Appendix I

Atlantic Herring Data Working Group meeting
January 30-February 3, 2012
Atlantic Herring Model Working Group meeting
April 9-April 13, 2012
Woods Hole, MA

Participants:

Jon Deroba – NEFSC - *Assessment Lead Scientist*
Gary Shepherd – NEFSC -*Working Group chair*
Mike Jech – NEFSC - *Acoustics*
Brian Smith – NEFSC - *Food Habits*
Laurel Col – NEFSC- *Marine Mammals*
Dave Richardson – NEFSC - *Icthyoplankton*
Larry Jacobson –NEFSC - *SS3*
Matt Cieri - ME DMF - *Catch*
Nick Markis – MIT –*OAWRS*
Jon Hare – NEFSC- *Oceanography*
Jason Link – NEFSC – *Ecosystems*
Steve Cadrin –SMAST –*Stock Structure*
Al Seaver - NEFSC
Andrew Cooper - Dept. of State
Bob Gamble – NEFSC
Chris Legault - NEFSC
Dan Hennen - NEFSC
Deb Palka - NEFSC
Fred Serchuk - NEFSC
Jeff Kaelin - Lund Fisheries
John Crawford - PEW
Julie Nieland - NEFSC
Kathy Sosebee -NEFSC
Liz Brooks - NEFSC
Loretta O'Brien – NEFSC
Lori Steele – NEFMC
Mark Terceiro – NEFSC
Mary Beth Tooley - O'Hara Fisheries
Micah Dean, MA DMF
Michael Fogarty - NEFSC
Michael Palmer - NEFSC
Paul Nitschke - NEFSC
Paul Rago - NEFSC
Peter Corkeron - NEFSC
Piera Carpi - SMAST
Purnima Ratilal - Northeastern Univ.
Rich McBride - NEFSC
Sarah Gaichas - NEFSC
Sean Lucey - NEFSC
Sigrid Lehuta - GMRI
Steve Weiner - CHOIR
Susan Wigley - NEFSC
Tom Dempsey - NEFMC
Tim Essington - Univ. Washington
Vincent Manfredi - MA DMF
Wendy Gabriel - NEFSC
Wenjiang Guan - Univ. Maine
Yong Chen - Univ. Maine

Appendix 2: Exploratory Stock Synthesis models for herring

Summary

Stock Synthesis (SS3) models were developed for herring to determine if incorporating length data directly into the assessment, modeling selectivity as a function of length and using other advanced features of SS3 would improve the stability and accuracy of stock size and mortality estimates for herring. We hoped that SS3 or a similar approach would facilitate modeling when age data are not available (e.g. in the terminal year or for an entire survey), help deal with changes in survey timing and growth and, in particular, reduce retrospective patterns. A large number of SS3 model runs were carried out but all SS3 estimates and results shown here are from a single demonstration run.1

These SS3 results shown here were not completely reviewed by the Coastal Pelagic Working Group (WG) and are not useful for management purposes. The best use of this information is in identifying modeling approaches that might be useful in future. Both SS3 and the current assessment model (ASAP) were originally intended for use in working group deliberations. However, the lead stock assessment scientist and Working Group were unable to review the SS3 model configuration, resolve all data and modeling questions or consider results in the available time.

Based on preliminary results, the focus in modeling on length data and SS3 model configuration appear promising because retrospective patterns were reduced without having to make assumptions about high natural mortality during recent years (Figure A2-1). Survey and fishery selectivity appear to be a function of size with the exception of young fish in coastal waters that are not found in offshore fisheries and surveys. It was possible to estimate time varying growth parameters that were similar to external estimates. Size data, time varying growth and estimation of size selectivity curves helped accommodate changes in survey timing and effects of changes in growth on selectivity. Fit to most data sources was good and it was possible to use survey data when ages were unavailable without assuming an age selectivity pattern.

SS3 configuration of SS3 for herring is summarized in Table A2-1. Data are summarized in Figure A2-2. Suggestions for future modeling and information about details with explanations follow.

Suggestions for future modeling

Historical catch data are required in SS3 and can be important because the model was originally designed for long-lived groundfish assumed to have been reduced from the virgin state to some initial level based on an average annual historical catch level. In this way, model stability was increased because the estimate of virgin biomass, the estimated spawner recruit curve (which can be used to independently calculate virgin biomass as in the ASAP model), MSY reference points (which are linked to the spawner-recruit curve and virgin biomass) and assumptions about historical catch are interdependent. This approach may be misleading and inappropriate for dynamic short lived fish like herring that experienced long periods of significant and variable amounts of fishing pressure prior to the onset of the modeled time period. The effect of this potential problem on preliminary SS3 estimates was not evaluated.

In future, it would be useful to try reducing the importance of historical catch data by

---

1 The SS3 run shown here was identified as the "Cadillac" run in working group meeting documents.

establishing very weak priors for historical fishing mortality parameters and by estimating recruitment offset parameter available in the model. The weak priors for fishing mortality parameters would effectively mean that the historical catch data were imprecise allowing the model to estimate initial stock size to maximize fit to the available data, rather than correspondence between virgin and initial stock size. The recruitment offset parameter effectively rescales the spawner-recruit curve during the historical period so that virgin and initial stock sizes are not directly linked by the spawner-recruit curve used elsewhere in the model and so that initial stock size is estimated to maximize fit to the available data.

These assumptions about ageing errors are based on recent QA/QC experiments and probably understate the actual imprecision of herring age data, particularly for older individuals and because they ignore possible changes in ageing criteria over time. It may be advisable to carry out historical and current age reader experiments that compare ages from the same otoliths collected by historical and current age readers.

A prior on the variance of spawner-recruit residuals from Overholtz et al. (2004) was used in SS3 but probably incorrectly. It might be advisable to assume more temporal variability in catchability or, perhaps, selectivity parameters when modeling the fall survey prior to 1985 when the survey doors changed (Figure A2-19 and see below). Historical catch estimates should be refined in possible.

<u>Details and additional explanation</u>

All of the likelihood weights used in fitting SS3 was zero. Some adjustments were made to assumed sample size and variances based on preliminary fits. A total of 190 parameters were estimated in SS3 (see below). Most of parameters were annual deviations in the von Bertalanffy growth parameters L_{max} and K. Selectivity curves required a relatively high number of parameters because there were seven surveys and four fisheries, length selectivity was often domed and because logistic selectivity at age was estimated in addition to selectivity at length for offshore fisheries and surveys that do not capture young herring of any size.

Parameter type	N parameters
Natural mortality and growth	5
Growth deviations (L_{max} and K)	78
Spawner-recruit	2
Recruit deviations	47
Historical fishing mortality	4
Survey catchability	4
Size and age selectivity	50
Total	190

“Exact” instantaneous fishing mortality rates during the modeled time period were calculated in SS3 using they hybrid method because Pope-type approximations may be inaccurate when mortality rates are high. With this approach, catch data are fit exactly (Figure A2-3). In contrast, SS3 uses fishing mortality rate parameters (one per fishery) to fit assumed levels of average historical catch that link virgin stock size to initial stock size in the model.

Four fisheries defined in SS3 were defined in terms of gear and season. In particular, we modeled the fixed gear (nearshore) semester 1 (January-June) and semester 2 (July-December), and mobile gear (offshore) semester 1 and semester 2 fisheries separately. Length and age data were available for all years in the mobile gear fisheries. Length and age data were used for the fixed gear fisheries if sampling was sufficient and included data from the US component. Commercial length data for herring appear to be informative (Figure A2-4).

The SS3 run shown here treated fall and spring surveys carried by the NOAA Research Vessel Albatross IV and Delaware II prior to 2009 and fall and spring surveys carried out by the

NOAA Research Vessel Bigelow during 2009-2011 as separate surveys, even though the Bigelow series were only three years in length. In the basecase ASAP run, Bigelow catches were calibrated to Albatross equivalents and used to extent the Albatross time series through 2011. The standard approach was not used in SS3 to determine the shape of Bigelow survey selectivity curves and if three years of data were sufficient to start a new bottom trawl survey time series. Results for size data in the Bigelow spring survey (see below) suggest that the Bigelow survey time series are too short (3 years) at this time to by analyzed separately as uncalibrated time series.

In addition to the spring and fall Albatross and Bigelow bottom trawl survey data series, we used the winter bottom trawl and shrimp survey time series. Length data were available for all surveys and fisheries and appear informative (Figure A2-5). Age composition data were available for all years and all surveys except for Bigelow fall survey during 2011 and in all years for the shrimp survey.

Based on NEFSC routine QA/QC age reader experiments, age data in SS3 were assumed to have unbiased measurement errors that increased with age (Figure A2-6). The standard deviation of errors in the age data was assumed to be 0 y at age zero and increased linearly from 0.09 y at age one to 0.83 y at ages 11+.

The NEFSC fall bottom trawl survey for herring is difficult to interpret because the fall survey does not cover the entire herring stock so that seasonal migration patterns and overlap between the stock and survey may be variable and time dependent. Mean Julian dates of the fall NEFSC bottom trawl survey tows used for herring increased by roughly 30 days during 1963-1984 while bottom temperatures increased by about 3^{o} C (Figures A2-7 and A2-8). Fall sea surface temperatures increased during 1963-1985 and declined afterwards (Figures A2-8). Mean length at age in the fall and spring surveys declined beginning in the mid-1980s as growth apparently slowed to relatively low levels in recent years. Herring grow quickly, particularly at small sizes, and a 30 day delay in survey timing, additional growth, migratory movements and changes in temperature may result in substantial and continuous changes to fall survey catchability and selectivity at age if these parameters are actually functions of size when the survey is conducted.

The changes in survey timing, water temperatures and growth correspond and are probably aliased with the switch from BMW to Polyvalent bottom trawl survey doors in 1984-1985. Based on visual examination of trends and model results, the door change had a major effect on fall and spring survey catchability. Potential door effects on survey selectivity are not clear.

Random walks were used in SS3 to deal with continuous or abrupt changes in growth, selectivity and catchability parameters, particularly in the fall survey. In particular, fall and spring survey catchability parameters were allowed to change abruptly in 1985 (assuming a large variance on the deviation for 1985) to account for the door change. We also experimented with letting the fall survey catchability parameter follow a slow random walk during 1968-2006.

It is very important to use good estimates of growth in models that use size data. We modeled the growth parameters K and L_{max} using a random walk during 1968-2006 because we hypothesized that the changes in size at age (growth) and size selectivity might be sufficient to capture many of the effects of changes in the fall survey and water temperatures on size and selectivity at age. SS3 was able to estimate complicated temporal growth parameters that matched estimates made externally from the same data (Figure A2-9 and A2-10). The growth parameter t_0 was constant and modeled as an estimated parameter.

At the outset, we tried to use estimate selectivity at size only when fitting the SS3 model to survey and fishery length and age composition data. In SS3, selectivity at age S_a is a function of selectivity at length S_L:

$$S_a = s_a \sum_L \frac{S_L N_{L,a}}{N_{+,a}}$$

where s_a is selectivity at age ignoring size, $N_{L,a}$ is the estimated population abundance of herring that are age a and length L in the current time step and$N_{+,a} = \sum_L N_{L,a}$. Thus, $\frac{N_{L,a}}{N_{+,a}}$ is one element in the estimated population age-length key and the term in the summation on the right is mean selectivity at size for age a. In SS3 modeling, we initially assumed A_a=1 for all ages in all surveys and fisheries so that only size selectivity was important. However, it proved necessary to estimate logistic selectivity at age curves as well for all of the fisheries and surveys (except shrimp with no age data) because virtually no age one herring of any size are taken in any fishery or survey.

We experimented with random walks for survey selectivity parameters in the fall survey prior to 1985 and abrupt changes in survey size selectivity parameters during 1984-1985 but these approaches did not appear necessary as long as the model allowed for temporal variation in size at age and door effects on survey catchability.

The commercial and survey size selectivity curves for herring were logistic or dome shaped (Figure A2-11) and the decision about which type of curve to use was usually obvious on inspection of the corresponding size and age composition data and after preliminary model runs. The offshore mobile gear fisheries as well as shrimp and winter bottom trawl surveys which catch very large herring in greatest numbers had logistic shape size selectivity while all other fisheries and surveys had dome shaped size selectivity indicating that large herring are hard to catch in survey bottom trawls. The estimated age selectivity curves in SS3 were all logistic with nearly 100% selectivity at ages two to four years (Figure A2-12).

With the exception of the spring Bigelow survey, the SS3 model fit commercial and survey size and age composition data well (Figure A2-13 and A2-14). The spring Bigelow survey had a surprisingly high number of small herring during 2010-2011 (Figure A2-15). We hypothesize that the data for 2010-2011 were anomalous and distort the average size composition for the short spring Bigelow survey. In contrast to the spring survey, relatively low numbers of small herring were taken in the fall Bigelow survey as well as in the original Albatross spring survey. Also, paired tow vessel calibration data collected by the two vessels did not show the same pattern. Additional years of survey data will probably be necessary to clarify the size composition and selectivity of the spring and possibly fall Bigelow surveys.

Very large changes in survey catchability during 1984 and 1985 were required to fit the spring and fall survey trends. Catchability increased from about 79 to about 325 (by 410%) in the spring survey and from about 3.6 to about 154 (by 4280%) in the fall survey (Figure A2-16). Thus, the remarkably low herring catches prior to the door change appear due primarily to very low survey bottom trawl catchability.

Fit to the spring bottom trawl survey trend was good (Figure A2-17). The SS3 model fit the spring and fall Bigelow surveys well although the short time series show different trends (Figure A2-18). The model fit fall bottom trawl survey trend reasonably well after accommodating the change in catchability but there was a tendency for the model to over predict the survey in the years prior to the door change (Figure A2-19). For the fall survey, it might be better to build more temporal variability in catchability or, perhaps, selectivity parameters during years prior to the door change. The observed and predicted winter survey values seem poorly correlated (Figure A2-20). The model fit the shrimp survey trends reasonably well with the exception of the three earliest years (1982 and 1985-1986, Figure A2-21).

Recruitment estimates from SS3 suggest that the high biomass and productivity during the early 1960s may have been to a few years of unusually good recruitment (Figures A2-22 and A2-23). The assumption of a Beverton-Holt recruitment curve appears reasonable.

Fishing mortality is complicated to quantify in the SS3 model for herring because there are four fisheries with markedly different selectivity patterns. For simplicity, fishing mortality was quantified as total annual catch biomass divided by age 1+ biomass on July 1 (Figure A2-24). This simple calculation accommodates differences in fishery selectivity, seasonal growth and seasonal population dynamics.

Spawning biomass estimates from SS3 differ markedly from the ASAP basecase estimates (Figure A2-25). Comparisons are difficult, however, because assumptions about natural mortality in recent years are very different in the two models.

References

Northeast Fisheries Science Center. 2008. 47th Northeast Regional Stock Assessment Workshop (47th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 08-12a; 335 p.

Northeast Fisheries Science Center. 2010. 49th Northeast Regional Stock Assessment Workshop (49th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 10-03; 383 p.

Overholtz, W.J.; Jacobson, L.D., Melvin, G.D., Cieri, M., Power, M., Libby, D., Clark, K. 2004. Stock assessment of the Gulf of Maine - Georges Bank Atlantic herring complex, 2003. Northeast Fish. Sci. Cent. Ref. Doc. 04-06, 290 p.

Table A2-1. Summary of SS3 model configuration for herring.

Item	Descriptor	Note
Years covered	1963-2011	All years with survey data
Seasons	2	Season 1 = January-June, Season 2 = July-December
Number areas	1	
Number sexes	1	
Number "morphs"	1	
Lengths	4-35 cm	
Length bins	1 cm	
Ages	0-15+ y	
Age bins	1 y	
Commercial fleets	4	Mobile gear season 1, Mobile gear season 2, Fixed gear season 1, Fixed gear season 2
Commercial selectivity at length	Mobile S1	Logistic
	Mobile gear (S2)	Logistic
	Fixed gear S1	Domed
	Fixed gear S2	Domed
Commercial selectivity at age	Mobile S1	Logistic
	Mobile gear (S2)	Logistic
	Fixed gear S1	Not used (one for all ages)
	Fixed gear S2	Not used (one for all ages)
Assumed historical catch (pre-1963)	96171 mt	Prorated by fleet based on proportions by mobile and fixed gear fleets during 1964 (US and Canada). Fleet values broken down by semester based on US&CA data (season 1) or US data only (season 2)
Fishing mortality	Instantaneous rates	Hybrid method
Survey data (mean N/tow, vessel correction factors applied but no Albatross-Bigelow calibration factors)	Winter	1992-2007
	Spring	1968-2008 (before the R/V Bigelow) with length and age data for all years
	Spring Bigelow	2009-2011 with length and age data for all years
	Shrimp	1983-2011 with length data for all years (no ages)
	Fall	1963-2008 (before the R/V Bigelow)
	Fall Bigelow	2009-2011 with length and age data except ages unavailable for 2011

Survey selectivity at length	Winter	Domed
	Spring	Domed
	Spring Bigelow	Domed
	Shrimp	Logistic
	Fall	Logistic
	Fall Bigelow	Domed
Survey selectivity at age	Winter	Logistic
	Spring	Logistic
	Spring Bigelow	Logistic
	Shrimp	Not used (one for all ages)
	Fall	Logistic
	Fall Bigelow	Logistic
Survey catchability	Winter	Median unbiased (calculated internally)
	Spring	Random walk (very low variance) except for 1984 (higher variance) to accommodate door change (breaks the time series trend while using the same selectivity curve for early and late periods), base and deviation parameters estimated
	Spring Bigelow	Median unbiased (calculated internally)
	Shrimp	Median unbiased (calculated internally)
	Fall	Same as spring
	Fall Bigelow	Median unbiased (calculated internally)
Ageing errors	Based on NEFSC ageing QA/QC experiments	Unbiased with standard deviations that increase with age from 0.09 y at age 1 to 0.838 y at ages 12+
Natural mortality	Average of natural mortality rates at age used in the ASAP model	Constant over time but increase at age from 0.66 y-1 at ages 0 and 1 to 0.22 y-1 at age 13+
Mean size at age (growth)	von Bertalanffy	t_0 estimated, K and L_{max} follow random walk during 1968-2006 with estimated deviations (sd=1)
Variability in size at age	Standard deviation a linear function of length at age	Standard deviation for size at age 1 and at L_{max} estimated
Maturity at age	Assumed	From earlier stock assessment
Spawner-recruit relationship	Beverton and Holt	R_0 estimated, steepness fixed at 0.85, variance estimated with lognormal prior (mean 0.904, sd=1.010, based on meta-analysis in Overholtz et al. 2006) - **This was probably not done correctly.**
Years with freely estimated recruitments	1959-2005	Earlier and later years from spawner-recruit model
Likelihood weights	All one (1.0)	Used to weight each term in the negative log likelihood

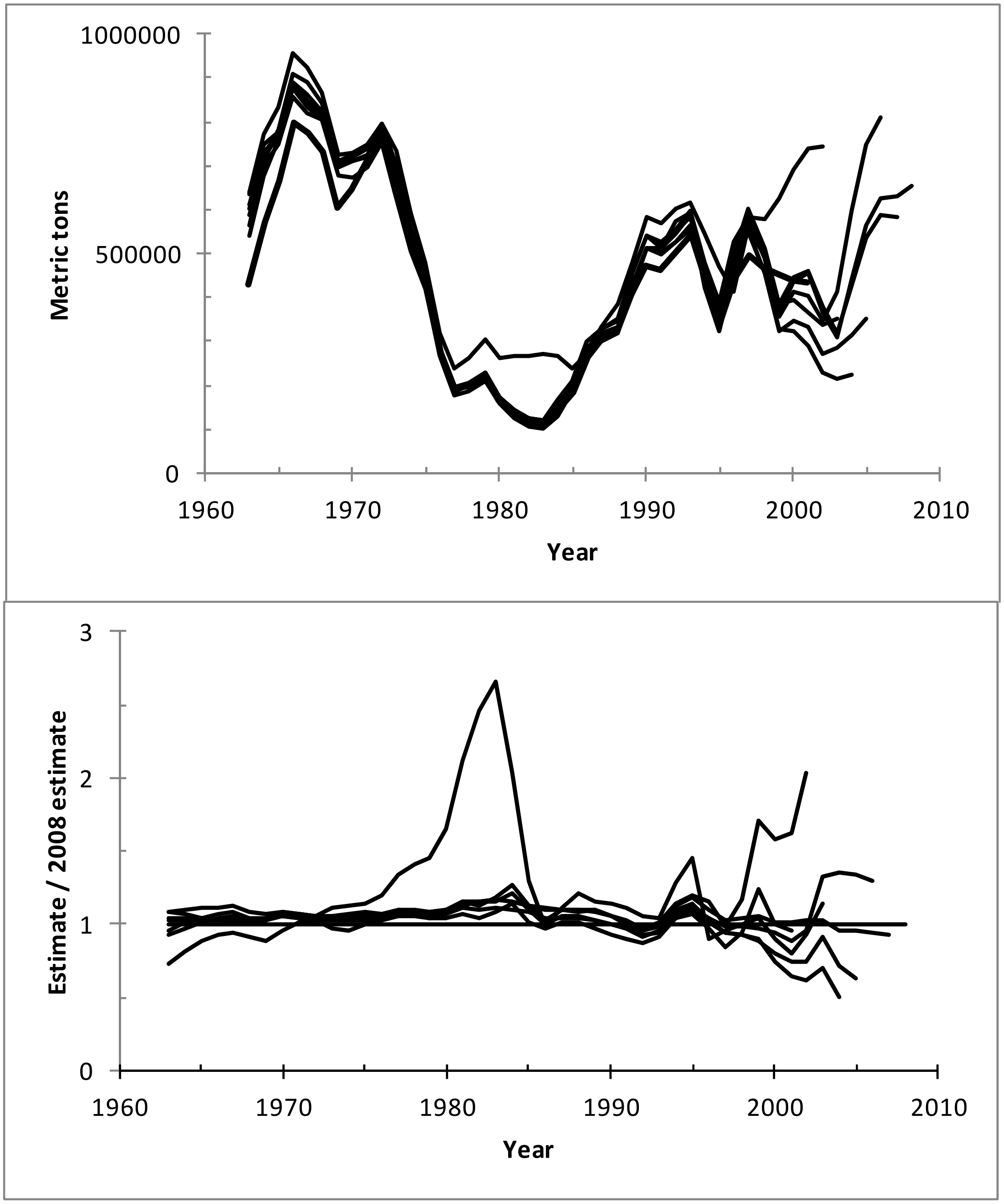


Figure A2-1. Retrospective analysis for herring spawning stock biomass estimates from SS3. The terminal year was 2008 to avoid inconsistencies using in the retrospective analysis due to the short 2009-2011 Bigelow surveys.

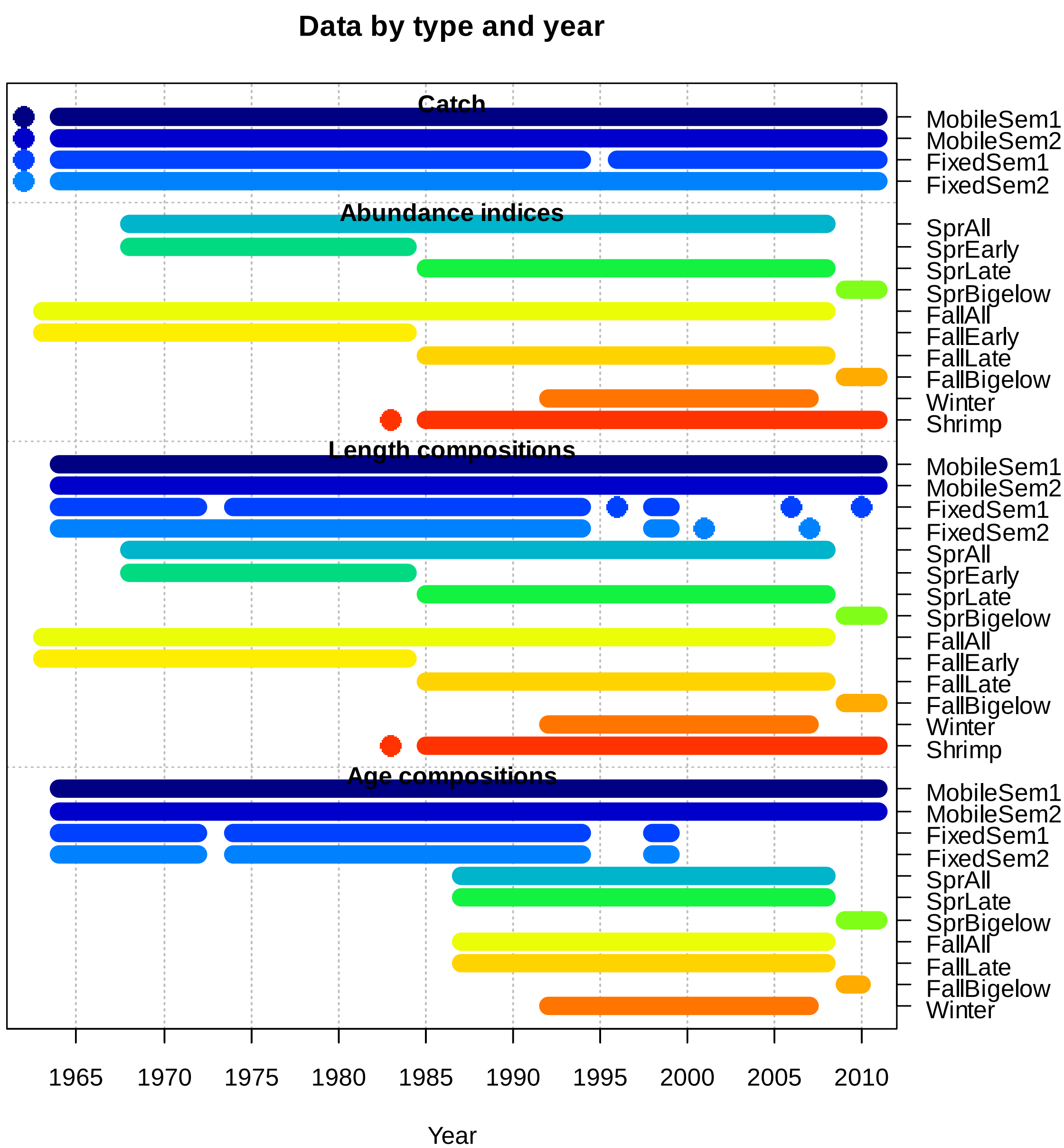


Figure A2-2. Summary of commercial and survey data for herring used in SS3. The surveys SprEarly, SprLate, FallEarly and FallLate (spring and fall surveys separated at 1984/1985 to accommodate survey door changes as in ASAP) were included in data files but were not used in the SS3 run shown here.

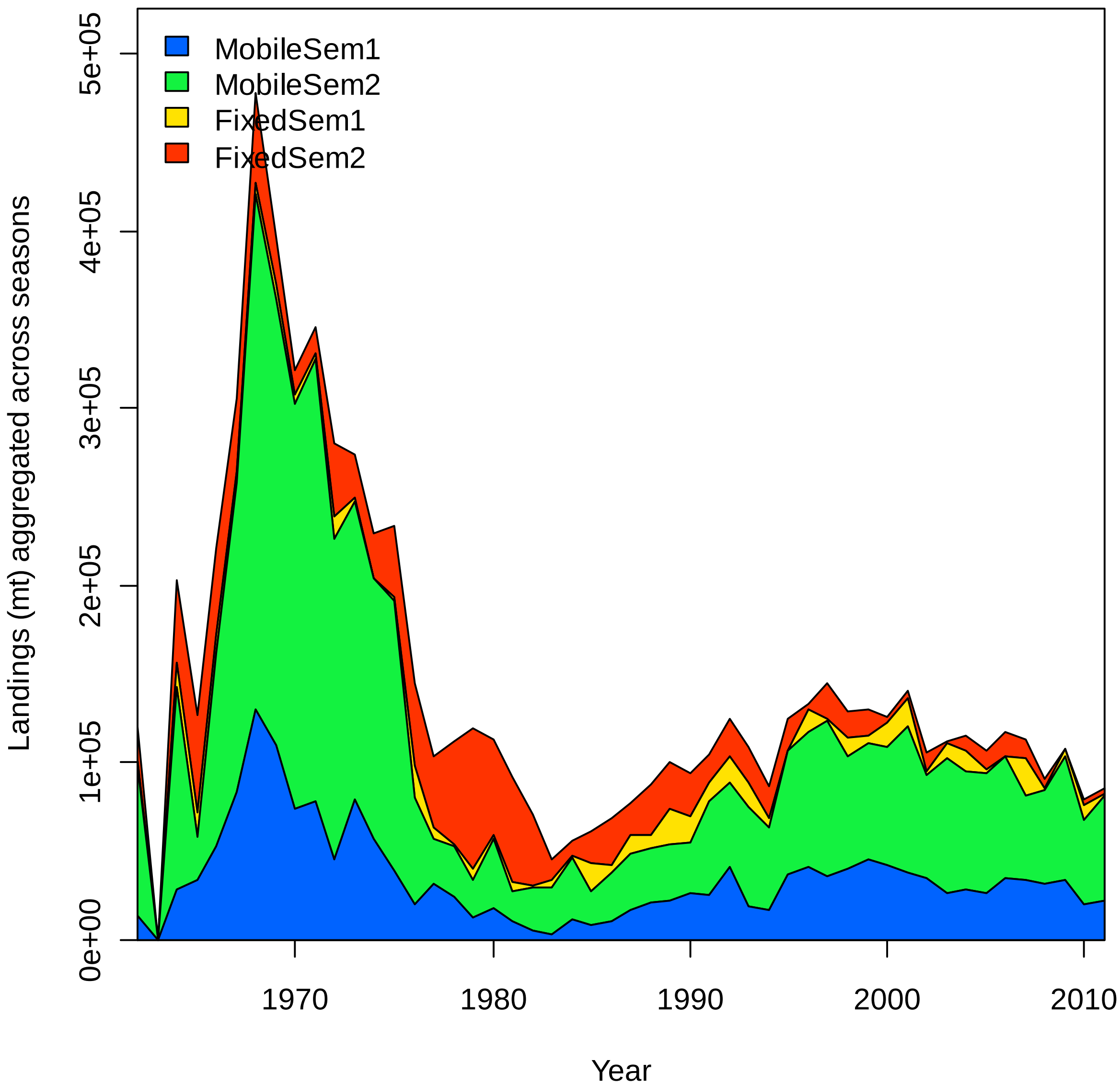


Figure A2-3. Commercial catch data for herring by fleet and season during 1963-2011 as used in the SS3 model.

