# Statistical catch-at-age analysis vs. ADAPT-VPA: the case of Gulf of Maine cod

Doug S. Butterworth and Rebecca A. Rademeyer

Butterworth, D. S., and Rademeyer, R. A. 2008. Statistical catch-at-age analysis vs. ADAPT-VPA: the case of Gulf of Maine cod. – ICES Journal of Marine Science, 65: 1717–1732.

In 2003, given an estimate of a spawning-stock biomass ($B^{sp}$) of 27% of the maximum sustainable yield (MSY) level ($B^{sp}_{MSY}$) based on an adaptive framework-virtual population analysis (ADAPT-VPA) assessment using data only after 1981, the Gulf of Maine cod (*Gadus morhua*) stock was deemed "overfished" under the US Magnuson–Stevens Act. However, an alternative statistical catch-at-age assessment (SCAA) at the time, using survey data from 1964, indicated $B^{sp}$ above $B^{sp}_{MSY}$. This is investigated, together with other (sometimes conflicting) suggestions made during a number of recent assessment reviews of this stock. The primary reason for the different result is that the ADAPT-VPA assessment imposed asymptotically flat selectivity-at-age when there was strong statistical evidence for dome-shaped selectivity. Once adjusted for this, either assessment method robustly estimates $B^{sp}$ relatively close to $B^{sp}_{MSY}$ rather than below the "overfished" threshold of 0.5 $B^{sp}_{MSY}$. SCAA allows the longer series of survey data available to be incorporated, providing a better basis to estimate MSY-related targets and doubling the related precision in some cases. As such targets are important when implementing the Magnuson–Stevens Act, SCAA seems preferable to ADAPT-VPA for assessing this stock. Some broader inferences to be drawn from this comparative process include the need for: (i) careful treatment of the plus-group, especially if selectivity may be dome-shaped; (ii) flexible parameterizations of selectivity-at-age in SCAA to avoid false perceptions of the precision of results; and (iii) care in the use of the Beverton–Holt stock–recruitment function, as it gives inappropriately low estimates of $B^{sp}_{MSY}$ if there is an overall negative trend in the estimates of recruitment plotted against $B^{sp}$.

**Keywords:** ADAPT-VPA, age-structured production model, cod, domed selectivity, Magnuson–Stevens Act, statistical catch-at-age analysis.

Received 27 May 2008; accepted 8 October 2008.

*D. S. Butterworth and R. A. Rademeyer: Marine Resource Assessment and Management Group (MARAM), Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch 7701, South Africa. Correspondence to D. S. Butterworth: tel: +27 21 6502343; fax: +27 21 6502334; e-mail: doug.butterworth@uct.ac.za.*

## Introduction

Broadly speaking, there are two approaches to the incorporation of catch-at-age information in fisheries assessments: virtual population analysis (VPA) and statistical catch-at-age analysis (SCAA). When catch-at-age data are among those used to fit an age-structured production model (ASPM), this approach can become equivalent to SCAA; therefore, these two names are sometimes used interchangeably. Interestingly, the VPA approach tends to be the preferred method in many marine resource assessments on either side of the North Atlantic, whereas SCAA/ASPM is more frequently applied on the west coast of North America and in many southern hemisphere countries [e.g. the CASAL package originally developed for assessments of New Zealand fisheries (Bull *et al*., 2005)], as well as in some international fisheries organizations (e.g. CCAMLR, CCSBT, and IWC).

VPA (e.g. Gulland, 1965) makes the assumption that catch-at-age data are exact (i.e. with negligible error) and requires these to be available for all the years covered by the assessment. As catch-at-age data alone do not provide sufficient information to uniquely determine abundance trends (e.g. Butterworth and Punt, 1990), VPA has to be "tuned" by the incorporation of some index of relative abundance in the estimation process. Among the most popular of such approaches is the adaptive framework-VPA (ADAPT-VPA) approach, as originally introduced by Gavaris (1988).

SCAA approaches, in their simplest form, make the assumption of an invariant fishing selectivity-at-age pattern over time that determines the true age distribution of the total catch taken each year. This pattern is then estimated in the model-fitting process by comparing this distribution with the observed catch-at-age data (e.g. Punt and Hilborn, 1997). Pope (1974) and Doubleday (1976) were the first to implement this concept of separability of annual fishing mortality-at-age into age (selectivity) and year (fully selected fishing mortality) components to assist in fitting models to catch-at-age data, although Agger *et al*. (1971) had applied it in a simpler form. The catch-at-age analysis (CAGEAN) package (Deriso *et al*., 1985) constituted an early implementation of this approach. Fournier and Archibald (1982) refined the formalism with particular emphasis on the stochastic aspects to allow estimation to be set in a likelihood framework (hence the "S" of SCAA) and to admit the simultaneous (internal) estimation of the parameters of a spawning stock–recruitment function. Methot (1990) and Schnute (1994) contributed further to the development of the approach. A particular advantage of SCAA is that, unlike VPA, it does not require that catch-at-age data be available for every year covered by the assessment.

Thus, VPA assumes that the observed catch-at-age data are exact, with the fishing selectivity pattern consequently varying from year to year, whereas SCAA approaches assume the selectivity pattern to be fixed in time and consider the differences between observed and (constant selectivity) model-predicted catch-at-age data to reflect errors associated with age reading and other sources. More sophisticated approaches (e.g. Fournier *et al.*, 1998; Butterworth *et al.*, 2003a) can span the range between these two extremes by allowing for the possibility of the selectivity pattern varying over time through the use of time-series models.

ASPMs (e.g. Hilborn, 1990, who termed them general age-structured models) were a development of simpler biomass ($B$) dynamics or age-aggregated production models (AAPMs), e.g. that of Schaefer (1957), which used the logistic form $rB(1 - B/K)$ for the production function. Extending the dynamics of such models to a full age-structured form has the advantage of properly accounting for time-lags, such as the period from birth to first reproduction, and expressing biomass in a form that relates directly to quantities estimated in absolute terms by survey methods (e.g. hydroacoustic survey estimates of abundance). When fitting the model to data, given the values of biological parameters such as natural mortality, the estimation of the Schaefer model $r$ and $K$ is effectively replaced by that of two parameters of the spawning stock–recruitment function. If that function has a stochastic component, and catch-at-age data (from either or both commercial or research survey catches) are included in the fitting process, the ASPM becomes an SCAA. A simpler form of the ASPM approach was first proposed by Kimura and Tagart (1982), who termed it "stock reduction analysis", and later generalized by Kimura *et al.* (1984).

In practical terms for appropriate fisheries management advice, however, does it matter much which of the VPA or SCAA/ASPM (from now on termed ASPM) approaches is used? Punt *et al.* (2002) and Radomski *et al.* (2005) conducted comparative simulation studies based on the southeast Australia fishery and the recreational walleye (*Sander vitreus*) fishery in Lake Mille Lacs, MN, USA, respectively. Broadly speaking, both studies suggested better performance by the ASPM approach, although there were exceptions depending on the underlying reality and precise form of the assessment approach used.

In this paper, assessments of the Gulf of Maine cod (*Gadus morhua*) stock are used to examine this question. Coincidentally, groundfish resources off the Northeast US coast are good candidates for the application of the ASPM methodology because research vessel bottom trawl surveys have been conducted with essentially unchanged methodology since 1964. Therefore, to the extent that fish distribution patterns have not changed, the age-specific estimates of abundance provided by these surveys satisfy exactly the constant selectivity assumption underlying the basic ASPM approach.

The issue of possible substantial differences in assessment results for this resource under the ADAPT-VPA and ASPM methodologies first arose during the NOAA-commissioned 2003 review of the US Northeast coast assessments by a panel from the Center of Independent Experts (CIE: www.rsmas.miami.edu/groups/cie). Table 1 contrasts the results obtained at that time using the two methods: (i) ADAPT-VPA (coupled to an externally fitted spawning stock–recruitment relationship), as detailed in NEFSC (2002) and (ii) ASPM by Butterworth *et al.* (2003b), which coincidentally implemented a recommendation by the US National Research Council (NRC, 1998) that such an approach be considered for this stock in particular. Key differences are that the ADAPT-VPA approach estimated the then current spawning biomass $B^{sp}$ to be only 27% of that required to harvest the maximum sustainable yield (MSY) ($B^{sp}_{MSY}$), whereas the ASPM approach estimated $B^{sp}$ to be above that level. Furthermore, the ADAPT-VPA estimated the then current fishing mortality rate ($F$) to be almost double $F_{MSY}$, whereas the ASPM estimated it to be below $F_{MSY}$. These differences are important because the National Standard Guidelines (e.g. Federal Register, 2005) associated with the Magnuson–Stevens Fishery Conservation and Management Act governing US fisheries requires fishing mortality to be reduced if it exceeds $F_{MSY}$. Moreover, if $B^{sp}$ drops below $B_{lim}$, for which $0.5B^{sp}_{MSY}$ can serve as a proxy, a stock is declared "overfished", and a fishery management plan amendment that aims to rebuild the stock to $B^{sp}_{MSY}$ within a specified period must be put into place.

**Table 1.** Key management quantities from NEFSC (2002) ADAPT-VPA-based and the Butterworth *et al.* (2003b) ASPM assessment of Gulf of Maine cod.

| Parameter | NEFSC (2002) | Butterworth *et al.* (2003b) |
|---|---|---|
| $M$ | 0.2 | 0.2 |
| $K^{sp}$ | 274 | 159 |
| $B^{sp}_{2001}$ | 22 | 47 |
| $B^{sp}_{2001}/K^{sp}$ | 0.08 | 0.30 |
| $B^{sp}_{MSY}$ | 83 | 40 |
| $B^{sp}_{2001}/B^{sp}_{MSY}$ | 0.27 | 1.17 |
| MSYL | 0.30 | 0.25 |
| MSY | 17 | 11 |
| $F_{MSY}$ | 0.23 | 0.30 |
| $F_{2001}$ | 0.57 | 0.20 |

Biomass units are thousand tonnes and MSYL $= B^{sp}_{MSY}/K^{sp}$.

The summary report of the CIE panel at that time (Payne, 2003) found that "most methodologies used by the Northeast Fisheries Science Center (NEFSC) to compute $F_{MSY}$ and $B_{MSY}$ are adequate" and that the ADAPT-VPA methodology provided a rigorous and adequate basis for evaluating possible fisheries management policies. However, in light of the different ASPM results for Gulf of Maine cod, the panel found that "there would definitely be value in investigating the ASPM and ADAPT-based approaches to better understand the differences between them".

In December 2003, following this review, the New England Fishery Management Council (NEFMC) adopted Amendment 13 to the Northeast Multispecies Fishery Management Plan (NEFMC, 2003), which declared the Gulf of Maine cod stock "overfished" in 2001 based on the ADAPT-VPA assessment. Consequently, limits on the fishery, particularly in terms of reductions in days-at-sea allocations, were put into place to reduce effort to achieve estimated rebuilding targets for this stock and other species in the groundfish complex to meet the requirements of the Magnuson–Stevens Act.

