C. Gulf of Maine (GOM) WINTER FLOUNDER STOCK ASSESSMENT FOR 2011

[SAW52 Editor's Note: The SARC-52 peer review panel concluded that no ASAP model run provided a suitable basis for management advice. A swept-area biomass method was accepted instead, and it is described in Appendix C1.]

The Southern Demersal Working Group (SDWG) prepared the stock assessment. The SDWG met during April 19-21, April 26-28, and May 3-5, 2011 at the Northeast Fisheries Science Center, Woods Hole, MA, USA.

The following participated in all or part of the meetings:

Name	Affiliation	email
Paul Nitschke	NEFSC	paul.nitschke@noaa.gov
Lisa Hendrickson	NEFSC	lisa.hendrickson@noaa.gov
Jon Hare	NEFSC	jon.hare@noaa.gov
Yvonna Rowinski	NEFSC	yvonna.rowinski@noaa.gov
Emilee Towle	NEFSC	emiliee.towle@noaa.gov
Katherine Sosebee	NEFSC	Katherine.sosebee@noaa.gov
Jay Burnett	Public	
Mark Wuenschel	NEFSC	mark.wuenschel@noaa.gov
Eric Robillard	NEFSC	eric.robillard@noaa.gov
David McElroy	NEFSC	dave.mcelroy@noaa.gov
Kiersten Curti	NEFSC	kiersten.curti@noaa.gov
Michael Palmer	NEFSC	michael.palmer@noaa.gov
Richard McBride	NEFSC	richard.mcbride@noaa.gov
Katie Almeida	REMSA	katie.almeida@noaa.gov
Bonnie Brady	LICFA	greenfluke@optonline.net
Chuck Weimar	Fisherman	star2017@aol.com
Matt Camisa	MADMF	matt.camisa@state.ma.us
Vin Manfredi	MADMF	vincent.manfedi@state.ma.us
Piera Carpi	SMAST	piera.carpi@an.ismar.cnr.it
Sally Sherman	MEDMR	sally.sherman@maine.gov
Linda Barry	NJ Marine Fish.	linda.barry@dep.state.nj.us
Susan Wigley	NEFSC	susan.wigley@noaa.gov
Tom Nies	NEFMC	tnies@nefmc.org
Scott Elzey	MADMF	scott.elzey@state.ma.us
Jeremy King	MADMF	jeremy.king@state.ma.us
Steve Cadrin	SMAST	scadrin@umassd.edu

Yuying Zhang SMAST yzhang2@umassd.edu
Anthony Wood NEFSC anthony.wood@noaa.gov
Dave Martins SMAST dmartins@umassd.edu
Larry Alade NEFSC larry.alade@noaa.gov
Gary Shepherd NEFSC gary.shepherd@noaa.gov
Jess Melgey NEFMC jmelgey@nefmc.org
Jim Weinberg NEFSC james.weinberg@noaa.gov
Paul Rago NEFSC paul.rago@noaa.gov
Lisa Kerr SMAST lkerr@umassd.edu
Maggie Raymond Assoc. Fish. Maine maggie.raymond@comcast.net
Mark Terceiro NEFSC mark.terceiro@noaa.gov
Doug Butterworth MARAM doug.butterworth@uct.ac.za

SAW 52 Terms of Reference

C. Winter flounder (Gulf of Maine Stock)

1. Estimate catch from all sources including landings and discards. Characterize the uncertainty in these sources of data.

2. Present survey data being considered and/or used in the assessment (e.g., regional indices of abundance, recruitment, state and other surveys, age-length data, etc.). Characterize uncertainty in these sources of data.

3. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series (integrating results from TOR-5), and estimate their uncertainty. Include area-swept biomass estimates. Investigate if implied survey gear or catchability estimates are reasonable. Include a historical retrospective analysis to allow a comparison with previous assessment results.

4. Perform a sensitivity analysis which examines the impact of allocation of catch to stock areas on model performance (in TOR-3).

5. Examine the effects of incorporating environmental factors in models of population dynamics (e.g., spring water temperatures in an environmentally-explicit stock recruitment function).

6. State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for BMSY, BTHRESHOLD, and FMSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

7. Evaluate stock status (overfished and overfishing) with respect to the "new" BRPs (from TOR 6), and with respect to the existing BRPs (from a previous accepted peer review) whose values have been updated.

8. Develop and apply analytical approaches and data that can be used for conducting single and multi-year stock projections and for computing candidate ABCs (Acceptable Biological Catch; see Appendix to the TORs) under a set of alternative harvest scenarios. If the stock needs to be rebuilt, take that into account in these projections.

a. Provide numerical short-term projections (3-5 yrs, or through the end of the rebuilding period, as appropriate). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. In carrying out projections, consider a range of assumptions about the most important uncertainties in the assessment (e.g., terminal year abundance, variability in recruitment).

b. Take into consideration uncertainties in the assessment and the species biology to describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming or remaining overfished, and how this could affect the choice of ABC.

c. Develop plausible hypotheses (e.g., mixing among the three stocks) which might explain any conflicting trends in the data and undertake scenario analyses to evaluate the consequences of these alternate hypotheses on ABC determination.

9. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in recent SARC reviewed assessments and review panel reports. Identify new research recommendations.

Executive Summary

1. Estimate catch from all sources including landings and discards. Characterize the uncertainty in these sources of data.

Commercial landings were near 1,000 mt from 1964 to the mid 1970s. Thereafter commercial landings increased to a peaked of 2,793 mt in 1982, and then steadily declined to 350 mt in 1999. Landings have been near 650 mt from 2000 to 2004 and about 300 mt from 2005 to 2009. Landings have declined to a record low of 140 mt in 2010. Recreational landings reached a peak in 1981 with 2,554 mt but declined substantially thereafter. Recreational landings have generally been less than 100 mt since 1994, with exception of 2008 were the landings was estimated at 103 mt. A discard mortality of 15% was assumed for recreational discards. Discards were estimated for the large mesh trawl (1982-2010), gillnet (1986-2010), and northern shrimp fishery (1982-2010). A discard mortality of 50% was assumed for commercial fishery. In general the total discards are a small percentage (time series average 11%) of the total catch. There has been a substantial decline in the total catch compared to the early 1980s (recent catch is roughly 5% of the 1980s catch).

2. Present survey data being considered and/or used in the assessment (e.g., regional indices of abundance, recruitment, state and other surveys, age-length data, etc.). Characterize uncertainty in these sources of data.

The spring and fall NEFSC, Massachusetts DMF (MDMF) and the Maine New Hampshire (MENH) surveys were used in the Gulf of Maine winter flounder assessment. In general the survey indices are relative flat over the time series in comparison to the catch trends. All of the indices generally show a slight decrease in the population in the late 1980s from a high in the early 1980s with low abundance remaining through the early 1990s. All of the indices show signs of increase abundance starting in 1998 and 1999. Since 2001 all indices indicate some decrease in abundance. However there have been recent increases in the indices at age for the older fish. Length base conversions were use in 2009 and 2010 when the new survey vessel was used in the NEFSC survey.

The SARC accepted GOM winter flounder assessment is based on an empirical swept-area model utilizing data from the 2010 NEFSC fall survey, the MADMF fall survey, and the Maine-New Hampshire fall inshore surveys. Using an efficiency value of 0.6 the estimated stock biomass in 2010 of fish greater than 30 cm was 6,341 mt (80% CI 4,230 - 8,800 mt) (Appendix C1).

3. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series (integrating results from TOR-5), and estimate their uncertainty. Include area-swept biomass estimates. Investigate if implied survey gear or catchability estimates are reasonable. Include a historical retrospective analysis to allow a comparison with previous assessment results.

The base and split VPA were updated from the GARM III assessment. The SDWG changed the *assumed natural mortality from 0.2 to 0.3 in this assessment. Diagnostics still imply major sources for concern surrounding the VPA model formulation for GOM winter flounder. The SDWG developed a new assessment in ASAP (Age Structured Assessment Program) which provides more flexibility in the weighting of data sources. The population models have difficulty with the conflicting data trends within the assessment, specifically the large decrease in the catch over the time series with very little change in the indices or age structure in both the catch and surveys. The scaling of the population estimates was sensitive to the weight imposed on the catch at age compositions. The ASAP model allowed errors in the fit to the catch at age and improved fit to the survey indices without the split. However this resulted in a lack of fit to the plus group in the catch at age composition. The combined survey 30+ biomass area swept estimate was used to inform the optimal weighting for the preferred model formulation. The resulting final SDWG model weighting formulation considered both the tradeoff between retrospective bias and feasible biomass estimates at the end of the time series. The within model uncertainty did not capture the uncertainty in this assessment considering how sensitive the results were to the model formulation and weighting.* ***The SARC concluded that the ASAP assessment model was too unreliable to be a basis for management.***

The accepted assessment of GOM winter flounder stock is based on an empirical swept-area model utilizing data from the 2010 NEFSC fall survey, the MADMF fall survey, and the Maine-New Hampshire fall inshore surveys. Using an efficiency value of 0.6 the estimated stock biomass in 2010 of fish greater than 30 cm was 6,341 mt (80% CI 4,230 - 8,800 mt). Exploitation rate in 2010 was estimated at 0.03 (80% CI 0.02 - 0.05) based on the ratio of 2010 catch (195 mt) to survey based swept area estimate of biomass for winter flounder exceeding 30 cm in length (6,341 mt). The biomass estimate for 2010 is 16% lower than that for 2009 using the same survey methods but this difference is not statistically significant (Appendix C1).

4. Perform a sensitivity analysis which examines the impact of allocation of catch to stock areas on model performance (in TOR-3).

The SDWG interpretation of TOR4 is that the variance of the commercial landings due to the 1995 and later area-allocation scheme should be used as the basis for the magnitude of landings that might be lost or gained from the stock-specific assessments, and then perform an exercise to run the assessment model with those potential biases and report the results. Additional work was done to estimate the error in the commercial landings due to misreporting of commercial landings to statistical area at allocation level A. Given the magnitude of these errors, the SDWG elected to run the final GOM winter flounder ASAP model, with an additional 5% PSE in commercial landings added to the estimated PSE over the 1995-2010 time series.

The commercial landings have a calculated Proportional Standard Error (PSE; due to the commercial landings area-allocation procedure; available for 1995 and later years, with the mean of those years substituted for 1982-1994) ranging from 5.3% to about 6.5%; the commercial discard (trawl and gillnet) PSEs range from 16-177% (available for 1994-2010, mean of those years substituted for 1982-1993); and the recreational landings PSEs range *from 17-50%. Because the PSEs for the commercial landings are low, and the commercial landings account for about two-thirds of the total catch, the total catch weighted-average annual PSEs range from 7-30%, and averages 11.7% (unweighted) for the 1981-2010 time series.*

The catch in the final assessment model was increased and decreased by the annually varying PSE and models were re-run to provide an additional measure of uncertainty of assessment estimates. For the final ASAP multi model, the fishing mortality estimate in 2010 did not change greatly (0.01 to 0.034). The 2010 SSB range was 4,700 to 6,900 mt, was similar to the MCMC estimate of uncertainty. However the assessment modeling was not accepted by the SARC as a basis for management.

5. Examine the effects of incorporating environmental factors in models of population dynamics (e.g., spring water temperatures in an environmentally-explicit stock recruitment function).

To develop environmentally-explicit stock recruitment relationships, three specific types of data are required: spawning stock biomass, recruitment, and environmental data. Spawning stock biomass and recruitment data from the final 2011 SAW 52 assessment models were used in the analysis. For the GOM stock, recruitment (lagged by 1 year) and spawning stock biomass pairs were used from the ASAP multi model. Two general types of temperature data were used: air temperatures and coastal water temperature. In addition to temperature, four large-scale forcing indices were included in the analyses. The North Atlantic Oscillation (NAO) is the dominant mode of winter climate variability in the North Atlantic region and has been related to numerous physical and biological variables across the North Atlantic (Ottersen et al. 2001, Visbeck et al. 2003). The Atlantic Multidecadal Oscillation (AMO) is a natural mode of climate variability and represents a detrended multi-decadal pattern of sea surface temperatures across the North Atlantic with a period of 60-80 years (Kerr 2005). Finally, the Gulf Stream index is a measure of the northern extent of the Gulf Stream south of the northeast U.S. shelf ecosystem. The Gulf Stream position is related to the larger basin-wide circulation, which in turn is related to NAO and AMO. Two Gulf Stream indices are used here (Joyce and Zhang 2010, Taylor and Stephens 1998).

