A. SOUTHERN NEW ENGLAND / MID-ATLANTIC (SNE/MA) WINTER FLOUNDER STOCK ASSESSMENT FOR 2011

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

SAW 52 Terms of Reference

A. Winter flounder (Southern New England 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 B_{MSY}, $B_{THRESHOLD}$, and F_{MSY}) 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

The Southern Demersal Working Group (SDWG) met in April and May of 2011 to develop stock assessments for the Southern New England/Mid-Atlantic (SNE/MA) stock of winter flounder. The SDWG met within the process of the Northeast Regional SAW 52 and addressed nine Terms of Reference, as follows:

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

Commercial fishery landings reached an historical peak of 11,977 metric tons (mt) in 1966, then decreased through the 1970s, peaked again at 11,176 mt in 1981, and then steadily decreased to 2,128 mt in 1994. Commercial landings then increased to 4,556 mt in 2001 and then decreased to only 174 mt in 2010. The Proportional Standard Error (PSE) of commercial landings has averaged less than 1%. Recreational fishery landings peaked in 1984 at 5,510 mt but decreased thereafter, with only 28 mt estimated for 2010. The PSE of the recreational landings has averaged about 27%. Commercial fishery discards for 1981 to 1993 were estimated from length frequency data from the NEFSC and MADMF trawl surveys, commercial port sampling of landings at length and Fishery Observer sampling of landings and discard at length. The Standardized Bycatch Reporting Method (SBRM) has been used for estimation of SNE/MA winter flounder commercial fishery discards for 1994 and later years. Commercial fishery discard losses peaked in the early 1980s at 1,000-1,500 mt per year and have decreased to less than 200 mt per year since 1997. A discard mortality rate of 50% was applied to the commercial live discard estimates. The PSE of the commercial fishery discards has averaged 27%. Recreational fishery discard losses peaked in 1984-1985 at about 700,000-750,000 fish or 150-200 mt and then decreased to less than 100,000 fish or 20 mt per year since 2000. A discard mortality rate of 15% was applied to recreational live discard estimates. The PSE of the recreational discards has averaged 30%.

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 NEFSC winter, spring and fall bottom trawl surveys provided long time series of fishery-independent indices for SNE/MA winter flounder. The strata set defined for SNE/MA winter flounder was revised in this assessment to use a consistently sampled strata set over the historical time series and into the future. NEFSC indices generally increased from a low point in the early to mid-1970s to a peak by the early 1980s. NEFSC survey indices reached near- or record low levels in the late 1980s-1990s. Indices from the three survey series generally increased during the late 1990s, but have since decreased again. The Fisheries Survey Vessel (FSV) Albatross IV (ALB) was replaced in spring 2009 by the FSV Henry B. Bigelow (HBB) as the main platform for NEFSC research surveys, including the spring and fall bottom trawl surveys. Calibration experiments to estimate these differences in fishing power between the vessels were conducted and peer-reviewed. Length-based calibration models were used to express 2009-2010 NEFSC indices in ALB units.
Several state survey indices were available to characterize the abundance of SNE/MA winter flounder. The Massachusetts Division of Marine Fisheries (MADMF) spring, Rhode Island Division of Fish and Wildlife (RIDFW) spring, University of Rhode Graduate School of Oceanography

(URIGSO), Connecticut Department of Environmental Protection (CTDEP) Long Island Sound Trawl Survey, and the New Jersey Division of Fish, Game and Wildlife (NJDFW) ocean and rivers research surveys provided indices of abundance at age used in the assessment. Numerous state recruitment surveys (MADMF, RIDFW, CTDEP, New York Department of Environmental Conservation (NYDEC), NJDFW, Delaware Division of Fish and Wildlife (DEDFW)) were also considered.

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 2011 SAW 52 assessment indicates that during 1981-1993, fishing mortality (fully recruited F, ages 4-5) varied between 0.61 (1982) and 0.95 (1993) and then decreased to 0.47 by 1999. Fishing mortality then increased to 0.70 by 2001, and has since decreased to 0.051 in 2010, generally tracking the decrease in fishery catch. SSB decreased from 20,100 mt in 1982 to a record low of 3,900 mt in 1993, and then increased to 8, 900 mt by 2000. SSB has varied between 4,500-8,000 mt during 2001-2009, and was 7,076 mt in 2010. Recruitment at age 1 decreased nearly continuously from 71.6 million age-1 fish in 1981 (1980 year class) to 7.5 million fish in 2002 (2001 year class). Recruitment has averaged 10.5 million during 2003-2010. The fishery selectivity pattern in the first time block (1981-1993) was estimated to be 0.01 at age 1, 0.24 at age 2, 0.75 at age 3, was fixed at 1.00 at age 4, was estimated at 1.00 at age 5, 0.99 at age 6, and 1.00 at age 7+. The pattern in the second time block (1994-2010) was estimated to be 0.01 at age 1, 0.19 at age 2, 0.70 at age 3, was fixed at 1.00 at age 4, was estimated at 0.97 at age 5, 0.89 at age 6, and 0.67 at age 7+.

The precision of the 2010 stock size at age, F at age and SSB was evaluated using MCMC techniques. There is an 80% probability that fully recruited F for ages 4-5 in 2010 was between 0.04 and 0.06. There is an 80% probability that SSB in 2010 was between 6,433 mt and 8,590 mt. Retrospective analysis for the 2003-2010 terminal years indicates retrospective error in fishing mortality (F) ranged from -38% in 2006 to -13% in 2009, retrospective error in SSB ranged from +42% in 2004 to +12% in 2009, and retrospective error in recruitment at age 1 (R) ranged from +78% in 2005 (2004 year class) to -11% in 2009 (2008 year class).

For the NEFSC Spring, Fall, and Winter surveys expressed as swept area numbers, aggregate survey catchability (q) was estimated at 0.126, 0.617, and 0.253, respectively. The other calibration surveys are of more limited geographic extent and were input in their original units, and therefore q estimates for those surveys ranged from 0.00001 (MADMF summer seine survey age 0 index) to 0.0017 (CTDEP trawl survey). A comparison between the results of the current assessment and the five previous assessments, or "historical retrospective," illustrates the underestimation of fishing mortality and overestimation of SSB that had been present between assessments since 1995. This pattern is in addition to the persistent "internal retrospective" that has been present in each of the assessments. The SDWG notes that the current assessment with assumed M = 0.3 is not consistent with those previous which assumed M = 0.2, and that much of the upward magnitude shift in numbers and biomass and downward shift in fishing mortality is due to this change.

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. The SDWG developed such an exercise using the 2008 GARM-III assessment data and ADAPT VPA model in an initial response to TOR4 and concluded that the application of a annually varying "bias-correction" in one direction in such an exercise provides stock size estimates and BRPs that scale up or down by about the same average magnitude as the gain or loss. After evaluation of the first exercise, the SDWG concluded that the calculated variance of the area-allocated commercial landings likely underestimates the true error. More 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). The SDWG elected to update the exercise using the final SNE/MA assessment ASAP model, with an additional 5% PSE in commercial landings added to the currently estimated 0.4 to 4.5% over the 1995-2010 time series. This increased the average commercial landings PSE from 0.9% to 3.7%, and increased the overall catch PSE from 8% to 10%, ranging from 4.9% in 1992 to 23.7% in 2010. 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. As in the previous version of the exercise, the application of a annually varying "bias-correction" in one direction in such an exercise provides stock size estimates that scale up or down by about the same average magnitude as the gain or loss. For the final ASAP mode, fishing mortality on average changed by +/- 0.3%, and the range in 2010 F was 0.05 to 0.04, comparable to the MCMC estimate of uncertainty. SSB on average changed by +/- 9.0%, and the range in 2010 SSB was 6,500 to 7,600 mt, within the MCMC estimate of uncertainty.

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).

Winter flounder spawn in winter and early spring in estuaries along the mid-Atlantic, southern New England and Gulf of Maine, as well as in continental shelf waters on Georges Bank. In southern New England, Manderson (2008) found that overall recruitment was linked to spring temperatures, presumably by acting on larvae, settlement stage, and/or early juveniles. Further, Manderson (2008) found that young-of-the-abundance among 19 coastal nurseries became more synchronized in the early 1990's and argued that increased frequency of warm springs was creating coherence in early life stage dynamics among local populations.

The best fit environmentally-explicit stock recruitment relationship for the Southern New England stock predicted higher recruitment at lower winter air temperatures. The variable in the best model was Southern New England air temperature in January and February. The best environmentally-model provided a similar function to the standard model at mean environmental conditions, but importantly the predicted asymptotic recruitment was lower with the environmental model. The environmentally-explicit models support the hypothesis that increased temperatures during spawning and the early life history result in decreased recruitment in the SNE/MA stock. 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 B_{MSY}, $B_{THRESHOLD}$, and F_{MSY}) 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.

FMSY, SSBMSY, and MSY BRPs from an external stock-recruitment model and proxy BRPs based on 40% MSP were estimated. For the final assessment model, the stock-recruitment model with a fixed value for steepness (h=0.61) was judged to fit best while providing feasible results. FMSY is estimated to be 0.290; SSBMSY is estimated to be 43,661 mt; MSY is estimated to be 11,728 mt; F40% is estimated to be 0.327; SSB40% is estimated to be 29,045 mt; MSY40% is estimated to be 8,903 mt.

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 Southern New England/Mid-Atlantic (SNE/MA) winter flounder stock complex was overfished but overfishing was not occurring in 2010. Fishing mortality (F) in 2010 was estimated to be 0.051, below FMSY = 0.290 (18% of FMSY) and below F40% = 0.327 (16% of F40%). SSB in 2010 was estimated to be 7,076 mt, about 16% of SSMSY= 43,661 mt and 24% of SSB40% = 29,045 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).

Catch of 842 mt in 2011 is projected to provide median F2011 = 0.100 and median SSB2011 = 9,177 mt. Projections at F = 0.000 in 2012-2014 indicate less than a 1% chance that the stock will rebuild to SSBMSY = 43,661 mt by 2014.

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. A considerable source of additional vulnerability is the continued weak recruitment and low reproductive rate (e.g., recruits per spawner) of SNE/MA winter flounder. If weak recruitment and low reproductive rate continues, productivity and rebuilding of the stock will be less than projected. Stock-recruit modeling suggests that warm temperatures are having a negative effect on recruitment of SNE/MA winter flounder.

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.

The SDWG has initiated further research pursuing use of a more complex model (i.e., Stock Synthesis) to maintain separate fishery and survey catch for the three current stock units, while allowing a small amount (a few percent) of exchange between the stock units based on information from historical tagging. However, development of that research has not progressed sufficiently to be made available for peer review at this time.

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.

Twelve of the previous 16 research recommendations have been addressed in full or in part. Four have not been addressed. Twelve new research recommendations have been developed by the SDWG for SAW52.

INTRODUCTION

Stock Structure

Winter flounder (*Pseudopleuronectes americanus*) is a demersal flatfish species commonly found in North Atlantic estuaries and on the continental shelf. The species is distributed between the Gulf of St. Lawrence, Canada and North Carolina, U.S., although it is not abundant south of Delaware Bay. Boundaries for four stock units were originally defined in the Atlantic States Marine Fisheries Commission (ASMFC) management plan (Howell et al. 1992): Gulf of Maine (GOM), Georges Bank (GBK), Southern New England (SNE; waters from coastal Massachusetts to eastern Long Island, New York), and Mid-Atlantic (MA; western Long Island, New York , New Jersey, and Delaware waters). A review of tagging studies for winter flounder for the 1995 SAW 21 assessment (Shepherd et al. 1996; NEFSC 1996) indicated that mixing has occurred among the Southern New England and Mid-Atlantic populations. Shepherd et al. (1996) noted that differences in growth and maturity among samples from Southern New England to the Mid-Atlantic could reflect discrete sampling along a gradient of changing growth and maturity rates over the range of a stock complex. Differences in growth rates within the Mid-Atlantic unit were observed to be greater than differences between Mid-Atlantic and Southern New England units (Shepherd et al. 1996). Therefore, since the 1995 SAW 21 assessment (NEFSC 1996), winter flounder populations in the Southern New England and Mid-Atlantic regions have been combined into a single stock complex for assessment purposes. Winter flounder in U.S waters are currently managed as three stock units: Gulf of Maine (GOM), Georges Bank (GBK), and Southern New England/Mid-Atlantic (SNE/MA; Figure A1). Within the SNE/MA stock complex, winter flounder undergo migrations from estuaries, where spawning occurs in the late winter and spring, to offshore shelf areas of less than 60 fathoms (110 meters).

Tagging studies (e.g., Howe and Coates 1975) indicate that there is limited mixing of fish among the three current stock units, with about 1%-3% between the GOM and SNE/MA, about 1% between GBK and SNE/MA, and <1% between GOM and GBK. Meristics studies based mainly on fin ray counts also indicate a separate GBK stock (Kendall 1912; Perlmutter 1947) or separate GOM, GBK, and SNE stocks (Lux et al. 1970; Pierce and Howe 1977). Growth and maturity studies also support the distinction of at least three stock areas (Lux et al. 1970; Howe and Coates 1975; Witherell and Burnett 1993), with GBK fish growing and maturing the fastest and GOM fish the slowest.

An interdisciplinary review of U.S. winter flounder stock structure was conducted for this assessment (DeCelles and Cadrin MS 2011). Information on morphology, tagging studies, genetics, larval dispersal, life history traits, environmental signals and meristics was considered. This work found "contingent groups" (localized populations) are likely present in several regions, and their coherence merits further research. Despite evidence for local population structure, information from tagging, meristics, and life history studies suggest extensive mixing within the current stock units, thereby supporting the current assessment and management structure.

The SNE/MA stock complex extends from the coastal shelf east of Provincetown, MA southward along the Great South Channel (separating Nantucket Shoals and Georges Bank) to the southern geographic limits of winter flounder off Delaware. Northeast Fisheries Science Center (NEFSC) commercial fishery statistical areas within this boundary are 521, 526, 533-539, and 611-639 (Figure A1). The corresponding recreational fishery areas are southern Massachusetts (the southern half of

Barnstable County; Dukes, Nantucket and Bristol counties), Rhode Island, Connecticut, New York, New Jersey, Delaware, Maryland and Virginia. NEFSC survey strata included for this stock extend from the waters of outer Cape Cod to the south and west, and include offshore strata 1-2, 5-6, 9-10, 25, 69-70 and 73-74 and inshore strata 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 45, 46 and 56.

Assessment History

The initial analytical assessment of the SNE/MA stock complex of winter flounder was completed in 1995 at SAW 21 (NEFSC 1996). The SAW 21 assessment included fishery catches through 1993, research survey abundance indices through 1995, catch at age analyzed by Virtual Population Analysis (VPA) for 1985-1993, and biological reference points based on Yield and Spawning Stock Biomass (SSB) per recruit models (Thompson and Bell 1934). The 1995 SAW 21 assessment concluded that the stock complex was over-exploited and at a record low level of SSB. SSB in 1993 was estimated to be 3,792 mt, about 11% of the Maximum Spawning Potential (MSP), and the fully recruited fishing mortality rate on ages 4-5 in 1993 was estimated to be F = 0.83, about four times F40% = 0.21.

The next benchmark assessment of the SNE/MA stock complex of winter flounder was completed in 1998 at SAW 28 (NEFSC 1999). The SAW 28 assessment included fishery catches through 1997, research survey abundance indices through 1998, catch at age analyzed by VPA for 1981-1997, and biological reference points based on a production model conditioned on VPA results. The 1998 SAW 28 assessment concluded that the stock complex was fully exploited and at a medium level of biomass. Total Stock Biomass (TSB) in 1997 was estimated to be 17,900 mt, about 64% of BMSY = 27,810 mt, and the fishing mortality rate on ages 4-5 in 1997 was estimated to be F = 0.31, just above F40% = 0.21, while the total biomass weighted F was 0.24, below FMSY = 0.37.

A benchmark assessment was completed in 2002 at SAW 36 (NEFSC 2003). The SAW 36 assessment included fishery catches through 2001, research survey abundance indices through 2002, and catch at age analyzed by VPA for 1981-2001. Biological reference points were based on stock-recruitment modeling conducted by the 2002 Working Group on Re-estimation of Biological Reference points for New England Groundfish (NEFSC 2002), which indicated that FMSY = 0.32, SSBMSY = 30,100 mt, and MSY = 10,600 mt. The SAW 36 assessment concluded that the stock complex was overfished and that overfishing was occurring. The SSB in 2001 was estimated to be 7,600 mt, about 25% of SSBMSY = 30,100 mt. The fishing mortality rate in 2001 was estimated to be F = 0.51, about 60% above FMSY = 0.32. The 2002 SAW 36 Review Panel noted that the 2002 assessment provided a much more pessimistic evaluation of stock status than the 1998 SAW 28 assessment, mainly due to the retrospective pattern of underestimation of F and overestimation of SSB during the late 1990s.

