Working Paper 2.1
February 18, 2008

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

Utility of statistical catch at age models for assessing Northeast groundfish stocks (a workshop report)

Larry Jacobson, Chris Legault, Loretta O'Brien and Kathy Sosebee
Northeast Fisheries Science Center
Woods Hole, MA

Introduction and summary

This report describes the outcome of a training workshop using the SS2 (Stock Synthesis, version 2) statistical catch at age stock assessment model for northeast groundfish. The workshop was held at the Northeast Fisheries Science Center (NEFSC) in Woods Hole, MA during 4-7 February 2005. Training was provided by Dr. Ian Stewart (Northwest Fisheries Science Center, Seattle, WA) who is expert in use of SS2 for west coast groundfish and has helped apply the model to other stocks around the world. Dr. Richard Methot (Northwest Fisheries Science Center, Seattle, WA), the creator of SS2, made a number of improvements to the program prior to the meeting to accommodate GARM stocks. Both colleagues provided very useful technical assistance before, during and after the training meeting.

The authors of this report were the primary participants at the workshop. However, colleagues from the Mid-Atlantic Fishery Management Council, Atlantic States Marine Fisheries Commission, School of Science and Marine Technology (University of Massachusetts, New Bedford, MA), and other NEFSC staff were invited to attended as time allowed. The total number of participants was roughly twenty. The workshop was "hands on" and focused on application of the model to three example GARM stocks (white hake, Georges Bank cod, and Georges Bank yellowtail flounder or GBYTF) and for one stock (summer flounder) not involved in the GARM.

The workshop was successful in developing potentially useful but preliminary SS2 models for all three stocks. In particular, a model for white hake was developed that spanned a period with missing fishery and survey age data. Previous attempts to assess white hake using a variety of models were problematic. Estimates from an SS2 model for Georges Bank (GB) cod appear relatively precise, presumably because of high quality of

the available data. Preliminary SS2 models for GB yellowtail flounder suffered from many of the same problems as VPA models in the last assessment and a number of different model configurations were explored in SS2 that help address these problems.

Experience at the workshop with all three stocks indicates that statistical catch at age models are applicable and promising for northeast stocks. SS2 and VPA estimates were generally similar, particularly when configured in similar ways. As exemplified by Georges Bank cod, the data available for northeast stocks are adequate for application of statistical catch at age models. In fact, data available for all three example stocks were more comprehensive that the data available for many stocks where statistical catch at age models are used routinely. Results for white hake show that statistical catch at age models may be useful for stocks with missing age data or stocks currently without an analytical assessment approach. The inherent flexibility of statistical catch at age models may be advantageous in dealing with difficult assessment problems and complex data, although results for GB yellowtail flounder indicate that there are no “silver bullets” and that some chronic issues are likely to persist.

Participants at the workshop agreed that SS2 was substantially more difficult to apply than the traditional ADAPT VPA with graphical user interface, because of the variety of data involved, novelty of the program, and because SS2 is very flexible with a large number of options. Our experience at the workshop indicates that training and technical support are absolutely necessary at the outset for new users. Events elsewhere show that even experienced users require technical support and access to the software development team on an ongoing basis.

Although it has been used around the world for a large number of stocks, SS2 was developed originally for relatively slow growing, long-lived and unproductive rockfish stocks on the west coast of North America with limited data and relatively low fishing mortality rates. A number of changes were made to the program to accommodate different circumstances in the northeast. Additional work along these lines will probably be required if SS2 is used for northeast groundfish. For example, substantial amounts of fishery and survey length data are available for northeast stocks that could be used directly for tuning in SS2, particularly for stocks with limited age data. However, most northeast stocks grow very quickly relative to rockfish on the west coast and it is important to calculate size at the time of the survey accurately in the model. The current version of SS2 can accomplish this by dividing the year up into seasons (surveys are assumed to occur at the middle of the season in which they occur). However, with multiple surveys through the year as many as four or more seasons can be required and run times increase substantially. A better approach for northeast stocks might be to account for growth exactly at the time of each survey, ignoring seasonal time steps.

Workshop results indicate that extensive data available in the northeast may enhance estimation of key parameters in statistical catch at age models. For example, it was possible to estimate spawner-recruit model parameters (maximum recruitment in a pristine stock, steepness and recruitment variance) for all three example stocks. Assessments for west coast rockfish and other stocks with less data typically have to fix one or more spawner-recruit parameters at assumed levels.

Some summary information about preliminary SS2 models for all three example stocks is given below. Additional details for white hake and Georges Bank cod are provided in other working papers. This report contains extra information about SS2

models for GBYTF because the first author used GB yellowtail flounder data during the workshop and because GBYTF is a good example of a difficult assessment. Difficulties are problems of interpretation and not due to limited data.

Comparisons between SS2 and ADAPT VPA (virtual population analysis) models are inevitable in this report because ADAPT VPA is the current standard for northeast groundfish. However, as described below, the purpose was not to decide between SS2 and ADAPT VPA, *per se*. Rather, SS2 and ADAPT should be considered working, well developed examples of the two general approaches (VPA and statistical catch at age modeling). Other statistical catch at age models are available. ASAP (A Stock Assessment Program), in particular, is another statistical catch at age model that has been used successfully for other northeast stocks. SS2, ADAPT and ASAP are all available in the NOAA Stock Assessment Toolbox.[1]

Terminology

Certain terms may be used differently by fishery biologists familiar with VPA and statistical catch at age models. In this report, catch is the total number or weight of fish, including discard if it occurs. Age composition data are the proportions of catch or stock in each age group for one fishery or survey during a single time period. Catch-at-age data, in contrast, are the corresponding number of fish caught in each age group for one fishery or survey during a single time period.

Survey and fishery catch-at-age data are important in VPA models, while statistical catch at age models (and SS2 in particular) typically use commercial and survey age composition data, along with total catch. There is overlap in this regard between the two types of models because catch-at-age is the product of age composition and total catch.

"Selectivity" in this report refers to a set of coefficients (one for each age or length) used to convert age composition of commercial or survey catches to age composition of the stock. Selectivity coefficients measure relative catching power for different age or length groups so that, for example, an age class with survey selectivity 50% will be underrepresented in survey catches by 50%, relative to the age classes with full (100%) selectivity. Selectivity includes the affects, which may be age- or size dependent) due to overlap between the fishing gear and the stock, probability of capture for fish which may enter the gear, the probability of escape after capture, and any other factors which determine relative catch rates for fish of different age or size. For convenience, selectivity coefficients are rescaled so that the largest value in a set is one. Selectivity parameters which can be used to calculate selectivity at age and which describe the shape of selectivity patterns (e.g. logistic selectivity curves) tend to be important in statistical catch at age modeling.

"Catchability" parameters are important in statistical catch at age models and relate total abundance or biomass of fully selected fish in the stock to total catch in the same age or size group. In particular, survey "catchability at age" is the product of catchability for fully selected fish and age specific selectivity in the survey or fishery.

Catchability at age parameters tend to be important in VPA models where survey data for different age groups are usually modeled as independent sources of information. Catchability at age parameters in VPA usually relate the total abundance of fish of a

[1] Contact Alan Seaver (Al.Seaver@noaa.gov) for access to NOAA Stock Assessment Toolbox software.

particular age in the stock to survey catches of the same age group. The distinction between survey selectivity and catchability in VPA is implicit because, as described above, catchability at age is the product of catchability and selectivity.

In VPA, fishery and survey selectivity are not represented explicitly. Survey selectivity and catchability estimates can be extracted *post-hoc* from ADAPT VPA output, however, because catchability at age estimates for different age groups in the same survey can be rescaled to a maximum value of one and because the highest catchability at age for a survey is an estimate of catchability. Similarly, fishery selectivity estimates can be extracted from ADAPT VPA output by rescaling fishing morality rates for each age in a single year to a maximum value of one.

In VPA age specific selectivity values for the commercial fishery is typically called the 'partial recruitment' vector reflecting the concept that the force of mortality may not be fully applied to age classes. Age groups that are vulnerable to the full force of mortality are “fully recruited”. VPA assumes no parametric form for commercial or fishery selectivity patterns. Instead, VPA implicitly assumes that the catches at age data are sufficient to characterize commercial fishery selectivity at age.

If the range of fully recruited ages includes the oldest age groups in VPA or stochastic catch at age models, then the model estimates or assumes a “flat-topped” selectivity pattern for the fishery or survey. When some intermediate range of ages are fully recruited but older ages are not, then the model estimates a “dome shaped” selectivity pattern with highest selectivity at intermediate ages.

In statistical catch at age models, flat topped and dome shaped selectivity curves are represented parametrically by smooth curves of appropriate shape. In VPA, selectivity curves are not an explicit part of the model. The overall shape of the implied selectivity curve for the commercial fishery in VPA models is, however, important because it is used to help define or estimate the fishing mortality rates on each age group in the terminal year of the assessment. Fishing mortality rates in the terminal year are usually among most important estimates from VPA assessment models because of their management significance and because they strongly affect stock abundance and biomass estimates for recent years

Finally, VPA and many statistical catch at age models use different but related and roughly comparable metrics to describe the force of fishing mortality. In VPA models for workshop example stocks, the overall level of annual fishing mortality is typically the average fishing mortality at age for a range of “fully recruited” ages. In stochastic catch at age models like SS2, the overall level of annual fishing mortality is the fishing mortality at the age or ages where selectivity at age is one (maximum).

Methods

All models used annual time steps. Total catch data were assumed accurate in SS2 for the sake of simplicity and comparability. Total catch and catch at age were assumed to include discards and were modeled as though from single fishery. The natural mortality rate (M=0.2) was the same as in previous assessments for all three stocks. Baranov’s catch equation, rather than Pope’s approximation, was used exclusively in all models.

Most SS2 models estimated a recruitment parameter for each year in the model while estimating an underlying Beverton-Holt stock-recruit curve which may have had some affect on stock estimates, particularly on initial and recent recruitment estimates, although substantial deviations from the curve were allowed. Recruit curve parameters estimated in the model were the maximum recruitment level at virgin biomass, steepness and the variance of log scale recruitments around the curve. Estimated variances were realistically large indicating that recruitment estimates were estimated relatively with relatively little constraint around the underlying recruit curve.

SS2 models were run assuming either "one" (females and males) or two (male and females separately) sexes. The two sex models were usually the same as one sex models in most workshop runs (except for GBYTF) because natural morality was the same for males and females in all runs and because any differences between growth for males and females was ignored. The two sex models used here may facilitate further development later and this is a promising area for further work with new assessment models.

Results from two or more SS2 configurations (alternate models) were produced for each example species and compared to results from VPA models from recent assessments. In addition a statistical catch at length model called "SCALE" was applied to white hake (SCALE is also available in the NOAA Stock Assessment Toolbox).

With the exception of runs used to mimic VPA (see below), SS2 models developed at the workshop were typical SS2 configurations using aggregate survey abundance indices and survey age composition data while estimating logistic or domed fishery and survey selectivity patterns. Fishery selectivity patterns in SS2 were allowed to change if necessary beginning in 1985 when the Hague line was established to separate US and Canadian waters and in 1995 to account for changes in mesh size and other management measures at that time. Survey length composition data were used in white hake SS2 models because of missing age data, but not for GB cod or yellowtail flounder for the sake of simplicity and comparability.

Survey abundance data (mean numbers per tow) for all three species were adjusted prior to use in all models to account for statistically significant historical changes in bottom trawl net design, door type and vessels. In some VPA models, survey abundance indices were "broken" in two or more series based on the timing of changes so that separate catchability parameters were estimated for each segment. Some SS2 runs used survey covariates that were dummy variables (with values of 0 or 1) to estimate different catchability parameters for a survey in different periods. In effect, the survey covariates were used to break the surveys in SS2.

Historical catch data without associated length or age composition are often used in SS2 models to "burn in" the assessment (i.e. to help estimate initial abundance and age composition), to link initial to virgin stock conditions, and to stabilize the overall scale of stock estimates. Historical catch data (other than as a long term average) proved unnecessary for the data rich example stocks at the workshop and were excluded from the SS2 models pending further analysis. In initial SS2 model runs, the effective sample size for length and age composition data was assumed the same as the number of positive tows for survey data and the number of trips sampled for commercial data. The assumed effective sample size in later runs was adjusted so that the assumed and apparent (based on goodness of fit) values for effective sample size were roughly the same. Similarly, log

scale standard deviations for survey trend data were adjusted in initial runs to roughly match the variance of residuals. This *ad-hoc* iterative process of adjusting assumed variance parameters is called "tuning" in many SS2 assessments. In contrast, "tuning" is usually synonymous with fitting the models and estimating parameters in VPA modeling.

