Report on April 2019 CIE Review for GOA rex, flathead, and Dover sole

Carey McGilliard

Reviewers: Patrick Cordue, Geoff Tingley, and Kurt Trzinski

Chair of review: Jim Ianelli

September 2019 GOA Plan Team Meeting

Terms of Reference

- Evaluation of the ability of the stock assessment model for GOA rex/flathead/Dover sole, with the available data, to provide parameter estimates to assess the current status of rex/flathead/Dover sole in the Gulf of Alaska

- Evaluation of the strengths and weaknesses in the stock assessment model for GOA rex/flathead/Dover sole

- Recommendations for improvements to the assessment model.

Quick review of the three assessments

- All transitioned from “roll-your-own” age-structured models to Stock Synthesis (flathead and Dover in 2013, rex in 2015) by way of matching exercises, each presented at September PT meetings

- In transition to Stock Synthesis, the following change were made to all:
  - Estimation of growth within the assessment model
  - Models start at age 0
  - Timing of survey and fishery refined within the model
  - Use of # of hauls as input sample size to length comp data
  - Use of McAllister-Ianelli data weighting approach (but each attempted Francis data weighting at some point post-2013)
  - Catchability fixed at 1
  - All use 1984 and 1987 survey data
  - All estimate early recruitment deviations

Unique aspects of each assessment

- Flathead: not super unique
- Rex:
  - Distinct spatial growth pattern (Western-Central fish get bigger than Eastern fish)
  - Two-area model with growth estimated in each area to account for spatial growth pattern
  - The two area model, along with newly aged fishery ages brought rex sole from Tier 5 to Tier 3 in 2017, as this assessment resolved a major uncertainty in fishery reference points
  - Fishery age data is a combo of haul and port data. No lat lon/haul info for port data, included after an analysis looking at whether age and length data come from same areas and seasons as for catches

Unique aspects of each assessment

- Dover:
  - Ontogenetic movement suggested by data, where all fish recruit inshore and only old fish appear in waters >500m
  - Data split by years the survey sampled to 500m, or beyond 500m such that separate selectivity patterns can be estimated for these different year-survey-depth blocks
  - The random effects model is used to fill in depth-area gaps in the survey biomass index, which is then associated with the full-coverage comp data
  - Hard to age, so 1990 excluded due to biased surface ageing method and ageing imprecision is incorporated (using West Coast estimates of imprecision for Dover)
  - Old cohorts did not grow as big as young cohorts already are

Data sources for each assessment

Source	Type	Years
Fishery	Catch biomass	1978-2013
Fishery	Catch length composition	1989-1999, 2001-2007, 2009-2013
GOA survey bottom trawl	Catch per unit effort	Triennial: 1984-1999, Biennial: 2001-2013
GOA survey bottom trawl	Catch length composition	Triennial: 1984-1999, Biennial: 2001-2013
GOA survey bottom trawl	Catch age composition, conditioned on length	Triennial: 1984-1999, Biennial: 2001-2013

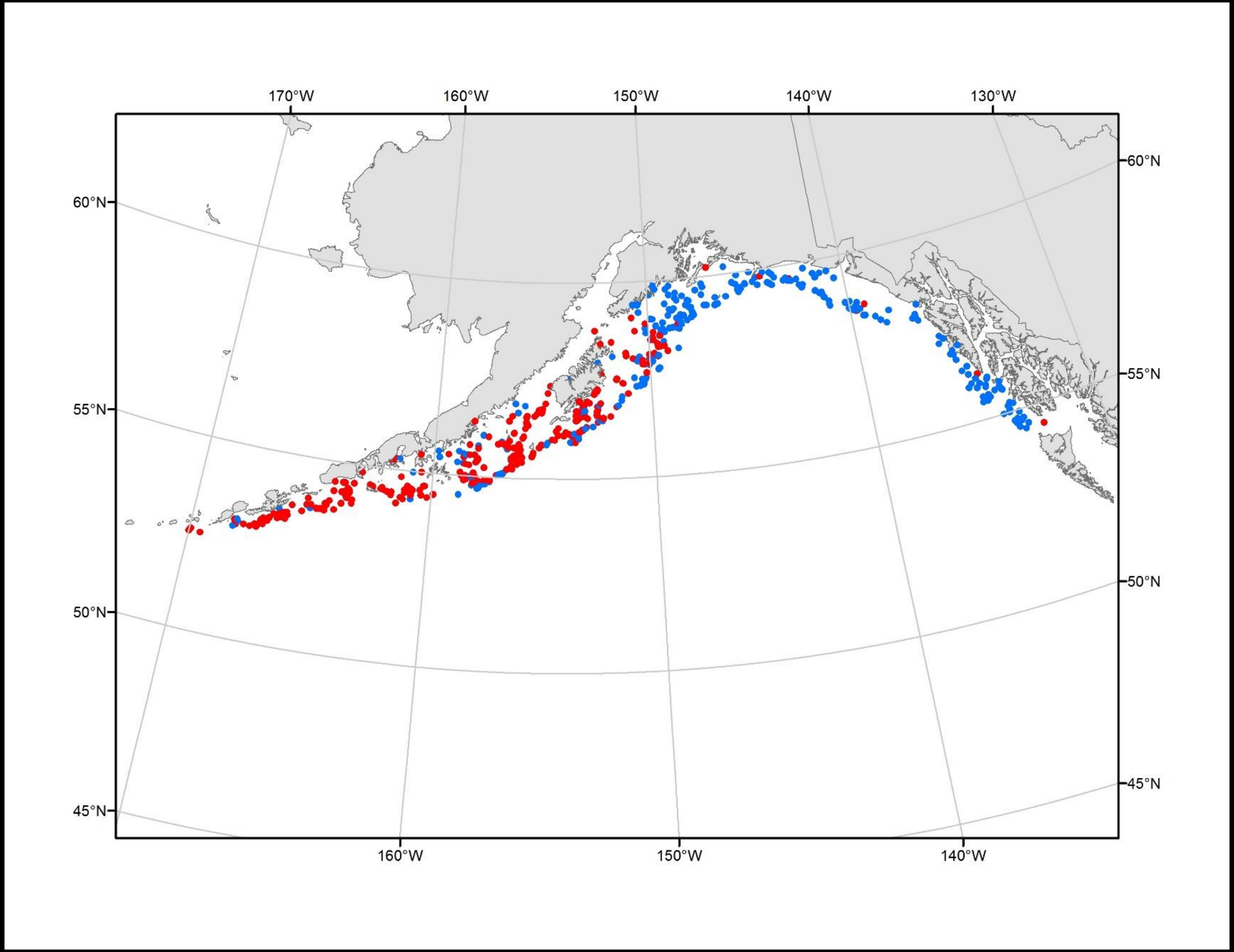
- Rex: + Fishery Age data:
  - 1992,1995,1999,2003,2005,2007,2009,2010, 2012 2014-2016
- Dover: -1990 age data (biased surface ageing)

Took a look at growth for all three species (given the findings for rex in 2017) prior to this review

GOA **rex sole** residuals from sex-specific von-Bertalanffy models fit to survey data 2001-2015 outside the assessment model.

The blue points are more than 1 residual standard error below the curve and the red points are more than 1 RSE above the curve.

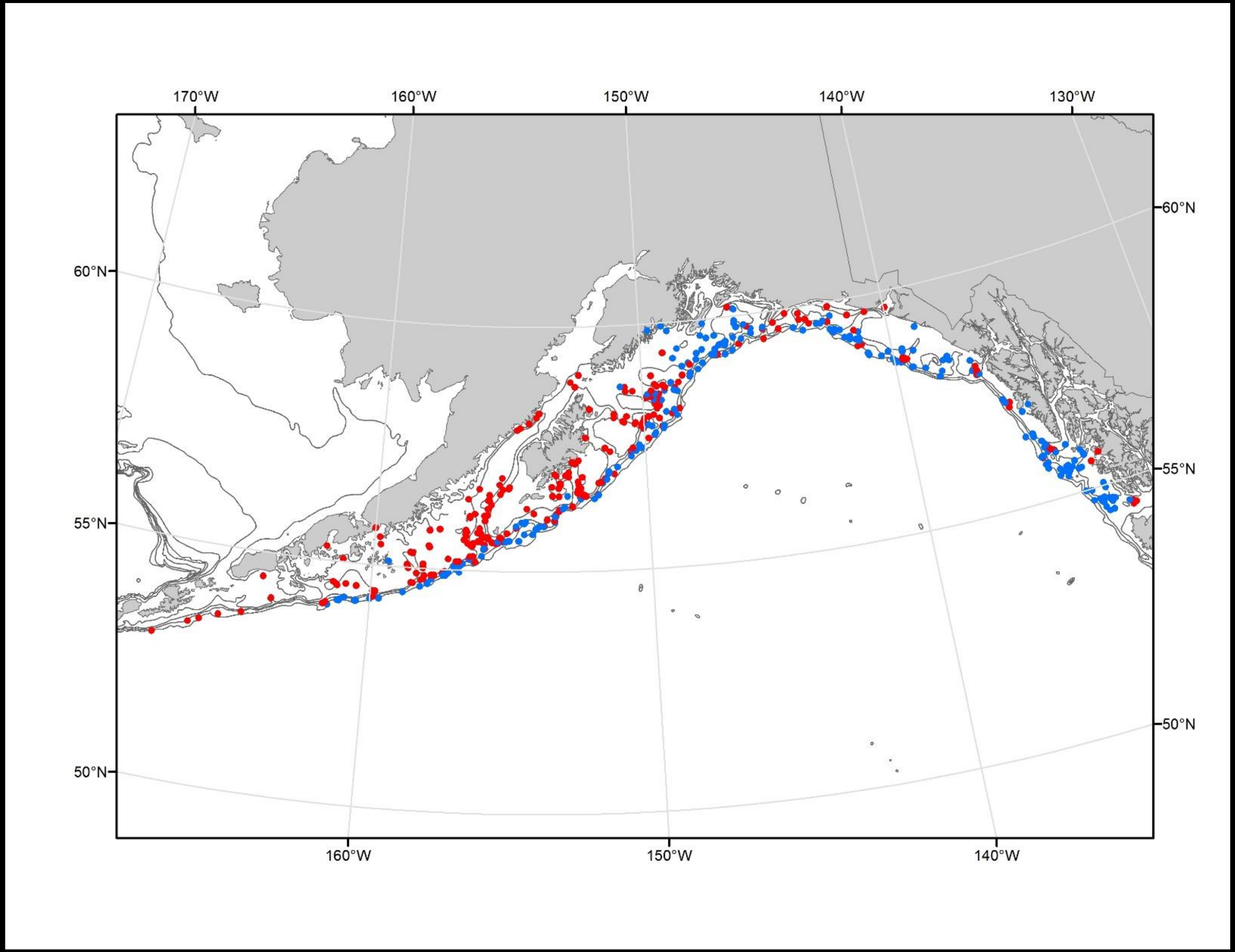
(courtesy of Beth Matta)



GOA **Dover sole** residuals from sex-specific von-Bertalanffy models fit to survey data 2001-2015 outside the assessment model.

The blue points are more than 1 residual standard error below the curve and the red points are more than 1 RSE above the curve.

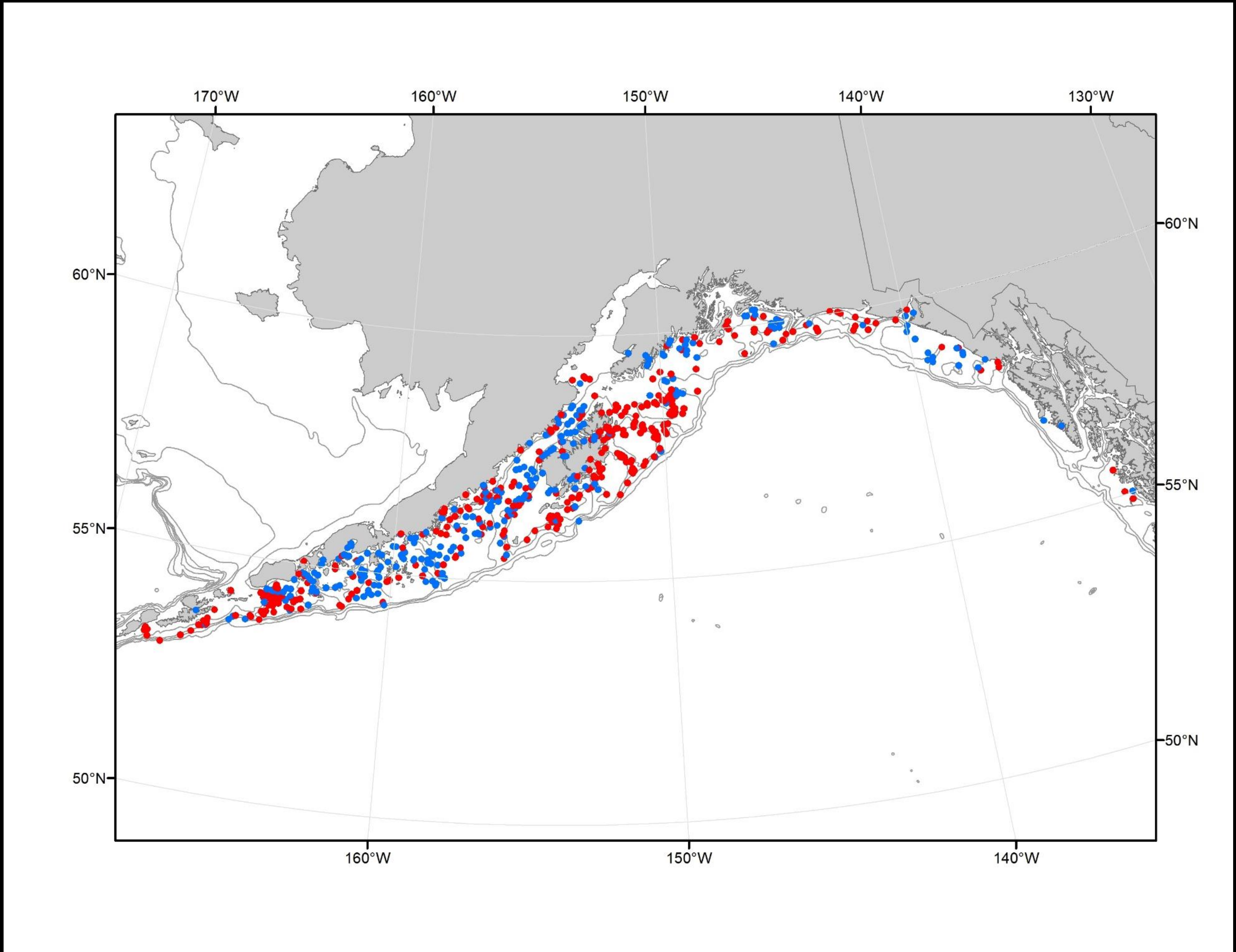
(courtesy of Beth Matta)



GOA **flathead sole** residuals from sex-specific von-Bertalanffy models fit to survey data 2001-2015 outside the assessment model.

The blue points are more than 1 residual standard error below the curve and the red points are more than 1 RSE above the curve.

(courtesy of Beth Matta)



General Comments from CIE Reviewers:

All reviewers agreed that the assessments were appropriate for use in management

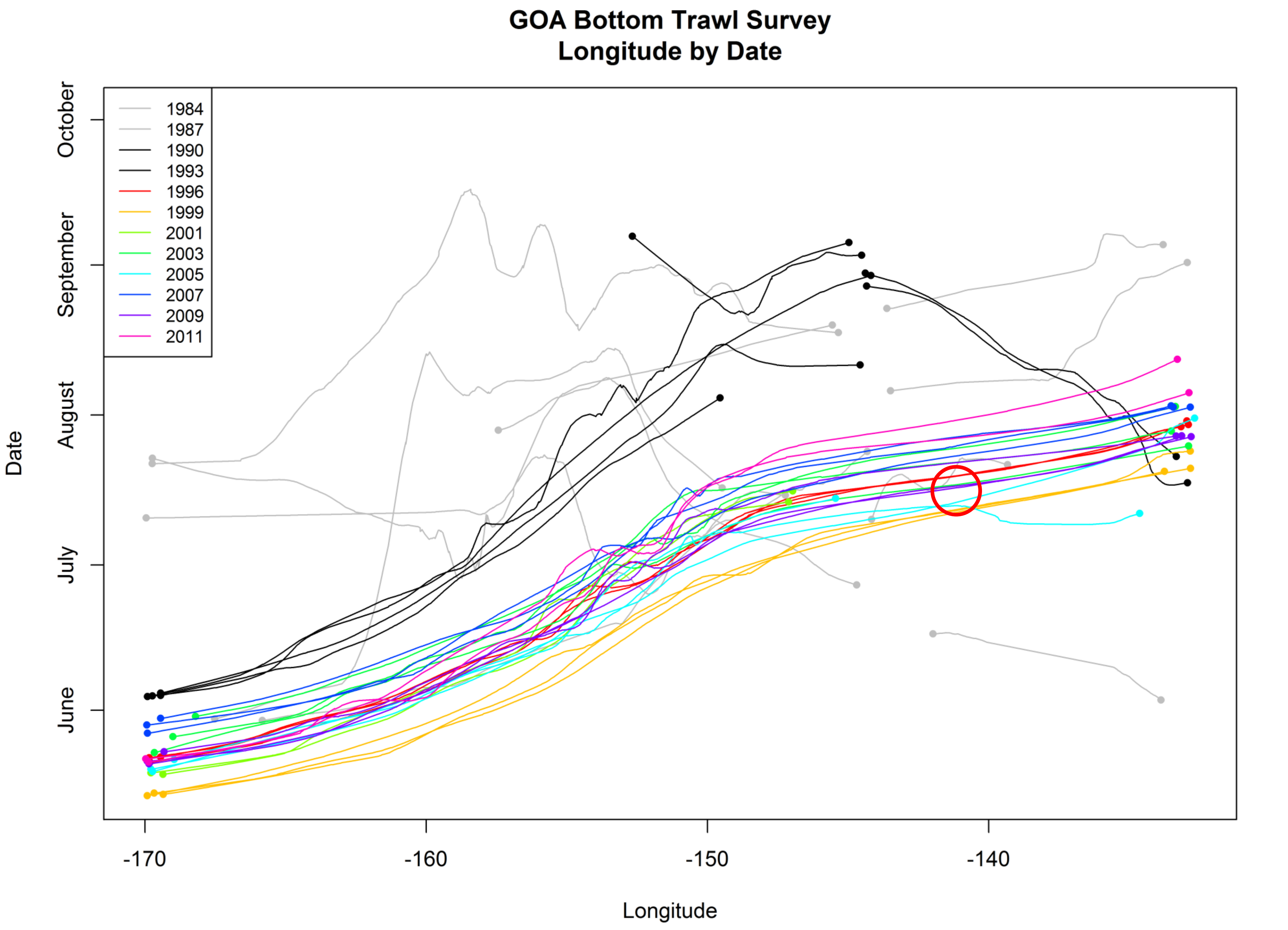
- "In general, the age-structured models were appropriate given the available biological, abundance, and composition data. A particular strength of the assessments is the availability of a consistent timeseries of biomass estimates from the GOA trawl surveys (in particular since 1996)."
- "The reviewers appreciated the excellent presentations by the NMFS staff, the hard work of the assessment author, and the collegial and constructive atmosphere under which the review meeting was conducted"

Overview of main CIE Reviewer recommendations and concerns

- Use of the 1984 and 1987 (and 1990 and 1993) survey data
- Estimate survey catchability
- Observer program sampling design does not sample some species (like Dover) very well
- Post-stratification of fishery length comp data (rex), leave out port data
- Revisiting whether early recruitment deviations should be estimated and how many
- Dover: one reviewer did not like the idea of using the random effects model to estimate biomass index depth-area gaps and variability, would rather split into 3 biomass indices and estimate catchability for each

CIE Reviewer Recommendations and Concerns

- "Common to all three assessments is the issue of the use of the 1980s trawl surveys and the use of the 1990 and 1993 surveys. The 1980s surveys should not be used (non-standard vessels and gear); the 1990 and 1993 surveys should probably be used in the base model, but a sensitivity should also be done which excludes them (they were conducted later in the year than the surveys since 1996)."



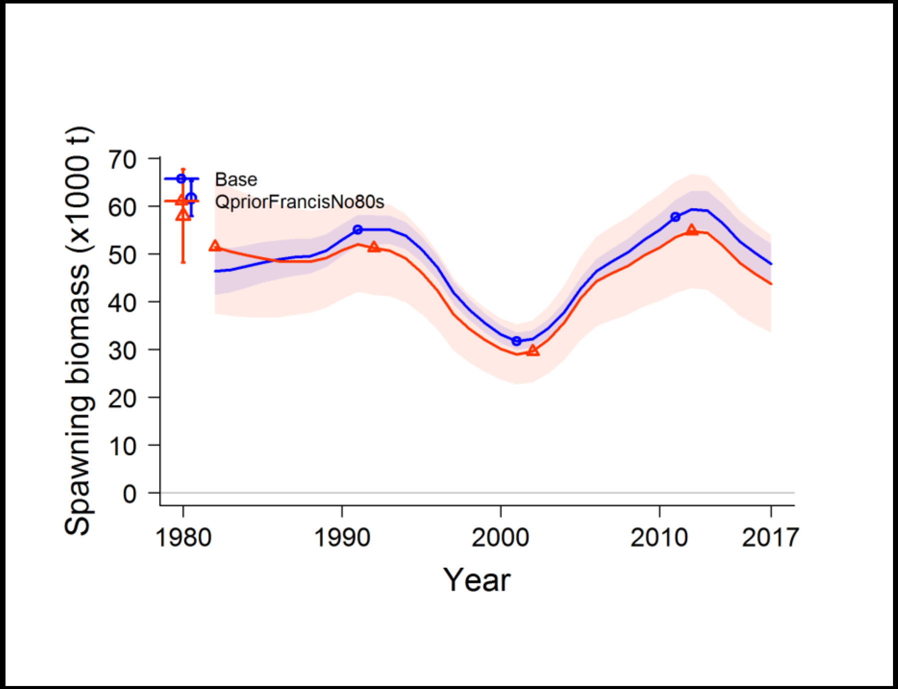
Also, 30 minute tows in 1984 and 1987, while more recent years: 15 min tows

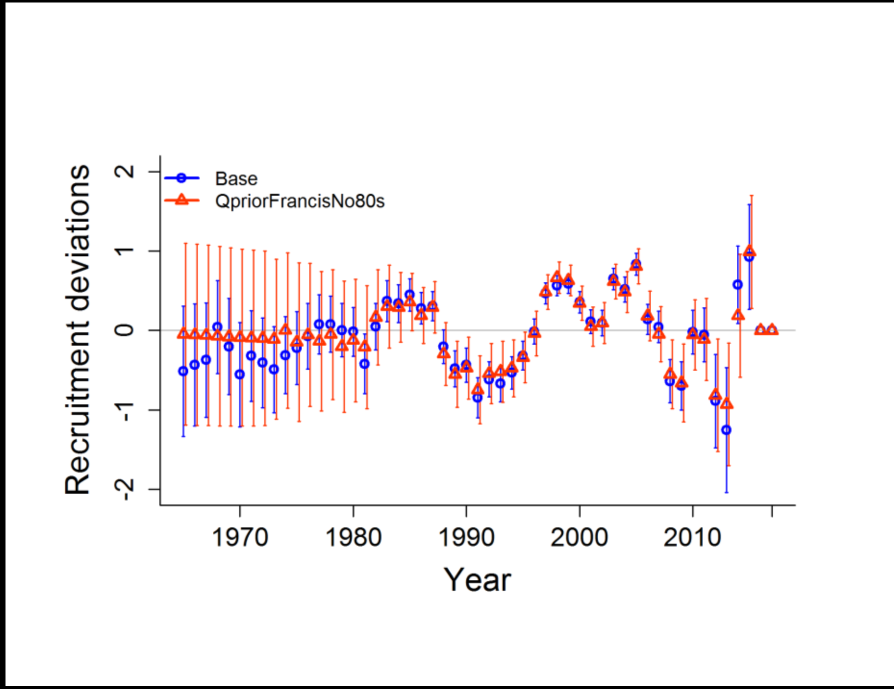
CIE Reviewer Recommendations and Concerns