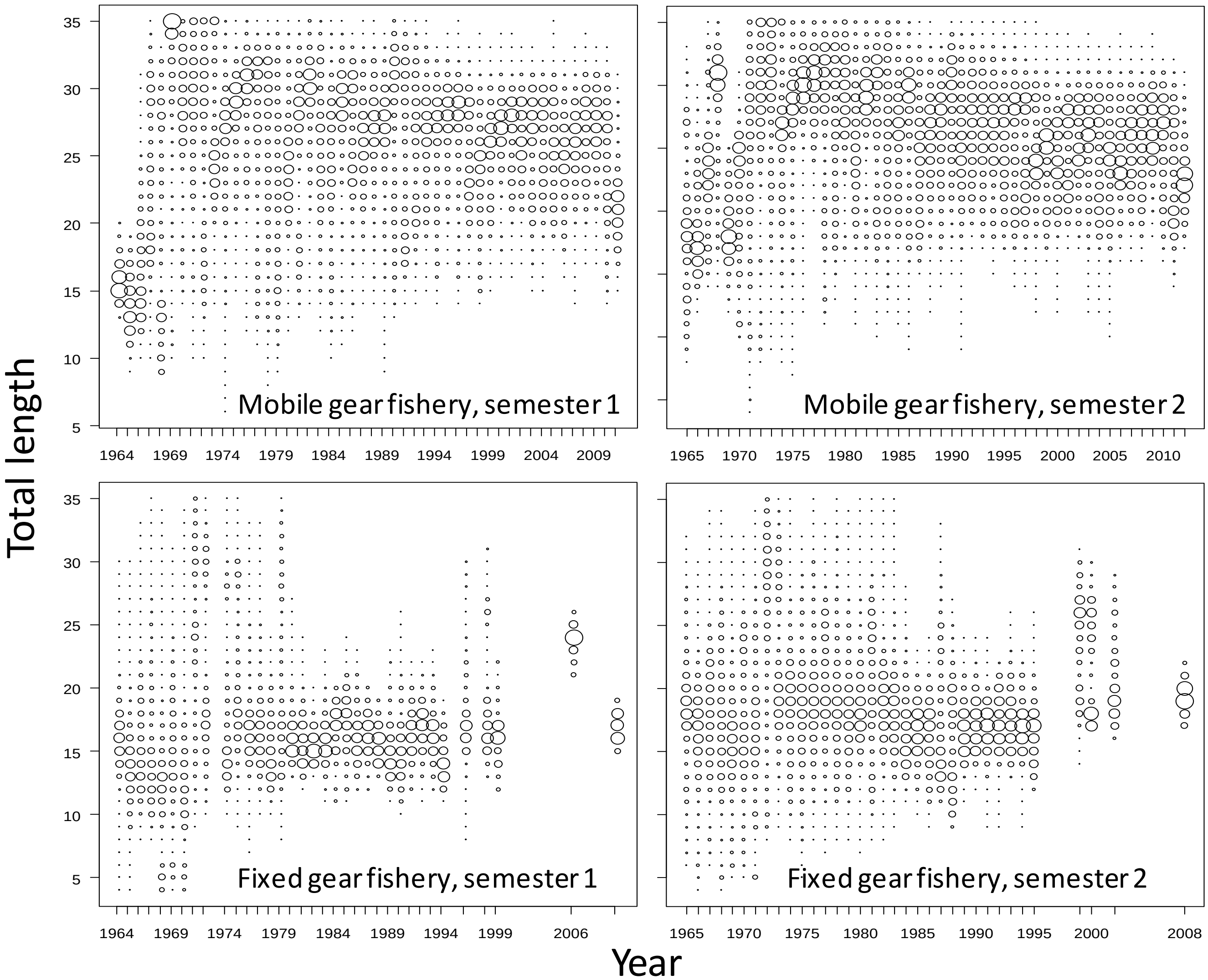


Figure A2-4. Commercial size composition data for herring used in SS3.

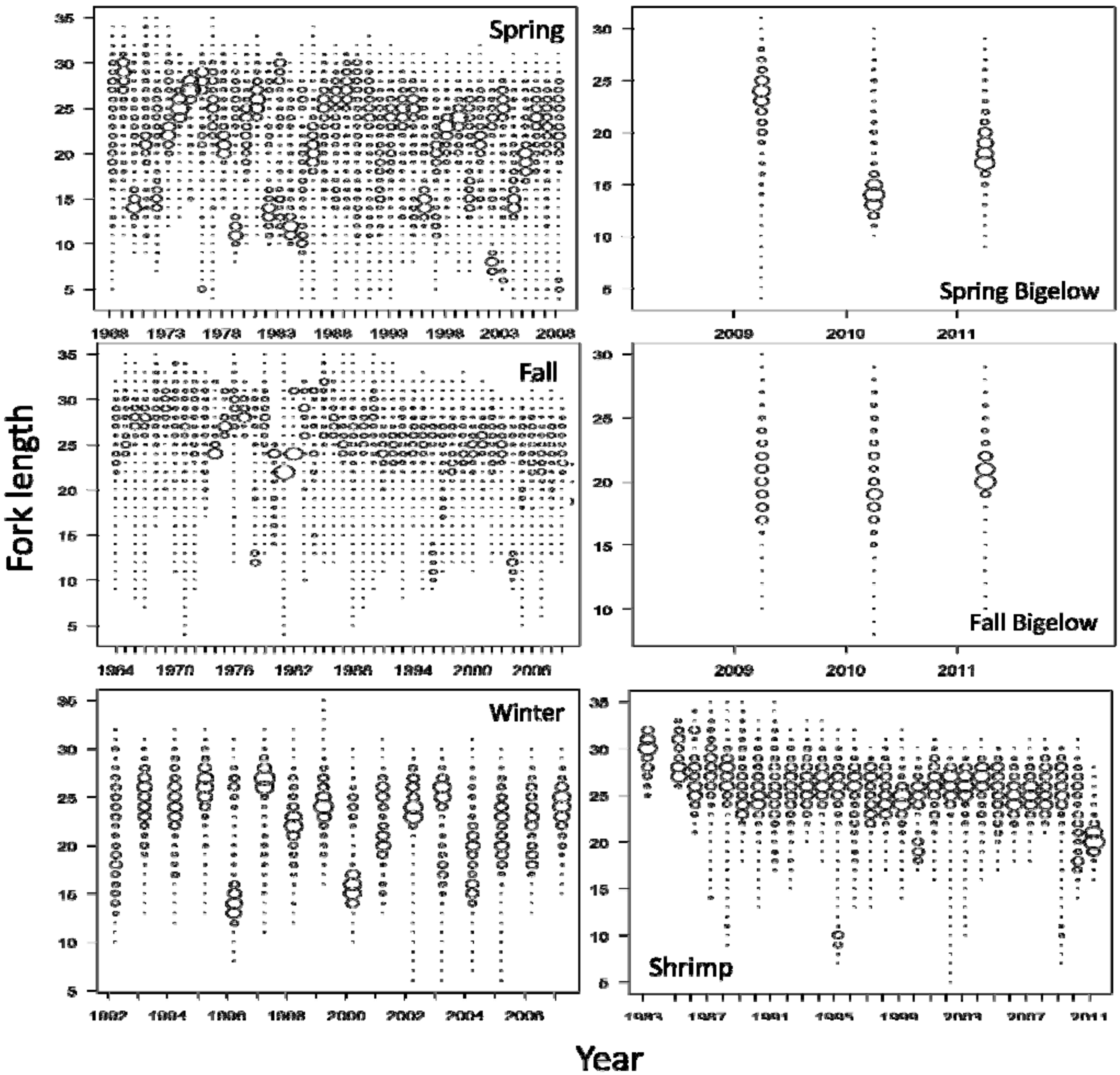


Figure A2-5. Survey size composition data for herring used in SS3.

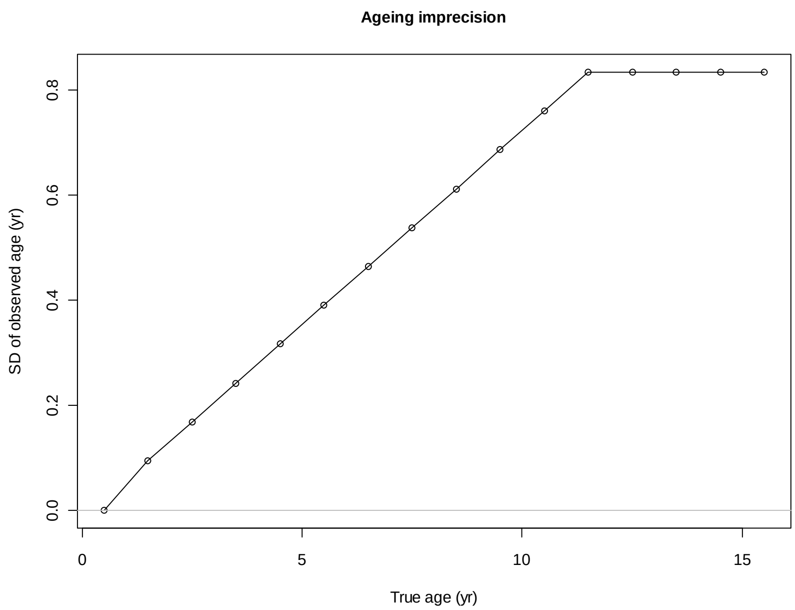


Figure A2-6. Assumed standard deviations for ageing imprecision in herring assumed in SS3.

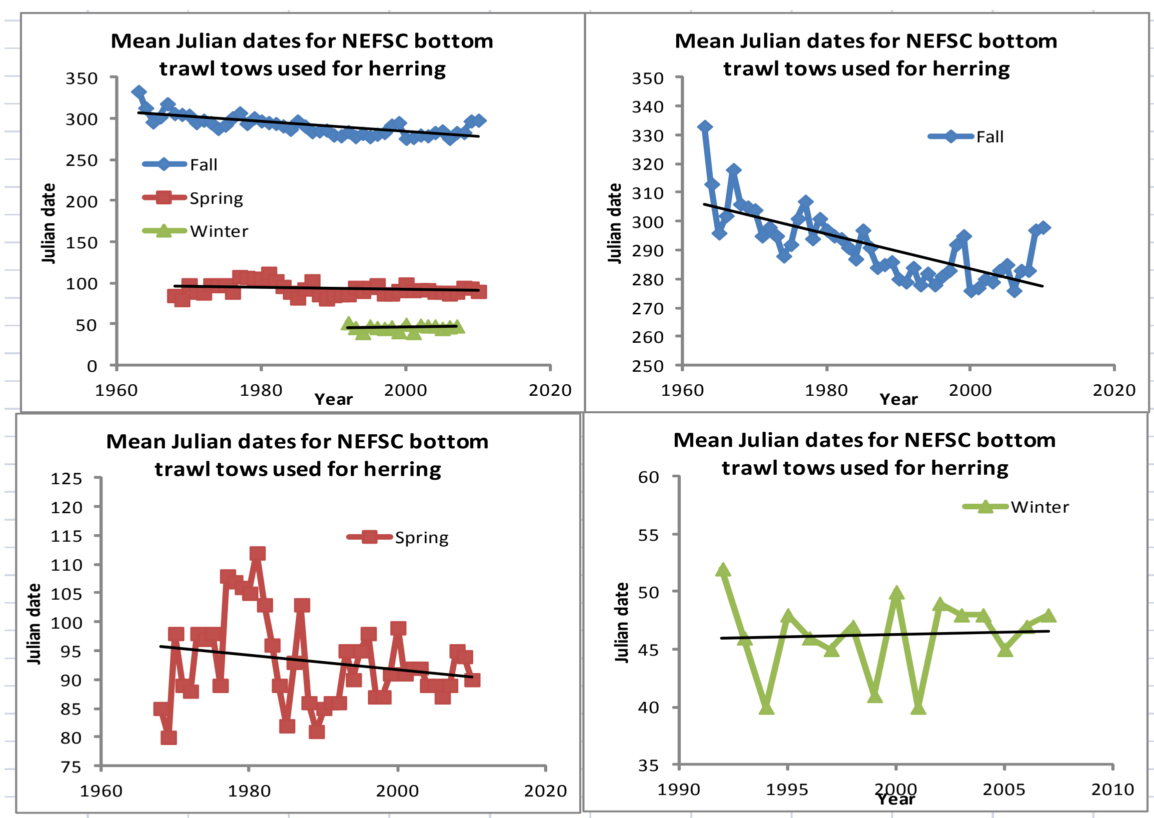


Figure A2-7. Mean annual Julian dates used for bottom trawl survey tows used for herring in SS3.

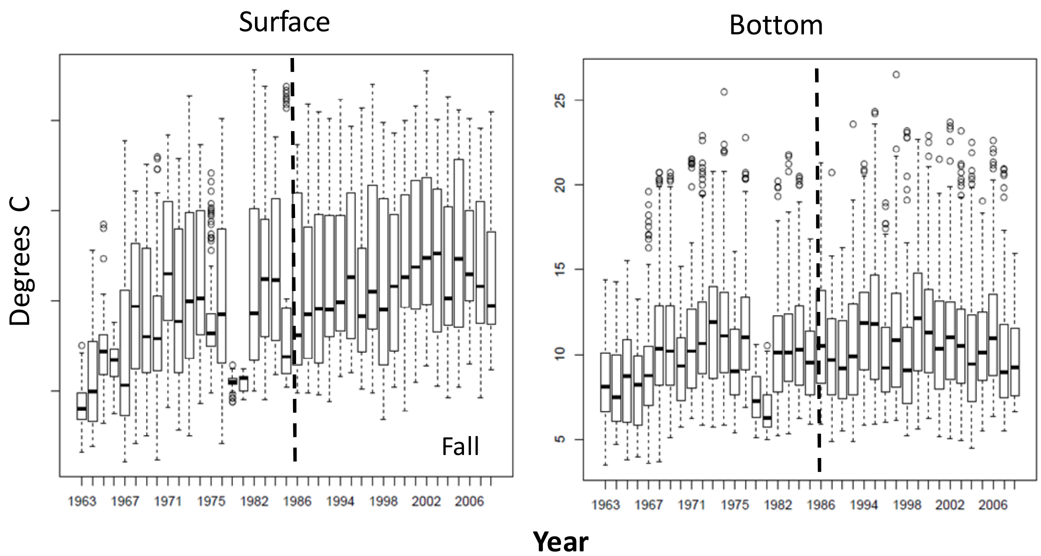


Figure A2-8. Surface and bottom temperatures for NEFSC fall survey tows used in the herring assessment. The short dark horizontal lines are the median temperatures. The dash vertical line shows the change in bottom trawl survey doors during 1984/1985.

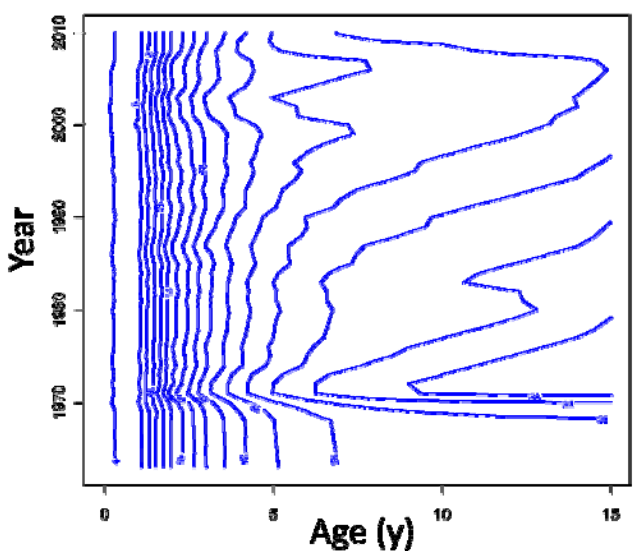


Figure A2-9. Estimated size at age in the SS3 model for herring during 1963-2011 based on von Bertalanffy growth curves with random walk parameters.

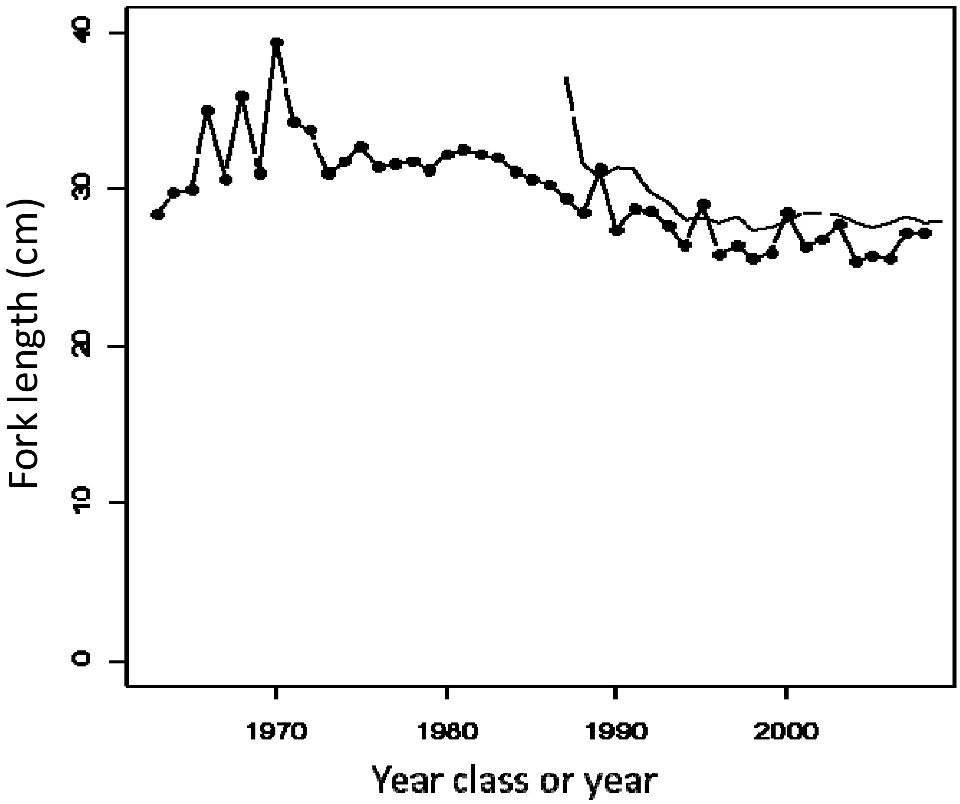


Figure A2-10. Von Bertalanffy L_{max} parameter estimates for herring from SS3 (January 1, solid symbols) and from growth curves fit externally to spring survey data. The SS3 estimates are by year class while the external estimates are by calendar year.

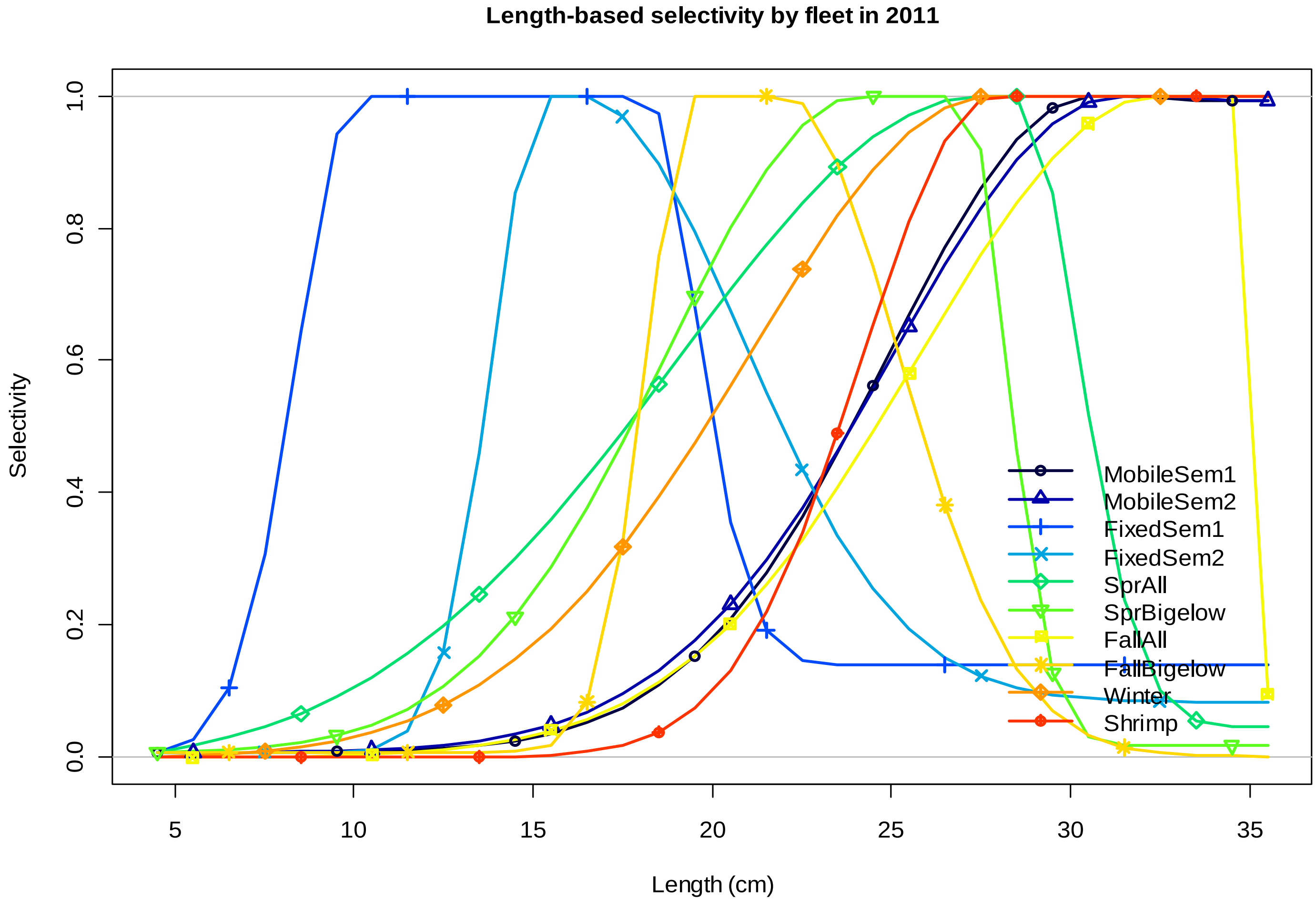


Figure A2-11. Selectivity at length curves for herring in commercial fisheries and surveys estimated in SS3.

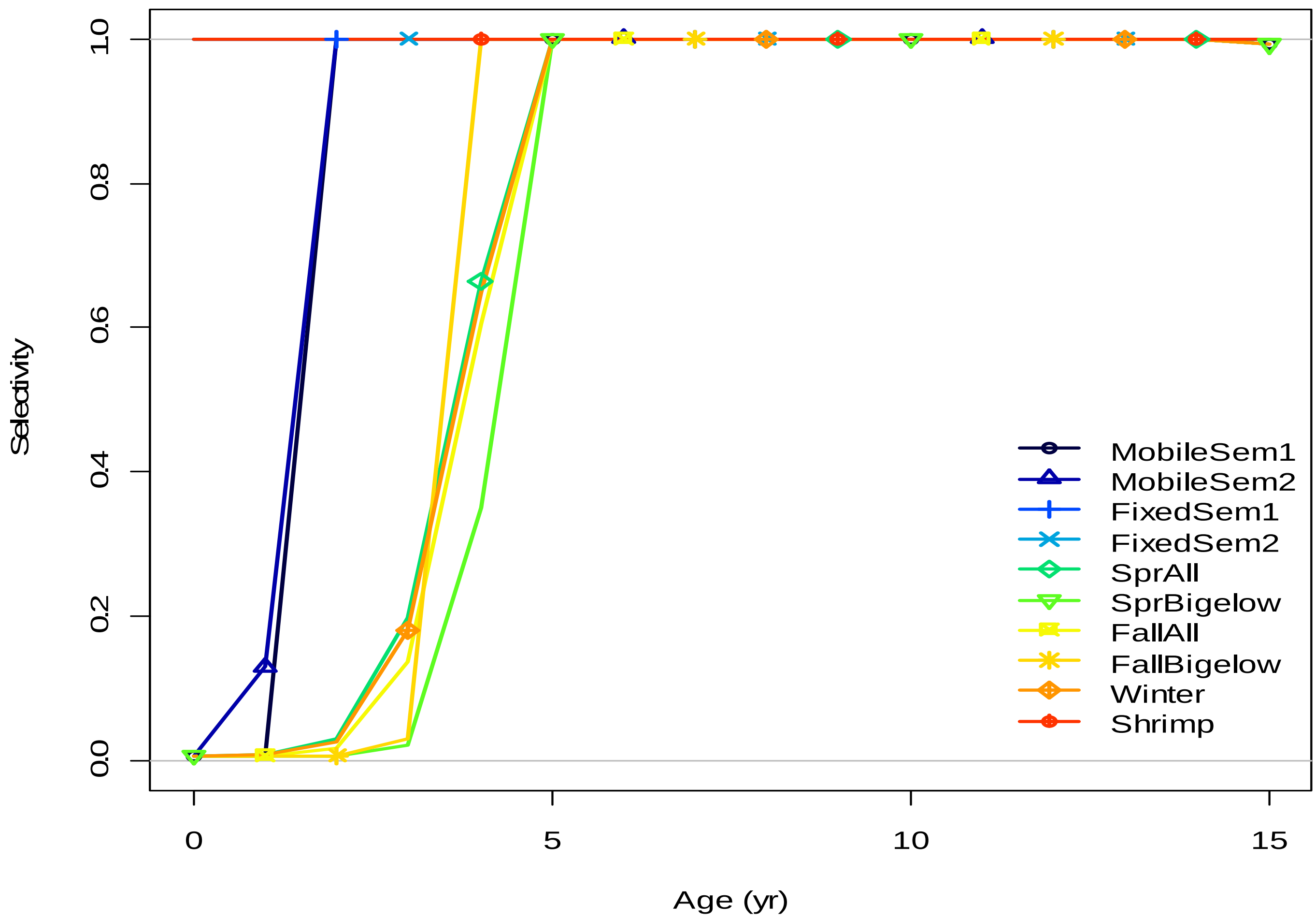


Figure A2-12. Selectivity at length curves for herring in commercial fisheries and surveys estimated in SS3.

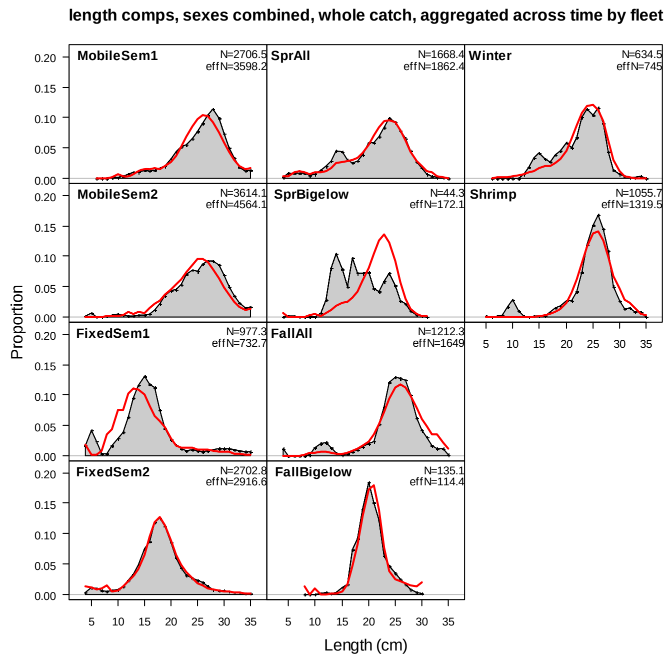


Figure A2-13. Average commercial and survey length composition data (in grey) and average predicted values (red line) for herring in the SS3 model.

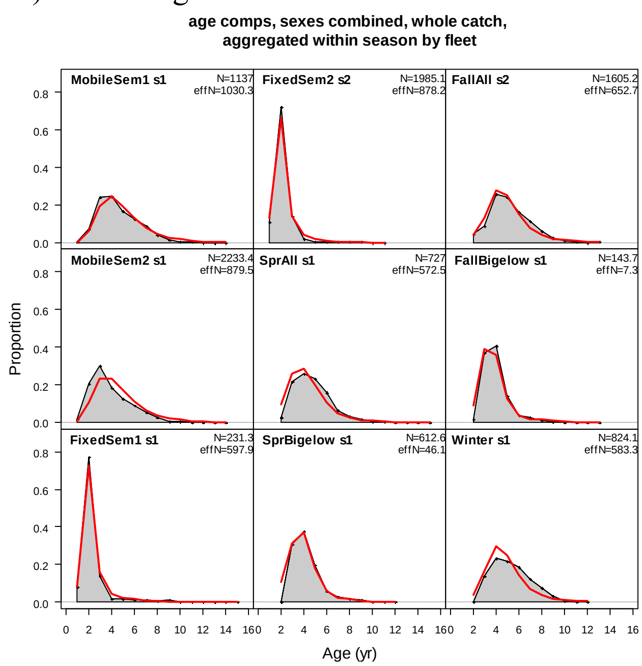


Figure A2-14. Average commercial and survey age composition data (in grey) and average predicted values (red line) for herring in the SS3 model.

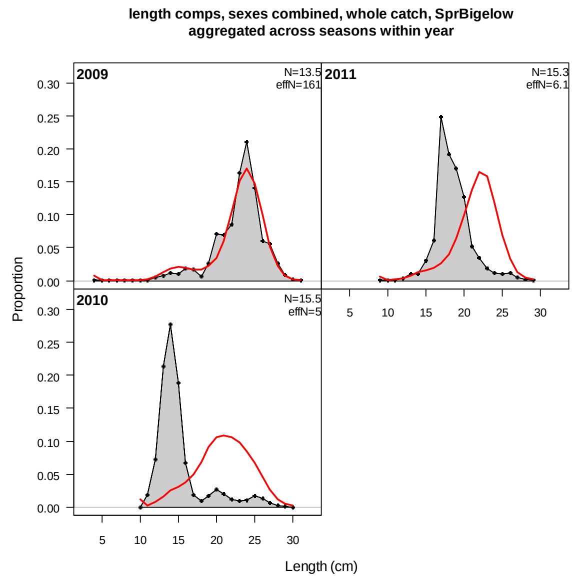


Figure A2-15. Annual observed spring Bigelow survey size composition data (in grey) for herring with predicted values (red line) from the SS3 model for herring.

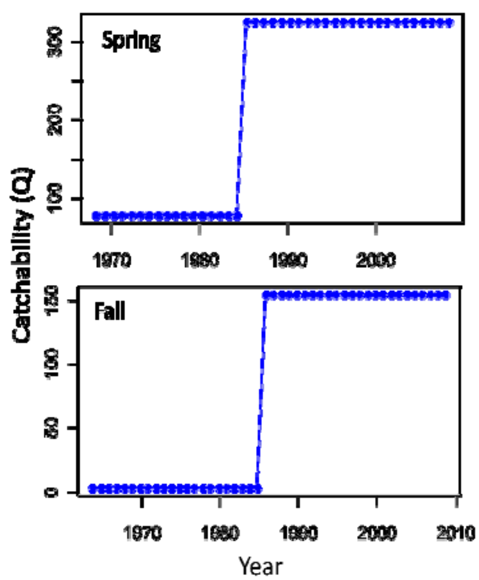


Figure A2-16. Changes in catchability for herring in the spring and fall bottom trawl surveys estimated in SS3.

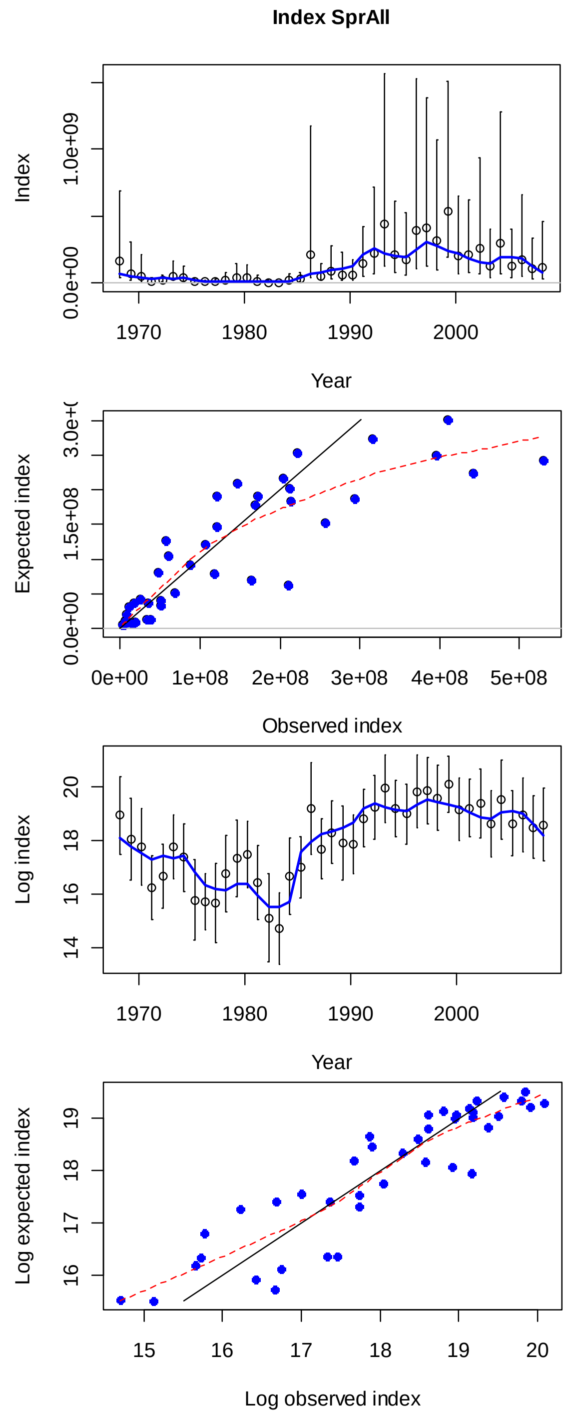


Figure A2-17. Goodness of fit plots for the SS3 model and herring in the NEFSC spring bottom trawl survey.

