SAW 53 Northern Demersal WG: Discussion and Decisions Leading to Consensus Model for the Gulf of Maine Cod Assessment

Presented to SSC

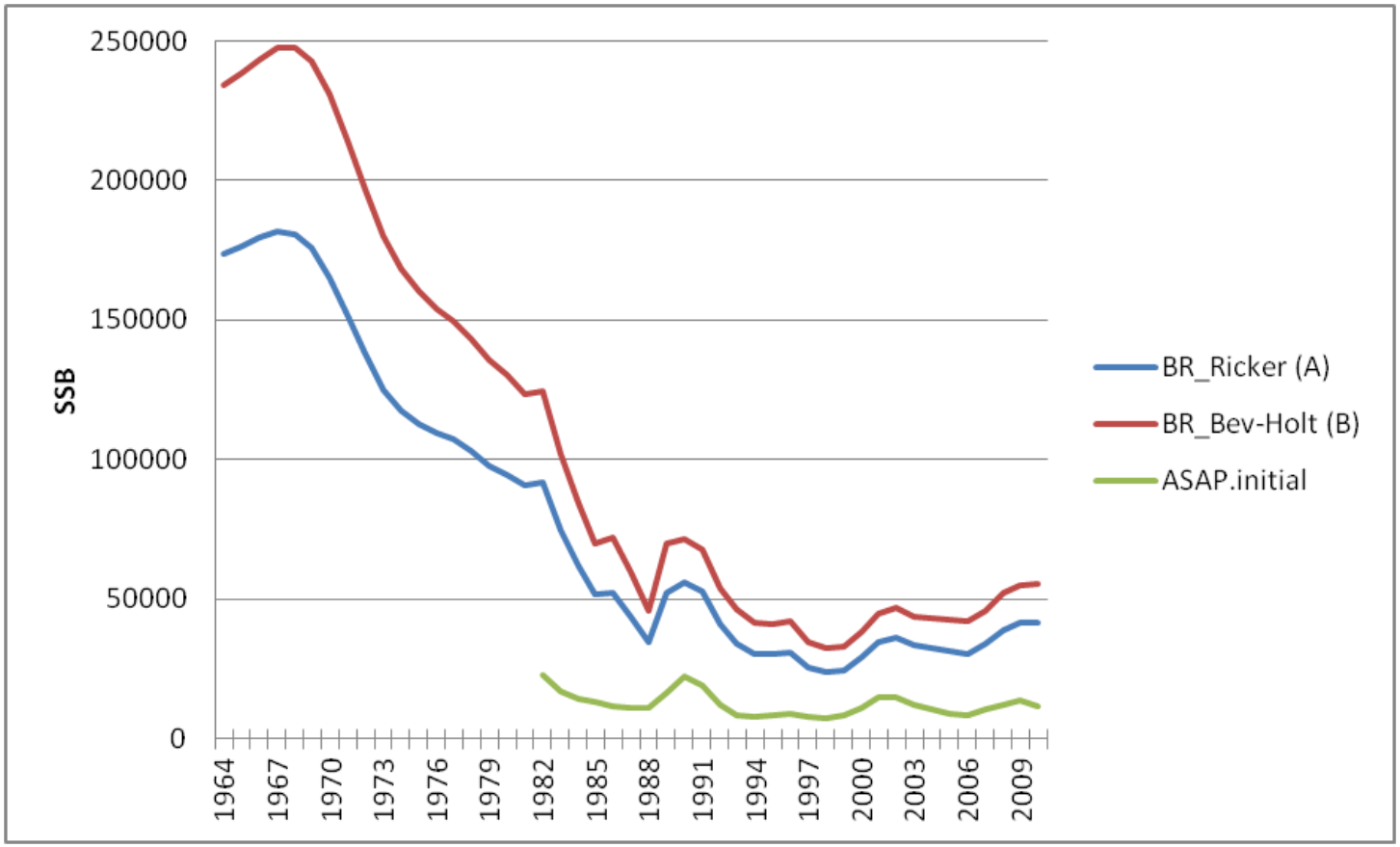
By the Chair of the Northern Demersal WG

28 March 2012

Path to WG Consensus on Assessment Model Configuration

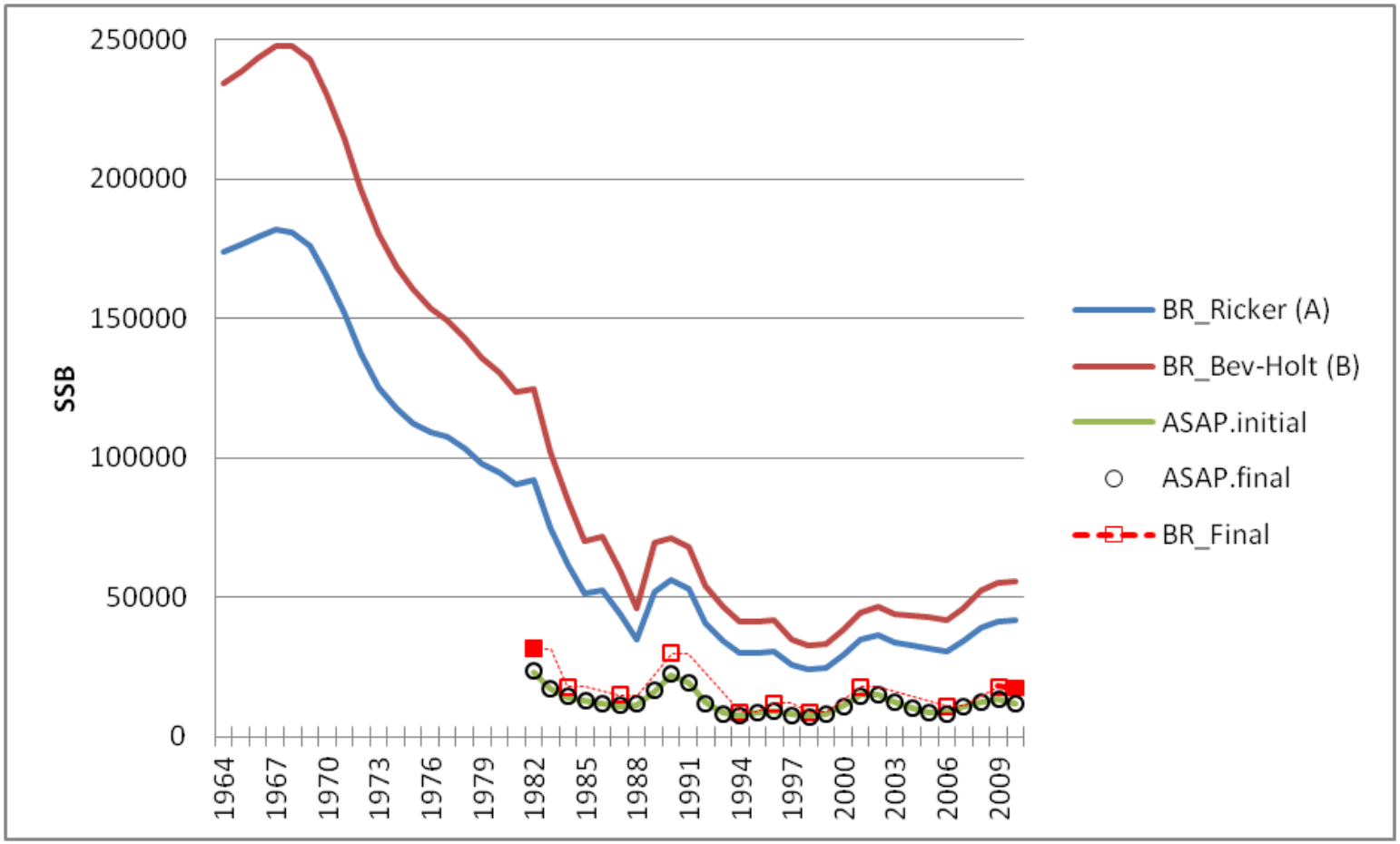
- Data meeting held Sept. 7-9, 2011
  - Data emailed to B&R Sept. 16, 12:05 EST
  - ➢ This is 1 month prior to the model meeting
- Model meeting held Oct. 17-21, 2012
  - This is 1 full week for 1 stock

Day 1 of model meeting



Note: “BR” indicates Butterworth and Rademeyer;
Source: document 1 and model input files for BR’s SCAA

Day 5 of model meeting



Point estimate (p.55 Full Assessment Report)

Approximate value, Fig. 1 in document 1

What happened between Day 1 and Day 5?