Further interim analyses addressing the reasons for this difference (Butterworth *et al.*, 2005) were considered through NMFS-commissioned postal reviews by four independent sources (R. M. Cook, R. Hilborn, R. K. Mohn, and Cefas, Lowestoft; NEFMC, 2005a). Three of these sources acknowledged utility in the ASPM approach, and the fourth considered that it should be preferred to ADAPT-VPA in this case. These reviews were considered, in turn, at the meeting of the Scientific and

Statistical Committee (SSC) of the NEFMC held in April 2005, which concluded that "while the ASPM approach is worth consideration in conjunction with, or as an alternative to, the current ADAPT-VPA approach, it is premature to make management recommendations based on it at this time" (NEFMC, 2005b). The SSC made certain suggestions regarding further ASPM investigations, but did not review the ADAPT-VPA approach. In light of this report, the NEFMC took no immediate related action, noting that the various methods could be further considered during the next major review of groundfish assessments planned for 2008.

This paper summarizes the results of further comparisons that have been conducted since the 2005 review of the ADAPT-VPA and ASPM approaches for the Gulf of Maine cod stock, which include addressing a number of comments made by the various reviewers. The dataset used for the comparison is first specified, followed by details of the ASPM and ADAPT-VPA approaches (the latter as implemented by the NEFSC of the NOAA National Marine Fisheries Service, together with a modification thereof advanced by the authors of this paper). Results using these approaches are presented which, it is argued, identify the primary reasons for the original difference, and the associated wider implications are discussed.

## Data

The data used for the analyses here comprise: (i) annual landings by mass from 1893; (ii) year-specific weights-at-age, maturity-at-age, and landings-at-age from 1982; and (iii) mean numbers-at-age per tow from various survey series, the earliest of which commenced in 1964. These are listed in the ADAPT-VPA assessment of the Gulf of Maine cod stock by Mayo and Col (2006), which updates the data used to obtain the results in NEFSC (2002), quoted in Table 1.

However, a slight adjustment has been made to the Mayo and Col (2006) landings-at-age matrix for the calculations in this paper, because the principal interest is a comparative analysis of methodologies applied to the same dataset. This is necessary, given the data available, to allow ASPM computations to take better account of dynamics within what the ADAPT-VPA treats as a 7+ group. The impact of this adjustment on results is discussed below. The details of this adjustment are set out in Section D of the Supplementary Material. Assessments extending farther back in time than 1982 assume the same maturity-at-age vector for those earlier years as in 1982, whereas earlier weights-at-age vectors are taken to be the averages over the post-1982 periods for which those data are available.

## Methodology

Section A of the Supplementary Material first sets out the ADAPT-VPA methodology as implemented by the NEFSC for the Gulf of Maine cod stock (Mayo and Col, 2006). It then points to a mathematical inconsistency in the manner that the plus-group abundance is calculated in this approach and indicates how this can be corrected in what is termed an "Alt-VPA" approach. This last approach also allows for flexibility in the shape of the selectivity-at-age function at larger ages through introduction of an estimable parameter $\alpha$ [see Equation (S16) of the Supplementary Material], which reflects the slope of the function at such ages. Note that $\alpha = 1$ corresponds (in an average sense) to the asymptotically flat selectivity assumed for the NEFSC ADAPT-VPA assessments of Gulf of Maine cod.

Section A concludes by specifying how MSY and associated quantities (e.g. the spawning biomass corresponding to MSY, $B^{sp}_{MSY}$, and the associated fully selected fishing mortality, $F_{MSY}$) are calculated in both the ADAPT-VPA and ASPM approaches.

Section B of the Supplementary Material details the ASPM methodology applied, including the penalized maximum likelihood criterion used to fit the model. Precision is evaluated by extending the approach to a fully Bayesian form in which the penalized maximum likelihood estimates correspond to posterior modes.

In the ASPM results that follow, total penalized negative log-likelihood ($-\ln L$) values and sometimes Bayesian probability intervals (PIs) are quoted for a number of applications of the approach. However, it needs to be remembered that the inclusion of a penalty term for residuals about the stock–recruitment relationship means that these $-\ln L$ values cannot strictly be used for AIC-based model selection. Although this does not compromise the Bayesian computations, for which these penalties serve as priors, there are probably some correlations among the data inputs that the Bayesian (as well as the frequentist) approach treats as independent, and these will introduce some bias into the Bayesian estimates of PIs. Therefore, the $-\ln L$ and PI values reported are only illustrative rather than definitive in a model comparison context. Nevertheless, a model option for which, for example, a (pseudo-) AIC value is much higher than that for others should not be accorded much weight.

## Results and discussion

### Age-structured production model

To provide a focus for consideration of a potentially substantial set of results for various model options, this section is structured to address the major concerns raised by the reviewers of earlier work (Butterworth *et al.*, 2005) in NEFMC (2005a). These were:

(i) the estimability of natural mortality ($M$);

(ii) the choice of functional form for a spawning stock–recruitment relationship, with concerns about the implications of high estimates of steepness $h$ for the Beverton–Holt form; and

(iii) selectivity-related questions concerning, particularly, the strength of evidence for a dome shape, with selectivity decreasing at older ages, and the assumption of temporal invariance for the fishery in the years, before 1982, for which landings-at-age data are not available.

Results are reported for a reference case (RC) application (RC-ASPM) and a number of sensitivities (Table 2 and Table S1 in Section C of the Supplementary Material). These sensitivities are drawn from a wider set that was evaluated, with selection being based on either impact on results or queries stressed by reviewers. Some key choices in the specification of RC-ASPM are natural mortality $M = 0.2$ year$^{-1}$, the use of a Ricker form for the spawning stock–recruitment relationship, and initiating the analysis from as early a date as data are available rather than only in 1982, as for the NEFSC ADAPT-VPA assessments. An RC assessment does not claim to be a "best" assessment, but rather a convenient choice to facilitate comparisons with alternative options. Nevertheless, an RC should sensibly be chosen to be reasonably close to a likely eventual "best" selection (or set of selections). The reasons underlying the choices listed above will become clear from the comparisons discussed below.

**Table 2.** Penalized maximum likelihood estimates (followed by Bayesian posterior medians and 95% PIs in parenthesis) of key management quantities for the ASPM reference case (RC-ASPM) and nine sensitivities.

| Parameter | (1) $M = 0.2$, Ricker Reference case | | | (2) $M = 0.3$, Ricker | | | (3) $M = 0.2$, Ricker, $\gamma$ estimated | | | (4) $M = 0.2$, Beverton–Holt | | | (5) Flat survey and commercial selectivity for age 5+ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MLE | Posterior median | 95% PI | MLE | Posterior median | 95% PI | MLE | Posterior median | 95% PI | MLE | Posterior median | 95% PI | MLE | Posterior median | 95% PI |
| $-\ln L$ overall | −46.3 | – | – | −48.7 | – | – | −46.3 | – | – | −39.9 | – | – | −11.4 | – | – |
| $M$ | **0.20** | **0.20** | – | **0.30** | **0.30** | – | **0.20** | **0.20** | – | **0.20** | **0.20** | – | **0.20** | **0.20** | – |
| $h$ | 1.67 | 1.41 | (1.06; 1.82) | 1.39 | 1.20 | (0.95; 1.57) | 1.66 | 1.65 | (1.11; 2.50) | 0.98* | 0.92 | (0.78; 0.98) | 3.02 | 2.81 | (2.40; 3.21) |
| $\gamma$ | **1.00** | **1.00** | – | **1.00** | **1.00** | – | 1.05 | 0.87 | (0.54; 1.15) | – | – | – | **1.00** | **1.00** | – |
| $K^{sp}$ | 127.3 | 150.3 | (121.8; 192.3) | 82.3 | 92.9 | (78.0; 108.3) | 126.5 | 159.8 | (126.3; 213.3) | 205.0 | 234.6 | (199.5; 288.1) | 75.3 | 77.7 | (73.6; 82.3) |
| $B^{sp}_{2004}$ | 37.1 | 45.3 | (33.5; 61.3) | 32.4 | 37.0 | (28.7; 46.9) | 37.3 | 46.2 | (34.2; 63.0) | 37.8 | 49.5 | (35.5; 69.1) | 26.1 | 27.9 | (22.3; 34.3) |
| $B^{sp}_{2004}/K^{sp}$ | 0.29 | 0.30 | (0.23; 0.38) | 0.39 | 0.40 | (0.31; 0.50) | 0.29 | 0.29 | (0.22; 0.37) | 0.18 | 0.21 | (0.16; 0.27) | 0.35 | 0.36 | (0.28; 0.45) |
| $B^{sp}_{MSY}$ | 46.9 | 56.2 | (45.0; 73.8) | 33.5 | 38.0 | (31.7; 44.7) | 47.5 | 56.8 | (45.8; 73.8) | 36.1 | 46.5 | (36.5; 69.4) | 31.3 | 32.4 | (30.2; 34.9) |
| $B^{sp}_{2004}/B^{sp}_{MSY}$ | 0.79 | 0.80 | (0.61; 1.03) | 0.97 | 0.98 | (0.75; 1.24) | 0.78 | 0.81 | (0.62; 1.06) | 1.05 | 1.04 | (0.70; 1.51) | 0.84 | 0.86 | (0.67; 1.07) |
| MSYL | 0.37 | 0.37 | (0.36; 0.39) | 0.41 | 0.41 | (0.40; 0.42) | 0.38 | 0.36 | (0.31; 0.39) | 0.18 | 0.20 | (0.17; 0.25) | 0.42 | 0.42 | (0.41; 0.43) |
| MSY | 13.4 | 13.5 | (12.0; 15.3) | 12.8 | 12.8 | (11.9; 13.8) | 13.5 | 13.0 | (11.1; 15.0) | 10.5 | 10.9 | (9.6; 12.7) | 13.0 | 12.9 | (12.5; 13.1) |
| $F_{MSY}$ | 0.62 | 0.52 | (0.39; 0.66) | 0.89 | 0.70 | (0.51; 1.05) | 0.61 | 0.50 | (0.38; 0.64) | 0.65 | 0.52 | (0.37; 0.70) | 0.69 | 0.61 | (0.51; 0.75) |
| $F_{2004}$ | 0.26 | 0.22 | (0.16; 0.30) | 0.27 | 0.23 | (0.17; 0.31) | 0.26 | 0.22 | (0.16; 0.30) | 0.28 | 0.22 | (0.16; 0.32) | 0.25 | 0.23 | (0.18; 0.30) |

