SSC risk table workshop: Assessment category

NOAA FISHERIES
Alaska Fisheries
Science Center

Grant Thompson

February 5, 2021

Focus on BSAI assessments as examples

- Total risk tables (assessments) = 23
  - 2 from 2019 (Alaska plaice, northern rockfish)
  - 21 from 2020 (everything else)
- Distribution of levels for the assessment category:
  - Level 3: n=2
    - sablefish, blackspotted/rougheye rockfish
  - Level 2: n=5
    - northern rock sole, POP, northern rockfish, other rockfish, sharks
  - Level 1: n=16
    - everything else
- Almost no correlation between level and tier (= -0.044)
  - Authors were comparing either to previous assessments, or assessments within the same tier?

Overview of rationales listed by authors

- Summary of rationales by category
  - "Positive" means that the rationale suggests *lower* risk
  - "Negative" means that the rationale suggests *higher* risk

	Total occurrences			Unique rationales		
Category	Positive	Negative	Total	Positive	Negative	Total
Data quality/availability	20	24	44	8	8	16
Retrospective bias	14	6	20	4	1	5
Goodness of fit	12	4	16	4	3	7
Model behavior	9	4	13	5	4	9
Model structure	2	8	10	2	5	7
Alternative models	2	2	4	2	1	3
Harvest control rules	0	2	2	0	1	1
Other	4	2	6	4	2	6
Total	59	50	109	25	23	48

Most common positive rationales

- 25 unique rationales supporting a lower risk level (“positive”)
- Unique rationales with no. occurrences ≥ 2:

Rationale	Count
Small retrospective bias	10
Good fits to data overall	9
Results from analysis of missing a survey indicate no major problems	8
Annual surveys through 2019	4
No convergence issues	3
Good availability of age data	2
Mis-ageing is not a concern	2
New data had little impact	2
Recruitment estimates are consistent with the data	2
Retrospective bias is improved relative to previous assessments	2

Most common negative rationales

- 23 unique rationales supporting a higher risk level ("negative")
- Unique rationales with no. occurrences ≥ 2:

Rationale	Count
Lack of 2020 summer trawl surveys	10
Large retrospective bias	6
Data conflicts exist	4
Lack of EBS slope surveys since 2016	4
Tight prior distributions cause uncertainty to be underestimated	3
Alternative models show disparate results	2
Shortcomings in harvest control rule	2
Shortcomings in structure of Tier 5 RE model	2
Strong residual patterns	2
Survey biomass estimates are relatively imprecise	2

Tangible steps toward quantifying risk (1 of 2)

- Thompson (2018) suggested that a multivariate logistic equation could be used to determine the appropriate reduction in ABC:

$$reduction = 2\left(1 + \exp\left(-\sum_{var=1}^{nvar} x_{var}\beta_{var}\right)\right)^{-1} - 1\,,$$

- where $\mathbf{x}$ is a vector, each element of which is either 0 or 1, indicating whether the rationale corresponding to that element applies; and $\boldsymbol{\beta}$ is a vector of non-negative coefficients
- Thompson tallied the rationales corresponding to $\mathbf{x}$ and then fit $\boldsymbol{\beta}$ to the 76 instances between 2003 and 2017 in which one or both GPTs recommended reductions, resulting in $R^2 = 0.824$
- SSC minutes (10/18): "**The SSC recommends not pursuing this analysis further.**"

Tangible steps toward quantifying risk (2 of 2)

- Suppose that the SSC is open to reconsidering its position
- Here are some initial steps toward developing a model that quantifies the importance of rationales that have been identified:
  - Average number of positive rationales: 2.783
    - Range: 0 – 7
  - Average number of negative rationales: 2.261
    - Range: 0 – 7
  - Indep. var. #1: $x_1 \equiv$ difference (# negative minus # positive)
  - Indep. var. #2: $x_2 \equiv$ presence of both "data conflict" and "large ρ"
  - Model: $level = 1.328 + 0.090x_1 + 1.267x_2$
    - $R^2 = 0.721$

Internalizing structural uncertainty (1 of 2)

- Using the risk table to account for external structural uncertainty
  - Run n models
  - Choose a preferred model
  - Note that, because the n models imply n different ABCs, the preferred model does not account for structural uncertainty
  - Raise the risk score for the assessment category accordingly
  - After considering all four risk categories, (perhaps) recommend an *ad hoc* reduction from the maxABC implied by the preferred model
- Using ensemble modeling to internalize external structural uncertainty
  - Run n models
  - Choose a set of model weights
  - Create an ensemble model as the weighted average of the n models
  - Recommend no reduction from the maxABC implied by the ensemble

Internalizing structural uncertainty (2 of 2)

- Using a factorial design to create an ensemble can help to avoid "stacking the deck"
- Some possible factors:
  - Data selection, for example:
    - Choice of data sets
    - Choice of data weighting
  - Parameterization, for example:
    - Choice of functional forms
    - Choice of fixed parameter values
  - Model complexity, for example:
    - Number of free parameters
    - Number of time-varying parameters

Double counting in the assessment category

- Do either of the following constitute double counting when used as rationales for an increased risk level in the assessment category?
  1. Signals in the data that are being fitted by the model
     - E.g., “The survey biomass data show a downward trend, so the assessment risk level should go up”
  2. Uncertainties in the data that are incorporated in the fitting process
     - E.g., “The variances associated with the survey biomass data are large, so the assessment risk level should go up”
- Perhaps the relevant question is not, “Do the estimates produced by the model already use this information when estimating maxABC?” but, “Does this information suggest that ABC should be less than the model’s estimate of maxABC (i.e., the default buffer is too small)?”
  - If so, then #1 constitutes double counting, but #2 might not