B: STOCK ASSESSMENT OF WITCH FLOUNDER FOR 2016

SAW-62 Editor's Note regarding witch flounder:

The SARC-62 peer review panel did not accept specific applications of the analytical assessment models (e.g., ASAP, VPA, Replacement Yield, SCAA) for witch flounder presented in this report or its appendices. SARC-62 does not recommend that the specific results from those models serve as a basis for fishery management decisions for 2017.

The analytical models and results are included in this report to document the material that the witch flounder SAW Working Group provided to the SARC-62 for peer review.

Text expressing SARC-62 panel opinions has been added *in italics* in the EXECUTIVE SUMMARY and TOR sections of this report to reflect the outcome of the SARC-62 peer review.

The SARC-62 panel is recommending that managers consider using the empirical approach discussed in this document as the basis for developing management advice. The SARC-62 did not fully evaluate the empirical approach during the peer review, and therefore expected additional evaluation of the method, calculations, and results to be made following the peer review.

TERMS OF REFERENCE

B. Witch flounder

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

2. Present available federal, state, and other survey data, indices of relative or absolute abundance, recruitment, etc. Characterize the uncertainty and any bias in these sources of data and compare survey coverage to locations of fishery catches. Select the surveys and indices for use in the assessment.

3. Investigate effects of environmental factors and climate change on recruitment, growth and natural mortality of witch flounder. If quantifiable relationships are identified, consider incorporating these into the stock assessment.

4. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series (integrating results from TOR-3 if appropriate), and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results and previous projections. Compare F's and SSB's that were projected during the previous assessment to their realized values.

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 (or possibly models, in accord with guidance in attached "Appendix to the SAW Assessment TORs") developed for this peer review. In both cases, evaluate whether the stock is rebuilt.
    a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the updated BRP estimates.
    b. Then use the newly proposed model (or possibly models, in accord with guidance in "Appendix to the SAW Assessment TORs") 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.
    a. Provide numerical annual projections (3 years) and the statistical distribution (e.g., probability density function) of the catch at F_{MSY} or an F_{MSY} proxy (i.e. the overfishing limit, OFL) (see "Appendix to the SAW Assessment TORs"). 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, magnitude and 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. Identify reasonable projection parameters (recruitment, weight-at-age, retrospective adjustments, etc.) to use when setting specifications.

c. Describe this stock's vulnerability to becoming overfished, and how this could affect the choice of ABC. The choice takes scientific uncertainty into account (see "Appendix to the SAW Assessment TORs").

8. Evaluate the validity of the current stock definition, taking into account what is known about migration, and make a recommendation about whether there is a need to modify the current stock definition for future stock assessments.

9. Review, evaluate and report on the status of research recommendations from the last peer reviewed benchmark stock assessment. Identify new research recommendations.

Appendix to the SAW Assessment TORs: **Clarification of Terms used in the SAW/SARC Terms of Reference**

Guidance to SAW WG about "Number of Models to include in the Assessment Report": In general, for any TOR in which one or more models are explored by the WG, give a detailed presentation of the "best" model, including inputs, outputs, diagnostics of model adequacy, and sensitivity analyses that evaluate robustness of model results to the assumptions. In less detail, describe other models that were evaluated by the WG and explain their strengths, weaknesses and results in relation to the "best" model. If selection of a "best" model is not possible, present alternative models in detail, and summarize the relative utility each model, including a comparison of results. It should be highlighted whether any models represent a minority opinion.

On "Acceptable Biological Catch" (DOC Nat. Stand. Guidel. Fed. Reg., v. 74, no. 11, 1-16-2009):

Acceptable biological catch (ABC) is a level of a stock or stock complex's annual catch that accounts for the scientific uncertainty in the estimate of Overfishing Limit (OFL) and any other scientific uncertainty…" *(p. 3208) [In other words, OFL ≥ ABC.]*

ABC for overfished stocks. For overfished stocks and stock complexes, a rebuilding ABC must be set to reflect the annual catch that is consistent with the schedule of fishing mortality rates in the rebuilding plan. *(p. 3209)*

NMFS expects that in most cases ABC will be reduced from OFL to reduce the probability that overfishing might occur in a year. (p. 3180)

ABC refers to a level of ''catch'' that is ''acceptable'' given the ''biological'' characteristics of the stock or stock complex. As such, Optimal Yield (OY) does not equate with ABC. The specification of OY is required to consider a variety of factors, including social and economic

factors, and the protection of marine ecosystems, which are not part of the ABC concept. (p. 3189)

On "Vulnerability" (DOC Natl. Stand. Guidelines. Fed. Reg., v. 74, no. 11, 1-16-2009):

"Vulnerability. A stock's vulnerability is a combination of its productivity, which depends upon its life history characteristics, and its susceptibility to the fishery. Productivity refers to the capacity of the stock to produce Maximum Sustainable Yield (MSY) and to recover if the population is depleted, and susceptibility is the potential for the stock to be impacted by the fishery, which includes direct captures, as well as indirect impacts to the fishery (e.g., loss of habitat quality)." (p. 3205)

Participation among members of a Stock Assessment Working Group:

Anyone participating in SAW meetings that will be running or presenting results from an assessment model is expected to supply the source code, a compiled executable, an input file with the proposed configuration, and a detailed model description in advance of the model meeting. Source code for NOAA Toolbox programs is available on request. These measures allow transparency and a fair evaluation of differences that emerge between models.

EXECUTIVE SUMMARY

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

USA commercial landings of witch flounder generally increased from the early 1960s, peaking in 1984 at 6,658 mt (14.68 million pounds). Subsequently, landings declined and fluctuated between 2,000 and 3,000 mt (6.6 million pounds) until 2003, when landings further declined. Landings in 2015 were 492 mt (0.11 million pounds), the lowest in the time series. The majority of USA commercial landings come from the western Gulf of Maine and central basin and from the deeper waters north of the South Channel. Witch flounder landings occur throughout the year, with a slight increase during March-July months when witch flounder form dense pre-spawning aggregations. The majority of the landings are taken by otter trawl gear. Trends in landings by market category reveal a sharp decline in LARGE market category fish in the late 1980s.

Total discards from six fisheries ranged between 93 mt (2015) and 774 mt (1991) with a time series mean of 301 mt; coefficients of variation (CVs) ranged between 7% and 40%. The discards estimated in this assessment differ slightly from those estimated in 2015 Operational Assessment that include only three fisheries and did not include the entire stock area. Over the time series, discards average 13% of the total catch. Discard mortality is assumed to be 100%.

There is no recreational fishery for witch flounder.

Annual sampling from 1982 onward ranged between 11 and 133 mt per 100 lengths, exceeding the informal sampling threshold criterion of 200 mt per 100 lengths. However, during the 1990's, sampling was insufficient in some quarters and market categories and it was necessary to pool some quarters for some market categories. Sampling improved by quarter and market category in the late 1990's in which an average 34 mt per 100 lengths was obtained. Commercial age data were available for 1982 to 2015; landed witch flounder ranged in age between 3 and 33 years. The majority of landed fish were ages 6 to 9. The CVs for ages 6 through 9 are generally less than 15%. The age composition reveals a truncated age structure since the mid-1980s, a short-lived expansion in the early 2000s was not sustained.

Total commercial landings in 2015 were 493 mt, commercial discards were estimated to be 93 mt, for a total catch of 585 mt.

Landings per unit effort is described under TOR 2.

TOR 2. Present available federal, state, and other survey data, indices of relative or absolute abundance, recruitment, etc. Characterize the uncertainty and any bias in these sources of data and compare survey coverage to locations of fishery catches. Select the surveys and indices for use in the assessment.

The Northeast Fisheries Science Center (NEFSC) has conducted annual research vessel stratified random bottom trawl surveys during autumn since 1963 and during spring since 1968. The NEFSC survey samples offshore waters between 27 m to 366 m from Cape Hatteras NC to Canadian border. Witch flounder are generally distributed throughout the Gulf of Maine, along the northern edge and southern flank of Georges Bank, and southward along the continental shelf as far south as Cape Hatteras, NC. While NEFSC spring survey indices tend to be more variable

due to pre-spawning aggregations of witch flounder, spring and autumn indices generally display similar trends. NEFSC survey indices generally declined from the early 1960s to record low levels in the late 1980s and early 1990s. There was an increasing trend in the late 1990s to early 2000s followed by a declining trend. In 2015, the survey indices were below their time series means. Age-specific relative abundance from NEFSC spring and autumn 1980-2015 and spring 2016 surveys are derived using annual, seasonal age length keys. A truncation of the age structure occurred in the late-1980s; a slight expansion in age structure occurred in early 2000s but was not sustained.

In 2009, *FS/V Henry B. Bigelow* replaced the *R/V Albatross IV* as the principle vessel conducting the bottom trawl survey. There are no significant vessel, door, or net conversion factors for witch flounder in the NEFSC bottom trawl survey between 1963 and 2008; however, there are significant vessel conversion factors between the *FS/V Bigelow* and the *R/V Albatross IV*. The vessel conversion factor for numbers and weight (spring and autumn) is 3.2572 and has been applied to the 2009 through 2016 NEFSC bottom trawl surveys.

The NEFSC survey strata set was evaluated and no additional strata were identified as contributing to the total catch or improving the precision of the estimates of mean weight per tow. Therefore, the existing strata set of 22-30, 36-40 was retained.

A field experiment was conducted to estimate the relative detectability of witch flounder in the net used in the NEFSC bottom trawl survey and a net with a chain sweep by using a twin net configuration. The goal was to estimate the sweep efficiency of the NEFSC survey trawl. The WG accepted the witch flounder catchability (q) estimates (0.056 for Albatross and 0.291 for Bigelow) from the sweep study and recommended their use in the stock assessment to convert the NEFSC survey relative abundance indices into swept area numbers.

The Commonwealth of Massachusetts Division of Marine Fisheries (MA DMF) began a spring and fall inshore trawl survey in 1978 which complements the NEFSC survey in coastal Massachusetts waters in depths less than 27 m. Witch flounder indices of abundance and biomass from this survey have been presented in previous stock assessments as corroborative information. The percentage of positive tows is low (< 50%) in the spring and fall survey series. Fall survey catch per tow indices varies over the time series with a stratified mean less than 1 fish interspersed with peaks of 2 or 3 fish per tow occurring. Given the sparse length frequency in these seasonal surveys, NEFSC age length keys were not applied. This survey has a spatial coverage confined to the inshore waters of Massachusetts and therefore used as corroborative information only (not included in model).

The Atlantic States Marine Fisheries Commission (ASMFC) has conducted an annual northern shrimp survey during August in the Gulf of Maine since 1983, with catch data for witch flounder available since 1984 onward. There is a high percentage of positive tows in the survey series. Witch flounder age structures are not collected on this survey; however, age-specific relative abundance indices were derived by applying annual NEFSC spring and fall (combined) age-length keys to the length frequency data from the ASMFC survey. Although this survey has limited spatial coverage, much of the juvenile range is covered. This survey appears to provide useful information on relative to trends in abundance, distribution, and recruitment of witch flounder in the Gulf of Maine and could be considered as tuning indices in models

Maine Department of Marine Resources and New Hampshire Fish and Game Department (MENH) have conducted spring and fall bottom trawl surveys along the New Hampshire and Maine coast since fall 2000. Similar to the ASMFC survey, the MENH survey is considered useful in providing a pre-recruit index rather than to characterize the population as whole given the MENH survey catches few witch flounder greater than 30 cm (approximately age 4). Witch

flounder age structures have been collected on this survey; however, ageing has not occurred. Age-specific relative abundance indices were derived by applying annual, season-specific NEFSC age-length keys to the length frequency of the MENH survey. Although this survey has limited spatial coverage, much of the juvenile range is covered. This survey appears to provide useful information on relative to trends in abundance, distribution and recruitment of witch flounder in the Gulf of Maine and could be considered as tuning indices in models.

The WG considered other surveys (NEFSC winter survey; Canadian bottom trawl survey; NEFSC's Cooperative Research Program longline survey; and the MA DMF Industry-Based cod survey) and agreed that the limitations associated with these surveys would prevent their use in models.

The NEFSC spring and fall surveys, MA DMF spring and fall surveys, and ASMFC summer survey exhibit similar trends in which indices are above their time series mean prior to the late 1980s, below the mean during the late 1980s to mid-1990s and remain below or near the mean in the late 1990's and 2000s onward. The MENH seasonal surveys, which have the shortest time series of all surveys considered here, appear to be more variable than the other surveys.

The WG evaluated the **Dealer Report**-based trawl gear landings per unit effort indices (1964-2015 Landings per Day Fished; LPUE), **Vessel Trip Report (VTR)** trawl gear fishery catch per unit effort indices (1994-2015 Catch per Day Fished; CPUE), the **Observer Program** trawl gear fishery catch per unit effort indices (1989-2015 Catch per Day Fished; CPUE), and the NEFSC Cooperative Research Partners **Study Fleet** trawl gear catch per unit effort indices (2007-2015 Catch per Day Fished; CPUE) for their utility as indices of abundance in the witch flounder benchmark stock assessment. For the Dealer data, the WG noted that the utility of the LPUE series as the basis for an index of abundance is limited in that a) it includes only landings and not the total catch including discard, so that the resulting landings rates could be biased low relative to the true abundance of fish, and b) the use of only positive trips that land witch flounder may bias the landings rate as well. For the VTR data, the WG noted that given the low discard rate reported, the total catch reported may be biased low relative to the true total catch. For the Observer data, the WG noted that there may be a bias in fishermen's behavior when an Observer is aboard the vessel that changes the total catch rate in an uncertain and possibly inconsistent manner. The WG also noted that for the Dealer, VTR, and Observer data in particular, the ratio of Days Fished per Trip has 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. During the respective time series there have been a number of spatial, temporal, and catch limit regulatory changes that could bias the commercial fishery catch rate indices relative to the true abundance of witch flounder. The information on these regulatory changes is not part of the commercial fisheries databases and must be developed independently and integrated within the standardization models, and has not been included here. Given the changing regulatory landscape, there likely have been a multitude of changes in fishermen's behavior and strategies over time, and those have not been accounted for here. The increasing sophistication of fishing technology (fish-finding electronics, positioning electronics, fishing gear) is likewise not easy to include in fishery dependent catch rate modeling, and has not been attempted here. The Observer data were further explored to help inform which Dealer LPUE series (10%, 25%, and 40% directed trips) would be the most appropriate to use in a model sensitivity run. Summaries of the Observer haul and trip level data for observed hauls for which witch flounder was reported as the target species indicated no strong justification to select one series over another. For consistency with previous assessments,

the most directed series, LPUE 40%, was selected. The WG agreed to include the commercial LPUE 40% series as a sensitivity run of the assessment model.

TOR 3. Investigate effects of environmental factors and climate change on recruitment, growth and natural mortality of witch flounder. If quantifiable relationships are identified, consider incorporating these into the stock assessment.

An examination of available NEFSC trawl survey data indicated that while there have been some "statistically significant" changes in mean temp and depth of juvenile and adult witch flounder distribution, the slopes of the linear change are not large and movement has been generally within the stock definition strata set. There is also evidence of increases in the mean temperature and depth of the habitat occupied by witch flounder, but a model of changes in "available witch flounder habitat" did not indicate a trend over the time series. Modeling revealed no evidence of temperature influencing recruitment.

TOR 4. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series (integrating results from TOR-3 if appropriate), and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results and previous projections. Compare F's and SSB's that were projected during the previous assessment to their realized values.

> ***SARC-62 Opinions****: The VPA, ASAP, and SCAA age-structured models applied to data for the witch flounder fishery were found to have major retrospective patterns that prevented their use for status evaluation and determination of catch advice.*
>
> *Based on an empirical area swept method (that includes the results of the NEFSC survey efficiency experiment) conducted by the SAW and examined at the SARC, stock biomass has declined since 2002 and appears to be 16,181 mt Jan-1 surveyable biomass in 2016 and 14,563 mt exploitable biomass in 2016.*
>
> *Survey population biomass was converted to exploitable biomass using a factor of 0.90 based on examination of survey and fishery selectivity patterns from the base ASAP assessment.*
>
> *The relative exploitation rate needed to achieve the recent observed catch is approximately 0.05.*
>
> *The fishery landings and survey catch by age indicate truncation of age structure and a reduction in the number of old fish in the population.*
>
> *The time series of recruitment is not available from the empirical area swept method. All the surveys indicate the 2013 year class is relatively strong. However, this year class is not expected to fully recruit to the*

directed fishery until 2020 at age 7 and should start to appear in the large mesh otter trawl discards in 2017 at age 4. The absolute size of the 2013 year class is uncertain since the estimate is based on only two years of survey observations.

The VPA model used in the 2015 assessment was updated to incorporate re-estimated discards and an additional year of catch, survey, and maturity data. The updated VPA estimated average fishing mortality (F_{8-9}) to be 0.43 and spawning stock biomass to be 3,044 mt in 2015. The 2015 assessment estimated 2014 F_{8-9} to be 0.42 and SSB to be 3,129 mt; however, the updated VPA now estimates 2014 F to be 0.64 and SSB to be 2,480 mt. Including an additional year of data in the VPA model did not change the overall perception of stock. The updated VPA model exhibited a major retrospective pattern that was similar in direction and magnitude as the previous VPA models in which F is underestimated and SSB and recruitment are overestimated. The retrospectively adjusted $F_{8-9,rho}$ in 2015 is estimated to be 0.69 (above the F_{msy} proxy of 0.20). The retrospectively adjusted SSB_{rho} in 2015 is estimated to be 2,035 mt (16% of the SSB_{msy} proxy of 12,499 mt). **The updated VPA is not the base model for this assessment.**

A new statistical catch at age model (ASAP) was developed in this assessment to better account for catch and survey data uncertainty. The ASAP Run 9_5_v2 estimates fully selected fishing mortality (F_{full}) to be 0.16 and SSB to be 5,479 mt in 2015. Age 1 recruitment averaged 9.8 million fish over the 1982-2015 time series. The ASAP model exhibited a major retrospective pattern in which fishing mortality is underestimated and SSB and recruitment are overestimated. The retrospectively adjusted $F_{full,rho}$ is estimated to be 0.29 and SSB_{rho} to be 3,335 mt in 2015.

The WG considered numerous sensitivity runs of the two analytical models (ASAP and another statistical catch at age [SCAA] model, See Appendix B3) and considered several empirical approaches (minimum swept area biomass, replacement yield model, and an empirical approach) in case analytical modeling approaches did not provide defensible results. The WG considered the pros and cons of each approach for determining stock status and catch advice. The age structured models in the accepted final configurations provided calculated NEFSC survey q of approximately 4, exhibited major retrospective patterns requiring rho adjustments, and require very large increase in catch or natural mortality (M) to remove the retrospective errors. The WG noted that these data conflicts have been addressed in the previous accepted and current age structured models explicitly through a rho adjustment, without having to apply potentially implausible scalars of M or catch. **The WG recommends the ASAP Run 9_5_v2 as the preferred model in which to evaluate stock status and provide catch advice.**

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 BMSY, BTHRESHOLD, FMSY 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.

***SARC-62 Opinions**: The age-structured models applied to data for the witch flounder fishery from 1982-2015 were found to have major retrospective patterns that prevented their use for status evaluation and determination of*

catch advice. Therefore biological reference points are not available.

In 2015 Operational Assessment, the VPA model results were accepted as the basis for the biological reference points and stock status determination for witch flounder (NEFSC 2015). Reference points were calculated using the non-parametric yield and SSB per recruit/long-term projection approach adopted for the New England groundfish stocks (NEFSC 2008). 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 SSB_{msy} (spawning stock biomass at maximum sustainable yield), the cumulative density function was re-sampled from estimates of Age 3 between 1982-2013 to provide future recruitment estimates. Recruitment of witch flounder is not well determined by traditional stock recruitment relationships. Given this, an MSY proxy was used for reference points. $F_{40\%}$ is the proxy used for the overfishing threshold (F_{MSY}). This is consistent with the choice of proxy in the previous assessments (NEFSC 2012, NEFSC 2008). The biological reference points estimated in the 2015 assessment were F_{msy}=$F_{40\%}$=0.28, SSB_{msy}=9,473 mt, and MSY=1,957 mt.

In this assessment, the yield and SSB per recruit calculations (NFT 2013a) used the most recent five year (2011-2015) averages for mean weights, fishery partial recruitment pattern, and maturity ogive. For the estimation of MSY and SSB_{msy}, the cumulative density function was re-sampled from estimates of Age 3 (VPA) or Age 1 (ASAP) between 1982-2015 to provide future recruitment estimates (Table B66 for VPA Run A2 input and Table B67 for ASAP Run 9_5_v2 input). Proxy reference points based on VPA Run A2 and ASAP Run 9_5_v2 are given in Table B68. Stochastic projections at Fmsy=F40% (0.19) were used to determine new biomass-related reference points (proxies for both SSB_{msy} and MSY) based on F40% (Table B69).

Using Run VPA Run A2 and F40%=0.20, biological reference points are estimated to be 12,499 mt for SSB_{msy}, 6,250 mt for $SSB_{threshold}$, and 1,512 mt for MSY (Table B69).

Using ASAP Run 9_5_v2 and F40% = 0.19, biological reference points are estimated to be 12,747 mt for SSBmsy, 6,374 mt for $SSB_{threshold}$, and 1,998 mt for MSY (Table B69). **The 2016 WG recommends the ASAP Run 9_5_v2 model results as the basis for the new biological reference points and stock status determination for witch flounder.**

TOR 6. Evaluate stock status with respect to the existing model (from previous peer reviewed accepted assessment) and with respect to a new model (or possibly models, in accord with guidance in attached "Appendix to the SAW Assessment TORs") developed for this peer review. In both cases, evaluate whether the stock is rebuilt.
a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the updated BRP estimates.
b. Then use the newly proposed model (or possibly models, in accord with guidance in "Appendix to the SAW Assessment TORs") and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-5).

> ***SARC-62 Opinions****: The status of the witch flounder stock is unknown with regards to biological reference points. The age-structured models applied to data for the witch flounder fishery from 1982-2015 were found to have major retrospective patterns that prevented their use for status evaluation and determination of catch advice.*

Based on an empirical analysis (referred to as "empirical area swept method") conducted by the SAW and examined at the SARC, stock biomass has declined since 2002 and appears to be 16,181 mt Jan-1 surveyable biomass in 2016 and 14,563 mt exploitable biomass in 2016. The relative exploitation rate needed to achieve the recent observed catch is approximately 0.05. The SARC-62 did not fully evaluate the empirical approach during the peer review, and therefore expected additional evaluation of the method, calculations, and results to be made following the peer review.

The fishery landings and survey catch by age indicate truncation of age structure and a reduction in the number of old fish in the population.

a) The existing VPA model updated with new data indicates witch flounder stock was overfished and overfishing occurred in 2015. This determination was made on retrospectively adjusted $F_{full,rho}$ and SSB_{rho} in 2015. The retrospective adjustment was required but did not lead to a change in status. The biological reference points used were F_{MSY}=$F_{40\%}$=0.20, SSB_{msy}=12,499 mt, and MSY=1,512 mt. The 2015 $F_{8\text{-}9}$ is estimated to be 0.43 (210% of $F_{40\%}$) and the retrospectively adjusted 2015 $F_{8\text{-}9,rho}$ to be 0.69 (337% of $F_{40\%}$). The 2015 SSB is estimated to be 3,044 mt (24% of SSB_{msy}) and the retrospectively adjusted 2015 SSB_{rho} to be 2,035 mt (16% of SSB_{msy}).

b) The new ASAP model indicates witch flounder stock was overfished and overfishing occurred in 2015. This determination was made on retrospectively adjusted $F_{full,rho}$ and SSB_{rho} in 2015. The retrospective adjustment was required and lead to a change in status of overfishing. The biological reference points used were F_{MSY} proxy=$F_{40\%}$=0.19, SSB_{msy}=12,747 mt, and MSY=1,998 mt. The 2015 F_{full} is estimated to be 0.16 (83% of $F_{40\%}$) and the retrospectively adjusted 2015 $F_{full,rho}$ to be 0.29 (154% of $F_{40\%}$). The 2015 SSB is estimated to be 5,479 mt (43% of SSB_{msy}) and the retrospectively adjusted 2015 SSB_{rho} to be 3,335 mt (26% of SSB_{msy}).

TOR 7. Develop approaches and apply them to conduct stock projections.

a. Provide numerical annual projections (3 years) and the statistical distribution (e.g., probability density function) of the catch at FMSY or an FMSY proxy (i.e. the overfishing level, OFL) (see "Appendix to the SAW Assessment TORs"). 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, magnitude and 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. Identify reasonable projection parameters (recruitment, weight-at-age, retrospective adjustments, etc.) to use when setting specifications.

c. Describe this stock's vulnerability to becoming overfished, and how this could affect the choice of ABC. The choice takes scientific uncertainty into account (see "Appendix to the SAW Assessment TORs").

SARC-62 Opinions: *The empirical area swept method was used to provide an F_{MSY} proxy for 2017 based on a relative exploitation rate. The SARC-62 did not fully evaluate the empirical approach during the peer review, and therefore expected additional evaluation of the method, calculations, and results to be made following the peer review.*

Stochastic projections were made to provide forecasts of stock size and overfishing limit (OFL) catches in 2017-2019 consistent with the 2016 SAW 62 assessment biological reference points. The WG conducted one projection by assuming recent patterns in the fishery and the most recent maturity ogive. A major source of uncertainty is the retrospective pattern. The past projection was optimistic (outside the confidence bounds for the rho adjusted biomass). Subsequent projections will be conducted by the Groundfish Plan Development Team, when 2016 catch is determined.

The 2016 stock assessment indicates that the stock was below the biomass target and being fished above the fishing mortality threshold in 2015. Witch flounder is slow growing and late maturing and is vulnerable without stringent management regulations in place.

TOR 8. Evaluate the validity of the current stock definition, taking into account what is known about migration, and make a recommendation about whether there is a need to modify the current stock definition for future stock assessments.

Witch flounder are assessed and managed as a unit stock within USA waters; this includes the Gulf of Maine and Georges Bank region as well as areas to the south as far as Cape Hatteras, North Carolina. No tagging studies have been conducted on witch flounder in Gulf of Maine-Georges Bank region. No studies on stock structure have been conducted in the Gulf of Maine-Georges Bank region. A study found biological and yield characteristics of deep water (depth > 366 m) witch flounder south and west of Georges Bank to be slower in growth and the otolith structure was different than the Gulf of Maine-Georges Bank witch flounder suggesting a separate population. Studies of witch flounder on the Grand Banks indicate small independent subpopulations in that region.