| | (6) Linear selectivity for NEFSC survey, $M = 0.2$ | | | (7) Exclude pre-1982 index data, $M = 0.2$, Ricker | | | (8) Exclude pre-1982 index data, $M = 0.2$, Beverton–Holt | | | (9) Alternative earlier commercial selectivity (pre-1982 = post-1991) | | | (10) Alternative earlier commercial selectivity: pre-1982, $S_1 = 0$, $S_2 = 0.35$, $S_3 = 0.85$, and $S_4 = 1.0$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MLE | Posterior median | 95% PI | MLE | Posterior median | 95% PI | MLE | Posterior median | 95% PI | MLE | Posterior median | 95% PI | MLE | Posterior median | 95% PI |
| $-\ln L$ overall | 19.5 | – | – | (−12.6) | – | – | (−8.6) | – | – | −30.6 | – | – | −56.5 | – | – |
| $M$ | **0.20** | **0.20** | – | **0.20** | **0.20** | – | **0.20** | **0.20** | – | **0.20** | **0.20** | – | **0.20** | **0.20** | – |
| $h$ | 3.74 | 3.46 | (2.63; 4.14) | 1.30 | 0.93 | (0.56; 1.41) | 0.90 | 0.78 | (0.53; 0.94) | 1.66 | 1.39 | (1.04; 1.79) | 1.70 | 1.41 | (1.09; 1.86) |
| $\gamma$ | **1.00** | **1.00** | – | **1.00** | **1.00** | – | **–** | **–** | – | **1.00** | **1.00** | – | **1.00** | **1.00** | |
| $K^{sp}$ | 73.6 | 76.1 | (69.0; 86.3) | 166.4 | 242.4 | (168.3; 382.8) | 268.7 | 354.5 | (283.4; 482.0) | 125.8 | 149.2 | (121.2; 189.6) | 128.7 | 152.7 | (124.0; 191.3) |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $B^{sp}_{2004}$ | 27.9 | 29.0 | (23.2; 36.2) | 50.6 | 92.7 | (54.6; 191.8) | 62.1 | 120.1 | (71.0; 217.9) | 36.3 | 44.5 | (32.9; 60.4) | 37.3 | 45.0 | (33.8; 60.3) |
| $B^{sp}_{2004}/K^{sp}$ | 0.38 | 0.38 | (0.30; 0.47) | 0.30 | 0.38 | (0.28; 0.58) | 0.23 | 0.34 | (0.23; 0.48) | 0.29 | 0.30 | (0.23; 0.38) | 0.29 | 0.29 | (0.23; 0.38) |
| $B^{sp}_{MSY}$ | 28.9 | 30.2 | (26.5; 34.0) | 62.4 | 95.0 | (62.7; 159.4) | 54.3 | 85.8 | (54.7; 148.8) | 46.1 | 55.8 | (44.7; 72.7) | 47.3 | 57.0 | (45.5; 73.3) |
| $B^{sp}_{2004}/B^{sp}_{MSY}$ | 0.97 | 0.96 | (0.75; 1.20) | 0.81 | 0.98 | (0.69; 1.51) | 1.14 | 1.38 | (0.84; 2.42) | 0.79 | 0.79 | (0.60; 1.02) | 0.79 | 0.79 | (0.60; 1.03) |
| MSYL | 0.39 | 0.39 | (0.38; 0.41) | 0.38 | 0.39 | (0.37; 0.42) | 0.20 | 0.25 | (0.17; 0.33) | 0.37 | 0.37 | (0.36; 0.39) | 0.37 | 0.37 | (0.36; 0.39) |
| MSY | 14.1 | 13.9 | (13.1; 14.9) | 13.6 | 13.4 | (10.4; 17.5) | 11.8 | 12.5 | (9.5; 16.8) | 13.0 | 13.0 | (11.4; 14.8) | 13.8 | 13.7 | (12.3; 15.4) |
| $F_{MSY}$ | 1.20 | 1.00 | (0.67; 1.61) | 0.51 | 0.36 | (0.20; 0.54) | 0.54 | 0.37 | (0.20; 0.59) | 0.58 | 0.38 | (0.31; 0.46) | 0.63 | 0.41 | (0.34; 0.49) |
| $F_{2004}$ | 0.25 | 0.24 | (0.19; 0.32) | 0.22 | 0.15 | (0.09; 0.23) | 0.21 | 0.13 | (0.08; 0.21) | 0.26 | 0.22 | (0.16; 0.30) | 0.26 | 0.22 | (0.17; 0.30) |

Biomass units are thousand tonnes. The estimates given for quantities such as $B^{sp}_{MSY}$ refer to the commercial selectivity function from 1992+. Values shown emboldened are fixed on input. Negative log-likelihoods are shown in parenthesis when not comparable with that for the RC-ASPM because of data differences. See Subsection B.4.1 of the Supplementary Material for specifications of Sensitivity 6. The asterisk indicates a constraint boundary. Definitions of variables not provided in the main text may be found in Sections A or B of the Supplementary Material.

## Estimability of $M$

In earlier ASPM implementations (e.g. Butterworth *et al.*, 2005), the likelihood profile results for $M$ had suggested that this could be estimated with reasonable precision. However, those assessments had assumed that the selectivity pattern in the NEFSC surveys was linear with age based on such indications from the ADAPT-VPA results. Such an assumption (Sensitivity 6 of Table 2) leads to a greatly inferior fit to the data compared with RC-ASPM ($-\ln L$ larger by $>65$ units); hence the fully flexible form now used (see Subsection B.4.1 of the Supplementary Material). However, when such linearity was not imposed, there was no indication that $M$ was estimable with reasonable precision. RC-ASPM therefore fixes $M$ at 0.2, as has been customary for NEFSC ADAPT-VPA assessments of this stock.

Bayesian posterior medians and 95% probability envelopes for historical spawning biomass ($B^{sp}$) trends for RC-ASPM are shown in Figure 1. The steep decline shown over the first decade in the series should not be considered particularly reliable, because the stock had been exploited before 1893, contrary to the assumption of unexploited equilibrium at this time made here. Sensitivity tests (see Table S1 in Section C of the Supplementary Material) show that the estimates of recent trends in abundance and quantities of importance for management are insensitive to this assumption. Starting the analysis as far back as 1893 adds the advantage of not having to specify or estimate the parameters $\theta$ and $\varphi$, which account for the non-equilibrium nature of the starting population age structure [see Equations (S36) and (S37) in Section B of the Supplementary Material]. Figures 2–4 show fits to the indices of abundance and age-structure information and do not show any obvious indications of model mis-specification. Figure 5 shows the estimates of selectivities-at-age for both commercial catches and NEFSC surveys; the dome shape is evident, with a steeper decline at greater age for commercial catches compared with surveys.

Sensitivity 2 (Table 3) explores the consequences of changing the value of $M$ to 0.3 (Figure 6). This higher value is slightly preferred in likelihood terms. It results in estimates of spawning biomass somewhat lower in absolute terms, but also an estimated current status of $B^{sp}_{2004}$ closer to the target MSY level $\left(\mathrm{MSYL} = B^{sp}_{MSY}/K^{sp}\right)$.

## Stock–recruitment relationship

Earlier ASPM results (e.g. Butterworth *et al.*, 2005) focused on the use of the Beverton–Holt spawning stock–recruitment relationship, because this had been preferred in the original NEFSC (2002) assessments. However, both sets of assessments yielded estimates of steepness $h$ close to the maximum of 1 that applies for this form, and concerns were raised that this yielded a very low estimate of MSYL, below most existing observations, which, if accepted, would see management targeting low abundance levels where inferred resource behaviour depended on extrapolation beyond the range of most available data.

Sensitivity 4 in Table 2 shows the results of replacing the Ricker form in RC-ASPM by a Beverton–Holt form; the associated fits to annual estimates of recruitment ($N_1$) and spawning biomass are shown in Figure 7 for each case. The Ricker form achieves a better fit to the data (some 6 log-likelihood units), or some 400 times more likely in AIC weighting terms, and also leads to an estimate of $B^{sp}_{MSY}$ which is close to the centre of the range of $B^{sp}$ values rather than to the lower end (and hence more reliably estimated).

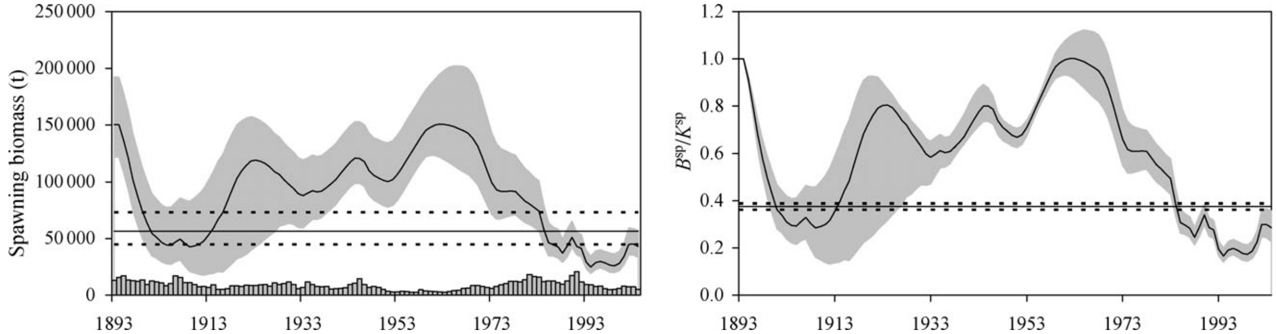


**Figure 1.** Posterior medians of spawning-biomass trajectories (in absolute terms and in terms of pre-exploitation level) for the RC-ASPM. The shaded areas represent the 95% PI envelopes. The estimated $B^{sp}_{MSY}$ and MSYL are also shown, with the 95% PI as dotted lines. The bar plot shows the annual total landings (t).

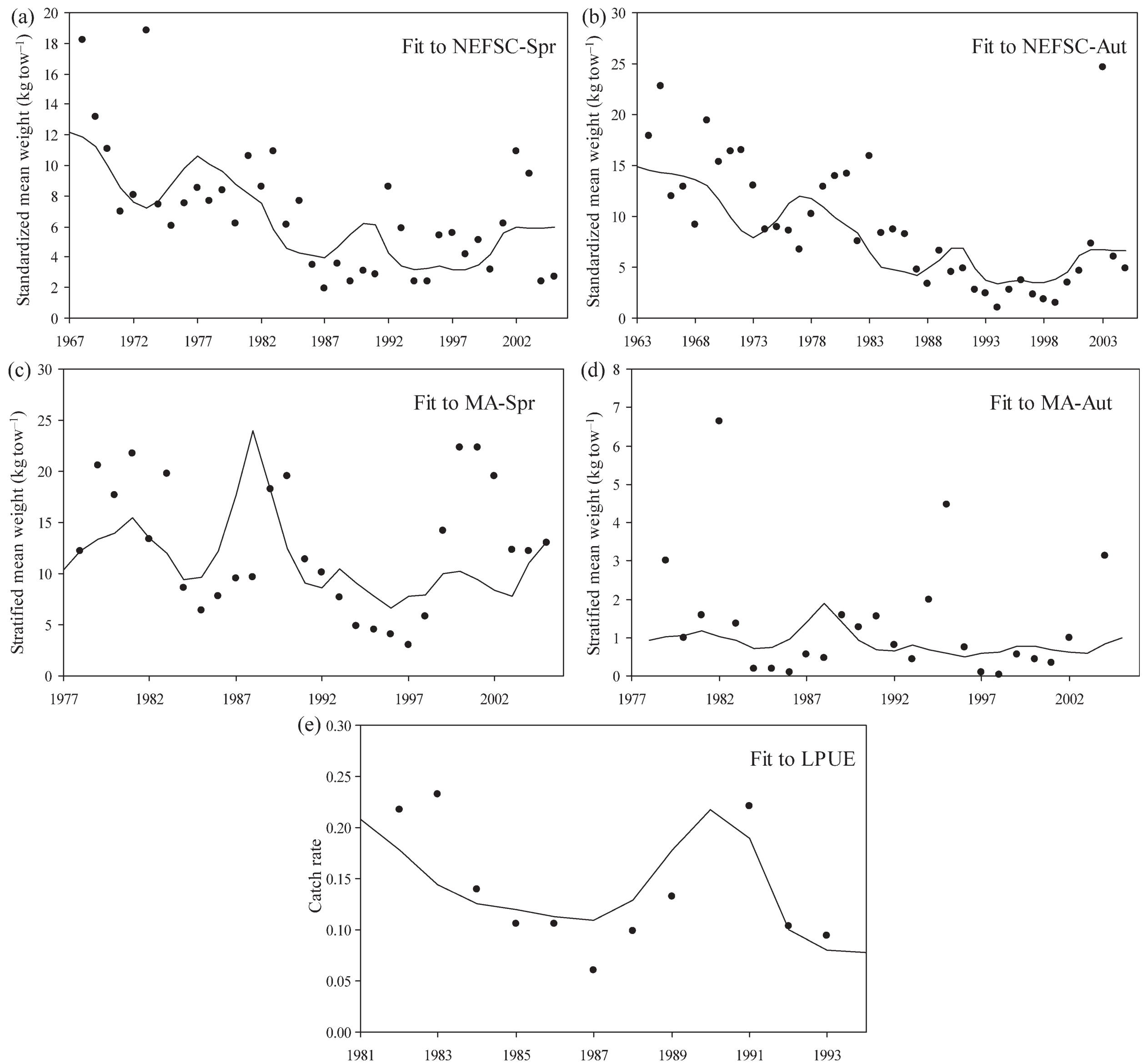


**Figure 2.** RC-ASPM assessment model fits to the abundance indices [survey and catch (landings) per unit effort, lpue].

