SCUP BENCHMARK STOCK ASSESSMENT FOR 2015

A1. Terms of Reference

1. Estimate catch from all sources including landings and discards. Include recreational discards, as appropriate. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the uncertainty in these sources of data.

2. Present the survey data being used in the assessment (e.g., indices of relative or absolute abundance, recruitment, state surveys, age-length data, etc.). Characterize the uncertainty and any bias in these sources of data.

3. Describe the thermal habitat and its influence on the distribution and abundance of scup, and attempt to integrate the results into the stock assessment.

4. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results and previous projections.

5. 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}, $B_{THRESHOLD}$, F_{MSY} and 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.

6. Evaluate stock status with respect to the existing model (from previous peer reviewed accepted assessment) and with respect to a new model developed for this peer review.
   a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.
   b. Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-5).

7. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) (see Appendix to SAW TORs for definitions).
   a. Provide numerical annual projections (3 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).
   b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.
   c. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

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

A2. Executive Summary

TOR 1. Estimate catch from all sources including landings and discards. Include recreational discards, as appropriate. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the uncertainty in these sources of data.

The otter trawl is the principal commercial fishing gear. Commercial landings of scup peaked in 1960 at 22,200 mt, then decreased during the 1960s and ranged between 5,000 and 10,000 mt until the late 1980s. Commercial fishery quotas were implemented in 1997, and landings then ranged between 1,200 mt and 8,100 mt and averaged 4,000 mt during 1997-2014. Reported 2014 commercial fishery landings were 7,228 mt = 15.935 million lbs, about 77% of the commercial quota, and 68% of the total catch.

The NEFSC Northeast Fishery Observer Program (NEFOP) has collected information on landings and discards in the commercial fishery since 1989. In previous assessments, a method using the Geometric Mean Discards-to-Landings Ratio (GMDL) was been used to estimate scup discards. The Observer data have provided evidence that the Gear Restricted Areas (GRAs) implemented in 2000-2001 have been effective in reducing the scup discard percentage. The current assessment absolute estimates of scup discards using the GMDL approach, however, are produced on a temporal and spatial scale that is too coarse to directly evaluate the effectiveness of specific discard reduction measures (e.g., on a specific area or season basis). This prompted a re-examination of the methods used to estimate commercial fishery scup discards using the Standardized Bycatch Reporting Method (SBRM), which was implemented in February 2008 to address the requirements of the Magnuson-Stevens Fishery Conservation and Management Act. The SBRM for the estimation of discards has now been adopted for most NEFSC stock assessments that have been subject to a benchmark review since 2009. In this assessment, newly developed SBRM estimates of scup discards are compared the current GMDL estimates. The new SBRM discard estimate time series is used in the 2015 SAW 60 scup assessment. Estimated 2014 commercial fishery live discards were 1,140 mt = 2.513 million lbs (CV = 14%), about 11% of the total catch. The commercial discard mortality rate is assumed to be 100%.

Scup is the object of a major recreational fishery, with the greatest proportion of catches taken in the states of Massachusetts, Rhode Island, Connecticut and New York. Estimates of the recreational catch in numbers were obtained from the NMFS Marine Recreational Fishery Statistics Survey (MRFSS) for 1981-2011, and from the NMFS Marine Recreational Information Program (MRIP) for 2004-2014. The estimated recreational landings during 1981-2014 averaged 2,300 mt per year. Estimated 2014 recreational fishery landings were 2,025 mt = 4.464 million lbs (CV = 13%), about 64% of the recreational harvest limit, and 19% of the total catch.

The estimated recreational live discard during 1984-2011 ranged from 43 mt in 1999 to a high of 2,120 mt in 2010, averaging 600 mt per year. A discard mortality rate in the recreational fishery of 15% has been used in this and previous assessments, resulting in a time series average discard mortality of about 126 mt per year. Estimated 2014 recreational fishery dead discards were 227 mt = 0.500 million lbs (CV = 14%), about 2% of the total catch.

In response to fishing industry (both commercial and recreational) comments that the utility of fishery dependent catch per unit effort (CPUE) should be evaluated as indices of abundance for scup, a subset of the 2015 SAW 60 Scup Working Group (SWG) with an interest in fishery dependent CPUE compiled data and conducted analyses from a number of sources. The SWG noted generally that 1) the utility of the fishery dependent data as the basis for indices of abundance is limited in that some of them include only landings and not the total catch including discards, and so the resulting LPUE could be biased low relative to the true abundance of fish, 2) the use of only positive trips that catch scup may bias the LPUE or CPUE as well, and may be influenced by management regulations, and 3) the ratio of catch to effort has generally changed over time, and it is unclear how this change reflects real changes over time in fishing behavior due to fish abundance, management regulations, or changes in data reporting systems. The SWG concluded that further analysis beyond the scope of the assessment is needed to standardize the complexity of factors influencing fishery catch rates.

TOR 2. Present the survey data being used in the assessment (e.g., indices of relative or absolute abundance, recruitment, state surveys, age-length data, etc.). Characterize the uncertainty and any bias in these sources of data.

Indices of stock abundance from the NEFSC winter, spring, and fall, Massachusetts DMF spring and fall, Rhode Island DFW spring and fall, University of Rhode Island Graduate School of Oceanography (URIGSO), Connecticut DEEP spring and fall, New York DEC, New Jersey DFW, and Virginia Institute of Marine Science (VIMS) Chesapeake Bay (ChesMMAP) and VIMS juvenile fish trawl surveys were used in the 2008 model calibration and in subsequent assessment updates through 2012. The NEAMAP spring and fall bottom trawl, RIDFW spring and fall survey age compositions, and RI Industry Cooperative trap survey data have been added to the 2015 SAW 60 assessment documentation. After a process of building the 2015 population model, the NEFSC spring, MADMF spring, RIDFW spring and fall, and VIMS ChesMMAP surveys were omitted from the model calibration.

TOR 3. Describe the thermal habitat and its influence on the distribution and abundance of scup and attempt to integrate the results into the stock assessment.

Some of the NEFSC winter, spring and fall trawl survey environmental data were summarized for the strata sets used for scup to investigate the correspondence between the environmental factors and the distribution of scup. The environmental factors were surface air temperature in degrees Celsius, surface and bottom water temperature in degrees Celsius, and bottom water salinity in parts per thousand (PPT). Examination of patterns in the survey catch, for spring and fall and day and night, confirms the irregular distributions of catch by temperature, salinity and depth and portend the difficulties of modeling the scup survey catch data. No well defined relationships are evident; i.e., small catches are as likely to be taken at shallow depths as large depths and at both warm and cold temperatures and large catches can occur over a relatively large range of depth and temperature (e.g, over a range of 70 meters or 10 degrees). Therefore, generalized linear model (GENMOD) and generalized additive model (GAM) based indices of abundance for the scup NEFSC seasonal survey data proved to be not useful, due to highly variable results owing from the inability of the models to adequately fit the variable and complex temporal and spatial properties of scup survey catches.

The NEFSC survey indices sometimes appear to mainly reflect the availability of scup to the survey, rather than true abundance, making it difficult to interpret large inter-annual changes in the indices. In particular, the spring 2002 and 2014 spring indices were unexpectedly much higher than adjacent indices, across all ages. In 2002, this 'availability event' appears to have been a response to higher than normal spring water temperatures, as large scup survey catches and bottom water with temperatures higher than 10°C were distributed further inshore on the shelf than usual. Near 'normal' bottom conditions were present in 2014, but catches of large scup occurred near mid-shelf in large-area strata, and the 2014 indices were among the largest of the spring time series. These two sequences of potential 'availability events' make clear the difficulty that is encountered when interpreting survey indices for scup – do high survey indices indicate high availability, high abundance, or (more likely) some combination of both?

Estimates of proportions of thermal habitat surveyed in the NEFSC and NEAMAP surveys were developed that could be used to account for errors in survey observations related to temperature dependent changes in geographic distribution and seasonal migration. Time varying estimates of the proportion of thermal habitat suitability for scup surveyed on the Northeast US shelf were calculated for the NEFSC and NEAMAP bottom trawl surveys from 1975-2012. An average of 63 % of the thermal habitat suitability available to scup within the model domain (Cape Hatteras to Nova Scotia) was sampled from 1973-2012 by the fall NEFSC bottom trawl survey, while 50% was sampled in the spring. In the 2008-2012 NEAMAP surveys 14% of available thermal habitat suitability on the Northeast US continental shelf was sampled during the fall, while 11% was sampled in the spring. Yearly estimates of the proportion of thermal habitat suitability surveyed did not exhibit systematic trends.

Logit-transformed annual values of the 'proportion of suitable scup thermal habitat sampled' – i.e., availability - were used in a version of the final assessment model run to provide annually varying estimates of relative survey catchability (q), where q is the product of availability and survey gear efficiency (assumed = 1). The NEFSC survey qs were estimated to be variable without long term trend; NEAMAP survey qs were variable over the short 7-8 year time series. Given the similarity of results and still preliminary nature of the 'varying q' model version (the version of the model and associated documentation have not yet been released to the public), the 'varying q' version of the final model was not used for status evaluation.

TOR 4. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results and previous projections.

The instantaneous natural mortality rate (M) for scup has been assumed to be 0.20 in all previous stock assessments. Given the historical maximum size and age of 41 cm and 15 years, recent observations of large fish (45 cm) up to age 12, the range of M (0.1 – 0.6) estimated by the empirical methods based on maximum age, and the likelihood profile of a preliminary assessment model run that indicated a best fit at 0.10 and of the final model at 0.15, the SARC decided there was no compelling reason to change from the previous assumption for M, and adopted a value of M = 0.20 for all ages and years in the 2015 SAW 60 assessment models.

The most recent benchmark peer review of the scup assessment was conducted by the 2008 Northeast Data Poor Stocks Working Group (DPSWG). The assessment model for scup changed in 2008 from a simple index-based model to a complex statistical catch at age model.
The fishery catch is modeled as four fleets: commercial landings, recreational landings, commercial discards and recreational discards. The time series of commercial discard and recreational catch estimates have been revised since the 2008 assessment.

Indices of stock abundance from NEFSC winter, spring, and fall, Massachusetts DMF spring and fall, Rhode Island DFW spring and fall, University of Rhode Island Graduate School of Oceanography (URIGSO), Connecticut DEEP spring and fall, New York DEC, New Jersey DFW, and Virginia Institute of Marine Science (VIMS) Chesapeake Bay (ChesMMAP) and VIMS juvenile fish trawl surveys were used in the 2008 model calibration and in subsequent assessment updates through 2012. The NEAMAP spring and fall bottom trawl, RIDFW spring and fall survey age compositions, and RI Industry Cooperative trap survey data have been added to the 2015 SAW 60 assessment documentation.

The ASAP model structural configuration and settings were significantly revised for the 2015 SAW 60 assessment. After a process of building the 2015 population model, the NEFSC spring, MADMF spring, RIDFW spring and fall, and VIMS ChesMMAP surveys were omitted from the model calibration. The general results (e.g., highest estimated stock size and low F in the last decade) are robust to all proposed alternative model configurations, including the length of the time series and a range of priors and likelihood component weightings. There is no consistent retrospective pattern in F, SSB, or recruitment evident in the scup assessment model. However, there are some indications of poor model fit from lack of correspondence among surveys (higher than expected variance when accounting for potential process error, some residual patterns), and there is uncertainty in the absolute magnitude of recent stock size estimates (although the terminal year estimates are calculated to be relatively precise with CVs less than or equal to 15%). Alternative survey catchabilities (e.g., relative, absolute using wing or door spread), starting years, commercial and recreational selectivity patterns (see note below), and time-varying survey catchability configurations can produce about a +/- 40% range of terminal year SSB. The SARC concluded, however, that the accepted model run provided the best balance between good retrospective diagnostics, acceptable fishery and survey fit diagnostics, and stability over most configurations, and recommended use of the ASAP model final run for status evaluation.

During the evaluation of the accepted model, sensitivities were examined which highlighted some additional risk. The main one of relevance to management is the choice of selectivity pattern. The base model has a strong domed selectivity pattern which could result in an increasing cryptic biomass given current stock trajectory. Conclusions regarding current stock status are robust to alternative selectivity patterns but decreased recruitment or increased F in the future could lead to divergence between domed and flattop selectivity model results.

Spawning stock biomass (SSB) decreased from about 68,000 mt in 1963 to about 5,000 mt in 1969, then increased to about 27,000 mt during the late 1970s. SSB declined through the 1980s and early 1990s to less than about 4,000 mt in the mid-1990s. With greatly improved recruitment

and low fishing mortality rates since 1998, SSB increased to about greater than 100,000 mt = 220 million lbs since 2003. SSB was estimated to be 182,915 mt = 403 million lbs in 2014. There is a 90% probability that SSB in 2014 was between 153,000 and 222,000 mt (337 and 489 million lbs). Fishing mortality estimated at the 'apical' age 3 (model age 4) where full selection occurs varied between F = 0.5 and F = 2.0 during the 1960s and 1970s. Fishing mortality next peaked at about F = 1.5 in the 1990s. Fishing mortality decreased after 1994, falling to less than F = 0.15 since 2000, with F in 2014 = 0.127. There is a 90% probability that F in 2014 was between 0.093 and 0.149. Recruitment at age 0 averaged 98 million fish during 1963-1983, the period in which recruitment estimates are tightly constrained (CV = 0.1 on recruitment deviations and stock-recruitment scaler with fixed h =1) to ensure near constant recruitment before 1984, when fishery catch at age are not available. Since 1984, recruitment estimates from the model are influenced mainly by the fishery and survey catches at age, and averaged 109 million fish during 1984-2014. The 1999, 2006, and 2007 year classes are estimated to be the largest of the time series, at 222, 222, and 218 million age 0 fish. After below average recruitment in 2012 and 2013, the 2014 year class is estimated to be above average at 112 million age 0 fish.

Despite changes in model assumptions, configurations, and estimation procedures, the 'historical' retrospective analysis indicates that the general trends in stock biomass, recruitment, and fishing mortality have been consistent for the last decade. Estimates of SSB are in line with previous 2009-2012 projections, F is lower than from the 2011-2012 projections, and catch is lower than from the 2011-2012 projections, with the fishery in 2014 taking about 75% of the ACL.

TOR 5. 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}, $B_{THRESHOLD}$, F_{MSY} and 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.

The 2008 Data Poor Stocks Working Group (DPSWG) Peer Review Panel accepted the ASAP model results as the basis for biological reference points and status determination for scup. Reference points were calculated using the non-parametric yield and SSB per recruit/long-term projection approach adopted for summer flounder and the New England groundfish stocks. For the estimation of MSY (Maximum Sustainable Yield) and SSBMSY (Spawning Stock Biomass at Maximum Sustainable Yield), the cumulative distribution function of the 1984-2007 recruitments (corresponding to the period of available fishery catches at age) was re-sampled to provide future recruitment estimates (mean = 117 million age 0 fish) for biomass reference point estimation. The existing reference points for scup are the 2008 DPSWG Peer Review Panel recommended F40% as the proxy for FMSY, and the corresponding SSBF40% as the proxy for SSBMSY. The F40% proxy for FMSY = 0.177, the proxy estimate for SSBMSY = SSB40% = 92,044 mt = 202.922 million lbs, and the proxy estimate for MSY = MSY40% = 16,161 mt = 35.629 million lbs (13,134 mt = 28.956 million lbs of landings and 3,027 mt = 6.673 million lbs of discards).

The SARC accepted the ASAP model S60_BASE_18 results as the basis for new biological reference points and status determination for scup. Reference points were again calculated using the non-parametric yield and SSB per recruit long-term projection approach. The cumulative distribution function of the 1984-2014 recruitments (corresponding to the period of available fishery catches at age) was re-sampled to provide future recruitment estimates (mean = 109 million age 0 fish) for biomass reference point estimation. The SARC recommended F40% as the proxy for FMSY, and the corresponding SSBF40% as the proxy for the SSBMSY biomass target. The F40% proxy for FMSY = 0.220; the proxy estimate for SSBMSY = SSB40% = 87,302 mt = 192.468 million lbs; the proxy estimate for the ½ SSBMSY biomass threshold = ½ SSB40% = 43,651 mt = 96.234 million lbs; and the proxy estimate for MSY = MSY40% = 11,752 mt = 25.909 million lbs (9,445 mt = 20.823 million lbs of landings and 2,307 mt = 5.086 million lbs of discards).

TOR 6. Evaluate stock status with respect to the existing model (from previous peer reviewed accepted assessment) and with respect to a new model developed for this peer review.
a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.
b. Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-5).

a) The existing model updated with new data indicated that the scup stock was not overfished and overfishing was not occurring in 2014 relative to the existing (old) biological reference points established in the 2008 Data Poor Stocks Working Group assessment (NEFSC 2009). The fishing mortality rate (F) was estimated to be 0.049 in 2014, below the fishing mortality threshold reference point = FMSY = F40% = 0.177. Spawning Stock Biomass (SSB) was estimated to be 219,066 metric tons (mt) = 483 million lbs in 2014, above the biomass target reference point = SSBMSY = SSB40% = 92,044 mt = 203 million lbs.

b) The scup stock was not overfished and overfishing was not occurring in 2014 relative to the new biological reference points recommended by the SARC. The fishing mortality rate (F) was estimated to be 0.127 in 2014, below the fishing mortality threshold reference point = FMSY = F40% = 0.220. Spawning Stock Biomass (SSB) was estimated to be 182,915 metric tons (mt) = 403 million lbs in 2014, above the biomass target reference point = SSBMSY = SSB40% = 87,302 mt = 192 million lbs.

TOR 7. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) (see Appendix to SAW TORs for definitions).
a. Provide numerical annual projections (3 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).
b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.

c. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

a) Stochastic projections were made to provide forecasts of stock size and overfishing level (OFL) catches in 2016-2018 consistent with the 2015 SAW 60 assessment biological reference points. The cumulative distribution function of the 1984-2014 recruitments (corresponding to the period of available fishery catches at age) was re-sampled to provide future recruitment estimates (mean = 109 million age 0 fish) for projections. The SWG conducted two sets of projections. Option A is proposed as the most realistic and assumes that given recent patterns in the fishery, it is likely that 75% of the 2015 ACL will be caught. Projection option B assumes that 100% of the 2015 ACL will be caught.

A) If the catch of scup in 2015 equals 75% of the specified ACL = 0.75 * 15,320 = 11,490 mt = 25.331 million lbs, the 2015 median (50% probability) landings are projected to be 10,058 mt = 22.174 million lbs and discards are projected to be 1,432 mt = 3.157 million lbs. The projected OFLs in 2016-2018 are 16,238, 14,556, and 13,464 mt (35.799, 32.090, and 29.683 million lbs).

B) If the catch of scup in 2015 equals 100% of the specified ACL = 15,320 mt = 33.775 million lbs, the 2015 median (50% probability) landings are projected to be 13,412 mt = 29.568 million lbs and discards are projected to be 1,908 mt = 4.206 million lbs. The projected OFLs in 2016-2018 are 15,745, 14,199, and 13,230 mt (34.712, 31.303, and 29.167 million lbs).

The biological inputs to the scup stock assessment are based on well-founded assumptions (e.g., for M, for discard mortality in the fisheries) and precisely estimated biological parameters (e.g., growth, age, maturity, and mean weights). Further, the research survey index CVs used in model calibration have been increased by 50-100% (depending on assessment model fit diagnostics) to account for process error. A broad set of model configurations produced a range about +/- 40% in the average estimate of terminal year SSB of about 180,000 mt (396 million lbs). The internal retrospective average error (for the terminal 7-years) of the assessment is low, at less than 10% for both SSB and F. The analytically derived CV for the 2014 SSB is 11%, the CV for the 2014 F is 15%, and the CV for the 2014 age 1 and older stock size total number is 15%. Given these properties of the 2015 scup stock assessment, it was concluded that an approximate doubling of the analytically derived 2016-2018 OFL CVs to 30% is a reasonable and sufficient adjustment to account for additional uncertainty in the assessment such as the magnitude of domed fishery selection, the magnitude of commercial fishery discards and recreational catch during the early part of the assessment model time series, and potential error in the aging process.

b) Both projection options have a realistic probability of being achieved and indicate there is zero percent chance that SSB will fall below the biomass threshold in 2016-2018 fishing at the OFL.

c) The scup stock has a low probability of becoming overfished in the short term (2016-2018) given recent trends in productivity and the responsiveness of the management regime.

TOR 8. Review, evaluate and report on the status of the SARC, SSC, and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

Nine of the 12 previously identified research recommendations were either addressed in full or significant progress was made. No progress has been made on a) quantifying contemporary discard mortality rates, b) quantifying the degree of bias in landings reporting and discard estimation including non-compliance, or c) development of a management strategy evaluation of alternative approaches to setting quotas. Six new research recommendations were developed.

A3. Working Group Process

The Stock Assessment Workshop (SAW) Scup Working Group (SWG) met during April 20-22, 2015 at the Northeast Fisheries Science Center (NEFSC) to develop the benchmark stock assessment of scup through 2014. The following people provided data, participated in the preparation, and/or where present for discussion of the assessment in the 2015 SWG:

Gary Shepherd	NEFSC Coastal/Pelagic Resources Task Leader; SWG Chair
Mark Terceiro	NEFSC Demersal Resources Task Leader, Scup Assessment Lead
Julia Beaty	Mid-Atlantic Fishery Management Council (MAFMC)
Mike Bednarski	Massachusetts Division of Marine Fisheries (MADMF)
Chris Bonzek	Virginia Institute of Marine Science (VIMS)
Steve Cadrin	University of Massachusetts-Dartmouth, School of Marine Science and Technology(SMAST), Science Center for Marine Fisheries (SCeMFiS)
Kirsten Curti	NEFSC Population Dynamics Branch
Peter Clarke	New Jersey Division of Fish and Wildlife (NJDFW)
Kiley Dancy	Mid-Atlantic Fishery Management Council (MAFMC)
Meaghan Lapp	Seafreeze Ltd.
Robert Leaf	University of Southern Mississippi (USM), Science Center for Marine Fisheries (SCeMFiS)
Chris Legault	NEFSC, Assessment Methods Task Leader
Jean-Jacques McGuire	Science Center for Marine Fisheries (SCeMFiS)
John Manderson	NEFSC Cooperative Research Sandy Hook Laboratory
John Maniscalco	New York Dept. of Environ. Conservation (NYDEC); ASMFC Technical Committee Chair
Jason McNamee	Rhode Island Division of Fish and Wildlife (RIDFW),
Alicia Miller	NEFSC Population Dynamics Branch
Tim Miller	NEFSC Population Dynamics Branch
Loretta O'Brien	NEFSC Population Dynamics Branch
Mike Palmer	NEFSC Population Dynamics Branch
Paul Rago	NEFSC Population Dynamics Branch
Kirby Rootes-Murdy	Atlantic States Marine Fisheries Commission (ASMFC)
Gregory Wojcik	Connecticut Department of Energy and Environmental Protection (CTDEEP)

A4. Introduction

A4.1 Biology

Scup (*Stenotomus chrysops*) is a schooling continental shelf species of the Northwest Atlantic that is distributed primarily between Cape Cod and Cape Hatteras (Morse 1978). Scup undertake extensive migrations between coastal waters in summer and offshore waters in winter. Scup migrate north and inshore to spawn in spring, with larger fish (age 2 and older) tending to arrive in spring first, followed by smaller fish (Neville and Talbot 1964; Sisson 1974). Larger scup are found during the summer near the mouth of large bays and in the ocean within 20 fathoms (120 feet = 37 meters), and often inhabit rough bottom areas. Smaller scup are more likely to be found in shallow, smooth bottom areas of bays during summer (Morse 1978). Scup migrate south and offshore in the fall as the water temperature decreases, arriving in offshore wintering areas by December (Hamer 1970; Morse 1978).

