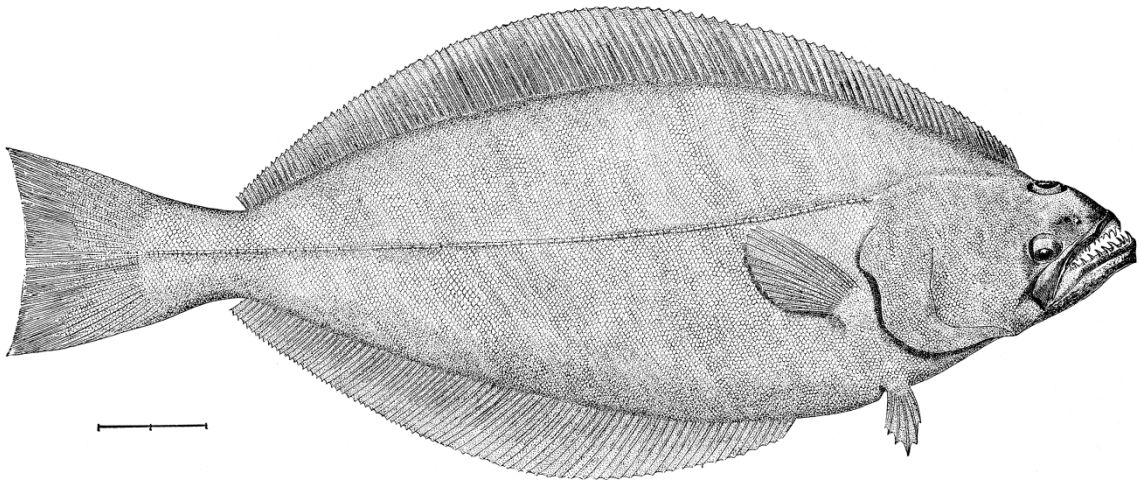
Model developments and changes to the 2012 BSAI Greenland turbot (*Reinhardtius hippoglossoides*) stock assessment



Steven J. Barbeaux and James Ianelli
Alaska Fisheries Science Center

Introduction

This year the BSAI Greenland turbot (*Reinhardtius hippoglossoides*) stock assessment will be lead by Dr. Steven Barbeaux. Although the stock will continue to be modeled using the same software as previous assessments (Stock Synthesis 3), there are a number of changes within the model. This paper is meant to guide you through changes in the Greenland turbot assessment dataset and model and identify the effects of these changes on model fit and results and relies heavily on the reader's understanding of the 2011 Greenland turbot stock assessment (Ianelli et al. 2011). **This paper is not meant as a final stock assessment and all results are preliminary and will change prior to the November plan team meeting.** The changes to the data and the assessment model are an attempt to better capture the complex population dynamics of this species due its unique life history and distribution across two geopolitical boundaries (the US-Russian EEZ and the Northern extent of the AFSC surveys).

Change in weight at length relationship

The 2011 Greenland turbot stock assessment model used the same weight at length relationship for males and females ($w = 2.44 \times 10^{-6} L^{-3.34694}$, where L = length in cm, and w = weight in kilograms). Given the high degree of sexual dimorphism observed in this species it was thought that having separate weight at length relationships for males and females would better capture the diversity in this stock. A new weight at length relationship has been developed using the combined weight and length data from all bottom trawl surveys conducted by the Alaska Fisheries Science Center in the Bering Sea and Aleutian Islands from 1983 to 2011. There were a total of 2,861 animals measured, 1,380 females, 1,383 males and 98 unidentified used in the analysis. A linear model was fit to the log transform of length and weight as $\log(L) = \alpha S_1(\log(w))+\beta S_2$, where L is the fork length in centimeters, w is the weight in kilograms α is the slope for females and β is the intercept for females, S_1 is the effect of sex on the slope and S_2 is the effect

This information is distributed solely for the purpose of pre-dissemination peer review under applicable information quality guidelines. It has not been formally disseminated by the National Marine Fisheries Service and should not be construed to represent any agency determination or policy.

of sex on the intercept. Results of the linear model are shown in Table 1. The model fit was highly significant with an R^2 of 0.997 and a P-value of less than 2×10^{-16}.

Table 1 Results from linear model on the weight to length relationship for Greenland Turbot.

Coefficient	Estimate	Std. Error	T-value	Pr(>|t|)
αFemale	-12.592	0.023	-549.804	< 2x10-16
βFemale	3.219	0.006	536.719	< 2x10-16
S1Male	-0.334	0.032	-10.303	< 2x10-16
S1Unident.	0.579	0.173	3.349	0.000823
S2Male	0.106	0.008	12.866	< 2x10-16
S2Unident.	-0.253	0.071	-3.582	0.000347
	$R^2 = 0.997$		P-value = $< 2 \times 10^{-16}$	

This year's model proposes to use $w = 2.43 \times 10^{-6} L^{3.325}$ for females and $w = 3.40 \times 10^{-6} L^{3.2189}$ for males (Fig 1), the conversion of the analysis results out of log space. This relationship is similar to the weight at length relationship observed by Ianelli et al. (1993) and used in the Greenland turbot stock assessment prior to 2002 (Fig. 2).

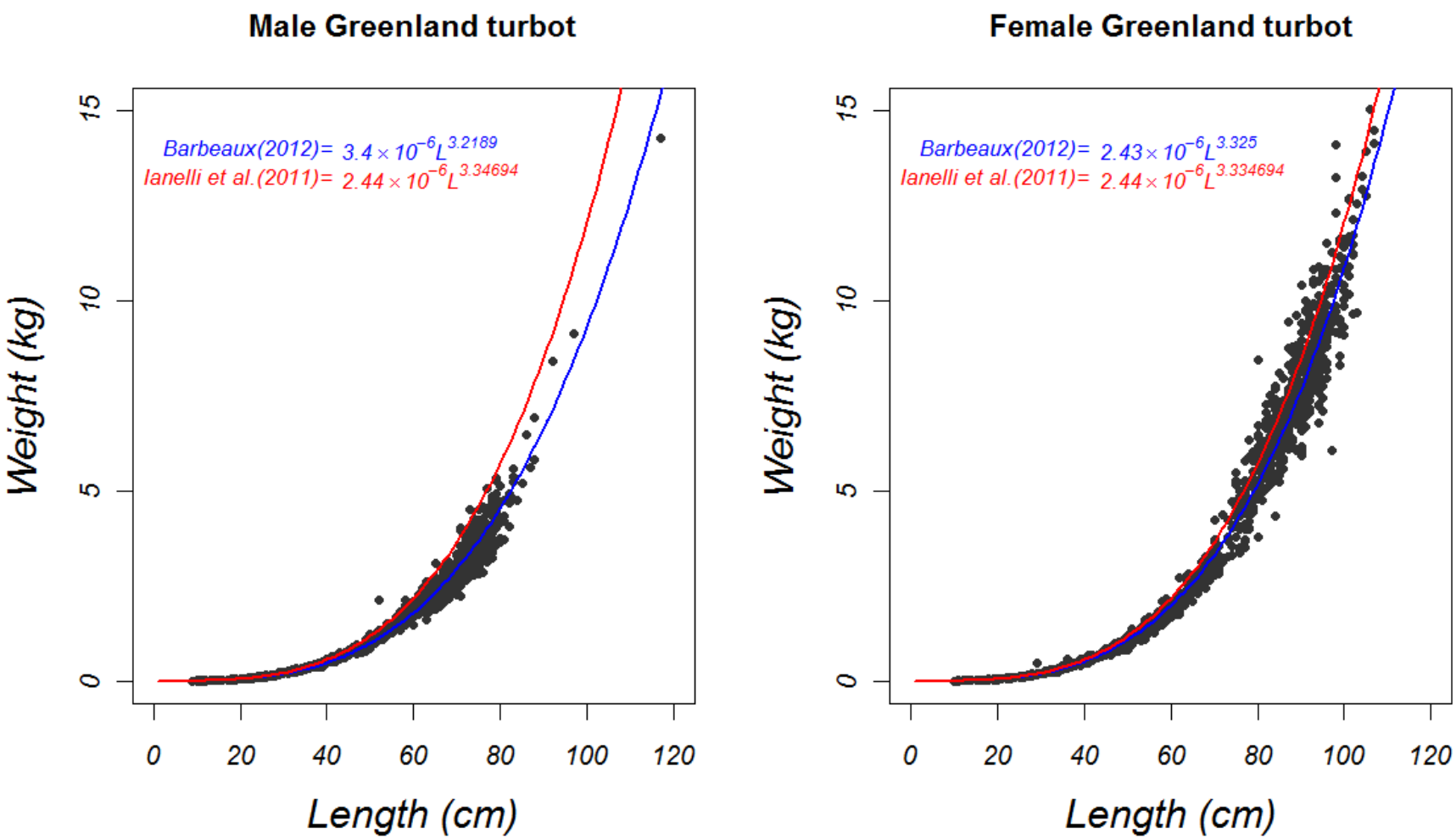


Figure 1 Comparison of weight at length relationship used in 2011 Reference model and proposed relationship. Data are compiled from all Bering Sea and Aleutian Islands trawl surveys 1983- 2011.

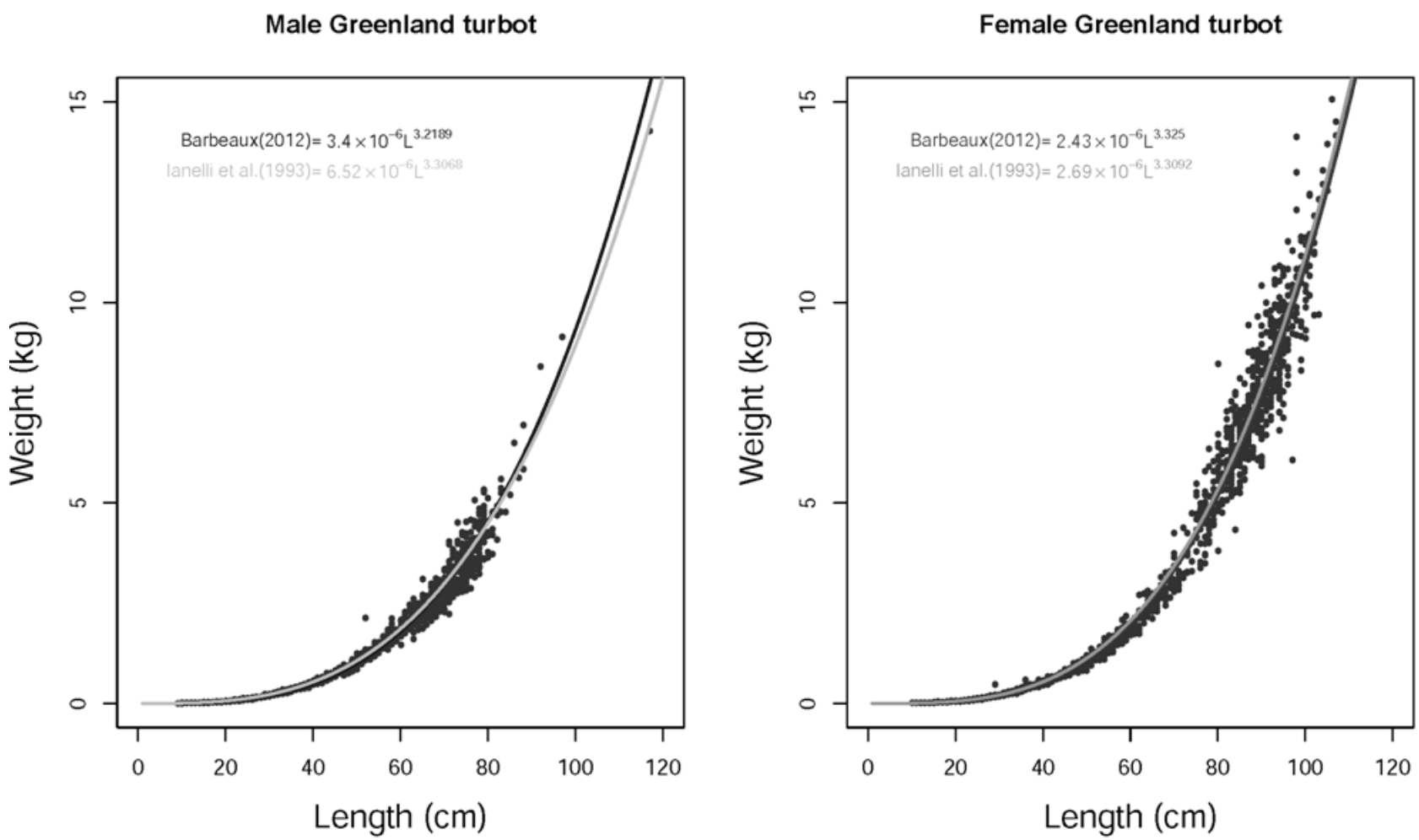


Figure 2 Comparison of weight at length relationship from Ianelli et al. (1993) and proposed relationship derived from fit to all Bering Sea and Aleutian Islands trawl survey data 1983-2010.

Effects of weight length relationship change

The 2011 reference model was run with the improved weight at length relationship to identify effects to assessment results. The model fit was improved in the new configuration with a lower negative log likelihood overall (LL; Table 2). Most of the improvement was in the fit to size at age (-514 LL) and length composition data (-43 LL). There was slight decrease in goodness of fit to the survey abundance index data (+3.5 LL). The change in weight at length resulted in lighter Greenland turbot at age than in the 2011 stock assessment (Fig. 3). As expected the change also resulted in smaller total and spawning stock biomass estimates (Fig. 4). Further, the change in weight at length resulted in differences in estimated recruitment, particularly for the 1978 year class and early recruitment from 1960 to 1970 (Fig. 4).

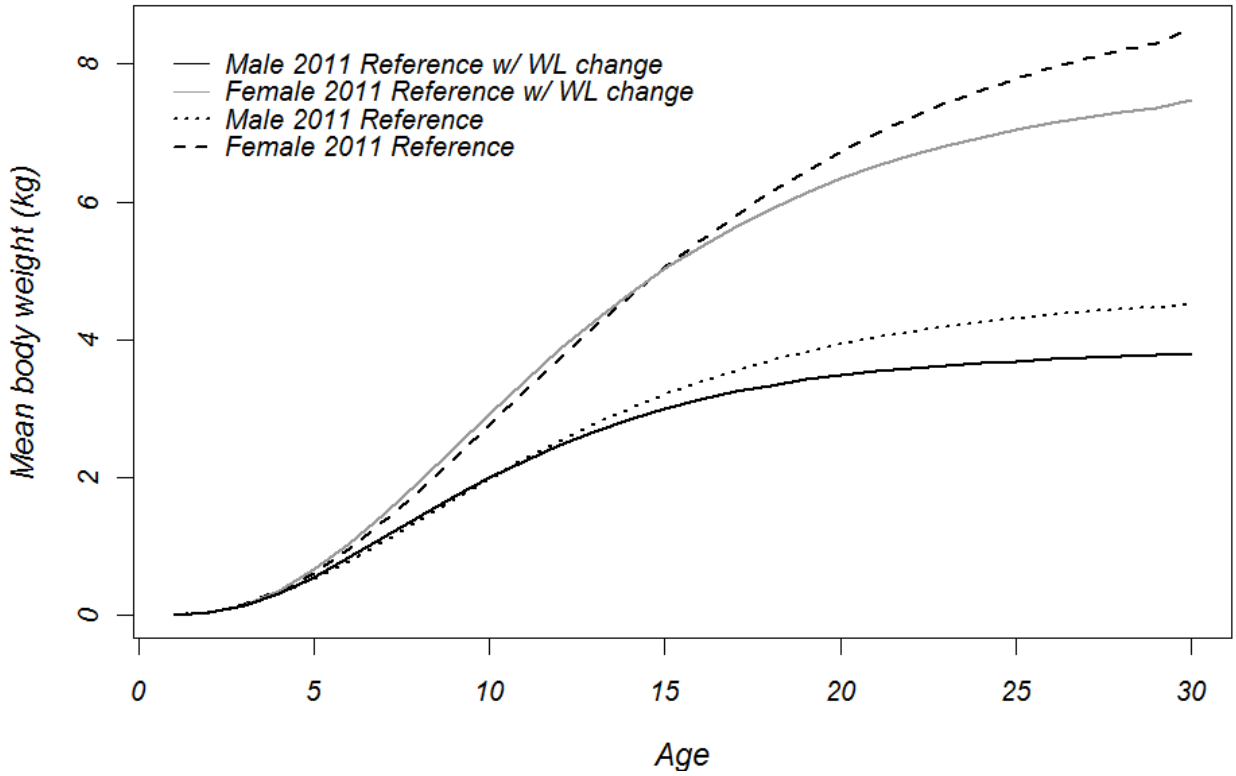


Figure 3 Difference in weight at age due to changing to the improved weight at length relationship.