Until tagging and/or stock structure analyses are undertaken that reveal subpopulations, the current stock definition should be retained for future stock assessments. Possible future research recommendations include: ageing larval fish to determine transport time; tagging and/or stock structure analyses.

TOR 9. Review, evaluate and report on the status of research recommendations from the last peer reviewed benchmark stock assessment. Identify new research recommendations.

SARC-62 Opinions: *The SARC recommends that witch flounder be put on a research track to address survey catchability and retrospective problems. Such an examination should include other stocks in the region exhibiting major retrospective patterns and survey catchability issues.*

Of the three identified research recommendations made since the last peer reviewed benchmark witch flounder stock assessment (2008 GARM), progress has been made on two recommendations (splitting the survey tuning series and exploring the dome selectivity in the catch and survey). No progress was made on examining the trends in mean weights at age and the possible underlying factors.

The WG developed nine new research recommendations.

WORKING GROUP PROCESS

The Stock Assessment Workshop (SAW) Witch flounder Working Group (WG) met during 13-16 September (Data Meeting) and 4-7 October 2016 (Models Meeting) at the Northeast Fisheries Science Center (NEFSC) to develop the benchmark stock assessment of witch flounder through 2015. The following people were present (in person or via webinar) for all or part of working group discussion.

Jessica Blaylock	NEFSC
Elizabeth Brooks	NEFSC (WG member)
Russell Brown	NEFSC
Doug Butterworth	Univ. of Cape Town (WG member)
Steve Cadrin	SMAST, NEFMC SSC
Richard Canastra	
Jamie Cournane	NEFMC staff (WG member)
Greg DeCelles	MA DMF
Libby Etrie	NEFMC
Vito Giacalone	Northeast Seafood Coalition
Mary Hudson	GMRI
Kevin Friedland	NEFSC
Chris Legault	NEFSC, NEFMC SSC
Brian Linton	NEFSC
John Manderson	NEFSC
Katey Marancik	NEFSC
Michael Martin	NEFSC
Dave McElroy	NEFSC
Tim Miller	NEFSC
Tom Nies	NEFMC staff
Paul Nitschke	NEFSC
Loretta O'Brien	NEFSC
Jackie Odell	Northeast Seafood Coalition
Mike Palmer	NEFSC (WG member)
Charles Perretti	NEFSC
Rebecca Rademeyer	
Maggie Raymond	Associated Fisheries of Maine
David Richardson	NEFSC
Chris Roebuck	F/V Karen Elizabeth
Katherine Sosebee	NEFSC
Mike Simpkins	NEFSC
Mark Terceiro	NEFSC (WG member and chair)
Harvey Walsh	NEFSC (WG member)
Jim Weinberg	NEFSC
Susan Wigley	NEFSC (WG member, assessment lead)
Brooke Wright	SMAST

INTRODUCTION

This assessment of witch flounder in the Gulf of Maine-Georges Bank region and southward presents a benchmark analytical assessment for the 1982-2015 period, estimates 2015 fishing mortality and spawning stock biomass for stock status, updates biological reference points, and provides short-term projections of median catch corresponding to the Overfishing Limit (OFL) for 2017-2019..

Biology

Witch flounder (*Glyptocephalus cynoglossus*) or grey sole is a deep water boreal flatfish occurring on both sides of the North Atlantic. In the Northwest Atlantic, witch flounder are distributed from Labrador to Georges Bank and in the continental slope waters southward to Cape Hatteras, North Carolina. In US waters, the species is predominately in the Gulf of Maine-Georges Bank region (defined by statistical areas [SA] 511-515, 521-522, 525-526, and 561-562; Figure B1), and, in the absence of any stock structure information, is assumed to comprise a single stock unit defined by SA 511-515, 521-22, 525-526, 533-539, 541-543, 561-562, 611-639.

Witch flounder do not undertake seasonal migrations (Bigelow and Schroder 1953; Burnett et al. 1992); however, witch flounder appear to form pre-spawning aggregations in the spring in the western and northern portions of the Gulf of Maine. Spawning is protracted from April to November, with peak during May – August (Bigelow and Schroeder 1953; Colton et al. 1979; Burnett et al. 1992). Using NEFSC bottom trawl survey data collected between 1963 and 1981, Burnett et al. (1992) evaluated witch flounder distribution relative to depth and temperature. Burnett et al. (1992) found that the distributions of juvenile and adult witch flounder were not significantly different for any season with respect to bottom temperature. However, distribution by depth differed for juveniles and adults for all seasons. Burnett et al. (1992) found that adults maintained an annual mean depth of 147 m (fathoms) while juveniles were found at shallower depths in winter and spring and greater depths in summer and fall and this seasonal-difference was attributed to differences in prey distribution associated with differences in diet. Burnett et al. (1992) did note that juveniles do not appear to be adequately sampled by the NEFSC survey gear.

There are no tagging data for witch flounder.

Age and Growth

Age and growth studies (including age validation) by Burnett et al. (1992) indicate that witch flounder is a long-lived, slow-growing, late-maturing species. Witch flounder exhibit sexual dimorphism, with females living longer and growing larger than males. Burnett et al. (1992) derived growth parameters from 1,120 fish that ranged age from 0 (5.3 cm) to 24 (56.8 – 62.1 cm) for males and females using a von Bertalanffy growth model and length-at-age data collected from the NEFSC bottom trawl surveys during 1980 – 1984. Estimated parameters were: for males (n = 626), L_{inf} = 58.04 and k = 0.15; for females (n= 571), L_{inf} = 61.99 cm and k = 0.15.

The largest and oldest fish sampled by the NEFSC surveys were a 73 cm age 24 fish sampled in the 1985 spring survey and a 64 cm age 30 fish sampled in the 1985 fall survey.

In the current assessment, a von Bertalanffy growth model and the 1982 – spring 2016 NEFSC trawl survey length-at-age data were used to derive growth parameters by sex (in which small fish of undetermined sex were grouped with both males and females) and combined sexes. The survey time series data provided parameters for males (n = 5,744) of L_{inf} = 60.7 cm, k = 0.12, with maximum length of 65 cm and age of 25; parameters for females (n = 6,070) of L_{inf} = 64.8 cm, k = 0.12, with maximum length of 73 cm and age of 30; and parameters for sexes combined (n =12,946, including small fish of undetermined sex) of L_{inf} = 59.2 cm and k = 0.15.

The growth curves are generally similar for sexes through about 40 cm and age 7, where the curves begin to diverge due to dimorphic growth (Figure B2). In the most recent stock assessments, ages are grouped together for ages 11 and older (age 11+ "plus group").

Age and growth information is also available from NEFSC commercial port sampling from 1982 – 2015. In the commercial landings, a 70 cm age 14 was sampled in 1986 and 61cm age 33 was sampled in 2005. Commercial length-at-age data are not collected by sex. The 1982-2015 commercial data provided parameters for sexes combined (n=29,865) of L_{inf} = 63.7 cm and k = 0.12. Fish taken from commercial landings are generally the faster growing individuals in the population than those sampled in the NEFSC surveys (Figure B2).

Tests to evaluate ageing precision are conducted by the NEFSC Fishery Biology Program on a regular basis (survey cruise or commercial quarter). There was a change in age reader following the GARM 2008 witch flounder assessment (NEFSC 2008). An extensive age reader comparison (N=251) was performed, resulting in 89% agreement between readers, with a low coefficient of variation (CV) and no bias. Precision tests for the current age reader are quite similar to those of the previous age reader, ranging between 85% and 94% agreement for survey samples, and 76% to 97% from commercial samples. Test results and statistical methods used are given at: http://www.nefsc.noaa.gov/fbp/QA-QC/index.html .

Length-Weight Relationship

Lux (1969) derived length-weight parameters from 188 witch flounder taken from 55 fathoms of water south of Nantucket Shoals in March 1962 by commercial trawlers and research vessels. Lux noted that the length-weight relationship was similar to that of Powles (1967) for witch flounder caught in February of southeast Nova Scotia.

Wilk et al. (1978) derived length-weight parameters, by sex, from 204 witch flounder between sizes of 12 and 55 cm taken from the New York Bight area during 1974 – 1975 from 700 randomly selected otter trawl stations.

Burnett et al. (1992) derived length-weight parameters from 1,652 witch flounder collected between March 1980 and June 1983 during NEFSC bottom trawl survey in the Gulf of Maine and Georges Bank region as well as two commercial sea sampling trips on Stellwagen Bank and in South Channel in spring of 1983. Burnett et al. (1992) reported the length-weight relationships derived in his study closely compared with those of Lux (1969) and Powles (1967).

Wigley et al. (2003b) updated the length-weight parameters used in audits of the NEFSC trawl survey data, using individual length and weight information from 3,778 fish for 1992-1999. There was no significant difference between females and male, however there was significant difference between winter/summer and autumn.

Sherman (ME DMR, pers. comm, July 2016) derived length-weight parameters using Maine-New Hampshire (MENH) inshore trawl survey data from spring 2003 and fall 2005 through 2015.

In the current assessment, individual length and weight information from 8,557 fish

(3,572 males, 4,985 females) sampled during 1992-2016 NEFSC bottom trawl surveys were used to estimate length-weight parameters for comparison with the earlier studies to judge whether changing from the Burnett et al. (1992) parameters would be justified.

A comparison among these studies indicates very little difference in the estimated length-weight relationships from Lux (1969), Wilk et al. (1978), Burnett et al. (1992), Wigley et al. (2003b), and the current examination of the NEFSC trawl survey data. The curves are almost identical through length 42 cm, below which 72% of the fishery catch has occurred, on average, since 1982 (Figures B3 and B4). A 42 cm fish represents roughly an age 7 fish.

WG recommendation: Based on the consistency of these L-W relationships through the majority of the length range of the fishery catch, the Burnett et al. (1992) length-weight parameters were retained for this assessment.

Relative Condition Factor

Individual fish weight collection for witch flounder began on NEFSC surveys in fall 1992. The NEFSC spring and autumn trawl survey sample data were examined for trends in relative condition factor by season using the ratio of observed weight to predicted weight. There are no long-term trends in relative condition factor by season (Figure B5).

Maturity

Witch flounder maturity observations have been collected in the NEFSC research bottom trawl survey since 1977. Macroscopic methods are used to determine maturity stage (see O'Brien et al. 1993 for criteria and maturity stages). The NEFSC spring surveys were used for maturity analyses as these surveys occur closest to and prior to spawning (Halliday 1978). Annually, the number of observations range between 27 and 402 over the 1982 to 2016 time period (Figure B6). Between 1987 and 1993, less than 100 fish were sampled during the spring surveys due to low witch flounder abundance.

Using a method developed and applied for Georges Bank cod (L. O'Brien, NEFSC, pers. comm.), a logistic regression and a 5-year moving window of pooled maturity data from the NEFSC spring survey is used to estimate median age (A_{50}) and length (L_{50}) at maturity. For example, the proportion mature in 1982 was estimated using NEFSC spring maturity data from 1980, 1981, 1982, 1983, and 1984. The 2015 maturity ogive was assumed to be equal to the 2014 ogives, the more recent annual maturity ogive with a full five years of data. In addition to annual maturity ogives, a single ogive using maturity data form the entire time series (1980 to 2016) was also calculated.

For females, median A_{50} has ranged between ages 4.2 and 6.3 years (see table below) and L50 between 29 cm and 38 cm (Table B1). The time series median age is 5.3 years for females and 3.7 years for males and median length is 33.8 cm for females and 27.1 cm for males (Table B2). The NEFSC survey maturity-at-age has remained generally stable over time; however, most recent annual estimates of median A_{50} and L_{50} are slightly below the long term mean (Figure B7). The 2014 median maturity for female witch flounder (using data from 2012 through 2016) is 5.0 years and 32.7 cm (Table B1; Figure B7).

Witch flounder maturity observations have been collected in the Massachusetts Division of Marine Fisheries (MA DMF) during the late 1980s and early 1990s; however, seasonal sample sizes are too sparse to estimate median maturity (Figure B6).

A total of 3, 448 samples (1,588 female and 1,859 male) have been collected in the Maine-New Hampshire inshore survey during spring 2001-2008 and fall 2005-2015 surveys. S. Sherman (ME DMR, pers. comm. July 2016) estimated L_{50} to be 32 cm for females and 23 cm for males (Figures B8 and B9). These estimates are slightly lower than NEFSC estimates.

WG recommendation: *In keeping with the approach from the previous two benchmark assessments (SARC 29 and GARM 2008), annual maturity ogives based on a 5-year moving window of NEFSC 1980- 2016 spring survey data are used with the fishery catch at age to compute SSB in the assessment model. The 5-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.*

Predators and Prey

The NEFSC trawl survey foods habits 1973-2015 database was investigated to identify the most frequent predators of witch flounder. Witch flounder was identified to species as a prey item in 81 predator stomachs of 15 fish species (B. Smith, NEFSC, pers. comm. August 2016). White hake was the predator in 28 cases (35%), followed by goosefish (13 cases; 16%), spiny dogfish (11 cases; 14%), cod (5 cases; 6%) and thorny skate (5 cases; 65). The remaining predators (10 cases; each 4% or less) included Atlantic halibut, red hake, pollock, longhorn sculpin, silver hake, smooth skate, sea raven, barndoor skate, winter skate, and summer flounder. The data are insufficient to calculate total absolute predator consumption of witch flounder.

Witch flounder diet is mainly polychaetes, accounting for more than 70% of the diet (Maurer and Bowman, 1975). The remaining components of the diet include echinoderms (mainly sea cucumbers), crustaceans (mainly amphipods), and mollusks (mainly squid). Bowman and Michaels (1984) and Bowman et al. (2000) have documented similar findings. See Link and Almeida (2000) for details on the food habits sampling program.

Instantaneous natural mortality rate (M)

The instantaneous natural mortality rate (M) for witch flounder has been assumed to be 0.15 in all previous stock assessments based upon an analysis by Burnett (MS 1987) in which estimates of instantaneous total mortality (Z; calculated from catch curves derived from NEFSC spring survey data collected during 1978 -1984, a time period before closed area and effort constraints) were regressed against commercial effort (identified as 'directed' towards witch flounder) in the previous year (calculated using landings per unit effort data of ton class 3 vessels). The slope, q, is the directed catchability coefficient and the intercept is an estimate of M (Gulland, 1983). The regression of annual Z estimates from 1978 – 1984 against commercial effort resulted in M = 0.16. This result was similar to M= 0.15 used by Halliday (1973) for Scotian Shelf witch flounder.

Alternative methods (age constant longevity based and age constant growth/life history based) for estimating M were conducted via an Excel spreadsheet developed by L. Lee and K. Drew (ASMFC, pers. comm. 2016). Longevity based estimators of M are sensitive to critical assumptions which include the value of p, or the small proportion of the population surviving to a given maximum age (t_{max}), and the maximum observed age under no exploitation conditions. Using a maximum age of 33 years, and p = 0.05 and p = 0.015 as the proportion of individuals that survive to t_{max}, age constant longevity based estimates of M range from 0.10 to 0.20 with an

average of 0.16 (Table B2a). Using L_{inf} = 619.9 mm, K= 0.1482, and mean water temperature of 7 degrees C, the age constant growth/life history based estimates of M range from 0.22 to 0.32 with an average of 0.25 (Table B2a). An overall average of M from all alternative methods was 0.20.

WG recommendation: Given the maximum size and age of 73 cm and 33 years, the range of M (0.13 – 0.20) estimated by the empirical methods (based on maximum age) as well methods based on growth/life history, the WG decided there was no compelling reason to change from the previous assumption of M and adopted a value of M=0.15 for all ages and years in the SAW 62 models.

Fishery Management

Witch flounder has been managed under the New England Fishery Management Council's Multi-species Fisheries Management plan since 1987. A brief summary of groundfish management regulations affecting witch flounder is presented in Table B3. Significant changes in regulations include increased minimum size in 1983 and 1987 followed by a decreased minimum size in 2013; increases in mesh size in 1982, 1983, 1994, 1999 and changes to mesh configurations in 1999 and 2002; effort reductions in 1996 and 2002; implementation of closed areas in 1994 and 1998; and implementation of annual catch limits and catch shares in 2010. The western Gulf of Maine area closure, Cashes Ledge area closure, the seasonal rolling closures, and the whiting small mesh exemption area all occur in the Gulf of Maine and overlap the witch flounder distribution (Figures B10-B12). A detailed description of groundfish regulations between 2001 and August 2015 are given in Table B4. The Northeast Multispecies regulations are available online at
https://www.greateratlantic.fisheries.noaa.gov/sustainable/species/multispecies/index.html
NE Multispecies Federal Register documents from 1994 to 2016 are available at:
http://www.greateratlantic.fisheries.noaa.gov/sustainable/species/multispecies/multifr/index.html

Management regulations for the northern shrimp fishery also impact witch flounder (Table B5): significant changes in the northern shrimp fishery include a monthly 10% by-catch limit which restricted the possession of groundfish to 10% by weight of shrimp in the mid-1980's to early 1990s; the implementation of the Nordmore grate to exclude groundfish in 1992; and fishery closures in 1978, 2014 - 2016.

Previous Stock Assessments

Burnett and Clark (1983) provided a preliminary assessment of witch flounder in the Gulf of Maine-Georges Bank region based on analysis of commercial landings and effort and NEFSC and MA DMF research vessel survey data. They reported that witch flounder was an intended component of the mixed trawl fishery and since the 1960s, USA landings had ranged between 1,000 and 6,000 mt and averaged 3,000 mt; total effort had increased and landings of small fish also appeared to be increasing. Survey indices suggested declines in abundance and biomass in the early 1980s with a consistent declining trend following peak landings in the early 1970s. Declines in abundance and biomass were expected in the immediate future if current levels of harvest rates were maintained.

In the early 1980's, Burnett was working on a master's thesis on the population biology of the witch flounder in the Gulf of Maine-Georges Bank region (Burnett MS, 1987; Burnett et

al. 1992) and was in the midst of conducting the first work on witch flounder in the Gulf of Maine-Georges Bank region. Since 1983, stock assessments for witch flounder have been conducted periodically as part of the Northeast Regional Stock Assessment Workshop.

In 1985, at 1st SAW (NEFC 1985) a summary of important assessment issues (terms of reference) were compiled for witch flounder that included: "*Problems of mixed-species effort confounds CPUE analyses for this species; discards of young witch should be examined more fully in the redfish and northern shrimp fisheries; estimates of Z are lacking; valid recruitment estimators have not been developed; ageing past c.a. 9 years is unreliable with scales, but ages 3-9 account for about 90% of landings. Ageing techniques need to be finalized and validated; evaluation of growth and mortality rates (natural, fishing, total) needs to be completed; studies of distribution by size and age may help to elucidate the recruitment mechanisms for this species; recent shifts of large vessel effort to witch may be due to a number of factors including displacement from traditional grounds, and decreases in abundance of other species."(NEFC 1985)*

In 1986 at 2nd SAW (NEFC 1986), witch flounder landings, catch and effort, and research vessel indices summarized by Burnett and Clark (1983) were updated through 1985. The witch flounder stock abundance "*appeared to be declining, based on declining indices of commercial catch-per unit effort and recent general downward trends in research survey abundance indices. Current historically high landing levels do not appear sustainable*" (NEFC 1986). It was noted that preliminary estimates of Z have been developed, ageing techniques using otoliths have been finalized and validated as far as possible, growth rates had been derived, and distribution of ichythoplankton, juveniles and adults has been completed as part of Burnett's thesis.

In 1994, the first virtual population analysis (VPA) was conducted by Wigley and Mayo (1996) and peer reviewed at the 18th SAW (NEFSC 1994). The 1994 assessment included commercial landings and discards at age data through 1993 and age-specific NEFSC spring and autumn abundance indices and age-specific landings per unit effort (LPUE) indices used for calibration. The stock was overexploited and at a low biomass level. Fishing mortality was estimated to be 0.45, and spawning stock biomass (SSB) had fallen to record-low levels in 1993, and the age-structure had become severely truncated.

In 1995 at 21st SAW (NEFSC 1996a 1996b), the northern demersal complex (25 stocks in the Gulf of Maine, Georges Bank and Southern New England region, including witch flounder) were examined using NEFSC and MA DMF research vessel bottom trawl survey biomass trends by species and stock and in aggregate for four species groups. Additional analyses were developed to examine habitat preference, spatial variability, concentration effects, and relationship between survey indices and VPA estimates of mature biomass. Of the 25 stocks examined, 18 (72%) have exhibited significant declines in biomass over the last 10-15 years. The biomass of 13 of these stocks is at or near record-low levels; the biomass of five other stocks remains well below historic levels.

In 1999 at 29th SAW (NEFSC 1999) an analytical assessment (calibrated VPA) of commercial catch (landings plus discards) at age data through 1998 was conducted. Additionally, an exploratory Stock Synthesis analysis was conducted and generally confirmed the VPA results (Wigley et al. 1999). A retrospective analysis was conducted and results indicated that: F was overestimated during 1991-1995, however, average F appeared to be underestimated in 1996-1997 (later described as a "flip-pattern"); estimates of SSB were overestimated in 1994 -1997, yet underestimated in 1991-1993 ("flip-pattern"); and recruitment was consistently estimated with the notable exception of two year classes (1992 and 1993) that were considerably overestimated. Yield and SSB per recruit analyses were updated with an assumed M = 0.15. Revised biological reference points (BRPs) were estimated: F_{msy} = 0.106, B_{msy} 25,000 mt, and MSY = 2,684 mt. Witch flounder was found to be near target biomass (above the overfished

threshold and near B_{msy}) and fishing mortality was projected to decline (slightly lower than the overfishing threshold) in 1999 based on part-year catch statistics.

In 2002, the Groundfish Assessment Review Meeting (GARM I ; NEFSC 2002) was held in which 20 groundfish stock assessments were reviewed, including witch flounder. An analytical assessment (calibrated VPA) of commercial catch (landings plus discards) at age data through 2001 was conducted using the same formulation as the 1999 stock assessment. The retrospective analysis indicated that average fishing mortality (F) was overestimated in the early to mid-1990's and underestimated in the late 1990s, but the 2000 F estimate was initially overestimated; SSB was consistently overestimated since 1994; and a pattern of relatively consistent estimates of the number of age 3 recruits, with the notable exception of the 1992, 1993 and 1996 year classes, which were overestimated. Based on yield and SSB per recruit analyses and the arithmetic mean of the VPA age 3 recruitment (NEFSC 2002), BPRs were estimated to be SSB_{msy} = 19,900 mt; $F_{msy} = F_{40\%}$ = 0.164, and MSY = 2,990 mt. In 2001, spawning stock biomass was slightly above ½ SSB_{msy} (9,950 mt), the overfished threshold, and fishing mortality (F = 0.45) was three times higher than F_{msy}, the overfishing threshold; therefore, witch flounder was not overfished but overfishing was occurring in 2001.

In 2003 at 37th SAW (NEFSC 2003a NEFSC 2003b; Wigley et al. 2003a), an analytical assessment (calibrated VPA) of commercial catch (landings plus discards) at age data was conducted through 2002. An alternative assessment using a statistical catch at age (SCAA) model was also evaluated. Despite differences in model configurations and estimation approach, the SCAA generally confirmed point estimates and trends in the VPA results. Biological reference points were re-estimated as in the NEFSC 2002 work. The yield and SSB per recruit analyses indicate $F_{40\%}$ = 0.23, MSY = 4,375 mt, and SSB_{msy} = 25,240 mt with an assumed M = 0.15. Fully recruited fishing mortality in 2002 was 0.41 (exploitation rate = 31%), nearly double F_{MSY} = 0.23, and SSB was estimated to be 18,300 mt in 2002, 72% of SSB_{msy} (25,200 mt). The witch flounder stock was not overfished, but overfishing was occurring in 2002.

In 2005, the GARM II (NEFSC 2005) was held in which 19 groundfish stocks were assessed and reviewed, including witch flounder. An analytical assessment (calibrated VPA) of commercial catch (landings plus discards) at age data through 2004 was conducted. VPA formulation was the same as the 2003 assessment. Similar to the 2003 stock assessment, the retrospective analysis indicated that: average F was underestimated in the late 1990s and early 2000s; SSB was consistently overestimated; and a pattern of relatively consistent estimates of the number of age 3 recruits, with the notable exception of the 1992, 1993 and 1996 year classes, which were considerably overestimated. Biological reference points were updated based on yield and SSB per recruit analyses and the arithmetic mean of the VPA age 3 recruitment NEFSC 2005), the biological reference points are: SSB_{msy} = 25,248 mt, $F_{msy} = F_{40\%}$ = 0.23, and MSY = 4,375 mt. Witch flounder was not overfished and overfishing was not occurring in 2004.