SS2 calculates survey catchability parameters, predicted values for survey trend data and associated negative log likelihood values (which measure goodness of fit) that are not actually used to estimate parameters. Similarly, it will calculate predicted values and negative log likelihood for age composition data not used to estimate parameters.

In all SS2 runs, except as noted and for special purposes, the likelihood weights for individual data sets, constraints, etc. were either one or zero, so that goodness of fit was quantified based on standard maximum likelihood theory. Using this method, data sets were either turned on (weight 1) and used to estimate parameters or turned off (weight 0) and ignored in estimating parameters. The single exception was catch data which was given a weight of 100 or 1000 so that model estimates and catch data would match exactly (i.e. catch data were assumed to be accurate).

For simplicity in preliminary workshop models, fishery and survey length composition data were assumed accurate for workshop other than GBYTF (see below). Information on age reader precision is collected routinely for northeast stocks and can be added later if required.

SS2 models to mimic VPA

One SS2 model for each stock was configured to mimic VPA (SS2MVPA configurations). The goal was to use the same survey data as in the VPA and increase emphasis on commercial catch at age data. The likelihood weight for commercial catch at age in SS2VPA runs was increased from 1 to 10, so that the model would fit the catch at age more precisely. Survey data for SS2MVPA runs were age specific abundance indices taken directly from VPA input files used in previous assessments (log scale standard deviations were assumed to be s=1.0). Age and length composition data and aggregate survey abundance indices were turned off.

Retrospective analyses

As described in a separate working paper, stock assessment estimates for GARM species may show retrospective patterns. Retrospective patterns are widespread and occur in many types of models (including VPA and statistical catch at age). Mohn's rho, which is a statistic designed to measure the strength of retrospective patterns, was calculated for most combinations of model, configuration and stock.

Retrospective evaluations involved five model runs with successively earlier terminal years, in addition to the runs with the full data set. In an analysis with a model having 30 years of data, the retrospective statistic for F used in this paper would be calculated:

$$\rho_F = \sum_{y=1}^{5} \frac{\left(f_{30-y} - F_{30-y}\right)}{f_{30-y}}$$

Where F_j is the estimated fishing mortality rate for year j from the reference model that includes all 30 years and f_j is the estimated fishing mortality for year j from the model with j as the terminal year. The statistic is calculated using the reference model and an additional five model runs with sequentially lower terminal years. Negative values of the

statistic indicate consistent under estimation relative to the reference model and vice-versa. Statistics closer to zero are better so that, for example, 1 is a better score than either -1.3 or 2. The distribution and statistical significance of a retrospective statistic can be evaluated by simulation but distributional properties were not investigated for this analysis.

Two sets of retrospective plots were made for fishing mortality and spawning stock biomass in each model. One type of plot shows the actual stock estimates from each model run. The other type (suggested by Steve Corriea, Massachusetts Department of Marine Fisheries, New Bedford, MA) shows the time series of $(f-F)/f$ values (proportional changes in initial estimates which are used in calculation of Mohn's rho) for each year and model run.

Different growth curves, but the same natural morality rates were used for males and females in SS2 models for white hake and GBYTF with two sexes. The sex ratio at age zero was assumed to be 50:50. Assumptions about growth in workshop SS2 models probably had little effect on growth because length data were ignored, as they were in VPA models.

Georges Bank yellowtail flounder

GBYTF is a sexually dimorphic and highly productive flatfish that supports an intense fishery. The maximum length of females is about 7 cm larger than male maximum size and females are heavier than males of the same length. The natural mortality rate typically assumed in stock assessment work ($M = 0.2\ y^{-1}$) is for both sexes although females live longer than males. Mortality rates are typically high for GBYTF due to intense fishing pressure and individuals older than 8 yr are relatively uncommon. There is little evidence of strong year classes in commercial and survey catch at age data for GBYTF.

Survey data indicate that GBYTF size at age varied over time with periods of high and low size at age lasting five or more years and with some annual variability that may be due to sampling error. Yellowtail flounder are relatively difficult to age for a short-lived species. As shown in the last GARM, the current age reader is relatively precise (age readings are repeatable) to about age 8. Preliminary data indicate, however, that differences between NEFSC age readers who have worked on GBYTF begin to accumulate at occur at about age 4 (J. Burnett, Northeast Fisheries Science Center, pers. comm.). For lack of more definite information, and based on preliminary data, age data for GBYTF were assumed to be unbiased at all ages. The standard deviation for age reader precision increased with age and was assumed to be 0 at age zero and to increase linearly to 0.8 y by age eight. Recent assessments grouped fish age 6+ into a single plus group to accommodate this and other potential problems.

Based on preliminary modeling, logistic selectivity at age was used for fishery and survey selectivity in all SS2 models for GBYTF. The commercial fishery selectivity pattern was allowed to change in 1994/1995 but not in 1985 because any change in the latter period appeared minor. For simplicity, the DFO winter bottom trawl and NEFSC spring bottom trawl surveys, which occur at about the same time of year, were modeled in SS2 using a single selectivity at age curve, but with different survey catchability parameters.

Survey abundance and age data used in SS2 models for GBYTF and other example stocks included data from the NEFSC spring, NEFSC fall and DFO winter surveys. The VPA model in the last assessment used a recruit index for age 0 GBYTF derived from sea scallop survey dredge data. For SS2, the recruit index was replaced by the total mean catch in numbers per tow from the scallop survey with corresponding length composition data. For simplicity in this analysis, sea scallop survey data were ignored in all SS2 runs.

In most runs for GBYTF, NEFSC spring survey abundance trend data (but not age compositions) for 1974-1981 were omitted because the spring survey used a Yankee #41 bottom trawl, instead of the Yankee #32 used in other years and in the fall survey. In effect, we assumed that survey catchability (but not selectivity) changed and that the routinely applied correction factor from field experiments was imprecise. These assumptions were for convenience in preliminary modeling because the current version of SS2 allows for only one covariate per survey that can be used to model changes over time in change survey catchability. We were interested in using the survey covariate for NEFSC bottom trawls as dummy variables to model potential effects of changes in trawl doors during 1984-1985. Omitting years when the alternate bottom trawl was used facilitated our investigations but there may be little scientific justification for the approach.

Recent stock assessments for GBYTF have been complicated by difficulties in interpreting relatively high survey abundance trends over the last decade (suggesting that stock abundance is increasing) given the scarcity of older fish in fishery and survey age composition data (suggesting that mortality rates are high, recruitment is average and stock abundance is relatively low). These anomalous patterns are the primary current problem in assessing the stock. A model used in the most recent assessment for GBYTF and shown below for comparison hypothesized a change in survey catchability for all three bottom trawl surveys during 1994 when closed area management began, although the mechanistic basis for the hypothesis was not clear. Several other hypotheses were explored with SS2 models developed for the workshop.

The goal in preliminary SS2 modeling for GBYTF during the workshop was to match the trends in survey abundance data during recent years and eliminate residual patterns due to lack of older fish in survey and fishery catch at age. In particular, we wanted to find a hypothesis other than changes in survey catchability during 1994-1995 due to closed area management and use the hypothesis to reinterpret the survey trends and catch at age data.

A number of modeling scenarios were considered for GBYTF but none were successful and only two are presented. The SS2MVPA scenario mimicked the “Major Change” ADAPT VPA model in a recent assessment. Scenario 1 was a typical SS2 application that allowed catchability in the NEFSC spring and fall surveys to change in 1985-1986 assuming, in effect, that the empirical adjustments for differences in fishing power due to the door change already applied were imprecise.

Results

This section presents a few general results for all three example stocks along with more detailed information for Georges Bank yellowtail flounder.

Retrospective patterns

Results for all three example stocks (Figures 1-3) show differences in retrospective patterns among models and model configurations. Based on Mohn's rho statistic (Table 1), retrospective patterns were relatively small in SS2 model runs except for GBKYT where VPA model estimates had much less retrospective pattern (see below). The SS2 models with retrospective problems for GBYTF were would not have been used for management due to severe residual patterns. The residual and retrospective problems are probably related to the same problems and incorrect modeling assumptions.

Retrospective patterns for GBYTF were particularly severe (Table 1 and Figure 3). This result suggests that statistical catch at age and SS2 models in particular may work best using aggregate survey abundance trend data and with survey catch at age data to help link and better estimate the effective survey catchability at age.

Trends

Surprisingly, the various models and model configurations estimated relatively similar spawning stock biomass and fishing mortality rates during most years (Figures 4-6). Differences in retrospective patterns seem more pronounced than differences in estimated trends.

Estimated fishing mortality trends from SS2 were smoother for GB cod than fishing mortality trends using VPA (Figure 4). Smoother trends in SS2 estimates may have been due to sensitivity of estimates from VPA to errors in catch at age for the oldest age groups which are may have fewer age samples. Further research is necessary to test this hypothesis, however.

Georges Bank yellowtail flounder

Residual patterns for the SS2 scenario 1 model illustrate the problems involved in assessing the stock. Residual plots show strong patterns for survey abundance trends (Figure 7). Residual plots for survey and commercial catch at age data are probably less pathological (Figure 8) indicating that the model "chose to believe" the age composition data at the expense of the survey trend data.

After tuning in preliminary runs, the standard deviations of log scale survey index residuals for the three bottom trawl surveys were around 0.65-0.7, compared to mean survey standard deviations of about 0.3-0.4, reflecting the overall lack of fit during recent years. After tuning, the effective sample size of age composition data ranged from 2-3 effective samples per positive survey tow or commercial trip sampled, reflecting reasonable or typical levels of fit to age composition data.

Estimated fishery and survey selectivity patterns seemed reasonable (Figure 9). In particular, the peak of commercial fishery selectivity shifted towards older ages when new management regulations (including larger minimum mesh size regulations) were implemented in 1994 (Figure 9a). In addition, the relatively high survey selectivity for age zero GBYTF in the fall survey reflects their availability to the survey due to increased size after growth since settlement during the spring (Figure 9c).

[[*Note, we have not given up so stay tuned….*]]

Acknowledgements

Ian Stewart and Rick Methot (Northwest Fisheries Science Center, Seattle, WA) made substantial contributions to the workshop and analyses for all three workshop stocks. Paul Nitschke Terceiro (Northeast Fisheries Science Center, Woods Hole, MA) provided SCALE model runs for white hake. Paul Rago, Dvora Hart and Mark Terceiro (Northeast Fisheries Science Center, Woods Hole, MA) and Steve Corriea (Massachusetts Department of Marine Fisheries, New Bedford, MA) made useful suggestions.

Table 1.

Retrospective analysis results for Georges Bank cod, Georges Bank yellowtail flounder and white hake using ADAPT VPA, SS2, and SCALE stock assessment models in various configurations. Retrospecificity is measured for terminal fishing mortality and spawning stock biomass estimates using Mohn's rho statistic (*F* rho or SSB rho) Values nearer zero indicate less retrospective pattern (e.g. -1 is better than 4 or -2). The lowest F and SSB rho values for each stock are shown in bold italic font. All retrospective analyses involved runs with five successive terminal years, in additon to the reference run. "Average *F* rho" and "Average SSB rho" are the overall rho statistics divided by 5 (the number of retrospective runs) and measure the average change in terminal year estimates. Rho statistics for different stocks or calculated using different years may not be comparable.