Historical tagging studies in the 1930s and 1950s (e.g., Neville and Talbot 1964; Cogswell 1960, 1961; Hamer 1970, 1979) have indicated the possibility of two stocks of scup, one in Southern New England waters and another extending south from New Jersey waters. However, the lack of definitive locations for tag return data coupled with distributional data from the NEFSC bottom trawl surveys support the concept of a single unit stock extending from Cape Hatteras north to New England (Mayo 1982). The NEFSC conducted a scup tagging program in cooperation with commercial and recreational fishermen in MA, RI, CT, and NY during 2005, tagging over 5,600 fish. The recapture rate was low at only 70 fish (1%) through 2008, with recoveries ranging from inshore waters off Southern New England to the edge of the shelf around Hudson Canyon.

Love and Chase (2009) compared morphology among scup populations by means of a geometric, landmark-based analysis of morphological and meristic traits for 180 individuals sampled in 2005 that were sexed and staged to maturity. They found morphological differences between a North Atlantic Bight (north of Cape Hatteras, NC) population and two South Atlantic Bight (south of Cape Hatteras) populations, at extremes of the scup's range in the northwestern Atlantic Ocean.

A4.2 Age and Growth

Historical studies of scup age and growth with reliable data include those of Finkelstein (1969a, b), Hamer (1970, 1979), Campbell et al. (1982), Dery and Rearden (1979), and Pentilla et al. (1989). These studies indicated that scup are relatively slow growing fish with maximum lengths of 37-41 cm and maximum ages of 13-15 years. Finkelstein (1969a, b) found both males and females to age 15, and noted that scup do not exhibit sexual dimorphism.

Age and growth information is available for full calendar years from NEFSC commercial port sampling from 1984-2014 and from NEFSC seasonal bottom trawl surveys from 1977-2014. The largest and oldest fish sampled by the NEFSC were a 46 cm age 10 fish sampled in 1973 and a 45 cm age 12 fish sampled in 2014; and 38-41 cm age 14 fish sampled in 1973, 1976, 1978, and 2014. For the NEFSC bottom trawl survey ages during 2008-2014, overall scup ageing precision, based on sample-size weighted intra- and inter-reader ageing agreement, averaged 90% with an overall Coefficient of Variation (CV) of 3%. For the NEFSC commercial

port sample ages during 2008-2014, overall scup ageing precision averaged 83% with an overall Coefficient of Variation (CV) of 2%.

Finkelstein (1969a) used data from 1,289 fish sampled from New York Bight in the 1960s to estimate the von Bertalanffy growth parameters for scup, finding Linf of about 34 cm for males and 37 cm for females, and k values of 0.27 and 0.22. Hamer (1979) used data from 1,429 fish sampled off New Jersey in the late 1950s and found a maximum age of 13 and estimated Linf for sexes combined to be about 34 cm and k to be 0.20.

The NEFSC trawl survey data for 1977-2014 were used to estimate growth parameters for males, females, and sexes combined. The full time series data provide parameters for males (n = 6,440) of Linf = 49.6 cm, k = 0.12, with maximum length of 38 cm and age of 10; parameters for females (n = 7,826) of Linf = 51.7 cm, k = 0.11, with maximum length of 41 cm and age of 14; and parameters for sexes combined (n = 20,197, including small fish of undetermined sex) of Linf = 46.6, k = 0.15, with maximum size of 41 cm and age of 14 (see table below). The growth curves are generally similar for all studies and sexes through about 30 cm and age 6, where they begin to diverge, due to the presence of larger fish of both sexes at ages 7 and older in the NEFSC survey data, compared to the same age fish in the Finkelstein (1969a) and Hamer (1970) data sets. In the most recent stock assessment update (Terceiro 2012), ages are grouped together for ages 7 and older (age 7+ 'plus group').

Study	N fish	Max age (M, F)	Linf (M, F, B)	k (M, F, B)
Finkelstein (1969a)	1,289	15,15	34.3,37.4	0.27,0.22
Hamer (1970)	1,429	13	34.1	0.29
NEFSC SVs	20,197	10,14	49.6, 51.7, 46.6	0.12, 0.11, 0.15

A4.3 Length-Weight Relationship

Morse (1978) used NEFSC trawl survey data from 2,234 New York Bight fish sampled during 1974-1975 to estimate the length weight parameters that are used for NEFSC commercial fishery length to weight conversions. Morse (1978) reported that an analysis of covariance showed no significant difference between males and females. Wigley *et al.* (2003) updated the length-weight parameters used in audits of the NEFSC trawl survey data, using individual length and weight information from 3,309 fish for 1992-1999. In the current work, individual length and weight information from 8,557 fish (3,572 males, 4,985 females) sampled during 1992-2013 were used to estimate length-weight parameters for comparison with the earlier studies to judge whether changing from the historical Morse (1978) parameters would be justified.

A comparison among these alternative compilations indicates very little difference in the estimated length-weight relationships from Morse (1978), Wigley *et al.* (2003), and the current examination for the NEFSC trawl survey data. The curves are virtually identical through a fork length of 30 cm at age 6, a threshold below which over 95% of the fishery catch has occurred. As noted earlier, larger fish of age 7 and older fish compose the assessment 'plus group.' Above 30 cm, the curves begin to diverge, with the Morse (1978) relationship providing mean weights at 35 cm and larger sizes that are about 10% higher than the current NEFSC survey combined relationship. Based on the consistency of these L-W relationships through 95% of the length range of the fishery catch, the Morse (1978) length-weight parameters were retained for this assessment.

A4.4 Condition Factor

Fulton's condition factor, K, is a measure of the relationship between fish length and weight that attempts to quantify the 'condition' of an individual or group of fish. Nash *et al.* (2006) note that it was Heincke (1908) who first used K as a measure of 'condition,' building on the 'cubic law' of growth in weight first introduced by Fulton (1904; K = x*weight / length**3, where x is a constant to scale K near 1). Nash *et al.* (2006) further point out that it was Ricker (1957) who first attributed the factor K to Fulton and coined the name 'Fulton's condition factor.' Froese (2006) reviewed the derivations of fish length-weight relationships and condition factors, and recommended use of a modern version of Fulton's K incorporating estimated length-weight relationship parameters as a better expression of 'relative condition factor.' The NEFSC spring and fall trawl survey sample data were examined for trends in relative condition factor by season and sex. Individual fish weight collection for scup began on NEFSC surveys in fall 1992. There are no long-term trends in condition factor by season or sex.

A4.5 Sex Ratio

The NEFSC winter, spring and fall trawl survey raw sample data were examined for trends in sex ratio by season and age, expressed as the proportion of females at age. The spring and fall series have sufficient data for the compilation beginning in 1977; the winter survey was conducted from 1992-2007. In all the series there are some years with no fish at ages older than 2.

In the winter survey, the proportion of females showed no trend for ages 1 and 2 and the proportion female generally varied from 0.4 to 0.8 (40 to 80% females), and the mean proportion was about 0.6. For age 3, the proportion increased from about 0.4 in the early 1990s to 1.0 by 1992, with a mean of about 0.6. For ages 4 to 6, the proportions are highly variable with no valid (i.e., ones that one would have confidence in, given the low sample sizes) trends due to low sample sizes.

In the spring survey, the proportion of females showed no trend for ages 1-3 and the mean proportion was about 0.6 for all three ages. For age 4, the proportion had an increasing trend, has been highly variable, and a mean of about 0.5. For ages 5 and 6, the proportions are highly variable with no valid trends, and mean proportions of 0.5-0.7.

In the fall survey, the proportion of females shows no trend for age 0 since 1981 and the mean proportion was 0.5. For age 1, the proportion has increased from about 0.5 in the 1980s to about 0.7 since the mid-2000s, with a mean of about 0.6. For age 2, the proportion has increased from about 0.5 in the 1980s to about 0.6 since the mid-2000s, with a mean of about 0.5. For age 3, the proportion was highly variable until about 2000, and has since varied from 0.4 to 0.7 with a mean of about 0.6. For ages 4 and 5, the proportions are highly variable with no valid trends, and mean proportions of about 0.6. Across all NEFSC surveys and ages, the proportion female has varied from 0.4 in 1981 to 0.7 in 2011, with a mean of 0.6.

A4.6 Maturity

Spawning occurs from May through August and peaks in June. Finkelstein (1969b) examined 849 male and 440 female scup and found the length and age at maturity for scup to be 16 cm and two years for both males and females, with spawning between May and July. Morse

(1978) found that about 50% of age-2 scup are sexually mature at about 17 cm total length while nearly all scup of age 3 and older are mature. O'Brien et al. (1993) used NEFSC spring trawl survey data for 1985 and 1987-1990 (516 total fish) and estimated L50% to be 15.6 cm for males and 15.5 cm for females.

For this benchmark assessment of scup, available maturity at age data from the NEFSC spring trawl survey for 1981-2013 (34 years) have been examined. The current data set consists of 1,472 males from age 1 to 10 and 1,828 females from age 1 to 11, for a total of 3,300 fish. The median length at maturity (50th percentile, L_{50}) was estimated at 15.6 cm (95% CI from 13.5 to 18.0 cm) for males, 16.3 cm (95% CI from 14.0 to 18.6 cm) for females, close to the Finkelstein (1969b), Morse (1978), and O'Brien et al. (1993) estimates noted above.

For the 1981-2013 NEFSC time series, the observed percent mature of males is 12% at age 1, 81% at age 2, 96% at age 3, and 100% for age 4 and older. The observed percent mature of females is 12% at age 1, 76% at age 2, 97% for age 3, and 100% for age 4 and older. The observed percent mature of sexes combined for the time series is 12% at age 1, 76% at age 2, 97% at age 3, and 100% for age 4 and older. Estimated maturity ogives for the time series indicate the maturity of both males and females to be 4% at age 1, 76% at age 2, and 100% at ages 3 and older, and for sexes combined to be 4% at age 1, 71% at age 2, 99% at age 3, and 100% at ages 4 and older.

The NEFSC spring survey data were pooled into three year time blocks (except for the first [1981-1984] and last [2009-2013] blocks) to look for trends or abrupt changes in the observed proportions mature over time. For many of the blocks, the male and female patterns are very similar, generally with age 1 observed maturity at 0-10%, age 2 at 60-80%, and age 3 at 90-100%. For some of the blocks (1991-1993, 1994-1996, 1997-1999) there is more divergence between the sexes at age 2. The most recent 2009-2013 block shows the lowest observed proportion mature for both sexes at age 2, with males at 63% and females at 61%, and sexes combined at 62%.

The next step was to estimate maturity ogives for three-year moving windows, in an attempt to stabilize the inter-annual variability and improve precision. Estimated three-year proportions mature for ages 1, 2, and 3 by sex provided a relatively smooth inter-annual pattern. Finally, in keeping with the approach from the previous benchmark assessment (NEFSC 2009), a sexes combined three-year moving window ogive was compiled from the NEFSC 1981-2014 spring survey data to be used with the fishery catch at age to compute SSB in the assessment model. The three-year moving window approach provides a) well-estimated proportions mature at age, b) estimated maturities at age that transition smoothly over the course of the time series, and c) reflect the recent trend of decreasing maturity at ages 1 and 2 (see table below). The average of the values for 1981-1983 (i.e., maturity at ages 0 and 1 = 0.00, maturity at age 2 = 0.83, maturity at ages 3+ = 1.00) was used in subsequent modeling for years before 1981.

MAT3	0	1	2	3	4	5	6	7+
1981	0.00	0.00	0.89	1.00	1.00	1.00	1.00	1.00
1982	0.00	0.00	0.83	1.00	1.00	1.00	1.00	1.00
1983	0.00	0.00	0.78	1.00	1.00	1.00	1.00	1.00
1984	0.00	0.01	0.68	1.00	1.00	1.00	1.00	1.00
1985	0.00	0.25	0.83	1.00	1.00	1.00	1.00	1.00
1986	0.00	0.21	0.77	1.00	1.00	1.00	1.00	1.00
1987	0.00	0.21	0.78	1.00	1.00	1.00	1.00	1.00
1988	0.00	0.06	0.67	1.00	1.00	1.00	1.00	1.00
1989	0.00	0.01	0.83	1.00	1.00	1.00	1.00	1.00
1990	0.00	0.01	0.90	1.00	1.00	1.00	1.00	1.00
1991	0.00	0.03	0.76	1.00	1.00	1.00	1.00	1.00
1992	0.00	0.03	0.68	1.00	1.00	1.00	1.00	1.00
1993	0.00	0.06	0.55	1.00	1.00	1.00	1.00	1.00
1994	0.00	0.06	0.70	1.00	1.00	1.00	1.00	1.00
1995	0.00	0.08	0.73	1.00	1.00	1.00	1.00	1.00
1996	0.00	0.05	0.89	1.00	1.00	1.00	1.00	1.00
1997	0.00	0.02	0.86	1.00	1.00	1.00	1.00	1.00
1998	0.00	0.01	0.89	1.00	1.00	1.00	1.00	1.00
1999	0.00	0.01	0.85	1.00	1.00	1.00	1.00	1.00
2000	0.00	0.02	0.81	1.00	1.00	1.00	1.00	1.00
2001	0.00	0.05	0.80	1.00	1.00	1.00	1.00	1.00
2002	0.00	0.08	0.80	1.00	1.00	1.00	1.00	1.00
2003	0.00	0.08	0.74	1.00	1.00	1.00	1.00	1.00
2004	0.00	0.06	0.70	1.00	1.00	1.00	1.00	1.00
2005	0.00	0.02	0.64	1.00	1.00	1.00	1.00	1.00
2006	0.00	0.04	0.79	1.00	1.00	1.00	1.00	1.00
2007	0.00	0.05	0.59	1.00	1.00	1.00	1.00	1.00
2008	0.00	0.06	0.61	1.00	1.00	1.00	1.00	1.00
2009	0.00	0.03	0.54	1.00	1.00	1.00	1.00	1.00
2010	0.00	0.02	0.58	1.00	1.00	1.00	1.00	1.00
2011	0.00	0.02	0.58	1.00	1.00	1.00	1.00	1.00
2012	0.00	0.02	0.51	1.00	1.00	1.00	1.00	1.00
2013	0.00	0.01	0.58	1.00	1.00	1.00	1.00	1.00
2014	0.00	0.01	0.52	1.00	1.00	1.00	1.00	1.00

A4.7 Predators and Prey

The NEFSC trawl survey foods habits 1973-2013 database was investigated to identify the most frequent predators and prey of scup. Scup was identified to species as a prey item in 527 predator stomachs. Spiny dogfish was the predator in 127 cases (24%), followed by summer flounder (119 cases, 23%), bluefish (59 cases, 11%), monkfish (45 cases, 9%), smooth dogfish (38 cases, 7%), and weakfish (28 cases, 5%), with other fish species accounting for the other 111 cases and 21%, including mostly species of rays, skates, and sharks. The data are insufficient to calculate total absolute predator consumption of scup.

The current investigation confirmed the work of Bowman et al. (2000), which indicated that scup below 25 cm in length consume mainly cnidarians, amphipods, mysids, and annelid and polychaete worms, while scup above 25 cm consume mainly squids and small fish including

silversides and butterfish.

A4.8 Fishery Management

The Mid-Atlantic Fishery Management Council (MAFMC) and Atlantic States Marine Fisheries Commission (ASMFC) jointly manage scup under Amendment 8 (1997) to the Scup, Scup, and Black Sea Bass Fishery Management Plan (FMP). The assessment and management unit includes all scup from Cape Hatteras, North Carolina north to the US-Canada border.

Amendment 8 to the FMP established a recovery plan for scup under which exploitation rates were to be reduced to 47% (F=0.72) during 1997-1999, to 33% (F=0.45) during 2000-2001, and to 21% (F=0.26) during 2002-2007. These goals were to be attained through implementation of a Total Allowable Catch (TAC) that included a commercial quota and a recreational harvest limit, commercial fishery trips limits, commercial fishery net minimum mesh sizes, fish trap minimum escape vent and fish sizes and closed areas, and recreational fishery minimum fish sizes, possession limits, and closed seasons.

Amendment 12 (1998) to the FMP established a biomass threshold (a proxy for one-half BMSY) for scup based on the three-year moving average of the NEFSC spring bottom trawl survey index of Spawning Stock Biomass (SSB) during 1977-1979, which was perceived to be a period when the stock was near one-half BMSY. The scup stock was considered to be overfished when the SSB index fell below a value of 2.77 SSB kg per tow. Amendment 12 defined overfishing for scup to occur when the fishing mortality rate exceeded the threshold fishing mortality of Fmax = 0.26 (as a proxy for FMSY).

Broad scale Gear Restricted Areas (GRAs) for scup were implemented in November 2000 under the framework provisions of the FMP to reduce discards of scup in the small mesh fisheries for *Loligo* squid and silver hake. Two Northern Areas off Long Island were implemented for November through January, while a Southern Area off the mid-Atlantic coast was implemented for January through April. The size and boundaries of the GRAs were modified in December 2000 and again in 2005 in response to commercial fishing industry recommendations.

Amendment 14 (2007) to the FMP defined the biomass target, implemented a stock rebuilding plan for scup, and made the GRAs modifiable through a framework adjustment. The stock was to fully rebuild to the biomass target by January 1, 2015. The proxy for BMSY was two times the 3-year moving average of the NEFSC spring index of SSB during 1977-1979 noted earlier, or 2*2.77 = 5.54 SSB kg per tow. A target fishing mortality rate of F = 0.10 was to be applied in each year of a 7 year rebuilding period beginning in 2008. A TAC of 4,491 mt = 9.901 million lbs and corresponding Total Allowable Landings (TAL) of 3,329 mt = 7.339 million lbs were established for 2008 to achieve the target F.

Amendment 15 (2011) established Annual Catch Limits (ACLs) and Accountability Measures (AMs) for scup to comply with the 2006 reauthorization of the Magnuson-Stevens Act (MSA); Amendment 16 (2013) revised the fishery AMs for each FMP species; Amendment 19 (2014) further modified the AMs for recreational fisheries.

The current overfished and overfishing definitions are based on revisions to the FMP through Framework 7 (2007) and use the values established in Amendments 12 (1998) and 14 (2007) as follows:

"The maximum fishing mortality threshold for each of the species under the FMP is defined as FMSY (the Fishing mortality producing Maximum Sustainable Yield or a reasonable

proxy thereof) as a function of productive capacity, and based upon the best scientific information consistent with National Standards 1 and 2. Specifically, FMSY is the fishing mortality rate associated with MSY. The maximum fishing mortality threshold (FMSY) or a reasonable proxy may be defined as a function of (but not limited to): total stock biomass, spawning stock biomass, total egg production, and may include males, females, both, or combinations and ratios thereof which provide the best measure of productive capacity for each of the species managed under the FMP. Exceeding the established fishing mortality threshold constitutes overfishing as defined by the Magnuson-Stevens Act.

The minimum stock size threshold for each of the species under the FMP is defined as one-half BMSY (or a reasonable proxy thereof) as a function of productive capacity, and based upon the best scientific information consistent with National Standards 1 and 2. The minimum stock size threshold (one-half BMSY) or a reasonable proxy may be defined as a function of (but not limited to): total stock biomass, spawning stock biomass, total egg production, and may include males, females, both, or combinations and ratios thereof which provide the best measure of productive capacity for each of the species managed under the FMP. The minimum stock size threshold is the level of productive capacity associated with the relevant one-half MSY level. Should the measure of productive capacity for the stock or stock complex fall below this minimum threshold, the stock or stock complex is considered overfished. The target for rebuilding is specified as BMSY (or reasonable proxy thereof) at the level of productive capacity associated with the relevant MSY level, under the same definition of productive capacity as specified for the minimum stock size threshold."

A4.9 Previous Stock Assessments

A peer-reviewed assessment including an analytical population model was accepted in 1995 by SAW 19 (NEFSC 1995). The assessment featured a virtual population analysis (VPA) modeled in the ADAPT framework (Conser and Powers 1990), with commercial and recreational landings and discards at age estimates, and with state and NEFSC abundance indices used for calibration. The 1995 SAW 19 assessment indicated that F in 1993 was 1.3, and SSB was 4,600 mt = 10.141 million lbs. A yield per recruit (YPR) analysis indicated that Fmax = 0.236.

The VPA was updated through 1996 and reviewed by the 1997 SAW 25 (NEFSC 1997), but due to concerns over the low intensity of fishery length sampling in the 1990s, uncertainty about the magnitude of commercial discards in the late 1990s, and the ongoing high variability and imprecision of survey indices, the VPA was not accepted as a basis for management decisions. Assessment conclusions were therefore based primarily on trends in NEFSC and state agency survey indices and catch curve analyses using those survey data. The 1997 SAW 25 was able to conclude that in 1996 scup were over-exploited and near record low abundance levels.

The scup assessment was next updated through 1997 and reviewed by the 1998 SAW 27 (NEFSC 1998). Several configurations of a surplus production model (ASPIC; Prager 1994) were reviewed in addition to an updated VPA, but like the VPA, the production model results were not accepted due to concerns over the validity of the input fishery and survey data. An updated YPR analysis was accepted and indicated that Fmax = 0.26. The 1998 SAW 27 concluded that a VPA or other analytical model formulation for scup would not be feasible until the quality of the input data, particularly the precision of discard estimates, was significantly improved and that scup was over exploited and at a low biomass level.

The 1998 SAW 27 Panel recommended the scup assessment be based on the long-term

time series of NEFSC trawl survey indices and fishery catches. The Panel noted that commercial landings were sustained at about 19,000 mt = 41.888 million lbs annually during the mid-1950s to mid-1960s, and concluded that the stock was likely near BMSY during that period (Figure A1). The nearest subsequent peak in NEFSC survey indices occurred in the late 1970s. Commercial and total fishery catches in the late 1970s were about one-half of those in the 1950s to 1960s, and so the late 1970s were identified as a period when the stock was likely to have been near one-half of BMSY. The Panel considered the NEFSC spring survey series to be most representative of SSB, since older ages were better represented in the age structure than in the NEFSC fall survey or other state agency surveys. The 1998 SAW 27 Panel recommended that the three-year moving average of the NEFSC spring bottom trawl survey index of SSB during 1977-1979 (2.77 SSB kg per tow) be used as the proxy biomass threshold (one-half BMSY) and that Fmax = 0.26 be used as the proxy fishing mortality threshold (FMSY). Those recommendations were subsequently adopted for the biological reference points in Amendment 12 to the FMP.

The scup assessment was next updated through 1999 and reviewed by the 2000 SAW 31 (NEFSC 2000). The assessment continued to be based on trends in research survey indices and fishery catches and indicated that the stock was overfished and that overfishing was occurring. The stock assessment was reviewed again by the 2002 SAW 35 and included fishery data through 2001 (NEFSC 2002). The assessment was again based on trends in research survey indices and fishery catches, but indicated that the stock was no longer overfished, although the 2002 SAW 35 Panel concluded that stock status with respect to the overfishing definition could not be evaluated due to the uncertainty of F estimates derived from research survey catch curve calculations. The 2002 SAW 35 Panel found sufficient evidence to conclude that the relative exploitation rates had declined in recent years and that survey observations indicated strong recruitment and some rebuilding of age structure.