An updated assessment was completed in 2005 at GARM2 (NEFSC 2005). The GARM2 assessment included fishery catches through 2004, research survey abundance indices through 2005, catch at age analyzed by VPA for 1981-2004, and biological reference points based on the NEFSC (2002) stock-recruitment model. The 2005 GARM2 assessment concluded that the stock complex was overfished and that overfishing was occurring. The SSB in 2004 was estimated to be 3,938 mt, about 13% of SSBMSY= 30,100 mt. The fishing mortality rate in 2004 was estimated to be F = 0.38, about 19% above FMSY = 0.32. The GARM2 Review Panel noted that the VPA exhibited a severe

retrospective pattern of underestimation of F and overestimation of SSB during the late 1990s and into 2001.

The most recent benchmark assessment was completed in 2008 at GARM-III (NEFSC 2008). The GARM-III assessment included fishery catch through 2007, research survey abundance indices through 2008, and catch at age analyzed by VPA for 1981-2007. The 2008 GARM-III Review Panel concluded that the "Base" VPA exhibited such a large retrospective pattern through the late 1990s and into 2001 that it required an adjustment. Splitting the time series of research survey data used in calibration was proposed to act as a proxy for fishery and biological factors that could have changed in the mid-1990s, resulting in the observed retrospective pattern. The VPA with most survey time series split at 1993/1994 appeared to reduce the retrospective pattern and this "Split" VPA was accepted as the best available estimate of stock status and a sufficient basis for management advice. Biological reference points were based on the non-parametric empirical Yield and SSB per recruit approach, which indicated that FMSY = F40% = 0.248, SSBMSY= SSB40% = 38,761 mt, and MSY = 9,742 mt. The 2008 GARM-III assessment concluded that the stock complex was overfished and that overfishing was occurring. The SSB in 2007 was estimated to be 3,368 mt, about 9% of SSBMSY= 38,761 mt. The fully recruited fishing mortality rate in 2007 was estimated to be F = 0.649, over twice FMSY= F40%= 0.248.

This 2011 SAW 52 benchmark assessment of the SNE/MA stock complex of winter flounder includes fishery and research survey catch through 2010.

Fisheries Management

Current management of the fisheries for winter flounder is coordinated by the Atlantic States Marine Fisheries Commission (ASMFC) in state waters and the New England Fishery Management Council (NEFMC) in federal waters. Winter flounder fisheries in state waters have been managed by Interstate Agreement under the auspices of the ASMFC Fishery Management Plan (FMP) for Inshore Stocks of Winter Flounder since 1992. The plan includes states from Delaware to Maine, with Delaware granted *de minimus* status (habitat regulations applicable but fishery management not required). Coastal states from New Jersey to New Hampshire have promulgated a broad suite of indirect catch and effort controls. State agencies have set minimum size limits for recreationally and commercially landed flounder, enacted limited recreational closures and bag limits, and instituted seasonal, areal, or state-wide commercial landings and fishing gear restrictions.

Winter flounder fisheries in the Exclusive Economic Zone (EEZ) are managed under the Northeast Multispecies Fishery FMP initially developed by the NEFMC in 1986. The principle catch of winter flounder in the EEZ has recently occurred as bycatch in directed trawl fisheries for Atlantic cod, haddock, and yellowtail flounder. The management unit encompasses the multispecies finfish fishery that operates from Maine through Southern New England. The FMP extends authority over vessels permitted under the FMP even while fishing in state waters if federal regulations are more restrictive than the state regulations. The initial FMP enacted codend minimum mesh size regulations, closed areas and seasons for haddock and yellowtail flounder, and an Exempted Fisheries Program allowing targeting of small-mesh species such as shrimp, dogfish, or whiting. In Southern New England waters, the groundfish bycatch on vessels fishing with small mesh was not limited in any way. There was an 11 inch (28 cm) minimum size for winter flounder which corresponded with the length at first capture (near zero percent retention) for 5.5 inch (140 mm) diamond mesh. Although the FMP was amended four times by 1991, it was widely recognized that many stocks, including winter flounder, were being overfished.

Time-specific stock rebuilding schedules were part of FMP Amendment 5 which took effect in May 1994. The rebuilding fishing mortality target for winter flounder was achievement of F20% within 10 years. Along with a moratorium on issuance of additional vessel permits, the cornerstone of Amendment 5 was an effort reduction program that required "large-mesh" groundfish vessels to limit their Days At Sea (DAS). There was an exemption from effort reduction requirements for vessels less than 45 feet in length and for "day boats." Vessels retaining more than the possession limit of groundfish (10% by weight, up to 500 lbs) were required to fish with either 5.5 inch (140 mm) diamond or square mesh in Southern New England or 6 inch (152 mm) mesh throughout the net in the regulated mesh area of Georges Bank-Gulf of Maine. The possession limit was allowed when using small mesh within the western Gulf of Maine (except for Jeffreys Ledge and Stellwagen Bank) and in Southern New England. Vessels fishing in the EEZ west of 72° 30' (the longitude of Shinnecock Inlet, NY) were required to abide by 5.5 inch (140 mm) diamond or 6 inch (152 mm) square codend mesh size restrictions consistent with the Summer Flounder FMP. The minimum landed size of winter flounder increased to 12 inches (30.5 cm), appropriate for the increased mesh size in order to reduce discards.

At the end of 1994, the NEFMC reacted to collapsed stocks of Atlantic cod, haddock, and yellowtail flounder on Georges Bank by recommending a number of emergency actions to tighten existing regulations to reduce fishing mortality. Prime fishing areas on Georges Bank (Areas I & II) and in the Nantucket Lightship Area were closed. The NEFMC also addressed an expected re-direction of fishing effort into Gulf of Maine and Southern New England waters while also developing Amendment 7 to the FMP. Under FMP Amendment 7, DAS controls were extended, and any fishing by an EEZ-permitted vessel required use of not less than 6 inch (152 mm) diamond or square mesh in Southern New England east of 72° 30'. Framework 27 in 1999 increased the square mesh minimum size to 6.5 inches (165 mm) in the Gulf of Maine, Georges Bank, and Southern New England mesh areas. FMP Amendment 9 revised the overfishing definitions for SNE/MA winter flounder as recommended by SAW 28 (NEFSC 1999).

During 2004-2009, formal rebuilding programs for many multispecies stocks, including winter flounder, were adopted to meet the requirements of the Magnuson-Stevens Act. The DAS allocations were reduced in 2004, 2006, and 2009 (FMP Amendment 13 and Framework 42). "Hard" (as opposed to target) quotas were adopted for a few programs and a few management units, although GBK yellowtail flounder was the only stock with a hard quota for all fishing.

The regulations of FMP Amendment 16 and Framework 44 were implemented in 2010, and the associated catch share program has resulted in most of the multispecies fishery being subject to hard quotas. A key component of the Amendment 16 catch share program was the formation of voluntary, self-selecting fishing organizations identified as "sectors." For SNE/MA winter flounder, Amendment 16 revised the overfishing definitions as recommended by the GARM-III (NEFSC 2008), established a target rebuilding date of 2014 under a target fishing mortality rate of F = 0.0, established an expected rebuilding date of 2017 given likely Fs, and specified Annual Catch Limits (ACLs) and Accountability Measures (AMs). Although the specified fishing mortality rate target for SNE/MA winter flounder for 2010-2012 is F = 0.0, and possession by federally permitted vessels is prohibited, the NEFMC and NMFS recognized that an incidental bycatch would be unavoidable. Framework 44 therefore established ACLs for SNE/MA winter flounder using the F expected to result from management measures designed to achieve F = 0.0, providing ACLs for the 2010-2012 Fishing Years (beginning May 1) of 605, 842, and 1125 metric tons.

Growth and Maturity

Winter flounder in the Gulf of Maine and Southern New England reach a maximum size of around 2.25 kg (5 pounds) and 60 cm. On Georges Bank fish may reach a maximum length of 70 cm and weight up to 3.6 kg (8 pounds; Bigelow and Schroeder 1953). An updated compilation and analysis of the NEFSC and Massachusetts Division of Marine Fisheries (MADMF) survey growth and maturity data for 1976-2010 for this assessment indicated the following maximum age, maximum length, and von Bertalanffy growth parameters that generally support the current stock structure (Figure A2):

GOM: 16,010 fish, maximum age 15 (55 cm); maximum length 61 cm;
Linfinity = 46.4 cm, k = 0.2727

GBK: 6,311 fish, maximum age 18 (50 cm), maximum length 70 cm;
Linfinity = 57.9 cm, k = 0.2829

SNE: 23,593 fish, maximum age 16 (51 cm), maximum length 60 cm;
Linfinity = 46.5 cm, k = 0.3184

The 1998 SAW 28 (NEFSC 1999) and previous assessments had used the maturity schedule as published in O=Brien et al. (1993) for winter flounder south of Cape Cod, based on data from the MADMF spring trawl survey for strata 11-21 (state waters east of Cape Cod, Nantucket sound, Vineyard Sound, and Buzzards Bay) sampled during 1985-1989 (n = 301 males, n = 398 females). Those data provided estimates of lengths and ages of 50% maturity of 29.0 cm and 3.3 yr for males, and 27.6 cm and 3.0 yr for females, and the following estimated proportions mature at age. The

female schedule (with the proportion at age 2 rounded down to 0.00) was used in the SAW 28 assessment (NEFSC 1999).

Age	1	2	3	4	5	6	7+
Males	0.00	0.04	0.32	0.83	0.98	1.00	1.00
Females	0.00	0.06	0.53	0.95	1.00	1.00	1.00

In the 1998 SAW 28 review of the SNE/MA winter flounder stock assessment (NEFSC 1999), the SAW recommended re-examination of the maturity schedule used in the yield per recruit analysis (YPR) and VPA to incorporate any recent research results in the next assessment. In response to the SAW 28 recommendation, the 2002 SAW 36 (NEFSC 2003) examined NEFSC spring trawl survey data for the 1981-2001 period in an attempt to better characterize the maturity characteristics of the SNE/MA winter flounder stock complex. Data from the NEFSC survey included those judged in the SAW 28 assessment to comprise the SNE/MA complex from Delaware Bay to Nantucket Shoals: NEFSC offshore strata 1-12, 25 and 69-76, and inshore strata 1-29, 45-56. This was a much larger geographic area than that included in the MADMF survey data used in O=Brien et al. (1993). Data were analyzed in 5-6 year blocks (1981-1985, 1986-1990, 1991-1995, and 1996-2001) and for the entire time period (1981-2001), for each sex and combined sexes. Observed proportions mature at age were tabulated, and from those data maturity ogives at length and age were calculated to provide estimated proportions mature at age.

In general, the 2002 SAW 36 examination of the NEFSC maturity data indicated earlier maturity than the MADMF data, with L50% values ranging from 22-25 cm, rather than from 28-29 cm, and with ~50% maturity for age 2 fish, rather than ~50% maturity for age 3 fish. To investigate the apparent inconsistency between the MADMF and NEFSC maturity data, the two data sets were further compared over the same time periods (1985-1989, 1990-1995, 1996-2001) for common/adjacent survey strata (MADMF strata 11-12; NEFSC inshore strata 50-56 and offshore strata 10-12 and 25). For comparable time periods and geographic areas, the NEFSC maturity data still consistently indicated a smaller size and younger age of 50% maturity than the MADMF data. NEFSC L50% and A50% values ranged from 22-26 cm and about 2.0 yr, while the MADMF values ranged from 27-30 cm and about 3.0 yr. The difference in values from this comparison was not as large as for the full NEFSC data set extending southward to Delaware Bay, which incorporates components of the stock complex that mature at smaller sizes and younger ages. However, the difference was still nearly a full age class difference at 50% maturity.

Given that both length and age varied in the same direction, it seemed unlikely that the differences could be attributed to aging differences between the two data sets. Since the MADMF and NEFSC geographic areas in this comparison did not match exactly, the difference in maturity rates may have been due to the extension of the NEFSC strata to somewhat deeper waters inhabited by fish that mature at a smaller size and younger age (inclusion of fish in offshore strata were necessary for sufficient sample size). Alternatively, for the size range of fish in question (20 to 30 cm length), it might have been that immature and mature fish are segregated by area, with mature fish in that size interval tending to occupy inshore areas during the spring, with immature fish tending to remain offshore. Finally, there may have been differences in the accuracy and consistency of the interpretation of maturity stage between MADMF and NEFSC survey staff.

The 2002 SAW 36 considered these data and analyses and the possible causes for the noted inconsistencies, concluded that more detailed spatial and temporal analyses were needed before revisions to the maturity schedule could be adopted, and made a number of research recommendations for future winter flounder maturity work. The O'Brien et al. (1993) maturity at age schedule used in the 1998 SAW 28 and 2002 SAW 36 assessments was retained in the 2005 GARM 2 (NEFSC 2005), and 2008 GARM 3 (NEFSC 2008) assessments.

The 2002 SAW 36 assessment Research Recommendations were to "Evaluate the maturity at age of fish sampled in the NEFSC fall and winter surveys" and "Examine sources of the differences between NEFSC, MA and CT survey maturity (validity of evidence for smaller size or younger age at 50% maturity in the NEFSC data). Compare NEFSC inshore against offshore strata for differences in maturity. Compare confidence intervals for maturity ogives. Calculate annual ogives and investigate for progression of maturity changes over time. Examine maturity data from NEFSC strata on Nantucket Shoals and near George=s Bank separately from more inshore areas. Consider methods for combining maturity data from different survey programs."

Some of these 2002 SAW 36 research recommendations are addressed in this assessment. However, the NEFSC winter survey (1992-2007) age structures have not been processed, and so the associated maturity stages are not available in computerized form. Maturity data from the CTDEP trawl survey have not yet been compiled and provided in computerized form to the SDWG; therefore, no analyses have been completed for those data. The current work responding to the 2002 SAW 36 Research recommendations focuses on the maturity schedule for female fish, which in the past has been adopted as a proxy schedule for all the fish in the catch at age. In all cases, probit regression models assuming lognormal error were fit to the maturity data to estimate proportions mature at age. Both the MADMF and NEFSC maturity data have been recompiled and updated schedules computed.

The MADMF Spring survey data for the SNE/MA stock strata (11-21) were updated through 2008, with year blocks for 1982-1984, 1985-1989 (corresponding to the data subset included in the O'Brien [1993] maturity schedule), 1990-1995, 1996-2001, 2002-2007, 2008, and all data combined for 1982-2008. The MADMF maturity data indicate a consistent pattern over the time series, with maturity at age 2 less than 10% across the time series, and some increase in maturity at age 3 (from about 50% to about 66%) in the 2002-2007 period (Figure A3).

Figure A3 and the table below show that when all the currently available MADMF Spring female maturity data are combined (1982-2008; 8208 in the plot legend) the resulting schedule is within 2-3% at age of the O'Brien (1993) schedule used in previous assessments.

Age	1	2	3	4	5	6	7+
O'Brien 1993	0.00	0.06	0.53	0.95	1.00	1.00	1.00
Current	0.00	0.08	0.56	0.95	1.00	1.00	1.00

The NEFSC Spring survey data for all SNE/MA stock complex strata (offshore 1-12, 25, 69-76; inshore 1-26, 45-56) were also updated through 2008, with year blocks for 1981-1984, 1985-1989 (corresponding to the data subset included in the O'Brien [1993] maturity schedule), 1990-1995, 1996-2001, 2002-2007, and 2008. The NEFSC Spring maturity data indicate a more variable pattern over the time series than the MADMF Spring data, with maturity at age 2 ranging from 28% to 70% across the time series, and maturity at age 3 at greater than 90% for the entire 1981-2008 period. The NEFSC Spring data continue to indicate an age of 50% maturity (A50) of about age 2 (Figure A4), compared to A50 = age 3 for the MADMF Spring data.

Data from the NEFSC Fall survey, the NEFSC Spring survey for Massachusetts waters inshore strata (55-56; Nantucket Shoals), and the NEFSC Spring survey for Massachusetts waters offshore strata (9-12 and 25) have also been compiled and analyzed in the same way as the NEFSC Spring and MADMF Spring survey full data sets, to respond to the Research Recommendations. Like the NEFSC Spring data, the NEFSC Fall data indicate an age of 50% maturity (A50) of about age 2 (Figure A5), compared to A50 = age 3 for the MADMF Spring data. The NEFSC Spring Massachusetts waters inshore strata maturity data indicate a more variable pattern over the time series than the full NEFSC Spring data set, with maturity at age 2 ranging from 0% to 74% across the time series, and maturity at age 3 from 89% to 100%. Like the full NEFSC Spring data set, the NEFSC Spring Massachusetts inshore data indicate an age of 50% maturity (A50) of about age 2 (Figure A6), compared to A50 = age 3 for the MADMF Spring data. Finally, the NEFSC Spring Massachusetts waters offshore strata maturity data indicate a more variable pattern over the time series than the full NEFSC Spring data set, with maturity at age 2 ranging from 6% to 86% across the time series, and maturity at age 3 from 73% to 100%. Like the full NEFSC Spring data set, the NEFSC Spring Massachusetts Offshore data indicate an age of 50% maturity (A50) of about age 2 (Figure A7), compared to A50 = age 3 for the MADMF Spring data.