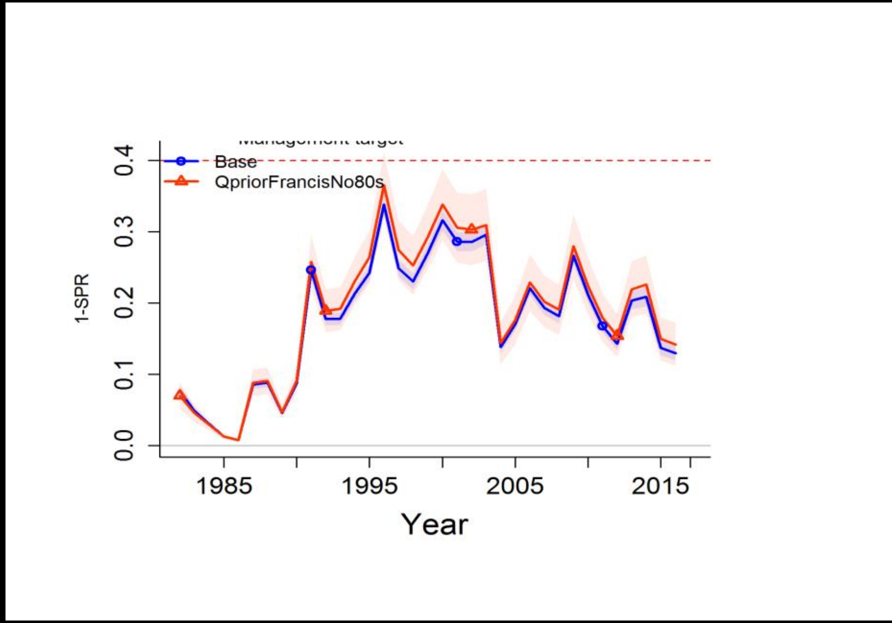
Reviewers requested runs for rex and flathead that made the following changes

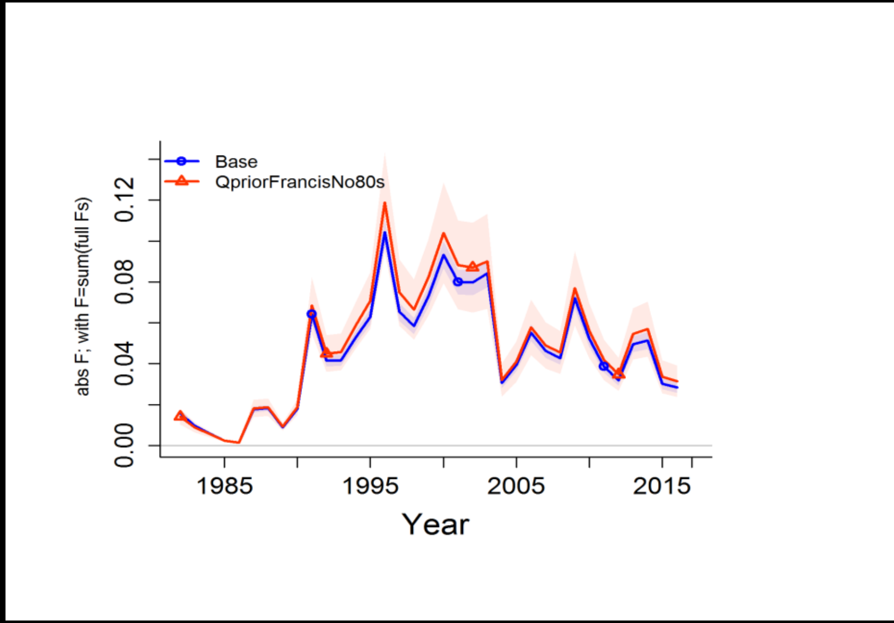
- Put a prior on catchability (normal prior was chosen)
- Conduct Francis data-weighting
- Remove 1984 and 1987 data
- Reduce the number of years of early recruitment deviations (flathead)

Rex: Comparing base case 2017 to a run with a normal prior on q, francis re-weighting, and removing 1984, 1987 data

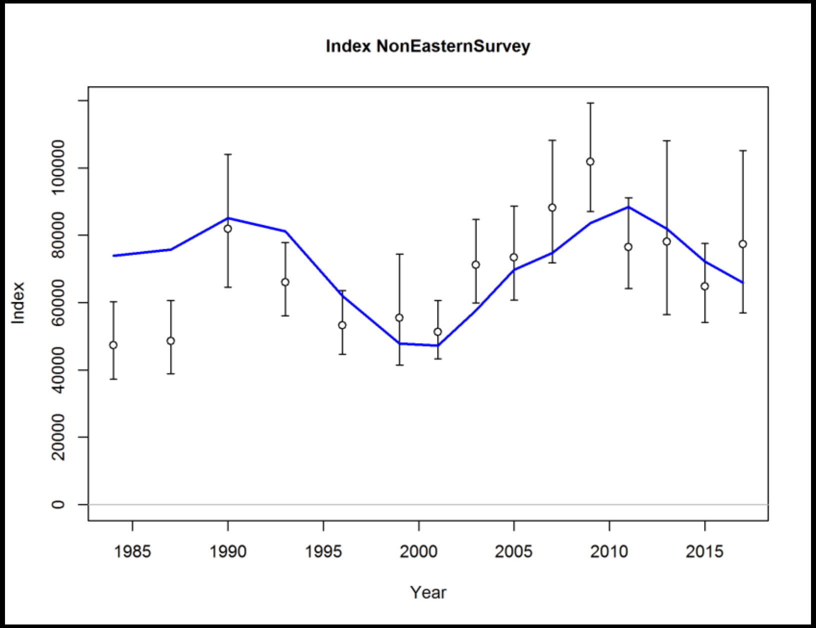


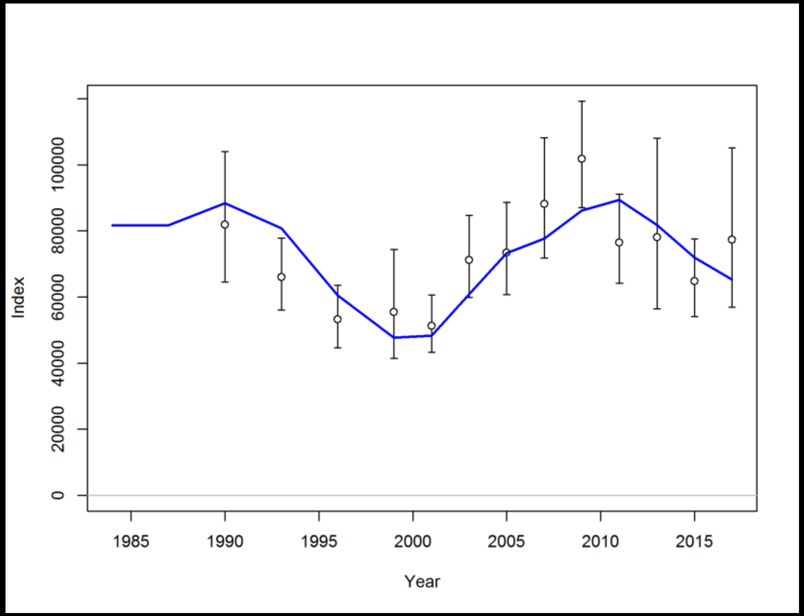


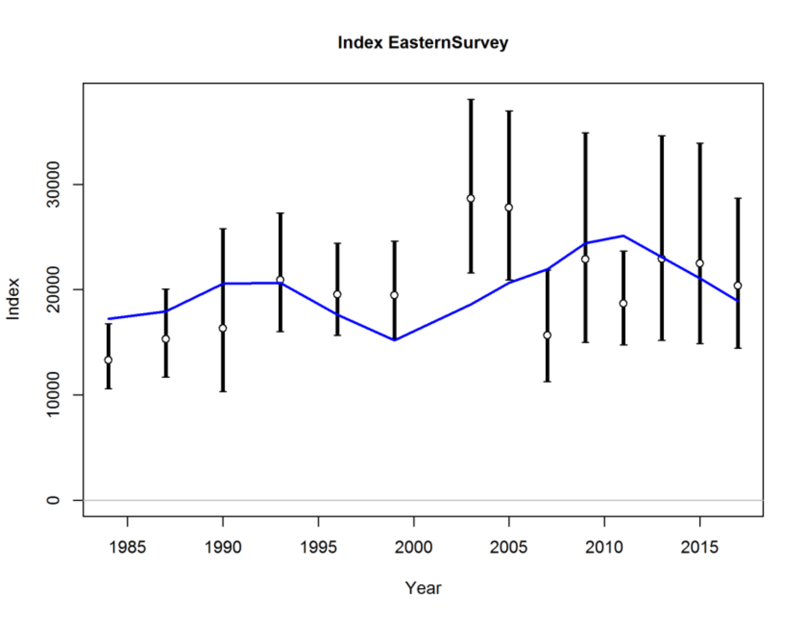


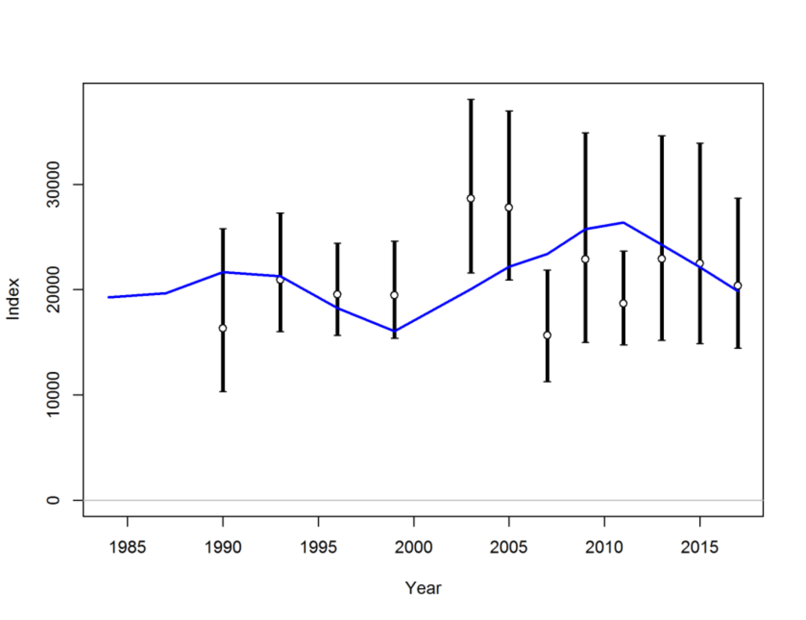


Rex: Comparing base case 2017 to a run with a normal prior on q, francis re-weighting, and removing 1984, 1987 data



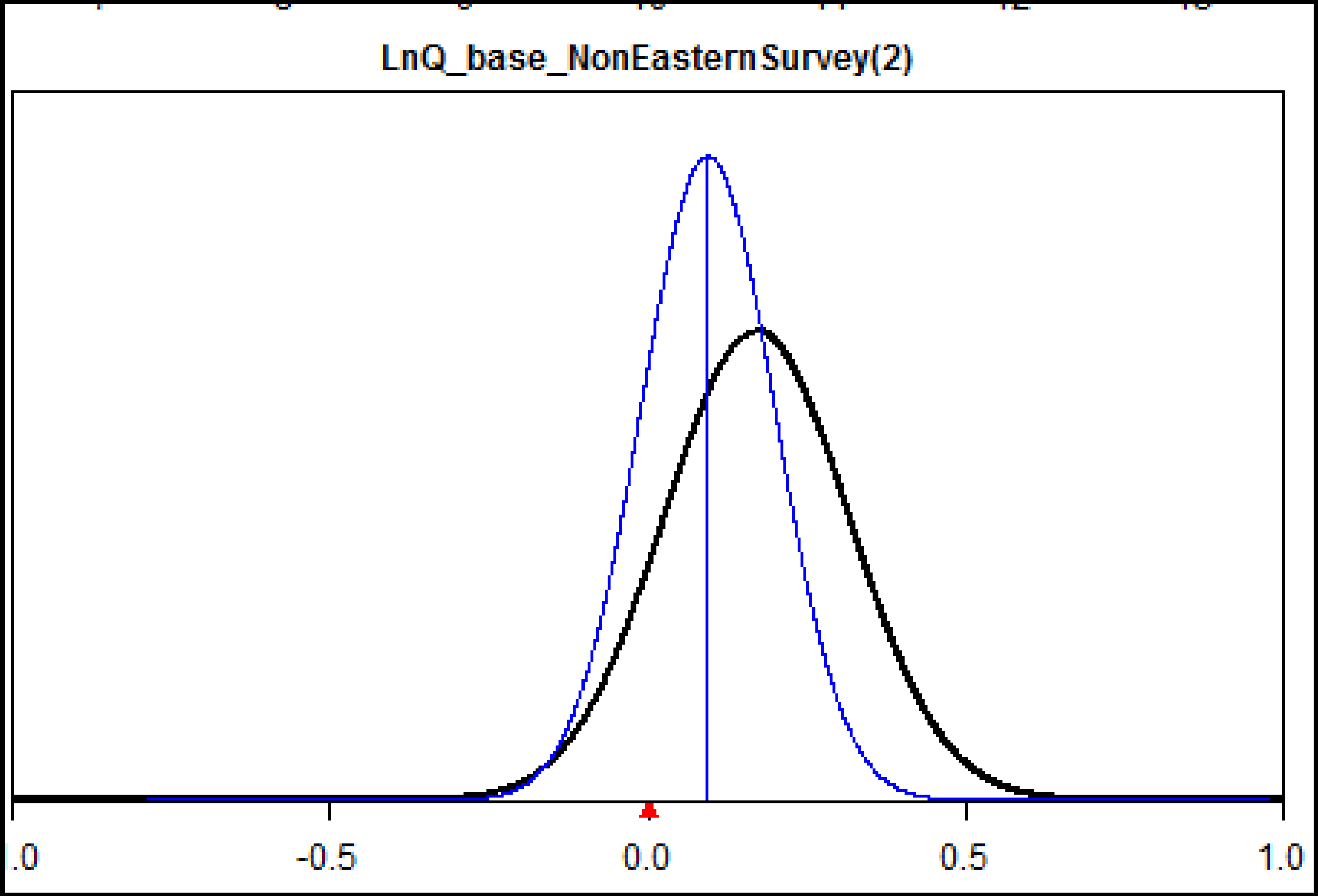






Rex:
Comparing base case 2017 to a run with a normal prior on q, francis re-weighting, and removing 1984, 1987 data

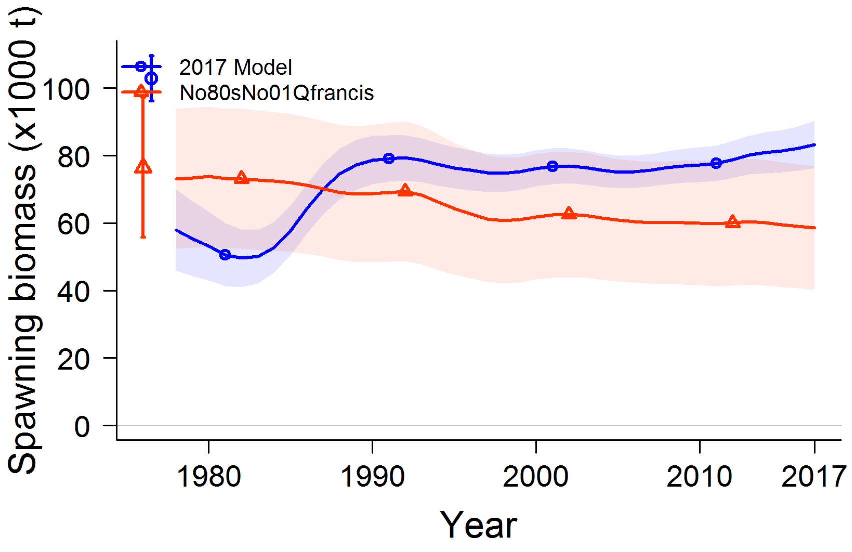
Prior in black, red triangle = initial value

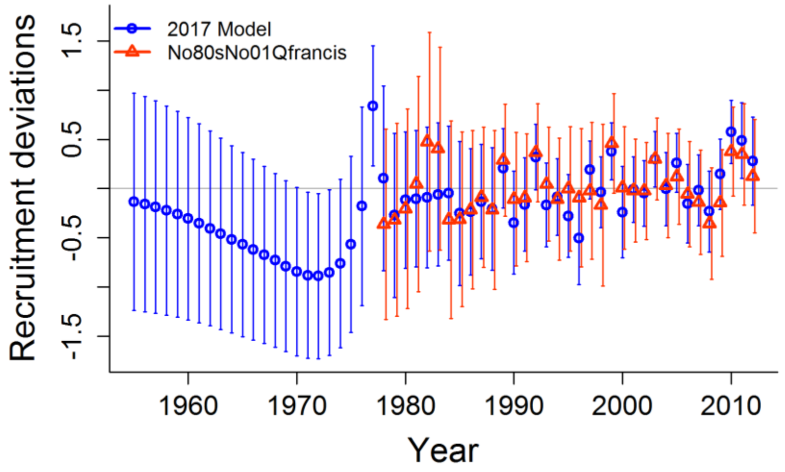


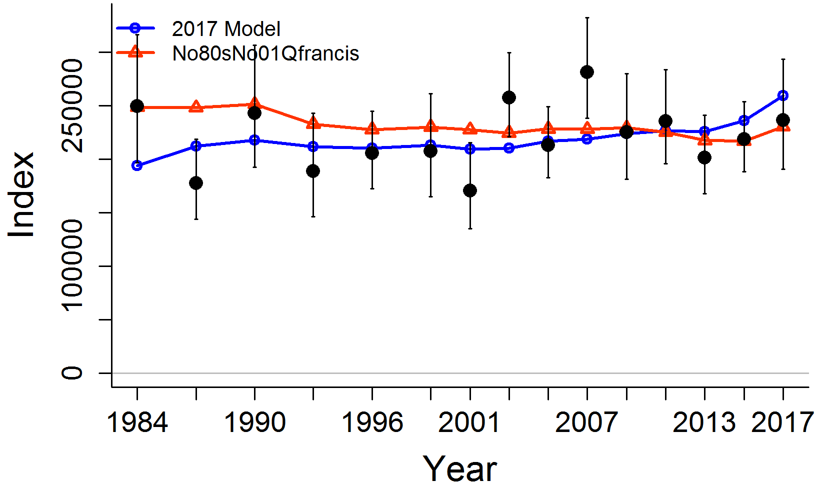
Flathead sole reviewer requested run

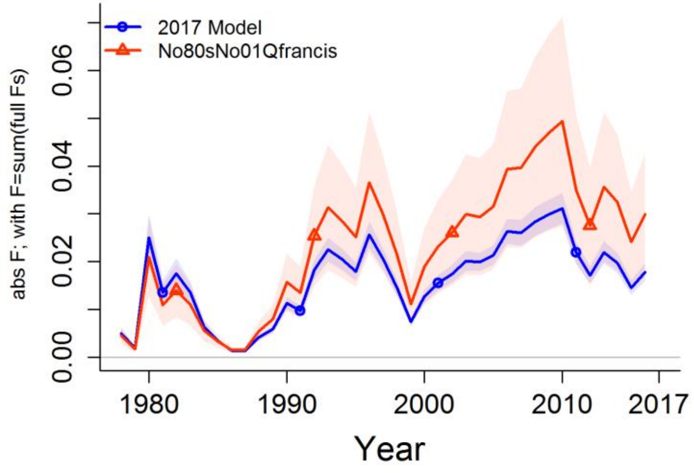
- Removed 1980s data and 2001 data
- Used a normal prior on q with mean = 1.2, SD = 0.175
- No early period recruitment:
  - Numbers drop off quickly around/after age 20
  - 3 years old before they are observed
  - First age data in 1990
  - Want to have observed them 5 times before including as a rec dev
  - 20-3-5 =12
  - 1990 – 12 = 1978 (model start year)
- Francis re-weighting (keeping input sample size = haul size)

