

Report of

Kenneth Patterson
51 Avenue des Glycines, 1950 Kraainem, Belgium

to

Centre for Independent Experts
University of Miami, RSMAS, FL USA

3 May 2005

1. Executive Summary

The benchmark assessment led to an appropriate perception of the problems in evaluating the state of the stock, and led to a reasonable strategy for resolving them. However, the credibility of the process may have been weakened by the non-attendance of the industry representatives, who may therefore be surprised by a new assessment and a changed advice in June. If possible, steps should be taken to communicate the outcome of the meeting to the industry sectors and to encourage their participation at future meetings.

The meeting was held in good order and completed within the set time. The benchmark process appears basically effective, but could be improved in some directions. In particular,

- More attention should be given to simple visual inspection of residual patterns during assessments. Data screening exercises would be a useful routine part of the assessment process.
- The assessment process could be assisted by obtaining more information from the fishery (e.g. observer reports concerning local catch rates, management agencies' assessment of the reliability of catch information, log-book reports, VMS information etc.)
- A distinction could be made between phases of identifying problems, estimating their magnitude and consequences, suggesting possible causative mechanisms, and drawing conclusions for the implementation of the assessment and management process. Some efficiency gains might be achieved by separating these elements of the work and assigning specific tasks to meeting participants.
- Model testing in a formal simulation framework would have been preferable, but sufficient resources were not available to set up such an ambitious programme. More resources directed in this direction would assist the process.

- Participants felt that adaptations of the harvest rule could only be proposed by the TRAC benchmark assessment within a management context. There is a need to resolve the institutional problem so that technical progress can be made.
- Biomass-dynamic models should not at present be used because the assumption of long-term stability in population parameters is apparently violated (and internal data inconsistencies may be missed).
- It should be encouraged that relevant scientists maintain a wide field of experience and expertise across different stocks by participating at other international meetings on fish stock assessment (e.g. ICES assessment and methodological meetings).

2. Background

The yellowtail flounder stock on the Georges Bank is a shared stock exploited by the USA and Canada. It is subject to management by TAC (shared according to an agreed percentage), by effort limits on the US, and is subject to several closed areas that were introduced in the mid 1990s.

Assessments are calculated using an age-structured assessment model with a long time-series of data from US surveys and a rather shorter series of Canadian surveys, as well as catch data. Substantial efforts have been made in standardising survey indices of abundance.

The stock is subject to an agreed target fishing mortality rate of $F = 0.25$ per year.

In the assessment calculated in 2004, substantial problems were identified (Legault and Stone, 2004). An important retrospective pattern was identified, and the previous perception of low exploitation rate was questioned. Compared with the assessment model projections, there were not enough older fish in the catches. Overall the assessment model fit was not considered sufficiently reliable to provide quantitative catch forecast advice.

The TRAC meeting of 25-29 April was convened with the terms of reference listed in Appendix I to address issues related to the reliable estimation of the state of the stock, including recommending a new assessment procedure and consequent methods for generating catch advice.

3. Description of review activities

Review activities were carried out as follows:

- a) Prior to the meeting, a data exploration exercise was carried out using data compiled by Overholtz et al. (2005). This was submitted as a working document to the meeting (Patterson, 2005a).
- b) Attendance at the meeting (from 27-29 April only, due to a flight cancellation), participation in discussions and contributions to the text of the final report. The substance of contributions so made is included under item 4.

c) During the meeting, a second working document was prepared in order to address some specific issues that had been the subject of discussions (Patterson, 2005b).

4. Summary of findings

4.1 Technical and Scientific Issues

4.1.1. Stock identity

Stock identity issues were reviewed by Stone and Legault (2005a) who identified no reason to change the assessment unit. This was accepted.

4.1.2. The Assessment

Considerable effort had been made in preparation of the meeting using a number of relatively complex modelling approaches. Arguably, simpler approaches and fewer analyses could have led to the same conclusions.