During 2002-2008, the status of the stock was evaluated by the MAFMC Monitoring Committee using trends in research survey indices and fishery catches. A relative exploitation index based on the annual total fishery landings and the NEFSC spring three-year average SSB index was used as a proxy for F to monitor status with respect to overfishing and provide guidance to the specification of the annual TAC. A projection of the NEFSC spring survey SSB index using assumptions about maturity, partial recruitment to the survey, and the level of future recruitment as indexed by the NEFSC spring survey at age 1 was used in Amendment 14 to the FMP to forecast stock rebuilding and set the F target for 2008-2015. An update to the status monitoring metrics was completed in 2008 to aid in the specification of fishery regulations for 2009. The update indicated that while the stock was overfished in 2007, the exploitation rate was at about the F target, suggesting that overfishing was not occurring in 2007. However, the stock rebuilding progress was slower than forecast by the Amendment 14 projection, with the NEFSC spring 2007 SSB index (three-year average = 1.16 kg per tow) at only 56% of the projected 2007 index (2.08 kg per tow).

The most recent benchmark peer review of the scup assessment was conducted by the 2008 Northeast Data Poor Stocks Working Group (DPSWG) Peer Review Panel (NEFSC 2009), which accepted an ASAP (A Stock Assessment Program; Legault and Restrepo 1988, NFT 2008) statistical catch at age (SCAA) model as the basis for status determination, with fishery and survey catch data through 2007. The new model of scup population dynamics was expected to provide a more stable tool for monitoring stock status and specifying annual fishery regulations than the previous single index-based model. The assessment indicated that the stock was not

overfished and overfishing was not occurring in 2007, relative to the revised biological reference points. Fishing mortality was estimated to have decreased rapidly after 1994, with F in 2007 = 0.054. With greatly improved recruitment and relatively low fishing mortality rates since 1998, SSB was estimated to have steadily increased to about 119,300 mt = 263 million lbs in 2007. There was no consistent retrospective pattern in F, SSB, or recruitment evident in the 2008 assessment model. Following the 2008 DPSWG stock assessment, the NMFS declared scup to be officially rebuilt in 2009.

The 2008 benchmark was last updated in 2012 (Terceiro 2012) using the same model configuration as the 2008 DPSWG (NEFSC 2009) benchmark and subsequent 2009-2011 assessment updates (Terceiro 2009, 2010, 2011). The updated population model included with fishery and survey catch information through 2011. The 2012 update found the stock was not overfished and overfishing was not occurring in 2011 relative to the 2008 biological reference points. The fishing mortality rate (F) was estimated to be 0.034 in 2011, below the fishing mortality threshold reference point = FMSY = F40% = 0.177. Spawning Stock Biomass (SSB) was estimated to be 190,424 metric tons (mt) = 420 million lbs in 2011, above the biomass target reference point = SSBMSY = SSB40% = 92,044 mt = 203 million lbs.

A5. TERM OF REFERENCE 1: Estimate catch from all sources including landings and discards. Include recreational discards, as appropriate. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the uncertainty in these sources of data.

A5.1 Commercial Fishery Landings

Commercial landings of scup peaked in 1960 at 22,200 mt, then decreased during the 1960s and ranged between about 5,000 and 10,000 mt until the late 1980s. Commercial landings averaged 4,900 mt annually during 1987-1996. Commercial fishery quotas were implemented in 1997, and landings then ranged between 1,200 mt and 8,100 mt and averaged 4,000 mt during 1997-2014, about 54% of the total catch. Reported 2014 commercial fishery landings were 7,228 mt = 15.935 million lbs, about 77% of the commercial quota (Figure A1). About eighty percent of the commercial landings of scup since 1979 were landed in Rhode Island (38%), New Jersey (26%), and New York (16%; Table A1). The otter trawl is the principal commercial fishing gear, accounting for about 65%-90% of the annual total commercial landings since 1979 (Table A2). The remainder of the commercial landings is taken by floating trap (~10%), hand lines (~5%), and fish pots (~5%), with paired trawl, pound nets, and other types of pots and traps each contributing between 1 and 4%.

The distribution of commercial landings by 3-digit statistical area indicated that scup were taken from 43 different areas, but with just 12 accounting for more than 1% of the cumulative total since 1964, lead by area 616 (20%) off northern NJ and western Long Island NY in the Hudson Canyon area, areas 537 (16%), 538 (12%), and 539 (9%) off RI and MA, area 622 (15%) off southern New Jersey and Delaware Bay, and area 613 (9%) off Long Island NY (Figure A2). The distribution of commercial fishing effort for scup expressed as days fished has a similar pattern of concentration, but areas 537-539 off RI and MA account for higher percentages than in the reported landings (Figure A3). It should be noted that not all states routinely reported all landings and effort data to the federal Dealer reporting system until the late 1980s. The distribution of landings by tonnage class (TC) indicated that about 60% of the landings were taken by tonnage class 3 vessels.

A5.2 Fishery Dependent Data Indices of Abundance (LPUE and CPUE)

In response to fishing industry (both commercial and recreational) comments that the utility of fishery dependent catch per unit effort (CPUE) should be evaluated as indices of abundance for scup, a subset of the 2015 SAW 60 Scup Working Group (SWG) with an interest in fishery dependent CPUE compiled data and conducted analyses from a number of sources. These sources include 1) the commercial Dealer reported data for trawl gear, 2) the commercial fishing vessel trip reports (VTR) data for trawl gear, 3) the Northeast Fishery Observer Program (NEFOP) data for trawl gear, 4) the recreational for-hire fishing vessel VTRs for rod-and-reel gear, and 5) the Marine Recreational Fishery Statistics Survey / Marine Recreational Information Program (MRFSS/MRIP) data for rod-and-reel gear, and 6) commercial Study Fleet detailed catch per tow information. This information was reported in 6 separate working papers that were considered during the winter of 2014-2015 by the SWG.

The SWG evaluated the fishery dependent landings or catch per unit effort indices and their utility as indices of abundance in the scup stock assessment. The SWG noted generally that 1)

the utility of the fishery dependent data as the basis for indices of abundance is limited in that some of them include only landings and not the total catch including discards, and so the resulting LPUE could be biased low relative to the true abundance of fish, 2) the use of only positive trips that catch scup may bias the LPUE or CPUE as well, and may be influenced by management regulations, and 3) the ratio of catch to effort has generally changed over time, and it is unclear how this change reflects real changes over time in fishing behavior due to fish abundance, management regulations, or changes in data reporting systems.

The SWG noted that over the long term, and especially since fishery quotas and harvest limits were instituted in 1997, there have been a number of associated regulatory changes, primarily seasonal trip limits and mesh regulations, which are different in timing and magnitude for each year. This information is not part of the fishery catch databases and must be developed independently and integrated within the Generalized Linear Models. This information generally could not be modeled adequately as classification variables within the generalized model framework (i.e., inability to develop a model which converges and produces valid parameter estimates).

At a conference call meeting in late March 2015, a subset of the SWG with an interest in fishery dependent CPUE recommended that the lead assessment scientist investigate the utility of 'directed scup trips' from the Dealer landings reports as the basis for an index of abundance. The SWG decided to move forward by using data for '75% scup trips' LPUE (trips for which scup account for 75% or more of the reported landings) in the hope that these strongly 'post-hoc directed' trips would prove a better candidate for the development of a useful fishery dependent index of abundance. The removal of ~200,000 'bycatch' trips for scup (those landing <75% scup) evidently increased the contrast of the cell means across classification strata sufficiently to allow successful estimation of classification effects for the management regulation effects of seasonal trip limits and mesh size. Thus, attempts to include the effects of management measures in the standardized of '75% scup trips' LPUE proved successful, from an estimation standpoint. The resulting '75% scup trip' nominal and model-based indices indicate a nearly flat linear trend in LPUE over the time series.

The SWG decided that the Dealer report standardized LPUE from >75% scup trips was the most appropriate information from which to attempt development of an index of abundance. However, the SWG noted that the resulting LPUE series was different than all other survey and CPUE stock indicators (e.g., slight peak in LPUE in mid 1990s). Figure A4 compares the trends in the fishery dependent nominal and model indices of abundance compiled for this assessment (no Study fleet model indices were compiled). The SWG concluded that further analysis beyond the scope of the assessment is needed to standardize the complexity of factors influencing fishery catch rates.

A5.3 Commercial Fishery Discards

A5.3.1 Current Geometric Mean Discards-to-Landings Ratio Estimates

The NEFSC Northeast Fishery Observer Program (NEFOP) has collected information on landings and discards in the commercial fishery since 1989. Quantifying discards from the commercial fishery is necessary for a reliable scup assessment, but low sample sizes in the past have resulted in estimates of uncertain and relatively low precision. Concern regarding the

uncertainty of discard estimates due to inadequate observer sampling has been expressed in previous SAW reviews of the scup assessment, and those reviews recommended increases in sampling intensity to increase the accuracy and precision of discard estimates (e.g., NEFSC 1995, 1997, 1998, 2000, 2002, 2009). Despite the uncertainty of the discard estimates, recent SAW panels have concluded that commercial discarding of scup has been high during most of the last 20 years, generally approaching or exceeding commercial landings, averaging 43% of the total commercial catch during 1989-2000. Since full implementation of the Gear Restricted Areas (GRAs) in 2001, estimated discards as a proportion of the total commercial catch have decreased, averaging about 33%.

In previous assessments, a method using the Geometric Mean Discards-to-Landings Ratio (GMDL) has been used to estimate scup discards for 1989 and later years. Data were sufficient to estimate directly discards for trawl gear only, and ratio of discards to landings was applied to total landings in order to get total commercial fishery discards. The ratios of discards to landings by trip landings level (for trip landings < 300 kg [661 lbs], the 'bycatch' fishery; or => 300 kg, the 'directed' fishery) and half year period are calculated and multiplied by the corresponding observed landings from the NEFSC Dealer report data to provide estimates of discards. Geometric mean rates (re-transformed, uncorrected, mean ln-transformed Discards to Landings [D/L] per trip) are used because the distributions of scup landings and discards and the ratio of discards to landings on a per-trip basis in the scup fishery are highly variable and positively skewed. Observed trips with both scup landings and discard are used to calculate the per-trip discards to landings ratios. Only trips with both non-zero landings and discards can be used for this approach to avoid division by zero. The number of trawl gear trips used to calculate the geometric mean discard-to-landings ratios (GMDL) by half year for 1997-2007 ranged from 1 to 104 for trips < 300 kg and from 1 to 35 for trips =>300 kg, with the best sampling occurring since 2003. No trawl gear trips were available for half year 2 in 1997 and 1999 for trips < 300 kg and for half year 2 in 1997-2001 for trips => 300 kg. The GMDL calculated for half year 1 was used to estimate discards for half year 2 when no trawl gear trips were available in half year 2. The GMDL ratios ranged from 0.03 in 2004 (half year 2, trips => 300 kg) to 121.71 in 1998 (half year 1, trips => 300 kg).

A large 1998 'directed' fishery discard ratio and subsequent very high annual discard estimate (111,973 mt) was based on one trawl gear trip. About 93% of the discard from that trip was attributable to a single tow in which an estimated 68 mt (~150,000 lbs) of scup were captured. This tow was not lifted from the water and the captain of the vessel estimated the weight of the catch. There has been debate concerning the validity of the catch weight estimate and whether or not it was representative of other vessels or trips in the fishery. However, the observation was reported by a trained NEFSC observer and was therefore included in the initial calculation of the GMDL estimate of scup discards. The 1998 discard estimate was considered infeasible, and replaced by the mean of the 1997 and 1999 GMDL estimates (3,331 mt) in subsequent tabulations of catch and in subsequent modeling (Table A3).

Since 1998 the GMDL approach discard estimates have been adopted by SAW review panels (NEFSC 1998, 2000, 2002) and the MAFMC Monitoring Committee to monitor trends in fishery catch and evaluate the status of the stock. The GMDL approach was accepted by the Data Poor Stocks Workshop peer review of the 2008 assessment as the best method to estimate scup discards (NEFSC 2009). The GMDL estimates were used for all subsequent modeling approaches considered in the 2008 and later assessments.

Broad scale Gear Restricted Areas (GRAs) for scup were implemented in November

2000 to reduce discards of scup in the small mesh fisheries for *Loligo* squid and silver hake. Initially two Northern Areas off Long Island were implemented for November through January, while a Southern Area off the mid-Atlantic coast was implemented for January through April. The size, boundaries, and other measures of the GRAs were modified in December 2000 and again in 2001 and 2005 in response to commercial fishing industry recommendations. Currently a Northern GRA restricts the use of codend mesh less than 5.0 inches (127 mm) during November and December, while a Southern GRA is in effect from January 1 through March 15.

Both the observed discards (as a function of both increased fishing activity for scup and increased sampled trip number) and the current assessment GMDL estimated fishery discards (as a function of increased fishery quotas and therefore increased fishing activity for scup) have generally increased as the fishery quotas have increased since 2005, although the observed discard percentage of total commercial catch has decreased. Scup commercial fishery estimated discards remain an important component of the commercial fishery removals and averaged about 25% of the estimated total commercial catch during 2010-2014.

The distribution of observed discards varies by statistical area, season, and mesh size. Within the nine important GRA 3-digit statistical areas that account for 84% of observed scup discards over the time series, 24% was observed in 'large' mesh tows (codend or liner < 4.5 [114 mm] or 5.0 in [127 mm], 35% in 'small' mesh tows (larger than 2.125 in [54 mm] and smaller than 4.5 or 5.0 inch), and 41% in 'squid' mesh tows (equal to or less than 2.125 inch).

The Observer data have provided evidence that the GRAs have been effective in reducing the scup discard percentage. The current assessment absolute estimates of scup discards using the GMDL approach, however, are produced on a temporal and spatial scale that is too coarse to directly evaluate the effectiveness of specific discard reduction measures (e.g., on a specific area or season basis). This has prompted a re-examination of the methods used to estimate commercial fishery scup discards using the Standardized Bycatch Reporting Method (SBRM).

A5.3.2 New Standardized Bycatch Reporting Method Discard Estimates

The SBRM Omnibus Amendment to the fishery management plans of the Northeast region was implemented in February 2008 to address the requirements of the Magnuson-Stevens Fishery Conservation and Management Act to include standardized bycatch reporting methodology in all FMPs of the New England Fishery Management Council and Mid-Atlantic Fishery Management Council. The SBRM for the estimation of discards (Wigley et al. 2008, 2011) has now been adopted for most NEFSC stock assessments that have been subject to a benchmark review since 2009. In this assessment, newly developed SBRM estimates of scup landings and discards are compared with Dealer reported landings and the current GMDL estimation approach estimates of discards as part of a re-examination of the estimation of commercial fishery scup discards.

Data are still sufficient to estimate discards for trawl gear only, the major commercial gear which has accounted for about 83% of commercial landings since 1989. Based on comments received from fishery managers and industry advisors since the 2008 assessment (NEFSC 2009), under the SBRM approach the trawl gear ratios of discards to landings have not been used to 'raise' trawl discards to account for discards from other gears. The remainder of the commercial gear includes floating traps, hand lines, fish pots, pound nets, and other types of pots and traps. All of these other gears are assumed to either have very low discard rates (e.g., traps,

pots, pound nets) and/or low discard mortality rates (e.g., hand lines), and so dead discards from those gears are assumed to be negligible.

In the SBRM, the sampling unit is an individual fishing trip. Live scup discards or landings were estimated using a stratified d/k ratio estimator (Cochran 1963) where *d* = observed discard or kept pounds of scup, and *k* = observed kept pounds of all species, raised by the trip landings of all species as reported by VTR or Dealer records, to provide estimates of scup discards or landings by stratum. Further computational details are provided in Wigley et al. (2011).

Three SBRM stratification alternatives were evaluated for scup discards and landings:

1) by calendar quarter for all areas and meshes, providing 4 strata annually (QTR4),

2) by calendar quarter for all areas and two mesh categories: 'large' (for codend or liner equal or larger than 4.5 [114 mm] or 5.0 inch [127 mm]) and 'small' (less than 4.5 or 5.0 inch, providing 8 strata (MESH8), and

3) by calendar quarter, statistical area, and three mesh categories: 'large' (for codend or liner equal or larger than 4.5 or 5.0 inch), 'small' (larger than 2.125 inch [54 mm] and less than 4.5 or 5.0 inch, and 'squid' (equal to or less than 2.125 inch), providing 240 strata (MESH240).

The three SBRM alternatives are compared with the current assessment GMDL estimates of discards for 1989-2013 in Table A4 and Figure A5 (note that 2014 data were not available when this work was conducted). Due to the influence of the 'infamous' 1998 tow, all 1998 estimates were replaced with the average of the adjacent years. Over the time series, the current GMDL estimates of discards have averaged 2,397 mt with PSE of 35%. The SBRM QTR4 estimates averaged 1,314 mt with PSE of 39%. The SBRM MESH8 estimates averaged 1,296 mt with PSE of 44%. The SBRM MESH240 estimates averaged 1,376 mt with PSE of 22%. Over the series, the three SBRM alternatives averaged about 1,300 mt, about 45% lower than the GMDL estimates.

The three SBRM alternatives are compared with the current assessment Dealer total and Trawl gear only landings as an additional means of evaluation (Figure A6). Over the 1989-2013 time series, the Dealer total landings have averaged 4,144 mt and the Trawl gear landings have averaged 3,245 mt. The SBRM QTR4 estimates averaged 2,529 mt (38% below the Dealer, 22% below the Trawl) with PSE of 35%. The SBRM MESH8 estimates averaged 1,757 mt (57% below the Dealer, 46% below the Trawl) with PSE of 44%. The SBRM MESH240 estimates averaged 1,831 mt (55% below the Dealer, 44% below the Trawl) with PSE of 18%. Over the series, the three SBRM alternatives averaged about 2,000 mt, about 50% lower than the Dealer landings and 35% lower than the Trawl gear landings. The SBRM MESH240 landings estimates correlate best with the Dealer total and Trawl gear reported landings, with a correlation coefficients (r) of 0.71 and 0.77 ($df = 24$, $p < 0.01$), compared to r values of 0.38 and 0.34 ($p < 0.5$) for the QTR4 estimates and 0.42 and 0.38 ($p < 0.5$) for the MESH8 estimates.

The final comparison made was for the SBRM MESH240 estimates apportioned to length and age (dead discards including the 100% discard mortality rate) with those using the current assessment GMDL estimates of discards. The SBRM estimates in absolute total numbers average 12.5 million fish per year during 1989-2013, about 62% of the GMDL estimate of 20.3

million. The largest difference in absolute total numbers was for 1992, with the GMDL estimate about 58.5 million fish larger than the SBRM estimate; the smallest difference in absolute total numbers was for 2005, with the SBRM estimate about 43,000 fish larger than the GMDL estimate. The largest difference in proportions at age was in 1993 at ages 0, 2, and 3, due to differences in the distribution of discards and subsequent allocation of lengths during the year. Comparable differences, generally at ages 0-2, were observed in 1990, 1992, 1993, 1994, 2001, and 2008.

The consideration of three SBRM discard estimators of scup discards and discards and comparison with the current GMDL method estimates indicates that the SBRM MESH240 estimator and stratification provides the best overall combination of feasible estimates of the scup discards and landings and good precision. The SBRM MESH240 discard estimator also provides the ability to evaluate the effectiveness management measures like the GRAs. The new SBRM MESH240 discard estimate time series (Table A5) is used in the 2015 SAW 60 scup assessment. The commercial fishery live discards of scup have averaged 1,375 mt during 1989-2014, the period for which direct estimates are available.

A5.4 Recreational Fishery Catch

Scup is the object of a major recreational fishery, with the greatest proportion of catches taken in the states of Massachusetts, Rhode Island, Connecticut and New York. Estimates of the recreational catch in numbers were obtained from the NMFS Marine Recreational Fishery Statistics Survey (MRFSS) for 1981-2011, and from the NMFS Marine Recreational Information Program (MRIP) for 2004-2014. These estimates were available for three categories: type A - fish landed and available for sampling, type B1 - fish landed but not available for sampling and type B2 - fish caught and released. The estimated recreational landings (types A and B1) in weight estimated by the programs during 1981-2014 averaged about 2,300 mt per year (Table A6). Since 1981, the recreational landings have averaged 32% of the commercial plus recreational landings total.

The commercial fishery VTR system provides an alternative set of reported recreational landings by the party/charter boat sector. A comparison of VTR reports and MRFSS estimates indicates that MRFSS estimates were on average about 57% higher over the 1995-2014 period, ranging from a factor of 0.34 in 1998 to 2.56 in 2013 (Table A7). It is unclear if this is due mainly to under-reporting of party/charter boat recreational landings in the VTR system, or a systematic positive bias of MRFSS landings estimates for the party/charter boat sector.

The estimated recreational live discard in weight during 1984-2011 ranged from 43 mt in 1999 to a high of 2,120 mt in 2010, averaging about 840 mt per year (Table A8). The weight of discards has been directly calculated only for those years (1984 and later) for which recreational catch at age has been compiled. In compilations of total fishery catch for earlier years, the recreational discards was assumed to be approximately 2% of the estimated recreational landings, based on the mean discard percentage for 1984-1996, the time period with catch at age estimates before the implementation of the FMP. The discard mortality rate in the recreational fishery has been reported to range from 0-15% (Howell and Simpson 1985) and from 0-14% (Williams, pers. comm.). Howell and Simpson (1985) found mortality rates were positively correlated with size, due mainly to the tendency for larger fish to take the hook deep in the esophagus or gills. Williams more clearly demonstrated increased mortality with depth of hook location, as well as handling time, but found no association with fish size. Based on these

studies, a discard mortality rate in the recreational fishery of 15% has been used in this and previous assessments, resulting in a time series average discard mortality of about 100 mt per year.

A5.5 MRIP Estimates of Recreational Fishery Catch

The NMFS Marine Recreational Fishery Statistics Survey (MRFSS) was replaced by the Marine Recreational Information Program (MRIP) in 2012 to provide improved recreational fishing statistics. The MRIP implemented a new statistical method for calculating recreational catch estimates, with many survey elements related to both data collection and analysis updated and refined to address issues such as data gaps, bias, consistency, accuracy, and timeliness. As part of the implementation of the MRIP, MRFSS recreational fishery catch estimates for 2004-2011 have been directly replaced by those using the MRIP estimation methods. For earlier years, a constant "ratio of means" of the MRFSS and MRIP estimates has been used to adjust the recreational catch estimates (Tables A6 & A8).

For the recreational fishery harvest number (catch types A + B1), the largest change was for the commonwealth of MA, with a cumulative 2004-2011 increase of about 4 million fish, about +67% and also the largest cumulative percentage increase amongst the states. The largest absolute decrease was for the state of RI with a cumulative 2004-2011 decrease of about 289,000 fish, or about -7%. The state of MD had the largest cumulative percentage decrease at -67%; however, MD's cumulative harvest (now about 3,600 fish) is only 0.1% of the coastal total. Over all states, the cumulative harvest in numbers increased by about 5.3 million fish (about +19%), ranging from a decrease of 174,000 fish in 2007 (-5%) to an increase of 2.5 million fish in 2004 (+52%; Table A9). Therefore, for the years 1963-2003 recreational harvest numbers were increased by 19% for this assessment (see TOTAL FISHERY CATCH section below for discussion of estimates before 1981).