New Flathead Run Compared to 2017 Model

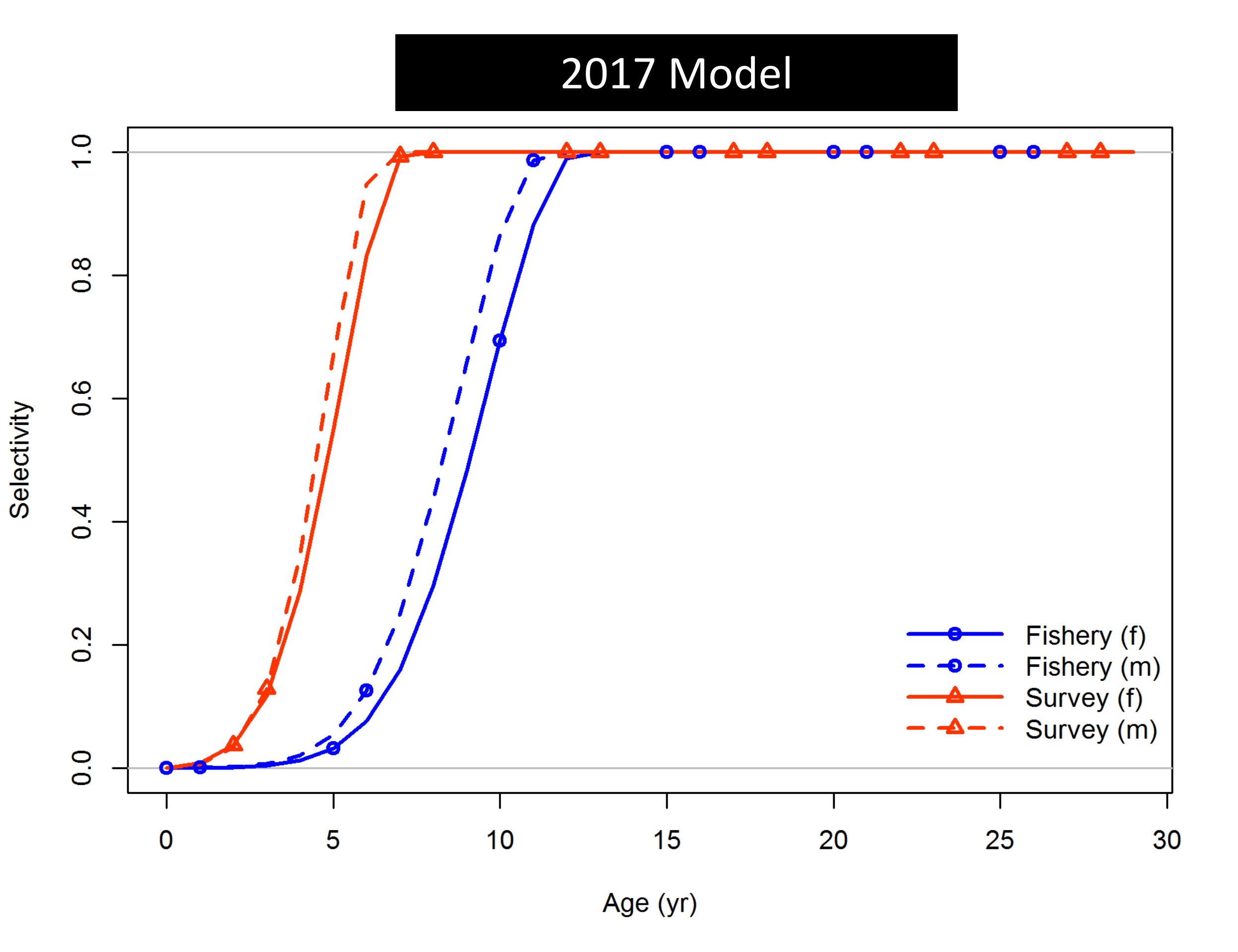


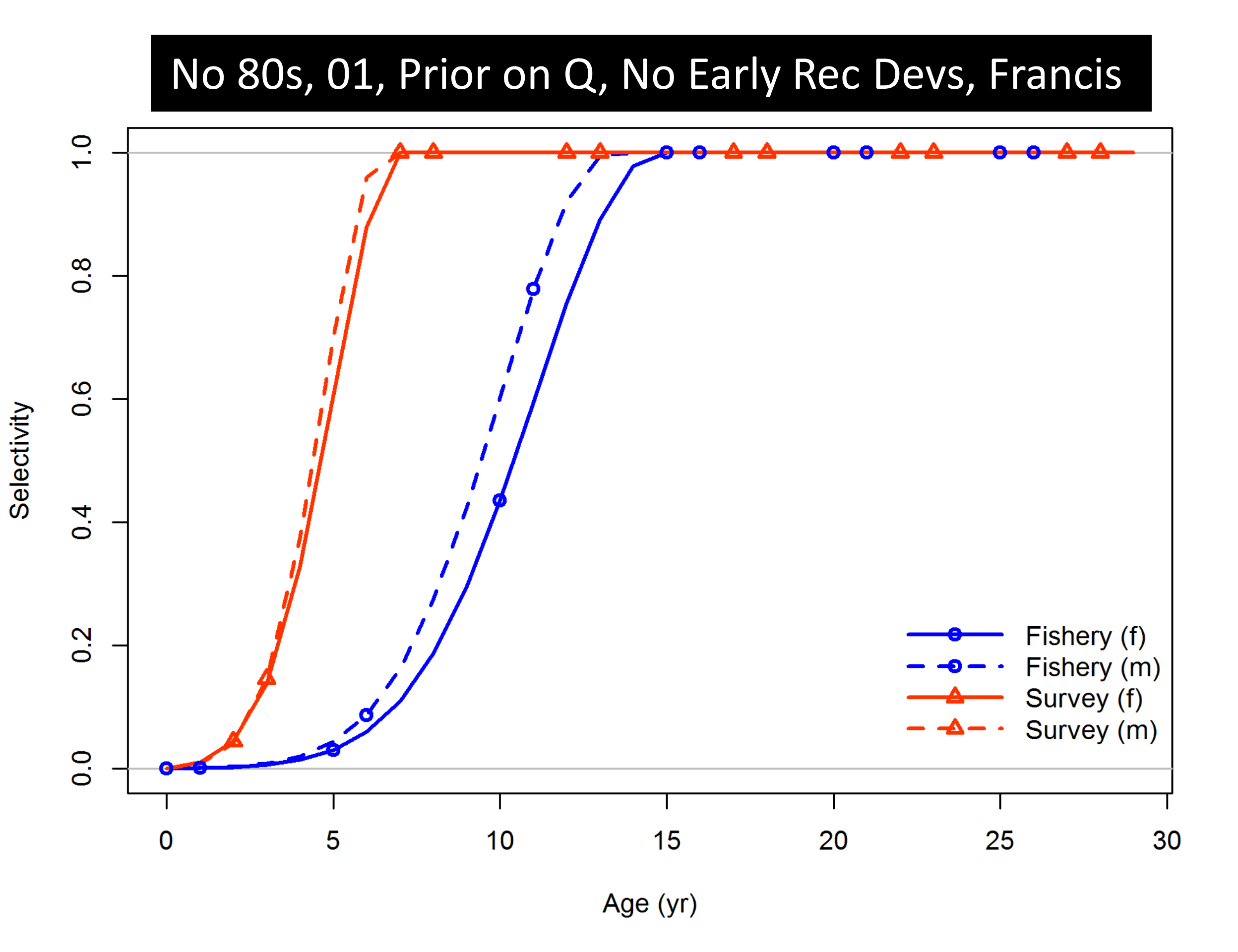




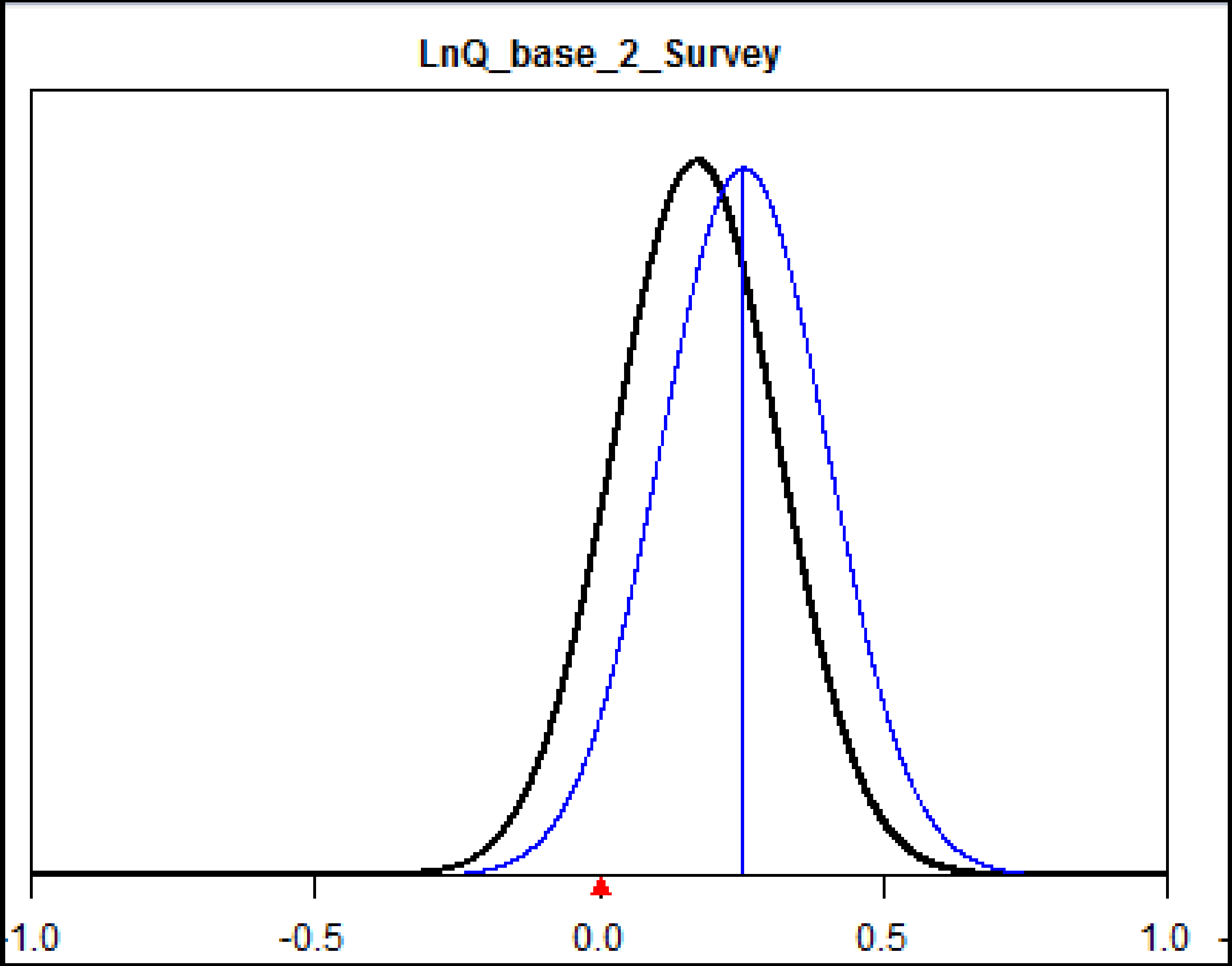


Selectivity Curves

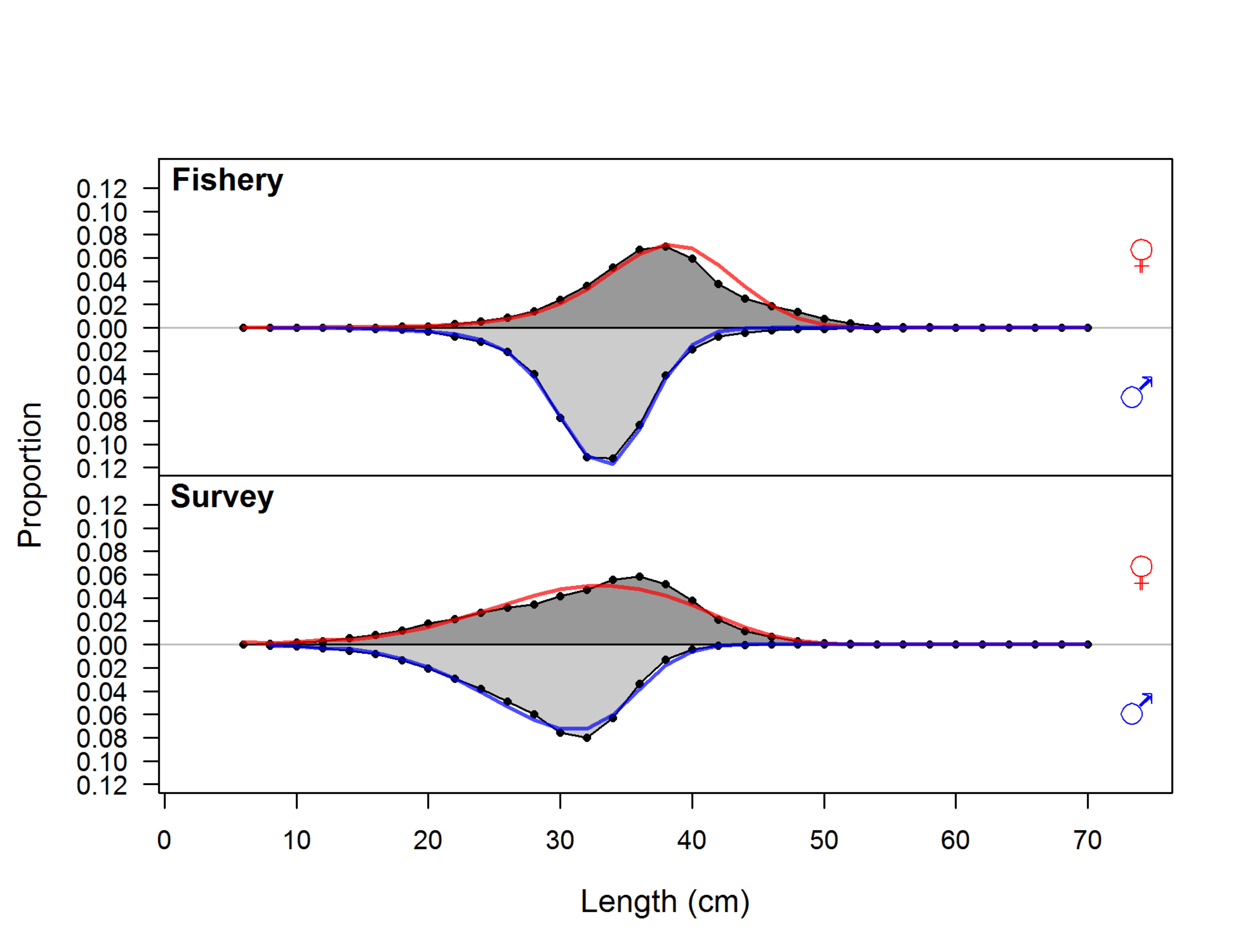




Reviewer requested run: q estimate



Reviewer Run



Revisiting: Challenges for the Dover assessment

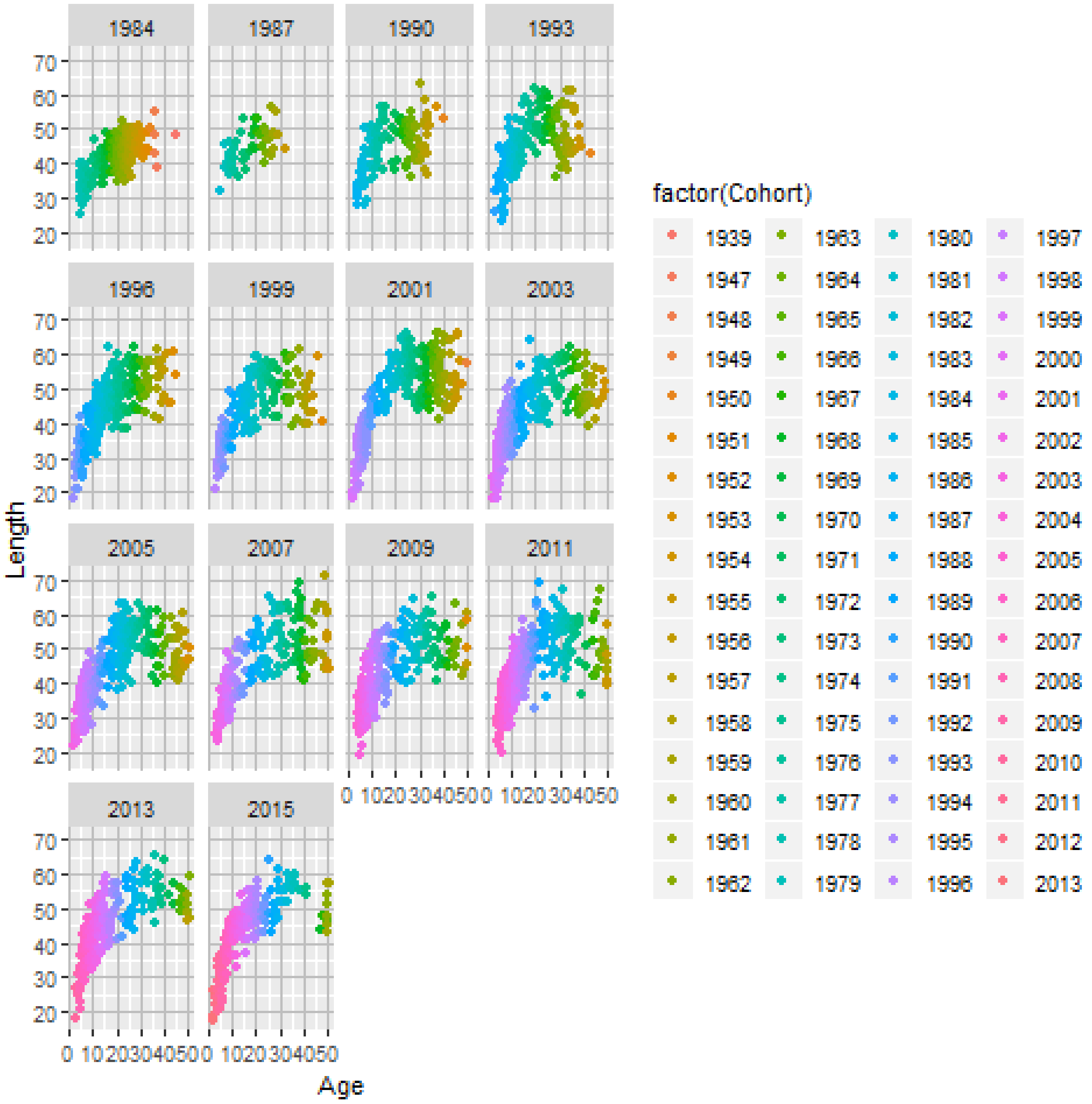
- Dover sole are long-lived and hard to age, especially as they get older
- There is a lot of variability in length-at-age
- The length-at-age relationship appears to have shifted over time, potentially, though there are other hypotheses as well

- Dover move ontogenetically from shallow to deep depths (up to 1,500 m) as they get older
  - The survey covers depths to 1,000m in some years and only 500 or 700m in other years
  - Prior to 2015, the survey was done on 2 boats (labeled “shallow” and “deep,” but ages were only collected on one boat (“shallow”)

- The fishery for Dover is very small (~3% of the catch limit is caught on average) and there are no fishery age data

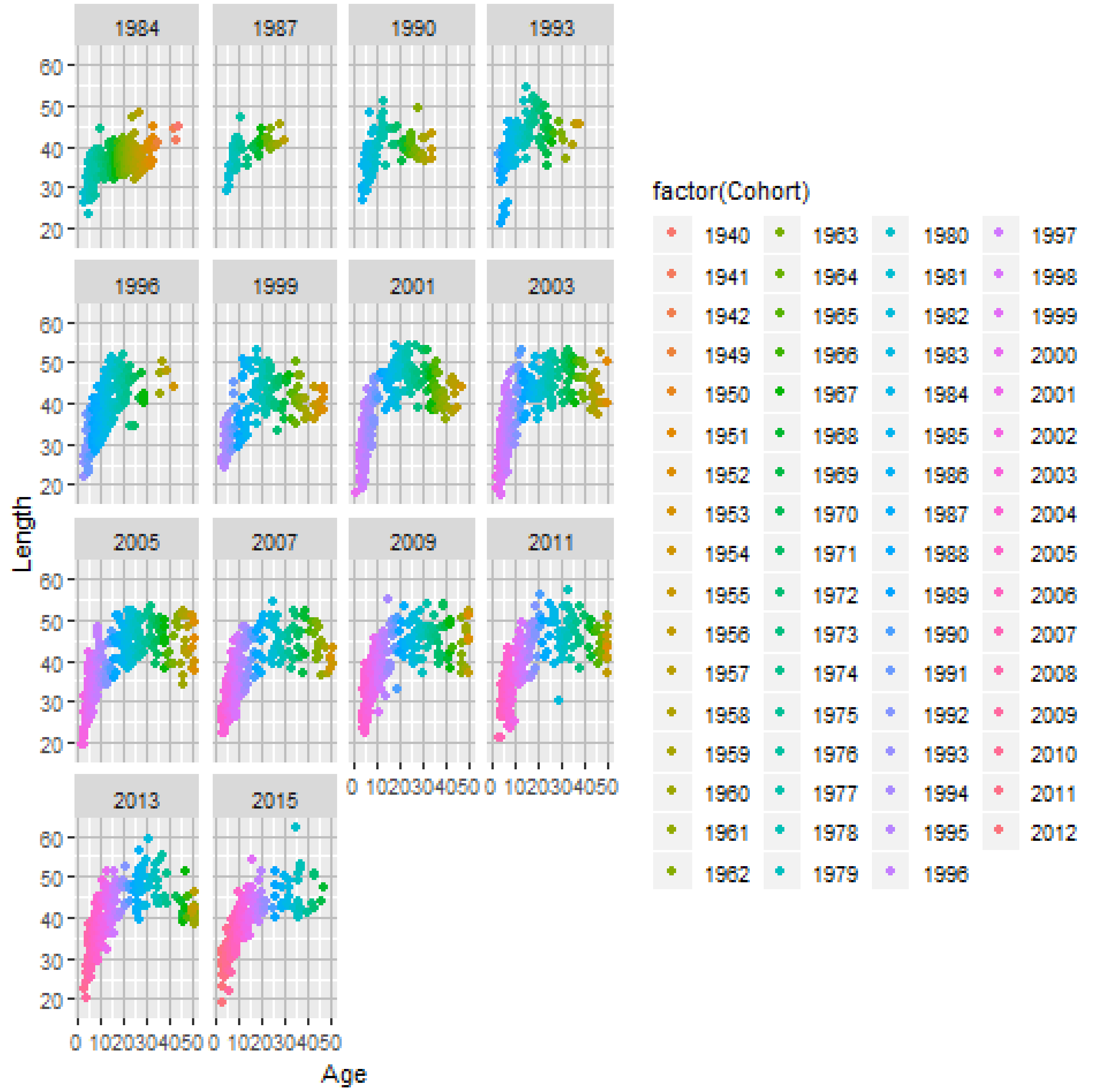
Female length-at-age by cohort and year

- High variation in length-at-age, especially at older ages
- Early cohorts were not as big at older ages as later cohorts are in middle-age
- Length-stratified data except for 2015 (which was random)
- Dover are hard to age, especially at older ages



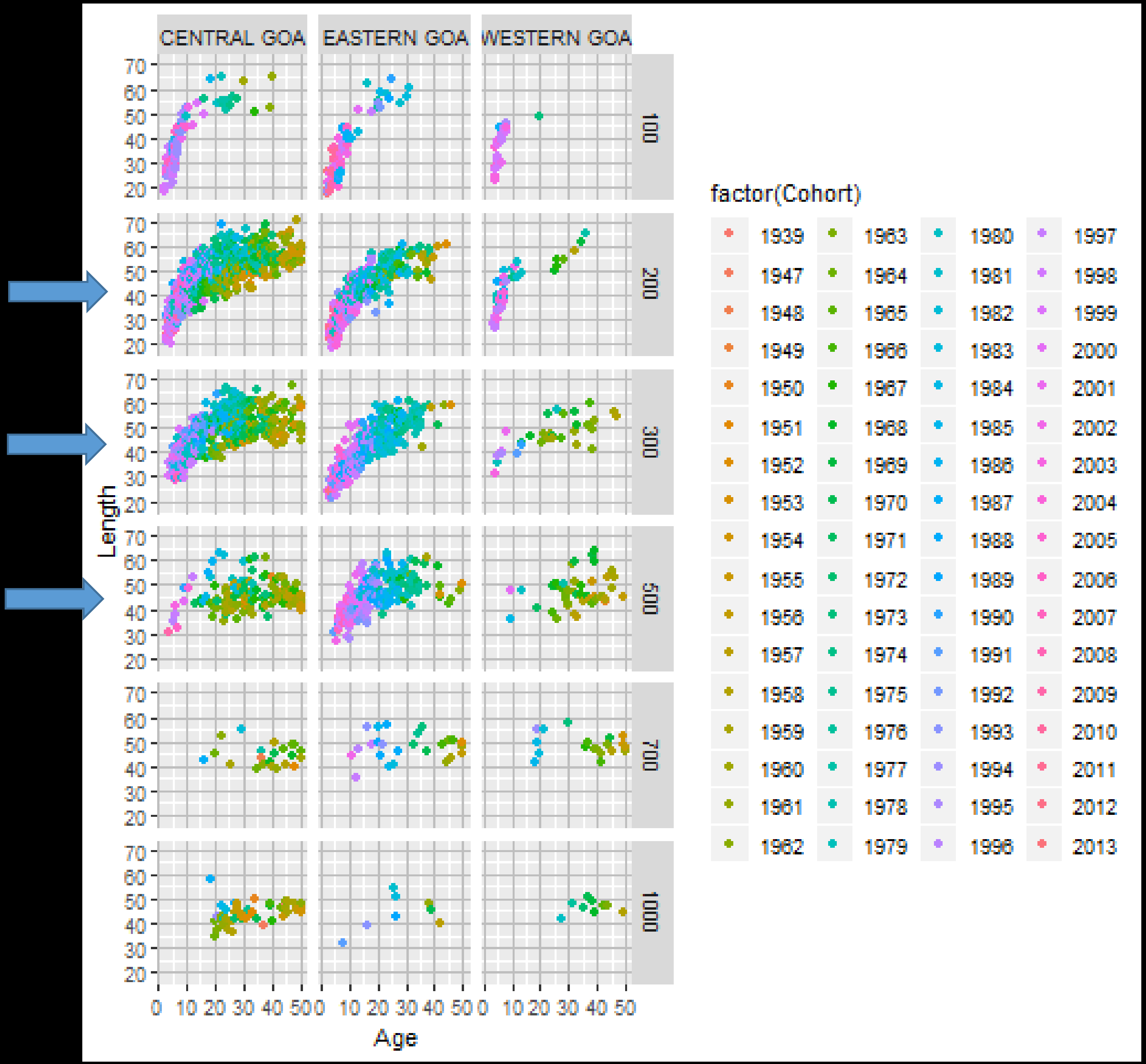
Male length-at-age by cohort and year

- High variation in length-at-age, especially at older ages
- Early cohorts were not as big at older ages as later cohorts are in middle-age
- Length-stratified data except for 2015 (which was random)
- Dover are hard to age, especially at older ages



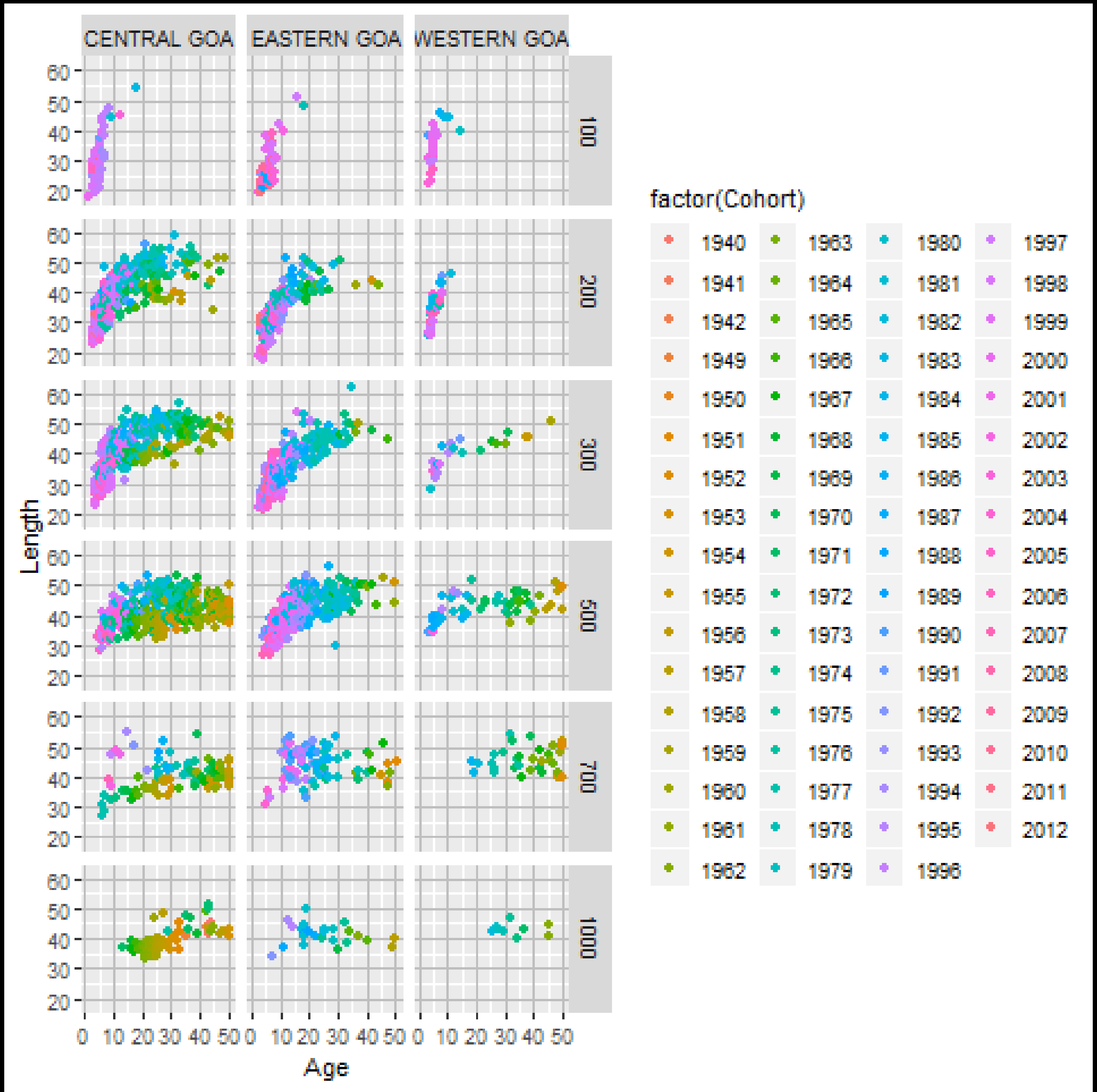
Female length-at-age by cohort, area, and depth