Stock	Model	Configuration or Scenario	Years	*F* rho	SSB rho	Average *F* rho	Average SSB rho	Notes
GB cod	SS2	Base	1978-2004	***0.51***	-0.80	**0.10**	-0.16	
		Mimic VPA		1.98	-2.53	0.40	-0.51	
	VPA	GARM-2		-4.53	***0.36***	-0.91	***0.07***	Same as GARM-2 but used catch equation, instead of Pope approximation (neglible effects); F was the average for ages 6-8
GB yelowtail	VPA	TRAC	1973-2006	***0.03***	0.63	0.01	***0.13***	Breaks all bottom trawl survey time series at 1994-1995
	SS2	Mimic VPA		-2028.37	4.93	-405.67	0.99	Breaks all bottom trawl survey time series at 1994-1995
		Scenario 1		-10.83	2.27	-2.17	0.45	Breaks NEFSC surveys in 1984-1985 for doors - severe residual patterns
White hake	SS2	Base	1964-2006	***-1.41***	0.82	***-0.28***	0.16	Years with catch at ageonly; change in commercial selectivity at 1994 not used in run with terminal year 1995
		Mimic VPA	1989-2000	-3.32	***0.78***	-0.66	***0.16***	
	SCALE	na	1964-2006	-2.88	0.79	-0.58	0.16	From 1+ biomass, rather than SSB.
	VPA	SARC	1989-2000	3.16	-4.16	0.63	-0.83	Model rejected by SARC-33

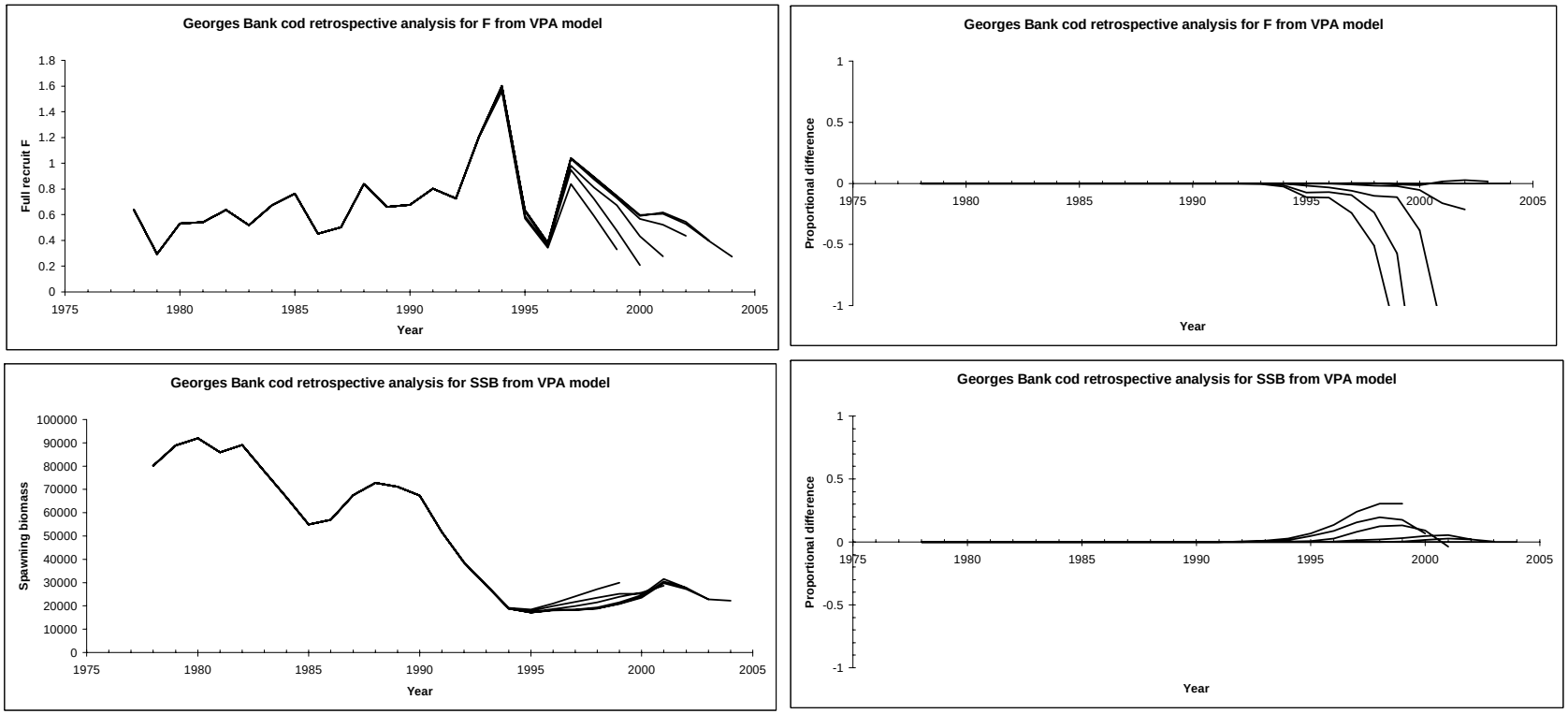


Figure 1a. Retrospective patterns for Georges Bank cod from ADAPT-VPA

.

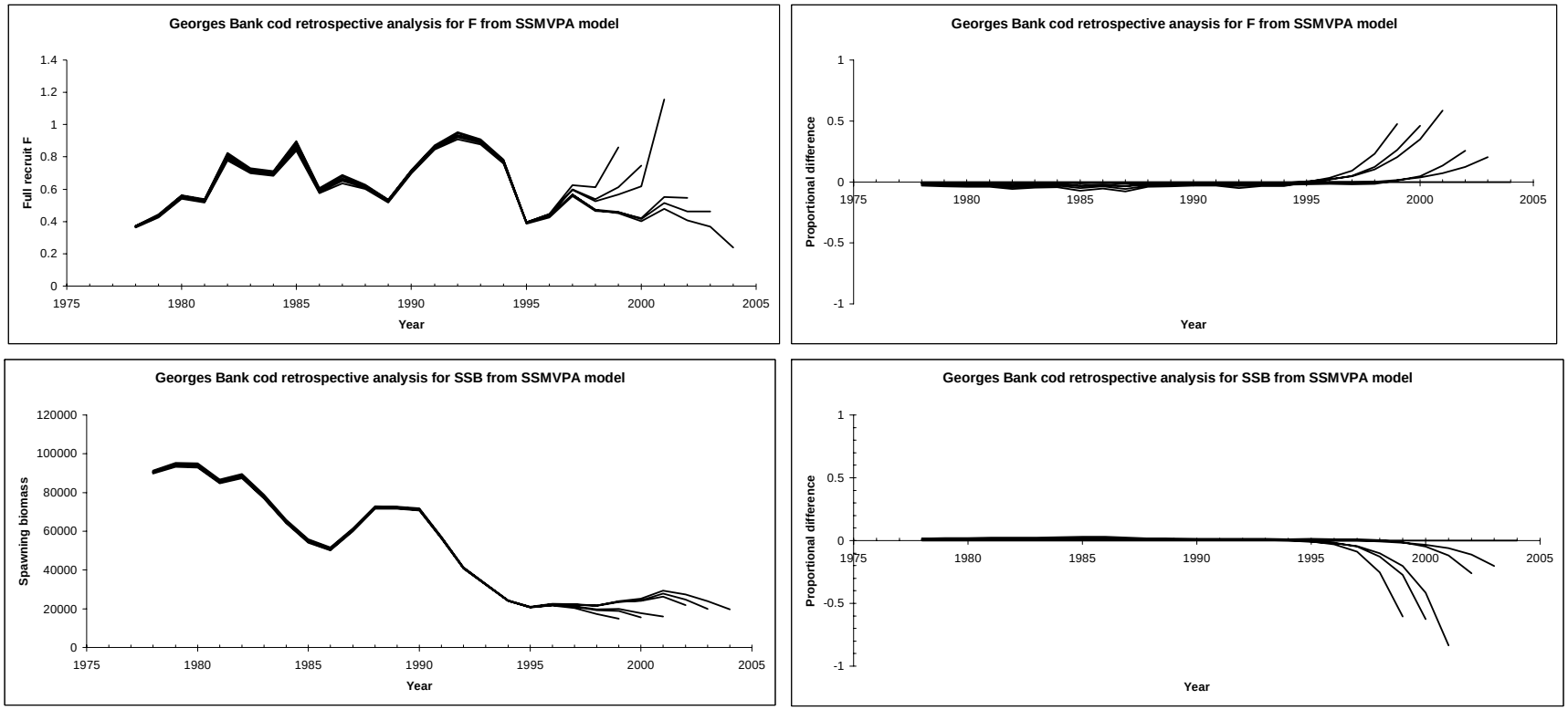


Figure 1b. Retrospective patterns for Georges Bank cod from an SS2 model configured to resemble ADAPT-VPA

.

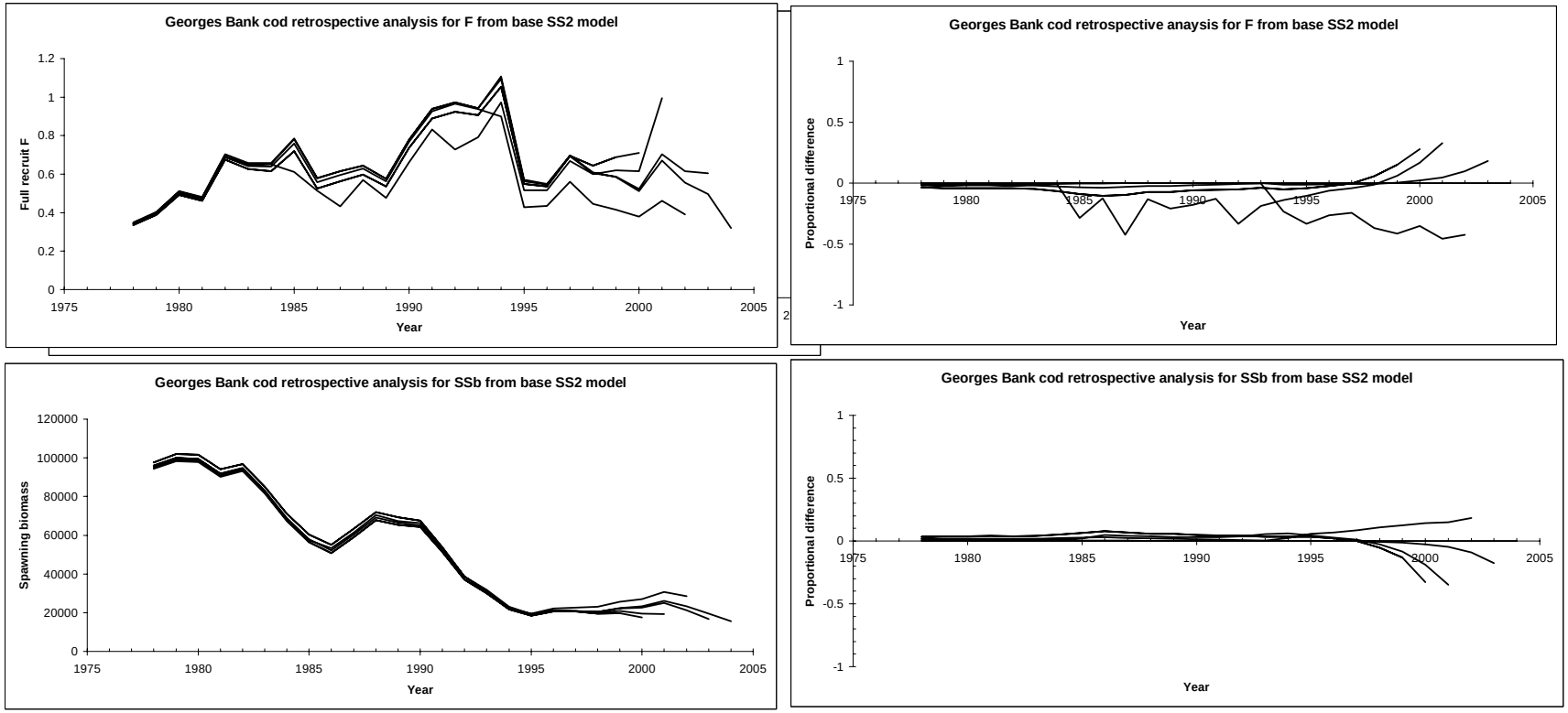


Figure 1c. Retrospective patterns for Georges Bank cod from a “base” SS2 model that was not configured to resemble ADAPT VPA.

.