Further, there is little indication of serial correlation in recruitment residuals about the Ricker curve where the data available allow these to be reasonably well estimated (Figure 8). Figure 7 indicates why the monotonically increasing Beverton–Holt form has difficulties in this case: with low recruitment having occurred at the highest biomass levels, the implied overall negative trend of

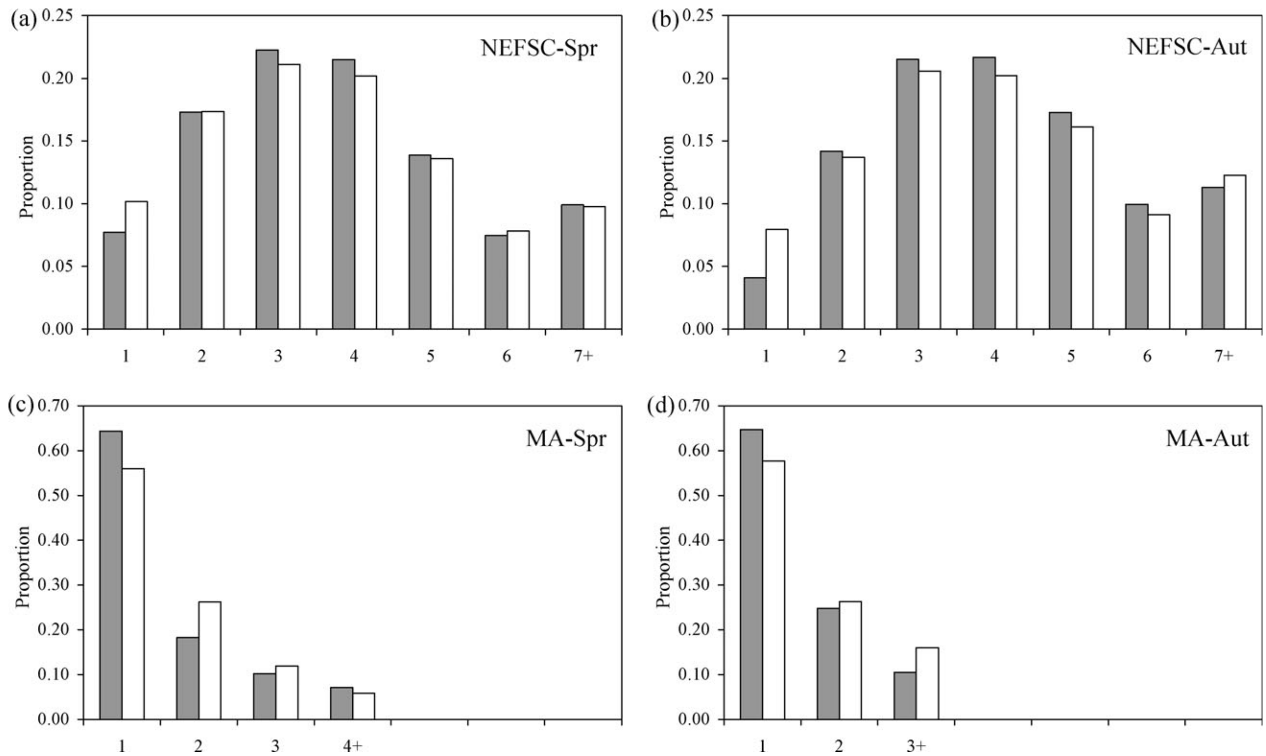


**Figure 3.** RC-ASPM assessment model fits to the catch-at-age data (survey and commercial averaged over all the years with data for each dataset).

recruitment with $B^{sp}$ can be accommodated only by setting $h$ as large as the form permits. All these considerations indicate that Ricker is the preferred of the two forms here.

Actually, the Ricker form is a special case of a more general form examined [see Section B, Equation (S30) in the Supplementary Material], which includes an additional shape parameter $\gamma$. However, attempts to estimate $\gamma$ (see Sensitivity 3 in Table 2 and Figure 6) offer neither sufficient improvement to the likelihood for the introduction of $\gamma$ to be justifiable in AIC terms nor meaningful differences to the fit.

## Dome-shaped selectivity

Figure 5 shows that the RC-ASPM maximum penalized likelihood estimates of selectivity-at-age are dome-shaped for both NEFSC surveys and commercial catches. If these selectivities are forced to be asymptotically flat (see Sensitivity 5, Table 2), $-\ln L$ deteriorates by 35 units, so that introduction of a dome shape is strongly preferred in AIC terms. Most of this deterioration takes place for the fit of proportion at ages 5–7+ for commercial catches and NEFSC surveys. Figure 9 shows residual plots for proportion-at-age fits for Sensitivity 5 and has to be compared with those for RC-ASPM shown in Figure 4. The model mis-specification in the former case (with asymptotically flat selectivity) is evident because most of the commercial catch-at-age residuals for ages 6 and 7+ are negative (i.e. fewer older fish are observed than are consistent with the flat selectivity assumption). This effect is also present, although not quite as evident, for the NEFSC autumn surveys. For RC-ASPM, the Bayesian posterior median and 95% PI estimates for $S_7/S_6$ selectivity ratios are 0.52 [0.41; 0.64] and 0.72 [0.66; 0.79] for the commercial and NEFSC survey catches. If $M = 0.3$, these estimates increase as expected: commercial = 0.68 [0.55; 0.83] and survey = 0.89 [0.82; 0.97], i.e. they still do not overlap the value of 1 that corresponds to flat selectivity.

In general, the estimation of age dependence in selectivity and in natural mortality, as well as the value of the latter if age-independent, tends to be confounded (Butterworth and Punt, 1990; Thompson, 1994; Clark, 1999). However, a value of $M = 0.3$ seems about as large as might enjoy general support as a realistic estimate of $M$ for the Gulf of Maine cod stock. A further possibility to explain the relative paucity of older fish is the increase of natural mortality at older ages, but there are no immediately obvious biological mechanisms that might lead to the substantial increases necessary. Therefore, the results above taken together strongly suggest that the available data are not compatible with the assumption of asymptotically flat selectivity, but rather evidence of a decline at higher ages.

## Selectivity prior to 1982

The ASPM requires some assumption concerning commercial selectivity-at-age before 1982. In the absence of landings-at-age data for any of that period, RC-ASPM sets this equal to that estimated for the period 1982–1991 (see Subsection B.4.1 of the Supplementary Material) making them consequently time-invariant.

This assumption is certainly not correct; for example, there were gear regulation changes during the pre-1982 period. The question, though, is whether incorporating such information into the analyses would substantially modify the key results. Some sensitivity tests to alternative (though also time-invariant) assumed commercial selectivities-at-age pre-1982 (see Sensitivities 9 and 10 in Table 2) suggest very little change to estimates of current $F$ and $B^{sp}$ levels relative to those at MSY. Considerable

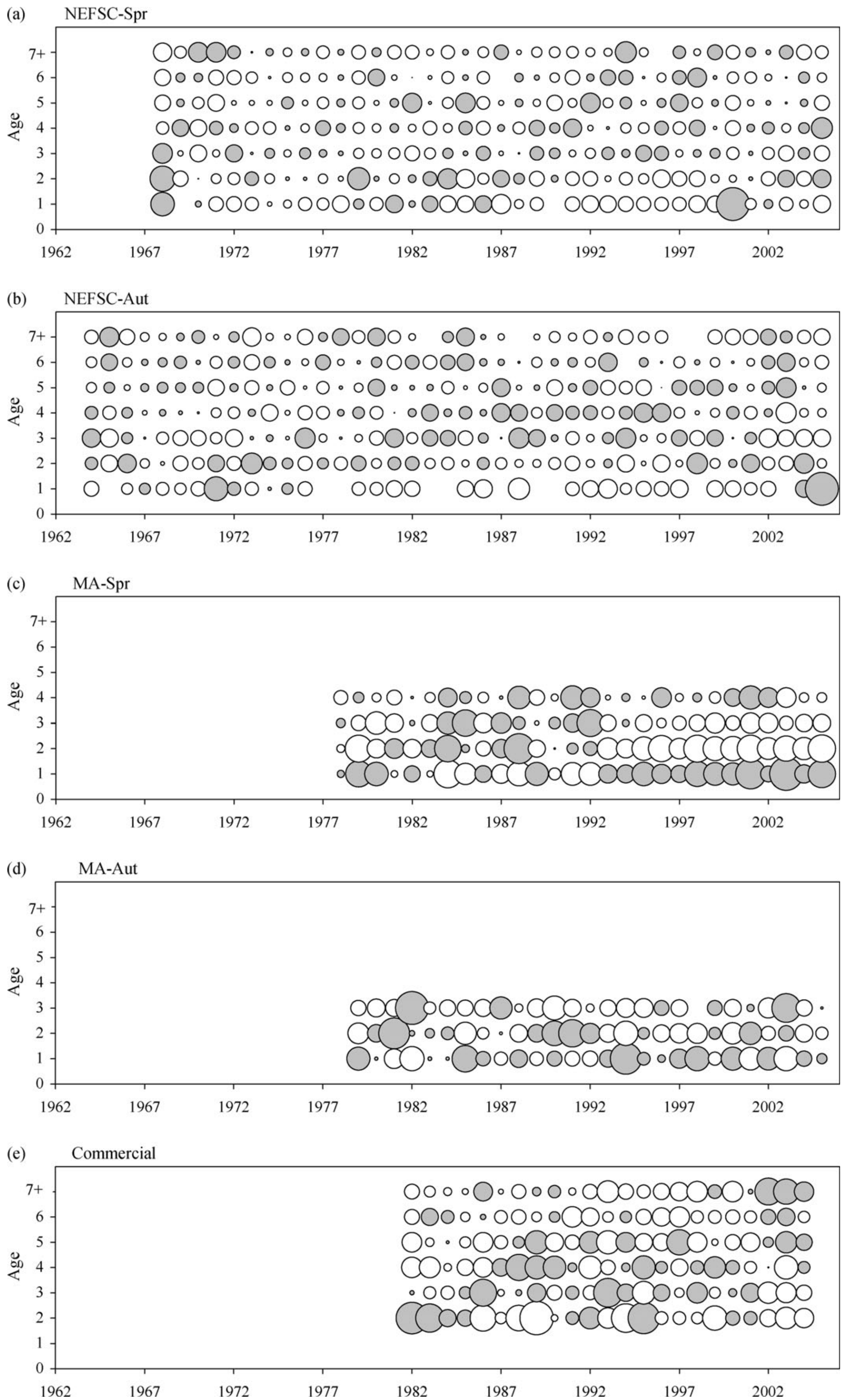


**Figure 4.** Bubble plots of the standardized residuals for the catch-at-age data for the RC-ASPM assessment. The size (area) of the bubbles represents the size of the residuals. Grey bubbles represent positive residuals and white bubbles represent negative residuals.