- High variation in length-at-age, especially at older ages
- Early cohorts were not as big at older ages as later cohorts are in middle-age
- Length-stratified data except for 2015 (which was random)
- Dover are hard to age, especially at older ages
- Few samples in Western GOA
- Few fish above 40 in Eastern GOA



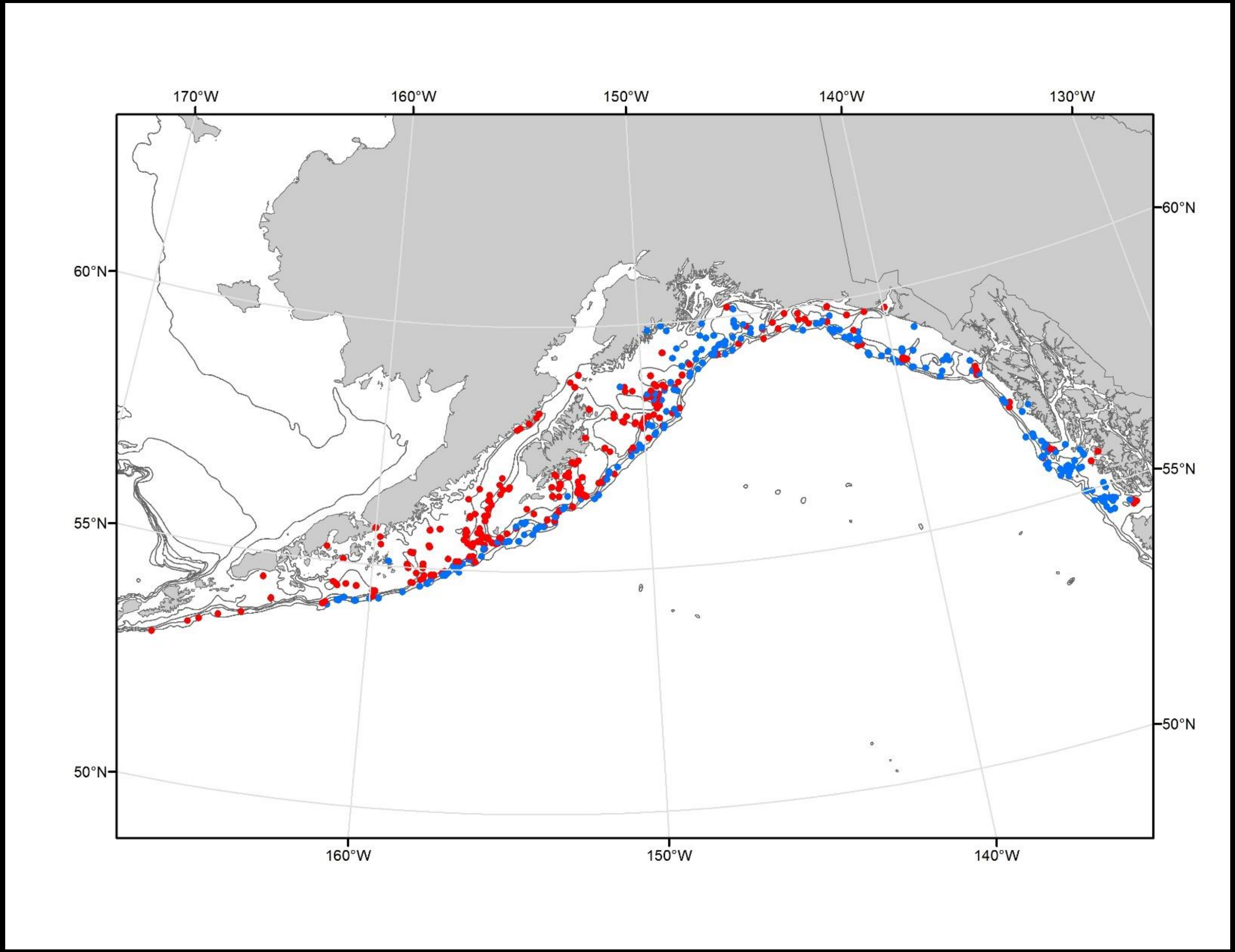
Male length-at-age by cohort, area, and depth

- High variation in length-at-age, especially at older ages
- Early cohorts were not as big at older ages as later cohorts are in middle-age
- Length-stratified data except for 2015 (which was random)
- Dover are hard to age, especially at older ages
- Few samples in Western GOA
- Few fish above 40 in Eastern GOA



GOA Dover sole residuals from sex-specific von-Bertalanffy models fit to survey data 2001-2015 outside the assessment model.

The blue points are more than 1 residual standard error below the curve and the red points are more than 1 RSE above the curve.

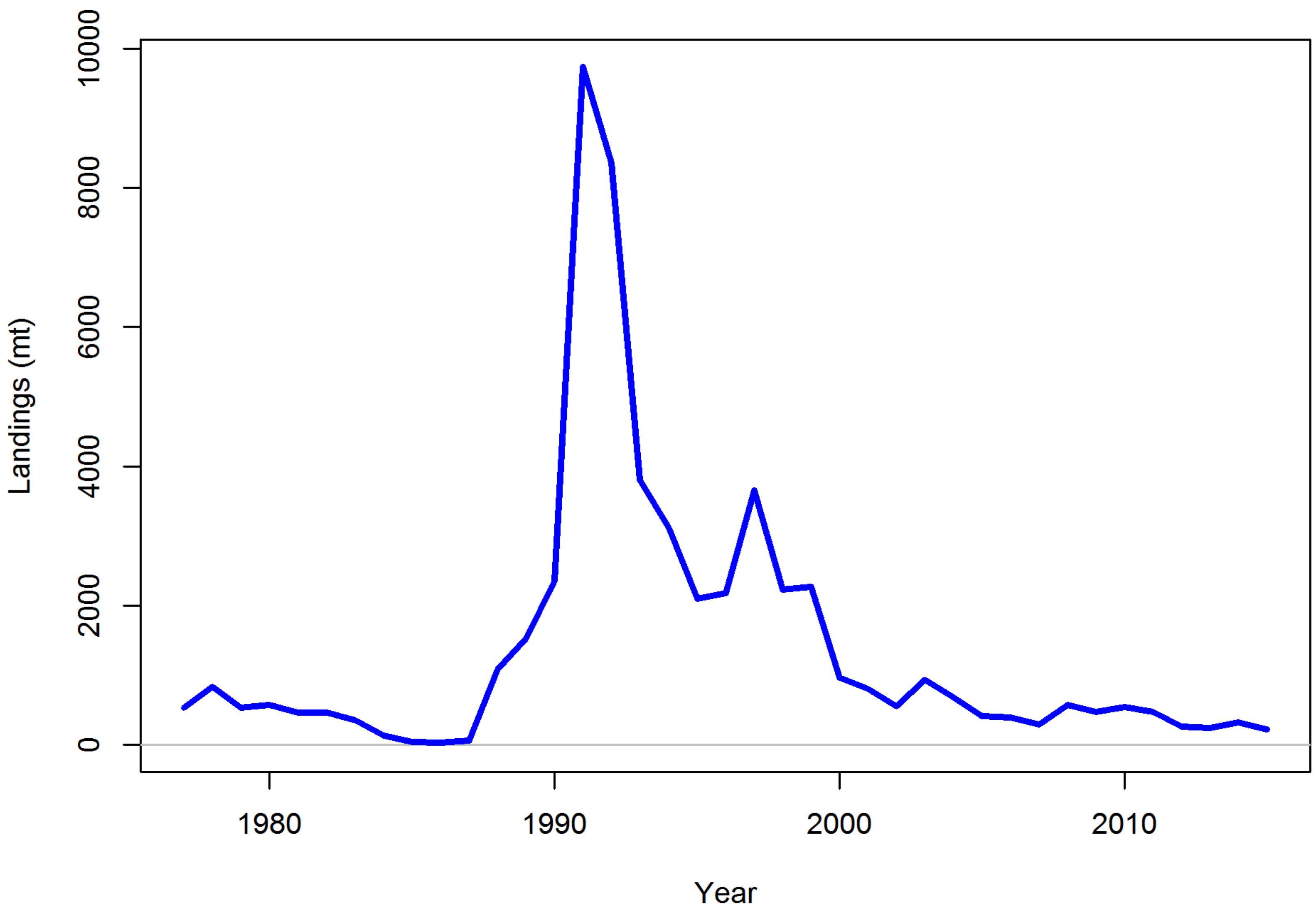


Dover sole: 2015 Assessment

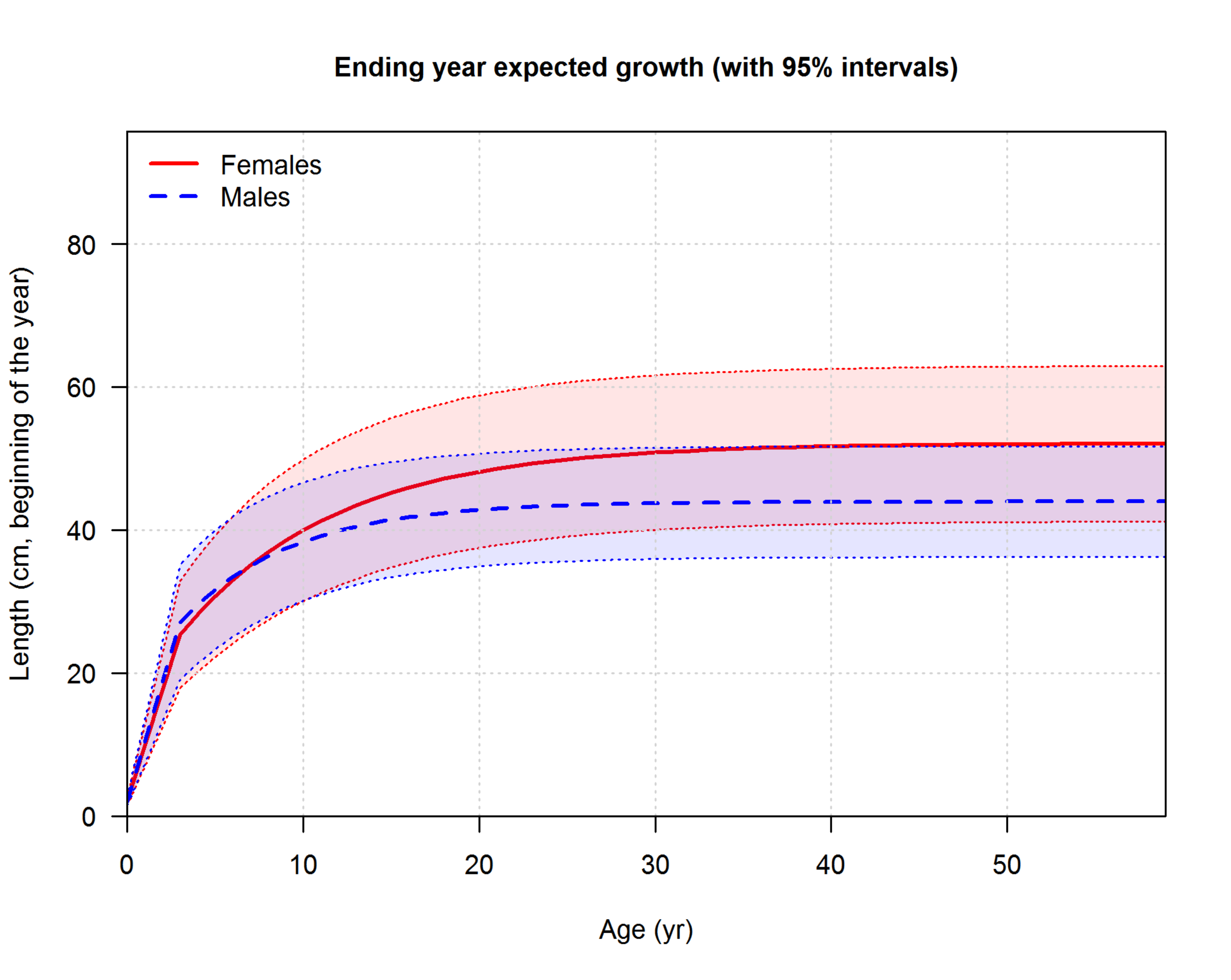
Dover Model Structure (unique among the 3 flatfish species

- Age- and sex-structured statistical catch-at-age model
- 2 surveys modeled: a “full coverage” survey for years where survey sampled deep depths (700+); a “shallow coverage” survey for years where survey sampled up to 500 meters in depth

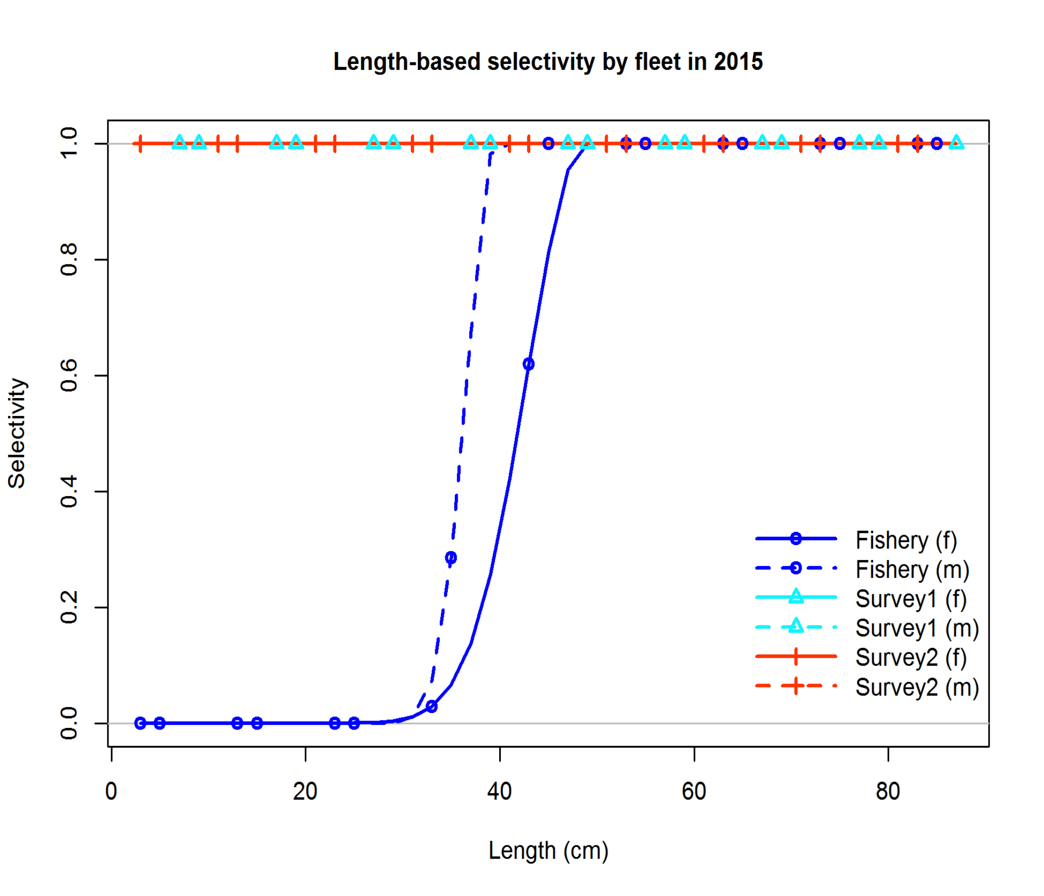
Time Series of Catches



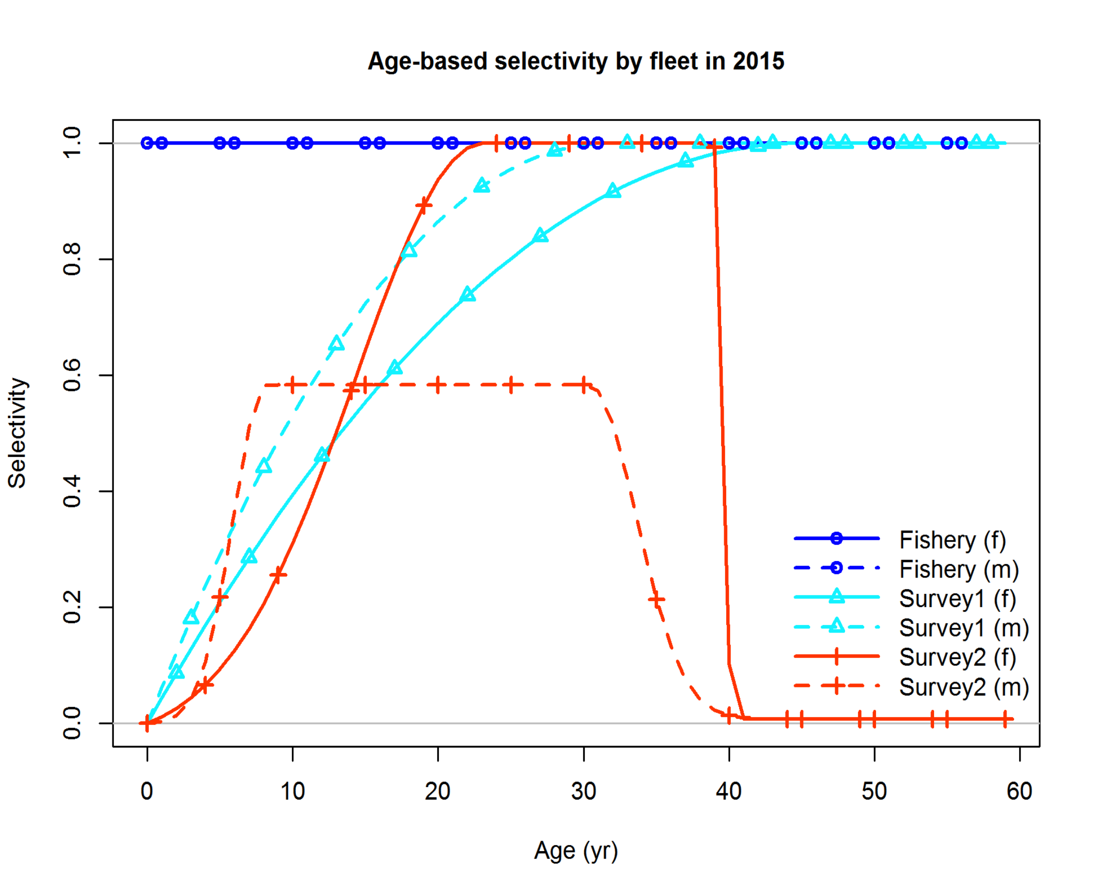
Estimating length-at-age relationship with 95% confidence bounds



Fishery Selectivity



Survey Selectivity



Selectivity Estimates:

Highlighted values correspond to a parameter on/near a bound

	Fishery		Full Coverage Survey		Shallow Coverage Survey	
Double-normal selectivity parameters	Est	Std. Dev.	Est	Std. Dev.	Est	Std. Dev.
Peak: beginning size for the plateau	48.81	1.27	**45.00**	0.09	23.16	1.80
Width: width of plateau	Fixed		Fixed		-0.28	0.25
Ascending width (log space)	4.26	0.24	**11.96**	1.21	5.06	0.22
Descending width (log space)	Fixed		Fixed		-0.73	14.80
Initial: selectivity at smallest length or age bin	Fixed		Fixed		-498	11236.20
Final: selectivity at largest length or age bin	Fixed		Fixed		**-4.99**	0.44
Male Peak Offset	-9.28	1.37	-13.35	1.41	**-15.00**	0.05
Male ascending width offset (log space)	-1.46	0.37	4.68	119.24	-2.74	0.65
Male descending width offset (log space)	Fixed		Fixed		3.75	14.12
Male "Final" offset (transformation required)	Fixed		Fixed		0.03	0.88
Male apical selectivity	Fixed		Fixed		0.58	0.06

Selectivity Estimates:

Highlighted values correspond to a parameter on/near a bound

	Fishery		Full Coverage Survey		Shallow Coverage Survey	
Double-normal selectivity parameters	Est	Std. Dev.	Est	Std. Dev.	Est	Std. Dev.
Peak: beginning size for the plateau	48.81	1.27	**45.00**	[illegible]	[illegible]	1.80
Width: width of plateau	[illegible]	[illegible]	Fixed	[illegible]	[illegible]	0.25
Ascending width (log space)	[illegible]	0.24	**11.96**	[illegible]	[illegible]	0.22
Descending width (log space)	[illegible]	[illegible]	Fixed		-0.73	14.80
Initial: selectivity at smallest length or age bin	Fixed		Fixed		-498	11236.20
Final: selectivity at largest length or age bin	Fixed		Fixed		**-4.99**	0.44
Male Peak Offset	-9.28	1.37	-13.35	1.41	**-15.00**	0.05
Male ascending width offset (log space)	-1.46	0.37	4.68	119.24	-2.74	0.65
Male descending width offset (log space)	Fixed		Fixed		3.75	14.12
Male "Final" offset (transformation required)	Fixed		Fixed		0.03	0.88
Male apical selectivity	Fixed		Fixed		0.58	0.06

Forces the curve to end up at 1 by age 45

Limits the shallowness of the curve between 0 and 1

Selectivity Estimates:

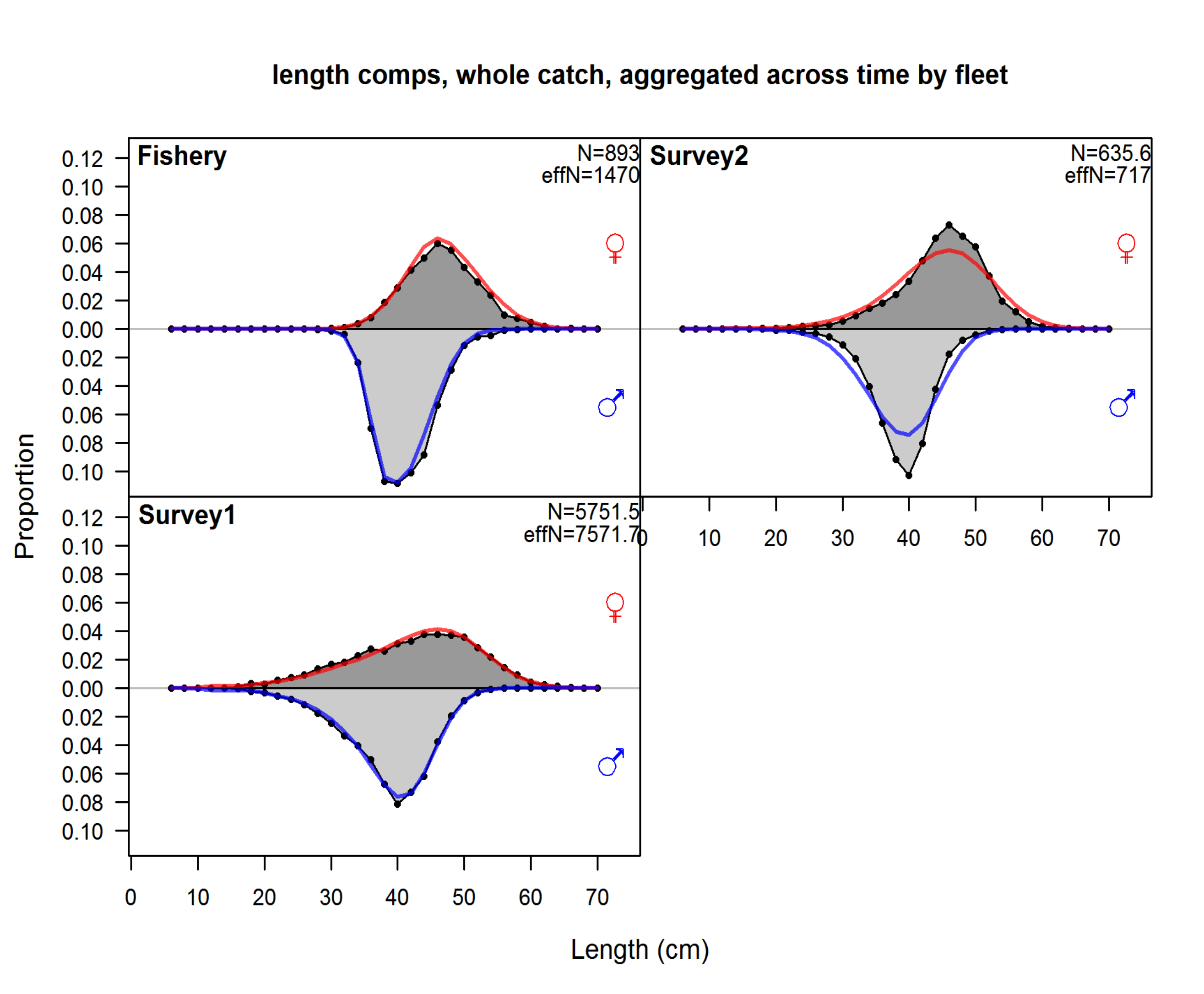
Highlighted values correspond to a parameter on/near a bound

	Fishery		Full Coverage Survey		Shallow Coverage Survey	
Double-normal selectivity parameters	Est	Std. Dev.	Est	Std. Dev.	Est	Std. Dev.
Peak: beginning size for the plateau	48.81	1.27	**45.00**	0.09	23.16	1.80
Width: width of plateau	Fixed		Fixed		-0.28	0.25
Ascending width (log space)	4.26	0.24	**11.96**	1.21	5.06	0.22
Descending width (log space)	Fixed		Fixed		-0.73	14.80
Initial: selectivity at smallest length or age bin	Fixed				-498	11236.20
Final: selectivity at largest length or age bin	Fixed				**-4.99**	0.44
Male Peak Offset	-9.28			1.41	**-15.00**	0.05
Male ascending width offset (log space)	-1.46	0.37	4.68	119.24	-2.74	0.65
Male descending width offset (log space)	Fixed				3.75	14.12
Male "Final" offset (transformation required)	Fixed				0.03	0.88
Male apical selectivity	Fixed		Fixed		0.58	0.06

Shallow survey catches none of the very oldest fish

Males reach peak selectivity more than 15 years before females?