1. Identified differences in model configuration
2. Evaluated model diagnostics and statistical support for model configuration
3. Reached consensus on best model

1. Model configuration for ASAP and BR's SCAA

ASAP initial run

- Age 9 plus group
- Baranov catch equation
- NAA(y=1) freely estimated
- Start year 1982
- No stock-recruit function
- Additional index variance input
- Indices in Number
- Multinomial fit to age composition data
- Domed fishery selectivity
- Flat NEFSC indices
- Plus group selectivity applies to all ages in plus group

BR initial base cases

- Age 9 plus group (11+ calc.)
- Pope's Approximation
- NAA(y=1) in equilibrium, 2 parameters estimated
- Start year 1964
- Ricker (Case A) stock-recruit fn.
- Additional Index variance estimated
- Indices in Biomass
- 'Adjusted' lognormal fit to age composition data
- Domed fishery selectivity
- Domed NEFSC indices
- Plus group selectivity allowed to continue decreasing

1. Model configuration for ASAP and BR's SCAA

ASAP initial run

- Age 9 plus group
- Baranov catch equation
- NAA(y=1) freely estimated
- Start year 1982
- No stock-recruit function
- Additional index variance input
- Multinomial fit to age composition data
- Domed fishery selectivity
- Flat NEFSC indices
- Plus group selectivity applies to all ages in plus group

BR initial base cases

- Age 9 plus group (11+ calc.)
- Pope's Approximation
- NAA(y=1) in equilibrium, 2 parameters estimated
- Start year 1964
- Ricker (Case A) stock-recruit fn.
- Additional Index variance estimated
- Indices in Biomass
- 'Adjusted' lognormal fit to age composition data
- Domed fishery selectivity
- Domed NEFSC indices
- Plus group selectivity allowed to continue decreasing

1. Identify differences in model configuration between ASAP and BR's SCAA

ASAP initial run

A. Baranov catch equation
B. NAA(y=1) freely estimated
C. i) Start year 1982
   ii) No stock-recruit function
D. Additional index variance input
E. Indices in Number
F. Multinomial fit to age composition data
G. Flat NEFSC indices
H. Plus group selectivity applies to all ages in plus group

BR initial base cases

A. Pope's approximation
B. NAA(y=1) in equilibrium, 2 parameters estimated
C. i) Start year 1964
   ii) Ricker (Case A) or Bev-Holt (Case B) stock-recruit fn.
D. Additional Index variance estimated
E. Indices in Biomass
F. 'Adjusted' lognormal fit to age composition data
G. Domed NEFSC indices
H. Plus group selectivity allowed to continue decreasing

1. Identify differences in model configuration between ASAP and BR's SCAA

ASAP initial run	BR initial base cases
A. Baranov catch equation	A. Pope's approximation
B. NAA(y=1) freely estimated	B. NAA(y=1) in equilibrium, 2 parameters estimated
C. i) Start year 1982 ii) No stock-recruit function	C. i) Start year 1964 ii) Ricker (Case A) or Bev-Holt (Case B) stock-recruit fn.
D. Additional index variance input	D. Additional Index variance estimated
E. Multinomial fit to age composition data	E. 'Adjusted' lognormal fit to age composition data
F. Flat NEFSC indices	F. Domed NEFSC indices
G. Plus group selectivity applies to all ages in plus group	G. Plus group selectivity allowed to continue decreasing

Not investigated during WG meeting

2. Evaluate model diagnostics and statistical support for model configuration

B. Estimation of starting numbers at age (NAA)

- ASAP: NAA in first model year are freely estimated
- BR: estimates 2 parameters as in B.1.4 (p27) of document 3 (Θ, ϕ):

$$B_{y_0}^{\text{sp}} = \theta \cdot K^{\text{sp}} \quad \text{(B10)}$$

with the starting age structure:

$$N_{y_0,a} = R_{\text{start}} N_{\text{start},a} \quad \text{for } 1 \le a \le m \quad \text{(B11)}$$

where

$$N_{\text{start},1} = 1 \quad \text{(B12)}$$

$$N_{\text{start},a} = N_{\text{start},a-1} e^{-M a-1} (1 - \phi S_{a-1}) \quad \text{for } 2 \le a \le m-1 \quad \text{(B13)}$$

$$N_{\text{start},m} = N_{\text{start},m-1} e^{-M_m - 1} (1 - \phi S_{m-1}) \Big/ (1 - e^{-M_m} (1 - \phi S_m)) \quad \text{(B14)}$$

where ϕ characterises the average fishing proportion over the years immediately preceding y_0.

2. Evaluate model diagnostics and statistical support for model configuration

B. Estimation of starting numbers at age (NAA)

- ASAP: NAA in first model year are freely estimated
- BR: estimates 2 parameters as in B.1.4 (p27) of document 3 (θ, ϕ):

θ controls amount of depletion (upper bound 0.95)

ϕ is average exploitation (lower bound 0.1)

2. Evaluate model diagnostics and statistical support for model configuration

B. Estimation of starting numbers at age (NAA)

- ASAP: NAA in first model year are freely estimated Model converged properly ✓
- BR: estimates 2 parameters as in B.1.4 (p27) of document 3 (Θ, ϕ):

Θ, ϕ not separately estimable; lack of model convergence; solutions on parameter boundary; highly sensitive to initial conditions; ✗

→ See Tables 2, 3 of document 3; This ties in with starting year of model (C)

2. Evaluate model diagnostics and statistical support for model configuration

C. i) Starting model for assessment year

- Candidates: 1982, 1970, 1964

2. Evaluate model diagnostics and statistical support for model configuration

C. i) Starting model for assessment year

- Candidates: 1982, 1970, 1964
- 1982: age composition in survey and fishery; recreational catch estimates available

  → ASAP and BR converge* ✓ ✓

*The 1982 start is the only Ricker model that converged (Table 2 in document 3) but it had a strong retrospective

2. Evaluate model diagnostics and statistical support for model configuration

C. i) Starting model for assessment year

- Candidates: 1982, 1970, 1964
- 1970: age composition in survey, no age composition in fishery; no recreational catch estimates, no commercial discard estimates; no catch weights

→ ASAP converges, results from 1982-2010 similar to 1982 model; had to make assumptions about catch and catch weights, selectivity ✓

→ BR did not converge for Ricker model; ✗

BR Beverton-Holt model hit boundary with domed survey selectivity; did not get Hessian with flat survey selectivity

2. Evaluate model diagnostics and statistical support for model configuration

C. i) Starting model for assessment year

- Candidates: 1982, 1970, 1964
- 1964: no age composition in survey or fishery; no recreational catch estimates, no commercial discard estimates; no catch weights;

→ ASAP converges, results from 1982-2010 similar to 1982 model, precision of NAA(y=1) and F(y=1) are terrible

→ BR with flat survey selectivity and Ricker did not converge; dome survey selectivity and Ricker hit boundary;

flat survey selectivity and Beverton-Holt did not get Hessian; dome survey selectivity and Beverton-Holt imprecise NAA(y=1)

2. Evaluate model diagnostics and statistical support for model configuration

- Given B and C. i), the best model configuration is to start in 1982
  - Models properly converge
  - No missing information in catch or weights, full age composition data for survey and fishery
  - Both methods to compute NAA(y=1982) appear to have acceptable diagnostics

2. Evaluate model diagnostics and statistical support for model configuration

C. ii) Estimation of stock recruit function

- ASAP: no stock recruit function estimated; data do not support estimation;
- BR: estimated Ricker or Beverton-Holt with h=0.98 (~similar to ASAP)

→ Ricker had strong retrospective pattern (Table 4a, Figure 5a in document 3); Ricker models only 'supported' by earlier model start (*but* those BR models did not converge; furthermore, we reached consensus that 1982 start was best approach)

Therefore, the Ricker model is not supported by consensus decision or model diagnostics

2. Evaluate model diagnostics and statistical support for model configuration

D. Treatment of additional index variance

- ASAP: initial runs used estimated index-specific CVs; additional CV factor added to each index to achieve acceptable fit (SAW 53 Assessment report, page 51)
- BR: amount of additional variance estimated within model

	NEFSC spring	NEFSC fall	MADMF spring	
ASAP	0.2	0.1	0.3	**CV factor**
BR	0.24	0.14	0.12	**Extra σ**

2. Evaluate model diagnostics and statistical support for model configuration

D. Treatment of additional index variance

- Both models appear to provide acceptable fits to the indices (whether in number or biomass)
- Consensus that ASAP formulation acceptable

	NEFSC spring	NEFSC fall	MADMF spring	
ASAP	0.2	0.1	0.3	**CV factor**
BR	0.24	0.14	0.12	**Extra σ**

2. Evaluate model diagnostics and statistical support for model configuration

E. Indices in number versus biomass

- ASAP: fits to indices in number
- BR: fits to indices in biomass
- For both models, using consensus formulation (1982 start year), fits to indices appear reasonable