Given the respective characteristics of the MADMF Spring and various strata set combinations of the NEFSC Spring and Fall maturity, and the indications from the McBride et al. (MS 2011) histological work that age 2 fish are likely not mature, the SDWG concluded that the MADMF Spring survey data continue to provide the best macroscopic evaluation of the maturity stage for SNE/MA winter flounder. The SDWG recommended that the MADMF Spring data 1982-2008 maturity estimates at age (age 1 - 0%; age 2 – 8%; age 3 – 56%; age 4 – 95%, age 5 and older – 100%) be used in the 2011 SAW 52 assessment.

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 used in all previous assessments was 0.20 for all ages and years, and was based on the ICES/FAO 3/Tmax "rule-of-thumb" (e.g., see Vetter 1988 and Quinn and Deriso 1999) 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 0.30 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 0.20. Dickie and McCracken (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) made equilibrium yield calculations for winter flounder from Rhode Island waters (Tmax = 12) using F values from Berry et al. (1965) and 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), 0.50 (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 MS 2011). 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 0.26 to 0.35.

Values derived from life history equations found in the fisheries literature also support a higher estimate of M for winter flounder. Three of these 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, Hewett and Hoenig 2005, and the ICES/FAO "rule-of-thumb" respectively). A recently proposed method from Gislason et al. (2010), based on the SNE/MA stock mean length at age (Ages 1-16) and associated von Bertalanffy growth parameters from NEFSC survey 1976-2010 age-length data (see Growth and Maturity above), estimated M to be 0.37 (see text table below).

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

Study	**Method**	**M**
ICES/FAO 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 MS 2011	Re-analysis of Howe and Coates 1975	0.30
Poole 1969	Analysis of Tagging Data from 1938	0.33
Dickie and McCracken 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 of mean squared residual for Preliminary ADAPT VPA SNE/MA stock models indicate best fits for M in the range of 0.20 to 0.30. The likelihood profile of initial ASAP SCAA model runs for the SNE/MA stock indicates a best fit for M= 0.60 (Figure A8). Model runs from Rademeyer and Butterworth (MS 2011 a, b) SCAA (ASPM) models at M equal to 0.20, 0.30, and 0.40 also reveal decreasing negative log-likelihood as M is increased for GOM and SNE/MA stock models (see text tables 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	0.20	0.30	0.40
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		
	0.20	0.30	0.40
-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) and 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) and usually less than 0.4%.

Contribution of winter flounder (percent by weight) to the diet of their major fish predators.

	% Diet composition of winter flounder	
	L95% CI	U95%CI
Spiny dogfish	0.107%	0.205%
Winter skate	0.145%	0.160%
Little skate	0.012%	0.016%
Atlantic cod	0.240%	0.317%
Sea raven	0.784%	0.883%
Goosefish	0.249%	0.260%

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 5,611 seals in 1999 (Waring et al. 2009) and a current survey is being conducted to estimate the size of the seal population. However, no time series of seal abundance or seal consumption of winter flounder are available.

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

Landings

Commercial fishery landings reached an historical peak of 11,977 metric tons (mt) in 1966, then decreased through the 1970s, peaked again at 11,176 mt in 1981, and then steadily decreased to 2,128 mt in 1994. Commercial landings then increased to 4,556 mt in 2001 but have generally decreased since then. Under a prohibition of commercial possession in the EEZ since May 2009, commercial landings decreased to 271 mt in 2009 and 174 mt in 2010 (Table A1, Figure A9). Since 1995, the procedure used to allocate the commercial landings to statistical area has allowed estimation of the variance in the landings due to this process. For the SNE/MA winter flounder commercial fishery landings, the Proportional Standard Error (PSE) has averaged less than 1% (Table A1). About 66% of the commercial landings have been allocated to statistical area based on a match of Dealer records and Vessel Trip Reports for each trip over the 1995-2010 time series, with lesser percentages allocated based on an increasingly broad stratification basis (Table A2).

Most of the commercial landings from the SNE/MA stock complex have historically been taken from statistical areas 521 and 526 (east and south of Cape Cod, MA), 537 and 539 (south of Rhode Island), and 611-613 (Long Island Sound and south of Long Island; Table A3 and Figures A10-A13 for the years 1983, 1993, and 2000). With the restrictions on EEZ landings beginning in 2009, the percentage of landings from area 521 decreased from about 40% in 2007-2008 to about 20% in 2009; however, that percentage rebounded to 58% in 2010 (Table A3 and Figures A10, A14-A15). In 2009 about 40% of the commercial landings were from areas 537 and 539 off Narragansett Bay, RI, and about 35% off the coasts of NY and NJ. In 2010 about 18% of the commercial landings were from areas 537 and 539 off Narragansett Bay, RI, and about 12% off the coasts of NY and NJ. The primary gear used in the commercial fishery is the otter trawl, which has accounted for an average of 98% of the landings since 1989. Scallop dredges, hand-lines, pound nets, fyke nets, and gill nets account for the remaining 2% of total landings. Most SNE/MA winter flounder

are landed as large and small market categories; additional, port-specific categories exist for medium, unclassified, and lemon sole (i.e., extra large and jumbo; Figure A16).

Recreational fishery landings in numbers and weight are directly estimated by the National Marine Fisheries Service (NMFS) Marine Recreational Fisheries Statistics Survey (MRFSS). Recreational landings peaked in 1984 at 5,510 mt, but declined substantially thereafter (Table A4, Figure A9). Recreational landings have been less than 1,000 mt since 1991, with only 28 mt estimated for 2010. The states of New York and New Jersey account for most of the recreational fishery landings (Figure A17), and the principal mode of fishing (>90%) is from private or rental boats, with most recreational landings occurring during January to June (Figure A18). The PSE of the recreational landings has averaged about 27% over the time series (Table A4).

Discards

In the review of the 1995 SAW 21 assessment of SNE/MA winter flounder (NEFSC 1996), the workshop concluded that there were too few NEFSC Fishery Observer Program sampled trips in which winter flounder were caught to adequately characterize the overall ratio of discards to landings in the commercial fishery. The Observer sample length frequency data, however, were judged adequate to help characterize the proportion discarded at length. Therefore, commercial discards for 1985 to 1993 were estimated from length frequency data from the NEFSC and MADMF trawl surveys, commercial port sampling of landings at length and Observer sampling of landings and discard at length. In this "mesh-selection" approach, survey length frequency data aggregated by half-years (MADMF survey in spring and NEFSC survey in fall, to maximize sample size) were smoothed using a three point moving average, then filtered through a mesh selection ogive for 4.5 inch (114 mm) mesh (1984-1989), 5.0 inch (127 mm) mesh (1990-1992, spring 1993) or 5.5 inch (140 mm) mesh (fall 1993). The choice of mesh sizes was based on the sizes and selection curves used in the yellowtail flounder assessments for southern New England (Rago et al. 1994) and comparison to length frequencies of commercial landings. The mesh filtering process resulted in a survey length frequency of retained winter flounder. A logistic regression was then used to model the percent discarded at length from 1989-1992 Observer data, and the resulting percentages at length were applied to the survey numbers at length to produce the survey-based equivalent of commercial kept and discarded winter flounder. The 1989-1992 average percentage discard at length was applied to 1981-1988. The survey numbers per tow at length "kept" were then regressed against commercial numbers landed at length. The linear relationship was calculated for those lengths common to both length frequencies and fitted with an intercept of zero. The slope of the regression provided a conversion factor to re-scale the survey "discard" numbers per tow at length to equivalent commercial numbers at length. The resulting vector of number of fish discarded at length was multiplied by a discard mortality rate of 50% (as averaged in Howell et al. 1992) to produce the vector of fish discarded dead at length per half year. The number of dead discards at length was adjusted by the ratio of weighout landings to total commercial landings and summed across seasons and lengths (and corresponding weight at length) to produce the annual total number and weight of commercial fishery discards for 1985-1993. In the SAW 28 assessment (NEFSC 1999), this same method using the 4.5 in mesh ogive and 1989-1992 average discard percentage at length was used to estimate commercial fishery discards for 1981-1984. These previously estimated values will be retained in the current assessment for estimates of the 1981-1993 commercial fishery discards (Table A5).

In the 1998 SAW 28 (NEFSC 1999), 2002 SAW 36 (NEFSC 2003), and 2005 GARM2 (NEFSC 2005) assessment, the SAW 21 survey length-mesh selection method, NEFSC Fishery Observer data (OB), NER

Vessel Trip Report (VTR), and Northeast Region Dealer Report (DLR) data were considered as sources of information to estimate commercial fishery discards, with a focus on the latter three sources. The characteristics of both the OB and VTR discard data (number of trip samples, frequency distributions of discards to landings ratio per trip, mean and variance of annual half-year discards to landings ratio) as a source for discard rates were examined, and the assessment reviews concluded that the VTR mean discard to landed ratio aggregated over all trips in annual half-year season strata (January to June, July to December) provided the most reliable data from which to estimate commercial fishery discards. VTR trawl gear fishery discards to landings ratios on a half-year basis (January to June; July to December) were applied to corresponding commercial fishery landings (all gears) to estimate discards in weight for 1994-2004. VTR discard ratios for winter flounder for other gears (scallop dredge, gillnet) were judged to be too variable to provide reliable estimates of discards.

In the 2008 GARM-III (NEFSC 2008) assessment, the Standardized Bycatch Reporting Method (SBRM) approach to the estimation of discards (Wigley et al. 2008) was applied for comparison with the OB and VTR discard rate estimation methods used in previous assessments. Discard rates by half-year were calculated for trawls and scallop dredges, and applied to the corresponding landings (winter flounder landings for the OB and VTR rates; landings of all species for the SBRM rates). OB discard rate estimates were found to be higher and more variable than discard estimates from the VTR and SBRM methods, which were generally of about the same order of magnitude. In particular, the 1999 and 2000 OB discard estimates appear to be infeasible.

When the VTR and SBRM discard estimates were examined by gear, it was apparent that the scallop dredge estimates generally made up a larger part of the SBRM estimate total when compared to the VTR estimates. The scallop dredge fishery lands a small amount of SNE/MA winter flounder (<35 mt annually) compared to the trawl fishery (1,000-5,000 mt annually, prior to 2009), and so even though the VTR scallop dredge discard rates can be high, the VTR discard estimates for the scallop fishery were relatively low. In previous assessments neither the OB nor VTR discard rate data were considered adequate for the estimation of discards specific to the scallop dredge fishery, due to sample size and inter-annual variability of the rates. In contrast, the SBRM scallop dredge discard estimates are quite variable and can be much larger than the trawl discard estimates, in spite of a low discard rate (discard of winter flounder to total landings of all species), because of the large magnitude of total fish landings in the fishery and the sensitivity of the discard estimate calculation to small inter-annual changes in the absolute discard rate. After reviewing the magnitude and precision of discard estimates from the VTR and SBRM approaches, the 2008 GARM-III panel adopted the SBRM as the best method for estimation of SNE/MA winter flounder commercial fishery discards for 1994 and later years. The PSE of the commercial discards has averaged 27% over the time series. A discard mortality rate of 50% was applied to the commercial live discard estimates, as assumed in Howell et al. (1992). Commercial fishery discard losses (i.e., dead fish) peaked in the early 1980s at 1,000-1,500 mt per year. Commercial fishery discard losses have since decreased to less than 200 mt per year since 1997 (Table A6).

Recreational fishery live discards in numbers of fish are directly estimated by the MRFSS (B2 category), and the estimated numeric variance has been assumed for the discard in weight, which is estimated in the assessment by allocation according to the length assumptions or samples. The PSE of the recreational discards has averaged 30% over the time series. A discard mortality rate of 15% was applied to recreational live discard estimates as assumed in Howell et al. (1992). Recreational fishery discard losses (i.e., dead fish)

peaked in 1984-1985 at about 700,000-750,000 fish or 150-200 mt. Discard losses have since decreased to less than 100,000 fish or 20 mt per year since 2000. (Table A4).

Length and Age Sampling and Estimated Age Compositions

Length samples of winter flounder are available from both the commercial and recreational landings. In the commercial fishery, annual length sampling intensity varied from 10 to 251 mt landed per 100 lengths measured during 1981-2010 (Table A7). Port sampling has generally been adequate to develop the annual commercial fishery landings at age on a half-year or quarterly, market category basis (Table A8). In the recreational fishery, annual length sampling intensity varied from 28 to 614 mt landed per 100 lengths measured during 1981-2010 (Table A9). Recreational fishery ages were determined on a half-year basis using NEFSC survey spring and fall age-length keys.

As noted above, prior to 1994 the NEFSC trawl survey length frequencies and commercial trawl fishery mesh selection data were used to estimate the magnitude and characterize the length frequency of the commercial fishery discard. For 1994-2010, NEFSC Fishery Observer trawl and scallop fishery winter flounder discards to total all-species landings ratio estimates (SBRM approach) were applied to corresponding commercial fishery all-species landings to estimate discards . The NEFSC Fishery Observer length frequency samples were applied on a half-year basis to characterize the proportion discarded at length for 1994-2010 (Table A10). The ages of the commercial fishery discards were determined using NEFSC survey spring and fall age-length keys.

Irregular sampling of the recreational fisheries by state fisheries agencies since 1997 has indicated that the recreational fishery discard is usually of fish below the minimum landing size of 12 inches (30.5 cm). For 2002-2010, discard length samples from the NYDEC sampling of the recreational for-hire fishery and from the CTDEP Volunteer Angling Survey (VAS) have been used to better characterize the recreational fishery discard. Ages were determined on a half-year basis using NEFSC survey spring and fall age-length keys.

Commercial and recreational fishery landings and discards at age are presented in Tables A11-A14. Total fishery catches and mean weights at age are summarized in Tables A15-A16 and Figures A19-A20. Aggregate fishery catches in weight and numbers are summarized in Table A17.

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.

The NEFSC spring and fall bottom trawl surveys provide long time series of fishery-independent indices for SNE/MA winter flounder. The NEFSC spring and fall surveys are conducted annually during March-May and September-November, ranging from just south of Cape Hatteras, North Carolina north to Canadian waters (Figures A21-A22). The NEFSC winter surveys were conducted during 1992-2007 from Cape Hatteras north to Georges Bank. Stratified mean indices for the NEFSC spring, fall, and winter surveys are presented in Table A18 and Figure A23.

The Fisheries Survey Vessel (FSV) *Albatross IV* (ALB) was replaced in spring 2009 by the FSV *Henry B. Bigelow* (HBB) as the main platform for NEFSC research surveys, including the spring and fall bottom trawl

surveys. The size, towing power, and fishing gear characteristics of the HBB are significantly different from the ALB, resulting in different fishing power and therefore different survey catchability. Calibration experiments to estimate these differences were conducted during 2008 (Brown 2009), and the results of those experiments were peer reviewed by a Panel of independent (non-NMFS) scientists during the summer of 2009 (Anonymous 2009, Miller et al. 2010). The terms of reference for the Panel were to review and evaluate the suite of statistical methods used to derive calibration factors by species before they were applied in a stock assessment context. Following the advice of the August 2009 Peer Review (Anonymous 2009), the all-seasons ratio estimator calibration factors were initially adopted to convert HBB survey catch number and weight indices to ALB equivalents. The aggregate catch number calibration factor for all seasons is 2.490; the aggregate catch weight factor for all seasons is 2.086.

The SDWG noted that the HBB will not routinely sampled the shallowest inshore strata in the standard set previously used for SNE/MA winter flounder (e.g. 47, 1, 3, 4, 12, 13, etc.), and also that winter flounder were rarely caught in the two deepest bands of offshore strata (e.g., 7-8, 11-12, etc.). The SDWG recommended that the NEFSC spring and fall survey time series be revised to reflect a strata set consistent with that being sampled by the HBB (i.e., using only the deepest band of inshore strata) and excluding the two deepest bands of offshore strata (i.e., generally consistent with the set used for the Winter survey series). The revised strata set includes offshore strata 1, 2, 5, 6, 9, 10, 25, 69, 70, 73, and 74, and inshore strata 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 45, 46, and 56, for the years 1976 and later.