Fits to length composition data, aggregated over years (more diagnostic slides as "extras" at end if we want)

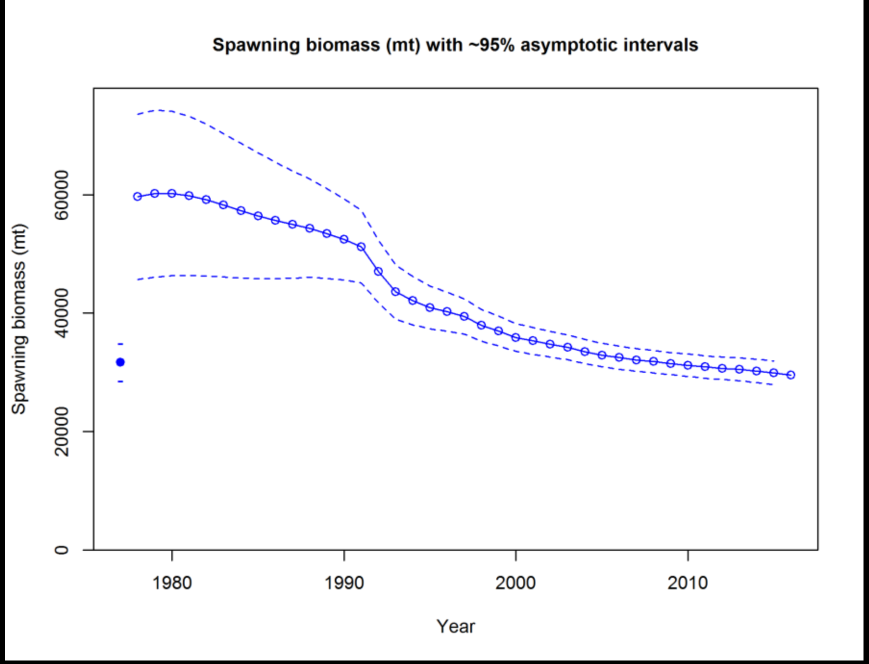


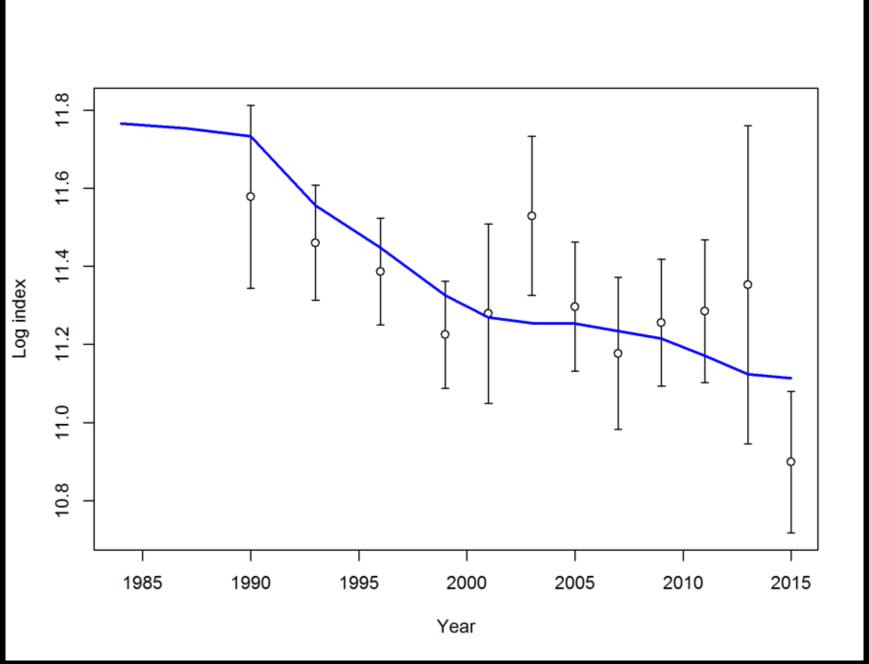
Dover sole: Cleaned up run

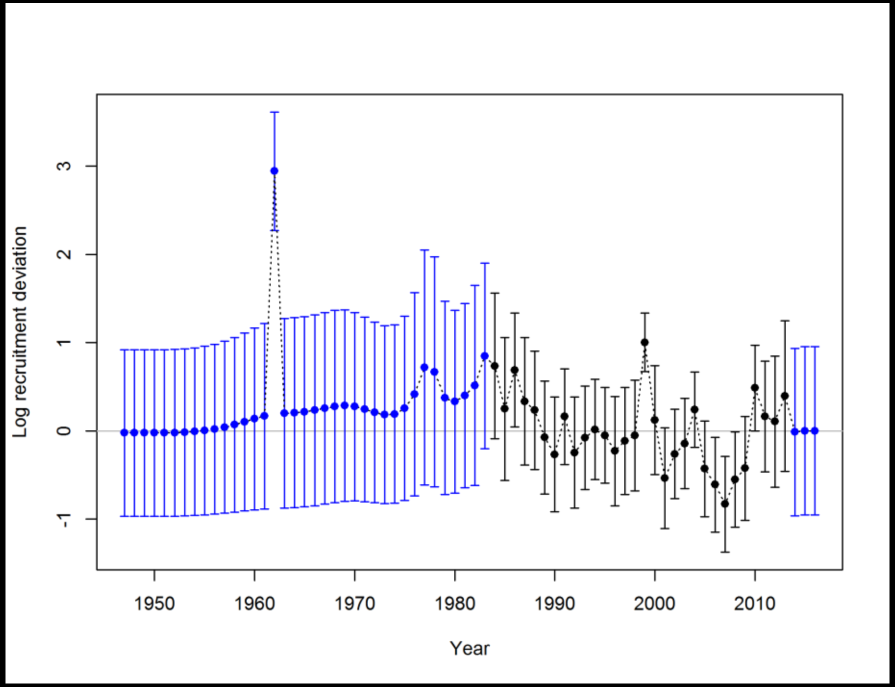
Where to go from here?

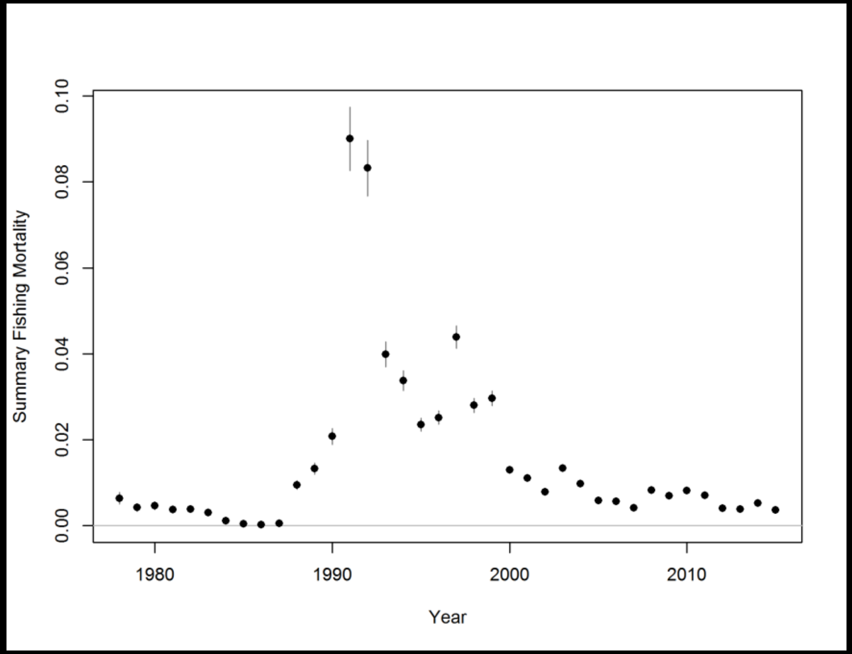
- Obvious fixes (did a “cleaned up run,” implementing the following):
  - Use age data bins for ages 1 and 2; data for ages 0-3 were aggregated within an age-3 bin; loss of info on how many age 1-2 fish were actually caught that could inform the selectivity curve (holdover from 2011 pre-SS model and 2013 matching exercise)
  - Removed survey biomass index from 1984 and 1987; methods were different in those years and survey length and age data were already removed
  - Survey timing should be month 6, not month 1 (this is a holdover from the 2011 pre-SS assessment and 2013 matching exercise)
- Biomass index is not a flat line + mostly light fishing: the model wants to put a huge recruitment in the early rec devs that is likely driving this

“Cleaned up run:” no parameters on bounds

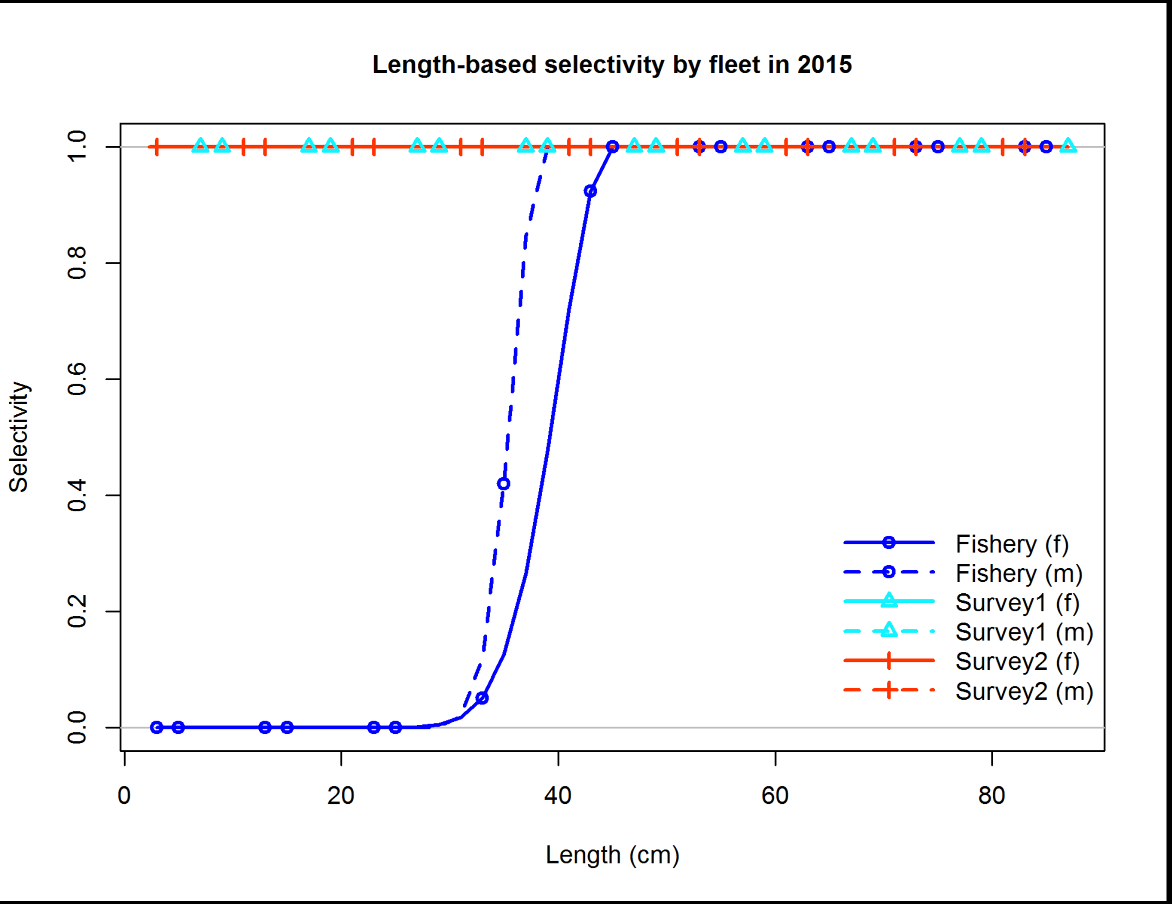


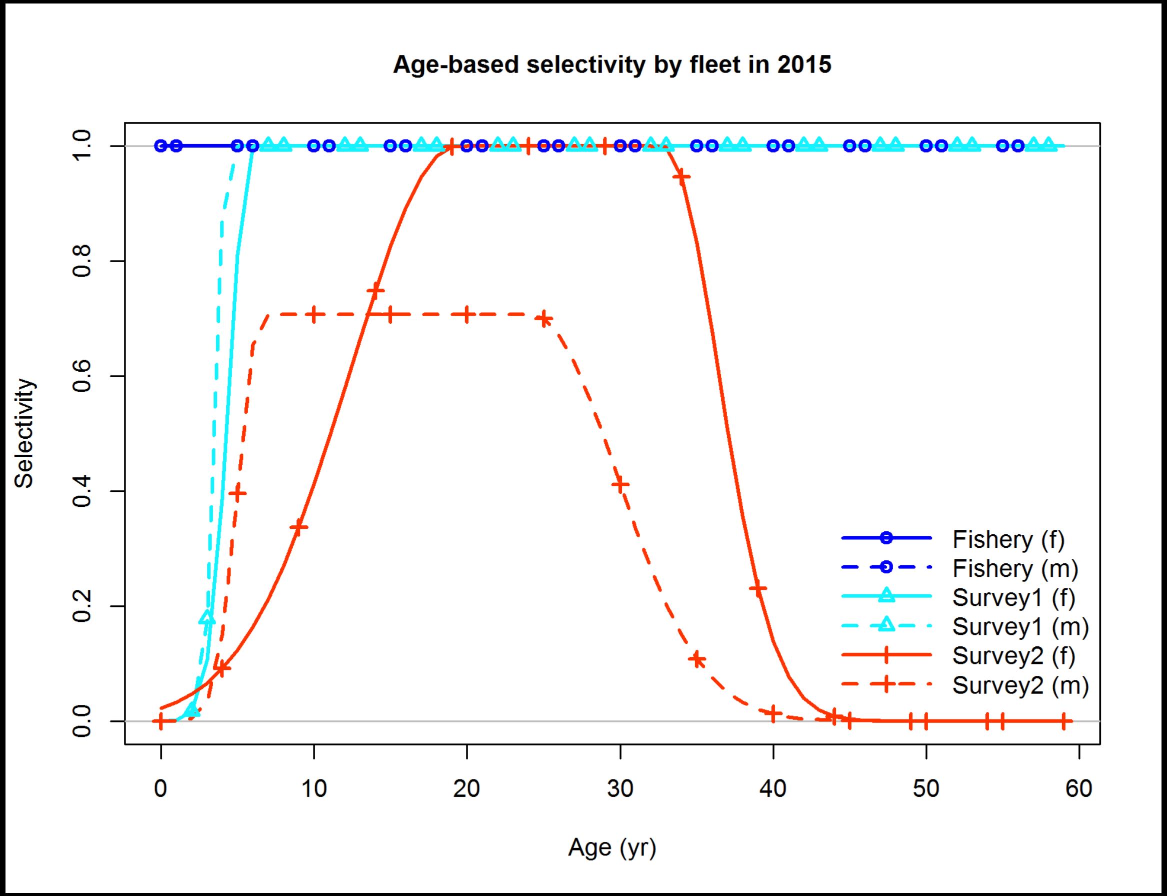




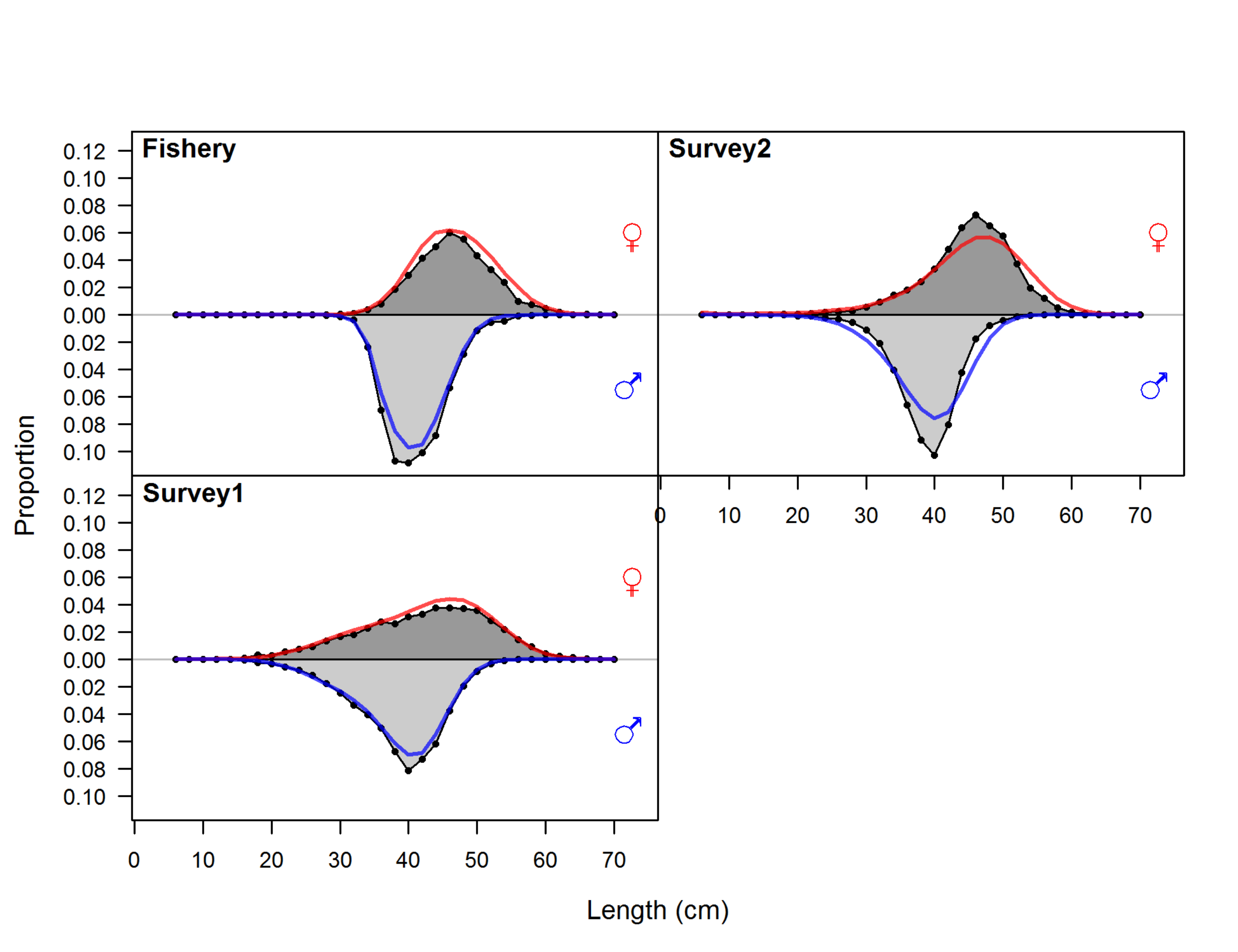


“Cleaned up run:” no parameters on bounds





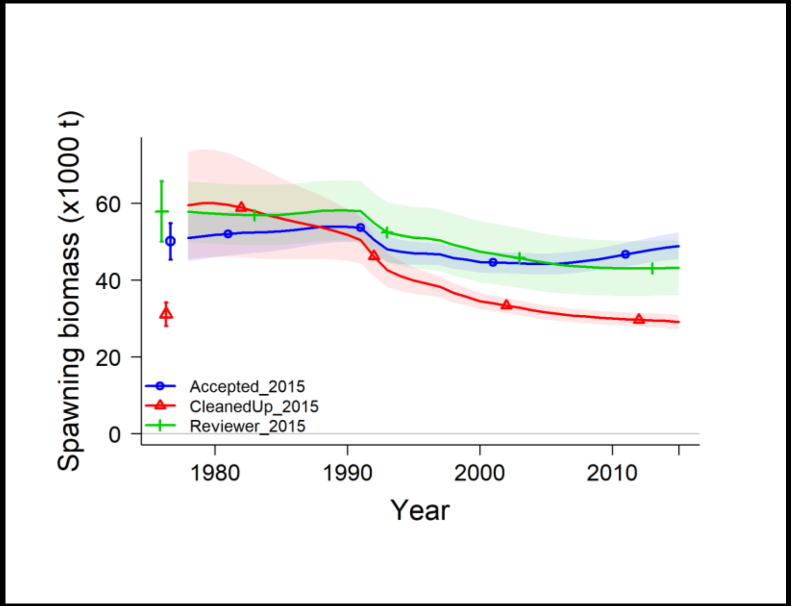
Fits to length compositions, aggregated across time