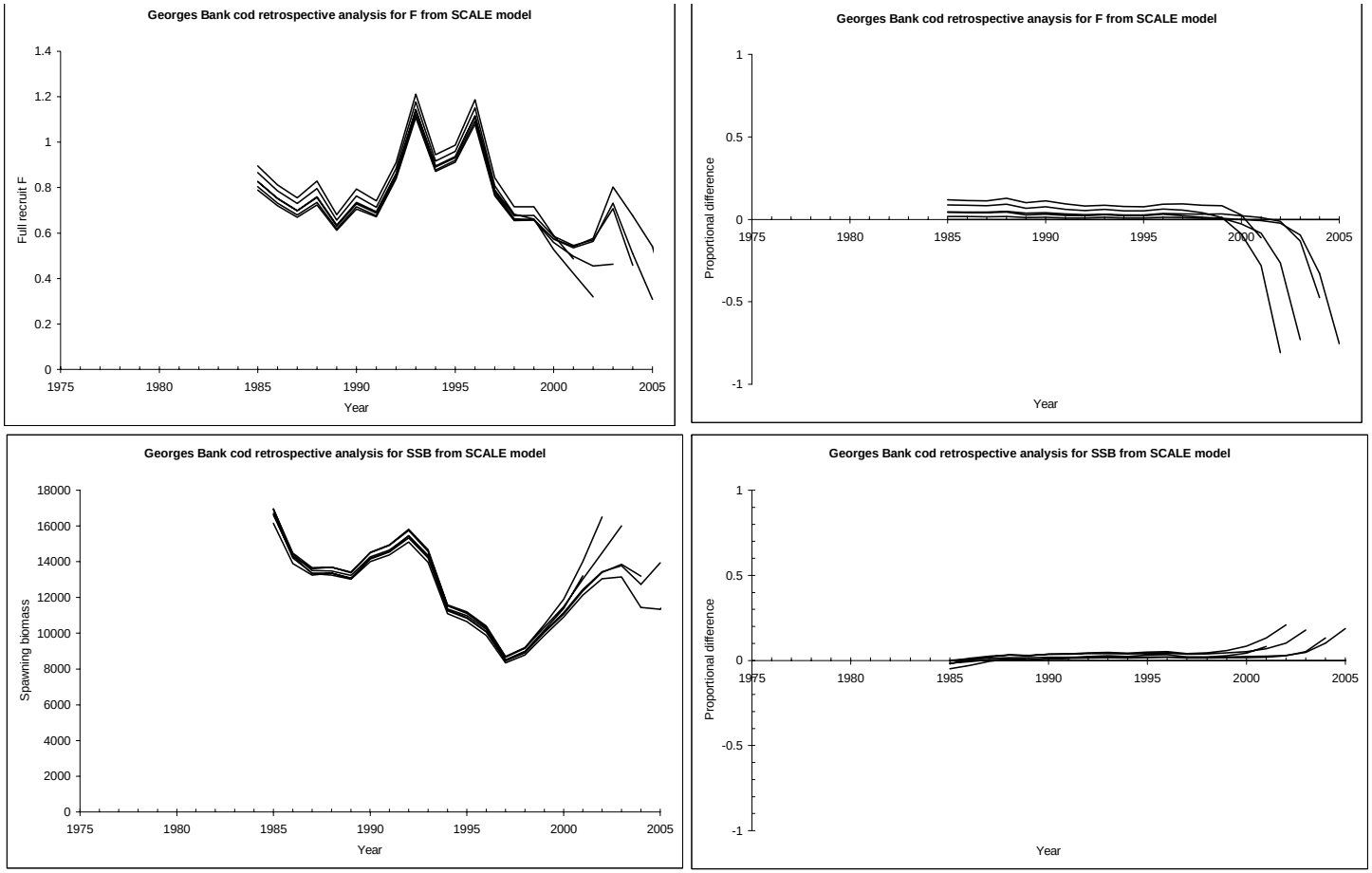


Figure 2a. Retrospective patterns for white hake from the SCALE model.

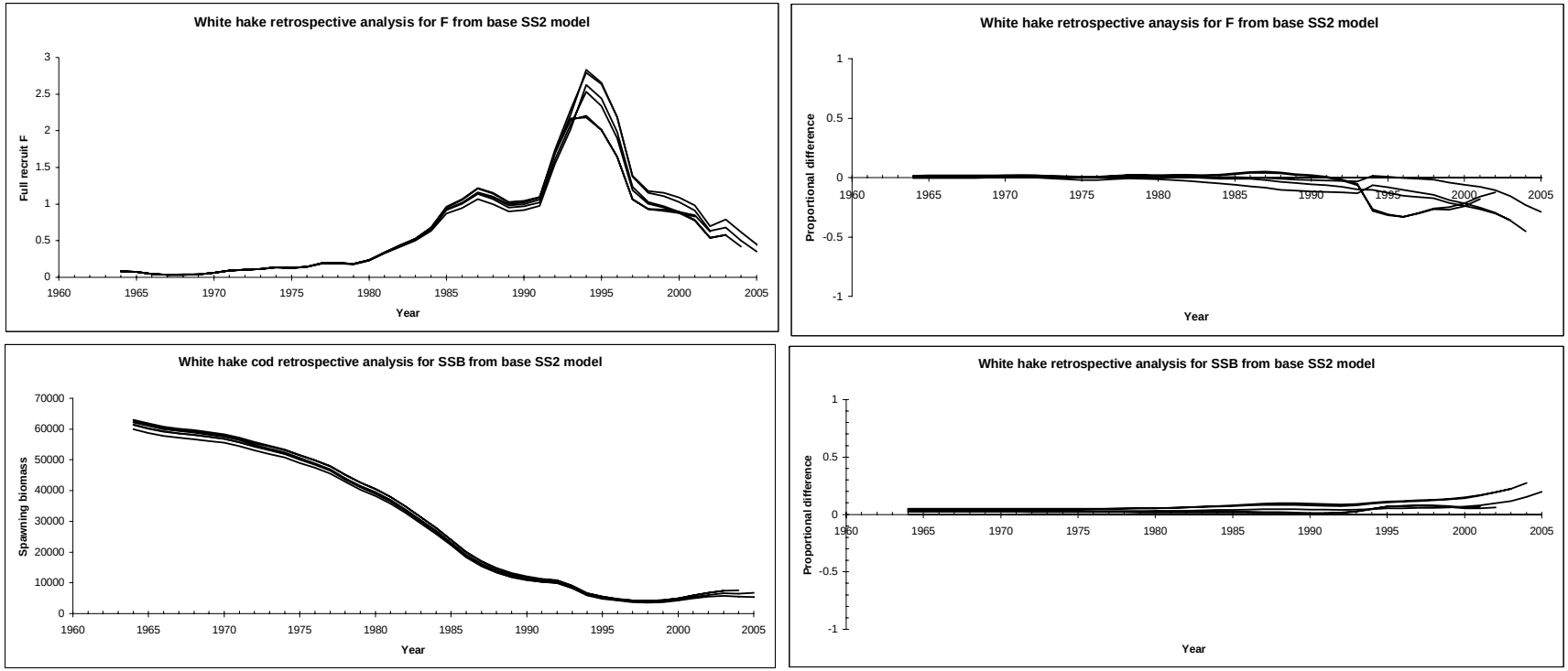


Figure 2b. Retrospective patterns for white hake from a “base” SS2 model that was not configured to resemble VPA.

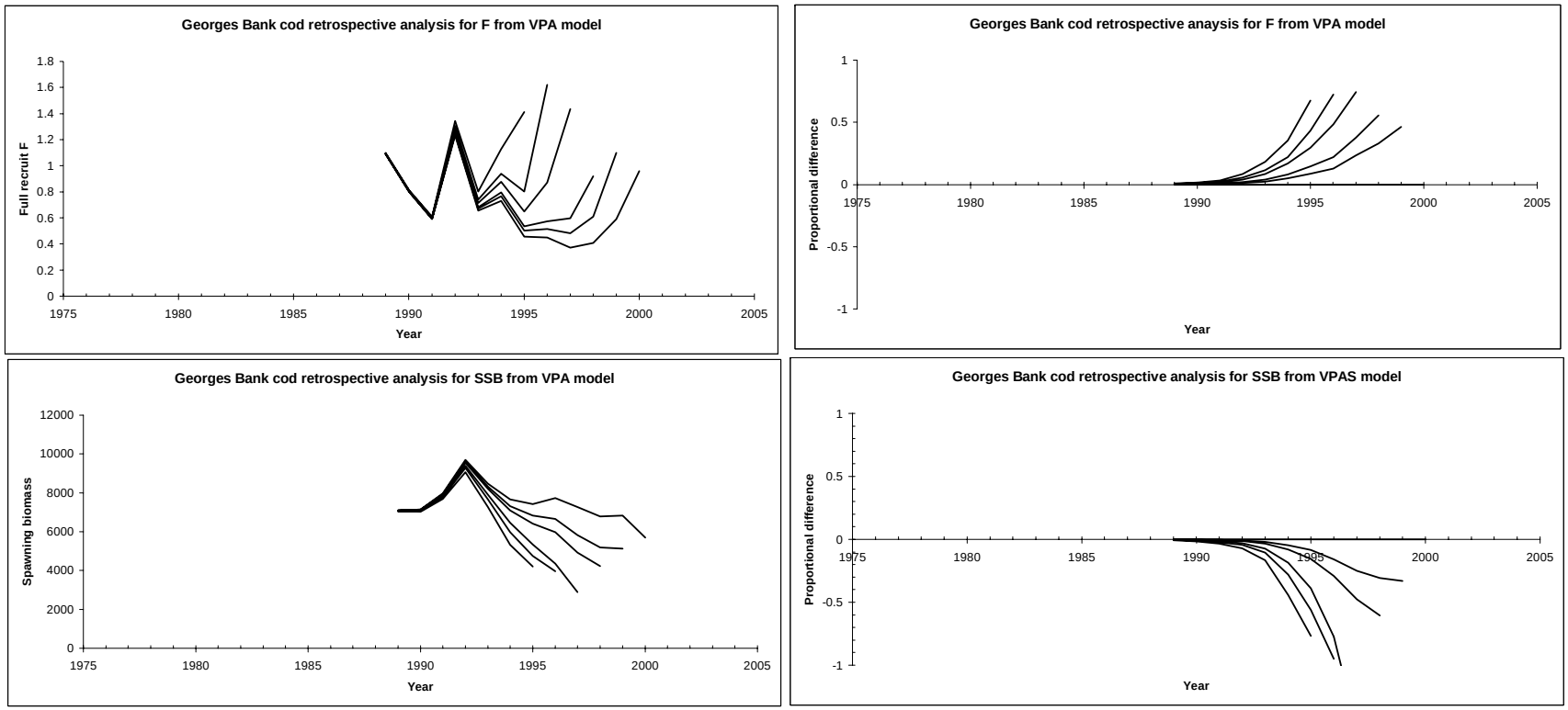


Figure 2c. Retrospective analysis for white hake fro an ADAPT VPA model.

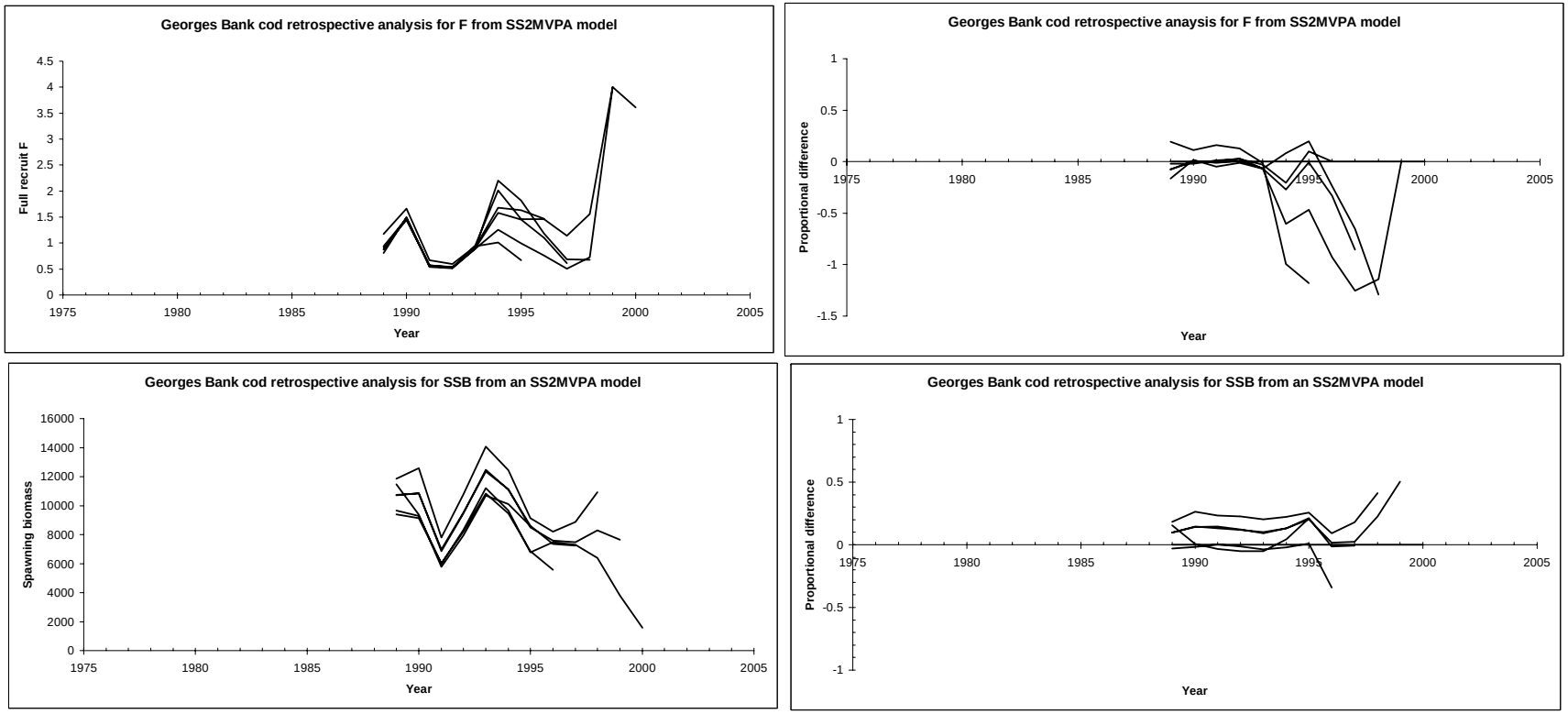


Figure 2d. Retrospective analysis for white hake from a SS2 model configured to like ADAPT-VPA.