In 2008, GARM III (NEFSC 2008), consisting of a series of meetings, was held during November 2007-August 2008 in which 19 stocks were reviewed, including witch flounder. An analytical assessment (calibrated VPA) of commercial catch (landings plus discards) at age data through 2007was conducted. VPA analyses were performed for a BASE case and a SPLIT case, where the survey time series were split between 1994 and 1995. This time split corresponded to changes in the commercial reporting methods as well as other regulatory management changes. NEFSC spring and autumn relative abundance indices at age were transformed into swept area absolute abundance indices and used as tuning indices to explore changes in survey catchabilities (q) between the BASE RUN and the SPLIT RUN. In the BASE RUN, the swept area survey q's ranged between 0.02 and 0.21. In the SPLIT RUN, the 1982-1994 series q's ranged between 0.01

and 0.24 and the 1995-2007 q's ranged between 0.05 and 0.30. The magnitude and pattern of increasing survey catchabilities at age for younger fish and a general level pattern at older ages in the BASE and SPLIT runs appear reasonable. The causes of the increased q's between the 1982-1994 and 1995-2007 series in the SPLIT RUN remain unknown. Both VPA formulations have retrospective patterns: the BASE RUN has a consistent pattern while the SPLIT RUN exhibits a 'flip' (change in direction) pattern. The combination of: (1) the contraction of the age structure observed in the survey indices at age and the commercial catch at age; (2) the low NEFSC survey abundance and biomass indices in recent years; and (3) the magnitude of the 2004 year class at age 3 relative to the age 3 abundance indices over the entire time series, indicates a strong 2004 cohort but not exceptional year class, all seem to suggest that the VPA SPLIT RUN more accurately characterizes the witch flounder population. Additionally, the Mohn rho statistics of the VPA SPLIT RUN indicate that the respective pattern is less severe than the VPA BASE RUN. The VPA SPLIT RUN is selected as the final run to use for biological reference point calculations and for stock status determination. The retrospective analysis of the VPA SPLIT RUN indicates a pattern of overestimation of average F prior to 2003 and then underestimation for average F from 2003 onward. A similar "flip" pattern is also evident for SSB, underestimated prior to 2001 and then overestimated from 2001 onward. The retrospective analysis for Age 3 recruits indicates an overestimation prior to 2001 and then an underestimation from 2002 onward. The SPLIT RUN Mohn rho statistics for F, SSB, and Age 3 was -0.02, 0.43 and -0.13, respectively. Biological reference points were estimated via a yield and SSB per recruit analysis using 5-year (2003-2007) averages for partial recruitment, stock weights, catch weights and maturity (2004-2008). Based on yield and SSB per recruit analyses, a proxy of F_{msy} is $F_{40\%MSP}$ = 0.20 for SPLIT RUN. Mean Age 3 recruitment was not used. Long-term (100 year) stochastic projections (AGEPRO v3.1.3) were performed to estimate spawning stock biomass and MSY under equilibrium conditions. The same partial recruitment vectors, mean weights at age, and maturity vectors used in the yield and SSB per recruit analysis were also used in the projections. A constant F scenario was used ($F = F_{msy} = 0.20$). Estimates of Age 3 recruitment used in the projections were derived by re-sampling the cumulative density function based on the empirical observations during 1982 to 2008 (1979 to 2005 year classes) from the SPLIT RUN. The proportions of F and M which occurs before spawning equals 0.1667 (March 1); M equals 0.15. Biological references points for the SPLIT RUN were estimated to be: SSB_{msy} = 11,447 mt; MSY = 2,352 mt. Based on the VPA SPLIT run, the 2007 SSB = 3,434 mt, 30% below SSBMSY (11,447 mt) and 2007 fishing mortality was 0.29, 45% above F_{msy} (F=0.20); therefore, witch flounder was overfished and overfishing occurred in 2007.

The Northeast Groundfish Stock Updates, 2012 Update (NEFSC 2012) was held in February 2012 in which assessments for 13 groundfish stocks were updated and reviewed through 2010, including witch flounder. A small working group was convened prior to the 2012 meeting to discuss candidate models and the application of length-based conversion factors for witch flounder. Length-based conversion factors have not been established for witch flounder. The group concluded that further work was needed before length-based factors were applied; however, the group agreed that a sensitivity run should be conducted using survey indices adjusted with length-based conversion factors derived from a second order polynomial. The VPA formulation was the same the previous assessment (SPLIT RUN configuration). The VPA runs had similar retrospective patterns indicating that average F was underestimated, SSB was overestimated, and Age 3 recruitment exhibited a "flip" (change in direction) pattern (overestimated than underestimated). For this assessment, yield and SSB per recruit analyses were updated using 5-year (2006-2010) averages for partial recruitment, stock weights, catch

weights and maturity (2007-2011). Based on yield and SSB per recruit analyses, a proxy of F_{msy} is $F_{40\%MSP}$ = 0.27 for the SPLIT Run. A long-term (100 year) stochastic projection was performed to estimate spawning stock biomass and MSY under equilibrium conditions. The same partial recruitment, mean weights at age and maturity vectors used in the yield and SSB per recruit analysis were also used in the projections. A constant F scenario was used ($F = F_{msy}$ = 0.27). Estimates of Age 3 recruitment used in the projections were derived by re-sampling the cumulative density function based on the empirical observations during 1982 to 2009 (1979 to 2006 year classes) from the SPLIT RUN. The proportions of F and M which occurs before spawning equals 0.1667 (March 1); M equals 0.15. Based on yield per recruit analyses, BRPs were estimated to be: $F_{msy} = F_{40\%MSP}$ = 0.27. SSB_{msy} = 10,051 mt and MSY = 2,075 mt were estimated using a long-term stochastic projection. The 2010 SSB age structure remains truncated compared to the conditions under MSY. The 2010 SSB was 4,099 mt, 41% below SSB_{msy} (10,051 mt) and 2010 fishing mortality was 0.47, 173% above F_{msy} (F=0.27); therefore, witch flounder was overfished and overfishing occurred in 2010.

Two sensitivity runs (Runs F and G) of the SPLIT RUN were performed during the 2012 Update (NEFSC 2012) where length-based conversion factors from the second order polynomial were applied to the 2009 through 2011 survey indices. Length-specific conversion factors were applied to the entire length range in one run (Run F) while another run (Run G) applied length-specific conversion factors to lengths between 20 and 40 cm and held the conversion factor constant for lengths less than 20 cm and greater than 40 cm (i.e., all lengths less than 20 cm used the 20 cm conversion factor and all lengths greater than 40 cm used the 40 cm conversion factor). The review panel concluded that length-based conversion coefficients require a fuller consideration than can be performed in an update assessment. A research recommendation applicable to several stocks suggested exploring the possibility of refining the calibration factors within the assessment model itself (e.g., splitting the survey tuning series and using the results from the calibration experiment as a prior).

The 2015 Northeast Groundfish Operational Assessments (2015 OA; NEFSC 2015) was held in September 2015 in which 20 groundfish stocks were assessed and reviewed, including witch flounder. An analytical assessment (calibrated VPA) of commercial catch (landings plus discards) at age data through 2014 was conducted. A VPA analyses was performed for the SPLIT case (survey time series was split between 1994 and 1995). The retrospective analysis indicated that model underestimated fishing mortality and overestimated SSB and the magnitude of the retrospective pattern had increased slightly for F and decreased slightly for SSB compared to the 2012 assessment. However, the retrospective pattern for this assessment was considered to be 'major' because the rho adjusted estimates of 2014 SSB (SSB_{rho} = 2,077 mt) and 2014 F (F_{rho} = 0.687) were outside the approximate 90% confidence region around SSB (2,643 - 3,864 mt) and F (0.321 - 0.603). Therefore, a retrospective adjustment was made for both the determination of stock status and for projections of catch in 2016. The retrospective adjustment changed the 2014 SSB from 3,129 mt to 2,077 mt and the 2014 F_{full} from 0.428 to 0.687. The retrospective adjustment did not change stock status; witch flounder was overfished and overfishing occurred in 2014.

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

COMMERCIAL FISHERY LANDINGS

USA commercial landings of witch flounder generally increased from the early 1960s, peaking in 1984 at 6,658 mt (14.68 million pounds; Table B6; Figure B13). Subsequently, landings declined and fluctuated between 2,000 and 3,000 mt (6.6 million pounds) until 2003, when landings further declined. Landings in 2015 were 492 mt (0.11 million pounds), the lowest in the time series (Table B6; Figure B13).

Canadian commercial landings from the Gulf of Maine-Georges Bank region have been minor (not more than 68 mt annually); 2015 landings are estimated to be 1 mt (Table B6).

The majority (> 75%) of USA commercial landings come from the western Gulf of Maine and central basin and from the deeper waters north of the South Channel, specifically, statistical areas 512, 513, 514, 515, 521, and 522 (Table B7; Figure B14). Over the 1982-2015 time series, a shift in statistical areas is apparent. A decline in the percentage of landings occur in statistical areas 513 and 515 while an increase in percentages occur in 514, 521, and 522 (Table B7; Figure B14). Witch flounder landings occur throughout the year, with a slight increase during March - July months when witch flounder form dense pre-spawning aggregations (Burnett et al. 1992; Table B8; Figure B15). The majority (>91%) of the landings are taken by otter trawl gear (Table B9). In recent years, between 3% and 6% of landings are taken by haddock separator trawl. The majority of landings occur in Maine and Massachusetts. Since the mid-1990s, an increasing percentage of witch flounder is landed in Massachusetts (Table B10; Figure B16). Trends in landings by market category reveal a sharp decline in LARGE market category fish in the late 1980s (Table B11; Figure B17). Although culling and grading practices vary by port, witch flounder have historically been landed as either SMALL or LARGE; however, three market categories (PEEWEE, MEDIUM, and JUMBO were added in some ports beginning in 1982. In the mid-1990s, the proportion of witch flounder landings from the PEEWEE market category peaked. In 1983, a minimum size of 33 cm (13 inches) was established for witch flounder. The minimum size was increased to 36 cm (14 inches) in 1987 and reduced back in 33 cm in 2013.

Percentage of landings by ten minute squares of latitude and longitude is given in Figure B18 for the entire time series, by calendar quarter in Figure B19, and in 5-year blocks in Figure B20. Landings generally occur in the western and central Gulf of Maine; seasonal patterns reveal more landings from the southern flank of Georges Bank and southward in quarters 1 and 2 than in quarters 3 and 4. The general pattern of witch flounder landings appears similar over the 5-year time blocks within the Gulf of Maine region. Over the time series, fewer landings are taken from the Mid-Atlantic region. The ten minute squares with the highest percentage of landings varies over time from single squares dispersed in the western and central Gulf of Maine in the 1980s to western Gulf of Maine and along Georges Bank in the late 1990s, to inshore Gulf of Maine between Cape Ann and Cape Cod in the 2000s.

Vessel Trip Report (VTR) commercial landings and number of trips by ten minute square latitude/longitude resolution by 5-year time block with survey weight per tow from NEFSC spring and autumn surveys are presented in Figures B21 and B22, respectively.

COMMERCIAL FISHERY DISCARDS

The Northeast Fisheries Observer Program (NEFOP; formerly known as Domestic Sea Sampling Program and later as Fisheries Observer Program) began in 1989 and has generated varying levels of observer coverage for different fisheries. Prior to the NEFOP, the NEFSC conducted sea sampling on an ad-hoc basis.

Discards were initially estimated in the 1994 assessment (Wigley and Mayo 1996) for the Gulf of Maine northern shrimp fishery and the large (>=5.5 inches) mesh otter trawl fishery in the Gulf of Maine and Georges Bank region only (USA statistical areas: 464, 465, 511, 512, 513, 514, 515, 521-526, 561, and 562; Figure B23). Discard estimation was expanded to include the small mesh (<5.5 inches) fishery in the 2008 witch flounder assessment (NEFSC 2008) and the discard estimation method was updated to use a combined ratio method (Cochran 1963) of d/k_{all} (discarded weight of witch flounder/kept weight of all species) for the large mesh otter trawl fishery. In the 2012 and 2015 assessments, discards were updated using the same methods as the 2008 assessment. Discard mortality is assumed to be 100%.

In the current assessment, a working paper containing several refinements to the discard estimation for 1982-2014 was evaluated by the WG, including: a spatial extent refinement to include the entire stock area; expansion of the discard estimation to include additional gear types (scallop dredge, gillnet, and haddock separator trawl); some gear-specific refinements for large-mesh otter trawl, small-mesh otter trawl and shrimp trawl, and presentation of discard estimation for gillnet, scallop dredge, haddock separator trawl gear.

A summary of the refinements to direct and indirect discard estimation approaches, by gear type and time period, is given in Table B12. Discards in 2015 were estimated using these refined methods.

The number of observed trips by gear type, mesh group and year used in the discard estimation for the current assessment (1982 – 2015) is given in Table B13 for large mesh otter, small mesh otter trawl, shrimp trawl, gillnet, scallop dredge, and haddock separator trawl fisheries. Discards and associated CVs, by gear type, are given in Table B14 and Figures B24.

Large Mesh (>= 5.5 inches) Otter Trawl (LMOT)

Direct Estimation (1989-2015): The NEFOP data from 1989 to 2015 and At-Sea Monitoring (ASM) from 2010 to 2015 are used to derive an annual combined ratio of witch flounder discard weight to kept weight of all species (d/k_{all}) for two regions, CORE (statistical areas: 464, 465, 511, 512, 513, 514, 515, 521-526, 561, and 562) and OTHER (statistical areas: 533, 534, 537-543, 611-639; Table B13; Figure B23). Total annual discard weight is derived by multiplying the d/k_{all} ratio by the kept weight of all species in the commercial landings (Dealer data) for each region and then summing over region.

Indirect Estimation (1982-1988): No observer data are available prior to 1989. Witch flounder discards were indirectly estimated by Wigley and Mayo (1996) for the large mesh otter trawl trips in the Gulf of Maine-Georges Bank (CORE) region based on a method developed by Mayo et al. (1992) that used survey and commercial landings at length data, commercial gear retention ogives, and information on culling practices. Research vessel length frequency data were filtered through commercial gear retention ogives corresponding to the predominant meshes employed in the large mesh fishery (130, 140 and 152 mm) and then through a culling

practice ogive. Due to sparse retention studies for witch flounder, mesh selection ogives were taken from Walsh et al. (1992) for American plaice. Given the high value and low abundance of this species, the culling practice of commercial fishermen was assumed to be knife-edge at the minimum landing size. A semi-annual ratio estimator of survey filtered "kept" index to semi-annual numbers landed was used to expanded the estimated discards survey index to obtain number of fish discarded at length. Estimated numbers at length for both landed and discarded witch flounder in the large mesh otter trawl fishery were derived by season (winter/spring and summer/autumn). A spreadsheet illustrating the method used is presented in Table 19 of Wigley et al. 1999 (and illustrated in Figure 9) for 1993 using the spring survey and commercial landings from quarters 1 and 2. Semi-annual numbers of discard fish at length were apportioned to age using the corresponding season NEFSC age-length key.

To estimate discard weight of witch flounder in the stock area, the annual indirect discard estimates for the CORE region were multiplied by 0.385 (the 1989-2014 average of the ratio of the CORE region discard weight to the OTHER region discard weight) and added to the CORE region discard weight.

Estimated discards over the entire time series are presented in Table B14 and Figure B24. Previously estimated discards are also shown in Appendix Figure B1. Current discard estimates average 198 mt over the time series and the CVs are generally below 30%. In most years, a slight increase in discards occurred when the spatial extent is expanded southward; however, in 1991 and 1992 there is a large increase which appear to be unrelated to observer coverage level in the OTHER region (Table B13).

Small Mesh (<5.5 inch) Otter Trawl (SMOT)

Direct Estimation (1989-2015): The NEFOP data from 1989 to 2015 (except 1994 and 1998) are used to derive an annual combined ratio of witch flounder discard weight to kept weight of all species (d/k_{all}) for each region, CORE and OTHER (Table B13). Total annual discard weight was derived by multiplying the d/k_{all} ratio by the kept weight of all species in the commercial landings (Dealer data) for each region and then summing over region. As mentioned previously, because there was a single observed trip for 1994 and 1998 in the CORE region (Table B13), the average of the discard weight from adjacent years for the CORE region (e.g., 1993 and 1995 were averaged to estimate 1994 for the CORE region) was added to the discard weight of the OTHER region to derive an annual discard estimate for the stock area for these two years.

Indirect Estimation (1982-1988): No observer data are available prior to 1989 (Table B13). Witch flounder discards were indirectly estimated by multiplying the region-specific 1989-2014 (excluding 1994 and 1998 for the CORE region) average of the combined ratio by the region-specific annual kept all weight. Annual region-specific discard estimates were then summed to derive annual discard estimates.

Estimated discards over the entire time series are presented in Table B14 and Figure B24. Previously estimated discards are also show in Appendix Figure B2. Current discard estimates average 78 mt over the time series; the CVs are generally greater than 30%.

Gillnet (GN)

Direct Estimation (1989-2015): The NEFOP data from 1989 to 2015 are used to derive an annual combined ratio of witch flounder discard weight to kept weight of all species (d/k_{all})

for each region, CORE and OTHER (Table B13). Total annual discard weight was derived by multiplying the d/k_{all} ratio by the kept weight of all species in the commercial landings (Dealer data) for each region and then summing over region.

Indirect Estimation (1982-1988): No observer data are available prior to 1989 (Table B13). Witch flounder discards were indirectly estimated by multiplying the region-specific 1989-2014 average of the combined ratio by the region-specific annual kept all weight. Annual region-specific discard estimates were then summed to derive annual discard estimates.

Estimated discards over the entire time series are presented in Table B14 and Figure B24. No previous discards have been used in the stock assessment given the negligible amount. Current discard estimates average 1 mt over the time series and the CVs are greater than 30% early in the time series but are below 30% in recent years.

Scallop Dredge (SD)

Direct Estimation (1992-2015): The NEFOP data from 1992 to 2015 are used to derive an annual combined ratio of witch flounder discard weight to kept weight of all species ($d/k_{al}l$) for each region, CORE and OTHER (Table B13). Total annual discard weight was derived by multiplying the d/k_{all} ratio by the kept weight of all species in the commercial landings (Dealer data) for each region and then summing over region.

Indirect Estimation (1982-1991): Only 2 observed trips are available prior to 1992 (Table B13). Witch flounder discards were indirectly estimated by multiplying the region-specific 1992-2014 mean of the combined ratio by the region-specific annual kept all weight. Annual region-specific discard estimates were then summed to derive annual discard estimates.

Estimated discards over the entire time series are presented in Table B14 and Figure B24. No previous discards have been used in the stock assessment. Current discard estimates average 50 mt over the time series and the CVs are below 30% since the late 1990s.

Haddock Separator Trawl (HS)

Direct Estimation (2010-2015): The NEFOP data from 2010 to 2015 are used to derive an annual combined ratio of witch flounder discard weight to kept weight of all species (d/k_{all}) for each region, CORE and OTHER (Table B13). Total annual discard weight was derived by multiplying the d/kall ratio by the kept weight of all species in the commercial landings (Dealer data) for each region and then summing over region. NEFOP data in 2009 could not be utilized because the commercial landings data does not contain the haddock separator trawl gear code.

Indirect Estimation (1982-2009): No discard estimation is needed for this time period as the haddock separator trawl is a recent gear type. In years leading up to the new gear code, there could be some incorrect gear codes reported in the Vessel Trip Report and used in the commercial landings data. However, a comparison conducted in which haddock separator trawl (negear 057) and large mesh otter trawl (negear 050) are combined reveals negligible impacts on discard estimation (Appendix Figure B3).

Estimated discards for the 5-year time series are presented in Table B14 and Figure B24. Current discard estimates average 1 mt over the time series and the CVs are at or below 30%.

Gulf of Maine Shrimp Trawl (GMSHR)

Discards from the shrimp trawl fishery (primarily associated with the northern shrimp fishery) in the Gulf of Maine have been derived for the CORE area only. In April 1992, a regulation requiring the use of the Nordmore Grate was implemented in the northern shrimp fishery. Small mesh exempted fisheries using shrimp trawls are not required to use a grate. The Vessel Trip Report logbook does not collect information regarding the use of a grate during commercial fishing trips. In 2014 and 2015, the northern shrimp fishery was closed; however, trips using shrimp trawls without grates fished in exempted areas within the Gulf of Maine (fishermen were contacted to confirm the gear used).

Direct Estimation (1989-1997; 2001, 2003-2013): The NEFOP data are used to derive an annual combined ratio of witch flounder discard weight to kept weight of all species (d/k_{all}) for the CORE region (Table B13). Total annual discard weight was derived by multiplying the d/k_{all} ratio by the kept weight of all species in the commercial landings (Dealer data).

Indirect Estimation (1982-1988; 1998-2000, 2002, 2014, 2015): No observer data are available prior to 1989 and between 1998 and 2000, 2002, 2014, and 2015 (Table B13). Witch flounder discards were indirectly estimated by multiplying an average of the combined ratio by the annual kept all weight. For years prior to 1993 and for 2014-2015, 1989-1992 time series was used (years when no grate was required or no northern shrimp fishery in 2014-2015). For years without observer data between 1998 and 2002, a 1993-2013 time series (excluding 2001 and 2011 in which insufficient trips occurred) average combined ratio was used.

The discard estimation methods described above differ from those used in the 2015 Operational Assessment (OA) where discards from the northern shrimp fishery were estimated using the following two methods. When no observer data were available, a regression of age 3 fish in the autumn NEFSC survey and observed discard rates was used to estimate ratios of discard weight to days fished (d/df) ratios. When observer data were available, d/df ratios were calculated by fishing zone (a surrogate for depth). To estimate discard weight, the mean discard ratio (weighted by days fished in each fishing zone) was expanded by the days fished in the northern shrimp fishery.

Discarded weight was translated into numbers of fish at age using observer discard length frequency and NEFSC spring survey age data. When no observer length frequency data were available, an average proportion at age and weighted mean weights over a 5-year time period were used to translate discarded weight into numbers of discarded fish at age (Appendix Figure B4).

Estimated discards over the entire time series from the Gulf of Maine shrimp fishery are presented in Table B14 and Figure B24. Previously estimated discards are also shown in Appendix Figure B5. The previously estimated discards are lower than the re-estimated discards for the1982-1988 time period and the higher for the mid-1990s but remain within the 95% confidence interval of the re-estimated discards.

Total Estimated Discards

Total discards from the six fisheries ranges between 93 mt (2015) and 774 mt (1991) with a time series mean of 301 mt; CVs range between 7% and 40% (Table B14; Figures B25 and B26). The discards estimated in the current assessment differ slightly from those estimated in previous assessments (Figure B26 and Appendix Figure B6) that include only three fisheries. A notably difference occurs in 1991 and is attributable to the spatial extent refinement for the large mesh otter trawl fisheries (the 1991 discard ratio in the OTHER region is 3 times higher than the penultimate discard ratio for this fishery and region).

RECREATIONAL FISHERY CATCH

There is no recreational fishery for witch flounder. The database of the NMFS Marine Recreational Information Program (MRIP) database contains no observations for witch flounder. The Vessel Trip Report database contains 377 witch flounder[1] from 26 party/charter trips during 1996 – 2015, a negligible amount of fish.

COMMERCIAL FISHERY LANDINGS AT LENGTH AND AGE

Commercial length frequency and age sampling data are summarized by market category and calendar quarter (Table B15). Some ports do not cull into "peewee" and "jumbo" categories, and so these categories have been combined into the "small" and "large" categories, respectively. Prior to 1982, sampling was sporadic. Annual sampling from 1982 onward ranged between 11 and 133 mt per 100 lengths, exceeding the informal sampling threshold criterion of 200 mt per 100 lengths. During the 1990's, sampling was insufficient in some quarters and market categories and it was necessary to pool some quarters for some market categories. Sampling improved by quarter and market category in the late 1990's in which an average 34 mt per 100 lengths was obtained. Commercial age data were available for 1982 to 2015; landed witch flounder ranged in age between 3 and 33 years. To estimate landings at age and mean weights at age, quarter, semi-annual or annual age-length keys were applied to corresponding commercial landings length frequency data by market category. Number of fish landed at age, mean weights at age, and mean lengths at age of landed fish are presented in Tables B16, B17, and B18, respectively for ages 0 to 11+ and in Figures B27 and B28 for ages 0 to 14+ and ages 0 to 11+, respectively. The majority of landed fish were ages 6 to 9. The coefficient of variation associated with the numbers of fish at age are given in Table B19; CVs for ages 6 through 9 are generally less than 15%.

COMMERCIAL FISHERY DISCARDS AT LENGTH AND AGE

A summary of the discarded length frequency data collected by the NEFOP is given in Table B20 and Figure B29 by gear type. Witch flounder age structures collected by the NEFOP are not aged; the NEFSC spring and autumn age length data are used.

<u>Large Mesh (>= 5.5 inches) Otter Trawl (LMOT)</u>

Based on NEFOP discard reasons, witch flounder are often discarded due to the minimum size regulation. There is a narrow mode of discarded witch flounder centered at the minimum size (Figure B29). Minimum size regulations for witch flounder have changed over

[1] Upon inspection of the VTRs with the 4 highest counts of witch flounder caught on recreational trips, 3 of the 4 trips had not reported witch flounder; it appears mis-interpretation of hand writing resulted in 171 (45% of 377) witch flounder incorrectly reported as recreational kept fish.

time. In 1983, a minimum size was established at 33 cm (13 inches). In 1987, the minimum size increased to 36 cm (14 inches) and then decreased back to 33 cm in July 2013.

With no new length frequency and age data since the 2015 OA, the numbers of discarded fish at age for the 1982-2014 time period reported in the 2015 OA were re-scaled to the current discard weight using an annual ratio of current discard weight to 2015 OA discard weight. The 2015 OA and the 2015 discards estimated in the current assessment used NEFOP discard length frequency data and NEFSC seasonal age-length keys to translate discarded weight into number of fish discarded at age for the large mesh otter trawl fishery. Witch flounder discarded in the large mesh otter trawl fishery range in age from 1 to 11+, with the majority between ages 4 and 6.

Small Mesh (<5.5 inch) Otter Trawl (SMOT)

In the small mesh otter trawl fishery, witch flounder are often discarded due to no retention regulations and size retention regulations. The overall proportion at length of discarded fish for the small mesh otter trawl fishery appears similar to the large mesh otter trawl fishery (Figure B29). There are some years with insufficient length measurements annually or when partitioned by half year; eight years during the 1989-2014 time period in which the numbers of discarded fish measured for length was less than 100 fish (Table B20).

With limited length frequency data over the entire 1982-2015 time series, the numbers of discarded fish at age from the large mesh otter trawl fishery were rescaled by the annual ratio of small mesh otter trawl discarded weight to the large mesh otter trawl discarded weight.

Gillnet (GN)

In the gillnet fishery, witch flounder are often discarded due to size regulations or for unknown reasons. The length frequency of discarded fish reveals a mode and range similar to large mesh otter trawl (Figure B29). The numbers of fish measured annually for length over the time series is very sparse (Table B20).

With very sparse length frequency data over the entire 1982-2015 time series, the numbers of discarded fish at age from the large mesh otter trawl fishery were rescaled by the annual ratio of gillnet discarded weight to the large mesh otter trawl discarded weight.

Scallop Dredge (SD)

In the scallop dredge fishery, witch flounder are often discarded due to no market or no retention regulations. The length frequency of discarded fish over all years combined reveals a mode and range that is shifted towards larger fish than the large mesh otter trawl (Figure B29). The numbers of fish measured annually for length over the time series varies annually (Table B20).

For years without discarded length frequencies or years with less than 100 fish measured (1982-1995, 1998, 2001, 2002), the annual numbers of discarded fish at age were derived using a combined length frequency over the time series and NEFSC spring and autumn age data combined with commercial landings age data to translate annual discarded weight into number of discarded fish at age for the scallop dredge fishery. Witch flounder discarded in the scallop dredge fishery range in age from 1 to 11+, with the majority at ages 4 to 8.