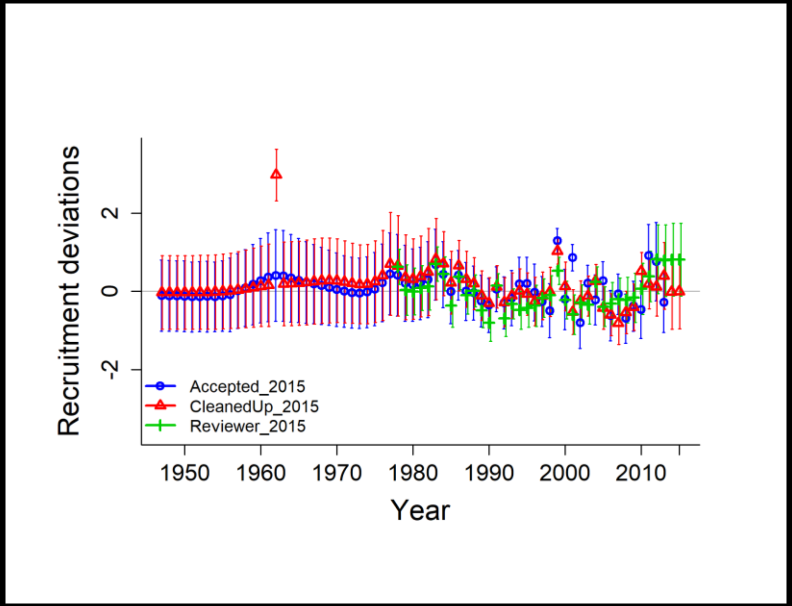


Reviewer Requested Run

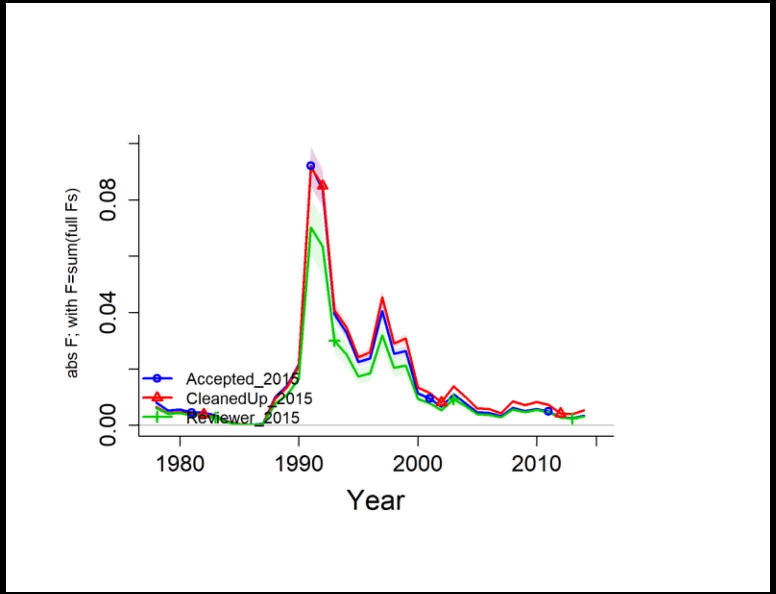
- Removed 1980s data and 2001 data
- Split survey into 3 surveys with their own q's
- Used a normal prior on q with mean = 1.2, SD = 0.175 for survey years going to 1000m
  - Used 1.17 for survey years to 700m, SD = 0.175
  - Used 1.08 for survey years to 500m, SD = 0.175
- No early period recruitment:
  - Numbers drop off around/after age 50
  - 3 years old before they are observed
  - First age data in 1990
  - Want to have observed them 5 times before including as a rec dev
  - 50-3-5 =12
  - 1990 – 42 = 1948
  - However, there is a lot of ageing error to consider, so can age more reliably when 20 or under, which then yields 1990 -20-3 = 1978 (model start year)
- Francis re-weighting (keeping input sample size = haul size)

Dover sole: Comparison: 2015 Accepted, Clean, Reviewer (all with 2015 data): both 2015 accepted model and reviewer models have selectivity parameters on bounds



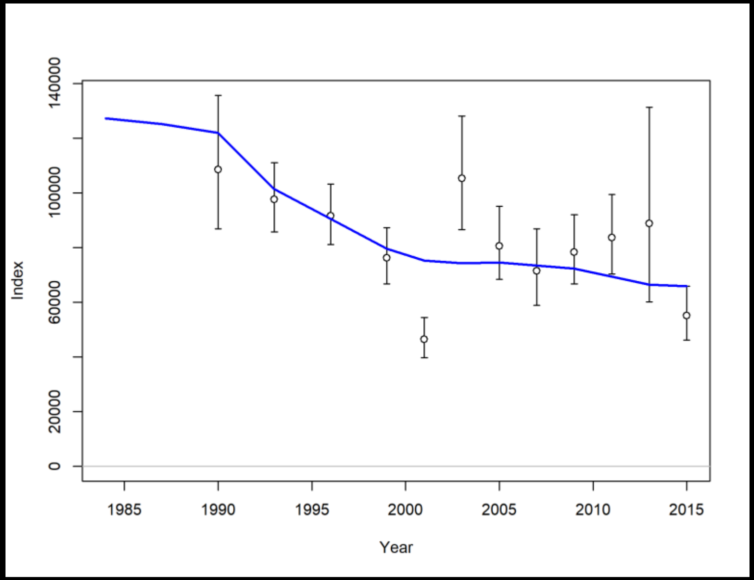


Reviewer run: 5 selectivity parameters on bounds (changes made on top of cleaned up run with extended bounds)

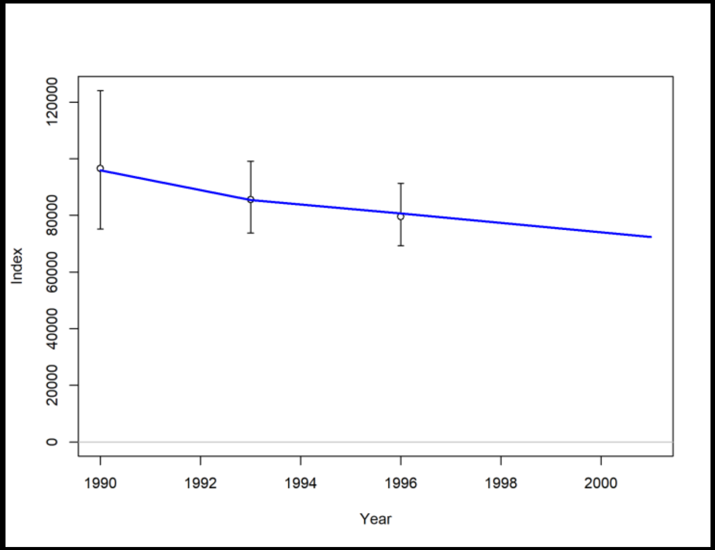


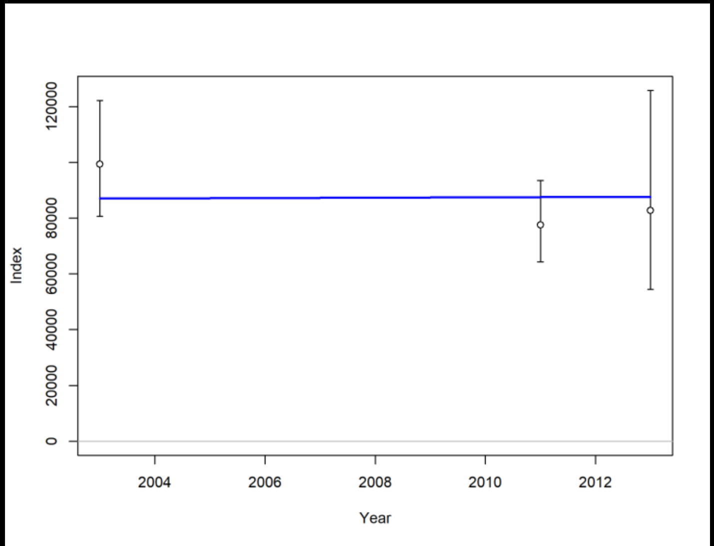
Fits to indices

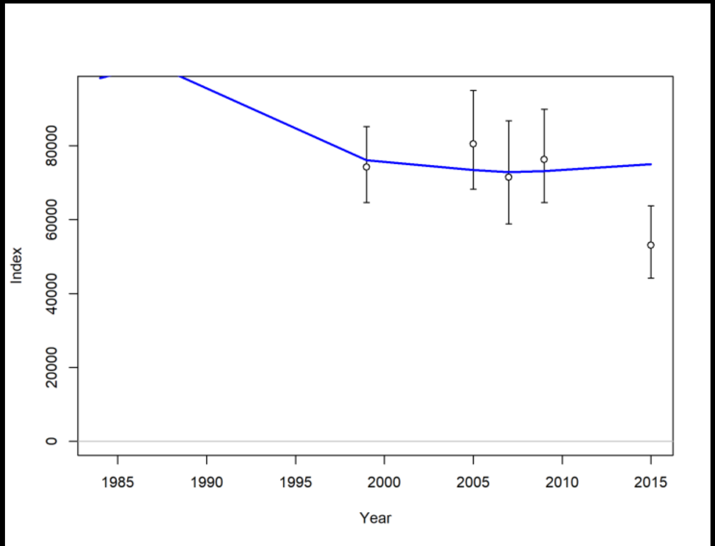
Cleaned up run ->



Reviewer run







What’s next for the Dover assessment?

- Cleaned up run could use more cleaning up, eliminating early recruitment deviations
  - Could estimate growth outside of the model so as to allow for running MCMCs in an efficient fashion, giving us more info on uncertainty
  - Could specify priors for some selectivity parameters

- Reviewer run could be simplified to include only <500m and >500m (leaving out the biomass index in years where the survey went to 700m)
  - This still divides the survey index into small pieces, not acknowledging the fact that we have a survey for a longer time period for <500m.
  - Not sure the extra info on uncertainty from estimating catchability is worth splitting up years of the biomass index

What’s next for the Dover assessment?

- Growth patterns are not accounted for (older fish are small, some Eastern fish are small)

- A two area model (shallow vs deep) would allow us to keep the biomass index for <500m for all years, and estimate growth separately for the old fish in the deep
  - Recruitment occurs only in the shallow – clear from data
  - May be able to estimate movement to deep without tagging data based on ontogenetic movement pattern evident in the data
  - Andrea will present this

Another idea

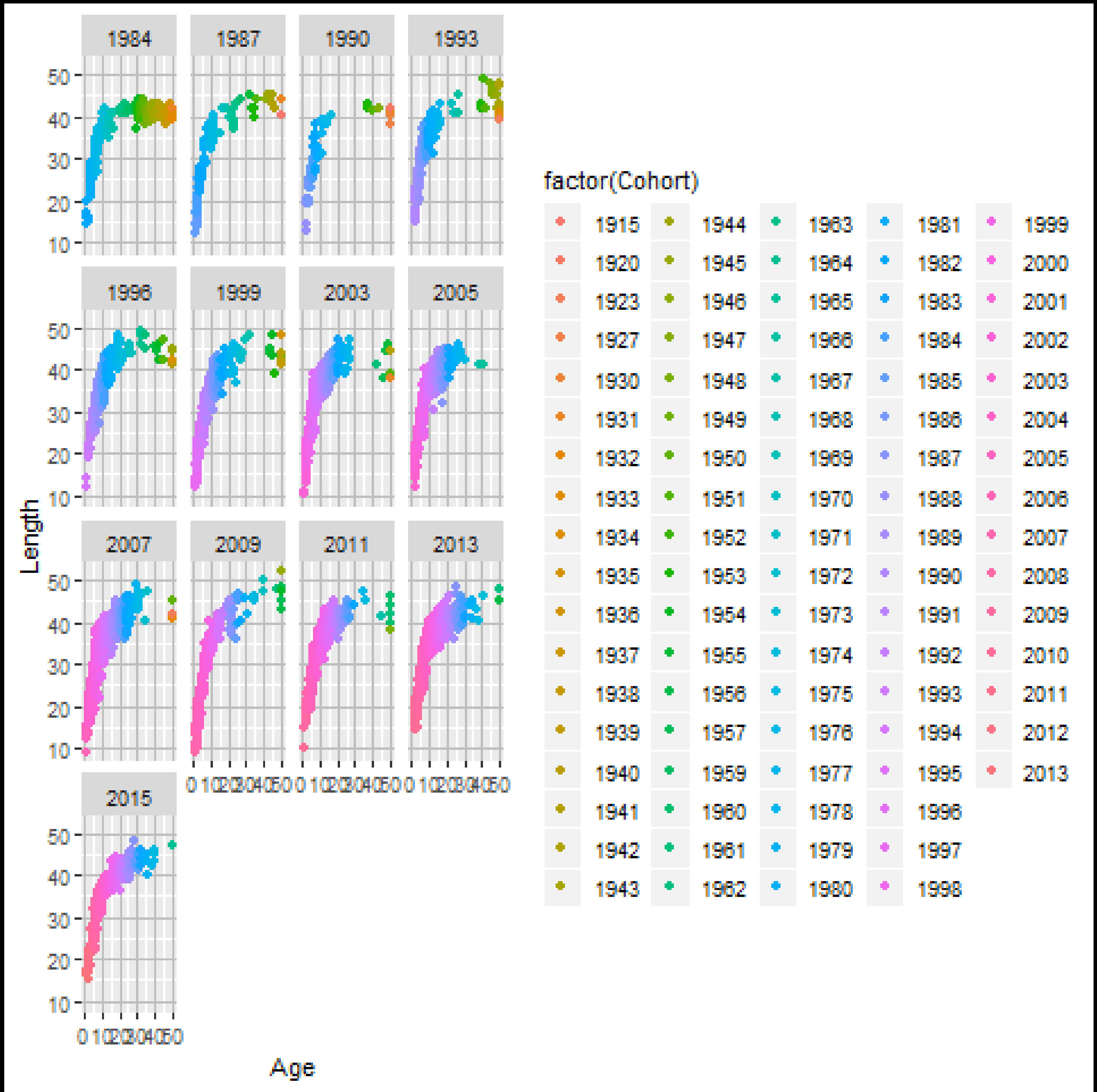
- A two area model (Western-Central vs Eastern) with separate growth estimates before the 1977 regime change: would this work?
  - Accounts for E-W pattern and time-varying pattern
  - Costly in terms of growth parameters that need to be estimated
  - Is the variability in growth in these single-area models causing problems for estimating selectivity parameters and model stability?

End

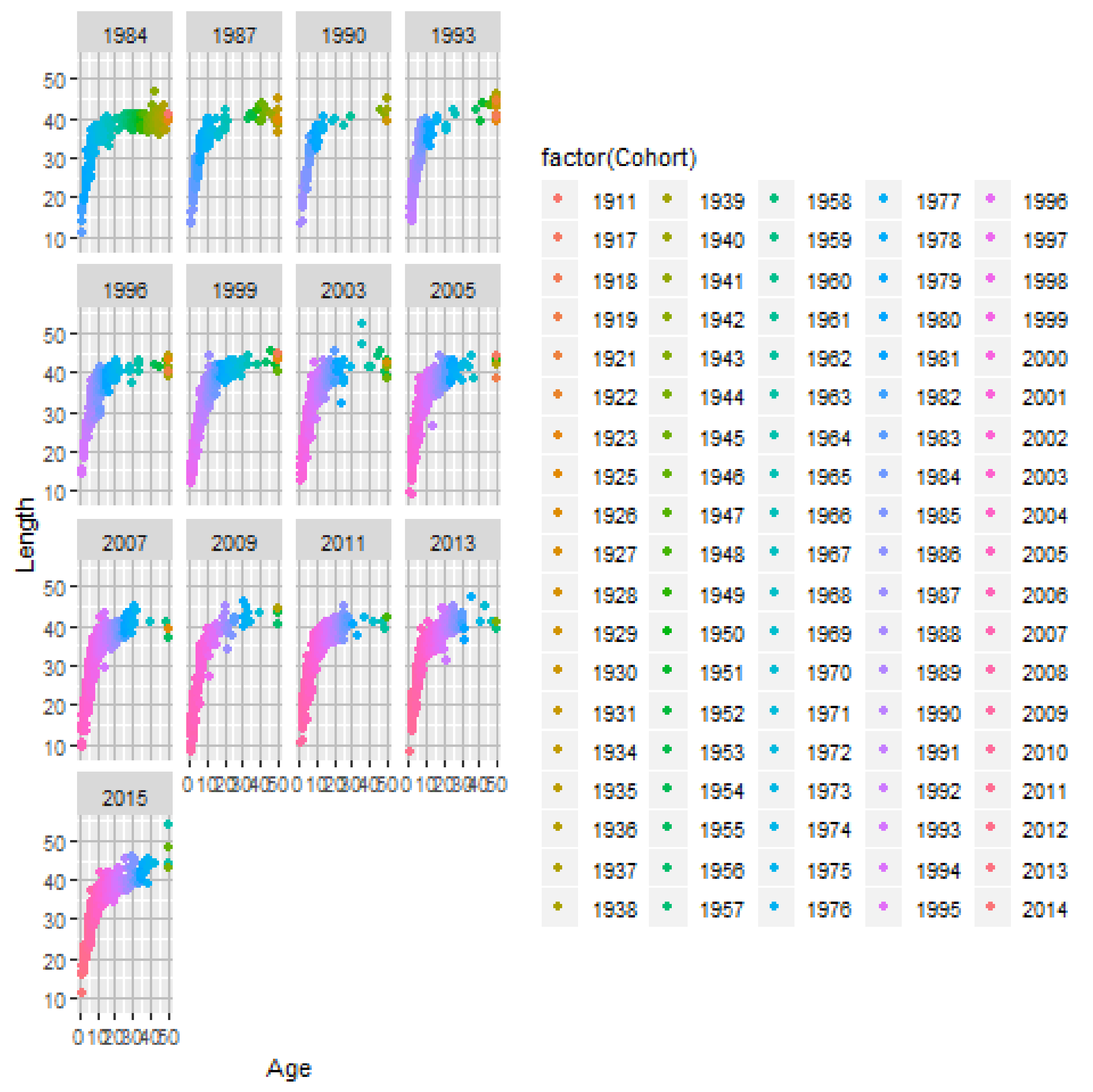
Extra slides if needed

How do these length-at-age plots compare to another long-lived GOA fish: Pacific Ocean Perch?

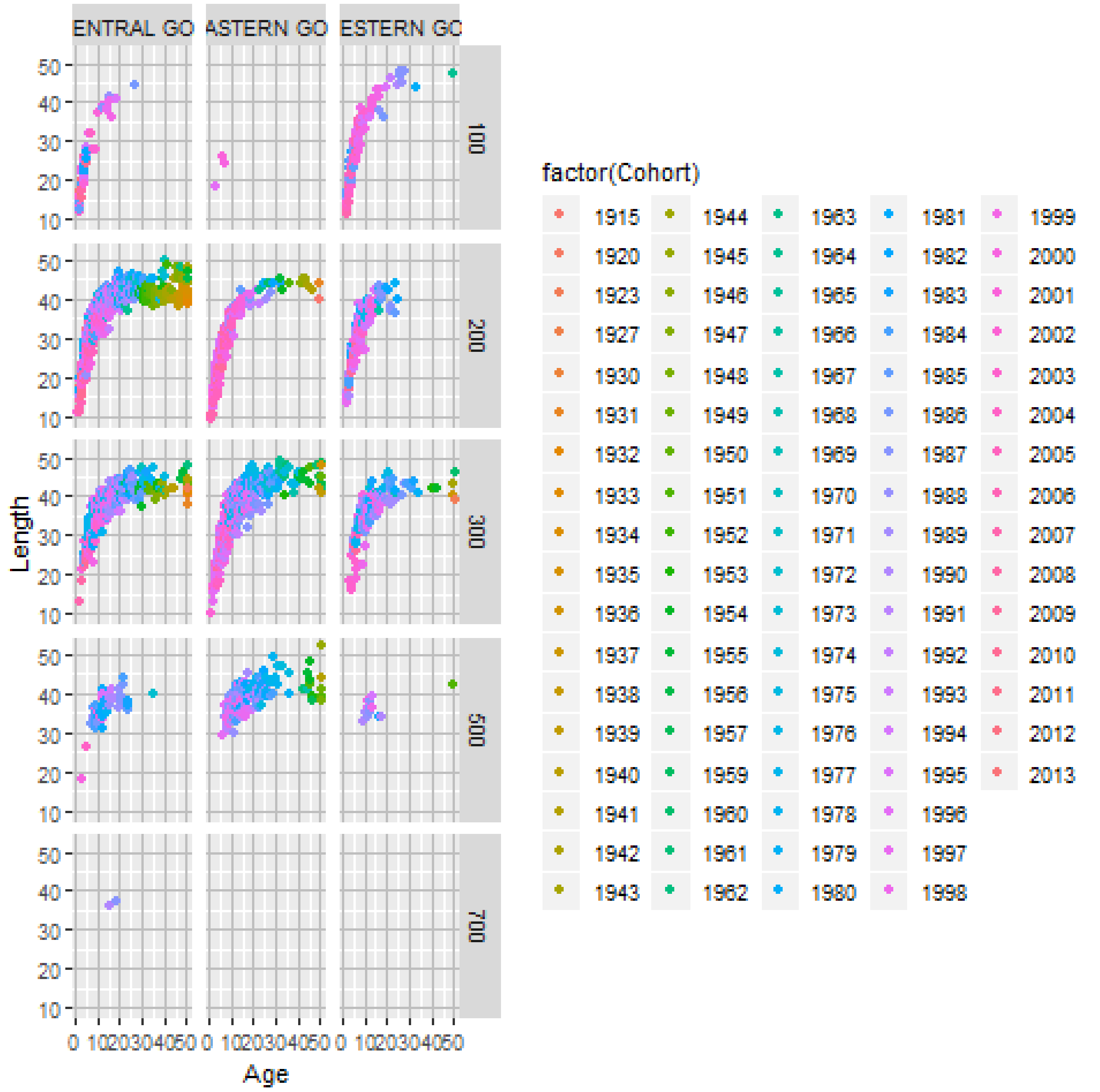
POP: Female length-at-age by cohort and year



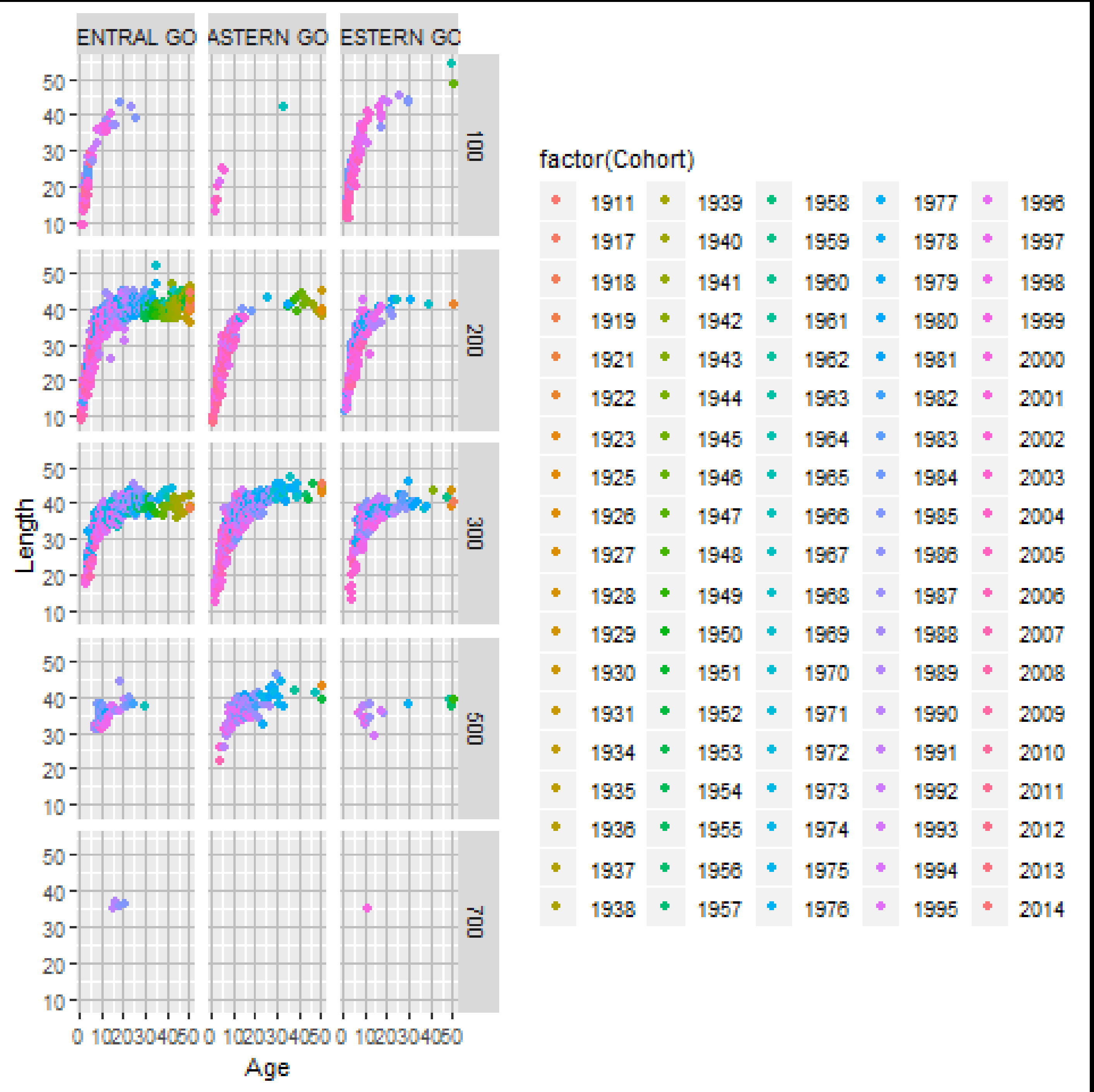
POP: Male length-at-age by cohort and year