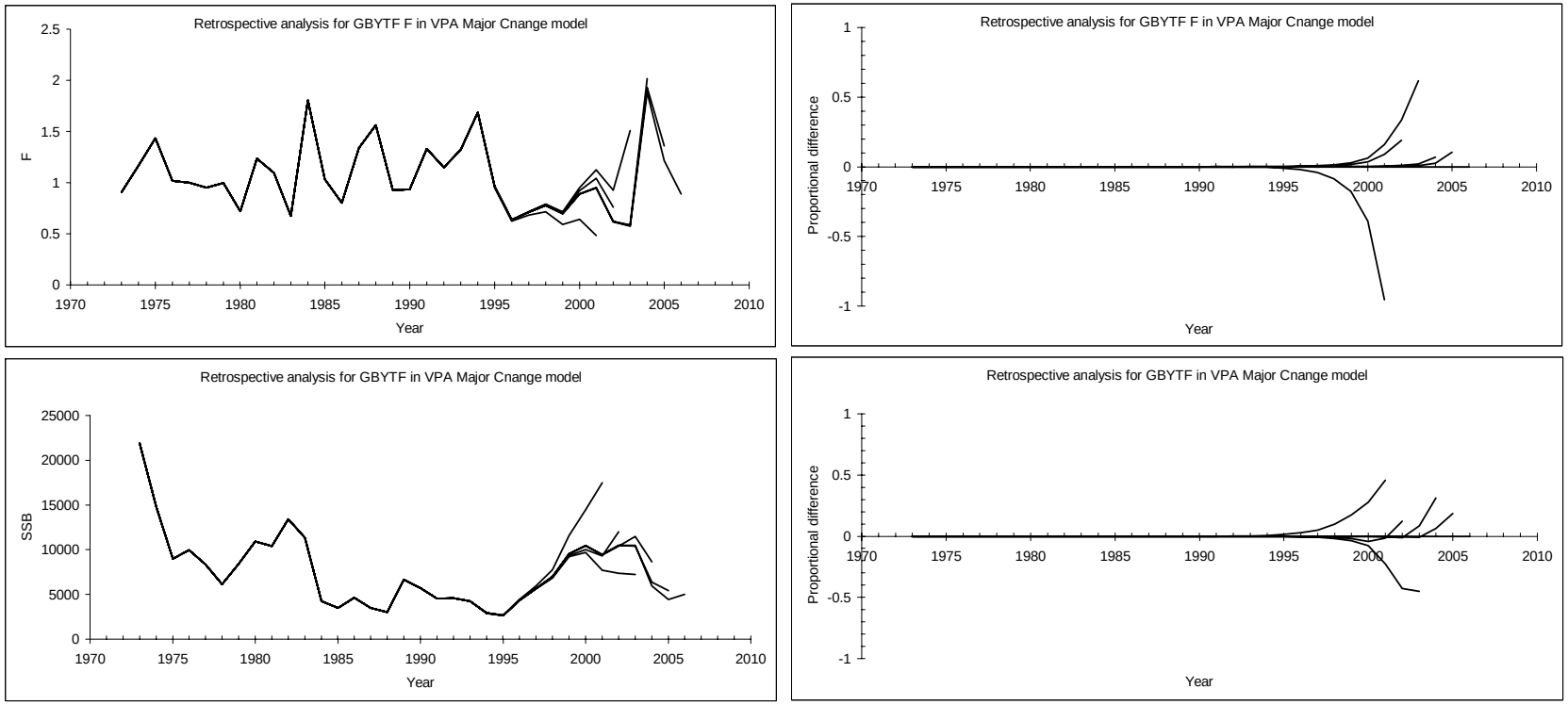


Figure 3a. Retrospective patterns for Georges Bank yellowtail flounder from the Major Change ADAPT-VPA model.

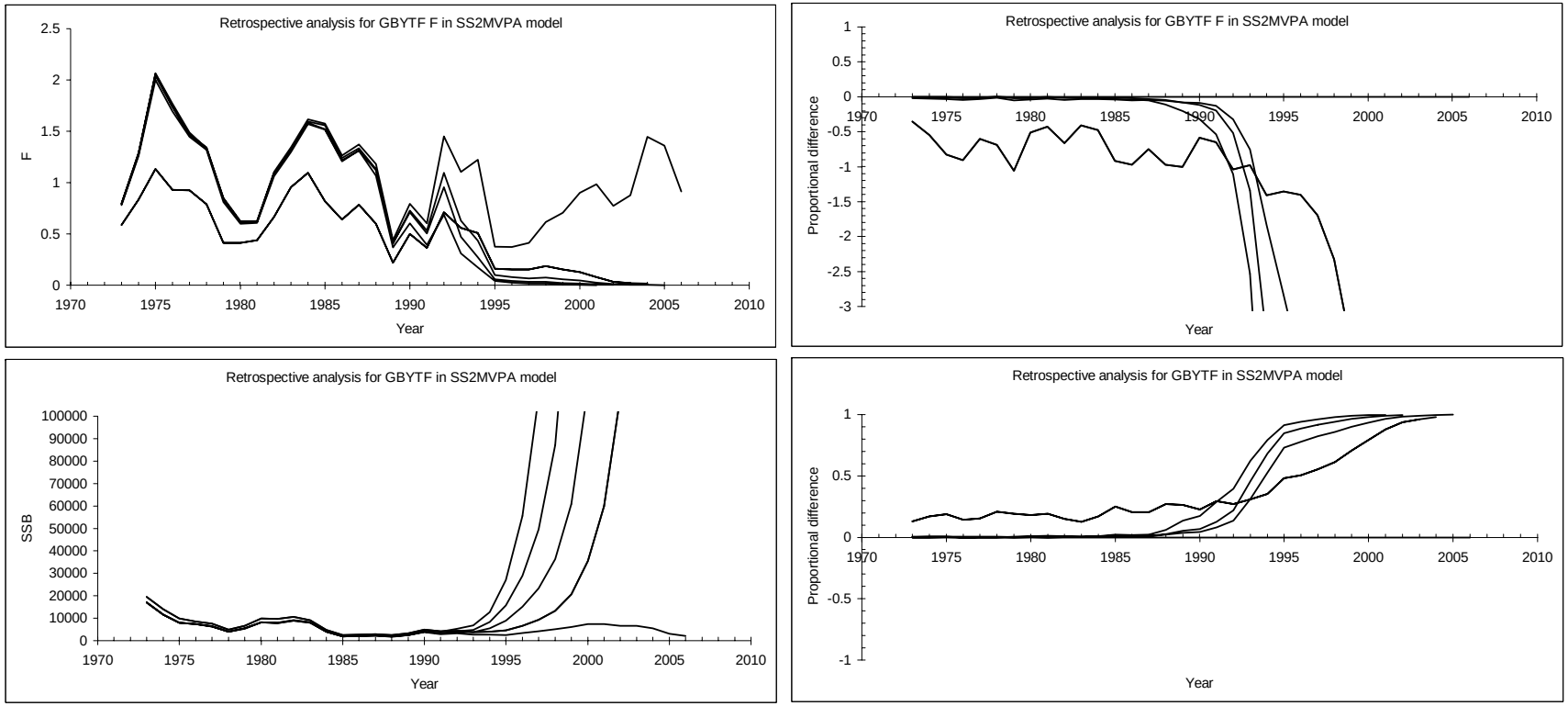


Figure 3b. Retrospective patterns for Georges Bank yellowtail flounder from an SS2 configured to resemble ADAPT-VPA.

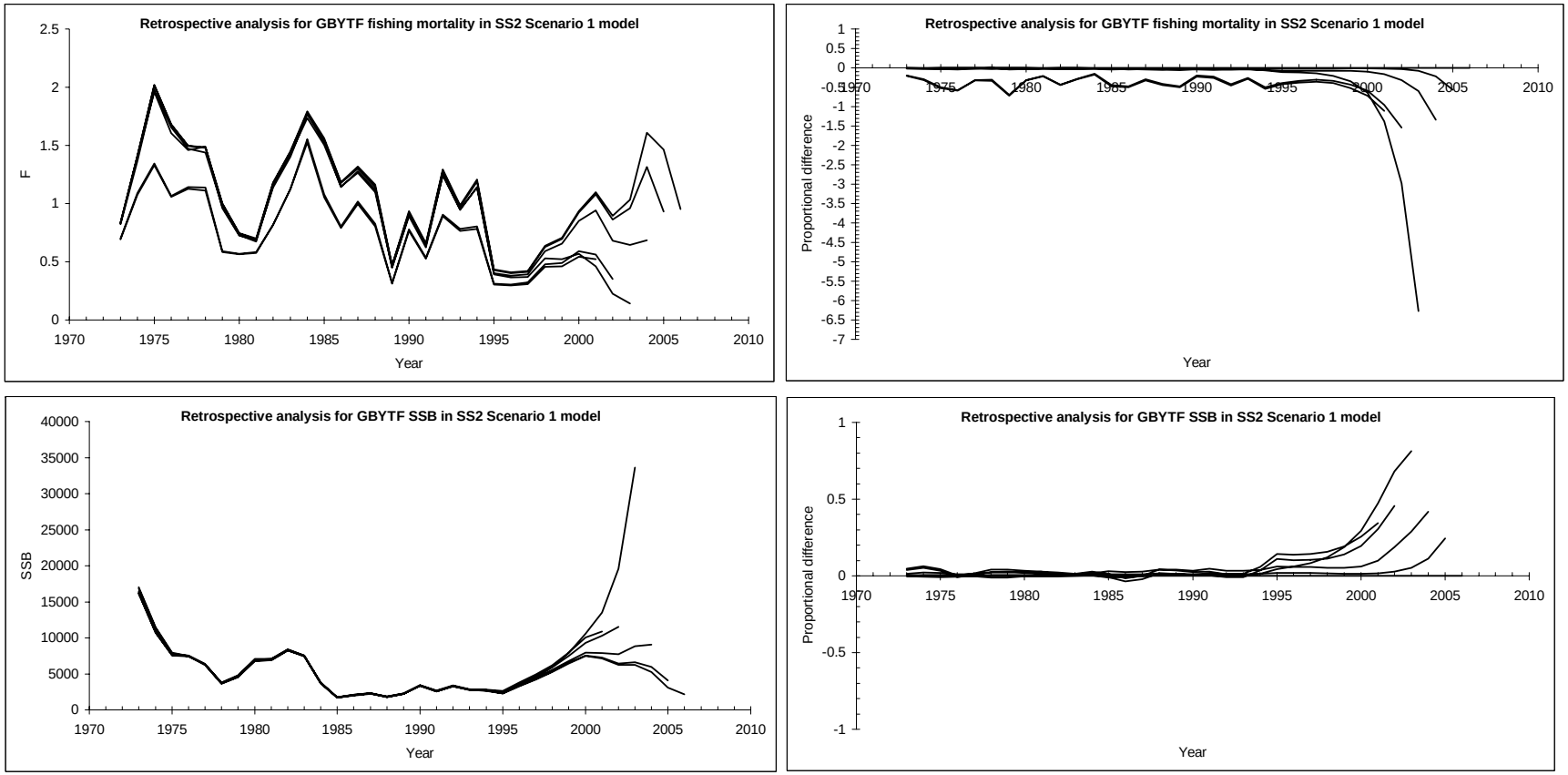


Figure 3c. Retrospective patterns for Georges Bank yellowtail flounder from the SS2 Scenario 1 model.

