

# Model averaging to estimate rebuilding targets for overfished stocks

Jon Brodziak and Christopher M. Legault

**Abstract:** Reducing overfishing and recovering overfished fish stocks is a challenging and important global problem. Rebuilding targets are essential ingredients for guiding overfished stocks to recovery. Having robust estimates of rebuilding targets is likely a necessary condition for effective long-term management of fishery resources. In this paper, we show how Bayesian model averaging can be applied to estimate rebuilding targets under alternative hypotheses about stock–recruitment dynamics. Using alternative hypotheses about stock–recruitment dynamics and environmental variation is important because using only a single hypothesis can lead to substantially different reference points and policy implications. The alternative hypotheses are augmented with prior information collected from meta-analyses of stock–recruitment data to construct a set of age-structured production models. We illustrate our approach using three overfished New England groundfish stocks. We fit alternative model likelihoods to observed data using Bayesian inference techniques. The Schwarz goodness-of-fit criterion was used to calculate model probabilities. Bayesian model averaging was used to estimate rebuilding targets that were robust to model selection uncertainty. Model-averaged estimates suggested that rebuilding targets for overfished stocks can be reasonably well determined if adequate prior information on stock productivity is available. Nonetheless, results had wide confidence intervals that reflected the underlying uncertainty in rebuilding targets.

**Résumé :** La réduction de la surpêche et la récupération des stocks de poissons surexploités sont des problèmes et des défis importants au niveau de la planète. Les objectifs de récupération sont essentiels pour guider le rétablissement des stocks surexploités. La détermination d'estimations robustes des objectifs de récupération est vraisemblablement une condition nécessaire pour une gestion efficace à long terme des ressources halieutiques. Nous démontrons, dans notre étude, que le calcul de moyennes de modèles bayésiens peut servir à estimer les objectifs de récupération sous diverses hypothèses de rechange au sujet de la dynamique du recrutement du stock. Il est important d'utiliser diverses hypothèses de rechange concernant la dynamique du recrutement du stock et la variation environnementale parce que l'utilisation d'une seule hypothèse peut mener à l'établissement de points de référence et de politiques de gestion très différents. Les diverses hypothèses de rechange sont modifiées sur la base d'informations a priori obtenues de méta-analyses des données de stock–recrutement pour obtenir une série de modèles de production structurés en fonction de l'âge. Nous illustrons notre méthodologie en l'appliquant à trois stock surexploités de poissons de fond de la Nouvelle-Angleterre. Nous ajustons les vraisemblances diverses des modèles au données observées à l'aide de techniques d'inférence bayésienne. Le critère d'ajustement de Schwarz sert à calculer les probabilités des modèles. La détermination des moyennes des modèles par techniques bayésiennes permet d'établir des objectifs de récupération qui sont robustes malgré l'incertitude associée à la sélection des modèles. Les estimations basées sur les moyennes des modèles laissent croire qu'il est possible d'établir assez bien des objectifs de récupération pour les stocks surexploités s'il existe une information préalable adéquate sur la productivité des stocks. Néanmoins, les résultats ont de grands intervalles de confiance, ce qui reflète l'incertitude sous-jacente aux objectifs de récupération.

[Traduit par la Rédaction]

## Introduction

Worldwide, many fishery resources have been or are now being overfished (Garcia and Newton 1997; Pauly et al. 2002). Reducing overfishing and recovering overfished fish stocks is a challenging and important global problem (Hutchings et al. 1997; Rosenberg 2003). Biological rebuild-

ing targets are an essential ingredient for guiding overfished stocks to recovery, since they provide the biological basis for managing fisheries to eliminate overfishing. Having robust estimates of rebuilding targets is likely a necessary condition for effective long-term management of fishery resources.

Estimating biological rebuilding targets can be a challenging task, however, because multiple factors can affect fish population dynamics. Why did a particular fish stock collapse? Was it due to recruitment overfishing or changing environmental conditions or some combination of both? These are fundamental questions that need to be addressed in any approach to estimate rebuilding targets. In the ideal scenario where the underlying population dynamics conform to a single known model, standard likelihood-based approaches can be used to draw inferences about rebuilding targets. More

Received 10 February 2004. Accepted 9 October 2004.  
Published on the NRC Research Press Web site at  
<http://cjfas.nrc.ca> on 13 April 2005.  
J17968

**J. Brodziak<sup>1</sup> and C.M. Legault.** National Marine Fisheries Service, Northeast Fisheries Science Center, 166 Water Street, Woods Hole, MA 02543, USA.

<sup>1</sup>Corresponding author (e-mail: [jon.brodziak@noaa.gov](mailto:jon.brodziak@noaa.gov)).

often, however, uncertainty exists about the best approximating model. For example, the stock–recruitment relationship is highly variable for many stocks and uncertain given short time series of data. Similarly, changes in environmental conditions that coincide with decreases in recruitment may suggest serial correlation in stock–recruitment dynamics. Another uncertainty arises when time series of stock–recruitment observations are limited to a period when recruitment overfishing was occurring. As a result, the range of stock–recruitment data is truncated to reflect only low spawning stock sizes. This can lead to the perception that an overfished stock is less productive than it was in the past, the so-called shifting baseline syndrome (Pauly 1995). In this case, there may be insufficient contrast in spawning stock size to statistically estimate a stock–recruitment relationship without using auxiliary information to determine model parameters. One direct approach to addressing model uncertainty is to develop multiple working hypotheses (e.g., see Hilborn and Mangel 1997) and then apply model selection or model averaging for statistical inference (Hoeting et al. 1999; Burnham and Anderson 2002).

In this paper, we show how Bayesian model averaging can be applied to estimate rebuilding targets for overfished stocks under alternative hypotheses about stock–recruitment dynamics. Using alternative hypotheses about the stock–recruitment relationship and environmental variation is important because using a single hypothesis about stock–recruitment dynamics can lead to substantially different reference points and policy implications (see Myers et al. 1994). The alternative hypotheses are augmented with prior information collected from meta-analyses of stock–recruitment data to construct a set of age-structured production models (Shepherd 1982; Sissenwine and Shepherd 1987). We fit the alternative models to observed data using Bayesian inference techniques (e.g., see Ellison 2004). Markov Chain Monte Carlo (MCMC) simulation was used to calculate the posterior distribution of rebuilding targets. The Schwarz goodness-of-fit criterion was used to rank the alternative models and assign them probabilities for model averaging (Kass and Raftery 1995). When only one model is highly probable, model averaging will produce results that are similar to selecting a single model based on goodness-of-fit. In contrast, when several competing models have similar probabilities, selecting a single model may produce misleading inferences and underestimate uncertainty. To alleviate this possibility, Bayesian model averaging was applied (Hoeting et al. 1999) to estimate rebuilding targets that were robust to model selection uncertainty.

We illustrate our approach using three overfished groundfish stocks from Georges Bank: Atlantic cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*), and yellowtail flounder (*Pleuronectes ferrugineus*). We begin by describing the construction of alternative age-structured production models to estimate rebuilding targets of  $S_{MSY}$  and  $F_{MSY}$ , the spawning biomass and fishing mortality rate associated with maximum sustainable yield (MSY), respectively. These rebuilding targets are a necessary component of harvest control rules for US marine fisheries. They also provide useful biological limits under a precautionary approach to fisheries management (Mace 2001). We then describe our statistical approach to estimate model parameters using Bayesian model averaging

of objectively selected models. Model-averaged rebuilding target estimates are compared with existing values (Northeast Fisheries Science Center (NEFSC) 2002a). The model-averaged results suggest that rebuilding targets for these overfished stocks can be reasonably well determined despite model uncertainty if sufficient prior information on stock productivity is available. The sensitivity of estimated rebuilding targets to the selection of informative priors is also investigated and discussed.

## Materials and methods

### Stock–recruitment models

The stock–recruitment relationship is of fundamental importance for determining many biological rebuilding targets (Sissenwine and Shepherd 1987; Quinn and Deriso 1999; Williams and Shertzer 2003). In this study, Beverton–Holt (BH) and Ricker (RK) stock–recruitment models were chosen for constructing alternative MSY-based rebuilding target models. This choice provided two competing hypotheses about stock–recruitment dynamics: compensatory under the BH model and overcompensatory under the RK model. The flat-topped BH model is commonly used when fish survival at one or more early life history stages is density dependent. In contrast, the dome-shaped RK model is commonly used when spawner abundance has a negative impact on an early life history stage survival through cannibalism or other interactions. These models can produce substantially different biological reference points when applied to the same data set (Myers et al. 1994). We did not include depensatory stock–recruitment models because there was no clear evidence of depensation for the three overfished stocks under consideration. In recent years, spawning biomasses of both the Georges Bank haddock and yellowtail flounder stocks have increased at low fishing mortality rates, while cod biomass has remained stable owing to continued overfishing (NEFSC 2002b). As a result, both BH and RK models were used to describe expected equilibrium stock–recruitment conditions for the three Georges Bank stocks.

A modified BH curve was used for estimation:

$$(1) \quad R = \frac{4z_{MAX} R_{MAX} S}{S_{MAX}(1 - z_{MAX}) + S(5z_{MAX} - 1)}$$

where  $S_{MAX}$  is the maximum observed spawning biomass or a proxy estimate,  $R_{MAX}$  is the recruitment expected at  $S_{MAX}$ , and  $z_{MAX}$  is the steepness of the modified BH curve computed as the ratio of  $R$  at 20% of  $S_{MAX}$  to  $R_{MAX}$ . In this context,  $S_{MAX}$  is simply a fixed spawning biomass at which to compute  $R_{MAX}$ . This reparameterization allowed us to choose a reasonable maximum observed spawning biomass for curve fitting in contrast with the asymptotic form of the BH curve  $R = \alpha S / (\beta + S)$ , where  $R \rightarrow \alpha$  as  $S \rightarrow \infty$ . Parameters of the asymptotic curve are related to those of the modified curve by

$$(2) \quad \alpha = \frac{4z_{MAX} R_{MAX}}{5z_{MAX} - 1} \quad \text{and} \quad \beta = \frac{S_{MAX}(1 - z_{MAX})}{5z_{MAX} - 1}$$

Given  $\alpha$  and  $\beta$ , equilibrium spawning biomass as a function of fishing mortality ( $F$ ) for the BH curve ( $S_{BH}$ ) was

$$(3) \quad S_{BH} = \alpha \text{SPR}(F) - \beta$$

Values of  $S_{MAX}$ , the reference spawning biomass for the parameter  $R_{MAX}$ , were computed from available assessment data for each stock. Assessment estimates of spawning biomass and recruitment are available for Georges Bank cod since 1978 and for Georges Bank yellowtail flounder since 1973. Since both stocks were relatively abundant during the 1960s, as evidenced by time series of mean biomass per tow from the NEFSC spring and autumn surveys and fishery yields (NEFSC 2002a), we computed values of  $S_{MAX}$  using assessment estimates and earlier survey data. In particular, values of  $S_{MAX}$  were computed as the product of average spawning biomass from stock assessments times the ratio of average NEFSC autumn survey biomass indices during 1963–1970 to the same average during 1990–1999 (NEFSC 2002a). For Georges Bank cod,  $S_{MAX} = 104\,200$  mt, while for yellowtail flounder,  $S_{MAX} = 36\,200$  mt. Since Georges Bank haddock had a long time series of assessment estimates, we set  $S_{MAX}$  to be the maximum spawning biomass value in the stock–recruitment time series and  $S_{MAX} = 199\,500$  mt for haddock.

The form of the RK model used for parameter estimation was

$$(4) \quad R = a S e^{-bS}$$

where  $a$  is the slope at the origin and  $b$  is the strength of density dependence in the relationship. Given  $a$  and  $b$ , equilibrium spawning biomass as a function of  $F$  for the RK curve ( $S_{RK}$ ) was

$$(5) \quad S_{RK} = \frac{1}{b} [\log(a \text{ SPR}(F))]$$

The rebuilding target for the three overfished stocks was defined to be the biomass that produces MSY ( $S_{MSY}$ ), as mandated by the US Sustainable Fisheries Act of 1996. MSY is typically interpreted to be the long-term average yield that can be taken from a stock under existing environmental conditions. In this context, uncorrelated errors may be an appropriate for estimating rebuilding targets if environmental factors affect annual stock productivity in a random fashion. Alternatively, serially correlated errors may be appropriate if stock productivity is affected by low-frequency environmental forcing. We investigate both possible error structures below.

In this application, we chose to only model uncertainty in the stock–recruitment relationship. Other potential sources of uncertainty, such as variation in life history parameters, that might affect equilibrium spawning biomass per recruit or yield per recruit calculations (Appendix A) were not included, since recruitment largely determines the magnitude of  $S_{MSY}$  under the Sissenwine–Shepherd model (Sissenwine and Shepherd 1987). To see this, observe that  $S_{MSY}$  is the product of the spawning biomass per recruit at  $F_{MSY}$  ( $\text{SPR}(F_{MSY})$ ) and recruitment that produces MSY ( $R_{MSY}$ ):

$$(6) \quad S_{MSY} = \text{SPR}(F_{MSY}) R_{MSY}$$

Next, observe that the variance of any equilibrium spawning biomass ( $S$ ) can be approximated to first order (see Appendix A) as a function of spawning biomass per recruit (SPR) and recruitment ( $R$ ) by

$$(7) \quad \text{VAR}[S] \approx \left[ \frac{\partial S}{\partial R} \right]^2 \sigma_R^2 + \left[ \frac{\partial S}{\partial \text{SPR}} \right]^2 \sigma_{\text{SPR}}^2 \\ = (\text{CV}[R]^2 + \text{CV}[\text{SPR}]^2)(R \cdot \text{SPR})^2$$

where  $\text{CV}[R]$  and  $\text{CV}[\text{SPR}]$  are the coefficients of variation for recruitment and spawning biomass per recruit, respectively. If the coefficient of variation of the recruitment process is much larger than the coefficient of variation for spawning biomass per recruit, then the uncertainty due to recruitment dominates the variance in equilibrium spawning biomass and in particular  $S_{MSY}$ . This is a likely scenario for many marine fish stocks that have highly variable recruitment but relatively stable life history parameters and fishery selectivity parameters. Thus, modeling uncertainty in recruitment strength is a parsimonious choice for estimating rebuilding targets.