2. Evaluate model diagnostics and statistical support for model configuration

F. Likelihood for age composition data

- ASAP: multinomial
- BR: ‘adjusted’ lognormal
- Detailed diagnostics examined for ASAP, results acceptable
- Strong patterning in age comp residuals for BR fit to MADMF Spring

2. Evaluate model diagnostics and statistical support for model configuration

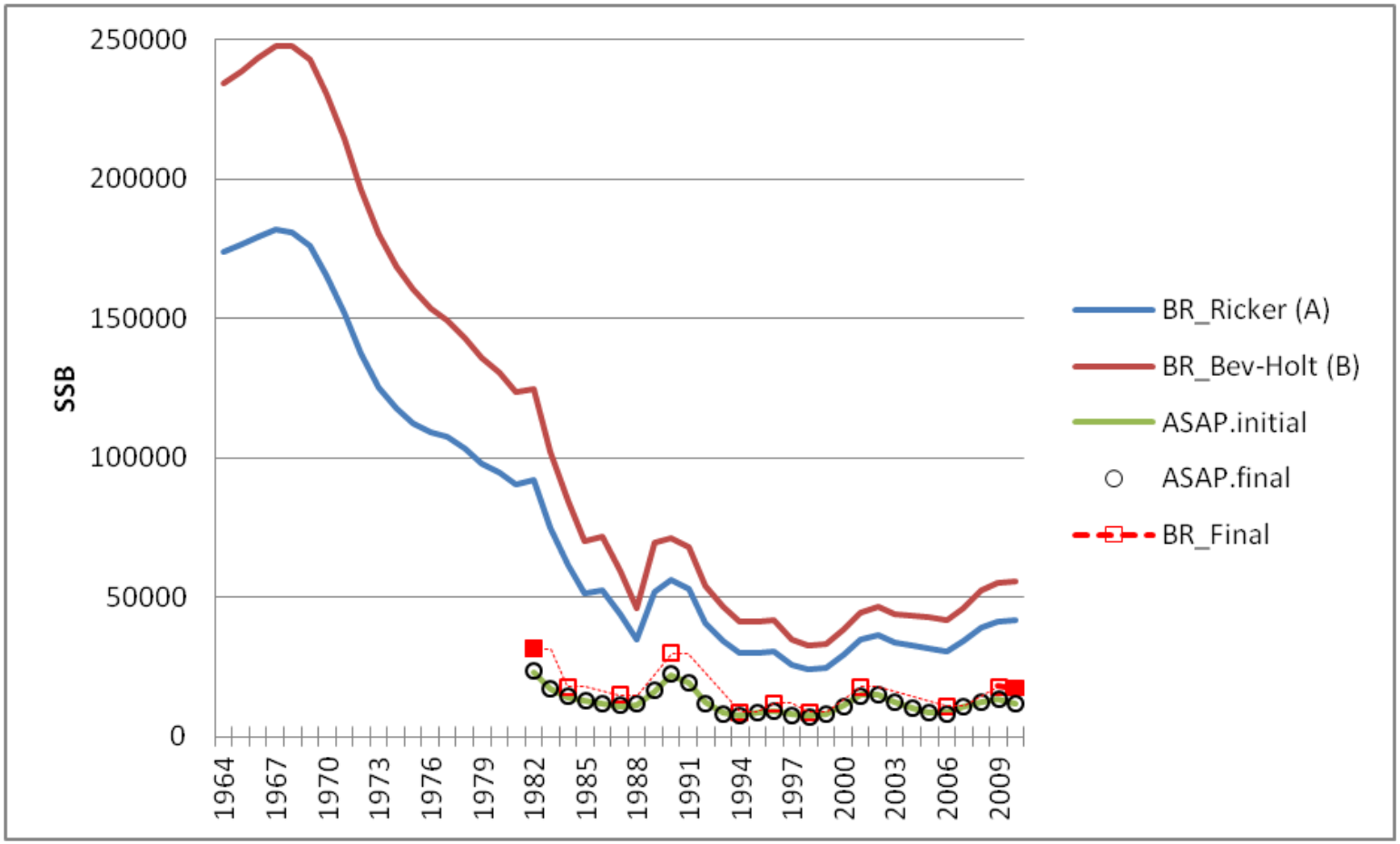
G. Dome versus Flat-topped NEFSC survey selectivity

- Dome sensitivity run explored for ASAP (pp52-53 of Assessment Report)
- 6 additional parameters estimated, only 3 point improvement in objective function
- Raw age proportion data for fully selected ages (6-9+) shows surveys catch more older fish than the fishery

3. Consensus on best model

- ASAP will be the assessment model
- Start year 1982, no stock recruit relationship
- Indices fit well, additional CV appropriate, no difference in ASAP whether indices fit in number or biomass—stick with numbers
- Flat NEFSC survey selectivity
- Agreed to drop MADMF fall and LPUE series

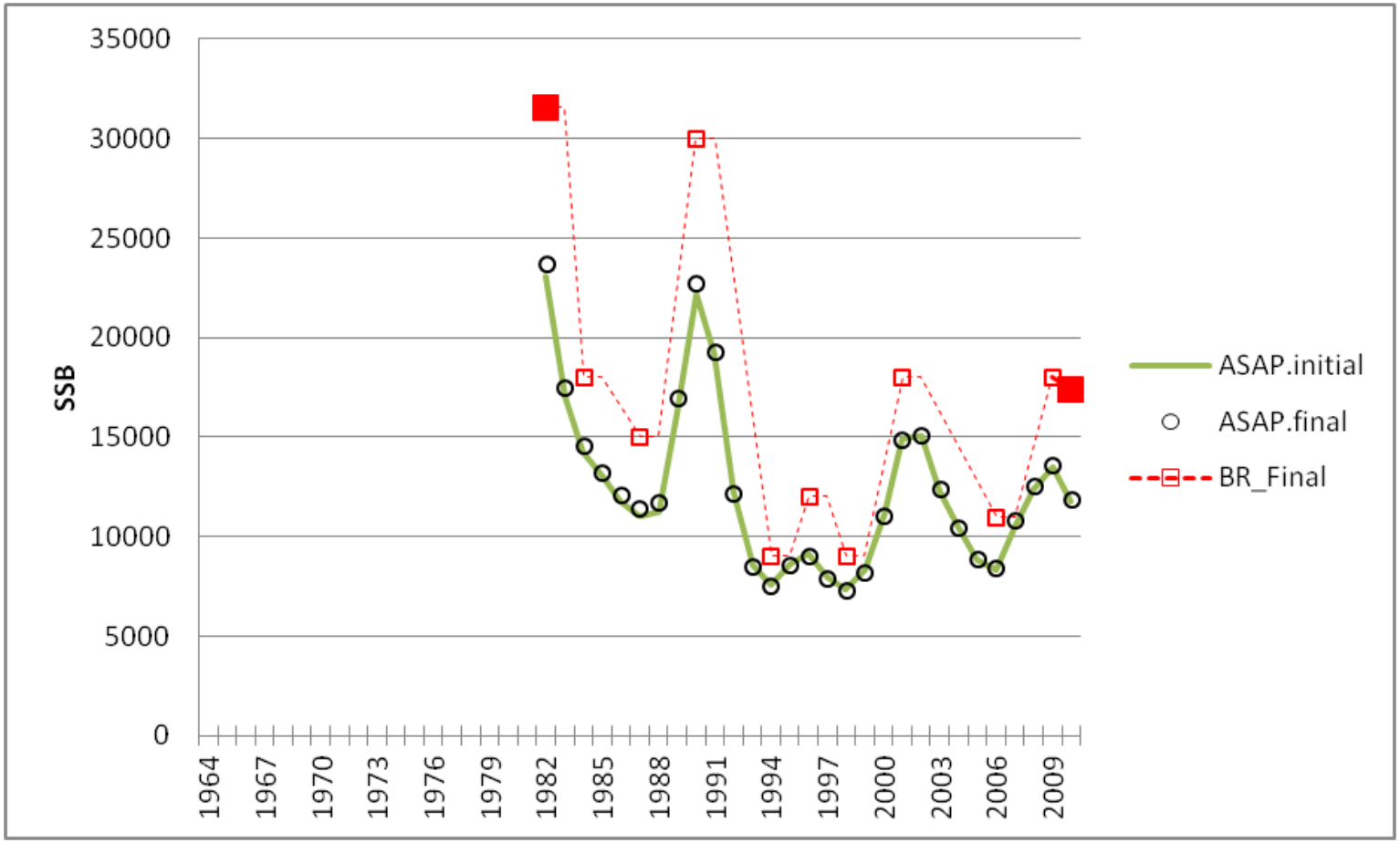
Day 5 of model meeting



Point estimate (p.55 Full Assessment Report)

Approximate value, Fig. 1 in document 1

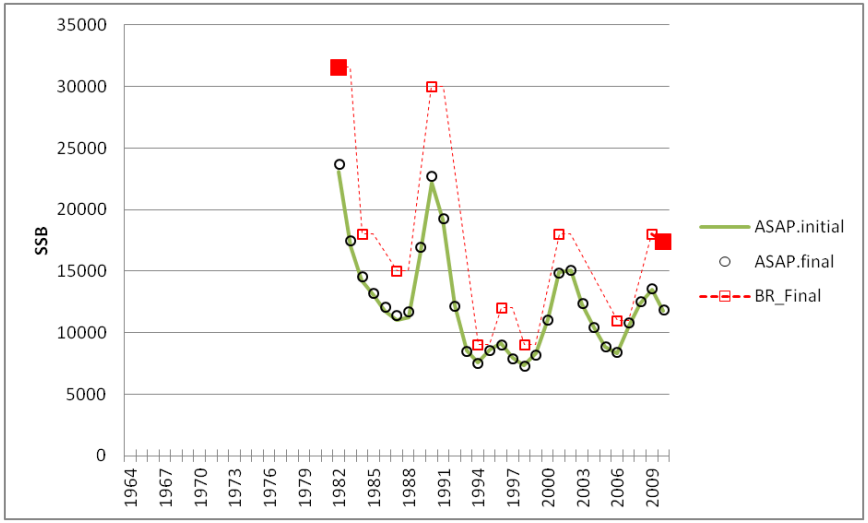
Day 5 models (enlarged view)



Point estimate (p.55 Full Assessment Report)

Approximate value, Fig. 1 in document 1

Remaining differences



Point estimate (p.55 Full Assmt. Report)

Approximate value, Fig. 1 in document 1

Both models were configured with the consensus decisions, but differences remained in the estimated trajectories. The WG noted (p.55, SAW 53 report) that most of the differences are believed to be due to:

1. Pope vs Baranov (Baranov preferred with high F values; Pope F estimates biased low)
2. B&R calculated SSB after 3/12 of the year while ASAP calculated SSB after 4/12 of year
3. Differences in model estimated selectivity, arising from likelihood form for age comp.

"Of these three items, the only one that would require further research is the form of the likelihood." --p.55, *SAW 53 report*

Summary

- An intensive month prior to the WG model meeting, and an intense week during the model meeting allowed an exhaustive examination of model configurations, sensitivities, diagnostics
- Consensus was reached on the most appropriate model structure
- A list of technical differences between ASAP and BR were listed to attempt to explain the remaining differences between models when configured similarly

Summary (cont.)

- As stated in the Full Assmt report, only the likelihood form for age composition data warrants further consideration
- As with all SAW/SARCs, the final report was sent to reviewers 2 weeks prior to review
- After the WG's report was submitted, with the consensus model, Butterworth and Rademeyer performed additional model runs that were not reviewed by the WG and that ignored the consensus decisions by the WG.
- Butterworth and Rademeyer requested their new model runs, and report, be reviewed by the SARC panel
- The SARC panel declined the request to review the Butterworth and Rademeyer report (dated November 2011)

Summary (cont.)

- The WG is the appropriate place to review model configurations, diagnostics, and reach consensus
- The WG model meeting is focused exclusively on the details of individual runs. This intense scrutiny is necessary to ensure that models intended for management advice are technically sound, converge properly, and are appropriate given the data

- The November 2011 BR report, as well as the updated January 2012 BR report (doc. 1) ignores WG decisions; the new BR models have the same problems with convergence, boundary solutions, unacceptable precision, and additionally an errors in variables bias
- In summary, there is nothing in the BR report (document 1) that is statistically preferred and it should not be considered as a basis for management advice

Detailed Response to Doc. 1:

An Investigation of Differences Amongst SCAA and ASAP Assessment (including Reference Point) Estimates for Gulf of Maine Cod

Doug S. Butterworth and Rebecca A. Rademeyer

January 2012

A. Pope vs. Baranov

- This was not investigated at working group meeting but was done afterwards for BR's model and results included in Assessment Report
- Results are well known: at high fishing mortalities, Pope's approximation underestimates F (p55 of Assessment Report: "For the estimation of F, Baranov is preferred when fishing mortality rates are high.")
- Table below is on p55 of Assessment Report

Biomass	**ASAP (BASE)**		**SCAA Pope**		**SCAA Baranov**	
SSB_{1982} (mt)	23,675	(20,760 - 26,958)	31,549	(19,831 - 43,267)	30,294	(19,642 - 40,946)
SSB_{2010} (mt)	11,868	(9,479 - 16,301)	17,373	(13,713 - 21,033)	16,481	(11,695 - 21,267)
SSB_0 (mt)	171,417	(136,351 - 218,992)	214,258	(7,481 - 421,035)	188,342	(59,499 - 317,181)
SSB_{MSY} (mt)	54,247	(41,394 - 72,462)	68,118	(59,626 - 76,609)	65,943	(53,936 - 71,446)
MSY (mt)	10,691	(8,012 - 14,687)	10,250	(8,891 - 11,609)	10,107	(8,462 - 10,754)

**Note that ASAP reference points were not estimated internally within the model but estimated through long term projections described in TOR. Also, confidence intervals (CI) presented for ASAP are 90% CI, while the B&R's SCAA are 95% CI.*

B. Estimation of starting NAA

- BR's new approach: estimate some ages freely, then calculate the remainder as exponential decline given ϕ estimate
- Table 2a claims 'statistical support' for estimating ages 0-6 ("…the AIC criterion…requires an improvement of at least one log-likelihood point for each extra parameter estimated from the data", p3)

1982 N vector:	-lnL: overall		
N_0 estimated	37.6		
N_0-N_1 estimated	37.6	Δ=0	+1 par: No statistical support. **FULL STOP.**
N_0-N_2 estimated	36.1	Δ=1.5	+2 par: No statistical support. **FULL STOP.**
N_0-N_3 estimated	25.8	Δ=11.8	+3 par: Models above not supported; illogical result.
N_0-N_4 estimated	25.6	Δ=12	+4 par: Models above not supported; illogical result.
N_0-N_5 estimated	24.9	Δ=12.7	+5 par: Models above not supported; illogical result.
N_0-N_6 estimated	21.3	Δ=16.3	+6 par: Models above not supported; illogical result.
N_0-N_7 estimated	20.6	Δ=17	+6 par: Models above not supported; illogical result.
N_0-N_8 estimated	20.6		
N_0-N_9 estimated	20.6		
N_0-N_{10} estimated	20.6		
N_0-N_{11} estimated	20.5		

B. Estimation of starting NAA

- Table 2b referenced to support number of ages estimable from Table 2a analysis (p15)
- Metric for decision if estimability is acceptable CV: ~0.5 (?)
- *Remember this CV level when examining other tables*

	Start in 1964		Start in 1965		Start in 1967		Start in 1970		Start in 1982	
	$N_{1964,0}$	CV	$N_{1965,0}$	CV	$N_{1967,0}$	CV	$N_{1970,0}$	CV	$N_{1982,0}$	CV
0	9137900	0.181	4556500	0.212	3988100	0.191	5681000	0.141	15792000	0.063
1	12832000	0.174	7455300	0.180	2085600	0.251	5889900	0.120	11476000	0.069
2 x	2276800	0.588	10414000	0.164	3031600	0.212	1754500	0.207	13012000	0.063
3	149*	21.262	1741600	0.563	4567300	0.176	1916900	0.189	5472000	0.094
4	149*	20.306	149*	12.764	5617300	0.155	823580	0.246	3269700	0.112
5	150	28.725	149*	14.905	808090	0.576	865010	0.207	1815300	0.139
6	150	34.019	1326	92.544	149*	7.985	1092000	0.180	182000	0.505
7	2659200	1.153	536	129.176	151	31.892	1275300	0.173	272340	0.494
8	149*	21.738	3113600	0.253	4396	41.655	192220	0.591	227910	1.054
9	150	28.269	149*	22.603	323	117.899	149*	3.629	157450	60.937
10	149*	19.490	919	124.098	1605600	0.261	149*	7.687	50871	42.793
11	1404600	1.660	6843	59.008	472	89.794	645440	0.240	114700	77.922

C. i) Selection of starting year

- Earlier start years were not acceptable at model meeting due to poor precision and lack of convergence
- Consider Table 1:
  - Models 1 is the "Day 5 BR's SCAA with Pope's approx."
  - Model 2 is "Day 5 BR's SCAA with Baranov"
  - Model 3 uses the new approach to estimate NAA, however the CV for ϕ is 1.28. Butterworth and Rademeyer have already established the unacceptable precision level of 0.5. The parameter ϕ is used to estimate ages 7-9+; <u>if the precision of ϕ is unacceptable, then the estimates of numbers at ages 7-9+ are unacceptable</u>.
  - Model 4 has 2 problems: i) ASAP base model <u>did use</u> additional variances for fitting indices; ii) the parameter ϕ hit the lower bound of 0.10; if ϕ hit a lower bound, then the model solution would prefer a smaller ϕ (less exploitation) and therefore <u>larger numbers at age</u>

C. i) Selection of starting year

- Earlier start years were not acceptable at model meeting due to poor precision and lack of convergence
- Consider Table 1:
  - Models 5-7: no apparent problem (but full diagnostics not reported); the CV on ϕ ranges from 0.3-0.4
  - Model 8: the parameter ϕ hit a lower bound of 0.1. This is referred to as the "New Base Case" but **it is untenable** due to lack of convergence (ϕ is used to estimate ages 3-9+)
  - Models 9-12: variants of Model 8 and have same problem: ϕ hit lower bound of 0.1. **These models are also untenable**.
  - Models 13-14: The CV on ϕ is 0.97-1.02. Unacceptable precision for estimating ages 3-9+; **untenable as well**. Furthermore, ASAP did not have this poor precision problem when fitting to indices in numbers.

C. i) Selection of starting year

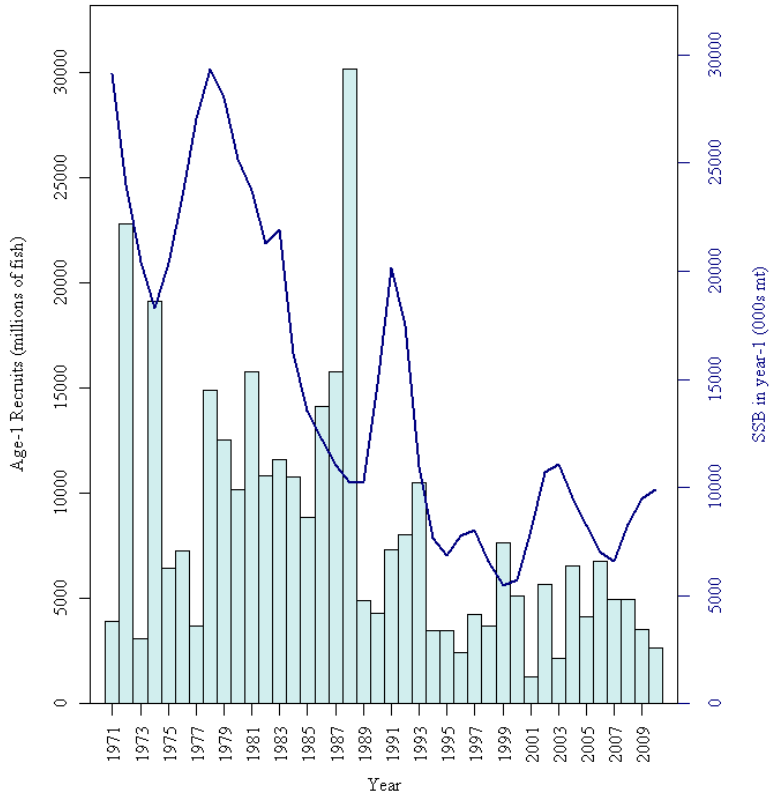
- Earlier start years were not acceptable at model meeting due to poor precision and lack of convergence
- Consider Table 5 (document 1):
  - Model 15: the CV on ϕ is 0.57, unacceptable based on Butterworth and Rademeyer criterion. **This model is also untenable**.
  - Model 16: the parameter ϕ hit a lower bound of 0.1 . **This model is also untenable**.
  - Models 17: the CV of 0.13 would indicate that ϕ did not hit a lower bound (CV is not 0.0), but ϕ appears to have a strong affinity for solutions of 0.1. **Is this models tenable?**
  - Models 18-21: The CV on ϕ ranges from 0.78-1.44. Unacceptable precision for estimating ages 3-9+. **These models are untenable as well**.

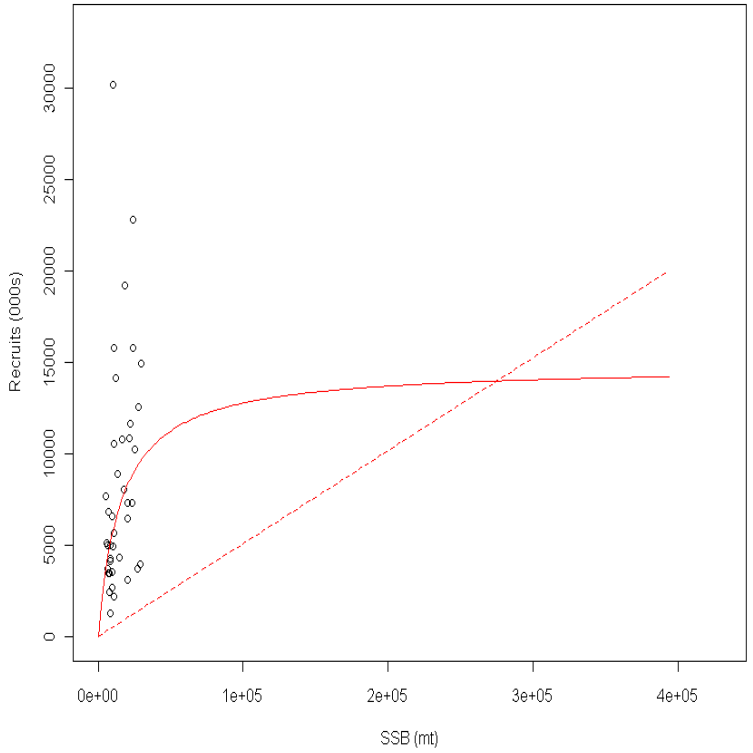
C. i) Selection of starting year

- P.4, Doc. 1: “The baseline ASAP assessment reported in NMFS (2011), however, extends back only to 1982, though reference point choices were based on inferences drawn from taking the assessment back to 1970. … **if recruitment estimates from 1970 onwards are deemed sufficiently reliable to inform reference point selection, those from the late 1960s must be as well.”**
- **They were not deemed sufficiently reliable by the SARC panel.**

C. i) Selection of starting year

- **They were not deemed sufficiently reliable by the SARC panel.**





C. i) Selection of starting year

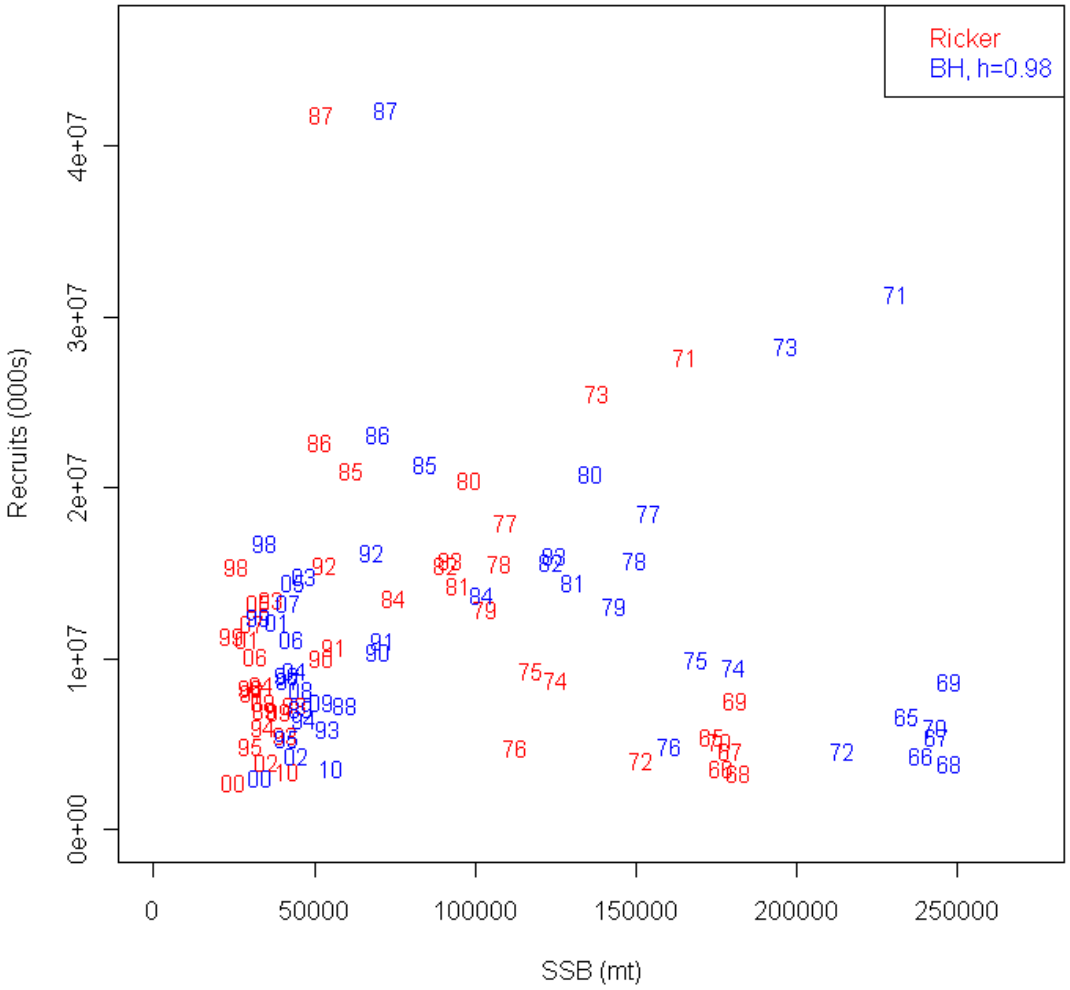
- P.5 “Table 3 and Fig. 2 contrast estimates for the 1970 numbers-at-age vector for two alternative assessments: case 5) commencing in 1970 and case 8) commencing in 1964.”
- The differences between cases 5 and 8 for ages 6-9+ are mostly negative. That bias is due to the boundary solution for ϕ in model 8 (estimate of ϕ=0.1). A smaller ϕ would have generated larger numbers at older ages. Consequently, inferred recruitments would also have been larger.
- Figure 3a: SSB trajectories for cases 5-7 (apparently no convergence problems) have lower SSB between 1964-1970; only model 8 (convergence problem) has higher SSB. ***
- The agreed upon model formulation used a plus group at 9, not 11.

*** important for later claims about support for form of stock recruit function.

C. ii) Estimation of stock-recruit function

- Estimation of a stock recruit function was not supported for the consensus model formulation
- Earlier attempts with BR model produced strong retrospective patterns when estimating Ricker function
- In Document 1, p.4: “Stock-recruitment …are now estimated externally to the assessment itself, rather than internally as in Butterworth and Rademeyer (2011), so that assumptions about the form of the relationship do not influence the assessment results quoted here.”
- This is the only way to avoid the retrospective problem with the Ricker model.

C. ii) Estimation of stock-recruit function



- Results for the Ricker and Beverton-Holt Cases A and B (doc. 3) clearly demonstrate that the form of the relationship *does* influence estimates of abundance.

C. ii) Estimation of stock-recruit function

- Estimating externally introduces an errors-in-variables problem because model estimates of SSB (the x-variable) are treated as being known without error. Consider Figure 4a (for model 8) as characterizing the magnitude of uncertainty in early SSB estimates.

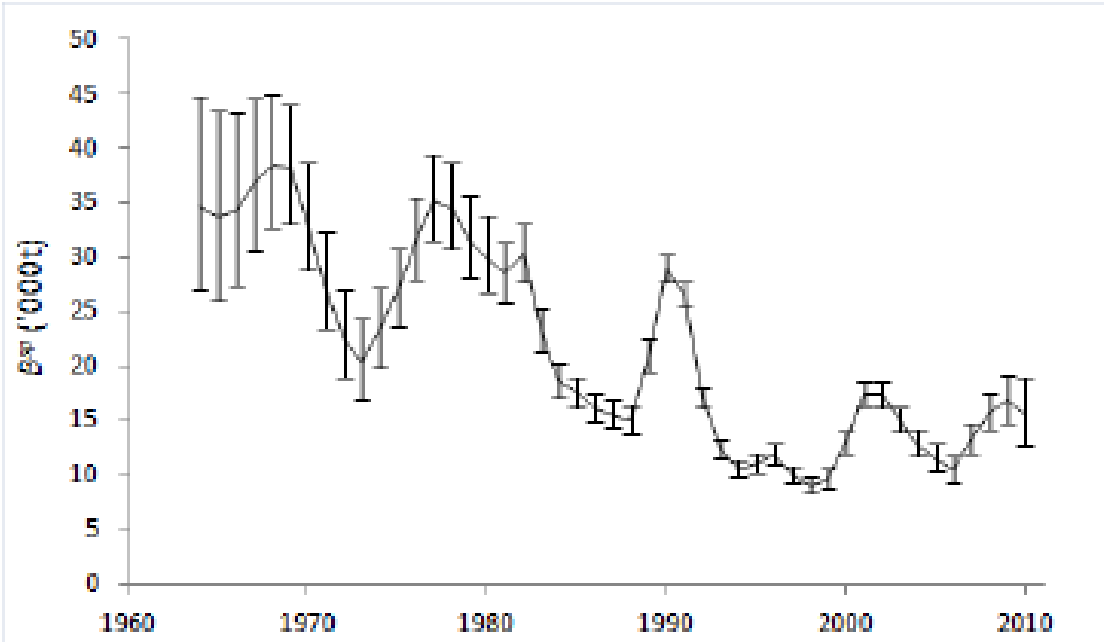
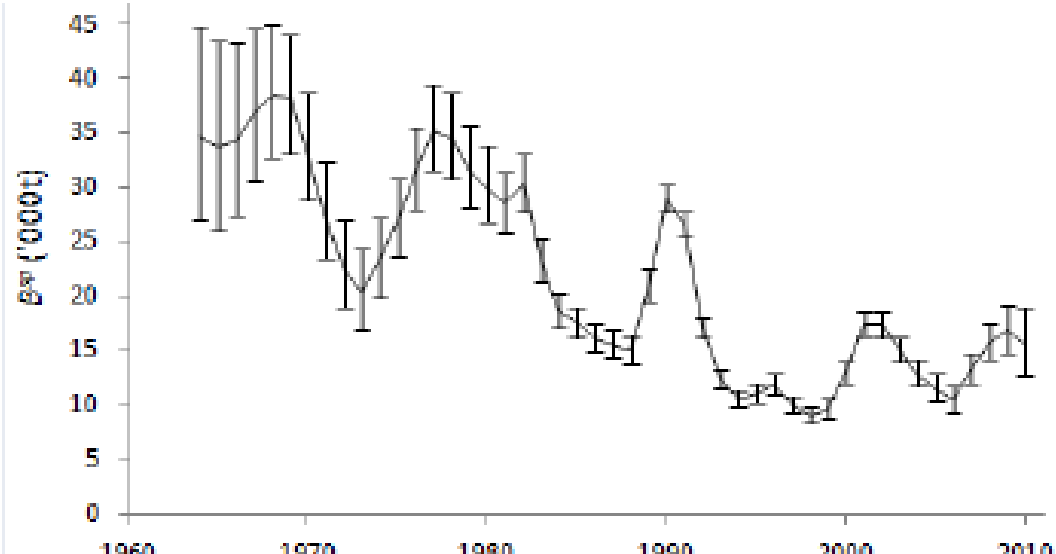


Fig. 4a: Spawning biomass trajectory for Case 8 (New Base Case), with Hessian-based 95% CIs, assuming lognormality.

C. ii) Estimation of stock-recruit function

- Estimating externally introduces an errors-in-variables problem because model estimates of SSB (the x-variable) are treated as being known without error. Consider Figure 4a (for model 8) as characterizing the magnitude of uncertainty in early SSB estimates.

Results for model 9 (where Ricker function was estimated within assessment model) are shown to infer lack of sensitivity to errors-in-variables problem. **However, model 9 did not converge, so this is a misleading comparison.**



C. ii) Estimation of stock-recruit function

- Beyond the errors-in-variables problem noted, the only models in Table 1 that did not have an obvious convergence/precision problem are models 5-7:

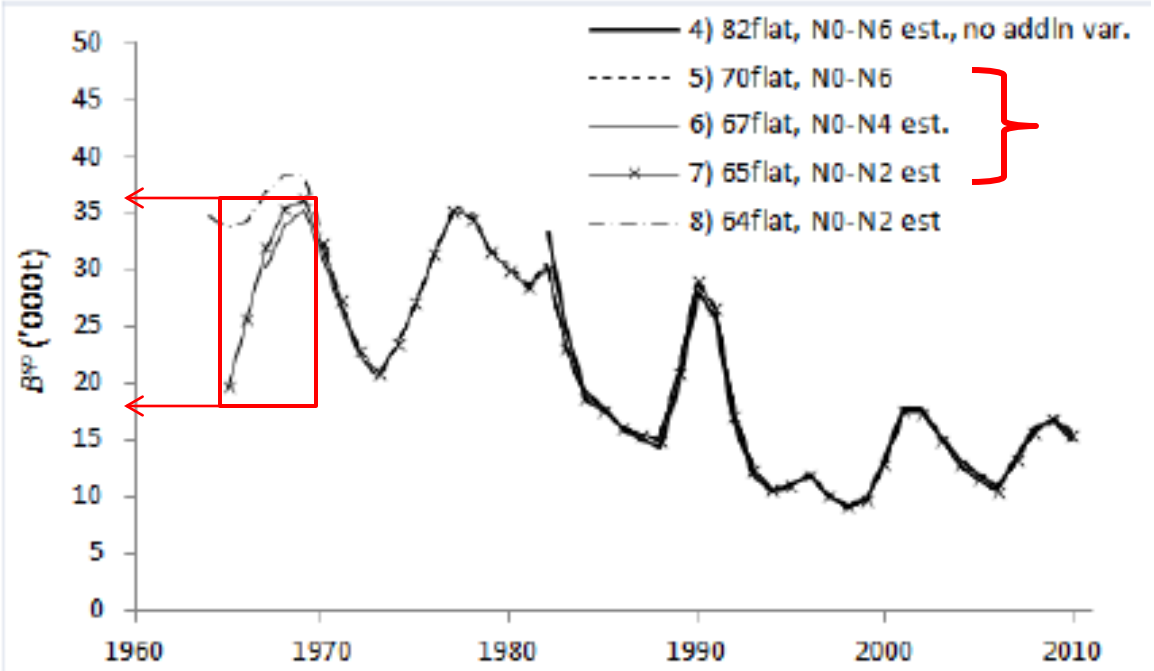
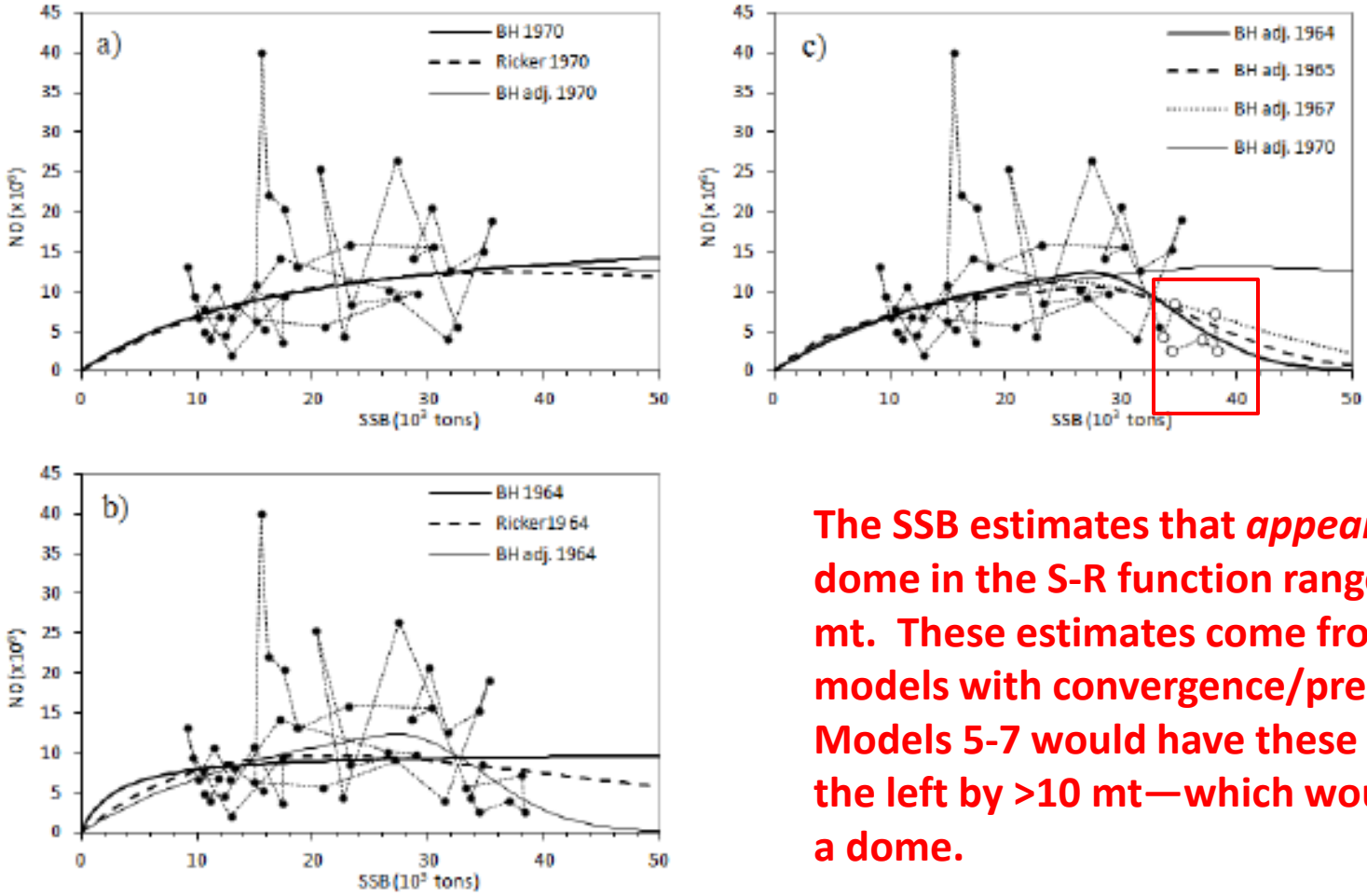


Fig. 3a: Spawning biomass trajectories cases 4 to 8.

Models 5-7 have SSB in the range of 18-35 mt for years pre-1970

C. ii) Estimation of stock-recruit function



The SSB estimates that *appear* to suggest a dome in the S-R function range from 30-40 mt. These estimates come from untenable models with convergence/precision problems. Models 5-7 would have these points shifted to the left by >10 mt—which would not support a dome.

Fig. 7: Fits to the stock-recruitment data for a) data from 1970, b) data from 1964 and c) Beverton-Holt adjusted curve for data from 1964, 1965, 1967 and 1970 (though the data shown in this plot is for the assessment starting in 1964).

C. ii) Estimation of stock-recruit function

- The stock recruit function cannot be estimated internally without retrospective problems.
- Estimation externally has errors-in-variables problem and lack of correspondence to model assumptions.
- The stock recruit functions are being fit to "data" (*output*) from models that did not converge. Models that did converge do not support a dome.
- Table 4 shows unacceptable precision for many of the estimated quantities for the adjusted Beverton-Holt model (CVs of 0.0, suggesting parameters at a boundary, or **CVs of 1.85-198.11!!**)
- The adjusted Beverton-Holt model has never been presented to the working group, has not been peer reviewed, and insufficient diagnostics are provided to evaluate it.

D. Additional variance when fitting to indices

- This point is wrong.
- P.51 of Full Assessment report: “Subsequent examination of the model fits to the survey indices resulted in adjustments to the survey CVs by adding the following constants to each of the survey CV vectors to account for additional process error: 0.2 (NEFSC spring), 0.1 (NEFSC fall), 0.3 (MADMF spring). It should be noted that these minor adjustments offered slight improvements to the statistical fit of the model but had little impact on the model results (e.g., see earlier models presented in Appendix 2 where survey CV vectors were not adjusted).”

	NEFSC spring	**NEFSC fall**	**MADMF spring**
ASAP	0.2	0.1	0.3
SCAA	0.24	0.14	0.12

CV factor
Extra σ

E. Fitting abundance indices in biomass or numbers

- Models 13 and 14 (Table 1) are BR models fitting to numbers rather than biomass and are analogous to models 3 and 8 (“New Base Case”)
- Model 3 is untenable because ϕ had a CV of 1.28
- Model 8 is untenable because ϕ hit the lower bound of 0.1
- As noted above, Models 13-14 are untenable because ϕ has a CV of 0.97 (13) or 1.02 (14)

E. Fitting abundance indices in biomass or numbers

- In the Summary for document 1 (p.1), it is claimed "A surprising result obtained in investigating such an ASAP surrogate is that fitting to abundance indices in terms of numbers rather than biomass radically changes perceptions for this particular scenario."
- The ASAP surrogate is not ASAP. Statements about an ASAP surrogate do not apply to any model but the surrogate. **These surrogate runs did not converge. ASAP did.**
- ASAP sensitivity runs that were fit to biomass did not result in a radical change in perception about stock status or SSB estimates.