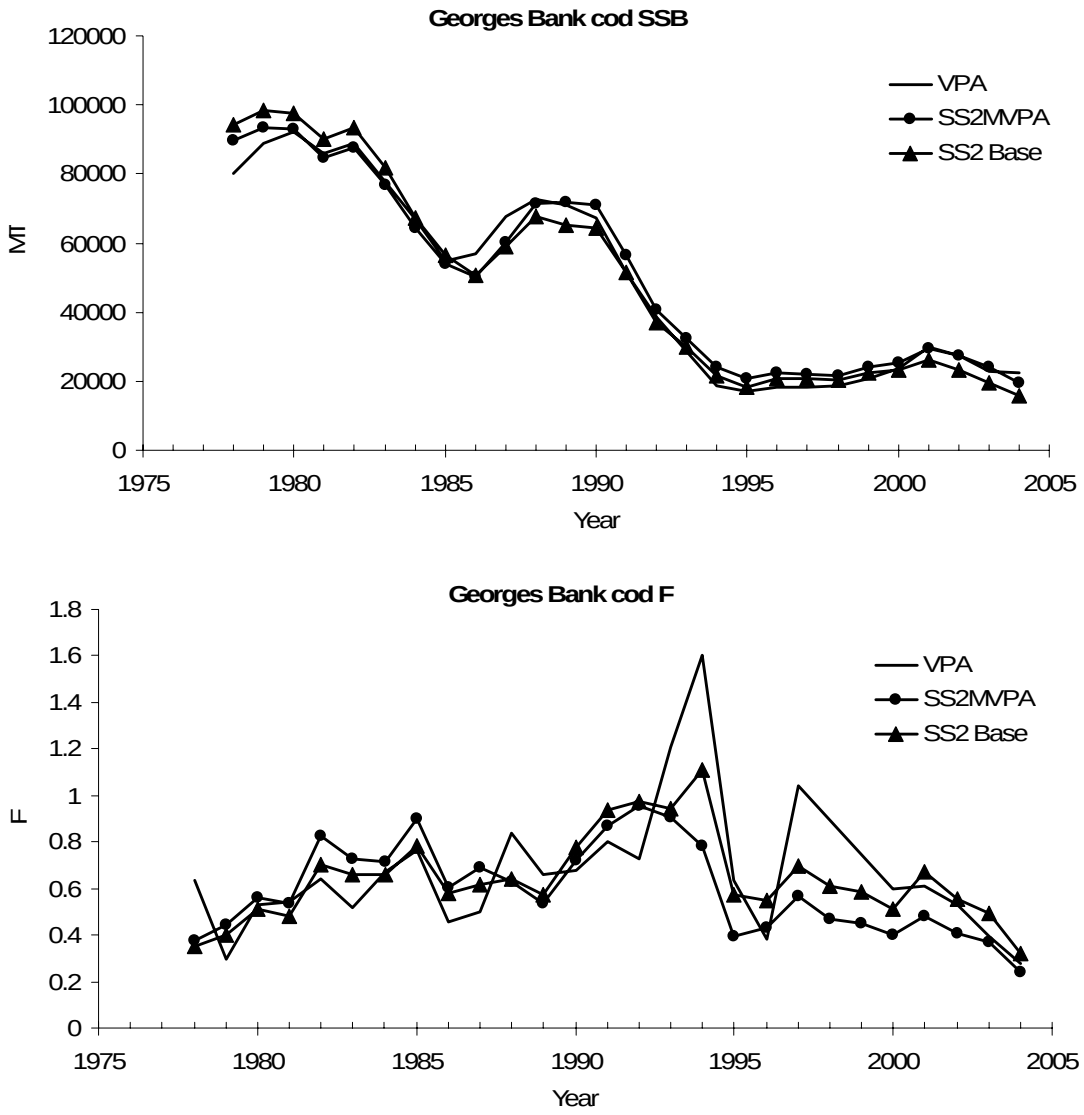
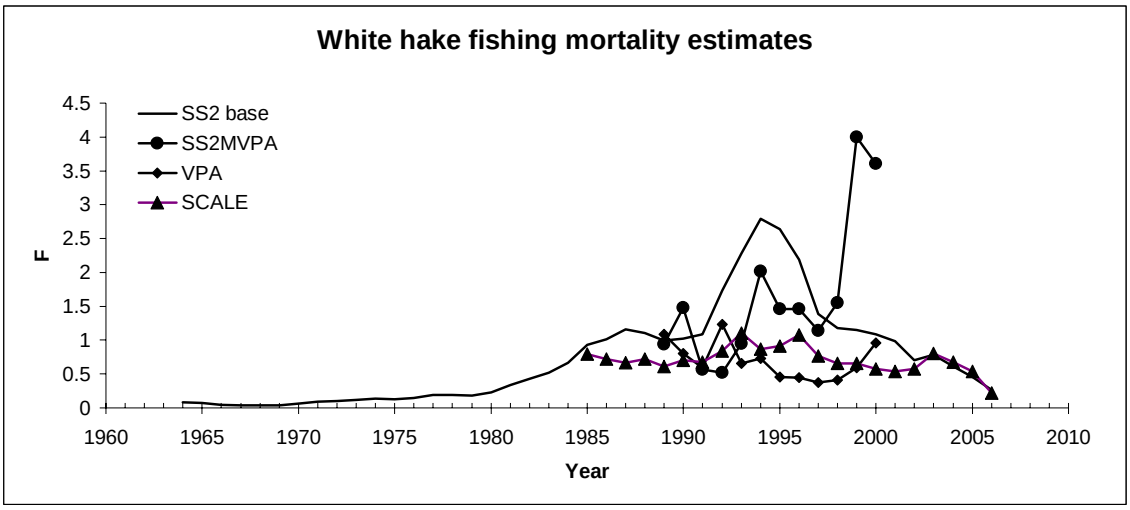


Figure 4. Trends in estimated spawning stock biomass and fishing mortality for Georges Bank cod estimated with various models.



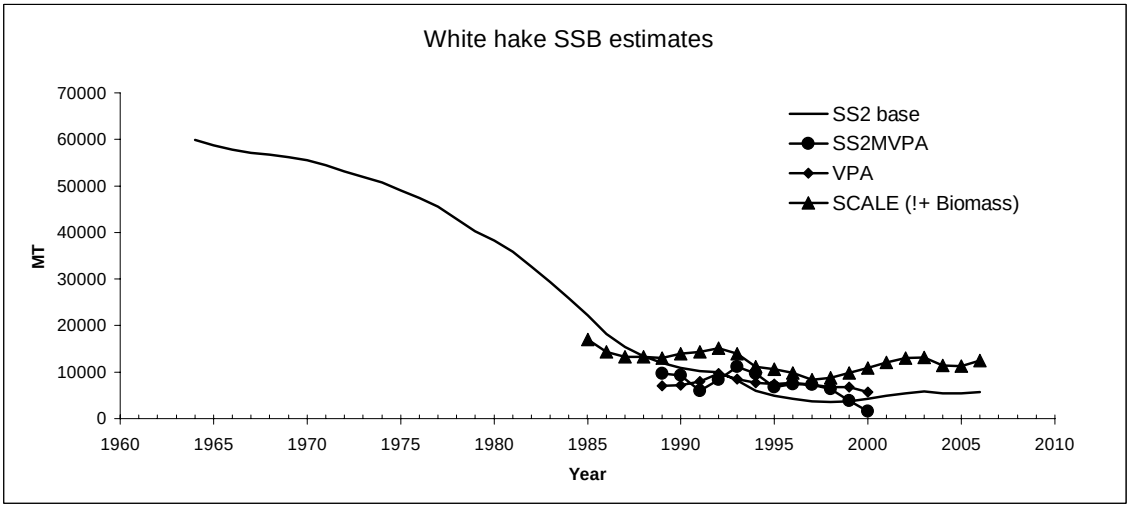


Figure 5. Trends in estimated stock biomass and fishing mortality for while hake estimated with various models. Estimates from the SCALE model are 1+biomass and estimates from other models are spawning stock biomass.

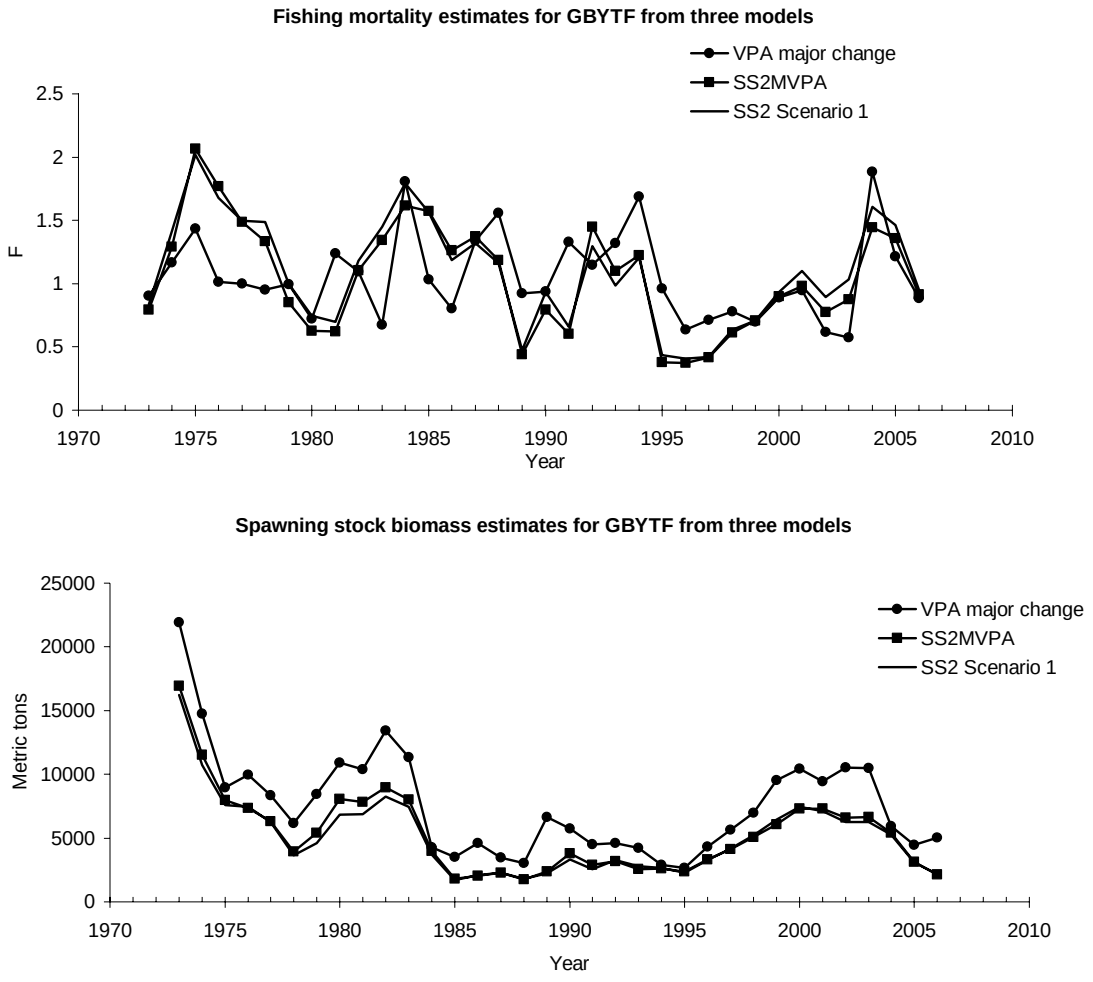


Figure 6. Trends in estimated spawning stock biomass and fishing mortality for Georges Bank yellowtail flounder estimated with various models.

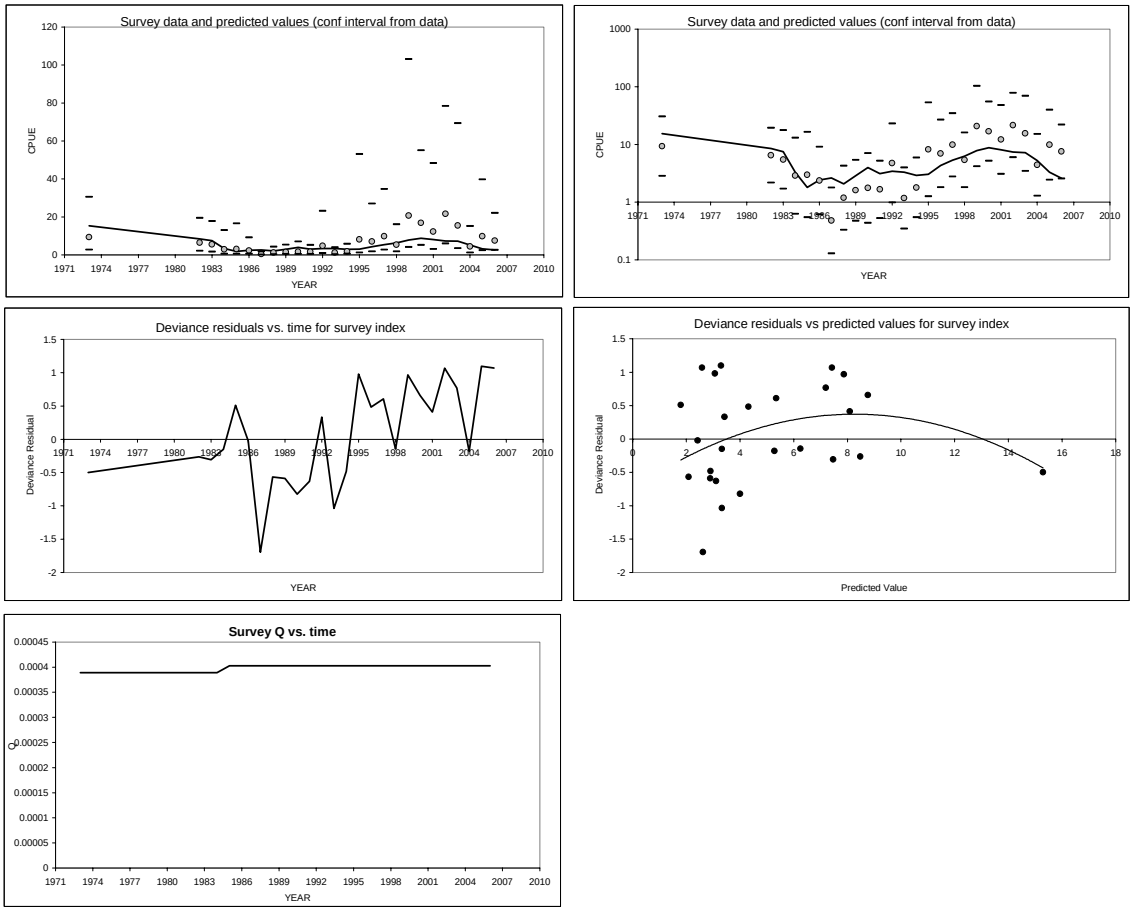


Figure 7a. Diagnostic plots for SS2 scenario 1 model fit to NEFSC spring survey abundance index data for GBYTF.

