Proposal for a Transboundary Resource Assessment Committee
Benchmark of the Georges Bank Yellowtail Flounder Empirical Approach

National Marine Fisheries Service
Northeast Fisheries Science Center
Population Dynamics Branch
Woods Hole, MA 02543

January 14, 2014

In December 2013 the Population Dynamics Branch of the Northeast Fisheries Science Center (NEFSC) finalized a proposal to develop an empirical approach for estimating abundance and setting catch limits for Georges Bank yellowtail flounder. The empirical approach will evaluate all relevant data sources with respect to their support for alternative hypotheses on stock status and, if possible, their directional impact on catch advice (Attachment #1). Drafts of the proposal have been considered by scientists within the NEFSC, staff at the Northeast Regional Office (NERO), members and staff of the New England Fishery Management Council (NEFMC), and colleagues at the Department of Fisheries and Oceans (DFO), Canada. Various concerns have been raised about the proposal including questions about the methodology, the process for review, and how it will be used to formulate catch advice. Management advice for this stock is determined by negotiation within the Transboundary Management Guidance Committee (TMGC), a bilateral understanding between the US and Canada and translated into fishery regulations by each country's authorized organizations. As such, it is important to have a mutually acceptable process for convening and vetting the scientific basis for the catch advice. Discussions between scientific staff and NEFSC and DFO have led to a proposal in which the Empirical Approach would be reviewed as a "diagnostic benchmark" within the Transboundary Resource Assessment Committee (TRAC).

Attachment #1 is a general outline of how a TRAC benchmark would be conducted and a discussion of the merits of such an approach. A TRAC Benchmark review would occur April 14th-18th, 2014 in Woods Hole with participation from US and Canadian scientists, academics, interested parties from both countries, and a number of external reviewers selected and supported by both countries. The Terms of Reference (TOR) for the meeting will be restricted to evaluation of information relevant to the estimation of biomass and age composition from various data sources. The TRAC benchmark will not be a forum for introduction of alternative stock assessment models. That review has already taken place at the 2013 International Council for the Exploration of the Seas (ICES) Strategic Initiative for Stock Assessment Methods (SISAM) meeting in Boston. Analyses by leading scientists from around the world demonstrated that further consideration of alternative stock assessment models was unlikely

to reveal the underlying causes for the lack of model fit. Lack of fit, presumably due to one or more changes in the data, or assumed or estimated parameters, was a common feature in all models. The SISAM review suggested that stock assessment models were not sufficient to uniquely identify such changes. Instead, a focus on external information would be an appropriate approach to explore problems in model diagnostics and retrospective patterns. The TRAC "diagnostic" benchmark would address these concerns directly but we acknowledge that this departs from the conventional understanding of benchmark assessments.

A diagnostic benchmark assessment through the TRAC will follow well-established and understood conventions for evaluating the scientific basis for catch advice within the US-Canada understanding (Attachment #2) and ensure participation by Canadian colleagues. A diagnostic benchmark also allows for a more thorough external review of the proposed approach and increase the likelihood that it can be used for management. In Attachment #3 we provide draft terms of reference for consideration.

Attachment #1

An Empirical Approach to Setting Catch Limits for Georges Bank Yellowtail Flounder

Problem Statement: The stock assessment for Georges Bank yellowtail flounder suffers from a severe retrospective pattern. Likely causes of the retrospective pattern include misreporting of landings, underestimation of discards, or increases in natural mortality. Unfortunately neither the model nor ancillary evidence is sufficient to distinguish among these competing hypotheses. In the absence of unequivocal evidence, there is no expectation that an update of the current assessment approach will alleviate any of the concerns raised about this assessment. Independent reviews and tests of alternative models by stock assessment scientists at the recent ICES World Conference on Stock Assessment Methods failed to find acceptable alternatives. All of the models suggested that a change in the underlying data or assumed magnitude of natural mortality had occurred, although none of the models could identify a proximate cause. Given the continuing need for stock assessment advice and the likely futility of identifying the perfect model, we propose a new approach that relies heavily on contemporary information. In pursuing this new path, it must be recognized that some of the desirable features of stock assessment models, such as biomass reference points, rebuilding strategies, and forecasting, will be given up. Instead, the approach will focus on a more narrowly defined question of "What is the appropriate level of harvest in the upcoming fishing year?"