For the Gulf of Maine stock, increased winter air temperatures are related to lower recruitment, but the strength of this environmental forcing is less than for the Southern New England stock. This result makes sense in the context of the distribution of winter flounder; the southern stock is most affected by warmer temperatures.

One use of the environmentally-explicit models is to develop short-term and long-term forecasting models. Based on this work, there is no trend in winter temperature over the past 30 years and thus short-term forecasts can be developed using the environmentally-explicit models assuming winter temperatures to be at their mean state. It may also be useful to develop short-term forecasts under warm temperatures and short temperatures to provide managers with a tangible understanding of the effect of temperature on the stocks. The environmentally-explicit models could also be used to develop longer-term forecasts following the approach of Hare et al. (2010). These forecasts would provide an assessment of the *sustainability of the winter flounder fishery on the 30-100 time scale. Work is underway within the SDWG to incorporate environmentally-explicit stock-recruitment models into the NFT standard software used to fit stock-recruitment models and to perform projections of stock and fishery catch. However, this work has not been developed sufficiently to be made available for peer-review at this time (see new Research Recommendation 10).*

6. State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for BMSY, BTHRESHOLD, and FMSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

The 2008 GARM III assessment was not accepted and the overfished and overfishing status of the GOM winter flounder stock is currently unknown. For the new 2011 assessment, the SDWG split VPA estimated higher percent maximum spawning potential (MSP) proxies relative to the ASAP model because the VPA estimated selectivity was shifted to older fish. The SDWG ASAP multi run estimated a F40% FMSY proxy at 0.34 using the 2006-2010 average mean weights and selectivity as input to the YPR analysis. The F40% SSBmsy was estimated from a long term projection (100 years) using the CDF of recruitment from the entire model time series (1982-2010) and the estimated YPR F40%. The SSBmsy using the FMSY = F40% proxy was estimated at 3,287 mt with a SSBmsy threshold estimate of 1,644 mt and MSY equal to 1,080 mt for the ASAP multi run. The Beverton Holt stock recruitment Fmsy using the Pleuronectids steepness prior from Myers et al. (1999; 0.8 mean and CV = 0.09) was estimated at 0.57. The stock recruit SSBmsy was estimated at 2,167, SSBmsy threshold = 1,084 mt, and MSY = 1,152. The MSY estimates did not vary greatly with SSBmsy from the mcmc in the stock recruitment analysis. The SDWG expressed concern with the stock recruitment estimate of SSBmsy being estimated in the lower end of the range of past SSB observations. However SARC 52 did not accept the SDWG model and the overfished status remains as unknown since biomass based reference points could not be estimated.

The SARC accepted a proxy value of the overfishing threshold which was derived from a length-based yield per recruit analysis that assumes all fish above 30 cm are fully recruited to the fishery and that natural mortality is 0.3. Using $F_{40\%}$ (0.31) as a proxy for Fmsy, the threshold exploitation rate is 0.23 and 75%$F_{40\%}$ exploitation was 0.17 with M=0.3. The reference points were converted to exploitation rates to be consistent with the swept area biomass approach.

7. Evaluate stock status (overfished and overfishing) with respect to the "new" BRPs (from TOR 6), and with respect to the existing BRPs (from a previous accepted peer review) whose values have been updated.

The 2008 GARM III assessment was not accepted and the overfished and overfishing status of GOM winter flounder stock is currently unknown. In the new 2011 assessment, stock status evaluation was consistent regardless of the model formulation (VPA and ASAP). Both the split

VPA model and the SDWG preferred ASAP multi model indicate that the stock is not overfished and overfishing is not occurring. However spawning stock biomass relative to the SSBmsy varied widely between the VPA and preferred ASAP multi model. SSB in 2010 to SSBmsy ratios varied from the stock recruit Split VPA estimate of 0.52 to the stock recruit estimate of 3.09 from the ASAP multi with no prior on steepness. All models show that fishing mortality in 2010 were well below their respective Fmsy reference points. Fishing mortality in 2010 to Fmsy ratios varied from the stock recruitment split VPA ratio estimate of 0.47 to the stock recruitment estimated ratio of 0.05 from the ASAP multi run with no prior on steepness. The SDWG ASAP multi run using the Fmsy = F40% proxy estimated the SSB2010/SSBmsy ratio at 1.77 and the F2010/F40 at 0.09. The stock recruitment priors did lower the estimated steepness which lowered the SSB2010/SSBmsy ratio to 2.74 and increased the F2010/Fmsy ratio to 0.06.

All GOM winter flounder models have diagnostic issues due to the conflicting signals in the data. The SDWG preferred the ASAP multi model as the best fit to all data sources including considerations for reasonable estimates of biomass in 2009 and 2010 in comparisons to the survey area swept biomass estimates. However the SDWG questioned the feasibility of the estimated SSB relative to the SSBmsy reference points for both the F40% proxy and the stock recruit estimates (1.77 to 2.68). In general the trends and biomass estimated by the model seem appropriate. Surveys and anecdotal feedback from fishermen suggest a shift in the population to deeper water which can help explain the lack of catch in the recreational fishery. However questions remain with the lack of higher catches as the stock rebuilds during the late 1990s and early 2000s when effort in the groundfish fishery was high. In addition, there is little evidence of a change in the size structure or stock range expansion to waters off the coast of Maine which traditionally had higher catches. Considerable uncertainty remains with regards to the comparison of the 2010 SSB relative to the SSBmsy biological reference points. The SARC concluded that the population models are too uncertain as a based from stock status determination.

The overfished status remains as unknown since an analytical model was not accepted and a biomass reference point could not be estimated. *The SARC concluded that in 2010 overfishing was not occurring for the stock. A proxy value of the overfishing threshold was derived from a length-based yield per recruit analysis that assumes all fish above 30 cm are fully recruited to the fishery and that natural mortality is 0.3. Using F40% (0.31) as a proxy for Fmsy, the threshold exploitation rate is 0.23. Exploitation rate in 2010 was estimated at 0.03 (80% CI 0.02 - 0.05) which was based on the ratio of 2010 catch (195 mt) to survey based swept area estimate of biomass for winter flounder exceeding 30 cm in length (6,341 mt).*

8. Develop and apply analytical approaches and data that can be used for conducting single and multi-year stock projections and for computing candidate ABCs (Acceptable Biological Catch; see Appendix to the TORs) under a set of alternative harvest scenarios. If the stock needs to be rebuilt, take that into account in these projections.

a. Provide numerical short-term projections (3-5 yrs, or through the end of the rebuilding period, as appropriate). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. In carrying out projections, consider a range of assumptions about the most important uncertainties in the assessment (e.g., terminal year abundance, variability in recruitment).

SDWG Ten year AGEPRO projections assumed that the ACL of 230 mt will be taken in 2011. Projections were done using 1000 bootstrap iterations from the split VPA and 1000 mcmc iterations from the preferred ASAP multi run. SSB, catch, and fishing mortality with 80 confidence intervals were estimated from the split VPA at the Fmsy proxy of F40% = 0.43 (derived from the updated split VPA) and 75% of the F40% proxy = 0.32. Projections for the ASAP multi model were also run assuming the F40% proxy = 0.34 and 75% of the F40% = 0.26. Short term projections using the stock recruit reference point with the prior on steepness for the ASAP multi run were also done at Fmsy = 0.57 and 75%Fmsy = 0.42. All projections show relatively high catch in 2012 compared to model time series of catches. The projected VPA SSB increases towards SSBmsy after lower estimates of SSB in 2013 and 2014. The low SSB estimate in 2013 and 2014 is due to the low recruitment estimated in 2009 and 2010 which was influenced by the length based survey calibration. Therefore substantial uncertainty exists with the estimated recruitment in 2009 and 2010. The ASAP multi short term projections result in fishing of the SSB down to SSBmsy. The estimated catch in 2012 shows a large increase relative to the assumed catch in 2011 of 230 mt for both the split VPA and ASAP formulations. The ASAP multi run estimated 2012 catch varies from 1,700 mt from the 75% F40 projection to the stock recruit Fmsy projection estimate of 3,080 mt. However catch declines quickly after 2012 as the stock approaches SSBmsy. Consideration could be given to the overestimation of the plus group in the ASAP model projections. For example a plus group residual adjustment within AGEPRO can be approximated using an assumed plus group discard proportion.

The SARC did not accept the analytical modeling. Therefore projections are not possible.

b. Take into consideration uncertainties in the assessment and the species biology to describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming or remaining overfished, and how this could affect the choice of ABC.

Uncertainties that were not accounted for by assessment and reference point models were evaluated using model diagnostics. Standard model diagnostics (e.g., residual analyses, retrospective analyses) were used for model validation. Vulnerabilities that were not accounted for by assessment and reference point models were evaluated using exploratory modeling, habitat observations and testing the influence of environmental factors on recruitment dynamics. Additional considerations of vulnerability and productivity are the implications of shifts in distribution, recruitment dynamics and increased natural mortality. Nye et al. (2009) found an annual increase in mean depth (0.8 m per year) of the winter flounder distribution, which may have productivity and vulnerability implications. Apparent decreases in estuarine spawning or shifts toward coastal spawning (e.g., DeCelles and Cadrin 2010) may also have implications for vulnerability (e.g., less availability to recreational fisheries) and productivity (less larval retention). Consumption of winter flounder by other fishes, birds and mammals may be increasing as these predator populations increase. The GOM assessment indicates that the stock is well above B_{MSY} *and experiencing low fishing* *mortality. However, the GOM assessment is the most uncertain of the three (from a "feasibility" perspective, if not from a "statistical precision" perspective). The apparent shift in distribution to deeper habitat may be adding uncertainty to the stock assessment reference points that assume stationarity in vital rates. Therefore, it may be vulnerable to overfishing if managed at a catch level close to the nominally projected catch in the near term.*

c. Develop plausible hypotheses (e.g., mixing among the three stocks) which might explain any conflicting trends in the data and undertake scenario analyses to evaluate the consequences of these alternate hypotheses on ABC determination.

Major conflicting signals exist between the catch at age data and survey data within the modeling work. The split VPA is weighted towards the catch at age information while the preferred ASAP multi run has a greater weight on the survey information. Survey trends may not reflect the population changes in response to the large decline in the catch over time if a greater proportion of the population historically remained within the estuaries in the early 1980s where there is no survey coverage. This hypothesis could possible explain why the survey indices are relatively flat with little apparent response to the change in catch. However there is very limited data on the extent of estuarine residing populations in the 1980s. Therefore this hypothesis remains simply as speculation. The consequences of the split VPA being a better reflection of the true dynamics can be evaluated by assuming the catch or ABC from the preferred ASAP projection is taken within the split VPA projection formulation.

9. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in recent SARC reviewed assessments and review panel reports. Identify new research recommendations.

About ten of the previous fourteen research recommendations have been partially addressed. Twelve new research recommendations have been developed by the SDWG for SAW52.

Introduction and Assessment History

Gulf of Maine winter flounder is the smallest of the three winter flounder stocks (Figure C1). Gulf of Maine winter flounder was first assessed in SARC 21 (1995) as an index based assessment. It was noted at that assessment that survey indices were low and relatively few large fish were seen in the survey size distributions. Survey Z estimates were high (1978-1993 mean of 1.21) and the stock was thought to be overexploited. The SARC 36/GARM 1 assessment in 2001 was the first analytical assessment (ADAPT VPA) for this stock. The stock was considered rebuilt and overfishing was not occurring. In GARM II the ADAPT VPA model was updated through 2004 (NEFSC 2005). The GARM II assessment also concluded that the stock is not overfished and overfishing is not occurring. Spawning stock biomass was estimated to be at 3,400 mt and fully recruited F = 0.13 in 2004. SSB at B_{msy} was estimated to be at 4,100 mt and F_{msy} = 0.43. The GARM II VPA developed a severe retrospective pattern in F and a large overestimation of SSB. GARM II concluded that VPA results were too uncertain as a basis for performing projections.