For the recreational fishery harvest weight (catch types A + B1, mt), the most important change was for the commonwealth of MA with a cumulative 2004-2011 increase of about 1,713 mt, or about +67%. The state of DE had the largest cumulative percentage increase at +112%; however, DE's cumulative harvest (now about 4 mt) is less than 0.1% of the coastal total. The largest absolute decrease was for the state of RI with a cumulative 2004-2011 decrease of about 108 mt, about -6%. The state of MD had the largest cumulative percentage decrease at -30%, a cumulative decrease of about 1 mt. Over all states, the cumulative harvest in weight (mt; metric tons) increased by about 2,433 mt (about +18%), ranging from a decrease of 122 mt in 2008 (-7%) to an increase of 1,356 mt fish in 2004 (+71%; Table A10). Therefore, for the years 1963-2003 recreational harvest weight was increased by 18% for this assessment.

For the recreational fishery live releases in numbers (catch type B2), the largest change was for the commonwealth of MA, with a cumulative 2004-2011 increase of about 3.1 million fish, about +38% and also the largest cumulative percentage increase amongst the states. The largest absolute decrease was for the state of NJ with a cumulative 2004-2011 decrease of about 410,000 fish, or about -12%. The state of MD had the largest cumulative percentage decrease at -47%, a cumulative decrease of about 45,000 million fish. Over all states, the cumulative live release in numbers increased by about 4.5 million fish (about +11%), ranging from a decrease of 239,000 fish in 2008 (-3%) to an increase of 1.7 million fish in 2004 (+36%; Table A11). Therefore, for the years 1963-2003 recreational live release and discard mortality estimates were increased by 11% for this assessment.

A5.6 Commercial Fishery Landings at Length and Age

The NER commercial fishery length frequency sampling is summarized in Table A12 and Figure A7. Annual sampling intensity has varied from 18 to 687 mt per 100 lengths, with sampling exceeding the informal threshold criterion of 200 mt per 100 lengths since 1995. For this assessment, commercial fishery landings at age beginning in 1984 have been updated through 2014, with samples for most of the series pooled by market category (pins/small, medium, large/mix, jumbo, and unclassified) and by half-year (January-June, July-December); samples were pooled on a regional (New England, Mid-Atlantic), quarterly basis (e.g., January-March) where possible since 2004. Estimates of commercial fishery landings at age (Figure A8) and mean weights at age are presented in Tables A13-A14.

A5.7 Commercial Fishery Discards at Length and Age

The intensity of length sampling of discarded scup from the NEFSC Fishery Observer Program declined in 1992-1995 relative to 1989-1991 (Table A15, Figure A7). Sampling intensity ranged from 489 to 335 mt per 100 lengths sampled in 1992-1995, failing to meet the informal criterion of 200 mt per 100 lengths. Sampling intensity improved to 100 mt per 100 lengths in 1996, but then declined to over 200 mt per 100 lengths in 1997-1999. Sampling intensity has generally met the 200 mt per 100 lengths threshold since 2000. The mean weight of the discard was estimated from length frequency data using a length-weight equation, total numbers discarded at length were then estimated by dividing total weight at length by mean weight at length. Discards at length were aged using a combination of commercial and survey age-length keys, with discards at age dominated by fish aged 0, 1, or 2, depending on the year under consideration. Estimated proportions at length and age for 1984-1988 (before the advent of the Observer sampling) were derived from irregularly collected NEFSC samples (NEFSC 1998) and the ratio of scup discards to scup landings during 1989-1991 (0.50 for the GMDL estimates; 0.46 for the SBRM estimates). Estimates of commercial fishery discards at age (Figure A9) and mean weights at age are presented in Tables A16-A17.

A5.8 Recreational Fishery Landings at Length and Age

For the recreational fishery, length sampling intensity has varied from 45 to 471 mt per 100 lengths. Sampling in all years except 1984 during 1981-1987 failed to meet the informal criterion of 200 mt per 100 lengths, but since 1988 the criterion has been met except for 1999-2000 (Table A6, Figure A7). Numbers at length for recreational landings were determined from recreational fishery length samples pooled by half-years (January-June; July-December) over all regions and fishing modes, and were converted to numbers at age by applying half-year age-length keys constructed from NEFSC commercial and survey samples. Age-length keys from spring surveys and first and second quarter commercial samples were applied to numbers at length from the first half of the year, while age-length keys from fall surveys and third and fourth quarter commercial samples were applied to numbers at length from the second half of the year. Estimates of recreational fishery landings at age (Figure A10) and mean weights at age are presented in Tables A18-A19.

A5.9 Recreational Fishery Discards at Length and Age

No length frequency samples of the scup discard were collected under the MRFSS program before 2005, so recreational discards were assumed to be fish aged 0 and 1, in the same relative proportions and with the same mean weight as the landed catch samples less than state regulated minimum fish sizes. An inspection of discard length frequency samples from the New York recreational fishery for 1989-1991 indicated that this assumption was reasonable. Since 2005, the MRFSS/MRIP For-Hire Survey discard samples have been used in concert with the MRFSS/MRIP sub-legal landed lengths to characterize the length frequency of the recreational discard. The informal sampling criterion of 200 mt per 100 lengths has been consistently met since 2007 (Table A8, Figure A7). Numbers at length were converted to numbers at age by applying half-year (January-June; July-December) age-length keys constructed from NEFSC commercial and survey samples. As noted earlier, a 15% discard mortality rate is assumed. Estimates of recreational fishery discards at age (Figure A11) and mean weights at age are presented in Tables A20-A21.

A5.10 Total Fishery Catch

Total commercial and recreational landings in 2014 were 9,253 mt = 20.399 million lbs and total commercial and recreational discards were 1,367 mt = 3.014 million lbs, for a total catch in 2014 of 10,620 mt = 23.413 million lbs (Table A22, Figure A12). Estimates of the total fishery catch at age and mean weights at age (Figure A13) for 1984-2014 (the time series is limited by the availability of sampled fishery ages) are presented in Tables A23-A24. An extended time series of the total catch of scup has been estimated to provide an historical perspective of the exploitation of scup in the years before a) the MRFSS/MRIP was implemented in 1981 to estimate recreational fishery catch, b) the Observer program was implemented in 1989 to provide estimates of commercial fishery discard, and c) fishery aging data became available in 1984 (Table A25). These estimates include commercial and recreational landings and discards. The recreational fishery catch for 2004-2014 has been estimated using the MRIP methods. For earlier years, a constant "ratio of means" of the MRFSS and MRIP estimates has been used to adjust the recreational catch estimates (see previous MRIP section).

The catches before 1981 are the less reliable due to uncertainty about a) the magnitude of domestic commercial fishery discards, b) the magnitude of the distant water fleet (DWF) catch and c) the uncertainty of assumptions made to estimate the recreational catch (50% reduction from estimates based on time-varying ratios to the commercial landings made in Mayo 1982 for 1960-1978; recreational discards assumed to be 2% of the adjusted recreational landings). For years in which no commercial fishery observer data were collected (1963-1988), commercial discards were computed using a constant "ratio of means" using landings and discards for 1989-2001 (0.50 for the GMDL estimates) as in previous assessments (NEFSC 2002; NEFSC 2009). This ratio for the SBRM estimates adopted for the 2015 SAW 60 assessment is 0.46.

A6.TERM OF REFERENCE 2: Present the survey data being used in the assessment (e.g., indices of relative or absolute abundance, recruitment, state surveys, age-length data, etc.). Characterize the uncertainty and any bias in these sources of data.

A6.1 Research Suvey Indices of Abundance

A6.2 Northeast Fisheries Science Center

The NEFSC spring and fall bottom trawl surveys provide long time series of fishery-independent indices for scup. The NEFSC spring and fall surveys are conducted annually during March-May and September-November, ranging from just south of Cape Hatteras, NC to Canadian waters. NEFSC spring and fall abundance and biomass indices for scup exhibit considerable inter-annual variability (Table A26, Figure A14). NEFSC spring survey catches are characterized mainly by scup of ages 1 and 2 (Figure A15), while the fall survey often captures large numbers of age 0 and 1 fish (Figure A16).

The Fisheries Survey Vessel (FSV) *Albatross IV* (ALB) was replaced in spring 2009 by the FSV *Henry B. Bigelow* (BIG) 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 BIG 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 three 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 methods proposed in Miller et al. (2010), and the precedents set in peer-reviews of stock assessments for haddock (Van Eeckhaute and Brooks 2010), yellowtail flounder (Legault et al. 2010), silver and red hake (NEFSC 2011a), and winter flounder (NEFSC 2011b), aggregate and length-based calibration factors were used to convert 2009-2014 spring and fall BIG survey catch number and weight indices to ALB equivalents for use in this stock assessment update (Tables A27-A30; Figure A14).

The NEFSC survey indices sometimes appear to mainly reflect the availability of scup to the survey, rather than true abundance, making it difficult to interpret large inter-annual changes in the indices. For example, the 2002 spring biomass index was about twice the second highest spring index, which was observed in 1977 (Figure A14). The spring numeric abundance indices are similar; the 2002 index is the highest observed in the series and about twice the 1970 index. These dramatic increases were evident across all ages in the estimated 2002 spring numbers at age (Table A31; Figure A15). However, the previous fall survey estimates of numbers at age in 2001 had not reflected relatively large values from which the corresponding 2002 spring numbers at age might have been expected to derive (Table A32, Figure A16) nor did they subsequently translate to exceptional indices of biomass in fall 2002 or spring 2003. A potentially similar ‘availability’ event appears to have occurred in spring 2014, with the largest biomass and numeric indices sampled since 2002, but with no follow-up apparent in the 2014 fall indices (Tables A26-A27).

The NEFSC winter survey was started in 1992 primarily as a flatfish survey, was conducted during February, and ranged from Cape Hatteras, NC to the southwestern part of Georges Bank. The winter survey 2002 abundance and biomass indices were, like the spring survey, the largest of the time series (Table A33, Figure A13). Similar to the spring estimates, numbers at age estimated for the 2002 winter survey were also exceptionally large (Table A34, Figure A17). The winter trawl series ended in 2007.

The large differences in the absolute magnitude of NEFSC survey catches of ages 0-2 compared to those of fish at ages 3 and older suggests a substantial difference in survey selection at age between these two aggregate age groups. In the 2008 DPS assessment (NEFSC 2009), aggregate biomass indices restricted to the lengths of fish ages 0-2 were constructed for calibration of those ages in the population model (maximum length of 22 cm in the winter, 20 cm in the spring, and 23 cm in the fall series). The 2009-2014 BIG values for these aggregate indices have also been converted to ALB equivalents using length calibration factors (Table A35). Both the NEFSC spring and fall indices indicate an increasing trend in scup abundance since the late 1990s.

Alternate NEFSC strata sets

Only about one-third (spring) to one-half (fall) of the 30 offshore strata included in the standard assessment long-term aggregate spring and fall (offshore strata 1-12, 23, 25, 61-76) strata sets account for large proportion of the scup catches. In the spring, these are the 'middle two' bands of offshore strata with depths from 56 to 185 meters (about 30 to 100 fathoms), and from North to South include strata 2, 3, 74, 75, 70, 71, 66, 67, 62, and 63. In the fall, these are the 'inner two' bands of offshore strata with depths from 27 to 110 meters (about 15 to 60 fathoms), and from North to South include strata 9, 10, 5, 6, 1, 2, 73, 74, 69, 70, 65, 66, 61, and 62. These two groups of seasonal strata were used to construct candidate 'Alternate' offshore strata sets for the long-term aggregate indices used for scup. The spring Alternate set of 10 strata includes 97.5% of the time series total catch, while the fall Alternate set of 14 strata includes 99.8% of the time series total catch. The goal of developing indices using the alternate sets was to explore if the inter-annual variability and occasional extreme 'outliers' (e.g., spring 2002) in the time series might be reduced, before attempting the development of model-based indices.

The alternate series indices for both seasons are, as expected, scaled higher as the strata that were omitted had low catches. When normalized to each respective time series mean, however, trends were very similar for both abundance and biomass indices for both seasons. The alternate series indices also had slightly higher variance, because the omitted strata catches generally had small or zero variance. The time series Proportional Standard Error (PSE: the ratio of the time series standard error to the time series mean) increased from 129% to 135% for the spring number per tow index, and from 95% to 97% for the fall. PSE magnitudes and changes were comparable for the seasonal biomass indices. More importantly, no significant reduction in inter-annual variation was realized. Given these results, the standard assessment NEFSC strata sets and stratified random indices of abundance were retained for use in the 2015 SAW 60 assessment.

Model-based NEFSC indices of abundance

Descriptive statistics indicate that the NEFSC survey scup catch distribution is highly contagious and overdispersed in relation to a normal distribution. For both spring and fall, examination of patterns in the survey catch, for both day and night, confirm the irregular distributions of catch by temperature, salinity and depth and portend the difficulties of modeling the survey scup catch data. No well defined relationships are evident; i.e., small catches are as likely to be taken at shallow depths as large depths and at both warm and cold temperatures and large catches can occur over a relatively large range of depth and temperature (e.g, over a range of 70 meters or 10 degrees). Generalized linear model (GENMOD) and generalized additive model (GAM) based indices of abundance for the scup NEFSC seasonal survey data proved to be not useful, due to highly variable results owing from the inability of the models to adequately fit the variable and complex temporal and spatial properties of scup survey catches.

A6.3 Massachusetts DMF

The Massachusetts Division of Marine Fisheries (MADMF) has conducted spring and fall bottom trawl surveys of Massachusetts territorial waters in May and September since 1978. Survey coverage extends from the New Hampshire to Rhode Island boundaries and seaward to three nautical miles, including Cape Cod Bay and Nantucket Sound. The study area is stratified into geographic zones based on depth and area. The MADMF spring survey catches are characterized mainly by scup of ages 1 and 2, while the fall survey often captures large numbers of age 0 fish. The spring biomass and abundance indices decreased sharply from a high in the early 1980s to relatively low levels through the 1990s, and have since exhibited a variable but increasing trend (Table A36, Figure A18). The MADMF fall abundance index can include large numbers of age 0 fish and therefore can be more variable as it reflects inter-annual variance in recruitment. The fall biomass index exhibits an increasing trend since the mid 1990s (Table A36, Figure A18).

A6.4 Rhode Island DFW

The Rhode Island Division of Fish and Wildlife (RIDFW) has conducted spring and fall bottom trawl surveys based on a stratified random sampling design since 1979. Three major fishing grounds are considered in the spatial stratification, including Narragansett Bay, Rhode Island Sound, and Block Island Sound. Stations are either fixed or randomly selected for each stratum. The spring index shows relatively low scup abundance and biomass through 1999 followed by a steep increase during 2000-2002, in common with the NEFSC and MADMF indices, and high variability since then (Table A37; Figure A19). The RIDFW spring survey catches a full age range of scup of ages 1 through 7+ (Table A38, Figure A20). The RIDFW fall survey indices show a general increase to a 1993 peak, followed by a steep decline until 1998, and a steady increase since then. The fall biomass series reached a time series peak in 2011 (Table A37, Figure A18). The RIDFW fall survey is dominated by age 0 scup (Table A39, Figure A21).

The RIDFW implemented a ventless trap survey in cooperation with commercial fishermen beginning in 2005 and ending in 2012 (Table A40, Figure A19). The cooperative trap survey has a fixed station format, and survey catches are expressed as catch per trap soak hour.

The RIDFW cooperative trap survey caught a full age range of scup of ages 1 through 7+ (Figure A22).

A6.5 University of Rhode Island Graduate School of Oceanography (URIGSO)

University of Rhode Island Graduate School of Oceanography (URIGSO) has conducted a standardized, year-round, weekly two-station trawl survey in Narragansett Bay and Rhode Island Sound since the 1950s, with consistent sampling since 1963. Irregular length-frequency samples for scup indicate that most of the survey catch is of fish from ages 0 to 2. The aggregate numbers-based index reached a peak in the late 1970s, was relatively low during the late 1990s, and has since generally increased. The 2014 index was the third highest of the time series, after the 1976 and 1989 indices (Table A41, Figure A23).

A6.6 Connecticut DEEP

The Connecticut Department of Energy and Environmental Protection (CTDEEP) trawl survey program was initiated in May 1984 and encompasses both the New York and Connecticut waters of Long Island Sound. The stratified random design survey is conducted in the spring (April-June) and fall (September-October). The CTDEEP spring index indicates relatively low abundance through most of the survey period, but has increased substantially since 1999 (Table A42, Figure A24). The CTDEEP fall survey, which often catches large numbers of age-0 scup, indicates that recruitment was relatively stable during most of the survey period, but the aggregate fall indices have also increased substantially since 1999. (Table A43, Figure A22) Due to vessel engine failure, a complete fall survey was not conducted in 2010. The CTDEEP spring and fall surveys catch scup from ages 0-7+ (Figures A25-A26).

A6.7 New York DEC

The New York Department of Environmental Conservation (NYDEC) initiated a small mesh trawl survey in 1985 to collect fisheries-independent data on the age and size composition of scup in local waters. This survey is conducted in the Peconic Bays, the estuarine waters which lie between the north and south forks of eastern Long Island. The NYDEC survey provides age 0, 1, and 2+ indices of scup abundance (Table A44). The index of age 2 and older fish indicates a substantial increase since the late 1990s (Figure A27). The age 0 indices indicate recruitment of strong cohorts since the late 1990s. In the early years of the survey, however, there often was not been a strong correspondence between the age 0 indices and age 1 and 2+ indices in the following years (Figure A28).

A6.8 New Jersey DFW

The New Jersey Department of Fish and Wildlife (NJDFW) conducts a stratified random bottom trawl survey of New Jersey coastal waters from Ambrose Channel south to Cape Henlopen Channel. Latitudinal strata boundaries correspond to those in the NEFSC trawl survey; longitudinal boundaries correspond to the 30, 60, and 90 foot isobaths. Each survey includes two tows per stratum plus one additional tow in each of nine larger strata for a total of 39 tows. The NJDFW survey indices exhibit variable patterns over the early part of the time

series. The biomass index reached a minimum in 1996 and then generally increased, peaking in 2007, but has since decreased (Table A45; Figure A29).

A6.9 Virginia Institute of Marine Science (VIMS)

A6.9.1Juvenile Fish Trawl Survey

The Virginia Institute of Marine Science (VIMS) has conducted a juvenile fish trawl survey in lower Chesapeake Bay during June-September since 1988. The VIMS age-0 scup indices indicate a general decline in recruitment from relatively high levels with peaks in the late 1980s to early 1990s, to relatively low levels from the late 1990s to early 2000s, and the indication of several recent strong year classes (Table A45).

A6.9.2 ChesMMAP Trawl Survey

The VIMS Chesapeake Bay Multispecies Monitoring and Assessment Program (ChesMMAP) trawl survey is designed to support stock assessment activities at both a single and multispecies scale. While no single gear or monitoring program can collect all of the data necessary for quantitative assessments, ChesMMAP was designed to fill data gaps by maximizing the biological and ecological data collected for several recreationally and commercially important species in the bay. Total abundance and biomass indices composed mainly of age 0 and 1 fish are available since 2002, and suggest strongest recruitment in 2005 and 2010 (Table A46, Figures A30-A31).

A6.9.3 NEAMAP Trawl Survey

The VIMS Northeast Area Monitoring and Assessment Program (NEAMAP) trawl survey was started in fall 2007, providing 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 survey data are included for the first time in the 2015 SAW60 scup assessment population model (Table A47-A48, Figures A29, A32-A33).

A6.10 Aggregate research survey trends

Figure A34 presents the trends in aggregate indices of numeric abundance for the 16 surveys used in the assessment (the 17th is the VIMS juvenile fish trawl survey). The historical peak in the 1970s is evident, as is the decrease to a minimum in the late 1990s. Most surveys indicate an increase in abundance since the late 1990s, some to historic highs.

Figure A35 presents the trends in scup recruitment at age 0 for the 8 surveys with significant catch of age 0 scup. Multiple surveys indicated good recruitment in the late 1980s, poor recruitment in the mid-1990s, and improved to historically high recruitment during the 2000s. Some surveys indicate decreased recruitment since about 2010.

A6.11 Integrated Indices of Abundance

A6.11.1 Aggregate and At-Age indices from General Linear Modeling (GLM)

Several of the Northeast United States fish stock assessments conducted by Northeast Regional Stock Assessment Workshop (SAW) Working Groups and Atlantic States Marine Fisheries Commission (ASMFC) Technical Committees incorporate abundance indices from several state and federal agency research survey programs (e.g., summer flounder, winter flounder, bluefish, black sea bass, striped bass, weakfish, tautog, scup, etc.). Typically, this information is provided to the assessment process as annual or seasonal aggregate indices of biomass or numbers, and sometimes as indices at age. These indices can be used in complex, age-structured analyses to calibrate population trends and relative cohort size.

The evaluation process of candidate indices for use in complex models has typically included looking for common trends (i.e., signal) by: a) examination of time series plots, b) analysis of correlation (of lack thereof) between survey indices and between survey indices and population dynamics model results, c) outlier analysis, and d) consideration of the magnitude and trend of residuals when indices are included in population dynamics models such as VPA and ASAP. Multiple analyses with different sets of indices are often conducted to examine the sensitivity of model results to inclusion of a given index series to determine the best analysis configuration to characterize stock status. Alternatively, all available abundance indices may be included in an analysis with the results most strongly influenced by those indices that statistically fit best within the analytical framework. Even given these approaches, with 50 or more indices of abundance at age from up to 15-20 surveys (as in this assessment of scup) to consider for inclusion in a complex age structured assessment, it can be difficult to qualitatively discern general trends in abundance from the battery of available indices. The decision to include a given index time series at age can therefore often be subjective, based on a loose set of decision rules that may vary from one assessment to another. SAW peer reviews have often recommended the investigation of methods to better integrate trends in stock abundance inferred from survey indices of abundance, prior to the inclusion of such indices in a population model calibration. A review of NEFSC data collection programs (NEFSC 2013a) recommended: "…better integration of NEFSC and state surveys. This could include planning efforts to standardize timing and methods, to improve comparability among surveys. On the stock assessment side, panelists questioned the appropriateness of giving equal weight to a survey covering the whole range, compared to a large set of geographically restricted surveys of unknown rigor."

The integration of survey indices collected by different research sampling programs can be viewed as analogous to the standardization of commercial fishing vessel catch rates in developing fishery-dependent indices of abundance (e.g., Robson 1960, Gavaris 1980, Kimura 1981, O'Brien and Mayo 1988). Viewed in that light, a Generalized Linear Model framework (GLM; Searle 1987, McCullaugh and Nelder 1989, SAS Institute 2011) or Generalized Additive Framework (GAM; Hastie and Tibshirani 1990, SAS Institute 2011) might be used in which deviations from the mean trend are modeled by defining various classification variables which are thought to account for the deviations. This general approach has been used in several North Atlantic Fisheries Organization (NAFO) groundfish stock assessments to integrate multiple fishery-independent survey indices of recruitment (e.g., Healey et al., MS 2001 and subsequent Greenland halibut assessments; Stansbury et al., MS 2001 and subsequent Grand Banks cod

assessments).

For this scup assessment, the GLM approach using lognormal error was used to calculate 'integrated' indices of abundance at age for use in model calibration. As noted above, this analytical approach is analogous to a GLM standardization analysis of commercial fishing vessel catch per unit effort data: the 'year' main effect classification variable serves as the index of abundance, while the 'survey' classification variable is analogous to a 'vessel' classification variable, each with its' own time series of catch per unit effort that has some relationship to the underlying true abundance of the stock. The mean index of abundance is modeled as a log-linear function of the classification variables. The analysis could be expanded by including additional classification variables, such as the sampling gear type, tow duration, temporal variables (e.g., day/night) or environmental variables (e.g., water temperature anomalies). However, such details typically are not immediately available for most assessments, as indices are most often presented to the assessment working group process as aggregate annual or seasonal indices at age. As configured here, the analysis provides average, or 'integrated,' aggregate indices of abundance.