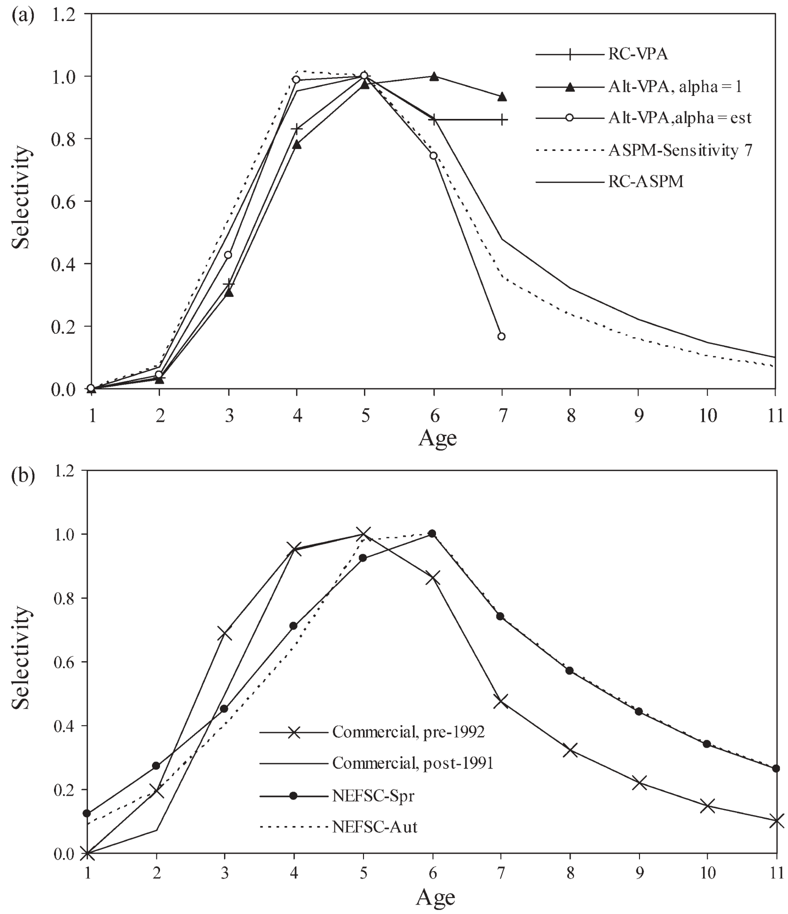


**Figure 5.** (a) Commercial selectivities-at-age (average over 1992–2004) for the ASPM reference case and Sensitivity 7 (RC except exclude pre-1982 index data), the RC-VPA, and two of the VPA sensitivities; (b) commercial (pre-1992 and post-1991) and NEFSC survey selectivities-at-age for the RC-ASPM.

experience with the ASPM approach in many other fisheries (e.g. at the International Whaling Commission) suggests that such effects are generally second order, with the historical sequence of total catch by weight being the much more influential factor. The probability that this is also the case here is supported by the trends shown in Figure 10 of the per-recruit contribution by age to cohort biomass, taking into account the effects of both natural mortality and somatic growth. These are relativity flat over ages 3–6 for which the commercial selectivity is relatively high, which suggests that limited changes in the age distribution of catches will have little impact on biomass and hence on future resource dynamics.

A related question pertains to the survey data before 1982 and whether concerns of their possible lack of comparability with later data should rather see analyses restricted only to the use of 1982+ abundance indices and proportion-at-age data. Sensitivities 7 and 8 in Table 2 address this, omitting pre-1982 data from the fitting criterion compared with the RC-ASPM (Ricker) and its Beverton–Holt counterpart (Sensitivity 4). In both cases, there is an appreciable decrease in the precision with which certain quantities can be estimated. The ranges of the 95% PI for MSY and importantly $B^{sp}_{2004}/B^{sp}_{MSY}$ roughly double, and those for $B^{sp}_{2004}$ in absolute terms roughly treble. Radomski *et al.* (2005) point to the possibility of large errors if commercial selectivity varies rather than remaining constant, as assumed for this ASPM implementation, but this possibility needs to be weighed against the fact that the NEFSC surveys are perhaps the longest in the world and have focused on maintaining the same methodology.

## Comparison of ADAPT-VPA and ASPM results

A number of ADAPT-VPA assessment results are reported in Table 3, together with those for related ASPM assessments. The different spawning-biomass trajectories are shown in Figure 11, with selectivity-at-age functions plotted in Figure 5. Note that the VPA results themselves are independent of the spawning stock–recruitment function form, which is fitted externally to VPA outputs, so the Beverton–Holt vs. Ricker distinction affects only certain of the quantities listed in Table 3.

The ADAPT-VPA RC assessment (RC-VPA) applies the same methodology as Mayo and Col (2006) (see Section A of the Supplementary Material), but to the slightly amended data as mentioned above and detailed in Section D of the Supplementary Material. There is little qualitative difference between the Mayo and Col (2006) results and the RC-VPA results. The Alt-VPA ($\alpha = 1$) method, which involves alternative treatment of the plus-group while maintaining the asymptotically flat selectivity assumption (in an average sense), produces a virtually identical spawning-biomass trajectory to RC-VPA, and also a similar estimate of $B^{sp}_{2004}$ relative to $B^{sp}_{MSY}$ (Table 3).

**Table 3.** Estimates of key management quantities for VPA assessments of Gulf of Maine cod.

| Parameter | Beverton–Holt | | | | | | Ricker | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mayo and Col (2006) | RC-VPA | Alt-VPA, $\alpha = 1$ | Alt-VPA, $\alpha =$ est | ASPM-Sensitivity 8, excluding pre-1982 data | ASPM-Sensitivity 4 | Mayo and Col (2006) | RC-VPA | Alt-VPA, $\alpha = 1$ | Alt-VPA, $\alpha =$ est | ASPM-Sensitivity 7, excluding pre-1982 data | RC-ASPM |
| VPA SR SS | (9.78) | 9.88 | 9.89 | 9.95 | – | – | (9.71) | 9.71 | 9.77 | 9.92 | – | – |
| VPA fit SS 1–6 | (165.99) | 166.01 | 169.33 | 163.36 | – | – | (165.99) | 166.01 | 169.33 | 163.36 | – | – |
| VPA fit SS 1–7+ | | 206.16 | 208.06 | 196.46 | – | – | – | 206.16 | 208.06 | 196.46 | – | – |
| Penalty $p_2$ | | | 4.09 | 9.36 | – | – | – | – | 4.09 | 9.36 | – | – |
| $M$ | **0.2** | **0.2** | **0.2** | **0.2** | **0.2** | **0.2** | **0.2** | **0.2** | **0.2** | **0.2** | **0.2** | **0.2** |
| $h$ | 0.80 | 0.84 | 0.74 | 0.83 | 0.90 | 0.98* | 2.08 | 2.18 | 1.90 | 1.73 | 1.30 | 1.67 |
| $K^{sp}$ | 224.0 | 193.2 | 250.5 | 186.2 | 268.7 | 205.0 | 70.1 | 63.0 | 73.2 | 79.0 | 166.4 | 127.3 |
| $B^{sp}_{2004}$ | 21.0 | 21.3 | 20.7 | 30.8 | 62.1 | 37.8 | 21.0 | 21.3 | 20.7 | 30.8 | 50.6 | 37.1 |
| $B^{sp}_{2004}/K^{sp}$ | 0.09 | 0.11 | 0.08 | 0.17 | 0.23 | 0.18 | 0.30 | 0.34 | 0.28 | 0.39 | 0.30 | 0.29 |
| $B^{sp}_{MSY}$ | 72.1 | 61.2 | 81.0 | 44.4 | 54.3 | 36.1 | 30.9 | 27.8 | 31.8 | 29.5 | 62.4 | 46.9 |
| $B^{sp}_{2004}/B^{sp}_{MSY}$ | 0.29 | 0.35 | 0.25 | 0.69 | 1.14 | 1.05 | 0.68 | 0.77 | 0.65 | 1.04 | 0.81 | 0.79 |
| MSYL | 0.32 | 0.32 | 0.32 | 0.24 | 0.20 | 0.18 | 0.44 | 0.44 | 0.43 | 0.37 | 0.38 | 0.37 |
| MSY | 15.0 | 13.1 | 15.8 | 9.9 | 11.8 | 10.5 | 10.9 | 10.1 | 10.8 | 9.6 | 13.6 | 13.4 |
| $F_{MSY}$ | 0.24 | 0.27 | 0.26 | 0.53 | 0.54 | 0.65 | 0.53 | 0.63 | 0.57 | 0.77 | 0.51 | 0.62 |
| $F_{2004}$ | 0.38 | 0.36 | 0.44 | 0.35 | 0.21 | 0.28 | 0.38 | 0.36 | 0.44 | 0.35 | 0.22 | 0.26 |

Biomass units are tonnes. The estimates given for quantities such as $B^{sp}_{MSY}$ refer to an average commercial selectivity function for 1992–2004 and MSYL $= B^{sp}_{MSY}/K^{sp}$. Values shown emboldened are fixed on input. Objective function (SS) values shown in parenthesis are not comparable with those for RC-VPA because of data differences. Note that the RC-VPA is fitted to SS over ages 1–6 as in Mayo and Col (2006), whereas Alt-VPA is fitted to SS over ages 1–7+. (Table S2 of Section C of the Supplementary Material provides a detailed breakdown by series and age of the contributions to SS for RC-VPA.) See Equation (S15) in Section A of the Supplementary Material for details of the penalty term $p_2$. The asterisk indicates a constraint boundary.