In GARM III the review panel was unable to determine the stock's status relative to the BRPs, but stated that trends in the population were very troubling (NEFSC 2008). The Review Panel generally agreed that the stock biomass was highly likely to be less than the B_{MSY} proxy, and that there is a substantial probability that it was below the minimum stock size threshold. The split VPA model estimated spawning stock biomass in 2007 at 1,100 mt or about 29% of the B_{MSY} proxy (3,792 mt) and fishing mortality in 2007 was 0.42 or about 147% of $F_{40\%}$ = 0.28. The base case VPA and a split forward projection model (SCALE) which put higher weight on the recruitment indices suggested that the stock was not overfished and overfishing was not occurring. However the base case VPA had a severe retrospective pattern. The VPA showed greater reductions in biomass than observed in the survey biomass trends. All models had difficulty fitting the relatively flat age 1 and age 2 recruitment indices and the decrease in adult indices with the large decline in the catch at the end of the time series. The models were not accepted as a basis for status determinations. Therefore the stock status is unknown.

TOR 1: Estimate catch from all sources including landings and discards. Characterize the uncertainty in these sources of data.

Commercial landings were near 1,000 mt from 1964 to the mid 1970s. Thereafter commercial landings increased to a peaked of 2,793 mt in 1982, and then steadily declined to 350 mt in 1999. Landings have been near 650 mt from 2000 to 2004 and about 300 mt from 2005 to 2009. Landings have declined to a record low of 140 mt in 2010 (Table C1, Figure C2). The primary gear used was the otter trawl from 1964-1985 that accounted for an average of 95% of the landings. Otter trawl accounted for an average of 74% of the landings from 1986- 2010 with an increase in the proportion of the landings coming from gillnets (26% from 1986-2010) (Table C2). Since 1999 around 95% percent of the landings are taken in Massachusetts from statistical area 514 (Figures C3 and C4). Winter flounder are landed throughout the year. However a greater proportion of the landings have been coming from quarter three over the last ten years (Figure C4). The proportion of the landings coming from the medium market category has decreased since 2004 (Figures C4).

Recreational landings reached a peak in 1981 with 2,554 mt but declined substantially thereafter (Table C4, Figure C5). Recreational landings have generally been less than 100 mt since 1994, with exception of 2008 were the landings was estimated at 103 mt. The PSE of the recreational landing averaged 29% over the time series. Recreational landing weight was re-estimated using the expanded numbers at length and the length weight relationship by half year for input to the VPA, SCALE, and ASAP models.

In the commercial fishery, annual sampling intensity varied from 6 to 310 mt landed per sample during 1982-2007. Overall sampling intensity was adequate, however temporal and market category coverage in some year was poor (Table C4). Samples were pooled by half year when possible. In 1982 mediums were pooled with unclassified by half year, in 1985, 1995, 2005, 2006, and 2007, smalls were pooled with mediums, and the large samples from adjacent years were used for the lack of samples in 1996, 1999, and 2001. Sampling coverage may have been poor but length frequency samples appeared relatively constant over time and there was a substantial amount of overlap between market categories which help justify the pooling used in the assessment. Lengths of kept fish from observer data were used to supplement length data of unclassified fish. Kept fish lengths taken from gillnet trips in the observer data were used to characterize the gillnet proportion of the landings (Table C5). In 2002 gillnet landings also shifted from occurring mostly in the first half of the year to a greater proportion coming from the second half. In general there has been an increase in the sampling intensity from the commercial ports. However the decline in landings has made it difficult to get samples from the medium and large market categories in recent years. As in GARM III catch at age and catch at length was estimated using observe kept length measurements by gear supplemented with unclassified port lengths by gear from 1999 to 2010. Characterization of the landings using the observer data produced expanded catch at length distributions similar to the length expansions using the port samples by market category for years which had relatively good port sampling (Figures C6 and C7). Size distributions of the landings have been very stable over the past 10 years (Figure C7).

Discards were estimated for the large mesh trawl (1982-2010), gillnet (1986-2010), and northern shrimp fishery (1982-2010) (Table C6 through C7). The survey method was used in estimating both the discard and proportion discards at length for the large mesh trawl fishery from 1982-1988 (Mayo et al. 1992). Observer discard to landings of all species ratios were applied to corresponding commercial fishery landings to estimate discards in weight from 1989 to 2010 for the large mesh trawl fishery. (Wigley et al. 2008)The Fishery Observer length frequency samples were judged inadequate to characterize the proportion discarded at length from 1989 to 1998 for the large mesh trawl fishery and the length proportion from the survey method was used to characterize the size distribution of discarded fish. Observer discard length sampling increased in 2001 and was used to characterize the large mesh trawl discards from 2001 to 2010 (Table C8). The observer sum discarded to landing of all species ratios were used for estimating gillnet discard rates. Observer sum discarded to days fished ratios were used for the northern shrimp fishery since landing of winter flounder in the shrimp fishery is prohibited. The observer length frequency data for gillnet and the northern shrimp fishery were used to characterize the proportion discarded at length. The sample proportion at length, converted to weight, was used to convert the discard estimate in weight to numbers at length. Data from the small mesh trawl fishery was judged as inadequate to estimate discards over the time series (Tables C7 and C9). Observer coverage has improved in the small mesh fishery over the last ten years. The small mesh discard estimates suggests that the discards are small from this fishery. However the estimate in 2010 did showed an increase. As in the southern New England stock (NEFSC 1999), a 50% mortality rate was applied to all commercial discard data (Howell et al., 1992). Numbers at ages were determined using NEFSC/MDMF spring and NEFSC fall survey age-length keys.

A discard mortality of 15% was assumed for recreational discards (B2 category from MRFSS data), as assumed in Howell et al. (1992). Discard losses peaked in 1982 at 140,000 fish. Discards have since declined to an average of about 8,000 fish from 2000 to 2010 (Table C3, Figure C5). Since 1997, irregular sampling of the recreational fisheries by state fisheries agencies has indicated that the discard is usually of fish below the minimum landing size of 12 inches (30 cm). For 1982-2006, the recreational discard has been assumed to have the same length frequency as the catch in the MDMF survey below the legal size and above an assumed hookable fish size (13 cm). Since 2007 lengths of B2 released catch have been collected by the MRFSS program on party charter vessels which were used to characterize the size of the B2 catch. The recreational discard for 1982-2010 is aged using NEFSC/MDMF spring and NEFSC fall survey age-length keys.

A summary of how the catch at age was constructed can be seen in Table C10. Predicted landings using the same discard method was used as a diagnostic of the discard estimates (Table C11). The predicted landings using the kept to landing of all species ratio are variable but on the same order of magnitude with the dealer landings (Table C1). Decreases in the catch and the catch at age components are shown in Table C12 through C16 and Figures C8 and C9. Mean weights at age and the total catch at age are given in Table C17 and Figure C10. Declines in the mean weights at age were observed for most ages in the catch at age over the last four years.

TOR 2: Present survey data being considered and/or used in the assessment (e.g., regional indices of abundance, recruitment, state and other surveys, age-length data, etc.). Characterize uncertainty in these sources of data.

Mean number per tow indices for the NEFSC and the Massachusetts Division of Marine Fisheries (MDMF) spring and fall time series are presented in Table C18 and Figures C9 through C15. 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). For most flatfishes there is evidence for differences in selectivity at length between the two survey vessels. The SDWG used the estimated length based calibration by stock to convert the survey indices in 2009 and 2010 into Albatross equivalent units (Figure C16). Details on the estimation of length based calibration coefficients at length is outlined in a working paper by Miller entitled "Winter Flounder Length-based Survey Calibration". Both the length based and published peer reviewed aggregate calibration effects can be seen in Figures C11 and C12 (Miller 2011, Miller et al., 2010). The survey length and calibrated lengths can be seen in Figures C17 and C18.

All of the indices generally show a slight decrease in the population in the late 1980s from a high in the early 1980s with low abundance remaining through the early 1990s. All of the indices show signs of increase abundance starting in 1998 and 1999. Since 2001 all indices indicate a decrease in abundance (Figure C15). The MDMF survey catchability is on the order of 60 to 100 fish per tow while NEFSC survey catchability is on the order of 4 to 14 fish per tow. Age data for the MDMF fall survey are not available. The NEFSC fall ages were used to age the MDMF fall index.

Maine and New Hampshire (MENH) have been conducting an inshore bottom trawl survey in the spring since 2001 and in the fall since 2000. These survey indices are relatively flat over the time series with slightly higher abundance in the fall of 2010 (Figure C19). The MENH survey catches relatively few fish over 30 cms (Figures C20 and C21). Age modes for the younger fish are also not clearly seen in the size data. However the increase in the fall of 2010 could be due to an incoming stronger year class. A more defined mode at 9cm can be seen in the fall of 2010 (Figures C21). The working group examined some preliminary age information from the spring MENH index. It was noted that growth from inshore Maine and New Hampshire appears to be slower relative to the MDMF and NEFSC surveys. The MENH indices at age were not included in the models for this assessment due to time constraints and missing age data for some years. However the MENH survey was used in the direct biomass area swept estimate.

Normandeau Associates, Inc. monitored entrainment of winter flounder larvae through the Pilgrim Nuclear power plant since 1975. In general this data suggests a higher abundance of winter flounder larvae since 1997 relative to the 1980s and early 1990s (Figure C22).

An examination of the survey catch per tow at length was conducted to determine the ability of the survey in tracking cohorts. Survey catch per tow at length were plotted with alternating spring and fall surveys over time (Figures C23 through C25). Year classes modes were approximated using growth information. The growth and tracking of cohorts in the younger ages can be seen in the MDMF spring and fall surveys. The younger length modes are more difficult to observe in the NEFSC survey which has a lower catchability for the smaller fish. The raw length frequency data suggests the occurrence of a strong 1998 yearclass evident in both the MDMF and NEFSC surveys. However the detection of this yearclass as it growths above legal size is more difficult to discern (Figure C23 and C24). The strong 1998 yearclass is not estimated in the VPA model. However the tracking of year classes is more difficult to observe in the indices at age (Figures C26 through C28).

Some evidence for a change in the spatial distribution can be seen in the MDMF and NEFSC surveys. There appears to be a shift in abundance for all sizes from shallow water in early 1980s to deeper strata at the end of the time series (Figure C29). Offshore stratum 26 which contains Stellwagon bank also shows increase abundance starting in 1999 while the northern offshore strata off the coast of Maine show no signs of recent increases (Figures C30 and C31). Input from fishermen at the SMAST Fishermen input meeting also reiterated this observation. It is not clear how this shift effects the interpretation of the survey indices. Speculation on a reason for why the survey trends are relatively flat over the time series could be due to a greater proportion of the population residing within the estuaries during the 1980s during the height of the recreational fishery. Fish that reside within the estuaries are not covered by any survey.

TOR 3: Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series (integrating results from TOR-5), and estimate their uncertainty. Include area-swept biomass estimates. Investigate if implied survey gear or catchability estimates are reasonable. Include a historical retrospective analysis to allow a comparison with previous assessment results.

[SAW52 Editor's Note: The SARC-52 peer review panel concluded that no ASAP model run provided a suitable basis for management advice. A swept-area biomass method was accepted instead, and it is described in Appendix C1. The ASAP model and results are included below in this report to document the ASAP modeling runs that the SAW Working Group provided to the SARC for peer review.]