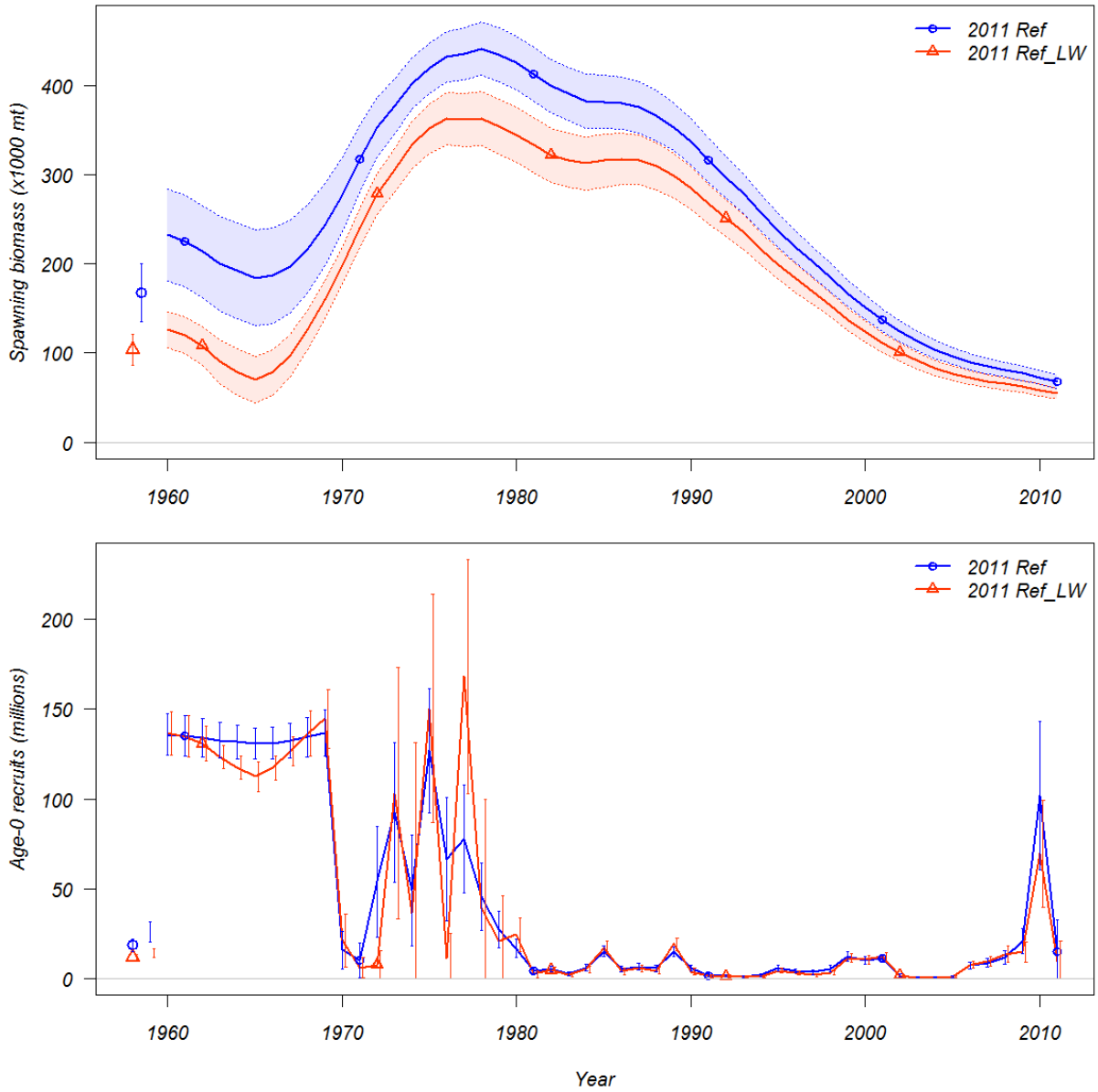


Figure 4 Change in spawning biomass estimates and age-0 recruitment from the 2011 Reference model (2011 Ref) and the 2011 Reference model with the improved weight at length relationship (2011 Ref_LW).

Table 2 Fit to the 2011 Reference model and 2011 Reference model with changes to the improved weight at length relationship (Ref_LW) , with changes to indices (Ref_Ind), changes to size and age composition data (Ref_Comp), and all three sets of data changes combined in the reference model (Ref_All). Shaded values are those where the underlying data have been changed and likelihoods should not be compared with the reference model.

		2011 Ref	Ref_LW	Ref_Ind	Ref_Comp	Ref_All
Likelihoods						
	Total	3879.21	3309.35	3834.43	4321.07	3570.63
	Survey	-30.19	-26.64	-34.20	-30.55	-31.33
	Length Composition	2219.61	2176.56	2179.42	2287.26	2322.36
	Age Composition	215.64	210.01	215.32	238.34	236.8
	Parameter priors	3.44	5.06	3.53	4.93	6.96
	Size at Age	1347.63	833.18	1347.42	1719.87	940.56
	Recruitment	123.09	111.18	122.94	101.23	95.28
Key Parameters						
	LN(R0)	9.83	9.38	9.82	9.58	9.20
	Q for Shelf Survey	0.73	1.00	0.73	0.51	0.64
	L at Amax Fem	87.31	87.20	87.31	87.31	87.83
	L at Amax Mal	72.93	74.31	72.93	72.93	72.63
	VonBert K Fem	0.12	0.14	0.12	0.12	0.13
	VonBert K Mal	0.10	0.10	0.10	0.10	0.18
Recruits 1960-1969						
Age-0 (1×106)	Mean	133.50	128.06	132.55	158.66	147.19
	Median	133.39	128.68	132.46	158.77	149.21
	CV	0.02	0.08	0.02	0.06	0.19
Spawning stock biomass						
(1,000 t)	1978	441.27	358.99	434.27	406.42	327.83
	2011	67.58	57.21	67.85	55.58	46.12

Changes in the Data

To ensure that all aspects of the assessment could be replicated and that all corrections made to the data in the database were carried forward in the model, all data used in the model (Fig. 5) were queried anew from their respective databases (see Ianelli et al. 2011). In addition, there are some improvements to how the size composition data are processed prior to inclusion in the proposed 2012 model configuration. All data queries and data processing were conducted in R. The differences in the 2012 Candidate dataset are discussed below and effects to the 2011 Reference model results due to each change are highlighted separately.

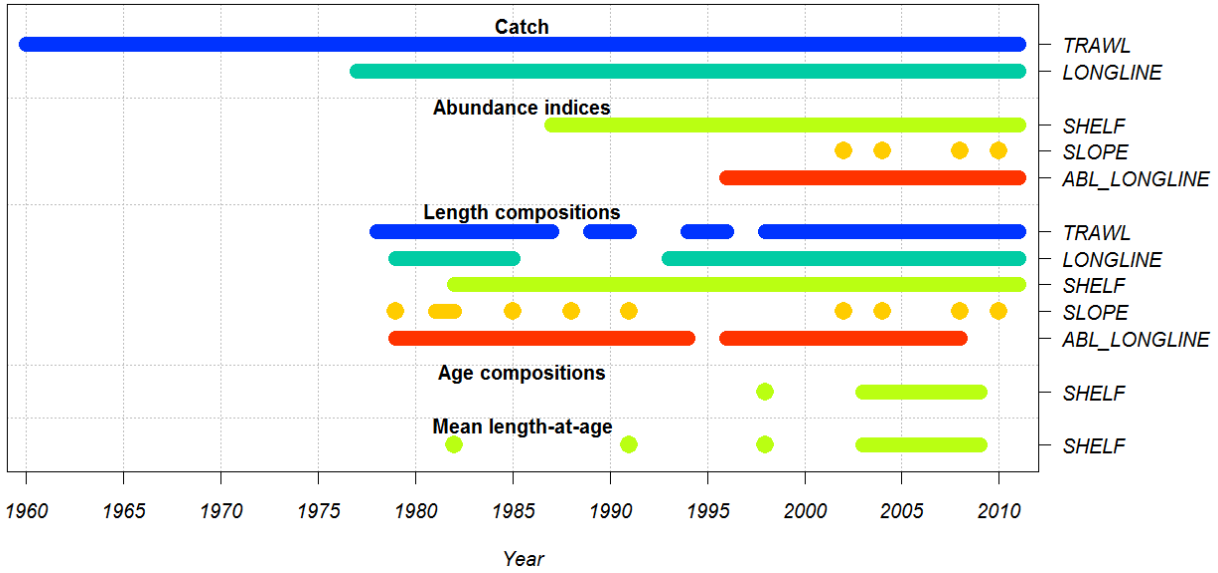


Figure 5 Data sources for the 2012 Greenland turbot stock assessment by type and year. From top to bottom, trawl fishery (Trawl) catch, longline fishery (Longline) catch, Bering sea Shelf trawl survey (Shelf) index, Bering Sea slope trawl survey (Slope) index, Auke Bay longline survey (ABL Longline) index, Trawl length composition data, Longline length composition data, Shelf length composition data, Slope length composition data, Shelf age composition data, and Shelf mean length at age data.

Catch and survey index estimates

Catch estimates were queried directly from the North Pacific Catch Accounting System (CAS) database maintained at the Alaska Fisheries Science Center by Terry Hyatt. There were differences in the most recent (post-2003) catch estimates from the 2011 Reference dataset (Fig. 6). The largest difference was in the 2010 Longline catch estimate, with an increase of 185 tons. The authors do not know why the estimates differ.

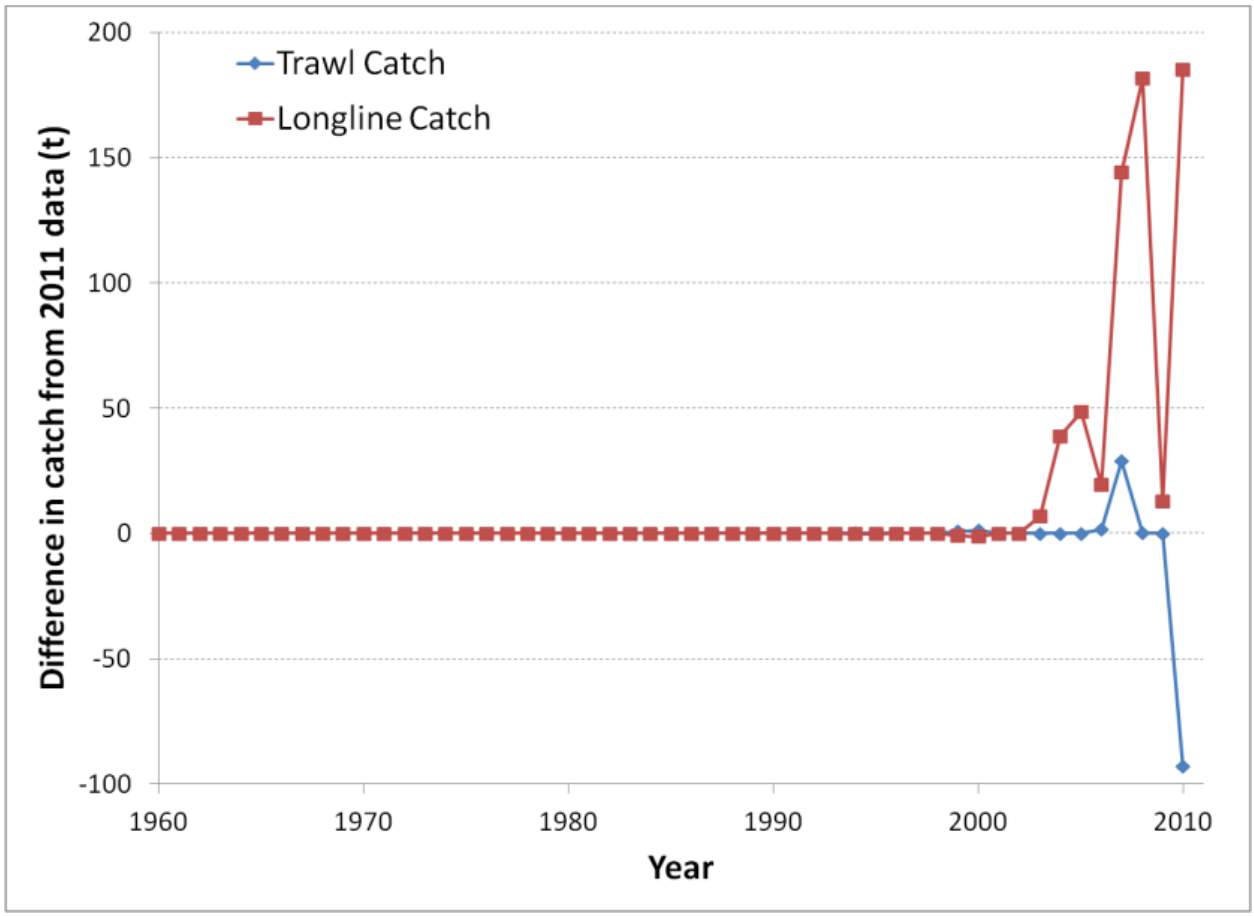


Figure 6 Difference between the 2012 candidate dataset and 2011 Reference dataset for Trawl and Longline catch estimates in tons.

The most substantial change to the assessment survey index dataset was the removal of the six Slope survey abundance index values for surveys conducted prior to 2002. These data were removed after discussions with Dr. Jerry Hoff, the current Slope survey Chief Scientist. Dr. Hoff stated that the older Slope survey data are not comparable to the most recent surveys, and may have not been conducted consistently enough in the early years to be considered a time series. The surveys differed in vessel power, in gear used, and in the ability of the surveyors to determine whether the gear was in contact with the bottom. There were only minor differences (±3%) in index values from the 2011 Reference dataset for the other two survey index values used in the 2012 Candidate dataset.

Effects of changes to indices of abundance and catch estimates on the assessment

Removal of the early Bering Sea slope survey index values and differences in the catch estimates and other survey index values made little impact on the fit or results of the 2011 Reference model (Table 1 and Fig. 7). The slight improvement (-4 LL) to the survey index likelihood can be attributed to fewer data points in the model.

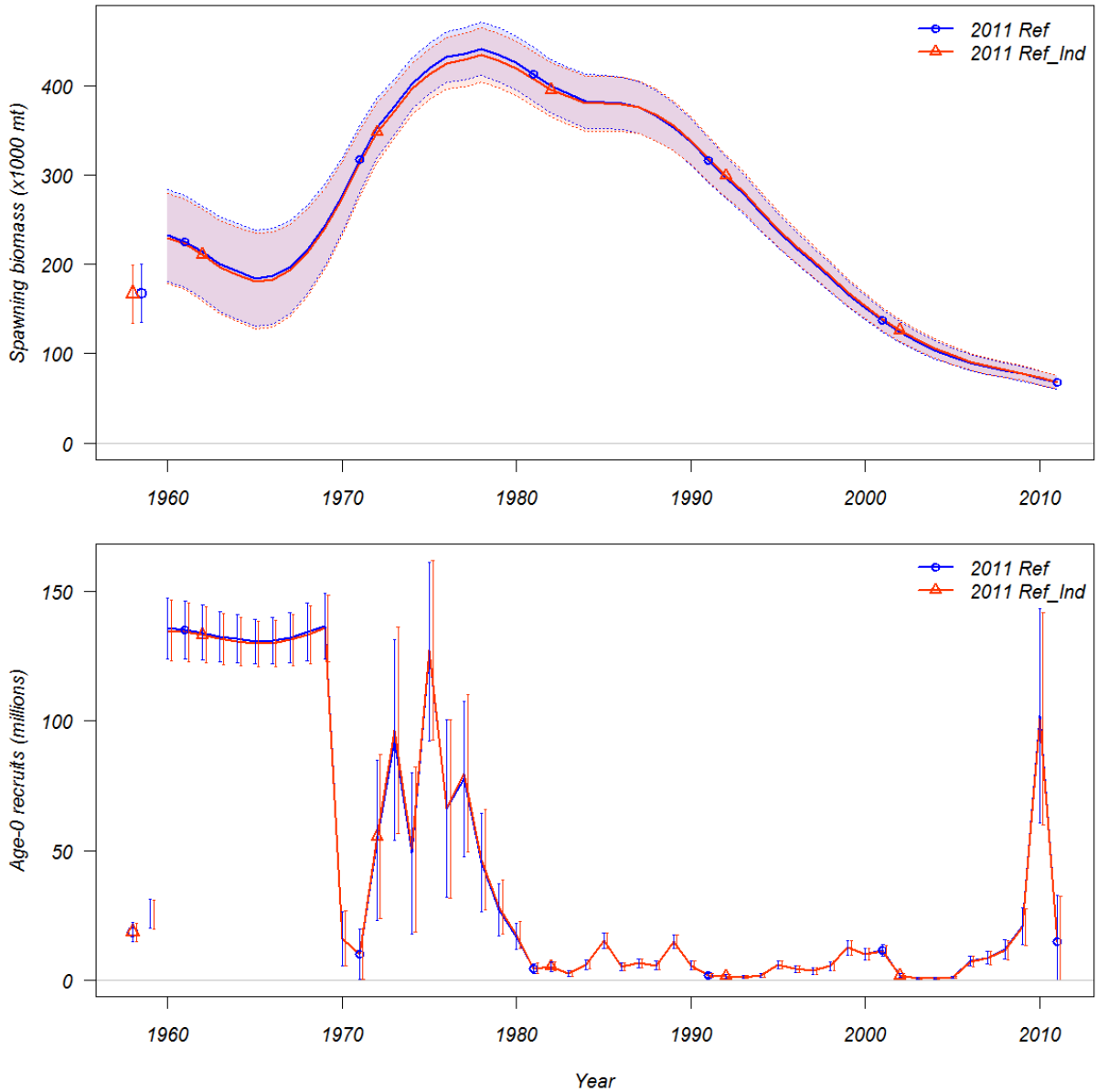


Figure 7 Spawning biomass and age-0 recruitment for 2011 reference model (2011 Ref) and 2011 Reference model with new index and catch estimates and removal of early Bering Sea slope data (2011 Ref_Ind).

Changes in length and age composition data

Fishery length composition data were treated differently this year than in previous years. The raw Trawl and Longline length composition data were proportioned to catch numbers by haul to obtain a more accurate representation of the catch composition. The proportion (P) of fish for a particular length bin (*l*) and year (*y*) was calculated as $P_{yl} = \frac{\sum\left(\frac{n_{ylh}}{n_{yh}} \times N_{yh}\right)}{\sum N_{yh}}$, where *n* is the number of fish in a length bin (*l*) for an individual year (*y*) and haul sample (*h*) and *N* is the total number of fish in a haul (*h*) for year (*y*) for each

fleet. This assumes that the length composition samples were representative of the length composition of each sampled haul and that observer haul samples were representative of overall effort in the fleet. Previous assessments assumed that the summed raw fishery length composition samples were representative of the fleet-wide catch length composition. The largest change was in the early trawl female length composition data. In the new dataset the larger females compose a much smaller proportion of the trawl catch for 1977 through 1980 (Fig. 8A). Although there were a large number of length samples taken from the directed trawl fishery in deeper waters where these large female fish were caught, the total number of Greenland turbot observed caught in deeper waters was small in comparison with the observed number of Greenland turbot caught in shallower waters. That is, although fewer samples were taken in the shallower waters, the shallower trawl hauls contained a larger number of smaller fish. The sexed length composition data from 1990 to 2004 in both the trawl and longline fisheries were combined as unsexed in the 2011 Reference dataset, but remained differentiated in the new dataset (Fig. 8a, Fig. 8b, and Fig. 8c). The catch at age composition data were proportioned to catch in the same manner as the length composition data, but there was little difference between the 2011 Reference dataset and the 2012 Candidate dataset (Fig. 9).

