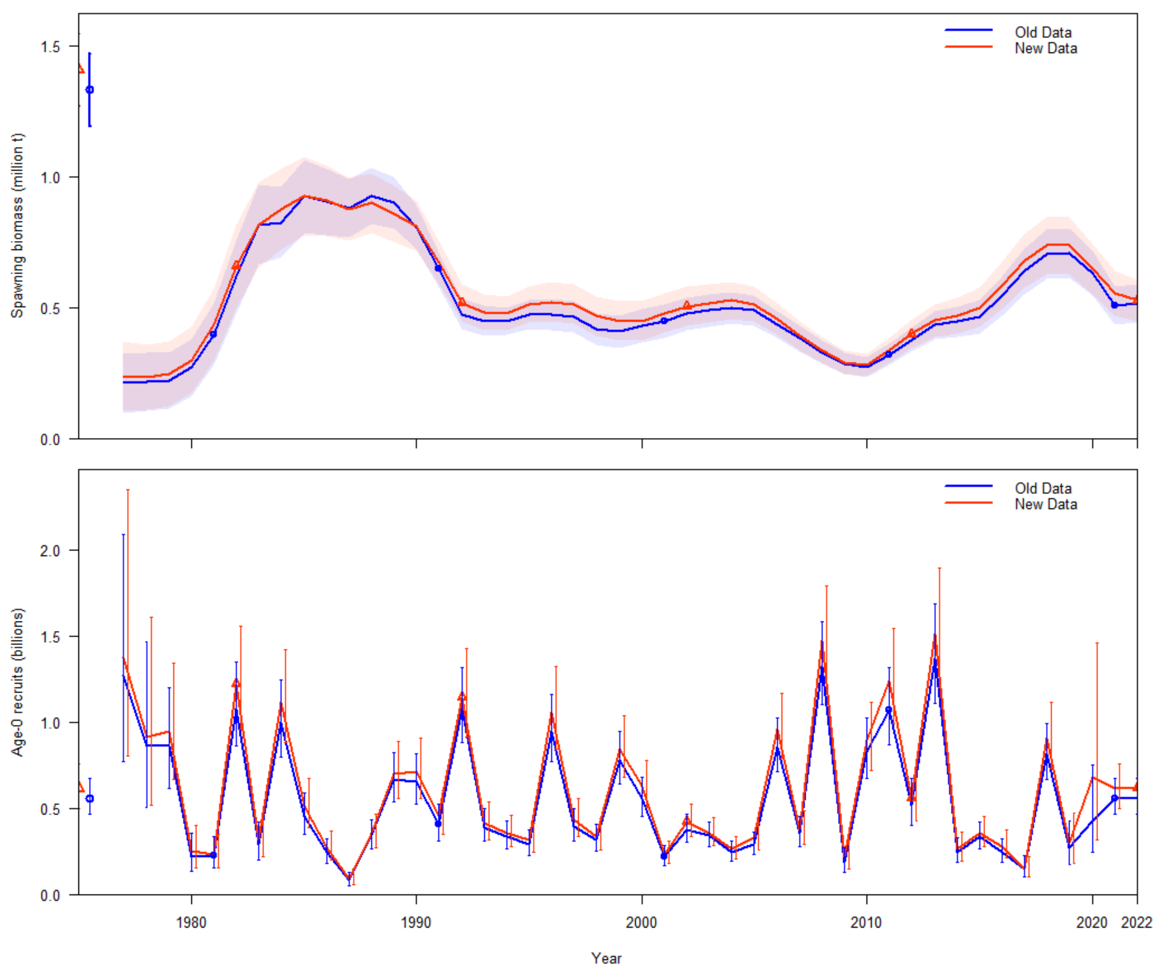


Figure 6. (Top) spawning biomass estimates (t x 10^9) and (bottom) age-o recruits (n x 10^{12}) from Model 19.12A with old and new length composition data and annual seasonally corrected weight at length relationship.

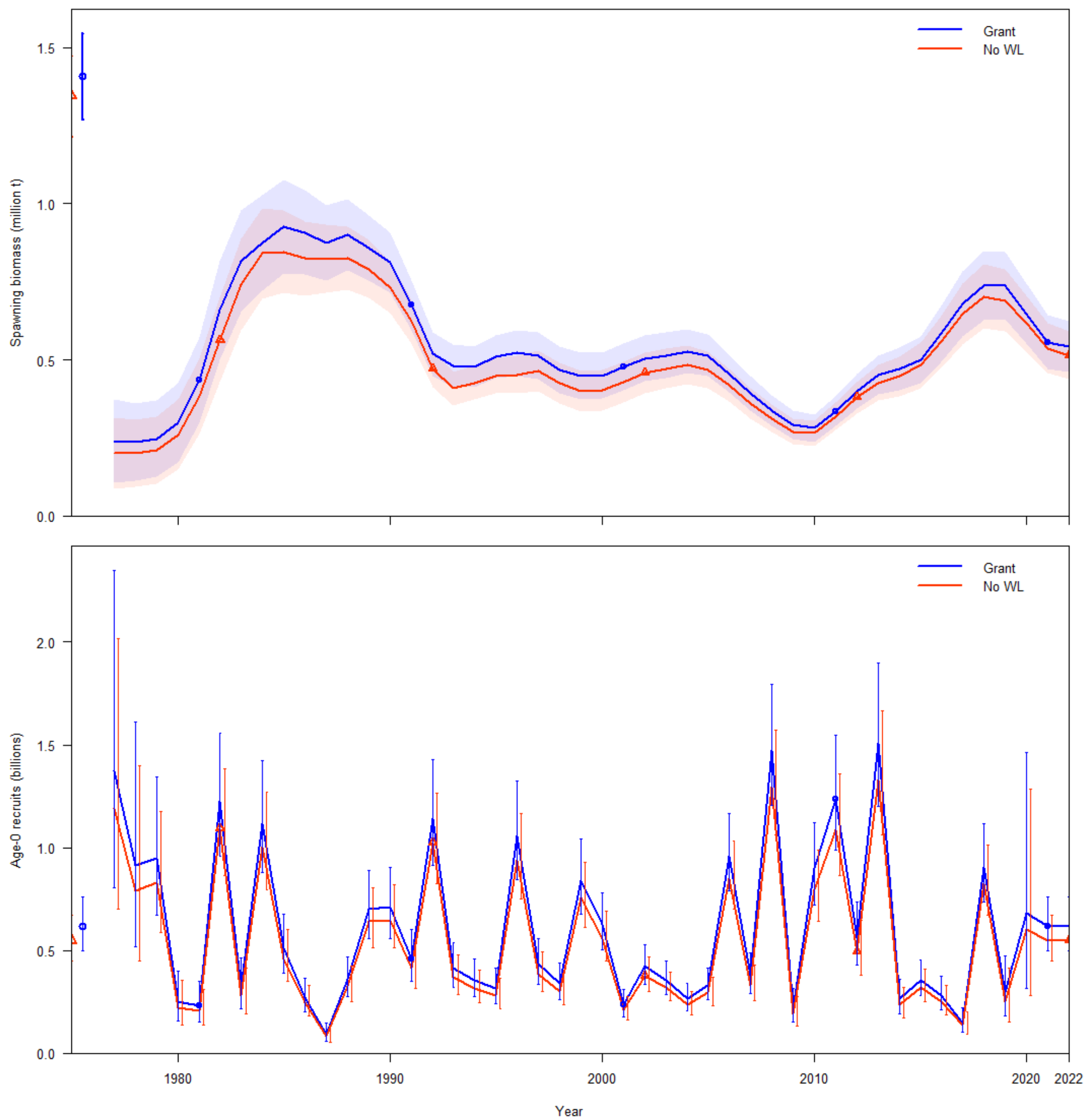


Figure 7. (Top) spawning biomass estimates (t x 10^9) and (bottom) age-0 recruits (n x 10^{12}) from Model 19.12A with new length composition data and with (GRANT) and without (No WL) annual seasonally corrected weight at length relationship.

Figure 8. Aging bias fit in all models and versions. X-axis is age in years, y-axis is average bias in years.