Instantaneous Natural Mortality (M)

The SDWG adopted a change in the instantaneous rate of natural mortality (M) for the winter flounder stocks. The value of M previously used in all assessments was 0.2 for all ages and years, and was based on the ICES 3/Tmax "rule-of-thumb" using observed maximum ages for winter flounder (Tmax) of about 15. The current observed Tmax values for the three stock units are GOM = 15 years, GBK = 18 years, and SNE/MA = 16 years (see Growth and Maturity section, above). The adopted change increases this rate to 0.3 for all stocks, ages and years. Evidence can be found in the literature and current model diagnostics to support the increase.

Literature values of M from tagging studies and life history equations indicate M for winter flounder is likely higher than 0.2. Dickie and McCraken (1955) carried out a tagging study in St. Mary Bay, Nova Scotia, Canada (GOM Stock) and estimated a percentage natural mortality rate to be 30% (M = 0.36). Saila et al. (1965) applied Ricker's equilibrium yield equation to winter flounder from Rhode Island waters (Tmax = 12) and using F values from Berry et al. (1965) calculated M to be 0.36. Poole (1969) analyzed tagging data from New York waters from five different years and estimated values for M of 0.54 (1937), 0.33 (1938), 0.5 (1964), 0.52 (1965), and 0.52 (1966). Finally, an analysis of tagging data from a large scale study along the coast of Massachusetts provided a percentage natural mortality rate of 27%, or M = 0.32 (Howe and Coates 1975). For this assessment, a re-analysis of the Howe and Coates (1975) tagging data was conducted using a contemporary tagging model to estimate natural mortality (Wood WP 15). The tagging model fit to the data was the instantaneous rates formulation of the Brownie et al. (1985) recovery model (Hoenig et al. 1998). This work provided an M of 0.30 with 95% confidence interval from 0.259 to 0.346.

Values derived from life history equations found in the literature also support a higher estimate of M for winter flounder. Three equations were used along with a maximum age (Tmax) of 16 to derive estimates of M equal to 0.28, 0.26, and 0.19 (the equations from Hoenig 1983, Hewitt and Hoenig 2005, and ICES, respectively). A newly proposed method from Gislason et al. (2010), based on SNE/MA stock mean size at age (Ages 1-16) and von Bertalanffy growth parameters, estimated M to be 0.37 (see text table below).

Values of Natural Mortality (M) for winter flounder found in the literature and derived using life-history equations.

Study	**Method**	**M**
ICES rule-of-thumb	Equation: 3/Tmax	0.19
Hewett and Hoenig 2005	Equation: 4.22/Tmax	0.26
Hoenig 1983	Equation: 1.44-0.982*ln(Tmax)	0.28
Howe and Coates 1975	Analysis of Tagging Data	0.32
Wood 2011 WP15	Re-analysis of Howe and Coates 1975	0.30
Poole 1969	Analysis of Tagging Data from 1938	0.33
Dickie and McCraken 1955	Analysis of Tagging Data	0.36
Saila et al. 1965	Ricker Equil. Yield Equation and Tmax	0.36
Gislason et al. 2010	Equation: Mean size at age and VBG	0.37
Poole 1969	Analysis of Tagging Data from 1964	0.50
Poole 1969	Analysis of Tagging Data from 1965	0.52
Poole 1969	Analysis of Tagging Data from 1966	0.52
Poole 1969	Analysis of Tagging Data from 1937	0.54

Preliminary assessment population model run diagnostics also in general support a higher value for M. Profiles in mean squared residual for ADAPT VPA SNE/MA stock models indicate best fits for M in the range of 0.2 to 0.3. The likelihood profile of initial ASAP SCAA model runs for the SNE/MA stock indicates a best fit for M= 0.6. Model runs from Rademeyer and Butterworth SCAA (ASPM) model (2011) at M equal to 0.2, 0.3, and 0.4 also reveal decreasing negative log-likelihood as M is increased for GOM and SNE/MA stock models (see text table below).

Results of SCAA for the **Gulf of Maine winter flounder** for each combination of 3 levels of natural mortality (*M*=0.2, 0.3 and 0.4, constant throughout the assessment period) and 3 weightings of the survey CAA likelihood (*w*=0.1, 0.3 and 0.5). The runs with w=0.3 and 0.5 have both commercial and survey selectivities flat at older ages, while the runs with w=0.1 have only the commercial selectivity flat. Displayed values are the negative log-likelihoods of each model.

		M	
Weighting	0.2	0.3	0.4
0.1	-123.2	-126.6	-129.1
0.3	-156.9	-177.2	-196.1
0.5	-255.6	-263.2	-280.8

Results of SCAA for the **SNE/MA winter flounder** for 3 levels of natural mortality for Base Case 2. Displayed values are the negative log-likelihoods of each model.

	M		
	0.2	0.3	0.4
-LL	-123.2	-126.6	-129.1

The SDWG also considered other evidence that might justify an increase in M for winter flounder. The NEFSC's food habits database (Smith and Link 2010) was examined to identify the major fish predators of winter flounder. These predators include Atlantic cod, sea raven, monkfish (goosefish), spiny dogfish, winter skate and little skate. A preliminary examination was undertaken to determine the prominence of winter flounder in the diets of these predators, across all seasons, years, size classes of predator, sizes of prey, and geographic locales. The overall frequency of occurrence of winter flounder in the stomachs is not a common or high occurrence (see text table below), always less than 0.15%.

Occurrence of winter flounder in their major fish predators.

	Number of stomachs	Occurrences of winter flounder	% Freq. of occurrence
Spiny dogfish	67,565	27	0.040%
Winter skate	17,708	6	0.034%
Little skate	28,725	6	0.021%
Atlantic cod	20,142	27	0.134%
Sea raven	7,968	10	0.126%
Goosefish	10,742	12	0.112%

Further, the contribution of winter flounder to the diets of these predators species is also notably small (see text table below), usually less than 0.4%.

Contribution of winter flounder to the diet of their major fish predators.

	% Diet composition of winter flounder,	95% CI
Spiny dogfish	0.2049%	0.10678
Winter skate	0.1454%	0.16008
Little skate	0.0124%	0.01618
Atlantic cod	0.3172%	0.24032
Sea raven	0.8831%	0.78407
Goosefish	0.2492%	0.25947

Understandably the temptation exists to evaluate these relatively low contributions of diet with respect to consumptive removals of winter flounder as compared to winter flounder stock abundance and (relatively low) landings, initially using *ad hoc* or proxy methods. Yet just as one would not do so when assessing the status of a stock without a fuller exploration of all the sensitivities, uncertainties and caveats of the appropriate estimators and parameters, the SDWG did not recommend doing so for scoping winter flounder predatory removals at this time. The SDWG also noted that for percentages as low as observed, when allocated to the three winter flounder stocks and explored seasonally or as a time series, there are going to be large numbers of zeroes and attendant uncertainties and variances that would logically offset any potentially high individual predator total population-level consumption rates. Thus, the SDWG does not provide comment as to the merit of exploring or relative magnitude of the issue, but recommends that the topic should be forwarded as an important research recommendation. Other sources of increased natural mortality may come from perceived increases in seal populations along the New England coast, which are known to be predators of winter flounder (Ampela 2009). Population size was estimated at 5600 seals in 1999 (Waring et al. 2007) and a current survey is being conducted to estimate the size of the seal population. However, no time series of seal abundance or consumption of winter flounder is available.

Stock Assessment Models

Abundance indices at age were available from several research surveys: NEFSC spring bottom trawl ages 1-8+, NEFSC fall ages 1-8+ (advanced to tune January 1 abundance of ages 2-8+), Massachusetts spring ages 1-8+, and Massachusetts fall ages 0-8+ (advanced to tune January 1 abundance of ages 1-8+) (Figures C32). The influence of the length based conversion on the indices at age can be seen in Figure C33. The survey mean lengths at age also showed a slight decline at the end of the time series (Figure C34).

There was little change in the female 3 year moving average maturity using MDMF spring survey (Figure C35). A logistic maturity estimate using all years combined (1982-2010) from the spring MDMF survey did not change from the maturity schedule estimated (1982-2007) from GARM III (Figure C36). A histological maturation study described in the working paper by McBride et al 2011 indicated that the MDMF survey macroscopic maturation estimate was appropriate for this stock.

The base and split VPA with assumed natural mortality equal to 0.2 was updated from the GARM III assessment. Differences between the split VPA m=0.2 and m=0.3 can be seen in Figure C37. There was little difference in retrospective pattern between the split model with m=0.2 to the split model m=0.3. All subsequent model runs were done with m=0.3 based on the SDWG conclusion above. As in GARM III the base case VPA run showed a severe pattern in the residuals (Figure C38) and exhibits a severe retrospective pattern in F, recruitment, and a large overestimation of SSB (Figures C39 and C40). Splitting the surveys allows the model to estimate further declines in abundance with higher Fs at the end of the time series. The split survey model is less constrained by the conflicting signals between the large decline in the catch and the survey abundance of the older fish (4+) at the end of the time series. As in GARM III, splitting all of the surveys between 1993 and 1994 did improve the retrospective pattern (Figures C41 and C42). The survey split in the updated assessment appears to have reduced the retrospective bias further then what was observed in the GARM III split VPA model. In addition the update split model estimates for 2007 was similar to the terminal year estimates from the GARM III split VPA which can be seen in the historical retrospective plots in Figures C43 and C44. However other diagnostics still imply major sources for concern surrounding the VPA model formulation for GOM winter flounder. 1) A significant residual pattern in the survey exists for the first half of the model (1982-1993), however the residual pattern seems to have improved for the second half (1994-2010) (Figure C38). 2) Forward and backward diagnostic calculations of the plus group suggest that the plus group estimates are not well determined (Table C19). 3) Area swept Q estimates suggest efficiencies greater than one in both the base and split model runs indicating that the area swept survey population estimate is higher than what is estimated by the model (Figures C45 and C46). 4) The split model results in a large change in the Q estimate. Many of the survey Qs more than tripled in the split VPA run. 5) Biological reasons for a strong dome shape pattern in the Q at age from the surveys is difficult to understand (Figures C45 and C46). However this dome shape concern in the surveys also exists in the forward projecting age structured models.

The SCALE model is a simple forward projecting model that tunes to age data for the younger recruitment ages (age 1, 2, and 3) and length data for the larger adult fish (30+ cm). The SCALE model assumes an overall time invariant growth curve with assumed input variation around the mean lengths at age. The model also assumes flat-topped selectivity in the surveys. The Base SCALE model run possessed a similar retrospective pattern as the VPA. The split SCALE model results were sensitive to the weighting on the recruitment indices. The SDWG did a brief exploration of the SCALE model for this assessment. The SCALE model appeared to possess similar diagnostic issues as observed during GARM III. The estimated selectivity and fishing mortality was sensitive to the assumed input variation on the growth (mean lengths at age). The SDWG concentrated on developing the assessment in ASAP (Age Structured Assessment Program) since there appeared to be greater dynamics present in the indices at age relative to the apparent lack of change in the size structure over time. In addition ASAP allows for the estimation of dome shape selectivity patterns in the surveys.

Preliminary runs were first developed in ASAP similar to the base and split VPA configuration. Indices were input as indices at age. This preliminary runs had a relatively high weight on fitting the catch at age compositions (150 effective sample size). The preliminary runs showed similar results as the VPA with similar diagnostic issues (Figure C47). However, the split ASAP model possessed a severe retrospective pattern (Figure C48). The split in ASAP did not reduce the retrospective pattern as observed with the split VPA model.

Reducing the weight on fitting the catch at age composition (50 effective sample size) in the ASAP base model allows a better fit to the survey indices. Trends in the estimated stock numbers at age can be seen in Figure C49. The estimated biomass over the last decade increases as the weight on the catch at age composition is lowered. This results in further overestimation of the plus group relative to the run with a higher weight (150) on the catch at age composition (Figures C50 and C51). The retrospective pattern with an effective sample size weight of 50 compared to a weight of 150 also showed a reduction in the retrospective pattern (Figure C52 and C53). Similar results were seen with the modeling of Gulf of Maine winter flounder done in an Age-Structured Production Model (ASPM) which is described in Rademeyer and Butterworth MS 2011. The 50 weight model showed a similar dome shaped pattern in Qs as the split VPA (Figure C54).