Since the 2009 Peer Review, it has become evident that accounting for size of individuals can be important for many species. If there are different selection patterns for the two vessels for a given species, the ratio of the fractions of the fish caught by the two vessels can vary with size. Since 2009, length-based calibration factors have been estimated for several stocks (cod, haddock, and yellowtail flounder through the Trans-boundary Resource Assessment Committee [TRAC] assessment process; silver, offshore, and red hakes during the 2010 SARC 51 and *Loligo* squid during the 2010 SARC 51 (Brooks et al. 2010, NEFSC 2011). For those length-based calibrations, the same basic beta-binomial model from Miller et al. (2010) was assumed, but various functional forms were assumed for the relationship of length to the calibration factor. Since then, Miller (submitted) has explored two types of smoothers for the relationship of relative catch efficiency to length and the beta-binomial dispersion parameter. The smoothers (orthogonal polynomials and thin-plate regression splines) allow much more flexibility than the functional forms previously considered for other stocks by Brooks et al. (2010) and NEFSC (2011).

The SDWG reviewed work by Miller (MS 2011) on winter flounder in greater detail, and compared the model results for all winter flounder to those from a model that accounted for effects of stock area (GOM, GBK, and SNE/MA). The SDWG also explored seasonal effects, but did not fully pursue those models due to a lack of samples in the Gulf of Maine stock region during the spring. The lead assessment scientists for each of the winter flounder stocks compared predicted indices in Albatross units based on the different fitted models to explore the degree of consistency between calibrated indices using the different models.

When fitting the fourth order polynomial with smoother models to data from each stock region, there were convergence issues for the GOM stock data, likely due to over-parameterization of the length effects. When the order of the polynomial was reduced to two for this region, these issues were resolved. The resulting model performed better than the best models that Miller (submitted) fit that did not account for effects of stock area. Inspection of residuals revealed no strong trend with predicted number captured by the HBB or

total number captured by station and no strong departure from normality. The predicted relative catch efficiency was lowest at intermediate size classes for all three stock areas, but the location of the minimum was at larger size for the GBK stock than for the other stock areas. For the SNE/MA stock, there were actually two minima with a slight rise in relative catch efficiency estimated between them.

When applying the relative catch efficiencies to surveys conducted in 2009 and 2010 with the HBB, there is an important caution to note. Lengths may be observed in these surveys that are outside of the range of lengths observed during the calibration study. This problem is exacerbated when the data are broken down into stock area subsets for the estimation of relative catch efficiency, because the limits of the range of sizes available in the subsets can be narrower than the range of the entire data set, and so caution must be taken in predicting catches in ALB units at these sizes. The SDWG also had some concerns with the asymptotically increasing estimates of relative catch efficiencies at the smallest and largest sizes for the winter flounder stocks, particularly when converting historic ALB indices to HBB equivalents. Sizes of fish outside of the ranges observed during the calibration study would potentially lead to extremely high HBB abundance indices at the extremes of the length composition for the historic data.

An adaptation of the regional model was explored that constrained lengths beyond a minimum and maximum length to have constant relative catch efficiencies. The minima and maxima were determined by specifying a maximum coefficient of variation (CV) of predicted relative catch efficiencies at these lengths. These CV criteria resulted in models that provided aggregate abundance indices that were very similar to the corresponding models without the CV criteria. Because no ad-hoc CV criteria were necessary in the initial regional length models, the SDWG found those to be preferable.

Lastly, the swept areas for each tow during the 2009 and 2010 surveys would ideally be used to predict ALB catches at each station, but if there is little variability in the swept areas, a mean can be used and the mean number per tow at length in HBB units can be converted to ALB units. The fourth order polynomial model fit to data for the SNE/MA stock region, incorporating a mean ratio of the vessel swept areas of 0.5868 (HBB to ALB), was used to calculate the factors at length (Figure A24) used to calibrate the 2009-2010 NEFSC HBB survey indices to ALB units for use in population model calibration (Table A19). After the application of age-length keys, the effective calibration factors at age (ratio of HBB to ALB indices at age) ranged from 6.86 at age 1 in spring 2009 to 2.50 at age 7+ in spring 2010, averaging 3.19 across all ages and seasons (Table A20).

Several state survey time series were available to characterize the abundance of SNE/MA winter flounder. The MADMF spring survey, Rhode Island Division of Fish and Wildlife (RIDFW) spring survey, University of Rhode Graduate School of Oceanography (URIGSO), Connecticut Department of Environmental Protection (CTDEP) Long Island Sound Trawl Survey (LISTS) spring, and the New Jersey Division of Fish, Game and Wildlife (NJDFW) ocean and rivers research survey trends are summarized in Tables A21-A22 and Figures A23 and A25. The numerous state recruitment surveys (MADMF, RIDFW, CTDEP, New York Department of Environmental Conservation (NYDEC), NJDFW, Delaware Division of Fish and Wildlife (DEDFW)) are summarized in Table A23 and Figures A26-A27.

The University of Rhode Island Graduate School of Oceanography (URIGSO) has conducted a standardized, two-station trawl survey in Narragansett Bay and Rhode Island Sound since the 1950s, with consistent sampling since 1963. The mean numbers per tow for the two stations, one in upper Narragansett Bay and

one at the mouth of the Bay, were averaged to provide annual aggregate and indices at age. The URIGSO indices for SNE/MA winter flounder peaked in the late 1960s and again in the early 1980s, and have since shown has decreasing trend, with a record low in 2007 (Table A24 and Figure A25).

The VIMS NEAMAP industry-cooperative survey was started in fall 2006 to provide research survey samples in the spring and fall seasons along the Atlantic coast from Rhode Island to North Carolina in depths of 20-90 feet (9-43 meters). The NEAMAP indices for SNE/MA winter flounder do not indicate a trend in the recent abundance of winter flounder (Table A25 and Figure A25).

Indices at age are available from most of the research surveys for use in model calibration and are presented in Tables A26-A36.

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.

2008 GARM-III ADAPT VPA Model selection process

The suite of research survey calibration indices developed for use in the 2002 SAW 36 assessment (NEFSC 2003) was retained in the 2005 GARM2 and 2008 GARM-III assessments (NEFSC 2005, 2008). The 2008 GARM-III VPA BASE case model exhibited a strong retrospective pattern, although it was less severe in the most recent terminal years than in the 2005 GARM2 assessment. Retrospective patterns in stock assessments result from structural errors in model, occurring when there has been a change during the model time series some inputs or estimated parameters that are assumed known (e.g., the catch) or constant (e.g., natural mortality or survey catchability). The 2008 GARM-III Panel (NEFSC 2008) considered that there are four potential causes of retrospective patterns in age structured stock assessments: 1) an unrecorded change in catches, 2) an undetected change in natural mortality, 3) an undetected change in survey calibration index catchability (q), or 4) an undetected change in fishery selectivity or partial recruitment. In all cases, either the biomass has changed (changes in natural mortality and unrecorded catch) or is perceived to have changed (changes in catchability or selectivity) in a way that cannot be explained by the catch-at-age data. Random noise is thought to be an unlikely cause of the retrospective pattern, based on simulation analyses considered by the 2008 GARM-III Panel, although those analyses raised the possibility of retrospective patterns being caused by mis-specification of the likelihood function or the impact of influential data points in the survey calibration series. The 2008 GARM-III Panel noted that while assuming dome-shaped fishery and survey partial recruitments may resolve retrospective patterns, these may also lead to what was termed "cryptic" biomass – biomass generated by the model that has not been observed in either the fishery or surveys. Throughout the 2008 GARM 3 review, the burden of proof was placed upon analysts to convincingly demonstrate that fish existed in the population when not observed in the fishery and surveys, even if the model fit with dome-shaped partial recruitment appeared superior. In some cases, additional information (data and/or assumptions) external to the model was considered (NEFSC 2008).

It was not possible to determine which single factor or combination of factors was responsible for the retrospective pattern observed in the SNE/MA winter flounder VPA model. However, the 2008 GARM-III

Panel judged that it was appropriate to adjust the model formulation to reduce the retrospective pattern (NEFSC 2008). In the SNE/MA winter flounder VPA model, the survey series were therefore split "pre and post 1994" (i.e., split between 1993 and 1994, given the change in commercial discard estimation and commercial landings reporting methods between these years), except for the NEFSC Winter, NJDFW Ocean, and NJDFW River survey series, which began in 1992, 1993, and 1995, respectively. Under this SPLIT run configuration, the retrospective pattern was reduced. No significant problems in residual patterns developed as a result of splitting the survey series, and the pattern for the NEFSC Fall survey appeared to be improved (less of a trend/blocking from negative residuals in the 1980s to positive residuals in the 1990s-2000s, likely corresponding to the change in retrospective patterns). There was little change in the pattern of the CTDEP Spring residuals, which continued to show a trend/ blocking in both the BASE and SPLIT run configurations. The precision of the SPLIT run terminal year estimates was comparable to the BASE run estimates. The Mohn's rho statistic calculated for the BASE and SPLIT runs ([retrospective year – terminal year]/terminal year; i.e., relative difference), either summed or averaged over the last seven retrospective years, was comparable in absolute magnitude but opposite in sign for F. The absolute value of the Mohn's rho for SSB was about 85% smaller for the SPLIT run; the value for recruitment at age 1 was about 30% smaller. The SPLIT configuration ADAPT VPA model was accepted as the basis for 2008 GARM-III SNE/MA winter flounder catch advice (NEFSC 2008).

2011 SAW 52 input data and Preliminary model configurations and results with M = 0.2

An initial population analysis was conducted using the NOAA Fisheries Toolbox (NFT) ADAPT VPA version 3.0.3. (NFT 2010) to provide a "bridge" from the 2008 GARM-III assessment (NEFSC 2008) to the current work by demonstrating updated results using the same general model configuration. The following NEFSC and state agency trawl survey abundance indices at age were input as candidate calibration indices: NEFSC spring trawl ages 1-7+, NEFSC fall trawl ages 1-6+ (advanced to calibrate January 1 abundance of ages 2-7+), NEFSC winter trawl ages 1-5, MADMF spring trawl ages 1-7+, RIDFW fall seine age 0 (advanced to age-1), RIDFW spring trawl ages 1-7+, URIGSO trawl ages 1-7+, CTDEP fall seine age 0 (advanced to age-1), CTDEP spring trawl ages 1-7+, NYDEC trawl age 0 (advanced to age-1) and ages 1-2, MADMF summer seine index of age-0 (advanced to age-1), DEDFW juvenile trawl age-0 (advanced to age-1), NJDFW Ocean trawl ages 1-7+, and NJDFW River trawl ages 1-7+ (Tables A26-A36). In all models, the NEFSC Winter, Spring and Fall indices were input as "area-swept" numbers (assuming 100% survey efficiency and area-swept of 0.0112 square nautical miles per tow). Both BASE (with all survey indices input as continuous series) and SPLIT (with some survey series split at 1993/1994, as in the 2008 GARM-III assessment) preliminary ADAPT VPA model configurations were considered.

As an alternative to the ADAPT VPA model used in the 2008 GARM-III assessment (NEFSC 2008), the same input catch and survey index data were used in the ASAP version 2.0.21 Statistical Catch At Age (SCAA) model (NFT 2011). Two model configurations of the survey calibration indices were constructed. In the first, Indices At Age (IAA), the survey indices were input as in the ADAPT VPA, with each index at age input as a separate series with a fixed selectivity at age of one (S = 1) and a characteristic Coefficient of Variation (CV) set at 0.4 (40%). In this configuration, a catchability coefficient (q) is estimated for each index at age; the CV of the q was set at 0.9 to allow flexibility from the starting value, and the weighting factor (Lambda) for each index at age was set equal to one (L = 1). Annual Effective Sample Sizes (ESS) for the fishery age compositions was set at 200. An internal stock-recruitment relationship was not estimated. Both BASE and SPLIT model ASAP IAA configurations were considered.

In the second ASAP configuration, the survey indices were input as the aggregate total and as a vector of indices at age for each year. In this configuration, each set of survey indices at age is modeled as a multinomial distribution (probabilities at age; MULTI) with an accompanying vector of fixed or estimated of selectivity at age, with the CV of selectivity set at 0.5. To ensure robust estimation, the selectivity was fix at one (S=1) for age 4, and selectivity at age estimated for the other ages in each series. A characteristic CV for each series aggregate total was set at 0.4 (40%). The CV of the catchability coefficient (q) for each series was set at 0.9 to allow flexibility from the starting value, and the weighting factor (Lambda) for each index series was set equal to one (L = 1). Annual Effective Sample Sizes (ESS) for the fishery age compositions was set at 200; annual ESS for all multinomial survey age compositions was set at 10. For single age recruitment index series, the surveys were modeled as in the IAA configuration. An internal stock-recruitment relationship was not estimated. Both BASE and SPLIT model ASAP MULTI configurations were considered.

The Preliminary ADAPT VPA BASE model run with M = 0.2 provided estimates of SSB that ranged from about 17,000 mt in 1982 to 2,300 mt in 2005, increasing to 4,200 mt in 2010. Estimates of F (ages 4-5) increased from about 0.54 in 1981 to 1.55 in 1993, decreasing to 0.09 in 2010. Recruitment at age 1 ranged from about 61 million in 1981 (1980 year class) to about 4 million in 2007 (2006 year class). The preliminary VPA BASE run exhibited a strong retrospective pattern, with the underestimation of terminal F ranging from -53% in terminal year 2005 to -29% for terminal years 2008-2009 and the overestimation of SSB ranging from +103% in 2007 to +30% in 2009.

The Preliminary ADAPT VPA SPLIT model run with M = 0.2 provided estimates of SSB that ranged from about 17,000 mt in 1982 to 1,900 mt in 2009, increasing to 2,900 mt in 2010. Estimates of F (ages 4-5) increased from 0.54 in 1981 to 1.55 in 1993, decreasing to 0.14 in 2010. Recruitment at age 1 ranged from about 61 million in 1981 (1980 year class) to about 3 million in 2007 (2006 year class). The SPLIT configuration resulted in a reduced retrospective pattern compared to the BASE run, with the retrospective error in terminal F ranging from -26% in terminal year 2005 to +16% for terminal year 2006 and the retrospective error in SSB ranging from +57% in 2007 to +6% in 2009.

Estimates from the Preliminary ADAPT VPA SPLIT model run with M = 0.2 are compared with previous assessment results in Figures A28-A30. In general, the historical trends in F and recruitment are very similar, but "historical retrospective" errors in both estimates are evident. Historical estimates of SSB during 1981-1985 are the most different in absolute terms; these differences are due mainly to changes in the ADAPT VPA calculations for the oldest true age, the "plus-group" age 7+, and the use of the exact catch equation (instead of Pope's approximation) in the current ADAPT VPA model, compared to versions used in previous assessments. Substantial "historical retrospective" errors in SSB are also evident for the 1997, 2001, 2005, and 2007 terminal years.

In the Preliminary ADAPT VPA SPLIT run configuration with M = 0.2, the retrospective pattern was reduced as the estimated survey catchability (q) generally decreased before the split (1981-1993) and increased after (1994-2010), by as much as +/- 40%-50% (e.g., NEFSC Fall survey). For several series (e.g., NEFSC Spring and Fall, RI Spring) the pattern in q at age also became more asymptotic (flat) after the split (Figure A31). For the CT Spring series, however, the changes were different, from a nearly flat pattern in the BASE configuration to one with a decreasing trend in q at age before the split at 1993/1994 but an increasing trend in

q at age after in the SPLIT configuration (Figure A31). For the NJ and URIGSO series changes in survey *q* were small (Figure A32). Of the YOY series, the largest proportional changes in *q* were in the MA and CT indices, generally following the pattern of reduced *q* before the split and increased *q* after the split (Figure A32). In the VPA, there was little change in the fishery selectivity patterns between the BASE and SPLIT configuration with both exhibiting a decrease on selectivity at ages 2-4 during 1994-2010, in line with expectations given changes in fisheries regulations (Figure A33).

The Preliminary ASAP IAA BASE model run with M = 0.2 provided estimates of SSB that ranged from about 11,500 mt in 1982 to 2,100 mt in 1993, increasing to 4,800 mt in 2000, and then decreasing again to 2,100 mt in 2008 before increasing to 3,600 mt in 2010. Estimates of F (ages 4-5) increased from 0.67 in 1982 to 1.32 in 1985 and then remained at about 0.6 or higher until peaking again at 1.14 in 2007, before decreasing to 0.08 in 2010. Recruitment at age 1 ranged from about 55 million in 1981 (1980 year class) to about 5 million in 2007. The ASAP IAA fishery selectivity patterns before and after the survey split (the runs were purposely configured with fishery selectivity blocks to coincide with the survey split) were similar to those from the VPA, but with a slight dome for the years before the split (1981-1993) at age 6-7+ (S = 0.7-0.8; Figure A33). The ASAP IAA BASE run exhibited a moderate retrospective pattern, with the underestimation of terminal F ranging from -65% in terminal year 2007 to -16% for terminal year 2009 and the overestimation of SSB ranging from +73% in 2007 to +8% in 2009.