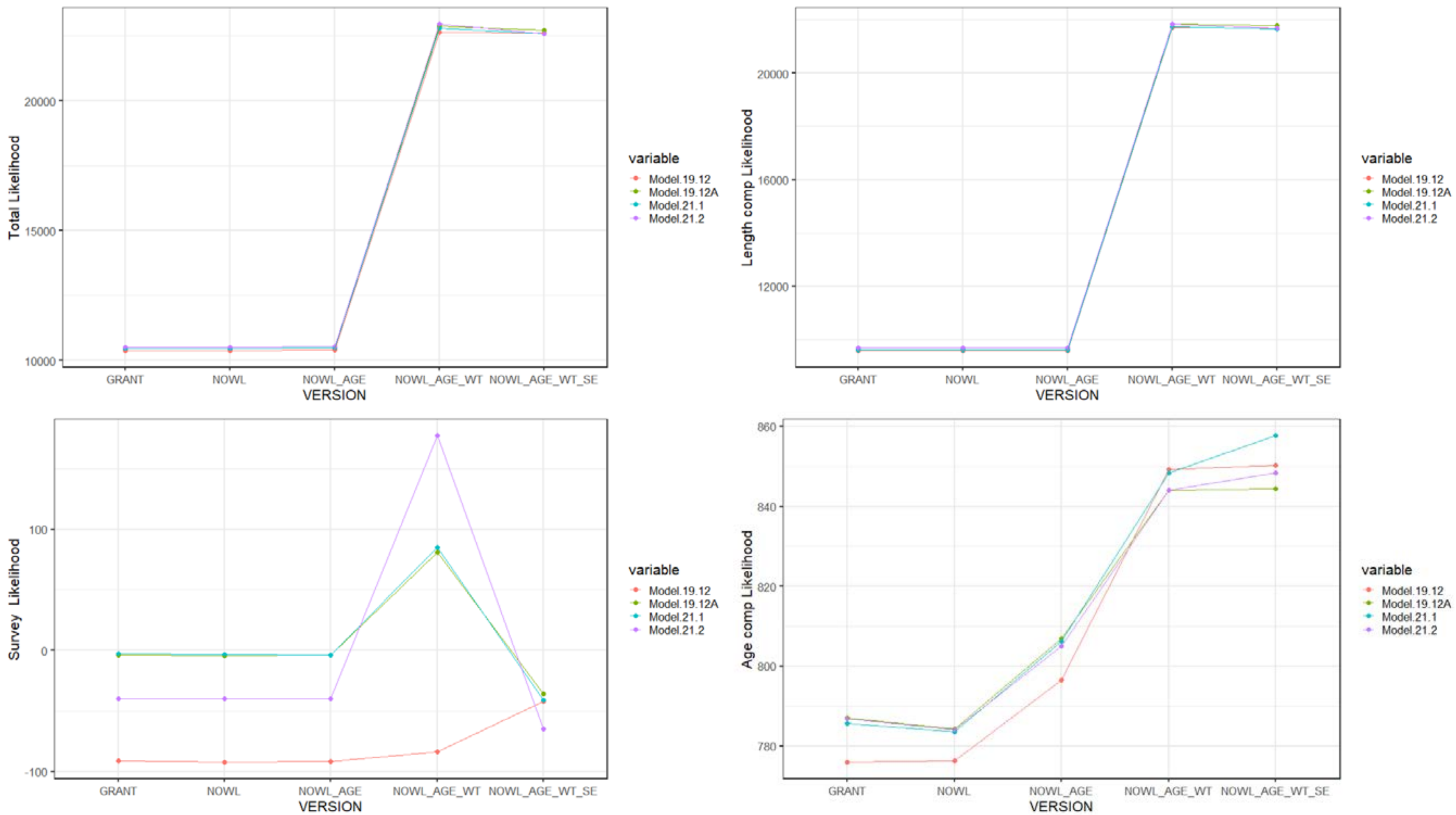


Figure 9. Comparison of likelihood elements from models with new data. Version is GRANT=2021 base models with new data, NOWL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey.

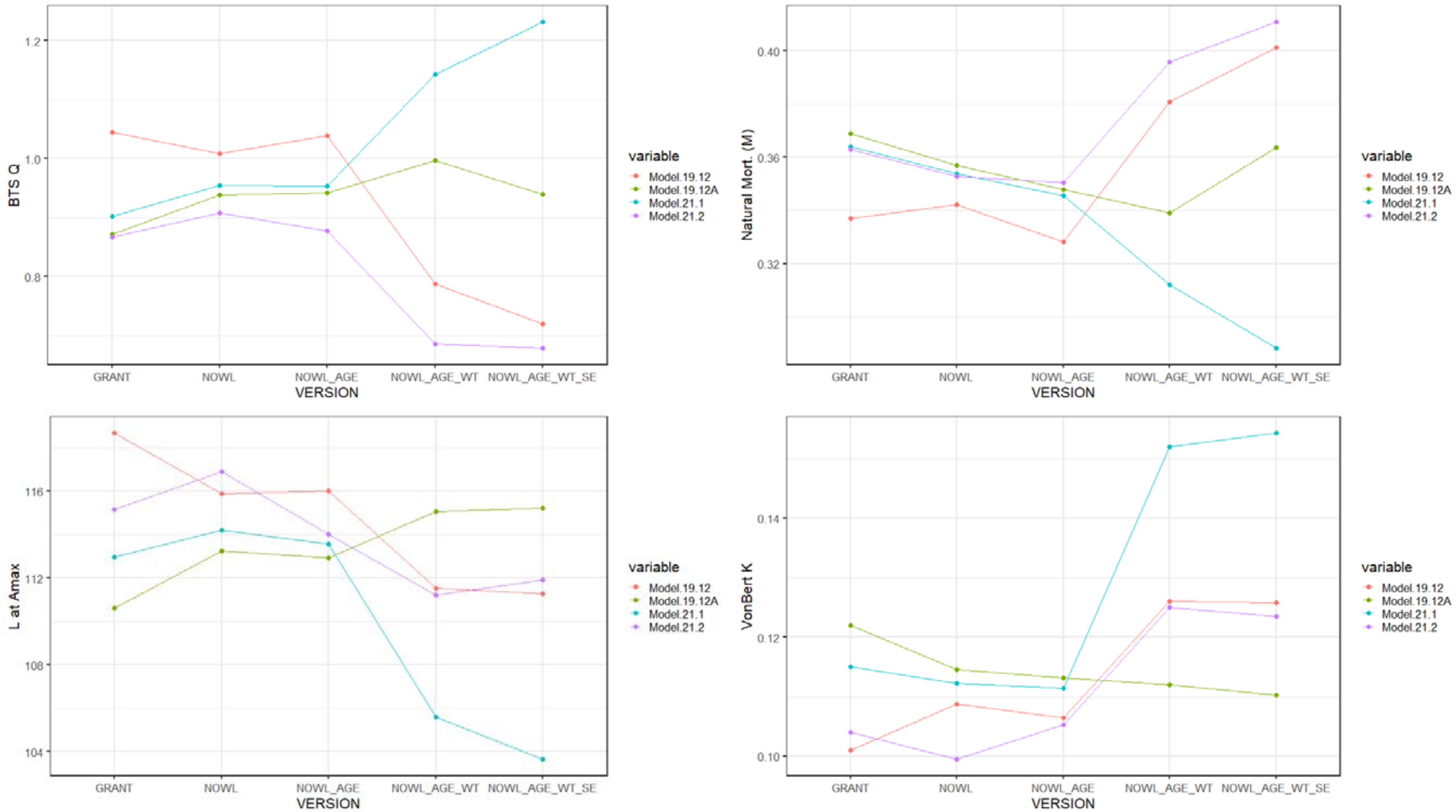


Figure 10. Comparison of key model results from models with new data. Version is GRANT=2021 base models with new data, NOWL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey.

