

Review Report on the 2002 Groundfish Assessment Review Meeting (GARM) and its findings and recommendations

By

Jon Helge Vølstad¹, Ph.D.
Versar, Inc.
9200 Rumsey Road
Columbia, Maryland 21045
USA

¹Representing the Center of Independent Experts, at the Rosenstiel School of Marine and Atmospheric Science, University of Miami.

October 29, 2002

Executive Summary

Assessments (through 2001) of the 20 stocks under the Northeast fisheries management plan updated and reviewed by GARM are of consistent high quality, based on models that are suitable for the available data for each stock. The GARM review meeting was competently chaired, and conducted in a spirit of cooperation and teamwork. The assessments, conducted by experienced stock assessment biologists, were subject to a rigorous and very open peer review process that identified the most likely sources of uncertainty and, in some cases, inconsistencies were discovered. In response to requests by GARM, inconsistencies were usually corrected in time for review the next day. Inaccurate information on catch-at-age, resulting from limited spatial and temporal sampling coverage of landings and limited, if any, information on discard was identified as a major source of uncertainty in some age-based assessments. Systematic patterns in the residuals of many VPA model fits (e.g., multiple years in a row with negative or positive residuals) strongly indicates bias in the catch-at-age data, for example resulting from significant discards not accounted for, or from biased catch sampling. The control of fishing mortality through trip limits, which has been implemented for some stocks in recent years, might increase the level of discards and, worse, cause a shift in the time series of catch-at-age that is difficult to correct for in stock assessments.

GARM conducted a rigorous and very thorough evaluation of potential effects of a trawl warp offset recently discovered on the NOAA research vessel *Albatross IV*. The offset, which applied to surveys conducted from winter and spring 2000-2002, and fall 2000 and 2001, could potentially affect recent stock assessments. A suite of studies related to gear configuration, fishing power, and survey indices of abundance and size composition over time strongly suggested that any change in trawling efficiency as a result of the offset was minor. Results from previous controlled fishing power experiments conducted with *Albatross IV* trawling alongside another vessel (*Delaware II*) showed very similar catching efficiencies for *Albatross IV* relative to *Delaware II* before and after the warp offset, thus precluding a major drop in trawl catching efficiency caused by the offset. Analyses of the series of *Albatross IV* survey indices of abundance (and size

composition) alone, or compared to Canadian independent surveys for some species, did not reveal a substantial shift related to drop in fishing power after the warp offset. An additional sensitivity analysis conducted by GARM using hypothetical increases of 10%, 25%, and 100% in survey abundances (corresponding to reductions in catching efficiency well beyond what is supported by available data) would not have sufficient impact on stock assessments to change the determination of status with respect to overfishing (fishing mortality rates) for 19 of the 20 species under the Northeast FMP. American plaice, which had a biomass close to the target of $\frac{1}{2}$ B-MSY in 2001, changed status from ‘overfishing’ to ‘no overfishing’ in the (unlikely) event of a 50% reduction in fishing power caused by the warp offset (corresponding to a 100% upward adjustment of the abundance index). The qualitative status based on the comparison of estimated biomass to the reference ($\frac{1}{2}$ B-MSY) changed for two species (American Plaice and Gulf of Maine haddock) with a hypothetical 10% (or higher) increase in abundance because their estimated biomass were close to the threshold. Clearly, when the point estimate of biomass is close to the threshold, even insignificant changes in catching efficiency (i.e., within the natural variability caused by weather conditions, variations in the vertical distribution of fish, and other factors) could change the qualitative status. Based on these considerations, I believe that it was demonstrated beyond a reasonable doubt that the offset of trawl warps on *Albatross IV* has not invalidated the stock assessments and the determination of fishing status with respect to overfishing.