### Error structures

Error structures are a fundamental component of our approach because they represent how one expects stock–recruitment data to vary about the expected relationship (see Likelihoods below). Observation errors for alternative stock–recruitment models were assumed to be multiplicative lognormal distributions. Multiplicative lognormal errors were used because this positively skewed distribution arises naturally when groundfish survival rates during early life history are affected by numerous independent random events represented as multiplicative log-scale effects. As the number of random events becomes large, the distribution of the mean of the log-scale multiplicative process approaches a normal random variable under the central limit theorem.

Observation errors for the stock–recruitment models were either a multiplicative lognormal or an autoregressive, multiplicative lognormal error structure with a lag of 1 year. The uncorrelated error term represented the common assumption that environmental forcing was independent through time. In contrast, the autoregressive error term represented serial correlation in random environmental forcing. The error terms were multiplied by the expected stock–recruitment curve, denoted as  $f(S_i)$  for the  $i$ th data point, to obtain the observation error model

$$(8) \quad R_i = f(S_i) e^{\varepsilon_i}$$

For uncorrelated errors, the random shocks  $\varepsilon_i$  were independent and identically distributed Gaussian random variables with mean equal to  $-\sigma^2/2$  and variance equal to  $\sigma^2$ . In this case, the mean was chosen to set the expected value of the multiplicative error term to unity and the error variance ( $\sigma^2$ ) was a parameter to be estimated. For autoregressive lag 1 errors, the random shocks  $\varepsilon_i$  were distributed as

$$(9) \quad \varepsilon_i = \phi \varepsilon_{i-1} + w_i$$

where  $|\phi| < 1$  and

$$w_i \sim N\left(-\frac{\sigma_w^2}{2}, \sigma_w^2\right)$$

In this case, the expected value of the multiplicative error term was unity and the autoregressive coefficient ( $\phi$ ) and error variance were additional parameters to be estimated.

**Likelihoods**

Likelihoods of the observed stock–recruitment data were constructed for the BH and RK models using the two com-

peting hypotheses for error structures. The sample likelihood was conditioned on the assumed prior distributions (see Prior information below) and stock–recruitment model. Based on a total of  $n$  stock–recruitment data points ( $R_i, S_i$ ), the log-likelihood ( $\log L_{SR}$ ), using uncorrelated errors, was (Brodziak et al. 2001)

$$(10) \quad \log L(\theta, \sigma^2) = -\frac{n}{2} \log(2\pi) - n \log \sigma - \sum_{i=1}^n \log R_i - \frac{1}{2\sigma^2} \sum_{i=1}^n \left( \log R_i - \log f(S_i) + \frac{\sigma^2}{2} \right)^2$$

Similarly, the log-likelihood for serially correlated errors (for a derivation, see Seber and Wild (1989) with  $\sigma_w^2 = (1 - \phi^2)\sigma^2$  was

$$(11) \quad \log L_{SR}(\theta, \sigma^2, \phi) = -\frac{n}{2} \log(2\pi) - n \log \sigma_w - \sum_{i=1}^n \log R_i + \frac{1}{2} \log(1 - \phi^2) - \frac{1}{2\sigma_w^2} \sum_{i=2}^n \left( \log R_i - \phi \log R_{i-1} - \log f(S_i) + \phi \log f(S_{i-1}) + (1 - \phi) \frac{\sigma_w^2}{2} \right)^2 - \frac{(1 - \phi^2)}{2\sigma_w^2} \left( \log R_1 - \log f(S_1) + \frac{\sigma_w^2}{2} \right)^2$$

**Prior information**

Using prior information is a key component of our approach to estimating rebuilding targets. In this context, prior information represents the accumulated scientific knowledge of the likely stock–recruitment dynamics of the three stocks. We chose to estimate rebuilding targets assuming either the absence or presence of such knowledge. To represent an absence of knowledge about the likely stock–recruitment dynamics, uninformative priors for stock–recruitment model parameters were chosen before inspecting the data. For each stock, we used very broad uniform prior distributions ( $\pi$ ) for the square root of the error variance  $\sigma$ , autoregressive parameter  $\phi$ , BH parameters  $\log(R_0)$  and  $z$  (defined below), and RK parameters  $\log(a)$  and  $b$ :

$$(12) \quad \begin{aligned} \pi_\sigma &\sim U[0.0001, 4], & \pi_\phi &\sim U[-1, 1] \\ \pi_{\log(R_0)} &\sim U[-10, 100], & \pi_z &\sim U[0.20001, 1] \\ \pi_{\log(a)} &\sim U[-10, 10], & \pi_b &\sim U[-1\,000\,000, 0] \end{aligned}$$

Inferences about rebuilding targets using the uninformative priors were, in effect, based solely on the likelihood function.

In contrast, informative priors represented accumulated scientific knowledge of the likely stock–recruitment dynamics of the three stocks. Based on previous analyses (Brodziak et al. 2001; Brodziak 2002), it was recognized that there may be insufficient contrast in the stock–recruitment data to estimate either the steepness and unfished recruitment parameters of the BH curve or the slope at the origin of the RK curve. To resolve the issue of estimability, we borrowed from the strength of meta-analyses of numerous fish populations (Hilborn and Liermann 1998; Myers and Mertz 1998) to determine these parameters in an empirical Bayesian estimation framework (Gelman et al. 1995; Carlin and Louis 2000). In particular, we developed informative priors for steepness, slope at the origin, and unfished recruitment for each stock.

To include auxiliary information on the likely values of the steepness parameter for each of the three stocks, results

from the Myers et al. (1999) meta-analysis of a large number of stock–recruitment data sets were used to determine appropriate prior distributions for the steepness parameter ( $z$ ). These priors provided auxiliary information on the likely rate of increase in expected recruitment as spawning stock increases from low levels. Steepness was calculated relative to the unfished spawning stock size ( $S_0$ ) and was related to unfished equilibrium recruitment ( $R_0$ ) and the parameters of the asymptotic BH curve by

$$(13) \quad z = \frac{\alpha S_0 0.2}{(\beta + S_0 0.2) R_0}$$

In general, values of  $S_0$  and  $R_0$  differed from  $S_{MAX}$  and  $R_{MAX}$ , since unfished stock size was not necessarily equal to the value of  $S_{MAX}$  as estimated from assessment data.

For each species, the prior on steepness for the BH curve was based on value of  $z$  reported for that species in Myers et al. (1999, table 1). The informative prior was assumed to be distributed as a Gaussian random variable with the constraint that  $z \in [0.2, 1]$  so that the logarithm of the prior on steepness ( $\pi_z$ ) was

$$(14) \quad -\log \pi_z = 0.5 \log(2\pi) + \log(\sigma_z) + \frac{(z - \mu_z)^2}{2\sigma_z^2}$$

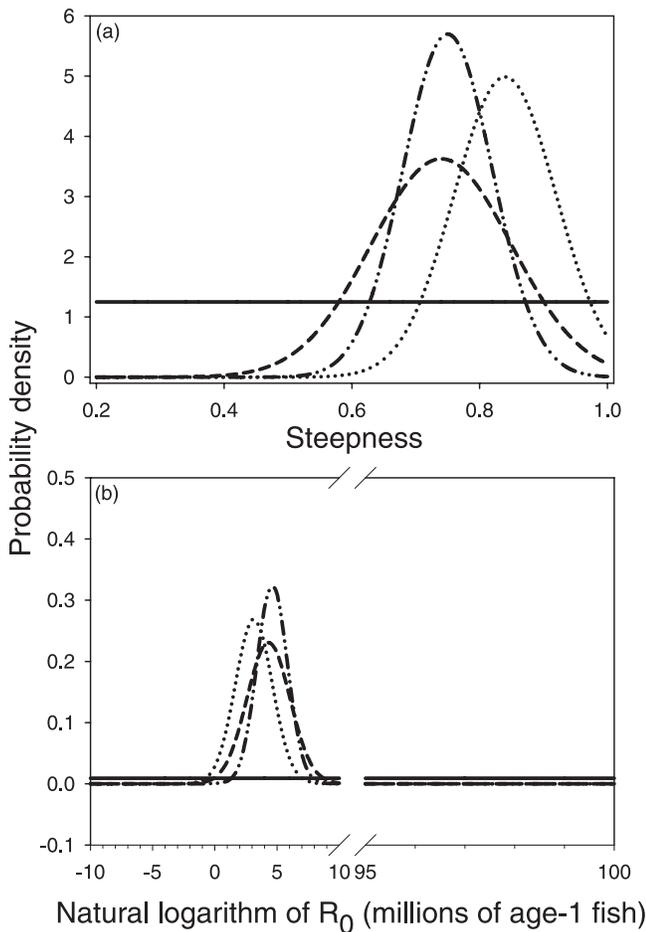
The mean of the informative prior ( $\mu_z$ ) was taken to be the median point estimate for the species. The standard error of the informative prior ( $\sigma_z$ ) was computed from the upper and lower values of the 60% confidence interval for steepness from Myers et al. (1999) and the assumption that the steepness was normally distributed. This produced informative priors for the steepness of Atlantic cod, haddock, and yellowtail flounder (Table 1; Fig. 1a).

We used the informative prior distributions for steepness to parameterize prior distributions for the mean slope at the origin parameter ( $a$ ) of the RK model. This choice ensured that comparable assumptions were made for slope at the ori-

**Table 1.** Values of the mean and standard error of informative prior distributions for steepness, slope at the origin, and unfished recruitment.

Species	Steepness		Slope		Unfished recruitment ( $\times 1\ 000\ 000$ )		
	Mean	SE	Mean	SE	Mean	SE	Recruitment data
Atlantic cod	0.84	0.08	0.78	0.07	23.2	4.4	Top quartile of spawning biomass, 1978–2001
Haddock	0.74	0.11	1.25	0.19	75.2	5.6	1931–1960
Yellowtail flounder	0.75	0.07	4.39	0.41	101.2	3.4	Top quartile of spawning biomass, 1963–1998

**Fig. 1.** Uninformative (solid line) and informative priors for (a) stock–recruitment steepness and (b) the natural logarithm of unfished recruitment of Georges Bank Atlantic cod (dotted line), haddock (dashed line), and yellowtail flounder (dotted–dashed line).



gin and steepness prior distributions under the alternative stock–recruitment hypotheses. The informative prior ( $\pi_a$ ) for the slope at the origin parameter ( $a$ ) of the RK curve was also assumed to be distributed as a normal random variable. For each stock, the mean of the informative prior on the slope at the origin was calculated from the mean of the informative prior for steepness. The relationship between steepness and slope was (see Myers et al. 1999)

$$(15) \quad a = \frac{4z}{(1-z) \text{SPR}(0)}$$

The standard error of the informative prior ( $\sigma_a$ ) was set by assuming that the informative priors for  $z$  and  $a$  had equal coefficients of variation. This determined the parameters of

the informative priors for the slope at the origin of Atlantic cod, haddock, and yellowtail flounder (Table 1).

It was also recognized that there was limited information on unfished recruitment because the stock–recruitment time series were short in comparison with the historic periods of exploitation. For example, Georges Bank cod has been fished since the 1600s but the available stock–recruitment time series begins in 1978. Further, all three stocks were overfished throughout much of the stock–recruitment time series with many observations at relatively low spawning stock sizes. To alleviate potential bias due to the lack of observations at high spawning stock sizes, we used an empirical Bayesian approach (Carlin and Louis 2000) to set an informative prior based on the distribution of unfished recruitment. The informative prior for unfished recruitment ( $\pi_{R_0}$ ) was also assumed to be normally distributed with the constraint that  $R_0 > 0$ . The mean ( $\mu_R$ ) and standard error ( $\sigma_R$ ) of the informative prior for  $R_0$  were determined using functional statistics of the empirical distribution of recruitment at high spawning stock size. In this case, we used what a panel of experts had previously judged to be an appropriate subset of observed stock–recruitment data (NEFSC 2002a). For Georges Bank haddock, recruitment values during 1931–1960 were used to determine the prior parameters (Table 1; Fig. 1b). This corresponded to the period of high fishery yields prior to the stock collapse in the early 1970s. For Georges Bank cod, recruitment data for spawning stock sizes in the top quartile of the spawning biomass distribution were used to determine the prior parameters. For Georges Bank yellowtail flounder, recruitment data for spawning biomass sizes in the top quartile of the hindcast spawning biomass distribution from Brodziak et al. (2001) were used to determine the prior parameters.

Informative priors were not developed for the density-dependence parameter  $b$  of the RK curve because there was no auxiliary information on the spawning stock size at which overcompensation would be expected to occur.