SAS software version 9 (SAS 2011) PROC GENMOD was used to develop models of the scup state and academic trawl survey data. The GENMOD procedure fits generalized linear models (GLM) that allow the mean of a population to depend on a linear predictor through a nonlinear link function, and allows the response probability distribution to be specified from a number of probability (error) distributions. These include the normal, lognormal, binomial, Poisson, gamma, negative binomial (negbin), and multinomial distributions (McCullagh and Nelder 1989). The GENMOD procedure fits the models by maximum likelihood estimation. There is generally no closed form solution for the maximum likelihood estimates of the parameters, so the procedure estimates the parameters of the model numerically through an iterative fitting process, with the covariances, standard errors, and p-values computed for the estimated parameters based on the asymptotic normality of maximum likelihood estimators (SAS 2011).

The time series of years for the scup ASAP model is 1963-2014, with fishery catch available for the entire series and fishery age compositions available for 1984 and later. The longest survey series is the University of Rhode Island Graduate School of Oceanography (URIGSO) aggregate index beginning in 1963; the shortest are the Northeast Monitoring and Assessment Program (NEAMAP) spring (2008) and fall (2007) trawl series, which have 'limited' age compositions. The state and academic survey series were grouped into spring and fall seasonal collections to develop seasonal standardized, or 'integrated,' aggregate indices. The spring collection includes the MADMF spring, RIDFW spring, CTDEP spring, and NEAMAP spring trawl survey aggregate numeric indices. The spring collection surveys index age 1 and older abundance. The fall collection includes summer and fall seasonal surveys; the MADMF fall, RIDFW fall, URIGSO, CTDEP fall, NYDEC, NJDFW, ChesMMAP, and NEAMAP fall trawl survey aggregate numeric indices. The fall collection surveys index age 0 and older abundance.

GLM main classification effects were limited to the year of sampling (1982, 1983...2014) and the identity of the survey (MASPR, RIFAL, etc.) The resulting year effect coefficients, corrected for lognormal-transformation bias and re-transformed to the original scale, serve as the seasonal indices of abundance. Models were constructed using lognormal, Poisson, negative binominal, and gamma error distributions with log-links where necessary. The estimates of- and changes in several goodness of fit statistics were used to evaluate the goodness of fit of the

model and the significance of the classification factors: a) the ratio of the deviance (twice the difference between the maximum attainable log likelihood and the log likelihood of the model) to the degrees of freedom (DF) – this statistic is a measure of "dispersion" and of fit of the expected probability distribution to the data (closer to 1 is better), b) the value of the log-likelihood (a measure of model fit), c) the computed AIC (a measure of model fit and performance, valid for a sequence of models within each distribution), d) whether or not the model converged (whether the negative of the Hessian matrix was positive definite, allowing valid estimation of the parameters and their precision), and e) the significance of the classification factors as indicated by the log-likelihood ratio statistics at the 5% level. A Type III analysis was used since it does not depend on the order in which the classification factors (i.e., the survey ID) are specified (SAS 2011). The seasonal 'integrated' aggregate numeric indices were then used as calibration indices and results compared with the existing (2008 model updated through 2014) and preliminary SAW 60 scup model (new surveys with full age 0-7+ compositions) configurations. The GLM seasonal state/academic survey indices of aggregate numeric abundance are shown in Figure A36.

There are insufficient seasonal state/academic survey indices at age to construct integrated indices at age for both seasons for the full range of ages, 0 to 7+. For example, there are only two spring age 2 series (CTDEEP and NEAMAP), and only one spring series each for ages 3, 4, and 5-7+ (from the CTDEEP spring survey). Therefore, standardized integrated indices at age were constructed using indices for both seasons to construct independent annual index series for ages 0, 1, 2, 3, 4, and 5-7+. Main classification effects were limited to the year of sampling (1982, 1983...2004) and the identity of the survey (CTDEEP fall age 0, CTDEEP fall age 1...CTDEEP fall age 5:7+). The resulting year effect coefficients, corrected for lognormal-transformation bias and re-transformed to the original scale, were used as six independent indices of abundance at ages 0, 1, 2, 3, 4, and 5-7+ that were input to the model calibration in place of the original, multiple (28) state/academic survey series at age. Survey selection was set at 1 for each age series. The construction of the six independent, annual 'integrated' indices at age suggested it could be useful to have a corresponding annual 'integrated' aggregate index, analogous to the way the 2008 assessment model was configured; one was constructed using all state/academic spring and fall indices, as in the previous section. The six independent, annual 'integrated' indices at age and the annual 'integrated' aggregate numeric index were then used in sequential fashion as calibration indices in the existing 2008 and preliminary SAW 60 scup model configurations.

A model using only seasonal 'integrated' aggregate indices indicated lower SSB over the last decade, about 40% in 2014, and higher F by 50-100% in 2014, compared to the existing 2008 and preliminary SAW 60 models. The 'integrated' indices model provided more uncertain estimates of 2014 SSB and F than the existing/preliminary models, with comparable precision of recruitment at age 0. A model using an integrated aggregate index for both seasons plus 'integrated' indices at age' for ages 0-2 provided the closest agreement between the existing 2008 and preliminary SAW 60 models. As 'integrated' indices at ages 3 and older were added, the estimates of SSB for 2010 and later years increased above the existing/preliminary models. The SWG viewed this work as a useful 'sensitivity' analysis of the existing and preliminary model configurations.

A6.11.2 Hierarchical Analysis (Conn 2010) Indices of Abundance

The 'hierarchical analysis' approach demonstrated in Conn (2010) was applied to the same collections of scup spring and fall research survey data from state agencies and academic institutions as used in the GLM 'integrated indices' work described earlier. In his paper Conn (2010) concluded "…I have shown how hierarchical analysis can be used to estimate a common population trend from multiple indices. This framework separates components of index variation into process error and sampling error. In this manner, analysts can calculate a single, ''most probable'' index prior to stock assessment analyses. Such an index may be of interest in its own right or may be advantageous in model fitting because it reduces the dimensionality of the likelihood and precludes numerical problems that can arise when fitting data to multiple, conflicting indices. It also has the potential to reduce the number of subjective decisions that are typically made about which indices to include in the analysis."

The result was construction of seasonal time series of relative abundance for use in scup model calibration. No hierarchical indices at age were constructed. The hierarchical seasonal indices of aggregate numeric abundance are shown in Figure A37.

A6.12 Comparative analysis and Conclusion

The 'GLM Integrated' and 'Hierarchical' spring and fall indices, with all 4 series scaled to their respective time series means, are shown in Figure A38. The 'Hierarchical' series are less variable, resulting in a stronger 'smooth' through the state and academic spring and index series. The 'GLM Integrated' and 'Hierarchical' seasonal indices of aggregate abundance were added to the preliminary SAW 60 ASAP model run referenced earlier in the GLM section, to examine the influence of each on the model results and compare to the preliminary SAW 60 'full' model. The SWG viewed this work as a useful 'sensitivity' analysis of the existing and preliminary model configurations, but it has not been carried forward in the assessment.

This work for scup suggests there are 'pros' and 'cons' to the construction of 'integrated' indices and their use in the calibration of population models. 'Pros' include the idea that the standardization procedures serve as objective statistically based 'smoothers' of survey indices with high inter-annual variability and relatively low precision. The resulting indices then serve as temporally and spatially synoptic 'integrated' metrics of aggregate abundance. 'Cons' include the notion that use of 'integrated' indices as calibration data in a model means that much of the characteristic variability of the original survey indices has been 'smoothed out' by the standardization procedure, although there is a trade-off with the decrease in degrees of freedom (fewer 'surveys' used in the calibration). The SWG concluded that the 'hierarchical' approach held more promise for future development, but that considerably more work is needed before these indices could be used in the scup assessment.

A7. TERM OF REFERENCE 3: Describe the thermal habitat and its influence on the distribution and abundance of scup, and attempt to integrate the results into the stock assessment.

A7.1 NEFSC Trawl Survey Environmental Data

Some of the NEFSC winter, spring and fall trawl survey environmental data were summarized for the strata sets used for scup to investigate the correspondence between the environmental factors and the distribution of scup. The environmental factors were surface air temperature in degrees Celsius, surface and bottom water temperature in degrees Celsius, and bottom water salinity in parts per thousand (PPT). Valid surface and bottom temperature data on a per tow basis are generally available for the entire 1968-2014 time series for the scup survey strata (Great South Channel to Cape Hatteras) in both spring and fall, with the exception of fall 2008, for which large numbers of observations are missing. Air temperatures are generally missing during the 1970s and during 2012-2014 in both spring and fall. Bottom salinities are generally available for 1997 and later years, except for fall 2008.

First, the cumulative distributions of the scup survey catches by tow and the environmental factors were compiled for the spring (offshore strata 1-12, 23, 25, 61-76) and fall (offshore strata 1-12, 23, 25, 61-76, inshore strata 1-61) strata sets. For this simple compilation, the cumulative totals over tows are not weighted by stratum area. In the spring survey strata, over the full 1968-2014 time series, scup were in general caught at stations (tow sites) that had a warmer surface temperature (Figure A39; median [50th %ile] catch at 8.5^{o}C, median tows at 6.3^{o}C), a warmer bottom temperature (Figure A40; median [50th %ile] catch at 9.8^{o}C, median tows at 6.8^{o}C), higher bottom salinity (Figure A41; median catch at 34.8 PPT, median tows at 33.6 PPT), and warmer air temperature (Figure A42; median catch at 10.0^{o}C, median tows at 6.0^{o}C) than the median environment of the spring scup strata set. In the fall survey strata, scup were in general caught at stations (tow sites) that had a warmer surface temperature (Figure A43; median catch at 22.1^{o}C, median tows at 19.9^{o}C), a warmer bottom temperature (Figure A44; median catch at 21.0 median tows at 13.4^{o}C), lower bottom salinity (Figure A45; median catch at 31.9 PPT, median tows at 32.5 PPT), and slightly warmer air temperature (Figure A46; median [50th %ile] catch at 19.0^{o}C, median tows at 18.7^{o}C) than the median environment of the fall scup strata set.

In a second compilation, the annual stratified mean values of the environmental factors for positive scup catch tows were compared with the annual stratified mean values of the environmental factors for all tows in the scup strata sets to investigate trends over time. Figure A46 shows that the mean surface temperature on NEFSC spring survey tows with positive scup catch (SCP_surftemp) was generally warmer than the mean surface temperature of all tows (All_surftemp) over the series. The solid trend lines show that the mean surface water temperature of both positive scup tows and all tows in the spring strata set has increased over time. Figure A48 shows the pattern for NEFSC fall survey tows, with the mean surface temperature on tows with positive scup catches generally close to the mean surface temperature of all tows over the series. The solid trend lines show that the mean surface water temperature of positive scup catch tows and all tows in the fall strata set has increased over time.

Figure A49 shows that the mean bottom temperature on NEFSC spring survey tows with positive scup catches (SCP_bottemp) was generally warmer than the mean bottom temperature of all tows (All_bottemp) over the series. The solid trend lines show that the mean bottom water temperature of both positive scup tows and all tows in the spring strata set has slightly increased over time. Figure A50 shows the pattern for NEFSC fall survey tows, with the mean bottom

temperature on tows with positive scup catches generally warmer than the mean bottom temperature of all tows over the series. The solid trend lines show that the mean bottom water temperature of scup tows in the fall strata set has increased more over time than the bottom temperature in all tows.

Figure A51 shows that the mean bottom salinity on NEFSC spring survey tows with positive scup catches (FLK_botsalin) was generally higher than the mean salinity of all tows (All_botsalin) since 1997. The solid blue trend line shows that the mean bottom salinity of all tows in the spring strata set has increased since 1997. Figure A52 shows the pattern for NEFSC fall survey tows, with the bottom salinity on tows with positive scup catches generally lower than the mean salinity of all tows since 1997. The solid trend lines show that the mean salinity of all tows in the fall strata set has a similar trend as the spring.

Figure A53 shows the mean air temperature on NEFSC spring survey tows with positive scup catches (FLK_airtemp) was slightly higher than the mean air temperature of all tows (All_airtemp) over the series. The solid trend lines show that the mean air temperature of all tows in the spring strata set has decreased over time. Figure A54 shows the pattern for NEFSC fall survey tows, with the air temperature on tows with positive scup catches generally comparable to the mean air temperature of all tows. The solid red trend line shows that the air temperature of all tows in the fall strata set has increased over the series.

As noted in the NEFSC surveys section under TOR 2, examination of patterns in the survey catch, for spring and fall and day and night, confirms the irregular distributions of catch by temperature, salinity and depth and portend the difficulties of modeling the survey scup catch data. No well defined relationships are evident; i.e., small catches are as likely to be taken at shallow depths as large depths and at both warm and cold temperatures and large catches can occur over a relatively large range of depth and temperature (e.g, over a range of 70 meters or 10 degrees). Therefore, generalized linear model (GENMOD) and generalized additive model (GAM) based indices of abundance for the scup NEFSC seasonal survey data proved to be not useful, due to highly variable results owing from the inability of the models to adequately fit the variable and complex temporal and spatial properties of scup survey catches.

The NEFSC survey indices sometimes appear to mainly reflect the availability of scup to the survey, rather than true abundance, making it difficult to interpret large inter-annual changes in the indices. As noted in the description of the NEFSC trawl survey indices above, the spring 2002 and 2014 indices were unexpectedly much higher than adjacent indices (Figure A14), across all ages. In 2002, this 'availability event' appears to have been a response to higher than normal spring water temperatures, as large scup survey catches and bottom water with temperatures higher than 10°C were distributed further inshore on the shelf than usual. Figures A55-A57 show the distribution of scup catches and temperatures during 2001-2003. In more recent years, the bottom temperature pattern in 2011 and 2013 was more 'normal' and large scup catches were restricted to the shelf edge (Figures A58 & A60). The bottom temperature in 2012 was similar to that in 2002, and scup catches were distributed across the shelf (Figure A59), resulting in a high biomass and abundance indices, although not as extreme as in 2002. Near 'normal' bottom conditions were present in 2014 (Figure A61), but catches of large scup occurred near mid-shelf in large-area strata, and the 2014 indices (especially in biomass per tow) were among the largest of the spring time series. These sequences of potential 'availability events' make clear the difficulty that is encountered when interpreting survey indices for scup – do high survey indices indicate high availability, high abundance, or (more likely) some combination of both? This issue has lead NEFSC investigators to pursue the work described in

the next section.

A7.2 Modeling annually varying suitable thermal habitat

The working paper of Manderson et al. (MS 2015; Working Paper A11) describes the development of estimates of proportions of 'thermal habitat suitability' for scup (Figure A62) surveyed in the NEFSC and NEAMAP surveys that could be used to account for errors in survey observations related to temperature dependent changes in geographic distribution and seasonal migration. The working paper described the development and evaluation of time series of varying estimates of the proportion of thermal habitat suitability for scup surveyed on the Northeast US shelf by the NEFSC and NEAMAP bottom trawl surveys from 1975-2012 in a manner that accounted for thermal habitat occurring outside the surveys and the relative motions of habitat and the survey vessel. The working paper estimated that an average of ~63 % of the thermal habitat suitability available to scup within the model domain (Cape Hatteras to Nova Scotia) was sampled from 1973-2012 by the fall NEFSC bottom trawl survey, while ~50% was sampled in the spring. In the 2008-2012 NEAMAP surveys approximately 14% of available thermal habitat suitability on the Northeast US continental shelf was sampled during the fall, while 11% was sampled in the spring. Yearly estimates of the proportion of thermal habitat suitability surveyed did not exhibit systematic trends (Figures A63-A65).

A8. TERM OF REFERENCE 4: Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results and previous projections.

A8.1 Instantaneous Natural Mortality Rate (M)

The instantaneous natural mortality rate (M) for scup has been assumed to be 0.20 (Crecco *et al.* 1981, Simpson *et al.* 1990) in all previous stock assessments. Longevity based estimators of M are sensitive to critical underlying assumptions which include the value of p, or the small proportion of the population surviving to a given maximum age (tmax), and the maximum observed age under no exploitation conditions. Using a maximum age of 15 years for scup, the 'Rule of Thumb' method of 3/tmax noted in Quinn and Deriso (1999) and the methods of Hoenig (1983) and Hewitt and Hoenig (2005), longevity based estimates of M for combined sexes range from 0.20 to 0.28. Age-specific and size variable estimates of M, based on the work Lorenzen (1996, 2000) and Gislason et al. (2010) range from 0.18 to 1.72, with the highest values associated with age 0 fish (fish at smallest lengths and weights).

Then et al. (2014) recently conducted a review of the performance of the best known empirical estimators of natural mortality. Then et al. (2014) recommended use of the updated Hoenig (1983) estimator when an estimator of tmax is available, or the updated Pauly estimator when a reliable estimate of tmax is not available. For a scup tmax of 15 years, the updated Hoenig method provides an estimate of 0.41, and for Linf = 51.6 cm and K = 0.16, the updated Pauly method provides an estimate of 0.30.

Alternative estimates of M for scup are presented in the table below. Given the historical maximum size and age of 41 cm and 15 years, recent observations of large fish (45 cm) up to age 12, the range of M (0.1 – 0.6) estimated by the empirical methods based on maximum age, and the likelihood profile of a preliminary assessment model run that indicated a best fit at M = 0.10 and of the final model at 0.15, the SWG decided there was no compelling reason to change from the previous assumption for M, and adopted a value of M = 0.20 for all ages and years in the 2015 SAW 60 assessment models.

Age	3/tmax Rule of Thumb	Hoenig (1983), Hewitt and Hoenig (2005)	Gislason et al (2010)	Lorenzen (1996, 2000)	Lorenzen Scaled to Rule of Thumb	Lorenzen Scaled to Hewitt & Hoenig	Then et al. (2014): Pauly	Then et al. (2014): Hoenig
0	0.20	0.28	1.72	1.38	0.82	0.68	0.30	0.41
1	0.20	0.28	0.96	1.03	0.61	0.51	0.30	0.41
2	0.20	0.28	0.59	0.77	0.46	0.38	0.30	0.41
3	0.20	0.28	0.44	0.65	0.38	0.32	0.30	0.41
4	0.20	0.28	0.36	0.57	0.34	0.28	0.30	0.41
5	0.20	0.28	0.31	0.53	0.32	0.26	0.30	0.41
6	0.20	0.28	0.28	0.50	0.30	0.24	0.30	0.41
7	0.20	0.28	0.27	0.48	0.28	0.23	0.30	0.41
8	0.20	0.28	0.25	0.46	0.27	0.23	0.30	0.41
9	0.20	0.28	0.24	0.42	0.25	0.20	0.30	0.41
10	0.20	0.28	0.20	0.42	0.25	0.20	0.30	0.41
11	0.20	0.28	0.21	0.40	0.24	0.20	0.30	0.41
12	0.20	0.28	0.20	0.40	0.24	0.19	0.30	0.41
13	0.20	0.28	0.19	0.39	0.23	0.19	0.30	0.41
14	0.20	0.28	0.19	0.38	0.23	0.19	0.30	0.41
15	0.20	0.28	0.18	0.38	0.22	0.18	0.30	0.41
Mean	0.20	0.28	0.34	0.57	0.34	0.28	0.30	0.41

A8.2 2015 SAW 60 Model Building

A8.2.1 Existing 2008 Assessment Model Updated through 2012

The most recent benchmark peer review of the scup assessment was conducted by the 2008 Northeast Data Poor Stocks Working Group (DPSWG) panel (NEFSC 2009), which accepted an Age Structured Assessment Program (ASAP) statistical catch at age model (Legault and Restrepo 1998, NFT 2008) with fishery and survey catch data through 2007 as the basis for status determination. The assessment indicated that the stock was not overfished and overfishing was not occurring in 2007 relative to the corresponding biological reference points. There was no consistent retrospective pattern in F, SSB, or recruitment evident in the assessment model.

ASAP is an age-structured model that uses forward computations assuming the separability of fishing mortality into year and age components to estimate population sizes given observed catches, catch-at-age, and aggregate and at-age indices of abundance. The separability assumption is partially relaxed by allowing for fleet-specific computations and by allowing the selectivity-at-age to change in blocks of time. Weights (emphasis factors) are input for different components of the objective function which allows for configurations ranging from relatively simple age-structured production models to fully parameterized statistical catch-at-age models.

The objective function is the sum of the negative log-likelihood of the fit to estimable model components. Catch at age and survey at age compositions are generally modeled assuming a multinomial distribution, while most other model components are assumed to have lognormal error. Specifically, lognormal error distributions were assumed for the total catch in weight, research survey catch at age calibration indices, selectivity parameters, annual fishing mortality parameters, survey catchability parameters, estimated stock numbers at age, and Beverton-Holt stock-recruitment parameters, when estimated. Recruitment deviations are also assumed to follow a lognormal distribution, with annual deviations estimated as a bounded vector to force them to sum to zero, thus centering the predictions on the expected stock-recruitment relationship. In the 2008 assessment ASAP model an instantaneous natural mortality rate of M = 0.2 was assumed for all ages and years. Additional initial model settings included specification of the likelihood component emphasis factors (weights or Lambdas, L), the size of deviation factors expressed as standard deviations (i.e., ln-scale CV), and the penalty functions for extreme fishing mortality estimates. These were set at consensus values by the 2008 DPSWG (NEFSC 2009) after multiple sensitivity runs to evaluate a range of inputs.

The 2008 ASAP model built on earlier Virtual Population Analysis (VPA) models for scup (NEFSC 1998), and the 2008 scup assessment was one of the first uses of the ASAP model in Greater Atlantic Region stock assessments. As such, the survey indices at age were configured as in the earlier VPA model, with indices input to the model as individual time series (e.g., NEFSC fall survey Age 0, 1984-2007; CTDEEP spring survey age 6, 1984-2007; VIMS age 0, 1987-2007). During the model building process for the 2008 assessment, additional aggregate survey biomass series were added to the model to provide more and longer time series of survey data and explicitly model aggregate population trends (e.g., NEFSC winter, spring and fall biomass series, MADMF spring and fall biomass series, RIDFW spring and fall biomass series, and NJ biomass and URIGSO aggregate numeric series). The addition of the long-term aggregate series helped stabilize the model estimates and ensured consistent convergence. Winter, spring, and mid-year survey indices and all survey recruitment (age-0) indices were calibrated to population numbers of the same age at the beginning of the same year. Fall survey indices were calibrated to population numbers one year older at the beginning of the next year. Lognormal error distributions were assumed for the survey catch at age calibration indices. This survey index configuration was retained in the 2008 and subsequent assessment updates.

Four fishery fleets were modeled in aggregate (metric tons; Tables A22 & A27) and at-age (in thousands of fish at ages 0-7+): commercial landings (Table A13), commercial discards with mortality rate of 100% (Table A16), recreational landings (Table A18), and recreational discards with mortality rate of 15% (Table A20). In ASAP, a single catch numbers-weighted mean weight at age matrix (Table A24) serves as the basis for mid-year catch and extrapolated (Rivard method) SSB mean weights at age. Fleet CVs were set at 0.10, 0.32, 0.10, and 0.12 and Fleet Effective Sample Sizes (ESS) were set at 22, 9, 31, and 4. Fishery selectivity (S) was modeled as 'at-age' selectivity (estimate individual S at age) by fleet and time block. Two time blocks were set: 1963-1996, before the implementation of quotas, and 1997 and later, after implementation. Commercial and recreational landings S was set fixed at 1 for (true) age 4 for both time blocks with L = 1 and CV = 0.1. Commercial discards S was set fixed at 1 for (true) age 2 and recreational discards S was set fixed at 1 for (true) age 1 for both time blocks with L = 1 and CV = 0.1. Survey selectivity (S) was set fixed at 1 for each individual index at age.