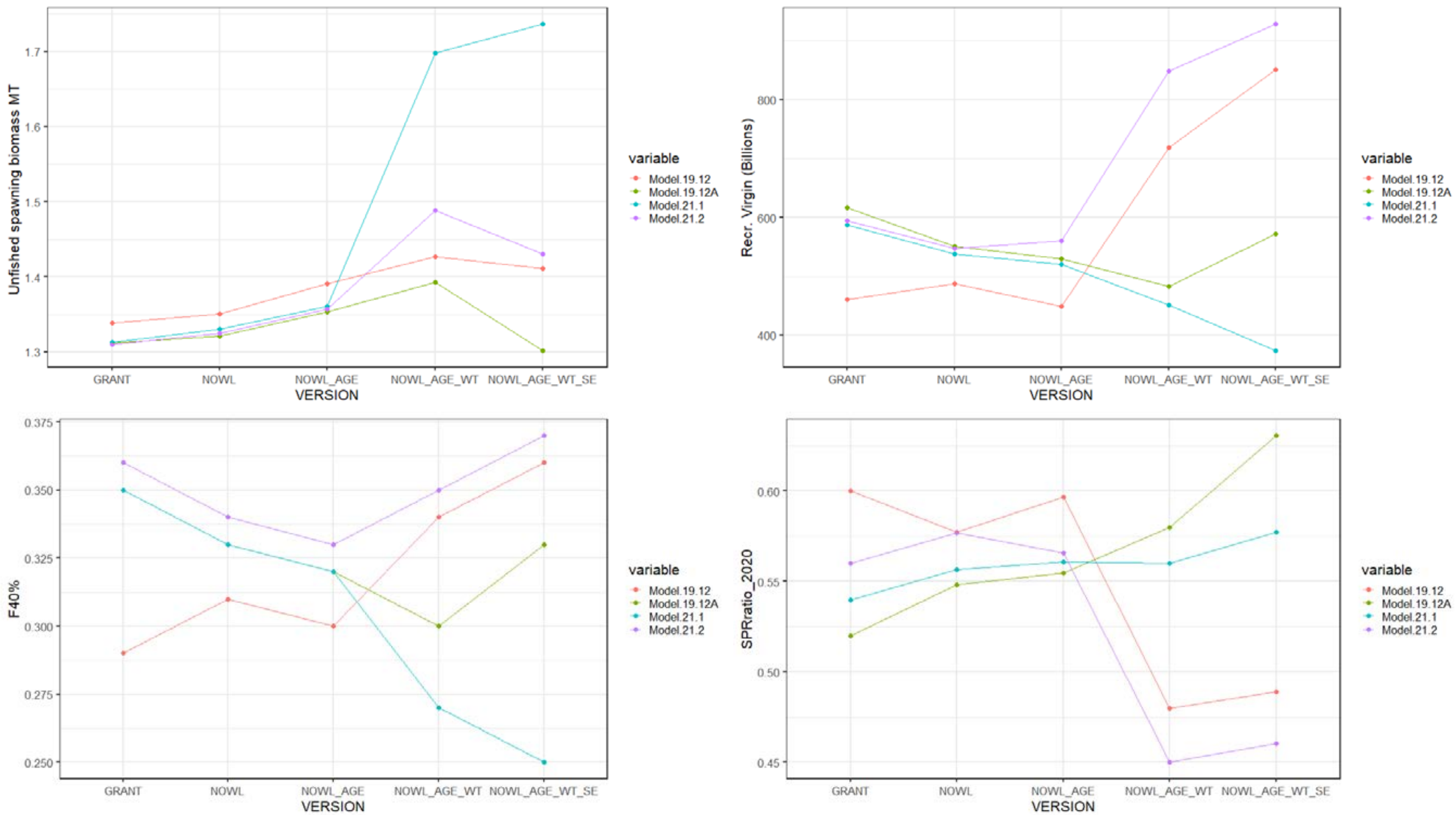
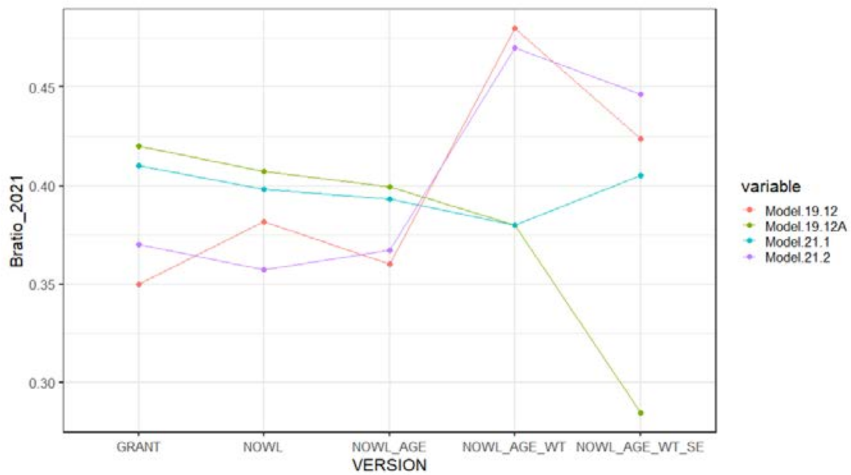


Figure 10 Cont. Comparison of key model results from models with new data. Version is GRANT=2021 base models with new data, NOWL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey.



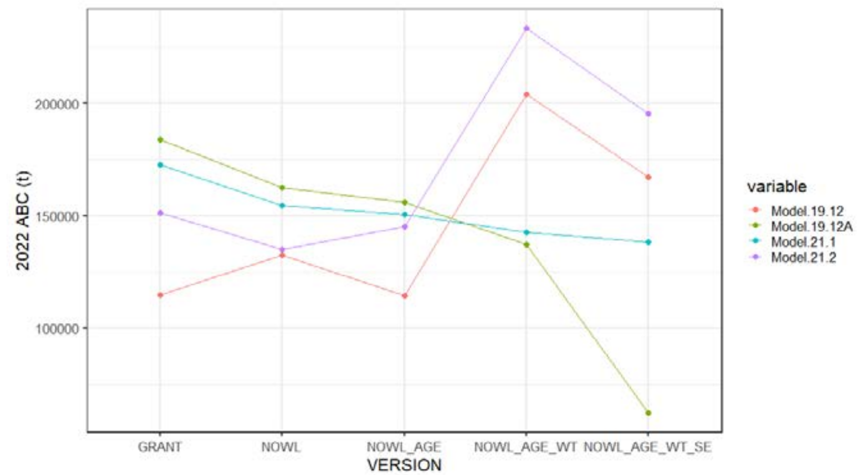


Figure 10 Cont. Comparison of key model results from models with new data. Version is GRANT=2021 base models with new data, NOWL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey.

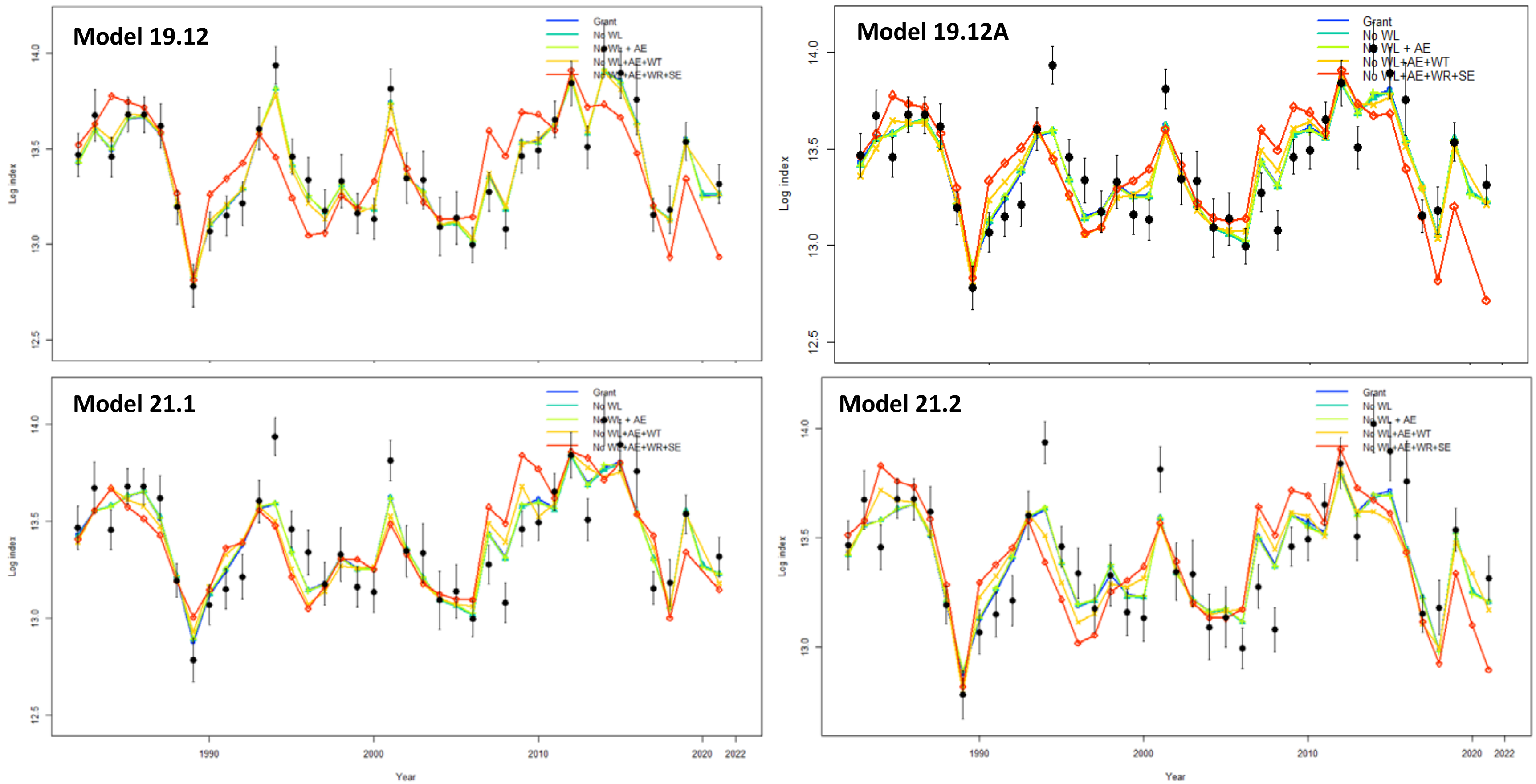


Figure 11. Fit to the VAST combined Bering Sea bottom trawl survey index (log numbers) for alternative models with new data with versions GRANT = 2021 base model, No WL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey.