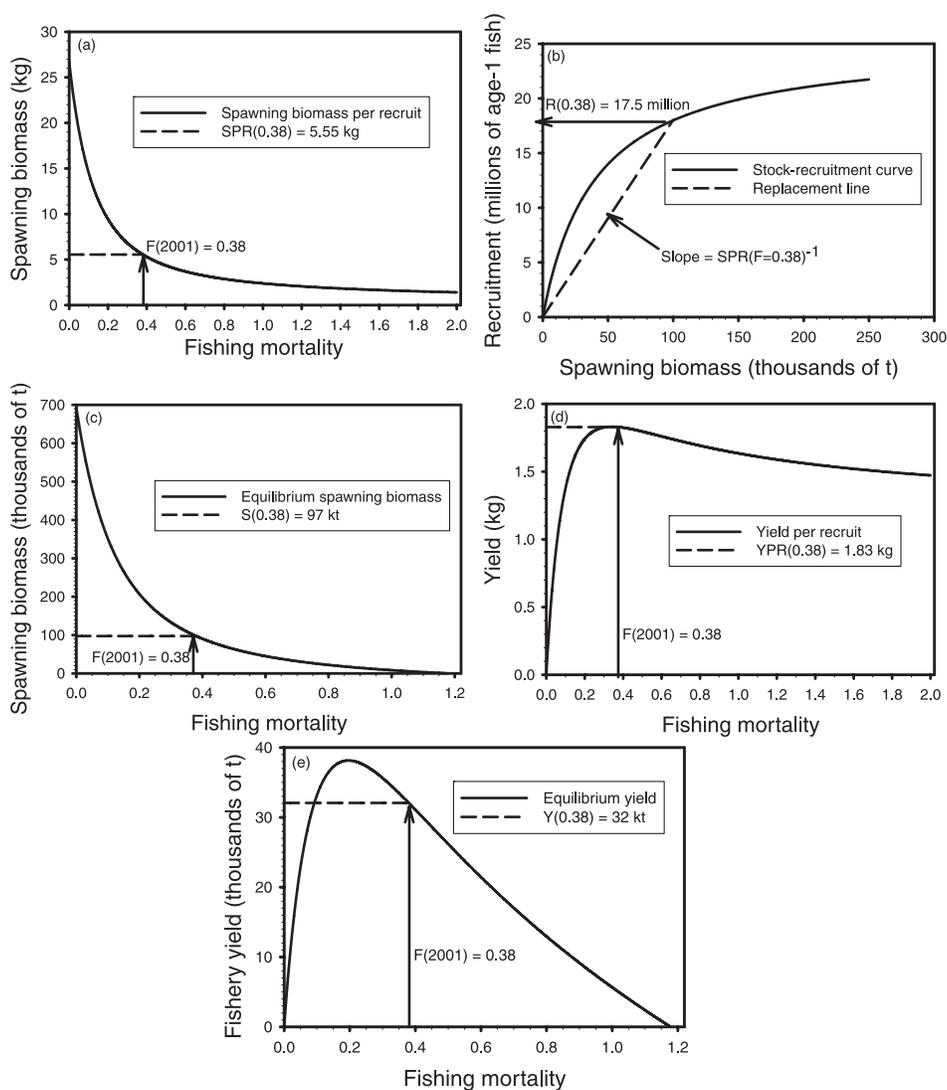
**Alternative models**

We constructed a set of 12 alternative probability models for estimating rebuilding targets using the stock–recruitment models, error structures, and assumptions about prior information. Each alternative represented a discrete hypothesis about stock dynamics and parameter estimability. The alternatives were formed as a Cartesian product of stock–recruitment curve, error structure, and assumptions about informative priors. For the BH curve, this led to eight alternative rebuilding target models (Table 2). For the RK curve, this led to four alternatives (Table 2). Under each of the alternative models, MSY-based rebuilding targets were calculated using the Sissenwine–Shepherd approach (Appendix A).

**Table 2.** Abbreviations of alternative rebuilding target models (model components indicated with a cross).

Model component	Model abbreviation											
	BH	ABH	RBH	RABH	ZBH	ZABH	RZBH	RZABH	RK	ARK	SRK	SARK
<b>Stock–recruitment curve</b>												
Beverton–Holt	×	×	×	×	×	×	×	×				
Ricker									×	×	×	×
<b>Error structure</b>												
Uncorrelated	×		×		×		×		×		×	
Correlated		×		×		×		×		×		×
<b>Informative prior</b>												
Unfished recruitment			×	×				×	×			
Steepness					×	×	×	×				
Slope at origin											×	×

**Fig. 2.** Schematic depiction of construction of (a) spawner per recruit, (b) recruit per spawner, (c) equilibrium spawning biomass, (d) yield per recruit, and (e) equilibrium yield curves using Georges Bank Atlantic cod as an example. Metric tons, t.



To illustrate the Sissenwine–Shepherd approach to computing rebuilding targets, consider the spawning biomass per recruit curve for Georges Bank cod (Fig. 2a) constructed using life history and fishery parameters in NEFSC (2002b). The point estimate of fishing mortality on this Atlantic cod stock

in 2001 was  $F_{2001} = 0.38$ . At this value of  $F$ , the  $SPR(0.38)$  is about  $5.55 \text{ kg-recruit}^{-1}$ . The inverse of this  $SPR$  value is the slope of the stock replacement line at equilibrium (Fig. 2b). The intersection of the replacement line and the stock–recruitment curve is the equilibrium recruitment,  $R(0.38)$ ,

which is roughly 17.5 million age-1 recruits at  $F_{2001}$ . Equilibrium recruitment  $\times$  spawning biomass per recruit equals equilibrium spawning biomass  $S(0.38) = R(0.38)SPR(0.38) = 17.5$  million recruits  $\times 5.55$  kg-recruit $^{-1} \approx 97\,000$  metric tons (t) (Fig. 2c). Yield per recruit at an  $F = 0.38$  is roughly 1.83 kg-recruit $^{-1}$  (Fig. 2d). Multiplying equilibrium recruitment and yield per recruit (YPR) equals equilibrium yield,  $Y(F)$ . In this case,  $Y(0.38) = R(0.38)YPR(0.38) = 17.5$  million recruits  $\times 1.83$  kg-recruit $^{-1} \approx 32\,000$  t (Fig. 2e). Repeating this calculation over a range of feasible values, one can numerically determine the  $F$  that maximizes equilibrium yield; this is  $F_{MSY}$ .

### Parameter estimation

Given the model likelihoods, priors, and observed stock–recruitment data, posterior distributions of rebuilding targets for statistical inference were calculated using standard numerical methods. We used a likelihood-based Bayesian simulation approach to compute model-averaged estimates of rebuilding targets. The AD Model Builder software package (Schnute et al. 1998; Otter Research Ltd. 2001) was used for maximizing the logarithm of the joint posterior distribution ( $L$ ), where  $L_{SR}$  denotes the sample likelihood and  $\Pi$  denotes the joint prior distribution comprising uninformative or informative priors:

$$(16) \quad \log L = \log L_{SR} + \log \Pi$$

Samples from the joint posterior distribution of rebuilding target estimates were generated in three steps. The first step was to compute most likely values (posterior modes) of model parameters for each of the 12 stock–recruitment models. The second step was to apply MCMC simulation to generate sequences of samples from the joint posterior distribution of parameters (Gilks et al. 1996). The AD Model Builder software performs MCMC sampling of model parameters using the Metropolis–Hastings algorithm (Gelman et al. 1995). In this implementation, the jumping kernel that determines the value of the next posterior sample has the same covariance matrix as the target (posterior) distribution. The covariance matrix is estimated from the observed information matrix evaluated at the posterior mode. For models that used uninformative prior distributions, the use of MCMC simulation was analogous to the use of a parametric bootstrap to estimate the sampling distribution of the rebuilding targets. We ran two MCMC simulations (chains) of length 300 000 for each model using different random seeds. Two chains were needed to assess whether the simulated values converged to the posterior distribution. Each chain was thinned by selecting every 50th sample to eliminate the possibility of auto-correlation. Of the 6000 samples remaining after thinning, we excluded the first 1000 to burn-in the chain, which effectively eliminated any dependence on the initial condition (the most likely estimates of the rebuilding targets). This left 5000 simulated posterior samples for inference. Convergence of the MCMC chains was verified for each rebuilding target ( $S_{MSY}$ ,  $F_{MSY}$ , and  $MSY$ ) using a convergence diagnostic called the potential scale reduction factor (PSRF) (Gelman et al. 1995) and further verified using another diagnostic called the second-order central-moment reduction. The latter diagnostic does not require a normality assumption (Brooks and Gelman 1998). In practice, a PSRF value

$<1.2$  provides confirmatory evidence that the chains have converged to the posterior distribution (Gelman et al. 1995).

The last step was to examine the results for credibility. In this case, three criteria were applied to determine whether the likelihood fits were consistent with auxiliary information and with respect to model goodness-of-fit measures. These criteria were used as a quality control check to ensure that the individual model outputs made sense (e.g., Schnute and Richards 2001). Although this required subjective judgment, it was necessary given the short time series of stock–recruitment data. A priori, we required that the estimated posterior mode from each model fit satisfy both the first- and second-order derivative conditions required for a strict maximum of  $L$ . Second, posterior samples with  $F_{MSY}$  values  $>2$  (81% exploitation rate with natural mortality of  $M = 0.2$ ) were not considered credible given the likely range of spawning biomass per recruit required for long-term sustainability (Mace and Sissenwine 1993). Last, we excluded models where the coefficient of variation of  $MSY$ ,  $S_{MSY}$ , or  $F_{MSY}$  derived from posterior sampling exceeded 100%. In this case, an extremely high coefficient of variation indicated that one or more model parameters were not identifiable. The remaining models formed the set of credible models.

### Goodness-of-fit criterion

We evaluated the relative likelihood of credible models using the Schwarz criterion (Schwarz 1978) (minus twice the Schwarz criterion is also called the Bayesian information criterion). This goodness-of-fit criterion allows one to compare the fits of alternative models to a data set. It is conceptually similar to the use of a multiple coefficient of determination to select a set of best predictor variables in multivariate regression. In our application, we used the Schwarz criterion to compare the fits of BH and RK models to stock–recruitment data. In general, if one wants to compare two alternative models  $M_j$  and  $M_k$  with parameter vectors  $\theta_j$  and  $\theta_k$  of dimension  $d_j$  and  $d_k$  using fixed data ( $D$ ) with sample size  $n$ , the Schwarz criterion ( $SC_{jk}$ ) is

$$(17) \quad SC_{jk} = \log(L_{SR}(D|\hat{\theta}_j, M_j)) - \log(L_{SR}(D|\hat{\theta}_k, M_k)) - \frac{1}{2}(d_j - d_k) \log(n)$$

where  $\hat{\theta}_j$  and  $\hat{\theta}_k$  are the most likely values of the parameter vectors. The model with the greatest likelihood, as adjusted by a parameter penalty, provides the best fit to the data. For example, in eq. 17, a negative  $SC_{jk}$  value would indicate that model  $M_k$  provided a better fit because it had a larger log-likelihood adjusted by the parameter penalty than model  $M_j$ . Using the Schwarz criterion provided a single objective measure to judge each model's goodness-of-fit to the data.

The Schwarz criterion is related to Bayesian inference with several models. The Schwarz criterion is approximately equal to the natural logarithm of the Bayes factor ( $B_{jk}$ ) (Kass and Raftery 1995). The Bayes factor provides a summary of the weight of evidence in favor of one scientific theory to another for a fixed data set. In particular, the Bayes factor is the ratio of the posterior odds of a hypothesis to its prior odds. When used for comparing competing models where each model has equal prior odds, the Bayes factor represents

the relative odds of one model versus another being true. One can approximate the Bayes factor for comparing alternative models  $M_j$  and  $M_k$  as

$$(18) \quad B_{jk} = \frac{\Pr(D|M_j)}{\Pr(D|M_k)} \approx \exp(\text{SC}_{jk})$$

where  $\Pr(D|M_k)$  is the integrated likelihood of  $D$  under model  $M_k$ .

The Bayes factor measures the relative success of alternative models at predicting the observed data. The approximation in eq. 18 gives a reliable, if not exact (the value of  $\text{SC}_{jk}/\log(B_{jk})$  approaches unity as  $n$  becomes arbitrarily large), measure of the weight of evidence in favor of one model versus another conditioned on the observed data (Kass and Raftery 1995). The strengths of the approximate Bayes factor are that it is easy to use and does not depend on the evaluation of prior distributions. In this context, minus twice the natural logarithm of the Bayes factor ( $-2\log(B_{jk})$ ) measures the strength of evidence of model  $M_j$  versus model  $M_k$  on the same scale as a deviance or a likelihood ratio statistic. The quantity  $-2\log(B_{jk})$  can be interpreted as the strength of evidence against model  $M_k$  (also known as the Bayesian information criterion; Kass and Raftery 1995): values of  $-2\log(B_{jk})$  between 0 and 2 provide no evidence against  $M_k$ , values between 2 and 6 provide positive evidence against  $M_k$ , and values between 6 and 10 and  $>10$  provide strong and very strong evidence against  $M_k$ , respectively. Given that it is the relative difference that determines the weight of evidence and not the actual value, we used the Schwarz criterion approximation to compute Bayes factors for all pairs of credible models.

### Model probabilities

The Bayes factors of the credible models were used to rank the models and compute their relative probability conditioned on the available data. The Bayes factor and prior odds of model  $M_k$  were calculated relative to the best fitting credible model ( $M_0$ ). In general, the prior odds ( $\alpha_k$ ) of  $M_k$  would be  $\alpha_k = \Pr(M_k)/\Pr(M_0)$ . In practice, we set the prior odds to be identically unity for all credible models, since there was a priori no reason to favor any one of the credible models. This choice implied that the relative model probabilities were determined by their approximate Bayes factors, which in turn depended on the likelihood functions. The posterior probability of each credible model was determined by its prior odds and Bayes factors via

$$(19) \quad \Pr(M_k|D) = \frac{\alpha_k B_{k0}}{\sum_j \alpha_j B_{j0}}$$

As a sensitivity analysis, the Akaike information criterion, was also used to compute model probabilities (see Brodziak et al. 2001; Burnham and Anderson 2002) and model-averaged rebuilding targets under the same numerical simulation approach.

### Model averaging

The posterior distributions from the set of credible models were used to calculate robust rebuilding target estimates us-

ing Bayesian model averaging (Kass and Raftery 1995; Hoeting et al. 1999). Under this approach, averaging over all credible models can typically provide better predictive accuracy than using a single best-fitting model. The resulting model-averaged estimates explicitly accounted for model selection uncertainty among the credible models. To do this, we computed the model-averaged value of a rebuilding target, say  $S_{\text{MSY}}$ , from its conditional expectation under each model and the associated model probabilities:

$$(20) \quad E[S_{\text{MSY}}|D] = \sum_k E[S_{\text{MSY}}|D, M_k] \Pr(M_k|D)$$

The conditional expectation of  $S_{\text{MSY}}$  under model  $M_k$  was derived from its MCMC chain. Similarly, the posterior probability of model  $M_k$  was computed as the expected value of the MCMC sequences of Bayes factors  $\{B_{jk}(1), B_{jk}(2), \dots, B_{jk}(5000)\}$  across models. In this case, the ordering of the samples of Schwarz criterion values  $\text{SC}_{jk}(m)$  by model did not affect the calculated posterior probability, since the MCMC chains were independent samples from each individual model's posterior distribution.