The SDWG also explored an ASAP model formulation which fit the aggregated survey indices and survey age structure as a multinomial. This formulation does allow for fixing the assumptions on survey selectivity. In general both ASAP formulations produced similar results. The multinomial (ASAP multi run) formation did produce some difference in estimated biomass trends at the start of the model (1980s) and a lower estimate of biomass at the end of the time series. The SDWG did some further refinement to the final multi run through the estimate of a separate selectivity block from 1998 to 2010. This did result in a slight shift in the selectivity to older fish for the second time block as observed in the catch at age. Fits to catch at age composition, estimated survey selectivity, fits to the aggregate survey indices, predicted stock numbers at age, and the retrospective pattern can be seen in Figures C55 to C60. The difficulties in estimating population scale can be seen when comparing the results from different models (VPA and ASAP) and for models with different weighting on the data sources (Figure C61).

The combined survey area swept 30+ biomass estimates are described in Appendix C1. The fall survey biomass estimates were judge more appropriate since a greater proportion of the population should be within the survey area during the fall because the fish are not spawning within the estuaries at that time. The area swept 30+ biomass for the fall between 2009 and 2010 ranged from 6,300 mt to 7,600 mt assuming a gear efficiency of 60 percent (q=0.6). This survey based biomass estimate was used to inform the weighting on the catch at age composition used in the model. Therefore the 30+ biomass estimate at the end of the time series was important for judging the feasibility of the model results. For example the ASAP which used dome shape fishery selectivity had desirable diagnostic properties but the biomass estimates were unfeasibly high at the end of the time series (over 20,000 mt). The 30+ cm biomass estimate from the survey estimate is comparable to the 4+ biomass, exploitable biomass, and the SSB in 2009 and 2010 from the SDWG multi ASAP run (Figures C62 and C63).

Age Structured Assessment Program (ASAP) Description

ASAP (Age Structured Assessment Program v2.0.20, Legault and Restrepo 1998) and the technical manual can be obtained from the NOAA Fisheries Toolbox (http://nft.nefsc.noaa.gov/). ASAP is an age-structured model that uses forward computations assuming separability of fishing mortality into year and age components to estimate population sizes given observed catches, catch-at-age, and indices of abundance. Discards can be treated explicitly. The separability assumption is partially relaxed by allowing for fleet-specific computations and by allowing the selectivity at age to change in blocks of years. Weights are input for different components of the objective function which allows for configurations ranging from relatively simple age-structured production models to fully parameterized statistical catch at age models.

The objective function is the sum of the negative log-likelihood of the fit to various model components. In the SDWG preferred ASAP multi run the catch at age and survey age composition are modeled assuming a multinomial distribution, while most other model components are assumed to have lognormal error. Specifically, lognormal error is assumed for: total catch in weight by fleet, survey indices, stock recruit relationship, and annual deviations in fishing mortality. Recruitment deviations are also assumed to follow a lognormal distribution, with annual deviations estimated as a bounded vector to force them to sum to zero (this centers the predictions on the expected stock recruit relationship).

ASAP Model Inputs and Formulation

The ASAP model formulation used a composite catch by a directed fleet starting in 1982. Commercial landings do exist prior to 1982. However recreational landings are unknown prior to 1981. All models included the NEFSC spring and fall as well as the Massachusetts state surveys for both the spring and fall. Minimum swept area abundances and an assumed CV of 0.4, as well as age composition for each survey were used in the model. The working group focused initial scrutiny on models that treated the survey indices by age, similar to a VPA model formulation but due to difficulties to reconcile model diagnostics, the multinomial formulation was preferred by the working group. The preferred ASAP multi model used a plus group at age 7. Exploratory runs examined model sensitivity to estimating a stock recruit function versus estimating an average recruitment with annual deviations; estimating age-specific selectivity for the surveys versus forcing the survey to have a flat-topped selectivity; "breaking" the survey time series into two separate series or maintaining a continuous time series; and adding or removing selectivity "blocks" to the directed and bycatch fleets. In considering these various model iterations, diagnostics were examined to determine if the fit improved. Specifically, the pattern of residuals in age composition for catch and indices, residuals in the fit to total catch and annual index values, components of the objective function in addition to total objective function and number of estimated parameters, as well as the "feasibility " of the estimated selectivity patterns were examined. With regard to the last criterion ("feasibility" of estimated selectivity), the models tended towards solutions with sharply domed selectivities for both the directed fleet and the surveys. As there was nothing biological to suggest that fish at ages 5 and beyond would have very low catchability (i.e., no known behavioral aspects, no strong swimming capabilities), nothing gear related that would suggest lower catchability (no outswimming otter trawls, no other known gear interactions), and no known market conditions that would favor smaller fish. The SDWG found it hard to reconcile selectivities of 0.10 on the 7+ group.

Model formulations for both the indices at age and the multinomial model were examined. Although the objective function values were not directly comparable between these two model treatments, owing to differences in the underlying data, residual diagnostics, overall fits, and retrospective patterns were compared. The working group agreed to the following preferred multi configuration: A model that did not split the survey indices, two selectivity blocks for the directed fleet (the break occurred between 1997 and 1998), forced with a selectivity = 1 for ages 4 and older. With all models considered, there was a strong correlation between the selectivity estimated for the directed fleet and the selectivity of the surveys. Forcing a flattop for the survey indices caused the selectivity estimates for the directed fleet to be also flattopped. Similarly, allowing a dome in the survey led to dome selectivity in the directed fishery. For this reason, a flattop was assumed for the directed fleet fishery. For this selectivity pattern, the age composition residuals showed some patterning, particularly in the plus age category and the overall index as well as the total catch showed some time trends in the fit to the residuals. In contrast, when a dome selectivity is estimated in the fishery, there was an improvement in both the residual age composition and residual fit the overall index and total catch. However, the estimates of spawning stock biomass and recruitment were unreasonably high due to cryptic biomass that was generated from the dome selectivity pattern. Although the flattop configuration did not provide the best diagnostics, the estimates of spawning stock biomass and recruitment were within reason. This is a fairly consistent trade-off seen in many of the model diagnostics, wherein improvements in the fit to the catch at age, including the total data (catch or total index values) results in a different perception of the stock. Thus, selecting the 'best' model depended to some extent on the amount of confidence that one had in the age composition data as well as the total catch and the indices. Complete diagnostic output plots can be found in the Appendix C2 ("Multi models diagnostics_2_Block_Fishery_Selectivity").

Preferred ASAP Multi model Retrospective Pattern

A retrospective analysis on the ASAP multi model using a seven year peel was conducted to examine the stability of the model estimate for fishing mortality, recruitment and spawning stock biomass (Figures C59 and C60). Due to the change in selectivity block beginning in 1998, it was difficult to interpret the earliest peels because there was an imbalance in the number of parameters being estimated versus number of years with additional data. However, it was noted that the model that estimated a dome in the directed fleet had the lowest retrospective while the preferred multi model exhibited higher retrospective averages.

Preferred ASAP Multi model Sensitivity Analyses

For completeness, sensitivity to the model decisions adopted in the base model are summarized in Table C20. Seven additional runs were explored including assuming survey flattop selectivity, lowering or increasing the age to fix survey selectivity, allowing a dome in the fishery and removing time blocks in the fishery selectivity. Due to convergence problems in some of the sensitivity runs, only four of the seven sensitivity runs were reported. Only one of the four runs reported assumed no time blocks in the directed fishery selectivity. The motivation for introducing selectivity blocks, and the year that they were introduced, was an attempt to account for changes in the fishery composition and pertinent regulations (mesh size and minimum sizes changes). While this model offered similar diagnostics as in the ASAP multi run, the retrospective estimates were improved for spawning stock biomass, recruitment and Fishing mortality. The overall objective function for the single block directed fishery was 3480 while for the base model, it was 3453. Thus, the multi run which estimated an additional block of selectivity improved the objective function by 27 points at the cost adding four additional parameters to the model.

The remaining three models were based on the two block selectivity in the fishery. Lowering the age to fix the survey selectivity suggested improvement in the likelihood components of the model. However, the model had problem converging, possibly due to parameter boundary issue as hessian was obtained for the model. Assuming flattop in the survey did not improve the overall objective likelihood function neither did it provide improved diagnostic in comparison to the ASAP multi run. Additionally, survey catchability for the Massachusetts state survey was greater than one and the retrospective estimates deteriorate substantially.

Preferred ASAP multi Model Results

Fishing mortality on ages 3+ varied between 0.359and 0.648 from 1982 to 1989 then decreased consistently since 1990 from 0.586 to 0.102 in 1999. Fishing mortality Varied slightly between 0.138 and 0.058 from 2000 to 2009. The fishing mortality rate in 2010 is estimated to be 0.032 (80% confidence interval 0.026 – 0.038; Figures C63 and C64).

Recruitment has been relatively stable throughout the time series. Mean recruitment was around 8.1million for age1 recruits. Several abundant year classes were produced in 1982-1983, 1985, 2004,-2007 ranging from 10 million to 11.9 million. Recruitment in 2010 is estimated to be 4.7 million, lowest in the time series (80% Confidence interval 3.2 million – 6.2million).

Spawning stock biomass declined substantially early in the time series from 12,506 metric tons in 1982 to 1,487 metric tons in 1993, lowest in the time series. Thereafter, SSB has steadily increased from 1,664 metric tons in 1994 to 5,817 metric tons in 2009. Spawning stock biomass in 2010 is estimated to be 5,803 metric tons (80% confidence interval 4,901 – 6,705 metric tons; Figures C63 and C64).

SDWG Stock Assessment Model Discussion and Conclusions

The population models have difficulty with the conflicting data trends within the assessment, specifically the large decrease in the catch over the time series with very little change in the indices or age structure in both the catch and surveys. These conflicting signals were identified in GARM III and results in a severe retrospective pattern in the modeling. Splitting of the survey indices did help reduce the retrospective bias in the models. However the magnitude of the change in q estimated from the split that was required for the model to fit the lack of older fish in the catch at age was no longer believable. Area swept q estimates (2-3 second half) which exceeded 1 suggested that model estimates of biomass was far lower then what was observed in the surveys. At GARM III stock status determination changed from not overfished and not overfishing to overfished and overfishing with the split. Examination of an alternative forward projecting model (SCALE) that tunes to length data produced similar results and had similar diagnostic issues as the VPA. Status determination from the SCALE model was also sensitive the weighing of different data components. The lack of fit to the survey indices in the GARM III VPA resulted in high uncertainty in the status determination which led to rejection of the models.

Conflicting trends in the data still exist in this assessment. However there are several changes that contribute to change in the estimated population trends and status determination relative to the GARM III models. 1) The change in assumed natural mortality from 0.2 to 0.3. 2) Trends at the end of the assessment during the GARM III where difficult to interpret due to the declining catch with declines in the survey index for older fish (4+) at the end of the time series in 2007. This assessment added three more years (2008-2010) to the GARM assessment. 3) There have been increases in the indices at age for the older fish (5,6 7+) since the GARM assessment. 4) The biggest change that contributed to the change in population trends was the switch in the modeling of the stock to ASAP which allowed errors in the fit to the catch at age and a better fit to the surveys indices without the split.

Population scale is poorly determined within the modeling due to the conflicting data trends. The scaling of the population estimates was sensitive to the weight imposed on the catch at age compositions. The conflicting trends in the data produce a bifurcation in the model results. This was observed in both the ASAP and ASPM modeling work from Rademeyer and Butterworth MS 2011. Forward projections models that are forced to fit the catch at age cannot fit the survey indices and result in similar trends as seen in the VPA. Tension within the model is lowered, retrospective bias is reduced, and population estimates are scaled higher with larger increases at the end of the time series as the fit on the catch at age composition in ASAP and ASPM models is relaxed. Preferred ASAP and ASPM models assumed a flat-topped selectivity pattern. This results in an overestimation of fish in the plus group as the fit to the catch at age composition is lowered. Allowing both models to fit a dome shape selectivity pattern further releases the tension within the model and allows the estimation of the strong dome shaped pattern with unrealistically high biomass estimates at the end of the time series.