The Preliminary ASAP IAA SPLIT model run with M = 0.2 provided estimates of SSB that ranged from about 15,000 mt in 1982 to 2,000 mt in 1993, increasing to 3,900 mt in 2000, and then decreasing again to 1,600 mt in 2008 before increasing to 2,600 mt in 2010. Estimates of F (ages 4-5) increased from 0.57 in 1982 to 1.22 in 1985 and then remained at about 0.6 or higher until peaking again at 1.22 in 2007, before decreasing to 0.12 in 2010. Recruitment at age 1 ranged from about 66 million in 1981 (1980 year class) to about 4 million in 2007 (2006 year class). The SPLIT configuration resulted in a reduced retrospective pattern compared to the BASE run, with the retrospective error in terminal F ranging from -55% in terminal year 2007 to -2% for terminal year 2008 and the retrospective error in SSB ranging from +56% in 2007 to 0% in 2008. The SDWG noted that the reduction in the retrospective pattern due to the SPLIT configuration was not as great for the ASAP IAA model as for the ADAPT VPA.

In the ASAP IAA SPLIT run configuration, the retrospective pattern was reduced as the estimated survey catchability (*q*), as in the ADAPT VPA SPLIT run, generally decreased before the split (1981-1993) and increased after (1994-2010), by about the same as in the VPA (e.g., NEFSC Fall survey; Figure A34). For several series (e.g., NEFSC Spring and Fall, RI Spring) the pattern in *q* at age also became slightly more asymptotic (flat) after the split (Figure A34). For the CT Spring series, however, the changes were again different, from a nearly flat pattern in the BASE configuration to one with a decreasing trend in *q* at age before the split but an increasing trend in q at age after in the SPLIT run; however, the changes were smaller than in the VPA (Figures A32, A34).

For the NJ and URIGSO series, there were only small changes in survey q. Of the YOY series, as in the VPA the largest proportional changes in *q* were in the MA and CT indices, generally following the pattern of reduced *q* before the split and increased *q* after the split. In the ASAP IAA SPLIT run, there was more of a change in the fishery selectivity patterns between the BASE and SPLIT configuration than in the VPA, with an increase in selectivity at ages 6-7+ during 1994-2010 (Figure A33).

The Preliminary ASAP MULTI BASE model run with M = 0.2 provided estimates of SSB that ranged from about 15,000 mt in 1983 to 3,400 mt in 1994, increasing to 7,100 mt in 2000, and then decreasing again to

4,300 mt in 2005 before increasing to 6,400 mt in 2010. Estimates of F (ages 4-5) increased from 0.73 in 1982 to 1.12 in 1991 and then decreased to 0.07 in 2010. Recruitment at age 1 ranged from about 66 million in 1981 (1980 year class) to about 5 million in 2003 (2002 year class). The ASAP MULTI BASE fishery selectivity pattern before the survey split (the runs were purposely configured with fishery selectivity blocks to coincide with the survey split) was similar to those from the VPA and ASAP IAA models. The ASAP MULTI BASE fishery selectivity pattern after the split (1994-2010) was different, however, with a more substantially domed shape and S ~ 0.6-0.8 at age 6 and ~0.2-0.4 at age 7+ (Figure A33). The ASAP MULTI BASE run exhibited a moderate retrospective pattern, with the underestimation of terminal F ranging from -35% in terminal year 2003 to -13% for terminal year 2009 and the overestimation of SSB ranging from +48% in 2007 to +12% in 2009.

The Preliminary ASAP MULTI SPLIT model run with M = 0.2 provided estimates of SSB that ranged from about 15,000 mt in 1983 to 3,000 mt in 1994, increasing to 6,200 mt in 2000, and then decreasing again to 2,000 mt in 2009 before increasing to 3,800 mt in 2010. Estimates of F (ages 4-5) were consistently high, from 0.74 in 1982 to 1.17 in 1991, remaining above 0.6 until 2007, and then decreasing to 0.09 in 2010. Recruitment at age 1 ranged from about 67 million in 1981 (1980 year class) to about 4 million in 2007 (2006 year class). The SPLIT configuration resulted in a very slightly reduced retrospective pattern compared to the BASE run, with the retrospective error in terminal F ranging from -38% in terminal year 2006 to -16% for terminal year 2009 and the retrospective error in SSB ranging from +56% in 2007 to +18% in 2008. In contrast to the ADAPT VPA and ASAP IAA models, the use of the SPLIT configuration in the ASAP MULTI run configuration was not effective in reducing the retrospective errors, and in fact the errors were generally larger for most of the terminal year "peels."

In the ASAP MULTI SPLIT run configuration, the estimated aggregate survey catchability (q), as in the ADAPT VPA and ASAP IAA SPLIT runs, generally decreased before the split (1981-1993) and increased after (1994-2010), but generally by less for most surveys (in relative terms) than the age-specific q in the VPA or ASAP IAA models (e.g., NEFSC Fall survey; Figure A35). More response was seen in the ASAP MULTI runs in the estimated survey and fishery selectivity patterns. In general, survey selectivity patterns were more asymptotic (flat) after the split (Figure A36), while the fishery selectivity pattern after the split became "less-domed" by about 10% for age 5, 30% for age 6 and 50% for age 7+ (Figure A33).

A second ASAP MULTI SPLIT run configuration (SELEX3) included a third fishery selectivity block, for the years 2006-2010, as a means to explore the sensitivity of the ASAP model retrospective patterns. The SELEX3 configuration resulted in nearly the same retrospective pattern as the MULTI SPLIT run with 2 fishery selection blocks run; the retrospective error in terminal F ranged from -35% in terminal years 2006-2007 to -17% for terminal year 2009 and the retrospective error in SSB ranged from +49% in 2007 to 12% in 2008. Therefore, adding a third selectivity block to the ASAP MULTI SPLIT model did not further reduce the retrospective pattern.

A third ASAP MULTI SPLIT run (S7P) was configured to explore the sensitivity of the model to fixing the fishery selectivity at age 7+ at S = 0.8, in line with the results from the ADAPT VPA and ASAP IAA models. Fixing S = 0.8 for age 7+ resulted in little change in fishery selectivity pattern at ages 5-6 between the two time blocks, and minor changes in the model population and F estimates.

The S7P configuration resulted in nearly the same retrospective pattern as the MULTI SPLIT run with estimated fishery selectivity for age 7+; the retrospective error in terminal F ranged from -38% in terminal year 2006 to -18% for terminal year 2009 and the retrospective error in SSB ranged from +52% in 2007 to +19% in 2009. Therefore, fixing the selectivity of age 7+ at S = 0.8 had little effect on either the model estimates or the retrospective pattern.

A comparison of these Preliminary ADAPT VPA and ASAP BASE and SPLIT configuration model results with M = 0.2 is presented in Figures A37-A39 for Fishing Mortality, SSB, and recruitment at age 1. Time series patterns in F were in general similar for the six model configurations, although annual estimates varied by as much as 2-3 fold (e.g., 2007), due mainly to differences in the estimated fishery selectivity patterns among models. Trends in SSB were likewise comparable, again with as much as a 2-3 fold difference. Trends in recruitment at age were the most consistent, with the greatest variation at the beginning of the time series.

2011 SAW 52 Developmental model configurations and results with M = 0.3

Besides providing a "bridge" back to the 2008 GARM-III assessment results, examination of the Preliminary ADAPT and ASAP model runs with M = 0.2 clarified the changes in survey *q* (both aggregate and at-age), survey selectivity, and fishery selectivity that occurred with different model configurations (i.e., BASE versus SPLIT; ASAP IAA versus ASAP MULTI). The SDWG elected to continue model development with the ADAPT VPA and ASAP MULTI models, dropping the ASAP IAA configuration from further consideration, since the MULTI configuration provided increased model flexibility (ability to weight and estimate both survey selectivity and aggregate catchability, and to weight fishery catch components) and was generally more in line with widely accepted Statistical Catch at Age (SCAA) modeling practice. The ADAPT VPA SPLIT configuration was carried forward since the retrospective pattern was reduced compared to the BASE configuration, which was dropped from further consideration. However, the ASAP MULTI BASE configuration was carried forward, since the SPLIT configuration was not effective in reducing the retrospective in the ASAP model.

All available survey indices had been used in the calibration in the Preliminary runs (see previous section). In the subsequent model development process, the SDWG reviewed the performance of survey indices used in the calibration and removed some indices from the models based on based on consideration of a) the partial variance in an initial VPA trial run including all indices, b) the precision of the survey series, c) residual error patterns from the various trial runs, and d) the significance of the correlation among indices and with ADAPT VPA abundance estimates from the preliminary BASE run configuration including all potential calibration indices. The SDWG discussed the relative merits of including all available indices in the models versus excluding some indices at age from multi-age time series due to poor performance, typically those at the youngest and oldest ages. The SDWG concluded that all age groups for multi-age surveys would be included in further Developmental models, with the exception of the NYDEC Peconic Bay Small Mesh Trawl Survey (Table A33), for which none of the indices exhibited acceptable diagnostics.

The following single age, YOY abundance indices were also excluded from Developmental model runs because of the presence of large partial variances (i.e., lack of fit), lack of correlation with model estimates, or trends in the residuals (i.e., indication of bias): RIDFW seine survey age 0 (advanced to age 1), NYDEC index of age-0 (advanced to age-1), and DEDFW juvenile trawl age-0 (advanced to age-1; Table A23).

The next step in model development was to increase M from 0.2 to 0.3, adopt the revised calibration survey set in the models, and investigate the Developmental ADAPT VPA SPLIT and ASAP MULTI BASE model

estimates and diagnostics. Time series in trends in F, SSB, and R were comparable for the VPA SPLIT M = 0.2, VPA SPLIT M = 0.3, and ASAP MULTI BASE M = 0.3 models. Increasing M in the VPA decreased the estimates of F and increased the estimates of SSB and R. The ASAP model estimates of F were about 25% lower over the time series than from the VPA with M = 0.2, and were higher at the start of the time series and lower since the late 1980s (Figure A40). ASAP model estimates of SSB averaged about 25% higher, and were lower at the start of the time series and higher since the late 1980s (Figure A41). ASAP recruitment estimates at age 1 averaged about 50% higher than the VPA with M = 0.2 for most of the time series (Figure A42). The range of retrospective errors in F and SSB from the VPA with M = 0.3 were comparable to the VPA with M = 0.2, with no "patterns" in F (Figures A43-A44). The ASAP model exhibited a retrospective pattern on underestimation of F and overestimation of SSB, with the range of retrospective errors in F and SSB (about 40%) comparable to but slightly less than those from the VPA models (40-50%) (Figure A45).

The next developmental step was the further investigation of configurations that would reduce the retrospective errors in the ASAP MULTI model, through changes in the weighting of likelihood components and selection of survey calibration indices. Five additional ASAP models were configured: a) reducing the weight on the fishery catch compositions from 200 to 50, still 5 times that for the survey age compositions, b) reducing the on the fishery catch compositions from 200 to 10, equal to that for the survey age compositions, c) fixing the fishery selectivity in both periods (1981-1993; 1994-2010) at S = 1.0 (flat topped) for ages 4 and older, d) removal from the model of the NEFSC Fall survey series, which exhibited a strong residual pattern in most model configurations and e) internal estimation of the stock-recruitment function. Of these configurations, reducing the annual fishery ESS from 200 to 10 (ASAP model CAT10) provided decreased retrospective errors in both F (ranging from -38% to -13%) and SSB (ranging +42% to +12%), and so this ESS setting was adopted for subsequent ASAP model development.

The SDWG noted that sensitivity run e) internal estimation of the stock-recruitment function, provided feasible estimates of steepness (h = 0.66) and reference points when using a steepness prior. However, the final model did not include internal stock-recruitment function estimation; instead, the stock-recruitment parameters were fit externally so that a consistent set of mean weights (most recent 5 year average) could be used in the calculation of FMSY and potential proxies, to ensure consistency with biomass reference point and fishery catch projections.

In addition to the ADAPT VPA and ASAP MULTI Developmental models, Rademeyer and Butterworth (MS 2011b) provided an implementation of an Age-Structured Production Model (ASPM), in which they explored approaches to the reduce the retrospective errors in the SNE/MA assessment. Rademeyer and Butterworth (MS 2011b) implemented both autocorrelation in survey q variability and a "ramped" increase in M over time (10% per year across all ages, from 1995-2005, increasing M from 0.3 in 1995 to 0.6 in 2005 and later years). This configuration in the ASPM greatly reduced the retrospective in SSB and R (Figure A46). Due to the combination of University of Cape Town (Republic of South Africa) intellectual property and NMFS policy issues, however, the Rademeyer and Butterworth (MS 2011b) ASPM model was not eligible to be used as the final assessment model.

The concept of an increasing trend in M over the assessment time series was incorporated into the ADAPT VPA and ASAP BASE models in several configurations, with the goal of reducing the retrospective patterns. The autocorrelation in *q,* however, was not able to be incorporated in ASAP in the time available, as it would require programming changes. The change in M in the ADAPT VPA and ASAP models was incorporated both as a "ramp" of 10% per year from 0.3 to 0.45 or 0.6, beginning in 1994 or 2000, and as a "step" in M from 0.3 to 0.45 and from 0.3 to 0.6 in the year 2000. The retrospective errors observed for each of these model configurations are summarized and compared with the ADAPT VPA and ASAP SPLIT survey configurations with comparable values of M for all ages and years in Table A37. Incorporation of the "ramps" and "steps" in M in the BASE model configurations was effective in reducing the retrospective errors from 40-50% in the ADAPT VPA SPLIT models to 25-35% in ADAPT VPA BASE models. For the ASAP models, the range retrospective errors were reduced from over 50% to 13-18% (Table A37).

Based on these results and diagnostics, along with the inspection of residual patterns, the SDWG adopted the ASAP MULTI BASE model configuration CAT10 as the preferred model to move forward for further consideration, as it provided a more advanced and flexible model when compared to ADAPT VPA. The SDWG had extensive discussions about the implications of incorporating either a "ramp" or "step" in M to 0.6 in final models used for estimation of reference points and status determination, and concluded that based on analogy to the VPA SPLIT survey model configuration, the "step" approach was a better alternative. The SDWG elected to provide the ASAP CAT10 configuration (MULTI survey configuration, BASE survey *q* configuration, annual fishery ESS = 10, annual survey ESS = 10, M = 0.3 for all years and ages, no internal stock-recruitment function estimation) as the preferred final, or "best," model for status determination. The retrospective pattern in this model is moderate, but comparable to that deemed acceptable in the 2008 GARM-III assessment (NEFSC 2008). The SDWG has also brought forward a model incorporating a "step" from M = 0.3 during 1981-1999 to M = 0.6 in 2000-2010 (the STEPM model) as an alternative that provides reduced retrospective errors, but that also provides a substantially different perception of stock productivity, or "state of nature," for SNE/MA winter flounder in 2010 and beyond if M = 0.6 is assumed in the future.

The three model configurations were carried through the calculation of reference points and calculation of Frebuild and ABCs for 2012, although the results of the STEPM model are presented in less detail in subsequent portions of this report. The trends in F, SSB, and R for the preferred CAT10 model and the alternative STEPM model are compared in Figures A47-A49. The STEPM model provides lower estimates of F during the mid-1990s and early 2000s and higher estimates of F since 2006, and higher estimates of SSB during the mid-1990s and early 2000s and lower estimates of SSB since 2005. The STEPM model provides higher estimates of recruitment at age 1 throughout the assessment time series.

2011 SAW52 Final Assessment Model and Results

The ASAP CAT10 model configuration serves as the basis for evaluating the status of the stock and providing catch advice. The assessment indicates that during 1981-1993, fishing mortality (F ages 4-5) varied between 0.61 (1982) and 0.95 (1993) and then decreased to 0.47 by 1999. Fishing mortality then increased to 0.70 by 2001, and has since decreased to 0.051 in 2010, generally tracking the decrease in fishery catch (Table A38, Figure A50). SSB decreased from 20,100 mt in 1982 to a record low of 3,900 mt in 1993, and then increased to 8,900 mt by 2000. SSB has varied between 4,500-8,000 mt during 2001-2009, and was 7,076 mt in 2010 (Table A38, Figure A51). Recruitment at age 1 decreased nearly continuously from 71.6 million age-1 fish in 1981 (1980 year class) to 7.5 million fish in 2002 (2001 year class).

Recruitment has averaged 10.5 million during 2003-2010 (Table A38, Figure A51). The fishery selectivity pattern in the first time block (1981-1993) was estimated to be 0.01 at age 1, 0.24 at age 2, 0.75 at age 3, was fixed at 1.00 at age 4, was estimated at 1.00 at age 5, 0.99 at age 6, and 1.00 at age 7+. The pattern in the second time block (1994-2010) was estimated to be 0.01 at age 1, 0.19 at age 2, 0.70 at age 3, was fixed at 1.00 at age 4, was estimated at 0.97 at age 5, 0.89 at age 6, and 0.67 at age 7+.

The precision of the 2010 fishing mortality (F ages 4-5) and SSB was evaluated using MCMC techniques. One thousand MCMC iterations were realized (200,000 calculations with a thinning rate of 200). There is an 80% probability that F ages 4-5 in 2010 was between 0.04 and 0.06 (Figure A52). There is an 80% probability that SSB in 2010 was between 6,433 mt and 8,590 mt (Figure A53).