Given the model-specific averages of  $S_{\text{MSY}}$ , we computed its model-averaged variance from its conditional variance and expectation under each model along with the associated model probability by

$$(21) \quad \text{VAR}[S_{\text{MSY}}|D] = \sum_k (\text{VAR}[S_{\text{MSY}}|D, M_k] + E[S_{\text{MSY}}|D, M_k]^2) \Pr(M_k|D) - E[S_{\text{MSY}}|D]^2$$

In eq. 21, the conditional variance of  $S_{\text{MSY}}$  under model  $M_k$  was computed from its MCMC chain and the other terms were taken from eq. 20.

### Sensitivity analyses

We conducted sensitivity analyses to assess the relative importance of the choice of mean values for the informative priors under each alternative model. These analyses were expected to show which stock-recruitment parameters had the largest effect on rebuilding target estimates. Four separate analyses of sensitivity to informative priors were conducted: mean value of steepness prior, mean value of unfished recruitment prior, mean value of slope at the origin prior, and combination of mean values of steepness and unfished recruitment priors. In each sensitivity analysis, we recomputed the posterior modes of  $\text{MSY}$ ,  $F_{\text{MSY}}$ , and  $S_{\text{MSY}}$  using an alternative set of values for the prior mean for each of the three stocks. This was done for both uncorrelated and serially correlated errors. For each analysis, the variance of the prior was set so that the coefficient of variation was constant and equal to the value in the original analysis of each stock. For the sensitivity analysis of the steepness prior, the set of values of  $\mu_z$  was  $Z = \{0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95\}$ . A corresponding set of values for the prior mean of the slope at the origin prior  $\mu_a$  was determined using eq. 15 along with the expected value of unfished spawning biomass per recruit ( $\text{SPR}(F=0)$ ) for each stock. For the sensitivity analysis of the unfished recruitment prior, we used a set of multipliers  $M$  ranging from  $-50\%$  to  $+50\%$  of the original mean  $\mu_R$ ; the set of multipliers was  $M = \{0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5\}$ . In the combined sensitivity analysis of the steepness and unfished recruitment priors, we

**Table 3.** Coefficients of variation (%) of maximum sustainable yield (MSY), spawning stock to produce MSY ( $S_{MSY}$ ), and fishing mortality to produce MSY ( $F_{MSY}$ ) from posterior sampling for Georges Bank Atlantic cod, haddock, and yellowtail flounder under each alternative rebuilding target model (credible models indicated in bold).

	BH	ABH	<b>RBH</b>	<b>RABH</b>	ZBH	ZABH	<b>RZBH</b>	<b>RZABH</b>	RK	ARK	<b>SRK</b>	SARK
Georges Bank cod												
MSY	795	644	15	19	428	1447	15	18	100	322	65	393
$S_{MSY}$	839	509	19	22	445	1463	18	20	101	314	66	393
$F_{MSY}$	9	19	13	21	13	17	11	14	8	29	1	1
Georges Bank haddock												
MSY	2470	803	10	15	649	886	10	13	235	312	250	296
$S_{MSY}$	2437	835	14	22	706	965	12	16	239	318	253	289
$F_{MSY}$	25	64	33	59	20	41	24	44	23	29	15	14
Georges Bank yellowtail flounder												
MSY	856	194	5	9	554	301	5	6	210	381	613	409
$S_{MSY}$	960	248	6	13	617	307	5	7	253	544	643	428
$F_{MSY}$	15	27	9	27	11	15	8	13	23	45	8	8

**Table 4.** Model-averaging results for MSY (thousands of metric tons),  $S_{MSY}$  (thousands of metric tons), and  $F_{MSY}$  (per year) based on Markov Chain Monte Carlo sampling for Georges Bank Atlantic cod.

Model	Posterior model probability $[-2\log(B_{jk})]^a$	MSY	$S_{MSY}$	$F_{MSY}$	Evidence against model
RBH	0.34 [1.4]	37.4 (5.6)	193.7 (36.2)	0.21 (0.03)	None
RABH	0.15 [3.5]	35.4 (6.8)	176.1 (39.3)	0.23 (0.05)	Positive
RZBH	0.33 [1.4]	37.4 (5.6)	188.7 (33.6)	0.22 (0.02)	None
RZABH	0.16 [3.4]	35.8 (6.4)	172.7 (34.6)	0.23 (0.03)	Positive
SRK	0.01 [8.9]	46.2 (30.3)	87.5 (57.4)	0.69 (0.01)	Strong
Model average		36.9 (6.8)	184.7 (38.2)	0.23 (0.06)	
80% credibility interval		(28.2, 45.6)	(135.8, 233.6)	(0.15, 0.31)	

**Note:** Conditional expectations by model and model-averaged values of MSY,  $S_{MSY}$ , and  $F_{MSY}$  appear with the associated standard error in parentheses along with 80% credibility intervals for model-averaged results. The average value of twice the natural logarithm of the approximate Bayes factor appears in square brackets after each posterior model probability. Models: RBH, informative recruitment prior with uncorrelated Beverton–Holt; RABH, informative recruitment prior with autocorrelated Beverton–Holt; RZBH, informative recruitment and steepness priors with uncorrelated Beverton–Holt; RZABH, informative recruitment and steepness priors with autocorrelated Beverton–Holt; SRK, informative slope at origin prior with uncorrelated Ricker.

<sup>a</sup>Values may not sum to unity due to rounding.

used the set of all possible combinations of  $Z$  and  $M$ ; this was the set  $Z \otimes M$  with a total of 88 elements.

## Results

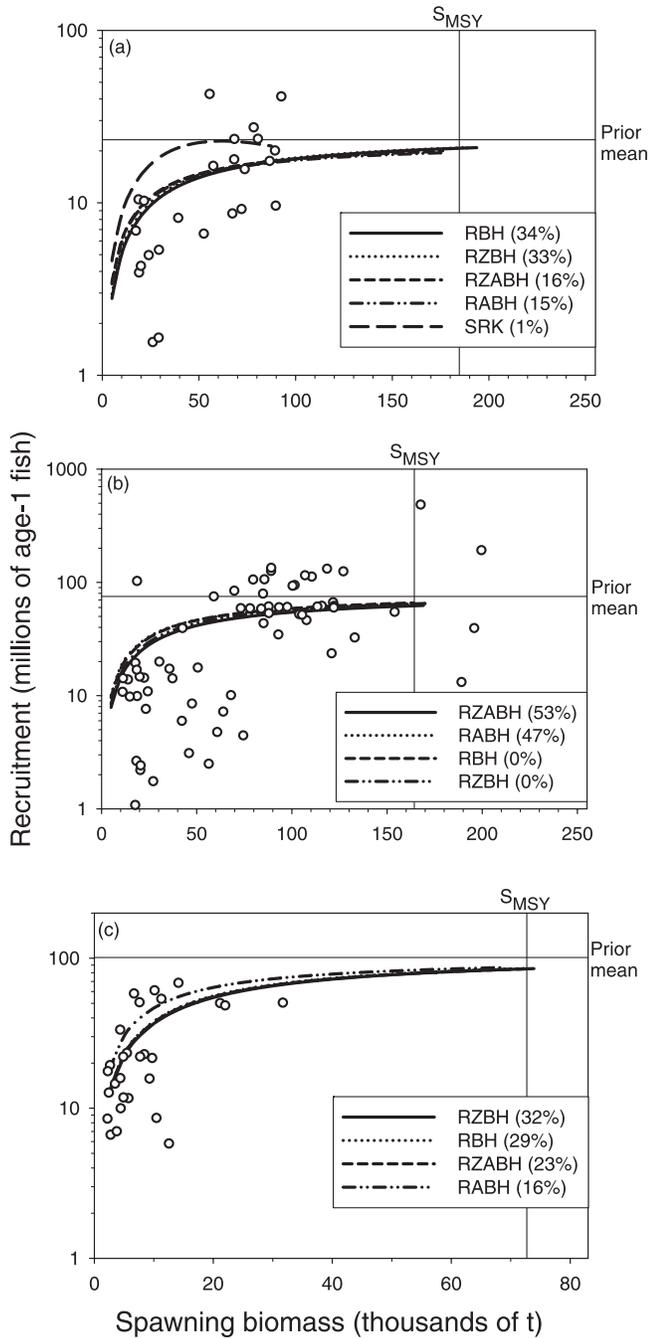
Convergence diagnostics indicated that the individual MCMC simulations converged for all of the alternative models. As a result, posterior modes were judged to be estimable for each alternative model and stock. The calculated PSRF values for the estimands MSY,  $S_{MSY}$ , and  $F_{MSY}$  were approximately 1.00 in all cases except for the  $F_{MSY}$  values of Georges Bank cod using the BH (PSRF = 1.59) and ABH (PSRF = 1.18) models. In these two cases, the models were not identifiable (see below) and were not used for model averaging. Overall, the convergence diagnostics indicated that the MCMC samples converged in probability to the joint posterior distribution of the rebuilding targets. Identical results were obtained using the alternative convergence diagnostic based on the second-order central-moment reduction factor.

Many of the alternative models did not have identifiable parameters based on the coefficients of variation of MSY,

$S_{MSY}$ , and  $F_{MSY}$  (Tables 2 and 3). For Georges Bank cod, the set of credible models was {RBH, RABH, RZBH, RZABH, SRK}. For Georges Bank haddock and yellowtail flounder, the set of credible models was {RBH, RABH, RZBH, RZABH}. Thus, virtually all of the identifiable models were BH. In contrast, virtually all of the RK models were found to be noncredible based on the coefficient of variation criterion.

Model-averaged estimates of MSY,  $S_{MSY}$ , and  $F_{MSY}$  were tabulated from the posterior means and model probabilities of credible models for each stock. For Georges Bank cod, most of the credible models were BH models with an informative prior on unfished recruitment (Table 3; Fig. 3a). None of the posterior samples had values of  $F_{MSY} > 2$ . Of the four credible models, the RBH and RZBH models were most probable (Table 4). These two uncorrelated models used uninformative and informative priors on steepness but produced very similar rebuilding targets. The corresponding RABH and RZABH models with serially correlated errors were approximately half as likely as the uncorrelated ones. The least probable model was the RK with an informative prior on slope at the origin. There was strong evidence

**Fig. 3.** Observed stock–recruitment data (circles), credible models and their approximate probabilities (%), model-averaged estimates of the spawning biomass to produce MSY ( $S_{MSY}$ , vertical line), and prior mean of unfished recruitment (horizontal line) for (a) Georges Bank Atlantic cod, (b) haddock, and (c) yellowtail flounder. Metric tons, t.



against this model, as it had roughly a 1% chance of being correct conditioned on the available data. Model-averaged values and standard errors of  $S_{MSY}$ ,  $F_{MSY}$ , and MSY were computed over the set of credible models (Table 4). For comparison, the calculated 80% credibility intervals of the model-averaged rebuilding targets overlapped the current rebuilding targets for Georges Bank cod (NEFSC 2002a) of  $S_{MSY}$  (216.8 kt),  $F_{MSY}$  (0.18), and MSY (35.2 kt).

Georges Bank haddock model-averaging results differed from those of cod. For this stock, the most probable models were serially correlated with informative priors on unfished recruitment (RZABH and RABH) (Table 3; Fig. 3b). A few of the posterior samples were excluded from model averaging because they had values of  $F_{MSY} > 2$ . The percentages of samples excluded were 0.14% for RBH, 2.42% for RABH, 0.04% for RZBH, and 0.38% for RZABH. These four BH models were almost equally likely, although the one with an informative prior on steepness was slightly more probable (Table 5). There was strong evidence against the corresponding uncorrelated models (RZBH and RBH), which had a very low probabilities of being true conditioned on the available data. For haddock, the model-averaged values and associated standard errors of  $S_{MSY}$ ,  $F_{MSY}$ , and MSY were computed and compared with current rebuilding targets. For Georges Bank haddock, the 80% credibility intervals overlapped the current rebuilding targets (NEFSC 2002a) of  $F_{MSY}$  (0.26) and MSY (52.9 kt) but did not overlap that for  $S_{MSY}$  (250.3 kt).

Model-averaging results for Georges Bank yellowtail flounder results were similar to those for cod since the credible models were the same for both stocks, except for the RK model deemed credible for cod but not for yellowtail flounder. For yellowtail flounder, none of the posterior samples had values of  $F_{MSY} > 2$ . Of the four credible models, the uncorrelated BH models with informative priors on unfished recruitment (RZBH and RBH) were most probable, as was the case for cod (Table 3; Fig. 3c). However, the difference between the model probabilities of the corresponding serially correlated models (RZABH and RABH) was smaller for yellowtail flounder than for cod (Table 6). There was some positive evidence against the autocorrelated models, which had similar odds of being true. In contrast with cod and haddock, the available yellowtail flounder data were relatively uninformative to discern among the credible models. The model-averaged values of  $S_{MSY}$ ,  $F_{MSY}$ , and MSY were similar to the individual model predictions (Table 6). In comparison with existing rebuilding targets for Georges Bank yellowtail flounder (NEFSC 2002a), the 80% credibility intervals of the model-averaged rebuilding targets did not overlap current rebuilding targets of  $S_{MSY}$  (58.8 kt),  $F_{MSY}$  (0.25), and MSY (12.9 kt).