Haddock Separator Trawl (HS)

In the haddock separator trawl fishery, witch flounder are often discarded due to size retention regulations. The length frequency of discarded fish reveals a mode and range similar to large mesh otter trawl (Figure B29). The numbers of fish measured annually for length over the short time series is limited (Table B20).

With limited length frequency data, the numbers of discarded fish at age from the large mesh otter trawl fishery were rescaled by the annual ratio of haddock separator trawl discarded weight to the large mesh otter trawl discarded weight.

Gulf of Maine Shrimp Trawl (GMSHR)

In the Gulf of Maine shrimp trawl fishery, witch flounder are often discarded due to unknown reasons or quota retention regulations. The length frequency of discarded fish over all years combined reveals a mode and range that is shifted towards smaller fish than fish discarded in the large mesh otter trawl or scallop dredge (Figure B29). The numbers of fish measured annually for length over the time series varies annually (Table B20).

With no new length frequency and age data since the 2015 OA, the numbers of discarded fish at age for the 1989-2013 time period reported in the 2015 OA were re-scaled to the current discard weight using an annual ratio of current discard weight to 2015 OA discard weight.

With no discard length frequency data in 2014 and 2015 from a shrimp trawl fishery that is not using a grate, the numbers of discarded fish at age from the large mesh otter trawl fishery were rescaled by the annual ratio of shrimp trawl discarded weight to the large mesh otter trawl discarded weight based on the assumption that discarded length frequency of witch flounder in shrimp trawl without a grate would be similar that of the small mesh otter trawl fishery.

Witch flounder discarded in the shrimp fishery ranges in age from 0 to 6, with the majority at ages 1 to 3. The estimate discard weight of with flounder from the shrimp fishery is small compared to the otter trawl fishery.

Total Estimated Discards

To form the total numbers of discarded witch flounder at age, the numbers of discarded fish at age in the large mesh otter trawl fishery (rescaled to account the discarded weight from the small mesh otter trawl, gillnet, 2014-2015 Gulf of Maine shrimp trawl, and haddock separator trawl fisheries) was added to the numbers of discarded fish at age in the 1982-2013 the Gulf of Maine shrimp trawl fishery and the numbers of discarded fish at age in scallop dredge fishery. Weighted mean weights and mean lengths at age were derived using numbers of fish as the weighting factor. Numbers of discarded fish at age, mean weight at age, and mean length at age are given in Tables B21, B22, and B23, respectively, and Figure B30.

TOTAL FISHERY CATCH

Total commercial landings in 2015 were 493 mt (1.1 million pounds), commercial discards were estimated to be 93 mt (0.2 million pounds), for a total catch of 585 mt (1.3 million pounds; Table B13).

To form the numbers of fish at age in the catch, the total discarded fish at age was

combined with the numbers of fish in the commercial landings. Weighted mean weights and mean lengths at age were derived using numbers of fish as the weighting factor. Numbers of witch flounder in the catch at age, mean weight at age, and mean length at age are given in Tables B24, B25, and B26, respectively, and Figure B31, B32, and B33.

The age composition data reveal strong 1979-1981 year classes (Figure B31). The 1989 and 1993 year classes also appear to have been strong, however, these cohorts were heavily discarded in both the shrimp and large mesh otter trawl fisheries. The age composition also reveals a truncated age structure since the mid-1980s, and a short-lived expansion in the early 2000s that was not sustained.

WG recommendation: WG members accepted the catch at age as derived.

TERM OF REFERENCE 2: Present available federal, state, and other survey data, indices of relative or absolute abundance, recruitment, etc. Characterize the uncertainty and any bias in these sources of data and compare survey coverage to locations of fishery catches. Select the surveys and indices for use in the assessment.

RESEARCH SURVEY INDICES OF ABUNDANCE

Northeast Fisheries Science Center (NEFSC)

The NEFSC has conducted annual research vessel stratified random bottom trawl surveys during autumn (generally between October and November in northern strata) since 1963 and during spring (generally between March and April) since 1968. The NEFSC survey samples offshore waters between 27 m [15 fm] to 366 m [200 fm] from Cape Hatteras NC to the Canadian border (Figure B34). Details on the survey sampling design are given in Grosslein (1969) and Azarovitz (1981), changes to the survey over time are described in Johnston and Sosebee (2014), and uses of survey data in stock assessments are given in Clark (1981). For witch flounder, strata 22-30, 36-40 (Figure B35) have been used for indices of relative abundance and biomass and have been used in previous peer-reviewed stock assessments as tuning indices in the analytical model (relative indices and/or swept area indices). Table B27 summarizes the square nautical miles of the NEFSC survey area and associated statistical areas.

The annual minimum, maximum, and weighted mean temperature and depth of the strata set over the time series are given in Figures B36 and B37. The weighting factor for mean temperature and mean depth was witch flounder abundance and biomass in tows with positive catches. In spring, bottom temperatures generally range between 3 and 9 degrees with a weighted mean of about 6 degrees. In autumn, bottom temperatures generally range 5 and 11 degrees with a weight mean of about 7.5 degrees. The reduced variability in maximum depth since the mid-1990s onward is attributed to a change in towing protocols for stations with depths greater than or equal to 183 m (Figure B37). Johnston and Sosebee (2014) document "*a more specific depth is randomly chosen among four depth intervals and then trawling was done along that depth contour.*"

The timing of the spring and fall surveys has varied slightly over the time series. In spring of 2014 and 2016, the spring survey was slightly later than the time series mean (Figure B38). The percentage of positive tows in the NEFSC spring and fall survey exhibit similar patterns (Figure B39). The percentage of positive tows was generally greater than 50% prior to a decline in the late 1980's and 1990's when percentages reached time series lows (23% in fall 1987; 24% in spring 1990). Since that time, the percentage of positive tows has increased to above 50% and since 2009 the numbers of positive tows are among the highest in the time series. The recent increase in positive tows is attributed to the change in survey vessel and trawl gear. Gini (concentration) indices of abundance and biomass for spring and fall exhibit similar trends indicating witch flounder are moderately concentrated among the NEFSC survey strata (Figure B40) with a declining trend in concentration over the time series more evident in the spring than fall.

Witch flounder are generally distributed throughout the Gulf of Maine, along the northern edge and southern flank of Georges Bank, and southward along the continental shelf as far south as Cape Hatteras, NC. Distributions of witch flounder abundance in spring 1968-2016 and fall 1963-2015 time series are given in Figures B41 and B42 (animated 5-year moving time block

plots are available on Data Portal). Juveniles are generally distributed in the western Gulf of Maine and along the Mid-Atlantic shelf (Figures B43) with concentrations along the western portion of the Gulf of Maine. As noted earlier, Burnett et al. (1992) found seasonal differences between juveniles and adults. Distribution of survey catches coincides with distributions of commercial landings and trips reported in the Vessel Trip Reports (Figures B21 and B22).

While NEFSC spring survey indices tend to be more variable due to pre-spawning aggregations of witch flounder, spring and autumn indices generally display similar trends. NEFSC survey indices generally declined from the early 1960s to record low levels in the late 1980s and early 1990s. There was an increasing trend in the late 1990s to early 2000s followed by a declining trend (Figure B44). In 2015, the survey indices were below their time series means (Table B28). Uncalibrated Bigelow indices are given in Figure B45.

NEFSC spring and autumn length frequency data are presented in Figures B46 and B47, respectively. The NEFSC survey does not consistently catch small witch flounder; however, occasional strong cohorts are evident in the fall survey between 10-15 cm. Strong cohorts are evident in the other surveys. The length frequency plots also reveal the NEFSC survey consistently caught large witch flounder (greater the 55 cm) in both spring and autumn early in the survey time series; however in recent years, few large fish are caught above 55 cm.

NEFSC spring and autumn age sampling is summarized in Table B29. The maximum age has declined over the time series (Table B29). In the 1980s, the maximum age in the autumn survey was between 20 and 30 years and in recent years, maximum age was between 10 and 15 years.

Age-specific relative abundance from NEFSC spring and autumn 1980-2015 and spring 2016 surveys are given in Table B30 and Figure B48. In the early to mid-1980s, older witch flounder were evident in the survey catches. A truncation of the age structure occurred in the late-1980s; a slight expansion in age structure occurred in early 2000s but was not sustained.

The NEFSC survey data indicate improved recruitment of age 3 fish in 2007, 2011 and 2013 (2004, 2008, and 2013 year classes, respectively; Figure B49); however, only the 2004 year class is evident as a strong year class in the following year.

NEFSC spring and autumn survey mean weights at age are given in Table B31 and Figure B50. Survey mean weights are variable, however, similar declines in mean weights for ages 6-9 were observed during the mid-1990s to 2002 in both the commercial landings and spring and autumn surveys.

Mean lengths at age from NEFSC spring and autumn surveys are presented in Table B32 and for ages 4 to 8 in Figure B51. Mean lengths at age for ages 5 to 7 appear to have increased (approximately 3-5 cm) from 1980 to the late 1980's, and then declined (Figure B1); however, von Bertalanffy growth analyses detected no significant changes in resulting growth parameters over the time period (Wigley et al. 2003a).

Additional NEFSC Survey data explorations

Day-night differences in witch flounder catches were explored; negligible day-night differences were found (Figure B52).

Survey Conversion Factors

There are no significant vessel, door, or net conversion factors for witch flounder in the NEFSC bottom trawl survey between 1963 and 2008. No conversion factors have been applied

to the survey indices for the 1963 through 2008 period.

In 2009, the Fisheries Survey Vessel (*FS/V*) *Henry B. Bigelow* replaced the *R/V Albatross IV* as the principle vessel conducting the bottom trawl survey. The vessel characteristics (size and towing power) as well as fishing gear characteristics differed between the two vessels resulting in different survey catchability. Calibration experiments were conducted to estimate catchability differences during 2008 (Brown 2009), and results of the experiments were peer reviewed in 2009 (Anonymous 2009, Miller et al. 2010). There are significant vessel conversion factors between the *FS/V Bigelow* and the *R/V Albatross IV* for witch flounder (Miller et al. 2010) in the NEFSC bottom trawl survey. The vessel conversion factor for numbers and weight (spring and autumn) is 3.2572 and has been applied to the 2009 through 2016 NEFSC bottom trawl surveys. The *FS/V Bigelow* survey indices were divided by the conversion factor to obtain indices consistent with the *R/V Albatross IV*.

Length-based conversion factors have not been established for witch flounder. A small working group was convened prior to the 2012 Update meeting to discuss candidate models and the application of length-based conversion factors for witch flounder. The group concluded that further work was needed before length-based factors were applied; however, the group agreed that a sensitivity run should be conducted using survey indices adjusted with length-based conversion factors derived from a second order polynomial. Two sensitivity runs (Runs F and G) of the SPLIT RUN were performed during the 2012 Update (NEFSC 2012) where length-based conversion factors from the second order polynomial were applied to the 2009 through 2011 survey indices. Length-specific conversion factors were applied to the entire length range in one run (Run F) while another run (Run G) applied length-specific conversion factors to lengths between 20 and 40 cm and held the conversion factor constant for lengths less than 20 cm and greater than 40 cm (i.e., all lengths less than 20 cm used the 20 cm conversion factor and all lengths greater than 40 cm used the 40 cm conversion factor). The review panel concluded that length-based conversion coefficients require a fuller consideration than can be performed in an update assessment. A research recommendation applicable to several stocks suggested exploring the possibility of refining the calibration factors within the assessment model itself (e.g., splitting the survey tuning series and using the results from the calibration experiment as a prior).

<u>WG recommendation</u>: Continue to use constant conversion factors for Bigelow series as tuning indices in Albatross units; explore use of a "stand alone" Bigelow series without conversion factors (see Term of Reference 4 for a summary of exploratory model configurations).

Alternate NEFSC strata sets

Analyses were conducted in 2003 to examine if the use of additional strata in the NEFSC bottom trawl survey might be appropriate. Burnett and Clark (1983) used NEFSC survey strata set 22, 24, 26-30, and 33-40 in the first witch flounder assessment; however, Burnett (MS 1987) suggested that fish from strata 33, 34 and 35 exhibited different growth rates indicating these fish may be from a different stock inhabiting the western Scotian shelf. Based on this information, Wigley and Mayo (1996) revised the witch flounder survey strata set excluding 33, 34, and 35, and included strata 23 and 25. Following a method developed by Cadrin and King (2003), witch flounder catches for the entire autumn bottom trawl survey time series were examined by individual stratum. The mean number per tow, expanded by the stratum area for each stratum was summed over the time period, and the percentage contribution of each stratum was calculated as well as the percentage of annual stratum sampling which produced no catch.

Results indicate that the current strata set (22-30, 36-40) accounts for approximately 93% of the survey catch and that only minor differences exist between the strata sets used in previous assessments. That analysis also indicated that stratum 6 contributed to the overall witch flounder catch. The stratified mean weight (kg) per tow was calculated for three strata sets: set 1 (22-30, 36-40); set 2 (22, 24, 26-30, 36-40); and set 3 (6, 22, 24, 26-30, 36-40). The trends of these biomass indices (and their variance) are indistinguishable. The inclusion of stratum 6 is not justified due to its geographical discontinuity with the core strata.

In the current assessment, an analysis comparing the current strata set (set 1: 22-30, 36-40) and an alternative set (set 2: 22, 24, 26-30, 36-40) was updated through 2015 and results were similar to those in 2003. The current strata set (22-30, 36-40) accounts for 94% of survey catch (in numbers and weight) and only minor differences exist between the alternative set in percentage of survey catch as well as relative biomass (and associated variance) trends (Table B33 and Figure B53).

The WG also reviewed three additional alternative NEFSC strata sets: (a) the inclusion of strata 33, 34, and 35/351 in the autumn survey series; (b) the inclusion of stratum 34 in the spring and autumn survey series; and (c) the inclusion of strata 2-4, 6-8, 10-18, and 74-76 in the spring and autumn survey series. Each additional alternative strata set was compared with strata set 1 (22-30, 36-40) used in recent assessments. Adding additional strata did not change the trend in abundance or biomass over the time series for the spring or fall surveys (Figures B53a, B53b, B53c).

The WG noted that Burnett et al. 1992 described witch flounder growth differences between Gulf of Maine-Georges Bank region and Scotian Shelf in which the warmer water of the Gulf of Maine (compared to Scotian Shelf) and year-around feeding were factors attributed to the faster growth of the Gulf of Maine-Georges Bank witch flounder. The WG members also noted the Scotian Shelf landings are not included in the catch used in the model, and this was another reason not to include the Scotian Shelf strata in the set.

<u>WG recommendations</u>: Since no additional strata were identified as contributing to the total catch, or improved the precision of the estimates of mean weight per tow, the strata set 22-30, 36-40 was retained.

Witch flounder catchability study

A field experiment was conducted to estimate the survey catchability of witch flounder in the net used in the NEFSC bottom trawl survey and a net with a chain sweep by using a twin net configuration. The goal was to estimate the sweep efficiency of the NEFSC survey trawl. A working paper entitled "Empirical estimates of maximum catchability of Witch Flounder *Glyptocephalus cynoglossus L.* on the Northeast Fisheries Science Center Fall bottom trawl survey" by Hare et al. 2016b describes the field study as well as the analyses used to derive the efficiency of the NEFSC survey net for witch flounder (See Appendix B1 for working paper and supplemental information.)

<u>WG recommendation</u>: The WG accepted the witch flounder catchability (q) estimates (0.056 and 0.291) from the sweep study and recommended their use in the stock assessment to convert the NEFSC survey relative abundance indices into swept area numbers (SWANs).

Massachusetts DMF

The Commonwealth of Massachusetts Division of Marine Fisheries (MA DMF) began spring and fall inshore trawl surveys in 1978 which complement the NEFSC surveys in coastal Massachusetts waters in that depths less than 27 m (the lower depth limit sampled by the NEFSC offshore survey) are sampled (for details see Howe et al. 1981). For witch flounder, strata 25-36 (region 4 and 5; Figure B54) have been used for indices of abundance and biomass and have been presented in previous stock assessments as corroborative information.

The MA DMF spring and fall survey time series has varied little in time of year over the time series (Figure B38). The percentage of positive tows is low (< 50%) in the spring and fall survey series (Figure B39). Gini (concentration) indices of abundance and biomass for the MA DMF surveys indicate witch flounder are generally concentrated and not evenly distributed among the survey strata (Figure B40). The fall gini indices appear similar over time while the spring gini indices indicate witch flounder have become more concentrated from the 1990s onward.

Distributions of witch flounder in the MA DMF surveys (1978 – 2015 time series and 5-year moving time block) are given in Figures B55 and B56.

Relative trends in abundance, biomass, mean length and individual weight of witch flounder in the MA DMF inshore spring and fall surveys are presented in Table B34, Figure B57. Spring survey catch per tow indices indicate declining trends from the late 1970's to time series lows in the mid-1990s; a slight increase is evident in the mid-2000s. The number of fish caught in the fall survey varies from a high of 171 in 2011 to a low of 2 fish in 1996 while the numbers of fish caught in the spring survey are sparser than the fall (Table B35). Fall survey catch per tow indices varies over the time series with a stratified mean less than 1 fish interspersed with peaks of 2 or 3 fish per tow occurring in 1983, 1995, 2002, and 2011 (Figure B57).

The mean length of witch flounder in the MA DMF spring and fall survey has declined from approximately 45 cm to 35 cm over the time series (Figure B58).

MA DMF length frequency data are sparse in the mid-1990s onward, with the exception of spring 2000 (1998 year class) and fall 1995 (1993 year class) and fall 2011; Figures B59 and B60.

Too few age samples are collected during the MA DMF surveys to reliably characterize the age composition (Table B35). Given the sparse length frequency in these seasonal surveys, NEFSC age length keys were not applied.

WG recommendation: This survey has a spatial coverage confined to the inshore waters of Massachusetts. Consider including MA DMF survey as corroborative information only (not included in model).

ASMFC Shrimp Survey

The Northern Shrimp Technical Committee of the Atlantic States Marine Fisheries Commission (ASMFC) has conducted an annual northern shrimp survey during August in the Gulf of Maine since 1983, with catch data for witch flounder available since 1984 onward (for details on the shrimp survey, see Northern Shrimp Survey, see Northern Shrimp Technical Committee MS 1984; Johnston and Sosebee 2014). This survey uses a stratified random sampling design and has consistently sampled strata 1, 3, and 5-8 (Figure B61). For witch flounder, strata 1, 3, 6, and 8 have been used for indices of abundance and biomass and have been presented in previous stock assessments as corroborative information. Comparison of an

alternative strata set (strata 1, 3, 5-8) revealed negligible differences between the original strata set and the alternative set (Figure B62), consequently the original strata set was retained.

The ASMFC survey time series has varied little in time of year over the time series (Figure B38). There is a high percentage of positive tows in the survey series (Figure B39). Gini (concentration) indices of abundance and biomass for this survey indicate witch flounder are distributed evenly among the survey strata (Figure B40).

Distribution of witch flounder survey catches for 1984 – 2015 time series combined is given in Figure B63. Five-year moving time blocks of survey catches are available on the data portal.

Trends in stratified mean abundance and biomass, mean length, and individual weight of witch flounder in the ASMFC shrimp survey are presented in Table B36 and Figure B64. Survey catch per tow indices indicate low abundance in the mid-1980s to mid-1990s, increased abundance in 1999 and early 2000s (a peak in 1999 associated with the strong 1998 and 1999 year classes), followed by a decline to a slightly higher mean than the mid-1980s. The peak in 1999 may also reflect increased availability rather than true abundance in that year.

Length frequency data from the ASMFC shrimp survey (Figure B65) suggests that incoming year classes can be identified prior to their appearance in the NEFSC survey. The ASMFC data might be more useful in providing a pre-recruit index than to characterize the population as whole given the ASMFC survey caught few witch flounder greater than 40 cm. The ASMFC survey data indicate improved recruitment (see length frequency modes at 12 cm, corresponding to age 1 fish) during 1991-1994, 1997, 1999, 2005, and 2014.

Witch flounder age structures are not collected on this survey; however, age-specific relative abundance indices, mean weight at age, and mean length at age were derived by applying annual NEFSC spring and fall (combined) age-length keys to the length frequency data from the ASMFC survey (Tables B37, B38, and B39, respectively; Figures B66 and B67). Any age-length key holes were manually filled in. Mean weight and mean length at age indices reveal similar trends as in other surveys in which mean weights and mean lengths for younger witch flounder fluctuate around the time series mean at age; however, in this survey age 4 and 5 fish exhibit a declining trend in mean weights and length from the mid-1980s to the early 2000s and then increases from the mid-2000s (Figure B68). These age groups represent the younger ages in the NEFSC survey.

WG recommendation: Although this survey has limited spatial coverage, much of the juvenile range is covered. This survey appears to provide useful information on relative to trends in abundance, distribution and recruitment of witch flounder in the Gulf of Maine and could be considered as tuning indices.

Maine-New Hampshire (MENH) Inshore Survey

Maine Department of Marine Resources and New Hampshire Fish and Game Department (MENH) have conducted spring and fall bottom trawl surveys along the New Hampshire and Maine coast since fall 2000. The survey has a stratified random sampling design with a fixed component. Sampling occurs in 5 regions and 4 depth zones that range from 5 fm to greater than 55 fm (Sherman et al. 2005; Figure B69). The distribution of witch flounder survey catches for spring and fall in 5-year time blocks are given in Figures B70-B75.

Relative trends in abundance and biomass of witch flounder in the MENH inshore survey are presented in Table B40 and Figure B76. In the spring, survey catch per tow indices averaged

about 6 fish per tow and 0.3 kg per tow until 2014 when a sharp increase in number of fish occurred (Figure B76). In the fall, a peak in both numbers and biomass occurred in 2001 and subsequently indices averaged about 12 fish per tow and 1 kg per tow.

Length frequency data from the MENH inshore spring and fall survey (Figures B77 and B78, respectively) suggest that incoming year classes may be identified prior to their appearance in the NEFSC survey. Similar to the ASMFC survey, the MENH survey might be useful in providing a pre-recruit index rather than to characterize the population as whole given the MENH survey catches few witch flounder greater than 30 cm (approximately age 4). The MENH inshore survey data indicate improved recruitment (see length frequency modes between 10-15cm, corresponding to age 1 fish) during 2001, 2005, and 2014.

Witch flounder age structures have been collected on this survey; however, ageing has not occurred. Age-specific relative abundance indices, mean weight at age, and mean length at age were derived by applying annual, season-specific NEFSC age-length keys to the length frequency of the MENH survey (Tables B41, B42, and B43, respectively; Figures B79 and B80). Any age-length key holes were manually filled in. The MENH survey catches ages 1 to 11+ ages, with most fish between the ages of 1 and 3. Normalized mean weights revealed a slight increase trend some ages (i.e., age 5) while other ages fluctuated about zero.

WG recommendation: Although this survey has limited spatial coverage, much of the juvenile range is covered. This survey appears to provide useful information on relative trends in abundance, distribution and recruitment of witch flounder in the Gulf of Maine and could be considered as tuning indices.

Other Surveys

- The NEFSC winter survey was conducted between 1992 and 2007 during February in the Georges Bank and Mid-Atlantic region using a stratified random sampling design. In general, there was poor coverage of Georges Bank and deep strata (183 -366 m) were not covered consistently. The following strata were consistently sampled over the 16 year time series: Strata 1-3, 5-7, 9-11; 61, 63, 65-67, 69-71, 73-75 (Figure B34; Johnston and Sosebee 2014). This survey does not sample the core distribution of witch flounder in the Gulf of Maine.

- The Canadian bottom trawl survey samples Georges Bank and does not sample the core distribution of witch flounder in the Gulf of Maine.

- The NEFSC's Cooperative Research Program longline survey uses a random sampling design within bottom trawl survey strata in the central basin of the Gulf of Maine. The longline survey was conducted in 2014 and 2015; no witch flounder have been caught (pers. comm. L. O'Brien August, 2016).

- The MA Division of Marine Fisheries (MA DMF) has conducted an Industry-Based cod survey (IBS) in the Gulf of Maine in 2003 - 2007 and 2016. Plots aggregating the April-May cruises were provided by M. Dean (MA DMF, pers. comm, July 13, 2016). Dean noted that the location of witch flounder appear to fairly similar between IBS1 and IBS2 study areas. A mean catch per tow (kg) was 4.8 kg/tow. Further information about this survey is available online at: http://www.mass.gov/eea/agencies/dfg/dmf/programs-and-projects/industry-based-survey-

for-gulf-of-maine-cod.html
This survey with a mix of random and non-random stations and has a short time series.

WG recommendation: The WG agreed that the limitations associated with these other surveys would prevent their use in the model.

Aggregate research survey trends

Trends in abundance and biomass (normalized to the mean within each survey) for 7 surveys are given in Figures B81 and B82 (by grouped by season). The NEFSC seasonal surveys (spring and fall), MA DMF seasonal surveys (spring and fall), and ASMFC summer survey exhibit similar trends in which indices are above their time series mean prior to the late 1980s, below the mean during the late 1980s to mid-1990s and remain below or near the mean in the late 1990's and 2000s onward. The MENH seasonal surveys, which has the shortest time series of all surveys considered here, is more variable than the other surveys, and fluctuates without trend around the mean for spring, with the fall indices exhibiting a slight downward trend.

WG recommendations: Survey indices appear to generally track one another. Length of time series impacts surveys with shorter time series.

FISHERY DEPENDENT DATA INDICES OF ABUNDANCE (LPUE or CPUE)

Introduction

Recent catch limits for witch flounder are much less than historical catch (NEFSC 2015). Therefore, the assessment relies on indices of stock size to determine whether low catches result from low fishing mortality or low stock size. Scientific surveys can inform relative stock size, but sample size (i.e., the number of survey stations) is often insufficient to detect trends in rebuilding stocks (Pershing 2015), and survey gear may be inefficient for sampling some species (e.g., Reid et al. 2007). Many fishermen are concerned that research trawl survey trends are different than their observed catch rates (NSC 2016).

Fishery catch rates can be evaluated as standardized catch per unit effort (CPUE) to consider fishery-dependent data in stock assessments and potentially supplement survey indices in stock assessment models. There are several potential problems associated with using fishery catch rates as indices of stock size (e.g., Harley et al. 2001, Maunder et al. 2006). However, some changes in fisheries can be addressed through effort standardization (Maunder and Punt 2004). Furthermore, the inspection of CPUE trends can help to reconcile different perceptions and provide valuable ecological information, even if these indices are not included in the assessment model.