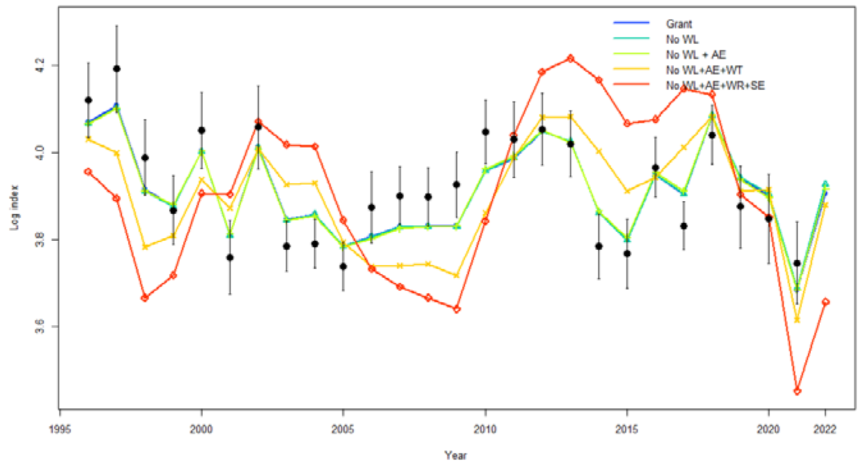


Figure 12. Model 21.2 fit to the winter longline fishery VAST CPUE index (log numbers) for alternative models with new data with versions GRANT = 2021 base model, No WL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey and CPUE Index.

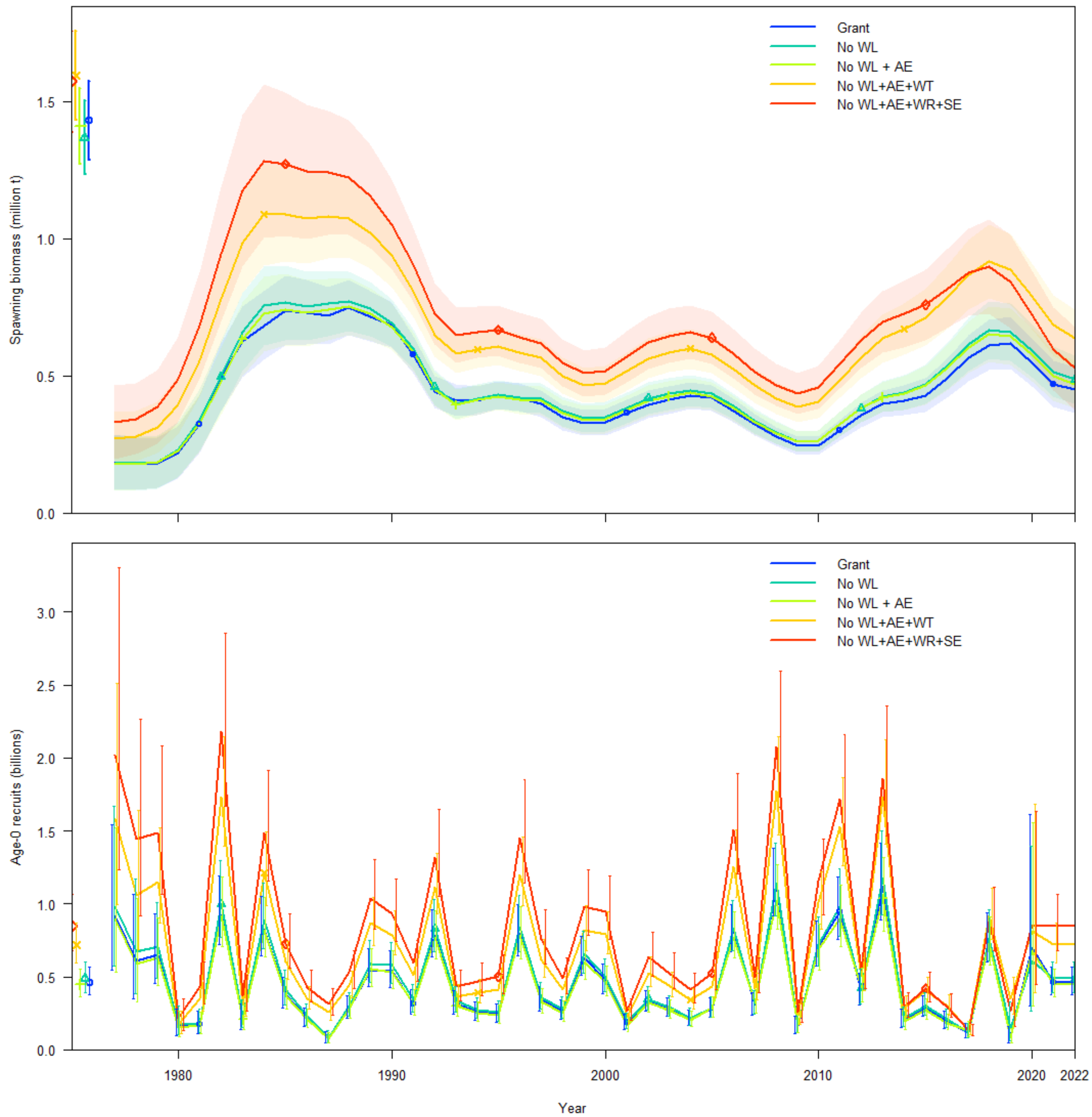


Figure 13. (Top) spawning biomass estimates (t x 10^9) and (bottom) age-o recruits (n x 10^{12}) from Model 19.12 with new data with versions GRANT = 2021 base model, No WL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey.