Sensitivity analyses showed that the rebuilding targets were substantially affected by the steepness prior (Fig. 4). Sensitivity analyses on the effect of the mean of the steepness prior for BH models showed different patterns across stocks as well. The posterior mode of MSY showed a non-linear decrease as steepness increased. In magnitude, estimates of MSY were over 100-fold greater than existing MSY estimates at low steepnesses below 0.5 for each stock. These estimates became more consistent with observed annual yields (30 000, 9000, and 4000 t during 1980–1999 for cod, haddock, and yellowtail flounder, respectively) and MSY estimates (35 000, 53 000, and 13 000 t for cod, haddock, and yellowtail flounder, respectively) as the mean steepness increased (Fig. 4). However, the steepness at which this occurred differed among stocks. Estimates of MSY for uncorrelated and autocorrelated errors were similar at high steepnesses, although uncorrelated estimates tended

**Table 5.** Model-averaging results for MSY (thousands of metric tons),  $S_{MSY}$  (thousands of metric tons), and  $F_{MSY}$  (per year) based on Markov Chain Monte Carlo sampling for Georges Bank haddock.

Model	Posterior model probability $[-2 \log(B_{jk})]^a$	MSY	$S_{MSY}$	$F_{MSY}$	Evidence against model
RBH	0.00 [19.6]	55.7 (5.6)	164.0 (23.3)	0.54 (0.17)	Very strong
RABH	0.47 [1.3]	53.8 (7.9)	158.5 (33.6)	0.59 (0.31)	None
RZBH	0.00 [19.3]	54.4 (5.2)	170.8 (20.4)	0.48 (0.11)	Very strong
RZABH	0.53 [0.9]	51.8 (6.5)	169.1 (26.2)	0.48 (0.19)	None
Model average		52.7 (7.3)	164.3 (30.2)	0.53 (0.26)	
80% credibility interval		(43.4, 62.0)	(125.6, 203.0)	(0.20, 0.86)	

**Note:** Conditional expectations by model and model-averaged values of MSY,  $S_{MSY}$ , and  $F_{MSY}$  appear with the associated standard error in parentheses along with 80% credibility intervals for model-averaged results. The average value of twice the natural logarithm of the approximate Bayes factor appears in square brackets after each posterior model probability. Models: RBH, informative recruitment prior with uncorrelated Beverton–Holt; RABH, informative recruitment prior with autocorrelated Beverton–Holt; RZBH, informative recruitment and steepness priors with uncorrelated Beverton–Holt; RZABH, informative recruitment and steepness priors with autocorrelated Beverton–Holt.

<sup>a</sup>Values may not sum to unity due to rounding.

**Table 6.** Model-averaging results for MSY (thousands of metric tons),  $S_{MSY}$  (thousands of metric tons), and  $F_{MSY}$  (per year) based on Markov Chain Monte Carlo sampling for Georges Bank yellowtail flounder.

Model	Posterior model probability $[-2 \log(B_{jk})]^a$	MSY	$S_{MSY}$	$F_{MSY}$	Evidence against model
RBH	0.29 [1.4]	22.5 (1.1)	73.0 (4.1)	0.36 (0.03)	None
RABH	0.16 [3.2]	23.5 (2.1)	68.5 (8.9)	0.42 (0.12)	Positive
RZBH	0.32 [1.1]	22.3 (1.0)	73.9 (3.8)	0.35 (0.03)	None
RZABH	0.23 [2.2]	22.4 (1.4)	73.4 (5.2)	0.36 (0.05)	Positive
Model average		22.6 (1.4)	72.7 (5.7)	0.37 (0.06)	
80% credibility interval		(20.8, 24.4)	(65.4, 80.0)	(0.29, 0.45)	

**Note:** Conditional expectations by model and model-averaged values of MSY,  $S_{MSY}$ , and  $F_{MSY}$  appear with the associated standard error in parentheses along with 80% credibility intervals for model-averaged results. The average value of twice the natural logarithm of the approximate Bayes factor appears in square brackets after each posterior model probability. Models: RBH, informative recruitment prior with uncorrelated Beverton–Holt; RABH, informative recruitment prior with autocorrelated Beverton–Holt; RZBH, informative recruitment and steepness priors with uncorrelated Beverton–Holt; RZABH, informative recruitment and steepness priors with autocorrelated Beverton–Holt.

<sup>a</sup>Values may not sum to unity due to rounding.

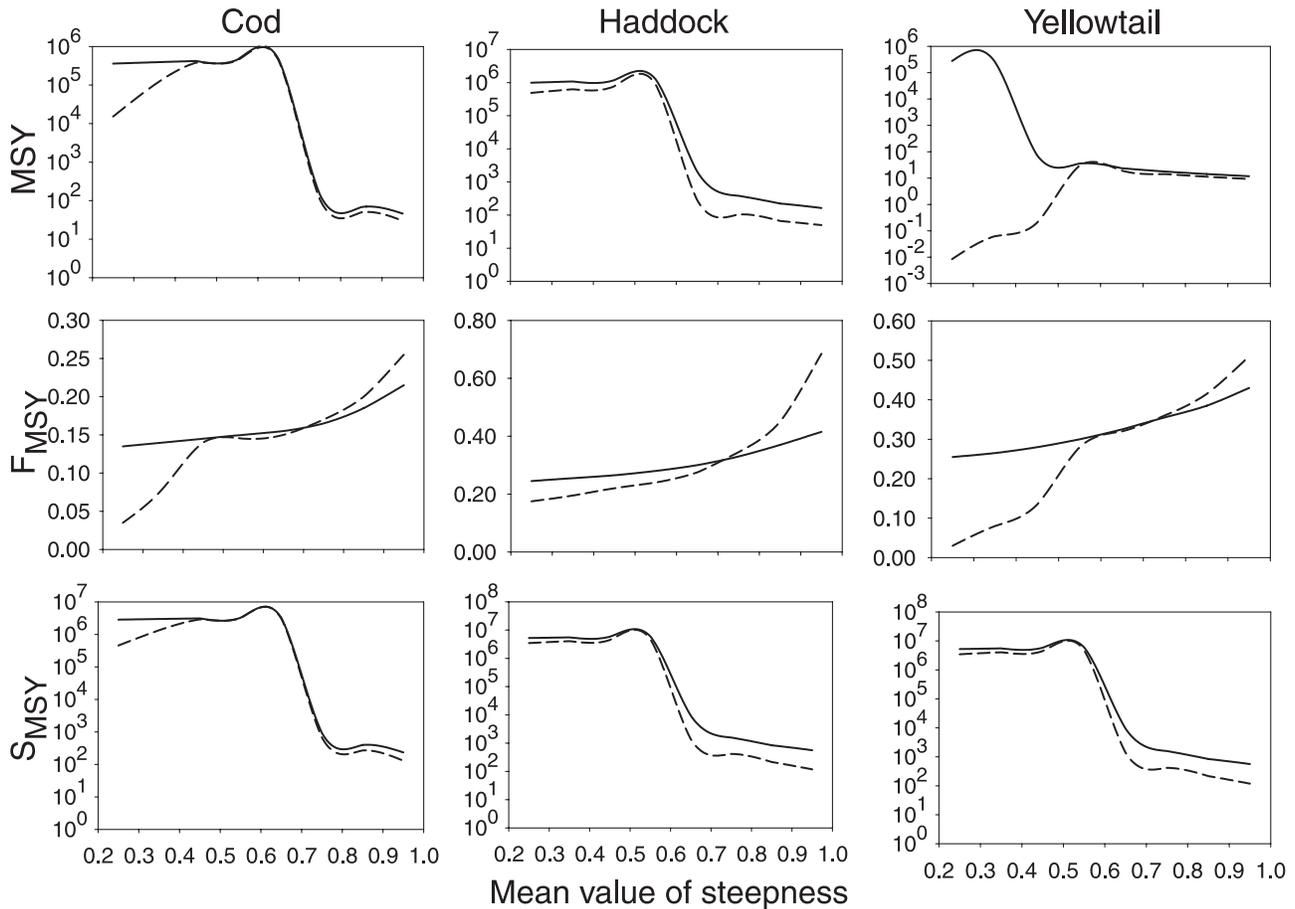
to be at or above autocorrelated ones. The sensitivity results for  $S_{MSY}$  were very similar to those for MSY. The  $S_{MSY}$  decreased as steepness increased and uncorrelated estimates tended to be at or above autocorrelated ones. In contrast, there was a more linear increase in the posterior mode estimate of  $F_{MSY}$  with mean steepness across stocks (Fig. 4). Further, the autocorrelated estimates of  $F_{MSY}$  showed a more pronounced response to changes in steepness than the uncorrelated estimates. Overall, MSY and  $S_{MSY}$  estimates were more sensitive to the steepness parameter.

Were rebuilding targets sensitive to the informative prior values for unfished recruitment under a BH model? Yes, the effect of the prior mean for unfished recruitment had an important effect on the rebuilding target estimates that varied among stocks (Fig. 5). The estimates of MSY and  $S_{MSY}$  generally increased with increasing unfished recruitment but the rate of increase differed among stocks. Uncorrelated estimates of both MSY and  $S_{MSY}$  were similar to autocorrelated estimates (Fig. 5). In contrast, estimates of  $F_{MSY}$  decreased with increasing unfished recruitment. Uncorrelated and autocorrelated estimates of  $F_{MSY}$  were similar for cod and yellowtail flounder but differed for haddock. Further, as unfished recruitment increased, MSY and  $S_{MSY}$  increased, while  $F_{MSY}$  decreased. Overall, the unfished recruitment parameter had a smaller effect on rebuilding target estimates than the steepness parameter.

Analyses also showed that estimated rebuilding targets were sensitive to the informative prior values for slope at the origin under an RK model. Sensitivity analyses of the effect of slope at the origin on rebuilding targets showed different patterns across stocks (Fig. 6), except for yellowtail flounder at low slope values. The estimates of MSY and  $S_{MSY}$  increased steadily with increasing slope, although the estimate of  $S_{MSY}$  for yellowtail flounder peaked at an intermediate slope value with autocorrelated errors. Uncorrelated and autocorrelated estimates of MSY and  $S_{MSY}$  were also similar, except for yellowtail flounder at low slope values (Fig. 6). The estimates of  $F_{MSY}$  tended to increase to the maximum allowable  $F$  ( $F = 2$ ) as slope increased for all stocks. Uncorrelated posterior modes of  $F_{MSY}$  were noncredible for mean slope priors of 10 or greater for both haddock and yellowtail flounder. There was very little difference between uncorrelated and autocorrelated estimates of  $F_{MSY}$  for cod and haddock. Overall, rebuilding target estimates were more sensitive to the slope at the origin priors than the unfished recruitment priors.

Were estimated rebuilding targets sensitive to the informative prior values for steepness and unfished recruitment priors under a BH model with uncorrelated errors? Yes, the rebuilding targets were sensitive to the joint effects of both priors. Sensitivity analyses of the joint effects of steepness and unfished recruitment priors using uncorrelated errors

**Fig. 4.** Sensitivity analyses of posterior modes of maximum sustainable yield (MSY, thousands of metric tons), spawning biomass to produce MSY ( $S_{MSY}$ , thousands of metric tons), and fishing mortality to produce MSY ( $F_{MSY}$ ) to choice of mean steepness for a Beverton–Holt stock–recruitment curve using uncorrelated (solid line) and autocorrelated (broken line) errors.



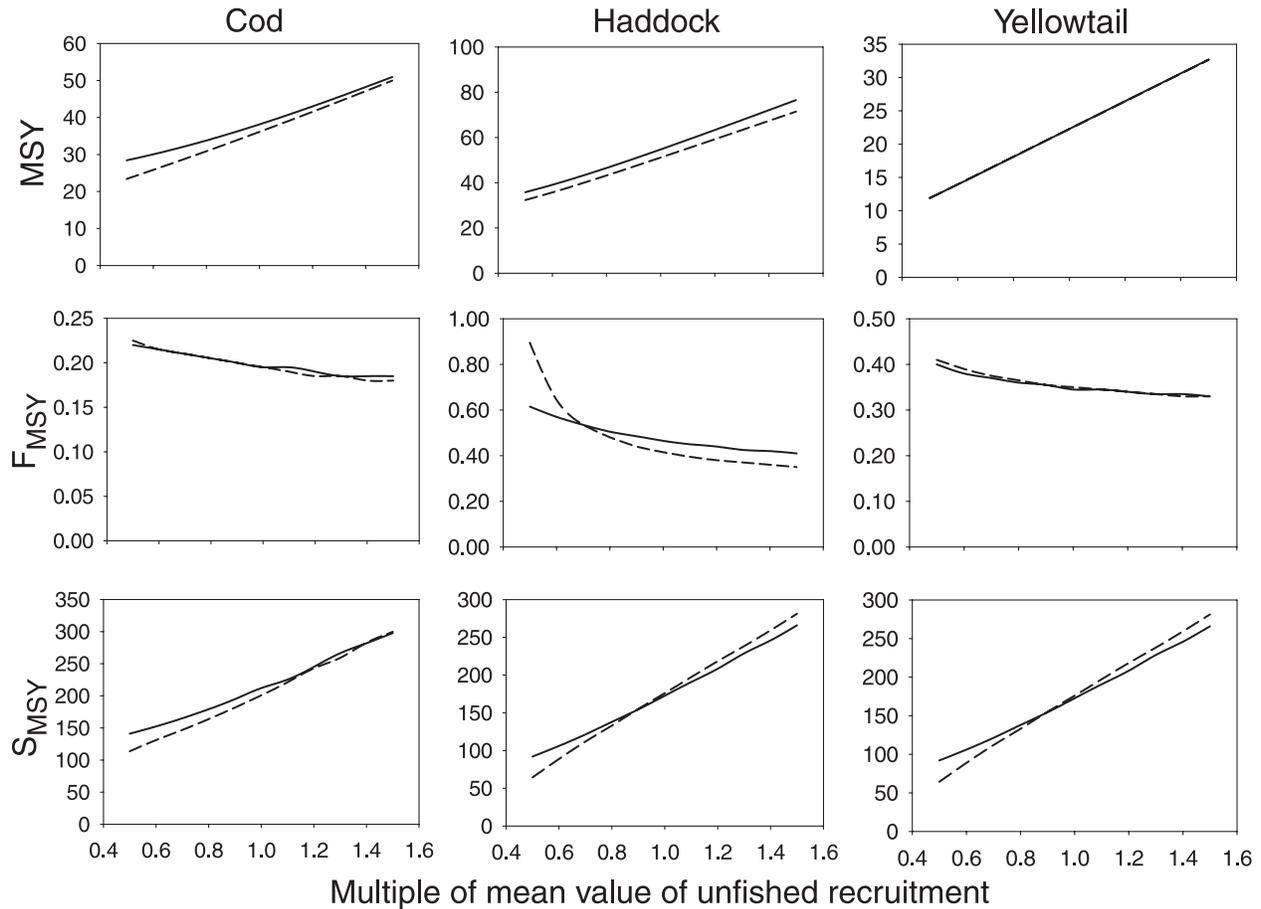
(Fig. 7) showed that the range of effects differed among stocks. Estimates of MSY and  $S_{MSY}$  were strongly affected by the unfished recruitment prior but were less sensitive to the choice of steepness prior. For fixed steepness, the estimated values of MSY and  $S_{MSY}$  were largely determined by unfished recruitment. MSY estimates had a slight negative correlation between steepness and unfished recruitment, while  $S_{MSY}$  estimates had a slight positive correlation (Fig. 7). In contrast,  $F_{MSY}$  estimates were strongly affected by the steepness prior. The  $F_{MSY}$  estimates appeared to have a moderate positive correlation between steepness and unfished recruitment. For a fixed value of unfished recruitment, steepness largely determined the estimated value of  $F_{MSY}$ . Thus, for uncorrelated errors, both MSY and  $S_{MSY}$  were determined by unfished recruitment, while  $F_{MSY}$  estimates were determined by steepness.