The management history of New England groundfish has several distinct periods (Murawski et al. 1999, Murphy et al. 2014):

1) before 1970, essentially unmanaged domestic and distant water fleets,
2) 1970-1981, a period of internationally allocated Total Allowable Catches that were not well enforced,
3) 1982-1995, an input control system (including limited entry, limited days at sea, fishing gear, minimum fish sizes and closed areas) for an exclusively domestic fleet that was based on a

socioeconomic definition of optimum yield (i.e., the Fishery Management Council considered short-term social and economic factors in determining fishery limits),

4) 1996-2009, an input control system for the domestic fleet that was based on achieving maximum sustainable yield (MSY), and
5) an output control and catch share system since 2010.

The fishery-dependent data collection system in the northeast U.S. changed in 1994, transitioning from fishermen interviews in a landings intercept program to self-reported logbooks (i.e., vessel trip reports) to obtain information on fishing effort and location. The frequent changes in management, switch in the fisheries-dependent data collection system, and the multispecies nature of the fishery have hindered the ability to develop useful indices of abundance from fishery data. However, a review of New England groundfish stock assessments recommended that "fishers have a greater trust in the data that they themselves provide, and therefore an effort should be made to validate and use CPUE data" (NRC 1998). Fishermen emphasize the need to examine CPUE data in comparison to survey data, and expressed a willingness to share fine-scale information with scientists and managers on fishing locations and catch rates (Hudson & Joyce 2015). Based on a review of historical use of CPUE in the groundfish assessment process, there are opportunities to reconsider this information for upcoming groundfish assessments which could help to reconcile what fishermen see on the water with the results of analytical analyses (O'Keefe et al. 2015).

Previous assessments of witch flounder included landings per unit effort (LPUE; Wigley & Mayo 1996), but LPUE was excluded from the stock assessment model in 1999 because of uncertainty associated with the effect of the 1994 change in effort reporting on CPUE (e.g., Wigley et al. 1999). The 2015 witch flounder stock assessment has severe retrospective patterns (NEFSC 2015), and fishermen are concerned that the research trawl survey trends do not accurately represent stock abundance trends, because the surveys are not efficient for sampling flatfish (NSC 2016).

The Working Group objective was to develop LPUE/CPUE indices of the witch flounder stock for consideration at the 2016 62nd Northeast Regional Stock Assessment Workshop. The benchmark assessment is expected to consider new information and models for determining stock status and providing catch projections. The consideration of LPUE/CPUE trends will help to quantify fishery-dependent perspectives and possibly to supplement research trawl surveys as indices of stock size.

In discussions following initial reviews of 5 commercial fishery LPUE/CPUE working papers (Cadrin and Wright 2016; Terceiro 2016a, b, c, d) using the NEFSC Dealer, Vessel Trip Report, Observer, and Study Fleet data, the Working Group decided to focus on the commercial Dealer landings data (due to the extended length of time series and large and synoptic sample, in spite of concerns about the lack of inclusion of discards and changing reporting systems over time) as the best candidate for a fishery dependent index of abundance to be used as an assessment model calibration index. The WG decided to revise the data and analysis to include only fish trawls and to consider various thresholds of "directed" of trips (trips with witch flounder accounting for =>10%, =>25%, and =>40% of total trip landings) in further work.

<u>Witch flounder Dealer Data</u>

For witch flounder taken in **"All"** trawl gear trips (all negear "050" to "059", to account for nearly all witch landed), the initial nominal data set for 1964-2015 consisted of 280 million lbs of witch flounder caught on 564 thousand trips, accounting for 678 thousand days fished (DF;

recorded fishing time divided by 24; Terceiro 2016a). The mean total landings lbs per DF (LPUE; LDF) was 416 lbs per DF, with a median of 189, mode of 50, variance of 3 million, standard deviation of 1,724, skewness of 56.4, kurtosis of 8,236, and range from 1 to 338,500 lbs per DF.

These descriptive statistics indicate that the Dealer report Trawl gear landings rate distribution is overdispersed in relation to a normal distribution, as the mean is larger than the mode, the variance is several orders of magnitude larger than the mean, and skewness is larger than zero. The distributions of the observed total landings and LPUE were modeled for three classification variables – calendar quarter (QTR; 1 = Jan-Mar, 2 = Apr-Jun, etc.), 3-digit statistical area (AREA), and vessel tonnage class (TC; binned for vessels < 5 gross registered tons [TC = 1], 5-50 [TC = 2], 51-150 [TC = 3], 151-500 [TC = 4], 501-1000 [TC = 5], and 1001 and larger [TC = 6]), expressed as the cumulative sum or mean of the total landings or LPUE for each class level (Terceiro 2016a).

Modeling Standardized Dealer Report Indices

SAS software version 9 (SAS 2011) was used to develop models of the Dealer Report Trawl Gear landings rate data for witch flounder. The independent variable in this case was the continuous variable total landings per day fished (LPUE; rounded to integer values to be 'discrete' where appropriate), while the classification factors considered were, as noted above, the discrete variables YEAR (the 'year' effect that in a main classification factors only model serves as the index of abundance), calendar quarter (QTR), 3-digit statistical area (AREA), and vessel tonnage class (TC).

The SAS GENMOD procedure was used to fit generalized linear models 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 error (probability) distributions. These include the normal, lognormal, binomial, Poisson, gamma, negative binomial (negbin), and multinomial (McCullagh and Nelder 1989). Given the properties of the Dealer trawl data (e.g., Figures B85-B86), the Dealer landings rate data were modeled using lognormal (for ln-transformed landings rates), gamma, Poisson, and negative binomial (for untransformed landings rates) distributions. The procedure fits a generalized linear model to the data 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 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 fit of the expected error distribution to the data (closer to 1 is better) and is comparable across models for a given error distribution, (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, and across models with the same type of data), (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 (Terceiro 2003, Dick 2004, Maunder and Punt 2004, SAS 2011).

A sequence of models, including from one factor to four factors, were fit and the differences/changes in the goodness of fit diagnostics used to determine the best model under each error distribution assumption. A Type III analysis was used since the results do not depend on the order in which the classification factors are specified. For the discrete variable Poisson and negative binomial error distributions, individual trip catch rate values were rounded to integer values. As noted in the 5 previous working papers, the full-factor (YEAR, QTR, AREA, TC) negative binomial models had the best overall diagnostics and were selected as 'best-fitting.' Therefore in this "Directed" trip work, only full-factor negative binomial models were considered.

Results

For "directed" fish trawl (negear code "050") trips where witch flounder accounted for **40%** or more of the trip reported landings, the data set for 1964-2015 consisted of 44.6 million lbs of witch flounder (**16% of the "All" trip total**) caught on 15,713 trips (**3% of "All" trips**), accounting for 17,109 thousand days fished (DF; recorded fishing time divided by 24; **3% of "All" trips DF**). The mean total landings lbs per DF (LPUE; LDF) was 2,609 lbs per DF (6.3 times the "All" trips lbs per DF), with a median of 2,000, mode of 1,250, variance of 26 million, standard deviation of 5,051, skewness of 21.3, kurtosis of 1,099, and range from 2 to 324,000 lbs per DF. Total reported landings (lbs), trips, days fished, and nominal annual 'directed' LPUE are presented in Table B44 and Figures B83-B84. Model indices (negbin model with Year, Quarter, Area, and Tonnage Class main effects) are presented in Table B45 and Figure B85.

For 'directed' fish trawl (negear code '050') trips where witch flounder accounted for **25%** or more of the trip reported landings, the data set for 1964-2015 consisted of 90.4 million lbs of witch flounder (**32% of the "All" trip total**) caught on 46,953 trips (**8% of "All" trips**), accounting for 45,913 thousand days fished (DF; recorded fishing time divided by 24; **7% of "All" trips DF**). The mean total landings lbs per DF (LPUE; LDF) was 1,969 lbs per DF (4.7 times the "All" trips lbs per DF), with a median of 1,492, mode of 1,250, variance of 15 million, standard deviation of 3,827, skewness of 23.4, kurtosis of 1,464, and range from 2 to 324,000 lbs per DF. Total reported landings (lbs), trips, days fished, and nominal annual 'directed' LPUE are presented in Table B46 and Figures B86-87. Model indices (negbin model with Year, Quarter, Area, and Tonnage Class main effects) are presented in Table B47 and Figure B88.

For "directed" fish trawl (negear code "050") trips where witch flounder accounted for **10%** or more of the trip reported landings, the data set for 1964-2015 consisted of 178.5 million lbs of witch flounder (**64% of the "All" trip total**) caught on 141,321 trips (**25% of "All" trips**), accounting for 147,777 thousand days fished (DF; recorded fishing time divided by 24; **22% of "All" trips DF**). The mean total landings lbs per DF (LPUE; LDF) was 1,969 lbs per DF (4.7 times the "All" trips lbs per DF), with a median of 1,492, mode of 1,250, variance of 15 million, standard deviation of 3,827, skewness of 23.4, kurtosis of 1,464, and range from 2 to 324,000 lbs per DF. Total reported landings (lbs), trips, days fished, and nominal annual 'directed' LPUE are presented in Table B48 and Figures B89-B90. Model indices (negbin model with Year, Quarter, Area, and Tonnage Class main effects) are presented in Table B49 and Figure B91.

Comparison among "Directed" trips LPUE and NEFSC indices

The 'Directed' trips nominal LPUE series are compared Figure B92, the model LPUE series are compared in Figure B93, and the model LPUE series are compared with the NEFSC trawls survey biomass indices in Figure B94. The nominal series have a high degree of

correspondence (Figure B92; r > 0.9), as do the model indices (Figure B93; r >0.9). The model series show similar trends from the early 1970s to the mid-1990s, with the exception of the previously mentioned 'directed' 1992 model index "spike" (Terceiro 2016d); further investigation indicates that this is due to the influence of a few high LPUE trips during a period of relatively low (compared to neighboring years) sample size. All LPUE series show a decrease from the 1960s to the late 1970s, an increase to a peak in the early 1980s, a decrease through the 1980s to the mid-1990s, and then an increase to a peak in the early 2000s. Since 2009, the model series diverge, with the less-directed, but more inclusive 10% model index increasing less (+32%) than the more directed 25% (+48%) and 40% (+64%) series (Figure B93).

When comparing the Dealer LPUE with the trawl survey indices, it should be noted that the survey indices include more young fish (<= 3 years old) than do the Dealer commercial indices based on landed fish, so peaks and valleys may not always exactly coincide due to this "fish growth" lag. The NEFSC trawl surveys series decrease more from the early 1980s to the mid-1990s than do the fishery model indices, while the peak in the early 2000s is comparable at about the respective time series mean (i.e., 1.00). Since the early 2000s peak, the NEFSC trawl survey indices have decreased (spring –71% since 2002; fall –69% since 2001), while the fishery model indices are now at about the same level as in the early 2000s (i.e., at about 1.00 to 1.30 their time series means). Since 2009, the trawl survey spring indices have decreased by –10%, while the fall indices have increased by +17%, and are now at about 40% of their time series means (Figure B94).

*<u>WG recommendation</u>: The Working Group evaluated the **Dealer Report**-based trawl gear landings per unit effort indices (1964-2015 Landings per Day Fished; LPUE, LDF), **Vessel Trip Report (VTR)** trawl gear fishery catch per unit effort indices (1994-2015 Catch per Day Fished; CPUE), the **Observer Program** trawl gear fishery catch per unit effort indices (1989-2015 Catch per Day Fished; CPUE), and the NEFSC Cooperative Research Partners **Study Fleet** trawl gear catch per unit effort indices (2007-2015 Catch per Day Fished; CPUE) for their utility as indices of abundance in the witch flounder benchmark stock assessment.*

For the Dealer data, the Working Group noted that the utility of the LPUE series as the basis for an index of abundance is limited in that a) it includes only landings and not the total catch including discard, so that the resulting landings rates could be biased low relative to the true abundance of fish, and b) the use of only positive trips that land witch flounder may bias the landings rate as well. For the VTR data, the WG noted that given the low discard rate reported, the total catch reported may be biased low relative to the true total catch. For the Observer data, the Working Group noted that there may be a bias in fishermen's behavior when an Observer is aboard the vessel that changes the total catch rate in an uncertain and possibly inconsistent manner. The Working Group also noted that for the Dealer, VTR, and Observer data in particular, the ratio of Days Fished per Trip has 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.

During the respective time series there have been a number of spatial, temporal, and catch limit regulatory changes that could bias the commercial fishery catch rate indices relative to the true abundance of witch flounder. The information on these regulatory changes is not part of the commercial fisheries databases and must be developed independently and integrated within the standardization models, and has not been included here. Given the changing regulatory landscape, there likely have been a multitude of changes in fishermen's behavior and strategies over time, and those have not been accounted for here. The increasing sophistication

of fishing technology (fish-finding electronics, positioning electronics, fishing gear) is likewise not easy to include in fishery dependent catch rate modeling, and has not been attempted here.

The Observer Program data were explored to help inform which LPUE series (10%, 25%, and 40%) would be the most appropriate to use in a model sensitivity run (presentation by WG member M. Palmer). Summaries of Observer Program haul and trip level data for observed hauls for which witch flounder was reported as the target species indicated no strong justification to select one series over another. For consistency with previous assessments, the most directed series, LPUE 40%, was selected.

The Working Group agreed to include the commercial LPUE 40% series as a sensitivity run of the assessment model.

CATCH CURVE ANALYSIS OF FISHERY AND SURVEY CATCH-AT-AGE DATA

The Working Group reviewed a catch curve analysis (Palmer 2016). Catch curves were constructed for the aggregate total fishery catch numbers-at-age (combined landings and discards) and the NEFSC spring and fall bottom trawl survey indices-at-age (number/tow) based on the methods of Robson and Chapman (1961). Catch curves were conducted on a cohort basis rather than an annual basis which removed the confounding effects of differential year class strength on the interpretation of catch curve results. For the total fishery catch-at-age, linear regressions were fit to the log transformed catches of ages 7 to 10 for the 1975 to 2006 year classes (Figure B95) and ages 5 to 9 for the 1975 to 2006 year classes for the NEFSC spring (Figure B96) and fall (Figure B97) surveys. While these age ranges may not precisely match the fully recruited ages, it offers a compromise between full selection and having sufficient ages to fit a reliable regression. The slope of the regressions provides a model-independent estimate of cohort Z. The analyses suggest time series Z estimates averaging around 0.4 to 0.5 in the NEFSC surveys with no clear time series trends while the Z estimates from the total fishery catch were marginally higher with some indication of an increasing trend, particularly after 1983 (Figure B98).

Catch curves can also be useful for making general inferences on the selectivity of both fisheries and surveys. While selectivities can be estimated from the fitting of stock assessment models, it is useful to have model-independent estimates of selectivity that can be used to validate model-based estimates and/or provide some *apriori* understanding of selectivity. A method described in Restrepo et al. (2007) uses the residuals from the log-transformed linear catch curve analysis to infer relative selectivity-at-age. Selectivities are estimated using the ratio of observed to predicted catch proportions and then rescaling the residuals from each curve so that the maximum positive residual equals 1. The distribution of selectivities-at-age from all cohorts was examined to evaluate the time series distributions of catch selectivity at age. While this approach masks any changes that may be occurring in the selectivity across time, it is useful for gaining a general understanding of catch and survey selectivities and evaluating whether there is strong evidence for the presence of domed-selectivity (i.e., lower selectivity at older ages). Examination of the residual patterns suggest full selectivity not occurring until ages 6-8 in the NEFSC surveys and age 8 in the total catch (Figure B99). Both the NEFSC survey and total catch residual distributions suggest that there may be some degree of doming beyond the age of full selectivity.

Comparing the ratio of catch-at-age of fishery catch to surveys can be helpful for gaining a qualitative understanding of the selectivities of each (e.g., is the fishery likely to have lower

selectivity at older ages relative to the survey). While these comparisons do not offer definitive estimates of overall selectivity, they are helpful for gaining an understanding of the relationships in a model-independent framework (Clark 2013). The proportion of fish age-6 to 11+ fish caught in the NEFSC surveys relative to the fishery (Table B50). The comparison suggests a higher proportion of age-6 fish and a lower proportion of older fish (with the exception of the 11+ group) caught in the survey relative to the fishery. These results are likely skewed by the fact that fish are not fully selected to the fishery until age-7 or 8 whereas they are selected to the survey around ages 5 or 6. While there is a general trend for lower selectivity of the older ages in the survey, this trend is not consistent with nearly half of the ratios for age 9 and 10 in the spring survey and age 7 and 10 in the fall survey greater than or equal to 1. The NEFSC surveys appear to capture a higher proportion of fish 11+ relative to the fishery. Overall, the results are inconclusive with no clear evidence of doming being greater in either NEFSC surveys or the fishery.

WG recommendations: *Based on this analysis, there was no clear evidence of doming. Model sensitivity to dome selectivity should be conducted.*

TERM OF REFERENCE 3: Investigate effects of environmental factors and climate change on recruitment, growth and natural mortality of witch flounder. If quantifiable relationships are identified, consider incorporating these into the stock assessment.

The WG considered three working papers: (1) changes in distribution of larval, juvenile and adult witch flounder; (2) estimated witch flounder habitat using random forest models; and (3) environmentally explicit stock-recruitment relationships in witch flounder.

Changes in the Distributions of Larval, Juvenile, and Adult Witch Flounder in the Northeast U.S. Shelf Ecosystem (Walsh et al. 2016)

Background

The spatial distribution of larval and adult witch flounder and seasonal occurrence of larvae in the Northeast U.S. Shelf Ecosystem were examined for distribution changes between two decades (Walsh et al. 2015). Larval witch flounder did not shift significantly spatially or seasonally, but adults shifted significantly deeper in the spring when comparing 1977 – 1987 to 1999 – 2008 (Walsh et al. 2015). The analyses were updated through 2015, and include new analyses of juvenile (< 30 cm) witch flounder.

Methods

Methods were described in Walsh et al. (2015). Briefly, the relative proportion (percent of annual sum) of estimated absolute number of larvae, juveniles, and adults within each of 47 Ecosystem Monitoring (EcoMon) strata were used to examine changes in distribution. The EcoMon strata, which combine some bottom trawl survey strata, were used to allow comparisons to be made among the life stages, and the larval strata were originally developed using the stratum design of the bottom trawl survey. Significant differences of stratum proportions between time periods were calculated using a Kruskal-Wallis chi-square. The stratum-specific Kruskal-Wallis H values were considered a measure of the magnitude of change. Three linear regressions were calculated using the strata Kruskal-Wallis H as the independent variable and along-shelf distance (km), cross-shelf distance (km), and depth (m) values as the dependent variables to test whether distribution changes were coherent in the along-shelf, cross-shelf, and depth dimensions.

For larvae, three bimonthly seasons representing the early (March – April), peak (May – June), and late (July – August) larval seasons were examined. We compared the Marine Resources Monitoring, Assessment and Prediction (MARMAP; 1977 - 1987) and EcoMon (1999 - 2015) programs (Figure B100). Timing of larval occurrence was also examined between the time periods to infer changes in timing of spawning, using similar methods used to examine changes in distribution.

For juveniles and adults, the number of witch flounder caught per tow during the NEFSC bottom trawl surveys were calibrated to Albatross units for collections from 2009 onward. No length based correction used, and the number per tow was divided by 3.257177 (Miller et al. 2010) to transform from Bigelow to Albatross units. For juvenile (< 30 cm) and adult (>= 30 cm) flounder, relative proportions of flounder for each NEFSC bottom survey were calculated: spring (March-April) and fall (September-October). Two 20 year time periods, 1970 - 1989 to 1996 – 2015, were compared.

Changes in the capture of witch flounder during the NEFSC bottom trawl survey by habitat (e.g., bottom water temperature and depth) in the Gulf of Maine region strata used in

previous assessments (NEFSC 2015) were examined. Annual stratified mean bottom temperature (^{o} C) and depth (m) was calculated using stratified mean catch per tow of juvenile and adult witch flounder for survey stations. Years with less than 90 % of stations with bottom temperature data were excluded from the mean bottom temperature analysis. Linear regressions were calculated for juveniles and adults during each survey season to determine if there was a significant change in stratified mean habitat. The p-val, r^2, and slope values reported were bootstraped 1000 times runs across years analyzed.

Results

Witch flounder larval distribution and timing of occurrence changed little over the past four decades (Figure B101). Most change in relative larval proportions among stratum occurred during the late-larval season (Figure B101, subplot C; July – August), with lower proportions in seven strata in the Mid-Atlantic Bight and Southern New England regions from 1999 to 2015. However, no significant change in along-shelf, cross-shelf, or depth dimensions occurred (Figure B101, subplots D-F).

Juvenile witch flounder distribution shifted significantly over the past forty years (Figure B102). Juveniles shifted northward and deeper in the spring, with higher proportions in the Gulf of Maine over the last 20 years (Figure B102, subplots A, C, and E). In the fall, juveniles shifted offshore, with higher proportions along the shelf edge in Southern New England, central Gulf of Maine, and Georges Bank (Figure B102, subplots B and D).

Adult witch flounder distribution shifted significantly over the past four decades (Figure B103). Adults shifted deeper in the spring, with higher proportions during 1999 to 2015 in the central Gulf of Maine and lower proportions in the eastern Gulf (Figure B103, subplots A and E). In the fall, adults shifted offshore, with higher proportions along the shelf edge in Southern New England, central Gulf of Maine, and Georges Bank (Figure B103, subplots B and D).

The changes in juvenile and adult witch flounder distributions resulted in significant changes stratified mean habitat. Water temperatures have generally increased 1 – 1.5 ^{o}C in the ecosystem over the survey time period (Fratantoni et al. 2015). Witch flounder stratified mean bottom water temperature did not significantly change during the spring survey, but linear trends have tracked the increasing average bottom water temperature sampled (Figure B104, subplot A). Adult witch flounder stratified mean bottom water temperature significantly changed (p-val ≤ 0.04, df = 37, $r^2 = 0.24$, slope = 0.03) in the fall survey, and witch flounder were caught at or near average temperatures sampled (Figure B104, subplot B). Witch flounder significantly increased stratified mean depth during the spring and fall surveys (Figure B105) for both juveniles (spring: p-val ≤ 0.001, df = 46, $r^2 = 0.33$, slope = -0.83; fall: p-val ≤ 0.001, df = 51, $r^2 = 0.33$, slope = -0.72) and adults (spring: p-val ≤ 0.002, df = 46, $r^2 = 0.27$, slope = -0.53; fall: p-val ≤ 0.02, df = 51, $r^2 = 0.20$, slope = -0.28). Juvenile and adult witch flounder were collected near or above the mean depth sampled during the bottom trawl surveys before about 1985, but now are caught mostly below the mean depth sampled, particularly for juveniles (Figure B105).

Estimated witch flounder habitat using random forest models by Friedland (NEFSC)

The working group considered a series of Random Forest models of witch flounder habitat based on catches in numbers and weight from the bottom trawl survey. Random Forest is a machine learning method associated with a group of classification and regression trees that are a collection of results from a large number of decision trees based on randomized subsamples of the data (Breiman 2001). It is this subsampling that produces models that are robust to the types

of data used, including highly correlated variables and missing data, resistant to over fitting, and provides measures of variable importance and node proximity (Cutler et al. 2007). In a comparison of supervised learning algorithms, Random Forest was found to be among the top performing models for classification given a range of performance criteria (Caruana and Niculescu-Mizil 2006). Increasingly, this type of classifier is being applied to estimate fish habitats, drawing upon environmental variables to inform model construction (Knudby et al. 2010; Lin et al., 2015).

Regression and classification Random Forest models were fit utilizing a mixture of static and dynamic variables. The regression models estimated habitat based on spring and fall biomass and abundance data. The same model formulation was used in all four models: Log10 (biomass/abundance + 1) ~ soft_sed, bpi_30_250_layer, DEPTH, BOTTEMP, chl_4km_month_1, chl_f_4km_month_1, sst_4km_month_1, sst_f_4km_month_1. The soft_sed, bpi_30_250_layer, DEPTH were static variables and BOTTEMP, chl_4km_month_1, chl_f_4km_month_1, sst_4km_month_1, sst_f_4km_month_1 were dynamic variables. The variables were described in the data meeting with the exception of the dynamic fields of bottom temperature used to develop habitat maps, which were estimated use spline interpolations of the spring and fall survey bottom temperatures by year. The same model formulation was in the classification models except the fits were to spring and fall absence presence (factor 0,1).

The regression models all had good fit diagnostics. All four models had pseudo r^2 greater than 0.86 and mean square error rates (MSE) of ~0.0033 for the biomass models and ~0.014 for the abundance models. DEPTH was the most important variable in all the models with BOTTEMP, bpi_30_250, and soft_sed being high contributors. With a model fit for each of the four models, retrospective data for the years 2000-2015 were assembled and spring/fall biomass/abundance were estimated each year. Figure B106 provides an example of a habitat map, in this case 2000 spring abundance estimates based on the spring model. The biomass model analysis suggests witch flounder habitat declined to low levels in the period 2008 to 2013 and showed some sign of recovery in the most recent years (Figure B107). The classification models fits were satisfactory as seen in the confusion matrices for the models. The OOB (out of bag sample used as classification test samples, which make the error rate a cross-validation estimate) classification error rates were 15.99% and 12.41% for the spring and fall models, respectively. Estimates of absence presence habitat time series were not available.

These types of habitat models and habitat estimates based on other procedures such as generalized additive models (GAMS) can be applied in a number of phases of this and other assessments. They provide a framework to analyze the availability of the species to survey gear based on changes in habitat and thus may provide estimates of catchability. The standard approach to the stratification of survey indices of abundance are based on strata that are in part based on ecological aspects of the ecosystem, i.e., depth. Habitat analyses provide alternative approaches to the way survey data are stratified that represent more natural grouping of tows by habitat. These stratification strategies may also be applied to empirical analyses such as swept area estimates of absolute abundance.

Environmentally explicit stock-recruitment relationships in Witch Flounder (Hare et al. 2016a)

Methods

Recruitment was derived from preliminary ASAP model runs. Three preliminary model runs were examined: Run 9 uses all survey indices (NEFSC, ASMFS, and MENH); Run 10 includes LPUE; Run 15 explores uncertainty in catch.

Fall bottom temperatures were derived from the deeper portions of the Gulf of Maine using temperature observations made on the fall trawl survey.

