#9a

Index Based Methods and Control Rules 2020 Research Track Assessment

Dr. Christopher M. Legault
Chair of the Index Based Methods and Control Rules Working Group

Presentation for
New England Fishery Management Council
28 January 2021

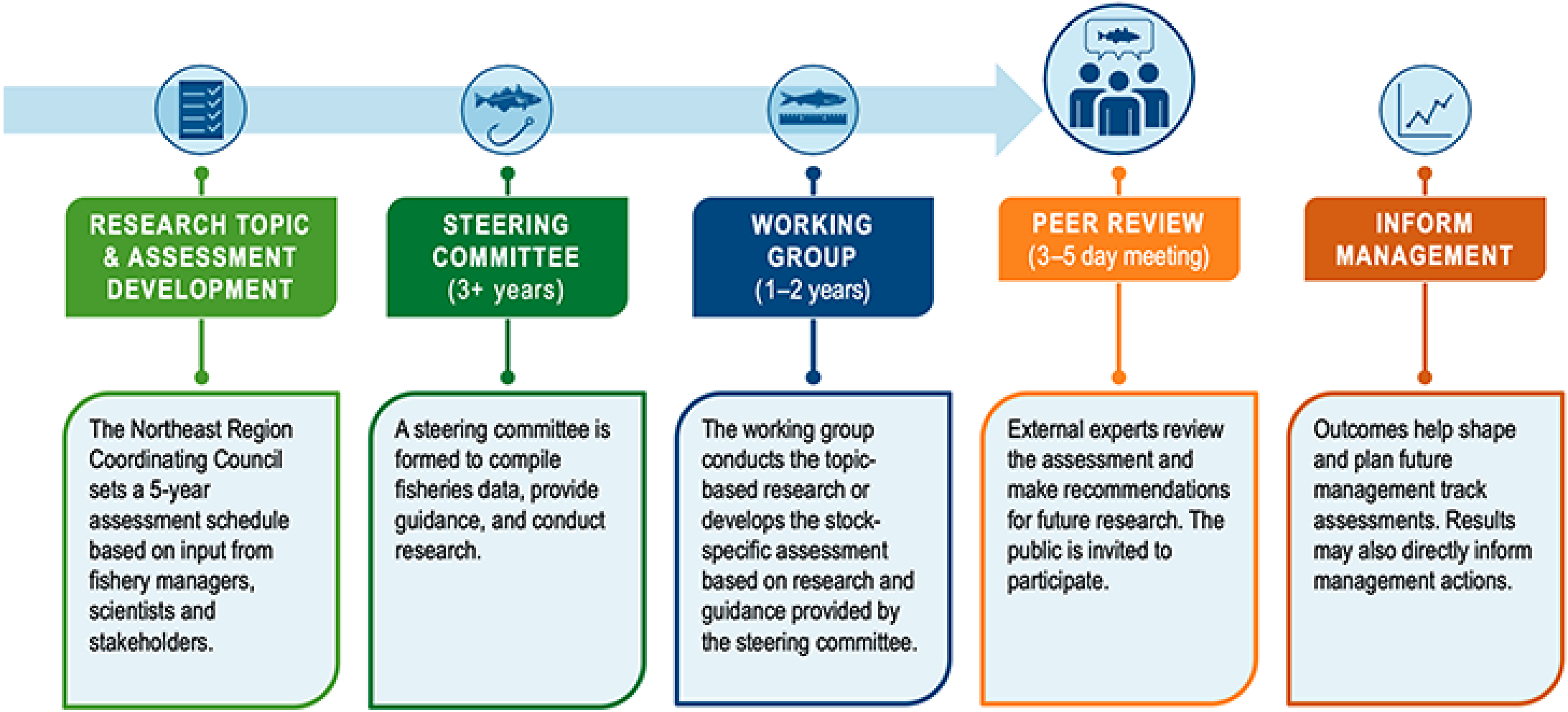
Take home

- No IBM better than SCAA
- Performance of IBMs differed
  - One group of IBMs good for rebuilding overfished stock
  - Another group of IBMs good for stocks in good condition
- Did not solve reference points for index-based stocks