Other 2008 assessment model settings included: total fishery catch weight lambda (L) = 1; fishing mortality (F) and stock size (N) in year 1 L = 1 and CV = 0.9; recruitment deviations L =

1, with CV = 0.1 during 1963-1983, and CV = 1.0 after 1983; S-R function and population scaler Ls = 1 with CV = 0.9, effectively 'turning on' the influence of the S-R function in the model and giving particular influence in years 1963-1983 before any fishery or survey age data were available; and survey catchability coefficients (q) estimated as a constant value (no deviations) with L = 1 and CV = 0.9.

Following the 2008 assessment, the NMFS declared scup to be officially rebuilt in 2009. The assessment was updated with new data under the same 2008 model configuration for 2009-2012. The 2012 update again found the stock was not overfished and that overfishing was not occurring in 2011 relative to the 2008 biological reference points (Terceiro 2012).

A8.2.2 Existing 2008 Assessment Model Updated through 2014

Model IAA-IND08 is the first of the 2015 SAW 60 models, with the same configuration and settings as the 2008-2012 models but with data updated through 2014. Surveys are configured as independent indices at age (IAA), the index set included in the model is the same as in the 2008-2012 models (IND08), and fishery and survey selection is modeled as 'at-age.' Model IAA-IND08 provides estimates appropriate to compare with the existing reference points, which are FMSY proxy = F40% = 0.177 and SSBMSY proxy = SSBMSY40% = 92,044 mt (TOR 6a). This model indicates that F in 2014 = 0.047 and SSB in 2014 = 232,673 mt, so the stock was not overfished and overfishing was not occurring (see TOR 6a). Summary results for 1984 and later years (the period when fishery age data are available and recruitment deviations can be estimated from fishery and survey catch at age) from the 2008 and 2012 assessments are compared with those from run IAA-IND08 in Figures A66-A68.

A8.2.3 2015 SAW 60 Assessment Model Updated through 2014

The subsequent model building occurred in three 'phases.' In phase 1, structural changes were made to the survey configurations (from individual indices-at-age modeled with lognormal error to catch-at-age matrices modeled with multinomial error, with full age compositions), several new survey series with full age compositions were added to the model, and new (revised) maturity and commercial discard estimates were added to the model. The end product of phase 1 was the BASE run with the most complete input data set to move forward.

In phase 2, the BASE run was tested to determine the likelihood components that are reliably estimable (e.g., starting N and F, fishery and survey selectivity, recruitment estimation, survey catchability, time series of F and N, etc.) , evaluate their statistical diagnostics (convergence, residuals, Root Mean Square Error [RMSE], etc.), and determine their influence on model results. Phase 2 determined the 'best' general model configuration to move forward.

In phase 3, the 'best' BASE run was 'tuned' by iterating survey CVs to allow RMSEs to approach the confidence intervals associated with a N(0,1) distribution (i.e., for a normal random variate) for that sample size, and by adjusting fishery and survey age composition ESS to near the time series means while accounting for 'outliers.' Subsequent 'final run' diagnostics included retrospective analyses, likelihood profiling over the assumptions for M and SSB0, sensitivity to the configuration of the NEFSC spring and fall survey series, and sensitivity to the length of the

modeled time series.

A8.2.4 Model Building Phase 1

The 2015 SARC 60 model building process started with the 2012 updated assessment model run with data through 2011 (Terceiro 2012). The 2012 model differed from the previous 2008 DPSWG benchmark assessment ASAP model (NEFSC 2009) only in minor changes to the values of the fleet Effective Sample Sizes (ESS). As noted above, the 2012 model has been updated with fishery and survey data through 2014 to create model IAA-IND08, with results compared to the existing 2008 reference points, in response to TOR 6A.

Since the 2008 assessment, the survey index configuration widely accepted as 'standard' in the ASAP model has evolved. In general, survey indices at age are now input as a 'catch-at-age' matrix modeled with multinomial error to calibrate population proportions at age, along with a corresponding aggregate numeric or biomass index modeled with lognormal error to calibrate aggregate population trends. Stand-alone recruitment indices can continue to be modeled as single-age indices, as can aggregate numeric biomass or numeric survey series for which no associated age composition data are available. Each model configuration change (step) in phase 1 generally builds on the previous step, unless noted. The model was first transitioned to the now 'standard' ASAP model survey index configuration using the same suite of indices as in 2008 and 2012 and given the name MULTI_IND08.

In the next step, new surveys and new ages [i.e., full age range] from previous surveys are added to the model, creating model NEWSVS. 'Full-catch-number-at-age' survey indices are available for the NEFSC spring, fall, and winter (ages 0-7+; Tables A31-A32, A34) and CTDEEP spring and fall (ages 0-7+; Tables A42-A43). 'Limited-catch-number-at-age' surveys are available for the NYDEC (ages 0-2; Table A44) and VIMS ChesMMAP (ages 0-1; Table A46). Aggregate numeric indices (no age compositions) are available for the MADMF spring and fall (Table A36), URIGSO (Table A41) and NJDFW surveys (Table A45) . The VIMS index of age 0 abundance is input as a stand-alone numeric index at age (Table A45). New 'Full-catch-number-at-age' survey indices from the RIDFW Industry Cooperative Trap Survey (ages 0-7+; Table A40) and 'Limited-catch-number-at-age' indices the NEAMAP spring and fall surveys (ages 0-2; Table A48) are also added. Late in the assessment process, too late to be added to the NEWSVS configuration, 'full-catch-number-at-age' survey indices became available for the RIDFW spring and fall surveys (Tables A38-A39). These new RIDFW indices replaced the previous aggregate indices (Table A37) and were evaluated in a later, phase 3 run. Finally, the fishery fleet ESS values were 'rounded' from [22, 9, 29, 4] to [30, 10, 30, 5] to provide a new ESS starting point given the addition of new ages for previous surveys and survey data series (it was noted that the estimated ESS values were starting to drift away in both directions from the initial 2012 assessment values).

The next step was to revise the commercial fishery discard estimates as described above in the COMMERCIAL FISHERY DISCARDS section, creating model NEWDISC. The final step in phase 1 was to adopt the revised maturity schedule using the 3 year moving window estimates as described above in the MATURITY section, creating model NEWMAT. Results from models the 2008 DPSWG, 2012 Update, and 2015 SAW 60 IAA_IND08 through NEWMAT are summarized in Tables A49-A50 and Figures A69-A71. Table A49 provides a summary of the initial steps in building the model configuration and settings, while Table A50 provides summary

results. Important changes in settings and estimates between modeling steps are highlighted with bold text. The largest changes occurred due to the use of the new survey configuration (MULTI_IND08) and the revision in commercial discards (NEWDISC). Retrospective analysis conducted for run NEWMAT found no pattern of large (i.e., > 30%) relative errors in SSB or F, which were both < 10%, with about +16% for age 0 (model age 1) recruitment.

A8.2.5 Model Building Phase 2

As in phase 1, each change in phase 2 generally builds on the previous step, unless noted. Model configuration NEWMAT was renamed S60_BASE_1 to begin phase 2. In addition to acceptance of survey indices at age input as a 'catch-at-age' matrix modeled with multinomial error as the standard ASAP configuration, a number of other settings have also became accepted as 'standard', mainly in the interest of allowing the input data to most strongly influence the model results and of reducing the influence of prior (initial) values, in the following general order:

1) Test the model sensitivity to the initial values of N in year 1 to minimize residuals and stabilize starting conditions, Ls set to 0 if possible

2) Test the model sensitivity to the initial values of F in year 1 (to minimize residuals and stabilize starting conditions) and F deviations in subsequent years; Ls set to 0 if possible

3) Ls for fishery and survey selectivity, Ls set to 0 if possible

4) If the internal S-R function will not be used for BRPs (e.g., if h ~1), 'turn off' S-R function (Ls set to 0)

5) Test the model for sensitivity to recruitment deviation priors, L set to 0 if possible

6) Test the model for sensitivity to use of likelihood constants, 'turn off' if possible

The first change was to iterate the initial guesses for N in year 1 from the very large values with exponential decline used in the 2008 assessment to values closer to the predicted 2008 values with simple deviations, creating run S60_BASE_2. This run provided results very close to S60_BASE_1.

The next change in phase 2 was to remove the prior (L=1 to L=0) for N in year 1 of the model, removing these parameters from the objective function. This run did not converge (no estimates), so the L was reset to 1, and the run continued to be called S60_BASE_2.

The next change in phase 2 was to remove the prior (L=1 to L=0) for F in year 1 of the model and for F deviations in subsequent years, removing these parameters from the objective function. The model performed somewhat better (more feasible F in year 1 estimate) when the L=1 for F in year 1 was retained, creating run S60_BASE_3. The changes from S60_BASE_1 to S60_BASE_3 resulting in only minor changes in the estimates of SSB, R, and F since 1984 (the first year in the model with both fishery and survey ages).

The next change was to remove the priors for fishery selectivities (L=1 to L=0), creating run

S60_BASE_4. Removing the constraint of the priors allowed the fishery landed catch selectivity patterns to become more domed, while the fishery discarded catch selectivity patterns became less domed. The landed catch dome in particular became extreme, to less than 10% selection for the plus group age in the second time block, which is likely not feasible. The overall effect on the general magnitude of SSB, R, and average F for adult fish (true ages 2 and older; model ages 3 and older) was relatively minor, however, for most of the time series.

The next change was to restore the priors for catch selectivities (L=0 to L=1) but increase the CV from 0.1 to 0.5, allowing moderate constraint, and creating run S60_BASE_5. This change provided intermediate results between runs 3 and 4, and was carried forward.

The next change was to remove the priors for survey selectivities (L=1 to L=0) for surveys with age compositions, creating run S60_BASE_6. Removing the constraint of the priors on survey selectivities allowed most of the selectivities to be estimated lower for ages 2 and older and to approach zero for ages 5 and older. This change had a relatively large effect. The overall effect on the general magnitude of R and SSB was an increase in recruitment during the 2000s and a stronger increase in SSB since 2000 which resulted in about a 20% increase in terminal year SSB compared to run S60_BASE_5 (Figures A72-A74). Some of the older age selectivities were imprecisely estimated or hit a boundary constraint. However, the run S60_BASE_6 survey selectivity settings were left as is until later in phase 2, where they would be re-examined.

Calculation of the S-R function parameters in runs 1-6 resulted in 'steepness' estimates ranging from 0.95 to 0.97, i.e., very close to 1.00. The next change was to change the Ls from 1 to 0 for 'Initial Steepness,' effectively 'turning off' the influence of the S-R function in the model, and thus relying only on the fishery and survey indices to estimate recruitment, constrained by L = 1 and CV = 0.1 during 1963-1983, increasing to CV = 1.0 during 1984-2014 for the annual recruitment deviations. These changes created run S60_BASE_7. 'Turning off' the S-R function mainly affected model estimates before 1984, which translated into about 10% lower F during the mid-1990s, but only very small changes in F or SSB since 2000 compared to run S60_BASE_6.

The next change was to remove the constraints on recruitment deviations, by changing L = 1 to L = 0, creating run S60_BASE_8. This resulted in an extremely variable pattern in estimated stock sizes at age in the years before 1984 (e.g., annual recruitment ranging from near 0 to about the post-1983 maximum of about 200 million), and infeasible estimates of F during the 1960s-1970s ranging to near the constraint of F = 5.0. With no apparent benefit to removing the recruitment deviations constraint that holds them near the mean for years before 1984, it was re-implemented by changing back to L = 1, and the S60_BASE_7 configuration was retained for moving forward.

The next change was to 'turn off' the 'likelihood constants' in the model, creating run S60_BASE_9. This change affects the way recruitment deviations are estimated in ASAP3. Ongoing ASAP model development work demonstrates that holding the value of the term constant can, in some cases, lead to underestimates of recruitment because the objective function can be reduced by lowering the estimated recruitment values, since one of the components sometimes is in fact not constant, with the degree of variation depending on the specific model configuration. For run 9, 'turning off' the likelihood constants resulted in a nearly uniform time series increase in recruitment of about 9% over the time series compared to run 7. One estimation difficulty re-emerged, however, as the run 9 model provided infeasible estimates of F during the 1960s-1970s ranging to near F = 3.0, due to the estimation of some transient but very large stock sizes at fully recruited ages early in the time series, similar to the DPSWG2008

assessment model and some of the earlier 2015 configurations. These 'odd' estimates do not generally persist for long, passing out of the population in 3-4 years, and so do not affect the population dynamics over the last 30 years when age compositions are available. 'Turning off' the 'likelihood constants' is now considered to be the preferred configuration for ASAP, so this change was retained in subsequent steps.

Some patterning in the fishery age composition residuals from the mid-2000s and later years had persisted through all the early S60_BASE run configurations. Run S60_BASE_10 built upon run 9, adding a third fishery selection block for 2006 and later years, with the fishery selection Ls = 1 and S = 1 for (true) age 4 for the landings and (true) age 2 for discards. This change slightly improved the fishery age composition residual magnitude and pattern, and the third selection block was retained.

Before moving to model 'tuning' in phase 3, a more detailed examination of diagnostics for run 10 was made, including those for fishery and survey selectivity parameter estimates, patterns in aggregate survey index residuals, and patterns in fishery and survey age composition residuals. Inspection of the estimated parameters of run S60_BASE_10 revealed that several of the fishery and survey selection parameters at age were poorly estimated (either constrained at a bound or with large standard error; although note that the survey selectivities are not part of the objective function as L = 0). In run S60_BASE_11, bounded fishery selection parameters at 1 were fixed at S = 1, generally true ages 4 or 5 adjacent to the S= 1 fixed at true age 3. Estimates from run S60_BASE_11 were nearly identical to those from run 10. Next, poorly estimated survey selection parameters at age (CV equal to or greater than 1.0), typically for the youngest or oldest ages, were fixed near the value of the nearest acceptably estimated age, resulting in run S60_BASE_12. Again, these change had little effect, and the results of S60_BASE_12 were nearly identical to those from run 11.

In summary, the largest changes in estimates over steps 1-12 of the BASE model were due to 1) changing the fishery selectivity prior CVs from 0.1 to 0.5 in run 5, 2) changing the survey selectivity Ls from 1 to 0 in run 6, 3) 'turning off' the recruitment likelihood constants in run 9, and 4) adding a third (2006 and later) fishery selectivity block in run 10. Except for the transient, starting condition-related extreme F early in the time series, the estimates change very little from run S60_BASE_9 through 12 (Tables A51-A52, Figures A75-A77).

A8.2.6 Model Building Phase 3

In phase 3, the following changes to the model configuration were made:

1) Iterate survey CVs to allow Root Mean Square Errors (RMSE) to approach the confidence intervals associated with a N(0,1) distribution for that sample size (i.e., +/- 2 se; see the 'normal random variate' diagnostic plot). For example, if RMSE is 'too low,' the CV can be reduced, while if the RMSE is 'too high,' the CV can be increased

2) Calibrate fleet ESSs to about the time series mean, one time, rather than Francis (2011) adjustment

3) Calibrate survey ESSs to about the time series mean, one time, rather than Francis (2011) adjustment

The first model 'tuning' step was undertaken in run S60_BASE_13. The input aggregate survey CVs, generally the means of the calculated time series averages, are intended to characterize the sampling error of those series. However, it is recognized that additional process (model) error may be present in the survey indices that are not reflected in the calculated CVs, as diagnosed by the distance of the Root Mean Square Error (RMSE) of each series from 1. Examination of the model diagnostics for the survey indices resulted in adjustments to the survey CVs, thereby allowing for larger deviations to bring their respective RMSEs within or close (sometimes) to the expected confidence intervals (CI) for the number of observations.

Most of the surveys included in the scup model have calculated CVs in the range of 0.2 to 0.9. Based on previous experience with winter (NEFSC 2011b) and summer (NEFSC 2013b) flounder assessment models in ASAP, the input CVs were initially set in the range of 0.5 to 0.6 to account for additional process error. Iterating survey SVs to reduce the RMSEs brought most of them to 0.8-0.9, but in some cases even a high CV of 1.2 still resulted in RMSE outside the N(0,1) confidence interval (RIDFW spring, MADMF spring, NEFSC spring, Figure A78). The next step might be to consider omission of some of those survey series from the model calibration. The input CVs and RMSEs for run S60_BASE_13 were as follows:

Index	Name	Initial CV	Adjusted CV	Run 13 RMSE
1	NECWIN	0.6	0.8	1.2
2	NECSPR	0.6	1.0	1.5
3	NECFAL	0.6	0.6	0.9
4	CTSPR	0.5	0.9	1.3
5	CTFAL	0.5	0.8	1.2
6	NYDEC	0.6	1.2	1.4
7	MASPRKG	0.5	1.2	1.4
8	MAFALKG	0.5	0.5	1.1
9	RISPRKG	0.5	1.2	1.6
10	RIFALKG	0.5	0.8	1.1
11	NJKG	0.5	0.8	1.3
12	URIGSO	0.5	0.7	1.2
13	ChesMMAP	0.6	1.0	1.4
14	VIMSYOY	0.6	1.2	1.2
15	NEAMAP SPR	0.5	0.7	1.3
16	NEAMAP FAL	0.5	0.5	1.2
17	RI Coop Trap	0.5	0.5	0.6
	Total			1.3

These adjustments in survey CVs resulted in lower recent stock sizes and higher recent F relative to the S60_ BASE_12 run (Figures A79-A81). The 'odd' large older age stock size estimates and corresponding unfeasible F estimates early in the time series were reduced. The larger survey CVs also resulted in more large residuals in the last 10-15 years of the model for

the CTDEEP spring, NYDEC, RIDFW spring and fall, and URIGSO indices.

The next change was to 'tune' the 4 fishery fleet age composition ESSs to about their time series means, roughly 'centering' them in the time series pattern. The ESSs were adjusted from the initial run 1 values of [30, 10, 30, 5] to [50, 20, 50, 5]. These 'centered' ESSs for three of the fleets were fairly close to the calculated Francis (2011) ESS values for this run (50 to 69, 50 to 46, 5 to 5), but diverged from the Francis values for the commercial discard fleet (20 to 4). These changes provided run S60_BASE_14. The estimates for run 14 were very similar to those from run 13.

The final changes was to 'tune' the 10 survey age composition ESSs to about their time series means, roughly 'centering' them in the time series pattern. These 'centered' ESSs all were significantly higher than the calculated Francis values. These changes provided run S60_BASE_15; the estimates for run 15 were very similar to those from runs 13 and 14. Tables A53-A54 summarize the changes due to the phase 3 model building steps through run S60_BASE_15. Figures A82-A84 summarize the changes in model estimates from the 2008 model updated through 2014 (IAA_IND08) to the initial 2015 BASE run (S60_BASE_1) through the phase 3 'tuning' steps (S60_BASE_15).

A8.2.7 Sensitivity to NEFSC trawl survey time series configuration

All the runs configured through S60_BASE_15 used continuous NEFSC trawl survey time series, with the years sampled by the FSV Albatross IV (ALB) and FSV Henry B Bigelow (BIG) joined by the use of length-based calibration factors. While the factors at length are constant over time, the 'effective' factors vary over time due to the inter-annual changes in the survey distribution at length. A sensitivity run of S60_BASE_15 was constructed by ending the ALB series in 2008 and adding two additional survey series for the BIG from spring 2009 onward (run S60_BASE_15_BIG).

The aggregate N q for the NEFSC spring survey ALB indices = 7.87e-5; the BIG spring indices q= 1.89e-4. The BIG spring aggregate N q is 2.40 times the ALB spring q. The spring effective calibration factor over all lengths has ranged from 0.89 to 2.36, averaging 1.59 (Table A29). The aggregate N q for the NEFSC fall survey ALB indices = 7.78e-4; the BIG fall indices q= 1.29e-3. The BIG fall aggregate N q is 1.66 times the ALB fall q. The fall effective calibration factor over all lengths has ranged from 2.08 to 4.33, averaging 3.05 (Table A30). Summary estimation results for the S60_BASE_15 and S60_BASE_15_BIG runs are presented in Figures A85-A87. The SWG concluded that the differences are minor, indicating that the NEFSC survey calibration factors are not a major source of uncertainty in the S60_BASE_15 model, and retained the NEFSC ALB-equivalent indices in subsequent runs.

A8.2.8 Sensitivity to Model Time Series Length

The 2008 DPSWG assessment (NEFSC 2009) adopted a model with a time series beginning in 1963, in spite of the need to extrapolate estimates of commercial fishery discards prior to 1989 and recreational fishery catches prior to 1981, in order to include the large catches of the early 1960s and peaks in survey indices in the late 1970s. Model configuration S60_BASE_15 (starting in 1963) was run with alternative time series lengths to evaluate the sensitivity of results

to the model time series length. Three alternatives were considered 1) start in 1977, the year with the earliest available age data (NEFSC spring), 2) start in 1984, when the fishery catch at age starts, and 3) start in 1989, when the Observer commercial fishery data start, and therefore none of the catch estimates rely on extrapolation from ratios.

All three alternative time series length models converged successfully. The SSB, R, and F estimates for the 1963, 1997, and 1984 time series are very similar. The 1989 model series has the fishery and several survey age composition series considerably shortened, which results in lower estimates of stock size (e.g., about 15% lower average recruitment than the 1963 run since 1989) and translates to lower SSB (25% lower average than the 1963 run since 1989) and slightly higher F (5% higher average than the 1963 run since 1989). Figures A88-A90 compare the S60_BASE_15_1963 summary results with the three alternatives.

Seven year retrospective 'peels' were run for the three alternative models and compared with the S60_BASE_15 run. The Mohn's rho (Mohn 1999, Legault at al. 2009) values expressed as average percent error are compared below. As the modeled time series is shortened, the retrospective error generally increases, although the differences are not large.

		Mohn's rho	
Run ID	SSB	R	F
S60_BASE_15_1963	-5%	-45%	-2%
S60_BASE_15_1977	-5%	-45%	-3%
S60_BASE_15_1984	-8%	-48%	+1%
S60_BASE_15_1989	-11%	-52%	-5%

An initial 1963 run with Monte Carlo Markov Chain (MCMC) estimates of uncertainty indicated some diagnostic problems. One thousand iterations with a thinning rate of 1,000 (one million total iterations of which 1,000 are saved) were conducted for one chain (random number seed). Ideally, the 'trace' of the MCMC chain should not show any trending or patterning, and the correlation between successive values in the chain should be low (e.g., less than 0.1 after year 0).

For the 1963 run, however, uneven patterning was evident in SSB and F estimates, especially for the 1963 estimates (Figure A91-A92). There was also evidence of high correlation between successive estimates of the chain for several years (lags; Figures A93-A94). These diagnostics indicate a fairly high level of uncertainty of the model estimates, especially at the beginning of the series. The 'transient' high stock sizes in the initial years of the model and associated very high Fs are a symptom of these issues (e.g., see models S60_BASE_9 and subsequent). The autocorrelation is also reflective of the near-constant recruitment assumed for the years before 1984 when no fishery age data are available (tightly constrained [CV=0.1] recruitment deviations and stock-recruitment scaler with fixed h = 1, by definition resulting in autocorrelated recruitment during this early period). The autocorrelation may also reflect the sequence of consecutive very strong (>25% above the time series average) year classes estimated for 1999-2001 and 2005-2008 that are reflective of the fishery and survey catches. The degree of uncertainty results in the 1963 point estimates for SSB and F not being 'centered' in the distribution of 1963 MCMC estimates (Figures A95-A96).