Environment-recruitment relationships were conducted per Hare et al. (2015). The underlying hypothesis examined was that fall temperature determine the amount of available juvenile habitat and warmer temperature reduces available habitat and recruitment (*sensu* Miller et al. 2016). The recruitment provided in the ASAP model output is for Age 1 fish and it is assumed that recruitment in the model output lagged by 1 year would align with the year of spawning and the corresponding SSB.

Results and Discussion

Based on the analyses conducted, there was no evidence of temperature influencing recruitment (Figures B108-B110).

In general, recruitment was high at low spawning stock biomasses and exhibited no relationship with temperature. Steepness in all model runs was quite high – approaching or equaling 1. The Beverton-Holt relationship without environment had the lowest AIC score for all three model runs.

The shape of the recruitment data suggests that a Ricker stock-recruitment relationship may be appropriate, but the density dependence of such a function would be quite high (depressing recruitment at high spawning stock biomasses).

WG recommendations: An examination of available NEFSC trawl survey data indicated that while there have been some 'statistically significant' changes in mean temperature and depth of juvenile and adult witch flounder distribution, the slopes of the linear change are not large and movement has been within the stock definition strata set. There is also evidence of increases in the mean temperature and depth of the habitat occupied by witch flounder, but a model of changes in 'available witch flounder habitat' did not indicate a trend over the time series. Modeling revealed no evidence of temperature influencing recruitment.

TERM OF REFERENCE 4: Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series (integrating results from TOR-3 if appropriate), and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results and previous projections. Compare F's and SSB's that were projected during the previous assessment to their realized values.

> ***SARC-62 Opinions****: The VPA, ASAP, and SCAA age-structured models applied to data for the witch flounder fishery were found to have major retrospective patterns that prevented their use for status evaluation and determination of catch advice.*
>
> *Based on an empirical area swept method (that includes the results of the NEFSC survey efficiency experiment) conducted by the SAW and examined at the SARC, stock biomass has declined since 2002 and appears to be 16,181 mt Jan-1 surveyable biomass in 2016 and 14,563 mt exploitable biomass in 2016.*
>
> *Survey population biomass was converted to exploitable biomass using a factor of 0.90 based on examination of survey and fishery selectivity patterns from the base ASAP assessment.*
>
> *The relative exploitation rate needed to achieve the recent observed catch is approximately 0.05.*
>
> *The fishery landings and survey catch by age indicate truncation of age structure and a reduction in the number of old fish in the population.*
>
> *The time series of recruitment is not available from the empirical area swept method. All the surveys indicate the 2013 year class is relatively strong. However, this year class is not expected to fully recruit to the directed fishery until 2020 at age 7 and should start to appear in the large mesh otter trawl discards in 2017 at age 4. The absolute size of the 2013 year class is uncertain since the estimate is based on only two years of survey observations.*

The WG considered four approaches for estimating annual fishing mortality, recruitment and stock biomass with which to determine stock status and catch advice: (1) minimum estimates of spawning stock biomass; (2) an empirical approach; (3) a replacement yield approach; and (4) three analytical models (VPA, SCAA, and ASAP). A brief overview of each approach is provided followed by the WG discussion on the pros and cons of each approach and why the WG accepted the ASAP model to provide estimates of annual fishing mortality, recruitment, and spawning stock biomass for determining stock status and catch advice.

MINIMUM ESTIMATES OF SPAWNING STOCK BIOMASS FROM NEFSC FALL

SURVEY

The WG reviewed minimum estimates of witch flounder spawning stock biomass using the 2016 experimental estimates of catchability on the NEFSC trawl survey (See Appendix B1for the Hare et al. 2016b working paper and supplement equations.)

A minimum estimate of spawning stock biomass using the survey data and the results of the catchability experiment were developed using the following equation:

$$SSB_t = \sum_{L=1}^{maxL} \frac{(N_{L,t} * W_L * P_{L,t})}{(\delta_L * \rho_L)} * \frac{A}{a}$$

where:

SSB_t = Spawning stock biomass in year t (kg)
N_L= Stratified Mean Number tow^{-1} at Length (L) in year t
W_L= Weight at Length (kg)
P_L= Proportion mature at Length
δ_L = detectability at Length
ρ_L = availability at Length (Set to 1 in this analysis)
A = Total stock area covered by survey (km^2)
a = area covered by a tow (km^2 tow^{-1})

Further information on basic biological parameters is required to implement this equation. The annual parameters used in the stock assessment for the logistic maturity at length equation were:

$$P_L = \frac{1}{1+e^{-(\alpha+\beta L)}}$$

These parameter values for maturity are given below (S.Wigley, NEFSC pers comm):

Maturity parameters used in the evaluation of Spawning Stock Biomass

	2009	2010	2011	2012	2013	2014	2015
a	-9.49773	-10.1677	-9.29966	-9.93273	-10.0779	-10.4224	-10.4224
b	0.289028	0.309791	0.28935	0.304885	0.308803	0.318461	0.318461
L50 (cm)	32.86	32.82	32.14	32.58	32.64	32.73	32.73

The length-weight equation for witch flounder in Burnett et al. 1992 was used ($W_L = aL^b$) where ln(a)=-13.553 and b=3.4384. These values were assumed to be constant across years.

Availability at length (ρ_L) was assumed to be 1 across all age classes. The stock

assessment considers witch flounder present in the Gulf of Maine to be a unit stock, separate from fish off Southern New England (beyond the NEFSC survey strata area). The abundance of witch flounder in the inshore areas of the Gulf of Maine not surveyed by the NEFSC is very low, and thus the assumption that the population occurs completely within the footprint of the trawl survey is reasonable.

Consistent with the stock assessment, the offshore NEFSC trawl strata 22-30 and 36-40 were used. The Area (A) covered by these strata is 75,728 km^2. The analyses focused on data collected by the NEFSC trawl survey since 2009 on the R/V Henry Bigelow. The area covered by an average tow on the NEFSC trawl survey is 0.024 km^2.

Stratified mean number at length (N_L) of witch flounder was calculated using standard approaches for only the fall trawl survey. The rationale for focusing on the fall trawl survey was presented in the September data meeting paper.

Detectability at length (δ_L) was calculated based on the results of the twin trawl experiment (aka sweep study) assuming that the detectability of the chain sweep gear was 1. Two analyses were performed on detectability at length from the catchability experiment, one using a beta-binomial analysis and the other using bootstrapping. The results of the two approaches were very similar. The detectability at length from the 1000 bootstrap runs was used in this analysis. As with previous analyses, we used a weighting factor of 0.57 for the nighttime detectability results and 0.43 for the daytime detectability results to account for the higher proportion of nighttime tows during the fall trawl survey in the Gulf of Maine.
Estimates of minimum spawning stock biomass for the Bigelow survey (2009 to 2015) ranged between 8,100 and 16,600 mt (Table B51).

AN EMPRICAL APPROACH

An empirical approach using witch flounder NEFSC bottom trawl survey data and the survey catchability estimate derived from the recent sweep study (Hare et al. 2016b) was considered by the WG as a backup method to provide catch advice in case modeling approaches did not provide defensible results. The empirical approach used is similar to that used in the Georges Bank yellowtail flounder stock assessment (Legault and Busawon 2016).

The NEFSC spring and autumn time series weight per tow indices (Table B28) were used in which the 2009 – 2016 survey indices were converted to Albatross (ALB) units using a calibration factor of 3.2572 (Miller et al. 2010). Minimum swept area biomass was derived using the witch flounder survey strata set area (22,321 nm^2 converted to 76558.89 km^2), the Albatross net footprint (wing spread; 0.038 km^2), and the Bigelow net footprint (wing spread; 0.024 km^2). Population biomass was derived assuming a Bigelow catchability of 1 and then assuming the chain sweep catchability of 1 (Tables B52 and B53; Figures B111 and B112).

For minimum swept area biomass, the relationship between survey area (*A*) and biomass correlates positively (lower *A* results in lower biomass and higher *A* results in higher biomass); however, the tow footprint (*a*) correlates negatively with biomass (lower *a* results in higher biomass and higher *a* results in lower biomass (Figures B111 and B112, subplots B and C). The Bigelow biomass does not change because the Albatross units account for both the BIG:ALB conversion and the relative size of the tow footprint (5.1572; Tables B52 and B53; Figures B111 and B112, subplots C). The Bigelow and Bigelow to ALB values are identical. The BIG:ALB conversion and the tow footprint impact the joining of the two time series (Figures B111 and B112, subplots D). In the sweep study, the Bigelow catchability is estimated to be 0.291 when

the chain sweep catchability is assumed to be 1. A lower catchability results in a higher biomass and conversely, a higher catchability results in a lower biomass.

To estimate beginning year biomass, a simple average of the NEFSC spring and autumn population biomass estimates is used. To utilize the most recent survey information, the autumn biomass in year t is paired with the spring biomass in year t+1 to provide catch advice for year t+2 (Table B54). Survey population biomass was converted to exploitable biomass using 0.90 based on examination of survey and fishery selectivity patterns from the base ASAP assessment (Table B54; Figure B113). An exploitation rate of 0.16 (based on the ASAP base run Fmsy = F40% = 0.19 and M = 0.15; Table B54) was used to convert exploitable biomass into estimates of the overfishing limit (OFL). The acceptable biological catch (ABC) was estimated using 75% (common practice) of the OFL (Table B54; Figures B114 and B115).

Estimated OFL and ABC are higher than actual annual OFL and ABC that are based on previous assessments and Science and Statistical Committee decisions (Figures B114 and B115). It should be noted that actual catch is calendar year, not fishing year. Historical exploitation rates indicate that an exploitation rate of 0.16 or higher has only occurred three times in the time series (Figure B116) indicating this is too high a value. The working group suggested a value around 0.05, the average of the most recent nine years, would be a more appropriate exploitation rate for this stock given the trends observed in the surveys and the estimated catches. Applying an exploitation rate of 0.05 to the most recent average biomass results in an OFL for 2017 of 728 mt (Table B55, Figures B117 and B118) compared to the OFL for 2017 of 2,330 mt when the exploitation rate is 0.16. Using historical exploitation rates is useful to "ground truth" the choice of exploitation rate. The magnitude (not the trend) of the catch advice will change when choices for q, area, conversions factors, etc. change.

Uncertainties associated with this approach include: the Bigelow to Albatross calibration, the area of the survey, tow footprint, chainsweep:rockhopper calibration, survey conversion between average biomass and exploitable biomass; the exploitation rate apply for catch advice, the appropriate exploitation rate for this stock, and whether any time series smoothing should be applied to the data or catch advice. This approach does not consider the age information and age-based projections cannot be performed using this approach. The working group did not discuss specifics about how to address any of these uncertainties.

REPLACEMENT YIELD MODEL

The "replacement yield" (RY) model is form of an age-aggregated dynamic surplus production model approach using a time series of catches and one or more indices of abundance with associated measures of precision, if available. When applied to witch flounder, the approach utilized catch, the NEFSC spring and autumn bottom trawl survey information and the 40% LPUE series and was conducted with a selection of 3 time periods, 4 values of q and with and without LPUE and NEFSC survey abundance series. See Appendix B2 for Butterworth and Rademeyer's working paper entitled "Replacement Yield Model Assessments of Gulf of Maine-Georges Bank Witch Flounder".

VIRTUAL POPULATION ANALYSIS (VPA) MODEL

The existing assessment model is VPA (Virtual Population Analysis v3.4.4), which can be obtained from the NOAA Fisheries Toolbox (NFT 2014b, http://nft.nefsc.noaa.gov/). As described at the NFT software website, population cohorts are estimated by a backward projection method. Catch is assumed to be measured without error. Population estimates are chosen so as to minimize the sum of squares difference between the population abundance and a set of one or more abundance indices. Catchability coefficients are estimated as functions of population estimates and observed indices. Bootstrapping is used to estimate the precision of all model parameters and all quantities that are functions of model parameters.

The current assessment used VPA v3.4.5 software, the same as used in the previous 2015 Operational Assessment. The VPA formulation was the same as the previous assessment and used USA total catch (landings and discards for ages 3 to 11+; Table B24) through 2015 and NEFSC spring and autumn survey abundance indices (ages 3 to 11+ through 2016 and 2015, respectively, to estimate stock sizes for ages 3 to 10. All indices are given equal weighting. Autumn survey indices are lagged forward one year and one age to calibrate with beginning year population sizes of the subsequent year.

For all VPA runs, NEFSC spring and autumn relative abundance indices at age (Table B30) were transformed into swept area absolute abundance indices (swept area numbers, SWANs in F/V Albatross (ALB) units using the witch flounder survey strata set area of 22,321 nm^2; a trawl swept area [wing spread] of 0.0112 nm^2; survey catchability (q) = 1 for Runs F1 and A1 and q = 0.056 for Run A2) and used as tuning indices to explore changes in model-estimated survey catchabilities. In previous assessments, the SWANs were derived using an assumed survey catchability of 1. The VPA runs for the current assessment used the survey time series with a split between 1994 and 1995 as was done in previous assessments to reduce the retrospective pattern. A flat-top partial recruitment vector is assumed, with full fishing mortality on ages 8 and older (F). The F on ages 10 and 11+ in all years prior to the terminal year is derived from the weighted estimates of Z on ages 8 and 9. Instantaneous rate of natural mortality is assumed to be 0.15. Spawning stock biomass is calculated at time of spawning (March) and mean weights at age calculated by the Rivard method (NFT 2008). Annual maturity ogives are estimated using NEFSC spring maturity at age data through 2016, pooled by 5-year moving time blocks.

Two VPA bridge runs were conducted to step through the changes from the previous assessment accepted run (VPA Run F): (1) VPA Run F1 used the revised catch at age (1982-2014) that incorporated the revised discard estimates; (2) VPA Run A1 incorporated an additional year of data to the catch at age (1982-2015) and maturity ogive. The VPA Run A2 was identical to VPA Run A1 except that it used NEFSC survey swept area numbers (SWANs) derived using the witch flounder survey catchability of 0.056. Although no change to the estimates of F or SSB were expected between VPA Run A1 and VPA Run A2, the comparison was performed to examine changes in the model estimated qs. See Appendix B4 for VPA Run A2 input. No VPA model sensitivities were conducted.

The Working Group reviewed the bridge runs and noted the discard estimation refinement had minimal impact on estimates of F and SSB in the terminal year (Table B56). The VPA diagnostics indicate similar patterns among the runs; however, there was a change in the model-estimated survey catchabilities between Run A1 and A2 as the result of using q=0.056 in the SWANs (Table B56); the qs exceeded 1 in Run A2.

VPA model diagnostics and results are given in Tables B56-B59 and Figures B119-B129. The VPA RUN A2 had a mean square residual of 0.676, the CVs for estimated stock size at age ranged between 29% and 60% (Table B56), and the CVs for survey catchability coefficients (q)

were consistent, ranging from 12% to 43%. In VPA RUN A2, the 1982-1994 series qs ranged between 0.1 and 4.3 and the 1995-2015 qs ranged between 1.3 and 6.8, exceeding the expected q =1 (Figure B119). The pattern of increasing survey catchabilities at age for younger fish and a general level pattern at older ages appeared reasonable; however, the magnitude of the qs indicates a model scaling issue. The cause of the increased q between the 1982-1994 and 1995-2015 series remains unknown.

The model-estimated partial recruitment pattern for the most recent five years is given in Figure B120 and reveals a flat-top selectivity at age pattern for the catch.

Residual patterns from the NEFSC survey tuning indices and indices at age from the VPA RUN A2 are given in Figures B121 and B122. The patterns appear random for most ages; however, for some ages (7 and 10) there appear to be blocks of positive and negative residuals.

Estimates of catch, SSB, F, Age 3 stock size are given in Table B57 and Figure B123 and beginning year stock sizes, F and SSB at age are given in Table B58. Results indicate average fishing mortality (ages 8-9, unweighted) increased from 0.26 in 1982 to 0.70 in 1988, declined to time series low of 0.23 in 1992, increased to a time series high of 1.14 in 1996 and fluctuated until 2012 and then subsequently declined to 0.43 in 2015 (Tables B57 and B58; Figure B123). Spawning stock biomass declined from a time series high of 16,677 mt in 1982 to 3,927 mt in 1996, increased to 6,827 mt in 2000 and then declined to time series low of 2,134 in 2013 mt; SSB remains low at 3,044 mt in 2015 (Tables B57 and B58; Figure B123). Since 1982, recruitment at age 3 has ranged from approximately 3.5 million fish (2007 year class) to 19.1 million fish (1980 year class) with a mean of 9.8 million fish (Tables B57 and B58; Figure B123. The addition of the recent year class to the stock-recruit data continued the negative trend observed in this relationship in the previous assessment (Figure B123). As in previous assessments, the Age 3 stock size in terminal year + 1 (2013 year class) is poorly estimated (60% CV; Table B56).

The retrospective pattern in previous assessment is still present in VPA RUN A2, in which average F was underestimated and SSB was overestimated, and the magnitude of the error was similar to the previous assessment as measured by the Mohn's rho statistic. The Mohn's rho statistic (Mohn 1999) was derived by taking the average of seven (2009 – 2015) relative differences between the quantity (e.g., F, SSB, and Age 3) from the reduced time series assessment and the same quantity from the full assessment. The VPA RUN A2 Mohn's rho statistics for F, SSB and Age 3 was -0.38, +0.50 and +0.69, respectively (Table B56; Figures B123 -B125).

The precision of the 2016 stock size at age, F at age in 2015, and SSB in 2015 from the VPA RUN A2 was evaluated using bootstrap techniques. Bootstrap results of the RUN A2 suggest that the estimates of F and spawning stock biomass are relatively precise with CVs of 19% and 11%, respectively. The 90% confidence interval (CI) for F_{2015}= 0.43 was 0.33 and 0.59, and for SSB_{2015} = 3,044 mt the 90% confidence interval was 2,601 mt and 3,740 mt (Table B59). The range of the bootstrap estimates and the probability of the individual values are presented in Figure B127.

An important source of uncertainty in the VPA Run A2 model is the major retrospective pattern wherein fishing mortality is underestimated and spawning stock biomass and recruitment are overestimated. The 7-year Mohn's rho, relative to SSB, was +0.51 in the 2015 assessment and was +0.50 in the VPA Run A2 and Mohn's rho, relative to F, was -0.38 in the 2015 assessment and was the same in the VPA Run A2 (Table B56). The VPA Run A2 retrospective pattern was considered to be a major retrospective pattern for this model because the rho adjusted estimates of 2015 SSB (SSB_{rho} = 2,035) and 2015 F (F_{rho} = 0.69) were outside the

approximate 90% confidence region around SSB (2,601 – 3,740) and F (0.33-0.59; Table B59; Figure B128).

Comparisons of time series estimates of F, SSB and Age 3 recruitments for VPA Runs F (2015 OA), F1 (bridge run), and A2 (updated model) reveal similar trends among VPA runs (Figure B129).

STATISICAL CATCH AT AGE (SCAA) MODEL

See Appendix B3 for Butterworth and Rademeyer's working paper entitled "Exploratory Catch-at-Age Assessments of Gulf of Maine-Georges Bank Witch Flounder".

AGE STRUCTURED ASSESSMENT PROGRAM (ASAP) MODEL

The new assessment model is ASAP (Age Structured Assessment Program v.3.0.17, Legault and Restrepo 1998), which can be obtained from the NOAA Fisheries Toolbox (NFT 2014a, http://nft.nefsc.noaa.gov/). As described at the NFT software website, ASAP is an age-structured model that uses forward computations assuming separability of fishing mortality into year and age components to estimate population sizes given observed catches, catch-at-age, and indices of abundance. Discards can be treated explicitly. The separability assumption is partially relaxed by allowing for fleet-specific computations and by allowing the selectivity at age to change in blocks of years. Weights are input for different components of the objective function which allows for configurations ranging from relatively simple age-structured production models to fully parameterized statistical catch at age models.

The objective function is the sum of the negative log-likelihood of the fit to various model components. Catch at age and survey age composition are modeled assuming a multinomial distribution, while most other model components are assumed to have lognormal error. Specifically, lognormal error is assumed for: total catch in weight by fleet, survey indices, stock recruit relationship, and annual deviations in fishing mortality. Recruitment deviations are also assumed to follow a lognormal distribution, with annual deviations estimated as a bounded vector to force them to sum to zero (this centers the predictions on the expected stock recruit relationship). For more technical details, the reader is referred to the technical manual (Legault and Resprepo 1998).

The WG considered a wide range of sensitivity runs to the base cases of the NEFSC ASAP and Butterworth/Radermayer SCAA models. These included (but were not limited to):

1) split in the NEFSC surveys (SVs) in 1993/1994, a configuration used in previous assessments to allow a change in catchability (q) to reduce the retrospective error in the VPA models used in previous assessments

   WG concluded that the NEFSC SV split in 1993/1994 was no longer effective in reducing the retrospective error, and dropped this aspect from further models.

2) flexibility of the models in fitting "time-blocks" for fishery selection: 1 block (1982-2015), 2 blocks (1982, 1993), and 3 blocks (1982, 1993, 2001-2005 [2005 selected])

WG concluded that the 3 blocks for fishery selectivity provided best model diagnostics, consistent results between ASAP and SCAA, and was roughly coincident with changes in the fishery and management.

3) split in the NEFSC SVs in 2008/2009, a configuration that removes the need to calibrate current Bigelow (BIG) survey indices to the historical Albatross (ALB) series, and to observe the model response in q

   WG considered pros and cons of retaining the ALB calibrated series vs. splitting the NEFSC SVs into independent ALB and BIG series.

   The WG noted several reasons to retain the ALB calibrated series: BIG series too short to stand alone; split perhaps allows too much flexibility, with some counterintuitive results in terms of ALB vs BIG q; limited age structure in short BIG series results in lack of precision in estimates of survey selectivity; effectively includes results of calibration experimental results (ALB to BIG 3.2572).

4) flexibility of the models in fitting fishery selectivity above the age of full (apical) selection, typically age 7 (block 1982-1992) or 8 (1993-2004, 2005-2015)

   WG concluded that the model diagnostics do not strongly support high flexibility to estimate fishery selection above age 7-8. Likewise, analyses external to the analytical models (catch curve analyses, distribution of proportion at length in different bottom types (Wigley 2016), consideration of size at age) do not provide strong evidence for dome shaped fishery selectivity. Going forward, WG elected to use flat top fishery selectivity, and will provide a sensitivity showing the effects of more flexibility for the final model configuration.

5) flexibility of the models in fitting survey selectivity above the age of full (apical) selection, typically age 7 in NEFSC SVs, and younger in the ASMFC shrimp (age 3-4) and MENH trawl surveys age (age 1)

   WG concluded that the model diagnostics do not strongly support high flexibility to estimate NEFSC survey selection above age 7. Likewise, analyses external to the analytical models (catch curve analyses, distribution of proportion at length in different bottom types, consideration of size at age) do not provide strong evidence for dome shaped survey selectivity. Going forward, WG elected to use flat top survey selectivity, and will provide a sensitivity showing the effects of more flexibility for the final model configuration.

6) for NEFSC SVs expanded to swept-area numbers using ALB q = 0.056 and BIG q = 0.291, fixing model aggregate catchability of the NEFSC SVs at 1, to observe the response of other model calculated/estimated parameters and residuals

   WG concluded that fixing NEFSC SV q = 1 was effective in demonstrating model responses, which were comparable for ASAP and SCAA, but did not provide plausible

results for some estimated parameters within the analytical models (i.e., implausible fishery and survey selection patterns).

7) inclusion of fishery LPUE (witch flounder "directed" trips) as an index of stock biomass in model calibration

   The WG included the Dealer data LPUE series to examine the response of the ASAP and SCAA models; models behaved similarly. The WG noted in the Data Meeting that (1) the utility of the Dealer Report data as the basis for an index of abundance is limited in that they 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 land witch flounder may bias the LPUE as well, and may be influenced by management regulations, and (3) the ratio of Days Fished per Trip has 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. Because of the unknown bias concerns, and issues in model diagnostics when the LPUE is included with survey series as calibration indices, the WG elected not to include the Dealer Directed Trips LPUE in final age-structured model development.

8) increase in landings since 1996 by varying scalar factors (up 5 times), to reduce to retrospective error and observe the model response in calculated qs

   The WG found that increasing the catch or M in the ASAP model was effective in reducing the retrospective error in the model. The magnitude of the increase in catch (3.0 to 5.0 times) or M (2.5 to 3.0 times) needed to reduce the retrospective error to a small amount also improved the residual diagnostics for the fishery catch components and marginally improved the residual diagnostics for all surveys, and reduced the NEFSC survey qs. The timing of the "step-increase" in fishery catch informed by large changes in model diagnostics in the early 2000s and was coincident with a change in catch reporting in 2004. Uncertainty in the catch of witch flounder has increased due to recent allegations of catch misreporting currently under litigation. Other than this information, the WG does not currently have access to any publically available independent evidence or justification for such a large unaccounted fishery catch or large increase in M over the last decade. The WG recognizes that such large removals could alias other unknown mortalities.

9) change in the "plus-group" from age 11 and older (11+) to age 9 and older (9+), to examine effects on "post-apical" age selectivity

   The WG examined the response of the ASAP model to this change, and found that the NEFSC SVs q decreased by a small amount, and the decrease in selectivity for the plus group was still large for the run that allowed for this flexibility (down from S = 1.0 at age 8 to about S=0.40 at age 9+). Therefore, this sensitivity run was not very helpful in informing the specification of the fishery selection pattern.

10) for NEFSC SVs expanded to swept-area numbers using ALB q = 0.056 and BIG q = 0.291, use door spread instead of wing spread in the area expansion (reduces SWAN), to observe the response of the model

The WG noted that expansion factor of wings to doors was not sufficient to reduce NEFSC SV qs to the expected value of 1.

The final age structured model configuration (ASAP Base Run) included the following:

1) Three fishery selection blocks: 1982-1992, 1993-2004, 2005-2015
2) CAA ages 1-11+
3) Flat fishery selection with S = 1.0 at age 7+ (1982-1992) and 8+ (1993-2004, 2005-2015)
4) Flat NEFSC SV selection with S = 1.0 at age 7+
5) Use NEFSC SV Albatross calibrated/catchability raised wing-spread SWAN indices in calibration, i.e., no splits and net wing spread
6) Include ASMFC Shrimp indices (with S=1.0 at ages 3 and 4) and ME/NH trawl survey indices (with S=1.0 at age 1) in calibration
7) i.e., ASAP Model 9_5, which became ASAP Run 9_5_v2 (with corrected month from 4 to 10 for MENH autumn; tweaked effective sample size [ESS; Francis 2011]: catch 67, NEFSC spring 46, NEFSC autumn 35, ASMFC summer 30, MENH spring 20, and MENH autumn 24)

See Appendix B5 for ASAP Run 9_5_v2 input.