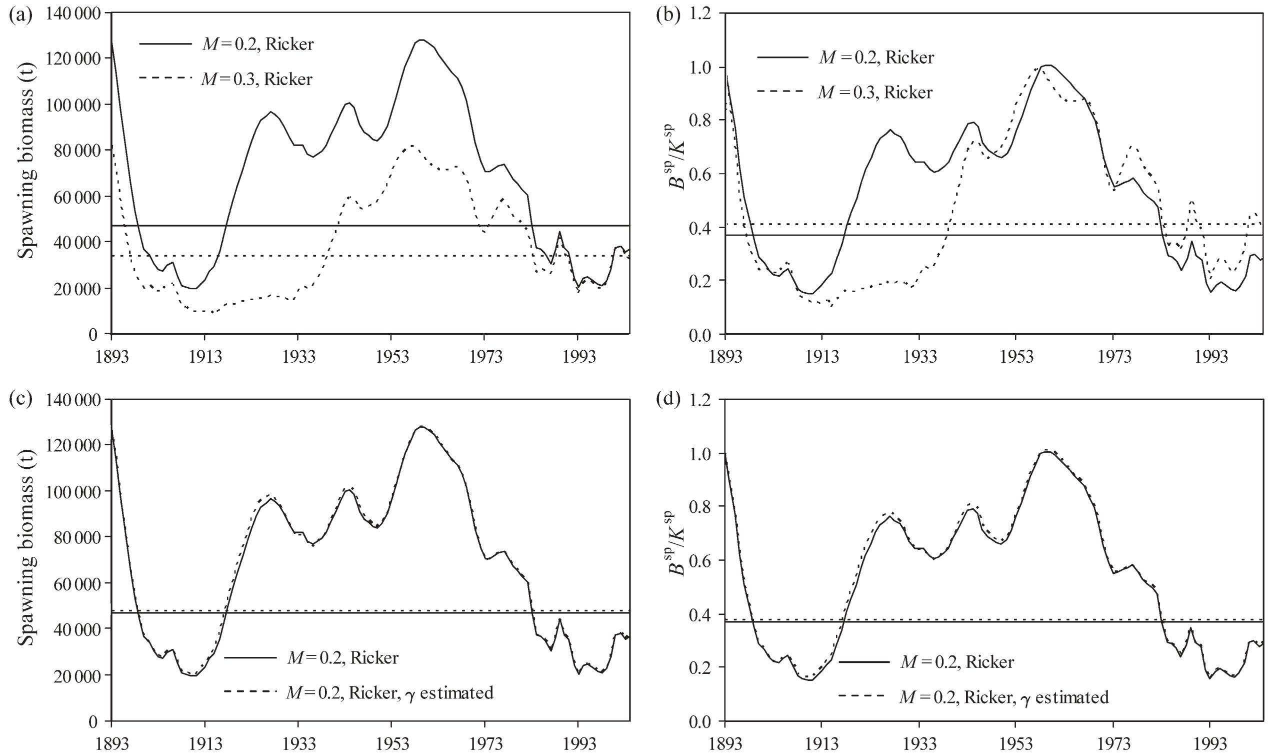


**Figure 6.** Comparison of MLE ASPM spawning biomass trajectories (in absolute terms and in terms of pre-exploitation level) for Sensitivities 1 (RC-ASPM) and 2 (RC with $M = 0.3$), and Sensitivities 1 and 3 (RC with $\gamma$ estimated). The estimated $B^{sp}_{MSY}$ and MSYL are also shown.

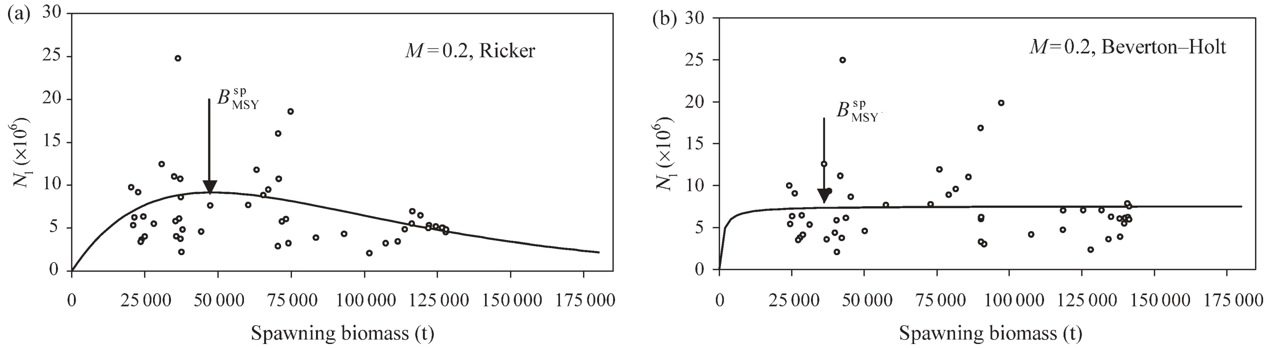


**Figure 7.** The estimated stock–recruitment curve and estimated recruitments each year over the period 1956–2004 for (a) RC-ASPM and (b) Sensitivity 4 (RC with Beverton–Holt). (Figure S2 of Section C of the Supplementary Material shows results for further sensitivities.)

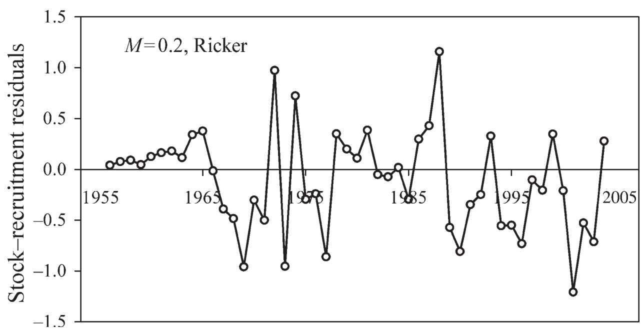


**Figure 8.** Estimated stock–recruitment residuals ($\varsigma_y$) for RC-ASPM.

Allowing the selectivity slope at higher age to be estimated ("$\alpha =$ est") within the Alt-VPA framework suggests dome-shaped selectivity (Figure 5a) and a higher biomass in absolute terms (Figure 11), together with a further increase in the estimate of $B^{sp}_{2004}$ relative to $B^{sp}_{MSY}$ (Table 3). When $\alpha$ is fixed at 1, the fit to the survey data is slightly worse in the Alt-VPA method compared with RC-VPA, but become better than those for RC-VPA when $\alpha$ is estimated. This holds whether the fitting criterion excludes (as for NEFSC assessments) or includes the 7+ group. However, these comparisons are not entirely even-handed, because the Alt-VPA method includes a penalty term $p_2$ [see Equation (S15) of the Supplementary Material] associated with variability about the relationship between fishing mortalities for the two oldest age

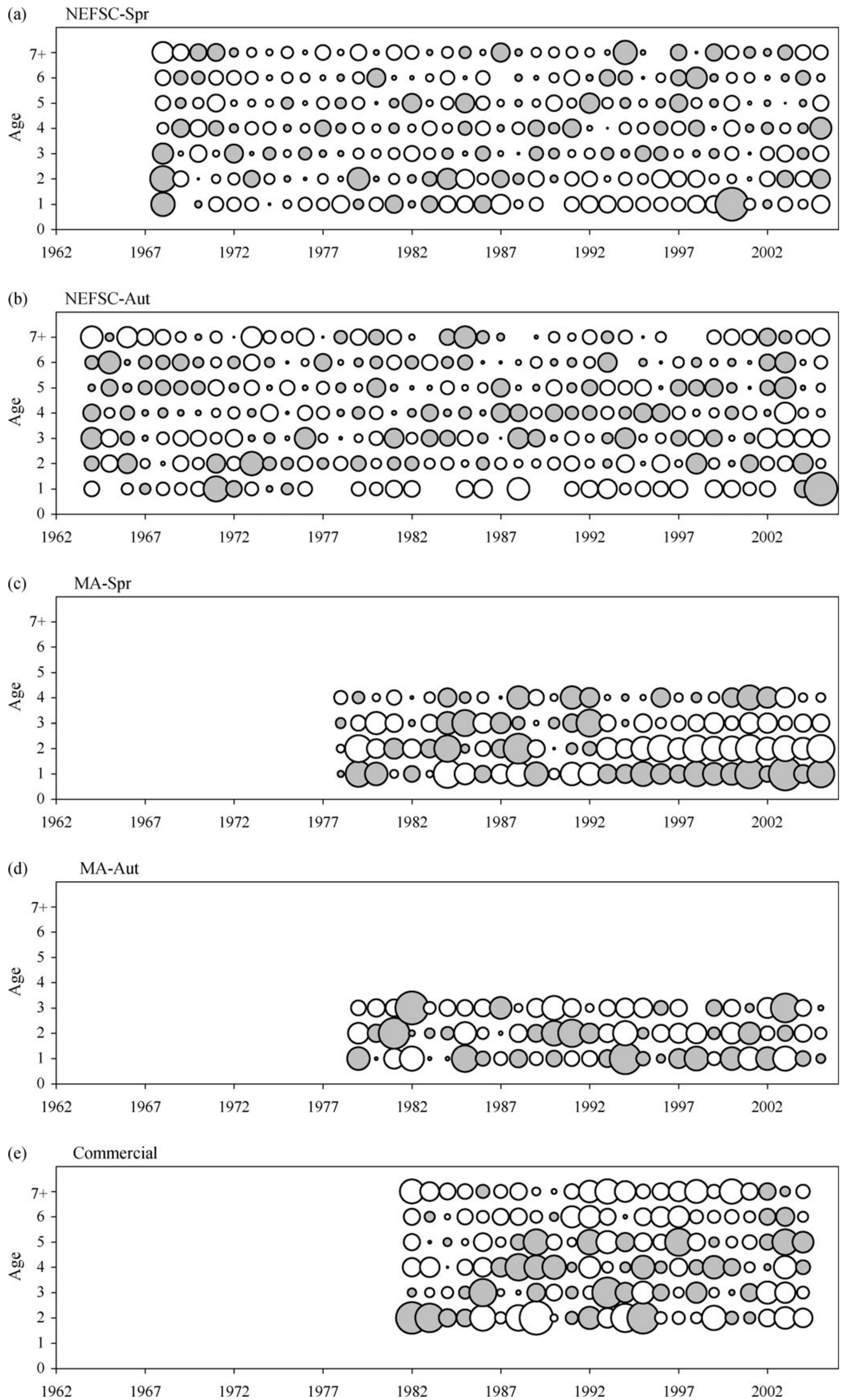


**Figure 9.** Bubble plots of the standardized residuals for the catch-at-age data for ASPM Sensitivity 5 (flat commercial and NEFSC survey selectivity for age 5+) assessment. The size (area) of the bubbles represents the size of the residuals. Grey bubbles represent positive residuals and white bubbles represent negative residuals.

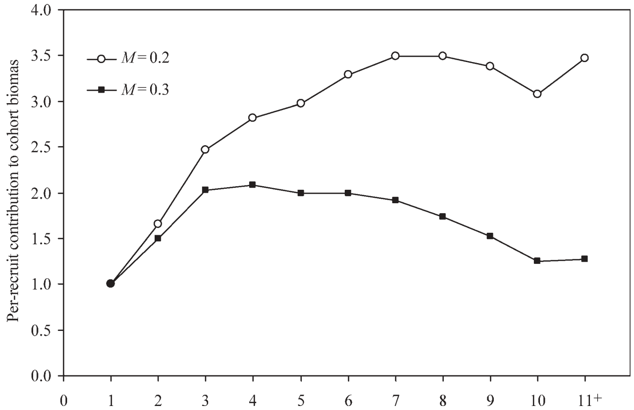


**Figure 10.** Per-recruit contributions by age to cohort biomass, taking natural mortality and somatic growth into account and expressed relative to the age 1 contribution, for $M = 0.2$ and $M = 0.3$. Start-of-the-year weights used are those for the last year available (2005). These contributions are evaluated as $w^{\mathrm{strt}}_{2005,a}e^{[-(a-1)M]}/w^{\mathrm{strt}}_{2005,1}$.

groups considered, whereas the corresponding relationship among ages 4–6 is forced to be exact for RC-VPA.