Technical Details: There are conflicting signals in the data. Survey trends indicate a rapid increase in the population from the mid 1990s through early 2000s followed by a slower decline. Age distributions from the surveys indicate high total mortality rates throughout the entire time period. Recent tagging studies confirm this high total mortality rate during 2003-2006. Catches have markedly declined in recent years. Dividing the catch time series by the survey time series produces a simple relative fishing mortality rate that shows high values in the 1970s, 1980s, and early 1990s then a sharp decline in 1995 to low levels since then. There is no evidence of a change in natural mortality rate, although fish condition (weight at length) has declined from the early 1990s through recent years. The conflict in the data arises because surveys suggest a high and steady total mortality (Z) despite a large sudden decline in relative fishing mortality (F) in recent years. When natural mortality (M) is assumed to be low and constant the current total mortality (Z) is much greater than F plus M, when in fact Z should exactly equal F plus M.

Proposed Solution: Given the aforementioned concerns, an entirely empirical approach could be used instead. This approach would be based strictly on the data observations: surveys, catch, and any empirical information available. No model would be used beyond Z = F + M.

Instead, the implications of different assumptions regarding recent natural and fishing mortality rate would be explored systematically to demonstrate the potential impact of different catch advice along with notes about the implications of these changes. The proposed method would use as much contemporary information as possible but require that proposed catch levels were logically consistent with the underlying hypotheses used to generate the abundance estimate.

Proposed Process for Georges Bank Yellowtail Flounder Empirical Approach

At its June 2013 meeting, TRAC agreed that a full conventional benchmark for yellowtail flounder was not feasible given the absence of any new data series and that the decision would be reconsidered pending the results of the July 2013 ICES World Conference on Stock Assessment Methods. The ICES Conference subsequently confirmed that none of the models tested provided unequivocal measures of stock abundance. The empirical approach presented here is considered a more complete analysis of the "trends in relative abundance and relative mortality rates derived from survey and fishery data" recommended as part of the benchmark formulation for this stock (Gavaris et al. 2005). However, since this approach is an expansion of the 2005 benchmark, further peer review is warranted**.** To meet this need we propose to conduct a diagnostic benchmark following the TRAC benchmark review process.

Development of the empirical approach will require close coordination with industry, academic and government partners. Prior to a TRAC integrated peer review, a series of informal meetings will be held with these partners to describe the proposed process and to more fully understand field experiments that may contribute to the formulation of biomass estimates. These meetings will be designed to explore the evidence with the same rigor applied when developing stock assessment models (See Term of Reference 1 in Attachment #3). The scope and timing of these meetings has not been determined but will be dictated by the current schedule of assessments for the Population Dynamics Branch and availability of Canadian colleagues. Following these informal meetings an integrated peer review will be held April 14th-18th, 2014 to examine the data analyses conducted for the empirical approach. Meeting participants will include TRAC members, NEFSC, DFO, and state scientists, academics, Council staff, industry stakeholders and invited external reviewers. The purpose of the peer review diagnostic benchmark is to determine if the empirical approach has correctly evaluated and summarized the available data for Georges Bank yellowtail flounder (See Term of Reference 1 in Attachment #3). One of the most challenging Terms of Reference is TOR 2 in which the consistency of alternative hypotheses will be evaluated. The meeting will address how the empirical approach could be used for catch advice but will not actually derive catch advice (See Term of Reference 3 in Attachment #3).

At the June 2014 TRAC meeting, the recommendations from the diagnostic benchmark meeting will be used to derive catch recommendations. Depending upon the outcome of the

benchmark, results will either be considered alone as a basis for catch advice or considered along with the current virtual population analysis (VPA) modeling results and relevant VPA sensitivity runs as have been conducted in the past[1]. The TRAC will synthesize all the available information to provide its recommendation on catch advice to the TMGC.

Gavaris, S., R. O'Boyle, and W. Overholtz. 2005. Proceedings of the Transboundary Resources Assessment Committee (TRAC) Benchmark Review of Stock Assessment Models for the Georges Bank Yellowtail Flounder Stock. TRAC Proceedings 2005/01. 36 p.