Sensitivity runs

1) Allow estimation of fishery selection for ages 8+/9+
2) Allow estimation of survey selection for NEFSC SV ages 8+
3) Include Dealer 40% Directed trips LPUE in calibration
4) Any adjustment of SV q prior CVs (lambda for catchability = 1 with CV = 0.001)
5) Effective sample size specification – apply Francis ESS (catch 27; NEFSC spring 21; NEFSC autumn 12; ASMFC summer 11; MENH spring 8, and MENH autumn 14)

Pros and cons of models for catch advice are described in a subsequent section.

ASAP run 9_5_v2 Model Fit Diagnostics

The objective function components that contribute to total likelihood and the fleet and index selectivity parameters (and their associated CVs) for ASAP Run 9_5_v2 are given in Tables B60 and Table B61, respectively. In all ASAP model runs, recruitment was modeled as random deviations from a mean (the stock-recruitment relationship for witch flounder is not well determined). In ASAP Run 9_5_v2, the root mean square errors (RMSE) for catch is close to 1 while the indices are outside the 80% confidence bounds of the median (Figure B13). The CVs for the catch and indices were increased to bring the RMSE closer to the bounds, but avoided excessive down weighting of the indices relative to catch. The CV scalars used are: x1.5 on

NEFSC spring index; x1.75 on NEFSC autumn index; x2.5 on ASMFC index; x2.5 on MENH spring index; and x4 on MENH fall index.

Total catch assumed CV = 0.1. The residuals of total catch revealed a block of positive residuals in early in the time series and a block of negative residuals late in the time series (Figure B131); however, the scale of residuals range between -2 and 2, and a roughly normal distribution of the probability density function of the residuals indicates an adequate fit despite the block pattern. For the age composition of total catch, the input effective sample size for total catch was derived using the median of the model estimated effective sample sizes from a penultimate model run (Figure B132). The total catch age composition residuals are given in Figure B133.

Model building started with one fleet selectivity block and additional blocks were added to break up the runs of positive and negative residuals. The year for the block break was based on the best reduction in the objective function value. As stated above, three fleet selectivity blocks were used: 1982-1992, 1993-2004, and 2005-2015 with full selectivity at ages 7+ for block 1 and 8+ for blocks 2 and 3 (Figure B134). CVs for the model estimated selectivity at age ranged between 10% and 86% for block 1, between 10% and 24% for block 2, and between 10% and 64% for block 3 (Table B61), with higher CV occurring at younger ages.

For ASAP Run 9_5_v2, five tuning indices were used: NEFSC spring (Index 1), NEFSC autumn (Index 2), ASMFC summer (Index 3), MENH spring (Index 4) and MENH autumn (Index 5). Model fit to total indices are given in Figures B135-B139. Generally, for the NEFSC spring and autumn and ASMFC summer indices, the residuals revealed a block of negative residuals in early in the time series and a block of positive residuals late in the time series (Figures B135-B137); however, the scale of residuals generally range between -3 and 3, and a roughly normal distribution of the probability density function of the residuals indicates an adequate fit. The MENH spring and autumn residuals also ranged between -3 and 3, and the pattern of positive and negative residuals blocks are present but not as evident due to the shorter time series of this survey (Figures B138-B139). The input effective sample size for each index was derived using the median of the model estimated effective sample sizes for each index from a penultimate model run (Figure B140). Model fit residuals for indices at age are given in Figures B141-B145. Index selectivity was fixed at 1.0 for ages 7+ for NEFSC spring and autumn, fixed at ages 3 and 4 for ASMFC summer and fixed at age 1 for MENH spring and autumn (Figure B146). The CVs for the model estimated selectivity at age ranged between 8% and 17% for spring and autumn NEFSC indices while CVs for ASMFC and MENH the model estimated selectivity at age ranged between 15% and 167%, with higher CV occurring at older ages (Table B61).

The expected model estimated survey catchability was 1 for the NEFSC spring and autumn indices because the swept area numbers were derived using the witch flounder survey catchability of 0.056; however, the model estimated qs were roughly 4 (Figure B147), indicating a scale issue. This persisted across all runs, except runs in which q was fixed (Table B60). Model estimated q for the other surveys were less than 1; the ASMFC and MENH indices used an assumed q=1 in the derivation of swept area numbers (Figure B147).

ASAP Run 9_5_v2 Model Results

ASAP Run 9_5_v2 model results are presented in Tables B62-B65 and Figures B148-B153. Estimated stock size at age (numbers of fish) from 1982 to 2015 is given in Table B62. Age 1 recruitment ranged between 2.8 and 46.4 million fish with a mean of 14.0 million fish

(Figures B149). A period of above average recruitment occurred in the 1990s and roughly average recruitment since 2009 except in 2014 in which the 2013 year class appears strong (Figure B153). The 2013 year class is uncertain since the estimate is based only on two survey data points. Recruitment CVs are generally less than 15%, except for the 2014 year class (Figure B152). Estimates of biomass (January 1, spawning stock, and exploitable) are given in Table B63 and Figure B153. Spawning stock biomass sharply declined in the 1980s to mid-1990s and has been somewhat stable since (Figure B148); CVs are less than 10% (Figure B154). Catches were 2 to 3 times greater earlier in the time series with older fish landed than later in the time series when fewer older fish are landed. The 2015 SSB is estimated to be 5,479 mt. Witch flounder stock-recruitment relationship is not well determined (Figure B152). Estimated fishing mortality has fluctuated between 0.16 and 1.09 over the time series (Table B64; Figure B148); in recent years, there is a declining trend. The fishing mortality CVs are less 17% (Figure B152). Fishing mortality (F_{full}) was estimated to be 0.16 in 2015.

ASAP 9_5_v2 Retrospective Analyses

An "internal" retrospective analysis for ASAP Run 9_5_v2 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 from terminal years back to 2009. A retrospective pattern was evident in which fishing mortality was underestimated and SSB and Age 1 recruitment was overestimated (Figure B154). The 7-year Mohn's rho was -0.46 for F and 0.64 for SSB. This was considered a major retrospective pattern because the rho adjusted estimates of 2015 F (F_{rho} = 0.29) and 2015 SSB (SSB_{rho}=3,335 mt) were outside the approximate 90% confidence region around F (0.12-0.20) and SSB (4,812-6,511; Table B65). A retrospective pattern has occurred in the previous assessments since 1999.

ASAP Run 9_5_v2 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 sizes in 2016 for future projections. The trace of 1982 and 2015 MCMC chains for both F and SSB (Figures B155-B156, respectively) appear random with negligible autocorrelation (Figures B157-B158) implying the thinning rate of MCMC chains was appropriate. The 1982 and 2015 MCMC distributions about the point estimates are generally centered about the point estimate of F and SSB and appear reasonable (Figures B159-B160).

ASAP Run 9_5_v2 Sensitivity Runs

The Working Group identified five sensitivity runs of ASAP Run 9_5_v2 to evaluate the impacts of: (1) domed fleet selectivity; (2) domed survey selectivity; (3) including LPUE 40% trips as an additional tuning index; (4) fixed q = 1 for NEFSC spring and autumn indices; and (5) down weighting the importance of the catch at age information (Francis ESS). The component contribution to the objective function changed minimally between the base run (Run 9_5_v2) and Sensitivity Runs 1-3 (fleet with domed selectivity [FLEETDOME], NEFSC surveys with domed selectivity [SVDOME], and inclusion of the LPUE 40% trips [LPUE40], respectively; Table B60). However, the contribution of the total catch, catch age composition, total index, and

index age composition differ for Sensitivity Run 4 (in which NEFSC survey q was fixed = 1 [QFIXED] and Sensitivity Run 5 (effective sample size reduced [FRANCIS]; Table B60). The objective function is largest in Sensitivity Run 4 and is smallest for Sensitivity Run 5 in which the importance of the catch at age information has been reduced by adjusting the effective sampling size adjusted downward. Model estimated selectivity at age, and their associated CVs are given in Table B61.

Comparisons of fishing mortality, spawning stock biomass, and stock size (Age 1 and total) estimated from ASAP Run 9_5_v2 and the five sensitivity runs are given in Figures B161-B163. Model results were robust to fishery and NEFSC survey selectivity (Sensitivity Runs 1 and 2). When the LPUE 40% index was included in Sensitivity Run 3, F was slightly higher than the base run between 1982-1999 and lower than the base run between 2000 and 2015, but remained within 2 standard deviations of F in the base run except for year 2010-2015. The SSB of the LPUE 40% Sensitivity Run 3 was lower than the base run early in the time series and higher than the base in later years but remained within 2 standard deviations of the SSB in the base run for all years except 1983-1987 and 2005-2015. When the effective sample size was adjusted downward in Sensitivity Run 5 to reduce the importance of the catch at age information, F was slightly lower and SSB was slightly higher than the base run but both F and SSB generally remained within 2 standard deviations of the base run.

In Sensitivity Run 4 the survey catchability (q) for the NEFSC survey SWAN indices were held fixed at 1 (calculated assuming an ALB efficiency of 0.056, providing freely estimated q of approximately 4). The Sensitivity Run 4 results were, as expected, significantly different from ASAP Run 9_5_v2, with F scaled significantly lower over the entire series and SSB, age 1 recruitment, and total stock numbers scaled higher, especially over the terminal 10 years (Figure B163). This difference in scaling, combined with the presence of the major retrospective pattern, led to further explorations (described below) of the response of the model to possible sources of missing mortality.

Two addition sensitivity analysis were conducted. A likelihood profile over assumed natural mortality was conducted using ASAP Run 9_5_v2 and explorations of missing mortality were conducted using ASAP Run 9_5 (a penultimate run of ASAP Run 9_5_v2).

Likelihood Profile over assumptions for Natural Mortality (M)

The ASAP Run 9_5_v2 model was run over a range of assumed M values from 0.05 to 0.50 (constant at all ages over time) to help determine which assumption for M fit best, given the diagnostic of total minimum log-likelihood (value of the total objective function). Figure B164 shows the likelihood did not find minima within the range of values explored. Given the occurrence of age 15 through 21 year old fish in the 2015 landings, it is unlikely that M would exceed the range of values explored.

Missing Mortality sensitivity runs

The Working Group considered an analysis exploring missing mortality and the response to the mean Mohn's rho value for spawning stock biomass (SSB_{rho}) based on 7 peels (a presentation by C Legault). The ASAP Run 9_5 (a penultimate run to ASAP Run 9_5_v2) was used to determine how the SSB_{rho} changed when morality changed by either increasing natural mortality or by increasing catch. The response of SSB_{rho} over a range of year breaks (1996-2005;

the year in which the increase begins) and over a range of multipliers (from 1.5 to 5.0 by 0.5 steps) for catch or M are summarized in Figures B165-B166. The catch proportions at age do not change as the multipliers changed. The M multiplier was applied to all ages.

The exploratory analysis revealed the SSB_{rho} can be reduced by a variety of combinations of increased catch or increased M over a range of year breaks; however, the NEFSC survey catchability (q) remained an issue. For example, when catch was increased by 4 times starting in 2002, the NEFSC SWAN indices qs ranged between 2.1 and 2.5 (still large compared to the expected qs = 1) and when M was increased by 3 times starting in 1997, the qs ranged between 3.2 and 3.6. Different sources of missing mortality (catch or mortality) produced different trends in SSB and F (Figure B167).

The Working Group concluded that there are many ways to eliminate the retrospective pattern that is present in the witch flounder model; however, all require large changes and while the survey q issue improved by increasing M or catch, a q different from the expected value of 1 still remained.

WORKING GROUP SELECTION OF MODEL for Status and Catch Advice

The Sweep Study document (estimation of NEFSC survey efficiency/catchability; Hare et al. 2016b) provided a useful example of how an absolute estimate of SSB (or another form of stock size) might be computed, and the results provided context for the age structured model results. Since no estimate of exploitable biomass (EB) was directly provided, there is no means from this approach alone to provide advice on the OFL 2017.

The "Empirical Assessment" documentation (i.e., the approach used in the Georges Bank yellowtail flounder TRAC assessment; Legault et al. 2014) incorporates the results of the Sweep Study and uses the two NEFSC survey series and the reported/estimated catch to provide a simple basis for the calculation of OFL for 2017. This approach does not provide biological reference points, does not use a "formal" model framework, does not use the available fishery and SV catch-at-age (CAA) and other biological data, and is strictly deterministic and so does not provide any metrics of uncertainty. Given these constraints, it is not possible to project the impact of catch on stock size.

The RY model is a survey and catch model framework that is simpler than the ASAP and SCAA age structured models that also incorporates the results of the Sweep Study and uses the two NEFSC survey series. This approach can provide model diagnostics to help determine the best model configuration to provide a simple basis for the calculation of catch OFL for 2017. Unfortunately, the WG was unable to sufficiently vet (compared to the work that was done for the ASAP and SCAA age structured models) the RY model performance; e.g., with regard to retrospective errors and impact of the inclusion of the ASMFC and MENH indices. This approach does not provide biomass or F reference points, appears to be sensitive to the times series of data included in the two examples that were examined, and does not use the available fishery and SV CAA and other biological data.

The WG noted that the latter two "simple" models (Empirical, RY) use a subset of the same fishery and survey data used in the more complex ASAP and SCAA age structured models. The Empirical and RY methods do not make some diagnostic issues as obvious, and do not highlight some conflicts in the data. Data conflicts are important for understanding sources of uncertainty and risk. The same conflicts/issues that are of concern in the age structured models (potentially unaccounted for catch, incorrect assumption in the swept area expansions, incorrect

assumption for M) may simply be "masked" in the simple models because they do not include the fishery and survey age structure and so don't explicitly account for fishery and survey selectivity. Witch flounder reaches full maturity at age 8-9, with full recruitment to the fishery at age 7. Age structure may be important for understanding annual dynamics, and for projecting future catches. The Empirical and RY approaches do not explicitly capture these important features.

The WG notes that in previous witch flounder assessments using VPA age structured models, using NEFSC SWAN indices assuming efficiency = 1 and a survey split in 1994/1995, freely estimated NEFSC SV qs were less than 1. These SWANs were proportionally smaller than those used in the current assessment using the Sweep Study results, where ALB efficiency = 0.056 and BIG efficiency = 0.291. With the benefit of the Sweep Study efficiency estimates, the ASAP and SCAA age structured final models examined in this assessment all provide freely estimated values of NEFSC SV q in the range of 4-5, depending on model configuration and weighting, and provided similar diagnostics. Numerous sensitivity runs examining changes in the magnitude of the fishery catch, M, selectivity patterns, and inclusion of fishery CPUE indicated that some configurations could provide a wide range of freely calculated estimates of NEFSC SV qs less than 4.

The WG notes that for many of the age structured model configurations examined, the assessment results do not differ when the survey q is simply a freely estimated parameter, but the resultant high qs and major retrospective errors are diagnostics that indicate some combination of the component input data (e.g., unaccounted catch, incorrect Swept Area Numbers, M) are incorrect. Sensitivity runs that force the models to approach or match the expected NEFSC survey SWAN index q = 1 provided diagnostics that show that models have difficulty reconciling patterns in model starting conditions, catch, aggregate survey trend, and the information on selectivity and magnitude of F provided by the fishery and survey CAA data. In summary, the age structured models in the accepted final configurations provide calculated NEFSC survey q ~ 4.0, exhibit major retrospective patterns requiring rho adjustment, and require very large increase in catch or M to remove the retrospective errors. The WG notes that these data conflicts have been addressed in the previous accepted and current age structured models explicitly through a rho adjustment, without having to apply potentially implausible scalars on M or catch (as has been done in this assessment as sensitivity examples).

The WG recognizes that the new information from the Sweep Study and resultant impacts on the estimation of NEFSC SV q can lead one to question the validity of the ASAP model assumptions and results. A minority of the WG views the ASAP final model freely estimated values of q ~ 4 in the face of the Sweep Study results as a "fatal error" that precludes its direct use for status evaluation and OFL advice. However, a majority of the WG does not feel this diagnostic is sufficient justification to reject the ASAP final model (Run 9_5_v2, adjusted for retrospective error in the terminal year) as the basis for determining stock status and providing advice on 2017 OFL. Therefore, **the WG has used the ASAP final model results as the basis for status evaluation and advice on the 2017 OFL**.

Comparison of VPA and ASAP model results

Trends in fishing mortality, spawning stock biomass for VPA Run A2 and ASAP Run 9_5_v2 (with associated 2 standard deviations) are given in Figure B169. The general pattern over the time series of fluctuating fishing mortality and declining spawning stock biomass are

similar between the two models; however, the VPA Run A2 estimates are not within 2 standard deviations of the ASAP Run9_5_v2 except for a few years throughout the time series.

A HISTORICAL RETROSPECITVE ANALYSIS

A historical retrospective analysis was conducted to allow a comparison with previous assessment results and previous projections using stock assessments from 2008, 20102, 2015 and current assessment model (ASAP Run 9_5_v2). The retrospective pattern is evident for F and SSB (Figure B169). Because the previous projection was conducted in 2015, there was only one year in which there was a projected F and SSB estimate and an estimate from this assessment. The projected estimates of F and SBB from the 2015 assessment are just outside the 90% CI of the retrospective adjusted 2015 estimates F_{rho} and SSB_{rho} from this assessment.

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 BMSY, BTHRESHOLD, FMSY 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.

> ***SARC-62 Opinions****: The age-structured models applied to data for the witch flounder fishery from 1982-2015 were found to have major retrospective patterns that prevented their use for status evaluation and determination of catch advice. Therefore biological reference points are not available.*

EXISTING: 2015 Operational Assessment Biological Reference Points

The VPA model results were accepted as the basis for the biological reference points and stock status determination for witch flounder (NEFSC 2015). Reference points were calculated using the non-parametric yield and SSB per recruit/long-term projection approach adopted for the New England groundfish stocks (NEFSC 2008). 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 SSB_{msy} (Spawning Stock Biomass at maximum sustainable yield), the cumulative density function was re-sampled from estimates of Age 3 between 1982-2013 to provide future recruitment estimates. Recruitment of witch flounder is not well determined by traditional stock recruitment relationships. Given this, an MSY proxy was used for reference points. $F_{40\%}$ is the proxy used for the overfishing threshold (F_{msy}). This is consistent with the choice of proxy in the previous assessments (NEFSC 2012, NEFSC 2008). The biological reference points estimated in the 2015 assessment were F_{msy}=$F_{40\%}$=0.28, SSB_{msy}=9,473 mt, and MSY=1,957 mt.

NEW: 2016 SAW 62 Biological Reference Points

In this assessment, the yield and SSB per recruit calculations (NFT 2013a) used the most recent five year (2011-2015) averages for mean weights, fishery partial recruitment pattern, and maturity ogive. For the estimation of MSY and SSB_{msy}, the cumulative density function was re-sampled from estimates of Age 3 (VPA) or Age 1 (ASAP) between 1982-2015 to provide future recruitment estimates (Table B66 for VPA Run A2 input and Table B67 for ASAP Run 9_5_v2 input). Proxy reference points based on VPA Run A2 and ASAP Run 9_5_v2 are given in Table B68. Stochastic projections at F_{msy}=$F_{40\%}$ (0.19) were used to determine new biomass reference points (proxies for both SSB_{msy} and MSY) based on $F_{40\%}$ (Table B69).

Using Run VPA Run A2 and F40%=0.20, biological reference points are estimated to be 12,499 mt for SSB_{msy}, 6,250 mt for $SSB_{threshold}$, and 1,512 mt for MSY (Table B69).

Using ASAP Run 9_5_v2 and $F_{40\%}$ = 0.19, biological reference points are estimated to be 12,747 mt for SSBmsy, 6,374 mt for $SSB_{threshold}$, and 1,998 mt for MSY (Table B69). **The 2016 WG accepted the ASAP Run 9_5_v2 model results as the basis for the new recommended biological reference points and stock status determination for witch flounder.**

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 (or possibly models, in accord with guidance in attached "Appendix to the SAW Assessment TORs") developed for this peer review. In both cases, evaluate whether the stock is rebuilt .
a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the updated BRP estimates.
b. Then use the newly proposed model (or possibly models, in accord with guidance in "Appendix to the SAW Assessment TORs") and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-5).

> ***SARC-62 Opinions****: The status of the witch flounder stock is unknown with regards to biological reference points. The age-structured models applied to data for the witch flounder fishery from 1982-2015 were found to have major retrospective patterns that prevented their use for status evaluation and determination of catch advice.*
>
> *Based on an empirical analysis (referred to as "empirical area swept method") conducted by the SAW and examined at the SARC, stock biomass has declined since 2002 and appears to be 16,181 mt Jan-1 surveyable biomass in 2016 and 14,563 mt exploitable biomass in 2016. The relative exploitation rate needed to achieve the recent observed catch is approximately 0.05. The SARC-62 did not fully evaluate the empirical approach during the peer review, and therefore expected additional evaluation of the method, calculations, and results to be made following the peer review.*
>
> *The fishery landings and survey catch by age indicate truncation of age structure and a reduction in the number of old fish in the population.*

2016 UPDATED STOCK STATUS

c) Existing VPA model (VPA Run A2)

Based on the VPA Run A2 model, the witch flounder stock was determined to be overfished and overfishing occurred in 2015. This determination was made on retrospectively adjusted $F_{full,rho}$ and SSB_{rho} in 2015. The retrospective adjustment was required but did not lead to a change in status (Table B69). The biological reference points used were $F_{MSY}=F_{40\%}=0.20$, $SSB_{msy}=12{,}499$ mt, and MSY=1,512 mt. The 2015 F_{8-9} is estimated to be 0.43 (210% of $F_{40\%}$) and the retrospectively adjusted 2015 $F_{8-9,rho}$ to be 0.69 (337% of $F_{40\%}$). The 2015 SSB is estimated to be 3,044 mt (24% of SSB_{msy}) and the retrospectively adjusted 2015 SSB_{rho} to be 2,035 mt (16% of SSB_{msy}).

d) New ASAP model (ASAP Run 9_5_v2)

Based on the ASAP Run 9_5_v2 model, the witch flounder stock was determined to be overfished and overfishing occurred in 2015. This determination was made on retrospectively adjusted $F_{full,rho}$ and SSB_{rho} in 2015. The retrospective adjustment was required and lead to a change in overfishing status (Table B69; Figures B170-B172). The

biological reference points used were F_{msy}=$F_{40\%}$=0.19, SSB_{msy}=12,747 mt, and MSY=1,998 mt. The 2015 F_{full} is estimated to be 0.16 (83% of $F_{40\%}$) and the retrospectively adjusted 2015 $F_{full,rho}$ to be 0.29 (154% of $F_{40\%}$). The 2015 SSB is estimated to be 5,479 mt (43% of SSB_{msy}) and the retrospectively adjusted 2015 SSB_{rho} to be 3,335 mt (26% of SSB_{msy}).

TERM OF REFERENCE 7: Develop approaches and apply them to conduct stock projections.

a. Provide numerical annual projections (3 years) and the statistical distribution (e.g., probability density function) of the catch at FMSY or an FMSY proxy (i.e. the overfishing limit, OFL) (see Appendix). 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, magnitude and 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. Identify reasonable projection parameters (recruitment, weight-at-age, retrospective adjustments, etc.) to use when setting specifications.

c. Describe this stock's vulnerability to becoming overfished, and how this could affect the choice of ABC. The choice takes scientific uncertainty into account (see Appendix).

> ***SARC-62 Opinions:*** *The empirical area swept method was used to provide an F_{MSY} proxy for 2017 based on a relative exploitation rate. The SARC-62 did not fully evaluate the empirical approach during the peer review, and therefore expected additional evaluation of the method, calculations, and results to be made following the peer review.*

A) NUMERICAL ANNUAL PROJECTIONS FOR 2017-2019

Short-term projections of median catch (total fishery yield) and spawning stock biomass (mt) for witch flounder were based on a harvest scenario of an assumed catch of 460 mt in 2016 and fishing at $F_{40\%}$ (0.19) in 2017-2019 to determine the Overfishing Limit (OFL) for 2017-2019. One hundred projections were made for each of the 1000 bootstrapped iterations of 2016 stock sizes from the ASAP Run 9_5_v2 model results using NFT AGEPRO version 4.2.2 (NFT 2013b). Future Age 1 recruitment was generated randomly from a cumulative density function of the updated recruitment series for 1982-2015 (year classes 1981–2014 with mean recruitment = 9.8 million fish). Partial recruitment was taken from the fleet selectivity for time block 2005-2015. Average 2011-2015 mean weights, and maturation ogive representing 2012-2016 maturity data were the same used in the biological reference points based on the ASAP Run 9_5_v2 (Table B67). Projection results based on the ASAP Run 9_5_v2 are shown below and in Figure B173. The SSB retrospective adjustment factor (0.6088) was applied to all ages (1 -11+) of the initial population from ASAP Run 9_5_v2 in these projections. Numbers in parentheses indicate 5th and 95th percentiles.

The probability of SSB falling below the biomass threshold in 2017-2019 fishing at the OFL ($F_{40\%}$ = 0.19), is 96%, 17% and 0%, respectively.

Year	Input	Catch (mt)	SSB (mt)		Harvest Strategy	Ffull	
2016	Catch input	460	4,212	(3,611 - 4,933)		0.14	(0.12 - 0.17)

2017	Projection	762	(650 - 895)	5,452	(4,670 - 6,337)	F40%	0.19
2018	Projection	911	(782 - 1,048)	6,940	(5,977 - 7,933)	F40%	0.19
2019	Projection	1,136	(978 - 1,295)	8,302	(7,136 - 9,528)	F40%	0.19

B) MOST REALISTIC PROJECTIONS

The Working Group conducted one projection by assuming recent patterns in the fishery and the most recent maturity ogive. A major source of uncertainty is the retrospective pattern. The past projection was optimistic (outside the confidence bounds for the rho adjusted biomass in this assessment). Subsequent projections will be conducted by the Groundfish Plan Development Team, when 2016 catch is determined.

C) STOCK VULNERABILITY

The 2016 stock assessment indicates that the stock was below the biomass target and being fished above the fishing mortality threshold in 2015. Witch flounder is slow growing and late maturing and is vulnerable without stringent management regulations in place.