POP: Female length-at-age by cohort, depth, and area

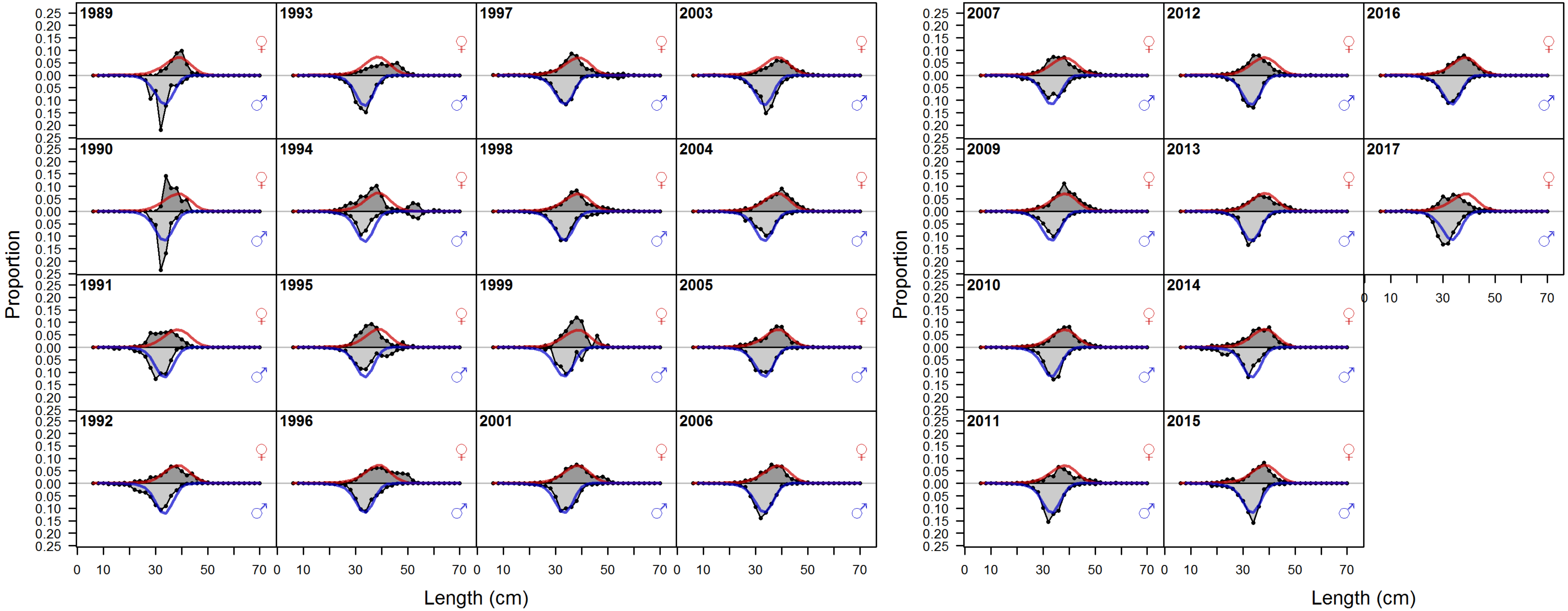


POP: Male length-at-age by cohort, depth, and area

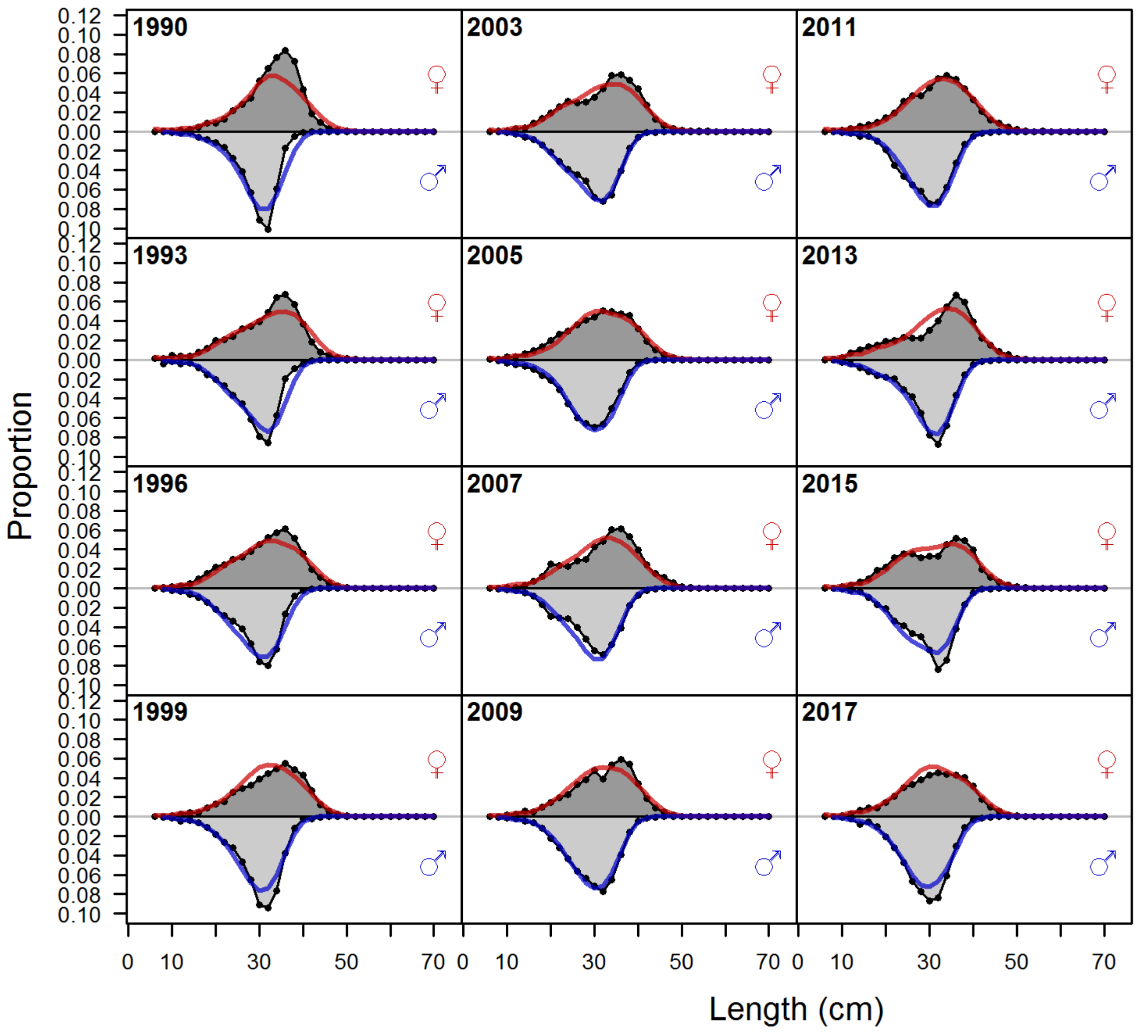


Additional flathead reviewer run diagnostics

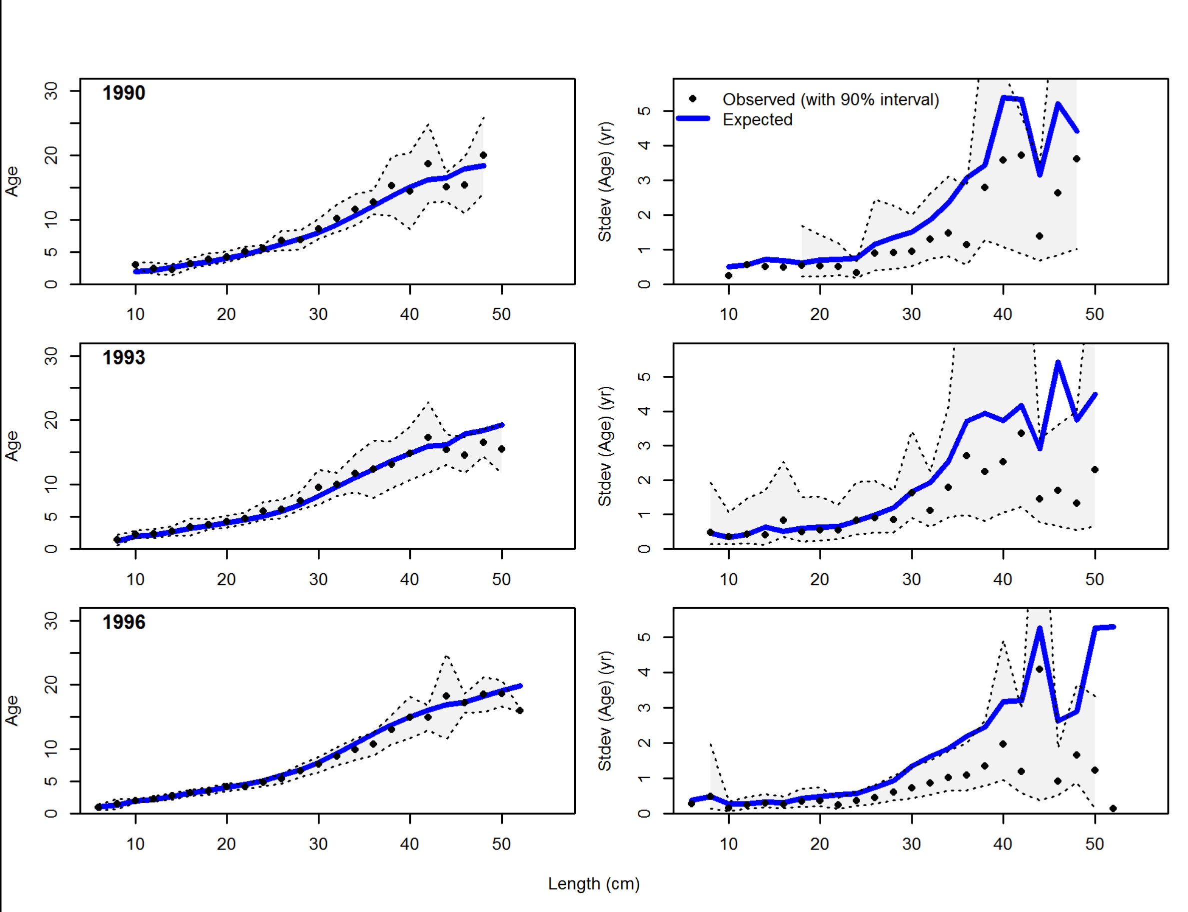
Fits to fishery length compositions (Reviewer run)

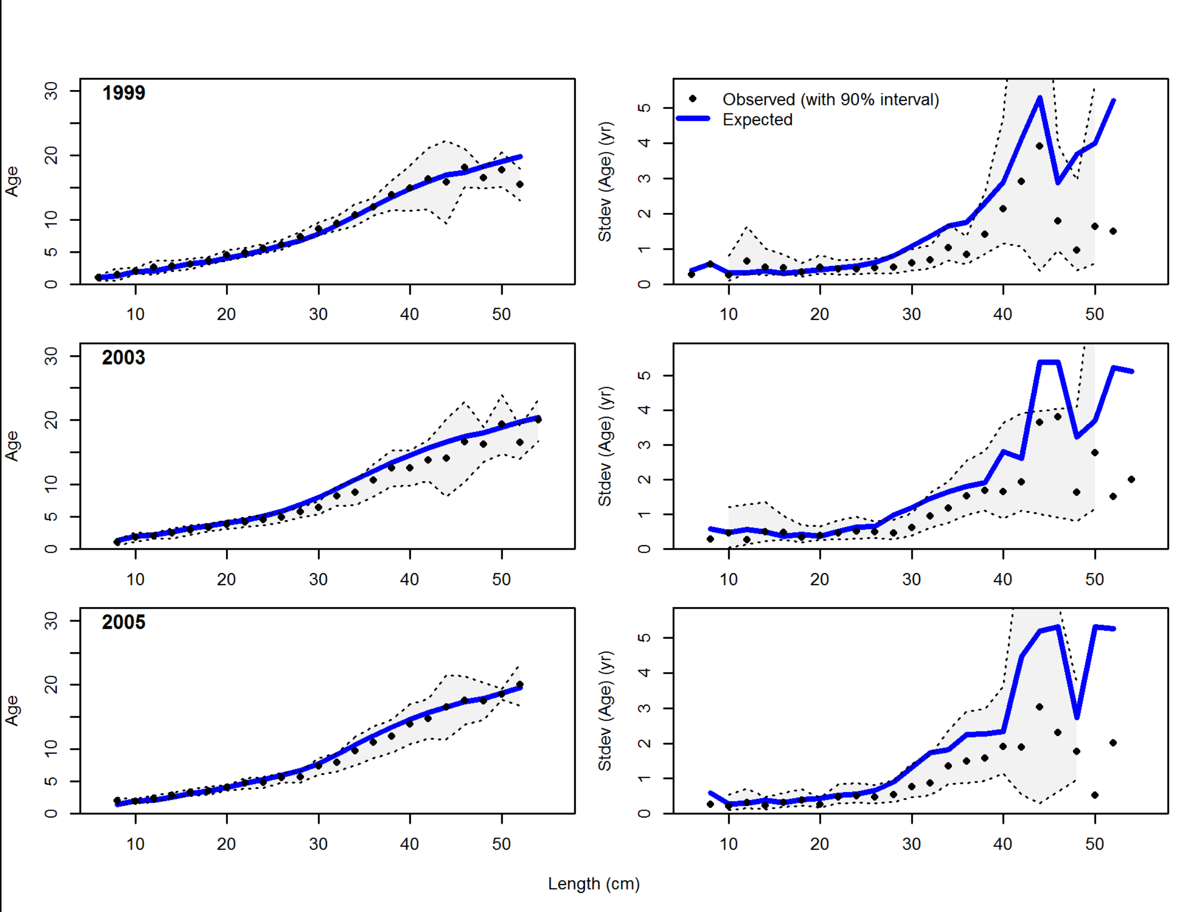


Fits to fishery survey length compositions (Reviewer run)

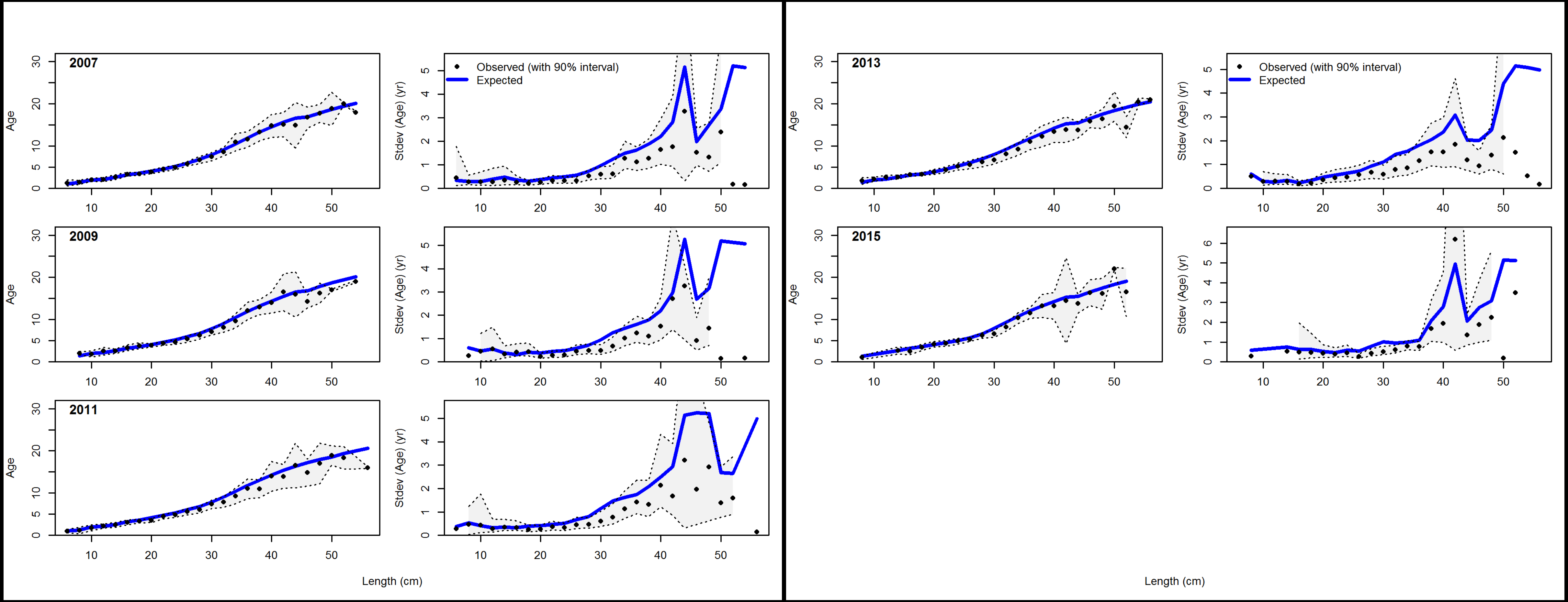


Reviewer Run





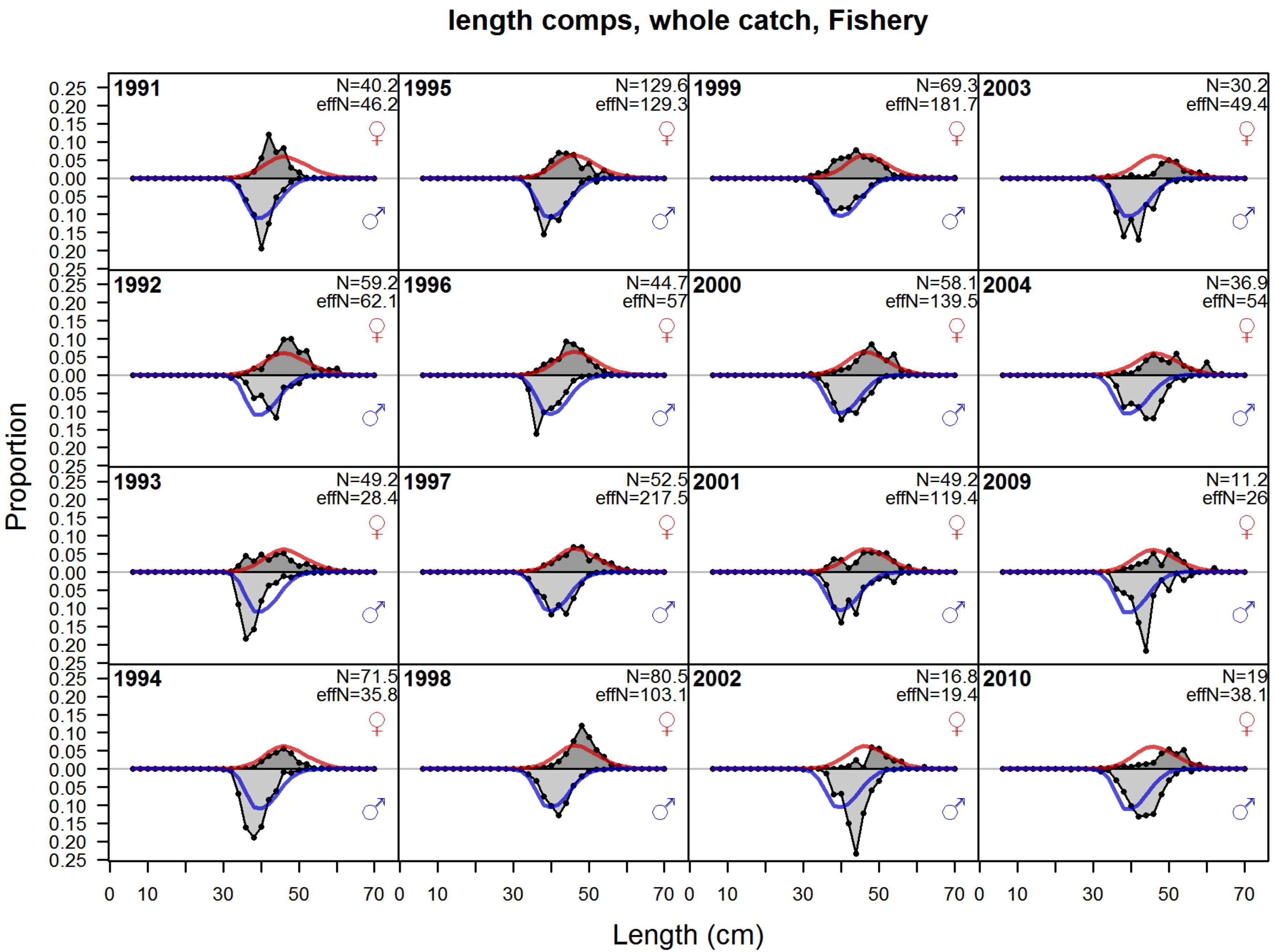
Reviewer Run



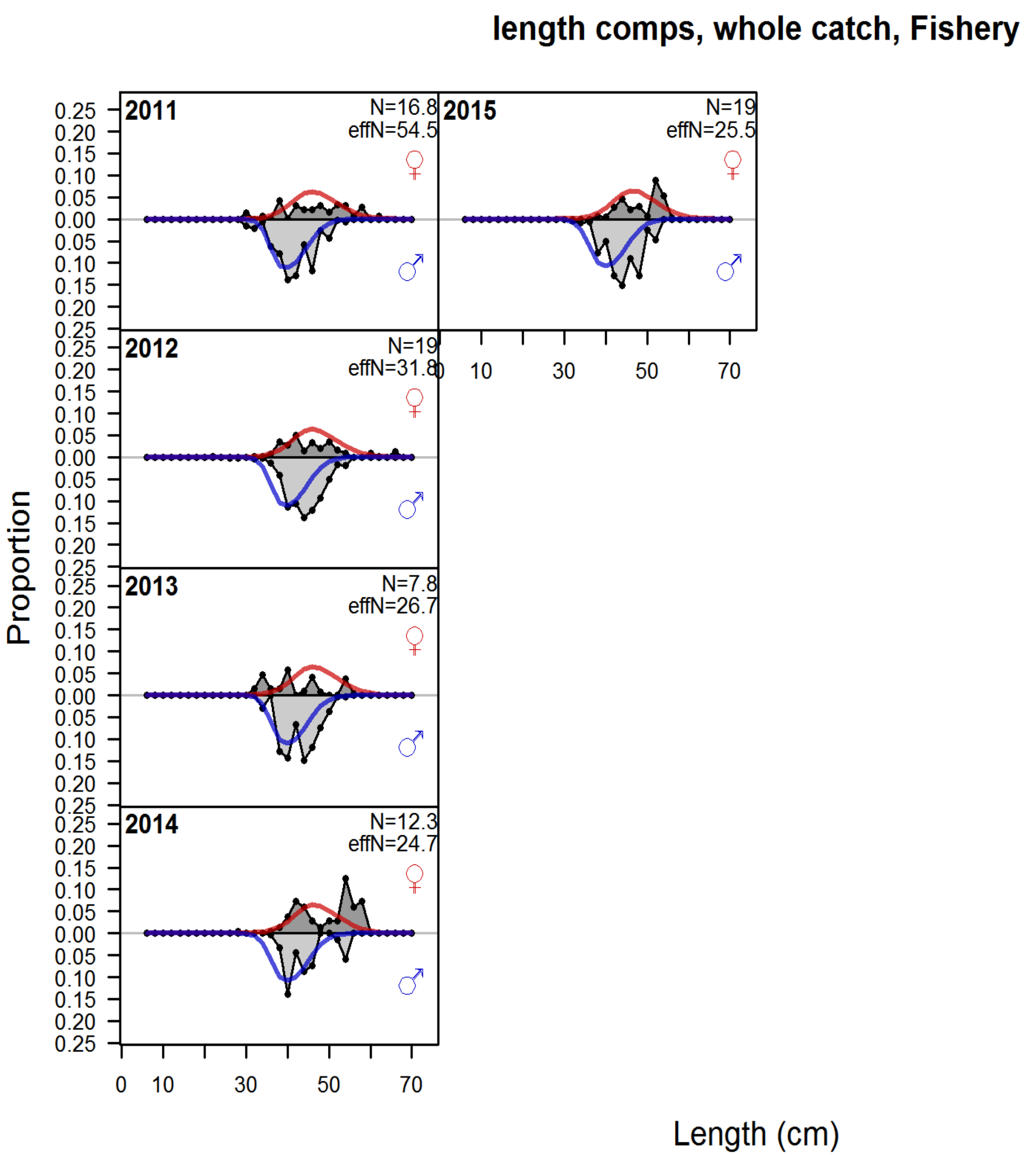
Parameters Estimated within the Dover model