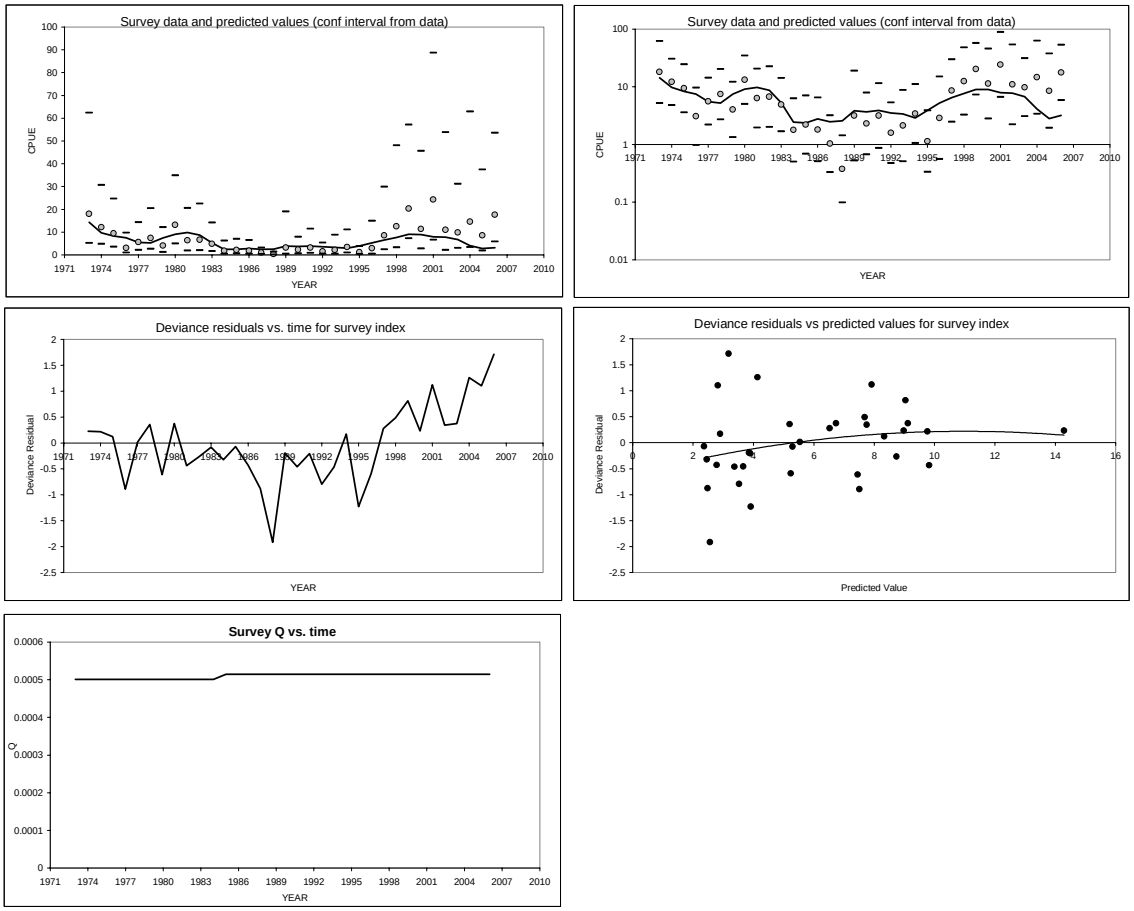


Figure 7b. Diagnostic plots for SS2 scenario 1 model fit to NEFSC fall survey abundance index data for GBYTF.

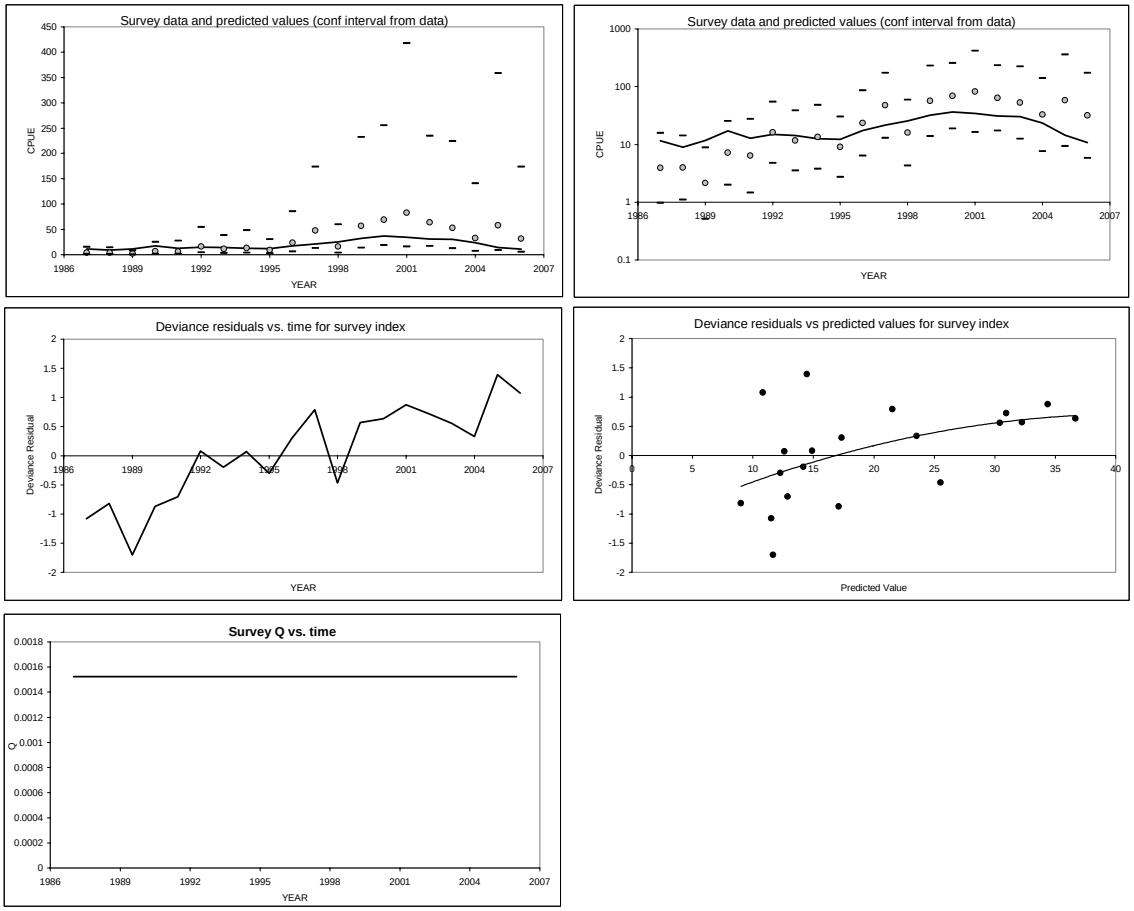


Figure 7d. Diagnostic plots for SS2 scenario 1 model fit to DFO winter survey abundance index data for GBYTF.

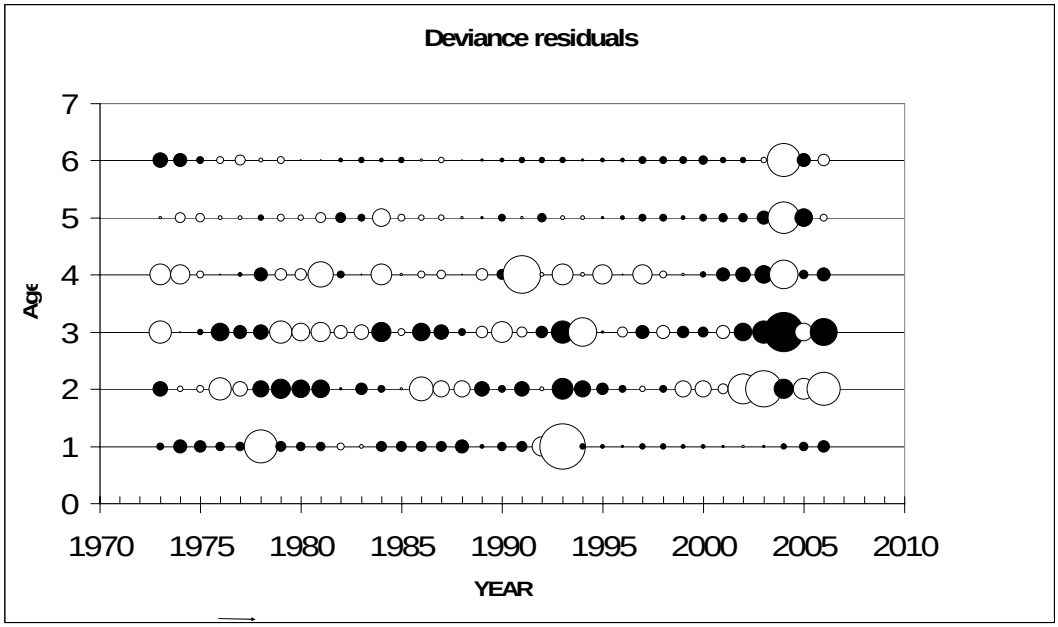


Figure 8a. Deviance residuals for SS2 scenario 1 model fit to commercial age composition data. Positive residuals (observed – precicted > 0) are shown as white bubbles and area of the bubble is proportional to the size of the residual.

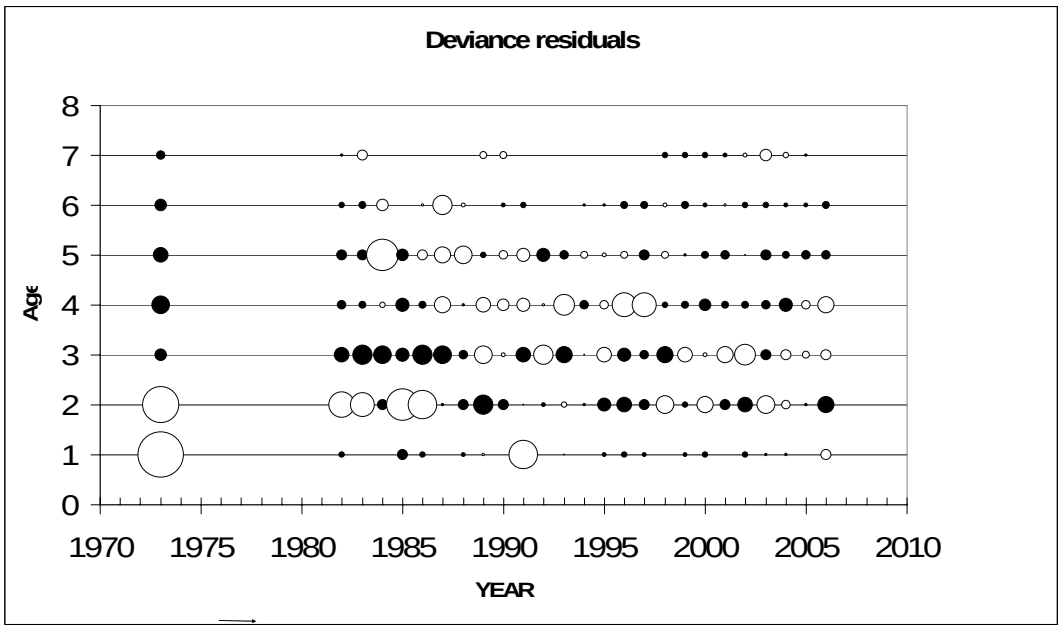


Figure 8b. Deviance residuals for GBYTF SS2 model scenario 1 fit to NEFSC spring survey age composition data. Positive residuals (observed – precicted > 0) are shown as white bubbles and area of the bubble is proportional to the size of the residual.