Compared with the ASPM results, the $B^{\mathrm{sp}}$ trajectory for the Alt-VPA ($\alpha = \mathrm{est}$) case is similar to that for RC-ASPM, particularly over the last 10 years. However, that is not an entirely appropriate comparison, because the fit of RC-ASPM also takes account of pre-1982 data. A better comparison is to ASPM-Sensitivity 7, which excludes these earlier data and results in $B^{\mathrm{sp}}$ values somewhat greater than those for Alt-VPA ($\alpha = \mathrm{est}$). Note that one would not expect exact agreement, because the ASPM takes account of dynamics within the plus-group, so average selectivity for the group as a whole changes over time because of the changing age structure within the group, whereas the ADAPT-VPA-based methods do not make allowance for this. Figure 12 compares some estimated fully selected fishing mortality time-series for the ASPM and ADAPT-VPA approaches; the latter are appreciably higher in some recent years.

Table 3 shows that only the combination of the asymptotically flat selectivity assumption and a Beverton–Holt spawning stock–recruitment relationship leads to VPA estimates showing a low value of $B^{\mathrm{sp}}_{2004}$ relative to $B^{\mathrm{sp}}_{\mathrm{MSY}}$ and a corresponding fishing mortality more than $F_{\mathrm{MSY}}$. If $\alpha$ is estimated, or a Ricker form assumed, $F_{2004}$ is consistently estimated to be well below $F_{\mathrm{MSY}}$, and all $B^{\mathrm{sp}}_{2004}$ relative to $B^{\mathrm{sp}}_{\mathrm{MSY}}$ estimates are well above the proxy 0.5 "overfished" threshold.

## Some broader issues

The multiple recent reviews of the assessments of US Northeast groundfish assessments, and particularly of the Gulf of Maine cod stock, have led to a variety of insights of broader pertinence.

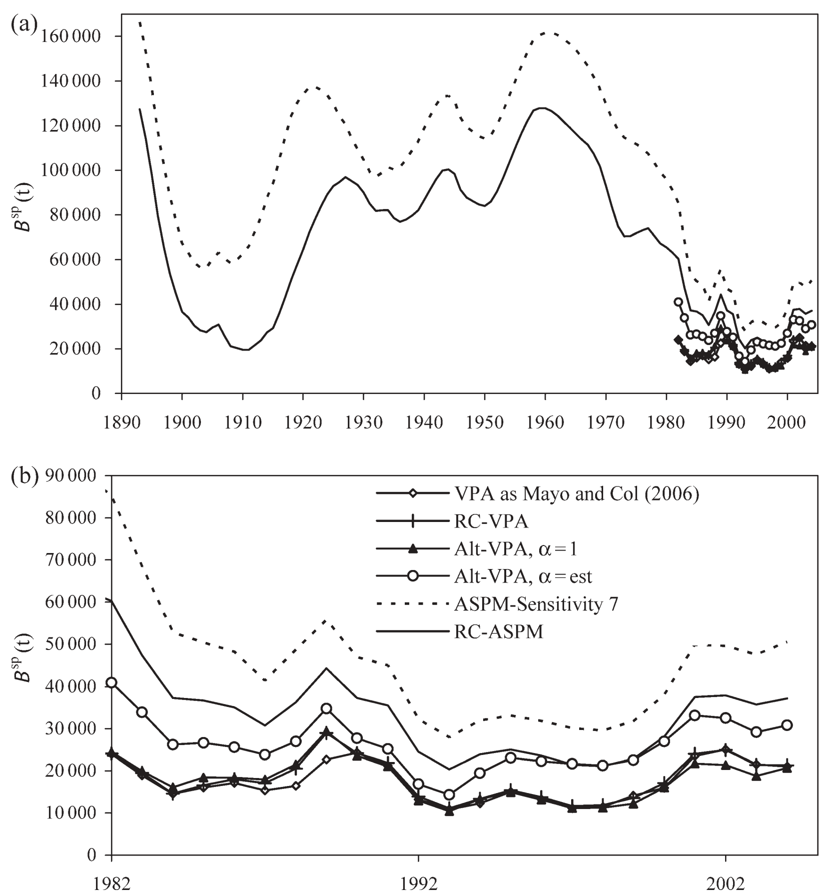


**Figure 11.** Time-series of spawning-biomass estimates for the VPA, as in Mayo and Col (2006), the RC-VPA, and two of the VPA sensitivities, as well as for RC-ASPM and ASPM-Sensitivity 7 (RC except exclude pre-1982 index data).

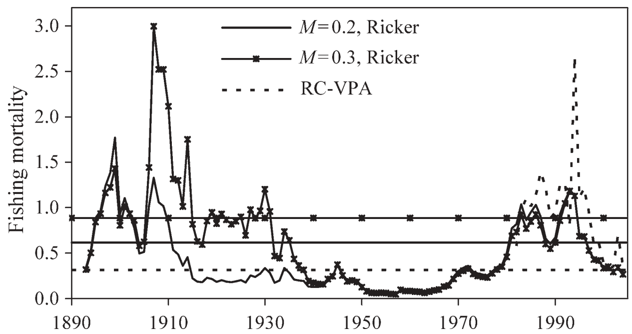


**Figure 12.** Comparison of fully selected fishing mortality trajectories for RC-ASPM and Sensitivity 2 (RC with $M = 0.3$) and the RC-VPA. Estimated $F_{MSY}$ values are shown by the flat straight lines. (Figure S1 of Section C of the Supplementary Material shows results for further sensitivities.)

## Period of data to consider in assessments

Perhaps the most interesting difference in views expressed by reviewers has related to how far back in time to incorporate data into an assessment. The NRC review (NRC, 1998) was unequivocal in querying the use of short time-series for assessments, stating that a longer-term view achieved through increased use of historical data was needed, and singling out the Gulf of Maine cod stock in this respect. Yet, some more recent reviews (Payne, 2003; NEFMC, 2005a, b) have appeared hesitant in this regard, expressing concerns about the necessary associated assumptions and the possibility of changes over time in underlying processes.

The two viewpoints seem to show some correlation with whether or not their exponents are closely involved in North Atlantic assessments. We posit that this may relate to personal experiences gained under circumstances of the great extent to which many North Atlantic stocks have been reduced, in contrast to the situation in some other areas. With highly depleted resources, the primary focus is to ensure that catch- or effort-related recommendations will lead to increased abundance. Therefore, using only data for more recent years to better ensure comparability of abundance indices and hence to obtain unbiased estimates of trend becomes paramount. Furthermore, a common solution to the presence of retrospective patterns in assessments of some of these stocks is to restrict the data used to tune VPA to more recent years only, so as to reduce such patterns and hence obtain short-term predictions of probably greater reliability; the fact that less recent data are inconsistent with such tuning is set aside. Continued high fishing mortalities mean that VPA-based estimates of abundance depend little on how these mortalities might be calculated for the oldest ages. In other circumstances, however, there tends to be a greater focus on medium-term targets, such as $B^{sp}_{MSY}$, and consequently greater emphasis on the use of longer time-series of data. For example, the International Whaling Commission's Revised Management Procedure stresses that account be taken of catch histories that extend as far back in time as possible (IWC, 1999).

## Estimation of stock–recruitment relationships

Should this be internal or external to the assessment, as in the ASPM or ADAPT-VPA approaches above, respectively? By the nature of its construction, ASPM must always involve internal estimation, whereas this could be external for SCAA implementations (although this might give rise to convergence difficulties for the SCAA). Proponents of the internal option will cite statistically self-consistent weighting of the various sources of information available. On the other hand, the external option insures against being misled by a possibly inappropriate choice of functional form for the relationship.

We suggest, instead, that the most important consideration is to check for any evidence of systematic lack-of-fit to both the stock–recruitment function and the various abundance indices and catch-at-age data. Given indications of such lack-of-fit, the internal option is not supportable; but in the absence of such indications, internal estimation seems to be the logical choice. Following simulation studies of the related question of estimating the effect of environmental factors on recruitment, Maunder and Watters (2003) conclude that the internal outperforms the external estimation approach, which can result in biased estimates when data are limited.

## The need to choose a "best" assessment

Some reviewers have queried whether there is a need to choose a "best" method, and hence to argue whether one method is better than the other, because all are approximations to reality, and the use of different methods adds value by providing different perspectives. Further, the merits of retaining the same method over time have been cited as a reason to maintain a methodological *status quo*.

The requirement for specific decisions concerning resource and fishing mortality levels under the National Standard Guidelines for the Magnuson–Stevens Act (e.g. Federal Register, 2005) would seem to necessitate prior agreement on decision rules, and hence on some associated "best" assessment (although this could be to take an average over a set of different assessments). Certainly, once an approach has been agreed upon, updates of management recommendations for the immediate future should be based on an unchanged approach so that impressions gained of resource trends upon which those recommendations would be based are not artefacts of methodological changes.

However, this should not be to the exclusion of medium-term review and possible change. The most important consideration is that the models used must be consistent with the available data (unless cogent reasons can be advanced to query the reliability of certain data and hence to exclude them from assessments).

## Simulation tests

Simulation testing has been suggested as a basis to resolve debate about the relative merits of ASPM and ADAPT-VPA for assessing the Gulf of Maine cod stock. However, the difficulty in such an approach is that the results will depend on the set of underlying realities chosen for inclusion in the simulation trials (e.g. Radomski *et al.*, 2005). Such an approach carries the overhead of first needing to get the debating parties around a table to attempt consensus agreement on trial specifications. This is a prerequisite to any possibility of obtaining a generally agreed interpretation of the results of such tests.

# Conclusions

At a more detailed level, three general observations that arise from the debates and analyses of Gulf of Maine cod stock data are:

(i) the need for care in consideration of and mathematically consistent treatment of the plus-group, particularly if there is the possibility of dome-shaped selectivity;

(ii) to err on the side of more flexible parameterizations of selectivity-at-age in SCAA/ASPM approaches, to avoid possibly misleading perceptions of the precision with which certain parameters (such as $M$) may be estimable; and

(iii) to take care when using the Beverton–Holt spawning stock–recruitment function, which will provide inappropriately low estimates of $B^{sp}_{MSY}$ if there is an overall negative trend in estimates of recruitment when plotted against those of $B^{sp}$.

More specific to the Gulf of Maine cod stock, important conclusions are:

(i) The primary reason for the differences shown in the results of the 2003 ADAPT-VPA and ASPM assessments shown in Table 1 that the CIE reviewers wanted understood (Payne, 2003) is that the former approach forced asymptotically flat selectivity, whereas the latter allowed this to be estimated from the data. The differences in question reduce substantially once this constraint on the ADAPT-VPA assessment is relaxed.