Genesis

- Retrospective patterns an issue for some (but not all) assessments in Northeast region
- Strong retrospectives led to 7 age-based models being rejected and replaced by index-based methods
- There are a range of index-based methods in the region
  - Both rejected age-based and always index-based
- Guidelines for picking an index-based method would help
- First Topic-based (instead of stock-specific) Research Track

RESEARCH TRACK STOCK ASSESSMENTS



https://s3.amazonaws.com/nefmc.org/Stock-assessment-process-June2020.pdf

Actual Timeline

- IBM Research Track approved 16 May 2019 by NRCC
- TORs finalized 10 Jan 2020
- WG formed 16 March 2020
- Weekly meetings began 26 March 2020
  - Total of 41 meetings
- Peer review 7-11 Dec 2020

10 months

8 months

https://www.fisheries.noaa.gov/new-england-mid-atlantic/population-assessments/stock-assessment-working-group-index-based-methods-and-control-rules

IBMWG

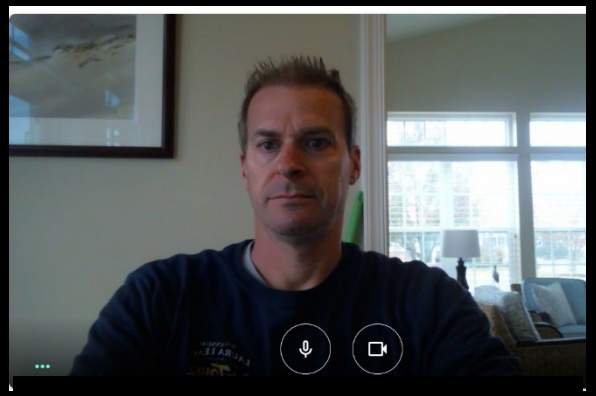

Andy Jones (NMFS)


John Wiedenmann (Rutgers)


Jon Deroba (NMFS)

HM =Honorary WG Member


Chris Legault (NMFS)


Brandon Muffley (MAFMC)


Gavin Fay (SMAST)


Liz Brooks (NMFS, HM)

Tim Miller (NMFS)


Rich Bell (TNC)


Jamie Cournane (NEFMC)


Joe Langan (URI, HM)

TORs

1. Develop methods to create data that if assessed with standard age-based approaches (e.g., VPA or ASAP) could exhibit a strong retrospective pattern.
2. Identify a number of index-based methods and a range of harvest control rules for use in closed-loop simulation, using index-based data resulting from ToR 1.
3. Identify metrics from the index-based assessment results that could be used in evaluations of trade-offs in performance among harvest control rules and index-based methods.
4. Evaluate the combinations of index-based methods and control rules using the metrics in ToR 3 to determine candidates for consideration by the Councils or other management authorities.
5. Provide guidance on specific situations that are and are not well-suited for a particular control rule or index-based method identified in ToR 4.
6. Create guidelines for setting biological reference points for index-based stocks.

Make Data

Pick IBMs

Select Metrics

Crank Sims

Advise

Ref Points

Make Data

- Closed loop simulations using WHAM
- Groundfish-like
- 50 year base period followed by 40 year feedback period
  - Assessments every other year in feedback period
- 8 conditions for base period
- 1,000 simulations each
- Mohn's rho of 0.5 for SSB when SCAA applied
  - Catch or M as source of retrospective

IBMs

1. AIM = An Index Method *#
2. CC-FM = Catch Curve
3. CC-FSPR = Catch Curve *
4. DLM = Dynamic Linear Model
5. Ensemble method
6. ES-FM = Expanded survey biomass
7. ES-Frecent = Expanded survey biomass *#
8. ES-FSPR = Expanded survey biomass *#
9. ES-Fstable = Expanded survey biomass
10. Islope = common trend based IBM *
11. Itarget = common level based IBM *
12. PlanB (PBS) = survey smoother *#
13. Skate = catch/B driven *#