Retrospective analysis for the 2003-2010 terminal years indicates retrospective error in fishing mortality (F) ranged from -38% in 2006 to -13% in 2009, retrospective error in SSB ranged from +42% in 2004 to +12% in 2009, and retrospective error in recruitment at age 1 (R) ranged from +78% in 2005 (2004 year class) to -11% in 2009 (2008 year class; Figures A54-A56).

Model fits to the aggregate survey indices (for those with multinomial age compositions) and recruitment indices are provided in Figures A57-A60. For the NEFSC Spring, Fall, and Winter surveys expressed as swept area numbers, aggregate survey catchability (q) was estimated at 0.126, 0.617, and 0.253, respectively. The other calibration surveys are of more limited geographic extent and were input in their original units, and therefore q estimates for those surveys ranged from 0.00001 (MADMF summer seine survey age 0 index) to 0.0017 (CTDEP spring trawl survey). Fishery age composition simple residuals (observed minus predicted proportions at age) are presented in Figure A61. There are some large positive residuals (about 15% in real terms) early in the time series, and some large negative residuals (10-15% in real terms) early in the time series at ages 2 and 4, and again in 2010 at age 3. However, there were no problematic, extensive "runs" of large residuals evident for the fishery catch proportions at age.

A comparison between the results of the current assessment and the five previous assessments is presented in Figures A62-A64. This "historical retrospective" illustrates the underestimation of fishing mortality and overestimation of SSB that has been present between assessments since 1995. This pattern is in addition to the persistent "internal retrospective" that has been present in each of the assessments. The SDWG notes that the current assessment with assumed M = 0.3 is not consistent with those previous which assumed M = 0.2, and that much of the upward magnitude shift in numbers and biomass and downward shift in fishing mortality is due to this change.

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 as a result of the allocation, and then perform an exercise to run the assessment model with those potential biases and report the results. For the SNE/MA stock the total catch consists of 4 components. The commercial landings have a calculated Proportional Standard Error (PSE; due to the aforementioned commercial landings area-allocation procedure; available for 1995 and later years, with the mean of those years substituted for 1981-1994) ranging from <1% to about 7%; the commercial discard PSEs range from 17-35% (available for 1994-2010, mean of those years substituted for 1981-1993); the recreational landings PSEs range from 17-40%; and the recreational discard PSEs range from 18-57%.

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 3.1-21.3%, and average 8% (un-weighted over years) for the 1981-2010 time series. The SDWG developed such an exercise using the 2008 GARM-III assessment data and ADAPT VPA model in an initial response to TOR4 and concluded that the application of a annually varying "bias-correction" in one direction in such an exercise provides stock size estimates and BRPs that scale up or down by about the same average magnitude as the gain or loss (Terceiro MS 2011a).

Since the initial exercise, the SDWG concluded that the calculated variance of the area-allocated commercial landings likely underestimates the true error. More 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 NEFSC Fishery 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 update the exercise using the final SNE/MA assessment ASAP model, with an additional 5% PSE in commercial landings added to the currently estimated 0.4 to 4.5% over the 1995-2010 time series. This increased the average commercial landings PSE from 0.9% to 3.7%, and increased the overall catch PSE from 8% to 10%, ranging from 4.9% in 1992 to 23.7% in 2010. The catch in the final assessment model was increased and decreased by the annually varying PSE and the adjusted models run to provide an additional measure of uncertainty of assessment estimates. As in the previous version of the exercise, the application of a annually varying "bias-correction" in one direction in such an exercise provides stock size estimates that scale up or down by about the same average magnitude as the gain or loss. For the final ASAP CAT10 model, fishing mortality on average changed by +/- 0.3%, and the range in 2010 F was 0.04 to 0.05, comparable to the MCMC estimate of uncertainty. SSB on average changed by +/- 9.0%, and the range in 2010 SSB was 6,500 to 7,600 mt, within the MCMC estimate of uncertainty (Figure A65).

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).

For the full presentation of the SDWG response to this TOR see Hare MS 2011 (SDWG52 WP13).

Winter flounder spawn in winter and early spring in estuaries along the mid-Atlantic, southern New England and Gulf of Maine coasts, as well as in continental shelf waters on Georges Bank (Able and Fahay 2010). There is also recent evidence of more coastal spawning in both Southern New England (Wuenschel et al. 2009) and in the Gulf of Maine (Fairchild et al. 2010). In southern New England, Manderson (2008) found that overall recruitment was linked to spring temperatures, presumably by acting on larvae, settlement stage, and/or early juveniles. Further, Manderson (2008) found that young-of-the-abundance among 19 coastal nurseries became more synchronized in the early 1990's and argued that increased frequency of warm springs was creating coherence in early life stage dynamics among local populations.

The specific mechanism linking temperature to recruitment was not defined by Manderson (2008), but temperature is an important parameter in many ecological processes affecting winter flounder. In a mesocosm study, Keller and Klein-MacPhee (2000) found that winter flounder egg survival, percent hatch, time to hatch, and initial size were significantly greater in cool mesocosms. Further, mortality rates were lower in cool mesocosms and related to the abundance of active predators. In the laboratory, Taylor and Collie (2003) found that consumption rates of sand shrimp were lower at lower temperatures implying lower predation pressure at colder temperatures. In the field, Stoner et al. (2001) found that settlement stage winter flounder prefer colder waters and that the importance of temperature in defining juvenile habitat decreases through ontogeny. Thus, temperature has multiple effects on the early life history of winter flounder and colder temperatures in general lead to higher survival and recruitment.

The relationship between winter flounder recruitment and temperature identified by Manderson (2008) did not include the effect of population size. The relationship between stock size and subsequent recruitment is generally poor in marine fishes (Rothschild 1986) but can have explanatory power. To examine the combined effect of environment and spawning stock biomass on recruitment, the goal here was to develop environmentally-explicit stock-recruitment relationships that include temperature and related environmental variables for the three U.S. stocks of winter flounder. As a basic framework, the approach of Hare et al. (2010) was followed. The resulting models could be used in short-term forecasts based on fishing and temperature scenarios (fixed patterns of temperature variability over several years) and long-term forecasts based on fishing and temperature projections from general circulation models.

To develop environmentally-explicit stock-recruitment relationships, three specific types of data are required: spawning stock biomass, recruitment, and environmental data. For the SNE/MA stock, recruitment (lagged by 1 year) and spawning stock biomass pairs were used from the ASAP CAT10 model (Table A38). Two general types of temperature data were used: air temperatures and coastal water temperatures (Table A39). Air temperature data from the NCEP/NCAR reanalysis (Kalnay et al. 1996) were used. This product combines observations and an atmospheric model to produce an even grid of atmospheric variables, in our case monthly mean surface air temperature. The spatial resolution is 2.5^{o} latitude by 2.5^{o} longitude. Air temperatures are closely related to estuarine water temperatures owing to efficient heat exchange in the shallow systems (Roelofs and Bumpus 1953, Hettler and Chester 1982, Hare and Able 2007). Data from representative grid points were averaged for each of three regions, and the monthly/regional averages were further averaged into annual estimates for three, two monthly periods (January-February, March-April, May-June).

Coastal water temperature data from Woods Hole, Massachusetts and Boothbay Harbor, Maine were available; the Woods Hole data were used for SNE/MA stock analyses (see Nixon et al. 2004 and Lazzari 1997 respectively). Monthly means were calculated from mostly daily data. These monthly means were then averaged into annual estimates for the three, two monthly periods (January-February, March-April, May-June). Temperature data were analyzed as annual averages for three, two month periods (January-February, March-April, May-June). These two monthly periods capture temperature variability from the late winter, through spring and into early summer. The spring period was identified as important by Manderson (2008). The broader seasonal range was chosen because of potential differences in the timing of winter flounder spawning and development among the three stocks (Able and Fahay 2010) and the uncertainty as to the stage where recruitment is determined.

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). Brodziak and O'Brien (2005) identified a significant effect of NAO on recruit-spawner anomalies of winter flounder in the Gulf of Maine. The mechanism is unspecified, but NAO is related to estuarine water temperatures in the region (Hare and Able 2007). The winter NAO index is used here (Hurrell and Deser 2010). 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). Nye et al. (2009) found the AMO was strongly related to distribution shifts of fishes in the northeast U.S. shelf ecosystem. 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. Work by Nye et al. (in review) shows the Gulf Stream index has explanatory power for the distribution of silver hake in the system, possibly through the large-scale linkages between the Gulf Stream, Labrador Current and hydrographic conditions on the northeast U.S. shelf. Two Gulf Stream indices are used here (Joyce and Zhang 2010 and Taylor and Stephens 1998). The two indices differ in their calculation, with the Joyce and Zhang (2010) index more associated with the Gulf Stream south of the northeast U.S. shelf and the Taylor and Stephens (1998) index more associated with the Gulf Stream across the North Atlantic. For all four large-scale forcing indices, annual values were obtained. Numerous studies have found lagged effects of the NAO on the northeast U.S. shelf ecosystem (Greene and Pershing 2003, Hare and Kane in press). In particular, a two year lag has been related to the remote forcing of the NAO on the northeast U.S. shelf through the Labrador Current system. In addition, a zero year lag has been related to direct atmospheric forcing on the northeast U.S. shelf. Zero, one, and two year lags of were included for NAO and zero year lags were used for the other three large-scale forcing variables. To understand the relations between the 21 environmental variables, a simple correlation matrix was calculated. Significant correlations were considered in the context of previous research in the region. Significance was based on standard p-values; no corrections for multiple comparisons were made. The purpose was exploratory with an aim of understanding the relation between variables before incorporating them into stock recruitment functions.

Ricker, Beverton-Holt, and Cushing stock recruitment models were used with and without the different environmental terms. The model forms followed Levi et al. (2003), who built upon the ideas of Neill et al. (1994) and Iles and Beverton (1998). The fits of the three standard models were all very similar for the SNE/MA stock. Owing to the general acceptance of the Beverton-Holt model for use in stock-recruitment relationships and the overall similarity in the fits of the three models, here only the analyses using the Beverton-Holt model are presented. Environmental variables were assigned *a priori* for consideration with

specific stocks. This was done to limit the number of environmentally-explicit stock recruitment relationships considered for each stock.

The standard stock-recruitment relationships were calculated first using the lsqcurvefit function in MatLab using the trust-region-reflective algorithm. A series of environmentally-explicit models also were fit using the same methods. The resulting models were compared using AICc and AICc weights, which represent the relative weight of evidence in favor of a model. The best environmentally-explicit model also was compared to the standard stock recruitment model using an evidence of weights procedure (Burnham and Anderson 1998). In this way the value of the environmentally-explicit stock recruitment functions relative to standard stock recruitment functions was judged. Model fitting included bounded parameters (or priors) to force realistic model forms.

Numerous relationships between environmental variables were evident based on the correlation analysis. The two Gulf Stream indices were related (r=0.54) but different enough to retain both in the analyses. Both Gulf Stream indices were related to the NAO with a 2 year lag (NAO leading). This relationship has been described before (Taylor and Stephens 1998). The Atlantic Multidecadal Oscillation exhibited relatively little relationship with other variables. The North Atlantic Oscillation was related to the two Gulf Stream indices as already noted. NAO was not related to winter temperatures which may result from non-stationarity in the NAO-winter temperature relationship (Joyce 2002). Woods Hole temperature is closely related to regional air temperatures. This link is not surprising based on previous studies. There is evidence of seasonal correlation in Woods Hole temperature, with values in January and February correlated to values in March and April, which in turn are correlated to values in May and June. However, the seasonal correlation is diminished after two months; temperatures in January and February are less related to temperatures in May and June.

The three air temperature series were all closely related indicating coherent air temperatures over the entire region. These analyses agree with the more comprehensive results of Joyce (2002). Correlations among regions over the same time (Jan-Feb) were higher than correlations within region between times (Gulf of Maine Jan-Feb compared to Gulf of Maine Mar-Apr). Seasonal correlation (Jan-Feb to Mar-Apr) were lower in the air temperature series compared to the water temperatures series as expected from the greater specific heat capacity of water.

The analyses suggest that the environmental forcing experienced by the three stocks differs in several important elements. The SNE/MA stock experiences coastal water temperatures that are strongly linked to local air temperatures. The GBK stock experiences water temperatures that are affected by both local air temperatures and more importantly, large-scale advective supply of relative cold, fresh water associated with the Labrador Current. Finally, the temperatures experienced by the GOM stock remain uncertain. If the Boothbay Harbor data is representative, then temperature is related to large-scale processes (AMO) and not local processes (air temperature). On the other hand, air temperature may be important, if early stage winter flounder are using shallower habitats.

Spawning stock biomass is comparable between the SNE/MA and GBK stock but recruitment is approximately four times greater for the SNE/MA stock at higher stock sizes (Figure A66). The stock recruitment functions for the GBK and GOM stock are similar, with near constant recruitment over a relatively broad range of spawning stock biomasses. Recruitment on Georges Bank is estimated to be higher than in the Gulf of Maine at a given spawning stock biomass.

The residuals of the stock-recruitment relationships for the three stocks appear to exhibit synchrony through time (Figure A67). Early in the time series, residuals between the stocks appear unrelated, but all residuals were positive in the mid 1990s and all were negative in the early 2000s. A formal analysis was conducted using serial correlation: calculating the correlation coefficient between two variables using a moving window. A similar analysis was used by Joyce (2002) to show that the relationship between NAO and east coast air temperatures has changed over the last 80 years and by Hare and Kane (in press) to show that the correlation between NAO and *Calanus finmarchicus* abundance has changed over the last twenty years. The serial correlation analysis demonstrated that early in the time series the residuals of the stock-recruitment functions were negatively or not correlated between the stocks (Figure A68). Then, during the early 1990s, the residuals became positively correlated. The trend is most evident for the SNE/MA and GOM stocks and less so for these two stocks compared to the GBK stock.

The timing in the synchrony between the SNE/MA and GOM stocks is similar to the timing in synchrony among local populations within the SNE/MA stock (Manderson 2008). This synchrony suggests that some large-scale forcing is responsible for creating variance in the stock recruitment relationships of winter flounder across the northeast U.S. shelf ecosystem. The synchrony is greater between the SNE/MA and GOM stocks suggesting that the large-scale forcing has greater coherence along the coastal areas of the northeast compared to the offshore waters of Georges Bank.

The best fit environmentally-explicit stock recruitment relationship for the Southern New England stock predicted higher recruitment at lower winter air temperatures (Table A40, Figure A69). The variable in the best model was Southern New England air temperature in January and February. This model had an evidence ratio of 106 compared to the standard model and explained an additional 14% of the variance (Table A41). Several other environmental variables were included in the top ten models (AMO, GS-J, and WH-JF), but three of the four top models included winter air temperatures over Southern New England. The best environmentally-model provided a similar function to the standard model at mean environmental conditions, but importantly the predicted asymptotic recruitment was lower with the environmental model.

The environmentally-explicit models support the hypothesis that increased temperatures during spawning and the early life history result in decreased recruitment in the SNE/MA stock. Winter temperature is correlated with spring temperature providing a potential bridge between this study and that of Manderson (2008). Using the same serial correlation approach to examine trends in winter air temperature shows an increase in correlation among the three regions starting in the late-1980's early-1990's. The correlation coefficients of Southern New England and Gulf of Maine air temperatures are correlated with the similar coefficients for recruitment. This result suggests that as regional air temperatures have become more coherent, winter flounder recruitment in the coastal stocks also has become more coherent.

To consider these environmentally explicit models stock recruitment models in the context of reference points, it is necessary to summarize model parameters. For the SNE/MA stock, an important issue in the standard stock recruitment model is the perceived need to bound the model parameters in both the prior stock assessment (NEFSC 2008) and in the current assessment. Specifically, the standard model estimates a high asymptotic recruitment (Table A42). Bounding asymptotic recruitment to the mean observed in a series of high recruitment years results in a very different model. At the mean environmental conditions, the unbounded environmentally-explicit model has a lower asymptotic recruitment (Table A42) and one benefit of this model is the lack of need for bounded parameters.

Another potential benefit for the environmentally explicit models is to forecast recruitment under different environmental conditions. Over the assessment record, there has been no change in winter air temperature (Figure A70). Further, the ability to forecast winter air temperatures in the 1-5 year range is limited at best. There is some skill in statistical seasonal forecasts with several months lead time (Cohen and Fletcher 2007) and developing forecast skill on the decadal scale is a major topic of research in the climate modeling community (Smith et al. 2007, Keenlyside et al. 2008), but interannual forecasts with demonstrated skill are few. Thus, the environmental models developed here can be used with a mean environment to calculate reference points. Additionally, scenarios could be evaluated calculating reference points under an assumption of warm winters and an assumption of cool winters to better inform management in the short-term.