The SDWG developed a table outlining the reasons why the ASAP multi model was the preferred model in this assessment (Table C21). The split VPA lack of fit to the overall survey indices with estimates on biomass far below the minimum area swept numbers made it difficult for the SDWG to accept this model formulation. The ASAP model formulation did not require a split survey configuration to adjust for the retrospective pattern. The combined survey 30+ biomass area swept estimate was used to inform the optimal weighting for the preferred model formulation. The resulting final model weighting formulation considered both the tradeoff between retrospective bias and feasible biomass estimates at the end of the time series. A retrospective pattern did exist in the preferred ASAP multi run but the SDWG noted that the last two years of the model appeared to be consistently estimated. The within model uncertainty will not capture the uncertainty in this assessment considering how sensitivity the results are to the model formulation and weighting.

The SARC concluded that the assessment model were too unreliable as a basis for management. The accepted assessment of GOM winter flounder stock is based on an empirical swept-area model utilizing data from the 2010 NEFSC fall survey, the MADMF fall survey, and the Maine-New Hampshire fall inshore surveys which is summarized in apppendex C1. Using an efficiency value of 0.6 the estimated stock biomass in 2010 of fish greater than 30 cm was 6,341 mt (80% CI 4,230 - 8,800 mt).

TOR 4: Perform a sensitivity analysis which examines the impact of allocation of catch to stock areas on model performance (in TOR-3).

The SDWG interpretation of TOR4 is that the variance of the commercial landings due to the 1995 and later area-allocation scheme should be used as the basis for the magnitude of landings that might be lost or gained from the stock-specific assessments, and then perform an exercise to run the assessment model with those potential biases and report the results. Additional work was done to estimate the error in the commercial landings due to misreporting of commercial landings to statistical area at allocation level A, the initial reporting level in mandatory Vessel Trip Reports (VTRs; Palmer and Wigley MS 2011). Vessel monitoring system (VMS) positional data from northeast United States fisheries for 2004-2008 were used to validate the statistical area fished and stock allocation of commercial landings derived from the VTRs. The accuracy of the VMS method relative to the VTRs was assessed using haul locations and catch data recorded by at-sea observers. This work was performed for several New England groundfish species. The perceived under-reporting of statistical areas in the VTR data led to minor (< 5%) differences in the overall species allocations; only nine stocks in the five year time-series exhibited differences in stock allocations exceeding 2.0% (2004: northern and southern silver hake, ± 3.0%; 2006: northern and southern windowpane flounder, ± 4.7%; 2007: Georges Bank winter flounder, 2.4%; 2008: Georges Bank winter flounder, 2.4%, Southern New England/Mid-Atlantic winter flounder, -3.2%, and northern and southern windowpane flounder, ± 3.4%). Given the magnitude of these errors, the SDWG elected to run the final GOM winter flounder ASAP model, with an additional 5% PSE in commercial landings added to the estimated PSE over the 1995-2010 time series.

For the GOM stock the total catch consists of 4 components. The commercial landings have a calculated Proportional Standard Error (PSE; due to the commercial landings area-allocation procedure; available for 1995 and later years, with the mean of those years substituted for 1982-1994) ranging from 5.3% to about 6.5%; the commercial discard (trawl and gillnet) PSEs range from 16-177% (available for 1994-2010, mean of those years substituted for 1982-1993); and the recreational landings PSEs range from 17-50%. Because the PSEs for the commercial landings are low, and the commercial landings account for about two-thirds of the total catch, the total catch weighted-average annual PSEs range from 7-30%, and averages 11.7% (unweighted) for the 1981-2010 time series.

The catch in the final assessment model was increased and decreased by the annually varying PSE and models re-run to provide an additional measure of uncertainty of assessment estimates. For the final ASAP multi model, the fishing mortality estimate in 2010 did not change greatly (0.01 to 0.034). The 2010 SSB range was 4,700 to 6,900 mt, was similar to the MCMC estimate of uncertainty (Figures C63 and C65).

TOR 5: Examine the effects of incorporating environmental factors in models of population dynamics (e.g., spring water temperatures in an environmentally-explicit stock recruitment function).

This TOR is addressed in a separate working paper from Hare MS 2011 entitled "Development of environmentally-explicit stock-recruitment models for three stocks of winter flounder (Pseudopleuronectes americanus) along the northeast coast of the United States**".**

To develop environmentally-explicit stock recruitment relationships, three specific types of data are required: spawning stock biomass, recruitment, and environmental data. Spawning stock biomass and recruitment data from the final 2011 SAW 52 assessment models were used in the analysis. For the GOM stock, recruitment (lagged by 1 year) and spawning stock biomass pairs were used from the ASAP multi model. Two general types of temperature data were used: air temperatures and coastal water temperature. In addition to temperature, four large-scale forcing indices were included in the analyses. The North Atlantic Oscillation (NAO) is the dominant mode of winter climate variability in the North Atlantic region and has been related to numerous physical and biological variables across the North Atlantic (Ottersen et al. 2001, Visbeck et al. 2003). The Atlantic Multidecadal Oscillation (AMO) is a natural mode of climate variability and represents a detrended multi-decadal pattern of sea surface temperatures across the North Atlantic with a period of 60-80 years (Kerr 2005). Finally, the Gulf Stream index is a measure of the northern extent of the Gulf Stream south of the northeast U.S. shelf ecosystem. The Gulf Stream position is related to the larger basin-wide circulation, which in turn is related to NAO and AMO. Two Gulf Stream indices are used here (Joyce and Zhang 2010, Taylor and Stephens 1998).

In summary, for the Gulf of Maine stock, increased winter air temperatures are related to lower recruitment, but the strength of this environmental forcing is less than for the Southern New England stock. This result makes sense in the context of the distribution of winter flounder; the southern stock is most affected by warmer temperatures.

One use of the environmentally-explicit models is to develop short-term and long-term forecasting models. Based on the this work, there is no trend in winter temperature over the past 30 years and thus short-term forecasts can be developed using the environmentally-explicit models assuming winter temperatures to be at their mean state. It may also be useful to develop short-term forecasts under warm temperatures and short temperatures to provide managers with a tangible understanding of the effect of temperature on the stocks. The environmentally-explicit models could also be used to develop longer-term forecasts following the approach of Hare et al. (2010). These forecasts would provide an assessment of the sustainability of the winter flounder fishery on the 30-100 time scale.

Work is underway within the SDWG to incorporate environmentally-explicit stock-recruitment models into the NFT standard software used to fit stock-recruitment models and to perform projections of stock and fishery catch. However, this work has not been developed sufficiently to be made available for peer-review at this time (see new Research Recommendation 10).

TOR 6: State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for BMSY, BTHRESHOLD, and FMSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

The 2008 GARM III assessment was not accepted and the overfished and overfishing status of the GOM winter flounder stock is currently unknown. However GARM III stated that it is highly likely that biomass is below B_{MSY}, and that there is a substantial probability that it is below the ½ B_{MSY} threshold. A rebuilding schedule was not developed for this stock since GARM III stock status was considered unknown and the GARM I and II assessments did not considered the stock overfished.

The estimated biological reference points are summarized in table C22. The split VPA estimated higher percent maximum spawning potential (MSP) proxies relative to the ASAP model because the VPA estimated selectivity was shifted to older fish (Figure C66). The preferred SDWG ASAP multi run estimated a F40 FMSY proxy at 0.34 using the 2006-2010 average mean weights and selectivity as input to the YPR analysis (Table C23; Figure C67). The F40 SSBmsy was estimated from a long term projection (100 years) using the CDF of recruitment from the entire model time series (1982-2010) and the estimated YPR F40. The SSBmsy using the FMSY = F40 proxy was estimated at 3,287 mt with a SSBmsy threshold estimate of 1,644 mt and a MSY equal to 1,080 mt for the ASAP multi run. Stock recruit relationships from the split VPA and ASAP multi run can be seen in figures C68 and C69. The split in the VPA results in a trend in the estimated recruitment which produces a lower steepness and a stronger relationship in the stock recruit curve. Performing a likelihood profile on steepness and a MCMC on the stock recruit model suggests that the steepness, SSBmsy and Fmsy was not well determined from the ASAP multi run (Table C24; Figure C70). The SDWG Beverton Holt stock recruitment Fmsy using the Pleuronectids steepness prior from Myers et al. (1999; 0.8 mean and CV = 0.09) was estimated at 0.57. The stock recruit SSBmsy was estimated at 2,167, SSBmsy threshold = 1,084 mt, and MSY = 1,152. The MSY estimate did not vary greatly with SSBmsy from the MCMC in the stock recruitment analysis (Figure C71). The SDWG expressed concern with the stock recruitment estimate of SSBmsy being estimated in the lower end of the range of past SSB observations.

The SDWG developed a unified response to TOR6, taking into consideration the assessment results for all three stocks. The fishing mortality and biomass Biological Reference Points (BRPs) discussed below are from the Final models accepted for the stocks. As defined in the Magnuson Act, 'overfishing' means "a rate or level of fishing mortality that jeopardizes the capacity of a fishery to produce the maximum sustainable yield on a continuing basis" (i.e., FMSY). The guidelines allow for the projected catch associated with the overfishing limit (OFL) to be based on FMSY proxies. Many proxies are used to define overfishing in situations when FMSY is not well determined. The SDWG interpreted these guidelines to mean that best practice is to use a FMSY estimate instead of a proxy when FMSY can be reliably estimated. The SDWG estimated FSMY for the winter flounder stocks as well as proxies in the form of

F40%. The SDWG developed consensus on some aspects of the FMSY estimates (relative magnitude across stocks), but also had some disagreement about the reliability of FMSY estimates (related to the perceived reliability of the respective assessments). The SDWG could not come to consensus on the preferred MSY reference points for the three winter flounder stocks. Updated estimates of F40% were provided as the existing overfishing definitions and as alternatives to FMSY and SSBMSY estimates. Estimates of F40% and SSB40% were provided as potential overfishing definitions based on the precedence offered by GARM-III (NEFSC 2008), instead of other potential %MSP alternatives.

Appropriateness of FMSY Estimates

The SDWG estimates of FMSY utilize data and prior information in a statistical framework. Estimation of the steepness parameters (*h*) in the stock-recruitment relationships used the available stock-recruitment estimates and a prior distribution of *h* from other Pleuronectid flatfishes (Myers et al. 1999), as was used in previous assessments of SNE/MA winter flounder (NEFSC 2002).

Steepness was estimated to be:

- 0.84 for Gulf of Maine winter flounder
- 0.85 for Georges Bank winter flounder
- 0.64 for SNE/MA winter flounder

The SDWG estimates of *h* for winter flounder stocks are realistic. They are compatible with both the estimates of *h* for Pleuronectids that were used as priors, and with the distribution of all of the estimates in Myers et al. (1999). Uncertainty in FMSY is estimable based on stock-recruitment relationships, but not all sources of uncertainty are included in the SDWG evaluation (e.g., uncertainty in assumed natural mortality, precision and accuracy of stock-recruit estimates are not considered).

Concerns about the reliability of the estimates FMSY

There are aspects of using a prior for steepness for these stocks that are problematic. If no prior is applied, two of the three resulting stock-recruit relationships are not theoretically feasible (e.g., the linear increase in SNE/MA recruitment as a function of spawning stock size; the constant recruitment even at low spawning stock size for GBK winter flounder). There are several concerns with the prior on *h* from Myers et al. (1999) meta-analysis for Pleurinectid flatfishes. The prior is not well understood, because the original data was not available at the SDWG. Many of the stocks used to form the prior have $M < 0.2$. The appropriateness of this prior for the U.S. winter flounder stocks, with assumed $M = 0.3$, is therefore unknown. The number of Pleuronectid stocks in the Myers et al. (1999) study is limited (n=14), and there were no winter flounder stocks included. Derivation of the precision estimate of *h* (0.09; NEFSC 2002) is not clearly documented. The assumed normal error structure for the prior may not be appropriate for a parameter bounded by 0.2 and 1. Myers et al. (1999) stated that "the family-level estimates (shown in boldface) should be used with caution." FMSY estimates depend on both mean and precision of steepness, but the SDWG did not have information on how well the Myers et al. (1999) values were estimated.