* indicates member of Ensemble method
# indicates used in regional assessment

FSPR = F40%SPR
FM = F set equal to M
Frecent = average of recent 5 years catch/B
Fstable = F to create stable population

Not Tuned

- All IBMs used in formulaic approach (hands off)
- Real assessments would examine diagnostics from methods
- Future research: dig into results to see if diagnostics would have rejected some catch advice
- Note: DLM < 1,000 simulations due to time constraints

2 Harvest Control Rules

A = applied catch advice directly (treat like ABC)

R = reduced (multiply catch advice by 0.75) (treat like OFL)

- Neither accounts for relative stock size
  - Some IBMs do not estimate relative stock size

50 Metrics

catch_a_iav_catch
catch_l_avg_catch
catch_l_avg_catch_msy
catch_l_iav_catch
catch_l_prop_g_msy_2_of_3
catch_l_sd_catch
catch_s_avg_catch
catch_s_avg_catch_msy
catch_s_iav_catch
catch_s_sd_catch

f_l_avg_f_fmsy
f_l_is_gr_f_dot_1_bmsy
f_l_is_gr_f_dot_5_bmsy
f_l_is_gr_fmsy
f_l_is_less_f_dot_1_bmsy
f_l_is_less_f_dot_5_bmsy
f_l_is_less_fmsy
f_l_n_gr_f_dot_1_bmsy
f_l_n_gr_f_dot_5_bmsy
f_l_n_gr_fmsy
f_l_n_less_f_dot_1_bmsy
f_l_n_less_f_dot_5_bmsy
f_l_n_less_fmsy
f_s_avg_f_fmsy
f_s_is_gr_f_dot_1_bmsy
f_s_is_gr_f_dot_5_bmsy
f_s_is_gr_fmsy
f_s_is_less_f_dot_1_bmsy
f_s_is_less_f_dot_5_bmsy
f_s_is_less_fmsy
f_s_n_gr_f_dot_1_bmsy
f_s_n_gr_f_dot_5_bmsy
f_s_n_gr_fmsy
f_s_n_less_f_dot_1_bmsy
f_s_n_less_f_dot_5_bmsy
f_s_n_less_fmsy

ssb_l_avg_ssb_ssbmsy
ssb_l_is_ge_bmsy
ssb_l_is_less_01_bmsy
ssb_l_is_less_05_bmsy
ssb_l_n_ge_bmsy
ssb_l_n_less_01_bmsy
ssb_l_n_less_05_bmsy
ssb_s_avg_ssb_ssbmsy
ssb_s_is_ge_bmsy
ssb_s_is_less_01_bmsy
ssb_s_is_less_05_bmsy
ssb_s_n_ge_bmsy
ssb_s_n_less_01_bmsy
ssb_s_n_less_05_bmsy

3 Sets of Simulations

- Base
  - 13 IBMs x 16 scenarios = 208 combinations
- No retro
  - 12 IBMs x 2 scenarios = 24 combinations
- SCAA
  - 1 “IBM” x 4 scenarios = 4 combinations

Total of
230,147 simulations
~300 GB results

Analyses

- Linear Models
- ANOVA
- Scorer App
- Graphical (lots)
- Risk/Status
- Grouping

Guide to Appendix 6

Figure Type	Report Example	Base	No retro	SCAA
Number of sims	N/A	1	1	1
Scores	4.1	2-14	188-200	269-281
Boxplots	4.9	15-32	201-218	282-299
Trade off (means)	4.15	33-38	219-224	300-305
1,000 points	4.11	39-96	225-260	306-341
Bagplots	4.12	97-125	N/A	342-359
Scenario panel sorted	4.3	126-131	261-266	360-365
Status	4.19	132-133	267-268	366-367
Confetti	4.13	134-187	N/A	N/A
ANOVA plots	N/A	368-445	N/A	N/A
Heatmaps	5.1	446-450	N/A	N/A