Given these issues with the early year estimates, the MCMC distributions for runs starting in 1977, 1984, and 1989 were examined for the same number of total and saved iterations. For the 1977 run there was less patterning evident in the SSB and F estimates than in the 1963 run,

although the pattern was still 'noisy' (Figures A97-A98). There was also still evidence of high correlation between successive estimates of the chain for several years (Figures A99-A100), although it is reduced compared to the 1963 run. The point estimates for SSB and F from the 1977 run are better 'centered' in the distribution of MCMC estimates than those from the 1963 run (Figures A101-A102).

For the 1984 and 1989 runs there was minor patterning evident in the SSB and F estimates, although the variability pattern was still 'noisy'. There was also still evidence of high correlation between successive estimates of the chain for 1-2 year lags. The point estimates for SSB and F from the 1984 and 1989 runs are further from the MCMC distribution mode for 2014 SSB than the 1997 run point estimate, as terminal year precision slightly decreases with the shorter series. The precision of the 2014 SSB and F estimates for the four different time series length runs are compared in the table below. The SWG concluded that using the full time series model starting in 1963, given an understanding of why the autocorrelation coefficients are high, caused no major technical issues in the S60_BASE_15 run that would hinder the evaluation of the status of the stock from terminal year results of the model, and retained the full time series in subsequent model development.

Run ID	MCMC CV% SSB 2014	MCMC CV% F2014
S60_BASE_15_1963	10.8	14.4
S60_BASE_15_1977	9.7	13.7
S60_BASE_15_1984	11.1	14.5
S60_BASE_15_1989	12.6	15.5

A8.2.9 Post run S60_BASE_15 revisions made in the SWG meeting

As noted earlier, the RIDFW supplied new spring and fall trawl survey aggregate numeric and indices-at-age, replacing the aggregate biomass indices used previously. The inclusion of the new RIDFW indices created run S60_BASE_16. Run 16 provided estimates of SSB and R slightly higher and F slightly lower in the terminal year compared to run 15 (Table A54).

Revisions to the 2014 NEFSC commercial ages were also made. The latest available 2014 fishery catch and age data were included in the model to create run S60_BASE_17. Run 17 provided estimates of SSB (-7%) and R (-1%) slightly lower and F slightly higher (+3%) in the terminal year compared to run 16 (Table A54).

The effect of several configuration changes to run 17 was examined. As noted in the description of run S60_BASE_13, iterating survey SVs to reduce the RMSEs brought most of them to 0.8-0.9, but in some cases even a high CV of 1.2 still resulted in RMSE outside the N(0,1) 95% confidence interval. Run S60_BASE_18 omitted five of the indices from the model calibration (NEFSC spring, MADMF spring, RIDFW spring and fall, and VIMS ChesMMAP), and the results and diagnostics examined in comparison to run 17. The run 18 SSB estimates are about 5-10% lower than the run 17 estimates over the terminal 5 years; recruitment at age 0 estimates are 2-5% lower; run F estimates are 10-20% higher (Figures A103-A105). The 'random normal variate' diagnostic plot of survey RMSE indicated that most of the surveys included in run 18 were now close to or inside the confidence interval of the theoretical N(0,1) distribution (Figure A106), indicating better overall survey index fit in the model.

It was noted again that estimates of the recreational fishery landings and discards and commercial fishery discards were based on ratio extrapolation from the commercial fishery landings for all years prior to 1981 or 1989, and that the CVs on those catches was based on the empirical CVs ranging from 13-22%. The CVs on those catches were increased to 30% for years before 1981, creating run S60_BASE_19, to examine the sensitivity of the model run 17 to that setting. Model 19 results were within a few percent of the run 17 results for the entire time series.

Finally, a run including only indices with age composition data, run S60_BASE_20, was examined. The run 20 SSB estimates are about 15-25% higher than the run 17 estimates over the terminal 10 years; recruitment at age 0 estimates are 2-5% lower; run F estimates are 15-25% lower (Table A54).

It was noted that run 18 results were more sensitive to time series length (1989 run start 2014 SSB estimate about 40% lower than the 1963 run start estimate and 2014 F estimate about 50% higher) than run 17 (2014 SSB about 30% lower, F about 45% higher). Run 18 was also more sensitive to the use of BIG indices than run 17, with the 2014 SSB estimate 10% higher and 2014 F 12% lower than when using all LAB equivalent indices; comparable run 17 results were 2014 SSB 5% higher and 2014 F 4% lower.

The SARC concluded that run S60_BASE_18 provided the information needed to meet TOR4 (estimate annual fishing mortality, recruitment and stock biomass for the time series, and estimate their uncertainty). The general results (e.g., record high stock size and low F in the last decade) are robust to the proposed alternative model configurations including alternative time series length and a range of priors and likelihood component weightings. However, there are some indications of poor model fit from lack of correspondence among surveys (higher than expected variance when accounting for potential process error, some residual patterns), and there is some uncertainty in the absolute magnitude of recent stock size estimates (although the terminal year estimates are calculated to be relatively precise with CVs equal to or less than 15%). Alternative survey catchabilities (e.g., relative, absolute using wing or door spread), starting years, and time-varying survey catchability configurations can produce about a +/- 40% range of terminal year SSB.

During the evaluation of the accepted model, sensitivities were examined which highlighted some additional risk. The main one of relevance to management is the choice of selectivity pattern. The base model has a strong domed selectivity pattern which could result in an increasing cryptic biomass given current stock trajectory. Conclusions regarding current stock status are robust to alternative selectivity patterns but decreased recruitment or increased F in the future could lead to divergence between domed and flattop selectivity model results (see Appendix 1). The SARC concluded, however, that the accepted model run provided the best balance between good retrospective diagnostics, acceptable fishery and survey fit diagnostics, and stability over most configurations, and recommended use of ASAP model run S60_BASE_18 for status evaluation.

Figures A107-A109 summarize the 1984 and later SSB, R, and F estimates for runs S60_BASE_1 to S60_BASE_20. Terminal year estimates of SSB range from about 159,000 mt (run 4) to 239,000 mt (run 11), or -13% to +31% of the final run 18 estimate of 183,000 mt. Terminal year estimates of R range from about 49 million (run 2) to 174 million (run 8), or -56% to +55% of the final run 18 estimate of 112 million. Terminal year estimates of F range from about 0.06 (run 11) to 0.14 (run 4), or -54% to +8% of the final run 18 estimate of 0.13.

A8.3 Final Run S60_BASE_18 Diagnostics

A8.3.1 Model Fit Diagnostics (R plots)

Figure A110 shows the distribution of objective function components contribution to total likelihood. The aggregate landings and discards catch and age composition fit diagnostics and residuals are presented in Figures A111-A118. The aggregate survey index and age composition fit diagnostics and residuals are presented in Figures A119-A138.

A8.3.2 Retrospective Analyses

An 'internal' retrospective analysis for the S60_BASE_18 was conducted to examine the stability of the model estimates as data were removed from the end of the time series. Retrospective runs were made for terminal years back to 2007. The scup stock assessment has historically not exhibited a strong retrospective pattern for SSB, F, or recruitment at age 0 (model age 1; R). Over the last seven years, the annual retrospective change in SSB has ranged from -8% in 2009 to -3% in 2007, with an average of -5% (Mohn's rho; Figure A139). The annual retrospective change in recruitment has ranged from -58% in 2011 to +40% in 2012, with an average of -26% (Figure A140). The annual retrospective change in fishing mortality has ranged from -25% in 2007 to +7% in 2013, with an average of -3% (Figure A141). The SWG concluded that these diagnostics indicate that the S60_BASE_18 model run does not exhibit a significant retrospective pattern.

The 2008 DPSWG benchmark assessment (NEFSC 2009), the 2012 assessment update (Terceiro 2012), and model run S60_BASE_18 (2015 SAW 60) results for 1984 and later years are compared in Figures A142-A144 to provide an 'historical' retrospective. The ASAP model has been used in the assessment during the 2008-2015 period, but due to changes in fishery selectivity estimation, 'fully-recruited' F is reported for ages 3-7+ in the 2008-2012 assessments, but only for age 3 ('apical' F where S = 1) in the 2015 assessment, and so is somewhat higher due to increased 'domed' selectivity since 2006 in model run S60_BASE_18. Despite changes in model assumptions, configurations, and estimation procedures, the 'historical' retrospective analysis indicates that the general trends in stock biomass, recruitment, and fishing mortality have been consistent for the last decade.

The estimation results of run S60_BASE_18 are compared with previous 2009-2012 assessment projections of SSB, F, and fishery catch in Figures A145-A147. Final model run S60_BASE_18 estimates of SSB are in line with previous 2009-2012 projections, F is lower than from the 2011-2012 projections, and catch is lower than from the 2011-2012 projections, with the fishery in 2014 taking about 75% of the ACL.

A8.3.3 MCMC Estimates of Uncertainty

Monte Carlo Markov Chain (MCMC) is a common approach to estimate uncertainty in models. A simple MCMC resampling procedure is implemented in ASAP to provide additional

estimates of model estimate uncertainty and an array of starting stock size in 2014 for future projections. For the S60_BASE_18 run, several chains of varying length and seed were examined, with the final one having 5 million iterations thinned by 5,000 to produce 1,000 final iterations for diagnostics and projections. Ideally, the 'trace' of the MCMC chain should not show any trending or patterning, and the correlation between successive values in the chain should be low (e.g., less than 0.1 after year 0).

For the S60_BASE_18 run, however (in fact, for all of the start in 1963 runs examined), uneven patterning was evident in SSB and F estimates, especially for the 1963 estimates (Figures A148-A149). There was also evidence of high correlation between successive estimates of the chain of the 1963 SSB and F for several years, although not for the 2014 estimates (lags; Figures A150-A151). These diagnostics indicate a fairly high level of uncertainty of the model estimates at the beginning of the series. The 'transient' high stock sizes in the initial years of the model and associated very high Fs are a symptom of these issues (e.g., see models S60_BASE_9 and subsequent). The autocorrelation is also reflective of the near-constant recruitment (tight constraint [CV = 0.1] on recruitment deviations and stock-recruitment scaler with fixed h =1 to ensure mean recruitment before 1984, by definition resulting in autocorrelated recruitment during this early period) assumed for the years when no fishery age data are available. The slight autocorrelation at the end of the time series may also reflect the sequence of consecutive very strong (>25% above the time series average) year classes from 1999-2001 and 2005-2008 that are indicated by the fishery and survey catches. The degree of uncertainty results in the point estimates for SSB and F not being 'centered' in the distribution of 1963 MCMC estimates (Figures A152-A153).

Estimates for 2014, in contrast, were well-centered. The 2014 SSB MCMC median was 186,000 mt, mean was 187,000 with CV = 11%, compared to the point estimate of 183,000 mt. The 2014 F MCMC median was 0.122, mean was 0.124 with CV = 15%, compared to the point estimate of 0.127.

Recognizing that these diagnostics in the early part of the series are due to the intentional model configuration and in the latter part of the series are due to stock sizes estimates that are well supported by the fishery and survey input data, it was concluded that there were no serious technical issues in the S60_BASE_18 run that would prevent its use in evaluation of the status of the stock.

A8.4 Profiles and Sensitivity Runs

A8.4.1 Likelihood Profile over assumptions for Natural Mortality (M)

Run S60_BASE_18 was run over a range of assumptions for M values from 0.05 to 0.50 (constant at all ages over time) to help judge which assumption for M fit best, given the diagnostic of total minimum log-likelihood (value of the total objective function). Figure A154 shows that likelihood was minimized for M = 0.15, with runs between 0.05 and 0.20 within 5 objective function total likelihood points. The current value of constant M= 0.20 was retained in the S60_BASE_18 model.

A8.4.2 Likelihood Profile over assumptions for unexploited SSB (SSB0)

A likelihood profile of run S60_BASE_18 over the population scaling parameter SSB0 (unexploited SSB with fixed steepness [h] = 1) with fixed values from 100 kmt to 300 kmt was constructed to help judge the behavior of other likelihood components of the model. Figure A155 indicates that the likelihood of most of the major objective function components is minimized at about 175 kmt (the calculated value for run S60_BASE_18 is 183 kmt with fixed h = 1). It was concluded that no further 'tuning' or other changes in likelihood component emphasis were necessary for the S60_BASE_18 model.

A8.4.3 Sensitivity to NEFSC and NEAMAP survey indices input as swept-area absolute estimates of abundance

All the runs configured through S60_BASE_15 used NEFSC and NEAMAP trawl survey time series of stratified mean numbers per tow with no efficiency assumption made (i.e., indices of relative abundance). In some New England groundfish assessments, assumptions about the efficiency of the trawl gear are made (typically 100%) and 'minimum swept-area numbers' based on area swept by the net wings and/or trawl doors are calculated and used as input to the assessment model (i.e., indices of absolute abundance). This does not result in changes to the estimates of population size and mortality, but does change the scaling of the catchability coefficients ('q') estimated for the surveys.

Some investigators prefer this treatment of the survey calibration data, contending that it serves as a 'check' of whether the scaling of the survey q in an assessment model is 'reasonable' or 'feasible'. Other investigators note that the validity of this 'check' rests on the validity of the assumptions behind the constants used in the simple swept-area calculation (i.e., the size of the trawl gear swept area, the assumption of trawl gear efficiency across lengths and ages, assumption about the uniform distribution of fish within strata, and assumptions about the total area included in the calculation). Experimental estimates of the NEFSC Albatross, NEFSC Bigelow, or NEAMAP trawl gear efficiency for scup are not available.

For the scup S60_BASE_18 model using relative indices for the NEFSC fall and NEAMAP spring and fall, the estimated aggregated N qs are 6.8e-4, 3.7e-5, and 2.4e-5, respectively. Using absolute indices based on wing spread (for NEFSC ALB specifications), the estimated aggregated N qs are 2.17, 0.02, and 0.08, respectively. Using absolute indices based on door spread, the estimated aggregated N qs are 1.02, 0.01, and 0.03, respectively. It was concluded that while it may be useful to look at q estimates using swept area indices to provide context for model estimates, the results should not be used to make reach conclusions about the accuracy of the 'scaling' of the assessment model until field experiments have been conducted to study the behavior of a particular species in reaction to the survey gear and better quantify survey catchability.

A8.4.4 Varying NEFSC and NEAMAP survey catchability

As described under TOR 3, the working paper of Manderson et al. (MS 2015; WP 11) provides time series of varying estimates of the proportion of thermal habitat suitability for scup

surveyed on the Northeast US shelf by the NEFSC and NEAMAP bottom trawl surveys from 1975-2012 in a manner that accounts for thermal habitat occurring outside the surveys and the relative motions of habitat and the survey vessel. Logit-transformed annual values of the 'proportion of suitable scup thermal habitat sampled' – i.e., availability - were used in an ASAP4 version of run S60_BASE_18 to provide annually varying estimates of relative survey catchability (q), where q is the product of availability and survey gear efficiency (assumed = 1).

The NEFSC survey qs were estimated to be variable without long term trend; NEAMAP survey qs were variable over the short 7-8 year time series. Compared to the ASAP3 version of run S60_BASE_18, there were changes in some SV residual patterns, with RMSEs generally larger. ASAP4 run 18 estimation results for 2014 were close to the ASAP3 results, with 2014 SSB estimated to be 3% lower, R 23% higher, and F 4% lower. Given the similarity of results and still preliminary nature of the ASAP4 model (the model and documentation have not yet been released to the public), the ASAP4 version of run 18 was not used for status evaluation.

A8.5 Annual Fishing Mortality, Recruitment, and Stock Size Estimates

Summary SSB, recruitment, and F estimates, estimated January 1 stock size at age in numbers, and estimated fishing mortality (F) at age from the final model (S60_BASE_18) for 1984-2014 (the years with input fishery catches at age) are provided in Tables A55-A56. Spawning stock biomass (SSB) decreased from about 68,000 mt in 1963 to about 5,000 mt in 1969, then increased to about 27,000 mt during the late 1970s. SSB declined through the 1980s and early 1990s to less than about 4,000 mt in the mid-1990s. With greatly improved recruitment and low fishing mortality rates since the late 1990s, SSB increased to greater than 100,000 mt = 220 million lbs since 2003. SSB was estimated to be 182,915 mt = 403 million lbs in 2014 (Figures A156-A157). There is a 90% probability that SSB in 2014 was between 153,000 and 222,000 mt (337 and 489 million lbs; Figure A158). Fishing mortality estimated at the 'apical' age 3 (model age 4) where full selection occurs (S=1) varied between F = 0.5 and F = 2.0 during the 1960s and 1970s. Fishing mortality next peaked at about F = 1.5 in the 1990s. Fishing mortality decreased after 1994, falling to less than F = 0.15 since 2000, with F in 2014 = 0.127 (Figure A159). There is a 90% probability that F in 2014 was between 0.093 and 0.149 (Figure A160).

Recruitment at age 0 averaged 98 million fish during 1963-1983, the period in which recruitment estimates are tightly constrained (CV = 0.1 on recruitment deviations and stock-recruitment scaler with fixed h =1) to ensure near constant recruitment before 1984, when fishery catch at age are not available. Since 1984, recruitment estimates from the model are influenced mainly by the fishery and survey catches at age, and averaged 109 million fish during 1984-2014. The 1999, 2006, and 2007 year classes are estimated to be the largest of the time series, at 222, 222, and 218 million age 0 fish. After below average recruitment in 2012 and 2013, the 2014 year class is estimated to be above average at 112 million age 0 fish (Figures A156-A157).

A9. TERM OF REFERENCE 5: 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}, $B_{THRESHOLD}$, F_{MSY} and 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.

A9.1 Existing: 2008 DSP Assessment Biological Reference Points

The 2008 DPSWG Peer Review Panel accepted the ASAP SCAA model results as the basis for biological reference points and status determination for scup (NEFSC 2009). Reference points were calculated using the non-parametric yield and SSB per recruit/long-term projection approach adopted for summer flounder (NEFSC 2008a) and the New England groundfish stocks (NEFSC 2008b). In the yield and SSB per recruit calculations, the most recent five year averages were used for mean weights and fishery partial recruitment pattern. For the estimation of MSY (Maximum Sustainable Yield) and SSBMSY (Spawning Stock Biomass at Maximum Sustainable Yield), the cumulative distribution function of the 1984-2007 recruitments (corresponding to the period of input fishery catches at age) was re-sampled to provide future recruitment estimates (mean = 117 million age 0 fish). The existing reference points for scup are the 2008 DPSWG Peer Review Panel recommended F40% as the proxy for FMSY, and the corresponding SSBF40% as the proxy for SSBMSY. The F40% proxy for FMSY = 0.177, the proxy estimate for SSBMSY = SSB40% = 92,044 mt = 202.922 million lbs, and the proxy estimate for MSY = MSY40% = 16,161 mt = 35.629 million lbs (13,134 mt = 28.956 million lbs of landings and 3,027 mt = 6.673 million lbs of discards).

A9.2 New: 2015 SAW 60 Biological Reference Points

The SARC accepted the ASAP SCAA model run S60_BASE_18 results as the basis for new biological reference points and status determination for scup. Reference points were again calculated using the non-parametric yield and SSB per recruit/long-term projection approach adopted for summer flounder (NEFSC 2008a) and the New England groundfish stocks (NEFSC 2008b). In the yield and SSB per recruit calculations, the most recent five year averages were used for mean weights and fishery partial recruitment pattern. For the estimation of MSY (Maximum Sustainable Yield) and SSBMSY (Spawning Stock Biomass at Maximum Sustainable Yield), the cumulative distribution function of the 1984-2014 recruitments (corresponding to the period of input fishery catches at age) was re-sampled to provide future recruitment estimates (mean = 109 million age 0 fish). The SARC recommended F40% as the proxy for FMSY, and the corresponding SSBF40% as the proxy for the SSBMSY biomass target. The F40% proxy for FMSY = 0.220. The proxy estimate for SSBMSY = SSB40% = 87,302 mt = 192.468 million lbs; the proxy estimate for the ½ SSBMSY biomass threshold = ½ SSB40% = 43,651 mt = 96.234 million lbs. The proxy estimate for MSY = MSY40% = 11,752 mt = 25.909 million lbs (9,445 mt = 20.823 million lbs of landings and 2,307 mt = 5.086 million lbs of discards).

A10. TERM OF REFERENCE 6: Evaluate stock status with respect to the existing model (from previous peer reviewed accepted assessment) and with respect to a new model developed for this peer review.

a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.

b. Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-5).

2015 UPDATED STOCK STATUS

a) The existing model updated with new data indicated that the scup stock was not overfished and overfishing was not occurring in 2014 relative to the existing (old) biological reference points established in the 2008 Northeast Data Poor Stocks Working Group (DPSWG; NEFSC 2009) assessment. The fishing mortality rate (F) was estimated to be 0.049 in 2014, below the fishing mortality threshold reference point = FMSY = F40% = 0.177. Spawning Stock Biomass (SSB) was estimated to be 219,066 metric tons (mt) = 483 million lbs in 2014, above the biomass target reference point = SSBMSY = SSB40% = 92,044 mt = 203 million lbs (Table A58).

b) The scup stock was not overfished and overfishing was not occurring in 2014 relative to the new biological reference points recommended by the 2015 SWG. The fishing mortality rate (F) was estimated to be 0.127 in 2014, below the fishing mortality threshold reference point = FMSY = F40% = 0.220. Spawning Stock Biomass (SSB) was estimated to be 182,915 metric tons (mt) = 403 million lbs in 2014, above the biomass target reference point = SSBMSY = SSB40% = 87,302 mt = 192 million lbs (Table A58, Figure A161).

A11. TERM OF REFERENCE 7: Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) (see Appendix to SAW TORs for definitions).

a. Provide numerical annual projections (3 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).

b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.

c. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

A11.1 Numerical Annual Projections for 2016-2018

Stochastic projections were made to provide forecasts of stock size and overfishing level (OFL) catches in 2016-2018 consistent with the 2015 SAW 60 assessment biological reference points. The projections assume that recent (2010-2014) patterns of discarding will continue over the time span of the projections. Different patterns that could develop in the future due to different trip and bag limits and fishery closures have not been evaluated. One hundred projections were made for each of the 1000 MCMC (Markov Chain Monte Carlo) realizations of 2014 stock sizes from the updated assessment results using NFT AGEPRO version 4.0.5 (NFT 2011). Future recruitment at age 0 was generated randomly from a cumulative density function of the updated recruitment series for 1984-2014 (mean recruitment = 109 million fish).

Two sets of projections were conducted. Option A is proposed as the most realistic and assumes that given recent patterns in the fishery, it is likely that 75% of the 2015 Allowable Biological Catch (ABC) will be caught. Projection option B assumes that 100% of the 2015 ABC will be caught.

Option A) If the catch of scup in 2015 equals 75% of the specified ABC = 0.75 * 15,320 = 11,490 mt = 25.331 million lbs, the 2015 median (50% probability) landings are projected to be 10,058 mt = 22.174 million lbs and discards are projected to be 1,432 mt = 3.157 million lbs. The table below shows the projected biomass and catch for Option A in 2015 if the stock is then fished at the fishing mortality threshold = FMSY = F40% = 0.220 in 2016-2018. The projected OFLs in 2016-2018 are 16,238, 14,556, and 13,464 mt (35.799, 32.090, and 29.683 million lbs).