1. Background

The Northeast Fisheries Science Center (NEFSC) conducts assessments of 20 major groundfish stocks (A-T, Appendix A) under the Northeast Multispecies Fishery Management Plan (FMP). These stock assessments are based on methods that use abundance indices from fisheries independent trawl surveys (conducted since 1963), either directly or indirectly. For stocks where sufficient information is available on catch at age over time (e.g., from port sampling, logbooks and other sources), the assessments are based on VPA (Virtual Population or Cohort Analysis) (A-G) or ADAPT (H). Independent survey indices of abundance (or biomass) are used for the tuning (calibration) of these age-based or integrated models. The assessments of nine stocks (L-T) are based directly on the fisheries independent survey indices of abundance (or biomass) and the estimated population characteristics of the stocks (e.g., size and age distribution), along with information on catch. Due to limitations in basic catch-at-age data, the assessment of two stocks (I, J) was based on a stock production model incorporating covariates (ASPIC), with spring and autumn survey indices of biomass used for tuning; while one stock (K) was assessed by a biomass dynamic model, using survey indices of biomass.

Abundance indices and estimates of population characteristics from fisheries independent surveys provide essential information for the assessments of the groundfish stocks under the Northeast FMP. The NOAA research vessel *Albatross IV* has been used by NEFSC to conduct stratified random surveys for the assessment and monitoring of groundfish stocks since 1963, representing an unparalleled continuous time series for tracking the status and trends of major stocks under the Northeast FMP. The normal operating area for *Albatross IV* is the Gulf of Maine, Georges Bank, and the continental shelf and slope from Southern New England to Cape Hatteras, NC. Concerns have been raised that the reliability of the surveys conducted by *Albatross IV* has been jeopardized in the most recent years because of an un-intended change in the sampling gear configuration in

2000, thus potentially compromising stock assessments for recent years. Measurements of NEFSC survey trawl warps in autumn 2002 suggested that right and left warps (the wires that attach the trawl gear to the vessel) may have been offset by up to several feet on the NOAA Ship *Albatross IV* during surveys conducted from winter 2000 through spring 2002. The offset was caused by biased measurements of the 50 m intervals for one warp, and as a result of these miss-markings, the offset increased proportionally with the length of cable deployed. The fishing industry and other constituents have postulated that the offset may have substantially reduced the trawl catching efficiency because of reduced bottom contact and lesser opening of the trawl. The possibility of the trawl collapsing at greater depths, which would result in zero catches, was also raised. A considerable reduction in trawling efficiency would introduce bias and reduce the reliability of the survey indices of abundance and estimates of population characteristics (e.g., length and age compositions) of groundfish species under the FMP.

These concerns were addressed in the Groundfish Assessment Review Meeting (GARM) from October 8-11, 2002 in Woods Hole, MA. GARM conducted a scientific review of assessment information and ancillary analyses including multiple studies to assess if the warp offset significantly affected trawling efficiency. GARM also assessed the potential implications of warp offset on stock assessment and resulting determination for overfished status (e.g., $<1/2$ BMSY) for the 20 species under the Northeast FMP, based on hypothesized reductions in the catching efficiency of the trawl in a series of sensitivity analyses.

2. Review of the GARM Activities and Findings

Dr. Steve Murawski chaired the GARM meeting, held at the Northeast Fisheries Science Center (NEFSC) in Woods Hole from October 8-11, 2002, in an organized and effective manner. The meeting was conducted in a spirit of cooperation and teamwork. Draft documents of most updated stock assessments were made available for review a few days

before the meeting. During the GARM meeting, the responsible assessment expert presented each stock assessment update, and the panel of experts reviewed it. The group of regional stock assessment scientists (both NMFS and non-NMFS people) and external experts conducted the review. The team of scientists was very diligent in the search for inconsistencies in the methods and results, and everyone was very open to critique from the panel of reviewers.

2.1. Updating of Stock Assessments

Most stock assessments reviewed at the GARM were routine updates of assessments previously reviewed in the SAW or elsewhere. All the assessments specified in the Terms of Reference for the GARM were covered. Estimates of fishing mortality rates (or proxies thereof) and biomasses in 2001, relative to management reference points 2001 fishing mortality (or its proxy), were provided for 19 stocks. For one stock (T), the assessment presented was developed for the first time (by the ASMFC Technical Committee), and has not yet been subject to standard peer review. Accordingly, the details of the analytical stock assessment modeling are not incorporated in the GARM report, pending that “benchmark” review to be conducted at SAW-36 in December 2002.