The evaluations made in Patterson (2005a,b) using separable models identified a number of specific problems with the assessment of this stock, which could also be found in other analyses:

1. A strong retrospective pattern exists with the existing model structure.
2. There is a large residual pattern indicating a change after the early 1990s.
3. Although the estimated biomass trend increases, commercial cpue decreases, or at least does not increase.
4. There are many fewer older fish in the data than are predicted by the assessment models.
5. Total mortalities estimated from survey data indicate an increasing trend, but this trend is not seen in the assessments.
6. The latest survey data (in 2004) are not consistent with previous assessments.

During the meeting, a number of different analytic approaches were applied.

Stone and Legault (2005b) recalculated the "traditional" assessment model using new data. They confirmed the identification of the problematic residual pattern beginning in the mid 1990s and coincident with the introduction of the closed areas, and also confirmed the persistence of the strong retrospective pattern.

Brodziak (2005) used a separable model with a multinomial error model. Eight hypotheses relating to age-related catchability were examined. All save one (which was judged non-credible) of the model fits reported estimated low fishing mortalities, and retained the feature of the large residual pattern identified in (2) above. Most model fits retained a mismatch between survey trends and modelled population trends in recent years.

Stone and Gavaris (2005) explored some issues for addressing the perceived problems, by extending the age-range in the assessment to 12 rather than 6 years, aggregating survey indices, and allowing a discontinuity in the estimate of survey

catchability. This also showed a discontinuity in catchability residuals around 1995, though this was palliated by assuming a higher rate of natural mortality. Model fits were rather sensitive to assumptions, and could 'flip' between perceptions of increasing or decreasing biomass in recent years. The discontinuity in catchability in the mid-1990s could be modelled. The authors also presented during the meeting some trials using a power-function catchability model for the younger ages, which reduced the trend in catchability residuals in the more recent years. Possible age-specific changes in natural mortality were investigated by Chouinard and found to be a possible, though not necessarily plausible, cause of retrospective effects.

Results of analyses attempting to draw inferences from retrospective patterns drew similar conclusions (Jacobson and Seaver, 2005) using a methodology due to Cadigan. These methods confirmed the discontinuity around 1995 and also identified another discontinuity around 2003 (but this was less reliably estimated). Hypotheses of missing catches or increased natural mortalities could be invoked to compensate for the observed patterns in retrospective analyses, but neither was considered wholly satisfactory.

Biomass-dynamic models were also tested (Cadrin, 2005) but after discussion it was concluded that it is problematic to omit age-structured information that yields valuable information about the state of the stock and the consistency of the model and the data. Furthermore, the assumption of stationarity in population parameters appears violated in this data set. An approach based on replacement ratios estimated from catch and survey data was presented by Rago, but it was concluded that this offers no advantages over the biomass-dynamic approach.

Overall, examination of the results of applying such methods led to a general perception in the meeting that the problems identified as 1. to 6. above were a reasonable representation of the statement of the problem. A full account of the models is to be found in the working documents and in the draft meeting report.

Considerable discussion continued on appropriate ways to resolve the problems, focussing on:

- a) Assumption of non-reported catches;
- b) Assumption of changing natural mortality rates;
- c) Changes in survey catchabilities, either over time or with respect to abundance.

Further analyses by Cadrin, Legault, Gavaris and in Patterson (2005b) suggested that problems 1 to 6 might be resolvable by both:

- Treating the relationship of catch to survey abundances in the period post 1995 as independent of the relationship in earlier years. However, no reason for such a difference could be identified.

- Treating the relationship of survey abundances to fitted populations as a power model for the years after 1995 and for ages 2 and 3 of the NMFS surveys. Again, no underlying causative mechanism could be identified.

Legault and Jacobsen also demonstrated that it is likely that the imposition of closed areas has probably caused an apparent dome-shaped exploitation pattern in the fishery. Analytic methods to address this were not presently available, save for the use of less restrictive assumptions in an "ADAPT" framework.

These issues were discussed and are to be reflected in the draft final report of the meeting.