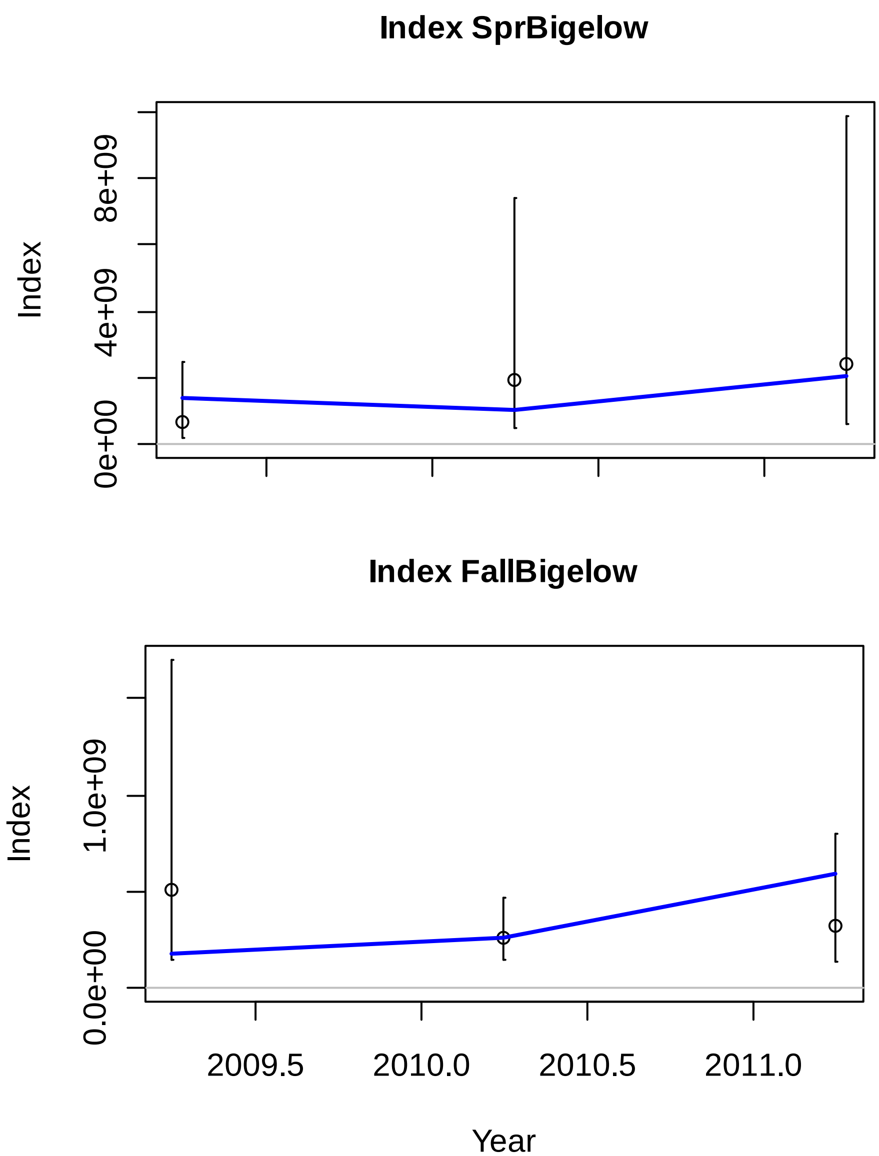


Figure A2-18. Goodness of fit plots for the SS3 model and herring in the NEFSC Bigelow spring and fall bottom trawl surveys.

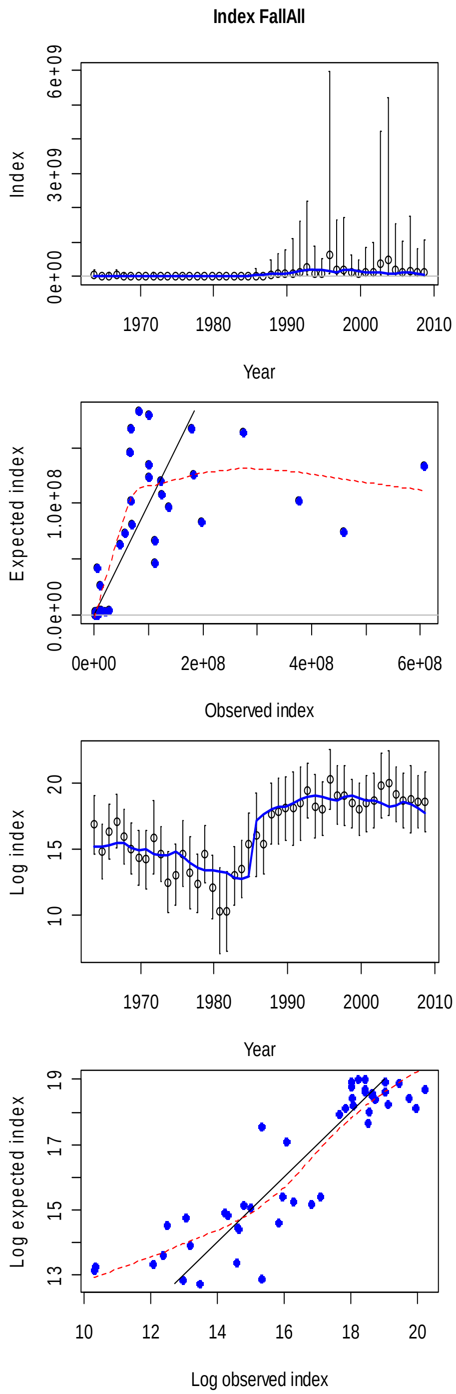


Figure A2-19. Goodness of fit plots for the SS3 model and herring in the NEFSC fall bottom trawl survey.

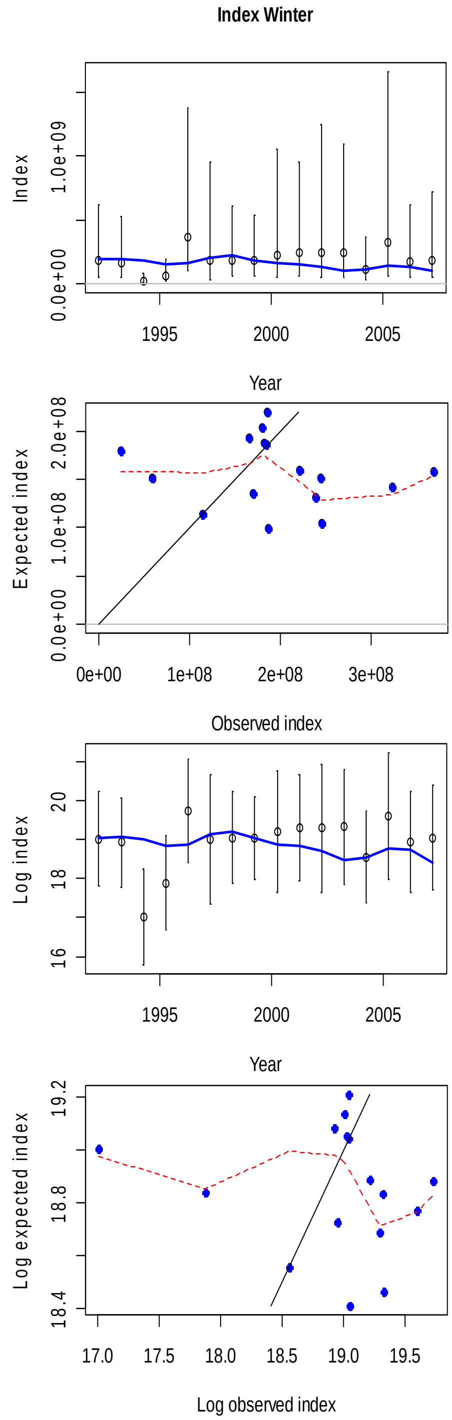


Figure A2-20 Goodness of fit plots for the SS3 model and herring in the NEFSC winter bottom trawl survey.

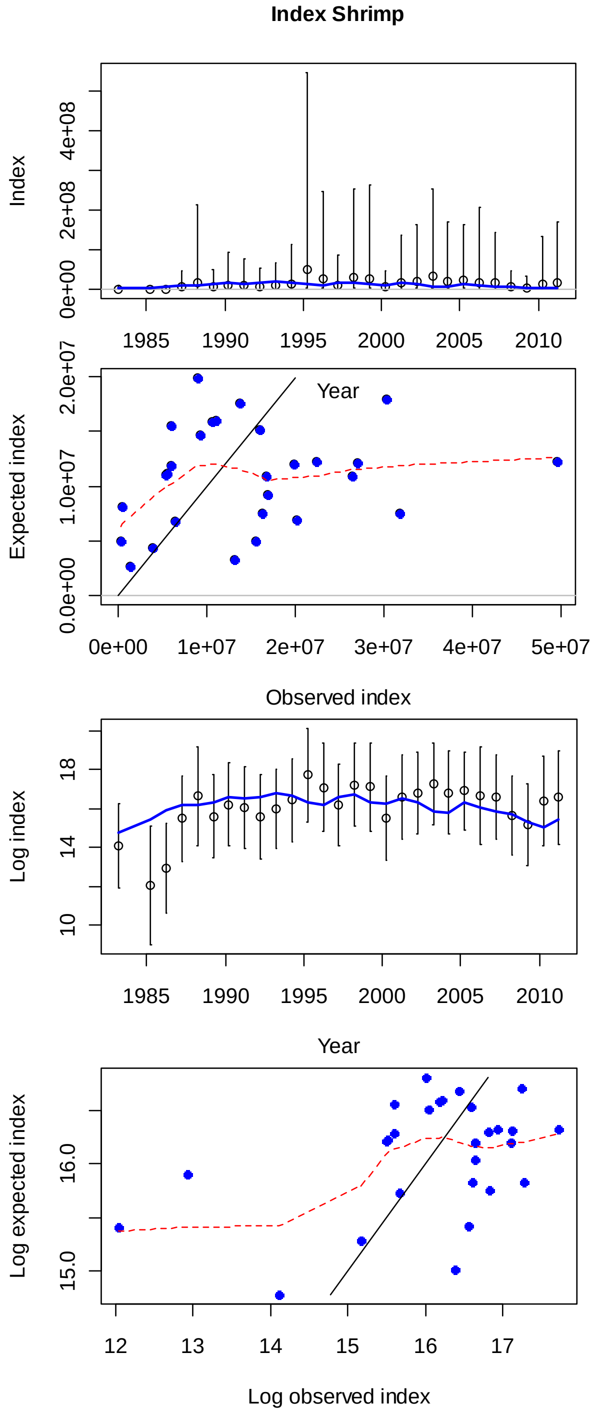


Figure A2-21. Goodness of fit plots for the SS3 model and herring in the NEFSC shrimp bottom trawl survey.

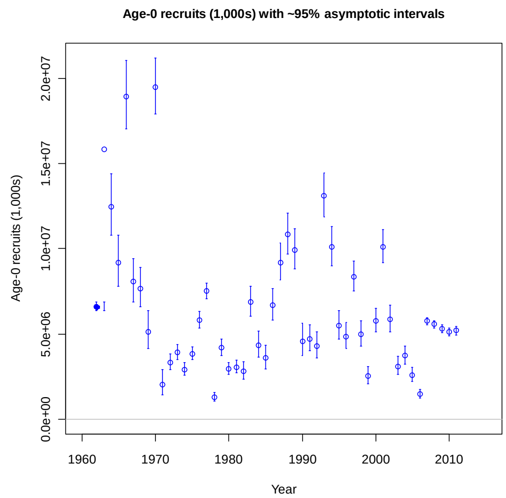


Figure A2-22. Recruitment estimates for herring from SS3. The first two estimates on the left are at the virgin and initial equilibrium recruitment levels. The third point from the left is the initial (1962) recruitment estimates. Other recruitments are estimates for 1963-2011. Recruitments were also estimated for 1959-1961 and used in initializing the population age and length composition. Recruitment estimates for 2006-2011 were from the model's estimated spawner-recruit curve.

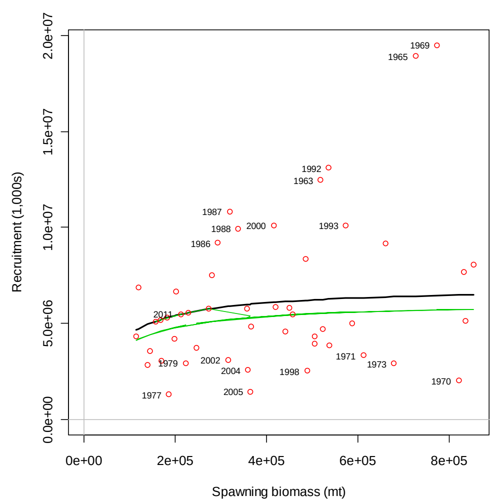


Figure A2-23. Spawner-recruit curve for herring estimated in SS3. The green line shows the geometric mean recruitment relationship and the black line shows the mean recruitment relationship. The 2006-2011 recruitments at spawning biomass levels of around2. 2.5 x 10^6 mt are expected values from the spawner-recruit curve.

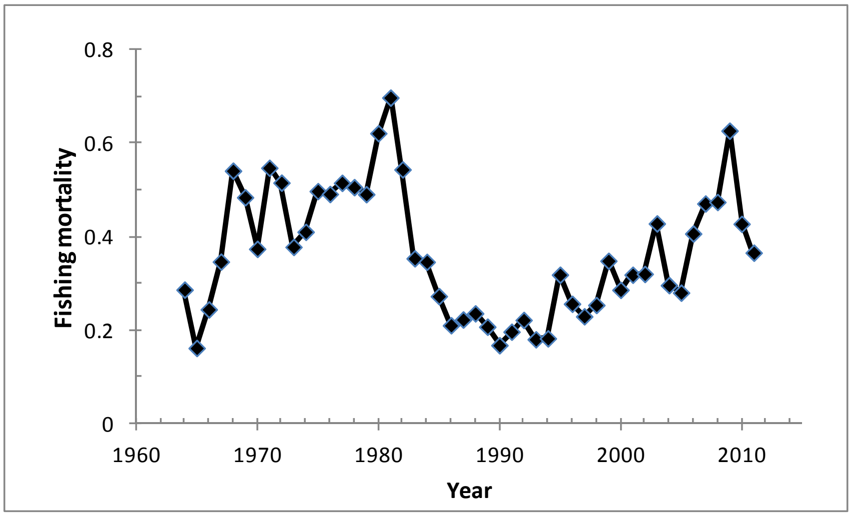


Figure A2-24. Approximate annual fishing mortality rate estimates for herring during 1964-2011 from SS3. The approximation for each year was computed as total annual landings divided by the biomass of herring age 1+ on July 1.

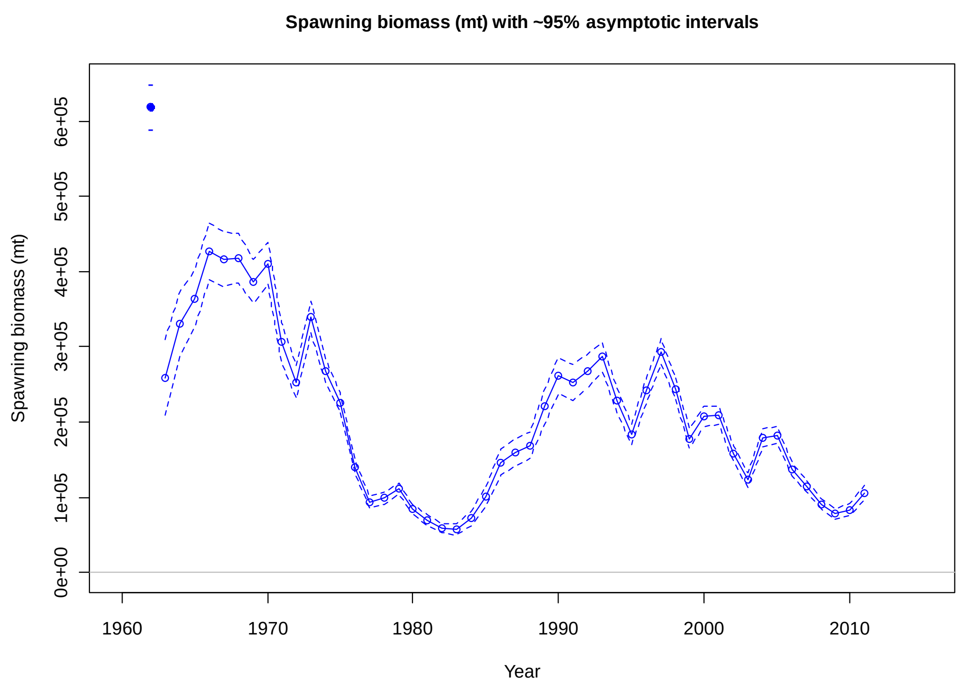


Figure A2-24. Approximate spawning stock biomass estimates (+ 95% CI) for herring during 1964-2011 from SS3.

SARC 54 Pelagics Working Group (SDWG)

"This information is distributed solely for the purpose of pre-dissemination peer review. It has not been formally disseminated by NOAA. It does not represent final agency determination or policy."

Atlantic Herring Length-based Bottom Trawl Survey Calibration
Tim Miller, NEFSC Population Dynamics Branch
May 15, 2012

Introduction

In 2009, the *NOAA SHIP Henry B. Bigelow* replaced the *R/V Albatross IV* as the primary vessel for conducting spring and fall annual bottom trawl surveys for the Northeast Fisheries Science Center (NEFSC). There are many differences in the vessel operation, gear, and towing procedures between the new and old research platforms (NEFSC Vessel Calibration Working Group 2007). To merge survey information collected in 2009 onward with that collected previously, we need to be able to transform indices (perhaps at size and age) of abundance from the *Henry B. Bigelow* into those that would have been observed had the *Albatross IV* still been in service. The general method for merging information from these two time series is to calibrate the new information to that of the old (e.g., Pelletier 1998, Lewy et al. 2004, Cadigan and Dowden 2010). Specifically we need to predict the relative abundance that would have been observed by the *Albatross IV* ($\hat{R}_A$) using the relative abundance from the *Henry B. Bigelow* (R_B) and a "calibration factor" (ρ),

$$\hat{R}_A = \rho R_B \,. \tag{1}$$

To provide information from which to estimate calibration factors for a broad range of species, 636 paired tows were conducted with the two vessels during 2008. Paired tows occurred at many stations in both the spring and fall surveys. Paired tows were also conducted during the summer and fall at non-random stations to augment the number of non-zero observations for some species. Protocols for the paired tows are described in NEFSC Vessel Calibration Working Group (2007).

The methodology for estimating the calibration factors was proposed by the NEFSC and reviewed by a panel of independent scientists in 2009. The reviewers considered calibration factors that could potentially be specific to either the spring or fall survey (Miller et al. 2010). They recommended using a calibration factor estimator based on a beta-binomial model for the data collected at each station for most species, but also recommended using a ratio-type estimator under certain circumstances and not attempting to estimate calibration factors for species that were not well sampled.

Since the review, it has become apparent that accounting for size of individuals can be necessary for many species. When there are different selectivity patterns for the two vessels, the ratio of the fractions of available fish taken by the two gears varies with size. Under these circumstances, the estimated calibration factor that ignores size reflects an average ratio weighted across sizes where the weights of each size class are at least in part related to the number of individuals at that size available to the two gears and the number of stations where individuals at that size were caught. Applying calibration factors that ignore real size effects to surveys conducted in subsequent years when the size composition of the available population is unchanged should not produce biased predictions (eq. 1). However, when the size composition changes, the frequency of individuals and number of stations where individuals are observed at each size changes and the implicit weighting across size classes used to obtain the estimated calibration factor will not be applicable to the new data. Consequently, the predictions from the constant calibration factor of the numbers per tow that would have been caught by the *Albatross IV* will be biased.

Length-based calibration has been performed for groundfish (cod, haddock, and yellowtail flounder through the Trans-boundary Resource Assessment Committee process and silver, offshore, and red hakes during SARC 51 and loligo squid during SARC 51 (Brooks et al. 2010, NEFSC 2011). For those length-based calibrations, the same basic beta-binomial model from Miller et al. (2010) was assumed, but various functional forms were assumed for the relationship of length to the calibration factor. Since then, Miller (submitted) has explored two types of smoothers for the relationship of relative catch efficiency to length and the beta-binomial dispersion parameter. The smoothers (orthogonal polynomials and thin-plate regression splines) allow much more flexibility than the functional forms previously considered for other species by Brooks et al. (2010) and NEFSC (2011). Catch efficiency at length, $q(L)$, as defined here relates the expected catch to the density of available individuals on a per unit swept area basis,

$$E\left(C_{ik}\left(L\right)\right)=q_k\left(L\right)f_{ik}A_{ik}D_i\left(L\right)$$

where $D_i\left(L\right)$ is the density of available fish at station i, and f_{ik} and A_{ik} are the fraction of the catch sampled for lengths and swept area for vessel/gear k. Relative catch efficiency is the ratio of the catch efficiencies for two vessels and is related to the calibration factor,

$$\rho\left(L\right)=\frac{E\left(C_{i1}\left(L\right)\right)}{E\left(C_{i2}\left(L\right)\right)}=\frac{q_1\left(L\right)}{q_2\left(L\right)}\frac{f_{i1}A_{i1}}{f_{i2}A_{i2}}.$$

Miller (submitted) analyzed data for six species and these methods were also used to estimate length-based calibration factors for each of the winter flounder stocks in the 2011 winter flounder assessment (Miller 2011). Here we use the same methods to estimate length-based calibration factors for Atlantic herring. We also explore differences in the effects of length on the models by season.

Methods

The data used in to fit the herring calibration models are numbers sampled by vessel, station, and 1 cm length class. Fish less than 12 cm in length were observed at a very small number of stations and some length classes are completely unobserved (Figure 1). However, substantial numbers of fish were caught at these few stations and most of them by the *Albatross IV* (Figure 2). Furthermore, when looking at spring and fall survey stations separately, it is apparent that most of the observations for these small fish and the largest numbers caught occurred in the spring (Figures 3 and 4). Because there was a large number of length classes without any observations between these small fish and larger sizes where most of the observations occurred, including these small fish caused difficulties in model fitting. Therefore, observations for fish less than 12 cm in length were excluded from further analysis.

I considered the orthogonal polynomial and thin-plate regression spline smoothers described by Miller (submitted). These models also allow for effects of swept area (SA) and sampling fraction (SF) on the beta-binomial dispersion parameter. I also considered models where effects on the relative catch efficiency and beta-binomial dispersion parameter differed for spring and fall seasons as well as the site-specific stations (outside the survey stations). I compared relative goodness-of-fit of the models using Akaike Information Criteria corrected for small sample size bias (AIC_c; Hurvich and Tsai 1989). I fit models in the R statistical programming environment (R Development Core Team 2010) and used the GAMLSSS package (Rigby and Stasinopoulos 2005, Stasinopoulos and Rigby 2007).

Results and Discussion

The best model without seasonal effects had a fifth order orthogonal polynomial smoother of the effects of length on the relative catch efficiency (Table 1). The best model also had a third order orthogonal polynomial smoother of the effects of length and effects of swept area and sampling fraction of each vessel on the beta-binomial dispersion parameter. All of the top 10 ranking models included the effects of swept area and sampling fraction on the dispersion parameter and the top four models all performed similarly with respected to AIC_c. The predicted relative catch efficiency from the best model is largest for the smallest and largest fish, but the uncertainty is also greatest for these sizes. The Henry B. Bigelow is estimated to be at least 2.5 times as efficient as the Albatross IV across all sizes between 12 and 31 cm (Figure 5 and Table 2). The dispersion parameter estimates are generally lower for all but the smallest size classes implying that there is less variability in the relative catch efficiency for smaller sizes from station to station (Figure 6). The residuals for this model show no concerning patterns (Figure 7) and there are substantial differences in the predicted relative catch efficiency between the best model with the orthogonal polynomial smoother and the best model with the thin-plate spline smoother (Rank 50) (Figure 8).

For data collected during the spring survey, the best model had no length effect on relative catch efficiency and a third order polynomial smoother for the effect of length on the dispersion parameter (Table 3). Effects of either swept area or sampling fraction or both were important in all of the top 10 ranking models and the fifth ranking model had a thin-plate spline smoother of the effects of length on relative catch efficiency and the dispersion parameter.

For fall data, the best model had a seventh order polynomial smoother for the effect of length on relative catch efficiency and a second order polynomial smoother for the effect of length on the dispersion parameter (Table 4). None of the top 10 ranking models had effects of sampling fraction on the dispersion parameter and four had an effect of swept area. Three of the top ten models had thin-plate spline smoothers for the effects of length on relative catch efficiency and the dispersion parameter. All of the top ten models performed similarly with respect to AIC_c.

Among site-specific stations, the one model with thin-plate spline smoothers and one with orthogonal polynomials performed identically as the best model (Table 5) The model with orthogonal polynomials had a first order smother (linear on the log scale) of length on the relative catch efficiency and a second order smoother for the effect on the dispersion parameter and the total number of estimated parameters was fewer. All of the top ten ranking models had effects of sampling fraction and swept area on the dispersion parameter.

The AIC_c (4111.32) obtained from the best fitted models for each of the subsets of data (spring, fall, site-specific) that was more than 100 units less than the best model (AIC_c = 4216.36) when the same model was fit to data from each subset. This substantial reduction in the performance measure would suggest using seasonal results for calibration. The dramatic difference in the length effects on relative catch efficiency for the spring (no length effect) and fall (high order polynomial) are reflected in the predicted values (Figure 9 and Tables 6 and 7). There is less difference in the length effects on the dispersion parameter (Figure 10). There are no concerning patterns in the residuals for the best spring and fall models (Figure 11) and the small differences between the best fitting orthogonal polynomial and thin-plate spline smoothers for the respective seasons reflects the small difference in their overall rank with respect to AIC_c (Figure 12).

When applying the relative catch efficiencies to surveys conducted in 2009 and beyond with the *Henry B. Bigelow*, there is an important caution to note. Lengths may be observed in these surveys that are outside of the range of lengths observed during the calibration study. Caution must be taken in predicting catches in *Albatross IV* units at these sizes. This problem can be exacerbated when the data are broken down into seasonal subsets for estimation of relative catch efficiency because the limits of the range of sizes available in the subsets can be narrower than the range of the entire data set, but this turned out to not be a concern for herring.

Lastly, the swept areas for tows during the 2009 and 2010 surveys would ideally be used to predict Albatross catches at each station, but if there is little variability in the swept areas a mean can be used and the mean number per tow at length in *Henry B. Bigelow* “units” can be converted to *Albatross IV* units (Table 8).

References

Brooks, E. N., Miller, T. J., Legault, C. M., O'Brien, L., Clark, K. J., Gavaris, S., and Van Eekhaute, L. 2010. Determining length-based calibration factors for cod, haddock, and yellowtail flounder. TRAC Ref. Doc. 2010/08.

Cadigan, N. G. and Dowden, J. J. 2010. Statistical inference about relative efficiency from paired-tow survey calibration data. Fish. Bull. 108: 15-29.

Hurvich, C. M. and Tsai, C.-L. 1989. Regression and time series model selection in small samples. Biometrika 76: 297-307.

Lewy, P., Nielsen, J. R., and Hovgård, H. 2004. Survey gear calibration independent of spatial fish distribution. Can. J. Fish. Aquat. Sci. 61: 636-647.

Miller, T. J. 2011. Winter flounder length-based calibration. SARC 52 Working paper 8.

Miller, T. J. Submitted. A hierarchical model for relative catch efficiency from gear selectivity and calibration studies.

Miller, T. J., Das, C., Miller, A. S., Lucey, S. M., Legault, C. M., Brown, R. W., and Rago, P. J. 2010. Estimation of *Albatross IV* to *Henry B. Bigelow* calibration factors. NEFSC Ref. Doc. 10-05.

NEFSC. 2011. 51st Northeast regional stock assessment workshop (51st SAW) assessment report. NEFSC Ref. Doc. 11-02.

NEFSC Vessel Calibration Working Group. 2007. Proposed vessel calibration for NOAA Ship *Henry B. Bigelow*. NEFSC Ref. Doc. 07-12. 26p.

Pelletier, D. 1998. Intercalibration of research survey vessels in fisheries: a review and an application. Can. J. Fish. Aquat. Sci. 55: 2672-2690.

R Development Core Team. 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Rigby, R. A. and Stasinopoulos, D. M. 2005. Generalized additive models for location, scale, and shape (with discussion). Applied Statistics 54: 507-554.

Stasinopoulos, D. M. and Rigby, R. A. 2007. Generalized additive models for location, scale and shape (GAMLSS) in R. Journal of Statistical Software 23: 1-46.

Table 1. Model type (thin-plate regression spline, SP, orthogonal polynomial, OP), relative catch efficiency, dispersion, and total degrees of freedom, dispersion covariates, and log-likelihood for best performing models based on AIC_c. Results are based on data for fish at least 12cm in length collected at all stations.

Rank	Model Type	# Total df	ρ df	ϕ df	ϕ Covariates	LL	AIC_c	Δ (AIC_c)
1	OP	12	6	6	SA, SF	-2096.07	4216.36	0.00
2	OP	13	7	6	SA, SF	-2095.06	4216.39	0.03
3	OP	14	7	7	SA, SF	-2094.05	4216.40	0.04
4	OP	13	6	7	SA, SF	-2095.13	4216.52	0.16
5	OP	9	3	6	SA, SF	-2099.78	4217.69	1.32
6	OP	15	8	7	SA, SF	-2093.90	4218.15	1.79
7	OP	14	8	6	SA, SF	-2094.96	4218.23	1.87
8	OP	10	3	7	SA, SF	-2099.17	4218.49	2.13
9	OP	15	9	6	SA, SF	-2094.50	4219.34	2.98
10	OP	16	9	7	SA, SF	-2093.48	4219.35	2.99

Table 2. Predicted relative catch efficiencies and coefficient of variation from the best fitted beta-binomial model with respect to AIC_c (see Table 1) based on data collected at all stations in 2008 for fish at least 12cm in length.