Quality of the input data, and the suitability of the VPA model specifications were evaluated through inspection of residual plots. Alternative model structures (e.g., ASPIC) were used in a few cases to examine the robustness of the VPA results. The two primary sources of uncertainty included in the projections (for 2002 and onwards) based on VPA assessments are: (1) the initial estimated stock abundance at age in year 2001, which is driven by the fisheries-independent survey indices used for tuning, and (2) future recruitment to the stock. The tuning minimized the differences between predicted numbers at age from the VPA and the fisheries-independent indices of abundance through parameter adjustments, and has most influence on estimates for recent years. The VPA is a recursive procedure that converges to yield robust estimates of number at age

back in time (under certain assumptions) provided that the input catch statistics are reliable over time.

Inaccuracy in catch-at-age for commercial (and recreational) fisheries resulting from limited spatial and temporal sampling coverage of landings, and limited, if any, information on discard, was recognized by the whole panel as a problem. The non-random residual pattern in the VPA model fits observed for many stocks, with residuals being negative or positive for a series of years, strongly indicates that substantial components of the catch are unaccounted for. The strong retrospective patterns of under-estimation of F also could results from discard unaccounted for. The recent implementation of trip limits to reduce fishing pressure for some stocks is likely to increase discard of target species (e.g., through high-grading), and could introduce significant bias in stock assessments.

2.2. Management Advice and Reference Points

The GARM maintained the reference points established in “The Report of the Working Group on Re-evaluation of Biological Reference Points for New England groundfish”. Sensitivity analyses were conducted to evaluate hypothetical effects of trawl warp offset on the status determination for each stock. Their analyses demonstrated that the determination of the status of stocks, relative to biological reference points set by management, was robust to small (10%), and moderate (25%) changes in the catching efficiency of the trawl used for tracking relative abundance over time. I agree with these findings.

Sampling variability in survey indices of abundance, and the variability in VPA estimates related to uncertainty in catch-at-age, is generally not taken into account in the determination of overfishing status. In my opinion, this introduces a risk, since true abundance could be substantially lower, or higher, than the value used in the determination of overfishing status. Effects of errors in catch-at-age, for example

resulting from poor information about discard, are likely to be of greater importance in the determination of fishing status than the minor change in trawl catching efficiency resulting from the warp offset on *Albatross IV*.

2.3. Trawling performance related to warp offset

The GARM reviewed the results of a series of 10 different studies to evaluate evidence of a reduction in trawl catch efficiency associated with the use of miss-calibrated trawl warps on *Albatross IV*. There were eight affected surveys (winter and spring 2000-2002; and fall 2000 and 2001). These studies covered three broad categories:

- a. Trawl geometry and performance in relation to the warp offsets as function of depth - direct observations of bottom contact and trawl configuration (wing-spread and headrope height measurements), as well as data on rates of gear damage from the *Albatross IV* surveys;
- b. Shifts in the time series of survey indices of abundance and estimated population characteristics (e.g., size distributions) resulting from reduced trawling efficiency;
- c. Fishing power studies - paired trawling conducted by *Albatross IV* alongside the *Delaware II* vessel before and after spring 2000.

I was very impressed by the thorough scientific analysis that had been conducted by NEFSC staff, using all available relevant data, to reveal any shift in trawl catching efficiency caused by the warp offset on *Albatross IV*. The analysts and methods experts, at the request of myself and other GARM members, conducted multiple additional analyses during the meeting. The results did not reveal any significant effects on the catching efficiency of the trawl related to the offset of the warps for depths where groundfish typically occur (warp offset up to about nine feet). In particular, a large number of parallel trawl hauls conducted by *Albatross IV* alongside *Delaware II* before

and after the warp offset showed virtually no change in relative fishing power between the vessels. The analyses were restricted to ten species (including cod, haddock, and yellowtail flounder) that had sufficient sampling coverage for a valid comparison of fishing power between vessels. The almost identical relative catch rates for flatfish before and after the warp offset clearly indicate that the trawl maintained good bottom contact despite the warp offset. In effect, this fishing power study can be considered a controlled experiment (although not intended at the time) to detect effects of the warp offset. The *Delaware II* served as a control because it used a fixed gear and trawling procedure for both time periods (before and after the warp offset on the *Albatross IV* gear). Thus, an appreciable reduction in catching efficiency for *Albatross IV* after the warp offset, as postulated by some, would have resulted in a change in the ratio of mean catches for the two vessels. This did not happen. Any substantial changes in the underlying fish abundance pre and post warp offset, which could be a confounding factor in the evaluation of *Albatross IV* survey indices alone, are essentially accounted for because of the parallel trawling (in space and time) with another vessel.

3. Conclusions and Recommendations

Assessments through 2001 of the 20 stocks under the Northeast fisheries management are of consistently high quality, based on models that are suitable for the available data for each stock. The assessments and updates, conducted by experienced stock assessment biologists, were subject to a rigorous review process. The potential effects of the warp offset on *Albatross IV* in recent years were thoroughly evaluated. Based on all available results, I firmly believe that the warp offset has had minimal effects on the stock assessments conducted in recent years. Because the reduction of trawl catching efficiency resulting from warp offsets appear to be very small in depths where groundfish typically occur (warp offset up to about 9 feet), an intensive and well designed parallel trawling experiment involving *Albatross IV* along with a control vessel (e.g., the *Delaware II* or an industrial fishing vessel) would be required to detect and calibrate such

effects. Such an experiment could also provide additional information on the fishing power in deep waters, with warps offsets up to 18 feet. Although recent trawl experiment studies showed that the net remained open at the maximum offset of 18 feet, the actual fishing power in this case can only be determined from further experimental trawling. It is my understanding that such an experiment is being planned by NEFSC in cooperation with the fishing industry. Depending on the results from such an experiment, additional evaluations of the effects of the warp offset on stock assessments might be warranted.

Age-based assessments (VPA or ADAPT) with appropriate tuning are generally considered to be robust, when assumptions about natural mortality and recruitment are reasonable. However, such age-based assessments are particularly sensitive to inaccurate information on catches at age, for example related limited sampling coverage (spatially and temporally) of landings, and unreported discards. I recommend that the variability in VPA (and ADAPT) assessments caused by sampling variability in estimated landings in number by age be evaluated, for example by applying boot-strapping to port sampling data in connection with the model runs. Also, biased assessments (of unknown magnitude) could occur when multiple survey indices used for tuning of VPA are assigned equal weights, regardless of spatial coverage and precision. Such bias can be severe when some surveys only cover a limited fraction of the distribution area of a species. One way to reduce or eliminate such bias is to combine the respective survey estimates by using a composite estimator with appropriate weighting of each series, before the series is applied in tuning of VPA models. Additional post-stratification might be appropriate when surveys overlap in a sub-area.

I also noted that index based methods for determining fishing status do not incorporate measures of uncertainty in relative fishing mortality rate in the determination of overfished status. Reference points are based on estimated relative F that support replacement of the stock, or a specified growth rate. In the index-based assessments, the relative fishing mortality rate of a species is estimated as the ratio of catch (or harvest for some species) to the relative estimate of abundance from the research trawl surveys. Uncertainty in this ratio estimate is due to sampling variability in the survey indices, as

well as in the estimated total catches. Also, bias would be introduced if the total catches were poorly estimated, for example due to unreliable or no estimates of discard. It is important that the precision and accuracy of the estimated fishing mortality (F) be quantified, so that risk of stock depletion (or of setting too harsh limits on catch size) under current management regimes can be evaluated. The research trawl surveys conducted by NEFSC are probability-based, and thus have the great advantage that precision in the survey estimates of abundance can be quantified. In fact, to my knowledge, the survey series started by NEFSC in 1963 (using *Albatross IV*) was the first example (worldwide) of applying stratified random sampling to trawl surveys. I recommend that NEFSC move towards using a more precautionary approach to determine status relative to reference points, for example based on confidence limits of the abundance estimates. Although the relative standard errors might be fairly large because the estimates are based on moderate sample sizes, it would be more in line with the pre-cautionary approach to take such uncertainty into account. This can be illustrated through an example: If a point estimate of F for a species is just below the overfishing threshold, but with an upper confidence interval that extends well above the threshold, this would clearly indicate that the risk of overfishing occurring is high. By contrast, an estimated F just above the threshold, but with tight confidence limits, suggest a lower risk of severe overfishing than in the former example. The comparison of sampling variability in survey indices of abundance to variability in estimates of catch-at-age can also be useful when planning allocation of resources among fisheries dependent and fisheries in-dependent sampling programs.

4. References

The Report of the Working Group on Re-evaluation of Biological Reference Points for New England groundfish Improving Fish Stock Assessments. National Academy Press, Washington DC, 1998. 177 p.

Improving Fish Stock Assessments. National Academy Press. Washington, DC, 1998. 176 pp.

Improving the Collection, Management, and Use of Marine Fisheries Data, 2000. Ocean Studies Board, National Research Council, 236 pp.

Appendix A

Bibliography of materials reviewed:

- Report of the Groundfish Assessment Review Meeting (GARM) Assessment of 20 Northeast Groundfish Stocks through 2001. (By the Northern Demersal Working Group; Southern Demersal Working Group; Assessment Methods Working Group).

The above report includes updated stock assessments for 20 groundfish species (Documents posted on www.nefsc.noaa.gov/garm):

- A. Georges Bank cod (Loretta O'Brien)
- B. Georges Bank Haddock (Jon Brodziak)
- C. Georges Bank Yellowtail Flounder (Chris Legault/Steve Cadrin)
- D. So. New England Yellowtail Flounder (Steve Cadrin/Chris Legault)
- E. Cape Cod Yellowtail Flounder (Steve Cadrin/Chris Legault)
- F. Gulf of Maine Cod (Ralph Mayo)
- G. Witch Flounder (Sue Wigley)
- H. American Plaice (Loretta O'Brien)
- I. Georges Bank Winter Flounder (Lisa Hendrickson)
- J. So. New England/Mid Atlantic Winter Flounder (Steve C/Mark)
- K. White Hake (Kathy Sosebee)
- L. Pollock (Ralph Mayo)
- M. Acadian Redfish (Ralph Mayo)
- N. Ocean Pout (Sue Wigley)
- O. Gulf of Maine/Georges Bank Windowpane (Lisa Hendrickson)
- P. So. New England/Mid-Atlantic Windowpane (Lisa Hendrickson)
- Q. Mid-Atlantic Yellowtail Flounder (Steve Cadrin/Chris Legault)
- R. Gulf of Maine Haddock (Jon Brodziak)
- S. Atlantic Halibut (Jon Brodziak)
- T. Gulf of Maine Winter Flounder (Paul Nitschke)

Results from the Methods Working Group Report on Examination of Possible Trawl Survey Time- Series Interventions beginning in 2000 were presented in power-point with additional hand-outs. The following issues were covered:

- Description of warp offset problem, and how it relates to fishing depth (Paul Rago)
- Gear Damage Studies (Larry Jacobson)
- Evaluation of Fish Size in Relation to Warp Offsets (Larry Jacobson w/ Anne Richards)
- Warp Experiment Information (Data provided by Lisa Hendrickson)
- Trawl Geometry and Related Issues (Paul Rago/Steve Cadrin)
- Mean/Variance Relationships in Fish Catch (Paul Rago)
- Catch-at-Depth Relationships (Paul Rago /Steve Cadrin)
- Changes in Abundance Indices Pre- and Post Warp Intervention (Steve Cadrin)
- Log Catch Ratios Between Affected and Unaffected Surveys (Larry/Steve)
- VPA Performance (Residuals, Retrospective Patterns) (Chris Legault)
- Evaluation of Fishing Power Experiments, 1980s vs. 2002 (Mike Fogarty/Steve Cadrin)

APPENDIX B: STATEMENT OF WORK

Subcontract between the University of Miami and Versar, Inc. (Dr. Jon Helge Vølstad)

Groundfish Assessment Review Meeting

The purpose of requesting outside peer reviewers from the Center of Independent Experts (CIE) is to provide input to the Groundfish Assessment Review Meeting (GARM) for northeast USA stocks. The Northeast Multispecies Fishery Management Plan (Multispecies Plan) includes 20 groundfish stocks. The GARM meeting (scheduled for 8-11 October, 2002, in Woods Hole, Massachusetts) will provide scientific review of assessment information and ancillary analyses. The CIE reviewers are requested to provide input on assessment results and forecasts, and to help construct the final report of the meeting.

The GARM meeting is a regional process for updating stock assessments using existing models, VPA formulations, and other assessment approaches. Specifically, the GARM will:

- A. Provide updated catch information (landings and discards, where appropriate) for the 20 stocks to be assessed (see list below), catch-at-age data (estimated based on port sampling, where applicable);

Cod	Gulf of Maine Georges Bank
Haddock	Gulf of Maine Georges Bank
Yellowtail flounder	Georges Bank Cape Cod Southern New England Mid-Atlantic
Winter flounder	Gulf of Maine Georges Bank Southern New England
Acadian redfish	
American plaice	
Witch flounder	
Pollock	
Windowpane flounder	Northern Southern
White hake	
Ocean pout	

Atlantic halibut

- B. Provide updated research vessel survey indices (through spring 2002) for all appropriate survey series, including NMFS spring and autumn series, Canadian series, and state surveys (as appropriate);
- C. Estimate fishing mortality rates (or appropriate proxies) for all 20 stocks (through 2001), and provide estimates of terminal year stock sizes;
- D. Evaluate stock status relative to applicable biological reference points (FMSY and BMSY) as provided in the Report of the Working Group on Re-Evaluation of Biological Reference Points for New England Groundfish (<http://www.nefsc.noaa.gov/nefsc/publications/crd/crd0204/>);
- E. Provide updated estimates of F-Rebuild (the fishing mortality rate required to rebuild biomasses to BMSY by 2009) for all applicable stocks; and
- F. Comment on the potential sensitivity of assessment results to trawl warp-marking discrepancies that occurred in surveys between spring 2000 and spring 2002.

Specific Responsibilities of the CIE Reviewer

The scientific expertise required is in the area of stock assessment and population dynamics.

The CIE reviewer's duties shall occupy no longer than 10 days: Several days prior to the GARM meeting for document review; four days to participate in the GARM meeting; one day following the GARM meeting to review the draft final workshop report; and several days to complete the report to be submitted to the CIE. No consensus opinion between the two CIE reviewers is sought.

Specific tasks and the schedule are itemized below.

1. Prepare for the GARM meeting by reviewing documents posted on the web prior to 8 October 2002. This web site is <http://www.nefsc.noaa.gov/garm/>. The login is "garm"; the password is "kingfish". Additional scientific information will be presented during the GARM meeting.
2. Serve as active participant in the GARM meeting from 8-11 October 2002, providing input, comment, and scientific overview of analyses, and actively participate in drafting the final report and conclusions of the GARM.
3. Review the draft GARM report during the week of 14-18 October 2002, so that the NEFSC can meet the deadline for completion of the final document by 21 October 2002. The review comments should be provided to the Northeast Fisheries Science

Center via Dr. Steven Murawski (508-495-2303, smurawsk@whsun1.wh.who.edu) no later than October 18, 2002.

4. No later than October 25, 2002, submit the written report¹ (see Annex I) addressed to the “University of Miami Independent System for Peer Review,” and sent to Dr. David Die, via email to ddie@rsmas.miami.edu, and to Mr. Manoj Shivlani, via email to mshivlani@rsmas.miami.edu. This report shall include the comments provided under task 3 above.

Signed _____

Date _____

¹ The written report will undergo an internal CIE review before it is considered final. After completion, the CIE will create a PDF version of the written report that will be submitted to NMFS and the consultant.

ANNEX I: REPORT GENERATION AND PROCEDURAL ITEMS

1. The report shall be prefaced with an executive summary of findings and/or recommendations.
2. The main body of the report shall consist of a background, description of review activities, summary of findings, and conclusions/recommendations.
3. The report shall also include as separate appendices the bibliography of materials reviewed for the GARM and a copy of the statement of work.