(ii) Population modelling indicates that the assumption of asymptotically flat selectivity is inconsistent with the available catch-at-age data. Either cogent reasons need to be advanced that current ageing of older cod is unreliable, or assessments based on the assumption of asymptotically flat selectivity must be rejected.

(iii) Once the constraint of asymptotically flat selectivity is relaxed, estimates of recent spawning biomass as a proportion of $B_{MSY}$ become substantially larger than the 27% of the NEFSC (2002) assessment that led to the classification of the stock as "overfished"/"depleted". Furthermore, perceptions that recent fishing mortality exceeds $F_{MSY}$ are reversed. These results hold for both ADAPT-VPA (Table 3) and over a wide range of sensitivities for ASPM (Table 2 and Table S1 in Section C of the Supplementary Material).

(iv) In circumstances where the implementation of the Magnuson–Stevens Act puts particular emphasis on the determination of $B^{sp}_{MSY}$, it is notable that (a) this benefits from a greater contrast in values of $B^{sp}$ by considering the post-1982 period alone (note from Figure 1, the much greater range of $B^{sp}$ values covered when this period is extended to post-1964), (b) precision of estimates of recent $B^{sp}$ relative to $B^{sp}_{MSY}$ is doubled through the inclusion of pre-1982 data in the estimation, and (c) concerns about possible changes in selectivity pre-1982 are offset by the NEFSC research survey series being perhaps the longest in the world still following the same methodology. Consequently, the ability of ASPM approaches to take pre-1982 data into account, unlike VPA, would seem to render the former preferable.

By way of a concluding note, it is again important to stress that this paper has focused on a question of *methodological* comparison. For that purpose, it accepts the summary data provided for the standard assessments of the Gulf of Maine stock as the best available and does not explore that aspect further. However, the RC-ASPM is not offered as conclusively the "best" possible assessment of the stock. Independent evidence should certainly be sought and considered for avoiding trawls or emigration that might give rise to the dome-shaped selectivity identified, so as to confirm the reliability of the enhanced stock status estimates that follow. Discussions are needed to determine which of the many options within the ASPM framework might be the best for the stock. Nevertheless, the broad inferences resulting from this work should facilitate the improvement of future assessments of this resource.

## Supplementary Material

The following supplementary material is available at *ICESJMS* online.

Section A provides details of the ADAPT-VPA models applied.

Section B provides details of the ASPM applied.

Section C includes results of some further sensitivity tests of the ASPM.

Section D details slight changes in the data analysed compared to those used by Mayo and Col (2006).

Four references not in the main text are cited in the Supplementary Material (Anon., 2003; Baranov, 1918; Pope, 1972; Walters and Ludwig, 1994), but are listed here for completeness.

## Acknowledgements

Éva Plagányi contributed to earlier development of this work. We thank various reviewers of earlier stages of this work for their comments, along with two anonymous reviewers of an earlier version of this manuscript, and Ewen Bell, Robin Cook, Chris Darby, Ray Hilborn, Murdoch McAllister, Bob Mohn, Andrew Payne, and scientists from the Northeast Fisheries Science Center and the New England Fishery Management Council's Scientific and Statistical Committee. Andre Punt and Mark Maunder also provided helpful comments. Financial support was provided by the Associated Fisheries of Maine and earlier by the Trawlers Survival Fund.

## References

Agger, P., Boëtius, I., and Lassen, H. 1971. On errors in the virtual population analysis. ICES Document CM 1971/H: 16. 10 pp.

Anon. 2003. VPA/ADAPT Version 2.0 Reference Manual. NOAA Fisheries Toolbox. http://nft.nefsc.noaa.gov.

Baranov, F. T. 1918. On the question of the dynamics of the fishing industry. Nauchnyi issledovatelskii iktiologisheskii Institut Izvestia, I: 81–128.

Bull, B., Francis, R. I. C. C., Dunn, A., McKenzie, A., Gilbert, D. J., and Smith, M. H. 2005. CASAL (C++ algorithmic stock assessment laboratory): CASAL User Manual v2.07-2005/08/21. NIWA Technical Report, 127. 272 pp.

Butterworth, D. S., Ianelli, J. N., and Hilborn, R. 2003a. A statistical model for stock assessment of southern bluefin tuna with temporal changes in selectivity. South African Journal of Marine Science, 25: 331–361.

Butterworth, D. S., and Punt, A. E. 1990. Some preliminary examinations of the potential information content of age-structure data from Antarctic minke whale research catches. Report of the International Whaling Commission, 40: 301–315.

Butterworth, D. S., Rademeyer, R. A., and Plagányi, É. E. 2003b. An age-structured production model based assessment and reference point evaluation for the Gulf of Maine cod stock. New England Fishery Management Council Document. 41 pp.

Butterworth, D. S., Rademeyer, R. A., and Plagányi, É. E. 2005. Report on an investigation of reasons for differences between outputs from ASPM and ADAPT-VPA assessment of the Gulf of Maine cod stock. New England Fishery Management Council Document. 4+98 pp.

Clark, W. G. 1999. Effects of an erroneous natural mortality rate on a simple age structured model. Canadian Journal of Fisheries and Aquatic Sciences, 56: 1721–1731.

Deriso, R. B., Quinn, T. J., and Neal, P. R. 1985. Catch-age analysis with auxiliary information. Canadian Journal of Fisheries and Aquatic Sciences, 42: 815–824.

Doubleday, W. G. 1976. A least squares approach to analyzing catch-at-age data. International Commission for Northwest Atlantic Fisheries, Research Bulletin, 12: 69–81.

Federal Register. 2005. Part II. Department of Commerce. National Oceanic and Atmospheric Administration. 50 CFR Port 600. Magnuson–Stevens Act Provisions; National Standard Guidelines; Proposed Rule 70 (119), Wednesday, June 22, 2005: 36240–36259.

Fournier, D. A., and Archibald, C. P. 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences, 39: 1195–1207.

Fournier, D. A., Hampton, J., and Sibert, J. R. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. Canadian Journal of Fisheries and Aquatic Sciences, 55: 2105–2116.

Gavaris, S. 1988. An adaptive framework for the estimation of population size. Canadian Atlantic Fisheries Scientific Advisory Committee, Research Document, 88/29. 12 pp.

Gulland, J. A. 1965. Estimation of growth and mortality rates. Annex to Arctic Fisheries Working Group Report, meeting in Hamburg, January, 1965. ICES Document CM 1965/Gadoid Fish Committee: 3. 9 pp.

Hilborn, R. 1990. Estimating the parameters of full age-structured models from catch and abundance data. International North Pacific Fisheries Commission Bulletin, 50: 207–213.

IWC. 1999. The Revised Management Procedure (RMP) for baleen whales. Journal of Cetacean Research Management, 1: 251–258.

Kimura, D. K., Balsiger, J. W., and Ito, D. H. 1984. Generalised stock reduction analysis. Canadian Journal of Fisheries and Aquatic Sciences, 41: 1325–1333.

Kimura, D. K., and Tagart, J. V. 1982. Stock reduction analysis, another solution to the catch equations. Canadian Journal of Fisheries and Aquatic Sciences, 39: 1467–1472.

Maunder, M. N., and Watters, G. M. 2003. A general framework for integrating environmental time series into stock assessment models: model description, simulation testing, and example. Fishery Bulletin US, 101: 89–99.

Mayo, R. K., and Col, L. A. 2006. The 2005 Assessment of the Gulf of Maine Atlantic Cod Stock. Northeast Fisheries Science Center, Reference Document, 06-02. 109 pp. http://www.nefsc.noaa.gov/nefsc/publications/crd/crd0602/.

Methot, R. D. 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. *In* Proceedings of the Symposium on Applications of Stock Assessment Techniques to Gadids, pp. 259–277. Ed. by L-L. Low. International North Pacific Fisheries Commission, Bulletin. 50 pp.

NEFMC. 2003. Amendment 13 to the Northeast Multispecies Fishery Management Plan. Related links: Final Amendment 13 – December 2003. http://www.nefmc.org/nemulti/index.html.

NEFMC. 2005a. Peer review of the Butterworth assessment of Gulf of Maine cod, as submitted by R. M. Cook, R. Hilborn, R. K. Mohn, and the Cefas Assessment and Modelling Team led by C. Darby, with report finalised by A. I. L. Payne and E. Bell. New England Fishery Management Council Documents: 8 pp., 8 pp., 11 pp., and 16 pp., respectively.

NEFMC. 2005b. NEFMC Scientific and Statistical Committee Report: evaluation of reports from the review panel concerning the Butterworth ASPM model for Gulf of Maine cod. New England Fishery Management Council document dated 9 June 2005. 1 pp.

NEFSC. 2002. Final Report of the Working Group on Re-Evaluation of Biological Reference Points for New England Groundfish. Northeast Fisheries Science Center, Reference Document, 02–04. 254 pp. http://www.nefsc.noaa.gov/nefsc/publications/crd/crd0204/.

NRC (National Research Council). 1998. Review of the Northeast Fishery Stock Assessments. The National Academic Press. 136 pp. http://books.nap.edu/catalog/6067.html.

Payne, A. I. L. (summariser). 2003. Report on the Groundfish Science Peer Review Meeting, 3–8 February 2003, University of New Hampshire, Durham. New England Fishery Management Council Document. 43 pp.

Pope, J. G. 1972. An investigation of the accuracy of virtual population analysis using cohort analysis. International Commission for the Northwest Atlantic Fisheries, Research Bulletin, 9: 65–74.

Pope, J. G. 1974. A possible alternative method to virtual population analysis for the calculation of fishing mortality from catch at age data. International Commission for Northwest Atlantic Fisheries, Annual Meeting June 1974. Research Document 74/20, Serial No. 3166. 16 pp.

Punt, A. E., and Hilborn, R. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. Reviews in Fish Biology and Fisheries, 7: 35–63.

Punt, A. E., Smith, A. D. M., and Cui, G. 2002. Evaluation of management tools for Australia's south east fishery. 2. How well can management quantities be estimated? Marine and Freshwater Research, 53: 631–644.

Radomski, P., Bence, J. R., and Quinn, T. J., II. 2005. Comparison of virtual population analysis and statistical kill-at-age analysis for a recreational, kill-dominated fishery. Canadian Journal of Fisheries and Aquatic Sciences, 62: 436–452.

Schaefer, M. B. 1957. A study of the dynamics of the fishery for yellowfin tuna in the eastern tropical Pacific Ocean. Inter-American Tropical Tuna Commission, Bulletin, 2: 247–285.

Schnute, J. T. 1994. A general framework for developing sequential fisheries models. Canadian Journal of Fisheries and Aquatic Sciences, 51: 1676–1688.

Thompson, G. G. 1994. Confounding of gear selectivity and the natural mortality rate in cases where the former is a nonmonotone function of age. Canadian Journal of Fisheries and Aquatic Sciences, 51: 2654–2664.

Walters, C. W., and Ludwig, D. 1994. Calculation of Bayes posterior probability distributions for key population parameters. Canadian Journal of Fisheries and Aquatic Sciences, 51: 713–722.

doi:10.1093/icesjms/fsn178