Length (cm)	$\hat{\rho}$	$CV(\hat{\rho})$
12	4.405	1.022
13	16.762	0.552
14	27.213	0.419
15	26.219	0.376
16	19.209	0.313
17	12.757	0.233
18	8.610	0.162
19	6.289	0.115
20	5.083	0.092
21	4.507	0.078
22	4.262	0.067
23	4.135	0.064
24	3.965	0.066
25	3.657	0.068
26	3.228	0.070
27	2.798	0.080
28	2.551	0.099
29	2.759	0.131
30	4.253	0.249
31	12.078	0.565

Table 3. For data collected during the spring survey, model type (orthogonal polynomial, OP or thin-plate spline, SP), relative catch efficiency, dispersion, and total degrees of freedom, dispersion covariates, and log-likelihood for best performing models based on AIC_c. Results are based on data for fish at least 12cm in length.

Rank	Model Type	# Total df	ρ df	ϕ df	ϕ Covariates	LL	AIC_c	Δ (AIC_c)
1	OP	7.00	1.00	6.00	SA,SF	-761.70	1537.58	0.00
2	OP	6.00	1.00	5.00	SA,SF	-763.12	1538.38	0.80
3	OP	11.00	5.00	6.00	SA,SF	-758.19	1538.80	1.22
4	OP	8.00	1.00	7.00	SA,SF	-761.37	1538.96	1.39
5	SP	7.94	2.00	5.94	SA,SF	-761.43	1539.05	1.48
6	OP	8.00	2.00	6.00	SA,SF	-761.42	1539.06	1.48
7	OP	7.00	2.00	5.00	SA,SF	-762.70	1539.57	1.99
8	OP	6.00	1.00	5.00	SA	-763.85	1539.83	2.26
9	OP	6.00	1.00	5.00	SF	-763.89	1539.90	2.33
10	OP	10.00	5.00	5.00	SA,SF	-759.86	1540.06	2.49

Table 4. For data collected during the fall survey, model type (orthogonal polynomial, OP or thin-plate spline, SP), relative catch efficiency, dispersion, and total degrees of freedom, dispersion covariates, and log-likelihood for best performing models based on AIC_c. Results are based on data for fish at least 12cm in length.

Rank	Model Type	# Total df	ρ df	ϕ df	ϕ Covariates	LL	AIC_c	Δ (AIC_c)
1	OP	11.00	8.00	3.00		-405.68	833.99	0.00
2	OP	10.00	8.00	2.00		-406.76	834.06	0.07
3	SP	7.96	6.96	1.00		-408.80	834.16	0.17
4	OP	12.00	8.00	4.00	SA	-404.71	834.17	0.18
5	OP	10.00	8.00	2.00	SA	-406.83	834.19	0.20
6	OP	9.00	8.00	1.00		-407.90	834.23	0.24
7	OP	11.00	8.00	3.00	SA	-405.83	834.30	0.32
8	SP	9.00	7.00	2.00	SA	-407.77	834.32	0.34
9	OP	10.00	7.00	3.00		-407.05	834.63	0.65
10	SP	9.16	7.16	2.00		-407.77	834.67	0.68

Table 5. For data collected from site-specific stations (outside of the fall and spring surveys), model type (orthogonal polynomial, OP or thin-plate spline, SP), relative catch efficiency, dispersion, and total degrees of freedom, dispersion covariates, and log-likelihood for best performing models based on AIC_c. Results are based on data for fish at least 12cm in length.

Rank	Model Type	# Total df	ρ df	ϕ df	ϕ Covariates	LL	AIC_c	$\Delta(AIC_c)$
1	OP	7.00	2.00	5.00	SA,SF	-862.73	1739.63	0.00
2	SP	10.45	2.00	8.45	SA,SF	-859.22	1739.80	0.00
3	OP	8.00	2.00	6.00	SA,SF	-862.10	1740.41	0.78
4	OP	9.00	2.00	7.00	SA,SF	-861.12	1740.50	0.88
5	OP	8.00	3.00	5.00	SA,SF	-862.25	1740.70	1.07
6	OP	9.00	3.00	6.00	SA,SF	-861.48	1741.21	1.59
7	OP	10.00	3.00	7.00	SA,SF	-860.50	1741.32	1.70
8	OP	12.00	3.00	9.00	SA,SF	-858.53	1741.52	1.89
9	OP	9.00	4.00	5.00	SA,SF	-862.04	1742.34	2.71
10	OP	11.00	4.00	7.00	SA,SF	-860.04	1742.46	2.84

Table 6. Predicted relative catch efficiencies and coefficient of variation from a fitted beta-binomial model with fourth degree orthogonal polynomials in length for the mean parameter and first degree (linear) polynomial in length for the dispersion parameter (best performing orthogonal polynomial model without gamma assumption) based on data collected during the spring survey for fish at least 12cm in length.

Length (cm)	$\hat{\rho}$	$CV(\hat{\rho})$
14	6.070	0.074
15	6.070	0.074
16	6.070	0.074
17	6.070	0.074
18	6.070	0.074
19	6.070	0.074
20	6.070	0.074
21	6.070	0.074
22	6.070	0.074
23	6.070	0.074
24	6.070	0.074
25	6.070	0.074
26	6.070	0.074
27	6.070	0.074
28	6.070	0.074
29	6.070	0.074
30	6.070	0.074
31	6.070	0.074

Table 7. Predicted relative catch efficiencies and coefficient of variation from a fitted beta-binomial model with fourth degree orthogonal polynomials in length for the mean parameter and first degree (linear) polynomial in length for the dispersion parameter (best performing orthogonal polynomial model without gamma assumption) based on data collected during the fall survey for fish at least 12cm in length.

Length (cm)	$\hat{\rho}$	$CV(\hat{\rho})$
12	2.430	1.323
13	14.515	0.699
14	35.491	0.595
15	33.642	0.578
16	16.701	0.630
17	6.513	0.592
18	2.835	0.473
19	1.705	0.347
20	1.496	0.258
21	1.760	0.195
22	2.351	0.149
23	2.973	0.137
24	3.125	0.140
25	2.663	0.138
26	2.035	0.148
27	1.708	0.166
28	1.957	0.183
29	3.277	0.280
30	5.745	0.433
31	3.511	1.063

Table 8. Mean swept area (sq. nm) per tow for each vessel at all offshore stations where herring at least 12 cm in length were observed, across all seasons or during spring and fall surveys. Note that swept area is not known for every tow.

	Albatross IV	*Henry B. Bigelow*
All stations	0.011668	0.007188
Spring	0.011644	0.006835
Fall	0.010966	0.007321

Figure 1. Number of stations where fish were observed by length class (top) and the proportions of stations where fish were observed aboard the *Henry B. Bigelow* only (black), *Albatross IV* only (white) or both vessels (gray).

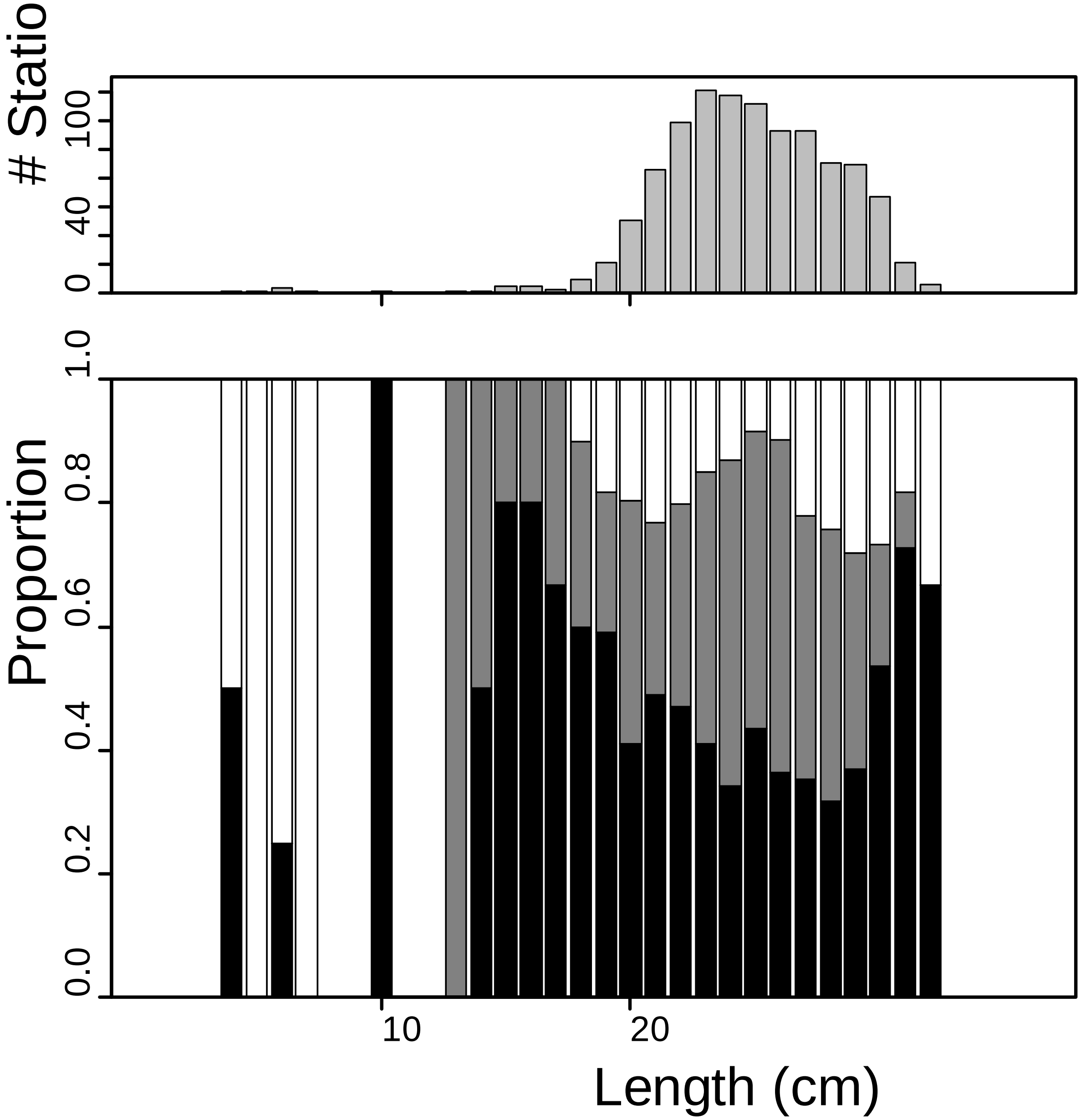


Figure 2. Total number of fish captured at each station in offshore strata (both vessels combined) at length (top) and proportions captured by the *Albatross IV* (white) and *Henry B. Bigelow* (gray) (bottom) from data collected at all stations in 2008 for fish at least 12cm in length.

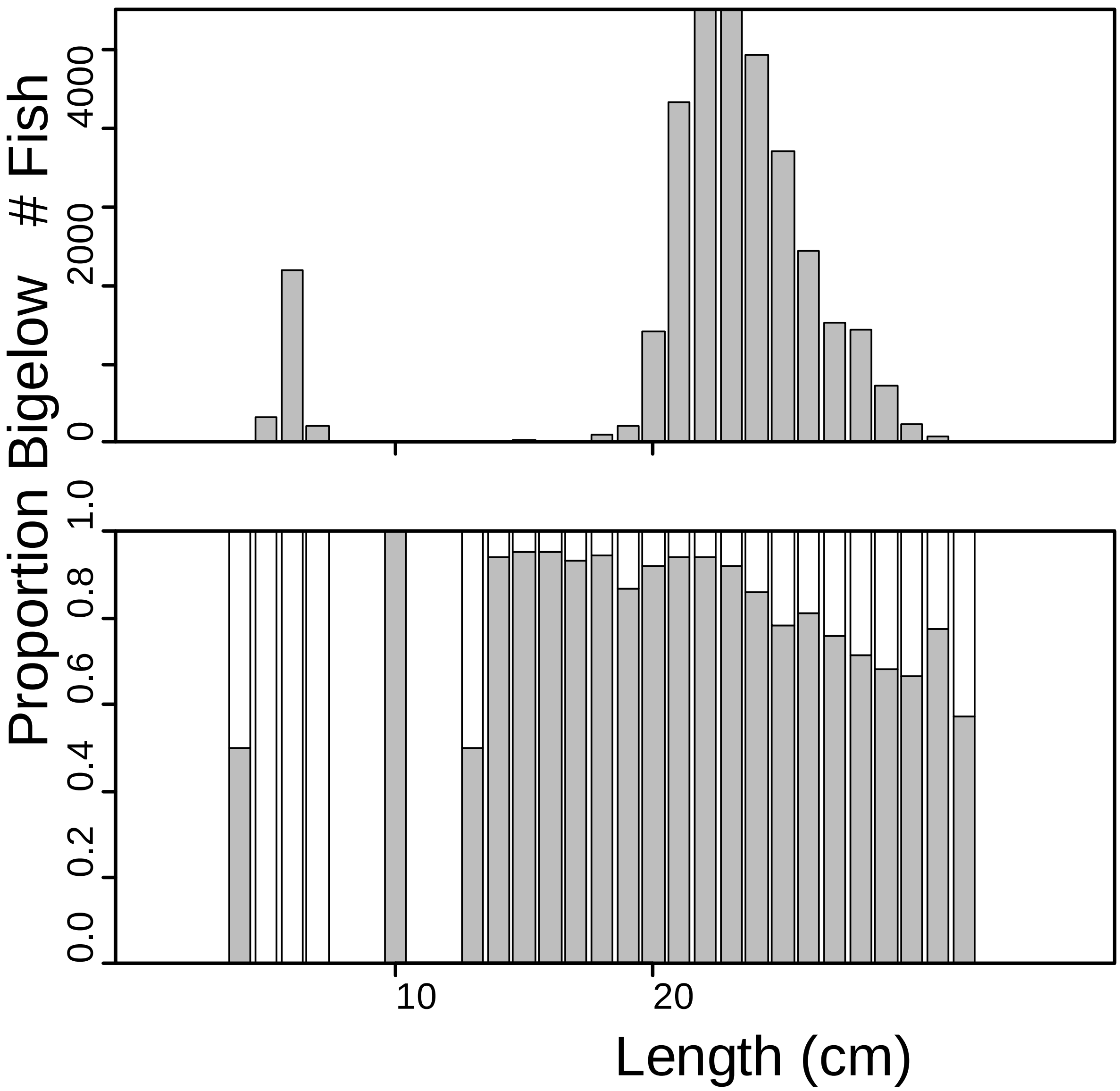


Figure 3. Number of stations where fish were observed by length class (top) and the proportions of stations where fish were observed aboard the *Henry B. Bigelow* only (black), *Albatross IV* only (white) or both vessels (gray) for data collected from stations during the spring (left) and fall (right) surveys in 2008.

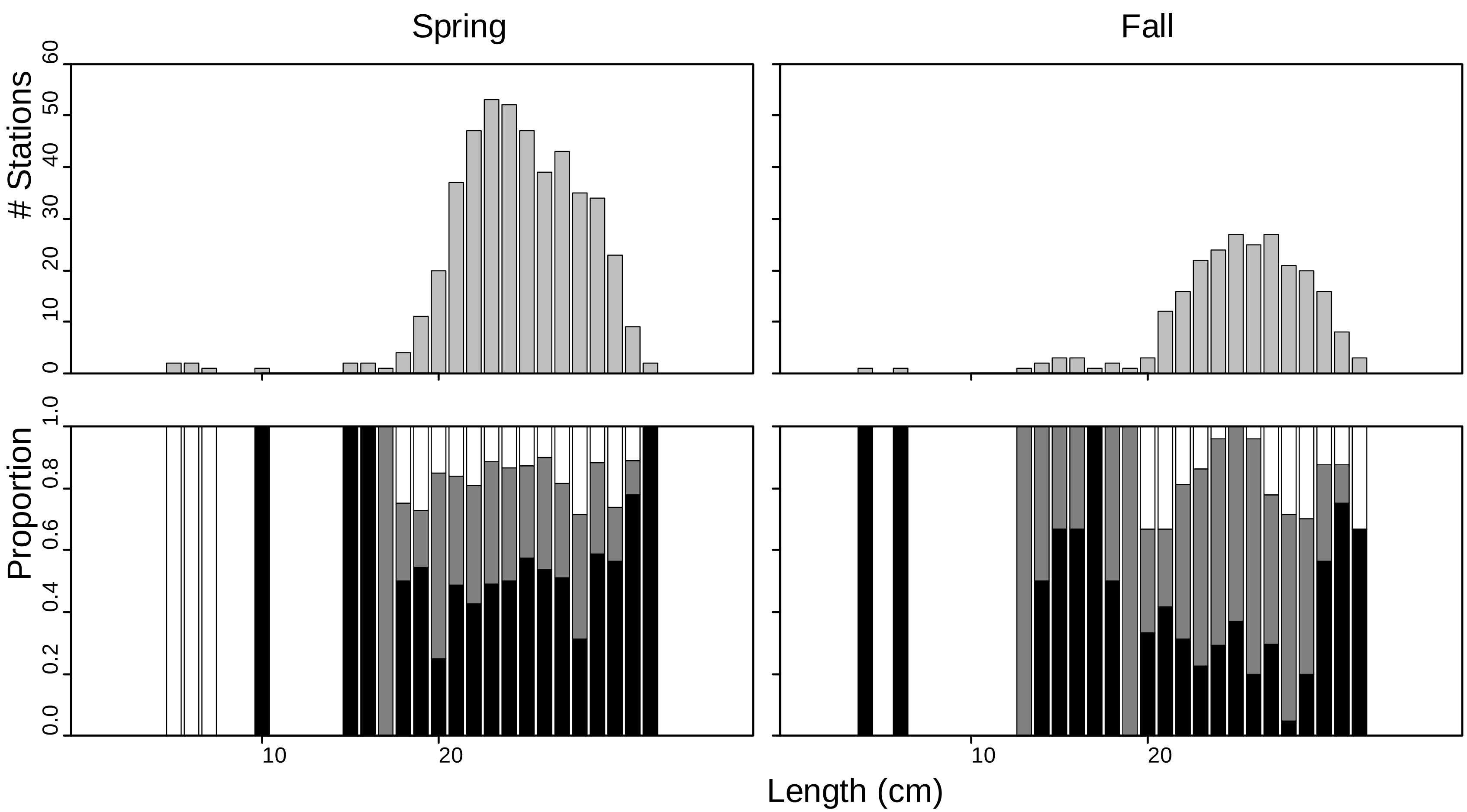


Figure 4. Total number of fish captured at each station (both vessels combined) at length (top) and proportions captured by the *Albatross IV* (white) and *Henry B. Bigelow* (gray) (bottom) for data collected from stations during the spring (left) and fall (right) surveys in 2008.

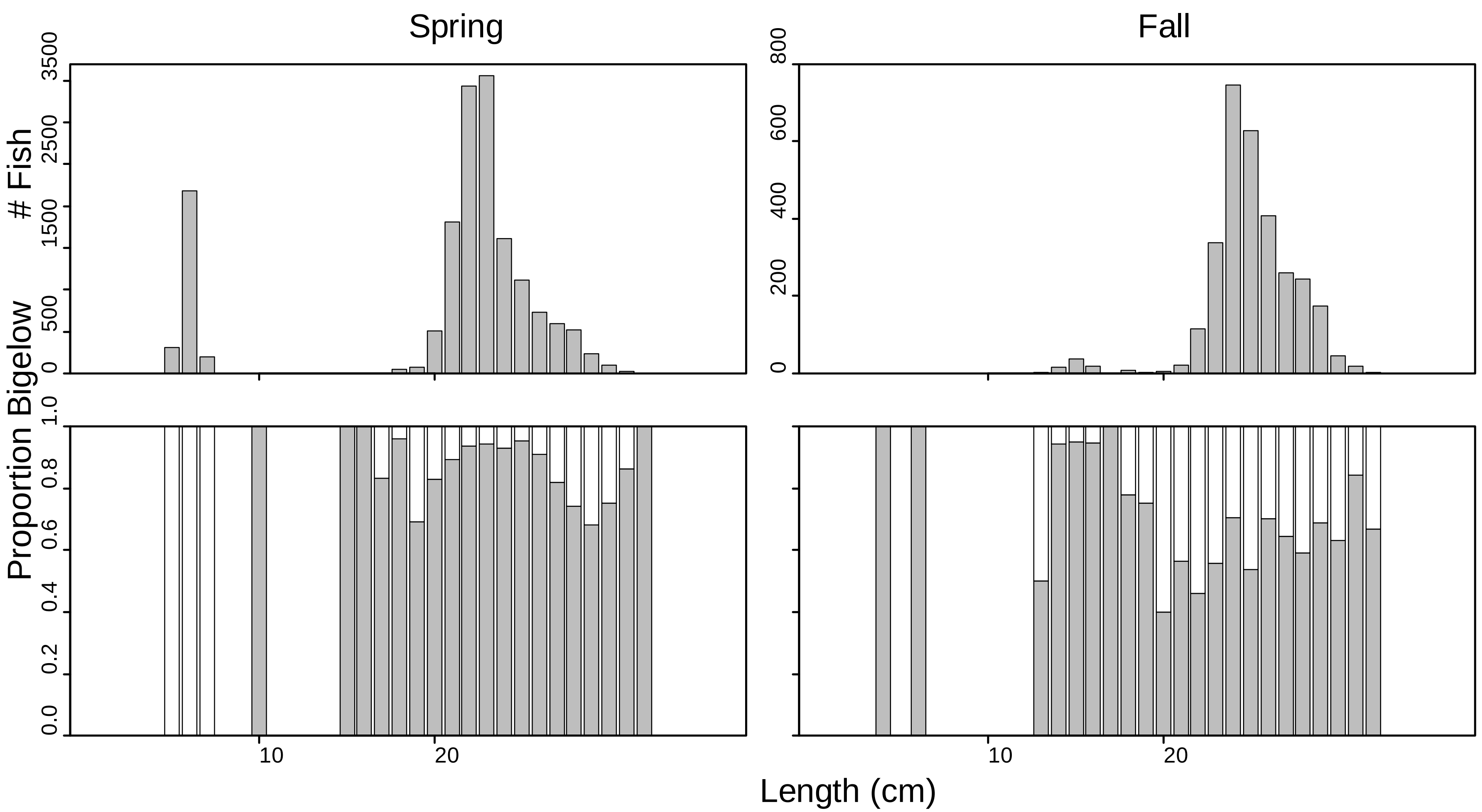


Figure 5. Predicted relative catch efficiency from the best performing model (red) and 95% confidence intervals (dashed lines) and predicted relative catch efficiency by length class (gray) with 95% confidence intervals (vertical lines). Results are based on data collected at all stations in 2008 for fish at least 12cm in length.

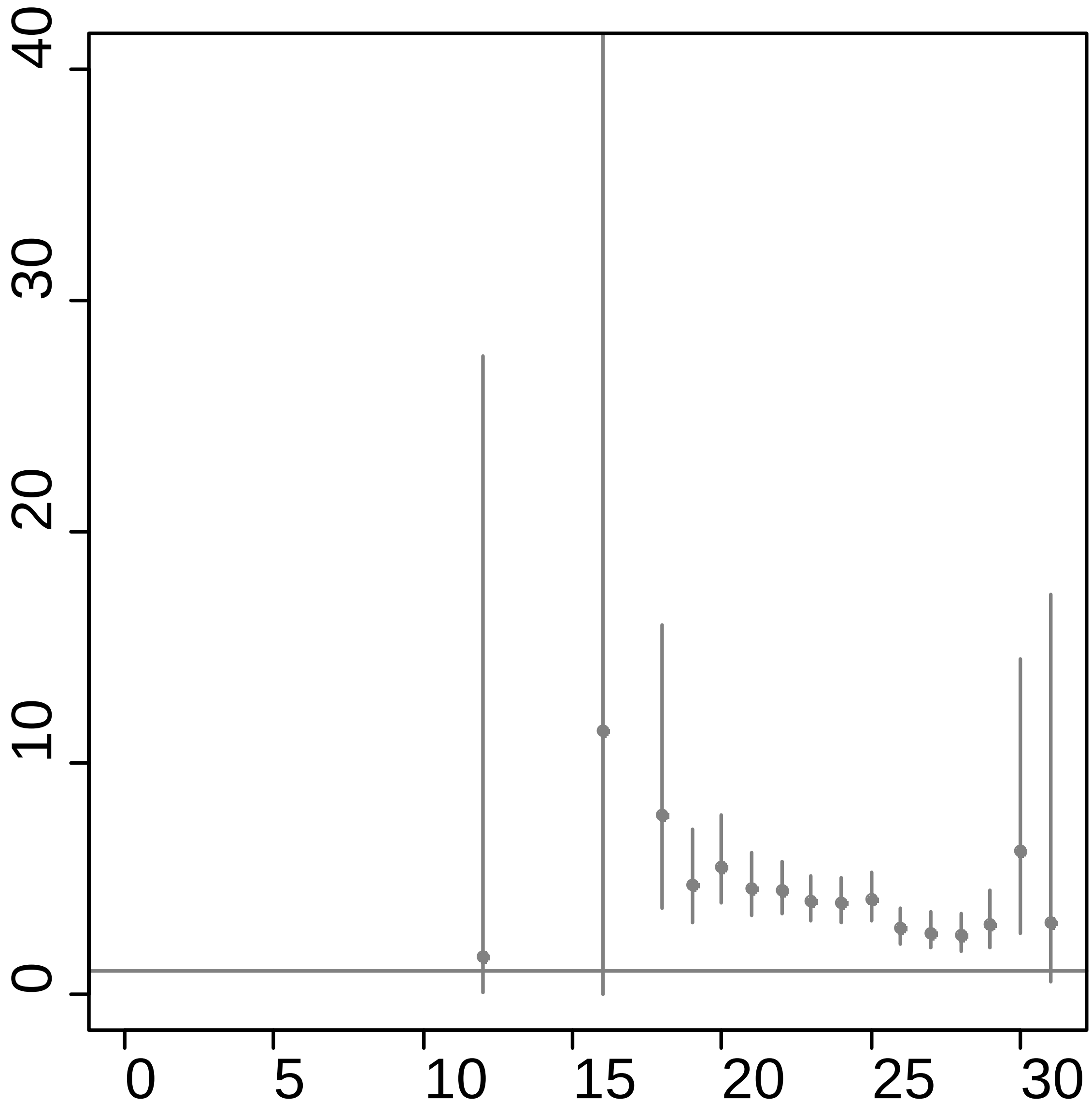


Figure 6. Predicted beta-binomial dispersion parameter from the best performing model (red) and 95% confidence intervals (dashed lines) and predicted dispersion parameter by length class (gray) with 95% confidence intervals (vertical lines). Results are based on data collected at all stations in 2008 for fish at least 12cm in length.

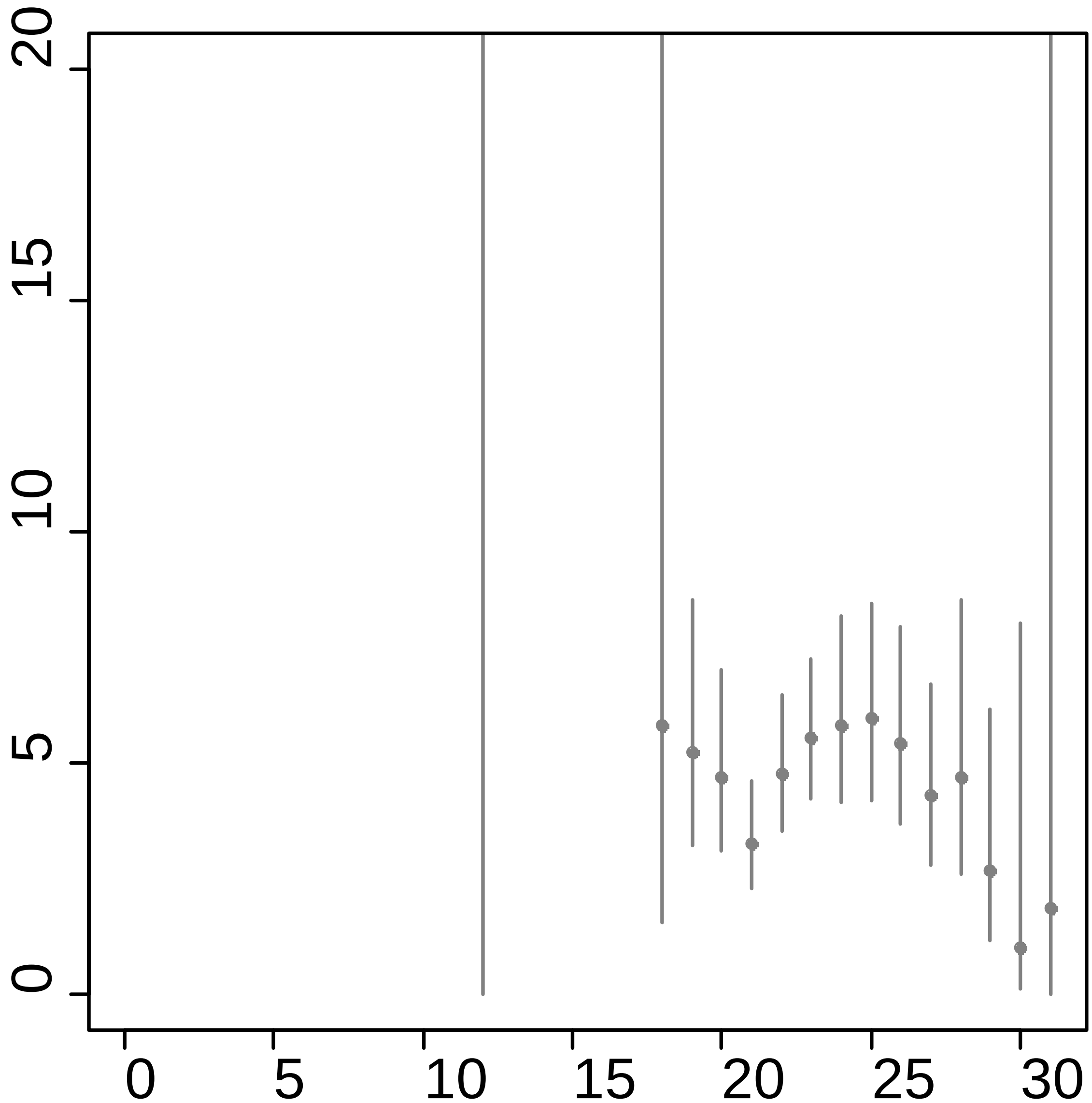


Figure 7. Randomized quantile residuals of the best performing model (as measured by AICc, see Table 1) in relation to the predicted number captured by the *Henry B. Bigelow* (left), the total number of fish captured at a station (middle), and their normal quantiles (right). Results are based on data collected at all stations in 2008 for fish at least 12cm in length.

