NOT TO BE CITED WITHOUT PRIOR
REFERENCE TO THE AUTHOR(S)

Northwest Atlantic Fisheries Organization

Serial No. N7316 **NAFO SCR Doc. 22/040**

SCIENTIFIC COUNCIL MEETING – JUNE 2022

A Mixed-Effects Model to Smooth and Interpolate Survey Weights-at-Age for 3NO Cod

by

N.G. Cadigan[1] and R.M. Rideout[2]

[1]Centre for Fisheries Ecosystem Research, Fisheries and Marine Institute - Memorial University of Newfoundland, St. John's, NL, Canada

[2]Northwest Atlantic Fisheries Centre, Fisheries and Oceans Canada,
St. John's, NL, Canada

Abstract

We propose a simple model to 1) smooth and reduce sampling error in time-series of survey estimates of fish body weights-at-age, and 2) fill in sampling gaps. We apply the model to average weights from both the Spring and Fall DFO bottom-trawl surveys in NAFO Divisions 3NO. The model is designed to fit the average weights closely when sampling precision is good, but to smooth values when the precision of estimates is low because sample sizes are low. We also use the model to estimate the average weight in a 3NO cod plus group starting at age 10. We demonstrate that age is the major source of variability in the weight-at-age data but there are some statistically significant cohort- and year-effects. The year-effects have an overall declining trend since about 2005 indicating overall declines in 3NO cod stock weight-at-age.

Introduction

Age-based stock assessment models typically estimate yearly (y) numbers-at-age *a* in the stock (i.e. N_{ay}) for a range of years. These models use estimates of the average weight-at-age in the stock (SW_{ay}) to estimate annual biomass,

$$B_y = \sum_a SW_{ay} N_{ay}. \tag{1}$$

Usually, the stock weights reflect fish body weights at the beginning of the year. In addition, using estimates of the proportion-mature-at-age (PM_{ay}), assessment models estimate mature biomass,

$$SSB_y = \sum_a PM_{ay} SW_{ay} N_{ay}. \tag{2}$$

Sometimes stock weights may reflect body weights at the time of spawning if this is substantially different than beginning-of-year body weights. Some assessment models also use estimates of the average weight-at-age of fishery catches (CW_{ay}) to estimate total fishery catch biomass.

Stock status evaluations as part of a precautionary approach usually involve SSB and it is therefore important for stock assessments to estimate SSB as precisely as possible. Measurement errors in SW's will contribute to uncertainty in biomass and SSB estimates, although to some extent these measurement errors cancel in the summations of Equations (1) and (2), which is one reason why estimation of SW's often does not receive detailed review in stock assessment. Nonetheless, errors in estimates of SW's will contribute to uncertainty in estimates of biomass, and even more to uncertainty in estimates of SSB because of the practically fewer age classes that contribute to SSB compared to biomass.

Sampling of commercial catches of 3NO cod has been very limited (Rideout et al. 2021) so here we focus on weights from RV surveys. Survey samples of fish weights and ages will usually provide more accurate estimates of body weight in the stock because the surveys are usually less length-selective compared to fishery catches. If the surveys occur during the first half of the year, then the survey weights are often used to measure beginning-of-year stock weights. If the surveys occur in the fall, then the weights should be adjusted or extrapolated to indicate body weights at the beginning of the year. It is often the case that $SW_{\mathrm{ay}} < CW_{\mathrm{ay}}$ for young ages because fisheries often only catch the larger sizes at young ages. This problem will also occur in surveys but to less of an extent than for commercial fisheries if the surveys use a small mesh sampling gear. However, low samples sizes in surveys for some species and ages is a problem. Some assessment ages may not be sampled at all in some surveys and years.

Cohort growth models have been used to fill in sampling gaps and to reduce measurement errors in SW's (e.g. Cadigan, 2016); however, the parametric growth models (e.g. Von Bertalanffy) do not always fit well and can produce biased estimates of weights for some ages. Cadigan (2020) proposed a more flexible model and used this to model SW's and CW's for cod in NAFO Subdivision 3Ps. In this paper we extend this approach to model SW's for 3NO cod. We use the model outputs to estimate SW's for ages 1 to 10+, where the 10+ SW's are population size-weighted averages for ages 10-22 based on a steady-state age-distribution approximation and total mortality rate $Z = 0.4$.

Methods

Data

Average weights-at-age from both the Spring and Fall DFO bottom-trawl surveys in NAFO Divisions 3NO are shown in Fig. 1. Clearly there is high between-year variability in the estimates and many missing values at older ages. As expected, at younger ages the weights estimated from fall survey samples are greater than those from spring survey samples. However, even for age 13 cod, fall survey weights are slightly greater on average than spring weights, although the difference is probably not statistically significant. Weights increase approximately linearly with age (Fig. 2), and we will model this relationship to improve estimation and fill in missing values.