Estimated rebuilding targets were also sensitive to the informative prior values for steepness and unfished recruitment using correlated errors. Sensitivity analyses of the joint impacts of steepness and unfished recruitment using autocorrelated errors showed some similarities across stocks (Fig. 8). The estimated posterior modes of MSY were affected by both steepness and unfished recruitment. There was a negative correlation between steepness and unfished recruitment effects on MSY and this relationship was non-

linear for cod and yellowtail flounder (Fig. 8). The effects of steepness and unfished recruitment on  $F_{MSY}$  estimates were similar across stocks. In this case, the steepness prior was much more important than the unfished recruitment prior. There was a slight positive correlation between steepness and unfished recruitment effects at high steepness and low unfished recruitment values. Estimates of  $S_{MSY}$  for all three stocks were strongly affected by unfished recruitment (Fig. 8) and there was a moderate positive correlation between steepness and unfished recruitment effects on  $S_{MSY}$ .

Comparison of the results of sensitivity analyses using uncorrelated and autocorrelated error structures showed that the choice of error structure was important for these stocks. MSY estimates using uncorrelated errors were dependent on the unfished recruitment prior and were generally larger than estimates using autocorrelated errors. Estimates of MSY using autocorrelated errors were dependent on both steepness and unfished recruitment (Fig. 8). They also exhibited a stronger negative correlation between steepness and unfished recruitment effects and had a nonlinear response surface for cod and yellowtail flounder. Estimates of  $F_{MSY}$  using uncorrelated and autocorrelated were both strongly affected by the steepness prior (Figs. 7 and 8). Estimates of  $F_{MSY}$  were also higher using uncorrelated errors than estimates using autocorrelated errors. There was a stronger positive cor-

**Fig. 5.** Sensitivity analyses of posterior modes of maximum sustainable yield (MSY, thousands of metric tons), spawning biomass to produce MSY ( $S_{MSY}$ , thousands of metric tons), and fishing mortality to produce MSY ( $F_{MSY}$ ) to choice of prior mean multiplier of unfished recruitment for a Beverton–Holt stock–recruitment curve using uncorrelated (solid line) and autocorrelated (broken line) errors.



relation between steepness and unfished recruitment effects using uncorrelated errors. Estimates of  $S_{MSY}$  using uncorrelated and autocorrelated errors were strongly dependent on the unfished recruitment prior (Figs. 7 and 8). Steepness and unfished recruitment effects were more positively correlated and showed some nonlinearity at high unfished recruitment using autocorrelated errors. The  $S_{MSY}$  estimates using uncorrelated errors also tended to be lower than estimates using autocorrelated errors. Overall, rebuilding target estimates were more sensitive to priors with correlated versus uncorrelated errors.

## Discussion

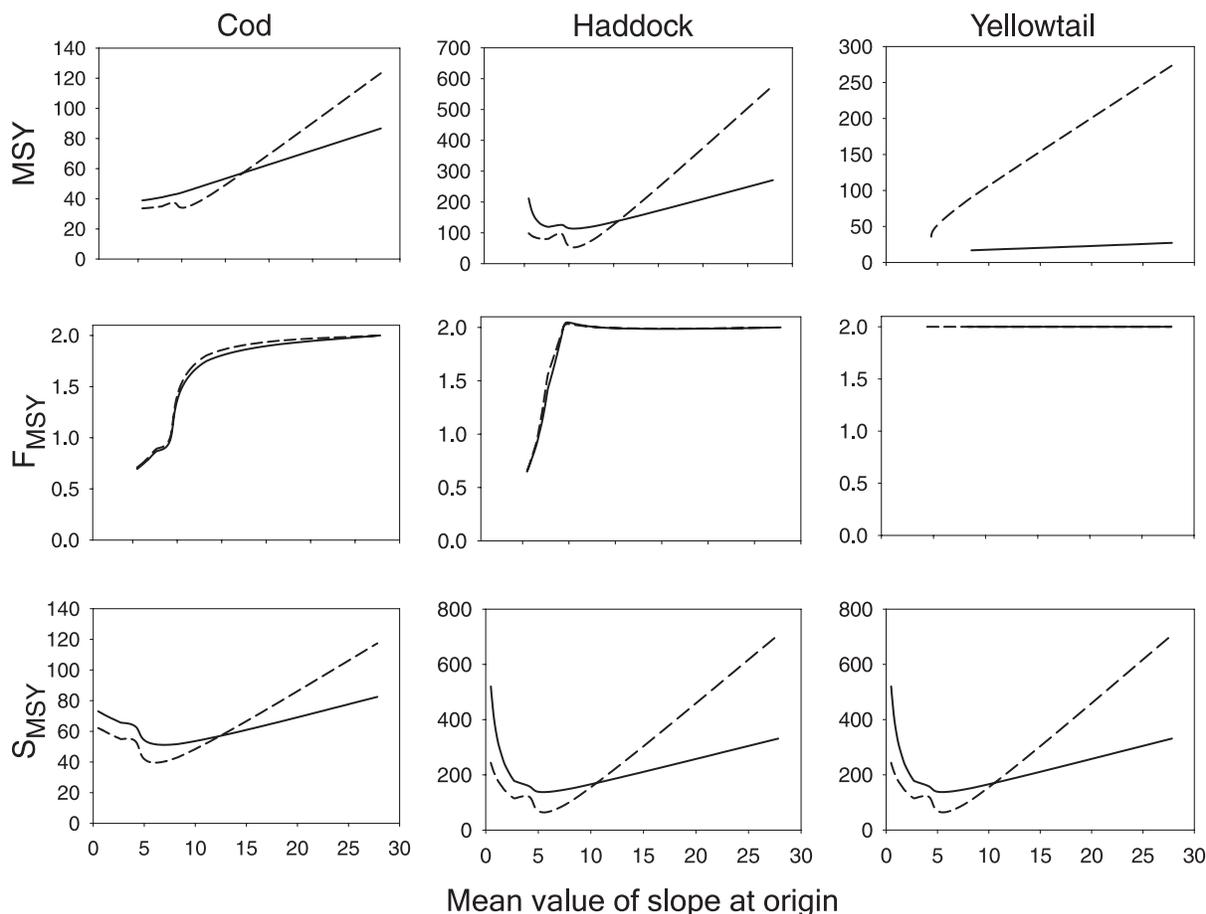
Model-averaging results placed almost all of the probability mass on the BH model for the three Georges Bank stocks. This suggested that the BH relationship was an adequate compensatory hypothesis given sufficient prior information on unfished recruitment. This relationship has a sound theoretical basis as a model of stock–recruitment dynamics and arises naturally when density-dependent effects are critical at some early life history stage (e.g., see Quinn and Deriso 1999). This model can also arise as a result of stock adaptation to balance predation and foraging risk in a variable environment (Walters and Korman 1999). Without sufficient prior information, however, none of the BH mod-

els were identifiable for any of the stocks. This emphasizes the importance of including as much relevant information as possible when evaluating rebuilding targets for overfished stocks.

In contrast, model-averaging results showed limited support for the overcompensatory stock–recruitment hypothesis for all three stocks. This finding is consistent with observed food habits information for these three stocks, which show very little incidence of cannibalism (one possible mechanism to generate overcompensation) for haddock and none for yellowtail flounder (NEFSC 2002a). Interestingly, Georges Bank cod, which exhibits an increasing incidence of cannibalism with ontogeny (Link and Garrison 2002), was the only stock that had positive probability for an RK model. However, that model was also the least probable (1%) model for cod.

Our illustration of model averaging to compute rebuilding targets was limited to contrasting a pair of two-parameter stock–recruitment models that can produce vastly different rebuilding target values with similar values of the goodness-of-fit criterion (Myers et al. 1994). When using model averaging in practice, it will be necessary to modify the hypotheses to match the specific situation. Investigating different stock–recruitment models, such as the hockey stick (Barrowman and Myers 2000) or curves that exhibit depensation, and using error structures that reflect expected dynamics un-

**Fig. 6.** Sensitivity analyses of posterior modes of maximum sustainable yield (MSY, thousands of metric tons), spawning biomass to produce MSY ( $S_{\text{MSY}}$ , thousands of metric tons), and fishing mortality to produce MSY ( $F_{\text{MSY}}$ ) to choice of mean slope at the origin for a Ricker stock–recruitment curve using uncorrelated (solid line) and autocorrelated (broken line) errors.



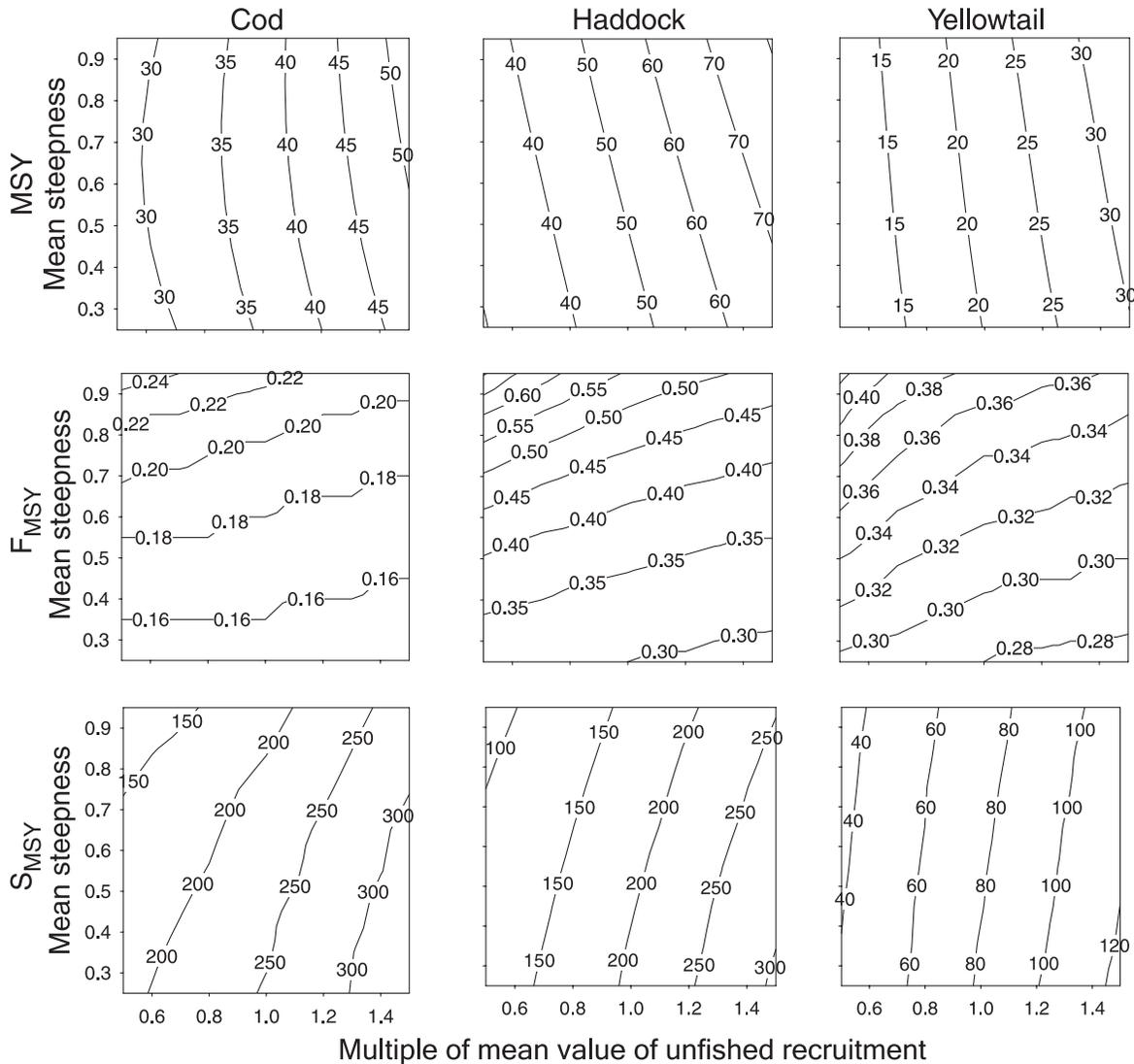
der alternative hypotheses may be necessary. In this context, it is the shape of the curves, their biological plausibility, and the estimability of model parameters that are of paramount importance. It may be difficult to identify stock–recruitment models that have similar shapes using short time series. This emphasizes the importance of choosing an appropriate set of models to average over when applying model-averaging techniques (Hoeting et al. 1999).