Base Results

- Two groups
  - CC-FSPR, CC-FM, DLM, PlanB, ES-Frecent, Islope: SSB and F ratios better
  - Skate, AIM, ES-Fstable, ES-FSPR, ES-FM, Ensemble, Itarget: Catch ratios better
- SSB ratio > 1 for all IBMs when M retro source
- SSB ratio >1 for 1st group of IBMs when catch retro source
- In real assessments, usually cannot determine retro source, so 1st group of IBMs recommended when rebuilding needed
- Smoother IBM (e.g. Plan B smooth) should not use HCR that reduces catch advice when stock size is high

No Retro Results

- No retrospective source causes long term SSB and F to be closer to MSY values than either the catch or M retrospective sources
- Still lots of variability among the simulations
- Alternative to bigger is better approach to metrics should be considered in the future

SCAA Results

- Overall, none of the IBMs outperformed SCAA with rho-adjustment
- When an SCAA is rejected due to retro, should not expect IBM to perform better than rejected model
  - Also lose the use of additional data, status determinations, and hypothesis testing with IBMs

Caveats

Caveat	Address in future using this framework?
• Groundfish-ish	• Yes
• Single source and magnitude of retro	• Yes
• Changing forcing function for retro over time	• Yes, with mods
• Assessed every other year	• Yes
• Hands off applications of IBMs	• No
• Limited methods to derive catch advice	• Yes, with mods
• Limited formulations of IBMs	• Yes, with mods

Ref Points

- Did not provide guidance
  - Lack of production function in most of the IBMs prevents analytical ref points
  - Insufficient time to explore ad hoc approaches to setting ref points

General conclusions and recommendation

- For stocks that have had an age-based assessment rejected due to a strong retrospective pattern, there is no expectation that an index-based assessment will perform better than a rho-adjusted statistical catch at age analysis.
- The performance of an index-based assessment in a specific situation can be analyzed through the framework developed for this project, but requires specific hypotheses about possible sources of the retrospective pattern.
  - The IBMWG recommends this framework be used for all assessments that have changed from age-based to index-based due to retrospective patterns, using biological and fishery settings appropriate for that stock to ensure the selected index-based method has a high probability of providing reasonable catch advice.
- The IBMWG recommends future research be conducted to both analyze the results of this study in more detail as well as build on this study to address other questions

Process conclusions and recommendations

- More time needs to be allocated to topic-based research tracks in order to fully address the TORs.
- Covid-19 travel limitations required weekly meetings with remote collaborations. This may be a useful approach for future topic-based research tracks, but may be less useful for stock-specific research tracks. The time certain weekly meetings helped track the programming and decision making progress necessary for the large simulation study. This may not apply well to stock-specific research tracks.
- GitHub was helpful for coordinating coding among multiple programmers.
  - Training session early on would ensure everyone able to work together efficiently
- This was a big project that required lots of computing power. The cooperation of network users not involved in the project to free up computing time was greatly appreciated.
- Fast internet speed an issue for moving large files and large numbers of files
  - Cloud computing would have been helpful
- Google docs with prompts before meeting allowed asynchronous contributions and then could build on it during meeting
- Google docs handy for meeting notes but not great for report writing
  - Need workflow all the way through to final report (508 compliance)
  - Would be helpful to be able to use Rmarkdown so don’t have to update tables and figures in report by hand
- Project management software (e.g., Jira) could be useful but would require training

Peer Review

- Panel
  - Drs. Paul Rago (chair), Yong Chen, Robin Cook, Paul Medley
- Fully met TORs 1-4, partially met TORs 5-6 (insufficient time)
- Recommended work continues on topic

Overview-General Perspective

- Impressive quantity and quality of work
- Operational model and scenarios well conceived and tailored to address retrospective problem
- IBMs appropriately matched to methods currently in use; new methods appear promising
- As in all simulation studies, conclusions are conditional on scenarios chosen.
- Management of work flow was unique and productive. Broad appeal to larger audience.
- Harvesting of model outputs will allow further insights and refinements
- Lack of time to complete all tasks.
- Research entails discoveries that necessitate refinement of the TOR.