4.2 Administrative Issues

The meeting was well conducted and held in a friendly and cooperative ambience, and with good facilities provided by NMFS. In an international context this is highly important. The terms of reference were addressed (as planned) by placing most emphasis on the stock assessment issues.

The discussion was relatively open and not tightly structured. It might have helped the process to move faster and be more efficient if there had been a clear distinction between issues, with decision-points identified at various stages in the meeting. Useful stages in the process might have been:

- Identification of the problem
- Hypotheses as to underlying causes
- Tests of such hypotheses
- Identification of future research activities
- New modelling or advisory procedures in the context of the problems identified.

There was also substantial duplication of effort, with different participants addressing substantially the same issue using slightly different assessment models. A better coordination of effort might have helped a wider range of issues being investigated.

Relatively little attention was paid to the advisory context and the methods for providing catch advice. The consensus in the meeting appeared to be that this was a management issue and not within the scope of the benchmark assessment procedure. This was regrettable, and (given that administrators and industry had been invited) an opportunity was missed to consider harvest rules and assessment uncertainties together in the same context. This appears driven by the institutional arrangements of the TRAC process.

It was disappointing that fishing industry sectors did not join the meeting nor nominate experts to the benchmark process. It seems likely that new assessment procedures will generate substantial need for dialogue with industry sectors, and here also an opportunity was missed to start this process.

4.4 Completion of Terms of Reference

ToR (1) (Assessment methods) took up the main business of the meeting. Biomass-dynamic and age-structured models of various types (including models assuming uncertainty in the catches-at-age) were widely used and tested. Delay-difference models were not tested, but this was not germane to the key problem issues. The methodological comparisons would have benefited from more rigorous testing against

clearly defined criteria and in a simulation framework, but the resources that would have been needed to develop this extended far beyond those available to the TRAC meeting.

ToR (2) was fully addressed and a strategy for addressing the next assessment of this stock was identified. Although not all issues had been resolved and a final assessment model was not chosen, good progress in defining the desired characteristics of a final assessment and in defining a method to reach such an assessment was reached. A definitive solution was not found and scope for disagreements remain, but progress can be considered satisfactory.

ToRs (3) and (4) were not explored in depth. It was felt by the participants that it was beyond the remit of the TRAC benchmark process to consider new or innovative methods for providing catch advice without the active participation of management agencies. This was regrettable but apparently unavoidable within the present structure of management arrangements for this stock. Further consideration might need to be given to this at a later stage.

5. Conclusions and Recommendations

- a. The benchmark assessment led to an in-depth investigation of the problems in evaluating the state of the yellowtail flounder stock, and led to a reasonable strategy for resolving such problems. As such, the benchmark process appears basically effective. Addressing strategies for providing catch advice was more difficult and this was not addressed in depth. Further comments refer to some possible improvements that could be made.
- b. More attention should be given to simple visual inspection of residual patterns during assessments. Age-specific plots of residuals with time and residuals with abundance are both useful, and could have been used to detect problems more quickly, at an earlier stage and more simply than some of the more laborious modelling exercised used during this meeting. Such plots should be routinely presented during assessments. A data screening exercise might be a useful routine part of the assessment process, eg. using multiplicative models on survey data.
- c. The assessment process could be improved by obtaining more fishery-dependent information. Although simply tuning assessments with unstandardised commercial catch rate information would be misleading, there could be valuable information in observer reports concerning local catch rates, the reliability of catch information, log-book reports, VMS information etc. Some time was spent at the meeting in speculating about fisheries activities when information should have been available (especially about the reliability of catch information).
- d. A distinction might be made between phases of identifying and clearly stating problems, estimating their magnitude and consequences, suggesting possible causative mechanisms, and drawing conclusions for the implementation of the assessment and management process. At times, different elements of these problems could get confused in the discussion.