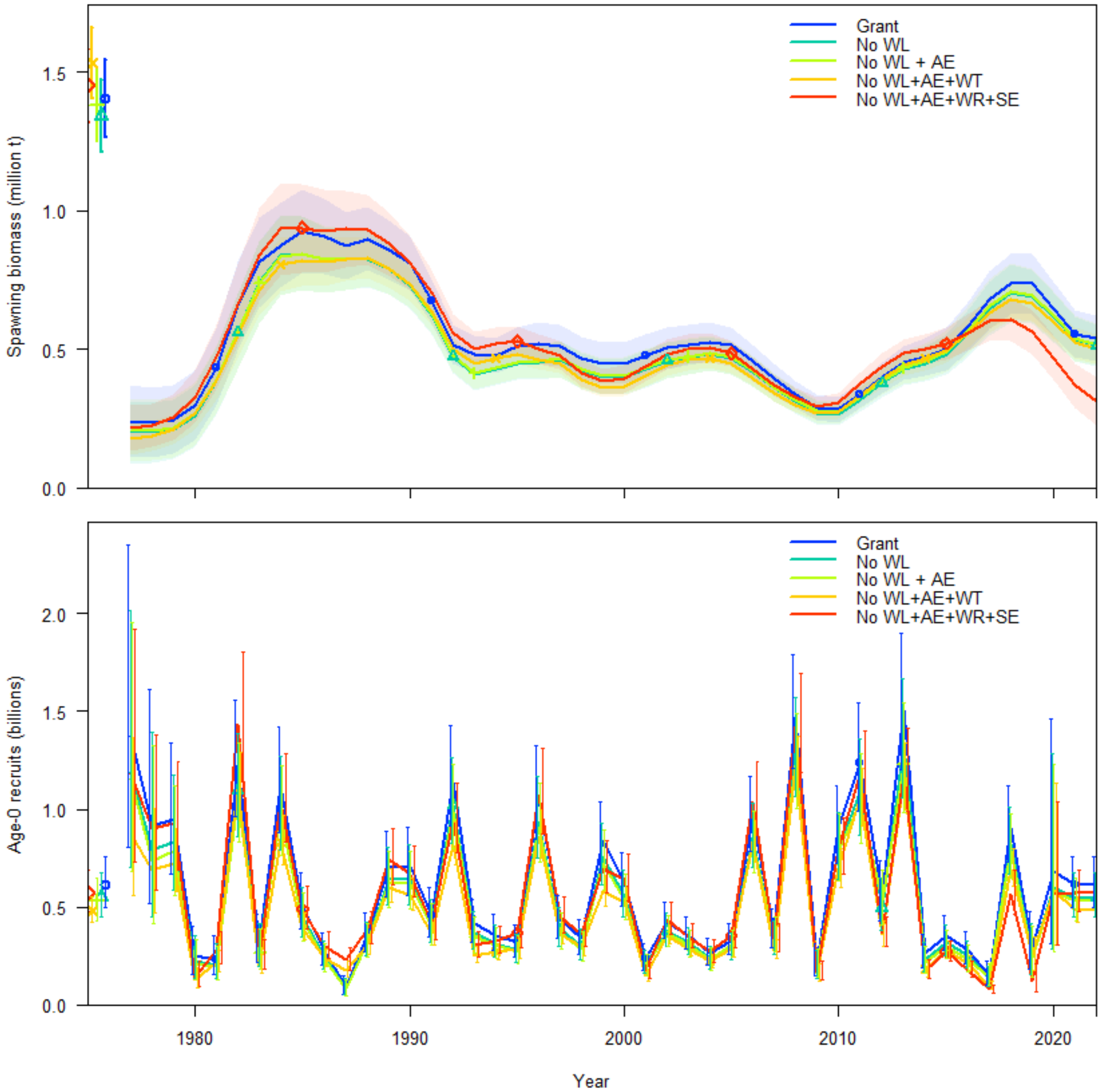


Figure 14. (Top) spawning biomass estimates (t x 10^9) and (bottom) age-o recruits (n x 10^{12}) from Model 19.12A with new data with versions GRANT = 2021 base model, No WL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey.

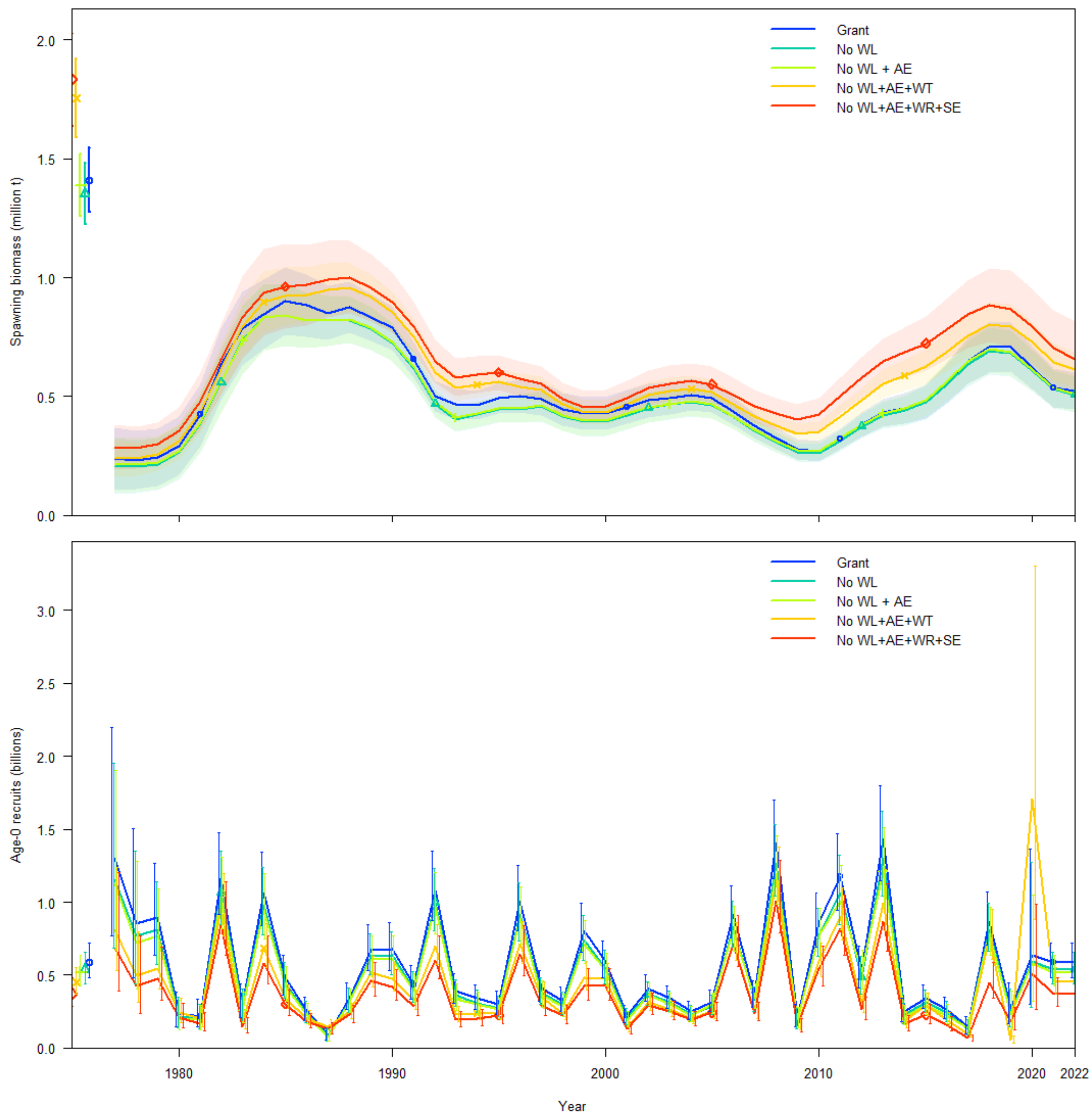


Figure 15. (Top) spawning biomass estimates (t x 10^9) and (bottom) age-o recruits (n x 10^{12}) from Model 21.1 with new data with versions GRANT = 2021 base model, No WL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey.

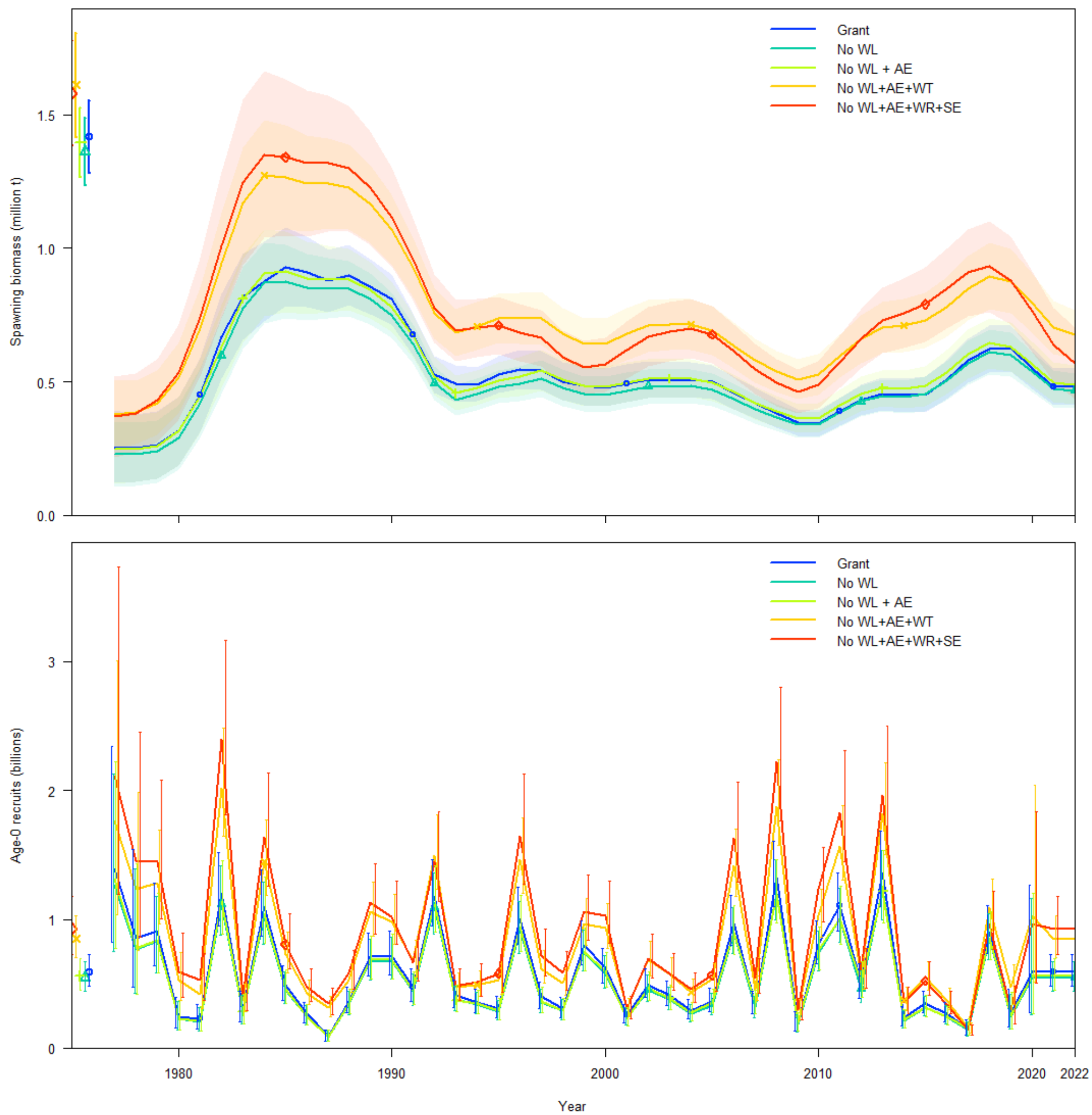


Figure 16. (Top) spawning biomass estimates (t x 10^9) and (bottom) age-o recruits (n x 10^{12}) from Model 21.2 with new data with versions GRANT = 2021 base model, No WL=No seasonally corrected weight at length relationship, +AGE = New Aging bias, +WT = new length composition data input sample sizes, +SE = Fit extra standard error for bottom trawl survey.