Slide from Peer Review Panel summary presentation

Comments on Process

- Distributed system for workgroup participation worked well. Local control facilitated rapid evolution of system and responsiveness to TOR
- Software system has generality and could be used not only for further investigation of existing models and control rules, but also other operating models, control rules, and fine tuning of index-based models.
- Saving outputs of model runs allows further analyses
- Encourage further work to synthesize current work and investigate other options
- Use of external advisory panel DURING process could be helpful for longer term Research Track assessments.
  - Could allow for modification of TOR.
  - Milestones for intermediate review.
  - Gain external perspective
- Modeling framework has potential to evaluate effects of missing 2020 survey data on model performance

Slide from Peer Review Panel summary presentation

Some General Concerns

- Report is rough in spots and needs refinement. Fold in information from presentations. Refine description of DLM to clarify.
- Inferences about ability of retro-adjusted SCAA to improve advice may be limited by presence of multiple real-world sources of retro pattern
- Conclusions strongly conditioned on assumption of no trend in survey catchability
- There seems to be some ambiguity in the desire for a stock status determination and a harvest control rule
- Closed-loop hands-off application of IBMs (i.e., no tuning) may restrict conclusions.
  - Does not fully replicate actual application.
  - Count # failures during simulations
- What is “better” for metrics of performance? More comments to follow.
- Conclusions are necessarily restricted to the domain of the simulated stock. Therefore generalization to real world should be suitably tempered.

Slide from Peer Review Panel summary presentation

General Recommendations (1 of 2)

- Strongly encourage additional work to interpret existing simulations and test alternative hypotheses and models.
- Investigate basis for differences amongst IBMs
- Some specific suggestions
  - Compute frequency of F=2 constraint in simulated runs
  - Distill F/F_msy vs B/B_msy plots using Odds ratio
    - Consider within simulation phase plane estimates
  - Consider var(F/F_msy) vs var(Catch/MSY) plots to examine IBM groupings
- Develop additional data extraction procedures, similar to Scorer App
- Consider various data reduction (e.g., PCA) approaches for analyses of performance metrics among models

Slide from Peer Review Panel summary presentation

General Recommendations (2 of 2)

- Short term(6 yr) and long term (20 yr) metrics characterize different facets.
  - Perhaps best to consider separately.
  - Long-term metrics will not necessarily characterize future state of resource.
- Decision Theory Perspectives
  - Implementation of IBMs via a decision tree approach would be useful to explore. Could build on results of Anova etc.
  - Consider various forms of utility functions including regret and satisficing.
- Pursue DLM approaches, but not fully exploited in this application.
- Appreciated use of ensemble models. Further refinement of candidate models and basis for weighting would be valuable.

Slide from Peer Review Panel summary presentation

Council Considerations

- No decision required – informational presentation
- Future Research Tracks
  - Timing
  - TORs
  - External advisory panel during process
- Continue IBMWG?

Take home

- No IBM better than SCAA
- Performance of IBMs differed
  - One group of IBMs good for rebuilding overfished stock
  - Another group of IBMs good for stocks in good condition
- Did not solve reference points for index-based stocks

Questions?

chris.legault@noaa.gov

Data Portal:

https://apps-nefsc.fisheries.noaa.gov/saw/sasi/sasi_report_options.php

Assessment Year: 2020

Species Name: Index Based Methods

Stock Area: Georges Bank

Review Type: Research Track