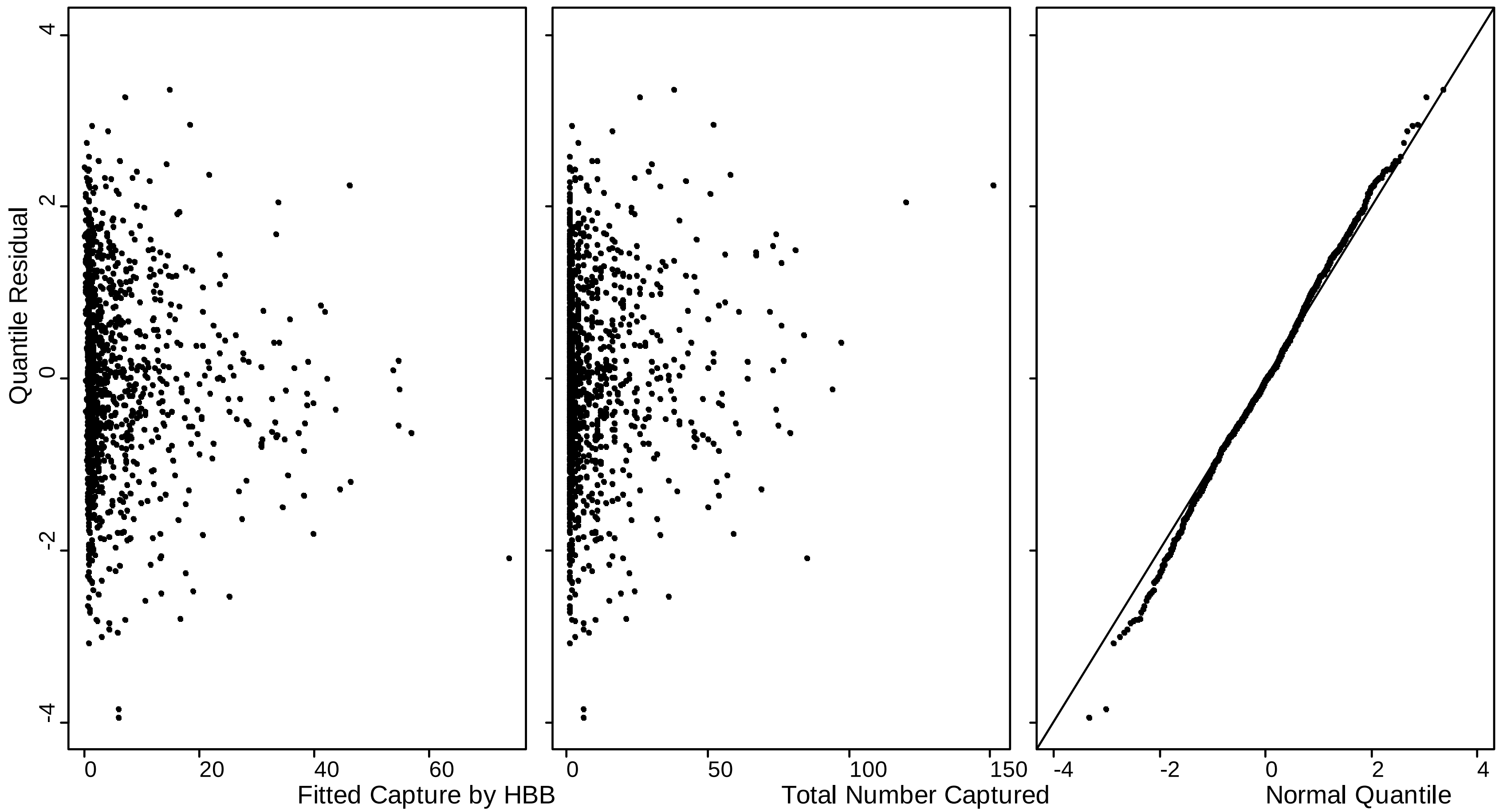


Figure 8. Predicted relative catch efficiency (left) and proportion captured by *Henry B. Bigelow* (right) from the best performing model and the best thin-plate regression spline smoother (Rank 50 with respect to AIC_c). Results are based on data collected across all stations in 2008 for fish at least 12cm in length.

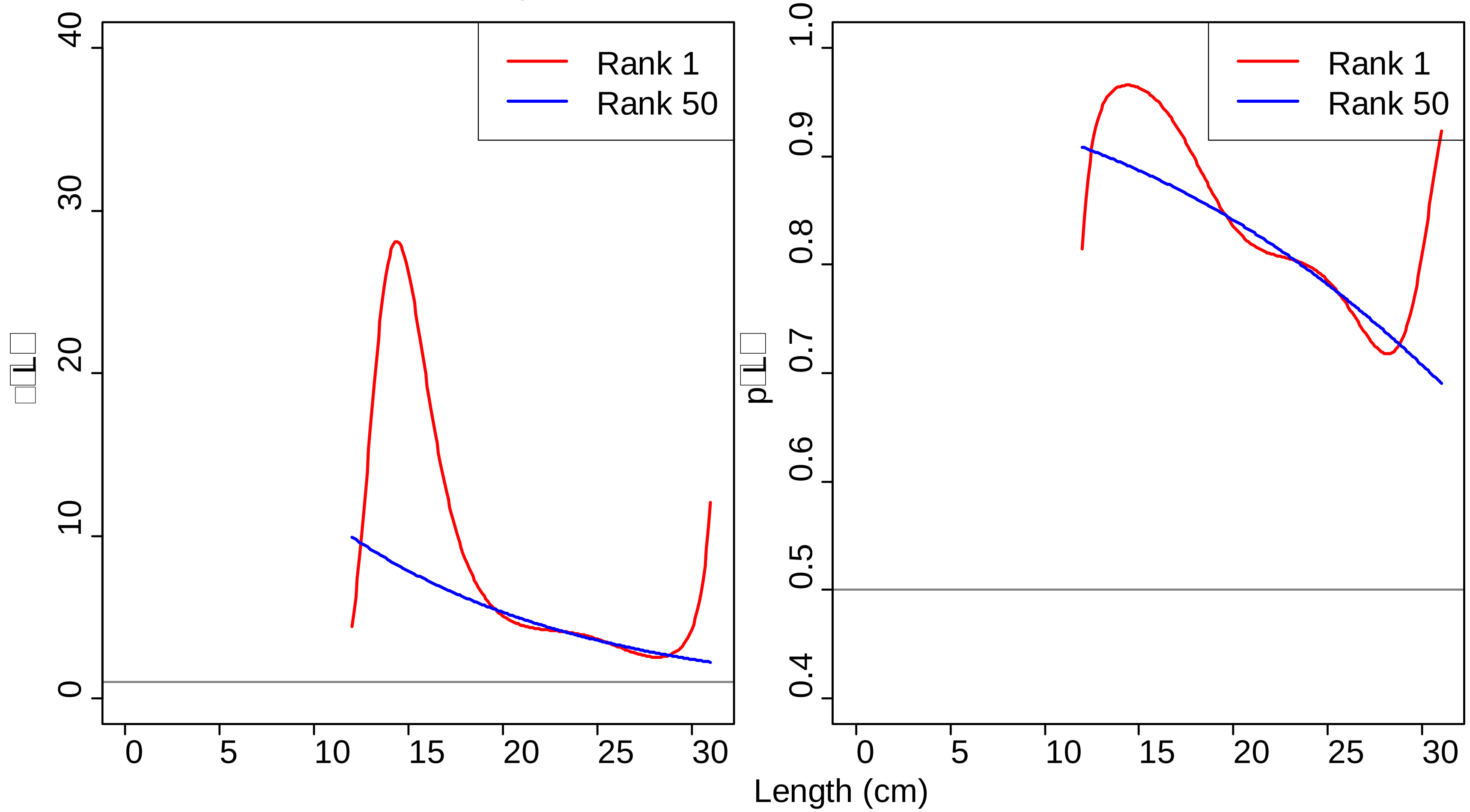


Figure 9. Predicted relative catch efficiency from the best performing orthogonal polynomial (without gamma assumption) model (red) and 95% confidence intervals (dashed lines) and predicted relative catch efficiency by length class (gray) with 95% confidence intervals (vertical lines). Results are based on data collected from stations during the spring (left) and fall (right) surveys in 2008 for fish at least 12cm in length.

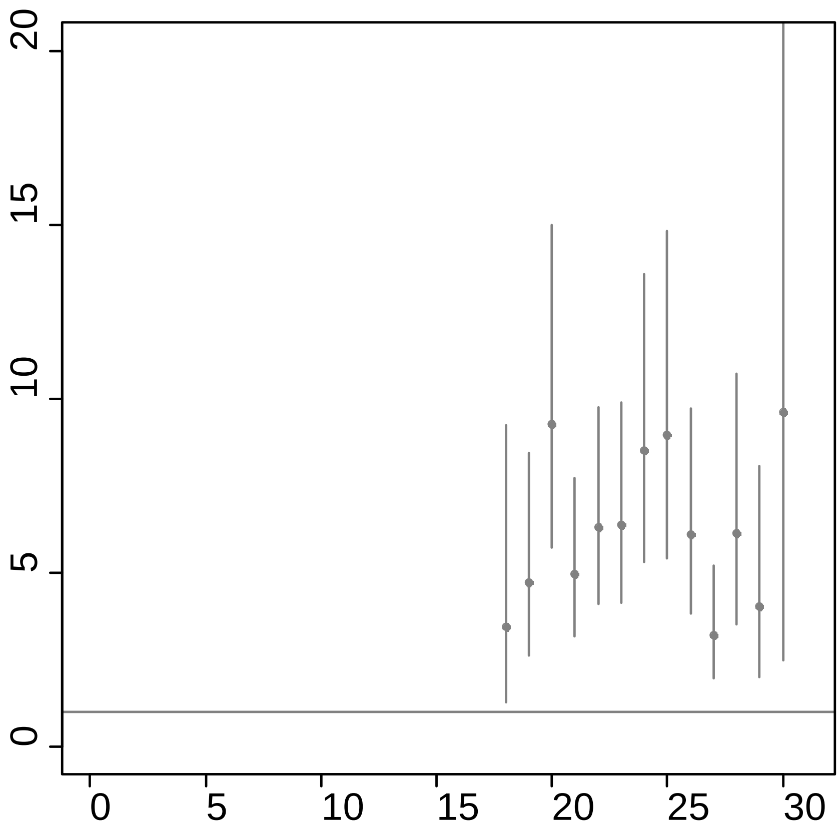


Figure 10. Predicted dispersion parameter from the best performing orthogonal polynomial model (red) and 95% confidence intervals (dashed lines) and predicted relative catch efficiency by length class (gray) with 95% confidence intervals (vertical lines). Results are based on data collected from stations during the spring (left) and fall (right) surveys in 2008 for fish at least 12cm in length.

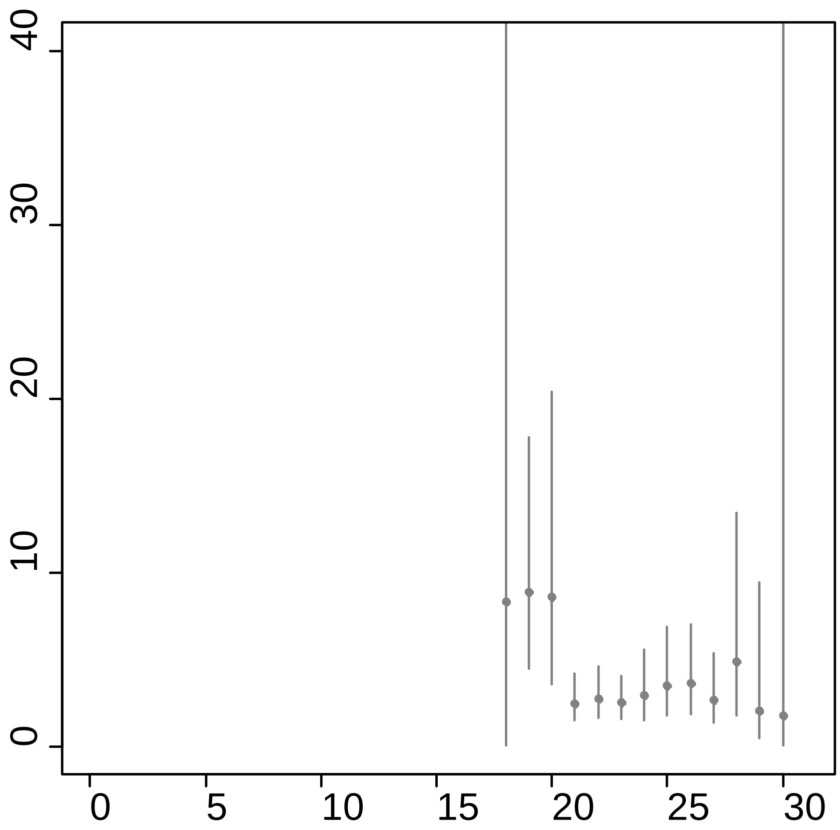


Figure 11. Randomized quantile residuals of the best performing (as measured by AICc) in relation to the predicted number captured by the *Henry B. Bigelow* (left), the total number of fish captured at a station (middle), and their normal quantiles (right). Results are based on data collected from stations during the spring (top) and fall (bottom) surveys in 2008 for fish at least 12cm in length.

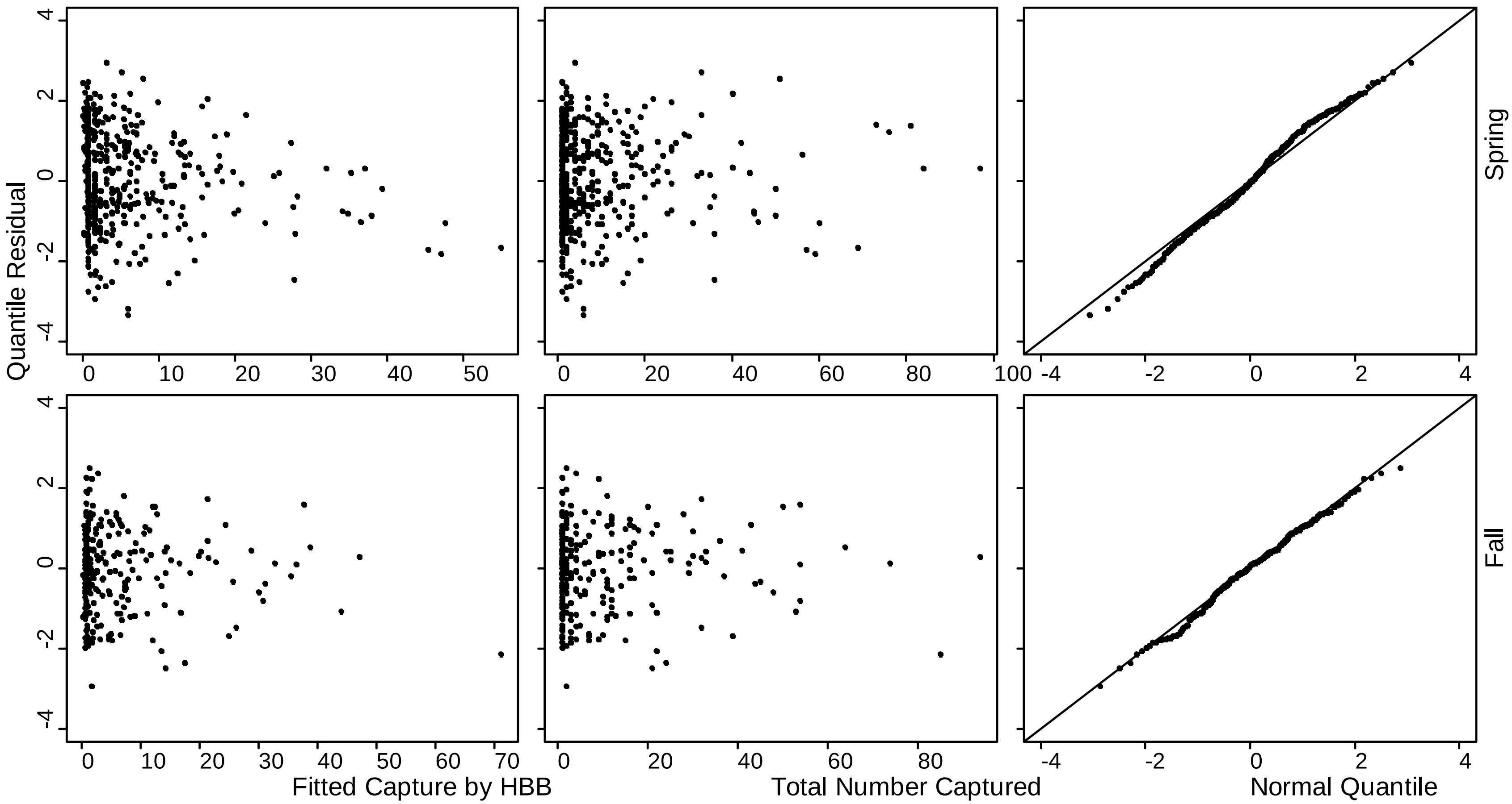
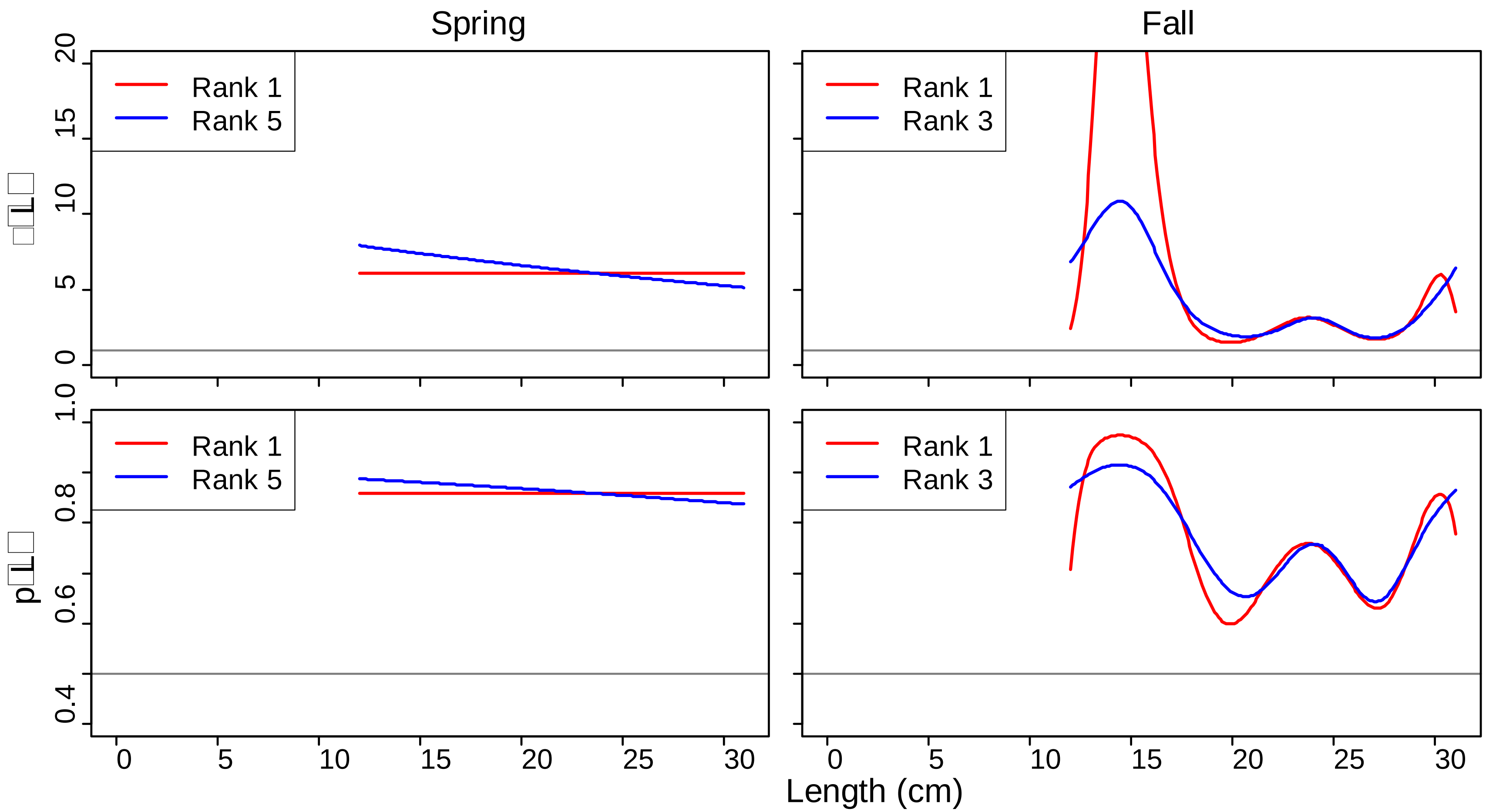


Figure 12. Predicted relative catch efficiency (top) and proportion captured by *Henry B. Bigelow* (bottom) from the best performing model (orthogonal polynomials, rank 1) and the best thin-plate spline smoother (Rank 12 for spring data, 11 for fall data) for data collected from stations during the spring (left) and fall (right) surveys in 2008 for fish at least 12cm in length.



An evaluation of whether changes in the timing and distribution of Atlantic herring spawning on Georges Bank may have biased the NEFSC acoustic survey

Preliminary results from a NOAA FATE funded project to:

Jonathan Hare[1], James Churchill[2], David Richardson[1], Michael Jech[1], Jonathan Deroba[1], and Harvey Walsh[1]

[1] - Northeast Fisheries Science Center

[2] - Woods Hole Oceanographic Institution

SUMMARY

At the 2009 TRAC assessment it was proposed that the NEFSC acoustic survey may not be sampling a fixed proportion of the Atlantic herring population year-to-year, resulting in a biased index. We used larval herring data collected by the NEFSC to evaluate changes in the timing and distribution of Atlantic herring egg hatching, which we use as a measure of spawning distributions. We did not find any evidence that herring spawning shifted from 2000 to 2003, the time period when the herring acoustic index declined substantially.

BACKGROUND

Acoustic surveys are used throughout the world to measure the size of stocks of pelagic species (Webb et al. 2008) and are generally the preferred method for surveying pelagic stocks (Simmonds & MacLennan 2005, McQuinn 2009). The NEFSC acoustic survey targets pre-spawning Atlantic herring on Georges Bank and was started in 1999 (Overholtz et al. 2006). However, during the 2009 TRAC assessment for Gulf of Maine/Georges Bank Atlantic herring, the abundance index derived from the NEFSC acoustic survey was excluded from the assessment model. During the assessment it was suggested that a change in the spatial-temporal overlap between the acoustic survey and herring spawning could have biased the index downward at the end of the time series. More generally, concern was raised that the dominant trend in the acoustic survey, a ≈70% decline between the 1999-2001 time period and the 2002-2004 time period (Figure 1), was not apparent in the NEFSC bottom trawl survey indices for Atlantic herring. In this working paper we evaluate changes in the timing and distribution of Atlantic herring egg hatching using larval herring data collected during the NEFSC ichthyoplankton surveys. The objective of this working paper is to evaluate the hypothesis that a change in overlap between the acoustic survey and the distribution of spawning on Georges Bank underlies the decline in the acoustic index

SAMPLING PROGRAMS

NEFSC ichthyoplankton sampling

NEFSC ichthyoplankton sampling is described in detail elsewhere (Richardson et al. 2010). Briefly, the NEFSC has performed 4-8 plankton surveys per year since 1971 using a 61-cm bongo net. Five different sampling programs (ICNAF, MARMAP, herring-sand lance interaction, GLOBEC, ECOMON) have occurred during this time period. Some of these programs have targeted specific species (e.g. GLOBEC, cod and haddock), while others were more general. The result is a consistent sampling method, but variability in the timing and spatial extent of sampling. The Ecosystem Monitoring (EcoMon) program started in its current form in 1999, the same year the acoustic survey was initiated. The EcoMon program is designed to sample twice during the fall spawning season of Atlantic herring. The first fall sampling is piggybacked on the fall trawl survey which generally occupies Georges Bank in early October. The second fall sampling occurs in early to mid November on a dedicated plankton survey. An additional Jan-Feb survey also provides useful information on larval herring abundance and distribution.

Data on the distribution of larval Atlantic herring from NEFSC plankton surveys have previously been used to describe the decline of the Georges Bank herring spawning in the late 1970s and the recolonization of Georges Bank in the late 1980s (Smith & Morse 1993). An index of larval herring abundance has also been developed for the Georges Bank spawning component of Atlantic herring (Richardson et al. 2010). This larval index incorporates functions describing the seasonality of spawning and larval mortality. Interannual variability in larval abundance on Georges Bank was recently proposed to be a function of both the abundance of adult herring spawning on Georges Bank and the survival of herring eggs from haddock predation (Richardson et al. 2011).

NEFSC Acoustic survey

The NEFSC initiated an acoustic survey for Atlantic herring in 1998, and established the current sampling design in 1999 (Overholtz et al. 2006). The details of the acoustic survey operations, equipment and data analysis are described elsewhere. The relevant information for this analysis is the spatial design of the sampling and the timing of the survey.

The acoustic survey samples evenly spaced parallel north-south transects (i.e. a systematic parallel design) off the northern edge of Georges Bank and the Great South Channel (Figure 2). The timing of the survey is designed to sample pre-spawning aggregations of Atlantic herring. The survey has consistently been performed during the last two weeks of September, with the exception of 2007 when the survey occurred during the last two weeks of October (Table 1). During 2003, the survey was repeated three times (Sept 4-12, Sep 18-25, Oct 3-10) with the middle survey used to calculate the index. In 2000 and 2001 Georges Bank was also sampled multiple times, using three different sampling designs (zig-zag, parallel systematic, parallel with random spacing).

METHODS

We first addressed the question of whether the spatial distribution of adult herring in the acoustic survey is consistent with the spatial distribution of larval herring in the EcoMon surveys. The spatial distribution of Atlantic herring in the acoustic survey was determined by first averaging the backscatter attributed to herring along a 0.22° longitude by 0.06° latitude grid for each year of the survey. The grid spacing in longitude was established to match the spacing of parallel transects along the survey. Higher resolution sampling occurs in the north-south direction thus allowing the finer latitudinal grid spacing. For each survey the proportion of the total herring backscatter in each grid cell was calculated; these proportional abundances were then averaged across years to generate the mean distribution map.

Larval herring distributions are a function of spawning locations and larval transport after hatching; larval distributions will tend to be broader than spawning distributions. We used a larval transport model to estimate the locations of egg hatching based on observed larval distributions in our EcoMon surveys. The larval transport model was run forward for 75 days. Initial release locations (N=327) were located on a 1/6th degree grid of stations <200 m depth in the western Gulf of Maine and Georges Bank. Particles were released every three days from mid-September to mid-December. Only 2008 and 2009 releases were available for this analysis; model runs from 1999-2007 are ongoing. An analytical technique was developed to estimate the magnitude of egg hatching at each of the 327 release locations given the observed abundance at age of herring larvae sampled on the EcoMon survey from 1999-2009. There is currently a mismatch between the sample years and model release years used in this analysis; this mismatch does contribute uncertainty to the analysis and will be corrected as more model output becomes available. Notably, many of the dominant circulation features on Georges Bank are consistent year to year.

Our second analysis addressed changes in the spatial distribution of spawning. In the Georges Bank region the spatial distribution of herring spawning primarily changes in the east-west direction. To capture spatial changes in egg hatching locations, we calculated the annual weighted mean longitude of Atlantic herring larvae <9 mm (about 10-15 days post-hatch) during October and November. Only Georges Bank and Southern New England samples were included in this index; samples from the western Gulf of Maine and the Scotian Shelf were excluded.

Finally we addressed changes in the timing of spawning. The temporal distribution of Atlantic herring egg hatching can be calculated based on the \ age distribution of larvae collected during sampling. The methodology we have used to estimate a larval index for Atlantic herring includes functions describing the seasonality of egg hatching and larval mortality (Richardson et al. 2010). Specifically a three parameter skew-logistic function was used to describe the average seasonality of hatching over the entire 41 year time series, while a two parameter Pareto function was used to describe larval mortality. We modified this larval index methodology to estimate inter-annual variability in egg hatching (versus a time-series mean). The skew-logistic hatching seasonality function was replaced with a two parameter normal curve. We further minimized the number of estimated parameters by only allowing the mean day of spawning to vary year-to-year; a single spawning season duration value was calculated for all years.

RESULTS

On average herring were in highest abundance in the acoustic survey at the northern edge of Georges Bank. An area between 68.5 W and 67.5 W contained the highest average abundances of herring in the acoustic survey. During the 1999-2009 period small (<9 mm and <10-15 days post hatch) larval herring were collected in highest abundances along the northeastern portion of Georges Bank, with fewer larvae collected along the western Great South Channel.

The analysis using the larval transport model and observed larval abundance-at-age data suggested a strong concentration of egg hatching at 67.2 W and 42 N for the years 1999-2009. For the years 1999 to 2009 combined, egg hatching was also predicted for the western Great South Channel and the western Gulf of Maine in proximity to Stellwagen Bank. For the period 1999-2009, 81% of egg hatching in the region was predicted to occur on the northern edge of Georges Bank, 12% in the western Great South Channel, and 7.5% in the western Gulf of Maine. Areas of the Gulf of Maine north of 43.5° N were not included in these calculations. In general, the location of highest herring acoustic backscatter corresponded well to the predicted location of highest egg hatching.

From 1977-present the weighted mean longitude of herring larvae varied (Figure 5). From 1980-1992 herring larvae were most abundant at the western edge of the Great South Channel with a mean longitude of 69.5 W. The recolonization of the northeastern edge of Georges Bank shifted the mean longitude of larvae to around 67 W in the mid 1990s (Figure 5). During the first 8 years of the acoustic survey (1999-2006) the mean longitude of larvae of herring larvae in the Georges Bank region remained stable, with a large majority of the larvae occurring on the eastern edge of George Bank (Figure 6). However, a westward shift occurred around 2007, as a higher proportion of larvae were collected along the western Great South Channel.

As with the weighted mean longitude of larvae the estimated mean day of egg hatching has varied over decadal time scales. During the 1980s and early 1990s the mean day of hatching was around day 300. Around 1994, concurrent with the shift in the spatial distribution of egg hatching, there was a shift to a mean day of hatching around day 288. From 1999-2005 the timing of egg hatching remained relatively stable, with certain years (2001, 2004) indicating earlier spawning and others (2005,2007) indicating later spawning (Figure 6).

Discussion

In order to provide a meaningful index of abundance the NEFSC acoustic survey must sample a relatively fixed proportion of the Atlantic herring population. If the timing or spatial distribution of herring spawning changes relative to the survey, the index could be biased. The acoustic index presented at the 2009 TRAC herring assessment declined substantially from 2001 to 2002, and was low for the remaining years. During the same 2001-2003 period, the spatial and temporal distribution of larval herring on Georges Bank remained relatively stable with a peak day of hatching around Oct 15th and a peak location of hatching along the northeastern portion of Georges Bank. Egg durations for Gulf of Maine Atlantic herring at 10° C were 11 days in laboratory studies (Lough et al. 1982), suggesting peak spawning during the beginning of October. With the exception of 2007 the spatial coverage and the timing of the acoustic survey has been relatively stable. This comparison of the acoustic survey design and the larval distribution data does not provide support for the hypothesis that a shift in the timing or distribution of spawning was responsible for the decline in the acoustic index in the early 2000s.

One consideration in evaluating larval herring data is that the relationship between the magnitude of Atlantic herring spawning and the number of eggs hatching into larvae is not fixed in time or space due to variability in egg mortality. On Georges Bank, substantial interannual variability in egg mortality has been suggested. Specifically, major declines in larval abundance on Georges Bank from 1975 to 1976 and 2003 to 2004 have been attributed to increased egg predation by the 1975 and 2003 year classes of haddock rather than reduced levels of spawning (Richardson et al. 2011). This raises a question of whether another scenario is possible, relatively stability in the spatial and temporal distribution of larval herring despite a substantial change in the pattern of spawning. We consider this scenario unlikely, as it requires a concurrent change in the distribution of egg predation and spawning distribution.

Overall, we did not find evidence that the spatial or temporal distribution of Atlantic herring spawning changed in the early 2000s, though there was year to year variability in our estimates of the timing of egg hatching. Our analysis did not provide any evidence that the acoustic survey has violated the requirement that it sample a fixed proportion of the herring population.