The precision of steepness (*h*) estimates show a moderate range of possible values and an associated moderate range in estimates of FMSY(see text table below):

Estimates of steepness (h), FMSY and %MSP with 80% confidence intervals and CVs.

Stock	h	CV	10%	90%	FMSY	CV	10%	90%	%MSP	10%	90%
GOM	0.84	0.08	0.75	0.92	0.565	0.19	0.43	0.77	28	34	21
GBK	0.85	0.08	0.75	0.94	0.500	0.22	0.39	0.69	29	35	22
SNE/MA	0.64	0.08	0.57	0.76	0.310	0.07	0.27	0.43	42	46	32

The implied maximum lifetime reproductive rate [4*h*/(1-*h*)] is quite variable among the stock (*h*=0.64 implies ahat=7.1 while *h*=0.84 implies ahat=21.0), where ahat represents the number of spawners produced by each spawner over its lifetime at very low spawner abundance (i.e., assuming absolutely no density dependence). With similar growth, maturity and natural mortality rates, it is not clear why the implied reproductive rates are so different.

The %MSP associated with the range of FMSYestimates suggests that F40% is compatible with FMSY for SNE/MA winter flounder, but those ranges suggest that F40% is not compatible with FMSY for the GOM and GBK stocks. The %MSP associated with FMSY estimates range from 28% to 42%, but it is again unclear why the %MSP values are up to 50% different for stocks with similar biology and fishery characteristic, when only the stock-recruitment steepness differs.

The SDWG had several concerns about the use of F40% as an overfishing definition. F%MSP ignores any information from stock and recruitment estimates, and therefore may be inconsistent with FMSY estimates that use such information. The performance of F40% for achieving MSY has not been evaluated specifically for winter flounder stocks. The SDWG recognized the logical difference between "data-based" inferences involved in estimates of FMSY vs. “hypothesis-based” expectations of inter-stock similarities, based on analogy to justify F40%.

In summary, from a comparative approach to MSY reference points, F40% is similar for all three stocks. The estimate of FMSY for GOM winter flounder is similar to that for the GBK stock but twice that for the SNE/MA stock. This two-fold range in FMSY among the three stocks is due to the differing patterns in the estimated stock-recruitment data (see text table below). The SNE/MA stock has a low steepness estimate that is driven by estimates of strong recruitment and high spawning stock size from the 1980s. Unlike the situation for SNE/MA winter flounder, for GOM and GBK winter flounder there is no pattern in the stock-recruitment estimates that supports inferences of lower steepness. The influence of environmental conditions that limit recruitment success (e.g., warmer temperatures and subsequent larval predation effects) is a possible explanation of the lower steepness of the SNE/MA stock (and subsequently lower FMSY). The SDWG noted that this explanation assumes no local and complete adaptation to environmental conditions among the stocks.

Stock	F_{MSY}	h	SSB_{MSY}	SSB_0	SSB_0/SSB_{MSY}	MSY	F_{40}	SSB_{40}	MSY_{40}
GOM	0.565	0.84	2,167	8,887	4.10	1,152	0.340	3,287	1,080
GBK	0.500	0.85	8,260	31,478	3.81	4,200	0.320	11,300	3,200
SNE/MA	0.310	0.64	33,820	92,657	2.74	9,763	0.327	29,045	8,903

Implications of Reference Point Decisions

Despite the uncertainty in reference point estimation for SNE/MA Atlantic winter flounder, the determination of stock status and rebuilding conclusions are robust. All candidate reference points lead to a conclusion that the stock cannot rebuild to Bmsy by 2014, even at F=0.

Major uncertainty persists in the GOM winter flounder stock assessment, and estimates of current biomass are much greater than all candidate estimates of BMSY or BMSY proxies. However, the relatively low estimates of F and conclusion that overfishing is not occurring are consistent with recent regulations and restrictions on catch. The estimate of SSBMSY corresponding to *h*=0.84 for GOM winter flounder is close to the lower end of the range of past SSB estimates, in contrast to the situation for GBK winter flounder, where it is close to the middle of this range. The minimum observed GOM SSB was 1487 mt, and the 80% confidence interval of SSB_{MSY} is 1640 to 2700 mt. Although the 80% confidence intervals for *h* for each of these two stocks are similar, this feature of the GOM estimates renders them less reliable than those for the GBK stock. While there were disagreements within the SDWG on the BRPs to use as the overfishing definition, the SDWG reached consensus that the current model and associated reference points for Gulf of Maine winter flounder were acceptable and the best that could be determined at this time.

SARC 52 did not accept the SDWG model and the overfished status remains as unknown since biomass based reference points could not be estimated. The SARC accepted a proxy value of the overfishing threshold which was derived from a length-based yield per recruit analysis that assumes all fish above 30 cm are fully recruited to the fishery and that natural mortality is 0.3. Using $F_{40\%}$ (0.31) as a proxy for Fmsy, the threshold exploitation rate is 0.23 and 75%$F_{40\%}$ exploitation was 0.17 with M=0.3. The reference points were converted to exploitation rates to be consistent with the swept area biomass approach (Appendix C1).

TOR 7: Evaluate stock status (overfished and overfishing) with respect to the "new" BRPs (from TOR 6), and with respect to the existing BRPs (from a previous accepted peer review) whose values have been updated.

BRPs for GOM winter flounder from the GARM-III assessment in 2008 (NEFSC CRD08-15) were based on F40%, a proxy for FMSY. SSBMSY and MSY were estimated with the AGEPRO projection model, including the model's CDF of age-1 recruitment and the estimate of F40%. Although BRP's were estimated in GARM-III (F40% = 0.283, SSBMSY = 3,792 mt, and MSY = 917 mt), the GARM-III Review Panel concluded that the assessment did not give "a clear picture of the status of the resource" and that "the proposed analysis could not be used to provide management advice nor stock projections". Therefore, the 2008 assessment was not accepted and the overfished and overfishing status of the GOM winter flounder stock is currently unknown.

Stock status evaluation was consistent regardless of the model formulation (VPA and ASAP) which is summarized in Table C22. Both the split VPA model and the SDWG preferred ASAP multi model indicate that the stock is not overfished and overfishing is not occurring. However spawning stock biomass relative to the SSBmsy varied widely between the VPA and preferred ASAP multi model. SSB in 2010 to SSBmsy ratios varied from the stock recruit Split VPA estimate of 0.52 to the stock recruit estimate of 3.09 from the ASAP multi with no prior on steepness. All models show that fishing mortality in 2010 were well below their respective Fmsy reference points. Fishing mortality in 2010 to Fmsy ratios varied from the stock recruitment split VPA ratio estimate of 0.47 to the stock recruitment estimated ratio of 0.05 from the ASAP multi run with no prior on steepness. The SDWG ASAP multi run using the Fmsy = F40 proxy estimated the SSB2010/SSBmsy ratio at 1.77 and the F2010/F40 at 0.09. The stock recruitment priors did lower the estimated steepness which lowered the SSB2010/SSBmsy ratio to 2.74 and increased the F2010/Fmsy ratio to 0.06.

All GOM winter flounder models have diagnostic issues due to the conflicting signals in the data. The SDWG preferred the ASAP multi model as the best fit to all data sources including considerations for reasonable estimates of biomass in 2009 and 2010 in comparisons to the survey area swept biomass estimates. However the SDWG questioned the feasibility of the estimated SSB relative to the SSBmsy reference points for both the F40 proxy and the stock recruit estimates (1.77 to 2.68). In general the trends and biomass estimated by the model seem appropriate. Surveys and anecdotal feedback from fishermen suggest a shift in the population to deeper water which can help explain the lack of catch in the recreational fishery. However questions remain with the lack of higher catches as the stock rebuilds during the late 1990s and early 2000s when effort in the groundfish fishery was high. In addition, there is little evidence of a change in the size structure or stock range expansion to waters off the coast of Maine which traditionally had higher catches. Considerable uncertainty remains with regards to the comparison of the 2010 SSB relative to the SSBmsy biological reference points.

The SARC concluded that the population models are too uncertain as a based from stock status determination. The overfished status remains as unknown since an analytical model was not accepted and a biomass reference point could not be estimated. The SARC concluded that in 2010 overfishing was not occurring for the stock. A proxy value of the overfishing threshold was derived from a length-based yield per recruit analysis that assumes all fish above 30 cm are fully recruited to the fishery and that natural mortality is 0.3. Using F40% (0.31) as a proxy for Fmsy, the threshold exploitation rate is 0.23. Exploitation rate in 2010 was estimated at 0.03 (80% CI 0.02 - 0.05) which was based on the ratio of 2010 catch (195 mt) to survey swept area estimate of biomass for winter flounder exceeding 30 cm in length (6,341 mt).

TOR 8: Develop and apply analytical approaches and data that can be used for conducting single and multi-year stock projections and for computing candidate ABCs (Acceptable Biological Catch; see Appendix to the TORs) under a set of alternative harvest scenarios. If the stock needs to be rebuilt, take that into account in these projections.

a. Provide numerical short-term projections (3-5 yrs, or through the end of the rebuilding period, as appropriate). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. In carrying out projections, consider a range of assumptions about the most important uncertainties in the assessment (e.g., terminal year abundance, variability in recruitment).

b. Take into consideration uncertainties in the assessment and the species biology to describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming or remaining overfished, and how this could affect the choice of ABC.

c. Develop plausible hypotheses (e.g., mixing among the three stocks) which might explain any conflicting trends in the data and undertake scenario analyses to evaluate the consequences of these alternate hypotheses on ABC determination.

A. Ten year AGEPRO projections assumed that the ACL of 230 mt will be taken in 2011. Projections were done using 1000 bootstrap iterations from the split VPA and 1000 mcmc iterations from the preferred ASAP multi run. Plots with 80 confidence intervals are shown for SSB, catch, and fishing mortality for the split VPA at the Fmsy proxy F40 = 0.43 and 75% of the F40 proxy = 0.32 (Figures C72 and C73). Projections for the ASAP multi model were also run assuming the F40 proxy = 0.34 and 75% of the F40 = 0.26 (Figures C74 and C75). Short term projections using the stock recruit reference point with the prior on steepness for the ASAP multi run were also done at Fmsy = 0.57 and 75%Fmsy = 0.42 (Figures C76 and C77). All projections show relativity high catch in 2012 compared to model time series of catches. The VPA SSB increases towards SSBmsy after low estimates of SSB in 2013 and 2014. The low SSB estimate in 2013 and 2014 is due to the low recruitment estimated in 2009 and 2010 which was influenced by the length based survey conversion. Therefore substantial uncertainty exists with the estimated recruitment in 2009 and 2010. The ASAP multi short term projections result in fishing of the SSB down to SSBmsy. The estimated catch in 2012 shows a large increase relative to the assumed catch in 2011 of 230 mt for both the split VPA and ASAP formulations. The ASAP multi run estimated 2012 catch varies from 1,700 mt from the 75% F40 projection to the stock recruit Fmsy projection estimate of 3,080 mt. However catch declines quickly after 2012 as the stock approaches SSBmsy.

Consideration in the projections could be given to the overestimation of the plus group in the ASAP model. For example a plus group residual adjustment within AGEPRO can be approximated using an assumed plus group discard proportion (Table C25). The effect of the plus group adjustment can be seen in table C25.

B. The Working Group accounted for vulnerability, productivity and susceptibility using conventional MSY reference points, and evaluated uncertainty using model estimates of precision and qualification of other uncertainties. Age-based analytical stock assessment models and associated MSY reference point evaluations provide a relatively comprehensive and synthetic evaluation of vulnerability that is entirely consistent with stock status determination and projection. Vulnerability and susceptibility were accounted for in both aspects of status determination (estimation of F and F_{MSY}) and projections as the magnitude of fishing mortality and recent selectivity at age. All components of productivity (reproduction, individual growth, and survival) were also explicitly accounted for in stock status determination and projections. Reproduction was monitored as age-1 recruitment, and projected as a function of SSB (the product of abundance, weight and maturity at age). Individual growth was monitored as empirical size at age, and projected as recent mean size at age. Survival was accounted for based on model estimates of fishing mortality and selectivity as well as assumed natural mortality, which was informed by tagging analysis. Uncertainties that were not accounted for by assessment and reference point models were evaluated using model diagnostics. Standard model diagnostics (e.g., residual analyses, retrospective analyses) were used for model validation. Retrospective inconsistencies that were outside the bounds of model precision estimates were addressed through selection of alternative models.

Vulnerabilities that were not accounted for from assessment and reference point models were evaluated using exploratory modeling, habitat observations and testing the influence of environmental factors on recruitment dynamics. All three winter flounder stocks are harvested in mixed-stock fisheries, but bycatch and discards are monitored and managed through Annual Catch Limits with Accountability Measures for exceeding those limits.

Additional considerations of vulnerability and productivity are the implications of shifts in distribution, recruitment dynamics and increased natural mortality. Nye et al. (2009) found an annual increase in mean depth of the winter flounder distribution, which may have productivity and vulnerability implications. Apparent decreases in estuarine spawning or shifts toward coastal spawning (e.g., DeCelles and Cadrin 2010) may also have implications for vulnerability (e.g., less availability to recreational fisheries, decreasing vulnerability to that fishery) and productivity (possibly less larval retention). Consumption of winter flounder by other fishes, birds and mammal predators may be increasing as those predator populations increase.

The GOM assessment indicates that the stock is well above B_{MSY} and experiencing low fishing mortality. However, the GOM assessment is the most uncertain of the three (from a “feasibility” perspective, if not from a “statistical precision” perspective). The apparent shift in distribution to deeper habitat may be adding uncertainty to the stock assessment reference points that assume stationarity in vital rates. Therefore, it may be vulnerable to overfishing if managed at a catch level close to the nominally projected catch in the near term.

C. Major conflicting signals exist between the catch at age data and survey data within the modeling work. The split VPA is weighted towards the catch at age information while the preferred ASAP multi run has a greater weight on the survey information. Survey trends may not reflect the population changes in response to the large decline in the catch over time if a greater proportion of the population historically remained within the estuaries in the early 1980s where there is no survey coverage. This hypothesis could possible explain why the survey indices are relatively flat with little apparent response to the change in catch. However there is very limited data on the extent of estuarine residing populations in the 1980s. Therefore this hypothesis remains simply as speculation. The consequences of the split VPA being a better reflection of the true dynamics can be evaluated by assuming the catch or ABC from the preferred ASAP projection as taken within the split VPA projection formulation. Figure C78 is an example of the consequence of the split VPA model being true and assuming the catch from the 75% Fmsy ASAP multi projection.

Tor 9: Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in recent SARC reviewed assessments and review panel reports. Identify new research recommendations.

SDWG SAW 52 New Research Recommendations

1) Update and investigate migration rates between stock and movement patterns. The most recent comprehensive tagging study was completed in the 1960s (Howe and Coates), and a new large scale effort is warranted. Further investigate localized structure/genetics within the stocks.

2) Investigate the feasibility of port samplers collecting otoliths from large and lemon sole instead of scales because of problems under-ageing larger fish.

3) Investigate use of periodic gonad histology studies as a check to make ensure maturity estimates are accurate, with particular attention to obtaining sufficient samples from the Georges Bank stock.

4) Investigate the skipped spawning percentage for each stock, and estimate interannual variation when sufficient data have been collected.

5) Investigate ways to improve compliance to help VTR reporting. Currently about 300 of the 1500 permitted vessels consistently under-report the number of statistical area fished.

6) Encourage support for Industry Based Surveys, which can provide valuable information on stock abundance, distribution, and catchability in research surveys that is independent of and supplemental to NMFS efforts.

7).Explore use of a more complex Stock Synthesis model with small rates of migration between stocks.

8) Develop time series of winter flounder consumption by the major fish predators of winter

flounder.

9) Conduct studies to better understand recruitment processes of winter flounder, particularly in the GOM and on GBK.

10) Revise the NEFSC assessment software to include the ability to model S-R functions including environmental factors with errors/probabilities.

11) Further explore the relationship between large scale environmental forcing (e.g., temperature, circulation, climate) for effects on life history, reproduction, and recruitment in the Georges Bank stock.

12) Explore development of an index of winter flounder larval abundance based on MARMAP, GLOBEC, etc. time series.

Research Recommendations from GARM III

Assessment approaches needs to be explored that consider all three Winter Flounder stocks as a stock complex within which there is significant interaction amongst the individual stock components.

Working paper addressed by Terceiro MS (2011.) examined

Research Recommendations from SARC 36

1) The MADMF fall survey does collect winter flounder otoliths and scales, so ageing such material should be undertaken.

The MADMF fall survey has not been aged.

2) Increase the number of tows and/or consistently sample inshore strata in the NEFSC bottom trawl survey.

The number of tows in inshore Massachusetts strata conducted with the RV Bigelow starting in 2009 has increased from 1-2 tows to about 2-3 tows per strata with the exception of the fall 2010 survey which lacked sampling in Cape Cod bay. In addition stratum 64 appears to be more consistently sampled with the RV Bigelow and could possibly be included in the index in the future. However depth constraints prevents the sampling of stratum 58.

3) Increase MRFSS length sampling intensity in the recreational fishery.

Length sampling of the winter flounder B2 catch now occurs in the recreational fishery and was used in this assessment.

4) Increase temporal and market category coverage of length sampling in the commercial landings.

Biological length sampling in the ports has improved for all species. However the decline in commercial landings for Gulf of Maine winter flounder has made length sampling coverage by market category difficult. Unclassified sampling in the ports and observer sampling of the kept fish appears to have provided an adequate characterization of the size structure. However, regardless of increased observer coverage in 2010 the length sampling appears to have suffered from the decline in landings in 2010.

5) Increase the intensity of observer sampling especially with small- and large-mesh trawl gear.

Observer sampling has improved in the small and large mesh trawl fishery.

6) Examine the sources of discrepancy between NEFSC and MA survey maturity estimates.

Reasons for the discrepancy between NEFSC and MA survey maturity was examined by McBride et a.l MS 2011.

7) Initiate periodic maturity staging workshops, involving State and NEFSC trawl survey staff.

A maturity staging workshop was done with state and NEFSC staff. Education on how to stage maturity for winter flounder will need to continue as an ongoing process in the maturity workshops.

8) Incorporate the results from the MEDMR research trawl survey (begun in 2001) into the assessment as they become available.

Preliminary ME/NH survey winter flounder age data was examined by the SAW 52 SDWG. The ME/NH survey was included in the area swept estimates of 30+ biomass for this assessment.

9) Investigate derivation of stock-specific parameters for the next assessment.

It is not entirely clear on the intension of this research recommendation. Sensitivity of the assumed natural mortality was explored in this assessment.

10) Attempt use of a forward projection (statistical catch at age model) in the next assessment.

The forward projection ASAP model was developed and used in this assessment.

Research Recommendations prior to SARC 36

1) Examine the implications of anthropogenic mortalities caused by pollution and power plant entrainment in estimating yield per recruit, if feasible.

This research recommendation was not done. It is not clear how this research *recommendation could be addressed.*

2) Examine growth variations within the Gulf of Maine, using results from the Gulf of Maine Biological Sampling Survey (1993-1994).

This research recommendation is perhaps not needed with the aging of the relatively new ME/NH survey.

3) Further examine the stock boundaries to determine if Bay of Fundy winter flounder should be included in the Gulf of Maine stock complex.

This research recommendation has not been done. The Bay of Fundy seems to be an appropriate natural break for the stock complex. See working paper by DeCelles MS 2011.

References

DeCelles GR, Cadrin SX. MS 2011. An interdisciplinary assessment of winter flounder (*Pseudopleuronectes americanus*) stock structure. Stock Assessment Workshop Southern Demersal Working Group 52 Working Paper 1.

DeCelles GR and Cadrin SX. MS 2010. Movement Patterns of winter flounder(*Pseudopleuronectes americanus)* in the southern Gulf of Maine: observations with the use of passive acoustic telemetry. Fish. Bull., 108: 407-419.

Hare JA. MS 2011. Development of environmentally-explicit stock-recruitment models for three stocks of winter flounder (*Pseudopleuronectesamericanus*) along the northeast coast of the United States. Stock Assessment Workshop Southern Demersal Working Group 52 Working Paper 13.

Hare JA, Alexander M, Fogarty MJ, Williams E, Scott J. 2010. Forecasting the dynamics of a coastal fishery species using a coupled climate-population model. Ecological Applications 20:452-464.

Howell, P., A. Howe, M. Gibson and S. Ayvasian. 1992. Fishery management plan for inshore stocks of winter flounder. Atlantic States Marine Fisheries Commission. Fisheries Management Report No. 21. May, 1992.

Mayo, R.K., L. O'Brien, and N. Buxton. 1992. Discard estimates of American plaice, *Hippoglossoides platessoides*, in the Gulf of Maine northern shrimp fishery and the Gulf of Maine-Georges Bank large-mesh otter trawl fishery. SAW 14 Res. Doc. 14/3. 40 pp.

McBride RS, Wuenschel MJ, McElroy D, Rowinski Y, Thornton G, Nitschke P. MS 2011. Classifying winter flounder maturity during NEFSC resource surveys: comparing at-sea, macroscopic classifications with results from a gonad histology method. Stock Assessment Workshop Southern Demersal Working Group 52 Working Paper 8.

Miller TJ. MS 2011. Winter flounder length-based survey calibration. Stock Assessment Workshop Southern Demersal Working Group 52 Working Paper 7.

Miller TJ, Das C, Politis PJ, Miller AS, Lucey SM, Legault CM, Brown RW, Rago PJ. 2010. Estimation of Albatross IV to Henry B. Bigelow calibration factors. NEFSC Ref Doc. 10-05. 233 p.

Myers RA, Bowen KG, Barrowman NJ. 1999. Maximum reproductive rate of fish at low population sizes. Can. J. Fish. Aquat. Sci. 56: 2404-2419.

NEFSC. 2003 Report of the 36th Northeast Regional Stock Assessment Workshop (36th SAW): Stock Assessment Review Committee (SARC) consensus summary of assessments. Ref. Doc 03-06. February 2003.

NEFSC. 2005. Assessment of 19 Northeast groundfish stocks through 2004. 2005 Groundfish Assessment Review Meeting (2005 GARM), Northeast Fisheries Science Center, Woods Hole, Massachusetts. 15-19 August 2005. Northeast Fish. Sci. Cent. Ref. Doc. 05-13. 508 p.

Palmer MC, Wigley SE. MS 2011. Using positional data from vessel monitoring systems (VMS) to validate the logbook-reported area fished and the stock allocation of commercial fisheries landings, 2004-2008. Stock Assessment Workshop Southern Demersal Working Group 52 Working Paper 9.

Rademeyer RA, Butterworth DS. MS 2011. Initial applications of statistical catch-at-age assessment methodology to the Gulf of Maine winter flounder resource. 28 p.

Terceiro M. MS 2011. Response to SNE/MA winter flounder TOR 4. Stock Assessment Workshop Southern Demersal Working Group 52 Working Paper 5.

Wigley SE, Palmer MC, Blaylock J, Rago PJ. 2008. A brief description of the discard estimation for the National Bycatch Report. NEFSC Ref. Doc. 08-02. 35 p.