

# Model averaging and probable status of North Pacific striped marlin, *Tetrapturus audax*

Jon Brodziak and Kevin Piner

**Abstract:** We show how model averaging can be applied to estimate the probable status of a fishery resource under assessment scenario uncertainty. This approach is applied to North Pacific striped marlin (*Tetrapturus audax*), an apex predator that may be vulnerable to recruitment overfishing in pelagic longline fisheries targeting tunas. In the current striped marlin assessment, two assessment scenarios were used to account for different hypotheses about the steepness of the stock–recruitment dynamics. Estimates of spawning stock and recruitment from these scenarios are used post hoc to fit age-structured production models that represent alternative hypotheses about the degree of compensation in stock–recruitment dynamics and the degree of serial correlation of environmental forcing. Model-averaged estimates of target spawning biomass to produce maximum sustainable yield ( $S_{MSY}$ ) and the associated limit fishing mortality ( $F_{MSY}$ ) characterize relative stock status ( $S/S_{MSY}$  and  $F/F_{MSY}$ ) under each scenario. Scenario-weighted averages of relative status determine probable stock status, with weightings reflecting the credibility of each scenario. Estimates of the variance of probable status account for both model selection and assessment scenario uncertainty in risk analyses. Using model averaging to estimate probable stock status from multiple assessment scenarios is analogous to using ensemble averages from multiple predictive models to make weather forecasts.

**Résumé :** Nous démontrons qu'on peut utiliser le calcul de la moyenne des modèles pour estimer le statut probable d'une ressource halieutique dont l'évaluation est en état d'incertitude. Nous utilisons cette méthode chez le makaire strié (*Tetrapturus audax*) du Pacifique Nord, un prédateur du sommet du réseau trophique, qui peut être vulnérable à la surpêche du recrutement dans les pêches pélagiques à la palangre qui ciblent les thons. Dans l'évaluation actuelle des makaires striés, nous avons utilisé deux scénarios d'évaluation pour tenir compte d'hypothèses différentes sur le degré d'inclinaison de la dynamique stock–recrutement. Les estimations du stock reproducteur et du recrutement tirées de ces scénarios ont servi à ajuster a posteriori les modèles de production structurés en fonction de l'âge qui représentent les hypothèses de rechange au sujet du degré de compensation dans la dynamique stock–recrutement et du degré d'autocorrélation du forçage environnemental. Les estimations obtenues par le calcul des moyennes des modèles de la biomasse ciblée des reproducteurs nécessaire pour produire le rendement maximal tolérable ( $S_{MSY}$ ) et la mortalité limite due à la pêche ( $F_{MSY}$ ) correspondante caractérisent le statut relatif du stock ( $S/S_{MSY}$  et  $F/F_{MSY}$ ) dans chacun des scénarios. Les moyennes du statut relatif pondérées en fonction du scénario déterminent le statut probable du stock dans lequel les pondérations reflètent la crédibilité de chacun des scénarios. Les estimations de la variance du statut probable expliquent à la fois le choix du modèle et l'incertitude du scénario d'évaluation dans les analyses du risque. Le calcul de la moyenne des modèles afin d'estimer le statut probable à partir de plusieurs scénarios d'évaluation est analogue à l'utilisation de la moyenne d'ensemble de plusieurs modèles prédictifs pour faire des prévisions météorologiques.

[Traduit par la Rédaction]

## Introduction

Uncertainty is inherent in fishery systems and is a key feature to characterize in quantitative assessments of fisheries resources (Hilborn and Walters 1992). Several types of uncertainty have been recognized in the stock assessment process (Francis and Shotton 1997; Hilborn 1997), including process, observation, model, and estimation uncertainties. Of these, the characterization of estimation uncertainty, which derives from the other types as a secondary source of uncer-

tainty, has typically received much more emphasis in stock assessments (e.g., Hilborn and Peterman 1996). The focus of this paper is on providing a framework to jointly characterize both model and estimation uncertainty within a given stock assessment in which multiple models or assessment scenarios have been developed for inference.

Model uncertainty is an important feature to explicitly consider when doing stock assessments (Hilborn 1997) or developing predictive models for ecosystem-based management of fisheries resources (Hill et al. 2007). Model uncer-

Received 21 August 2009. Accepted 8 March 2010. Published on the NRC Research Press Web site at [cjfas.nrc.ca](http://cjfas.nrc.ca) on 16 April 2010. J21357

Paper handled by Associate Editor Ray Hilborn.

**J. Brodziak.**<sup>1</sup> Pacific Islands Fisheries Science Center, Honolulu, HI 96822-2396, USA.

**K. Piner.** Southwest Fisheries Science Center, La Jolla, CA 92037-1508, USA.

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

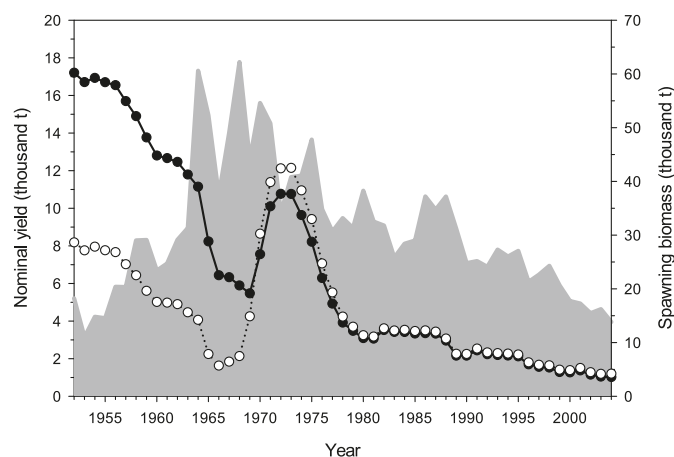
tainty is a generic issue in any approximation and arises due to a lack of complete knowledge of the state of nature and its dynamics (Fogarty et al. 1996). One recommended approach for dealing with model uncertainty is to identify a set of plausible causal mechanisms and associated models that could explain the observed data or processes. The explanatory power of the alternative models representing competing hypotheses is then compared in a likelihood-based framework for inference (Hilborn and Mangel 1997; McAllister et al. 1999). This is analogous to the so-called method of multiple working hypotheses (see, for example, Hilborn and Mangel 1997). In contrast, estimation uncertainty is the direct quantification of parameter precision within a particular model using standard statistical techniques such as likelihood profiling, bootstrapping, Markov chain Monte Carlo simulation, and sampling importance resampling (e.g., Patterson et al. 2001).

Model selection and model averaging techniques (Burnham and Anderson 2002; Claeskens and Hjort 2008) are powerful tools that can provide an objective means to judge the goodness of fit of alternative models with fixed data. In this paper, we show how model averaging can be used to estimate the probable status of a fishery resource under stock assessment scenario uncertainty. In particular, alternative age-structured production models of the compensatory behavior of recruitment dynamics of North Pacific striped marlin were fit post hoc to estimates of spawner abundance and recruitment from alternative stock assessment scenarios. Model-averaged results under each scenario were then weighted using the relative credibility of each scenario and averaged to ascertain the probable status of North Pacific striped marlin with respect to biological reference points.

From a biological perspective, striped marlin (*Tetrapturus audax* or *Kajikia audax*) are a fast-growing billfish species distributed in tropical and temperate waters of the Pacific and Indian oceans. This epipelagic predator is primarily harvested as a by-product of longline fisheries targeting tunas. Striped marlin are also targeted in coastal fisheries off Japan and Taiwan and support valuable recreational fisheries off the coasts of Australia, New Zealand, Mexico, and the United States. Striped marlin are late-maturing (females mature at ~6 years) and have a relatively long life span ( $\geq 12$  years) for a tuna-like species. As a result, striped marlin may be vulnerable to recruitment overfishing in longline fisheries that target tunas and in other fisheries.

The status of North Pacific striped marlin is not clear because of assessment scenario uncertainty and a lack of agreed-upon biological reference points. Nominal catches of striped marlin in the North Pacific have had a significant declining trend of roughly 1.8% per year since 1965 (Fig. 1;  $R^2 = 0.78$ ,  $P < 0.001$ ). The two alternative assessment scenarios from the most recent International Committee on Tuna and Tuna-Like Species Billfish Working Group (WG) stock assessment (International Scientific Committee for Tuna and Tuna-like species in the North Pacific Ocean (ISC) 2007) were based on alternative assumptions about the steepness of the stock–recruitment relationship; these were a moderately resilient spawner–recruit curve and recruitment randomly fluctuating about a constant mean. Under these two scenarios, spawning biomass of the North

**Fig. 1.** North Pacific striped marlin (*Tetrapturus audax*) nominal annual catches (shaded area) and estimates of spawning biomass for two scenarios of what governs recruitment strength: scenario 1, spawner–recruit hypothesis (solid circles); scenario 2, environmentally driven recruitment hypothesis (open circles), 1952–2004. t, tonnes.



Pacific striped marlin stock was estimated to have been fished down to between 6% and 16% of its 1952 abundance (Fig. 1). This low range of spawner abundance suggested that the stock was currently in a depleted condition. Similarly, estimates of current fishing mortality corresponded to an average spawning potential ratio of about 9% under both assessment scenarios (ISC 2007), suggesting the stock may be experiencing overfishing as well (e.g., Mace and Sissenwine 1993). However, because no overfishing thresholds have been established for this highly migratory stock, current stock status was considered to be uncertain. Overall, the WG concluded that biological reference points for North Pacific striped marlin needed to account for uncertainty about stock resilience, as reflected by the steepness of the stock–recruitment relationship.

As a result, two primary uncertainties needed to be resolved to evaluate the probable status of North Pacific striped marlin based on the most recent stock assessment. First, there was uncertainty in the appropriate model structure and error assumptions to be used in the age-structured production model for estimating reference points to quantify the risk of overfishing. In the stock assessment scenarios used for North Pacific striped marlin, a Beverton–Holt (BH) curve was used to fit the observed data and estimate spawning stock and fishing mortality. Although both stock assessment scenarios assumed the BH curve with lognormal multiplicative error and differing values of steepness, it was also logically possible that striped marlin recruitment may be reduced at high stock size as a result of intraspecific cannibalism or other strongly compensatory factors such as density-dependent juvenile growth. To address the possibility of overcompensatory recruitment dynamics, a Ricker (RK) curve provided an alternative hypothesis to the BH curve. In addition, serially correlated and uncorrelated error terms were used to evaluate whether or not annual variability about expected recruitment is subject to low-frequency forcing. Thus, one key uncertainty in the stock assessment was the choice of the most appropriate stock–recruitment

curve and error structure for modeling recruitment dynamics.

Second, there was uncertainty about the appropriate assessment scenario to be used for developing target or limit reference points for this stock. In the most recent stock assessment of North Pacific striped marlin (Piner et al. 2007), two assessment scenarios were provided to address the issue of what regulates the magnitude of recruitment. The first scenario represented the hypothesis that the striped marlin stock was moderately resilient to exploitation, assuming a BH stock–recruitment relationship. In the second scenario, striped marlin recruitment was assumed to be fluctuating about a constant mean as a result of changing oceanographic conditions, a common assumption for Pacific tuna stock assessments. According to the WG consensus, each scenario represented an equally plausible hypothesis about resource dynamics over the range of observed data, and each generated a time series of spawning stock and recruitment estimates. Thus, a framework for dealing with assessment scenario uncertainty was needed.

## Materials and methods

In this section, a general framework to describe the probable status of North Pacific striped marlin is developed that accounts for uncertainty in model structure and assessment scenario. An age-structured production model is applied to calculate external estimates of maximum sustainable yield (MSY) based biological reference points using stock–recruitment estimates from the two assessment scenarios and a Bayesian model averaging approach developed by Brodziak and Legault (2005). MSY reference points are the focus of this analysis because they are currently used to guide the management of several tuna stocks in the Pacific and also because they provide a natural biological limit for assessing the issue of overfishing as an indicator for ecosystem-based fishery management (Mace 2001; Brodziak and Link 2002). Model-averaged estimates of target spawning biomass to produce maximum sustainable yield ( $S_{MSY}$ ) and the associated limit fishing mortality ( $F_{MSY}$ ) characterize relative stock status ( $S/S_{MSY}$  and  $F/F_{MSY}$ ) under each scenario. Scenario-weighted averages of relative status are used to determine probable stock status, with weightings reflecting the credibility of each scenario. Estimates of the variance of probable stock status that account for both model selection and assessment scenario uncertainty are developed for risk analyses. Sensitivity analyses for the relative credibility of assessment scenarios provide insight into the robustness of estimates of MSY reference points and probable stock status to alternative scenario weightings.

### Estimates of spawning stock and recruitment

Estimates of spawning biomass and recruitment were obtained from two stock scenarios judged to be equally plausible in the most recent stock assessment (ISC 2007; Piner et al. 2007). These scenarios were (i) the spawner–recruit hypothesis (scenario 1) for a moderately resilient stock with stock–recruitment steepness  $h = 0.7$ , assuming a Beverton–Holt recruitment curve, and (ii) the environmentally driven recruitment hypothesis in which recruitment fluctuated about a constant mean (scenario 2) with steepness  $h = 1$  (Fig. 2a;

$h$  is defined as the fraction of unfished recruitment,  $R_0$ , expected when spawner abundance,  $S$ , is at 20% of unfished spawner abundance,  $S_0$ ). A total of 40 stock–recruitment estimates during 1965–2004 were used for fitting parameters of the age-structured production model to estimate MSY reference points external to the stock synthesis assessment model (version SS2; Methot 2000), which was used for assessment modeling. The initial year of 1965 was selected to match when recruitment deviations were freely estimated in scenario 2. The ending cutoff of 2004 was selected to match the most recent period that showed complete catch data. As a result, there were two sets of stock–recruitment estimates for evaluating MSY-based reference points along the primary axis of assessment uncertainty, the steepness parameter.

### Life history parameters

Estimates of weight at age and probability of maturity at age (Fig. 2b), as well as aggregate fishery selectivities at age from the two stock assessment scenarios, were used to fit the age-structured production model (ISC 2007; Piner et al. 2007). Striped marlin population age structure was approximated with a total of 21 age classes representing ages 0 through 19 fish and a plus group consisting of fish ages 20 and older. However, the probability of survival to ages 20 and older would be very low (<0.3%) in an unfished stock under a natural mortality rate of  $M = 0.30$ , and the relative contribution of the plus group to spawning biomass was relatively minor.

### Fishery selectivity

Under each assessment scenario, aggregate fishery selectivities for the set of international fisheries that capture striped marlin were calculated for two time periods as assumed in the stock assessment: 1965–1979 and 1980–2004 (Figs. 2c and 2d). Individual fishing fleet selectivities were modeled as constant during these two periods to reflect changes in the targeting of tuna species by Japanese distant-water longline vessels that occurred during the mid-1970s to 1980 (Ward and Hindmarsh 2007). Under each scenario, the overall aggregate fishery selectivity for striped marlin reference point estimation was calculated as the average of annual aggregate selectivities during each period. Annual aggregate selectivities were computed as a catch-weighted average of the individual fleet selectivities (ISC 2007). In this case, individual fleet selectivities were constant within each time period, whereas catch proportions by fleet and aggregate fleet selectivities varied annually.

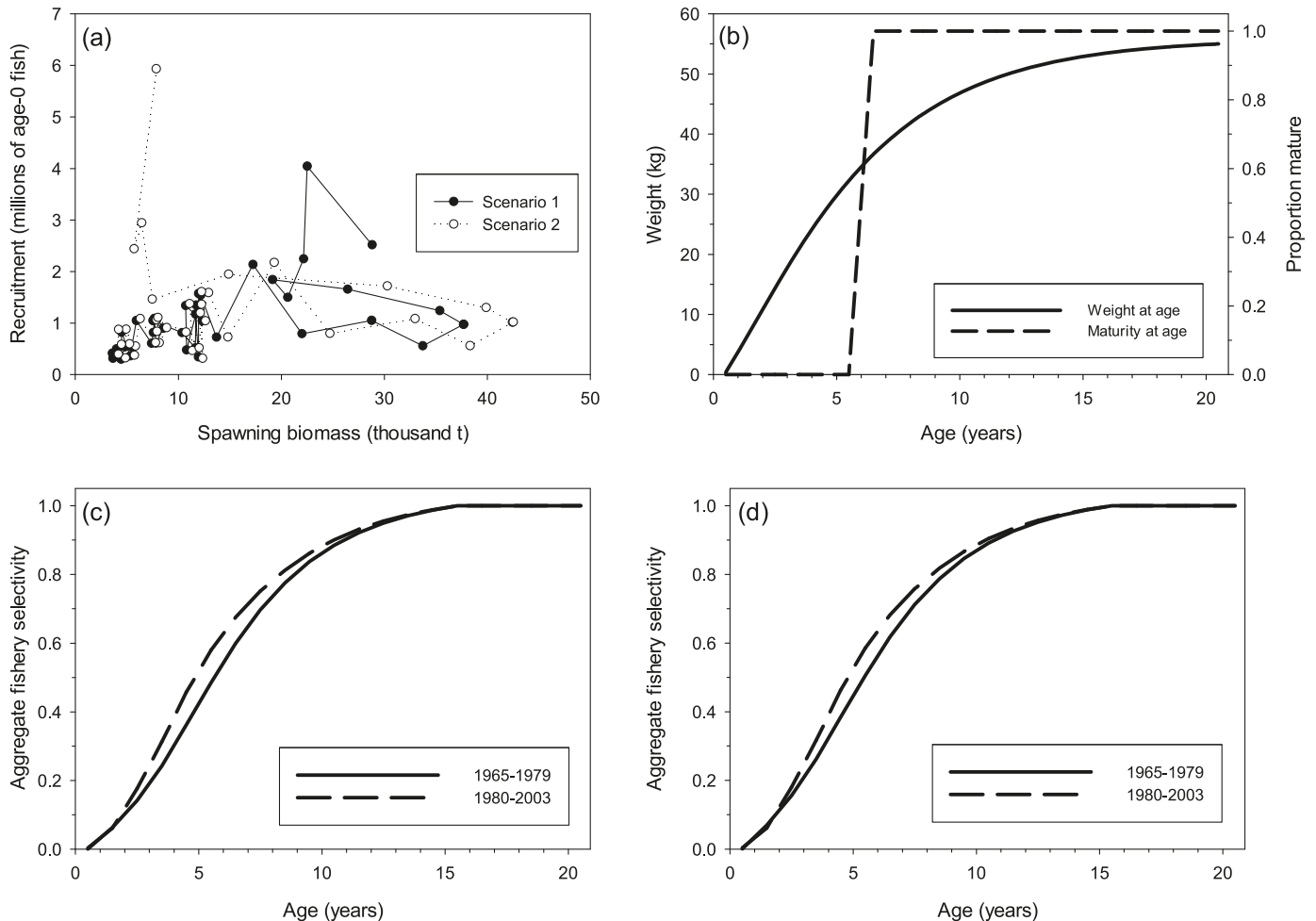
### Stock–recruitment models

Two alternative structural models for striped marlin recruitment dynamics were fit under alternative assumptions about the error terms describing deviations from the structural models. One model was the compensatory, flat-topped Beverton–Holt curve, which relates expected recruitment ( $R$ ) to spawner abundance ( $S$ ) via

$$(1) \quad R = \frac{4hR_0S}{S_0(1-h) + S(5h-1)}$$

where steepness ( $h$ ) and unfished recruitment ( $R_0$ ) were parameters to be estimated, with  $h$  in the interval [0.2, 1]

**Fig. 2.** (a) Stock–recruitment for scenario 1 (solid circles) and scenario 2 (open circles), (b) weight at age (continuous line) and maturity at age (broken line) data, and (c–d) time-varying aggregate fishery selectivity estimates during 1965–1979 (continuous line) and 1980–2003 (broken line) under assessment scenarios 1 (c) and 2 (d) used for age-structured production model analyses of North Pacific striped marlin (*Tetrapturus audax*). t, tonnes.



and  $R_0 > 0$ . This was the assumed model for the average stock–recruitment relationship used in the assessment. The other model was the overcompensatory, dome-shaped Ricker curve for which expected recruitment decreases at high spawning biomasses via

$$(2) \quad R = aS \cdot \exp(-bS) \\ = \frac{S}{\Phi_0} \exp\left(\frac{5}{4} \log(5h) \left(1 - \frac{S}{R_0 \Phi_0}\right)\right)$$

where parameters  $a$  and  $b$  are to be estimated, with  $a > 0$  and  $b > 0$ , respectively, and the second term on the right-hand side of eq. 2 represents the steepness parameterization of the Ricker curve (Michielsens and McAllister 2004), with steepness  $h \geq 0.2$  and unfished spawning stock per recruit ( $\Phi_0$ ). These models represented two alternative hypotheses about recruitment strength at high stock sizes, with recruitment approaching an asymptote as spawner abundance increases under the Beverton–Holt assumption or with recruitment eventually declining under the Ricker assumption.

### Error assumptions

The multiplicative error terms for fitting the age-structured production model to the annual stock–recruitment estimates from the stock assessment model were assumed to have lognormal distributions with an expected value of unity and constant variance. The assumption of a lognormal error term provided a positively skewed distribution for recruitment deviations from the expected model. In particular, the expected stock–recruitment value for the  $i$ th data point, denoted by  $f(S_i)$ , was multiplied by a lognormal term with random shock  $\varepsilon_i$  to obtain the observation error model

$$(3) \quad R_i = f(S_i) \exp(\varepsilon_i)$$

The annual error terms were independent and identically distributed (iid) uncorrelated or autocorrelated random variables with a 1-year lag. In the uncorrelated case, the temporal dependence of deviations from the stock–recruitment curve was assumed to be negligible across years. The random shocks  $\varepsilon_i$  were normally distributed, with mean of  $-0.5\sigma^2$  and variance of  $\sigma^2$ . This choice set the mean of the lognormal error term at 1, whereas the error variance  $\sigma^2$  was



a parameter to be estimated. In contrast, assuming correlated errors provided an explicit model for temporal dependence in annual recruitment deviations. Such dependence might be expected if striped marlin recruitment was strongly influenced by low-frequency environmental forcing such as the Pacific Decadal Oscillation. Under the correlated error assumption, the random shocks  $\varepsilon_i$  had a lag-1 autocorrelation in which

$$(4) \quad \varepsilon_i = \varphi \varepsilon_{i-1} + w_i \quad \text{where } |\varphi| < 1$$

$$\text{and } w_i \sim N(-0.5\sigma_w^2, \sigma_w^2)$$

In the correlated case, the expected value of the multiplicative error was also unity and the error variance ( $\sigma_w^2$ ) and autoregressive coefficient ( $\varphi$ ) were parameters to be estimated. The uncorrelated error assumption is a special case of the correlated error assumption with autoregressive coefficient  $\varphi = 0$ . Testing for differences in model fits between these nested models is one of the common uses of model selection or averaging, that is, to select a parsimonious set of predictive variables that can be estimated with reasonable precision (Burnham and Anderson 2002).

### Age-structured production model

An age-structured production model (see, for example, Sissenwine and Shepherd 1987; Brodziak and Legault 2005) was fit to the two sets of stock–recruitment estimates. This model produces estimates of MSY reference points, including MSY, the fishing mortality to produce MSY ( $F_{\text{MSY}}$ ), and the spawning biomass to produce MSY ( $S_{\text{MSY}}$ ) conditioned on the estimation of the parameters of the assumed stock–recruitment curve, including the parameters that determine the observation model error structure, e.g., the error variance and the autocorrelation coefficient, if appropriate. The MSY reference point estimates were also dependent on the input size at age and maturity at age vectors, natural mortality, and the fishery selectivity at age vectors. Because fishery selectivity estimates differed for the two time periods used in the stock assessment, separate estimates of MSY reference points were calculated for each selectivity period (1965–1979 and 1980–2003) under each scenario. As a result, there were two sets of MSY-based reference points for each steepness scenario.

### Candidate models

The stock–recruitment models and error assumptions resulted in a set of four candidate models of recruitment dynamics under each stock assessment scenario: the Beverton–Holt curve with uncorrelated errors (BHU), the Beverton–Holt curve with correlated errors (BHC), the Ricker curve with uncorrelated errors (RKU), and the Ricker curve with correlated errors (RKC). As a result, a total of four age-structured production models were fit to each set of stock–recruitment estimates. This gave a total of eight alternative models with four competing models under each steepness scenario for the model averaging analysis.

### Parameter estimation

Model parameters were estimated using a Bayesian model averaging approach (Brodziak and Legault 2005) as implemented in the NOAA Fisheries Toolbox Module SRFIT version 7.0 (available at [nft.nefsc.noaa.gov/](http://nft.nefsc.noaa.gov/)). This Bayesian

approach requires the specification of prior distributions for model parameters. Broad uniform priors ( $\pi$ ) were assumed for each model parameter, including the error standard deviation ( $\sigma$ ), the autoregressive coefficient ( $\varphi$ ), the log-transform of unfished recruitment of the Beverton–Holt curve ( $\log(R_0)$ ), the steepness value of the Beverton–Holt curve ( $h$ ), the log-transform of slope at the origin of the Ricker curve ( $a$ ), and the density dependence parameter of the Ricker curve ( $b$ ):

$$(5) \quad \pi_\sigma = U[10^{-4}, 4], \quad \pi_\varphi = U[-1, 1]$$

$$\pi_{\log(R_0)} = U[-10, 100], \quad \pi_h = U[0.2, 1]$$

$$\pi_{\log(a)} = U[-10, 10], \quad \pi_b = U[10^{-6}, 0]$$

Given these uninformative priors, inferences about striped marlin reference points were essentially based on the stock–recruitment data under each steepness scenario.

After estimating the posterior mode of each candidate model, Markov chain Monte Carlo (MCMC) simulation (Gilks et al. 1996) was applied to numerically sample two chains of length 600 000 from the posterior distribution using SRFIT. Each chain was thinned by 100 iterates to eliminate the within-chain autocorrelation, and the first 1000 thinned iterates were excluded to eliminate potential dependence on initial conditions. Five thousand iterates remained for numerical inference using the first chain. Thinned iterates from the two chains were used to assess convergence of MCMC sampling based on the potential scale reduction factor (e.g., Gelman et al. 1995), and in all cases, this factor was approximately unity, indicating support for convergence to the posterior distribution.

### Posterior model probabilities

For each model scenario, the likelihoods of the four candidate models were compared using the Bayesian information criterion (BIC; Schwarz 1978) to approximate the Bayes factor of each posterior sample and then averaging the resulting model probabilities. The BIC value for the  $i$ th model with maximized value  $L$  for the stock–recruitment likelihood function,  $p$  parameters, and  $n$  data points was

$$(6) \quad \text{BIC}_i = -2 \cdot L + p \cdot \log(n)$$

The BIC values for each model were the sum of the model deviance plus a parameter penalty that increased with the logarithm of the number of stock–recruitment estimates. Each model was assigned an equal prior weight of  $\Pr(M_i) = 1/4$  to reflect an a priori lack of preference for the candidate models. The exponential of minus one-half times the difference in BIC values between the  $i$ th and the best-fitting model ( $M_0$ ) at the  $k$ th iterate ( $\Delta_i^{(k)}$ ) provided a first-order approximation of the Bayes factor ( $B_{i,0}^{(k)}$ ), which measured the relative odds that model  $M_i$  versus  $M_0$  was the true model (see, for example, Brodziak and Legault 2005) as

$$(7) \quad B_{i,0}^{(k)} \approx \exp\left(-\frac{1}{2}(\text{BIC}_i^{(k)} - \text{BIC}_0^{(k)})\right) \\ = \exp\left(-\frac{1}{2}\Delta_i^{(k)}\right)$$

Given the estimates of the Bayes factors calculated from the set of 5000 MCMC samples, the posterior probability that model  $M_i$  was the true model under assessment scenario  $j$ , denoted as  $(\Pr(M_i|D_j))$ , was then calculated as

$$(8) \quad \Pr(M_i|D_j) = \frac{\Pr(M_i) \cdot B_{i,0}}{\sum_m \Pr(M_m) \cdot B_{m,0}} \\ \approx \frac{1}{5000} \sum_{k=1}^{5000} \frac{\exp(-0.5\Delta_i^{(k)})}{\sum_{m=1}^4 \exp(-0.5\Delta_m^{(k)})}$$

Thus, the posterior model probabilities were a function of the Bayes factors given that the prior model probabilities were equal.

$$(10) \quad V_M[S_{MSY,j}|D_j] = \left[ \sum_{i=1}^4 \Pr(M_i|D_j) \cdot \sqrt{V_M[S_{MSY,j}|M_i, D_j] + (E_M[S_{MSY,j}|D_j] - E_M[S_{MSY,j}|M_i, D_j])^2} \right]^2$$

The model-averaged variance estimate was the square of the weighted average of the square root of two terms: (i) the conditional model variance estimates and (ii) the squared difference between the model-averaged expected value and the conditional expected values. The first term measured the variability resulting from within-model variance, whereas the second term measured the variability resulting from between-model differences in expected value.

### Relative stock status

The relative status of North Pacific striped marlin under each steepness scenario was then determined as the ratio of assessment estimates to the model-averaged estimates of MSY reference points for biomass and fishing mortality. Relative status under steepness scenario  $j$  was based on the relative spawning biomass ( $S_j^*$ ) and relative fishing mortality ( $F_j^*$ ) in each year, where relative spawning biomass and fishing mortality from scenario  $j$  in year  $t$  were

$$(11) \quad S_j^*(t) = \frac{S_j(t)}{E_M[S_{MSY,j}|D_j]} \quad \text{and} \\ F_j^*(t) = \frac{F_j(t)}{E_M[F_{MSY,j}|D_j]}$$

The values of  $S_{MSY,j}$  and  $F_{MSY,j}$  corresponded to the fishery selectivity time period of year  $t$ , i.e.,  $t$  was either in the interval [1952, 1979] or in [1980, 2003]. Assessment estimates of the average spawning biomass and fishing mortality estimates during 2001–2003 were used to characterize the current relative status of the stock based on the WG consensus from the most recent assessment (ISC 2007). Estimates of relative spawning biomass and relative fishing mortality were also used to construct Kobe plots showing the time trends of these indicators of stock status under each scenario.

### Model-averaged estimates

Given the posterior model probabilities, the model-averaged expectations of parameter estimates ( $E_M$ ), such as the spawning biomass to produce MSY under scenario  $j$ , denoted as  $S_{MSY,j}$ , were computed as the weighted average of the four conditional model expectations in which the weights were the posterior model probabilities via

$$(9) \quad E_M[S_{MSY,j}|D_j] = \sum_{i=1}^4 \Pr(M_i|D_j) \cdot E_M[S_{MSY,j}|M_i, D_j]$$

Similarly, model-averaged variances ( $V_M$ ) of derived parameters such as spawning biomass to produce MSY under scenario  $j$ ,  $S_{MSY,j}$ , were computed from the four conditional model variance estimates and expected values (see, for example, Hoeting et al. 1999) as

### Probable stock status

Given the model-averaged estimates of relative status under each scenario, the probable status of North Pacific striped marlin was assessed by computing the scenario-averaged expected value of relative stock status ( $E_S$ ). In this case, we use the term “probable status” to emphasize the subjectivist or Bayesian view of probability as a measurement of the degree of belief in the model scenarios. The scenario probabilities ( $p_j$ ) for computing the scenario-averaged estimate of stock status were based on the relative credibility that the WG expressed for the alternative steepness scenarios. This led to equal scenario probabilities of  $p_1 = p_2 = 1/2$ , given the lack of preference expressed by the WG. Scenario-averaged relative biomass ( $S^*$ ) and relative fishing mortality ( $F^*$ ) in year  $t$  were then estimated by

$$(12) \quad S^*(t) = E_S[S_j^*(t)] = \sum_j p_j S_j^*(t) \quad \text{and} \\ F^*(t) = E_S[F_j^*(t)] = \sum_j p_j F_j^*(t)$$

Scenario-averaged estimates of the variance ( $V_S$ ) of relative spawning biomass ( $S^*$ ) were approximated using the delta method applied to relative spawning biomass under each scenario. The scenario-averaged variance in year  $t$  accounted for uncertainty as a result of the variability from model averaging ( $V_M$ ) and from parameter estimation ( $V_P$ ) for the estimate of spawning biomass

$$(13) \quad V_S[S^*(t)] = \sum_j p_j^2 \cdot V_{M,P}[S_j^*(t)] \\ \approx \sum_j p_j^2 \cdot \left( \frac{V_P[S_j(t)]}{E_M[S_{MSY,j}]^2} + \frac{S_j(t)^2 \cdot V_M[S_{MSY,j}]}{E_M[S_{MSY,j}]^4} \right)$$

In this first-order variance approximation, the variance of scenario-averaged spawning biomass is a weighted average of the variance of spawning biomass estimated from the assessment and the variance of the spawning biomass to produce MSY estimated from model averaging. The average annual coefficient of variation (CV) of spawning biomass was roughly 25% under scenario 1 and about 38% under scenario 2; these CVs were used to determine the parameter estimation uncertainty. Similarly, scenario-averaged variances of relative fishing mortality were calculated using the same approximation with fishing mortality substituted for spawning biomass and average annual CVs of roughly 7% and 11% under scenarios 1 and 2, respectively.

### Sensitivity to scenario probabilities

A sensitivity analysis was conducted to understand the importance of the choice of scenario probabilities on trends in probable status. In this analysis, time series of relative biomass and fishing mortality status were calculated for alternative values of the probability of scenario 1 being true of  $p_1 = \{0, 0.01, 0.02, \dots, 1\}$ . In each case, the value of  $p_2$  was determined as the complement of  $p_1$ , that is,  $p_2 = 1 - p_1$ . This sensitivity analysis was conducted to investigate the effects of emphasizing one of the two assessment scenarios versus the other in a systematic manner.

## Results

### Posterior probabilities of stock–recruitment models

Posterior model probabilities indicated that the Ricker curve was more likely under the spawner–recruit hypothesis (Table 1) during both fishery selectivity time periods (Table 1), with an evidence ratio (i.e., the ratio of the sum of posterior model probabilities representing the hypotheses) of approximately 3:1 in favor of the Ricker curve. In contrast, under the environmentally driven recruitment hypothesis, the Ricker and Beverton–Holt curves were equally likely, with an evidence ratio of roughly 1:1 during each selectivity period (Table 1). Overall, the posterior model probabilities indicated that overcompensatory recruitment dynamics were more likely under the spawner–recruit hypothesis for North Pacific striped marlin, all else being equal.

### Posterior probabilities of error assumptions

Posterior model probabilities also provided support for the competing error structure assumptions. In particular, the uncorrelated error assumption was more likely than the autocorrelated error assumption in both selectivity periods under each assessment scenario (Table 1). Evidence ratios in favor of the uncorrelated error assumption were roughly 7:1 and 3:1 under the spawner–recruit and environmentally driven recruitment scenarios, respectively. Mean estimates of the autoregressive parameter  $\phi$  were not significantly different from zero at a 5% credibility level and imprecisely estimated under the correlated error assumption, with coefficients of variation ranging from 52% to 1305%. Thus, there was consistent evidence favoring the uncorrelated error assumption in comparison with the correlated error assumption for North Pacific striped marlin recruitment dynamics and the effect of uncorrelated errors was more important under the spawner–recruit hypothesis.

### Model-averaged residuals

Model-averaged residuals suggested that, under each of the scenarios, the age-structured production model fits to the stock–recruitment data were nonrandom during 1965–1975 (Fig. 3). This time period corresponded to the initial estimation of recruitment deviations in the stock synthesis assessment model when no length frequency data were available to contrast year-class strength. Subsequent to 1975, the residual patterns of the age-structured production model fit under the spawner–recruit scenario appeared to be random (Fig. 3a). In contrast, the corresponding residual pattern under the environmentally driven recruitment scenario did not appear to conform to the assumption of random errors (Fig. 3b). Overall, the moderately resilient steepness scenario produced the better-fitting, age-structured production model.

### Model-averaged reference points

Comparing the model-averaged reference points across steepness scenarios showed that lower steepness values led to higher values of  $S_{MSY}$  and lower values of  $F_{MSY}$  (Table 1). In particular, scenario 1 ( $h = 0.7$ ) had the highest biomass and lowest fishing mortality reference points, whereas scenario 2 had the lowest biomass and highest fishing mortality reference points for both fishery selectivity time periods. In fact, model-averaged expected values of  $S_{MSY}$  and MSY were nearly identical for the two fishery selectivity time periods under both steepness scenarios (Table 1). In contrast, the variances of model-averaged estimates of  $S_{MSY}$  and MSY were generally higher for the 1980–2003 fishery selectivity period. Model-averaged expected values of  $F_{MSY}$  were slightly lower during 1980–2003 in comparison with 1965–1979 under each scenario, but variances of the  $F_{MSY}$  estimates were roughly equal across time periods. Expected values of  $F_{MSY}$  during 1965–1979 corresponded to roughly 46.6% and 28.6% of unfished spawning biomass per recruit under steepness scenarios 1 and 2, respectively. In comparison, estimates of  $F_{MSY}$  during 1980–2003 corresponded to 47.1% and 28.7% of unfished spawning biomass per recruit under scenarios 1 and 2, respectively. Overall, the differences between model-averaged MSY reference points from the two fishery selectivity periods were not consequential.

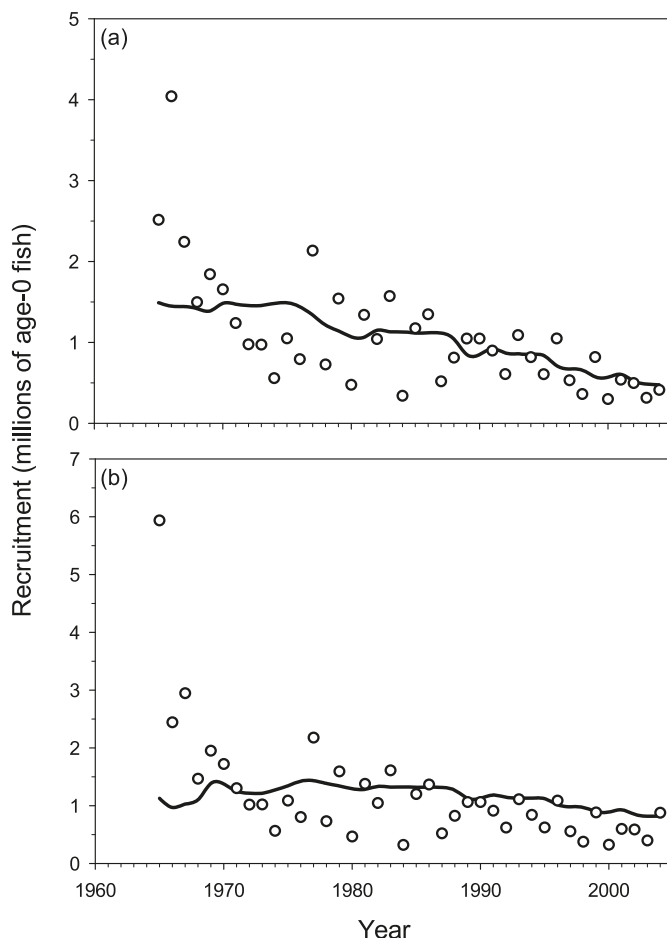
### Model-averaged estimates of relative spawning biomass and fishing mortality

Time trajectories of the model-averaged relative spawning biomass ( $S_j^*$ ) and relative fishing mortality rate ( $F_j^*$ ) exhibited similar patterns under each assessment scenario. Under scenario 1, model-averaged estimates of relative biomass (Fig. 4a) indicated that the North Pacific striped marlin spawning biomass remained above  $S_{MSY}$  during 1952–1977. Spawning biomass dropped below  $S_{MSY}$  in 1978 and  $S_1^*$  has continued to decline since that time, with a current value of roughly  $S_1^* \approx 0.3$  (Fig. 4a). Similarly, relative fishing mortality under scenario 1 was below unity until the early 1960s, when fishing mortality first exceeded  $F_{MSY}$ . Relative fishing mortality increased steadily during 1963–1978, then increased rapidly during the 1980s, before fluctuating around  $F_1^* \approx 3.5$  since the 1990s. Under steepness scenario 2, estimates of relative biomass fluctuated during the 1950s through the 1970s, whereas relative fishing mortality re-

**Table 1.** Posterior model probabilities and model-averaged estimates of spawning biomass to produce maximum sustainable yield,  $MSY$  ( $S_{MSY}$ , kilotonnes (kt)), fishing mortality rate to produce  $MSY$  ( $F_{MSY}$ , year<sup>-1</sup>), and  $MSY$  (kt) along with model-averaged standard deviations ( $\sigma$ ) under two assessment steepness ( $h$ ) scenarios assuming 1965–1979 and 1980–2003 average aggregate fishery selectivities.

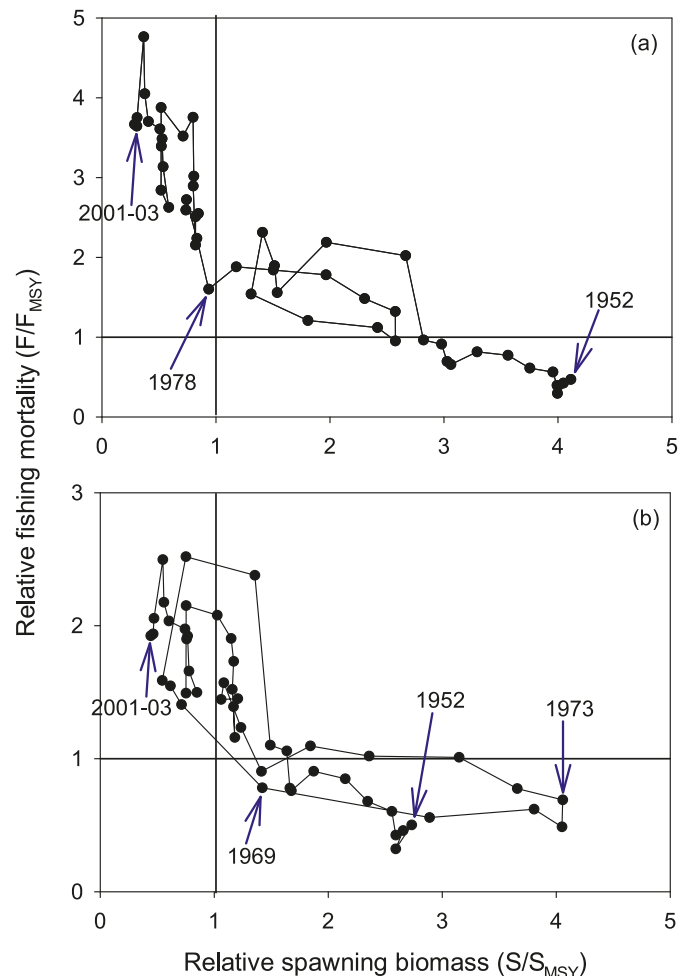
Steepness scenario	Model probabilities				Model-averaged reference points		
	BHU	BHC	RKU	RKC	$S_{MSY}$	$F_{MSY}$	$MSY$
<b>(a) Average fishery selectivity from 1965 to 1979</b>							
Scenario 1 ( $h = 0.7$ )	0.23	0.03	0.64	0.10	14.6 (7.6)	0.22 (0.04)	4.4 (1.3)
Scenario 2 ( $h = 1$ )	0.34	0.14	0.38	0.14	10.5 (8.3)	0.38 (0.15)	6.1 (5.3)
<b>(b) Average fishery selectivity from 1980 to 2003</b>							
Scenario 1 ( $h = 0.7$ )	0.24	0.03	0.63	0.10	14.6 (5.3)	0.19 (0.03)	4.4 (1.1)
Scenario 2 ( $h = 1$ )	0.35	0.14	0.37	0.14	10.5 (8.0)	0.34 (0.13)	6.1 (3.0)

**Fig. 3.** Predicted (line) and observed recruitments (open circles) of North Pacific striped marlin (*Tetrapturus audax*) under (a) model scenario 1, a moderately resilient stock, and (b) model scenario 2, a stock with recruitment fluctuating about a constant mean, during 1965–2004.



maintained at or below unity (Fig. 4b). During the 1980s and 1990s, however, fishing mortality generally increased as spawning biomass decreased. Since the late 1990s, the stock has experienced excess fishing mortality with  $F_2^* \approx 2$  and has been approaching a depleted condition with  $S_2^* \approx 0.5$  (Fig. 4b). Although relative fishing mortality was low and relative spawning biomass was high during the 1950s

**Fig. 4.** Trajectories of relative fishing mortality and relative biomass from model-averaged estimates of  $F_{MSY}$  and  $S_{MSY}$  (solid circles) during 1952–2000 and the recent average estimates during 2001–2003 under (a) scenario 1 and (b) scenario 2.

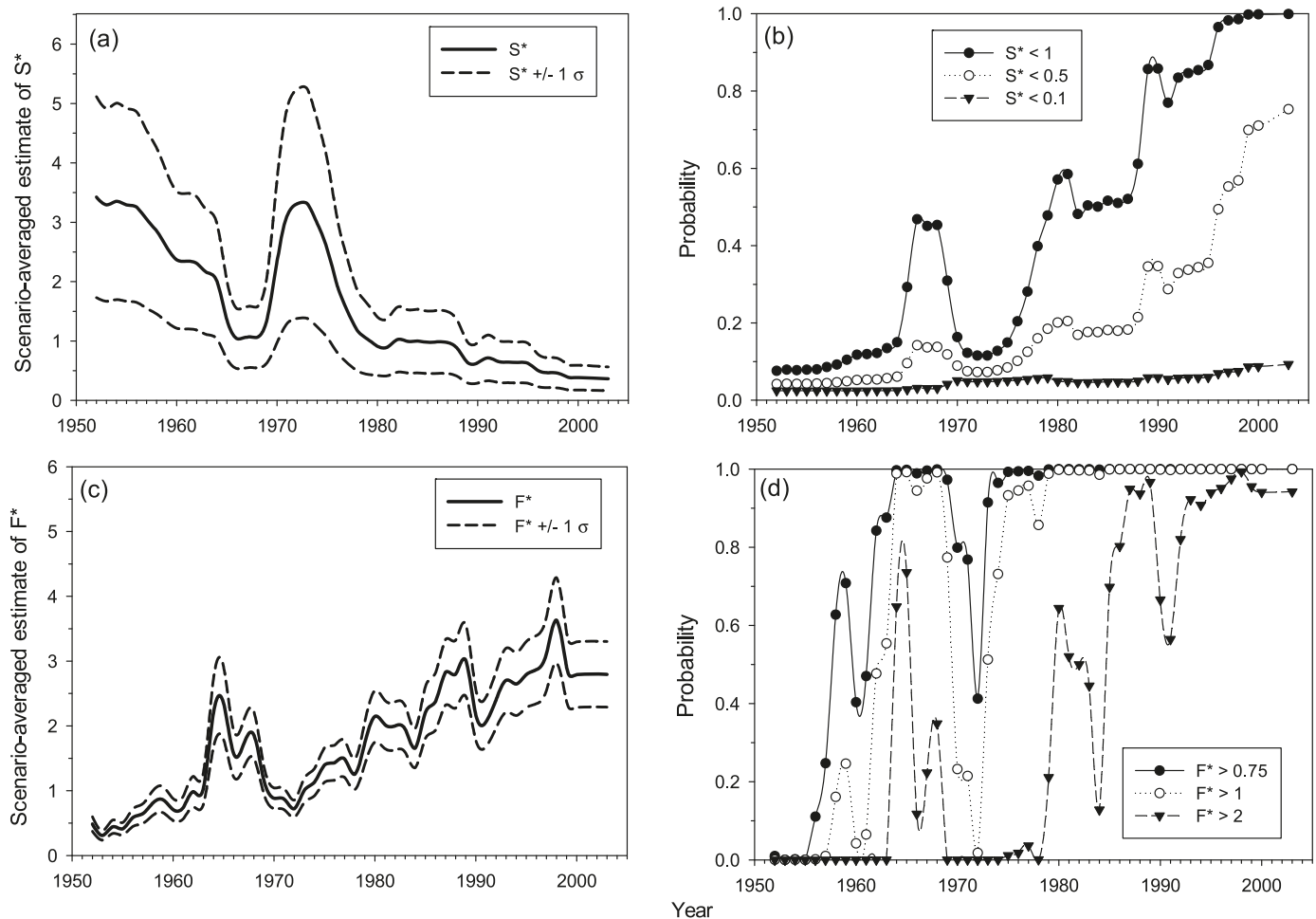


through the 1970s, this pattern was reversed by the late 1990s under each assessment scenario.

Model-averaged estimates of current relative status indicated that North Pacific striped marlin was experiencing overfishing under each steepness scenario. Expected values of relative fishing mortality rates during 2001–2003 were 367% of  $F_{MSY}$  under scenario 1 and 190% of  $F_{MSY}$  under



**Fig. 5.** North Pacific striped marlin (*Tetrapturus audax*) time trajectories of (a) scenario-averaged relative spawning biomass ( $S^*$ )  $\pm 1$  standard error, (b) probabilities that relative spawning biomass was less than threshold values of one (solid circles), one-half (open circles), and one-tenth (solid triangles), (c) scenario-averaged relative fishing mortality ( $F^*$ )  $\pm 1$  standard error, and (d) probabilities that relative fishing mortality was greater than threshold values of three-quarters (solid circles), one (open circles), and two (solid triangles).



scenario 2. Model-averaged estimates of relative biomass during 2001–2003 indicated that North Pacific striped marlin biomass was well below  $S_{MSY}$  and ranged from 29% of  $S_{MSY}$  under scenario 1 to 44% of  $S_{MSY}$  under scenario 2. If these  $MSY$ -based reference points were interpreted as thresholds and not as targets (e.g., Mace 2001), then the stock would also be considered depleted in the most recent period. Overall, the current status of North Pacific striped marlin was robust to model assumptions about the form of stock–recruitment curve used to fit the age-structured production model and was also robust to the steepness scenario assumed in the stock assessment model.

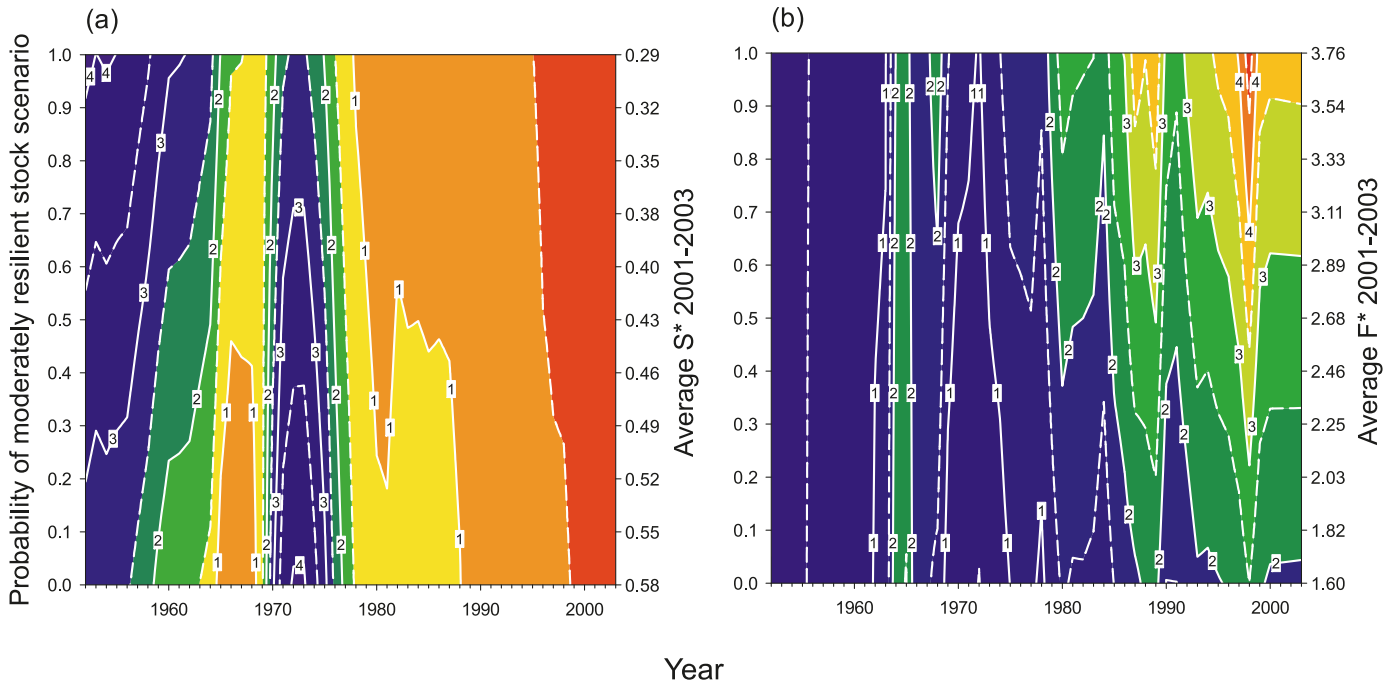
### Probable stock status

Scenario-averaged estimates of relative spawning biomass showed that North Pacific striped marlin abundance had declined substantially from relatively high abundances during the 1950s and 1970s (Fig. 5a). Current average spawning biomass was estimated to be 36% of  $S_{MSY}$ , with a CV of 55%. Thus, although recent abundance was low, there was considerable variability in the scenario-averaged estimate based on differences in the underlying steepness scenarios. Comparing the scenario-averaged, relative spawning biomasses to puta-

tive threshold values (Fig. 5b) indicated a very high probability that current spawner abundance was less than  $S_{MSY}$  (>99%) and a high probability that current abundance was less than  $1/2 S_{MSY}$  (75%). Nonetheless, the probability of severe depletion to less than one-tenth of  $S_{MSY}$  was relatively constant since the 1980s and remained around 10% (Fig. 5b). The scenario-averaged results indicated that spawning biomass of North Pacific striped marlin has been substantially reduced since the 1970s and that it was very likely that current spawner abundance was below  $S_{MSY}$ . However, it is unknown whether the rapid increase in relative spawning biomass during the 1970s was due to overcompensatory recruitment dynamics or positive environmental conditions for recruitment and survival, including changes in the target species of longline fleets (e.g., Ward and Hindmarsh 2007), or some combination of these factors.

The scenario averages of relative fishing mortality showed that fishing mortality on North Pacific striped marlin has increased steadily since the early 1970s (Fig. 5c). The current average fishing mortality was estimated to be 280% of  $F_{MSY}$ , with a CV of 18%. In contrast to recent average spawning biomass, it was relatively certain that current fishing mortality was much higher than desirable de-

**Fig. 6.** Contour plots of scenario-averaged estimates of (a) relative spawning biomass ( $S^* = S/S_{MSY}$ ) and (b) relative fishing mortality ( $F^* = F/F_{MSY}$ ) of North Pacific striped marlin (*Tetrapturus audax*) as a function of alternative probabilities of model scenarios 1 and 2 during 1952–2000 (left axis) along with the 3-year average of  $S^*$  and  $F^*$  during 2001–2003 (right axis). Each value of the probability of the moderately resilient stock scenario being true (y-axis values) represents a single scenario-averaged time series of values of  $S^*$  (a) or  $F^*$  (b), in which the probability of scenario 2 being true is the complement of the probability of scenario 1 being true, the moderately resilient stock scenario (left axis). For annual relative spawning biomass (a), values of  $S^* > 1$  (blue–green colors) indicate positive spawning biomass status, whereas values of  $S^* < 1$  (red–orange colors) indicate negative spawning biomass status. For annual relative fishing mortality (b), values of  $F^* > 1$  (red–orange colors) indicate negative fishing mortality status, whereas values of  $F^* < 1$  (blue–green colors) indicate positive fishing mortality status.



spite differences in the underlying steepness scenarios. In particular, the scenario-averaged estimate of current fishing mortality had a 100% chance of exceeding a potential target harvest rate of 75% of  $F_{MSY}$  (Fig. 5d) and had high probabilities of exceeding  $F_{MSY}$  (>99%) and  $2F_{MSY}$  (94%). Overall, the scenario-averaged results for fishing mortality indicated that North Pacific striped marlin had very likely experienced fishing mortalities that exceeded  $F_{MSY}$  since the 1980s.

#### Sensitivity of probable stock status to scenario probabilities

Sensitivity analysis of the impact of different combinations of scenario probabilities for scenarios 1 and 2 showed that probable status was robust to scenario uncertainty (Figs. 6a, 6b). For relative biomass, the general pattern of high biomass in the 1950s and lower biomass in the 1960s followed by an increase in the 1970s and a substantial decline in the 1980s through 2000s was apparent across scenario probabilities (Fig. 6a). If greater emphasis was placed on the moderately resilient stock scenario 1, the impact was that there was a lower estimate of biomass increase in the 1970s and a steeper decline in the 1980s–2000s. Scenario-averaged estimates of recent average biomass ranged two-fold across the range of scenario probabilities (Fig. 6a), from 29% to 58% of  $S_{MSY}$  depending on whether scenario 1 received a high degree or low degree of belief. Nonetheless, the conclusion that spawning biomass has been less than

$S_{MSY}$  since the early 1990s was robust to the degree of belief in either scenario.

The time series of probable biomass status exhibited a consistent pattern across scenario probabilities (Fig. 6b). Fishing mortality was below  $F_{MSY}$  until the mid-1960s when  $F^*$  exceeded unity and then declined in the late 1960s. In the early 1970s, fishing mortality again exceeded  $F_{MSY}$  and fluctuated about an increasing trend since then (Fig. 6b). Higher belief in the fluctuating about a constant mean recruitment scenario generally reduced the magnitude and fluctuations in relative fishing mortality. Nonetheless, the conclusion that fishing mortality has exceeded  $F_{MSY}$  since the 1980s was robust to the pattern of relative belief in scenarios 1 and 2. Furthermore, current fishing mortality ranged from 192% to 376% of  $F_{MSY}$  (Fig. 6b), which indicated that current probable status with respect to fishing mortality was robust to the degree of belief in the assessment scenarios. As a result, one would conclude that the North Pacific striped marlin stock was currently experiencing overfishing.

#### Discussion

Model averaging provides a general method to handle structural uncertainty when alternative models provide contrasting fits to a single data set. In our post hoc analysis of the two North Pacific striped marlin stock assessment scenarios, we analyzed one source of structural uncertainty, the

degree of compensation in the stock–recruitment relationship, and one source of model fitting uncertainty, the degree of serial correlation in the stock–recruitment residuals. In this case, the structural uncertainty was about the magnitude of striped marlin recruitment at high spawning stock sizes that would be predicted to fluctuate around a mean under the Beverton–Holt assumption and would be predicted to fluctuate around a declining trend under the Ricker assumption. The choice of stock–recruitment model form can have an important influence on the values of biological reference points (Myers et al. 1994), although it was not strongly influential for North Pacific striped marlin. In future assessments, it will be useful to investigate methods to incorporate uncertainty about the value of steepness using an informative prior distribution based on the reproductive ecology of the species (Mangel et al. 2009) and then to evaluate the structural uncertainty in the stock–recruitment relationship through model selection or model averaging techniques.

Although model uncertainty typically deals with structural differences between alternative models, it can also be used to handle alternative assessments when there are differences of opinion about the relative credibility of the data and model assumptions used in the alternative assessment scenarios. This situation can commonly occur when government and industry groups conduct competing stock assessments to provide management advice to inform decisions. In general, contested stock assessments can have a positive impact on the assessment and review process by fostering thorough evaluation of data and models, investigation of alternative models, and complete checking of computations (Starr et al. 1998). This was our experience with the west coast sablefish stock assessment conducted in 1998 in which two age-structured assessment models were evaluated, leading to an improved treatment of data and model assumptions. After the sablefish stock assessment review, the two competing model results were averaged, based on the opinions of the Pacific Fishery Management Council's Groundfish Management Team, to provide management advice on suitable total allowable catch levels. The contested assessment review process can also lead to a single modeling approach being selected as most credible above others (Butterworth and Rademeyer 2008). This outcome suggests that absolute certainty exists that one modeling approach is superior, which is typically not well supported unless simulation modeling is used to test the alternative approaches with known results. Instead, a more pragmatic approach would be to assign probabilities to alternative model scenarios using a Delphi approach, which elicits expert opinions in an anonymous manner, or other structured protocol to quantify beliefs in alternative scenarios (Morgan and Henrion 1990). In general, assigning probabilities to alternative assessment scenarios or models will most likely have to be based on subjective criteria because the competing models may use different data sources, model structures, and estimation approaches (Hill et al. 2007).

The practical implications of our post hoc analyses of biological reference points and probable status of the North Pacific striped marlin stock are straightforward. First, there is substantial evidence that the stock is depleted or is rapidly approaching a depleted condition. Second, it is highly likely

that the stock has been experiencing excessive fishing mortality ( $F > F_{MSY}$ ) since the 1980s. These conclusions about probable stock status suggest that it may be useful to consider how fishing mortality could be reduced by decreasing catch or fishing effort in fisheries that harvest striped marlin. Reduction in fishing mortality from recreational fisheries might be accomplished by changing patterns of catch and release, although the effectiveness of such policies critically depends on fish survival rates after being hooked and released (Domeier et al. 2003). In the case of North Pacific striped marlin, however, the recreational catches are not substantial and it is unlikely that significant reduction in total fishing mortality could be achieved through regulation of the recreational fisheries alone.

Commercial pelagic longline and driftnet fisheries take the vast majority of the North Pacific striped marlin catch. It is assumed that the commercial catch is the result of incidental takes of marlin during fishing operations targeting other species. Reducing fishing mortality from commercial fisheries might be accomplished by decreasing total fishing effort, changing commercial longline hook design, or changing the average depth of longline deployment (Boggs 1992). Time and area closures are another potential tool for reducing the bycatch of marlin in hotspot areas or seasons. However, such measures may have negative impacts on fishery yields of other species, primarily tunas, and may be politically difficult to implement for this reason.

Reducing North Pacific striped marlin mortality to more sustainable levels will require effective conservation measures that may reduce catches of other highly prized species. In the North Pacific Ocean, albacore and Pacific bluefin tuna are the primary targets of fisheries that indirectly take marlins. The tunas are faster growers (Shimose et al. 2009) and mature at earlier ages (Chen et al. 2006) and therefore appear to be more resilient to fishing pressure than striped marlin. Economics may favor the continuing overfishing of vulnerable stocks to maximize yields in more resilient stocks taken in mixed assemblages (Parker et al. 2000). Therefore, additional economic incentives may be necessary to offset the loss of yield associated with the protection of the vulnerable members of the assemblage.

The Western and Central Pacific Fisheries Commission (WCPFC) has tasked a committee made up of biologists, economists, and fishery-gear experts to develop measures to reduce fishing mortality of striped marlin. Economic and ecosystem effects from reduction in fishing mortality could be evaluated through additional fishery economics analysis and management strategy evaluations (Sainsbury et al. 2000), but that is beyond the scope of this paper. However, without formally adopted biological reference points or a target range of pretty good yields (i.e., Hilborn 2010), it remains unclear to what levels the current fishing mortality should be reduced by the WCPFC. In this paper, we provide results relative to  $MSY$ , which is an often used to characterize stock status of billfishes in other oceans (Restrepo et al. 2003). Overall, the probable stock status results appear to be robust even though they account for model selection uncertainty in reference point estimation and assessment scenario uncertainty in stock–recruit dynamics.

Model selection uncertainty has not commonly been considered in making inferences, although it is often an impor-

tant component of statistical uncertainty (Buckland et al. 1997). Our use of model averaging to estimate biological reference points for North Pacific striped marlin shows that the technique has important fisheries applications. In particular, model averaging allowed us to investigate plausible alternative models of striped marlin recruitment dynamics and incorporate these competing hypotheses directly into the estimation process. In turn, this allowed us to avoid the issue of defending a single recruitment dynamics hypothesis as being judged the best, a priori. Instead, the stock–recruitment data drove the process of model selection and weighting of results. In general, this provided a more objective basis for deriving biological reference points than the typical approach of assuming that a single model was most appropriate.

The additional complexity needed to include model uncertainty in providing management advice may be difficult, in some cases, to justify to stakeholders and fisheries managers. This concern can be addressed by emphasizing the posterior model probabilities and the model-averaged estimates and their precision in the construction of decision tables (Hilborn and Peterman 1996). Focusing on the relative likelihood of the alternatives can clarify how the competing hypotheses were supported by the data. For example, in this study, the evidence that striped marlin exhibit overcompensatory stock–recruitment dynamics was strong but not conclusive. Although this may seem counterintuitive given the longevity, size, and trophic niche of striped marlin in the pelagic ecosystem, it is more convincing to let the data inform us given the limited specific knowledge of the importance of density dependence for apex predators. Similarly, a priori one might have a higher degree of belief in the possibility that recruitment dynamics were environmentally driven, which is a common but ecologically optimistic assumption made in some assessments of tuna and tuna-like species (see, for example, Mangel et al. 2009). Although uncorrelated environmental noise was more probable in this study, the alternative hypothesis that low-frequency environmental forcing was important could also not be eliminated with certainty. Overall, the conclusion that each of the alternative age-structured production models had some credibility under the alternative assessment scenarios reflects the state of knowledge regarding processes influencing recruitment of striped marlin.

The evaluation of alternative scenarios is relatively common in stock assessment, although management advice is often predicated on one scenario being accepted as the best estimate of the true state of nature (see, for example, Butterworth and Rademeyer 2008). In contrast, our use of scenario averaging provides a way to combine the results of alternative assessment models to assess probable status. To do this, it is important to have common indicators of stock status that can be applied across scenarios. Two common measures are relative spawning biomass and relative fishing mortality, in particular, spawning biomass as a fraction of the target biomass and fishing mortality as a fraction of the limit in fishing mortality. The use of relative biomass and fishing mortality enables one to average assessment results across scenarios, which may be based on different data sets, models, or assumptions and, for this reason, are not directly comparable using model selection techniques (e.g., see Burnham and Anderson 2002). The development of esti-

mates of variance for scenario-averaged results provides a direct means to conduct risk analysis across scenarios, such as the evaluation of the probability that overfishing would occur given a set of management options for the total allowable catch of striped marlin.

Accounting for scenario uncertainty through scenario averaging may also provide a direct means to address the dilemma of being unable to take management action based on the uncertainty resulting from alternative assessments. In particular, scenario averaging requires that one be explicit about beliefs about the relative probability of alternative scenarios. This, in turn, leads to a more explicit accounting of risk preferences and the rationale for supporting one scenario over another. In the case of North Pacific striped marlin, the WG expressed indifference between the two competing scenarios. As a result, it implicitly chose to weight the two assessment scenarios equally. The point of scenario averaging is that the logical implications of equal plausibility or other beliefs can be carried forward to evaluate the probable status of fishery resources when competing assessment scenarios are available. In this context, the emphasis is ultimately on achieving robust predictive accuracy, which, in effect, requires high quality data and a thorough evaluation of alternative hypotheses. Overall, scenario-averaged estimates of relative stock status provide a direct measure of the probable status of the North Pacific striped marlin stock that accounts for both model and assessment scenario uncertainty. Using model averaging to estimate probable stock status from multiple assessment scenarios is analogous to using ensemble averages from multiple predictive models to make weather forecasts.

## Acknowledgments

We thank Chris Legault, Keith Bigelow, Gerard DiNardo, and two anonymous reviewers for their constructive comments on the draft manuscript. We also thank the members of the ISC Billfish Working Group for their help in preparing this document.

## References

- Boggs, C. 1992. Depth, capture time, and hooked longevity of longline-caught pelagic fish: timing bites of fish with chips. *Fish. Bull.* (Washington, D.C.), **90**: 642–658.
- Brodziak, J., and Legault, C. 2005. Model averaging to estimate rebuilding targets for overfished stocks. *Can. J. Fish. Aquat. Sci.* **62**(3): 544–562. doi:10.1139/f04-199.
- Brodziak, J., and Link, J. 2002. Ecosystem-based fishery management: what is it and how can we do it? *Bull. Mar. Sci.* **70**(2): 589–611.
- Buckland, S., Burnham, K., and Augustin, N. 1997. Model selection: an integral part of inference. *Biometrics*, **53**(2): 603–618. doi:10.2307/2533961.
- Burnham, K., and Anderson, D. 2002. Model selection and multi-model inference. 2nd ed. Springer-Verlag, New York.
- Butterworth, D.S., and Rademeyer, R.A. 2008. Statistical catch-at-age analysis vs. ADAPT-VPA: the case of Gulf of Maine cod. *ICES J. Mar. Sci.* **65**(9): 1717–1732. doi:10.1093/icesjms/fsn178.
- Chen, K.-S., Crone, P., and Hsu, C.-C. 2006. Reproductive biology of female Pacific bluefin tuna *Thunnus orientalis* from southwestern North Pacific Ocean. *Fish. Sci.* **72**(5): 985–994. doi:10.1111/j.1444-2906.2006.01247.x.



- Claeskens, G., and Hjort, N.L. 2008. Model selection and model averaging. Cambridge University Press, New York.
- Domeier, M., Dewar, H., and Nasby-Lucas, N. 2003. Mortality rate of striped marlin (*Tetrapturus audax*) caught with recreational tackle. *Mar. Freshw. Res.* **54**(4): 435–445. doi:10.1071/MF01270.
- Fogarty, M., Mayo, R., O'Brien, L., Serchuk, F., and Rosenberg, A. 1996. Assessing uncertainty in exploited marine populations. *Reliab. Eng. Syst. Saf.* **54**(2-3): 183–195. doi:10.1016/S0951-8320(96)00074-9.
- Francis, R.I.C.C., and Shotton, R. 1997. "Risk" in fisheries management: a review. *Can. J. Fish. Aquat. Sci.* **54**(8): 1699–1715. doi:10.1139/cjfas-54-8-1699.
- Gelman, A., Carlin, J., Stern, H., and Rubin, D. 1995. Bayesian data analysis. Chapman and Hall, New York.
- Gilks, W., Richardson, S., and Spiegelhalter, D. 1996. Markov chain Monte Carlo in practice. Chapman and Hall, New York.
- Hilborn, R. 1997. Uncertainty, risk, and the precautionary principle. In *Global trends: fisheries management*. Edited by E. Pikitch, D. Hupert, and M. Sissenwine. American Fisheries Society Symposium No. 20, Bethesda, Maryland. pp. 100–106.
- Hilborn, R. 2010. Pretty good yield and exploited fishes. *Mar. Policy*, **34**(1): 193–196. doi:10.1016/j.marpol.2009.04.013.
- Hilborn, R., and Mangel, M. 1997. The ecological detective: confronting models with data. Princeton University Press, Princeton, New Jersey.
- Hilborn, R., and Peterman, R. 1996. The development of scientific advice with incomplete information in the context of the precautionary approach. *FAO Fish. Tech. Pap.* **350**(2): 77–102.
- Hilborn, R., and Walters, C. 1992. Quantitative fisheries stock assessment. Chapman and Hall, New York.
- Hill, S., Watters, G., Punt, A., McAllister, M., Le Quére, C., and Turner, J. 2007. Model uncertainty in the ecosystem approach to fisheries. *Fish Fish.* **8**: 315–336.
- Hoeting, J., Madigan, D., Raftery, A., and Volinsky, C. 1999. Bayesian model averaging: a tutorial. *Stat. Sci.* **14**: 382–401.
- International Scientific Committee for Tuna and Tuna-like species in the North Pacific Ocean. 2007. Report of the Marlin and Swordfish Working Group Joint Workshop. Intersessional Workshop, 19–26 March 2007, Institute of Oceanography, National Taiwan University, Taipei, Taiwan.
- Mace, P. 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. Smith, J. Hunt, and D. Rivard. *Can. Spec. Publ. Fish. Aquat. Sci.* No. 120.
- Mangel, M., Brodziak, J., and DiNardo, G. 2009. Reproductive ecology and scientific inference of steepness: a fundamental metric of population dynamics and strategic fisheries management. *Fish Fish.* doi:10.1111/j.1467-2979.2009.00345.x.
- McAllister, M.K., Starr, P.J., Restrepo, V.R., and Kirkwood, G.P. 1999. Formulating quantitative methods to evaluate fishery-management systems: what fishery processes should be modelled and what trade-offs should be made? *ICES J. Mar. Sci.* **56**(6): 900–916. doi:10.1006/jmsc.1999.0547.
- Methot, R.D. 2000. Technical description of the stock synthesis assessment program. NOAA Tech. Memo. No. NMFS-NWFSC-43.
- Michielsens, C., and McAllister, M. 2004. A Bayesian hierarchical analysis of stock–recruit data: quantifying structural and parameter uncertainties. *Can. J. Fish. Aquat. Sci.* **61**(6): 1032–1047. doi:10.1139/f04-048.
- Morgan, M.G., and Henrion, M. 1990. Uncertainty: a guide to dealing with uncertainty in qualitative risk assessment and policy analysis. Cambridge University Press, Cambridge, UK.
- Myers, R.A., Rosenberg, A.A., Mace, P.M., Barrowman, N., and Restrepo, V.R. 1994. In search of thresholds for recruitment overfishing. *ICES J. Mar. Sci.* **51**(2): 191–205. doi:10.1006/jmsc.1994.1020.
- Parker, S.J., Berkeley, S.A., Golden, J.T., Gunderson, J.T., Heifetz, J., Hixon, M.A., Larson, R., Leaman, B.M., Love, M.S., Musick, J.A., O'Connell, V.M., Ralston, S., Weeks, H.J., and Yoklavich, M.M. 2000. Management of Pacific rockfish. *Fisheries*, **25**(3): 22–30. doi:10.1577/1548-8446(2000)025<0022:MOPR>2.0.CO;2.
- Patterson, K., Cook, R., Darby, C., Gavaris, S., Kell, L., Lewy, P., Mesnil, B., Punt, A., Restrepo, V., Skagen, D., and Stefánsson, G. 2001. Estimating uncertainty in fish stock assessment and forecasting. *Fish Fish.* **2**: 125–157.
- Piner, K., Conser, R., DiNardo, G., and Brodziak, J. 2007. Stock synthesis 2 sensitivity runs for striped marlin assessment WG 2007. International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific/MARWG-SWORWG 1/Working Paper 2.
- Restrepo, V., Prince, E.D., Scott, G.P., and Uozumi, Y. 2003. ICCAT stock assessments of Atlantic billfish. *Mar. Freshw. Res.* **54**(4): 361–367. doi:10.1071/MF02057.
- Sainsbury, K.J., Punt, A.E., and Smith, A.D.M. 2000. Design of operational management strategies for achieving fishery ecosystem objectives. *ICES J. Mar. Sci.* **57**(3): 731–741. doi:10.1006/jmsc.2000.0737.
- Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Stat.* **6**(2): 461–464. doi:10.1214/aos/1176344136.
- Shimose, T., Tanabe, T., Chen, K.-S., and Hsu, C.-C. 2009. Age determination and growth of Pacific bluefin tuna, *Thunnus orientalis*, off Japan and Taiwan. *Fish. Res.* **100**(2): 134–139. doi:10.1016/j.fishres.2009.06.016.
- Sissenwine, M.P., and Shepherd, J.G. 1987. An alternative perspective on recruitment overfishing and biological reference points. *Can. J. Fish. Aquat. Sci.* **44**(4): 913–918. doi:10.1139/f87-110.
- Starr, P., Annala, J., and Hilborn, R. 1998. Contested stock assessment: two case studies. *Can. J. Fish. Aquat. Sci.* **55**(2): 529–537. doi:10.1139/cjfas-55-2-529.
- Ward, P., and Hindmarsh, S. 2007. An overview of historical changes in the fishing gear and practices of pelagic longliners, with particular reference to Japan's Pacific fleet. *Rev. