Progress in Investigating Differences Amongst SCAA and ASAP Assessment (including Reference Point) Estimates for Gulf of Maine Cod

Doug S. Butterworth and Rebecca A. Rademeyer

November 2011

Summary

Various differences between SCAA and ASAP assessments of the Gulf of Maine cod, and their implications, are investigated further. Amongst the results are that neither all nor very few elements of the starting numbers-at-age vector for the assessment should be estimated, but rather an intermediate number which depends on the data available for years close to the starting year selected. This in turn leads to a demonstration that assessments commencing in 1970 incorporate reliable information about recruitment levels for the preceding 5 years, which therefore ought also to be taken into account in fitting stock-recruitment relationships and estimating MSY related reference points. Stock-recruitment relationships incorporating a downturn in recruitment at higher biomass levels are favoured in terms of the AIC statistical model selection criterion for analyses incorporating those earlier years, and suggest lower values for B_{MSY} as well as that the current status of the stock is not overfished. Investigations of the adjusted lognormal and multinomial distributions assumed for fitting proportions-at-age data for the SCAA and ASAP approaches respectively show that neither is appropriate, with each resulting in overweighting of data for younger compared to older ages, and hence to results of lesser precision than need be the case. Further consideration of the domed *vs* flat survey selectivity issue depends on appropriate modelling of the proportions-at-age distributions, and therefore needs to await further work on that topic. Estimation of additional variances in fitting to abundance indices is clearly justified on statistical grounds, while use of numbers rather than mass in fitting to abundance indices decreases the precision of estimates of current (2010) spawning biomass. For most of the SCAA model variants considered, this 2010 biomass is estimated in the 15 – 17 thousand ton range, though under the multinomial surrogate distribution for proportions-at-age data this drops to 14 thousand, while increasing to 20 thousand if numbers rather than mass are used in fitting to abundance indices (though reasons are offered for preferring the mass-based approach).

Introduction

As is always the case with complex assessments requiring many specifications, the Gulf of Maine (GoM) cod workshop held in October 2011 (NMFS, 2011), while making considerable progress, had to refer certain issues for further evaluation to better understand why different methods (ASAP and SCAA as then implemented) were giving different results. Furthermore, decisions had to be reached based on analyses that could be prepared before or during the meeting so that, again as always, such decisions might merit reconsideration based on further analyses whose completion during the meeting was not possible for reasons of time.

This document provides a progress report on addressing many of those issues, specifically:

A) Pope *vs* Baranov dynamics
B) Estimation of the starting numbers-at-age vector
C) The selection of the starting year for the assessment
D) Allowance for additional variance in fitting to the time series of abundance indices
E) The form of the term for catch-at-age proportions in the log-likelihood
F) Fitting abundances indices expressed in terms of mass or of numbers
G) Domed *vs* flat selectivity-at-age for the NEFSC surveys.

The starting point for the analyses reported in respect of data selections and certain assumptions reflects consensus reached to facilitate comparisons during the October workshop (NMFS, 2011), so that there are some differences from the specifications of the SCAA assessments reported in Butterworth and Rademeyer (2011), namely:

- Minor corrections have been effected to the tables for mean weights-at-age
- Spawning is taken to occur three rather than two months into the year
- The Massachusetts autumn and LCPUE abundance indices are omitted when fitting the assessment model to data
- The stock-recruit residuals penalty term is omitted from the objective function used when fitting the assessment model to data, except for the 2009 and 2010 recruitments to stabilise their estimates (this does not affect the estimates of spawning biomass reported below)
- Selectivities-at-age in the NEFSC surveys are fixed flat for ages 6+
- Although the underlying population model takes ages to 11+, when fitting no distinction is made for ages 9 and above which are grouped as 9+, both as regards data and assessment model assumptions (e.g. with respect to selectivities-at-age)
- Increases of pre-1982 catches by 25% to allow for levels of discards suggested by more recent analyses.

Approaches and Results

The starting point for these analyses is the Pope dynamics based SCAA assessment commencing in 1982, the results of which are reported in NMFS (2011). This is shown as case 1) in Table 1.

A) Pope vs Baranov dynamics

For existing assessments, ASAP has used Baranov and SCAA used Pope dynamics. While results for the two approaches will not differ greatly in most circumstances, differences can become important if fishing mortalities are high as may occur for GoM cod. Results replacing Pope by Baranov dynamics for case 1) are shown as case 2) in Table 1, with the spawning biomass trajectories compared in Fig. 1.

There is little difference between the results, with the Baranov form yielding a slightly lower estimate for current spawning biomass. As the Baranov form does not give rise to possible problems at high fishing mortality, and does not add unduly to the computational burden in this case, it has been retained for the further investigations below.

B) Estimation of the starting numbers-at-age vector

ASAP obtains separate estimates to each element of the numbers-at-age vector for the starting year of the assessment, whereas the SCAA assessments of Butterworth and Rademeyer (2011) reduce the number of estimable parameters to two, the spawning biomass at that time expressed as a proportion (θ) of the pristine level (K^{sp}), and a parameter ϕ which mimics recent average fishing mortality - see equations B10 to B14 of Butterworth and Rademeyer (2011).

This is a model selection question: how many estimable parameters will the available data support? Table 2a reports the negative log likelihood value commencing from the original SCAA formulation for the case 2) assessment (start year 1982) (indicated by "N_0 estimated"), and successively estimating additional elements of the stating vector, leaving only the remaining estimates linked through the estimable parameter ϕ. Under the AIC criterion, which requires an improvement of at least one log-likelihood point for each extra parameter estimated from the data, it is clear that there is justification for estimating some, though not all of the elements of the starting vector – in this case estimation to about age 6 is justified.

Table 2b reports results for a related analysis: estimating all elements of the starting vector and showing the associated Hessian-based CVs. For an assessment starting in 1982, it is evident that the CV starts to indicate unacceptably imprecise estimates at about the same age as the AIC approach of Table 2a suggests that independent estimation is no longer justified. Given this close relationship, this second approach (which is less onerous to perform) has been used to select the number of elements of the starting vector whose separate estimation can be justified for different starting years for the assessment. Thus for the cases shown in Table 2b, selections of age 6 for 1970, age 4 for 1967, and age 2 for 1965 and 1964 have been made. The reason this number drops for these earlier starting years is the absence of proportions-at-age data for the NEFSC surveys prior to 1970.
Thus the best approach is neither the ASAP "maximalist" nor the SCAA "minimalist", but rather an intermediate choice of the number of elements of the vector to be estimated, with the result depending on the data available for years close to the starting year. This approach has been followed for subsequent results reported in this document, for example case 3) which is a variant of case 2) of the assessment starting in 1982 but with ages up to 6 estimated separately for the numbers-at-age vector for 1982. Results in Table 1 and Fig. 1 show that this change has little impact on spawning biomass estimates.

C) The selection of the starting year for the assessment

The primary SCAA results of Butterworth and Rademeyer (2011) commenced assessments in 1964, co-incident with the first year for which survey data are available. This is in the general spirit of the SCAA approach which does not require values for (in particular) catch proportions-at-age every year, but instead makes use of assumptions about the selectivity-at-age vectors. A particular motivation for this is to be able to extend assessments further back in time to achieve better contrast and hence make allowance for better informed estimates of, for example, reference points related to MSY.

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. This last decision was because the results from the assessment starting in 1970 (unlike that starting in 1982) made clear that more recent recruitments corresponding to lower spawning biomasses tended to be lower than recruitments for the higher spawning biomasses of the 1970s. Hence any stock-recruitment relationship informing reference point selection would need to take account of a drop in expected recruitment as spawning biomass reduced.

Two reasons were advanced for preferring this approach to that of a 1964 start as preferred for the original SCAA analyses. The first was that proportions-at-age data were not available for commercial catches to inform commercial selectivities-at-age prior to 1982, or for surveys prior to 1970 to inform survey selectivities-at-age prior to 1970, so that to an increasing extent for assessments started further back in time, the SCAA estimates were more reliant on assumptions and less on data.

The decision to commence the assessment in 1970 rather than in 1964 to inform reference point selection has important consequences. The SCAA assessment starting in 1964 estimates the recruitments of the late 1960s, at relatively high spawning biomasses, to have been low, and these values are particularly influential in estimating the stock-recruitment relationship. The argument not to consider them because the absence of any age data prior to 1970 renders them uncertain, appears sound and reasonable at first sight.

However, it needs to be remembered that there is information about recruitment strength in the late 1960s from the proportions of older animals in the commercial catches and particularly the surveys of the early 1970s, and the newer SCAA procedure adopted here of estimating some of the elements of the numbers-at-age vector for the start year allows such information to be utilised. 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. What is immediately evident is that up to age 5, these two estimated vectors are effectively identical (and with very similar CVs). In turn this means that the assessment starting in 1970 already incorporates information sufficient to demonstrate (at a reasonable level of precision) that the recruitments of the late 1960s were low, so that **if recruitment estimates from 1970 onwards are deemed sufficiently reliable to inform reference point selection, those from the late 1960s must be as well.**

Results for SCAA assessments for a range of alternative starting years from 1970 to 1964 are reported as cases 5) to 8) in Table 1. These results are contrasted for recruitment and for spawning biomass in Fig. 3, with precision for a “New Base Case” starting in 1964 illustrated in Fig. 4, with its associated fit diagnostics shown in Fig. 5, and selectivity-at-age estimates in Fig. 6.

Stock-recruitment relationships have been fitted to the estimates of recruitment and spawning biomass provided by these various assessments to provide a basis to estimate reference points. Note that these 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 achieved by minimising the following negative log-likelihood:

$$-\ln L = \sum_{y=y1}^{2008} \left[\frac{\left(\ln(N_{y,0}) - \ln\left(\hat{N}_{y,0}\right)\right)^2}{2\left((\sigma_R)^2 + \left(CV_y\right)^2\right)} + \ln\left(\sqrt{(\sigma_R)^2 + \left(CV_y\right)^2} \right) \right] \tag{1}$$

where

$N_{y,0}$ is the "observed" (assessment estimated) recruitment in year *y*,

$\hat{N}_{y,0}$ is the stock-recruitment model predicted recruitment in year *y*,

σ_R is the standard deviation of the log-residuals, and

CV_y is the Hessian-based CV for the "observed" recruitment in year *y*.

Note that the differential precision of the assessment estimates of recruitment, which is lower for earlier years (e.g. see Fig. 4b), is taken into account, and that the summation ends at 2008 because little by way of direct observation is as yet available to inform estimates of recruitment for 2009 and 2010.

Some parameters of the various stock-recruitment curves fitted are reported in Table 4, with the fits of some of these to the "data" from the assessments shown in Fig. 7. In addition to the familiar Beverton-Holt and Ricker forms for a stock-recruitment relationship, results are also shown for an "Beverton-Holt adjusted" relationship:

$$\hat{N}_{y,0} = \begin{cases} \dfrac{\alpha B_y^{sp}}{\beta + B_y^{sp}} & \text{if } B_y^{sp} \le B^* \\ \dfrac{\alpha B^*}{\beta + B^*} \exp\left(-\left(\dfrac{B_y^{sp} - B^*}{\sigma_N} \right)^2 \right) & \text{if } B_y^{sp} > B^* \end{cases} \qquad (2)$$

where

α, β, $B^*$and σ_N are spawning biomass-recruitment relationship parameters, all estimated in the model fitting process of equation (1), and

B^{sp}_y is the spawning biomass from the assessment for year *y*.

The reason for including this last relationship is to have a form for which the shape at low spawning biomass as determined by data in that region of spawning biomass does not influence the shape at high biomass (distinct from what would occur for the Ricker relationship, for example). With this Beverton-Holt adjusted form, inferences about any decrease in recruitment at higher biomasses is determined entirely by the parameter σ_N , which in turn depends on the data at high biomass only.

Table 4 includes estimates for MSY related reference points based on these stock-recruitment relationships. Hessian-based CV's are given which are conditioned on the point estimate for F_{MSY} (though the CV associated with that will be low because *M* is assumed known and the commercial selectivity at age is precisely estimated). Where ratios of the 2010 spawning biomass to reference points are shown, the associated CVs assume independence of numerator and denominator. This will not be exactly true, and furthermore the estimates of precision take no account of the estimation errors associated with the assessment-based spawning biomass estimates that are input to calculations using equation (1). For this reason results for case 9) (which estimates the Ricker relationship internally within the assessment) have been added to Table 4, as those do take specific account of both those aspects. The results for CVs for case 9) do not differ greatly from those for the comparative assessment with external stock-recruitment relationship estimation, which suggests that any errors arising from the approximations/assumptions in the procedure adopted are not very large.

It is clear from the results in Table 4 and the plots of Fig. 7 that once recruitment estimates from the late 1960s are taken into account in the reference point computations, a very different picture emerges. There is clear statistical evidence supporting a downturn in recruitment at higher biomasses for starting years of 1967 and earlier (in fact this holds also for a start in 1968), with AIC favouring Ricker over Beverton-Holt, and Beverton-Holt adjusted over Ricker. Even for Beverton-Holt, starting the assessment in 1964 results in a pristine spawning biomass estimate reduced by nearly 50% from that estimated for the assessment starting in 1970 (see Fig. 8). Fig. 8 also shows how estimates of the current status of the resource depend on the start year for the assessment and the stock-recruitment relationship assumed. Note that unlike for the Beverton-Holt form, for assessments starting in 1967 or earlier, both the Ricker and the Beverton-Holt adjusted form indicate that the GoM cod stock is **NOT overfished**.

The key results from these investigations are therefore that some pre-1970 recruitments are well estimated from the available data despite the absence of proportions at age data prior to 1970; further when these are taken into account, statistical model selection approaches favour models showing a downturn in recruitment at higher biomasses, which has important implications for inferences concerning the current status of the resource.

Three reservations might be raised about these conclusions.

i) Assumptions have had to be made about the commercial (note effectively including recreational and discard) selectivity-at-age over the pre-1982 period, for which no associated proportions-at-age data are available. To address this sensitivities 12) and 13) reflecting considerable differences in this selectivity vector over this period were also run (see Fig. 9); the results are shown in Fig. 10, and indicate hardly any sensitivity of the resultant estimates of spawning biomass and recruitment to such differences. While other sensitivities of this nature could also be run (and a limited set of suggestions would be welcome), these results already suggest that it is hardly likely that they could result in qualitative changes to the conclusions above.

ii) Results in Table 3 and Fig. 3a suggest that estimates of recruitment and spawning biomass for 1964 are not as reliable as those for immediately following years. Nevertheless estimates of reference points and current stock status do not change meaningfully for assessments that start in 1965 compared to those that start in 1964 (see Table 4 and Fig. 8), so that this point has no real bearing on the reliability of the overall inferences.

iii) Comments have been made at GoM cod meetings/workshops that simulation studies have shown that estimates based on an assumed Ricker stock-recruitment function are biased, for example along the lines that assuming Ricker when Beverton-Holt holds will result in a negatively biased estimate of B_{MSY}. At the simplest level, one could respond that equally assuming a Beverton-Holt form when a Ricker applies will result in a positively biased estimate of B_{MSY} (precautionary considerations may be pertinent here, but they apply only in respect of decisions by management authorities, and should not be a consideration in selecting an assessment required to be based on the best available science.) There is a potential bias that arises with estimates for Ricker-like forms which is related to the simple argument that the highest spawning biomass observed can only have been so because it produced a recruitment below the expectation for that spawning biomass (and *vice versa* for the lowest spawning biomass), but that is not universally valid, as for example having the highest biomass occur in a particular year might rather be a consequence of a very large catch later that year. In any case in this instance of GoM cod, the inference about lower recruitment at the highest biomasses is not based on estimates for a single year, but at least four years in the late 1960s. Ultimately if arguments based on simulation studies are to be raised, those simulation studies need to be tabled so that first checks can be made as to whether they correspond sufficiently closely to the situation under consideration to bear any relevance. In any case, for reliable inference, simulation studies need to be conditioned on the situation at hand in a manner whose details are agreed by the scientists in debate on the issue before the studies are conducted, so that any final agreed inferences can be arrived at objectively and efficiently.

D) Allowance for additional variance in fitting to the time series of abundance indices

ASAP assumes additional variances associated with indices of abundance are zero, whereas SCAA estimates values separately for each of the three series of abundance indices used in fitting the GoM cod assessment model. Cases 4) and 10) (see Table 1) set additional variances to zero for SCAA assessments starting in 1982 and 1964 respectively. In both cases the negative log likelihood deteriorates by some 60 points with the loss of estimability of three parameters.

Statistical selection criteria thus overwhelmingly favour estimation of these parameters for the GoM cod assessment. The impacts on estimates of spawning biomass and recruitment are however not particularly large (see Table 1 and Fig. 10).

E) The form of the term for catch-at-age proportions in the log-likelihood

ASAP assumes a multinomial form for the distribution of proportion-at-age residuals in constructing the negative log likelihood to be minimised, whereas SCAA assumes an “adjusted” lognormal. The contribution of the proportions-at-age data (whether from catches or surveys) under the latter assumption is given by:

$$-\ell n L^{\mathrm{CAA}} = \sum_y \sum_a \left[\ell n \left(\sigma_{\mathrm{com}} / \sqrt{p_{y,a}} \right) + p_{y,a} \left(\ell n\, p_{y,a} - \ell n\, \hat{p}_{y,a} \right)^2 / 2 \left(\sigma_{\mathrm{com}} \right)^2 \right] \tag{3}$$

where

$p_{y,a}$ is the observed proportion of fish caught in year *y* that are of age *a*,

$\hat{p}_{y,a}$ is the model-predicted proportion of fish caught in year *y* that are of age *a*,

and

σ_{com} is the standard deviation associated with the catch-at-age data, which is estimated in the fitting procedure by:

$$\hat{\sigma}_{\mathrm{com}} = \sqrt{\sum_y \sum_a p_{y,a} \left(\ell n\, p_{y,a} - \ell n\, \hat{p}_{y,a} \right)^2 / \sum_y \sum_a 1} \tag{4}$$

where the summations exclude instances where the observed proportion is zero because the numerator contributions are structurally exactly equal to zero in such instances.

To compare this to the multinomial, it is convenient to approximate the latter in a manner that allows the same framework to be used. For the small proportions involved in this case, the multinomial approximates a Poisson (with variance equal, or at least proportional, to mean), and this in turn is well approximated, on taking square roots of the proportions concerned, by a normal distribution with constant variance.

For this multinomial-equivalent "sqrt(p)" form then:

$$-\ell n L^{\mathrm{CAA}} = \sum_y \sum_a \left[\ell n \left(\sigma_{\mathrm{com}} \right) + \left(\sqrt{p_{y,a}} - \sqrt{\hat{p}_{y,a}} \right)^2 / 2 \left(\sigma_{\mathrm{com}} \right)^2 \right] \tag{5}$$

and

$$\hat{\sigma}_{\mathrm{com}} = \sqrt{\sum_y \sum_a \left(\sqrt{p_{y,a}} - \sqrt{\hat{p}_{y,a}} \right)^2 / \sum_y \sum_a 1} \tag{6}$$

where these summations include all years and ages as no numerator contribution is structurally zero.

Thus variance is conveniently estimated within the ML process without the need for external specification of effective sample size.

For minimum variance estimates, residuals should be homoscedastic, so that a first comparative test of the appropriateness of the adjusted lognormal to the multivariate assumption for this case is provided by checking which better achieves such homoscedasticity. This in turn has been checked by using equations (4) and (6) **without** their summations over ages to provide residual variance estimates by age (effectively estimates of σ_{com} which should, for homoscedasticity, show no trend with age).

Results for this check are shown in Fig. 11. Contrary to the trendlessness sought, there are very clear downward trends for all three surveys and for the catch. Both distributional assumptions are near equally bad, and it is clear that both are effectively overweighting the data for smaller ages, and underweighting those for larger ages which for some reason are inherently less variable.

This though does **NOT** mean that the assessment results under either method are fatally flawed. Since there are no unequivocal signs of model misspecifications in the diagnostic plots of Fig. 5, for example, the resultant estimates are not necessarily biased. They are however reflecting greater variance than need be the case. Work is ongoing to develop an improved approach to achieve homoscedasticity in a parsimonious manner, and hence more precise results from the assessment.

Clearly though the assumption of a multivariate distribution for the proportions-at-age data for GoM cod seems wrong (and also the adjusted lognormal), and needs improvement.

F) Fitting abundance indices expressed in terms of mass or of numbers

ASAP routinely fits to abundance estimates expressed in terms of numbers, whereas SCAA instead uses mass. Cases 14) and 15) in Table 1 show results for the SCAA assessments starting in 1982 (case 3)) and 1964 (case 8)) respectively, with the corresponding estimated spawning biomass trajectories compared in Fig. 12.

This change to numbers results in slightly higher estimates of current spawning biomass approaching 20 thousand tons, though these are less precisely estimated than their counterparts based on mass. It should be noted that the move to numbers is associated with changes in estimates of additional variance. These lead to greater weight being accorded to the NEFSC surveys, and less to the Massachusetts survey, which is a possible reason for the change in the point estimate for current spawning biomass. The lesser precision is not unexpected given that the relative contributions of the different ages to the overall index are more skewed towards the younger ages for numbers. This together with the results from section E), which suggest greater variance associated with the younger ages, would seem to point towards a preference for the use of mass rather than numbers when fitting abundance indices for GoM cod assessment models, at least.

G) Domed vs flat selectivity-at-age for the NEFSC surveys

Settling the debate on whether the data favour domed over flat selectivity-at-age for the NEFSC surveys rests heavily on the use of model selection approaches to fits to the data in each case, and in particular appropriately structured log likelihoods. In turn this first requires resolution of the issue in section E) about how best to construct the likelihood for the proportions-at-age data. Thus further work on this point will first await first further progress on this last issue.

Conclusions at this Time

A striking feature of the results in Table 1 for the estimates of the current (2010) spawning biomass for GoM cod is their closeness over a wide range of “sensitivity” tests, with variation essentially between 15 and 17 thousand tons. Pope *vs* Baranov dynamics, the number of elements of the starting numbers-at-age vector, the selection of the starting year for the assessment, and taking additional variance in the time series of indices of abundance into account all make little difference. Two instances that lead to estimates outside this range are use of the multinomial surrogate (sqrt(p)) approach for the proportions-at-age contributions to the log likelihood (14 thousand tons), and the replacement of mass by numbers in fitting to abundance indices (20 thousand tons). The reason for the difference with the corresponding ASAP estimate (NMFS 2011) of about 12 thousand tons has thus not been resolved. However changes from mass to numbers in fitting the abundance indices also leads to changes to the additional variance estimates (and hence the relative weights) accorded to these indices, so that there are interactions amongst the effects of some of the ASAP-SCAA differences, and further explorations of different combinations of choices might still yield a case of closer agreement. With the exception of cases fitted to numbers rather than mass, the SCAA estimates of survey *q* in Table 1 (corresponding to the *Bigelow*) lie between 0.71 and 0.86, and consequently do not require any need to postulate gear herding effects.

The main result from this work to date is concerns MSY-related reference point estimation. If assessments starting in 1970 are acceptable for this purpose, so must be the (low) estimates of recruitment in the late 1960s already evidenced in that assessment. Those estimates in turn indicate much reduced estimates of pristine spawning biomass compared to a Beverton-Holt form fitted to the 1970+ spawning biomass and recruitment estimates only (as in NMFS 2011), and further provide statistical justification for preferring a dome-shaped relationship which decreases at higher biomasses. These in turn suggest that the status of the GoM cod stock is NOT overfished.

Both approaches currently in use to model the distributions of proportions-at-age residuals are sub-optimal, overweighting younger compared to older ages, and hence providing estimates with variances greater than need be the case. Work on an improved formulation is progressing, and further debate on the presence or otherwise of a dome in NEFSC surveys selectivities-at-age should first await this formulation.

Acknowledgements

We thank Michael Palmer for provision of the data upon which the analyses reported in this paper are based.

References

Butterworth DS and Rademeyer RA. 2011. Applications of statistical catch-at-age assessment methodology to Gulf of Maine cod. Document submitted to the 17-21 October, 2011 workshop on the assessment of Gulf of Maine cod, Falmouth. 31pp.

NMFS 2011. Stock Assessment Workshop (SAW 53). A. Gulf of Maine cod (*Gadus morhua*) stock assessment updated through 2010. GoM cod WP#1 (GoM cod Assessment) 11/14/11.

Table 1: Estimates of abundance and related quantities for the Gulf of Maine cod for a series of assessment sensitivities. Values in parentheses are Hessian based CV's. Mass units are '000 tons. y1 refers to the start year for the assessment. $N_{y1,0}$ is in millions. Refer to Appendix B of Butterworth and Rademeyer (2011) for definition of some of the symbols used. Note that the estimation procedure used bounds σ_{Add} above by 0.5.

	1)		2)		3)		4)		5)		6)		7)		8)		9)		10)		11)		12)		13)		14)		15)	
	Pope, start in 1982, flat survey sel., θ and ϕ estimated		As 1) Baranov		As 1) $N_{y1,0}$- $N_{y1,6}$ estimated		As 3) survey additional variances set to 0		As 3) start in 1970		As 3) start in 1967, $N_{y1,0}$- $N_{y1,4}$ estimated		As 3) start in 1965, $N_{y1,0}$- $N_{y1,2}$ estimated		As 3) start in 1964, $N_{y1,0}$- $N_{y1,2}$ estimated (New Base Case)		As New Base Case, Ricker SR fit within assessment		As New Base Case, survey additional variances set to 0		As New Base Case, sqrt(p) for CAA error		As New Base Case, with option 1 for pre-1982 commercial sel.		As New Base Case, with option 2 for pre-1982 commercial sel.		As 3) Fit to Numbers		As 8) Fit to Numbers	
-lnL: overall	41.9		37.6		21.3		79.3		-20.0		-27.0		-25.9		-30.6		5.6		31.7		-2349		-28.3		-24.4		62.3		23.2	
-lnL: survey	-16.4		-16.3		-16.8		32.2		-31.1		-34.2		-35.2		-36.4		-35.9		13.8		-35.7		-36.1		-36.5		25.6		19.9	
-lnL: comCAA	-79.3		-85.9		-96.4		-94.1		-93.4		-93.2		-93.2		-93.2		-92.9		-90.2		-685.5		-93.2		-92.1		-96.2		-91.9	
-lnL: survCAA	136.5		138.9		133.4		140.2		103.7		99.5		101.6		98.1		99.7		106.9		-1629		100.1		103.3		131.9		94.2	
-lnL: RecRes	1.1		1.0		1.0		1.1		0.9		0.9		1.0		0.9		34.7		1.2		1.4		0.8		0.9		0.9		0.9	
θ	0.21	(0.07)	0.16	(0.35)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	
$N_{y1,0}$	15.9	(0.04)	15.7	(0.04)	15.8	(0.06)	15.3	(0.06)	5.7	(0.14)	3.9	(0.19)	4.0	(0.21)	8.5	(0.18)	8.2	(0.17)	8.5	(0.17)	8.5	(0.20)	9.6	(0.17)	7.6	(0.17)	15.9	(0.06)	8.7	(0.18)
ϕ	0.49	(0.05)	0.54	(0.06)	0.13	(1.28)	0.10	(0.00)	0.55	(0.40)	0.52	(0.30)	0.55	(0.33)	0.10	(0.00)	0.10	(0.00)	0.10	(0.00)	0.15	(1.44)	0.10	(0.00)	0.10	(0.01)	0.17	(0.97)	0.21	(1.02)
B^{sp}_{2010}	17.4	(0.10)	16.5	(0.15)	16.4	(0.15)	15.1	(0.09)	15.7	(0.11)	15.5	(0.10)	15.5	(0.11)	15.5	(0.10)	16.2	(0.10)	15.3	(0.09)	13.8	(0.09)	15.7	(0.12)	15.6	(0.10)	19.4	(0.19)	19.9	(0.19)
B^{sp}_{1982}	31.5	(0.05)	30.3	(0.05)	31.1	(0.07)	33.2	(0.05)	30.4	(0.05)	30.3	(0.05)	30.1	(0.04)	30.3	(0.04)	30.7	(0.04)	31.1	(0.04)	26.7	(0.04)	31.7	(0.04)	29.5	(0.04)	30.4	(0.06)	30.4	(0.05)
B^{sp}_{y1}	31.5	(0.05)	30.3	(0.05)	31.1	(0.07)	33.2	(0.05)	33.5	(0.08)	30.2	(0.10)	19.7	(0.20)	34.7	(0.13)	35.1	(0.12)	35.9	(0.10)	33.5	(0.29)	32.9	(0.11)	38.2	(0.10)	30.4	(0.06)	25.0	(0.50)
	q	σ_{Add}	q	σ_{Add}	q	σ_{Add}	q	σ_{Add}	q	σ_{Add}	q	σ_{Add}	q	σ_{Add}	q	σ_{Add}	q	σ_{Add}	q	σ_{Add}	q	σ_{Add}	q	σ_{Add}	q	σ_{Add}	$qx10^3$	σ_{Add}	$qx10^3$	σ_{Add}
NEFSC spring	0.75	0.24	0.82	0.26	0.84	0.26	0.85	0.00	0.73	0.20	0.74	0.19	0.74	0.19	0.74	0.19	0.74	0.19	0.71	0.00	0.86	0.20	0.72	0.18	0.72	0.19	1.06	0.16	0.96	0.14
NEFSC fall	0.52	0.14	0.57	0.13	0.58	0.13	0.59	0.00	0.60	0.11	0.61	0.11	0.62	0.10	0.62	0.10	0.63	0.09	0.64	0.00	0.62	0.10	0.62	0.10	0.59	0.09	0.54	0.06	0.61	0.14
MADMF spring	0.15	0.12	0.15	0.12	0.15	0.12	0.15	0.00	0.15	0.10	0.15	0.10	0.15	0.10	0.15	0.10	0.15	0.10	0.15	0.00	0.13	0.10	0.15	0.10	0.15	0.10	0.46	0.50	0.46	0.50

Table 2a: Overall negative log-likelihood for different estimations of the starting numbers-at-age vector for a 1982 start year for the assessment.

1982 N vector:	-lnL: overall
N_0 estimated	37.6
N_0-N_1 estimated	37.6
N_0-N_2 estimated	36.1
N_0-N_3 estimated	25.8
N_0-N_4 estimated	25.6
N_0-N_5 estimated	24.9
N_0-N_6 estimated	21.3
N_0-N_7 estimated	20.6
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

Table 2b: Start year numbers-at-age vectors (all ages estimated) and Hessian-based CVs for assessments starting in different years.

	Start in 1964		Start in 1965		Start in 1967		Start in 1970		Start in 1982	
	$N_{1964,a}$	CV	$N_{1965,a}$	CV	$N_{1967,a}$	CV	$N_{1970,a}$	CV	$N_{1982,a}$	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	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

* lower boundary

Table 3: 1970 numbers-at-age vectors with Hessian-based CVs for Cases 5 (start in 1964) and 8 (start in 1970).

	Case 5) Start in 1970		Case 8) Start in 1964	
a	$N_{1970,a}$	CV	$N_{1970,a}$	CV
0	5603600	0.142	5638400	0.141
1	5766400	0.121	5837300	0.120
2	1707400	0.208	1739700	0.206
3	1879300	0.189	1905700	0.188
4	803200	0.246	817840	0.245
5	843430	0.207	855840	0.206
6	1433100	0.142	1080300	0.178
7	803410	0.123	1258700	0.170
8	433260	0.161	163010	0.227
9	190980	0.232	106890	0.291
10	136210	0.214	100100	0.237
11	338810	0.250	518790	0.181

Table 4: Estimates of reference points from fits to stock-recruitment data. Values in parentheses are Hessian based CV's (for σ_R these are typically of the order of 0.004). Mass units are '000 tons. Note that F refers to fishing mortality on age 5, and MSY is as calculated for the most recent commercial selectivity-at-age vector.

	Beverton-Holt								Ricker										Beverton-Holt adjusted							
	1964		1965		1967		1970		Case 9		1964		1965		1967		1970		1964		1965		1967		1970	
-lnL	7.2		8.2		4.8		0.5		34.7		5.2		7.7		4.3		0.5		1.9		7.4		1.6		0.5	
h	0.95	(0.06)	0.93	(0.06)	0.91	(0.06)	0.86	(0.06)	2.66	(0.17)	2.83	(0.16)	2.53	(0.17)	2.49	(0.17)	2.16	(0.18)	0.40	(0.37)	0.56	(0.32)	0.43	(0.44)	0.64	(61.70)
σ_R	0.66	(0.00)	0.68	(0.00)	0.64	(0.00)	0.60	(0.00)	0.61	(0.03)	0.65	(0.00)	0.67	(0.00)	0.64	(0.00)	0.60	(0.00)	0.61	(0.00)	0.67	(0.00)	0.64	(0.00)	0.60	(0.00)
K^{sp}	219.95	(0.19)	239.45	(0.23)	275.63	(0.26)	406.72	(0.39)	83.04	(0.17)	69.64	(0.12)	76.24	(0.16)	82.04	(0.19)	112.22	(0.32)	41.72	(0.14)	44.96	(0.14)	50.10	(0.15)	106.47	(198.11)
F_{MSY}	0.33		0.31		0.29		0.26		0.39		0.59		0.54		0.53		0.47		0.63		0.55		0.64		0.48	
$MSYL^{sp}$	0.22	(0.06)	0.23	(0.06)	0.24	(0.05)	0.26	(0.05)	0.33	(0.13)	0.33	(0.12)	0.34	(0.13)	0.34	(0.13)	0.36	(0.13)	0.65	(0.00)	0.61	(0.00)	0.48	(0.00)	0.38	(0.00)
B^{sp}_{MSY}	48.89	(0.14)	55.91	(0.18)	67.13	(0.21)	107.34	(0.34)	27.53	(0.08)	23.09	(0.07)	25.97	(0.08)	27.98	(0.10)	39.89	(0.21)	27.24	(0.14)	27.51	(0.14)	24.23	(0.15)	40.98	(198.11)
B^{sp}_{2010}	15.50	(0.10)	15.51	(0.11)	15.51	(0.10)	15.69	(0.11)	16.17	(0.10)	15.50	(0.10)	15.51	(0.11)	15.51	(0.10)	15.69	(0.11)	15.50	(0.10)	15.51	(0.11)	15.51	(0.10)	15.69	(0.11)
$B^{sp}_{2010}/B^{sp}_{MSY}$	0.32	(0.18)	0.28	(0.21)	0.23	(0.23)	0.15	(0.36)	0.59	(0.13)	0.67	(0.13)	0.60	(0.14)	0.55	(0.14)	0.39	(0.23)	0.57	(0.18)	0.56	(0.17)	0.64	(0.18)	0.38	(198.11)
B^{sp}_{2010}/K^{sp}	0.07	(0.22)	0.06	(0.25)	0.06	(0.28)	0.04	(0.41)	0.19	(0.21)	0.22	(0.16)	0.20	(0.19)	0.19	(0.21)	0.14	(0.34)	0.37	(0.18)	0.34	(0.17)	0.31	(0.18)	0.15	(198.11)

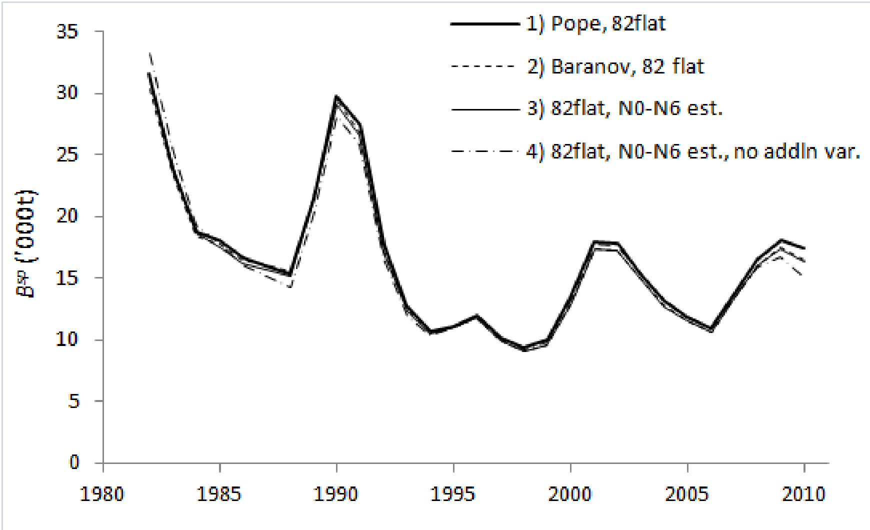

Fig. 1: Spawning biomass trajectories cases 1 to 4.

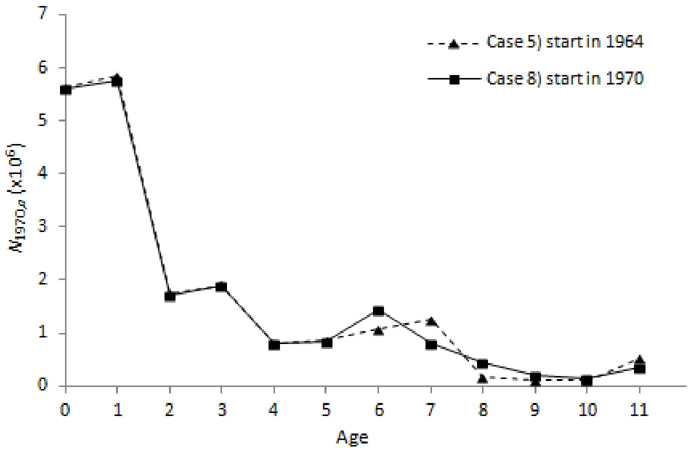

Fig. 2: Numbers-at-age vector for 1970 for Cases 5 (start in 1964) and 8 (start in 1970).

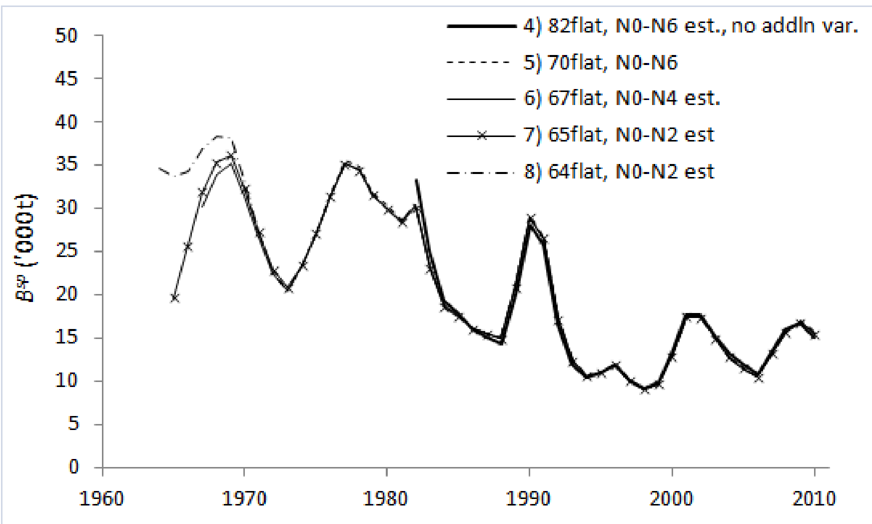

Fig. 3a: Spawning biomass trajectories cases 4 to 8.

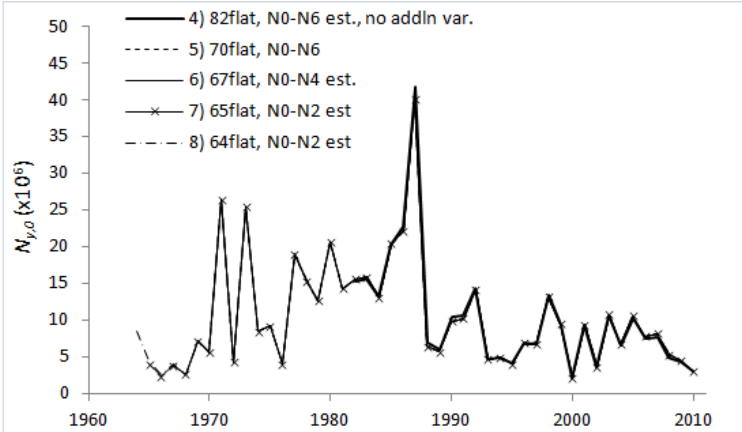

Fig. 3b: Trajectories of recruitment ($N_{y,0}$) for Cases 4 to 8.

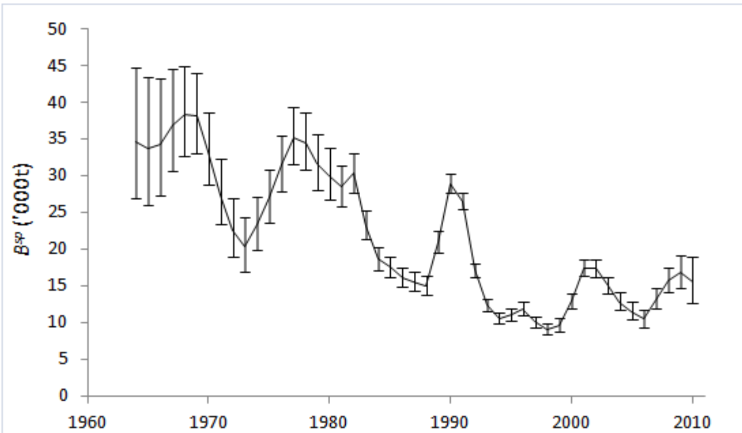

Fig. 4a: Spawning biomass trajectory for Case 8 (New Base Case), with Hessian-based 95% CIs, assuming lognormality.

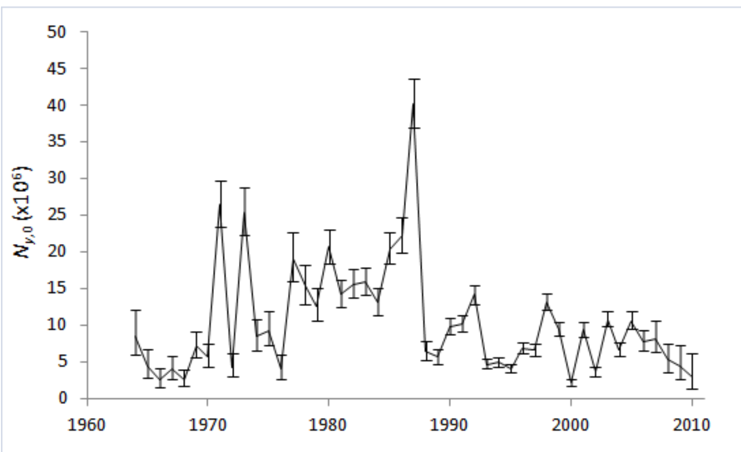

Fig. 4b: Trajectory of recruitment ($N_{y,0}$) for Case 8 (New Base Case), with Hessian-based 95%CIs, assuming lognormality.

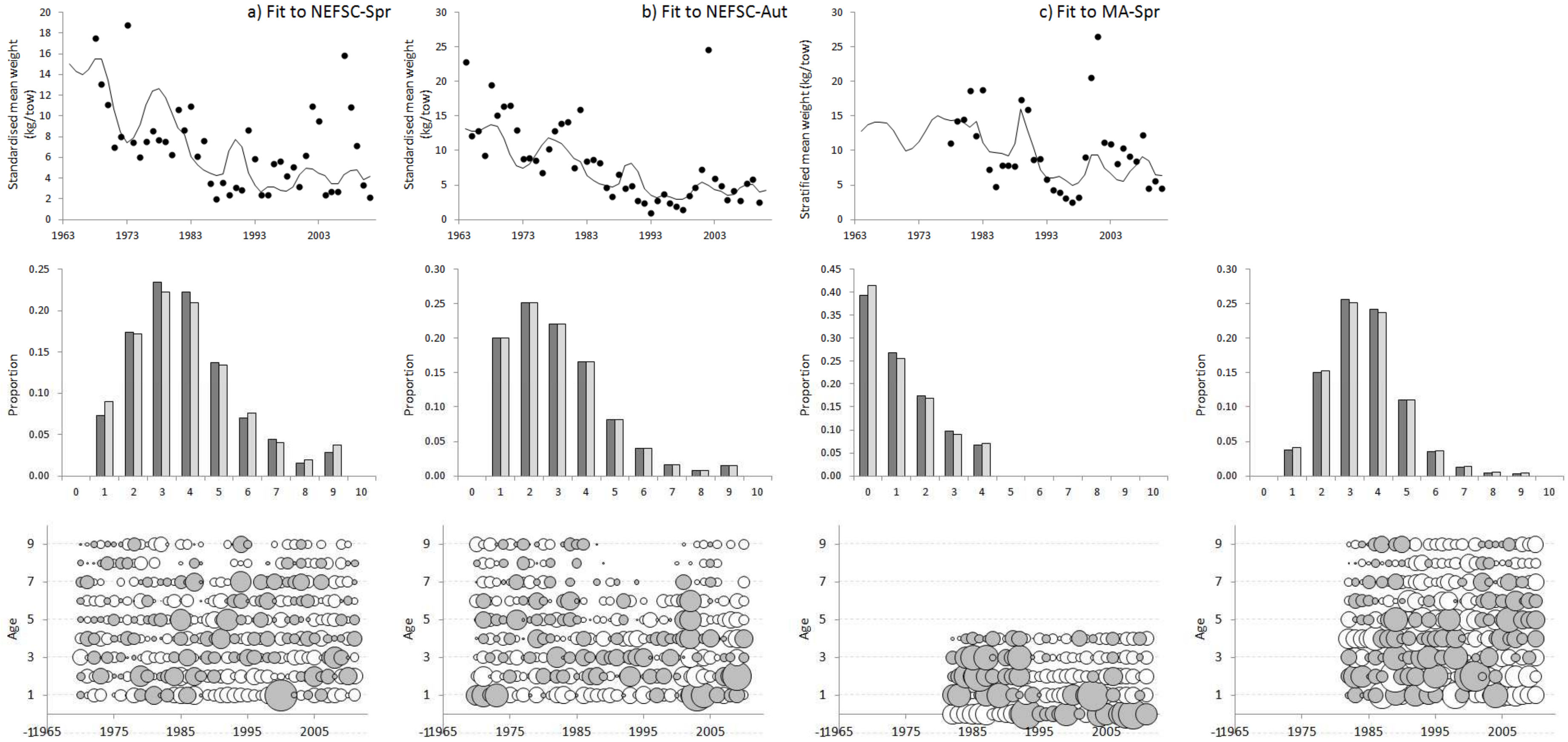

Fig. 5: Fits to the abundance indices (top row) and to the survey and commercial catch-at-age data for Case 8 (New Base Case). The middle row plots compare the observed and predicted CAA as averaged over all years for which data are available, while the bottom row plots show the standardised residuals, with the size (area) of the bubbles being proportional to the magnitude of the corresponding standardised residuals. For positive residuals, the bubbles are grey, whereas for negative residuals, the bubbles are white.