The results of the analyses support Manderson's (2008) earlier finding. Recruitment in coastal stocks of winter flounder is related to temperature during the spawning season. Importantly, recruitment is also dependent on spawning stock biomass and the environmentally-explicit stock-recruitment models capture the combined effect of environment and stock size. The temperature effect is strongest in the Southern New England stock, where the species is at the southern extent of its range. The signal is less pronounced in the Gulf of Maine, but recruitment is still linked to winter temperatures. The effect of environment on recruitment of Georges Bank winter flounder is less clear. There is a lot of variability in the stock-recruitment relationship and none of this variability is explained with the environmental terms considered here. Whether other environmental factors play a role in Georges Bank winter flounder recruitment is an important question requiring future research.

The closer link to air temperatures for the Southern New England stock is explained by the argument that water temperatures in estuarine winter flounder spawning, larval, and juvenile habitats are more closely related to air temperature than to coastal water temperatures. Prior studies have found a close link between air temperature and estuarine water temperature (Hare and Able 2007). Future studies should explicitly treat the spatial dynamics of winter flounder in more detail (see Manderson 2008); such an approach could better examine the effect of environmental forcing on local populations.

One use of the environmentally-explicit models is to develop short-term and long-term forecasting models. Based on the above analyses, 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 year time scale.

Work is underway within the SDWG to incorporate environmentally-explicit stock-recruitment models into the standard NFT software used to fit stock-recruitment models and to perform stock and fishery projections. 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 B_{MSY}, $B_{THRESHOLD}$, and F_{MSY}) 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.

In addition to the SNE/MA stock results presented below, the SDWG developed a unified response to TOR6 taking into consideration the assessment results for all three stocks, as presented in SDWG Working Paper D. 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 FSMY 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 if FSMY can be reliably estimated. The SDWG therefore estimated FSMY as well as proxies in the form of F40%. The SDWG developed consensus on some aspects of the FMSY estimates in terms of their relative magnitude across stocks, but also had some disagreement about the reliability of FMSY estimates that were related to the perceived reliability of the respective assessments. The SDWG could not come to consensus on the preferred reference points, and 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 Percent Maximum Spawning Potential (%MSP) alternatives.

The Working Group on Re-Evaluation of Biological Reference Points for New England Groundfish (NEFSC 2002) estimated biological reference points for SNE/MA winter flounder using Yield Per Recruit (YPR) and SSB per Recruit (SSBR) analyses (Thompson and Bell 1934) and Beverton-Holt stock-recruitment models (Beverton and Holt 1957, Brodziak et al. 2001, Mace and Doonan 1988) based on the SAW 28 assessment results (NEFSC 1999). A Beverton-Holt stock-recruitment model fit with a prior on unfished recruitment (R0) equal to the average of the five largest year classes (1981-1985) in the VPA time series was selected as the best stock-recruitment model. The YPR and SSBR analyses indicated that F0.1 = 0.25 and F40%= 0.21. The NEFSC (2002) stock-recruitment model indicated that MSY = 10,600 mt, FMSY= 0.32, and SSBMSY= 30,100 mt.

Both the parametric Beverton-Holt stock-recruitment model and the "non-parametric empirical" approach (YPR and SSBR model combined with VPA recruitment estimates and long-term projections) were considered in the 2008 GARM-III assessment to estimate biological reference points for SNE/MA winter flounder, based on the BASE and SPLIT VPA results. Stock-recruitment data were modeled for the 1981-2007 year classes (1981-2007 SSB; 1982-2008 recruitment at age 1). In the non-parametric empirical approach, a long-term (100 year) stochastic projection using the cumulative distribution function of the year classes produced when SSB exceeded 5,700 mt was used to estimate MSY and SSBMSY.

The 2008 GARM-III Biological Reference Point Review Panel (NEFSC 2008) concluded that the prior on unfished recruitment used to fit the parametric Beverton-Holt stock-recruitment model in the NEFSC (2002) work was inappropriate. The Beverton-Holt stock-recruitment model fit without the prior or with a prior on steepness (*h*) did not provide feasible results. The Panel therefore recommended the non-parametric empirical approach be used to estimate biological reference points for SNE/MA winter flounder based on a) the GARM-III SPLIT VPA results, b) the estimate of F40% as a proxy for FMSY, and c) a long-term (100 year)

stochastic projection using the cumulative distribution function of the year classes produced when SSB exceeded 5,700 mt (1981-1988 year classes; mean R = 35.239 million fish at age 1) to estimate MSY and SSBMSY. The 2008 GARM-III BRPs were F40% = 0.248 (proxy for FMSY and the fishing mortality threshold for overfishing), SSB40% = 38,861 mt (proxy for SSBMSY), and MSY40% = 9,742 mt (proxy for MSY). The biomass threshold was therefore 19,381 mt (proxy threshold for overfished).

In the current assessment for SNE/MA winter flounder, FMSY, SSBMSY, and MSY BRPs were estimated from an external stock-recruitment model for both the final CAT10 model and the alternative STEPM model estimates with future M = 0.3 or future M = 0.6 (Figure A71). Stock-recruitment parameters using no prior, a prior on steepness (h = 0.8; CV = 0.09; as in NEFSC 2002, as derived from Myers et al. 1999), and a prior on unfished recruitment (R0; mean of the five largest estimated recruitments [1981-1985] as in NEFSC 2002) were estimated. Proxy BRPs based on 40% MSP were also estimated for the models. Table A43 summarizes the stock-recruitment model fit results, and Table A44 summarizes the YPR and SSBR calculation results. For the final CAT10 model, the stock-recruitment model with a prior for steepness (h) was judged to fit best while providing feasible results (Figures A72-A73); for the two STEPM models, the fits with no priors were judged to fit best while providing feasible results (Figures A74-A77). YPR and SSBR calculations were used with fishery selectivity estimates for all three model configurations to provide 40%MSP based proxy BRPs.

The SARC 52 review panel concluded that steepness should be similar between the three winter flounder stocks in Northeast U.S waters. Therefore, FMSY, SSBMSY, and MSY were estimated from a stock-recruitment model using a range of values for steepness (slope of the stock recruitment curve near the origin) which was consistent with the stock-recruitment data. In computing the BRPs, values of steepness were chosen which were constructed to be as similar as possible between stocks, while also providing good fits to the stock recruitment data for each stock. For the SNE/MA stock, steepness was therefore set at 0.61, based on the likelihood profile over a range of fixed steepness values. The final recommended biological reference points for SNE/MA winter flounder are FMSY = Fthreshold = 0.290, SSBMSY = Btarget = 43,661 mt, 1/2 SSBMSY = Bthreshold = 21,831 mt, and MSY = 11,728 mt. For comparison, F40% computed using the same biological and fishery characteristics is 0.327, with SSB40% = 29,045 mt and MSY40 = 8,903 mt (Figures A78-A80).

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.

Table A45 summarizes the existing 2008 GARM-III BRPs for SNE/MA winter flounder (NEFSC 2008) and the recommended BRPs from the current assessment. In the current assessment, the assumed value for M has been increased from 0.2 to 0.3, and so the SDWG concluded that comparison of current assessment F and SSB estimates with the existing 2008 GARM-III reference points was not appropriate. The summary stock status statements below are based on the three assessment models and associated BRP configurations.

ASAP CAT10 M = 0.3

The Southern New England/Mid-Atlantic (SNE/MA) winter flounder stock complex was overfished but overfishing was not occurring in 2010 (Table A45, Figures A81-A83). Fishing mortality (F age 4-5) in 2010 was estimated to be 0.051, below FMSY = 0.290 (18% of FMSY) and below F40% = 0.327 (16% of F40%).

SSB in 2010 was estimated to be 7,076 mt, about 16% of SSMSY= 43,661 mt and 24% of SSB40% = 29,045 mt.

The SDWG recommends the ASAP CAT10 M = 0.3 model with stock-recruitment model based MSY BRPs as the basis for current and future stock status. The SDWG acknowledged the persistent retrospective pattern in this model, but does not recommend any adjustment to the 2010 assessment estimates.

ASAP STEPM M = 0.3

The Southern New England/Mid-Atlantic (SNE/MA) winter flounder stock complex was overfished but overfishing was not occurring in 2010 (Table A45). Fishing mortality (F age 4-5) in 2010 was estimated to be 0.087, below FMSY = 0.325 (27% of FMSY) and below F40% = 0.327 (27% of F40%). SSB in 2010 was estimated to be 4,144 mt, about 10% of SSMSY= 42,770 mt and 13% of SSB40% = 31,311 mt.

The SDWG provides the STEPM M = 0.3 model and associated BRPs as an alternative that reduces the persistent retrospective pattern in the model, while projecting that M, as a proxy for the factors that cause the retrospective patterns, will return to the base value of 0.3 in the future. The SDWG acknowledges that some retrospective pattern remains in this model, but does not recommend any adjustment to the 2010 assessment estimates.

ASAP STEPM M = 0.6

The Southern New England/Mid-Atlantic (SNE/MA) winter flounder stock complex was not overfished and overfishing was not occurring in 2010 (Table A45). Fishing mortality (F age 4-5) in 2010 was estimated to be 0.087, below FMSY = 0.145 (60% of FMSY) and below F40% = 0.652 (13% of F40%). SSB in 2010 was estimated to be 4,144 mt, about 60% of SSMSY= 6,899 mt and 60% of SSB40% = 6,926 mt.

The SDWG provides the STEPM M = 0.6 model and associated BRPs as an alternative that reduces the persistent retrospective pattern in the model, while projecting that M, as a proxy for the factors that caused the retrospective patterns, will remain at an elevated value of 0.6 in the future. The SDWG notes that the ASAP STEPM M = 0.6 model configuration and associated BRPs with future M = 0.6 provides substantially different perceptions of stock productivity, or "state of nature," for SNE/MA winter flounder both historically and in 2010 and beyond if M = 0.6 is assumed in the future, compared to assessment models and BRPs with M = 0.3. The SDWG did not come to consensus on whether the STEPM M=0.6 configuration provides a feasible assessment of SNE/MA winter flounder stock status in 2010 or into the future.

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.

8a. Projections of future stock status were made based on the current assessment results for both the CAT10 and STEPM models and corresponding BRPs. Mean weight, maturity and fishery selectivity patterns at age estimated for the most recent 5 years in the assessment (2006-2010) were used to reflect current conditions in the stock and fishery. Recruitment was projected using stock-recruitment models for the MSY-based BRPs, while two-stage recruitment models (resample the cumulative density function [cdf] of the lowest 23 year classes [1986-2010] for SSB less than 10,000 mt; resample the cdf of the highest 5 year classes [1981-1985] for SSB greater than 10,000 mt) were used for the 40%MSP based BRPs, to ensure that the magnitude of short-term recruitment would be appropriate for the magnitude of SSB. The projections assumed the FMP Framework 44 fishing year (May 1) catch of 842 mt would be landed as a calendar year (Jan 1) catch in 2011.

ASAP CAT10 M = 0.3

A catch of 842 mt in 2011 is projected to provide median F2011 = 0.100 and median SSB2011 = 9,177 mt. Projections at F = 0.000 in 2012-2014 indicate less than a 1% chance that the stock will rebuild to SSBMSY = 43,661 mt by 2014, and less than a 4% chance that the stock will rebuild to SSB40% = 29,045 mt by 2014.

ASAP STEPM M = 0.3

A catch of 842 mt in 2011 is projected to provide median F2011 = 0.174 and median SSB2011 = 4,720 mt. Projections at F = 0.000 in 2012-2014 indicate less than a 1% chance that the stock will rebuild to SSBMSY = 42,770 mt by 2014, and less than a 1% chance that the stock will rebuild to SSB40% = 31,311 mt by 2014.

ASAP STEPM M = 0.6

A catch of 842 mt in 2011 is projected to provide median F2011 = 0.202 and median SSB2011 = 4,429 mt. Projections at F = 0.000 in 2012-2014 indicate less than a 4% chance that the stock will rebuild to SSBMSY = 6,899 mt by 2014, and a 31% chance that the stock will rebuild to SSB40% = 6,926 mt by 2014.

8b. 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 consistent with stock status determination and projection. Vulnerability and susceptibility were accounted for in both aspects of status determination (estimation of F and FMSY) and projections as the magnitude of fishing mortality and recent fishery 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 by 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 (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.

A considerable source of additional vulnerability is the continued weak recruitment and low reproductive rate (e.g., recruits per spawners) of SNE/MA winter flounder (Figure A84). If weak recruitment and low reproductive rate continues, productivity and rebuilding of the stock will be less than projected. Stock-recruit modeling suggests that warm temperatures are having a negative effect on recruitment of SNE/MA winter flounder.

8c. The primary Research Recommendations from the 2008 GARM-III assessments for winter flounder were: "Assessment approaches needs [*sic*] 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." and "The Panel also had concerns about the unit stock, not only for this stock, but for all of the Winter Flounder stocks assessed. It recommended an analysis of Winter Flounder as a stock complex, rather than as individual stocks, be undertaken" (NEFSC 2008).

As noted earlier, the stocks are defined as they are now based on a) historical tagging studies that show low rates of exchange (a few percent) between the stock areas (Howe and Coates 1975; Pereira *et al*. 1999), b) differences in the growth rates between the stocks, with GBK fish growing faster, GOM fish

growing slower, and SNE fish growing at an intermediate rate (How and Coates 1975; Lux 1973; NEFSC 2008), c) differences in the rates of maturation (NEFSC 2008), d) differences in meristics, mainly fin ray counts (Lux *et al*. 1970), and e) fishery "integration" of catches from potential bay/estuarine specific-stocks in the GOM and the SNE "complexes."

Terceiro (MS 2011b) provided an exercise which responded to the GARM-III Research Recommendations aggregating all 3 stocks together in an "All Stocks" winter flounder ADAPT VPA (back-calculating model) - i.e., to assume 100% "interaction". Stock size and fishing mortality rate estimates from the combined analysis were a "blend" of the three GARM assessment results, as might be expected. Aggregation of the three stock units resulted in a larger aggregate spawning stock biomass reference point and MSY estimate, while the aggregate stock status remained overfished with overfishing occurring in 2007. The combined analysis exhibited a reduced retrospective pattern compared to those in the GARM-III GOM and SNE assessments (recent overestimation of SSB ranging from 8-15%; underestimation of F ranging up to 22%).

However, the SDWG notes that the exercise violated the existing assumptions of stock structure based on information about the biology, migration patterns, and fishing patterns for winter flounder. The SDWG concludes that the information available on winter flounder stock structure provides strong support for the current three stock units, and that attempts to model those units as a single complex are not worth pursuing further. The SDWG does not believe that the benefits from the single-stock analysis (a single analysis instead of three; reduced retrospective pattern; ability to model the Gulf of Maine unit within the complex) are sufficient to ignore the observed differences in biological traits (growth, maturity, fecundity) that affect the interpretation of the spawning stock reproductive potential of the three current units.

The SDWG has initiated further research pursuing use of a more complex model (i.e., Stock Synthesis) to maintain separate fishery and survey catch for the three current stock units, while allowing a small amount (a few percent) of exchange between the stock units based on information from historical tagging. However, development of that research has not progressed sufficiently to be made available for peer review at this time.

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.

Previous from 2002 SAW 36:

1) Evaluate the maturity at age of fish sampled in the NEFSC fall and winter surveys.

Fall survey data have been evaluated; winter survey samples have not been processed

2) Consider fieldwork to record ovary weights along with maturity stage data from 20-30 cm fish in the NEFSC and State agency surveys for 1-2 years to help resolve age/size at maturity differences between State and NEFSC surveys.

See McBride et al MS 2011

3) Conduct periodic maturity staging workshops involving State and NEFSC trawl survey staff.
Not addressed, but recommended as new RR3

4) Examine sources of the differences between NEFSC, MA and CT survey maturity (validity of evidence for smaller size or younger age at 50% maturity in the NEFSC data). Compare NEFSC inshore against offshore strata for differences in maturity. Compare confidence intervals for maturity ogives. Calculate annual ogives and investigate for progression of maturity changes over time. Examine maturity data from NEFSC strata on Nantucket Shoals and near George=s Bank separately from more inshore areas. Consider methods for combining maturity data from different survey programs.

Significant work completed for this assessment, and see McBride et al MS 2011

5) Increase the intensity of commercial fishery discard length sampling.

Completed for 2008 GARM 3 - adopted SBRM algorithm and increased sample request

6) Consider post-stratification of NEFSC survey offshore stratum 23, to facilitate inclusion of survey catches from this stratum (east of Cape Cod) in the SNE-MA winter flounder assessment.

See GBK winter flounder assessment – stratum 23 used in GBK assessment based on characteristics of age samples

7) Incorporate State samples (e.g. NY DEC Party Boat Survey and CT DEP Volunteer Angler Survey) in the estimation of recreational fishery landings and discards, if possible.