While we used a Bayesian approach for model averaging, a frequentist approach based on multiple-model inference and model averaging using bootstrapped Akaike weights could also have been applied (Anderson et al. 2001; Burnham and Anderson 2002). In this context, our use of informative priors is conceptually equivalent to applying a penalty function to a likelihood to constrain parameter estimates (e.g., Edwards 1992). For comparison, a sensitivity analysis using the Akaike criterion instead of the Schwarz criterion produced virtually identical model-averaged rebuilding targets for all three stocks. This was not surprising, since both criteria are based on the sample likelihood adjusted with a parameter penalty. Although our choice of a cutoff coefficient of variation of 100% on key model parameters to eliminate nonidentifiable models was subjective, sensitivity analyses using a higher cutoff of 200% produced identical rebuilding targets for haddock and yellowtail flounder. For cod, the uncorrelated RK model became credible using a 200% cut-

off. However, the model-averaged rebuilding target estimates changed by less than 0.5%. Overall, this suggested that the model-averaging approach was robust to the choice of goodness-of-fit and model identifiability criteria.

Although the concept of MSY has its limitations (Larkin 1977), the expected fishing mortality that produces MSY has proven to be a useful limit rebuilding target for precautionary single-species management (Mace 2001). Similarly, the expected spawning biomass that produces MSY provides a useful goal for implementing an ecosystem approach to fisheries management with multiple performance measures (Brodziak and Link 2002). We also note that the model-averaging approach can be applied with other reference points that depend on stock–recruitment models, such as biomass limits for recruitment overfishing (Myers et al. 1994) and fishing limits defined by the slope of the stock–recruitment relationship at the origin (ICES 1997). While it has been argued that it may be useful to eventually move away from model-based estimates of reference points (e.g., Hilborn 2002), we believe that rebuilding targets can provide satisfactory management advice provided that uncertainties about the stock–recruitment relationship are explicitly addressed. Furthermore, we recommend using multiple working hypotheses to address fundamental uncertainties. In this context, model averaging offers a structured alternative to estimating rebuilding targets when there are several compet-

**Fig. 7.** Sensitivity analyses of posterior modes of maximum sustainable yield (MSY, thousands of metric tons), spawning biomass to produce MSY ( $S_{MSY}$ , thousands of metric tons), and fishing mortality to produce MSY ( $F_{MSY}$ ) to choice of mean steepness and multiplier of mean unfished recruitment for a Beverton–Holt stock–recruitment curve using uncorrelated errors.



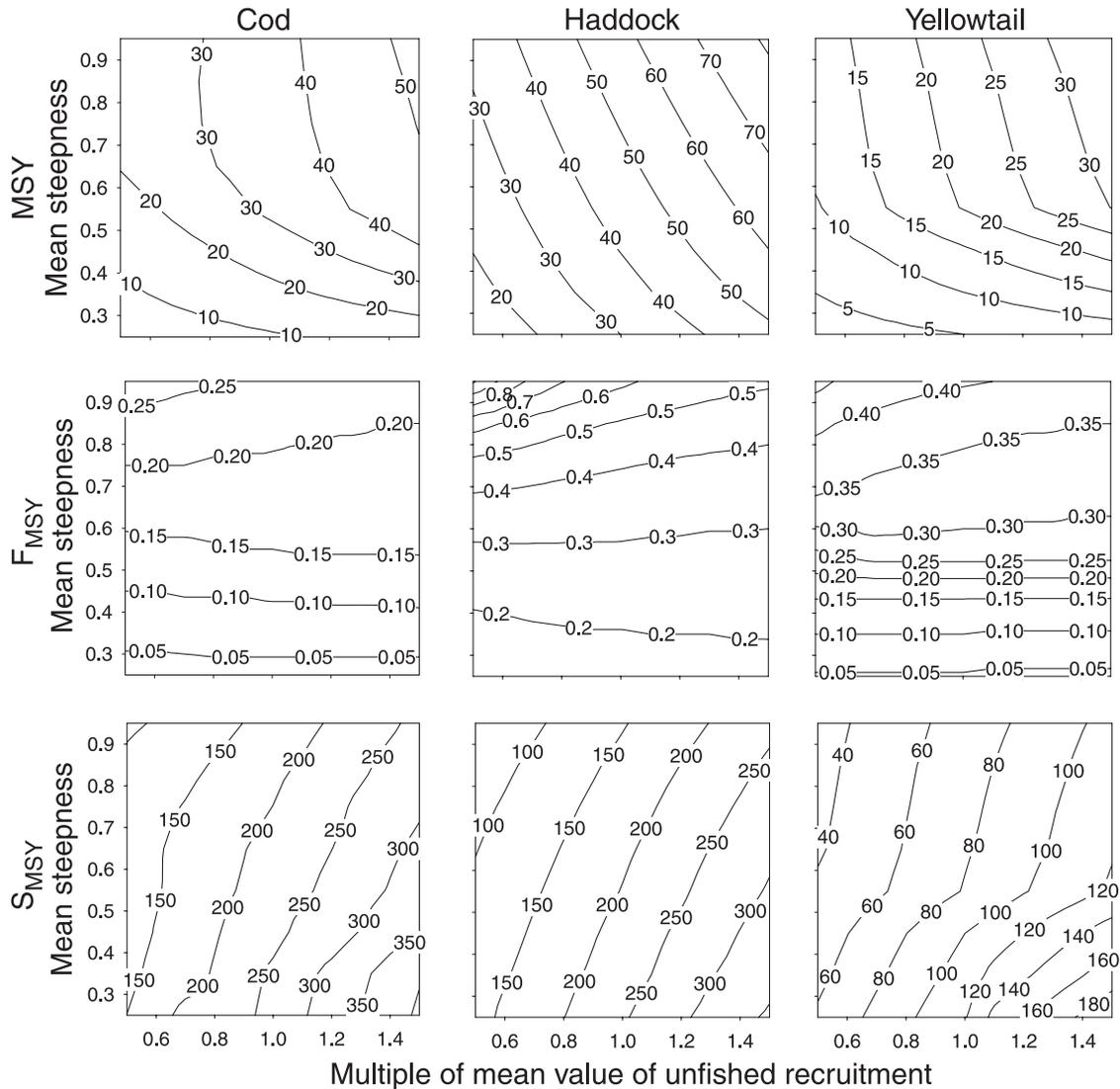
ing hypotheses about stock dynamics in an objective manner.

Model averaging results for the three groundfish stocks had wide confidence intervals that reflected the underlying uncertainty in rebuilding targets. Although some differences existed between the results of this study and existing biological rebuilding targets (e.g., haddock  $F_{MSY}$  and  $S_{MSY}$  and yellowtail flounder MSY,  $F_{MSY}$ , and  $S_{MSY}$ ), it is important to recognize that our results are only an illustration of the technical approach. In particular, our results could not logically supplant the existing biological rebuilding targets, which have undergone extensive peer review and stakeholder scrutiny and, as a result, have been judged to constitute the best available science for the current fishery management plan to manage New England groundfish. Nonetheless, it is clear that management advice for these three stocks would be robust to apparent differences. Fishing mortality on the Georges Bank cod stock needs to be reduced and the stock needs to be rebuilt to a larger spawning biomass (NEFSC 2002b). In contrast, the Georges Bank haddock and yellowtail floun-

der stocks were rebuilding in recent years at historically low fishing mortality rates (NEFSC 2002b). Furthermore, although the model-averaged estimate  $F_{MSY}$  for Georges Bank haddock in this study is roughly twice the current rebuilding target, it is imprecisely determined with a coefficient of variation of roughly 50%. It is interesting that this stock produced an average annual yield of 50 kt when fished at an average  $F$  of roughly 0.5 over three decades even though smaller mesh was utilized (Sissenwine and Shepherd 1987). At least historically, the haddock stock was resilient and productive at relatively high fishing mortalities. Nonetheless, the fundamental uncertainty in the rebuilding target estimates emphasizes the need to manage adaptively by monitoring progress towards stock recovery and reevaluating rebuilding targets as appropriate. This uncertainty would be reduced by observations of recruitment at high stock sizes if stock recovery plans are successful.

Our model-averaging results could also be applied to estimate rebuilding trajectories for these overfished stocks using a standard projection framework (e.g., Brodziak et al. 1998)

**Fig. 8.** Sensitivity analyses of posterior modes of maximum sustainable yield (MSY, thousands of metric tons), spawning biomass to produce MSY ( $S_{MSY}$ , thousands of metric tons), and fishing mortality to produce MSY ( $F_{MSY}$ ) to choice of mean steepness and multiplier of mean unfished recruitment for a Beverton–Holt stock–recruitment curve using autocorrelated errors.



in a straightforward manner. In particular, one can randomly select a credible stock–recruitment model in proportion to its model probability and then draw a set of model parameters from the joint posterior distribution of the selected model. If low-frequency environmental forcing is important, it will be incorporated into the projections through one or more autocorrelated models. In this way, the simulation directly reflects the model-averaging estimates — a useful simplification when addressing the large number of uncertainties that may be important for forecasting (e.g., Patterson et al. 2001). On the other hand, it is important to not overinterpret the reliability of recruitment predictions, especially if non-stationarity in stock–recruitment dynamics due to either environmental forcing or depensatory effects appears evident. In practice, it is advisable to assess whether there are trends in observed production of recruits per spawner in recent years. If strong time trends are apparent, a single stock–recruitment model fit with autocorrelated errors may not provide a satisfactory short-term approximation and other

structural approaches, such as randomly resampling recent recruitment observations, may have better predictive accuracy.

The results of the sensitivity analyses provide some guidance on the potential effects of choices for informative priors on steepness, slope at the origin, or unfished recruitment. In general, we found that steepness priors with mean values below 0.5 (well outside the probable range of the informative prior) led to noncredible results for MSY and  $S_{MSY}$ , regardless of whether an uncorrelated or autocorrelated error structure was used. In contrast, changes in unfished recruitment priors produced a more linear response in rebuilding target estimates. Slope at origin priors also produced noncredible results (e.g.,  $F_{MSY} > 2$ ) at higher mean slopes. This is likely due to an intrinsic lack of model identifiability when an overcompensatory stock–recruitment curve is combined with a high slope at the origin. In particular, RK and similarly shaped curves can produce deterministic chaos. Combining this deterministic feature with short time series

of highly variable stock–recruitment data may make parameter identification difficult. Our results also showed that  $MSY$  and  $S_{MSY}$  were strongly dependent on unfished recruitment, while  $F_{MSY}$  was strongly dependent on steepness. Thus, the overall magnitude of recruitment had a more pronounced effect on the magnitude of yield and stock biomass, while the rate of increase of the stock–recruitment curve near the origin had a greater effect on the fishing mortality rate. In comparison, the choice of error structure had a less pronounced effect in most cases, although autocorrelation introduced more nonlinearity in the response to steepness and unfished recruitment. Last, it was evident that the sensitivity of rebuilding targets differed among stocks, which emphasizes the importance of measuring differences in life history parameters and the time series of available stock and recruitment observations.

Model averaging is a general technique for ecological investigations that can be used to produce robust parameter estimates when model uncertainty is important (Johnson and Omland 2004). This approach could be applied to rebuilding target estimation for many commercially exploited stocks that have age-structured assessment data. One common problem of rebuilding target estimation is lack of knowledge about the steepness parameter of the stock–recruitment relationship (Williams and Shertzer 2003). We believe that the use of meta-analyses such as Myers et al. (1999) provides the best alternative when time series of stock–recruitment data are short and have high intrinsic variance, a common situation in assessment data worldwide. Another common problem with rebuilding target estimation is the lack of stock–recruitment data at high spawning stock sizes. Our use of informative priors for unfished recruitment based on observed recruitment distributions directly addresses this problem. In fact, our results show that it may not be possible to reliably estimate rebuilding targets for stocks with a long history of overfishing unless prior information on steepness or unfished recruitment is available. Overall, this underscores the importance of being explicit about the treatment of model and parameter uncertainty, especially when conflicts arise over appropriate rebuilding targets for overfished stocks (e.g., see Greene 2002).

## Acknowledgments

We thank Jim Bence, Fred Serchuk, and two anonymous referees for their helpful comments on the draft manuscript.