References Cited

Lough RG, Pennington M, Bolz GR, Rosenberg AA (1982) Age and growth of larval Atlantic herring, Clupea harengus, in the Gulf of Maine-Georges Bank region based on otolith growth increments. Fish Bull (Seattle) 80:187-199

McQuinn IH (2009) Pelagic fish outburst or suprabenthic habitat occupation: legacy of the Atlantic cod (*Gadus morhua*) collapse in eastern Canada. Can J Fish Aquat Sci 66:2256-2262

Overholtz WJ, Jech JM, Michaels W, Jacobson LD, Sullivan PJ (2006) Empirical comparisons of survey design in acoustic surveys of the Gulf of Maine-Georges Bank Atlantic herring. J Northwest Atl Fish Sci 36:55-63

Richardson DE, Hare JA, Fogarty MJ, Link JS (2011) The role of haddock egg predation in the decline of an Atlantic herring population. Proc Natl Acad Sci U S A 108:13606-13611

Richardson DE, Hare JA, Overholtz WJ, Johnson DL (2010) Development of long-term larval indices for Atlantic herring (*Clupea harengus*) on the northeast US continental shelf. ICES J Mar Sci 67:617-627

Simmonds J, MacLennan D (2005) Fisheries acoustics: theory and practice, Vol. Blackwell Publishing, Oxford, UK

Smith WG, Morse WW (1993) Larval distribution patterns: early signals for the collapse/recovery of Atlantic herring *Clupea harengus* in the Georges Bank area. Fish Bull 91:338-347

Webb J, Fay R, Popper A, Mann D, Hawkins A, Jech J (2008) Active and Passive Acoustics to Locate and Study Fish. In: Fay R, Popper A (eds) Fish Bioacoustics, Vol 32. Springer, New York, p 279-309

Table 1. NEFSC Atlantic herring acoustic surveys from 1999 to 2010. Surveys are numbered and labeled based on the survey design (prlll: systematic parallel design; Syszz: systematic zig zag; Rndpl: random parallel) .Transect lines labeled in red are the ones used to calculate the index for the assessment.

DATE/ CRUISE	Sept. 1st week	Sept. 2nd week	Sept. 3rd week	Sept. 4th week	Oct. 1st week	Oct. 2nd week	Oct. 3rd week	Oct. 4th week
DE199909					prlll16			
DE200008		syspl05	rndpl06	syszz07	prlll08, prlll09			
DE200109			prlll05	rndpl01	zigzg02			
DE200208			prlll06					
DE200308	prlll01		prlll03		prlll05			
DE200413			prlll03			prlll05		
DE200512			prlll02					
DE200615			prlll03					
DE200710							prlll02	
DE200809			prlll01					
DE200910				prlll02				
DE201010				prlll03				

Figure 1: Acoustic survey index for Atlantic herring from the 2009 TRAC assessment.

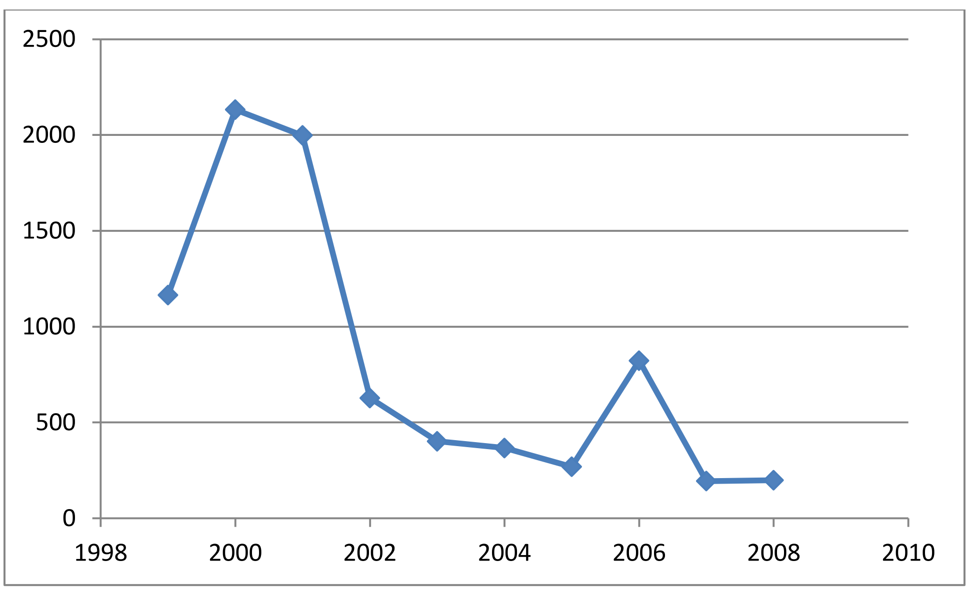


Figure 2. Spatial coverage of the acoustic survey with the systematic parallel sampling design.

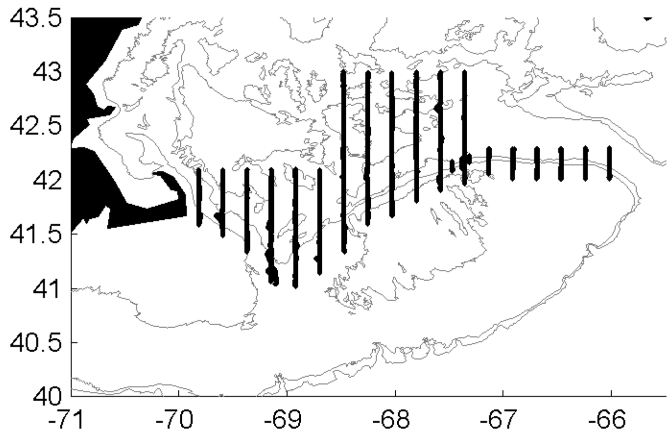


Figure 3. Distribution of small larval herring (< 9 mm) from the October and November ECOMON surveys for 1999-2010. Red x's indicate sampling locations where no small larvae were collected. Circle diameter is proportional to the square root of abundance. The larval distribution is a function of spawning location and larval drift, which is generally clockwise around Georges Bank. Acoustic survey track is overlaid on the figure.

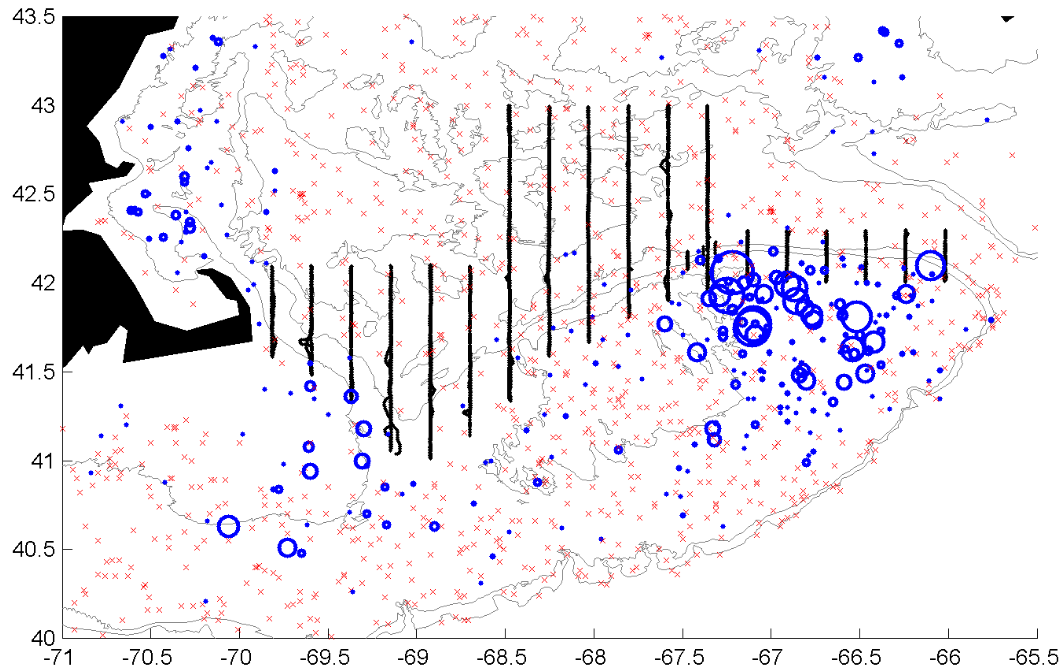


Figure 4. Predicted locations of herring egg hatching (circles) and measured abundances of herring on the acoustic survey (surface) for the years 1999-2009. The egg hatching locations are estimated using a larval transport model and the observed abundances of larval Atlantic herring at age; results are preliminary until further transport model runs are complete.

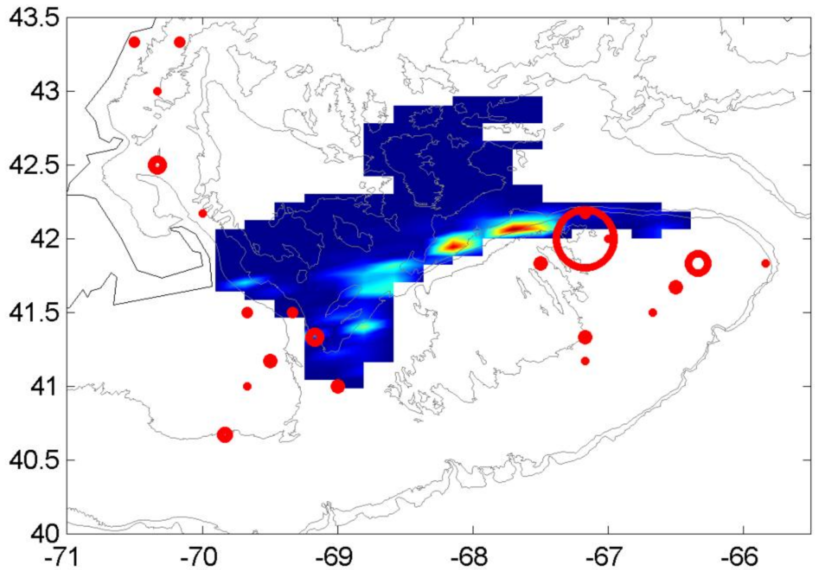


Figure 5. Estimated timing of mean hatch day of larval herring and average longitude of recently hatched larval herring on Georges Bank. Mean hatch day was determined on an annual basis using the approach used to develop a larval index in Richardson et al (2010). A two parameter normal distribution of spawning was substituted for the three parameter skew-logistic curve used in that manuscript. Average longitude of larvae is based on larvae <9mm sampled on either Georges Bank or the broader Nantucket Shoals area during October and November. Values are not calculated during years when the Oct/Nov time period was not sampled. A three year moving average is plotted for each value.

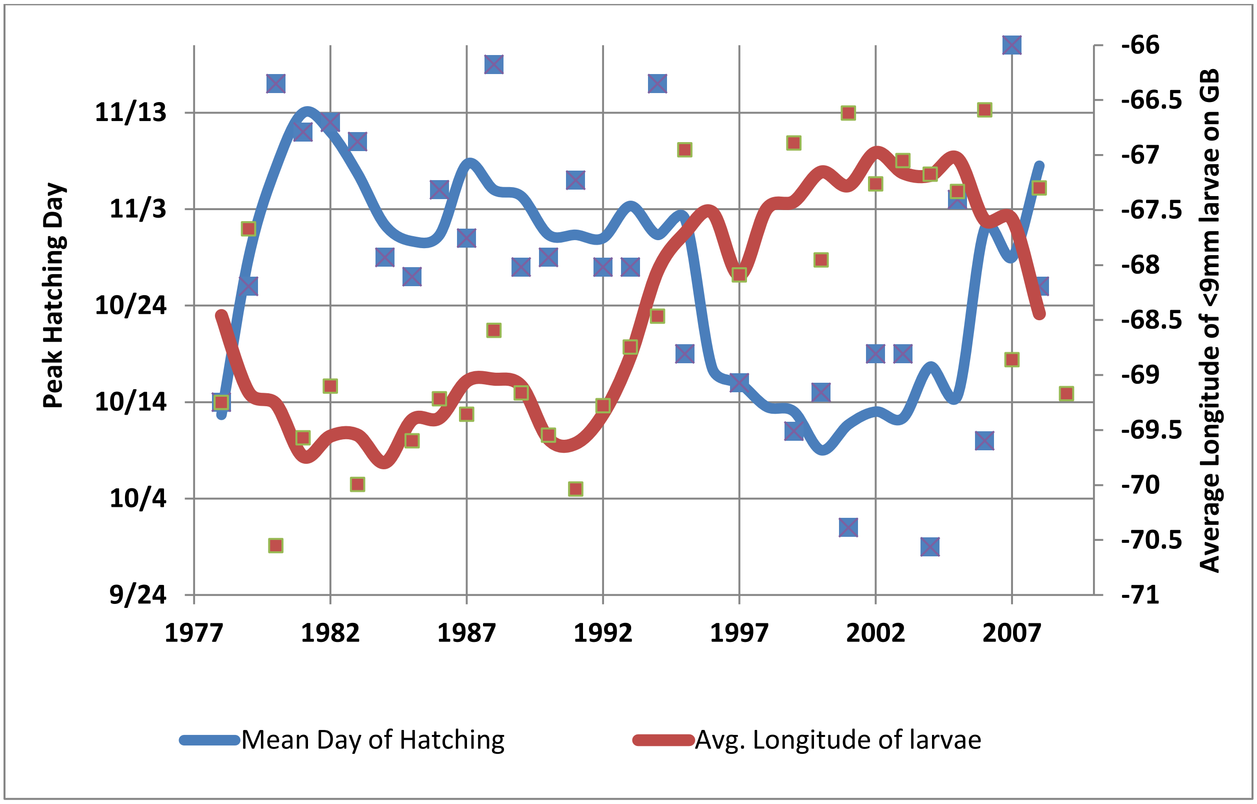


Figure 6. Same as figure 5, but with a focus on the 1999-2009 period of the acoustic survey.

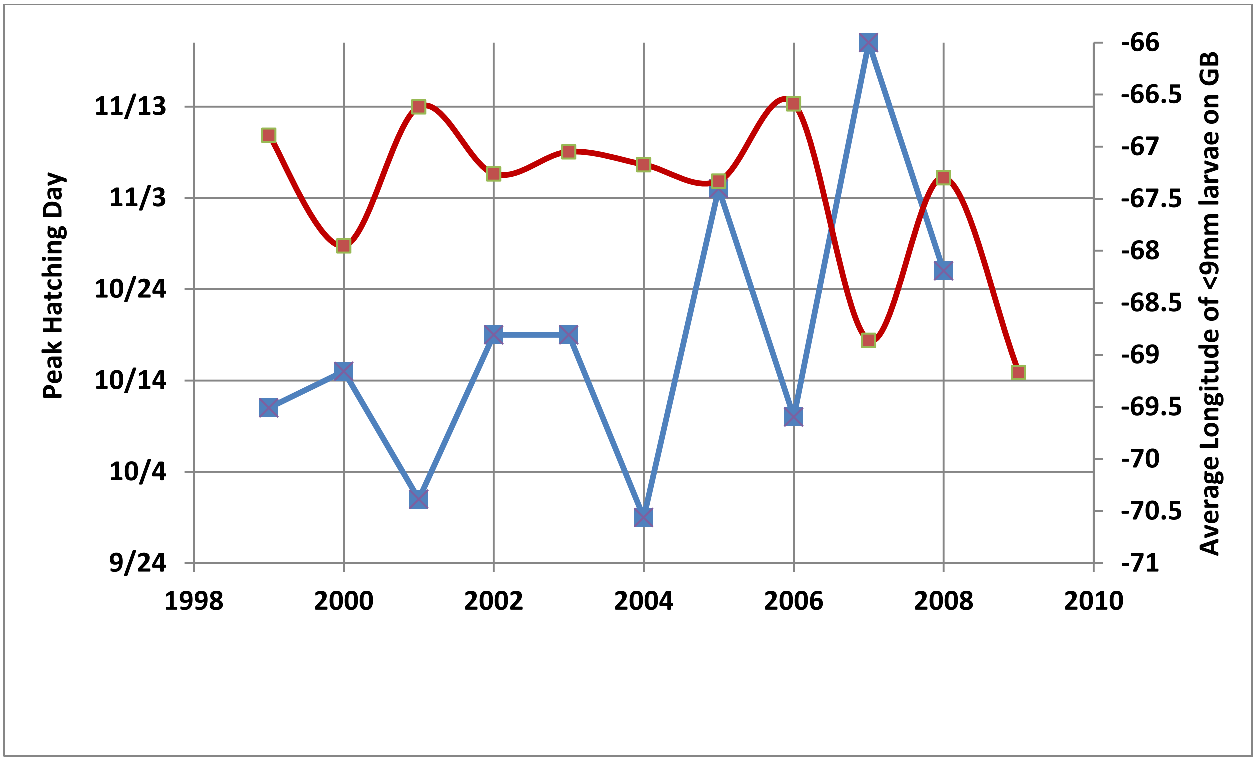
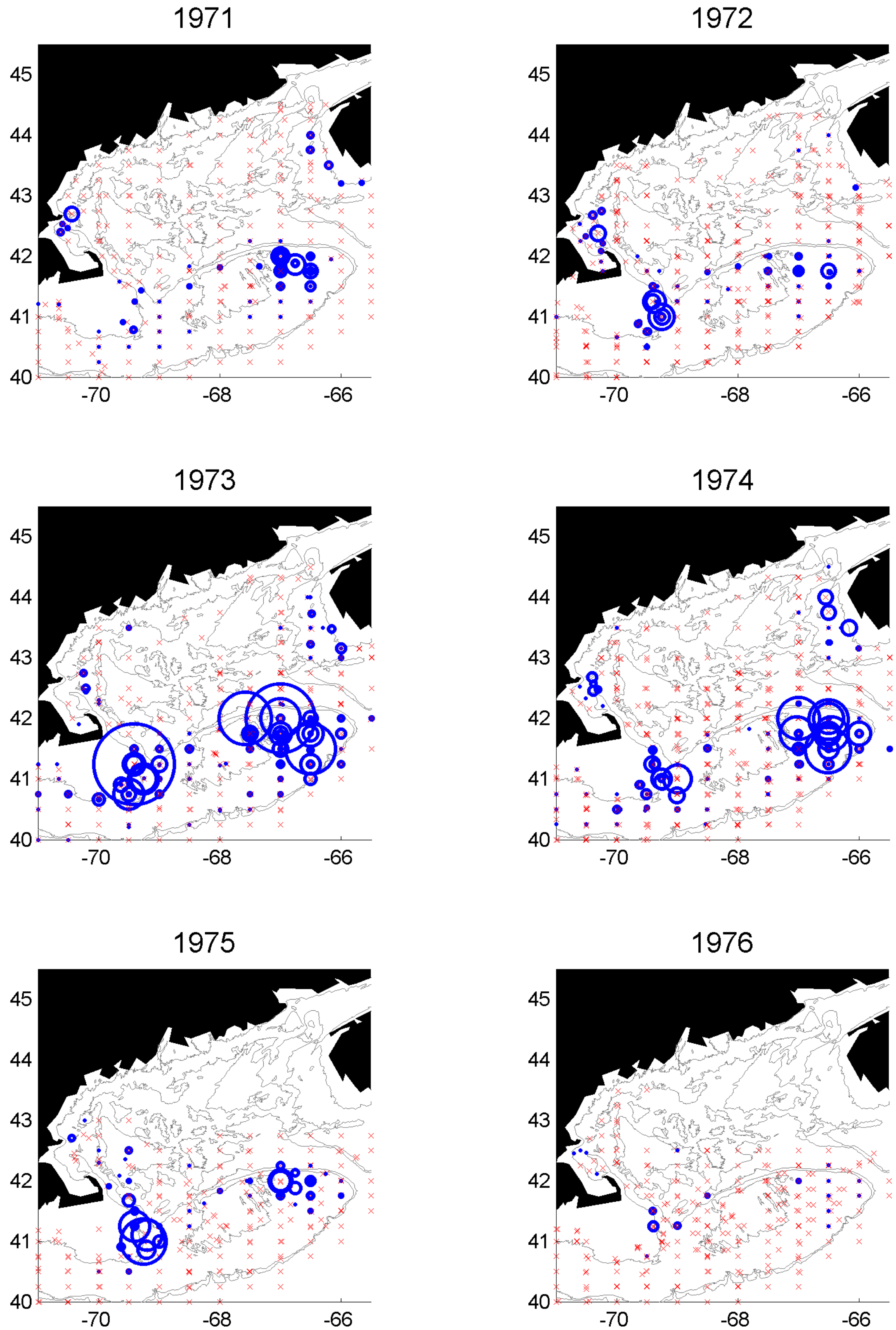
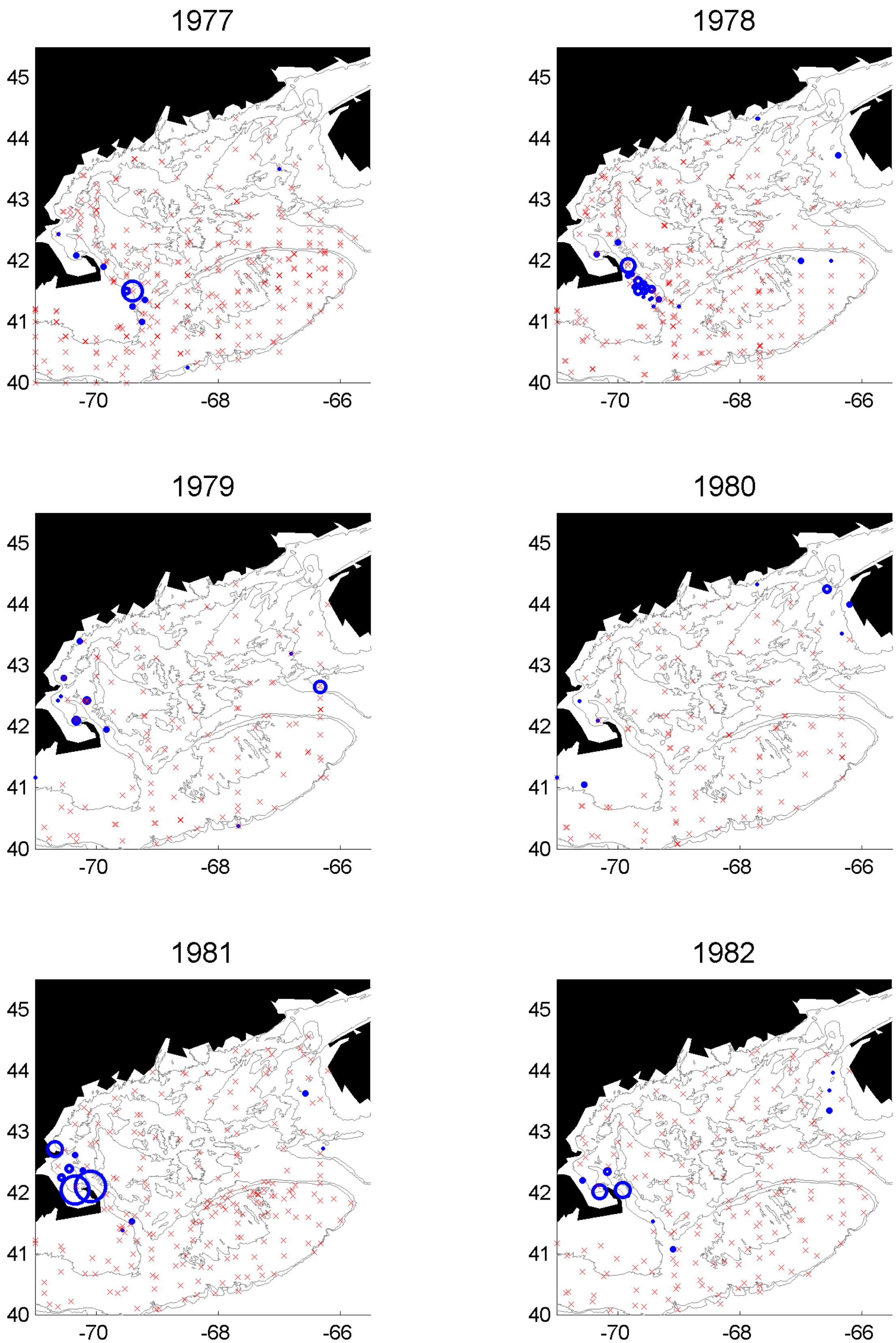
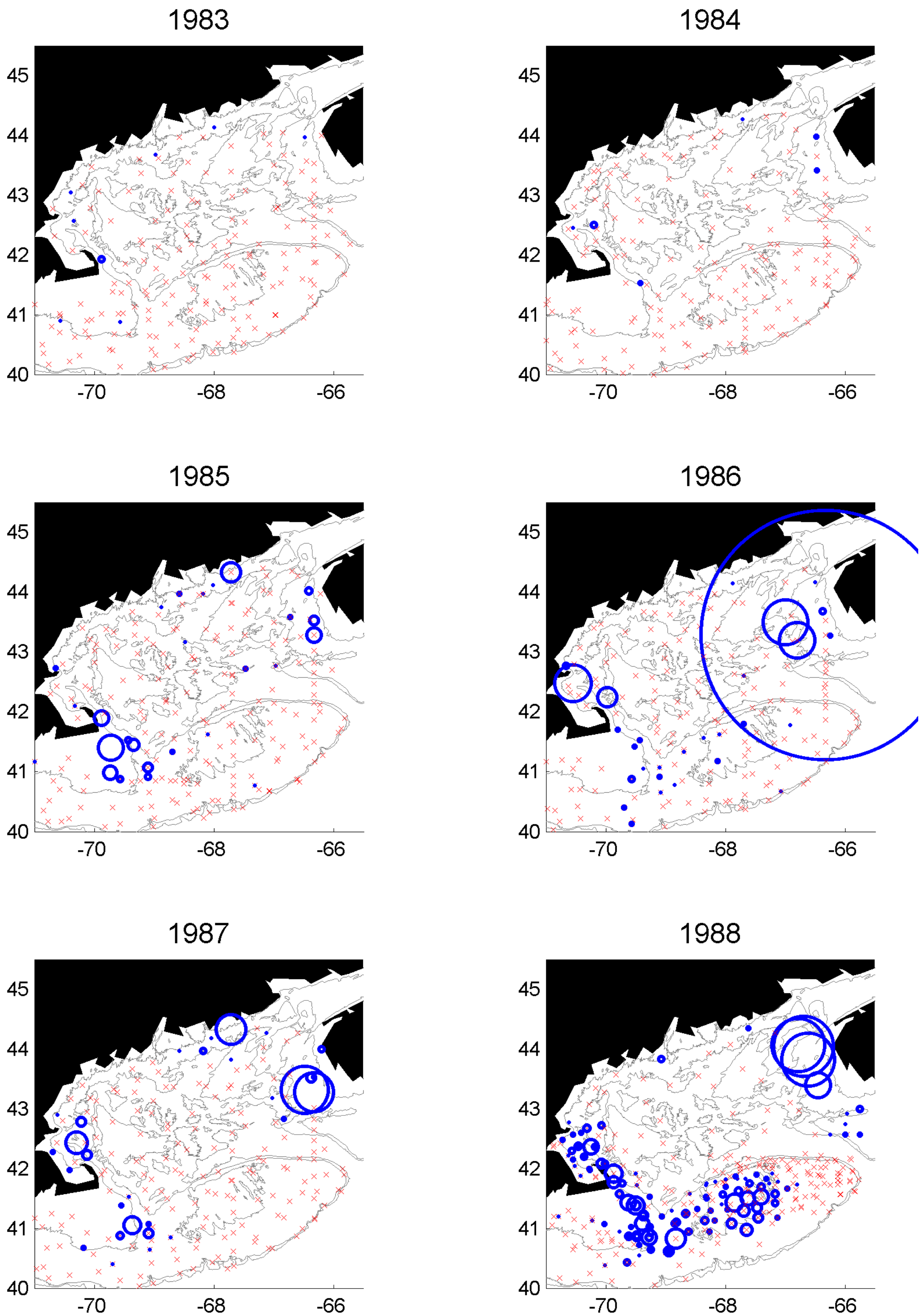
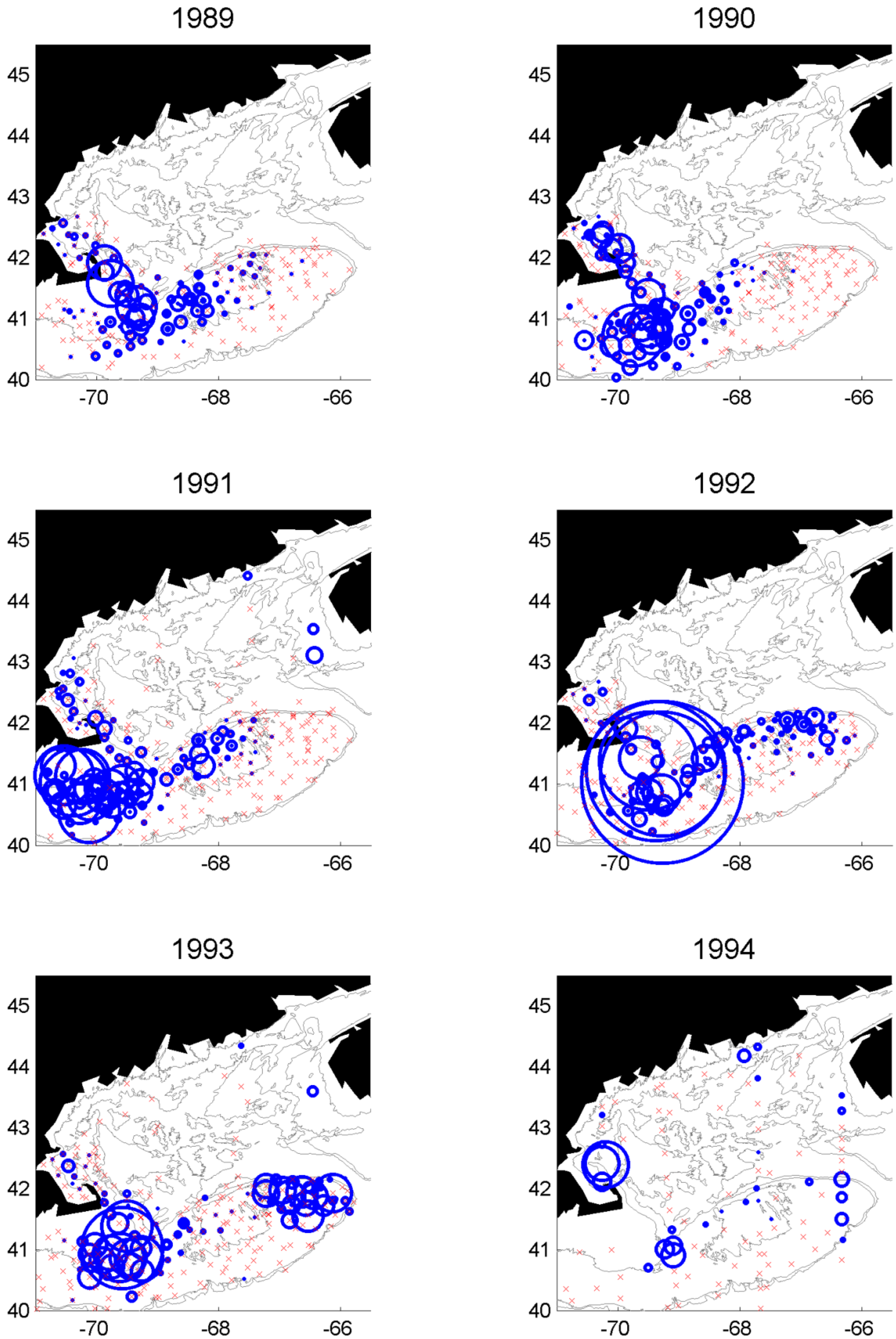


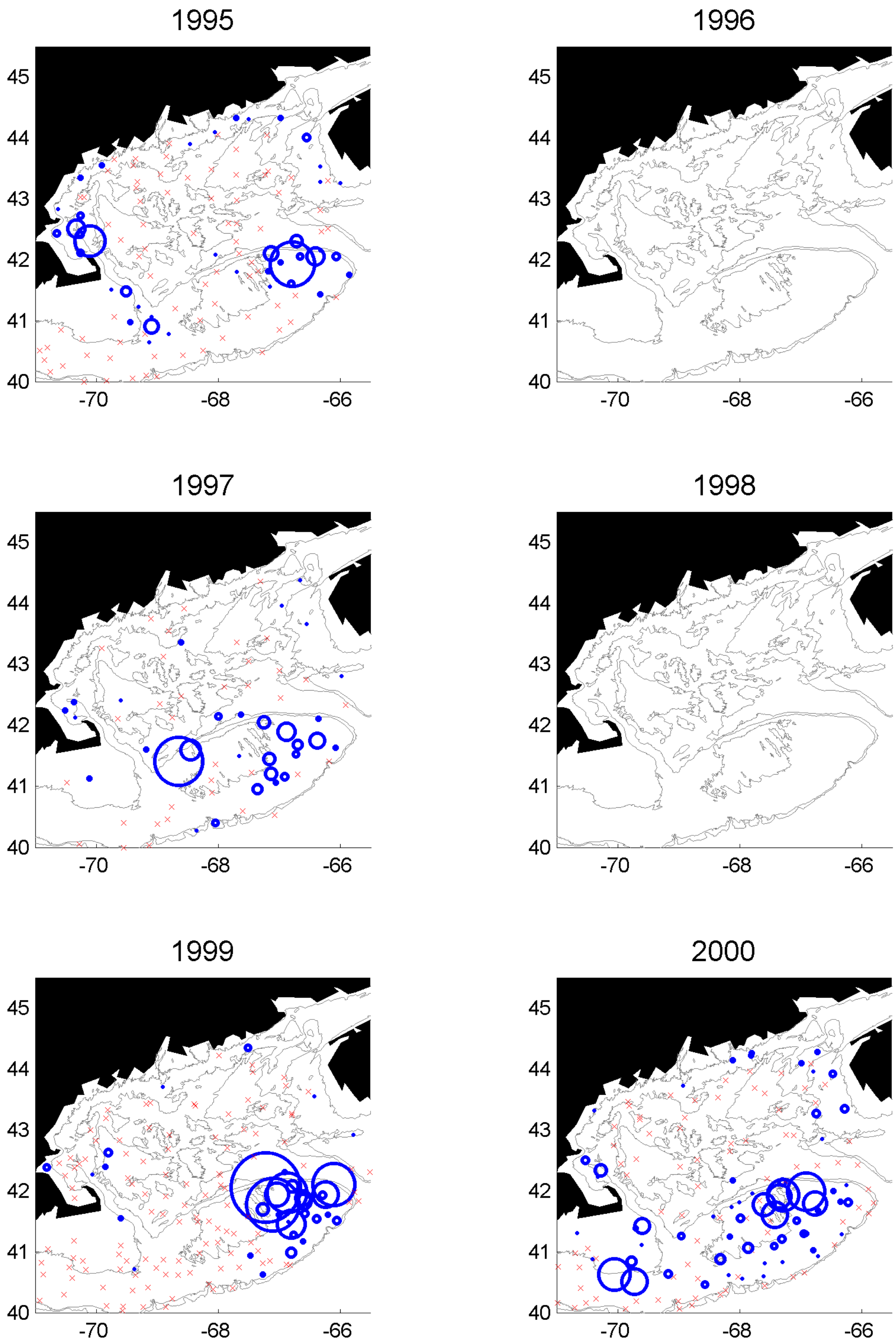
Figure 6 Annual distribution of small larvae (<9mm) during sampling in Oct-Dec.