Completed for 2008 GARM 3

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

Completed for 2008 GARM 3; see TOR3 current assessment final model

9) Continue to consider the effects of catch-and-release components of recreational fishery on discard at age (i.e., develop mortality estimates from the American Littoral Society tagging database, if feasible).

Not addressed

10) Compare commercial fishery discard estimates from the Mayo survey/mesh algorithm with those from VTR data for comparable time periods.

Completed for 2008 GARM 3 - adopted SBRM algorithm; see TOR 1

11) Maintain or increase sampling levels (currently supported by individual state funding) and collect age information from MRFSS samples.

Not addressed

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

Not directly addressed - although the power plant on Mount Hope Bay in MA has built two large cooling towers in part to reduce larval fish mortality

13) Examine the implications of stock mixing from data from Great South Channel region.

See Terceiro MS 2011b

14) Expand sea sampling for estimation of commercial discards.

Completed for 2008 GARM 3 - adopted SBRM algorithm and increased sample request; see TOR 1

15) Revise the recreational fishery discard estimates by applying a consistent method across all years, if feasible (i.e., the Gibson 1996 method).

A consistent method has been applied following approaches adopted for Mid-Atlantic species (although not the Gibson 1996 method)

Previous from 2008 GARM-III:

1) Assessment approaches needs [*sic*] 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. The Panel also had concerns about the unit stock, not only for this stock, but for all of the Winter Flounder stocks assessed. It recommended an analysis of Winter Flounder as a stock complex, rather than as individual stocks, be undertaken.

See Terceiro MS 2011a

New from 2011 SAW 52:

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. Explore options to conduct periodic maturity staging workshops involving State and NEFSC trawl survey staff.

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, and 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.

References

Able KW, Fahay MP. 2010. Ecology of Estuarine Fishes: Temperate Waters of the Western North Atlantic. John Hopkins University Press, Baltimore.

Ampela K. 2009. The diet and foraging ecology of gray seals (*Halichoerus grypus*) in United States waters. Ph. D. dissertation. The City University of New York. 176 p.

Anonymous. 2009. Independent Panel review of the NMFS Vessel Calibration analyses for FSV *Henry B. Bigelow* and R/V *Albatross IV*. August 11-14, 2009. Chair's Consensus report. 10 p.

Berry RJ, Saila SB, Horton DB. 1965. Growth studies of winter flounder *Pseudopleuronectes americanus* (Walbaum), in Rhode Island. Trans. Am. Fish. Soc. 94: 259-264.

Beverton RJH, Holt SJ. 1957. On the dynamics of exploited fish populations. Chapman and Hall, London. Facsimile reprint 1993.

Bigelow HB, Schroeder WC. 1953. Fishes of the Gulf of Maine. Fishery Bulletin of the Fish and Wildlife Service 74 Volume 53. 577 p.

Brodziak J, O'Brien L. 2005. Do environmental factors affect recruits per spawner anomalies of New England groundfish? ICES J. Mar. Sci. 62:1394-1407.

Brodziak JTK, Overholtz WJ, Rago P. 2001. Does spawning stock affect recruitment of New England groundfish? Can J Fish Aquat Sci. 58: 306-318.

Brooks EN, Miller TJ, Legault CM, O'Brien L, Clark KJ, Gavaris S, Van Eekhaute L. 2010. Determining length-based calibration factors for cod, haddock, and yellowtail flounder. TRAC Ref. Doc. 2010/08.

Brown R. 2009. Design and field data collection to compare the relative catchabilities of multispecies bottom trawl surveys conducted on the NOAA ship *Albatross* IV and the FSV *Henry B. Bigelow*. NEFSC Bottom Trawl Survey Calibration Peer Review Working Paper. Northeast Fisheries Science Center, Woods Hole, MA. 19 p.

Brownie C, Anderson DR, Burnham KP, Robson DS. 1985. Statistical inference from band recovery data. Fish and Wildlife Serv. Resource Publ. 156.

Burnham K P, Anderson DR. 1998. Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York.

Cohen J, Fletcher C. 2007. Improved Skill of Northern Hemisphere Winter Surface Temperature Predictions Based on Land–Atmosphere Fall Anomalies. Journal of Climate 20:4118-4132.

DeCelles GR, Cadrin SX. 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: 408-419.

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.

Dickie LM, McCracken FD. 1955. Isopleth diagrams to predict equilibrium yield of a small flounder fishery. J. Fish. Res. Bd. Canada 12: 187-209.

Fairchild EA, Sulikowski JA, Rennels N, Howell WH, Tsang PCW. 2010. Effects of moving acclimation cages before release of cultured fish: alternative release strategies for a juvenile winter flounder *Pseudopleuronectes americanus* stock enhancement effort. Aquacult. Res. 41: 602-606.

Gislason H, Daan N, Rice JC, Pope JG. 2010. Size, growth, temperature and the natural mortality of marine fish. Fish and Fisheries 11: 149-158

Greene C, Pershing A. 2003. The flip-side of the North Atlantic Oscillation and modal shifts in slope-water circulation patterns. Limnology and oceanography 48:319-322.

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, Able KW. 2007. Mechanistic links between climate and fisheries along the east coast of the United States: explaining population outbursts of Atlantic croaker (*Micropogonias undulatus*). Fisheries Oceanography 16:31-45.

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.

Hare JA, Kane J. In press. Zooplankton of the Gulf of Maine – a changing perspective. American Fisheries Society Science Symposium.

Hettler WF, Chester AJ. 1982. The relationship of winter temperature and spring landings of pink shrimp, *Penaeus duorarum*, in North Carolina. Fishery Bulletin 80:761-768.

Hewitt DA, Hoenig JM. 2005. Comparison of two approaches for estimating natural mortality based on longevity. Fish. Bull. 103: 433-437.

Hoenig JM. 1983. Empirical use of longevity data to estimate mortality rates. Fish. Bull. 82: 898-903.

Hoenig JM, Barrowman NJ, Hearn WS, Pollock KH. 1998. Multiyear Tagging Studies Incorporating Fishing Effort Data. Can. J. Fish. Aquat. Sci. 55:1466-1476.

Howe AB, Coates PG. 1975. Winter flounder movements, growth and mortality off Massachusetts. Trans. Am. Fish. Soc. 104: 13-29.

Howell P, Howe A, Gibson M, Ayvasian S. 1992. Fishery management plan for inshore stocks of winter flounder. Atlantic States Marine Fisheries Commission. Fish Mgmt Rep 21. May 1992.

Hurrell JW, Deser C. 2010. North Atlantic climate variability: the role of the North Atlantic Oscillation. Journal of Marine Systems 79:231-244.

Iles TC, Beverton RJH. 1998. Stock, recruitment and moderating processes in flatfish. Journal of Sea Research 39:41-55.

Joyce TM. 2002. One hundred plus years of wintertime climate variability in the eastern United States. Journal of Climate 15:1076-1086.

Joyce TM, Zhang R. 2010. On the Path of the Gulf Stream and the Atlantic Meridional Overturning Circulation. Journal of Climate 23:3146-3154.

Kalnay EC, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J. 1996. The NCEP/NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society **77**:437-471.

Keenlyside NS, Latif M, Jungclaus J, Kornblueh L, Roeckner E. 2008. Advancing decadal-scale climate prediction in the North Atlantic sector. Nature 453:84-88.

Keller AA, Klein-MacPhee G. 2000. Impact of elevated temperature on the growth, survival, and trophic dynamics of winter flounder larvae: a mesocosm study. Can. J. Fish. Aquat. Sci. 57:2382-2392.

Kerr RA. 2005. Atlantic climate pacemaker for millennia past, decades hence? Science 309:43-44.

Kendall WC. 1912. Notes on a new species of flatfish from off the coast of New England. Bulletin of the U.S. Bureau of Fisheries 30: 391-394.

Lazzari M. 1997. Monthly and annual means of sea surface temperature: Boothbay Harbor, Maine 1905–1996. Research Reference Document.

Levi D, Andreoli M, Bonanno A, Fiorentino F, Garofalo G, Mazzola S, Norrito G, Patti B, Pernice G, Ragonese S. 2003. Embedding sea surface temperature anomalies into the stock recruitment relationship of red mullet (Mullus barbatus L. 1758) in the Strait of Sicily. Scientia Marina 67:259-268.

Lux FE. 1973. Age and growth of the winter flounder, *Pseudopleuronectes americanus*, on Georges Bank. Fish. Bull. 71: 505-512.

Lux FE, Peterson AE Jr, Hutton RF. 1970. Geographical variation in fin ray number in winter flounder, *Pseudopleuronectes americanus* (Walbaum), off Massachusetts. Trans. Am. Fish. Soc. 1970(3): 483-488.

Mace PM, Doonan IJ. 1988. A generalized bioeconomic simulation model for fish population dynamics. NZ Fish Ass Res Doc. 88/4.

Manderson JP. 2008. The spatial scale of phase synchrony in winter flounder (Pseudopleuronectes americanus) production increased among southern New England nurseries in the 1990s. Can. J. Fish. Aquat. Sci. 65:340-351.

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. Submitted. A hierarchical model for relative catch efficiency from gear selectivity and calibration studies.

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. 1996. Report of the 21st Northeast Regional Stock Assessment Workshop (21st SAW): Stock Assessment Review Committee (SARC) consensus summary of assessments. NEFSC Ref Doc. 96-05d. 200 p.

NEFSC. 1999. Report of the 28th Northeast Regional Stock Assessment Workshop (28th SAW): Stock Assessment Review Committee (SARC) consensus summary of assessments. NEFSC Ref Doc. 99-08. 304 p.

NEFSC. 2002. Final report of the Working Group on re-evaluation of biological reference points for New England groundfish. NEFSC Ref Doc. 02-04. 123 p.

NEFSC. 2003. Report of the 36th Northeast Regional Stock Assessment Workshop (36th SAW): Stock Assessment Review Committee (SARC) consensus summary of assessments. NEFSC Ref Doc. 03-06. 453 p.

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. NEFSC Ref Doc. 05-13. 508 p.

NEFSC. 2008. Assessment of 19 Northeast groundfish stocks through 2007. Report of the 3rd Groundfish Assessment Review Meeting (GARM III), Northeast Fisheries Science Center, Woods Hole, Massachusetts. August 4-8, 2008. NEFSC Ref Doc. 08-15. 884 p.

NEFSC. 2011. 51st Northeast regional stock assessment workshop (51st SAW) assessment report. NEFSC Ref. Doc. 11-02. 856 p.

NOAA Fisheries Toolbox (NFT). 2010. Virtual Population Analysis Model (VPA/ADAPT) version 3.0.3. Internet website http://nft.nefsc.noaa.gov/VPA.

NOAA Fisheries Toolbox (NFT). 2011. Age Structured Assessment Program (ASAP) version 2.0.21. Internet website http://nft.nefsc.noaa.gov/ASAP.

Neill WH, Miller JM, Van Der Veer HW, Winemiller KO. 1994. Ecophysiology of marine fish recruitment: a conceptual framework for understanding interannual variability. Netherlands Journal of Sea Research 32:135-152.

Nixon SW, Granger S, Buckley BA, Lamont M, Rowell B. 2004. A one hundred and seventeen year coastal water temperature record from Woods Hole, Massachusetts. Estuaries 27:397-404.

Nye JA, Joyce TM, Kwon YO, Link JS. In review. Silver hake tracks changes in Northwest Atlantic circulation. Nature Communications.

Nye JA, Link JS, Hare JS, Overholtz WJ. 2009. Changing spatial distribution of fish stocks in relation to climate and population size on the Northeast United States continental shelf. Mar. Ecol. Prog. 393: 111-129.

O'Brien L, Burnett J, Mayo R. 1993. Maturation of nineteen species of finfish off the northeast coast of the United States, 1985-1990. NOAA Technical Report NMFS 113. 66 pp.

Ottersen GB, Planque B, Belgrano A, Post E, Reid P, Stenseth N. 2001. Ecological effects of the North Atlantic Oscillation. Oecologia 128:1-14.

Roelofs EW, Bumpus DF. 1953. The hydrography of Pamlico Sound. Bulletin of Marine Science of the Gulf and Caribbean 3:181-205.

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.

Pereira JJ, Goldberg R, Ziskowski JJ, Berrien PL, Morse WW, Johnson DL. 1999. Winter flounder, *Pseudopleuronectes americanus*, life history and habitat characteristics. NOAA Tech. Memo. NMFS-NE-138. 39 p.

Perlmutter A. 1947. The blackback flounder and its fishery in New England and New York. Bulletin of the Bingham Oceanographic Collection, 11: 1- 92.

Pierce DE, Howe AB. 1977. A further study on winter flounder group identification off Massachusetts. Trans. Am. Fish. Soc. 106: 131-139.

Poole JC. 1969. A study of winter flounder mortality rates in Great South Bay, New York. Trans. Am. Fish. Soc. 98: 611-616.

Quinn TJ II, Deriso RB. 1999. Quantitative fish dynamics. Oxford Univ. Press, New York, NY.

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

Rademeyer RA, Butterworth DS. MS 2011b. Initial applications of statistical catch-at-age assessment methodology to the Southern New England/Mid-Atlantic winter flounder resource. 29 p.

Rago PJ, Gabriel W, Lambert M. 1994. Assessment of yellowtail flounder *Pleuronectes ferrugineus*, 1993. NEFSC Ref. Doc. 94-02.

Roelofs EW, Bumpus DF. 1953. The hydrography of Pamlico Sound. Bulletin of Marine Science of the Gulf and Caribbean 3:181-205.

Rothschild BJ. 1986. Dynamics of marine fish populations. Harvard University Press, Cambridge, Massachusetts.

Saila SB, Horton DB, Berry RJ. 1965. Estimates of the theoretical biomass of juvenile winter flounder, *Pseudopleuronectes americanus* (Walbaum) required for a fishery in Rhode Island. J. Fish. Res. Bd. Canada 22: 945-954.

Shepherd G, Cadrin S, Correia S, Gabriel W, Gibson M, Howe A, Howell P, Grout D, Lazar N, Lambert M, Ling W. 1996. Assessment of winter flounder in Southern New England and the Mid-Atlantic. NEFSC Ref. Doc. 96-05b.

Smith BE, Link JS. 2010. The trophic dynamics of 50 finfish and two squid species on the Northeast U.S. continental shelf. NOAA Tech. Memo. NMFS-NE-216. 640 p.

Smith DM, Cusack S, Colman AW, Folland CK, Harris GR, Murphy JM. 2007. Improved surface temperature prediction for the coming decade from a global climate model. Science 317:796-799.

Stoner A, Manderson J, Pessutti J. 2001. Spatially explicit analysis of estuarine habitat for juvenile winter flounder combining generalized additive models and geographic information systems. Marine Ecology Progress Series 213:253–271.

Taylor AH, Stephens JA. 1998. The North Atlantic oscillation and the latitude of the Gulf Stream. Tellus A 50:134-142.

Taylor DL, Collie JS. 2003. Effect of temperature on the functional response and foraging behavior of the sand shrimp Crangon septemspinosa preying on juvenile winter flounder *Pseudopleuronectes americanus.* Marine Ecology Progress Series 263:217-234.

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

Terceiro M. MS 2011b. Response to 2008 GARM-III research recommendations winter flounder TOR 8c. Stock Assessment Workshop Southern Demersal Working Group 52 Working Paper 4.

Thompson WF, Bell FH. 1934. Biological statistics of the Pacific halibut fishery. 2. Effect of changes in intensity upon total yield and yield per recruit of gear. Rep Int Fish (Pacific halibut) Comm. 8: 49 p.

Vetter EF. 1988. Estimation of natural mortality in fish stocks: a review. Fish. Bull. 86: 25-43.

Visbeck M, Chassignet EP, Curry RG, Delworth TL, Dickson RR, Krahmann G. 2003. The ocean's response to North Atlantic Oscillation variability. Page 279 *in* JW Hurrel, Y Kushnir, G Ottersen, and M Visbeck, editors. The North Atlantic Oscillation: Climatic significance and environmental impact. American Geophysical Union, Washington, D.C.

Waring GT, Josephson E, Fairfield CP, Maze-Foley K. 2009. US Atlantic and Gulf of Mexico marine mammal stock assessments 2008. NOAA Tech. Memo. NMFS-NE-210. 429 p.

Witherell DB, Burnett J. 1993. Growth and maturation of winter flounder, *Pseudopleuronectes americanus,* in Massachusetts. Fish. Bull. 91: 816-820.

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.

Wood A. MS 2011. Re-analysis of Howe and Coates (1975): winter flounder natural mortality derived from data in Howe and Coates (1975) using instantaneous rates tagging models. Stock Assessment Workshop Southern Demersal Working Group 52 Working Paper 15.

Wuenschel M, Able K, Byrne D. 2009. Seasonal patterns of winter flounder Pseudopleuronectes americanus abundance and reproductive condition on the New York Bight continental shelf. Journal of Fish Biology 74:1508-1524.