Model

The model in Cadigan (2020) is

$$\log(SW_{ay}) = \beta_a + \delta_y + \delta_c + \delta_{ay} + \varepsilon_{ay}, \tag{3}$$

where β_a is an age-effect, and δ_y, δ_c, δ_{ay} are random effects for year, cohort, and age x year interactions, respectively. The ε_{ay} are measurement errors. The δ_y are assumed to be multivariate normal (MVN) distributed with mean zero and AR(1) covariance with correlation φ_Y and stationary variance $\sigma_Y^2/(1-\phi_Y^2)$. The Y subscript for σ_Y^2 and φ_Y only indicates a parameter for the year effect and does not indicate a specific year. Similarly, δ_c is MVN with mean zero, AR(1) lag one correlation φ_C and stationary variance $\sigma_C^2/(1-\phi_C^2)$. The δ_{ay} are MVN with a separable covariance matrix Σ with elements

$$Cov(\Sigma_{ay}, \Sigma_{a-i,y-j}) = \frac{\sigma_{AY}^2 \rho_A^{|i|} \rho_Y^{|j|}}{(1-\rho_A^2)(1-\rho_Y^2)}. \tag{4}$$

The ε_{ay} are assumed to have independent normal distributions with mean zero and user-specified variances, $\sigma_{\varepsilon,ay}^2$. Ideally these variances would be derived from the sampling variances obtained during the surveys. However, this information was not available. Similar to Cadigan (2020), we assume that the coefficient of variation in the distribution of length-at-age in the stock is 0.3. This implies that the standard deviation of log-length is approximately 0.3 for each age class in the population. If the slope of log-weight versus log-length is 3 then the standard deviation of log-weight is 0.9 for each age class, which is an approximation we assume. Hence, if there is no age measurement error and if fish are sampled randomly for age and weight from the survey catches then $\sigma_{\varepsilon,ay} \approx 0.9/n_{ay}$where n_{ay} is the number of fish sampled in year y that were age a. These sample sizes were available. This is the approximation we use for $\sigma_{\varepsilon,ay}$.

We modified how the β_a parameters are estimated compared to Cadigan (2020), who freely estimated these parameters. We assume the age effects are monotone increasing so that $\beta_{a+x} > \beta_a$ if $x > 0$. This is what the data suggests (Fig. 2). We also used fractional ages for spring (*a + 5/12*) and fall (*a + 10/12*) surveys. Hence, we estimated a total of 44 age effects for the 22 age classes times 2 surveys. We freely estimated an effect for spring age one, but for other ages we used a monotone regression model:

$$\beta_{a_i} = \begin{cases} \exp(\gamma_i), & i = 1, \\ \beta_{a_{i-1}} + \exp(\gamma_i) & i = 2, \ldots, 44. \end{cases} \tag{5}$$

The γ_i parameters were freely estimated except that we constrained the effects to be equal for all ages greater than 13. Hence, we estimated a total of 25 γ effects.

Plus group weights

We derive plus group weights (e.g. age 10+) using an abundance weighted average based on a steady-state age distribution approximation and a total mortality rate of Z = 0.4,

$$SW_{a+,y} = \frac{\sum_{i=a}^{22} SW_{iy} \exp\{-\mathrm{Z}(\mathrm{i}-\mathrm{a})\}}{\sum_{i=a}^{22} \exp\{-\mathrm{Z}(\mathrm{i}-\mathrm{a})\}}. \tag{6}$$

Results

The age-effects account for much of the variation in the weights (Fig. 3), followed by cohort effects. Year effects and year-age interactions (Fig. 4) are smaller, as indicated by their lower estimates of σ (see Table 1). The cohort effects ranged from -0.49 to 0.25, while the year effects ranged from -0.20 to 0.19 and the interactions ranged from -0.26 to 0.18. The "sawtooth" pattern in the age-effects for ages 1-8 indicate the seasonal pattern of growth for 3NO cod, with less growth between the fall and spring than between the spring and fall. The predicted interactions (Fig. 4) have no strong patterns and the age and year correlation estimates are low (Table 1) for these effects.

Model fits to the data are reasonably good (Figs 5 and 6), especially for year and ages with larger sample sizes (compare Fig. 5 and Fig. 2). There are no patterns in residuals (Figs 7 and 8), although the residual variation is somewhat lower than expected, especially at older ages, which may indicate mis-specification of the variance model. However, we estimated the model assuming a 20% CV in the variation in size at age and this affected the results very little (see below).

Spring predictions will better reflect body weights at the beginning of the year because of the more rapid increase in body weights between the Spring and Fall than between the Fall and Spring of the following year.

Therefore, Spring predictions are better for *SW*'s and SSB calculations. We compare our model predicted *SW*'s with assessment values in Fig. 9. The assessment values were derived from commercial sampling. As expected, the survey model predicted weights are lower at age 3 and to some extent ages 4 and 5. At ages 8 and 9 the model-based values are usually greater than the assessment stock weights. The model-based average weights for the age 10+ group are much greater than the assessment values, in part because the model-based average included ages 10-22 whereas the assessment average was only for ages 10-12.

The measurement error standard deviation approximation, $\sigma_{\varepsilon,ay} \approx 0.9/n_{ay}$, was a subjective choice. Robustness to subjectivity is always a stock assessment concern. We repeated the model estimation using $\sigma_{\varepsilon,ay} = 0.3/n_{ay}$, based on the assumption that the CV for length-at-age was 0.1, which is a low but plausible value. Changes in model predictions were usually small (Fig. 10). The predictions based on CV=0.1 had more between-year variation, especially at ages 7 and older compared to CV=0.3, but this variation was about the same as at ages 5 and 6 which had better sample sizes. We prefer the smoother predictions based on CV=0.3.