- e. Model testing in a formal simulation framework would have been preferable, but sufficient resources were not available to set up such an ambitious programme.
- f. Errors and uncertainties in the assessment process are not necessarily best addressed in the assessment process itself. Instead, a management system can be constructed that has a built-in robustness to assessment error. Measures such as adopting a percentage limit on TAC changes, or setting a more cautious target fishing mortality, have been found to be useful in other areas. In this case however it appeared that adaptations of the harvest rule could not be proposed by the TRAC benchmark assessment outside of a management context. There is a need to develop a way forward on this issue.
- g. Biomass-dynamic models should not at present be used because the assumption of long-term stability in population parameters is apparently violated.
- h. Participants at the meeting were not fully aware of methodological developments in the long-term management context that are being made at ICES and other international scientific fora. It should be encouraged that relevant scientists maintain a wide field of experience and expertise across different stocks e.g. by participating in more international meetings.
- i. Useful topics for further research could include tagging studies to improve mortality rate estimates, spatially-structured simulation modelling to assess the effects of the closed areas, and simulation modelling of long-term assessment and management strategies.
- j. Two specific recommendations for the assessment of the stock are:
 - Methods that assume a stability of population parameters pre- and post- 1995 should not be used. Data contradict this assumption.
 - Methods that model the catchabilities of ages 2 and 3 in the NMFS surveys as linear relationships to the fitted populations also appear to be inappropriate.

APPENDIX 1. Background documents and references

Brodziak, J. (2005). Some STATCAM analyses of Georges Bank yellowtail flounder, 1935-2003. TRAC Working paper 2005/5, 26 April 2005.

Cadrin, S. (2005). Biomass dynamics modelling of Georges Bank Yellowtail Flounder: past, present and future. TRAC Working paper 2005/2.

Jacobson, L. and A. Seaver. (2005) Retrospective patterns in yellowtail flounder.

Legault, C.M. and H.H. Stone (2004). Stock assessment of Georges Bank (5Zhjmn) Yellowtail Flounder for 2004. Transboundary Resource Assessment Committee/Comité d'évaluation des ressources transfrontalières Doc. 2004/03.

Patterson, K.R. (2005a) Yellowtail Flounder in Georges Bank : Assessment Issues. Working Document to the Transboundary Resources Assessment Committee Georges Bank Yellowtail Flounder Benchmark Assessment Review 26-29 April 2005, Woods Hole MA, USA.

Patterson, K.R. (2005b) Yellowtail Flounder: Further explorations to reduce retrospective inconsistencies. Working Document to the Transboundary Resources Assessment Committee Georges Bank Yellowtail Flounder Benchmark Assessment Review 26-29 April 2005, Woods Hole MA, USA.

Stone, H. H. and C.M. Legault (2005a). Review of current management unit for Georges Bank (5Zhjmn) Yellowtail Flounder. TRAC Working paper 2005/1.

Stone, H. and C. Legault (2005b). A comparison of Georges Bank Yellowtail Flounder VPA results using the 2004 assessment input data and the revised input data from the benchmark assessment review. TRAC working paper 2005/1.

APPENDIX 2. Statement of Work

Statement of Work

Consulting Agreement between the University of Miami and Dr. Kenneth Patterson

March 21, 2005

Background

Since 1998, the Transboundary Resources Assessment Committee (TRAC) has reviewed stock assessments and projections necessary to support management activities for shared resources across the USA Canada boundary in the Gulf of Maine-Georges Bank region. These assessments are necessary to advise decision makers on the status of these resources and likely consequences of policy choices. The TRAC employs a two-tiered review process in which each of the stocks periodically undergoes an intensive peer review of the assessment model and assumptions. This is termed a benchmark assessment review. The benchmark assessment framework is applied as required, generally on an annual schedule, to provide the peer reviewed assessment of the resource status to fisheries managers.

The scope of a benchmark assessment can be categorized under four broad themes, a) definition of management unit, b) best approach for estimation of current status, c) determination of harvest strategy reference points and d) establishing projection procedures for evaluation of management tactic options. In any benchmark, not all themes may be subjected to evaluation. The extent of evaluation of a theme is defined by the terms of reference of the benchmark assessment. However, the basis of accepted practice for all themes should be documented.

The need to review the yellowtail benchmark assessment framework was raised during the June 2004 TRAC meeting. The review will focus on themes b and d, addressing theme c to the extent necessary as a result of model formulation changes and will simply document the basis for the established management unit, theme a. Two sets of meetings were deemed necessary to adequately review the framework, the first to consider the data input and summarization and the second to consider the most appropriate model formulation for synthesis of available information. These will be conducted during October 2003 – April 2004 in time to be applied for the June 2005 TRAC assessment review meeting. The assessment methods meeting to be held in April 2005 will have participation from NEFSC and DFO Stock Assessment teams and other laboratory scientists, invited external (not from NEFSC or DFO Scotia Fundy) reviewers (including one CIE expert), representatives from U.S. and Canadian management agencies, US State and Canadian Provincial representatives, and U.S. and Canadian fishing industry participants.

Specifics and Responsibilities

A data compilation meeting was held in Woods Hole, MA on January 25-27, 2004 to review methods to process the data inputs for use in the framework synthesis. The terms of reference for the data compilation meeting were:

1. Review the documentation on rationale for the current management unit.
2. Review fishery landings and estimates of discards from all fisheries.
3. Determine most appropriate methods for calculating fishery landings at age.
4. Determine most appropriate methods for calculating discards at age from all fisheries.
5. Determine most appropriate methods for calculating weights at age for the fishery catch.
6. Determine most appropriate methods for calculating survey indices of abundance at age.
7. Determine most appropriate methods for calculating weights at age for the population.
8. Explore fishery catch per unit effort indices for use as tuning indices.
9. Examine tagging data for application in stock assessment.
10. Set date for availability of finalized data to be used in benchmark assessment.

The assessment methods meeting to be held in Woods Hole, MA on April 25-29, 2005 will determine analytical procedures for synthesizing all available information to determine the current status of the resource will be reviewed. The Terms of Reference for this meeting are:

1. Explore full range of assessment methods for estimating current abundance and exploitation rate such as, but not limited to, catch curves, separable VPA, index based approaches, surplus production, delay-difference, calibrated VPA, errors in catch at age models.
2. Determine assessment approach that will be used in next assessment.
3. If required, update reference points for harvest strategy based on agreed assessment approach.
4. Formulate projection procedures for harvest advice based on agreed assessment approach.

The CIE consultant's duties shall occupy a maximum of 14 workdays; a few days prior to the meeting for document review; the TRAC meeting; and a few days following the meeting to prepare a Review Report. The CIE consultant will participate, as a peer, in meeting discussions on available assessment methods for estimating current abundance and exploitation rate, and will specifically address the terms of reference for the meeting. The CIE consultant will participate in the development of the TRAC documents that provide details of the assessment framework and the technical basis for the recommendations.

Following the meeting the CIE consultant will prepare a CIE report, which provides an executive summary, a review of activities and a summary of findings and recommendations in response to the Terms of Reference (see Annex 1 for report format). No later than **May 13, 2005**, this report shall be submitted to the CIE for review. The CIE reports shall be addressed to "University of Miami Independent

System for Peer Review,” and sent to Dr. David Sampson, via e-mail to David.Sampson@oregonstate.edu and to Mr. Manoj Shivlani via e-mail to mshivlani@rsmas.miami.edu.

ANNEX 1: Contents of Panelist Report

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, conclusions/recommendations, and references.
3. The report shall also include as separate appendices the bibliography of all materials provided and a copy of the statement of work.

APPENDIX 3. Reviewer's Schedule of Activities

22 April	Background reading and preliminary data analysis
23 April	Background reading and data analysis
24 April	Preparation of Working Document
25 April	Attempt to travel to meeting; Preparation of Working Document
26 April	Travel to meeting
27 April	Attend TRAC Benchmark meeting
28 April	Attend TRAC Benchmark meeting
29 April	Attend TRAC Benchmark meeting
30 April	Report preparation + Travel to Belgium
1 May	Arrive in Belgium
2 May	Report preparation
3 May	Final editing of report

NB: Flight from Brussels to Amsterdam on 25 April was cancelled and the outbound trip could not be completed as scheduled.