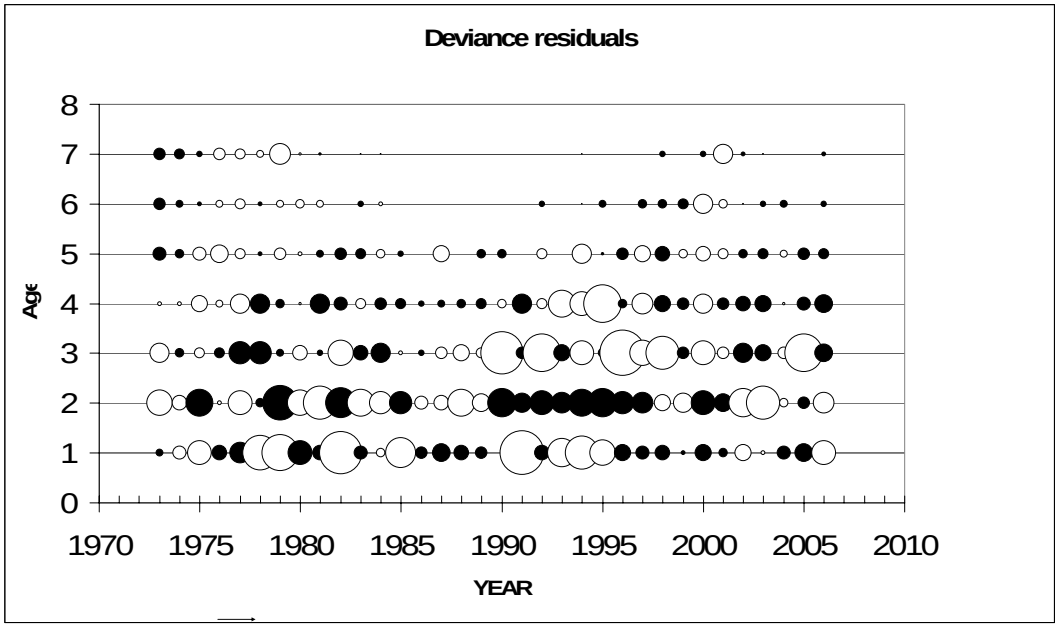


Figure 8d. Deviance residuals for GBYTF SS2 model scenario 1 fit to NEFSC fall survey age composition data. Positive residuals (observed – precicted > 0) are shown as white bubbles and area of the bubble is proportional to the size of the residual.

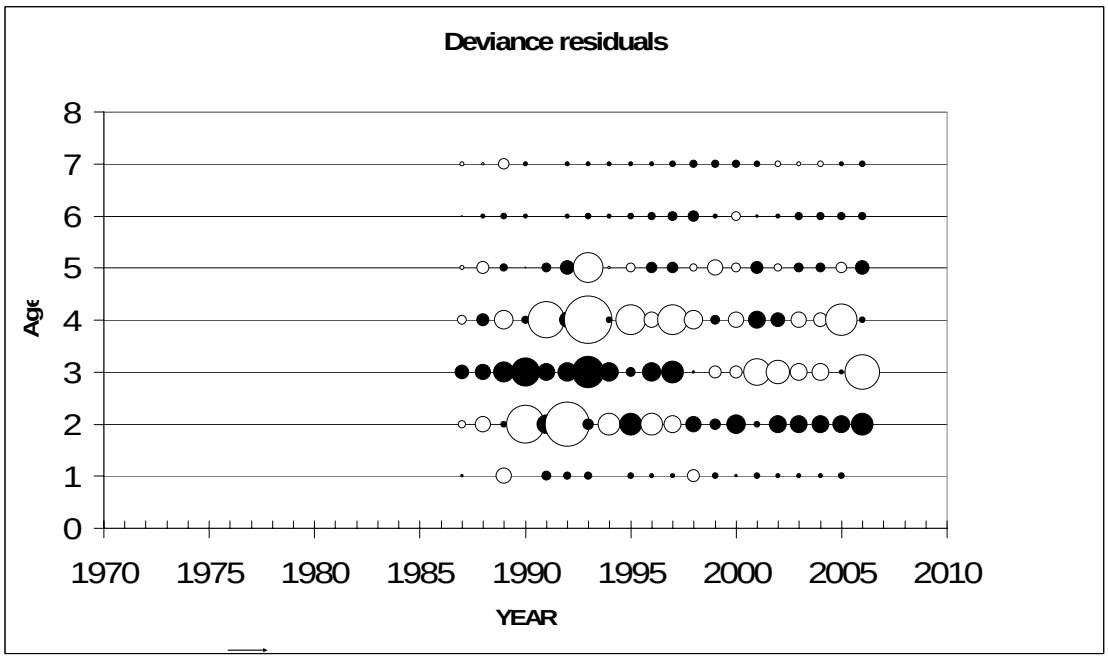


Figure 8e. Deviance residuals for GBYTF SS2 model scenario 1 fit to DFO winter survey age composition data. Positive residuals (observed – precicted > 0) are shown as white bubbles and area of the bubble is proportional to the size of the residual.

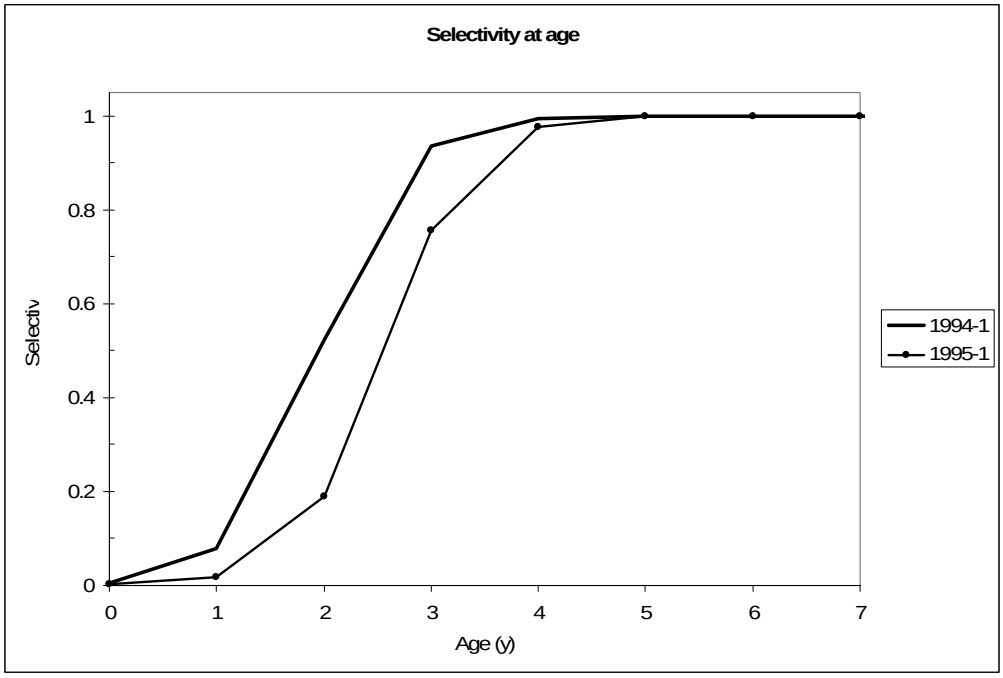


Figure 9a. Estimated commercial fishery selectivity curves for GBYTF from the SS2 scenario 1 model.

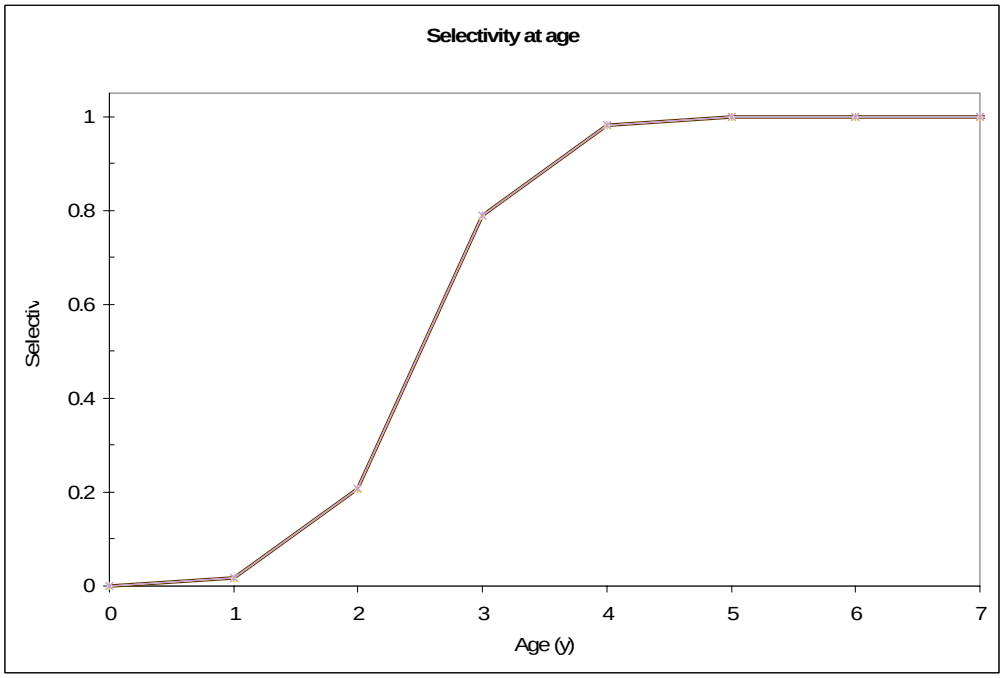


Figure 9b. Estimated fishery selectivity curves for GBYTF in the NEFSC spring and DFO winter surveys from the SS2 scenario 1 model.

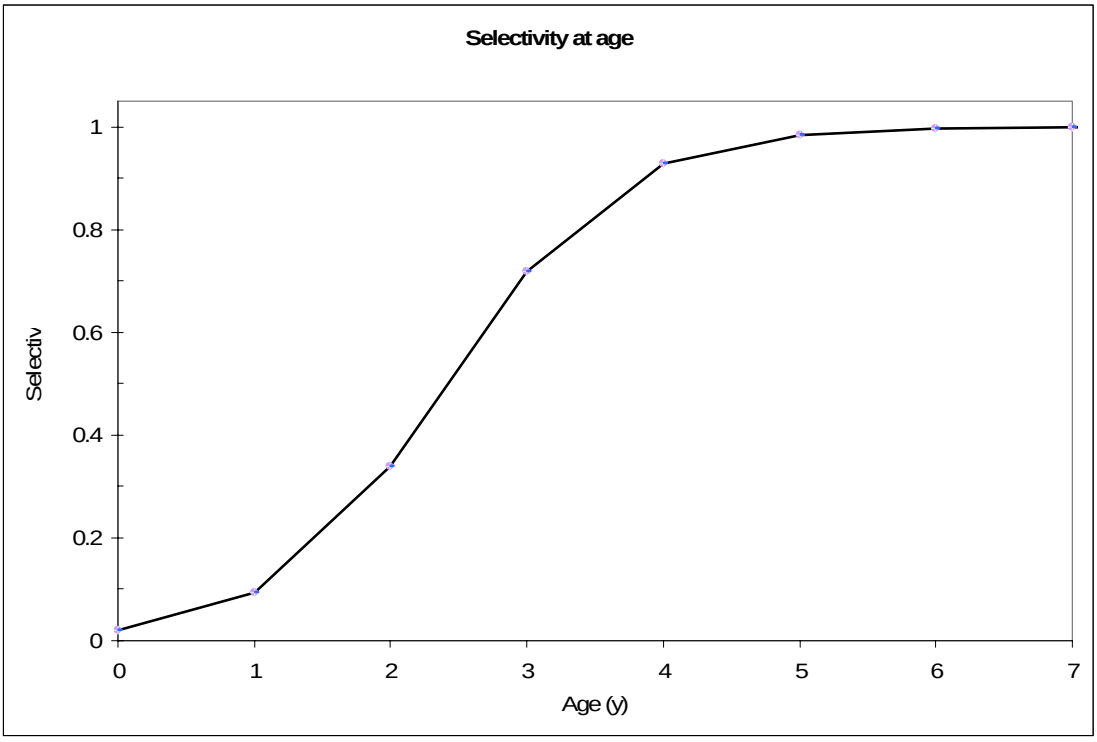


Figure 9c. Estimated fishery selectivity curves for GBYTF in the NEFSC fall and DFO winter surveys from the SS2 scenario 1 model.