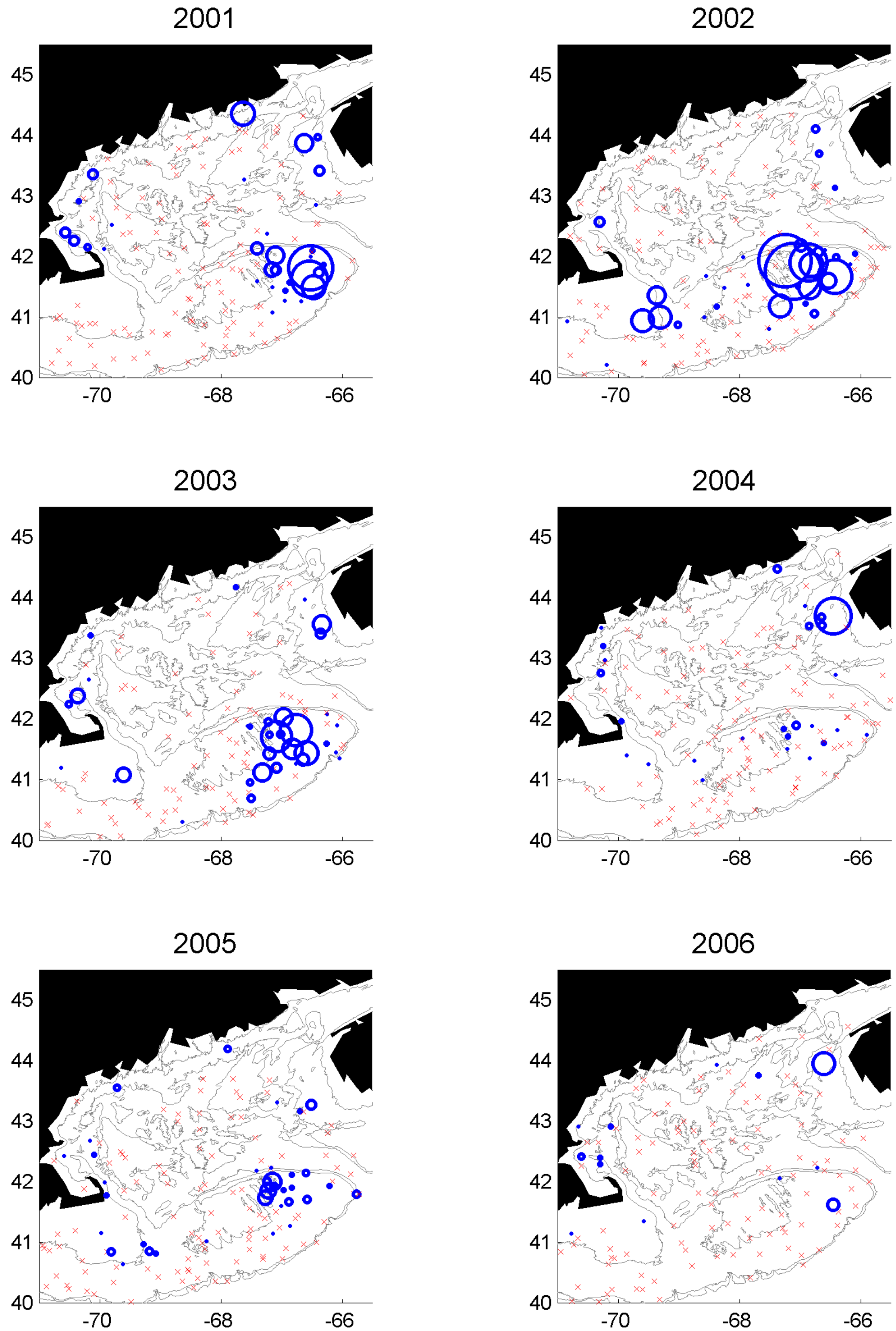


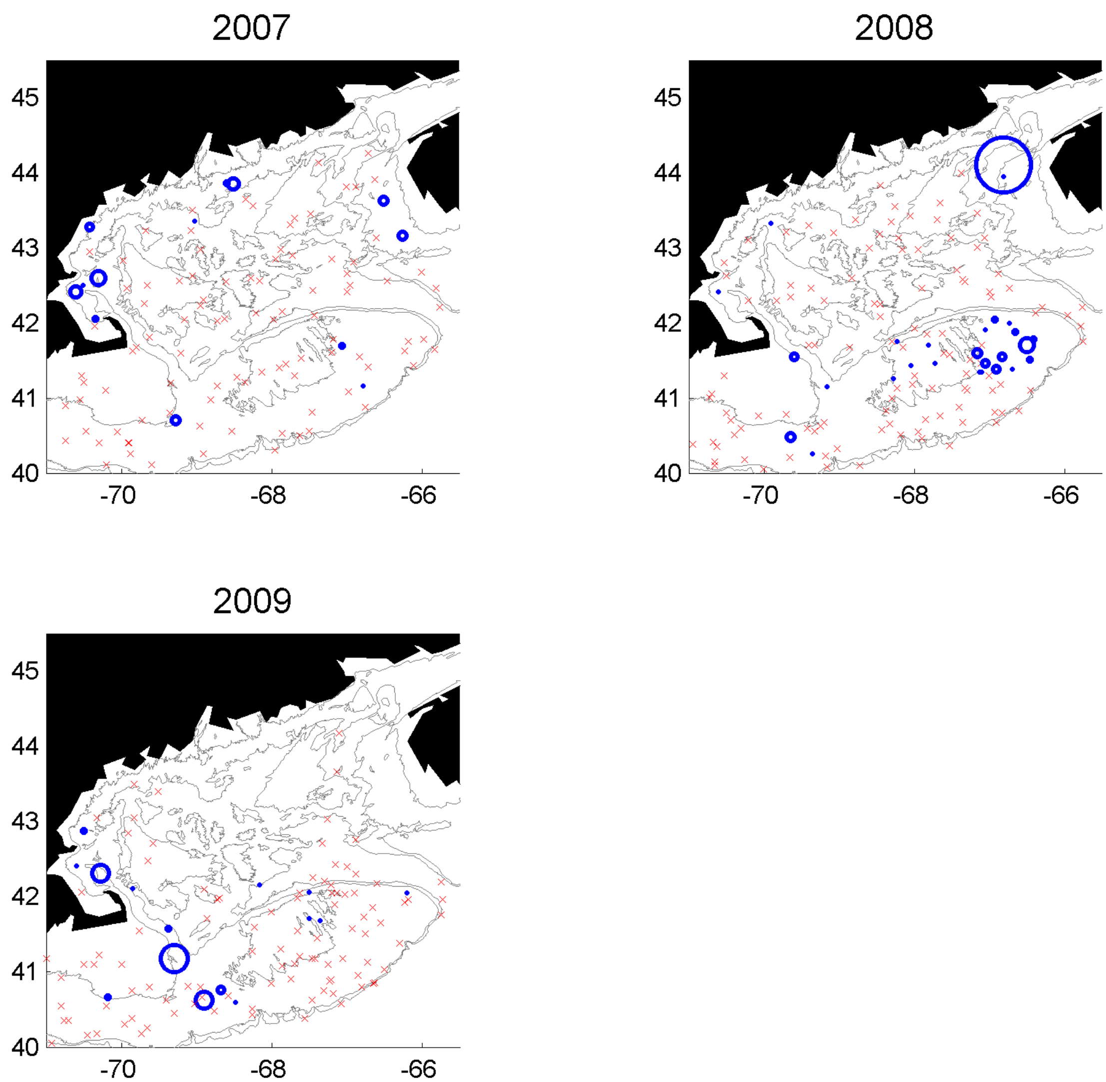

1977
1978
1979
1980
1981
1982
45
44
43
42
41
40
-70
-68
-66


1983
1984
1985
1986
1987
1988
45
44
43
42
41
40
-70
-68
-66


1989
1990
1991
1992
1993
1994
45
44
43
42
41
40
-70
-68
-66


1995
1996
1997
1998
1999
2000
45
44
43
42
41
40
-70
-68
-66


2001
2002
2003
2004
2005
2006
45
44
43
42
41
40
-70
-68
-66


2007
45
44
43
42
41
40
-70
-68
-66
2008
45
44
43
42
41
40
-70
-68
-66
2009
45
44
43
42
41
40
-70
-68
-66

An implementation of ASAP that allows modeling of environmental covariate effects on stock-recruit parameters and application to Atlantic herring

Timothy J. Miller

Northeast Fisheries Science Center, National Marine Fisheries Service, Woods Hole, MA 02543 USA

Introduction

The objective of this working paper is to both present details of an extension of the age-structured assessment model ASAP (ASAP 2008) to allow estimation of covariate effects on stock-recruitment ($ASAP_e$) and investigate models for Atlantic herring that incorporate effects in the stock-recruit relationship.

Methods

Beverton-holt stock-recruit relationship

The Beverton-Holt stock-recruit relationship in ASAP models recruitment at the beginning of year y as a function spawning biomass (S) and unfished spawning biomass per recruit (ρ_0) at time of spawning in year $y-1$ and steepness (h) and, in the next version to be released, unfished recruitment (R_0) rather than unfished spawning biomass,

$$R_y = \frac{\alpha S_{y-1}}{\beta + S_{y-1}} = \frac{4hR_0 S_{y-1}}{\rho_{0,y-1}R_0(1-h)+(5h-1)S_{y-1}}.$$

The unfished spawning biomass per recruit can change from year to year due to inter-annual changes in weight, maturity or natural mortality at age.

The stock-recruit relationship can be modified in various ways to account for effects of auxiliary variables. In this implimentation of ASAP, I allow four alternative modifications. First, transformations of unfished recruitment and steepness are allowed to be to be linear in the covariates,

$$R_0 = e^{\mathbf{X}_{R_0}\boldsymbol{\beta}_{R_0}}$$

$$h = 0.2 + \frac{0.8}{1+e^{-\mathbf{X}_h\boldsymbol{\beta}_h}}$$

This approach is analogous to the way link functions are used in generalized linear models and is helpful in avoiding parameter boundary issues. The other modifications now allowed in the stock recruit relationship involve scalar multipliers to either predicted recruitment (f) or spawning biomass (g). These scalars are modeled as functions of covariates identical to unfished recruitment,

$$f = e^{\mathbf{X}_f\boldsymbol{\beta}_f}$$

and

$$g = e^{\mathbf{X}_g\boldsymbol{\beta}_g}.$$

The resulting general Beverton-Holt stock recruit relationship is

$$R_y = f(\boldsymbol{\beta}_f)\frac{4h(\boldsymbol{\beta}_h)R_0(\boldsymbol{\beta}_{R_0})g(\boldsymbol{\beta}_g)S_{y-1}}{\rho_{0,y-1}R_0(\boldsymbol{\beta}_{R_0})(1-h(\boldsymbol{\beta}_h))+(5h(\boldsymbol{\beta}_h)-1)g(\boldsymbol{\beta}_g)S_{y-1}}$$

where each of the parameters can now change annually depending on the annual values of the covariates.

The f multiplier is intended to model effects of covariates on the recruitment predicted from the stock-recruit relationship whereas the SSB multiplier g is intended to model covariates that change the effective spawning biomass in the stock-recruit relationship. Lastly, there is also an option to use g instead of spawning biomass in the "stock-recruit" relationship. In all cases, the data $\mathbf{X}$ is a design matrix where there is at least one column of 1 for each year of the model and potentially additional columns for covariates. It is probably not advisable to attempt to fit the stock-recruit relationship with covariates in each of the various ways possible simultaneously because there will likely be some confounding of effects. In the absence of user-specified covariates, the default will be to either fix parameters (for f and g) or estimate a single parameters at constant values (for h and R_0) to retain the traditional constant Beverton-holt relationship. Note that the model can be configured to allow effects on expected recruitment through the R_0 parameter without assuming a stock-recruit relationship by setting $h = 1$.

Years where a covariate is unavailable, is a common practical difficulty in fitting these models. This is dealt with by providing an indicator vector of when the covariate is available and allowing the recruitment to influence the objective function only in those years where the covariate is available. This can be useful in evaluating whether the covariate is helpful by comparing fits of a null model (no effect) or the model with the effect estimated where the same years influence the objective function in both cases. The objective function and its components can be inspected for differences between the models. When the objective function is much lower when the parameters are estimated this may suggest that there is an improvement to the overall fit of the model, but there is no real justifiable statistical method of comparison for this type of model.

Atlantic Herring Application

The covariates that I considered were the herring larval index from the data group working paper by Miller et al., the summer temperature series from the Hare data working group paper and the fall Georges Bank haddock biomass index from the most recent assessment (NEFSC 2012). The larval index and summer temperature were investigated based on the results of Hare's working paper and the haddock index was considered based on the results of (Richardson et al. 2011) which found haddock to be an important predator of herring eggs.

For all of these results I take the input file for one of the earlier ASAP models (run51) that Jon Deroba evaluated for Atlantic herring and augment it for use in the ASAP_E version. I fit several models that include the larval index as an explanatory variable affecting steepness, unfished recruitment, and the scalar multipliers f and g. I also fit models without a stock-recruit relationship (steepness = 1) and effects of larval index on f which effectively models the effect of the larval index on annual recruitment. I compared these models to the null models without the effect of larval index on any parameter, but including the same years of recruitments in the objective function (all models described in Table 1). For summer temperature, I fit models with effects on steepness or unfished recruitment and compared them to the null model without the effects, but including the same years of recruitments in the objective function (described in Table 2). For haddock abundance, I fit models with effects on the scalar multiplier g and compared them to the null model without the effects, but including the same years of recruitments in the objective function. The haddock index was included in this way to allow the abundance to change the effective spawning biomass in the stock-recruit relationship. Larval and haddock abundance indices were log-transformed and centered at their mean values for all analyses (described in Table 3).

Results and Discussion

None of the covariates in any of the parameterizations investigated here appeared to provide more than a negligible improvement to the overall fit for run51. For all of the models that included the larval index, the minimized objective function was between 0.67 units less and 2.54 units greater than that of the base (null) run51 model that did not include larval index effects, but only included recruitments in the likelihood for years where the larval index was available (see Table 1). For summer temperature, the largest decrease in the minimized objective funtion was 1.23 for model st_1 where it was assumed to affect steepness (Table 2). Lastly, including the fall Georges-Bank haddock biomass index effects on a modifier of spawning biomass in the stock-recruit relationship results in a minimized objective function 0.22 units lower than the null model.

Of the models fit, st_1 with summer temperature affecting steepness provided the largest reduction in the minimized objective function. Although this model would have an AIC value 0.46 units lower than the null model, there is no justification for using AIC with statistical catch at age models. The estimated coefficient (1.83) had a standard error estimate of 1.27 which would result in a non-significant difference from zero for the coefficient, but again, statistical tests of significance may not be appropriate.

References

ASAP. 2008. Age structured assessment program, version 2.0. NOAA Fisheries Toolbox. NEFSC, Woods Hole, MA. Available at http://nft.nefsc.noaa.gov.

NEFSC. 2012. Assessment or data updates of 13 northeast groundfish stocks through 2010. NEFSC Ref. Doc. 12-06. 789 p.

Richardson, D. E., Hare, J. A., Fogarty, M. J., and Link, J. S. 2011. The role of egg predation by haddock in the decline of an Atlantic herring population. Proceedings of the National Academy of Sciences USA **108**: 13606–13611.

Table 1. All models investigated for Atlantic herring that incorporated the larval index are based on the model configuration run51 provided by Jon Deroba.

Model Name	Description	Difference in # of parameters from li_0	Minimized Objective function
li_0	Larval index null model with no effects, but SRR for years of index is included in objective function	0	3372.73
li_1	Larval index effect on g through slope parameter, $log(g) = \beta_1 \log(LI)$	1	3372.46
li_2	Larval index in place of spawning biomass, $gS = LI$	0	3375.27
li_3	Larval index effect on f through slope parameter, $log(f) = \beta_1 \log(LI)$	1	3372.43
li_4	larval index effect on steepness, $\log((h - 0.2)/(1 - h)) = \beta_0 + \beta_1 \log(LI)$	1	3372.41
li_5	larval index effect on unfished recruitment, $log(R_0) = \beta_0 + \beta_1 \log(LI)$	1	3372.06
li_6	No effect of larval index or spawning biomass, steepness = 1	-1	3374.73
li_7	larval index effect on average recruitment, $log(R_y) = log(R_0) + \beta_1 \log(LI)$	0	3374.19

Table 2. All models investigated for Atlantic herring that incorporated summer temperature (from Jon Hare's working paper) are based on the model configuration run51 provided by Jon Deroba.

Model Name	Description	Difference in # of parameters from st_0	Minimized Objective function
st_0	Summer temperature null model with no effects, but SRR for years of index is included in objective function	0	3452.68
st_1	Summer temperature effect on steepness, $\log((h - 0.2)/(1 - h)) = \beta_0 + \beta_1 \log(ST)$	1	3451.45
st_2	Summer temperature effect on unfished recruitment, $log(R_0) = \beta_0 + \beta_1 \log(ST)$	1	3452.48

Table 3. All models investigated for Atlantic herring that incorporated haddock abundance indices (from NEFSC (2012)) are based on the model configuration run51 provided by Jon Deroba.

Model Name	Description	Difference in # of parameters from hi_0	Minimized Objective function
hi_0	Haddock index null model with no effects, but SRR for years of index is included in objective function	0	3635.17
hi_1	Haddock index effect on g through slope parameter, $log(g) = \beta_1 \log(HI)$	1	3634.95

Appendix 6

Comparison of Atlantic herring acoustic abundance estimates with catch at age model results

May 5, 2012

Acoustic estimates of herring on Georges Bank were conducted in the fall of 2006 by two systems, the NEFSC herring acoustic survey and the MIT OAWRS system. The details were previously described. The Georges Bank stock is one component of the exploited mixed stock complex evaluated in the catch at age model. The percent of fish present on Georges Bank during the acoustic surveys was estimated using the ratio of the NEFSC fall survey results of Georges Bank strata and the entire stock complex. Ratio of number and biomass of the survey expanded population estimates for herring 15 cm and greater were compared. The percentage by number and weight for 2006 as well as the 2005-2007 average is provided in Table 1. These percentages were used to expand the acoustic estimates to the total stock complex for comparison to the catch at age model results.

Various estimates from the acoustic surveys were expanded using both the 2006 ratio and the 3 year average. The candidates were the minimum and maximum values from the two OAWRS integreated methods, the minimum, average and maximum daily OAWRS estimates, and the NEFSC acoustic estimates. Acoustic estimates in number were multiplied by average weight of 0.099 kg in samples during the NEFSC survey. These were compared to the ASAP number and biomass estimates for fish age 2 and greater. Acoustic estimates were conducted in autumn, so for comparisons ASAP January 1 stock sizes for 2006 and 2007 are provided. Two ASAP models are provided; the base model with increased M and the model with only Lorenzen M.

In general the daily estimates from OAWRS under-estimated stock sizes compared to NMFS acoustic and model results. However, the integrated numbers and biomass from OAWRS were quite similar to the ASAP base run. The NEFSC was consistanly less than OAWRS and ASAP base runs, but similar to the ASAP Lorenzen model. The integrated OAWRS, NEFSC acoustic and ASAP models were all similar in scale for 2006.

Table 1. Expansion of acoustic abundance estimates for 2006 using 2006 ratio and 2005-2007 average ratio.

2006 proportion

GB= 14.5%

3 yr avg. = 27%

2006 expanded total number

	OAWRS integrated	% GB	Age 2+	millions
	method 1			
min	1,680,000,000	15%	11,586,206,897	11,586
max	1,770,000,000	15%	12,206,896,552	12,207
	method 2			
min	1,350,000,000	15%	9,310,344,828	9,310
max	1,450,000,000	15%	10,000,000,000	10,000

	OAWRS integrated	% GB	Age 2+	millions
	method 1			
min	1,680,000,000	27%	6,222,222,222	6,222
max	1,770,000,000	27%	6,555,555,556	6,556
	method 2			
min	1,350,000,000	27%	5,000,000,000	5,000
max	1,450,000,000	27%	5,370,370,370	5,370

OAWRS daily	% GB	Age 2+	millions
average			
154,000,000	15%	1,062,068,966	1,062
154,000,000	27%	570,370,370	570
minimum			
52,100,000	15%	359,310,345	359
52,100,000	27%	192,962,963	193
maximum			
325,200,000	15%	2,242,758,621	2,243
325,200,000	27%	1,204,444,444	1,204

	% GB	Age 2+	millions
NEFSC acoustic			
693,000,000	15%	4,779,310,345	4,779
693,000,000	27%	2,566,666,667	2,567

ASAP - total number		Age 2+	millions
Base Run	1-Jan-06	9,193,008,000	9,193
	1-Jan-07	11,988,033,000	11,988
Lorenzen M	1-Jan-06	5,642,008,000	5,642
	1-Jan-07	7,287,197,200	7,287

Table 1. Expansion of acoustic biomass estimates for 2006 using 2006 ratio and 2005-2007 average ratio.

2006 proportion
GB= 18.5%
3 yr avg. = 30.7%
avg wt -acoustic
0.099 kg

2006

	kg		expanded total kg	
	OAWRS integrated	% GB	Age 2+	mt
	method 1			
min	166,320,000	19%	899,027,027	899,027
max	175,230,000	19%	947,189,189	947,189
	method 2			
min	133,650,000	19%	722,432,432	722,432
max	143,550,000	19%	775,945,946	775,946

	OAWRS integrated	% GB	Age 2+	mt
	method 1			
min	166,320,000	31%	541,758,958	541,759
max	175,230,000	31%	570,781,759	570,782
	method 2			
min	133,650,000	31%	435,342,020	435,342
max	143,550,000	31%	467,589,577	467,590

OAWRS daily	% GB	Age 2+	mt
average			
15,246,000	19%	82,410,811	82,411
15,246,000	31%	49,661,238	49,661
minimum			
5,157,900	19%	27,880,541	27,881
5,157,900	31%	16,800,977	16,801
maximum			
32,194,800	19%	174,025,946	174,026
32,194,800	31%	104,869,055	104,869

NEFSC acoustic	% GB	Age 2+	mt
68,510,000	19%	370,324,324	370,324
68,510,000	31%	223,159,609	223,160

ASAP - biomass		Age 2+	mt
Base Run	1-Jan-06	789,864,729	789,865
	1-Jan-07	1,090,800,651	1,090,801
Lorenzen M	1-Jan-06	510,558,758	510,559
	1-Jan-07	692,982,794	692,983

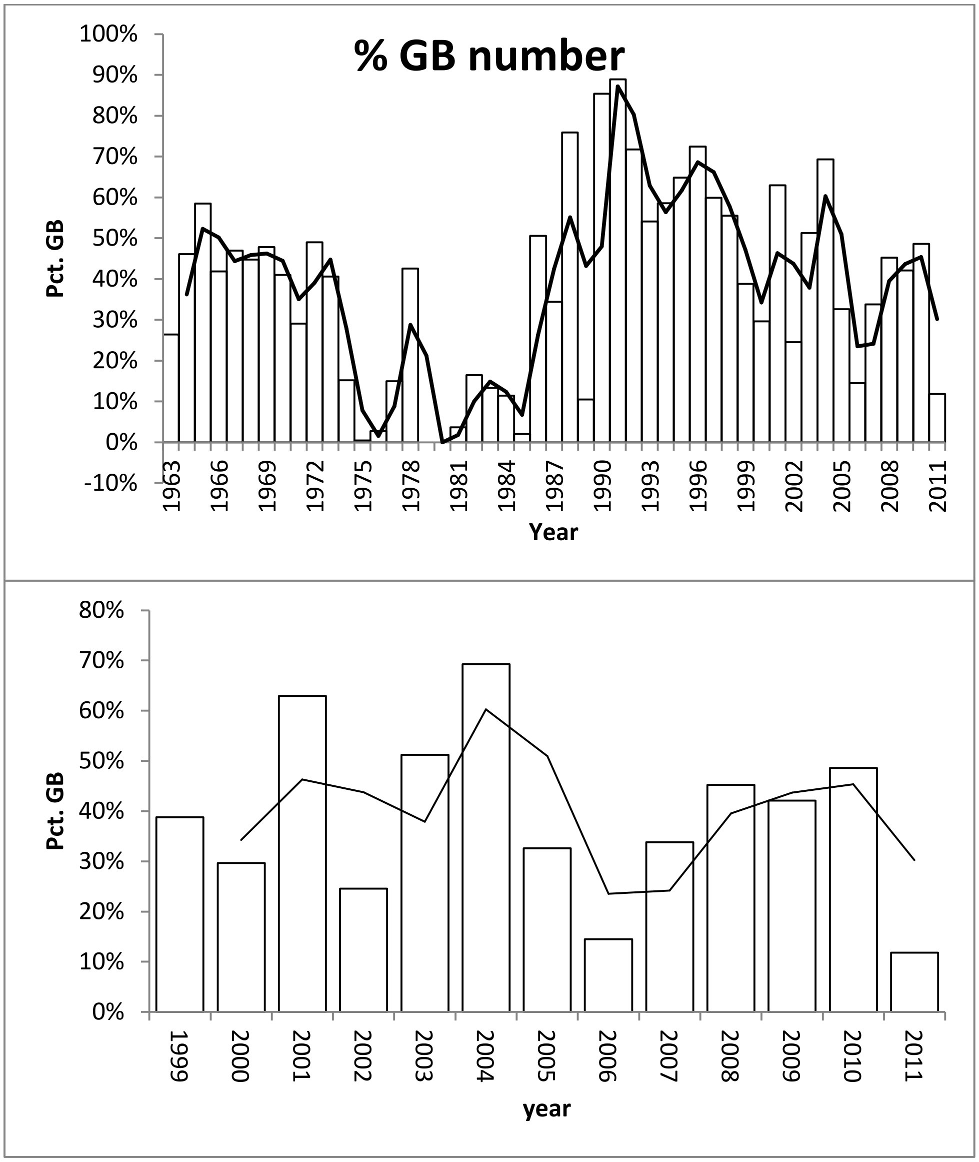


Figure 1. Proportion of herring abundance (>= 15 cm) on Georges Bank from NEFSC bottom trawl survey.