## References

- Anderson, D.R., Burnham, K.P., and White, G.C. 2001. Kullback–Leibler information in resolving natural resource conflicts when definitive data exist. *Wildl. Soc. Bull.* **29**: 1260–1270.
- Barrowman, N.J., and Myers, R.A. 2000. Still more spawner–recruitment curves: the hockey stick and its generalizations. *Can. J. Fish. Aquat. Sci.* **57**: 665–676.
- Brodziak, J. 2002. In search of optimal harvest rates for west coast groundfish. *N. Am. J. Fish. Manag.* **22**: 258–271.
- Brodziak, J., and Link, J. 2002. Ecosystem-based fishery management: what is it and how can we do it? *Bull. Mar. Sci.* **70**: 589–611.
- Brodziak, J., Rago, P., and Conser, R. 1998. A general approach for making short-term stochastic projections from an age-structured fisheries assessment model. *In Proceedings of the International Symposium on Fishery Stock Assessment Models for the 21st Century. Edited by F. Funk, T. Quinn II, J. Heifetz, J. Ianelli, J. Powers, J. Schweigert, P. Sullivan, and C.-I. Zhang.* Alaska Sea Grant College Program, University of Alaska, Fairbanks, Alaska.
- Brodziak, J.K.T., Overholtz, W.J., and Rago, P.J. 2001. Does spawning stock affect recruitment of New England groundfish? *Can. J. Fish. Aquat. Sci.* **58**: 306–318.
- Brooks, S.P., and Gelman, A. 1998. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* **7**: 434–455.
- Burnham, K.P., and Anderson, D. 2002. Model selection and multimodel inference. Springer-Verlag, New York.
- Carlin, B.P., and Louis, T.A. 2000. Bayes and empirical Bayes methods for data analysis. Chapman and Hall, New York.
- Edwards, A.W.F. 1992. Likelihood. Johns Hopkins University Press, Baltimore, Md.
- Ellison, A.M. 2004. Bayesian inference in ecology. *Ecol. Lett.* **7**: 509–520.
- Garcia, S.M., and Newton, C. 1997. Current situations, trends and prospects in world capture fisheries. *In Global trends: fisheries management. Edited by E.K. Pikitch, D.D. Huppert, and M.P. Sissenwine.* Am. Fish. Soc. Symp. **20**: 3–27.
- Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. 1995. Bayesian data analysis. Chapman and Hall, New York.
- Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. 1996. Markov Chain Monte Carlo in practice. Chapman and Hall, New York.
- Greene, K. 2002. Bigger populations needed for sustainable harvests. *Science (Wash., D.C.)*, **296**: 1229–1230.
- Hilborn, R. 2002. The dark side of reference points. *Bull. Mar. Sci.* **70**: 403–408.
- Hilborn, R., and Liermann, M. 1998. Standing on the shoulders of giants: learning from experience in fisheries. *Rev. Fish Biol. Fish.* **8**: 273–283.
- Hilborn, R., and Mangel, M. 1997. The ecological detective. Princeton University Press, Princeton, N.J.
- Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T. 1999. Bayesian model averaging: a tutorial. *Stat. Sci.* **14**: 382–401.
- Hutchings, J.A., Walters, C., and Haedrich, R.L. 1997. Is scientific inquiry incompatible with government information control? *Can. J. Fish. Aquat. Sci.* **54**: 1198–1210.
- ICES. 1997. Report of the study group on the precautionary approach to fisheries management. ICES CM 1997/Assess 7.
- Johnson, J.B., and Omland, K.S. 2004. Model selection in ecology and evolution. *Trends Ecol. Evol.* **19**: 101–108.
- Kass, R.E., and Raftery, A.E. 1995. Bayes factors. *J. Am. Stat. Assoc.* **90**: 773–795.
- Larkin, P. 1977. An epitaph for the concept of maximum sustained yield. *Trans. Am. Fish. Soc.* **106**: 1–11.
- Link, J.S., and Garrison, L.P. 2002. Trophic ecology of Atlantic cod *Gadus morhua* on the northeast US continental shelf. *Mar. Ecol. Prog. Ser.* **227**: 109–123.
- Mace, P.M. 2001. A new role for  $MSY$  in single-species and ecosystem approaches to fisheries stock assessment and management. *Fish Fish.* **2**: 2–32.
- Mace, P.M., and Sissenwine, M.P. 1993. How much spawning per recruit is enough? *In Risk evaluation and biological reference points for fisheries management. Edited by S.J. Smith, J.J. Hunt, and D. Rivard.* Can. Spec. Publ. Fish. Aquat. Sci. No. 120. pp. 101–118.
- Myers, R.A., and Mertz, G. 1998. Reducing uncertainty in the biological basis of fisheries management by meta-analysis of data from many populations: a synthesis. *Fish. Res.* **37**: 51–60.

- Myers, R.A., Rosenberg, A.A., Mace, P.M., Barrowman, N.J., and Restrepo, V.R. 1994. In search of thresholds for recruitment overfishing. *ICES J. Mar. Sci.* **51**: 191–205.
- Myers, R., Bowen, K.G., and Barrowman, N. 1999. Maximum reproductive rate of fish at low population sizes. *Can. J. Fish. Aquat. Sci.* **56**: 2404–2419.
- NEFSC. 2002a. Working Group Report, re-evaluation of biological reference points for New England groundfish. Ref. Doc. 02-04. National Marine Fisheries Service, Northeast Fisheries Science Center, Woods Hole, Mass. Available at <http://www.nefsc.nmfs.gov/nefsc/publications/crd/crd0204/>.
- NEFSC. 2002b. Assessment of 20 northeast groundfish stocks through 2001. A report of the Groundfish Assessment Review Meeting (GARM), Woods Hole, Massachusetts, 8–11 October 2002. Ref. Doc. 02-16. National Marine Fisheries Service, Northeast Fisheries Science Center, Woods Hole, Mass. Available at <http://www.nefsc.nmfs.gov/nefsc/publications/crd/crd0216/>.
- Otter Research Ltd. 2001. An introduction to AD Model Builder 6.0.2 for use in nonlinear modeling and statistics. Otter Research Ltd., Box 2040, Sidney, BC V8L 3S3, Canada.
- Patterson, K.P., Cook, R., Darby, C., Gavaris, S., Kell, L., Lewy, P., Mesnil, B., Punt, A., Restrepo, V., Skagen, D.W., and Stefansson, G. 2001. Estimating uncertainty in fish stock assessment and forecasting. *Fish. Fish.* **2**: 125–157.
- Pauly, D. 1995. Anecdotes and the shifting baseline syndrome of fisheries. *Trends Ecol. Evol.* **10**: 430.
- Pauly, D., Christensen, V., Guénette, S., Pitcher, T.J., Sumaila, U.R., Walters, C.J., Watson, R., and Zeller, D. 2002. Towards sustainability in world fisheries. *Nature (Lond.)*, **418**: 689–695.
- Quinn, T.J., II, and Deriso, R.B. 1999. Quantitative fish dynamics. Oxford University Press, New York.
- Rosenberg, A.A. 2003. Managing to the margins: the overexploitation of fisheries. *Front. Ecol. Environ.* **1**: 102–106.
- Schnute, J.T., and Richards, L.J. 2001. Use and abuse of fishery models. *Can. J. Fish. Aquat. Sci.* **58**: 10–17.
- Schnute, J.T., Richards, L.J., and Olsen, N. 1998. Statistics, software, and fish stock assessment. *In Proceedings of the International Symposium on Fishery Stock Assessment Models for the 21st Century. Edited by F. Funk, T. Quinn II, J. Heifetz, J. Ianelli, J. Powers, J. Schweigert, P. Sullivan, and C.-I. Zhang.* Alaska Sea Grant College Program, University of Alaska, Fairbanks, Alaska.
- Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Stat.* **6**: 461–464.
- Seber, G.A., and Wild, C.J. 1989. Nonlinear regression. Wiley, New York.
- Shepherd, J.G. 1982. A versatile new stock–recruitment relationship for fisheries, and the construction of sustainable yield curves. *J. Cons. Int. Explor. Mer.* **40**: 67–75.
- Sissenwine, M.P., and Shepherd, J.G. 1987. An alternative perspective on recruitment overfishing and biological reference points. *Can. J. Fish. Aquat. Sci.* **44**: 913–918.
- Walters, C.J., and Korman, J. 1999. Linking recruitment to trophic factors: revising the Beverton–Holt model from a life history and multispecies perspective. *Rev. Fish Biol. Fish.* **9**: 187–202.
- Williams, E.H., and Shertzer, K.W. 2003. Implications of life-history invariants for biological reference points used in fishery management. *Can. J. Fish. Aquat. Sci.* **60**: 710–720.

## Appendix A

### Spawning biomass per recruit and yield per recruit

Spawning biomass and yield per recruit calculations are two basic components of the rebuilding target models. We used a standard approach for computing spawning biomass and yield per recruit curves (Gabriel et al. 1989). Life history and fishery data for the spawning biomass and yield per recruit calculations were collected from NEFSC (2002b) for all three stocks.

Spawning biomass per recruit (SPR) is the sum of expected spawning biomass at age per recruit as a function of survival at age over a cohort's life span. Given a fish stock with  $A - 1$  true ages and a plus-group consisting of fish age  $A$  and older, SPR depends on spawning weight at age ( $W_{S,a}$ ), fraction mature at age ( $D_a$ ), natural mortality ( $M$ ), fishery selectivity at age ( $P_a$ ), fishing mortality ( $F$ ), and total mortality at age ( $Z_a$ ). Note that  $Z_a$  is related to  $M$ ,  $P_a$ , and  $F$  by  $Z_a = M + FP_a$ . The probability of a fish surviving to spawn at age  $a$ , where a fraction of natural mortality ( $C_M$ ) and fishing mortality ( $C_F$ ) occurs before spawning, is the probability of surviving to age  $a$  times the probability of surviving to the beginning of the spawning season. Summing the product of the survival to spawn probabilities and the average spawning weights over all age-classes gives an expression for SPR:

$$(A1) \quad \text{SPR} = \sum_{a=1}^{A-1} D_a W_{S,a} \exp\left(-C_M M - (a-1)M - C_F P_a F - F \sum_{k=1}^{a-1} P_k\right) + D_A W_{S,A} \frac{\exp(-C_M M - (A-1)M - C_F P_A F - F \sum_{k=1}^{A-1} P_k)}{1 - \exp(-Z_A)}$$

Similarly, yield per recruit (YPR) is the sum of expected yields at age per recruit as a function of survival at age over a cohort's life span. YPR is calculated in an analogous manner as the sum of the product of the probability of survival to age  $a$ , the probability that a fish is captured at age  $a$ , and the average landed weight at age ( $W_a$ ):

$$(A2) \quad \text{YPR} = \sum_{a=1}^{A-1} \frac{W_a P_a F}{Z_a} (1 - \exp(-Z_a)) \exp\left(-C_M M - (a-1)M - F \sum_{k=1}^{a-1} P_k\right) + \frac{W_A P_A F}{Z_A} \exp\left(-C_M M - (A-1)M - F \sum_{k=1}^{A-1} P_k\right)$$

### Rebuilding target model

We used the Sissenwine–Shepherd age-structured production model to estimate rebuilding targets (Shepherd 1982; Sissenwine and Shepherd 1987). This deterministic equilibrium model requires estimates of spawning biomass per recruit, yield per recruit, and stock–recruitment curves to construct estimates of  $S_{\text{MSY}}$ . To apply this model to a given fish stock, output recruitment ( $R$ ) and input spawning stock ( $S$ ) under a stock–recruitment curve ( $f$ ) are assumed to be in equilibrium with

respect to a constant fishing mortality rate  $F$ . In particular,  $R$  is a function of spawning stock,  $R = f(S, \theta)$ , where  $\theta$  is the parameter vector of the stock–recruitment curve. At equilibrium, the product of spawning biomass per recruit at  $F$ , denoted by  $\text{SPR}(F)$ , and  $R$  equals  $S$ , i.e.,  $S(F) = \text{SPR}(F)f(S, \theta)$ . Solving this equation for  $S$  in terms of  $F$  and  $\theta$  implicitly determines equilibrium recruitment as  $R(F) = f(S(F), \theta)$ . Thus, equilibrium yield ( $Y$ ) at fishing mortality  $F$  is the product of yield per recruit at  $F$  ( $\text{YPR}(F)$ ) and  $R$ , i.e.,  $Y(F) = \text{YPR}(F)R(F)$ . The maximum value of  $Y(F)$  can be numerically determined by searching over a grid of  $F$  values. This determines  $\text{MSY}$ ,  $S_{\text{MSY}}$ , and  $F_{\text{MSY}}$ .

### Approximate variance of equilibrium spawning biomass

In this section, an expression for the approximate variance of equilibrium spawning biomass is derived. Equilibrium spawning biomass ( $S^*$ ) is the product of equilibrium recruitment ( $R^*$ ) and spawning biomass per recruit ( $\text{SPR}^*$ ) at the equilibrium fishing mortality:

$$(A3) \quad S^* = R^* \text{SPR}^*$$

A first-order Taylor series expansion of  $S^*$  as a function of  $R^*$  and  $\text{SPR}^*$  approximates the variance of equilibrium spawning biomass ( $\text{VAR}[S^*]$ ) in terms of the variances of equilibrium recruitment ( $\text{VAR}[R^*]$ ) and spawning biomass per recruit at the equilibrium fishing mortality ( $\text{VAR}[\text{SPR}^*]$ ) as

$$(A4) \quad \text{VAR}[S^*] = \left( \frac{\partial S^*}{\partial R^*} \right)^2 \text{VAR}[R^*] + \left( \frac{\partial S^*}{\partial \text{SPR}^*} \right)^2 \text{VAR}[\text{SPR}^*]$$

From eq. A3, the partial derivatives of equilibrium spawning biomass with respect to equilibrium recruitment and spawning biomass per recruit are

$$(A5) \quad \frac{\partial S^*}{\partial R^*} = \text{SPR}^* \quad \text{and} \quad \frac{\partial S^*}{\partial \text{SPR}^*} = R^*$$

At equilibrium, the expected values of recruitment and spawning biomass per recruit are constant and equal to  $R^*$  and  $\text{SPR}^*$ . As a result, the coefficients of variation of equilibrium recruitment ( $\text{CV}[R^*]$ ) and spawning biomass per recruit at equilibrium fishing mortality ( $\text{CV}[\text{SPR}^*]$ ) are

$$(A6) \quad \text{CV}[R^*] = \frac{\sqrt{\text{VAR}[R^*]}}{R^*}$$

and

$$\text{CV}[\text{SPR}^*] = \frac{\sqrt{\text{VAR}[\text{SPR}^*]}}{\text{SPR}^*}$$

Solving for  $R^*$  and  $\text{SPR}^*$  in eq. A6 and substituting this result along with eq. A5 into eq. A4 gives the approximate variance of equilibrium spawning biomass in terms of the coefficients of variation of  $R^*$  and  $\text{SPR}^*$  in eq. 7.

### Appendix references

- Gabriel, W., Sissenwine, M., and Overholtz, W. 1989. Analysis of spawning stock biomass per recruit: an example for Georges Bank haddock. *N. Am. J. Fish. Manag.* **9**: 383–391.
- NEFSC. 2002b. Assessment of 20 northeast groundfish stocks through 2001. A report of the Groundfish Assessment Review Meeting (GARM), Woods Hole, Massachusetts, 8–11 October 2002. Ref. Doc. 02-16. National Marine Fisheries Service, Northeast Fisheries Science Center, Woods Hole, Mass. Available at <http://www.nefsc.nmfs.gov/nefsc/publications/crd/crd0216/>.
- Shepherd, J.G. 1982. A versatile new stock–recruitment relationship for fisheries, and the construction of sustainable yield curves. *J. Cons. Int. Explor. Mer.* **40**: 67–75.
- Sissenwine, M.P., and Shepherd, J.G. 1987. An alternative perspective on recruitment overfishing and biological reference points. *Can. J. Fish. Aquat. Sci.* **44**: 913–918.