[1] An earlier draft document generated some concern regarding how the empirical approach might be used. This is due to the following statements in that draft: *"This would require TRAC rejecting the benchmark assessment model formulations and relying on the benchmark recommendation of using survey and catch information to generate catch advice. Thus, a benchmark would not be required…".* This wording reflects only one possible way in which the TRAC could synthesize all the information according to the benchmark formulation, but is not the only way (as noted by use of the words could and would in this section of the document). Other outcomes include placing more emphasis on the VPA results for catch advice, or using a blended approach of relying on some aspects of the VPA results but not the exact numbers, as has been done by TRAC in the last two years. The expectation is that the empirical approach will more clearly demonstrate the conflicts among the data sources for the Georges Bank yellowtail flounder stock and lead to a better understanding by scientists and managers about why modeling this stock has been so challenging.

Appendix: Technical Details for Proposed Solution

Given the uncertainties described above, current catch data and assumed levels of natural mortality can no longer be used to compute stock size estimates consistent with abundance measures derived from synoptic surveys. A new approach is proposed that relies on analyses of contemporary data and evaluation of alternative hypotheses. These hypotheses will be evaluated with respect to their internal consistency and with respect to their implications for other factors. A mass balance approach will be used to illustrate the implications of alternative estimates of stock size on the likely magnitudes of unreported landings, discard mortality, non-catch mortality, and natural mortality. For parameters that can be bounded but not estimated (e.g. trawl efficiency, post release survival of tagged fish, fraction of stock in Canada, etc.) sensitivity analyses will be used to construct profiles of stock sizes consistent with plausible hypotheses.

Some approaches that may prove useful include

1. Synoptic swept area estimates of abundance from multiple NEFSC and DFO surveys
2. Swept area estimates of abundance over a limited spatial domain, (e.g., 2013 cooperative survey)
3. Gear comparison studies of roller vs. cookie sweep gear conducted under Cooperative Research program
4. Cohort and static catch curves to estimate total Z
5. Long term tagging studies (Wood and Cadrin) to estimate survival rates
6. Short term tagging studies (Peterson estimate by Melgey) to estimate abundance
7. Analyses of condition factor and weights at age as predictors of natural mortality.
8. Sensitivity analyses of potential impacts of mortality from disease (*Ichthyophonus*) or predators (seals?).
9. Seasonal variation in abundance from Cooperative Research/RSA projects.
10. HabCam-based estimates of relative density.

The basic model of abundance would be based on empirical measures of abundance and assumed parameters as follows

$$N = \frac{A_d}{a p_d} \frac{I_t}{e}$$

Where **N** is the estimated total population , $\mathbf{I_t}$ is the index of abundance expressed as numbers or weight per tow, $\mathbf{A_d}$ is the total area within the sampling domain, a is the average area swept per tow, $\mathbf{p_d}$ is the fraction of the total area within the population domain, (*i.e.,* $\mathbf{p_d = A_d/A}$ where **A** is the total area where the stock resides), and **e** is the efficiency of the gear, expressed as probability of capture given encounter.

Some of these parameters are unknown or poorly known. An objective of the evaluation would be to develop realistic empirical bounds on efficiency derived from comparative experiments and to use survey indices to derive estimates of the fraction of total stock within US waters. The uncertainty in the unknown parameters and the sampling variability in the observations would be fully incorporated into overall abundance estimates.

Mortality estimates derived from catch curves and tagging studies would be compared to estimated catches and assumed values of M to create a similar range of population estimates. It is hoped that this piecewise construction of population estimates can be used to identify a range of plausible values for unknown parameters. A mass balance approach will be used to identify the magnitude of missing removals consistent with the swept area biomass estimates and the known removals via landings and discards.

The sampling distribution of population size would be carried through to create a distribution of catches consistent with the population estimates. Several approaches could be used. One approach would be a status quo method that multiplies the estimated abundance by the ratio of catch to relative biomass in recent years. Uncertainty in the estimate of the relative F could be propagated to develop a broad measure of uncertainty in the suitable catch level. Another approach that may be useful is to use an F derived from a yield per recruit analysis. An important aspect of this analysis would be the uncertainty in the discard rate and natural mortality rate. One would focus on predicted magnitude of landings, discards, and natural deaths to gauge their plausibility. Thus, catch advice would not be provided based on a standard assessment approach formula, but rather have to be agreed to by the TRAC (and SSC) given a range of possible quotas and plausible outcomes associated with each possible quota. Feedback from one year to the next in terms of responses in the fishery catch, survey time series, survey age structure, and other pieces of information would be an important component of this approach.