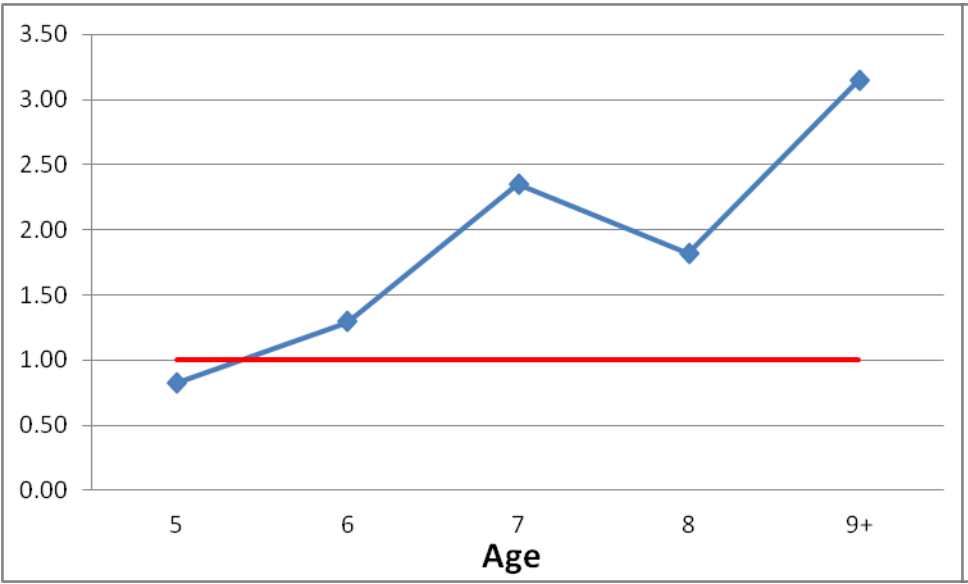
F. Form of likelihood for age composition data

- This point was identified by the working group as the only structural difference between ASAP and SCAA that would warrant further consideration.
- ASAP assumes a multinomial distribution
- BR assume an ‘adjusted lognormal’
- Document 1 is not an informative analysis on this point, because comparisons to the multinomial were not made.

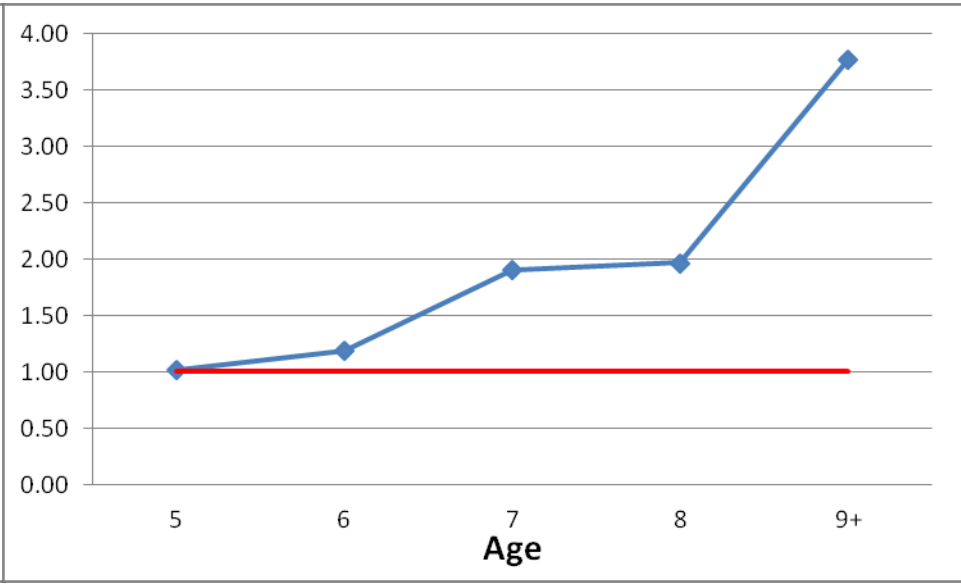
G. Domed vs flat selectivity for NEFSC surveys

- It is claimed that a domed NEFSC survey selectivity is supported based on point (F). However, point F is irrelevant as it did not compare the likelihood assumed in ASAP.
- Consideration of the raw age proportions (surveys and catch) outside of the model clearly demonstrates that for ages ≥6 (fully selected) **the survey catches more older fish than the fishery**

SPRING proportions / CATCH proportions



FALL proportions / CATCH proportions



G. Domed vs flat selectivity for NEFSC surveys

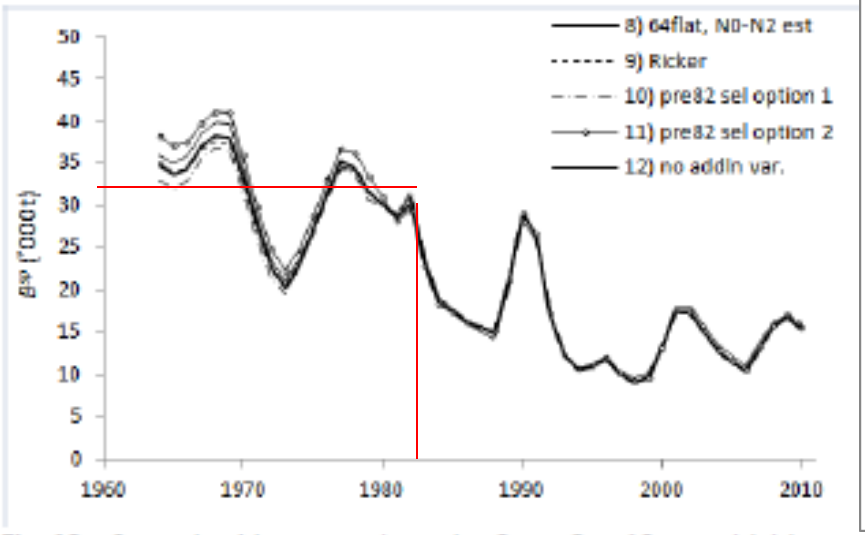
- A sensitivity run assuming a dome selectivity for NEFSC surveys was conducted and reviewed in the working group.
- This is also documented in the Assessment Report (pp.52-53)
- Assuming a dome meant estimating 6 additional parameters; the likelihood only improved by 3 points; therefore, there is no statistical support for domed survey selectivity

Summary

- The working group meeting is the appropriate place for models to be evaluated and consensus reached.
- Document 1 by Butterworth and Rademeyer ignores the WG consensus decisions, and reverts back to “Day 1” model configurations that are untenable.
- There is no statistical support for the claims made by Butterworth and Rademeyer (document 1).
- The SSC should uphold the SAW/SARC process as the appropriate route to developing a consensus base model for management advice.
- Ample time was available before and during the model meeting to explore model structure and sensitivity.

Take home point #1

- Despite the numerous models presented in document 1, the SSB trajectories for the period 1982-2010 are ***very*** similar to the "Day 5 model" trajectories and agree rather closely with ASAP results



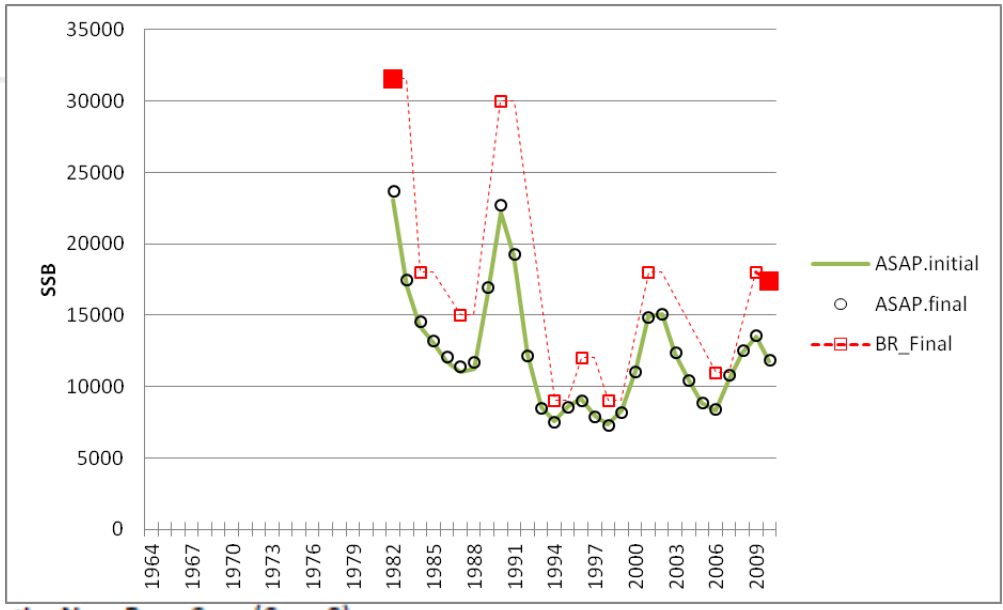


Fig. 10a: Spawning biomass trajectories Cases 8 to 12 - sensitivities on the New Base Case (Case 8).

Take home point #2

- Only models 5-7 (1970, 1967, 1965 start year) appear to converge for B&R. These models do not support a dome.
- **None** of the other models converged or had acceptable precision, therefore there are no model runs that support estimation of a Ricker or adjusted Beverton-Holt stock recruit model

THE END