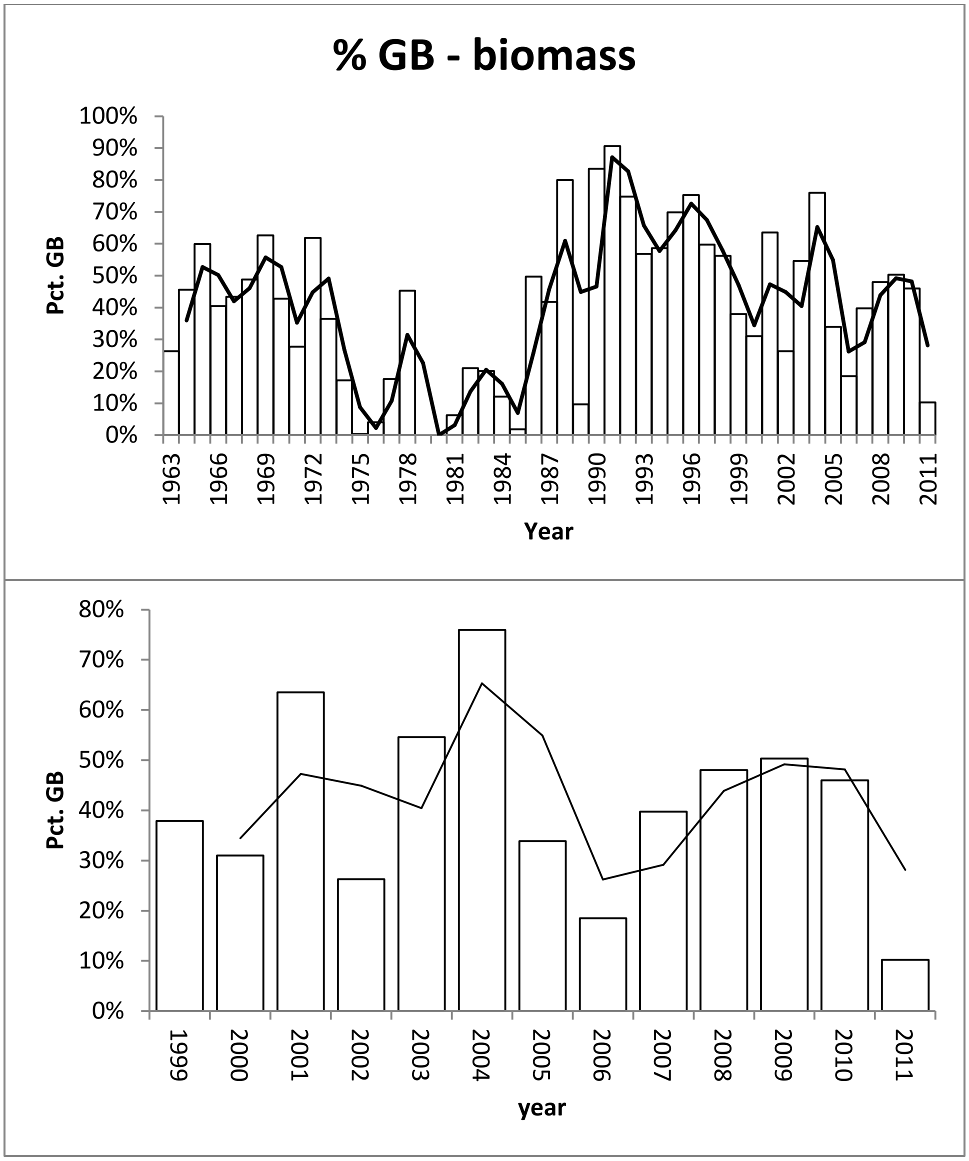
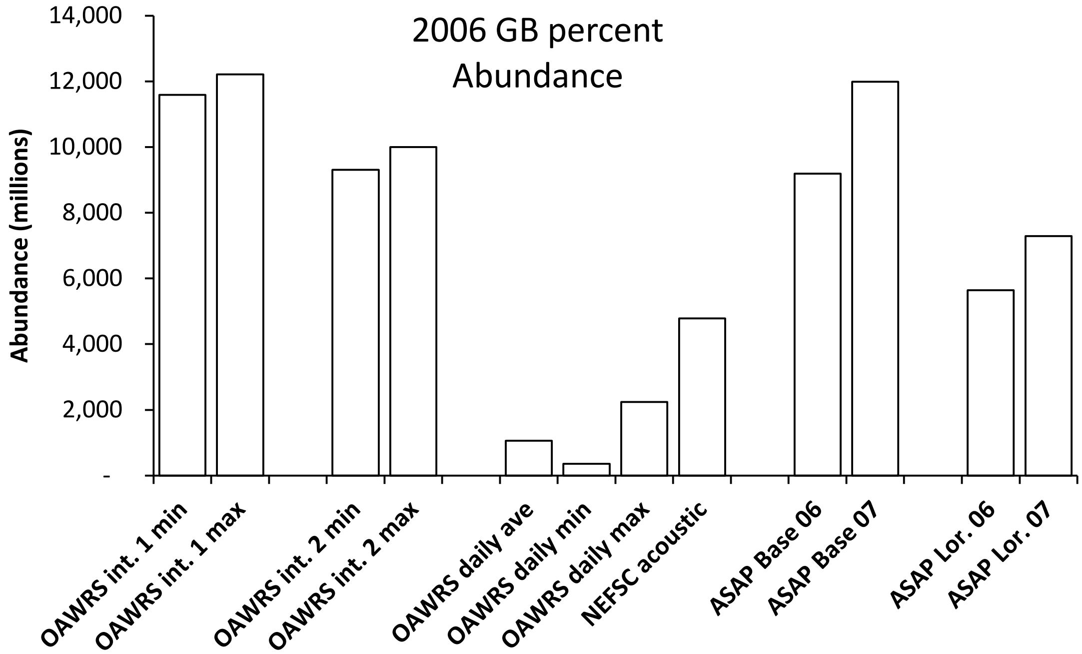


Figure 1. Proportion of herring biomass (>= 15 cm) on Georges Bank from NEFSC bottom trawl survey.



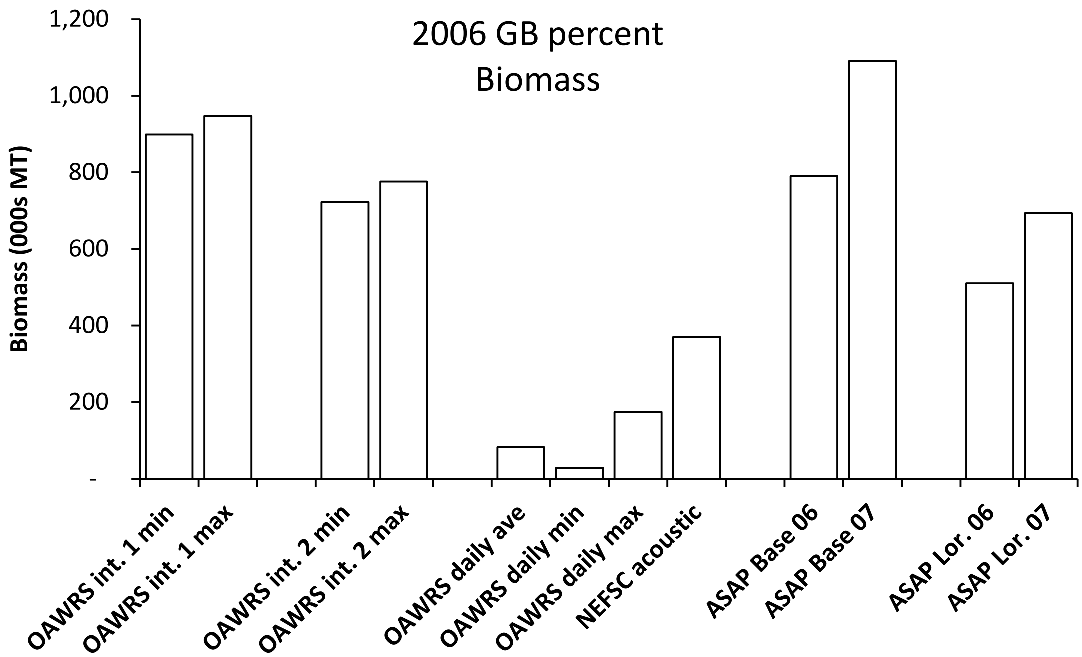


Figure 3. Comparison of abundance and biomass among methods based on 2006 survey ratio.

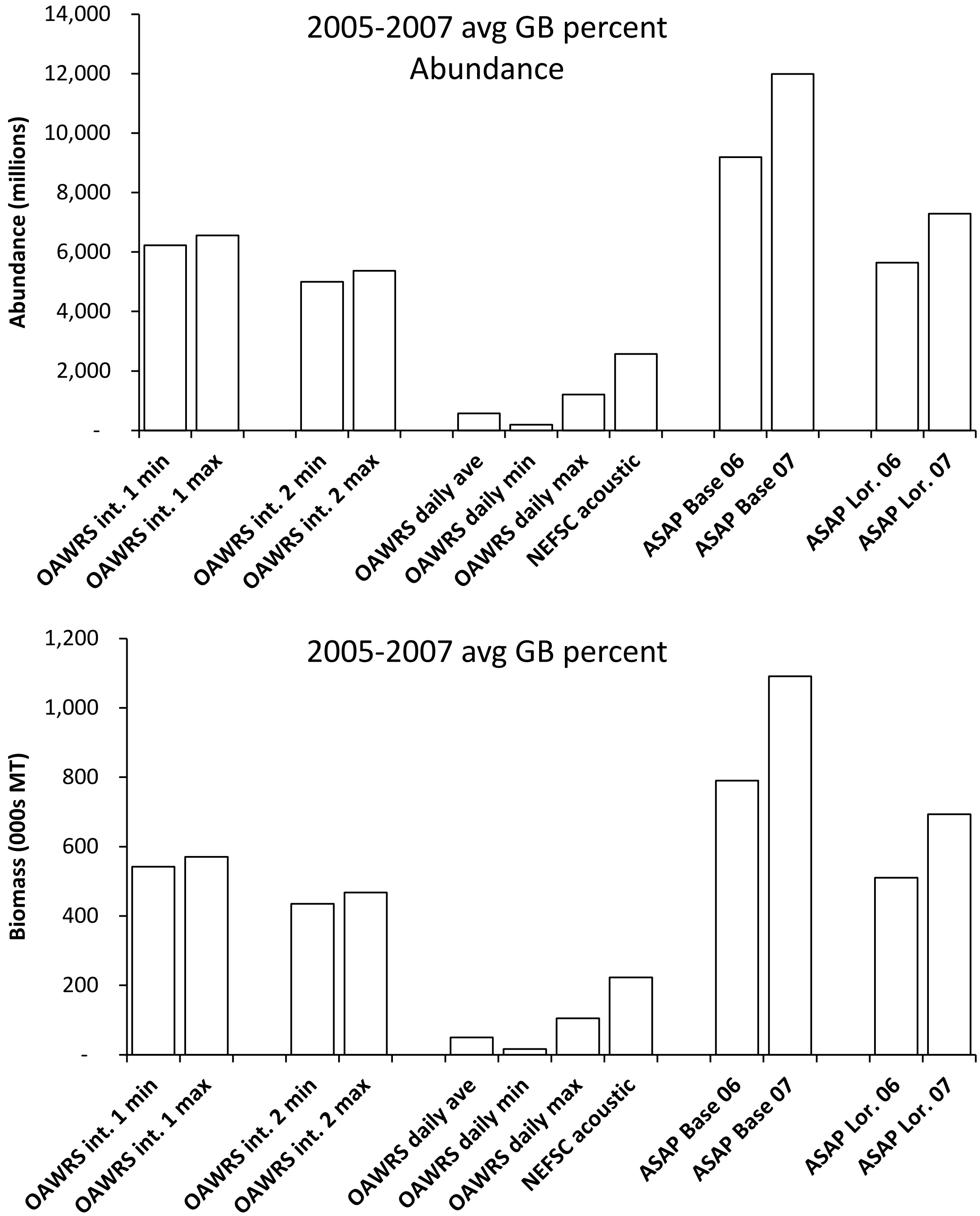


Figure 4. Comparison of abundance and biomass among methods based on 2005-2007 survey ratio.

Appendix 7

A summary of analysis done during the SAW/SARC 54 meeting

Jonathan J. Deroba

Throughout the course of the SAW/SARC meeting several analyses were undertaken to evaluate the uncertainty and robustness of the assessment model to various parameters. These analyses are summarized in this appendix.

Evaluating the 50% increase in natural mortality during 1996-2011

The 50% increase in natural mortality (M) beginning in 1996 in the base model was evaluated using alternative increases of 0%, 30%, 40%, 60%, and 70%. Furthermore, the sensitivity of the model to rescaling the Lorenzen M rates to the average value of 0.3 produced by the Hoenig method was tested by reducing the average M among ages in each year to 0.2 (Hoenig 1983; Lorenzen 1996). The value of 0.3 was produced by using the maximum age herring observed in commercial or survey catches (age 14). Age data, however, was only collected after several years of significant exploitation. So, the maximum age may actually be greater than 14. A maximum age greater than 14 would generate a lower M using the Hoenig method. Consequently, only a reduction in the average M was explored. The value of 0.2 was arbitrary, but is a conventional value used for stock assessment and was sufficient to address the sensitivity analysis. The 1996-2011 M values in the M=0.2 sensitivity analysis were increased by 90%, which produced a Mohn's rho similar to that of the base ASAP run.

Each of the sensitivity runs were compared to the base model using fit to data, degree of retrospective pattern, and similarity between levels of implied consumption and estimates of consumption. Fit to data was compared using the negative log likelihood values for fits to survey trends and age composition. The degree of retrospective pattern was evaluated using the Mohn's rho estimated for spawning stock biomass using the average of a 7-year peel. The similarity between implied levels of consumption and estimates of consumption was compared using the ratio of the geometric mean of the implied consumption values to the geometric mean of the consumption estimates. These ratios were calculated separately for the periods before and after 1996 when the 50% increase in M was used in the base model (i.e., 1968-1995 and 1996-2010). Because the estimates of consumption do not fully account for all sources of natural mortality, ratios greater than 1.0 were preferred, which would suggest that the implied levels of consumption are slightly greater than the estimates of consumption.

Based on the comparisons to the sensitivity runs, the base model 50% increase in M during 1996-2011 seemed appropriate. For all data sources, the base assessment model provided the best fit or within two likelihood values of the best fit (Table 1). Only 60% and 70% increases in M during 1996-2011 produced smaller Mohn's rho values than the base model (Table 1). These two runs, however, produced implied levels of consumption during 1996-2011 that were higher than estimates of consumption, and less consistent than the implied levels of consumption from the base model (Table 1).

Projections

Several sensitivity runs of projections through 2015 were conducted.

1) The results of projections from the base run were compared to the reference points from an assessment run with no increase in M during 1996-2011 (i.e., original Lorenzen values; 0% increase). This comparison was intended to evaluate the sensitivity of the probability of overfishing/overfished to the reference points produced using different assumptions about M during 1996-2011. For all the harvest scenarios projected, the probability of overfishing and for the stock to become overfished equaled zero (Table 2). These results are similar to the projections done exclusively with the base model, suggesting that stock status and the probability of overfishing/overfished are robust to the assumptions about M during 1996-2011 and the subsequent reference points.

2) Projections were conducted at F_{MSY} for the sensitivity assessment run described above with the average M in each year equal to 0.2 and a 90% increase in the underlying average M values during 1996-2011. This sensitivity was intended to evaluate the robustness of the probability of overfishing/overfished to an alternative assumption about M. Numbers-at-age in 2012 were drawn from 1000 vectors of numbers-at-age produced from MCMC simulations of this assessment sensitivity run. The projection results were compared to reference points estimated for this sensitivity run. The probability for the stock to become overfished equaled zero, suggesting robustness to alternative assumptions about M (Table 3 and 4).

3) Projections were conducted at F_{MSY} with the base assessment model reconfigured so that steepness in the stock recruitment model was fixed at 0.35 or 0.85, which approximate the 95% probability intervals of this parameter in the base model. This sensitivity was intended to test the robustness of the probability of overfishing/overfished to a range of steepness values, which was an uncertain parameter in the base model. Numbers-at-age in 2012 were drawn from 1000 vectors of numbers-at-age produced from MCMC simulations of each assessment sensitivity run. The projection results were compared to reference points estimated for each sensitivity run. The probability for the stock to become overfished equaled zero for both values of steepness, suggesting robustness to alternative assumptions about steepness (Table 3 and 4).

4) The robust nature of the assessment model results in the sensitivity runs for projections described above may be driven by the 2009 age 1 cohort, which was estimated to be the largest recruitment on record. To test the sensitivity of the probability of overfishing/overfished to the presence of this cohort, projections using the base assessment model through 2015 at F_{MSY} were conducted with the size of that cohort cut in half, which made the 2009 age 1 cohort approximately equal to previous high recruitments. The probability of the stock becoming overfished remained at zero, suggesting robustness to the size of the 2009 age 1 cohort (Table 3 and 4). Furthermore, an assessment model sensitivity run was conducted with the variation of the annual recruitments from the underlying Beverton-Holt stock recruitment model more restricted than in the base model. In the base model, the coefficient of variation (CV) that partially defined how much the recruitment deviations could vary from the underlying Beverton-Holt relationship equaled 1, but in the sensitivity run the CV equaled 0.67. The value of 0.67

was the CV of the recruitment deviations estimated in the base assessment model. This sensitivity suggested that even with these additional restrictions on recruitment variation, the age 1 2009 cohort would still be the largest on record.

Assessment model sensitivities

The base assessment model was tested for sensitivity to the way in which age composition data were weighted in model fitting. More specifically, the input effective sample sizes (ESS) were iteratively reweighted as described in Francis (2011). The input ESS used in the base assessment model for the mobile gear fishery, fixed gear fishery, spring survey during 1985-2011, and fall survey during 1985-2011 were multiplied by 0.37, 0.44, 0.63, and 0.28, respectively. The base assessment model and the results from the sensitivity run with the ESS values reweighted produced generally similar results (Figure 1).

The base assessment model was tested for robustness to age variation in the input M values. An assessment model was fit without the age varying M values that were used in the base model. More specifically, in this sensitivity run the M for all ages during 1965-1995 equaled 0.3 and during 1996-2011 equaled 0.45. Fits to the data were similar between the base model and the sensitivity run and the two models produced generally similar results (Table 5; Figure 2). So, although age variation in M may be justified using biological or theoretical arguments (Chen and Watanbe 1989; Lorenzen 1996; Chu et al., 2008), such additional realism does not necessarily lead to pragmatic differences in model results and may not be parsimonious. Age variation in M can, however, improve fits to data relative to using a constant M.

References

Chen, S., Seiichi Watanabe. 1989. Age dependence of natural mortality coefficient in fish population dynamics. Nippon Suisan Gakkaishi 55(2): 205-208.

Chu, C.Y.C., Hung-Ken Chien, and R.D. Lee. 2008. Explaining the optimality of U-shaped age-specific mortality. Theoretical Population Biology 73(2): 171-180.

Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences 68:1124-1138.

Hoenig, J.M. 1983. Empirical use of longevity data to estimate mortality rates. Fishery Bulletin 82(1): 898-903.

Lorenzen, K. 1996. The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. Journal of Fish Biology 49: 627-647.

Figure 1.—Time series estimates of spawning stock biomass, fishing mortality, and recruitment for the base model and a model with effective sample sizes adjusted as in Francis (2011).

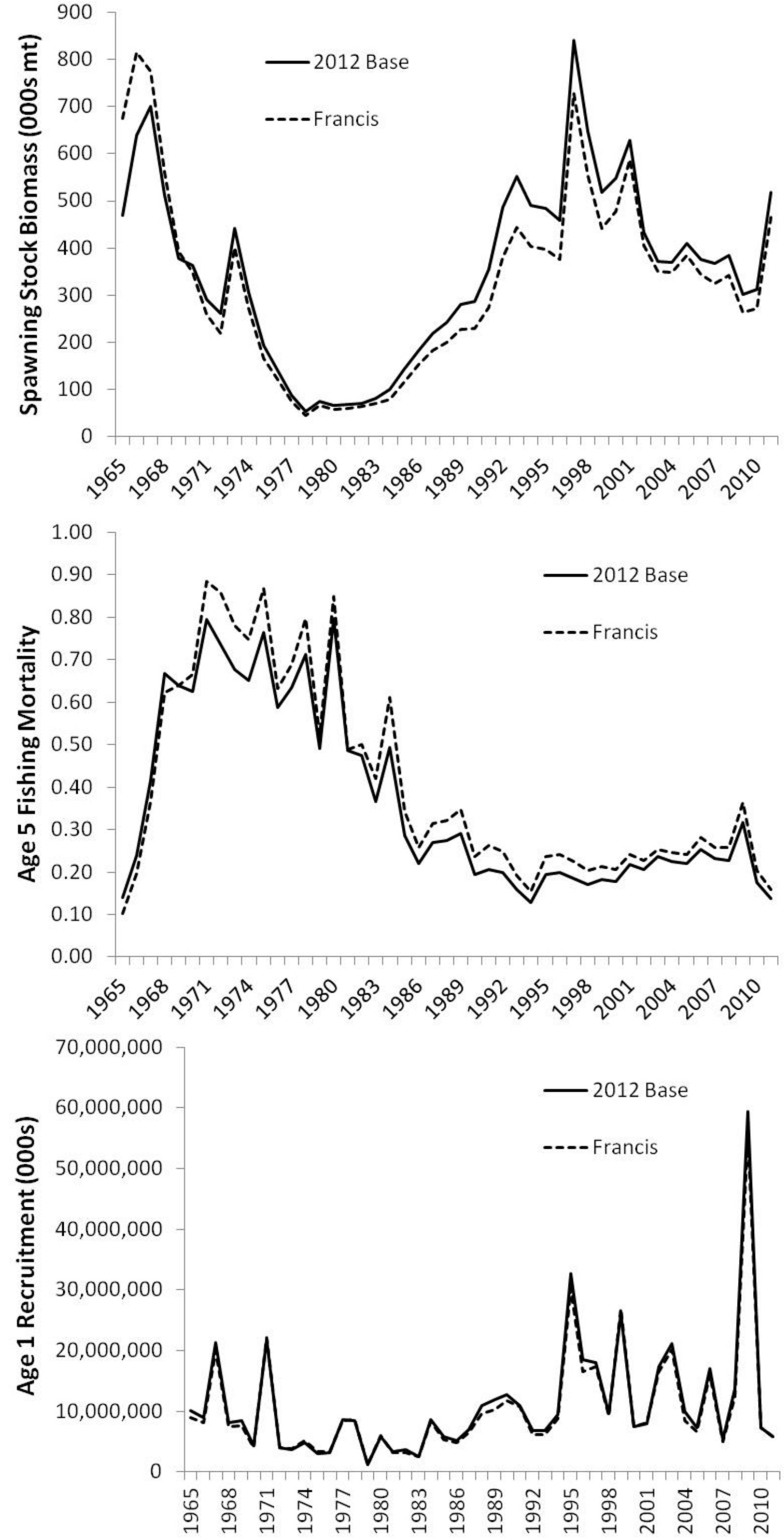


Figure 2. Time series estimates of spawning stock biomass, fishing mortality, and recruitment for the base model and a model without age variation in natural mortality.

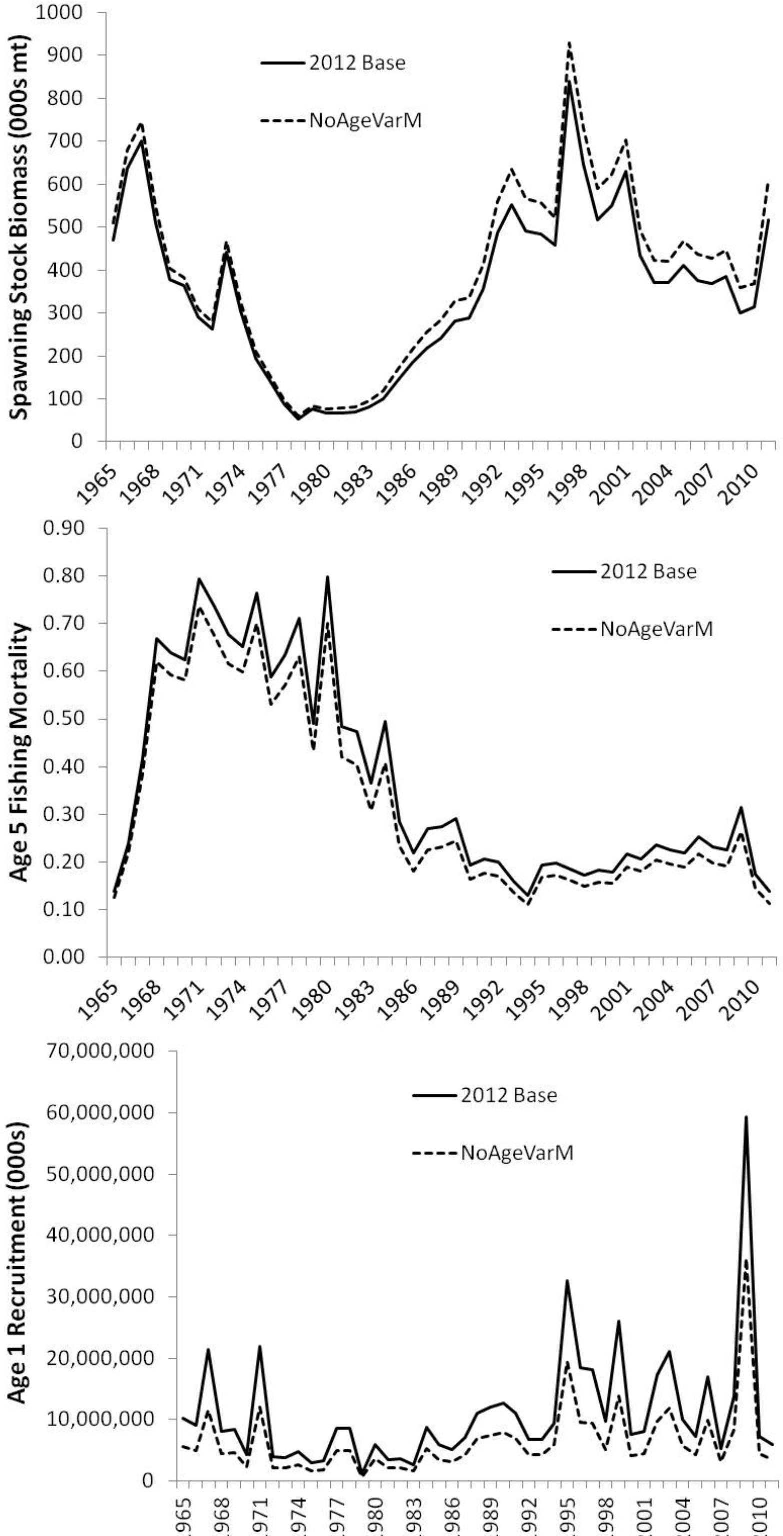


Table 1.—Negative log likelihood values for various data sources, the Mohn's rho for spawning stock biomass (SSB) estimated as the average of a 7-year peel, and the ratio of the geometric means for levels of implied consumption from each run (Imp.) to estimated consumption (Est.) for two time periods, reported for the base assessment model and various sensitivity runs. The Total row is the sum of all the likelihoods in the table for each run.

Comparison Metric	Percent Increase in M during 96-11						
	0% (Lorenzen)	30%	40%	50% (base)	60%	70%	0.2/90%
Spring 68-84	41	41	41	41	41	41	41
Fall 65-84	17	16	16	16	17	20	17
Spring 85-11	117	114	112	111	111	109	111
Fall 85-11	115	115	114	114	114	114	114
Shrimp	111	109	109	109	108	108	108
Catch_Age_Comps	816	815	815	815	815	813	816
Survey_Age_Comps	470	487	471	472	473	473	472
Total	1688	1696	1679	1678	1678	1678	1679
SSB Mohn's Rho	0.85	0.20	0.25	0.13	0.04	-0.08	0.14
Geo Mean Ratio 96-11 (Imp./Est.)	0.54	1.06	1.15	1.40	1.67	2.15	0.83
Geo Mean Ratio 68-95 (Imp./Est.)	0.77	0.87	0.83	0.85	0.87	0.91	0.42

Table 2.—Probabilities of overfishing/overfished estimated by comparing results of projections from the base run to the reference points from a run without an increase in natural mortality during 1996-2011 (original Lorenzen values) using various harvest scenarios.

Lorenzen Ref Points			
Fmsy = 0.41	SSBmsy = 236,428 mt		MSY = 121,580
2012 catch = quota	**2013**	**2014**	**2015**
	F_{msy}		
F	0.267	0.267	0.267
SSB	496,064 mt	368,501 mt	308,949 mt
80% CI	362,965 - 688,585 mt	275,695 - 517-815 mt	237,755 - 411,808 mt
Prob < SSBmsy/2	0	0	0
catch	168,775 mt	126,589 mt	104,430 mt
80% CI	124,868 - 230,764 mt	95,835 - 171,145 mt	79,505 - 139,925 mt
	$F_{75\%\ msy}$		
F	0.2	0.2	0.2
SSB	523,243 mt	409,309 mt	354,559 mt
80% CI	382,573 - 723,975 mt	306,011 - 574,128 mt	272,751 - 473,021 mt
Prob < SSBmsy/2	0	0	0
catch	130,025 mt	102,470 mt	87,574 mt
80% CI	96,216 - 177,894 mt	77,476 - 138,665 mt	66,739 - 117,318 mt
	$F_{status\ quo}$		
F	0.14	0.14	0.14
SSB	548,788 mt	450,496 mt	402,551 mt
80% CI	401,571 - 760,028 mt	336,594 - 631,502 mt	309,334 - 537,414 mt
Prob < SSBmsy/2	0	0	0
catch	93,159 mt	76,823 mt	67,912 mt
80% CI	68,954 - 127,518 mt	58,022 - 104,055 mt	51,752 - 91,001 mt
	MSY		
F	0.08	0.09	0.1
80% CI	0.06 - 0.11	0.07 - 0.12	0.07 - 0.14
Prob > Fmsy	0	0	0
SSB	576,092 mt	492,162 mt	448,725 mt
80% CI	413,046 - 813,298 mt	351,530 - 716,931 mt	321,209 - 633,132 mt
Prob < SSBmsy/2	0	0	0
catch	53,000 mt	53,000 mt	53,000 mt
	Status quo catch		
F	0.13	0.16	0.19
80% CI	0.1 - 0.18	0.11 - 0.23	0.13 - 0.27
Prob > Fmsy	0	0	0
SSB	551,686 mt	446,496 mt	385,995 mt
80% CI	388,989 - 789,568 mt	306,349 - 669,721 mt	259,178 - 569,560 mt
Prob < SSBmsy/2	0	0	0
2012 quota	87,683 mt	87,683 mt	87,683 mt

Table 3. Probabilities of overfishing/overfished at the fishing mortality rate associated with maximum sustainable yield for the base model and various sensitivity runs.

	Base Model		
	2013	**2014**	**2015**
F	0.267	0.267	0.267
SSB	496,064 mt	368,501 mt	308,949 mt
80% CI	362,965 - 688,585 mt	275,695 - 517-815 mt	237,755 - 411,808 mt
Prob < SSBmsy/2	0	0	0
catch	168,775 mt	126,589 mt	104,430 mt
80% CI	124,868 - 230,764 mt	95,835 - 171,145 mt	79,505 - 139,925 mt
	Average M = 0.2 with 90% Increase 1996-2011		
F	0.29	0.29	0.29
SSB	396,643 mt	301,811 mt	254,490 mt
80% CI	283,749 - 545,038 mt	219,886 - 411,460 mt	193,777 - 332,169 mt
Prob < SSBmsy/2	0	0	0
catch	142,085 mt	108,898 mt	90,773 mt
80% CI	102,392 - 192,607 mt	80,695 - 144,607 mt	68,361 - 119,094 mt
	Steepness = 0.35		
F	0.12	0.12	0.12
SSB	605,335 mt	513,679 mt	482,295 mt
80% CI	428,135 - 824,517 mt	369,059 - 707,783 mt	352,699 - 650,573 mt
Prob < SSBmsy/2	0	0	0
catch	90,530 mt	77,524 mt	70,985 mt
80% CI	64,223 - 122,488 mt	56,138 - 103,752 mt	51,441 - 96,428 mt
	Steepness = 0.85		
F	0.7	0.7	0.7
SSB	339,734 mt	179,453 mt	119,242 mt
80% CI	244,841 - 458,585 mt	135,762 - 239,971 mt	92,918 - 161,063 mt
Prob < SSBmsy/2	0	0	0
catch	356,988 mt	192,046 mt	127,255 mt
80% CI	262,388 - 479,137 mt	147,502 - 250,723 mt	96,720 - 174,479 mt
	2009 Age 1 Cohort Reduced by Half		
F	0.267	0.267	0.267
SSB	325,668 mt	268,161 mt	246,368 mt
80% CI	232,900 - 461,216 mt	197,151 - 381,017 mt	187,995 - 332,871 mt
Prob < SSBmsy/2	0	0	0
catch	110,377 mt	92,273 mt	81,708 mt
80% CI	81,128 - 157,019 mt	69,290 - 126,034 mt	61,183 - 111,824 mt

Table 4.Maximum sustainable yield reference points for the base model and various sensitivity runs.

	Base	0.2/90%	Steepness=0.35	Steepness=0.85	2009 Age 1 Halved
F at MSY	0.27	0.29	0.12	0.7	0.27
SSB at MSY	157,000	140,803	277,371	73,305	157,000
MSY	53,000	50730	40051	78,104	53,000

Table 5.— Negative log likelihood values for various data sources from the base assessment model and a model without age variation in natural mortality.

	Base	No Age M
Catch Total	884	884
Index Fit Total	391	392
Catch Age Comps	815	813
Survey Age Comps	472	473