One potential advantage of this approach is that it might give participants a better understanding of the piecewise components of the assessment model. It might also create buy-in and acceptance from constituents who otherwise feel disenfranchised. It is also possible that none of these goals will be achieved, but we do not expect much acceptance from solely pursuing another update or, convening a conventional benchmark.

Attachment #2

Criteria for Evaluation and Modification of TRAC Benchmark Assessments [2]

At the April 2013 TRAC Benchmark meeting the following term of reference was addressed:

"Discuss criteria to determine:
1) When a benchmark assessment should be conducted and
2) What degree of modification is acceptable to make to benchmark model formulation during an update assessment."

The TRAC concluded the following:

"Without new information or modeling approach, requesting a benchmark would not be productive. During a TRAC update, changes to a benchmark model formulation would be presented as a sensitivity run and evaluated to see if a future benchmark would be required based on points outlined below. In all future TRAC assessments, a cumulative summary of changes to the current benchmark model will be included in the assessment research document.

1. Accumulation of data changes result in substantial change in catch advice relative to the benchmark formulation.
2. Change in either data or model results in substantial change in perception of stock size or stock structure.
3. On a regular basis, e.g. every five years, evaluate whether a benchmark review would be justified.
4. New data becomes available, e.g., new survey, that would affect model results.
5. Model results are inconsistent with observations; poor diagnostics.

In a TRAC update, if a sensitivity run suggests that a benchmark is required, the TRAC will present catch advice for both models with rationale as to why the sensitivity run would be preferred in the interim."

[2] *Based on excerpt from :* Claytor R. and L. O'Brien. editors. 2013. Transboundary Resources Assessment Committee Eastern Georges Bank cod benchmark assessment and TRAC Benchmark Criteria Discussion. TRAC Proceedings 2013/x, in review

Attachment #3

Draft Terms of Reference for TRAC Georges Bank Yellowtail Flounder Diagnostic Benchmark 2014

In the 2013 TRAC Status Report (TSR) the following Special Comments were provided:

> *The TRAC acknowledges that the assumptions made about population dynamics in the model do not fully capture the trends in the data. However, the model's conclusion that stock conditions are poor is valid.*
>
> *There is a continued need to conduct research to limit the possible causes for the retrospective bias exhibited in this assessment.*

In response to these comments, the 2014 benchmark meeting is designed to explore all the data available for Georges Bank yellowtail flounder, including data that cannot easily or feasibly be incorporated in a stock assessment model. The purpose of this exploration is to evaluate possible sources of the poor diagnostics exhibited by the current Virtual Population Analysis (VPA). The work to be reviewed during this 2014 benchmark extends the 2005 benchmark assessment which recommended consideration of "trends in relative abundance and relative mortality rates derived from survey and fishery data" (Gavaris et al. 2005). The 2014 diagnostic benchmark will not examine alternative stock assessment models. Such an examination was conducted during the ICES World Conference on Stock Assessment Methods (July 2013, Boston, MA) where no model was found that performed well relative to all the data. As such, the following terms of reference are strictly limited to exploration of the data.

Terms of Reference

1) Summarize all available data for Georges Bank yellowtail flounder which can be used to explore possible causes of the poor diagnostics in the current VPA for this stock.
2) Determine which pieces of information are consistent with alternative hypotheses regarding current stock status (e.g., current population is near carrying capacity, current population is near a desired amount, and current population is well below a desired amount).
3) If possible, describe how catch advice could be provided based only on the data (e.g. without relying on a stock assessment model).

Date of the benchmark meeting: Week of April 14, 2014.

All individuals interested in presenting a working paper for this meeting must contact the US and Canada Co-Chairs no later than February 10, 2014 to indicate their intention to present and to identify their intended topic. Working papers will be due 2 weeks prior to the meeting, so the deadline to submit working papers will be March 28, 2014. Authors must be present at the meeting or via webex to present their working papers. Failure to adhere to these TRAC

protocols will result in the working paper being excluded from the meeting agenda. These protocols are designed to allow sufficient time for meeting participants to review the material and to ask questions of the authors during the meeting.

Gavaris, S., R. O'Boyle, and W. Overholtz. 2005. Proceedings of the Transboundary Resources Assessment Committee (TRAC) Benchmark Review of Stock Assessment Models for the Georges Bank Yellowtail Flounder Stock. TRAC Proceedings 2005/01. 36 p.