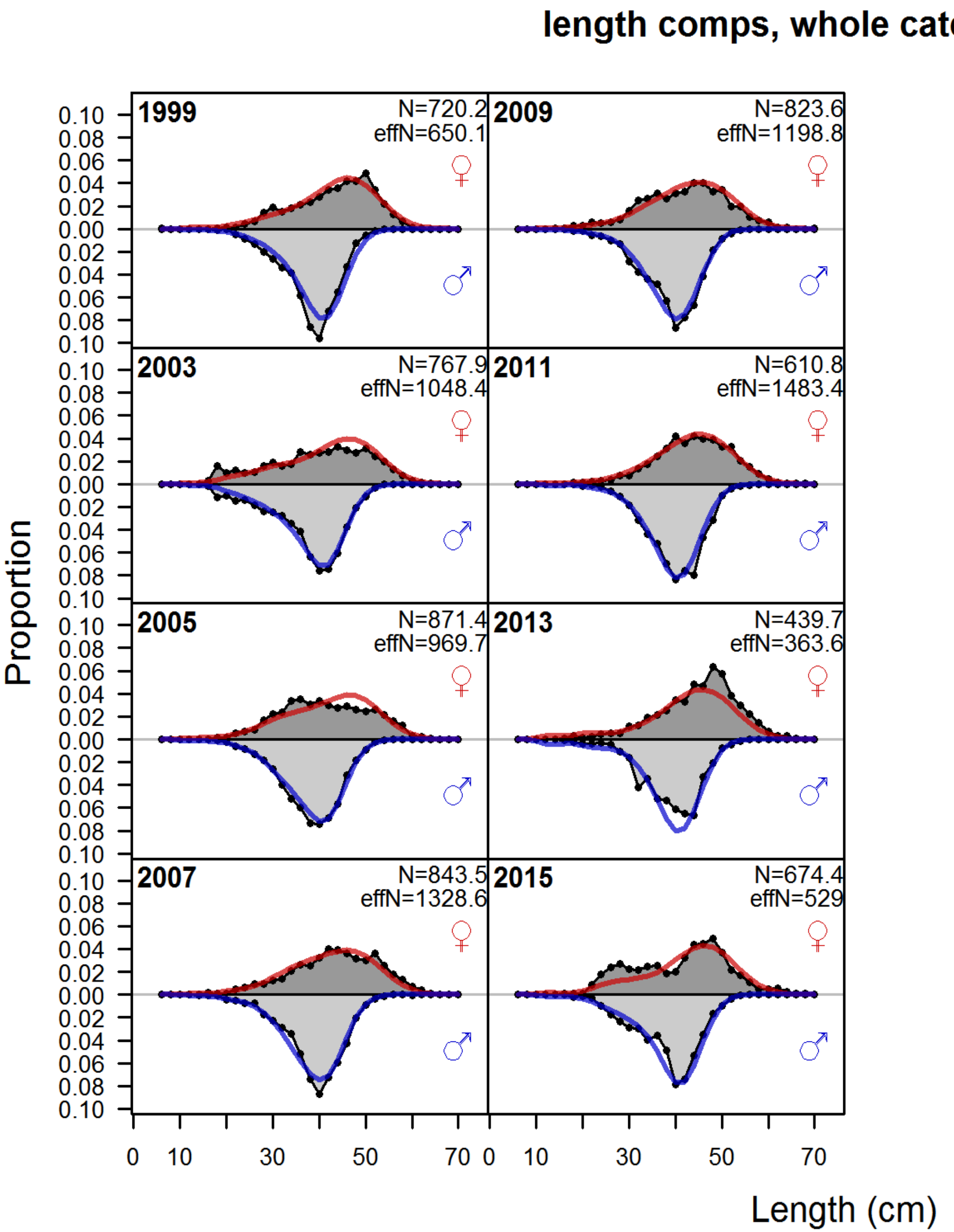
- Ln(R0)
- Length-based, asymptotic fishery selectivity
- Age-based double-normal shallow and full coverage survey selectivity (separately), full coverage survey selectivity restricted to be asymptotic and to reach 1 at a reasonable age
- Recruitment deviations (1965-2012) (simple deviations, no SR curve)
- Yearly fishing mortality rates
- Parameters of the von-Bertalanffy growth curve
- CV of length-age relationship for youngest and oldest fish

2015 Dover Run

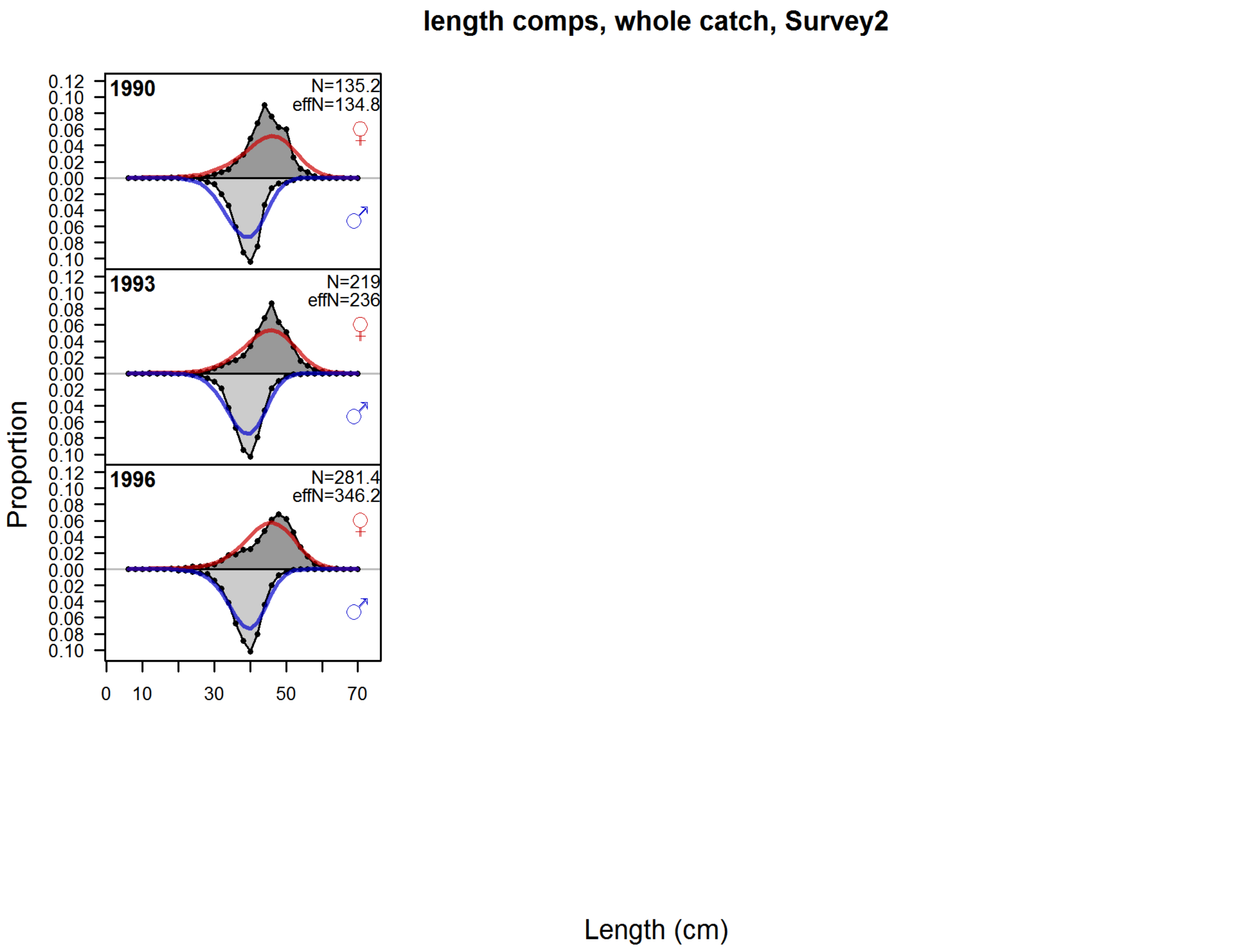


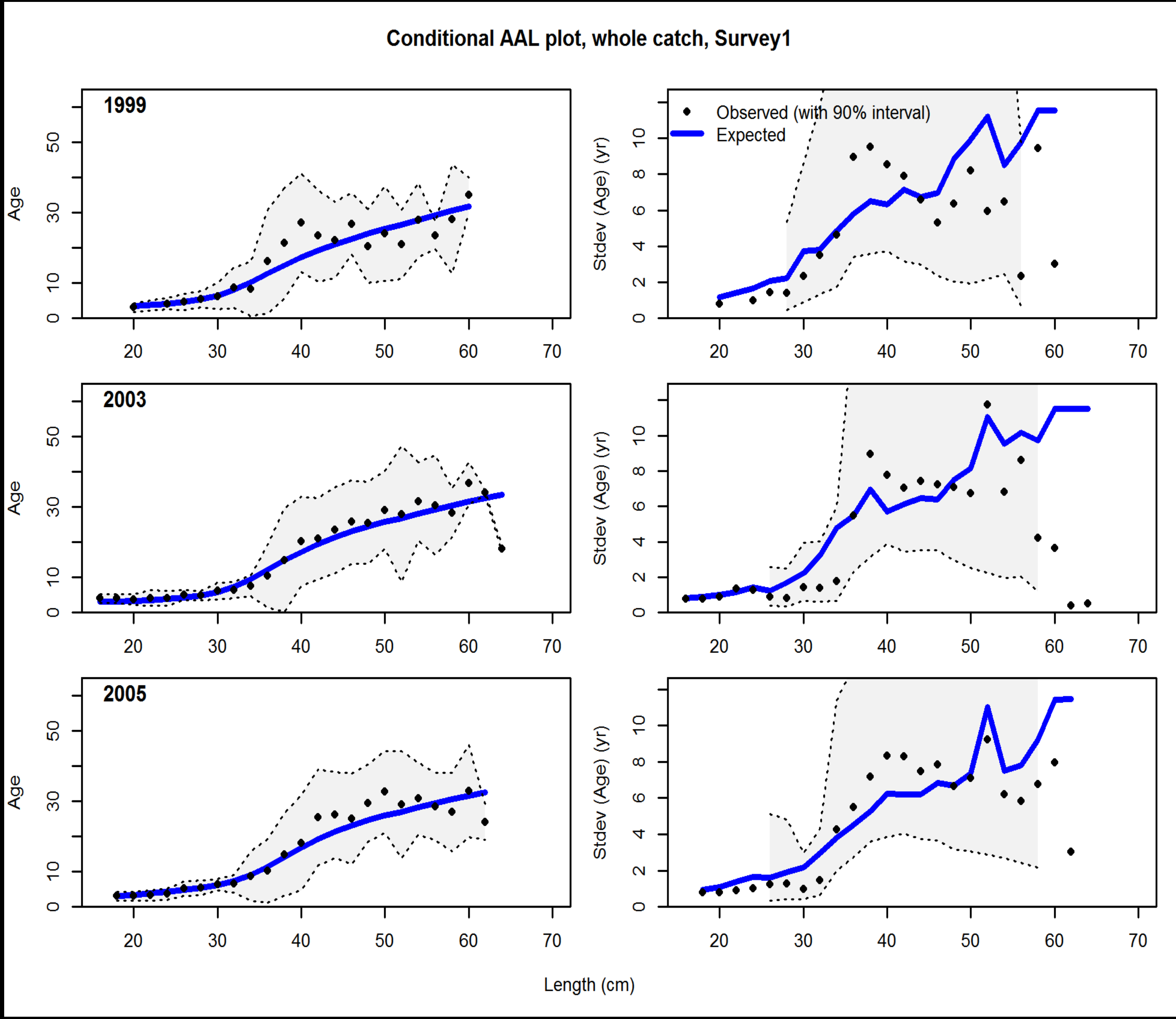
length comps, whole catch, Fishery
1991 N=40.2 effN=46.2
1995 N=129.6 effN=129.3
1999 N=69.3 effN=181.7
2003 N=30.2 effN=49.4
1992 N=59.2 effN=62.1
1996 N=44.7 effN=57
2000 N=58.1 effN=139.5
2004 N=36.9 effN=54
1993 N=49.2 effN=28.4
1997 N=52.5 effN=217.5
2001 N=49.2 effN=119.4
2009 N=11.2 effN=26
1994 N=71.5 effN=35.8
1998 N=80.5 effN=103.1
2002 N=16.8 effN=19.4
2010 N=19 effN=38.1
Proportion
Length (cm)


length comps, whole catch, Fishery
2011
N=16.8
effN=54.5
2015
N=19
effN=25.5
2012
N=19
effN=31.8
2013
N=7.8
effN=26.7
2014
N=12.3
effN=24.7
Proportion
0.25
0.20
0.15
0.10
0.05
0.00
0
10
30
50
70
Length (cm)

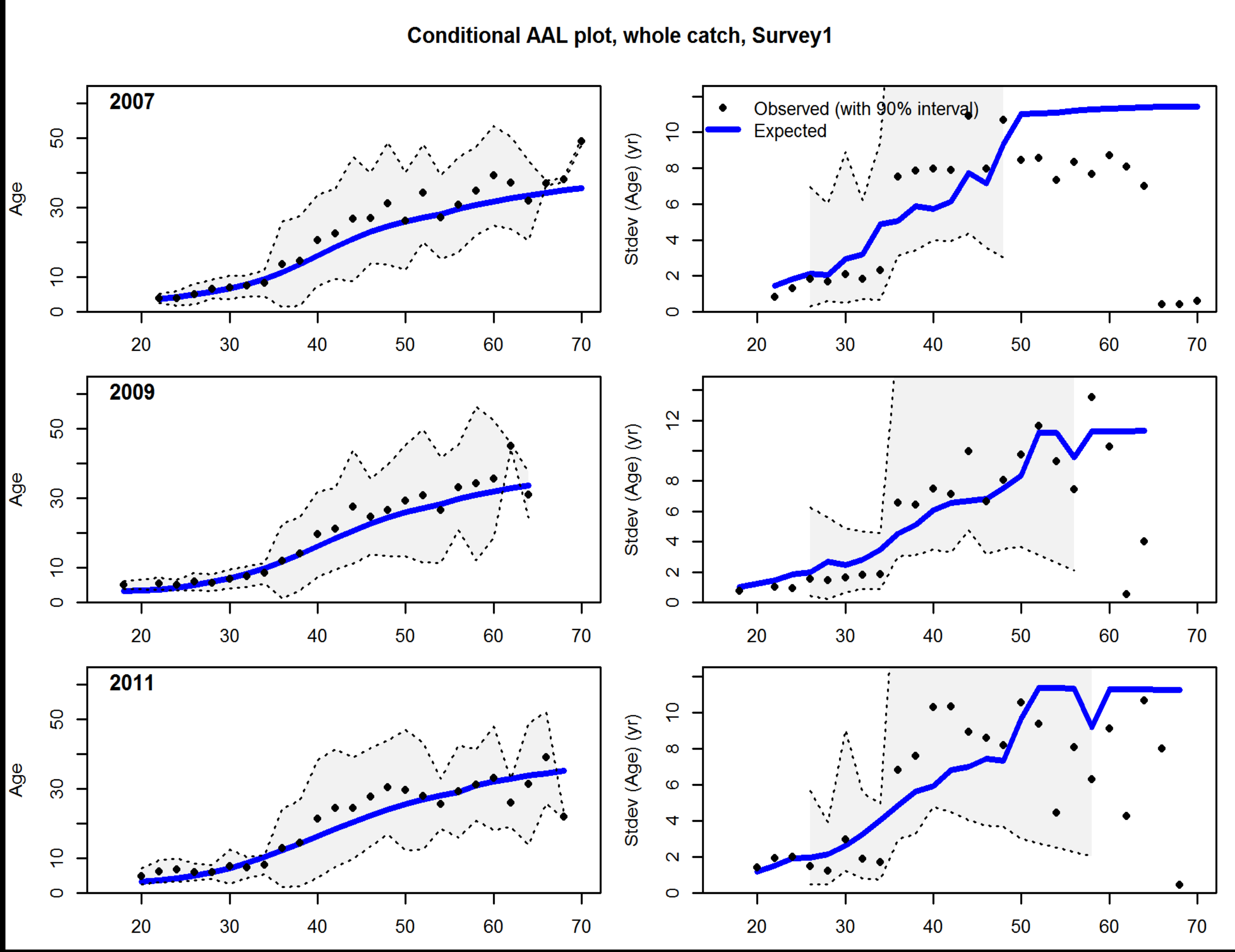


length comps, whole catch, Survey1
Proportion
Length (cm)
1999
N=720.2
effN=650.1
2009
N=823.6
effN=1198.8
2003
N=767.9
effN=1048.4
2011
N=610.8
effN=1483.4
2005
N=871.4
effN=969.7
2013
N=439.7
effN=363.6
2007
N=843.5
effN=1328.6
2015
N=674.4
effN=529
0.10 0.08 0.06 0.04 0.02 0.00 0.02 0.04 0.06 0.08 0.10
0 10 30 50 70

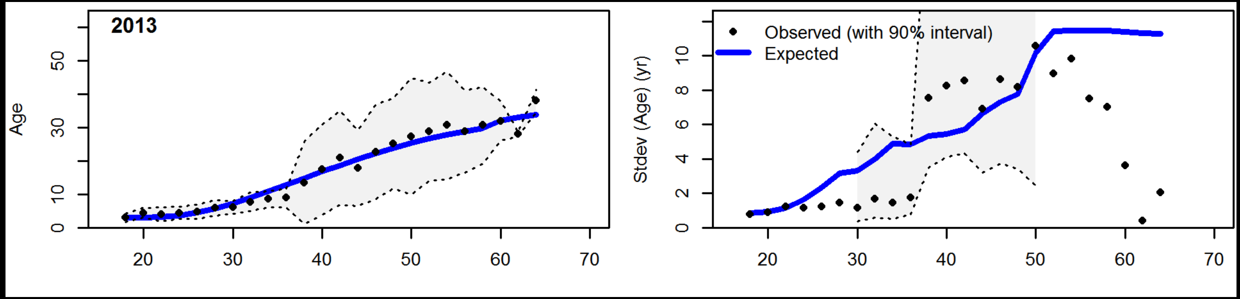

length comps, whole catch, Survey2
1990
N=135.2
effN=134.8
1993
N=219
effN=236
1996
N=281.4
effN=346.2
Proportion
Length (cm)
0
10
30
50
70
0.12
0.10
0.08
0.06
0.04
0.02
0.00



Conditional AAL plot, whole catch, Survey1
1999
2003
2005
Age
Stdev (Age) (yr)
Observed (with 90% interval)
Expected
Length (cm)

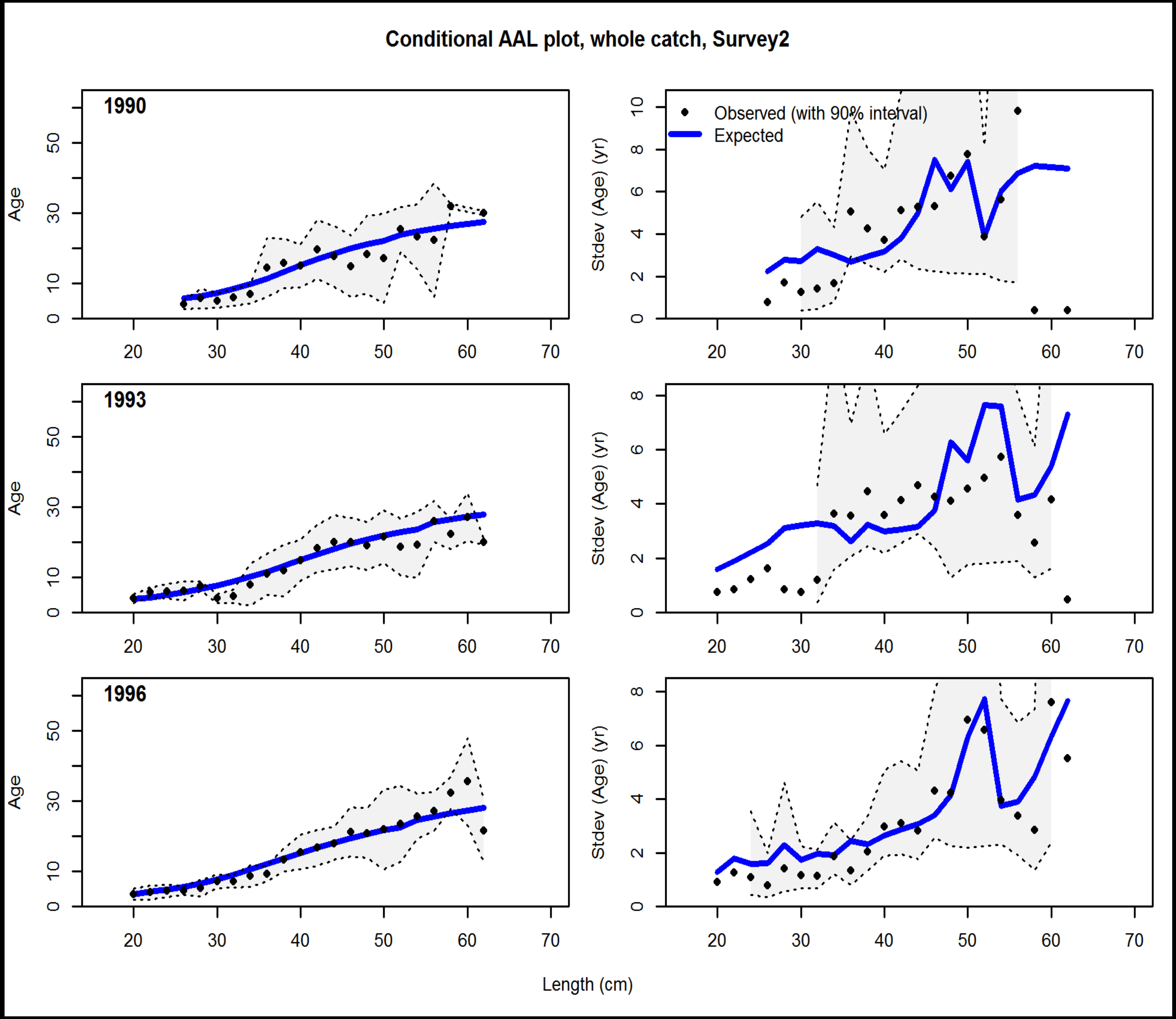


Conditional AAL plot, whole catch, Survey1
2007
2009
2011
Age
Stdev (Age) (yr)
Observed (with 90% interval)
Expected



2013
Age
Stdev (Age) (yr)
Observed (with 90% interval)
Expected

Length



Conditional AAL plot, whole catch, Survey2
1990
1993
1996
Age
Stdev (Age) (yr)
Observed (with 90% interval)
Expected
Length (cm)

Conditional age-at-length standard deviation plots

- Observed standard deviations are often low (or 0) for larger length bins because there are few samples (or 1 sample) in those bins
- Expected standard deviations at larger length bins are a direct function of the modeled numbers at age and length.
  - standard deviations reflect the model's interpretation of the population variability in ages within a length bin and not a standard deviation calculated from a sample.
- Variability in expected standard deviation can occur from year to year due to fluctuations in recruitment and fishing mortality

Francis (2011) Data Weighting Method

- Purpose:
  - Initial: to investigate whether effective sample sizes of fishery length comps were reasonable relative to effective sample sizes of survey composition data
  - To assign weights to composition data sources that account for the influence of intra-year correlations in length or age comps that are not explicitly modeled, to avoid preventing the model from fitting the biomass index well
- Examples of correlations not in the model: time-varying selectivity, time- and age-varying natural mortality
- Background:
  - Length and age comp data are often overdispersed relative to the variance assumed by the multinomial likelihood in the model
  - McAllister and Ianelli (1997), Appendix 2: calculates weights to account for overdispersed data relative to variance of the multinomial, ignores correlations
  - Pennington and Volstad (2004): Intra-haul correlation lowers effective sample size
    - E.g. fish of similar ages or lengths are often caught together in a haul
    - The precision of the mean lengths or ages based on a sample of fish from marine surveys is much lower relative to the precision of the mean length or age based on a random sample of the population
    - Precision for some marine surveys is close to the number of hauls, not number of fish
  - Francis (2011):
    - Same concept as for Pennington and Volstad, (measuring precision of means), except applied to intra-year correlations, rather than intra-haul correlations
    - Same idea as McAllister and Ianelli, but accounts for correlations by comparing variation in mean lengths or ages relative to expected means by year (where means are assumed to be normally distributed)
- Potential alternative: explicitly model time-varying effects that influence proportions at length and age so that residuals are not as correlated