Option A: Total Catch (OFL), Landings, Discards, Fishing Mortality (F)
and Spawning Stock Biomass (SSB) in 2015-2018
Catches and SSB in metric tons

Year	Total Catch (OFL)	OFL CV (%)	Landings	Discards	F	SSB
2015	11,490	fixed	10,058	1,432	0.143	187,477
2016	16,238	14	13,840	2,398	0.220	170,002
2017	14,556	13	12,214	2,342	0.220	154,083
2018	13,464	13	11,156	2,308	0.220	141,077

Option B) If the catch of scup in 2015 equals 100% of the specified ABC = 15,320 mt = 33.775 million lbs, the 2015 median (50% probability) landings are projected to be 13,412 mt = 29.568 million lbs and discards are projected to be 1,908 mt = 4.206 million lbs. The table below shows the projected biomass and catch for Option B in 2015 if the stock is then fished at the fishing mortality threshold = FMSY = F40% = 0.220 in 2016-2018. The projected OFLs in 2016-2018 are 15,745, 14,199, and 13,230 mt (34.712, 31.303, and 29.167 million lbs).

Option B: Total Catch (OFL), Landings, Discards, Fishing Mortality (F)
and Spawning Stock Biomass (SSB) in 2015-2018
Catches and SSB in metric tons

Year	Total Catch (OFL)	OFL CV (%)	Landings	Discards	F	SSB
2015	15,320	fixed	13,412	1,908	0.194	185,916
2016	15,745	13	13,398	2,347	0.220	166,355
2017	14,199	12	11,883	2,316	0.220	150,702
2018	13,230	12	10,935	2,295	0.220	138,072

The biological inputs to the scup stock assessment are based on well-founded assumptions (e.g., for natural and discard mortality) and precisely estimated parameters (e.g., growth, age, maturity, and mean weights). Further, the research survey index CVs used in model calibration have been increased by 50-100% (depending on assessment model fit diagnostics) to account for process error. Twenty-five alternative configurations of the assessment base model were examined to evaluate robustness, including starting years, impact of NEFSC calibration factors, natural mortality, fishery selectivity, and time-varying survey catchability. This broad set of configurations produced a range about +/- 40% in the estimate of terminal year SSB of about 180,000 mt (= 396 million lbs). The internal retrospective average error (for the terminal 7-years) of the assessment is low, at less than 10% for both SSB and F. The analytically derived CV for the 2014 SSB is 11%, the CV for the 2014 F is 15%, and the CV for the 2014 age 1 and

older stock size total number is 15%. Given these properties of the 2015 scup stock assessment, it was concluded that an approximate doubling of the analytically derived 2016-2018 OFL CVs to 30% is a reasonable and sufficient adjustment to account for additional uncertainty in the assessment such as the magnitude of domed fishery selection, the magnitude of commercial fishery discards and recreational catch during the early part of the assessment model time series, and potential error in the aging process.

A11.2 Most Realistic Projections

The commercial and recreational fisheries have landed about 75% of the landings quota over the last two years, suggesting that the 2015 ACL may not all be caught. The SWG concluded that a projection assuming that 75% of the 2015 ABC will be caught was more realistic than assuming 100% will be caught, and this scenario is identified as 'Option A.' An Option B projection assuming 100% of the 2015 ABC will be caught is also provided.

A11.3 Stock Vulnerability

The 2008 DPSWG Peer Review Panel (NEFSC 2009) advised that a gradual increase in the ABC toward the MSY level would facilitate an evaluation of the performance of the new assessment model and reference points in monitoring stock status, while reducing the risk to the stock due to rapidly increased catch.

The 2015 assessment indicates that the stock was well above the biomass target and being fished at well below the fishing mortality threshold in 2014. The high level of 2014 stock abundance is the result of historically low fishing mortality rates and historically high levels of recruitment since the late 1990s. The MSY proxy in terms of total catch is 11,752 mt (25.909 million lbs; CV = 19%), with total landings of 9,445 mt (20.823 million lbs) and total discards of 2,307 mt (5.086 million lbs). Total fishery catch is estimated to have averaged about 34,000 mt (~75 million lbs) during 1960-1965, while reported commercial landings alone averaged about 19,000 mt (~42 million lbs) in that period. Therefore, the MSY estimate appears feasible given historical evidence from the fishery.

Both projection options have a realistic probability of being achieved and indicate there is zero percent chance that SSB will fall below the biomass threshold in 2016-2018 fishing at the OFL. The scup stock has a low probability of becoming overfished in the short term (2016-2018) given recent trends in productivity and the responsiveness of the management regime.

A12. TERM OF REFERENCE 8: Review, evaluate and report on the status of the SARC, SSC, and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

Nine of the 12 previously identified research recommendations were either addressed in full or significant progress was made. No progress has been made on a) quantifying contemporary discard mortality rates, b) quantifying the degree of bias in landings reporting and discard estimation including non-compliance, or c) development of a management strategy evaluation of alternative approaches to setting quotas. Six newly developed research recommendations are listed below.

A12.1 Previous Research Recommendations

A12.1.1 DPWG 2008 (NEFSC 2009)

Short term analytical tasks

1) Evaluation of indicators of potential changes in stock status that could provide signs to management of potential reductions of stock productivity in the future would be helpful.

Some progress in SSC work on 'rumble strip' analysis – used in 2013.

The 2015 assessment explored the potential use of the Conn (2010) hierarchical method to combine indices across time and space; more developmental work is needed.

2) A management strategy evaluation of alternative approaches to setting quotas would be helpful.

No progress.

Long term data and analytical needs

3) Current research trawl surveys are likely adequate to index the abundance of scup at ages 0 to 2. However, the implementation of new standardized research surveys that focus on accurately indexing the abundance of older scup (ages 3 and older) would likely improve the accuracy of the stock assessment.

The RI Industry Cooperative Trap survey was implemented during 2005-2012. This survey had a higher catch rate for larger and older fish of age 3+ than the bottom trawl surveys. A peer review indicated that some of the design elements should be modified and this advice was followed; however, funding was halted after 2012.

4) Continuation of at least the current levels of at-sea and port sampling of the commercial and recreational fisheries in which scup are landed and discarded is critical to adequately characterize the quantity, length and age composition of the fishery catches.

Adequate sampling has been maintained (see assessment tables and figures).

5) Quantification of the biases in the catch and discards, including non-compliance, would help confirm the weightings used in the model. Additional studies would be required to address this issue.

No progress.

6) The commercial discard mortality rate was assumed to be 100% in this assessment. Experimental work to better characterize the discard mortality rate of scup captured by different commercial gear types should be conducted to more accurately quantify the magnitude of scup discard mortality.

No progress.

A12.1.2 MAFMC SSC July 2012

1) Improve estimates of discards and discard mortality for commercial and recreational fisheries

SBRM estimates of commercial fishery discards, which exhibit a less variable time series pattern and improved precision compared to previous estimates, were developed and accepted for this assessment.

No progress on discard mortality rates.

2) Evaluate indices of stock abundance from new surveys

The RI Cooperative Trap (ended in 2012), NEAMAP spring and fall surveys, indices at age from the RIDFW spring and fall surveys, and indices at age from the NYDEC survey are now included in the assessment documentation.

3) Quantify the pattern of predation on scup

The limited NEFSC survey food habits data for scup were reviewed and it is not possible to calculate absolute estimates of consumption of scup by predators due to sample size considerations (~500 identifiable scup in the ~40 year time series).

4) Conduct biological studies to investigate maturity schedules and factors affecting annual availability of scup to research surveys

The NEFSC maturity schedule for scup was updated.

GLM and GAM modeling and GIS investigation of NEFSC bottom trawl survey data on scup distribution, temperature preference, and salinity preference did not reveal strong effects that could be directly linked to a trend in availability.

Changes in scup distributions with respect to bottom temperature, body size and abundance within the NEFSC survey were examined to identify potential effects on availability. A thermal habitat model was developed to estimate proportions thermal habitat suitability for scup sampled during fall and spring NEFSC and NEAMAP surveys. These habitat based estimates of availability were used to inform catchability in sensitivity evaluations of the final ASAP model.

5) Explore the utility of incorporating ecological relationships, predation, and oceanic events that influence scup population size on the continental shelf and its availability to resource surveys into the stock assessment mode

GLM and GAM modeling and GIS investigation of NEFSC bottom trawl survey data on scup distribution, temperature preference, and salinity preference did not reveal strong effects that could be directly linked to a trend in availability.

Changes in scup distributions with respect to bottom temperature, body size and abundance within the NEFSC survey were examined to identify potential effects on availability. A thermal habitat model was developed to estimate proportions thermal habitat suitability for scup sampled during fall and spring NEFSC and NEAMAP surveys. These habitat based estimates of availability were used to inform catch ability in sensitivity evaluations of the final ASAP model.

6) Evaluate alternate forms of survey selectivity in the assessment to inform indices of abundance at higher ages

The multinomial approach to inclusion of fishery and survey catch at age was used in the assessment model, allowing use of low and variable indices at older ages and, where possible, estimation of selectivity at age.

A12.2 New Research Recommendations

1) A standardized fishery dependent CPUE of scup targeted tows, from either NEFOP observer samples or the commercial study fleet, might be considered as an additional index of abundance to complement survey indices in future benchmark assessments

2) Explore additional sources of length/age data from fisheries and surveys in the early parts of the time series to provide additional context for model results

3) Explore experiments to estimate the catchability of scup in NEFSC and other research trawl surveys (side-by-side, camera, gear mensuration, acoustics, etc.)

4) Refine and update the Manderson et al. availability analysis when/if a new ocean model is available (need additional support). Explore alternative niche model parameterizations including laboratory experiments on thermal preference and tolerance.

5) Explore the Study fleet data in general for information that could provide additional context and/or input for the assessment

6) A scientifically designed survey to sample larger and older scup would likely prove useful in improving knowledge of the relative abundance of these large fish.

Acknowledgments

Special thanks to Daphne Cassidy, Grace Thornton, Nicole Calabrese and the staff of the NOAA Fisheries NEFSC Population Biology Branch for their timely preparation of the 2014 scup ages used in this benchmark assessment.

Literature Cited

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.

Bowman RE, Stillwell CE, Michaels WL, Grosslein MD. 2000. Food of Northwest Atlantic fishes and two common species of squid. NOAA Tech Memo NMFS-F/NE-155. 138 p.

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.

Campbell MJ, Pentilla PA, Nichy FE. MS 1982. Growth of scup (*Stenotomus chrysops*). Unpublished Northeast Fisheries Science Center Ref Doc.

Cochran WG. 1963. Sampling Techniques. J. Wiley and Sons. New York.

Cogswell SJ. 1960. Summary of tagging operations, July 1, 1959 through June 30, 1960. US Bur. Comm. Fish. Woods Hole Laboratory. Lab Ref No. 60-1.

Cogswell SJ. 1961. Summary of tagging operations, July 1, 1960 through June 30, 1961. US Bur. Comm. Fish. Woods Hole Laboratory. Lab Ref No. 61-12.

Conn PB. 2010. Hierarchical analysis of multiple noisy abundance indices. Can J Fish Aquat Sci. 67: 108-120.

Conser RJ, Powers JE. 1990. Extension of the ADAPT VPA tuning method designed to facilitate assessment work on tuna and swordfish stocks. ICCAT Coll Vol Sci Pap 32: 461-477.

Crecco V, Maltezos G, Howell-Heller P. 1981. Population dynamics and stock assessment of the scup, (*Stenotomus chrysops)*, from New England waters. Conn Dept Environ Protect. Mar. Fish. Completion Rep. No. 3-328-R-2 CT, 62 p.

Dery L, Rearden C. 1979. Report of the state-federal scup (*Stenotomus chrysops*) age and growth workshop. National Marine Fisheries Service, Northeast Fisheries Center, Woods Hole Laboratory Lab Ref Doc. 79-57.

Finkelstein SL. 1969a. Age and growth of scup in the waters of eastern Long Island. NY Fish and Game Jour 16(1): 84-110.

Finkelstein SL. 1969b. Age at maturity of scup from New York waters. NY Fish and Game Jour 16(2): 224-237.

Francis RICC. 2011. Data weighting in statistical fisheries stock assessment models. Can J Fish Aquat Sci. 68: 1124-1138.

Froese R. 2006. Cube law, condition factor, and weight-length relationships: history, meta-analysis and recommendations. J Appl Ichthyol 22: 241-253.

Fulton TW. 1904. The rate of growth of fishes. 22nd Annual Report of the Fishery Board of Scotland 1904 (3):141-241.

Gavaris S. 1980. Use of a multiplicative model to estimate catch rate and effort from commercial data. Can. J. Fish. Aquat. Sci., 37: 2272-2275.

Gislason H, Daan N, Rice JC, Pope JG. 2010. Size, growth, temperature and the natural mortality of fish. Fish and Fisheries 2: 149-158.

Hamer PE. 1970. Studies of the scup, (*Stenotomus chrysops)*, in the Middle Atlantic Bight. NJ Div. Fish Game and Shellfish. Misc Rep No. 5M. 14 p.

Hamer PE. 1979. Studies of the scup, (*Stenotomus chrysops)*, in the Middle Atlantic Bight. NJ Div. Fish Game and Shellfish. Misc Rep No. 18M. 67 p.

Heincke F. 1908. Bericht über die Untersuchungen der Biologischen Anstalt auf Helgoland zur Naturgeschichte der Nutzfische. Die Beteiligung Deutschlands an der Internationalen Meeresforschung 1908:4/5:67-155.

Howell PT, Simpson DG. 1985. A study of marine recreational fisheries in Connecticut. March 1, 1981 - February 28, 1984. CTDEP, Fed Aid to Sport Fish Restoration F54R. Final Rep. 60 p.

Hastie TJ, Tibshirani RJ. 1990. Generalized additive models. Chapman and Hall, New York. 352 p.

Healey BP, Cadigan NG, Brodie WB. MS 2001. Analysis of pre-recruit data from surveys for Greenland halibut in NAFO subarea 2 and divisions 3KLMNO. NAFO SCR Doc. 01/44, Serial No. N4422.

Hewitt DA, Hoenig JM. 2005. Comparison of two methods 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. 81: 898-902.

Kimura D.K. 1981. Standardized measures of relative abundance based on modeling log (c.p.u.e.), and their application to Pacific ocean perch (*Sebastes alutus*). J.Cons. Int. Explor. Mer 39: 211-218.

Legault CM, Restrepo VR. 1998. A flexible forward age-structured assessment program. ICCAT Col Vol Sci Pap. 49: 246-253.

Legault CM, Chair. 2009. Report of the Retrospective Working Group, January 14-16, 2008, Woods Hole, Massachusetts. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 09-01; 30 p.

Legault CM, Alade L, Stone HH. 2010. Assessment of Georges Bank yellowtail flounder for 2010. TRAC Reference Document 2010/06. 97 p.

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

Lorenzen K. 2000. Allometry of natural mortality as a basis for assessing optimal release size in fish-stocking programmes. Can J Fish Aquat Sci. 57:2374-2381.

Love JW, Chase PD. 2009. Geometric morphological differences distinguish populations of scup in the northwest Atlantic ocean. Marine and Coastal Fisheries 1:1: 22-28.

Manderson JP, Schmidt A, Palamara L, Richardson D, Kohut J, Bonzek C. MS 2015. TOR 3: Describe the thermal habitat and its influence on the distribution and abundance of scup, and attempt to integrate the results into the stock assessment. 2015 SAW 60 Scup Working Group Working Paper A11. 52 p.

Mayo RK. 1982. An assessment of the scup, (*Stenotomus chrysops)* (L.), population in the Southern New England and Middle Atlantic regions. NMFS NEFC, Woods Hole Lab. Ref Doc No. 82-46. 60 p.

McCullagh P, Nelder JA. 1989. Generalized linear models. Chapman and Hall, London. 511 p.

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. Northeast Fisheries Science Center Ref Doc. 10-05. 233 p.

Mohn R. 1999. The retrospective problem in sequential population analysis: an investigation using cod fishery and simulated data. ICES J Mar Sci. 56: 473-488.

Morse WW. 1978. Biological and fisheries data on scup, (*Stenotomus chrysops)* (Linnaeus). NMFS NEFC. Sandy Hook Lab Tech Ser Rep. No 12. 41 p.

Nash DM, Valencia AH, Geffen AJ. 2006. The origin of Fulton's condition factor – setting the record straight. Fisheries 31(5): 236-238.

NEFSC. 1995. Report of the 19th Northeast Regional Stock Assessment Workshop (19th SAW): Stock Assessment Review Committee (SARC) consensus summary of assessments. Northeast Fish Sci Cent Ref Doc. 95-08: 221 p.

NEFSC. 1997. Report of the 25th Northeast Regional Stock Assessment Workshop (25th SAW): Stock Assessment Review Committee (SARC) consensus summary of assessments. Northeast Fish Sci Cent Ref Doc. 97-14: 143 p.

NEFSC. 1998. Report of the 27th Northeast Regional Stock Assessment Workshop (27th SAW). Stock Assessment Review Committee (SARC) consensus summary of assessments. Northeast Fish Sci Cent Ref Doc. 98-15: 350 p.

NEFSC. 2000. Report of the 31st Northeast Regional Stock Assessment Workshop (31st SAW). Stock Assessment Review Committee (SARC) consensus summary of assessments. Northeast Fish Sci Cent Ref Doc. 00-15: 400 p.

NEFSC. 2002. Report of the 35th Northeast Regional Stock Assessment Workshop (35th SAW). Stock Assessment Review Committee (SARC) consensus summary of assessments. Northeast Fish Sci Cent Ref Doc. 02-14: 259 p.

NEFSC. 2008a. 47th Northeast Regional Stock Assessment Workshop (47th SAW) Assessment Report & Appendixes. by Northeast Fisheries Science Center. CRD 08-12. 339 p.

NEFSC. 2008b. Assessment of 19 Northeast Groundfish Stocks through 2007: Report of the 3rd Groundfish Assessment Review Meeting (GARM III). Northeast Fisheries Science Center (NEFSC) Woods Hole, Massachusetts. August 4-8, 2008. NEFSC. CRD 08-15. 867 p.

NEFSC 2009. The Northeast Data Poor Stocks Working Group Report, December 8-12, 2008 Meeting. Part A. Skate species complex, Deep sea red crab, Atlantic wolffish, Scup, and Black sea bass. US Dept Commerce, Northeast Fish Sci Cent Ref Doc. 09-02; 496 p.

NEFSC 2011a. 51st Northeast Regional Stock Assessment Workshop (51st SAW) Assessment Report. US Dept Commerce, Northeast Fish Sci Cent Ref Doc. 11-02, 856 p.

NEFSC 2011b. 52nd Northeast Regional Stock Assessment Workshop (52nd SAW) Assessment Summary Report. US Dept Commerce, Northeast Fish Sci Cent Ref Doc. 11-11, 51 p.

NEFSC. 2013a. Northeast Fisheries Science Center Science Data Collection Program review Summary Report. 31 p. http://www.nefsc.noaa.gov/program_review/reports2013.html

NEFSC. 2013b. 57th Northeast Regional Stock Assessment Workshop (57th SAW) Assessment Summary Report. US Dept Commerce, Northeast Fish Sci Cent Ref Doc. 13-16, 967 p.

Neville WC, Talbot GB. 1964. The fishery for scup with special reference to fluctuations in yield and their courses. US Fish Wildl Serv. Spec Sci Rep - Fish. No 459. 61 p.

NFT 2008. NOAA Fisheries Toolbox (NFT). ASAP version 2.0.20. [Internet address http://nft.nefsc.noaa.gov].

NFT 2013a. NOAA Fisheries Toolbox (NFT). Yield per recruit (YPR) version 3.2.1 [Internet address http://nft.nefsc.noaa.gov].

NFT 2013b. NOAA Fisheries Toolbox (NFT). AGEPRO version 4.2.2.

NFT 2014. Age Structured Assessment Program (ASAP) version 3.0.17. (Internet address: http://nft.nefsc.noaa.gov).

O'Brien LS, Mayo RK. 1988. Sources of variation in catch per unit of effort of yellowtail flounder, *Limanda ferruginea* (Storer), harvested of the coast of New England. Fish. Bull. 86(1):91-108.

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

Pentilla JA, Nelson GA, Burnett III JM. 1989. Guidelines for estimating lengths at age for 18 Northwest Atlantic finfish and shellfish species. NOAA Technical Memorandum NMFS-F/NEC-66. 39 p.

Prager MH. 1994. A suite of extensions to a non-equilibrium surplus production model. Fish Bull US. 92:374-389.

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

Ricker WE. 1975. Computation and interpretation of biological statistics of fish populations. Bulletin of the Fisheries Research Board of Canada 191:1-382.

Robson DS. 1960. Estimation of the relative fishing power of individual ships. Comm. N.W. Atl. Fish. Res. Bull. 3: 5-14.

SAS Institute. 2011. SAS OnlineDoc, version 9. SAS Institute Inc., Cary NC.

Searle SR. 1987. Linear models for unbalanced data. John Wiley and Sons, Inc., New York.536 p.

Simpson DG, Howell PT, Johnson MW. 1990. Section 2 Job 6: Marine finfish survey in the State of Connecticut; a study of marine recreational fisheries in Connecticut, 1984-1988. CTDEP Fed Aid to Sport Fish Restoration. F54R Final Rep. 265 p.

Sisson RT. 1974. The growth and movement of scup (*Stenotomus chrysops*) in Narragansett Bay, RI and along the Atlantic coast. RI Division of Fish and Wildlife Completion Report. 3-138-R-3. 21 p.

Stansbury DE, Shelton PA, Murphy EF, Healey BP, and Brattey J. MS 2001. An assessment of the cod stock in NAFO divisions 3NO. NAFO SCR Doc. 01/72. Serial No. N4450.

Terceiro M. 2009. Stock assessment of scup for 2009. US Dept. Commerce, Northeast Fish Sci Cent Ref Doc. 09-18. 82 p.

Terceiro M. 2010. Stock assessment of scup for 2010. US Dept. Commerce, Northeast Fish Sci Cent Ref Doc. 10-16. 86 p.

Terceiro M. 2011. Stock assessment of scup for 2011. US Dept. Commerce, Northeast Fish Sci Cent Ref Doc. 11-21. 98 p.

Terceiro M. 2012. Stock assessment of scup for 2012. US Dept. Commerce, Northeast Fish Sci Cent Ref Doc. 12-25. 104 p.

Then AY, Hoenig JM, Hall NG, Hewitt DA. 2014. Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. ICES J Mar Sci.doi:10.1093/icesjm/fsu136.

Van Eeckhaute L, Brooks EN. 2010. Assessment of Eastern Georges Bank Haddock for 2010. TRAC Reference Document - 2010/05. 104 p.

Wigley SE, McBride HM, McHugh NJ. 2003. Length-weight relationships for 74 fish species collected during NEFSC research vessel bottom trawl surveys, 1992-99. NOAA Technical Memorandum NMFS-NE-171.

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

Wigley SE, Blaylock J, Rago PJ, Tang J, Hass HL, Shield G. 2011. Standardized Bycatch Reporting Methodology 3-year review report – 2011 Part 1. Northeast Fisheries Science Center Ref Doc 11-09. 285 p.

Williams E. Personal communication. University of Rhode Island, Department of Fisheries and Aquaculture. Kingston, RI. November 1, 1994.