Models No WL+AE+WT | **Models No WL+AE+WT+SE**

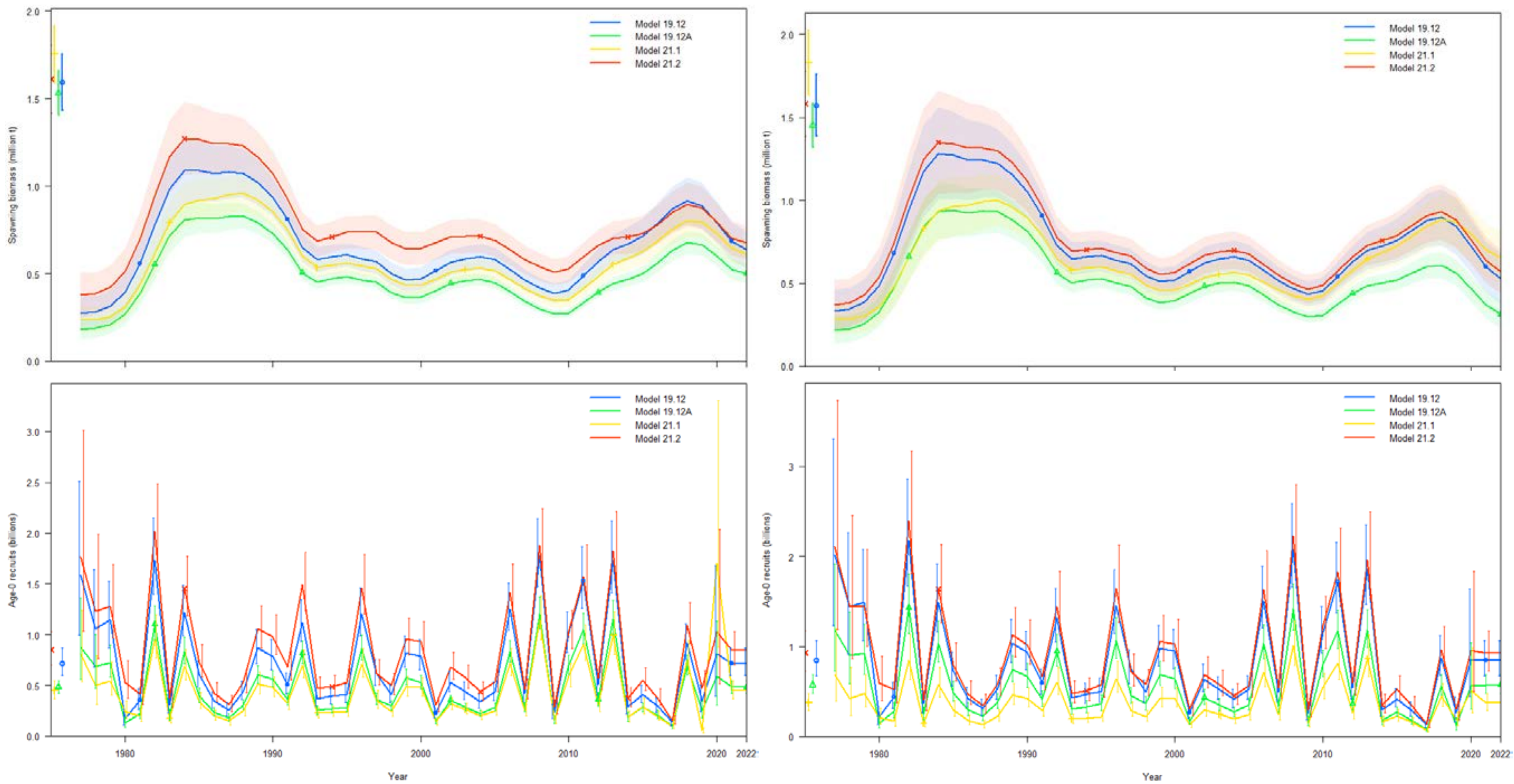


Figure 17. (Top) spawning biomass estimates (t x 10^9) and (bottom) age-o recruits (n x 10^{12}) from alternative models with new data for (left) no seasonally corrected weight at length relationship, new aging bias and new length composition data input sample sizes and (right) for no seasonally corrected weight at length relationship, new aging bias, new length composition data input sample sizes and fit with extra standard error for bottom trawl survey

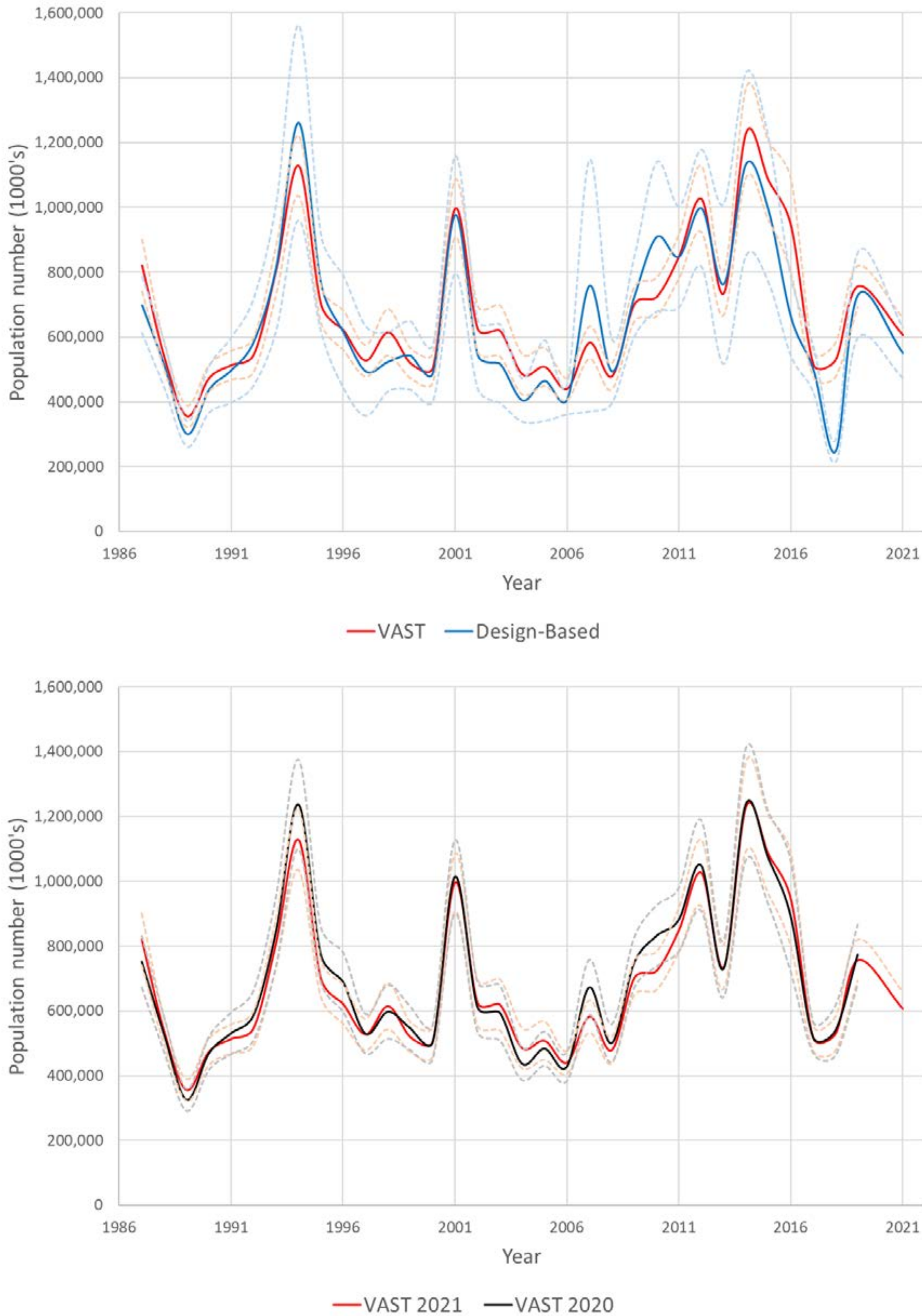


Figure 18. Eastern Bering Sea plus Northern Bering Sea survey indices for (top) the design-based in blue and 2021 VAST derived estimates in red and (bottom) the 2021 VAST derived estimates in red and 2020 VAST derived estimates in black.