TERM OF REFERENCE 8: Evaluate the validity of the current stock definition, taking into account what is known about migration, and make a recommendation about whether there is a need to modify the current stock definition for future stock assessments.

Witch flounder are assessed and managed as a unit stock within USA waters; this includes the Gulf of Maine and Georges Bank region as well as areas to the south as far as Cape Hatteras, North Carolina.

No tagging studies have been conducted on witch flounder in Gulf of Maine-Georges Bank region.

No studies on stock structure have been conducted in the Gulf of Maine-Georges Bank region. Burnett and Clark (1983) noted that NEFSC spring and autumn survey catch-per-tow data indicate witch flounder in this region are relatively discrete from the western Scotian Shelf or further south.

In the Newfoundland region, Fairbairn (1981), using biochemical systematics for stock separation, suggests that witch flounder in the Newfoundland region exist in relatively small, independent subpopulations with little inter-stock migration.

Bowering and Misra (1981) applied a multivariate statistical method to witch flounder meristics data in the Newfoundland-Labrador area and results indicated seven distinct populations of witch flounder throughout the areas studied (Hawk Channel, Funk Island Bank, Northern Grand Bank, Southwest Slope of Grand Bank, Burgeo-St Pierre Banks, Fortune Bay, and Gulf of St Lawrence). Bowering and Misra (1981) results agreed with those made by Fairbairn (1981) using entirely different techniques.

Wigley and Burnett (2003) compared biological and yield characteristics of deep water (depth > 366 m) witch flounder south and west of Georges Bank and found these fish had slower growth and that otolith structure was different than the Gulf of Maine-Georges Bank witch flounder, suggesting a separate population.

WG recommendation: Until tagging and/or stock structure analyses are undertaken that reveal subpopulations, the current stock definition should be retained for future stock assessments. Possible future research recommendations include: ageing larval fish to determine transport time; tagging and/or stock structure analyses.

TERM OF REFERENCE 9. Review, evaluate and report on the status of research recommendations from the last peer reviewed benchmark stock assessment. Identify new research recommendations.

> ***SARC-62 Opinions:*** *The SARC recommends that witch flounder be put on a research track to address survey catchability and retrospective problems. Such an examination should include other stocks in the region exhibiting major retrospective patterns and survey catchability issues.*

Research recommendations from the last peer reviewed benchmark stock assessment (2008 GARM) are summarized below. Of the 3 research recommendations, progress has been made on two recommendations. The WG developed nine new research recommendations.

Previous Research Recommendations since 2008 GARM benchmark assessment

<u>2008 GARM Benchmark (NEFSC 2008)</u>

The Panel had no specific research recommendations for this stock.

<u>2012 Update (NEFSC 2012)</u>

Two research recommendations, applicable to several stocks were suggested:

1) explore the possibility of refining the calibration factors within the assessment model itself (e.g, splitting the survey tuning series and using the results from the calibration experiment as a prior);

   The WG explored splitting the NEFSC survey tuning series into two separate independent series (Albatross and Bigelow). The WG noted several reasons to retain the ALB calibrated series: BIG series is too short to stand alone; split perhaps allows too much flexibility, with some counterintuitive results in terms of ALB vs BIG q; limited age structure in short BIG series results in lack of precision in estimates of survey selectivity; effectively includes results of calibration experimental results (ALB to BIG 3.2572).

2) continue to examine the trends in mean weights at age and their possible underlying factors.

   No progress.

<u>2015 Operational Assessment (NEFSC 2015)</u>

Research Needs: The Panel recommends that the sources of the retrospective pattern need to be addressed. Considering that retrospective patterns are a common problem, the generic problem may be most appropriately addressed in a research track topic, and all possible sources of the retrospective problem should be investigated (mis-specified natural mortality, changes in natural mortality, under-reported catch, changes in survey catchability and mis-specified

selectivity, etc.). For the next benchmark assessment, the Panel recommended exploring a statistical catch-at-age model to investigate possible doming in the catch selectivity.

> *VPA and ASAP models were considered in this assessment. Sensitivity runs exploring mis-specified M, under-reported catch, survey catchabilities, and mis-specified selectivity in the fleet and survey were conducted but were not accepted as a base run because model diagnostics do not strongly support high flexibility to estimate fishery or survey selection above age 7-8 and the missing mortality (see TOR 4 for a detailed list).*

New Research Recommendations developed by WG

1) Age the archived age structures from the MENH inshore trawl survey for use in the future assessments.

2) Conduct a stock identification study. Components could include: tagging (studies designed for stock id), larval dispersion (transport), morphometrics, meristics, genetics, etc.

3) Conduct larval growth studies to generate larval index of SSB.

4) Conduct tagging studies designed to decrease the uncertainty of estimates/assumption for M, if feasible.

5) Consider a further examination of the appropriate survey area, including environmental and/or habitat preferences (e.g., Friedland working paper), to be used in swept-area expansions.

6) Consider further investigation of the influence of catch-at-age information, as compared to aggregate fishery and survey catches, in the assessment model. Areas of research to consider could include (a) inclusion of ageing error estimates in the assessment model and (b) direct use of age-length keys and length distribution data as data inputs into the assessment model. This is a generic recommendation applicable to all SARC stock assessments.

7) Consider further investigation to maximize the information contained in the survey data, such as the use of new geostatistical /spatial modeling approaches to improve estimates of precision. This is a generic recommendation applicable to all SARC stock assessments.

8) The WG notes that unaccounted catch has been examined as a possible cause for the major retrospective errors in SARC stock assessments. The WG recommends that efforts should be made in all Northeast region fisheries to determine the veracity of currently reported catch statistics and the magnitude of possible unaccounted catch. The WG should then consider using any resulting “hard” evidence of unaccounted catch, as it becomes available, to inform further investigations (e.g., in terms of model performance and diagnostics) of this factor on stock assessment results. This is a generic recommendation applicable to all SARC stock assessments.

9) The WG notes that changes in natural mortality has been examined as a possible cause for the major retrospective errors in SARC stock assessments. The WG recommends that efforts should be made to investigate how M might change or has changed over time in response to biological and environmental factors. This is a generic recommendation applicable to all SARC stock assessments.

ACKNOWLEDGMENTS

The WG members thank C Legault, A Miller, P Nitschke, L O'Brien, C Tholke, S Sherman, and B Smith for their assistance and contributions to this report. We also thank the authors of the working papers for their work and consultations during the meetings. Special thanks to S Emery for the timely preparation of the witch flounder age data.

LITERATURE CITED

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

Azarovitz TR. 1981. A brief historical review of the Woods Hole Laboratory trawl survey time series. Canadian Special Publication of Fisheries and Aquatic Sciences. 58:62-67.

Bigelow HB, Schroeder WC. 1953. Fishes of the Gulf of Maine. US Fish and Wildlife Service, Fish Bull. 74, Vol 53, 577 p.

Bowering WR, Misra RK. 1981. Comparisons of witch flounder (*Glyptocephalus cynoglossus*) stocks of the Newfoundland – Labrador area, based upon a new multivariate analysis method for meristic characters. Can. J. Fish. Aquat. Sci. 39: 564-570.

Bowman RE, Michaels WL. 1984. Food of seventeen species of Northwest Atlantic fish. NOAA Tech Memo. NMFS-F/NEC-28; 183 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.

Breiman, L. 2001. Random forests. Machine Learning, 45: 5-32.

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.

Burnett J, Clark SH. 1983. Status of witch flounder in the Gulf of Maine - 1983. NMFS/NEFC, Woods Hole Laboratory Ref. Doc. No. 83-36, 31 p.

Burnett J, O'Brien L, Mayo RK, Darde J, Bohan M. 1989. Finfish maturity sampling and classification schemes used during the Northeast Fisheries Center bottom trawl survey. NOAA Tech. Mem. NMFS-F/NEC-76. 14 p.

Burnett J, Ross MR, Clark SH. 1992. Several biological aspects of the witch flounder, *Glyptocephalus cynoglossus* (L), in the Gulf of Maine-Georges Bank region. J. Northw. Alt. Fish. Sci. 12: 15-25.

Burnett JM. MS 1987. The population biology of the witch flounder, *Glyptocephalus cynoglossus* (L.), in the Gulf of Maine-Georges Bank region. M.Sc. Thesis, Department of Wildlife and Fisheries Biology, University of Massachusetts, Amherst, MA, 116 p.

Cadrin S, B Wright. 2016. Fishery Catch Rates of Witch Flounder. SAW62 Witch Flounder Working Paper. September 5, 2016

Cadrin SX, King, J. 2003. Stock assessment of yellowtail flounder in the Cape Cod - Gulf of Maine area. U.S. Dep. Commer. Northeast Fish. Sci. Cent. Ref. Doc. 03-03.

Caruana, R., and Niculescu-Mizil, A. 2006. An empirical comparison of supervised learning algorithms. In Proceedings of the 23rd international conference on Machine learning, pp. 161-168. ACM, Pittsburgh, Pennsylvania, USA.

Charnov EL, Berrigan D. 1990. Dimensionless numbers and life history evolution: Age of maturity versus the adult lifespan. Evolutionary Ecology,**4**, 273–5.

Clark SH. 1981. Use of trawl survey data in assessments. *In:* Doubleday, W.G. and D. Rivard (eds.), Bottom Trawl Surveys, p. 82-92, Can. Spec. Publ. Fish. Aquat. Sci. 58.

Clark WG. 2013. Direct calculation of relative fishery and survey selectivities. Fisheries Research. 158:135-137.

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

Colton JB, Smith WG, Kendall AW, Berrien PL, Fahy MP. 1979. Principal spawning areas and times of marine fishes, Cape Sable to Cape Hatteras. Fish. Bull., US. 76(4):5 pp.

Cutler DR, Edwards TC, Beard KH., Cutler A., and Hess, K. T. 2007. Random forests for classification in ecology. Ecology, 88: 2783-2792.

Dick EJ. 2004. Beyond 'lognormal versus gamma': discrimination among error distributions for generalized liner models. Fisheries Research 70: 351-366.

Fairbairn DJ. 1981. Which witch is which? A study of the stock structure of witch flounder (*Glyptocephalus cynoglossus*) in the Newfoundland Region. Can. J. Fish. Aqua. Sci. 38: 782-794.

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

Fratantoni PS, Holzworth-Davis T, Taylor MH. 2015. Description of oceanographic conditions on the Northeast US Continental Shelf during 2014. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 15-21; 41 p. doi: 10.7289/V5SQ8XD2

Grosslein M. 1969. Groundfish survey program of BCF Woods Hole. Comm. Fish. Rev. 31(8-9): 22-30.

Gulland JA. 1983. Fish Stock Assessment: A manual of basic methods. J. Wiley and Sons. New York.

Halliday RG. 1973. The flatfish fisheries of the Scotian Shelf. Int. Comm. Northw. Atl. Fish., Res. Doc. 73/103, Ser. No. 3064 (mimeo), 46 p.

Halliday RG. 1978. Size and age at sexual maturity of Atlantic argentine, *Argentina silus*: a critique. Environ. Biol. Fishes 19(2):165-261.

Hare JA., Brooks EN, Palmer MC, and Churchill JH. (2015). Re-evaluating the effect of wind on recruitment in Gulf of Maine Atlantic Cod (Gadus morhua) using an environmentally-explicit stock recruitment model. Fisheries Oceanography, 24(1), 90-105.

Hare JA, Fratantoni PC, Walsh, HM. 2016a. In situ temperature and salinity data for use in stock assessments. SAW 62 working paper, 4 p.

Hare J, Hoey J, Manderson J, Martin M, Politis P, Richardson D, Roebuck C. 2016b. Empirical estimates of maximum catchability of Witch Flounder *Glyptocephalus cynoglossus* L. on the Northeast Fisheries Science Center Fall bottom trawl survey. Witch Flounder Working Group working paper. 27 p.

Harley SJ, Myers RA, Dunn A. 2001. Is catch-per-unit-effort proportional to abundance? Canadian Journal of Fisheries and Aquatic Sciences 58: 1760-1772.

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.

Howe, AB, Germano FJ, Buckley JL, Jimenez D, Estrella BT. 1981. Fishery resource assessment, coastal Massachusetts. Completion Report, Massachusetts Division of Marine Fisheries, Commercial Fisheries Review Div. Project 3-287-R-3.

Hudson M, Joyce J. 2015. Recommendations from the Fish Tank port meetings. (available online: http://www.gmri.org/fisheries-convening/fish-tank/meeting-materials)

Jensen AL. 1996. Beverton and Holt life history invariants result from optimal trade-off of reproduction and survival. Can. J. Fish. Aqua.Sci. 53(4):820–822.

Johnston R, Sosebee K. 2014. History of the United States bottom trawl surveys, NAFO Subareas 4-7. NAFO SCR Doc 14/024. Serial No. N6319. 41 p.

Kimura DK. 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.

Knudby A., LeDrew E, Brenning A. 2010. Predictive mapping of reef fish species richness, diversity and biomass in Zanzibar using IKONOS imagery and machine-learning techniques. Remote Sensing of Environment, 114: 1230-1241.

Lange AMT, Lux FE. 1978. Review of the other flounder stocks (winter flounder, American plaice, witch flounder, and windowpane flounder) off the northeast United States. NMFS, NEFC, Woods Hole Lab. Ref. Doc. No. 78-44, 53 pp.

Legault CM, Alade L, Gross WE, Stone HH. 2014. Stock Assessment of Georges Bank Yellowtail Flounder for 2014. TRAC Reference Document 2014/01. http://www.nefsc.noaa.gov/saw/trac/TRAC_GBYT_2014_WP.pdf

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

Legault, C.M. and D. Busawon. 2016. Stock assessment of Georges Bank yellowtail flounder for 2016. TRAC Reference Document 2016/01. 63 p. available at http://www.nefsc.noaa.gov/saw/trac/

LinYP, Lin WC, Wu WY. 2015. Uncertainty in Various Habitat Suitability Models and Its Impact on Habitat Suitability Estimates for Fish. Water, 7: 4088-4107.

Link JS, Almeida FP. 2000. An overview and history of the food web dynamics program of the Northeast Fisheries Science Center, Woods Hole, Massachusetts. NOAA Tech Memo NMFS-F/NE-159. 60 p.

Lux, F. 1969. Length-Weight Relationships of Six New England Flatfishes. Trans. Amer. Fish. Soc., 1969, No. 4. 617-621.

Maunder MN, AE Punt. 2004. Standardizing catch and effort data: a review of recent approaches. Fisheries Research 70:141-159.

Maunder MN, JR Sibert, A Fonteneau, J Hampton, P Kleiber, SJ Harley. 2006. Interpreting catch per unit effort data to assess the status of individual stocks and communities. ICES Journal of Marine Science 63(8): 1373-1385.

Maurer R, Bowman R, 1975. Food Habits of marine fishes of the northwest Atlantic – data report. US. Natl. Mar. fish. Serv., Northeast Fish. Center., Woods Hole Lab. Ref. Doc, 75-3. 90 p. http://www.nefsc.noaa.gov/publications/series/whlrd/whlrd7503.pdf

Mayo RK, O'Brien L, Buxton N. 1992. Discard estimates of American plaice, *Hippoglossoides platessoides*, in the Gulf of Maine northern shrimp fishery and the Gulf of Maine-Georges Bank large-mesh otter trawl fishery. Appendix to NEFSC Ref. Doc. 92-07, Res. Doc. SAW 14/3, 40 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 Fish Sci

Cent Ref Doc. 10-05; 233 p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at: http://www.nefsc.noaa.gov/publications/crd/crd1005/

Miller TJ, Hare JA, Alade L. A. 2016. A state-space approach to incorporating environmental effects on recruitment in an age-structured assessment model with an application to southern New England yellowtail flounder. Canadian Journal of Fisheries and Aquatic Sciences, 73(999), 1-10.

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.

Murawski, SA, RW Brown, SX Cadrin, RK Mayo, L O'Brien, WJ Overholtz, KA Sosebee. 1999. New England groundfish. In: Our Living Oceans. NOAA Tech. Mem. NMFS-F/SPO-41:71-80.

Murphy, T, A Kitts, D Records, C Demarest, D Caless, J Walden, S Benjamin. 2014. 2012 Final Report on the Performance of the Northeast Multispecies (Groundfish) Fishery (May 2012-April 2013. NEFSC Ref. Doc. 14-01. 111 pp.

NEFC [Northeast Fisheries Center]. 1985. Rapporteur's Summary First Northeast Fisheries Center Stock Assessment Workshop. 8-2 July 1985, Falmouth MA. http://www.nefsc.noaa.gov/publications/classics/pdfs/1saw1985.pdf

NEFC. 1986. Report of the Second NEFC Stock Assessment Workshop (Second SAW). Woods Hole Laboratory Reference Document No. 86-09. 114 p. http://nefsc.noaa.gov/publications/series/whlrd/whlrd8609.pdf

NEFSC [Northeast Fisheries Science Center]. 1994. Report of the 18th Northeast Regional Stock Assessment Workshop: the plenary. NOAA/NMFS/ NEFSC: Woods Hole, MA. *NEFSC Ref* Doc. 94-23. http://www.nefsc.noaa.gov/publications/crd/pdfs/crd9423.pdf

NEFSC. 1996a. Report of the 21st Northeast Regional Stock. Assessment Workshop (21st SAW): Stock Assessment Review Committee (SARC) consensus summary of assessments. *Northeast Fish. Sci. Cent. Ref Doc.* 06-05d; 200 p.

NEFSC. 1996b. Report of the 21st Northeast Regional Stock Assessment Workshop (21st SAW): Public Review Workshop. *Northeast Fish. Sci. Cent. Ref Doc.* 96-05h~ 50 p. http://www.nefsc.noaa.gov/publications/crd/pdfs/crd9605h.pdf

NEFSC. 1999. Report of the 29th Northeast Regional Stock Assessment Workshop (29th SAW): Public Review Workshop. Northeast Fish Sci Cent Ref Doc. 99-13: 37 p. http://www.nefsc.noaa.gov/publications/crd/pdfs/crd9913.pdf

NEFSC. 2002. A Report of the Groundfish Assessment Review Meeting (GARM), October

2002. NEFSC Ref. Doc. 02-16, 511p.

NEFSC. 2003a. Report of the 37th Northeast Regional Stock Assessment Workshop (37th SAW): Stock Assessment Review Committee (SARC) consensus summary of assessments. Northeast Fish. Sci. Cent. Ref. Doc. 03-16; 597 p. http://www.nefsc.noaa.gov/publications/crd/crd0316/

NEFSC. 2003b. Report of the 37th Northeast Regional Stock Assessment Workshop (37th SAW): advisory report. Northeast Fish. Sci. Cent. Ref. Doc. 03-17; 52 p. http://www.nefsc.noaa.gov/publications/crd/crd0317/

NEFSC. 2005. Mayo RK, Terceiro M, editors. 2005. Assessment of 19 Northeast groundfish stocks through 2004. 2005 Groundfish Assessment Review Meeting (2005 GARM), Northeast Fisheries Science Center, Woods Hole, Massachusetts, 15-19 August 2005. U.S. Dep. Commer., Northeast Fish. Sci. Cent. Ref. Doc. 05-13; 499 p.

NEFSC. 2008. Assessment of 19 Northeast Groundfish Stocks through 2007: Report of the 3rd Groundfish Assessment Review Meeting (GARM III). Northeast Fisheries Science Center (NEFSC) Woods Hole, Massachusetts. August 4-8, 2008. NEFSC. CRD 08-15. 867 p. http://www.nefsc.noaa.gov/publications/crd/crd0815/

NEFSC. 2012. Assessment or Data Updates of 13 Northeast Groundfish Stocks through 2010. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 12-06; 789 p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at http://nefsc.noaa.gov/publications/

NEFSC. 2015. Operational Assessment of 20 Northeast Groundfish Stocks, Updated Through 2014. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 15-24; 251 p.

NFT. 2008. NOAA Fisheries Toolbox (NFT). Rivard Weights Calculator version 2.0. Online at: http://nft.nefsc.noaa.gov

NFT. 2013a. NOAA Fisheries Toolbox (NFT). Yield per recruit (YPR) version 3.3. Online at: http://nft.nefsc.noaa.gov

NFT. 2013b. NOAA Fisheries Toolbox (NFT). AGEPRO version 4.2.2. Online at: http://nft.nefsc.noaa.gov

NFT. 2014a. NOAA Fisheries Toolbox (NFT). ASAP version 3.0.17. Online at: http://nft.nefsc.noaa.gov

NFT. 2014b. NOAA Fisheries Toolbox (NFT). VPA version 3.4.4. Online at: http://nft.nefsc.noaa.gov

Northern Shrimp Technical Committee. MS 1984. Results of the 1983 northern shrimp survey in the western Gulf of Maine, August 1983. Unpublished Report, Woods Hole, MA, 16 p.

NRC [National Research Council]. 1998. Review of Northeast Fishery Stock Assessments. National Academies Press, 1998.

NSC [Northeast Seafood Coalition]. 2016. Letter to New England Fishermen Management Council Scientific and Statistical Committee. January 15 (www.nefmc.org/calendar/jan.-20-2016-ssc-meeting)

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.

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, Brown RW. 1996. Assessment of the Georges Bank Haddock stock for 1994. NEFSC Ref. Doc. 95-13, 108 p.

O'Keefe CE, SX Cadrin, J Wiersma. 2015. Fishery dependent data in New England groundfish stock assessments. Unpublished report to EDF (summary available online: http://www.gmri.org/fisheries-convening/fish-tank/meeting-materials).

Palmer, M. 2016. Catch curve analysis of witch flounder fishery and survey catch-at-age data. 2016 SAW 62 Witch Flounder Working Paper, p. 9.

Pauly, D. 1980. On the interrelationships between natural mortality, growth parameters, and mean environmental temperature in 175 fish stocks.

Pershing A. 2015. How fish abundance and habitat preference impact trawl survey performance. (available online: http://www.gmri.org/fisheries-convening/fish-tank/meeting-materials).

Powles, PM. 1967. Length-weight relationships for American plaice, witch, and yellowtail, in ICNAF Subarea 4. Int. Comm. Northw. Atlant. Fish, Res. Bull. No. 4: 121-123.

Ralston, S. 1987. Mortality rates of snappers and groupers. Pages 375–404 In: J.J. Polovina and S. Ralston (eds.), Tropical Snappers and Groupers: Biology and Fisheries Management. Westview Press, Boulder Colorado. 659 p.

Reid DG, Annala J, Rosen S, Pol M, Cadrin SX, Walsh SJ. 2007. Current status of mobile and static sampling gears used in resource surveys. ICES Journal of Marine Science 64: 1607-1609.

Restrepo VR, Ortiz de Urbina J, Fromentin JM, Arrizabalaga H. 2007. Estimates of selectivity for eastern Atlantic bluefin tuna from catch curves. Col. Vol. Sci. Pap. ICCAT 60:937-348.

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

Rikhter, VA, Efanov VN. 1976. On one of the approaches to estimation of natural mortality of fish populations. ICNAF Research Document 76/VI/8. 12p.

Robson DS, Chapman DG. 1961. Catch curves and mortality rates. Trans. Am. Fish. Soc. 90:181-189.

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 Online Doc, version 9. SAS Institute Inc., Cary NC.

SAS. 2011. SAS/STAT User's Guide, Version 9.3. The SAS Institute Inc., Cary, NC, USA.

Sherman S, Stepanek K, Sowles J. 2005. Maine-New Hampshire Inshore Groundfish Trawl Survey Procedures and Protocols. Maine Department of Marine Resources, Research Reference Document 05/01. 42 p. http://www.maine.gov/dmr/rm/trawl/reports/proceduresandprotocols.pdf

Terceiro M. 2003. The statistical properties of recreational catch rate data for some fish stocks off the northeast U.S. coast. Fish Bull U.S., 101:653-672.

Terceiro, M. 2016a. TOR 1: Description of commercial fishery Dealer Report trawl gear landings and effort and modeling landings rate (LPUE) data for witch flounder. 2016 SAW 62 Witch Flounder Working Paper: June 8, 2016

Terceiro M. 2016b. TOR 1 & 2: Description of Vessel Trip Report trawl gear catch and effort data and modeling catch rates (CPUE) for witch flounder. 2016 SAW 62 Witch Flounder Working Paper: 5/25/2016

Terceiro M. 2016c. TOR 1: Description and modeling of NEFOP (Observer) fish trawl gear catch rate (CPUE) data for witch flounder. 2016 SAW 62 Witch flounder Working Paper: May 26, 2016

Terceiro M 2016d. TOR 1: Description of commercial fishery Dealer Report trawl gear landings and effort and modeling landings rate (LPUE) data for witch flounder: 'Directed' Trips (=>40% of trip landings). 2016 SAW 62 Witch Flounder Working Paper: August 16, 2016

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.

Walsh HJ, Richardson DE, Marancik KE, Hare JA. 2015. Long-Term Changes in the

Distributions of Larval and Adult Fish in the Northeast U.S. Shelf Ecosystem. PLoS ONE 10(9): e0137382. doi: 10.1371/journal.pone.0137382.

Walsh SJ, Millar RB, Cooper CG, Hickey WM. 1992. Codend selection in American plaice: diamond versus square mesh. Fish. Res., 13:235-254.

Wigley SE. 2016. Rough vs Smooth Bottom Type: An Initial Exploration. SAW 62 Witch Flounder Working Paper. August 29, 2016. 15 p.

Wigley SE, Brodziak JKT, Cadrin SX. 1999. Assessment of the witch flounder stock in Subareas 5 and 6 for 1999. Northeast Fish. Sci. Cent. Ref. Doc. 99-16. 153 p.

Wigley SE, Brodziak JKT, Col L. 2003a. Assessment of the Gulf of Maine-Georges Bank witch flounder stock for 2003. Northeast Fish. Sci. Cent. Ref. Doc. 03-14. 186 p.

Wigley SE, Burnett JM. 2003. Preliminary estimates of biological and yield characteristics of deep-water witch flounder (*Glyptocephalus cynoglossus*) in the Georges Bank – Mid-Atlantic Bight region. J. Northw. Atl. Fish. Sci. Vol 31: 181-194.

Wigley SE, Mayo RK. 1996. Assessment of the Gulf of Maine-Georges Bank witch flounder stock for 1994. Northeast Fish. Sci. Cent. Ref. Doc. 94-17. 86 p.

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

Wilk, S.J, W. Morse, and D.E. Ralph. 1978. Length-weight relationships of fishes collected in the New York Bight. Bulletin of the New Jersey Academy of Science. 23(2)58-64.