Discussion

The purpose of our model was to smooth survey weights-at-age and reduce sampling error in the estimates. We applied the model to average weight-at-age statistics obtained from both the Spring and Fall DFO bottom-trawl surveys in NAFO Divisions 3NO. The model was designed to fit the average weights closely when sampling information was good, but to smooth values when sample sizes were low or zero. We made simple assumptions about the precision of the averages, which was that the CV was proportional to the number of fish measured for each age. The real precision may be different because the average weight-at-age statistics were adjusted to account for the length-stratified sampling designs (e.g. Morgan and Hoenig, 1997; Bettoli, and Miranda, 2001; Echave et al., 2012) and there may be some clustering (i.e. hauls) of samples. More statistically rigorous approaches are possible (e.g. Zheng et al., 2020) but they are not simple and not yet amenable for routine use in stock assessment.

The model accounts for fixed season-age-effects (i.e, years since birth plus fraction of year) and correlated random cohort- and year-effects. The model does not assume a parametric form for weight as a function of age, unlike the Von Bertalanffy model or other alternatives (e.g. Flinn and Midway, 2021). This increases the flexibility of the model and keeps it easy to implement. The model can also be used to interpolate values for some age classes that were not sampled in some years. However, our model cannot be directly used to extrapolate weights for ages outside of the range in the data. For example, we cannot use the model to predict the weight of a 30-year-old fish because there is no fixed-effect for this age. In fact, we cannot use the model to predict the weights for any ages other than *a + 5/12* and *a + 10/12* for a =1,…,22 which we have model fixed-effects for. In particular, the model cannot be used directly to predict beginning-of-year weights for any age *a+0* because there are no beginning-of-year ages in our data. We suggested that weight at age *a+0* be approximated with weight at age *a + 5/12*, but some other extrapolation methods could be used. The model could be generalized so that age effects are a flexible continuous parametric function of age (i.e., a monotone increasing spline smoother) or a correlated stochastic process to extrapolate beyond the values of ages in the model. This is useful future research.

We used the model to approximate plus group (ages 10+) weights-at-age by averaging model predicted weights at ages 10-22. The average was abundance-weighted, assuming a steady-state age-distribution based on a total mortality rate of Z = 0.4. Other choices are possible. A lower value for Z will produce a greater age 10+ average weight and a higher value for Z will produce a lower average weight. Change in the mortality rates of age 10+ cod and variation in cohort sizes at these ages will also affect the age 10+ average weight. These are all issue that can be addressed in our modelling approach if better information is available on the distribution of 3NO cod abundance at ages 10+. Ideally the 3NO cod assessment model would run for a plausible large range of ages and model results would be aggregated to match the levels of data aggregation. In this case there would be no need to aggregated weights in a plus group. This is also a useful area for future research.

We demonstrated that age was the major source of variability in the weight-at-age data, which is not surprising. However, we also found some statistically significant cohort and year effects. We did not explore what might be the drivers of these cohort and year effects; that was not our purpose. However, it is useful to have a better understanding of this. Of particular concern for stock assessment is the potential that the cohort and year effects are only adjusting for some other sampling artifact, such as a change in the age and growth sampling design, and that these changes do not reflect changes in the weight-at-age in the stock. This is a useful area for future research.

References

Bettoli, P.W. and Miranda, L.E., 2001. Cautionary note about estimating mean length at age with subsampled data. North American Journal of Fisheries Management, 21(2), pp.425-428.

Cadigan, N. 2016. Weight-at-age growth models and forecasts for Northern cod (*Gadus morhua*). DFO Can. Sci. Advis. Sec. Res. Doc. 2016/016. v + 19 p.

Cadigan, N. 2020. A Simple Random-Effects Model to Smooth and Extrapolate Weights-at-Age for 3Ps Cod. DFO Can. Sci. Advis. Sec. Res. Doc. 2020/xxx. v + 25p. In press.

Echave, K.B., Hanselman, D.H., Adkison, M.D., and Sigler, M.F. 2012. Interdecadal change in growth of sablefish (*Anoplopoma fimbria*) in the Northeast Pacific ocean. Fish. Bull. 110(3): 361–374.

Flinn, S.A. and Midway, S.R., 2021. Trends in growth modeling in fisheries science. Fishes, 6(1), p.1.

Morgan, M.J. and Hoenig, J.M., 1997. Estimating maturity-at-age from length stratified sampling. Journal of Northwest Atlantic Fishery Science, 21.

Rideout, R.M., Rogers, R. and Ings, D.W. 2021. An updated assessment of the cod stock in NAFO Divisions 3NO. NAFO SCR Doc. 21/031.

Zheng, N., Cadigan, N. and Morgan, M.J., 2020. A spatiotemporal Richards–Schnute growth model and its estimation when data are collected through length-stratified sampling. Environmental and Ecological Statistics, 27(3), pp.415-446.

Tables

Table 1. Parameter estimates (Est), standard errors (SD), and negative loglikelihood (nll) gradients (GRD). nll = 42.02, number of parameters = 32, AIC = 148.03, BIC = 302.65.