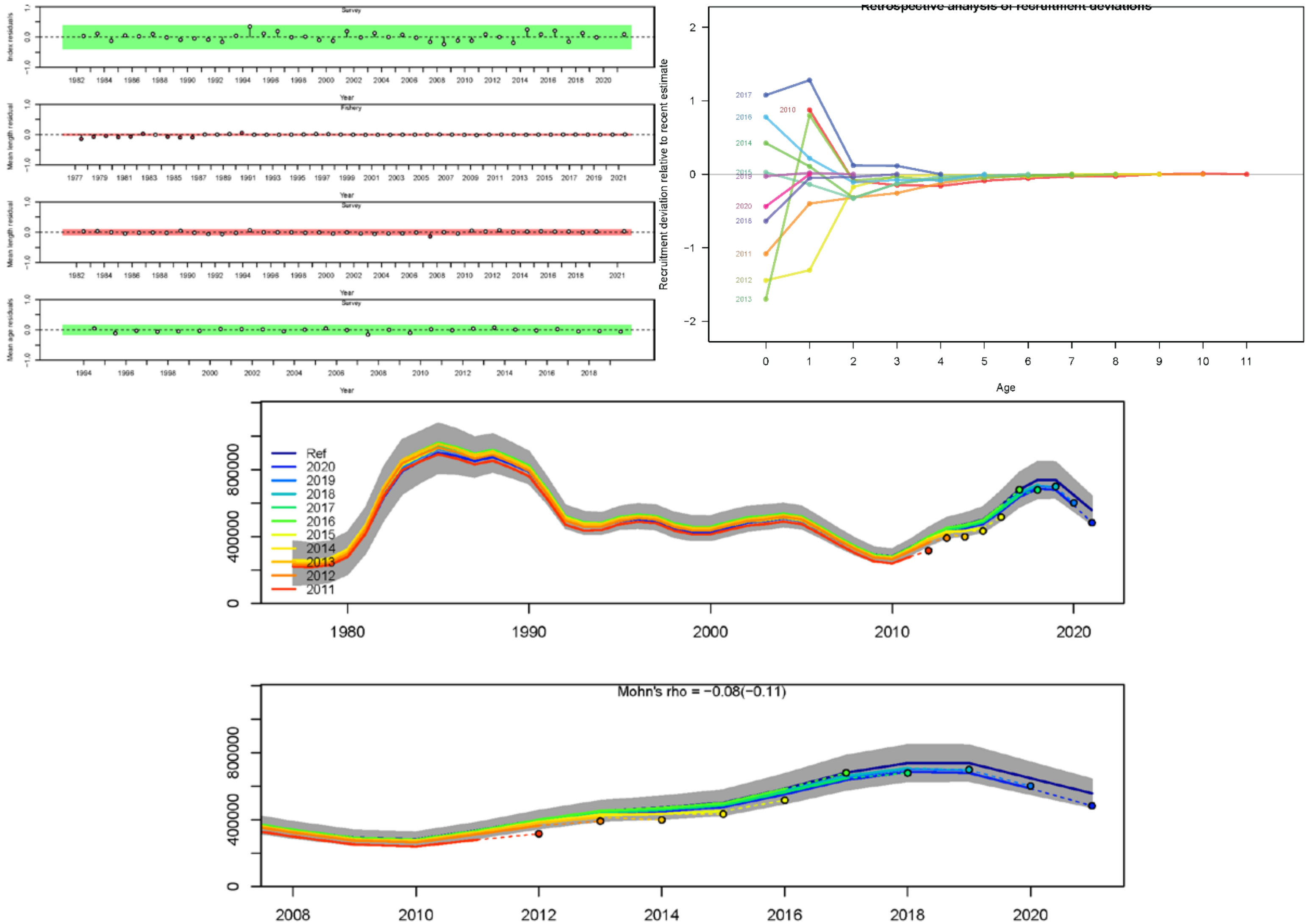


Figure 19. Model19.12A version 2021 result graphs from from Carvalho et al. (2021) (top left) residual run tests for correlated residuals, (top right) retrospective examination of year classes 2011-2020, (bottom) retrospective test showing spawning stock biomass (t).

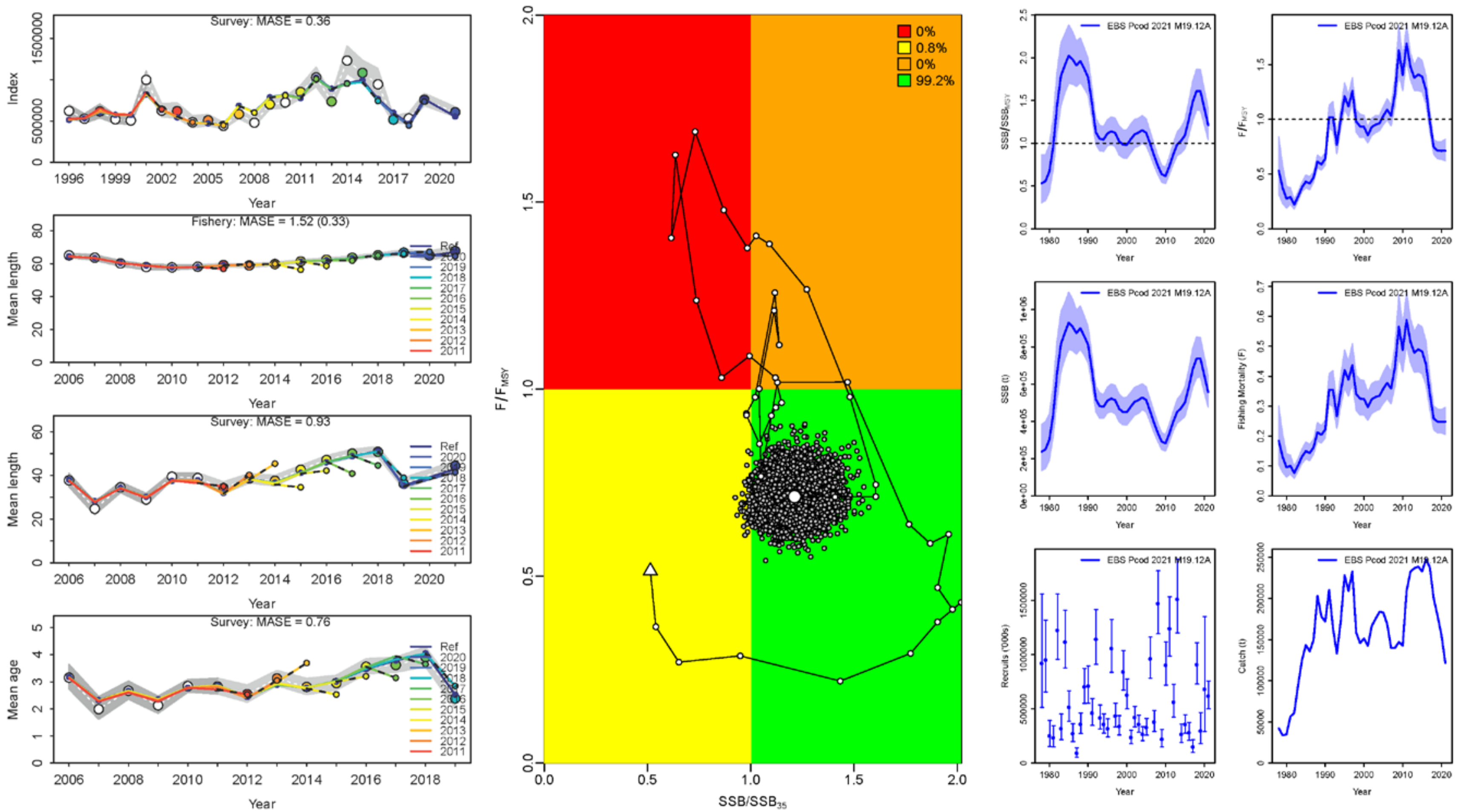


Figure 19 cont. Model19.12A version 2021 analysis results from Carvalho et al. (2021) (left) MASE analysis, (center) Kobe phase plot showing delta-Multivariate lognormal approximation Kobe probability distributions, (right) plots of various model results.

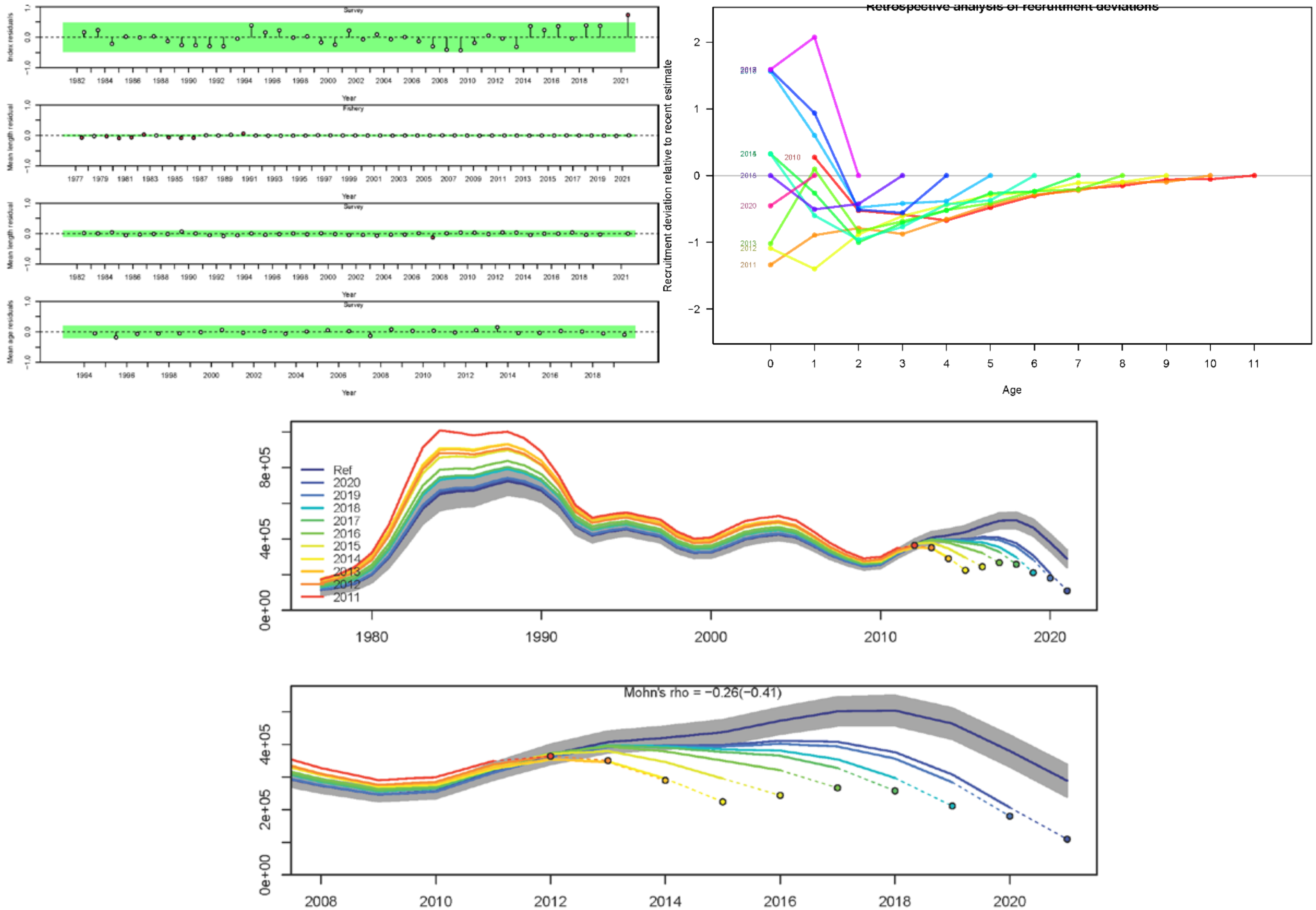


Figure 20. Model19.12A version NOWL+AGE+WT+SE result graphs from from Carvalho et al. (2021) (top left) residual run tests for correlated residuals, (top right) retrospective examination of year classes 2011-2020, (bottom) retrospective test showing spawning stock biomass (t).

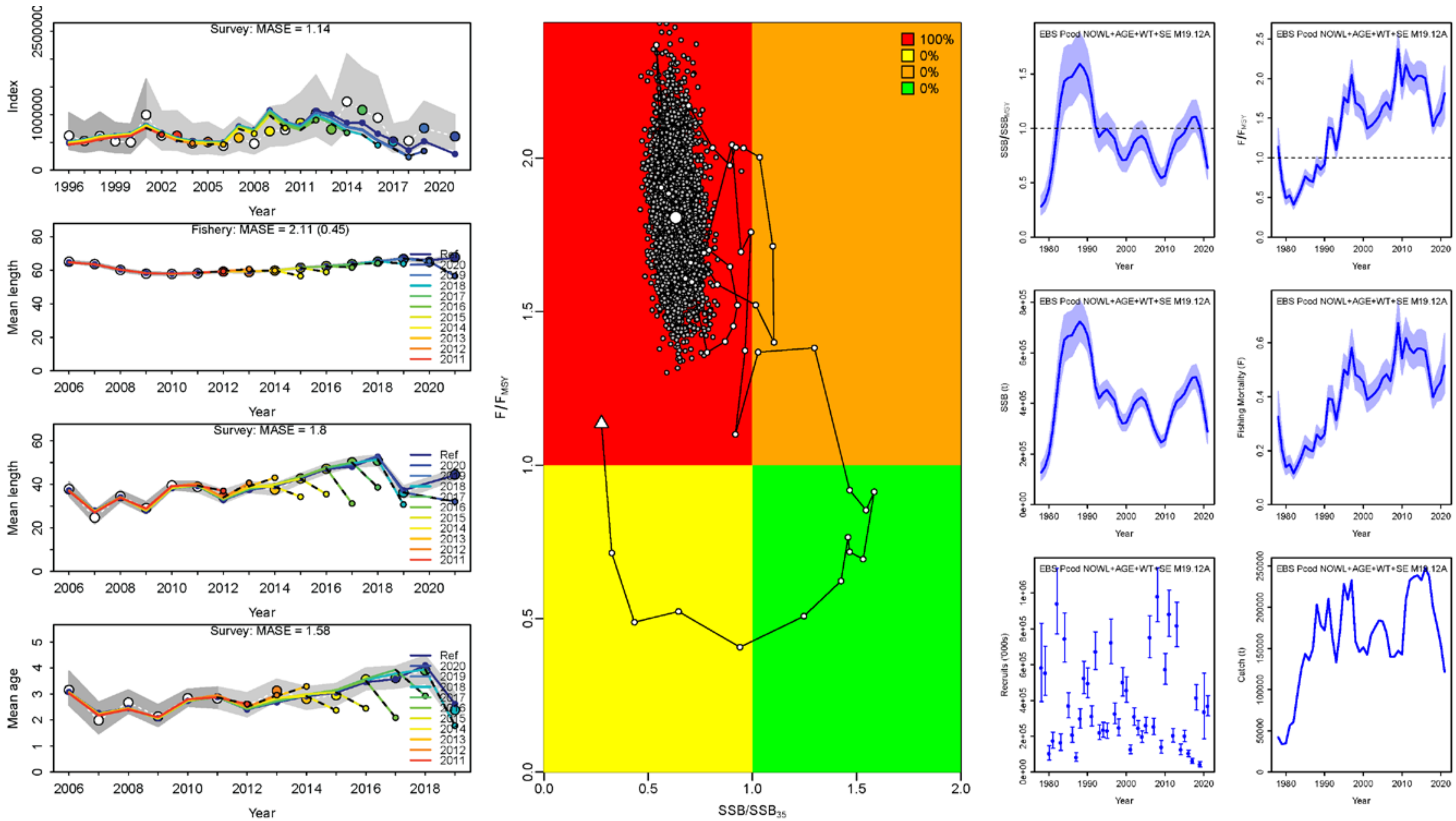


Figure 20 cont. Model19.12A version NOWL+AGE+WT+SE analysis results from Carvalho et al. (2021) (left) MASE analysis, (center) Kobe phase plot showing delta-Multivariate lognormal approximation Kobe probability distributions, (right) plots of various model results.