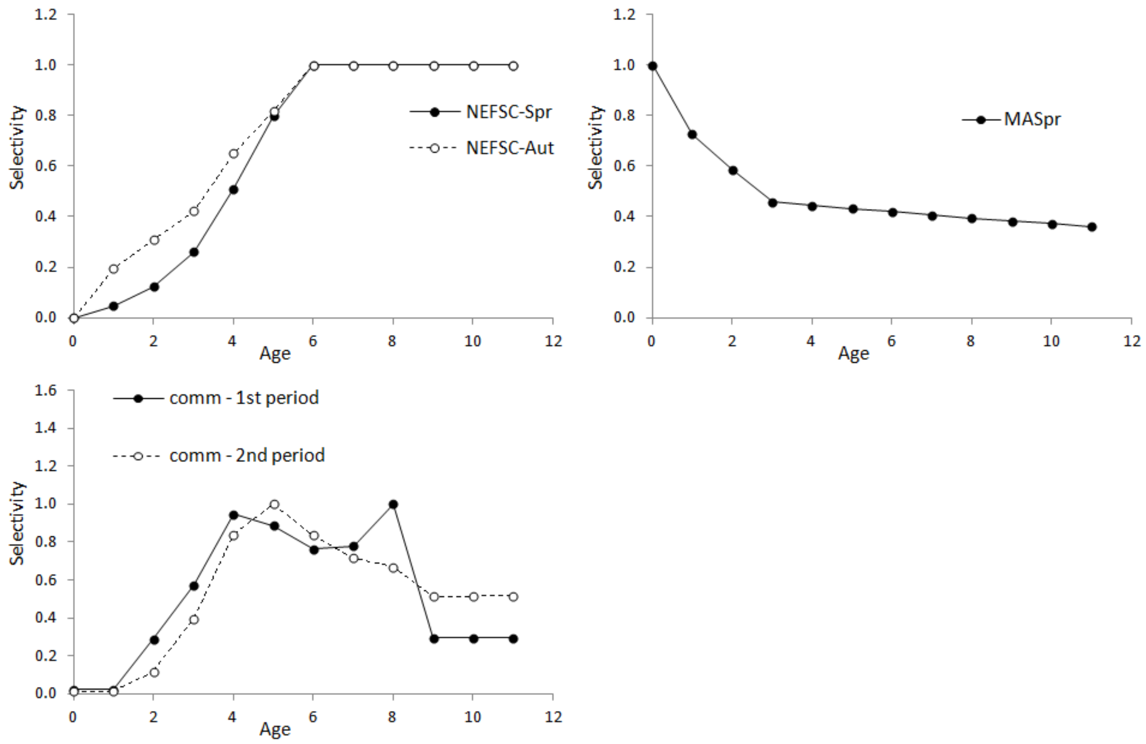

Fig. 6: Survey and commercial selectivities-at-age estimated for Case 8 (New Base Case).

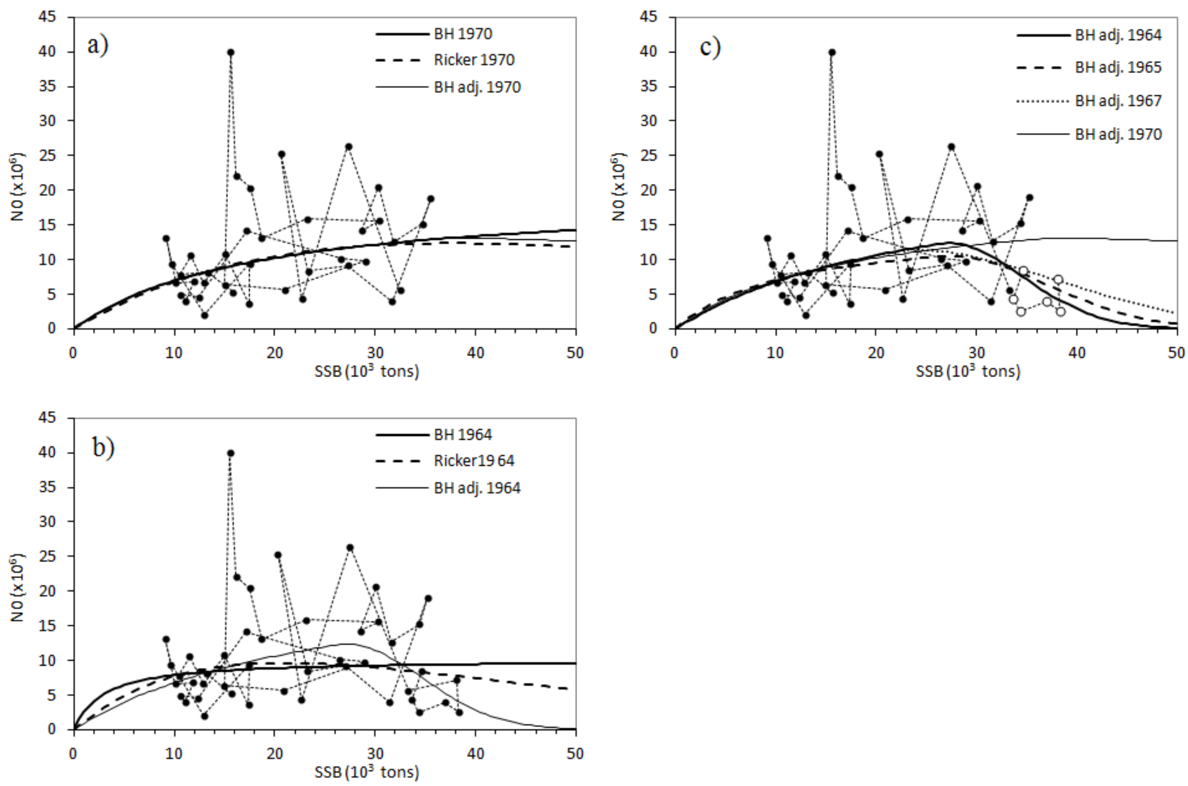

Fig. 7: Fits to the stock-recruitment data for a) data from 1970, b) data from 1964 and c) Beverton-Holt adjusted curve for data from 1964, 1965, 1967 and 1970 (though the data shown in this plot is for the assessment starting in 1964).

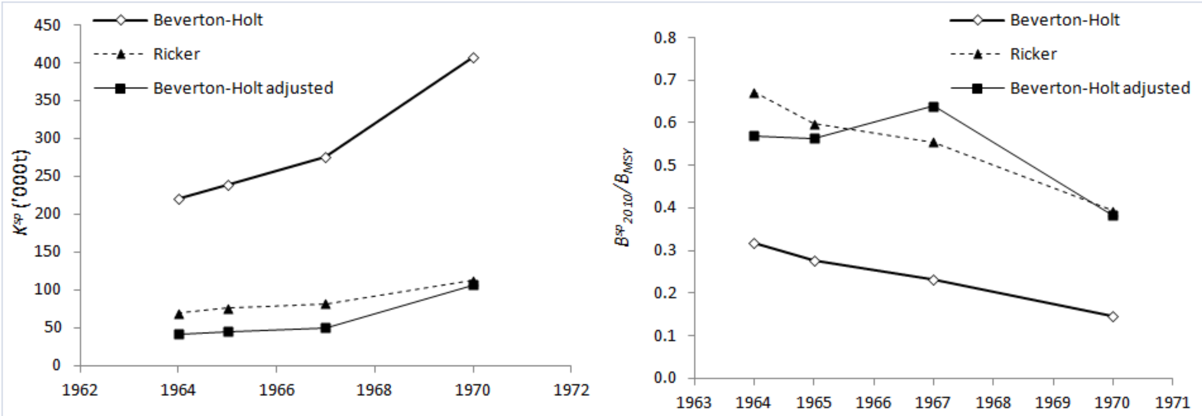

Fig. 8: K^{sp} (left-hand plot) and $B^{sp}{}_{2010}/B_{\text{MSY}}$ (right-hand plot) from fits to the stock-recruitment data for different starting years.

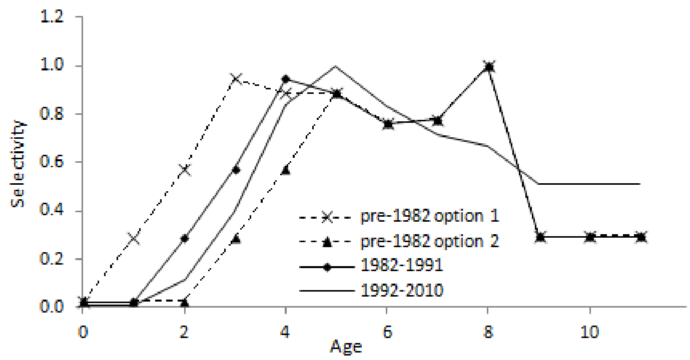

Fig. 9: Commercial selectivities-at-age for Cases 12 (pre-1982 option 1) and 13 (pre-1982 option 2).

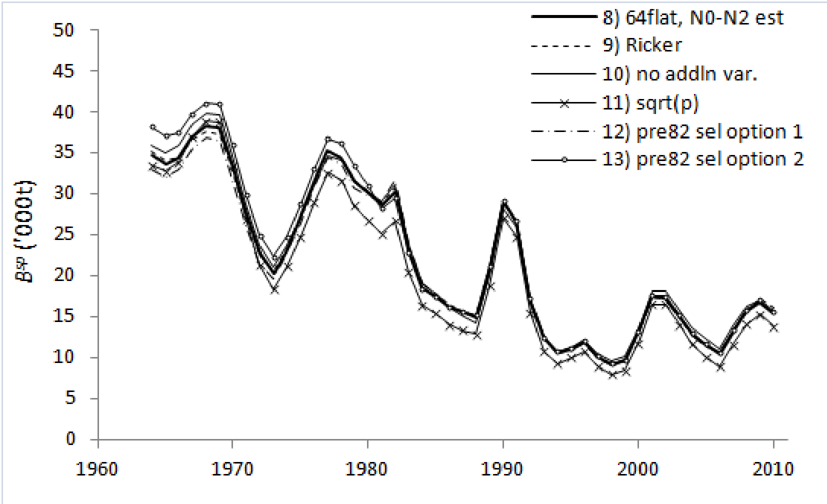

Fig. 10a: Spawning biomass trajectories cases 8 to 13 - sensitivities on the New Base Case (Case 8).

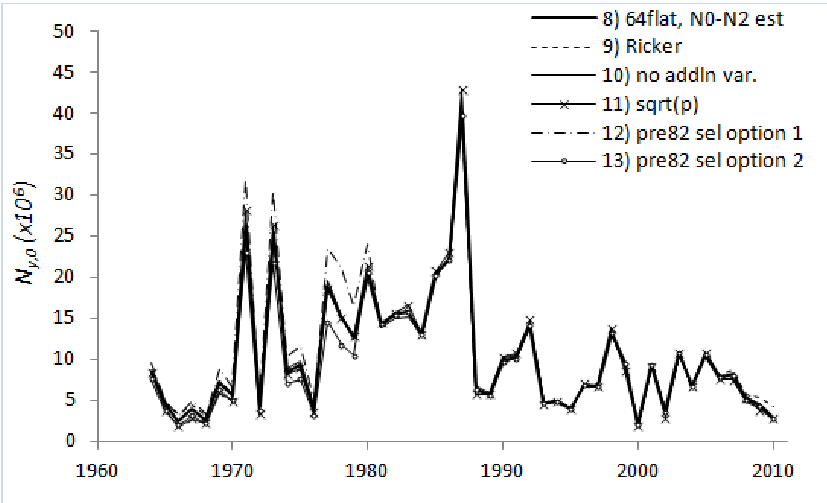

Fig. 10b: Trajectories of recruitment ($N_{y,0}$) for Cases 8 to 13 - sensitivities on the New Base Case (Case 8).

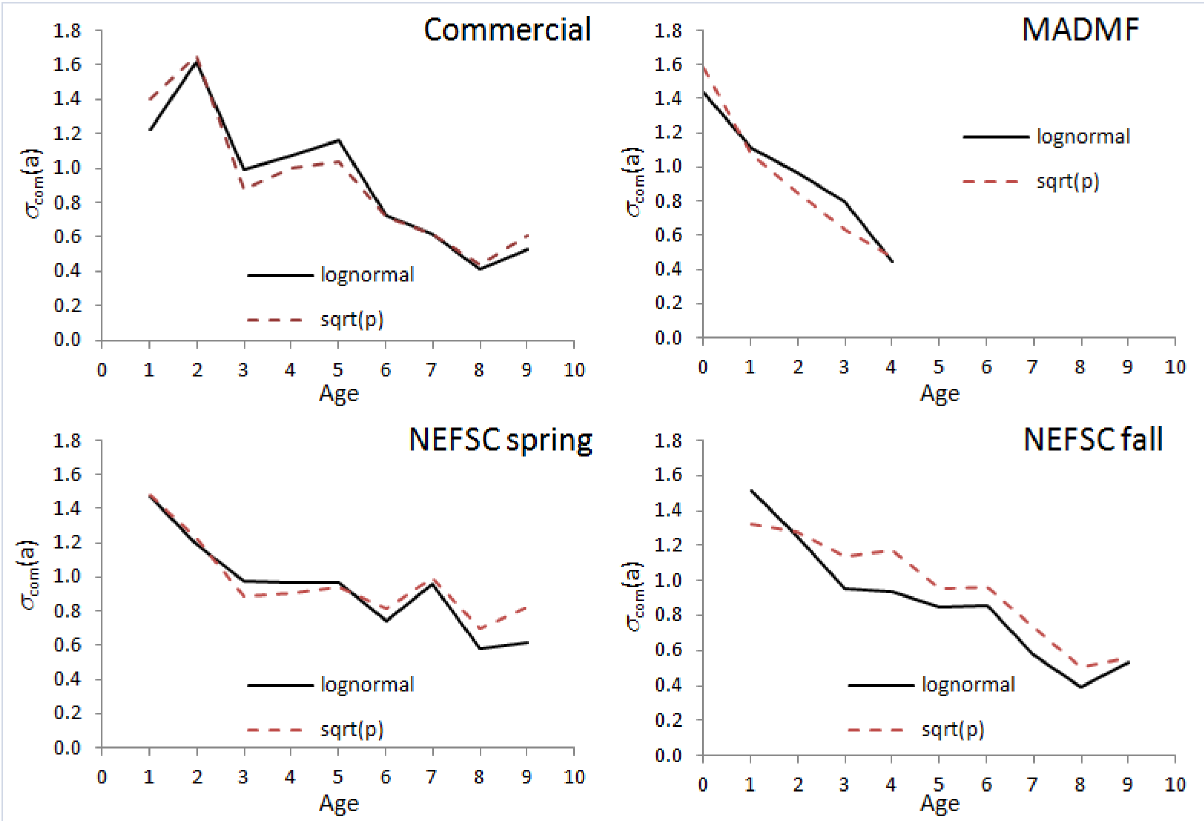

Fig. 11: σ_{com} for each age (relative to overall σ_{com} (equation 6) in each case) for the adjusted lognormal (Case 8) and sqrt(p) error distributions (Case 11) for proportions-at-age data.

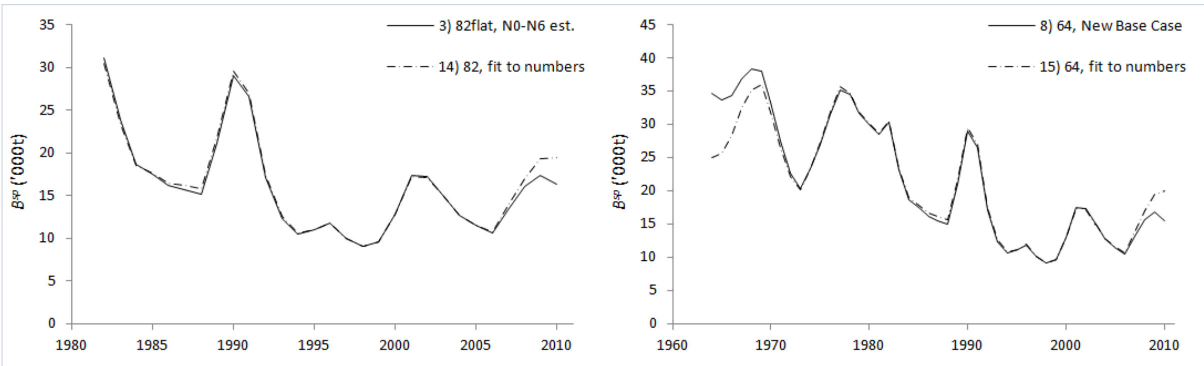

Fig. 12: Spawning biomass trajectories for cases 3 and 14 (left-hand plot: start in 1982, fitting to survey biomasses (3) or numbers (14)) and cases 8 and 15 (left-hand plot: start in 1964, fitting to survey biomasses (8) or numbers (15)).