2011 Reference dataset 2012 Candidate dataset

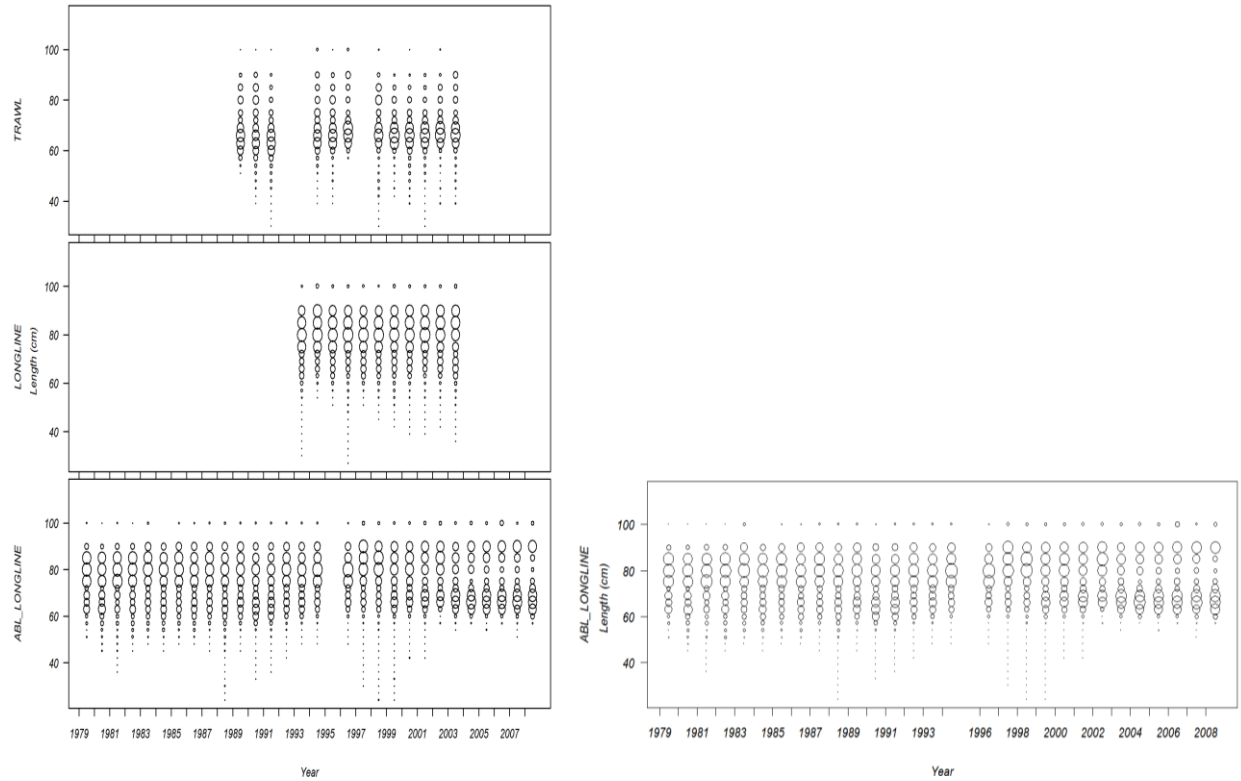


Figure 8a. Length composition data from the 2011 Reference dataset (left) and the 2012 Candidate dataset (right) for sexes combined.

2011 Reference dataset 2012 Candidate dataset

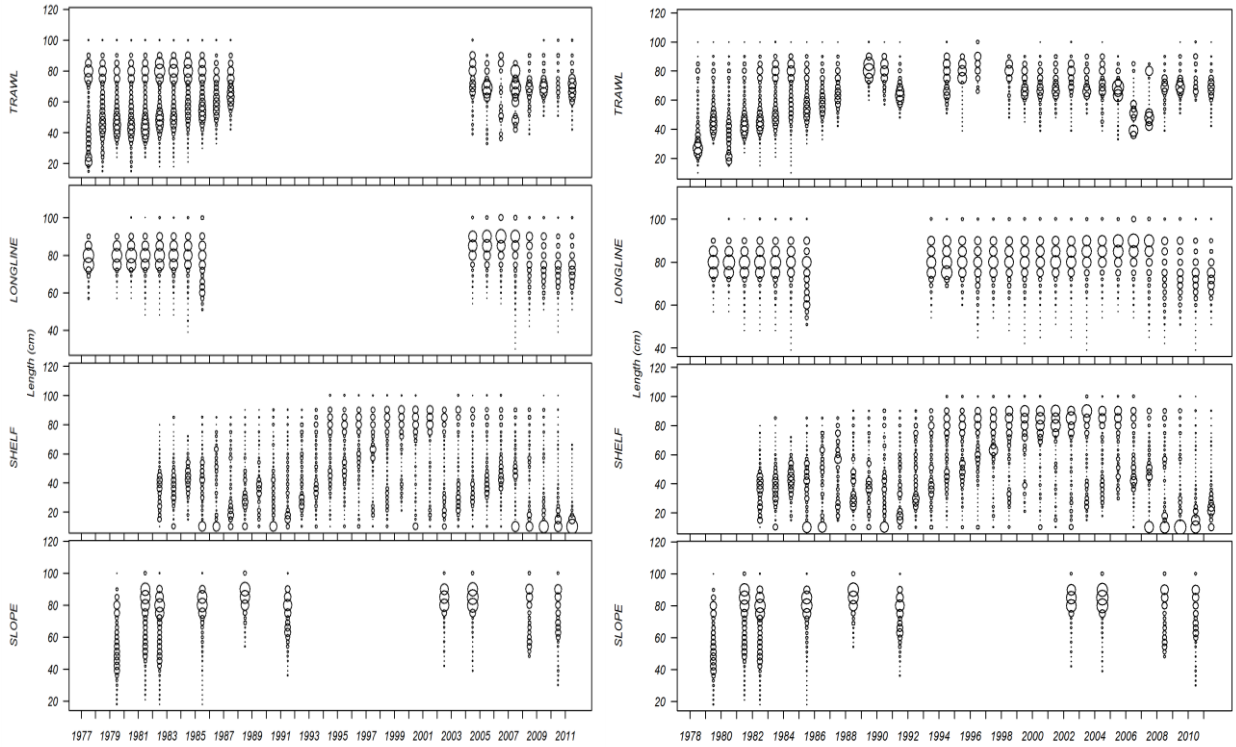


Figure 8b Length composition data from the 2011 Reference dataset and the 2012 Candidate dataset for females.

2011 Reference dataset 2012 Candidate dataset

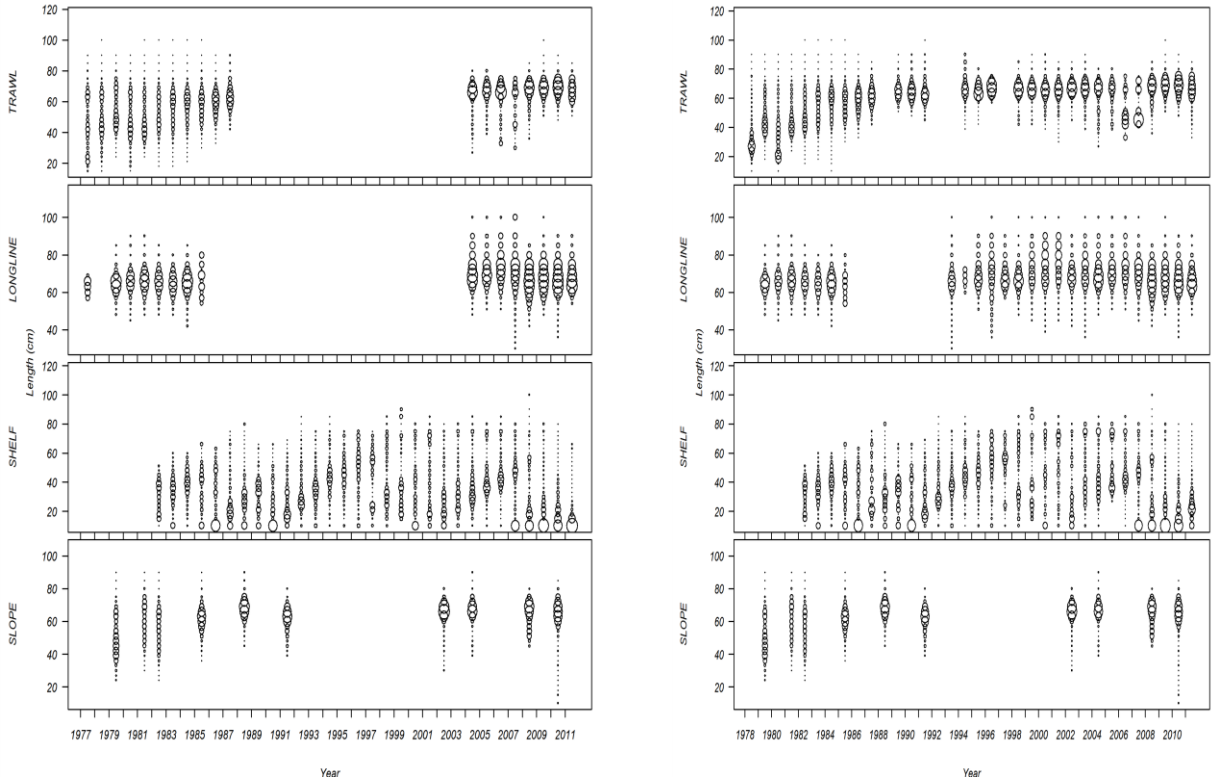
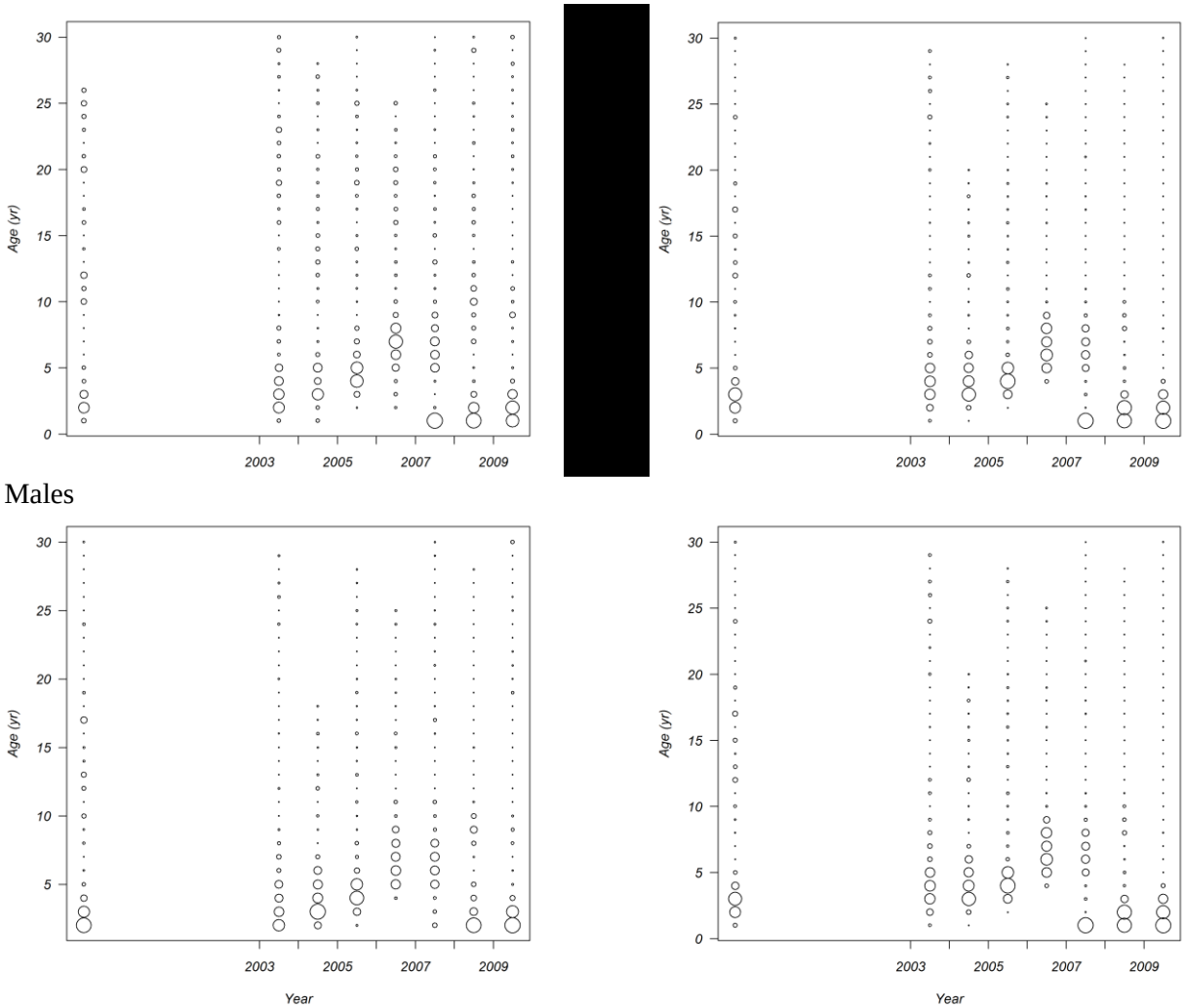


Figure 8c Length composition data from the 2011 Reference dataset and the 2012 Candidate dataset for males.

2011 Reference dataset 2012 Candidate dataset

Females



Males

Figure 9 Bering Sea shelf trawl survey age composition data from the 2011 Reference dataset (left) and the 2012 Candidate dataset (right) for females (top) and males (bottom).

For the 2012 assessment we also propose tuning the size composition sample sizes to the output effective sample sizes. Initial sample sizes for the two fisheries for each year was determined as the minimum of 100 + (number of hauls sampled/mean number of hauls sampled/100) or the number of hauls sampled. This schema was meant to reduce the influence of within sample and across haul autocorrelation in very large, single year, sample sizes on model fit. The largest differences in sample size are in the longline fishery (Fig. 10). The maximum sample size in the 2011 Reference dataset was 500 in the 1999 and 2000 longline fishery. The mean sample size in the 2011 Reference dataset for the longline and trawl fisheries was 251 and 114. In the 2012 Candidate dataset the means were 96 and 92 for the longline and trawl fisheries, respectively. The initial annual size composition sample sizes for the surveys were set at the same values as those used in the 2011 Reference dataset. The shelf trawl survey sample sizes were set at 100, the 2002 through 2010 slope survey sample sizes were set at 50, while those prior to 2000 were set at 25. The ABL longline sample sizes in the 2012 Candidate dataset were set at 60 following prior

assessments. The age composition sample sizes in the 2012 Candidate dataset were set at 100, following prior assessments.

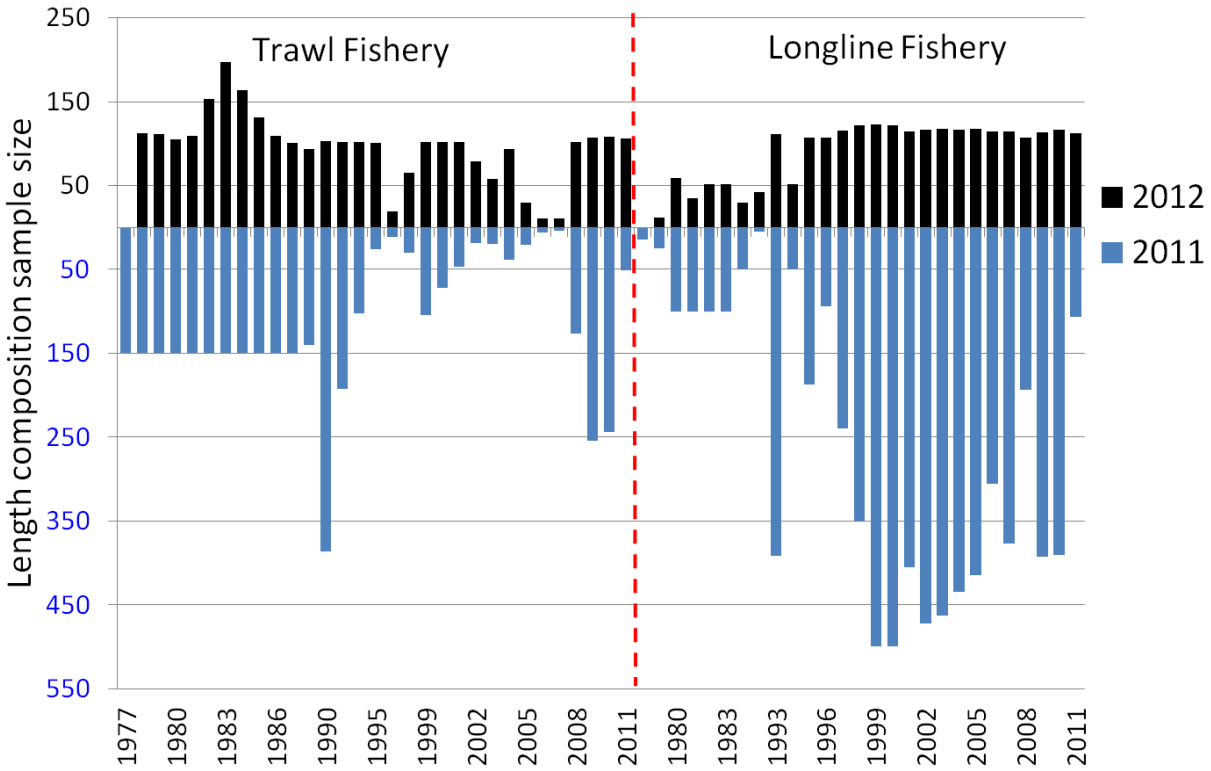


Figure 10 Initial length composition sample sizes for the trawl and longline fisheries in the 2011 Reference dataset (bottom) and the proposed 2012 Candidate dataset (top).

Effects of changes to length and age composition data and input sample sizes

The 2011 Reference model was run with only changes made to the length and age composition data to identify their effects on model results. Because we changed both the underlying data and the multinomial sample size, likelihoods between the 2011 Reference model fit to the old dataset and the model fit to the new dataset are not comparable (Table 2). The largest effect of these data changes was to reduce the estimated spawning biomass (Fig. 11) in spite of a reduction in the Shelf survey catchability (q) estimate from 0.73 to 0.51. Catchability for the Shelf trawl survey in the 2011 reference model was fit with a log uniform, non-informative prior with bounds at -2 and 2. Estimated recruitment in the early period (1960 - 1970) increased due to the smaller proportion of larger, older, females in the 1977 through 1980 trawl fishery length composition data which reduced the estimated 1972 and 1973 year class strength. In addition, the lower estimate of small fish for the 2011 survey in the proposed dataset compared to the 2011 reference dataset greatly reduced the estimate of the 2010 year class strength in the model.

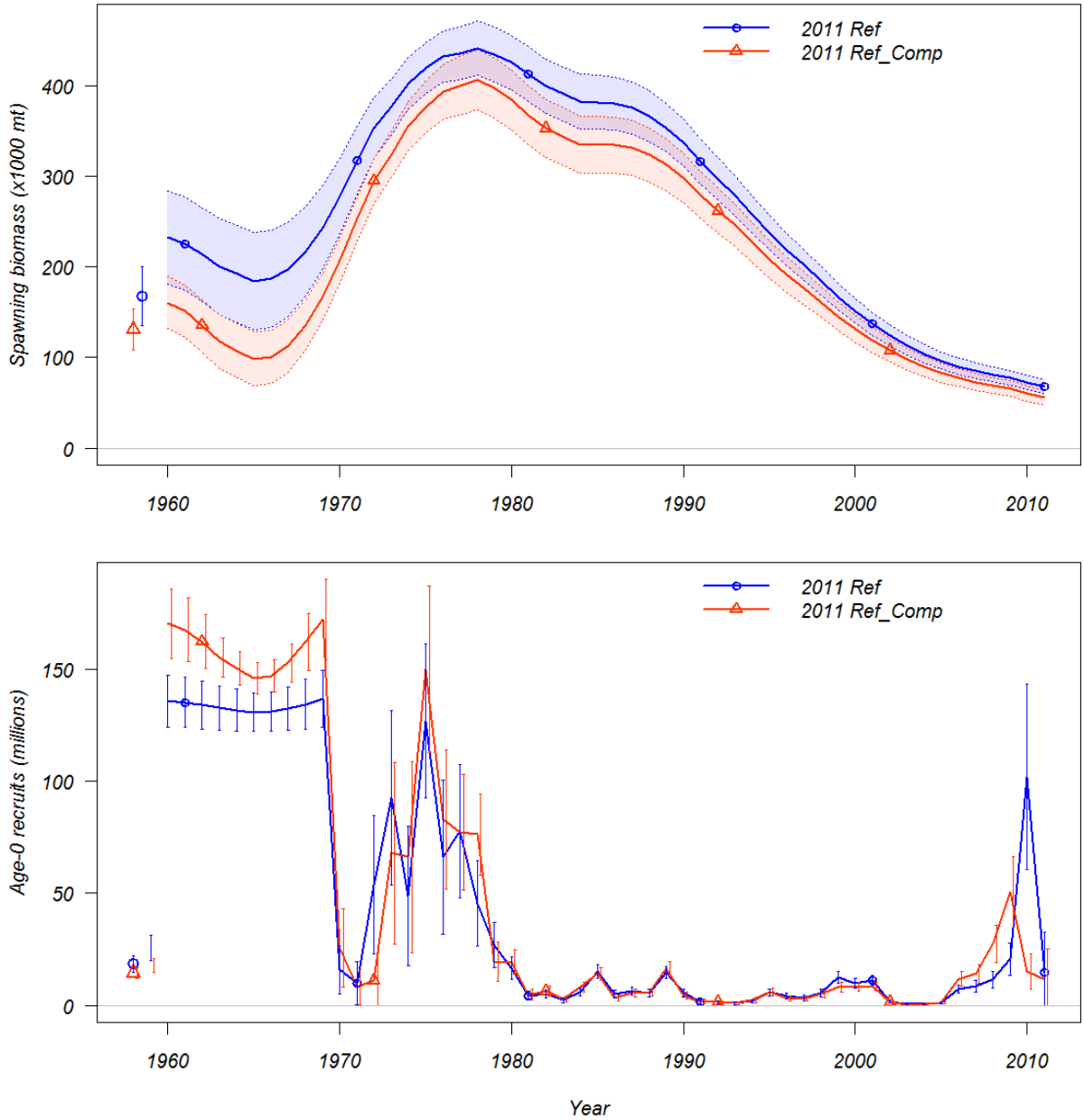


Figure 11 Spawning biomass (top) and age-0 recruitment (bottom) for 2011 Reference model and 2011 Reference model with new size and age composition data and estimated multinomial sample sizes (2011 Ref_Comp).

Cumulative effects of all data changes to model results

The cumulative effects of all the changes to the 2011 Reference dataset to the model results show an overall reduction in the estimated spawning biomass in spite of a reduction in estimated catchability for the shelf survey from 0.73 to 0.64, due to the change in the weight at length relationship (Table 2 and Fig. 12). The large 1977 year class in the 2011 Reference model is shifted to 1978. Recruitment in the 1972-1974 year classes is greatly reduced, but the 1975 and 1978 year classes are larger to account for the large Greenland turbot in the 1980's fisheries and surveys. Assessment of the model fit to these new data is not

really relevant at this point. Model changes to accommodate these new data and better selectivity curves for these new data will be addressed below.

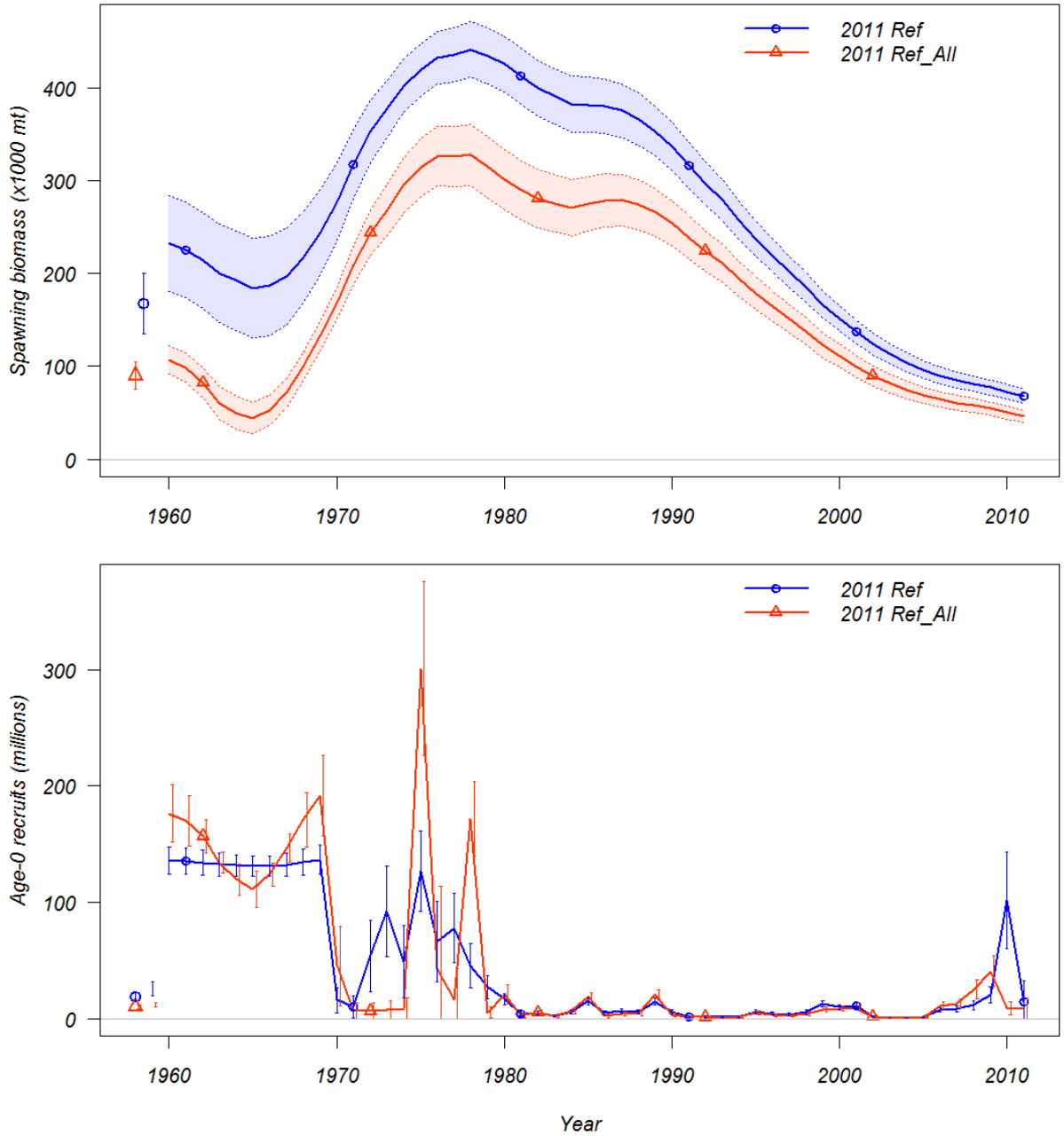


Figure 12 Spawning biomass (top) and age-0 recruitment (bottom) for 2011 Reference model and 2011 Reference model fit to the 2012 Candidate dataset with changes to the weight at length relationship, catch and index values, and length and age composition data applied.

Effects of tuning the length composition sample size

Once the 2011 Reference model was fit to the new data we proportionally reduced the sample size for each fishery or survey until the mean input sample size was close to the mean effective sample size. Candidate models were not precisely tuned as these are preliminary proposed models and more effort was placed on ensuring the models had the proper structure. More effort in precisely tuning the models will

be expended for the final models presented in November. Table 3, Figure 13, and Figure 14 show the adjustment done to the 2011 Reference model with all changes (2011 Ref_All), the tuned version of this model (2011 T_Ref_All), and one of the 2012 Candidate models (2012 RS_1Q) and their effects on spawning biomass and recruitment. Comparing fits to the length composition data between the 2011 Reference model with all data changes and the tuned version is not possible using likelihood as the overall weighting of the data is changed. Residual patterns in the length and age composition data remain similar and there appears to be little difference in the fit. As the weighting was reduced for the length composition data, the fit (Table 2 and Table 5) to all other data was slightly improved (-41 LL; -4%), mostly in the fit to the size at age data (-28 LL; -3%) and some improvement to the fit to the survey indices (-3 LL; -10%).

Table 3 Mean effective and adjusted input sample sizes and the sample size adjustment factor for three model configurations.

	TRAWL	LONGLINE	SHELF	SLOPE	ABL_LONGLINE
2011_Ref_ALL					
Mean Effective N	53	58	85	48	65
Mean adjusted input	96	92	100	35	60
Adjustment	1.00	1.00	1.00	1.00	1.00
2011 T_Ref_All					
Mean Effective N	55	57	80	55	67
Mean adjusted input	48	74	72	35	40
Adjustment	0.50	0.80	0.72	1.00	0.66
2012 RS_1Q					
Mean Effective N	57	70	80	40	49
Mean adjusted input	48	74	72	35	40
Adjustment	0.50	0.80	0.72	1.00	0.66

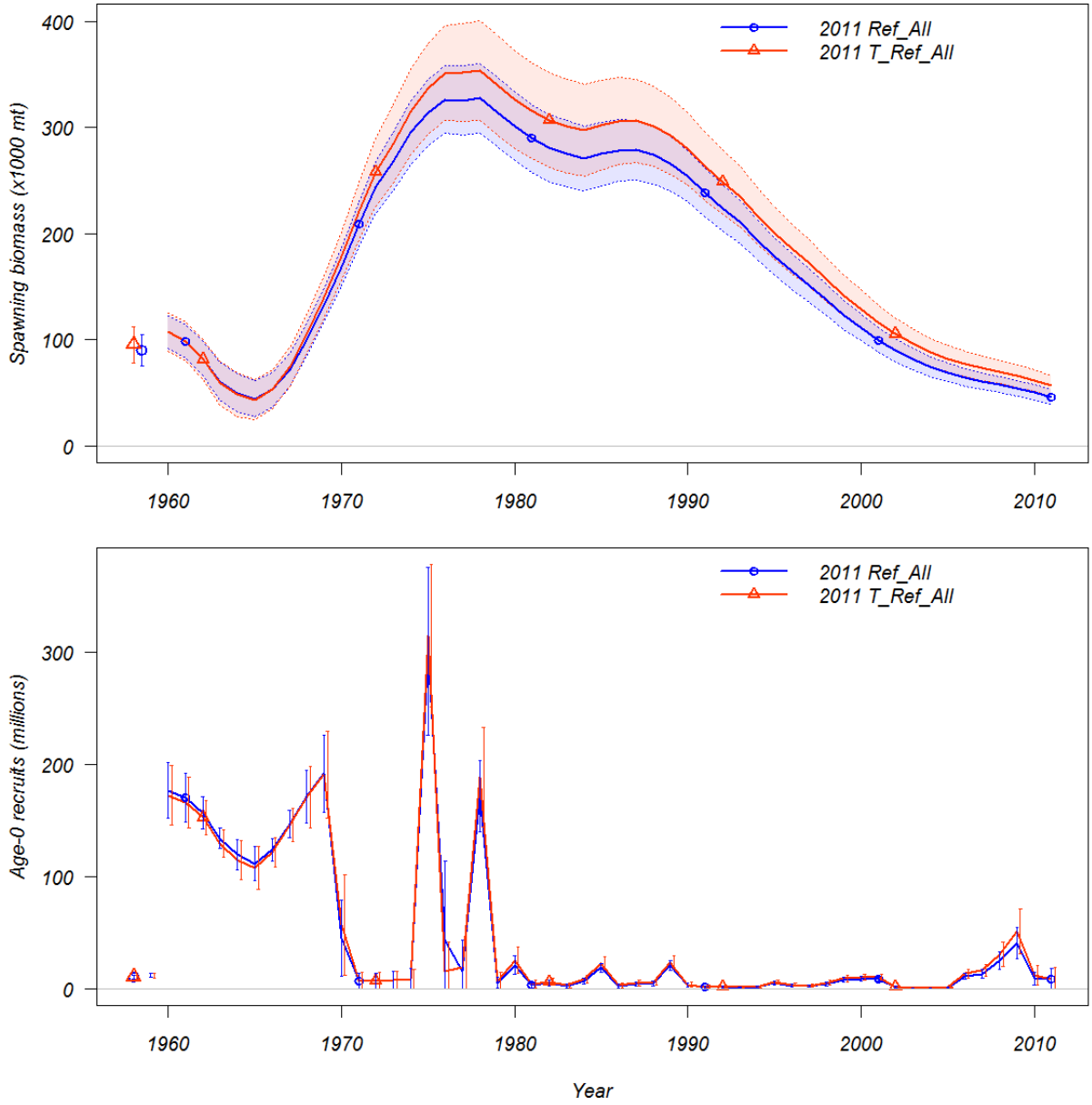
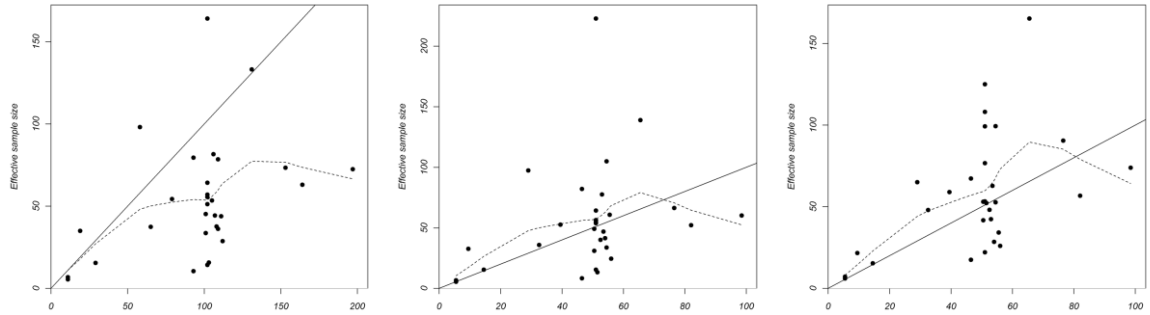


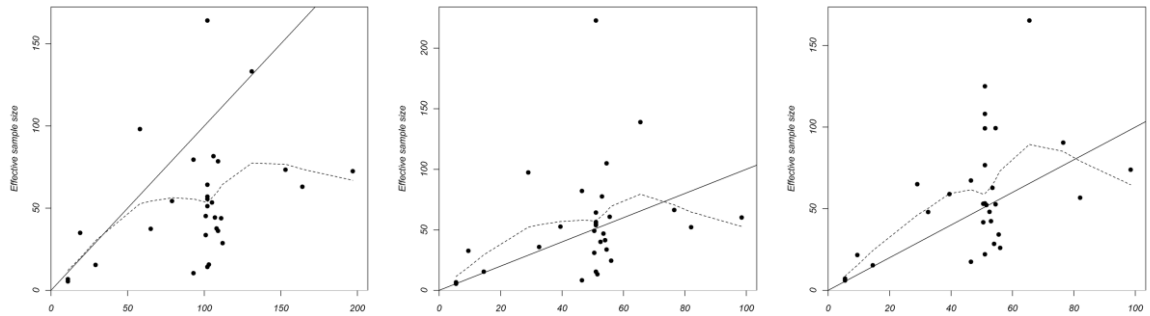
Figure 13 Spawning biomass (top) and age-0 recruitment (bottom) for 2011S Reference model fit to the 2012 Candidate dataset with changes to the weight at length relationship, catch and index values, and length and age composition data applied (2011 Ref_All) and this model tuned (2011 T_Ref_All).

011 Ref_All 2011 T_Ref_All 2012 RS_1Q

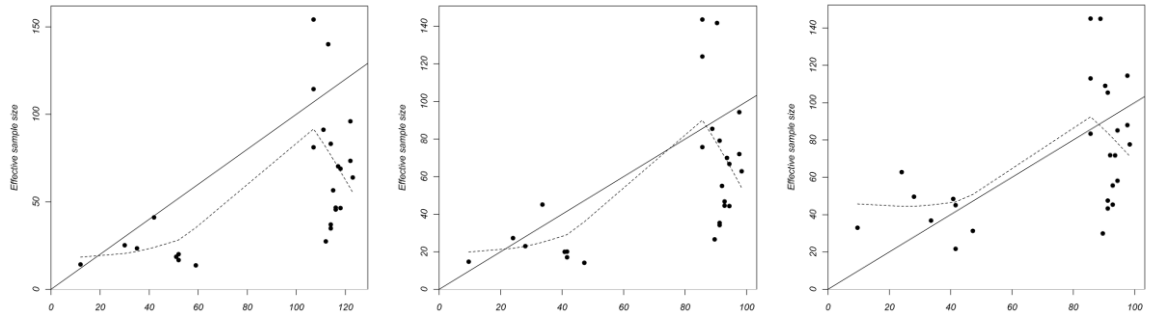
Trawl Fishery Females



Trawl Fishery Males



Longline Fishery Females



Longline Fishery Males

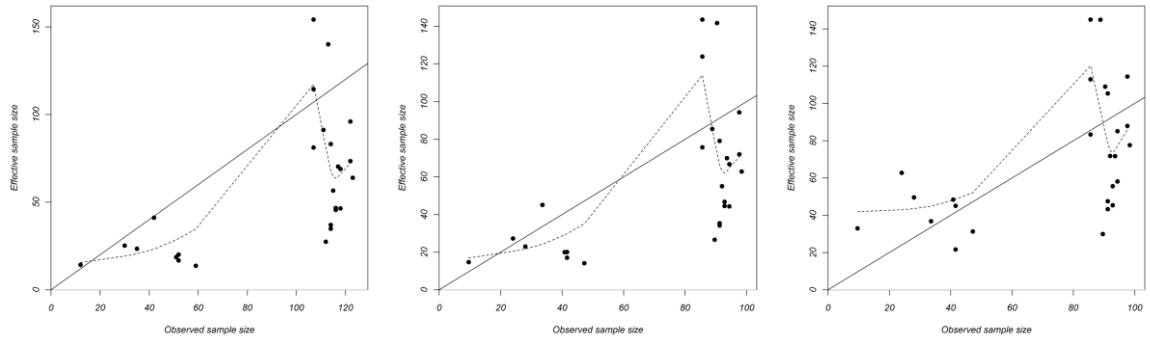


Figure 14 Expected sample size for the 2011 Reference model with all data changes (2011 Ref_All), the 2011 Reference model with all data changes and tuned sample size (2011 T_Ref_All) and one of the 2012 Candidate models with changes to recruitment, catchability, and selectivity (2012 RS_1Q). The solid line is the 1:1 line while the dotted line is a loess smoother fit.

Changes in the Assessment model

Recruitment and generating initial conditions

Because there was a large fishery on this stock prior to there being size or age composition data available (1960 – 1979; Fig.15), assumptions need to be made on the composition of the population for these early years if the early catches are to be included in the model. In the past when selecting the most parsimonious model in SS3 using maximum likelihood, a size/age distribution with a single, seemingly random, large recruitment event is selected. This was not deemed satisfactory by the previous stock assessment author. Recruitment in the 2011 Reference model was assumed to follow a Beverton-Holt stock recruitment curve with steepness at 0.9 and sigma R at 0.6. In order to generate a more diverse size/age structure in the population at the time data become available and to support the early fishery, the 2011 Reference model was fit with a higher R_0 in the years 1960 through 1969 than in following years. Recruitment pre-1970 was assumed to follow a spawner-recruit curve with no deviation. Recruitment deviations post-1970 were assumed to be simple lognormal deviations bounded between -7 and 7. The 2011 Reference model therefore assumed a higher productivity prior to 1970 and different recruitment relationships for years 1960 through 1969 than years 1970 through 2010 (Fig.16).

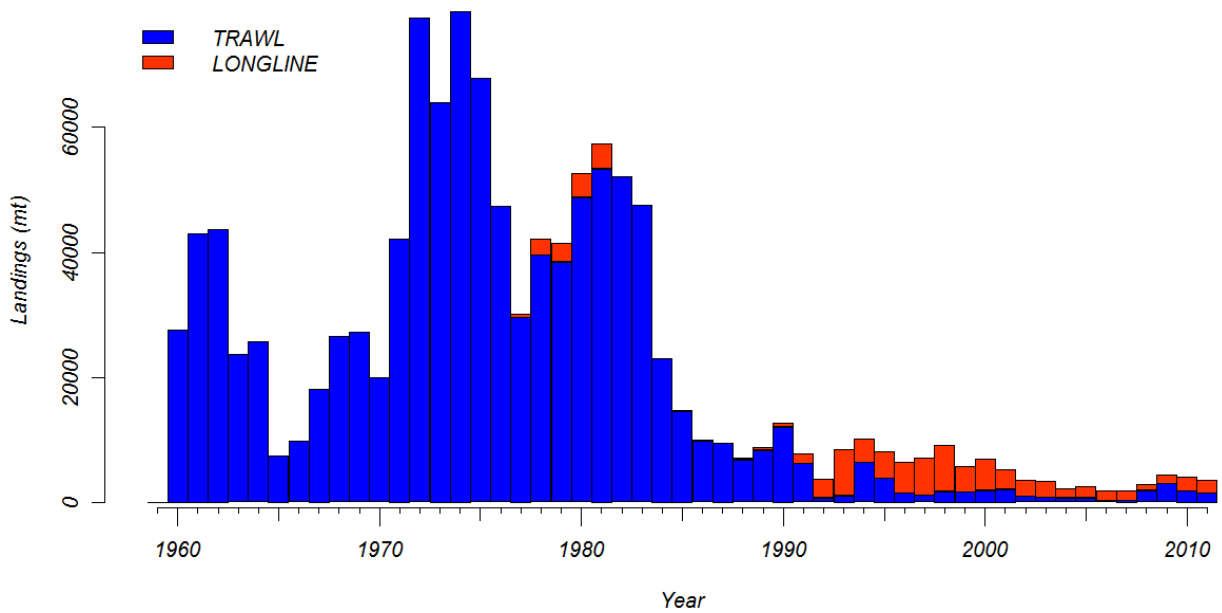


Figure 15 Greenland turbot Longline and Trawl catch in the Bering Sea and Aleutian Islands area from 1960 through 2011. This data includes targeted catch and bycatch.

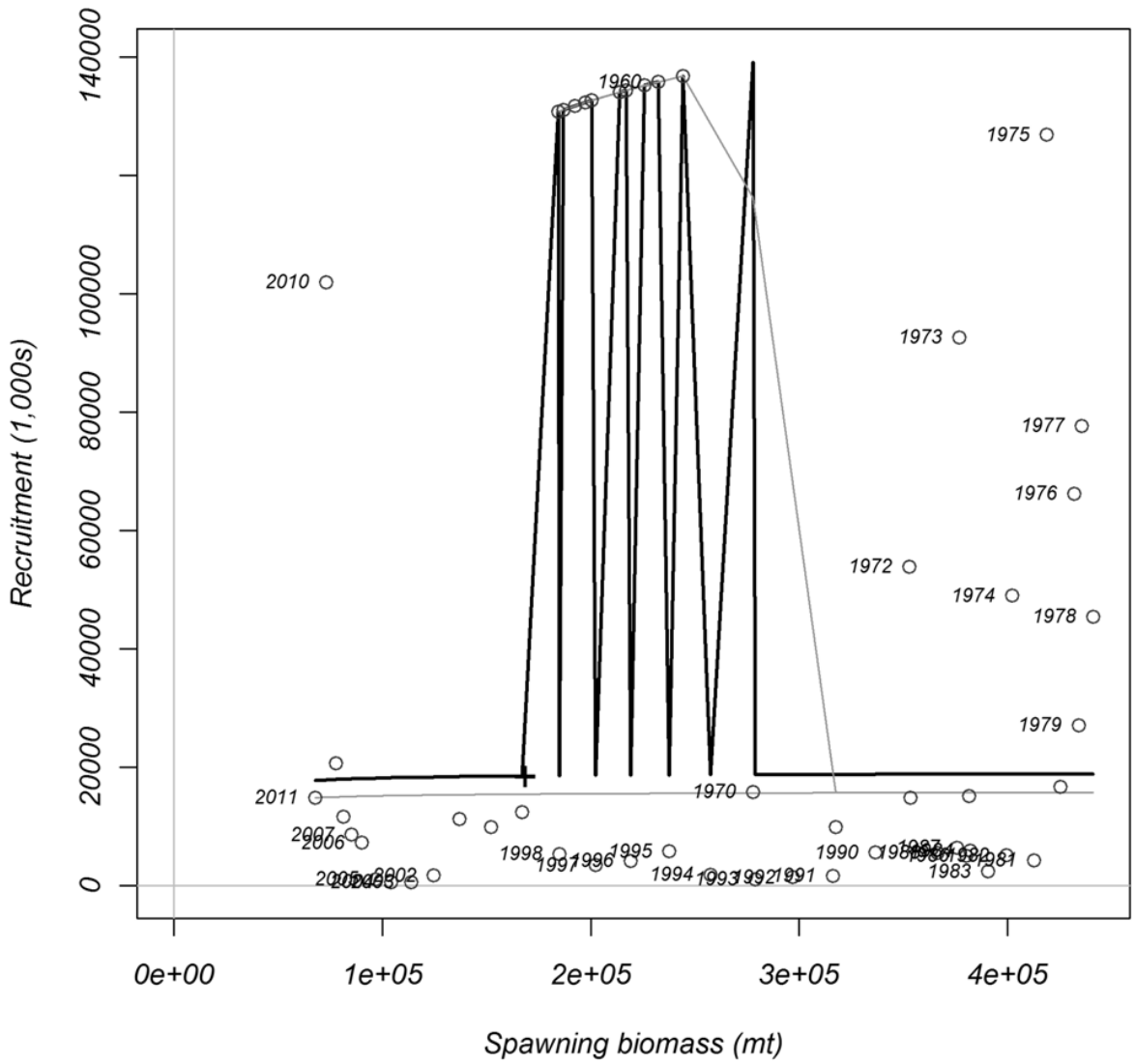


Figure 16 Combined early and late period stock recruitment relationship for 2011 Reference model showing the two different Beverton-Holt recruitment curves combined (grey line) and higher assumed productivity in the years 1960 -1970 than in later years.

In the 2012 Candidate models a single R_0 is assumed for all years. The models are fit to Beverton-Holt stock recruitment curve with a steepness (h) of 0.79 consistent with values found for Greenland turbot stocks in the North Atlantic and Arctic Ocean (Mertz and Myers 1996, Myers et al. 1999) An autocorrelation parameter was also investigated where the prior component due to stock-recruitment residuals (ε_i) is $\pi_R = \frac{\varepsilon_1^{\,2}}{2\sigma_R^2} + \sum_{i=2}^{n} \frac{\left(\varepsilon_i - \rho\varepsilon_{i-1}\right)^2}{2\sigma_R^2\left(1-\rho^2\right)}$, where ρ is the autocorrelation coefficient, and σ_R^2 is the assumed stock recruitment variance term. Although different ρ-values were explored, ρ was fixed at 0.7 for all models presented in this document. The starting year in all models was pushed back from 1960 to 1945 to allow the model more time to build a diverse population size distribution as expected from a

species with an assumed natural mortality of 0.112 (Fig. 17). Recruitment deviations for 1945 through 2011 were assumed to be simple lognormal deviations bounded between -5 and 5.

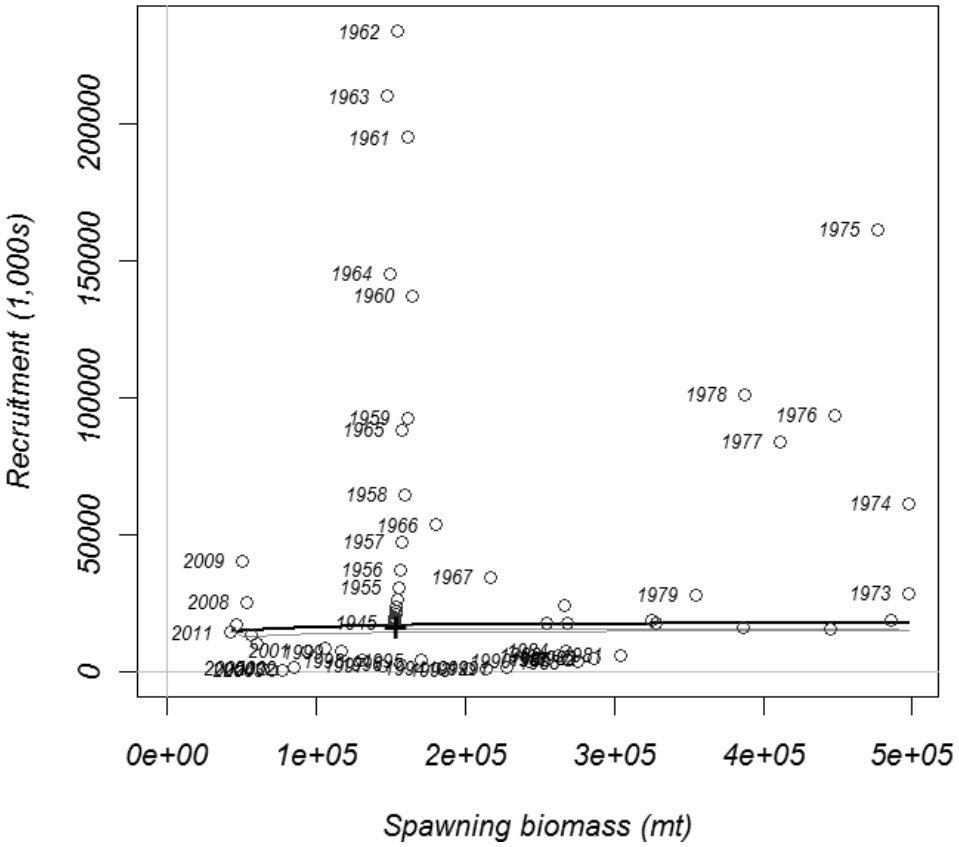


Figure 17 Stock recruitment assumed in the 2012 Candidate model 2012 R which is the 2011 T_Ref_All model with changes to recruitment. Higher productivity in the 1960s to account for high early catches is modeled as deviations from a single stock recruitment curve (grey line).

Effects of changes to recruitment and initial conditions

The changes to recruitment improved overall model performance (Table 5) with a lower negative log likelihood (- 42 LL, -2%). This improvement was made in fitting both the survey indices (-1.5; -5%) and the length composition (-58.7; -4%) with some slight degradation in fit to the size at age (+6.4; 1%) and an increase in the penalty on the priors (+18.4, +275%). The 2011 tuned reference model fit to the new data (2011 T_Ref_All) had a mean recruitment for 1960 through 1970 of 147 million age-0 fish and a coefficient of variation (CV) of 0.19. The estimates the 2012 Candidate model with only a change in recruitment (2012 R) are much lower (mean 114 million age-0 fish), but with a much higher CV (0.71). The result of the change in how the early recruitments are handled is a smooth curve of recruitment peaking in 1962 with a large degree of uncertainty around the estimates compare to the nearly flat and highly certain recruitment assumed in the 2011 Reference model (Fig. 18). The lack of uncertainty in the 2011 Reference model early recruitment translates into overconfident estimates of early biomass and current stock status (Fig. 18). The 2012 Candidate recruitment model more accurately reflects the degree of uncertainty in these early estimates and better reflects uncertainty in the early estimates of biomass and current stock status.

Table 4 Changes from 2011 Reference model for 2012 Candidate models and number of parameters. Model changes from the reference are indicated by shading.

Model	2011 T_Ref_All	2012 _R	2012 RS_1	2012 RS_2	2012 RS_1Q	2012 RS_2Q	2012 RS_1QV	2012 RS_2QV
Slope Q	0.75	0.75	0.75	0.75	0.579	0.548	0.597	0.564
ABL selectivity	Double Normal	Double Normal	Logistic	Spline	Logistic	Spline	Logistic	Spline
M	0.112	0.112	0.112	0.112	0.112	0.112	0.112	0.112
Shelf survey Age Comp	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Recruitment bounds	(-7,7)	(-7,7)	(-5,5)	(-5,5)	(-5,5)	(-5,5)	(-5,5)	(-5,5)
Female M	0.112	0.112	0.112	0.112	0.112	0.112	0.112	0.112
Shelf Q	0.572	0.700	0.778	0.771	0.705	0.682	0.725	0.702
sigma R	0.6	0.6	0.6	0.6	0.6	0.6	0.9	0.9
h - steepness	0.9	0.79	0.79	0.79	0.79	0.79	0.79	0.79
Autocorr	0	0.7	0.7	0.7	0.7	0.7	0.7	0.7
n parameters	111	129	129	132	130	133	130	133

Table 5 Fit to the 2011 Reference model with and all three sets of data changes combined (2011 Ref_All), Ref All with tuned composition data (2011 T_Ref_All) and 2012 models with changes in early recruitment (R), alternative selectivity curves (S_1 and S_2), changes in catchability assumptions (Q), and an alternative assumption on Sigma R (V) .

		2011 T_Ref_All	2012 R	2012 RS_1	2012 RS_2	2012 RS_1Q	2012 RS_2Q	2012 RS_1QV	2012 RS_2QV
Likelihoods									
	Total	2799.53	2757.43	2713.68	2660.87	2708.83	2653.99	2664.90	2609.93
	Survey	-28.30	-29.81	-25.11	-24.88	-25.67	-27.12	-26.75	-27.86
	Length Composition	1606.11	1547.42	1439.42	1386.66	1442.19	1389.24	1424.02	1370.61
	Age Composition	221.47	221.77	215.34	216.79	215.02	216.42	212.81	214.24
	Parameter priors	6.69	25.12	34.39	33.99	33.82	34.39	34.08	34.43
	Size at Age	912.10	918.51	972.78	971.14	968.63	966.12	970.33	967.91
	Recruitment	81.46	74.41	76.87	77.17	74.84	74.95	50.41	50.60
Key Parameters									
	SR_LN(R0)	9.24	9.73	9.68		9.73	9.73	9.70	9.71
	H – steepness	0.9	0.79	0.79	0.79	0.79	0.79	0.79	0.79
	Sigma R	0.6	0.6	0.6	0.6	0.6	0.6	0.9	0.9
	ρ - autocorrelation	0	0.7	0.7	0.7	0.7	0.7	0.7	0.7
	Q for Shelf Survey	0.57	0.70	0.78	0.77	0.70	0.68	0.73	0.70
	Q for Slope Survey	0.75	0.75	0.75	0.75	0.58	0.55	0.60	0.56
	L_at_Amax_Fem	87.99	87.87	88.54	88.57	88.50	88.51	88.50	88.51
	L_at_Amax_Mal	73.76	72.48	72.63	72.77	72.61	72.77	72.59	72.74
	VonBert_K_Fem	0.14	0.13	0.12	0.12	0.13	0.12	0.12	0.12
	VonBert_K_Mal	0.18	0.19	0.18	0.18	0.18	0.18	0.18	0.18
Recruits 1960-1969									
Age-0 (1×10^9)	Mean	147.20	114.09	104.78	101.31	104.57	103.02	113.18	112.47
	Median	149.21	112.75	88.45	83.93	88.75	84.12	81.32	74.60
	CV	0.19	0.71	0.60	0.62	0.59	0.61	0.79	0.82
Spawning Stock Biomass									
(1,000 t)	1978	353.92	387.18	197.87	174.62	206.03	182.45	206.36	183.22
	2011	56.75	42.86	19.73	20.02	25.39	27.25	24.00	25.79

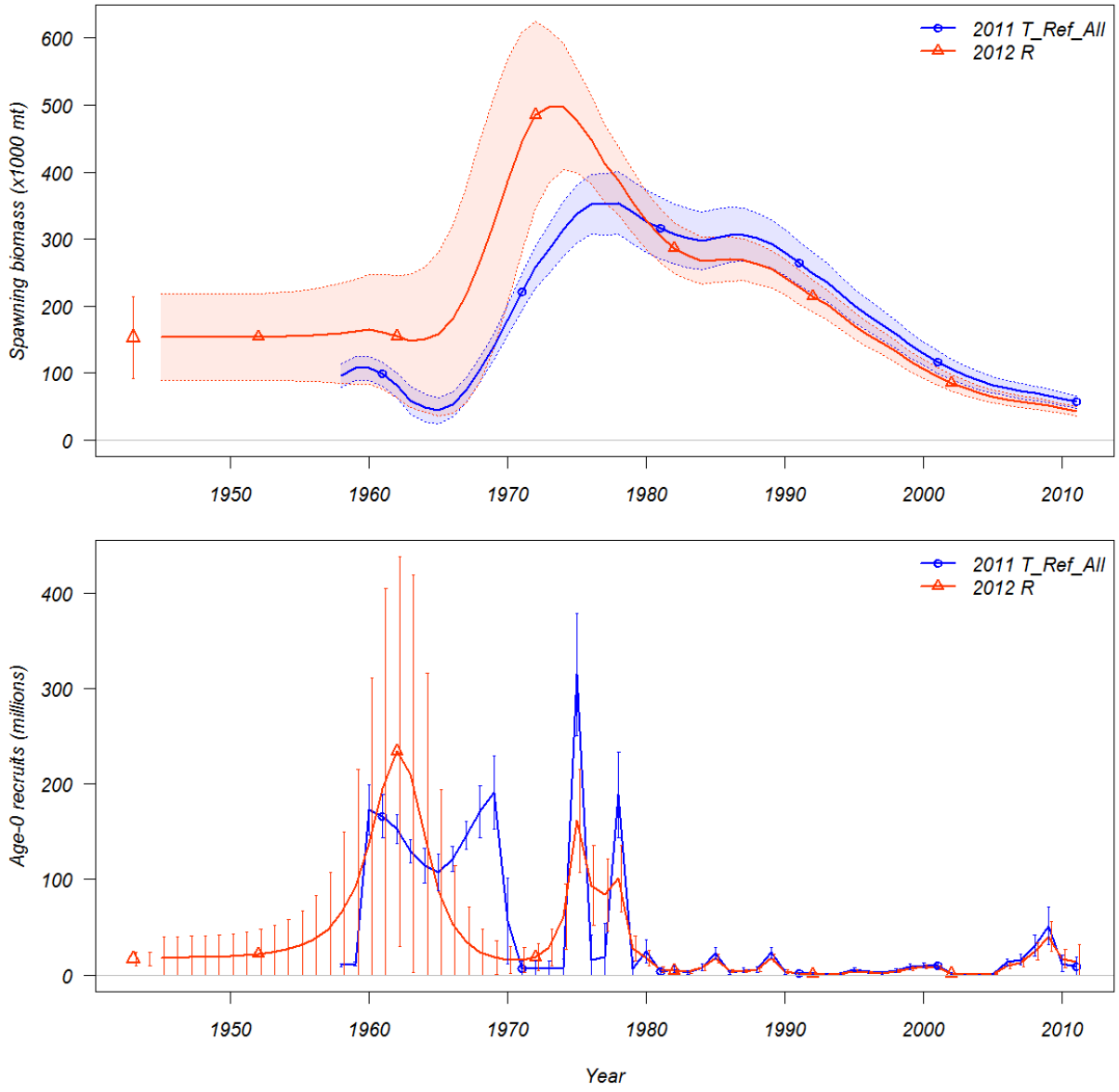


Figure 18 Spawning biomass (top) and age-0 recruitment (bottom) for 2011 tuned reference model with all proposed data changes (2011 T_Ref_All) and 2012 candidate model with alternative early recruitment assumptions (2012 R).

Selectivity

There was a lot of effort expended on exploring appropriate selectivity curves for the 2012 assessment. In this document we present two options for the 2012 Candidate model (S_1 and S_2). The only difference between the two selectivity options is in the selectivity curve chosen for the ABL longline survey.

- In S_1 the ABL Longline length composition data are fit with logistic curve.
- In S_2 the ABL Longline length composition data are fit with a four node spline. All other selectivity curves are the same between the two options.

The main difference between the 2011 Reference model selectivity and the 2012 Candidate selectivity is in how the male and female selectivity curves are allowed to differ. A new method for fitting curves that differ between male and females was implemented in the latest version of SS3 (V 2.24). In previous SS3 versions the male and female selectivity curves took the same underlying shape, but the curve could be altered between males and females using four parameters :

P1 – size at which a dogleg occurs
P2 – log(relative selectivity) at the minimum size
P3 – log(relative selectivity) at the dogleg
P4 – log(relative selectivity) at maximum size

These options do not allow the model to fit curves in which, for example, the male selectivity is higher than females in smaller fish, but lower than females in larger fish. This is the case for Greenland turbot in which males are much smaller than females reaching maturity at a smaller size, where both migrate to deeper waters as they mature, and where the fisheries and surveys are spatially distinct, targeting or encountering different parts of the population.

This problem was addressed in the latest version of SS3 (3.24) in which more flexibility in fitting the selectivity curves of the opposite sex is available. If the size selectivity pattern is logistic, then SS3 requires 3 parameters to differentiate the curve from the opposite sex:

p1 is added to the first selectivity parm (inflection)
p2 is added to the second selectivity parm (width of curve)
p3 is the asymptotic selectivity

If the size selectivity pattern is the double normal, then five parameters are needed to differentiate from the opposite sex:

p1 is added to the first selectivity parameter (peak)
p2 is added to the third selectivity parameter (width of ascending side)
p3 is added to the fourth selectivity parameter (width of descending side)
p4 is added to the sixth selectivity parameter (selectivity at final size bin)
p5 is the apical selectivity

This new method was explored for all fisheries and surveys with separate sex data. In addition, the longline and slope survey selectivity was simplified to a single logistic curve since the curve fit in last year's assessment, although fit as a double normal, was in effect a simple logistic curve.

The ABL longline catch at size data is somewhat difficult to fit since the data are from combined sexes and appear to have a bimodal distribution, one for males and another mode for females. Although a simple logistic model can be fit to the data, patterns in the residuals suggest some deficiencies in the fit. For this reason we explored fitting a four node spline in the S_2 selectivity option presented here. The four node selectivity spline may better take into account the odd shape expected from the combined sex data (Methot 2011).

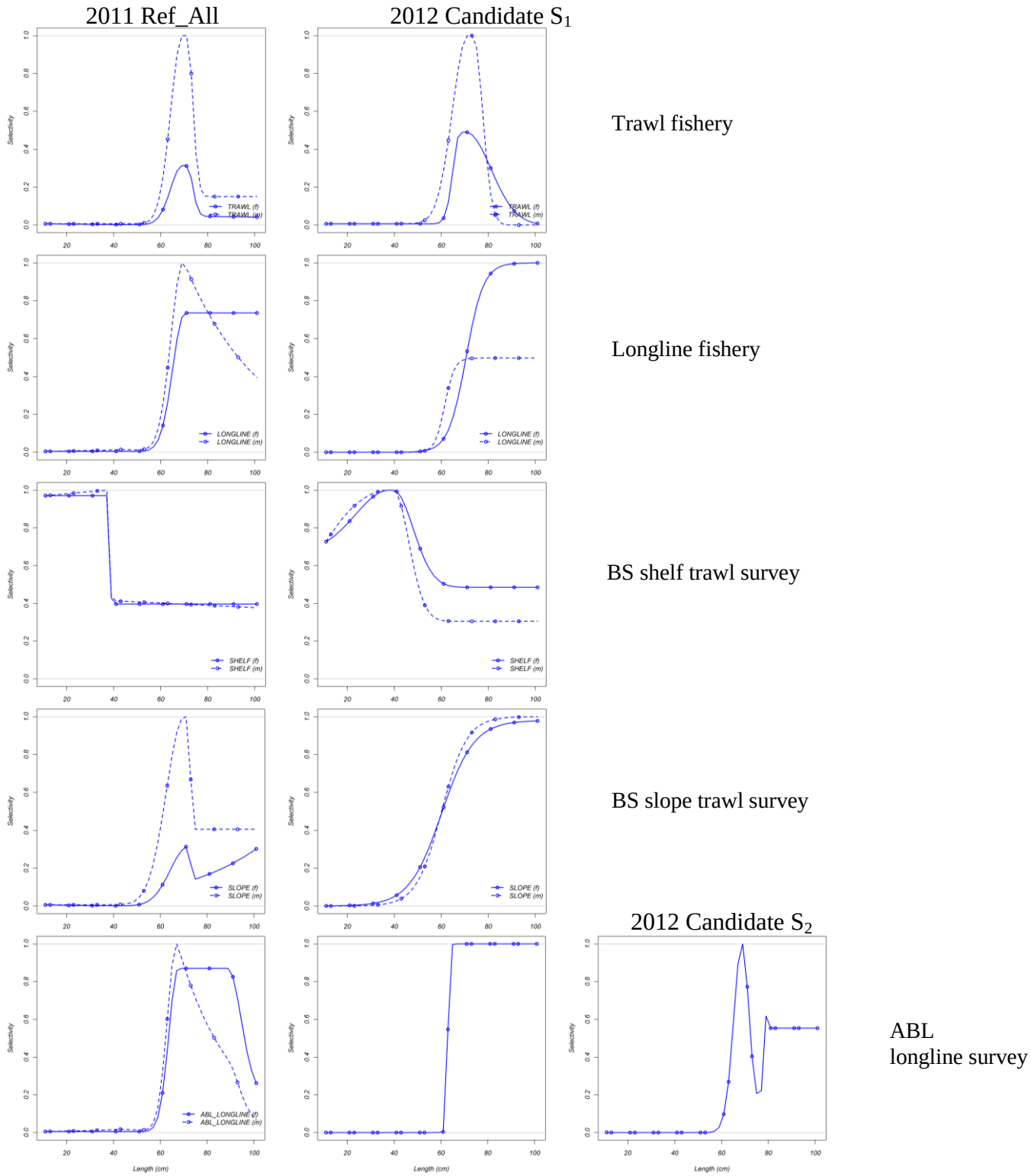


Figure20 2011 selectivity for 2011 Reference model and 2012 Candidate models for both male (solid) and female selectivity (dashed). Note only the ABL Longline survey selectivity changes between S_1 and S_2 models.

Effects of changes to selectivity curves

The changes in selectivity curves from the 2011 Reference model to the S_1 candidate selectivity provided some improvement to the overall model fit (Table 5) and residual pattern in the longline fishery (Fig. 21). The greatest improvement was in the fit to the size composition data (-166.69 LL; -10%) with better fits in both fisheries and all three surveys. The fit was most improved in the longline and trawl fisheries (-80 and -60.5 LL, respectively; -27% and -14%). The fit to the two trawl survey abundance indices was degraded (Shelf +5 LL or +20% and Slope +3.6 LL or +69%), while the fit to the ABL longline survey index was much improved (-5 LL; -320%). The fit to the size at age was degraded with an overall increase of +60 LL (+7%). The fit to the Shelf age composition data was slightly improved with a decrease in negative loglikelihood of -6 or -3%. The 2012 S_2 candidate selectivity option improved the model fit primarily to the ABL longline length composition data (Fig 22). The fit to the ABL Longline length composition data was improved by -50.3 LL or -39% from the S_2 over the S_1 configuration, with an overall improvement to the total model fit of -52.7 LL (-2%). Fits to all other data components in the model were changed individually by less than ±2 LL.

Even with the additional flexibility in fitting the two sexes in the latest version of SS3 there remains patterns in the residuals that are problematic in the early years of the trawl length data (1979-1989). The trawl fishery length composition data are pooled from the directed fishery and from fish caught in other fisheries. The directed fishery targeted the larger fish (predominantly females) on the slope, while the bycatch fishery mostly caught smaller fish (predominantly males) on the shelf, resulting in very different expected selectivity patterns for the two sexes. Currently SS3 can't handle such a large difference in selectivity patterns between sexes for the same fishery. In the future the authors would like to try to separate out the bycatch trawl data from the targeted trawl fishery data to see if the patterns in the size composition data for these early years can be rectified and perhaps present the results as a competing model in November. Since target is not included in these older data, this task may be difficult to accomplish.

The greatest changes in the model results (Table 5) from the 2011 Reference model selectivity configuration in the 2012 R Candidate model to the 2012 S_1 selectivity or 2012 S_2 configuration are a change in the Shelf survey catchability from 0.70 to 0.78 for S_1 and 0.77 for S_2 and a shift in the peak of the early period (1945-1974) recruitments from 1962 in the 2011 reference configuration to a slightly smaller peak in 1965 (Fig. 24) in both the S_1 and S_2 configurations (from 233.9 to 207.1 and 203.9 billion (1×10^9) age-0 fish, respectively). There is also a reduction in the point estimate of recruitment for the 1977 and 1978 age-0 recruitment in both the S_1 and S_2 configurations compared to the 2011 Reference configuration from 83.8 and 101.2 billion (1×10^9) fish to 77.65 and 88.25 billion fish for S_1 and 80.12 and 85.46 billion fish for S_2. Mean recruitment across all years (1945-2011) dropped from 36 billion age-0 fish using the 2011 reference selectivity to 32 billion age-0 fish for both the S_1 and S_2 selectivity options. The reduction in mean recruitment and increases in Shelf survey catchability resulted in smaller spawning and total biomass estimates for all years using either of the candidate selectivity options. The 1978 spawning biomass point estimate drops from 387 thousand tons using the 2011 Reference selectivity configuration to 198 and 175 thousand tons using the S_1 and S_2 configurations. The 2011 spawning biomass point estimate drops from 43 thousand tons using the 2011 Reference configuration to 20 thousand tons for both S_1 and S_2.

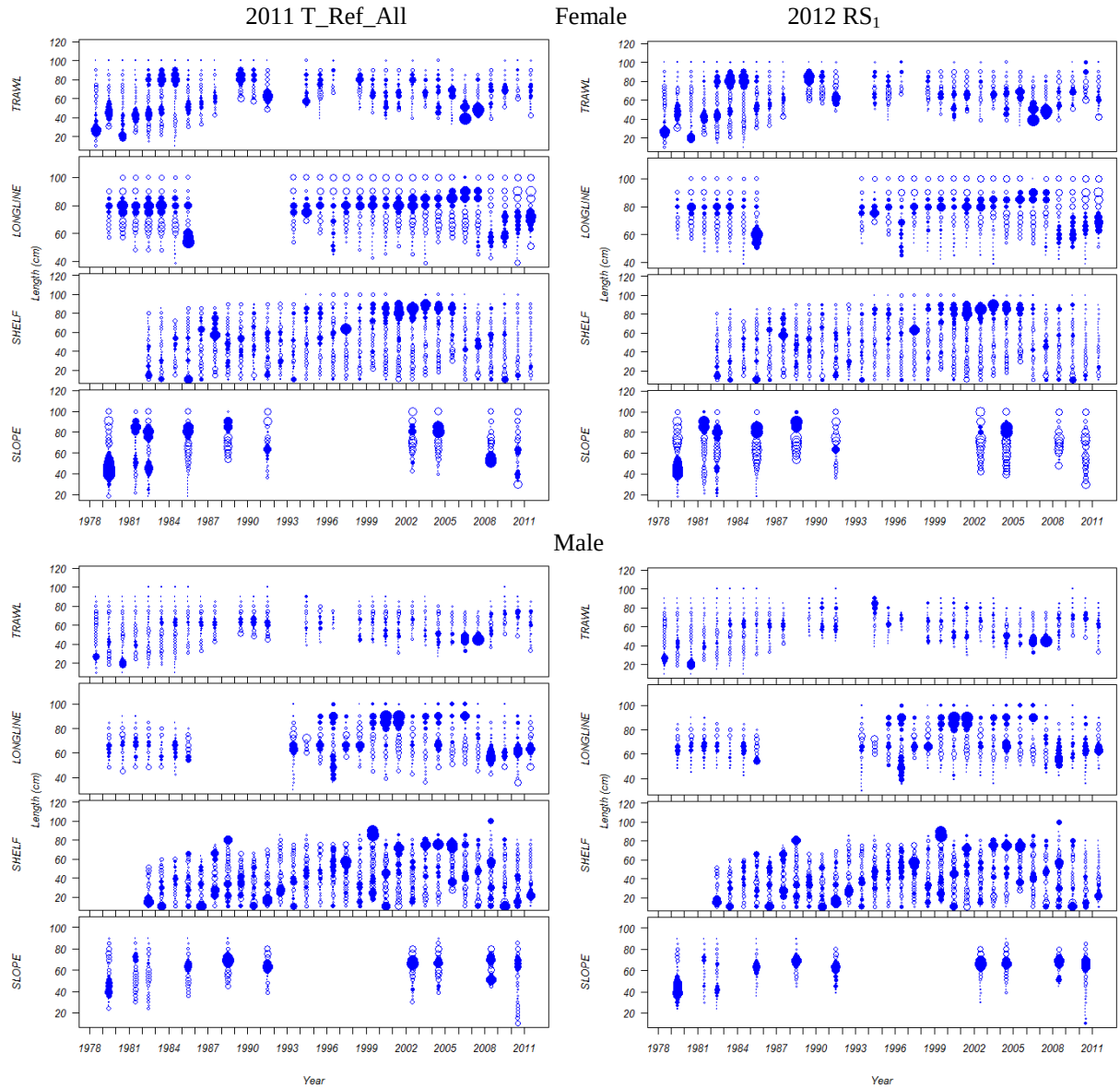


Figure 21 Residuals from length composition data fits for Model R which employs the 2011 reference model selectivity configuration and for Model RS_1. **Note that the scales for all of the plots differ. These plots are meant to help in examining possible patterns in the residuals and not the exact fits.** In each plot the maximum value for that data type is the same size across all data types.

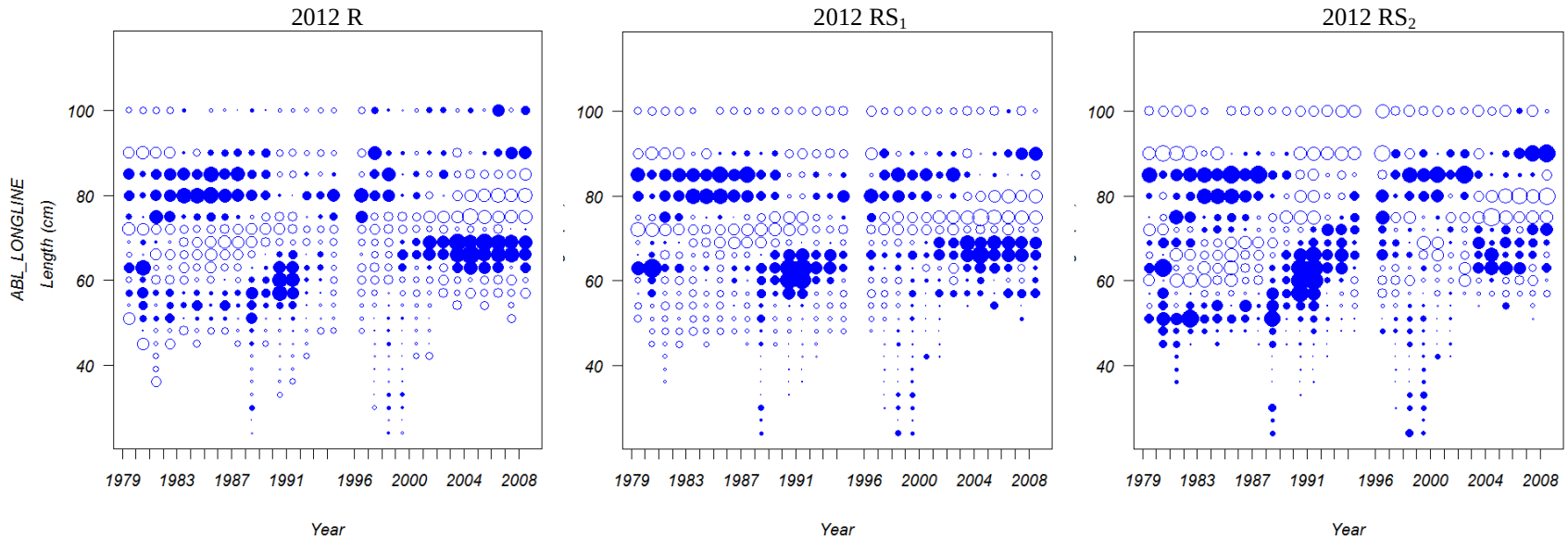


Figure 22 Residuals for fits to ABL longline data for the 2011 Reference selectivity configuration (2012 R) and S1 and S2 configureations. Note that the max value for the three plots differ (max 2012 R =3.07, 2012 RS1=3.04, and 2012 RS2 = 2.02)

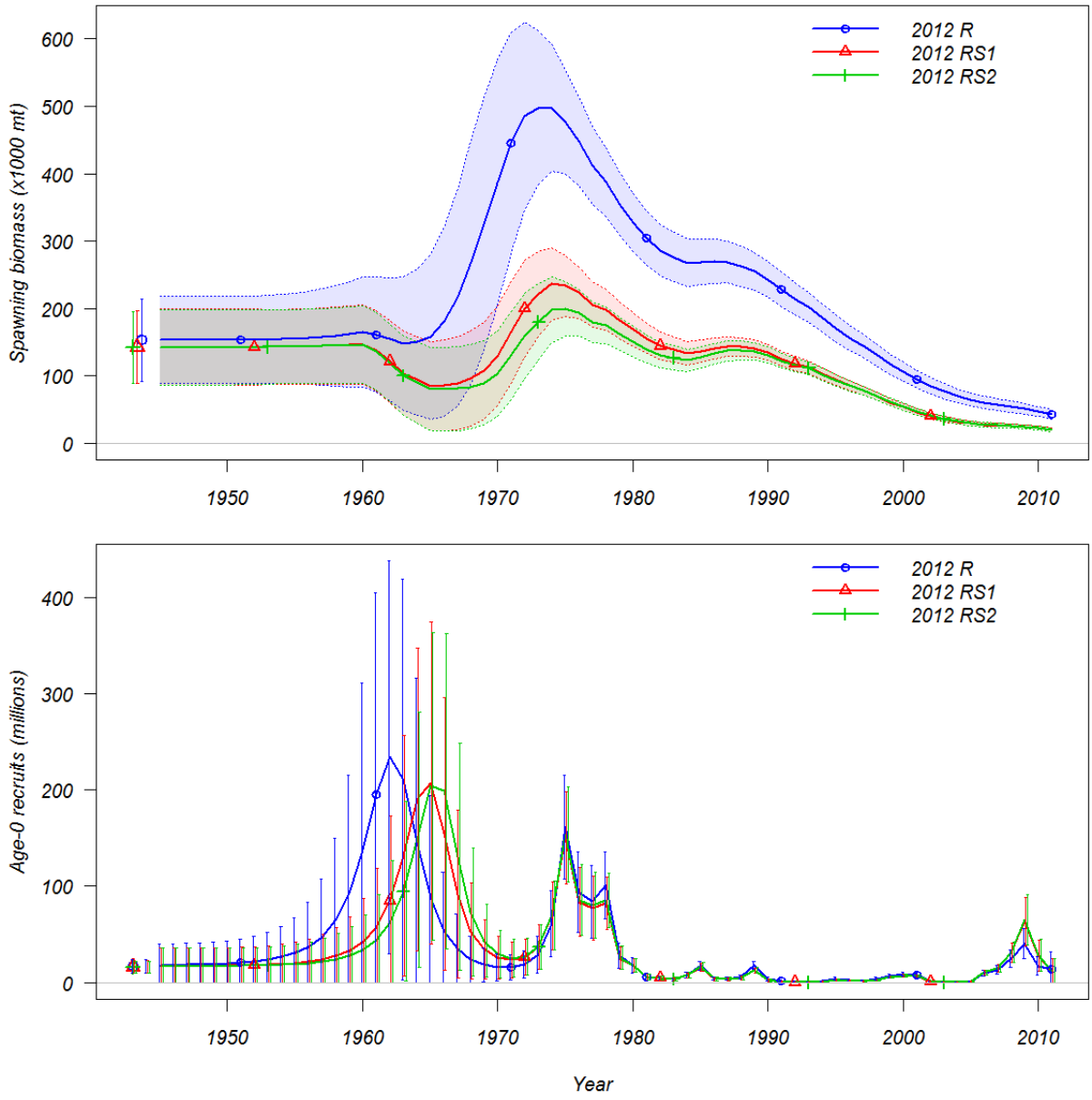


Figure 23 Spawning biomass (top) and age-0 recruitment (bottom) for 2012 Candidate model with changes to recruitment from the 2011 Reference model, with catchability fitted for both the Bering Sea slope and shelf trawl surveys using informative lognormal priors and sigma R at 0.6 (2012 RS1Q) and the same model with a four node spline fit for the Auke Bay Laboratory longline survey (2012 RS2Q).

Catchability in the Slope Survey

The 2011 Reference model assumes a fixed catchability for the slope trawl survey (q_{Slope}) of 0.75 and estimates catchability for the Bering Sea shelf survey (q_{Shelf}) with a log uniform, non-informative prior bounded between -2 and 2. There is no strong evidence to support the assumption that q_{slope} is exactly 0.75. Models were explored loosening this assumption with both the slope and shelf trawl survey catchability estimated using informative, lognormal priors [$\log(q_{slope}) \sim \mathbb{N}(-0.28768, 0.1)$ and $\log(q_{shelf}) \sim \mathbb{N}(-0.69385, 0.1)$].

Effects of loosening assumptions on Bering Sea slope trawl survey catchability

When assumptions on q_{Slope} are loosened and an informative prior distribution is place on q_{Shelf}, lower estimates of q for both the Bering Sea slope and shelf trawl surveys are obtained (Table 5 and Fig. 24). Shelf survey catchability changes from 0.78 to 0.7 while the Slope survey catchability changes from a fixed value of 0.75 to 0.68. Only marginally better fits to the survey indices (< -1 LL) and marginally poorer fits to the length composition data (>+3) are achieved in the alternative Q configuration. The changes do improve the fit to the shelf survey size at age data fit to a small degree amount (-4 LL). None of these changes to the model fit are substantial, but the change to the model estimates of spawning stock biomass is noticeable (Fig.25). The 2011 estimate of spawning stock biomass in the fixed q_{Slope} model (2012 RS_1) is 19.7 thousand tons while the spawning stock biomass estimate in the model where q_{Slope} is fit with an in formative prior (2012 RS_1Q) is 25.4 thousand tons, a 29% increase in the estimate. Similarly in the S_2 models the estimate changes from 20.0 thousand tons (2012 RS_2) to 27.2 thousand tons (2012 RS_2Q) when the assumptions on q_{Slope} are loosened. Although the variance on the MLE estimates are quite small the change to the overall likelihood is small and suggests that catchability is not well defined in the available data. This is likely due to the problem of this stock straddling the US-Russia border, migration of the stock between areas, and the surveys not consistently covering the same proportion of the stock each year.

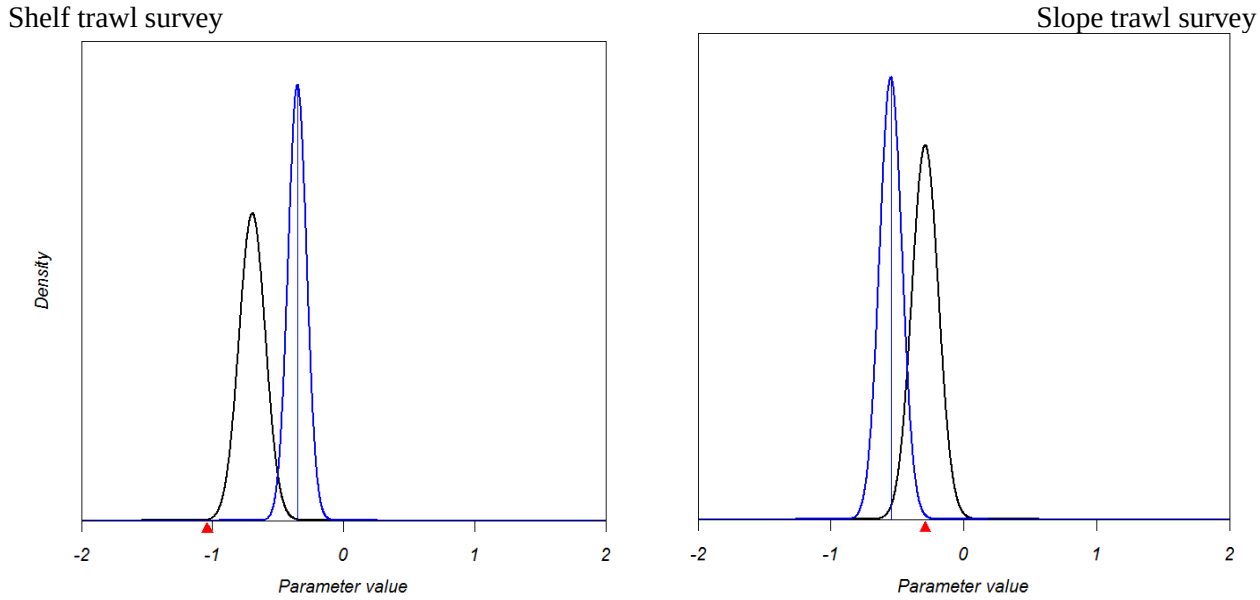


Figure 24 Prior distribution (black lines), initial value (red arrow), and MLE estimate with asymptotic variance estimate (blue lines) for Shelf trawl survey (left) and Slope trawl survey (right) catchability in the 2012 RS_1Q candidate model.

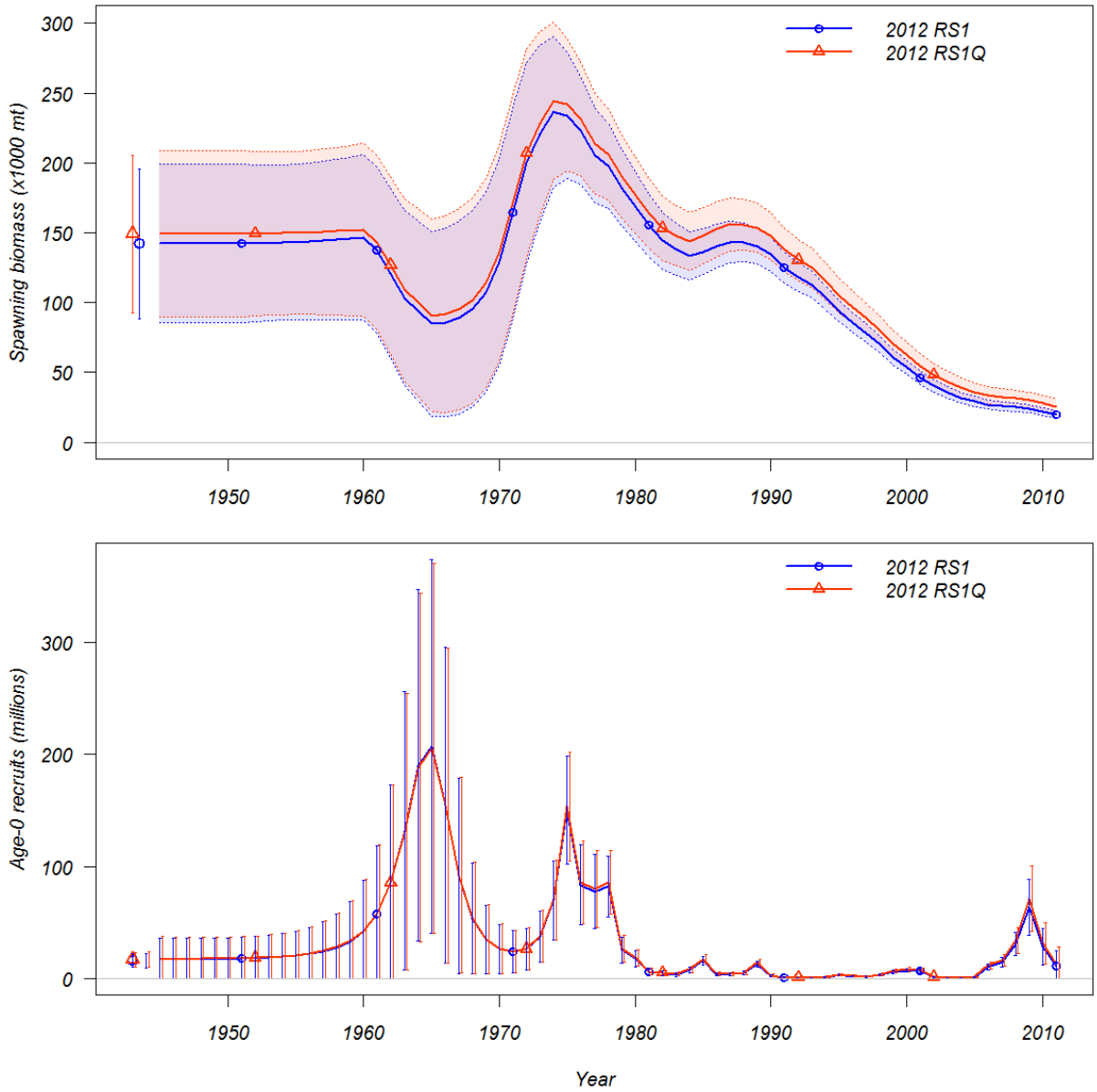


Figure 25 Spawning biomass (top) and age-0 recruitment (bottom) for 2012 Candidate model with only changes to recruitment (2012 RS1) and the 2012 Candidate model with changes to recruitment and catchability fitted for both the Bering Sea slope and shelf trawl surveys (2012 RS1Q) using informative lognormal priors.

Sigma R – Recruitment variability

The 2011 Reference model set recruitment variability (sigma R) of the Beverton-Holt stock recruitment curve to 0.6. Values for sigma R range from 0.15 to 1.0 for stock assessments of this species in the North Atlantic and Arctic Oceans (Myers et al. 1999). We tested the sensitivity of the model to changes in Sigma R with values ranging from 0.6 to 1.69.

Effects of varying sigma R

Increasing Sigma R improved model fit in all categories except size at age (Table 5). A Sigma R of 1.69 would be selected as the most parsimonious using likelihood as a goodness of fit criterion. This value is unreasonably high compared to assessments of this species in other areas and similar species. This high Sigma R is only selected due to the model attempting to create a single large recruitment event (log recruitment deviation > 5) in the 1960's to account for the early catch and large fish in the earliest length composition data where there is no data to direct the model.

Increasing Sigma R causes the model to fit a higher mean recruitment, to compensate for the higher recruitment the model fits a higher catchability for the shelf and slope surveys. Post-1978 this increase in recruitment and increase in catchability results in nearly the same values for the spawning stock biomass (Table 5). The total difference in estimated spawning biomass for 1978 from the Sigma R = 0.5 to 1.69 is +1,200 t or a decrease of < 1%. The total difference in spawning biomass for 2011 from the Sigma R = 0.5 to 1.69 is -3,300 t or a decrease of 12%. Although a single extreme recruitment event may be the most parsimonious model, it is not biologically reasonable. In the author's judgment a value of 0.6 would be most reasonable and consistent with recruitment variability observed in other species with similar life history characteristics.

From Methot (2011), "for each year in the total time series the contribution of that year to the LL is equal to: $dev^2/(2sigmaR^2)$+offset*log(sigmaR); where dev is the recruitment deviation from the expected for that year and where offset is the magnitude of the adjustment between the arithmetic and geometric mean of expected recruitment for that year. With this approach, years with a zero or small offset value do not contribute to the second component." Because of how the recruitment deviation likelihood is specified in SS3, where there are no data to inform the model, the likelihood will always be lower with many small recruitment deviations and a single or a few very large deviation, which equates to a high Sigma R (>1.0), rather than several mid-range deviations with the same mean recruitment and a lower sigma R (<1.0). Therefore for the Greenland turbot model where there are no early data to inform the model on Sigma R, for the early recruitment the model will select the highest Sigma R it can while changing the two survey catchability parameter estimates to keep the biomass estimates consistent with the available data in the later years.

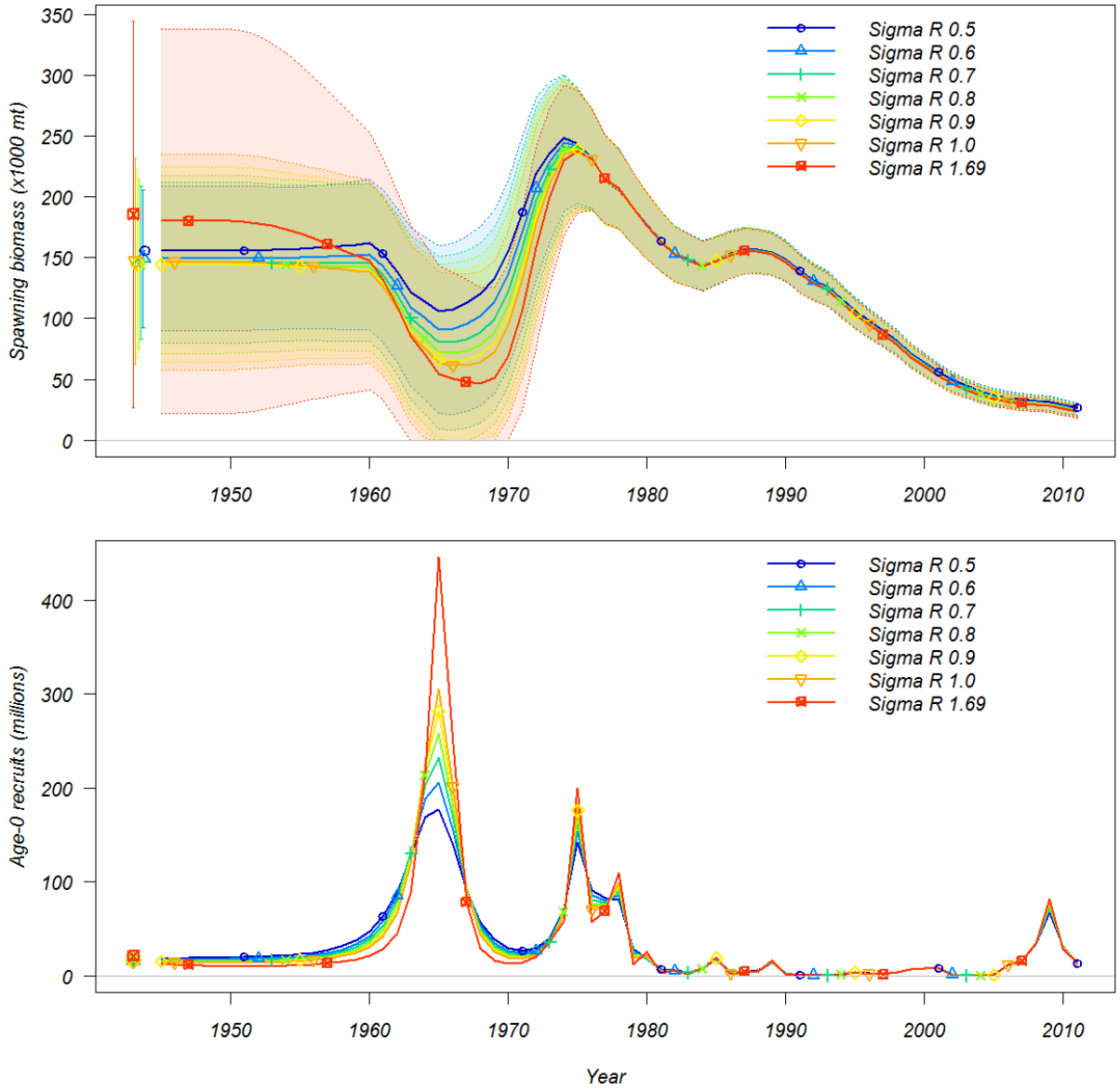


Figure 26 Effects of changing Sigma R on Age-0 recruitment in 2012 candidate model with changes to recruitment and fitting Bering Sea shelf and slope survey catchability (2012 RS_1Q).

Table 5 Effects of changing Sigma R in 2012 Candidate model with changes to recruitment and fitting Bering Sea shelf and slope survey catchability (2012 RS_1Q). The models here with sigma R = 0.6 and sigma R = 0.9 are the same as models 2012 RS_1Q and 2012 RS_1QV in Table 3.

	Sigma R	0.5	0.6	0.7	0.8	0.9	0.999	1.69
Likelihoods								
	Total	2740.95	2708.83	2688.16	2674.39	2664.90	2658.55	2645.68
	Survey	-24.70	-25.67	-26.22	-26.54	-26.75	-26.89	-27.31
	Length Composition	1455.35	1442.19	1433.66	1428.05	1424.02	1421.09	1411.70
	Age Composition	217.13	215.02	213.94	213.23	212.81	212.54	212.04
	Parameter priors	33.80	33.82	33.88	33.98	34.08	34.14	34.35
	Size at Age	967.35	968.63	969.42	969.96	970.33	970.64	971.56
	Recruitment	92.02	74.84	63.49	55.71	50.41	47.03	43.34
Key Parameters								
	SR_LN(R0)	9.77	9.73	9.70	9.70	9.70	9.71	9.94
	Q for Shelf Survey	0.69	0.70	0.71	0.72	0.73	0.73	0.74
	Q for Slope Survey	0.57	0.58	0.59	0.59	0.60	0.60	0.61
	L_at_Amax_Fem	88.50	88.50	88.50	88.50	88.50	88.50	88.51
	L_at_Amax_Mal	72.63	72.61	72.60	72.59	72.59	72.58	72.57
	VonBert_K_Fem	0.13	0.13	0.13	0.12	0.12	0.12	0.12
	VonBert_K_Mal	0.18	0.18	0.18	0.18	0.18	0.18	0.18
Recruits 1960-1969								
Age-0 (1×10^6)	Mean	99.97	104.57	108.11	110.88	113.18	115.11	121.49
	Median	90.04	88.75	86.55	83.95	81.32	78.73	61.07
	CV	0.51	0.59	0.66	0.73	0.79	0.84	1.16
Recruits 1975-2011								
Age-0 (1×10^6)	Mean	19.97	20.11	20.24	20.33	20.41	20.48	20.71
	Median	6.09	5.61	5.32	5.10	4.93	4.88	4.94
	CV	1.62	1.67	1.71	1.75	1.78	1.81	1.92
Spawning Stock Biomass								
(1,000 t)	1978	205.65	206.03	206.20	206.35	206.36	206.46	206.81
	2011	26.53	25.39	24.71	24.29	24.00	23.81	23.24

Authors' note to the NPFMC Groundfish Plan Team

There are two areas where the authors are seeking guidance from the Plan Team for the 2012 Greenland turbot stock assessment. First the authors are seeking the acknowledgment by the plan team that they understand and accept the changes to the underlying data and the effects of these changes on the results of the 2011 Reference model. These include:

1) the new weight to length relationship developed for the 2012 Candidate models,
2) the differences between catch and survey indices in 2011 Stock assessment and those queried for 2012,
3) the difference in how the fishery length composition data were proportioned to haul catch numbers for use in the 2012 assessment as opposed to using the raw composition data,
4) the new method for calculating the sample size for fishery length composition data,
5) the method for tuning the sample size to effective sample size for length composition data.

The changes to the dataset used in the model make substantial changes to the results of the stock assessment (Fig. 27).

Second, the authors are seeking guidance on the exploration of alternative model configurations and on what models the Plan Team would like to see in November.

1) The change in how early recruitments are handled in the proposed 2012 Candidate models is considered by the authors to be an improvement over last year because it does not presuppose a change in productivity in the stock and provides a more accurate representation of the high degree of uncertainty in these early recruitment values. The authors are seeking suggestions by the Plan team for possible alternatives to the recruitment model assessed here.
2) Both selectivity configurations proposed for 2012 provide a better fit to the data than the 2011 Reference model configuration and better capture the differences in selectivity between males and females in the fisheries and surveys. The author's would propose presenting models with both the S_1 and S_2 selectivity configurations for the final stock assessment review in November.
3) Fitting the Slope survey catchability in the Greenland turbot model is problematic as there is little difference in the likelihood for very large differences in catchability. Fixing the values is also problematic because there is no data on how much of the stock is represented by each of the surveys and it likely varies with oceanic conditions and stock size. The stock likely straddles the US-Russian EEZs and the northernmost boundary of the Bering Sea Shelf Survey. The Authors would like to obtain feedback from the plan team on whether they want to see models with fitted catchability given the issues with these fits and if so, whether they believe using a constrained prior is appropriate.
4) Sigma R cannot be fit in the model, should the authors consider alternative values (other than 0.6) for Sigma R in model configurations for November?

References

Cooper, D.W., K.P. Maslenikov, and D.R. Gunderson. 2007. Natural mortality rate, annual fecundity, and maturity at length for Greenland halibut (*Reinhardtius hippoglossoides*) from the northeastern Pacific Ocean. Fishery Bulletin, 105(2): 296-304.

Ianelli, J.N., T.K. Wilderbuer, and D. Nichol. 2011. Stock assessment of Greenland turbot. *In* Stock assessment and fishery evaluation document for groundfish resources in the Bering Sea/Aleutian Islands region as projected for 2000. Section 4. North Pacific Fishery Management Council, Anchorage, AK.

Mertz, G. and R. A. Myers. 1996. Influence of fecundity on recruitment variability of marine fish Can. J. Fish. Aquat. Sci. 53: 1618.1625

Methot, R. D. Jr. 2012. User Manual for Stock Synthesis Model Version 3.23b. NOAA Fisheries Seattle, WA

Myers, R.A., K. G. Bowen, and N.J. Barrowman. 1999. Maximum reproductive rate of fish at low population sizes. Can. J. Fish. Aquat. Sci. 56: 2404–2419

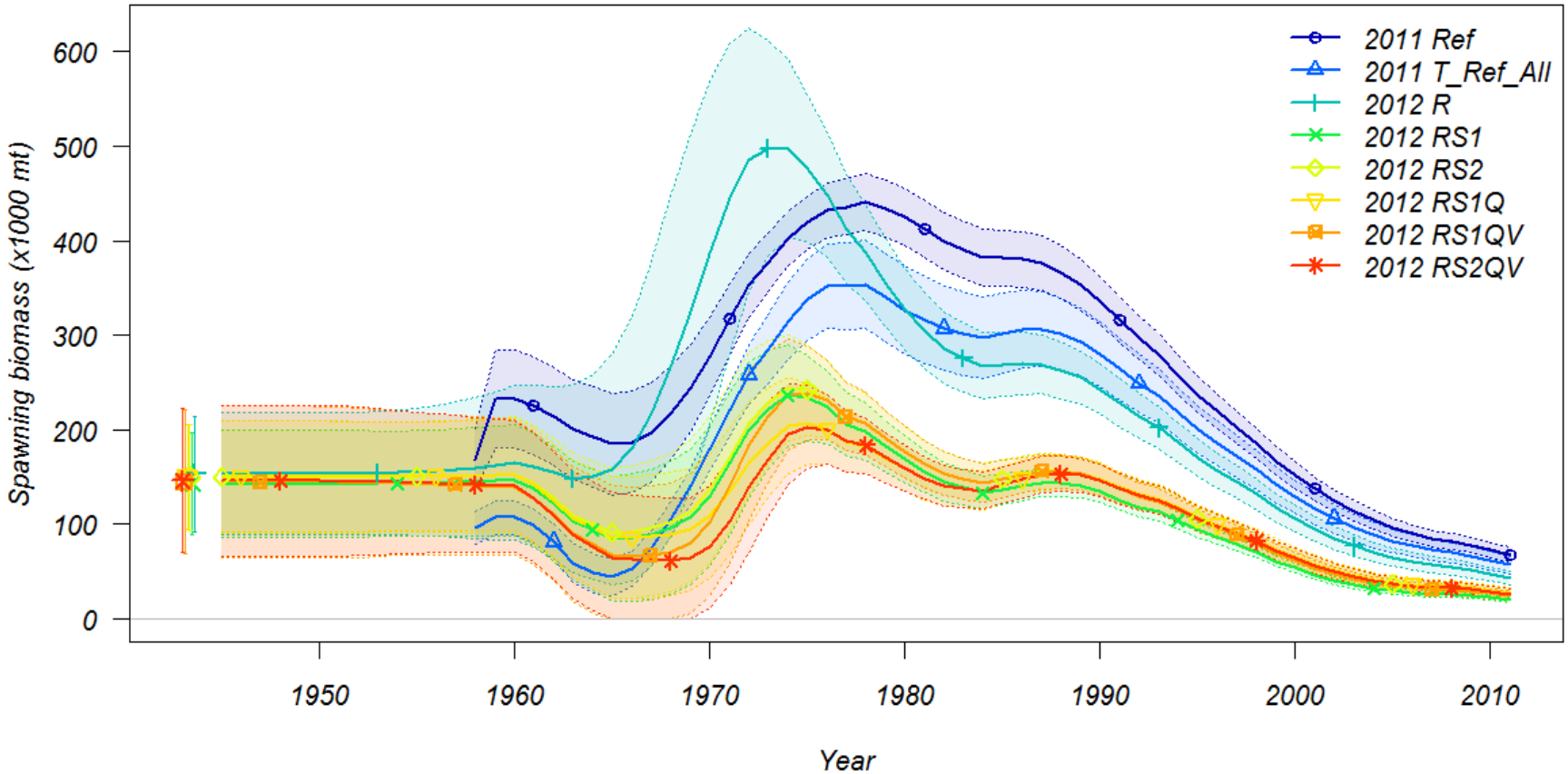
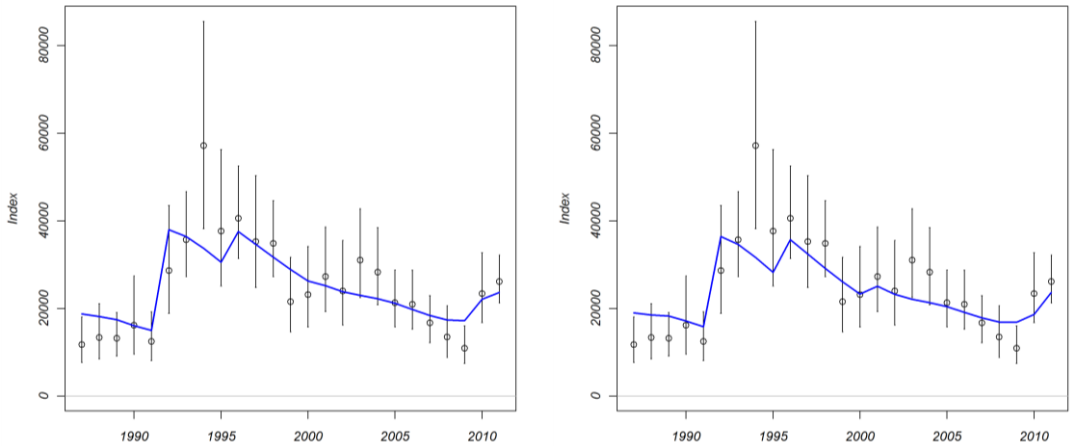


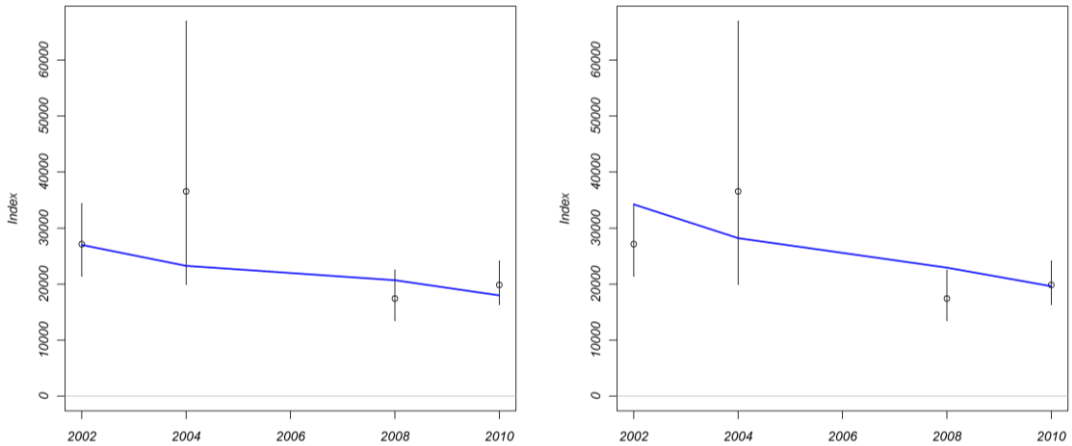
Figure 27 Spawning biomass estimates for the 2011 Reference models and selected 2012 Candidate models.

2011 T_Ref_All 2012 RS_2QV

Bering Sea Shelf Trawl Survey



Bering Sea Slope Trawl Survey



Auke Bay Longline Survey

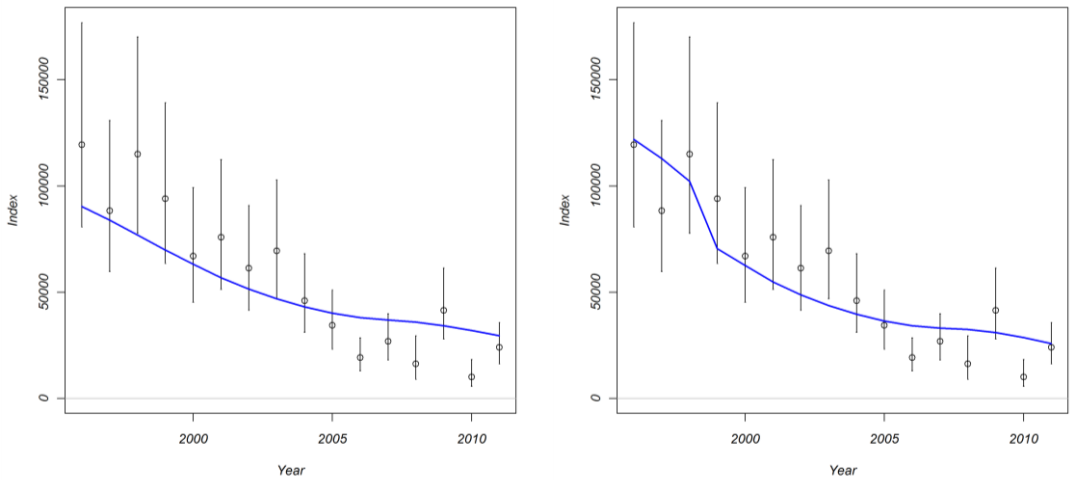


Figure (not referenced in the document) Index surveys and fits for 2011 Reference model with all data changes (2011 T_Ref_All, left) and 2012 RS_2QV candidate model (right).