Effect	Est	SD	GRD
$\gamma_{1.4}$	-3.961	0.086	0.000033
$\gamma_{1.8}$	0.179	0.042	0.000060
$\gamma_{2.4}$	-0.965	0.114	0.000059
$\gamma_{2.8}$	-0.160	0.047	-0.000027
$\gamma_{3.4}$	-1.640	0.200	-0.000017
$\gamma_{3.8}$	-0.413	0.058	-0.000046
$\gamma_{4.4}$	-2.288	0.395	-0.000017
$\gamma_{4.8}$	-0.611	0.076	-0.000118
$\gamma_{5.4}$	-2.695	0.636	-0.000014
$\gamma_{5.8}$	-0.890	0.108	-0.000019
$\gamma_{6.4}$	-2.744	0.732	-0.000001
$\gamma_{6.8}$	-0.921	0.125	-0.000057
$\gamma_{7.4}$	-3.989	2.800	-0.000004
$\gamma_{7.8}$	-1.089	0.173	-0.000045
$\gamma_{8.4}$	-2.282	0.606	-0.000017
$\gamma_{8.8}$	-1.678	0.361	-0.000013
$\gamma_{9.4}$	-1.806	0.419	-0.000010
$\gamma_{9.8}$	-1.878	0.490	-0.000005
$\gamma_{10.4}$	-2.030	0.591	-0.000001
$\gamma_{10.8}$	-1.728	0.439	-0.000016
$\gamma_{11.4}$	-2.785	1.306	0.000000
$\gamma_{11.8}$	-2.213	0.827	-0.000012
$\gamma_{12.4}$	-2.635	1.326	-0.000011
$\gamma_{12.8}$	-2.032	0.542	-0.000016
$\gamma_{13.4+}$	-3.592	0.293	-0.000037
$\log(\sigma_Y)$	-2.192	0.343	0.000015
$\log(\sigma_C)$	-1.841	0.202	-0.000005
$\log(\sigma_{AY})$	-2.476	0.095	0.000035
$logit(\varphi_Y)$	1.622	0.921	-0.000003
$logit(\varphi_C)$	0.318	0.710	-0.000002
$logit(\rho_A)$	-10.000	44.774	0.000499
$logit(\rho_Y)$	-1.038	0.762	0.000004

Figures

Figure 1. Time-series of average weights-at-age from Spring (green) and Fall (red) DFO bottom-trawl surveys in NAFO Divisions 3NO. Each panel is for an age class. Horizontal dashed lines indicate the series. The circle areas indicate the number of fish sampled for age and weight.

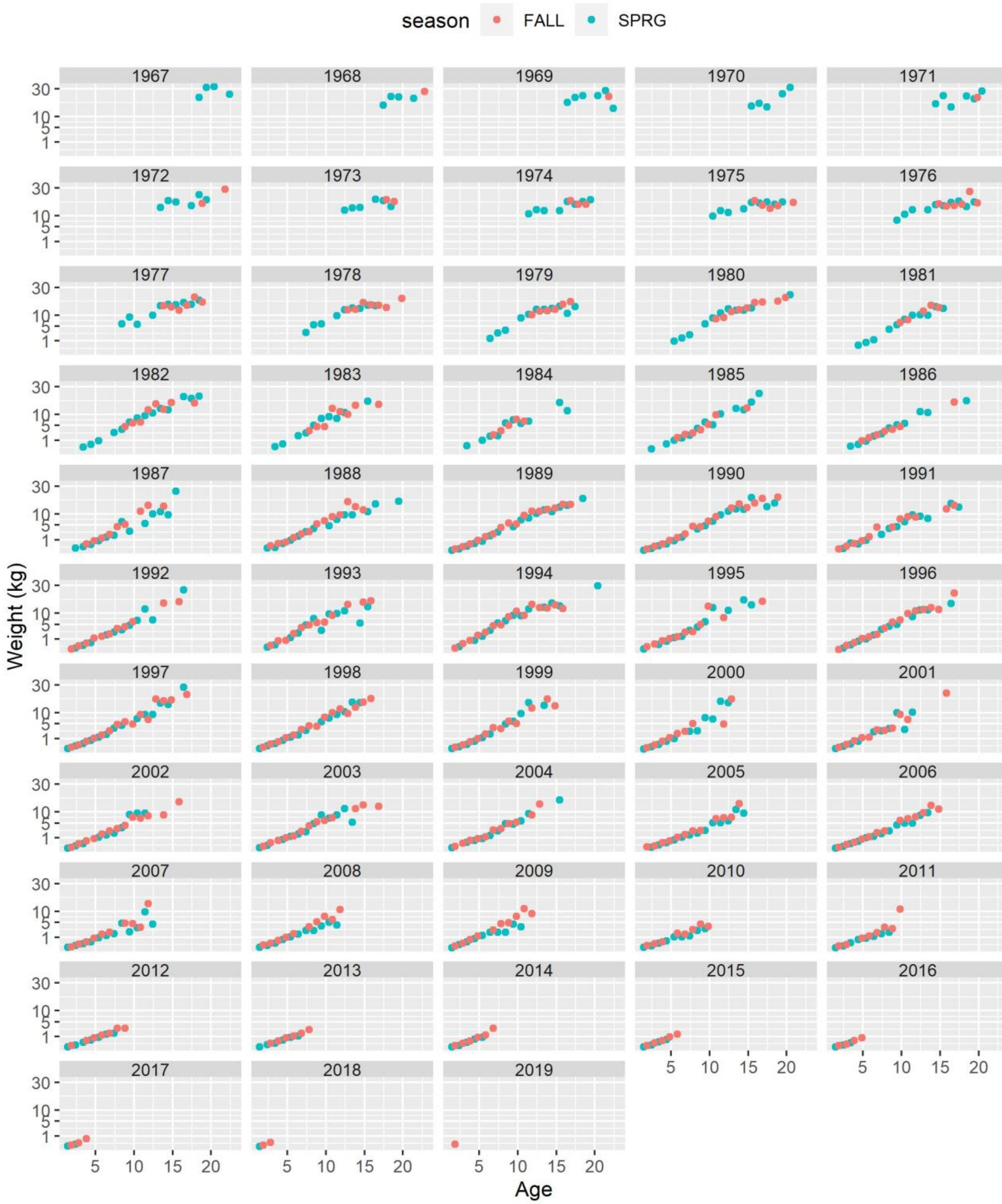

Figure 2. Average weights-at-age from Spring (green) and Fall (red) DFO bottom-trawl surveys in NAFO Divisions 3NO. Each panel is for a cohort.

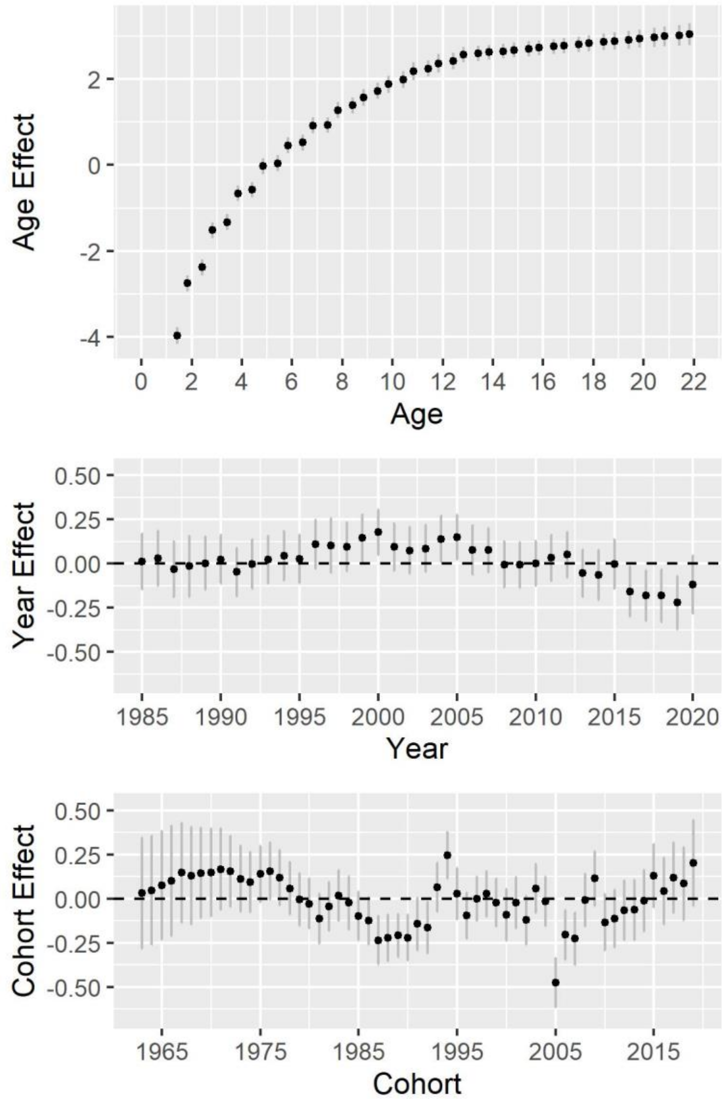

Figure 3. Estimates of the main effects in the weight-at-age model. Vertical line segments indicate 95% confidence intervals.

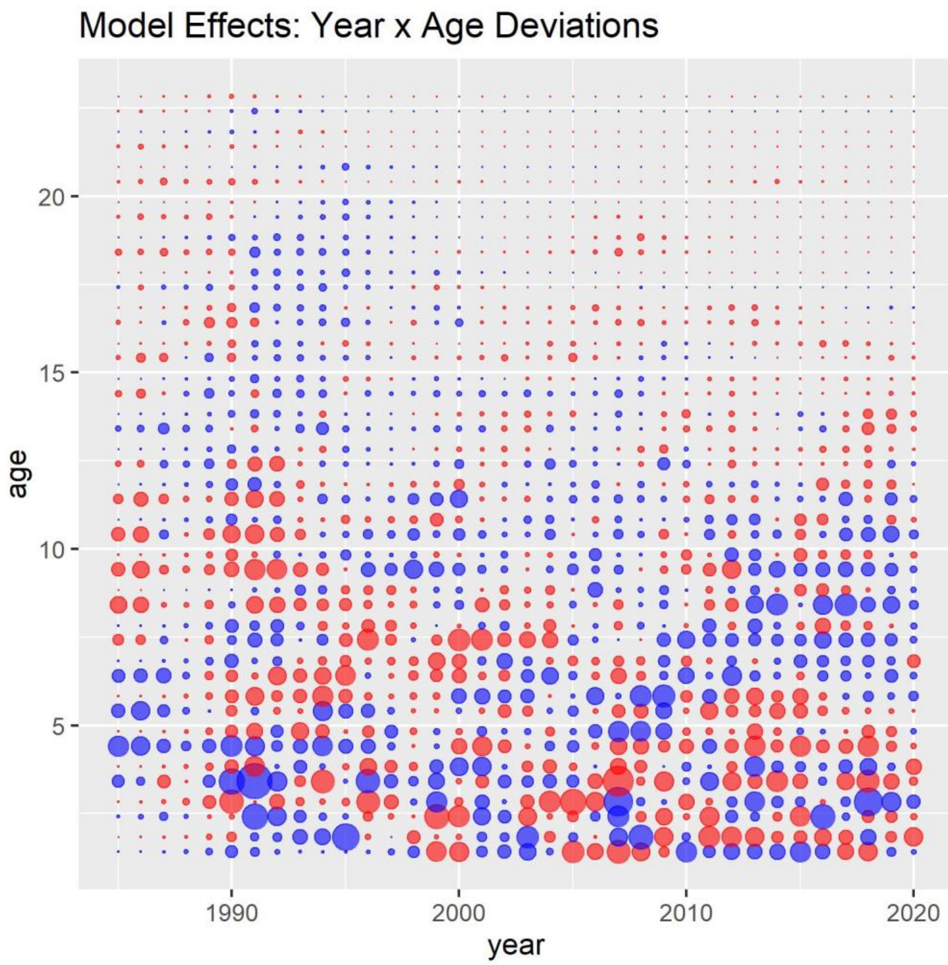

Figure 4. Estimates of the age x year interactions effects. The area of the circles indicates the absolute value of the effect, and the color indicates the sign (red +; blue -).

Figure 5. Time-series of observed (points) and model-predicted (lines) average weights-at-age from Spring (green) and Fall (red) DFO bottom-trawl surveys in NAFO Divisions 3NO. Each panel is for an age class. Shaded regions indicate 95% confidence intervals.

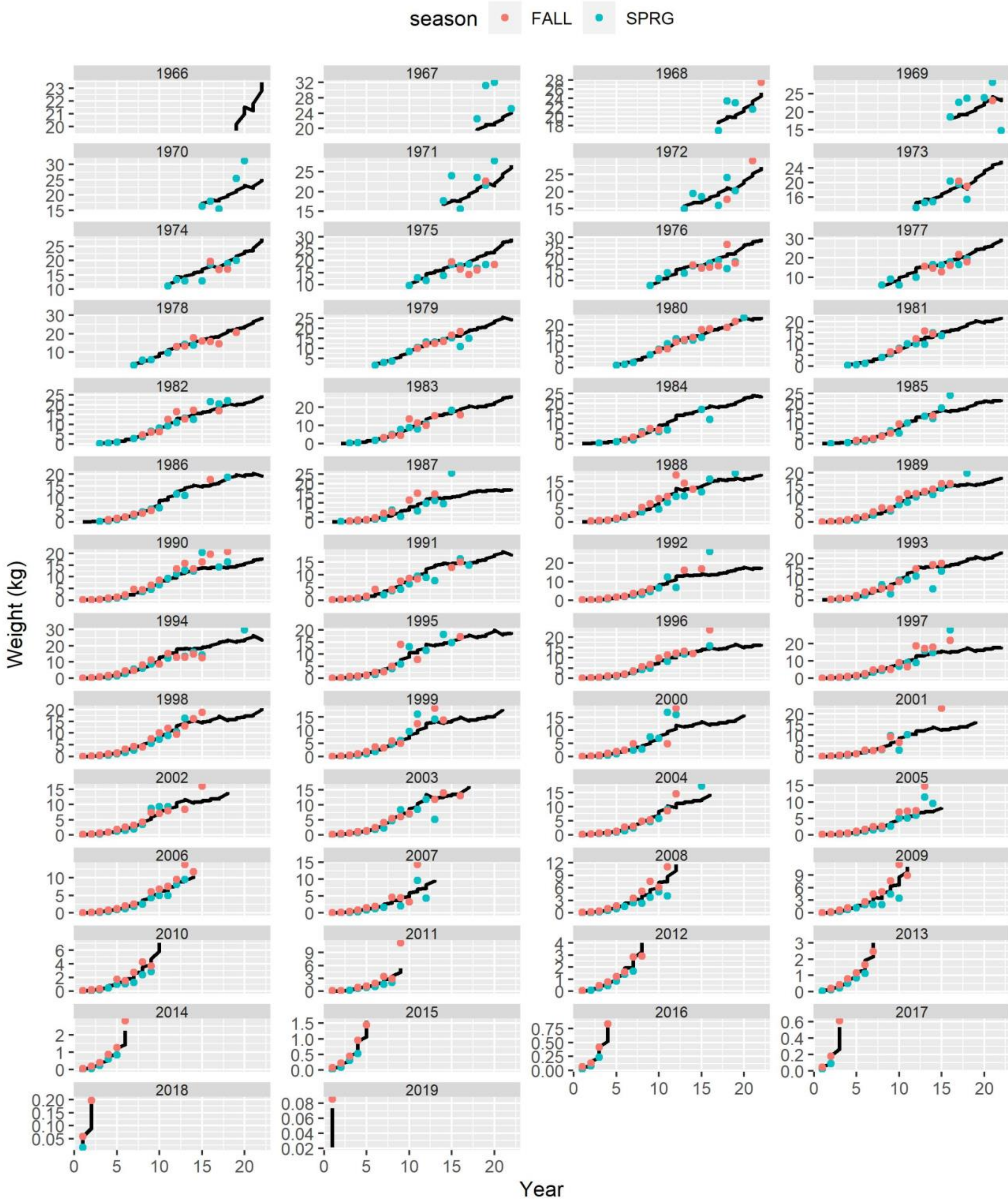

Figure 6. Observed (points) and model-predicted (lines) average weights-at-age from Spring (green) and Fall (red) DFO bottom-trawl surveys in NAFO Divisions 3NO. Each panel is for a cohort.

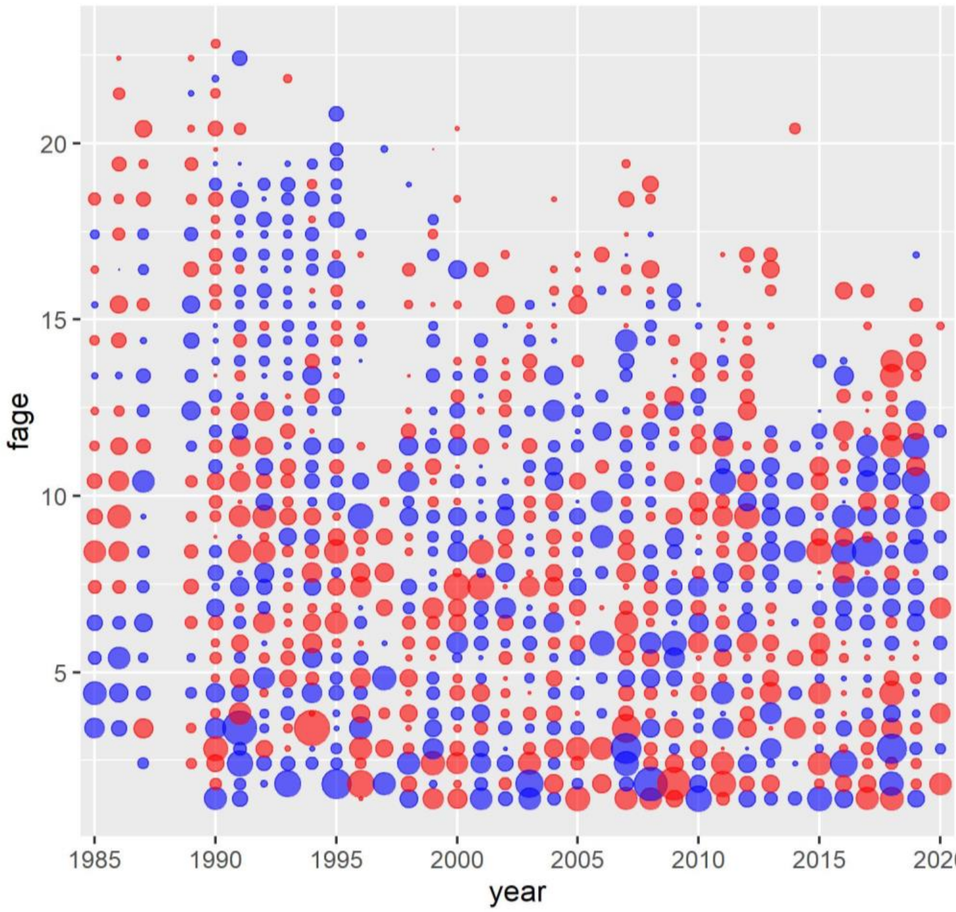

Figure 7. Model standardized residuals. The area of the circles indicates the absolute value of the residual and the color indicates the sign (red +; blue -).

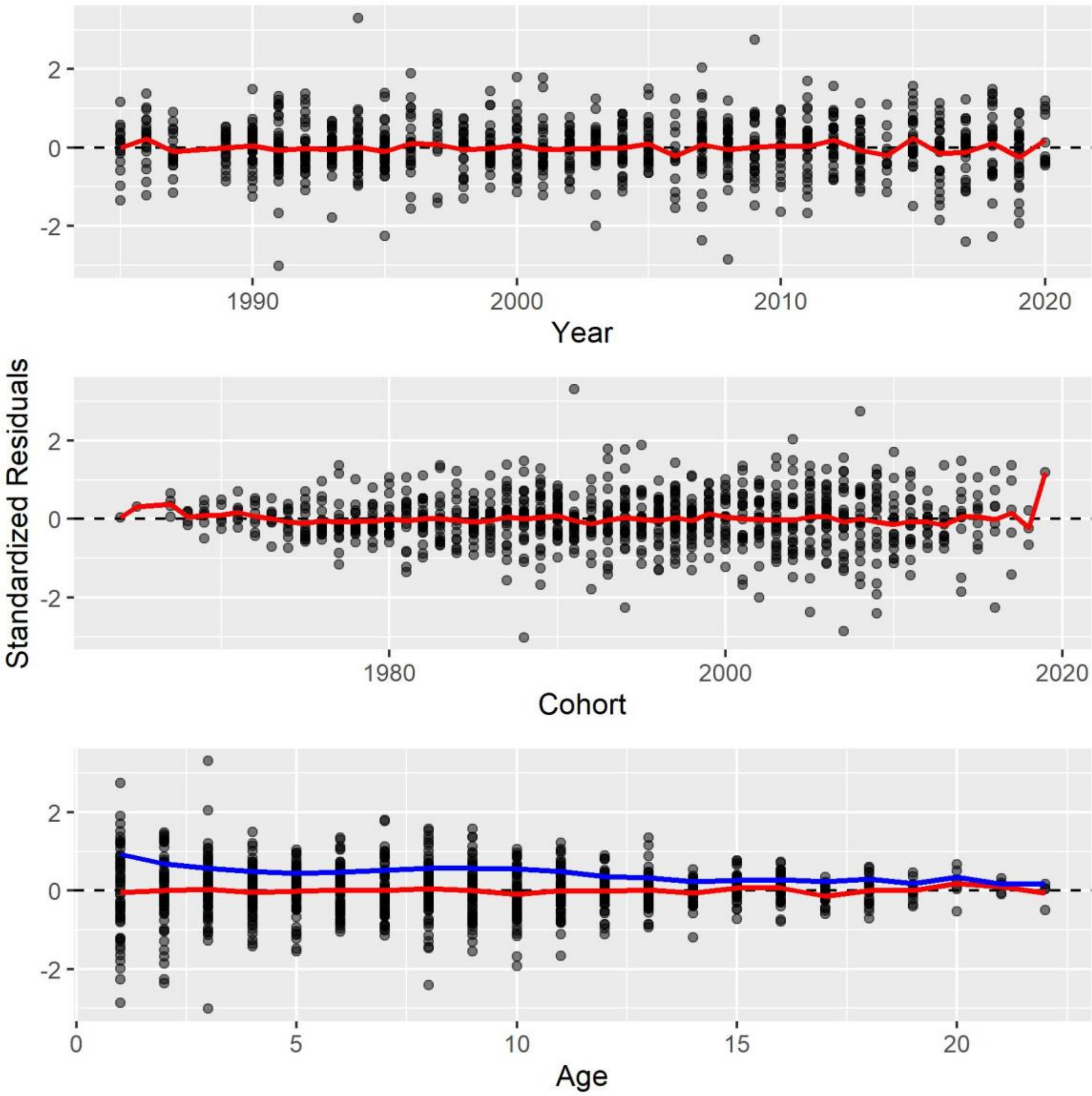

Figure 8. Standardized residuals versus year (top), cohort (middle), and age (bottom). Red lines indicate the average residual, and the blue line indicates the average absolute residual.

Figure 9. Comparisons of stock weights-at-age used in the 2020 assessment of 3NO cod (red lines) and the smoothing model estimates (green lines). Note that the age 10 weights are 10+ average weights for ages 10-22 for the smoothing model. They are averaged for ages 10-12 from the assessment.

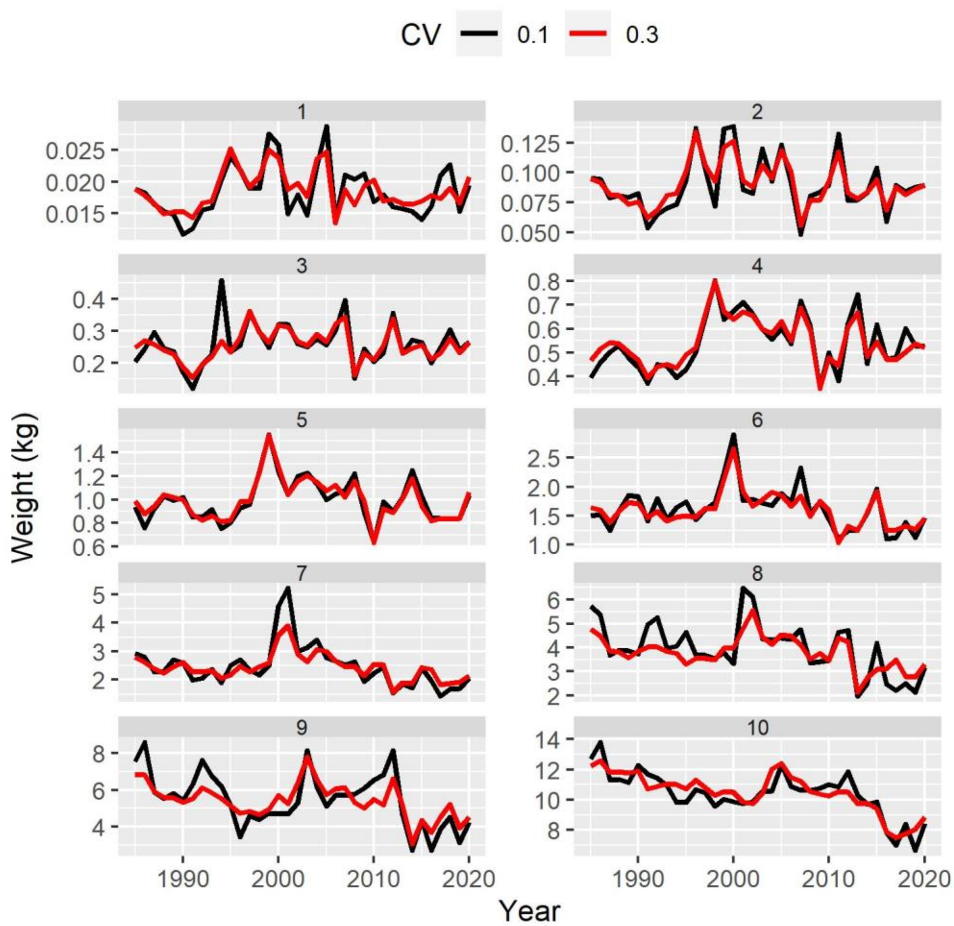

Figure 10. Comparison of model predictions for two choices of the length-at-age CV (0.1 or 0.3).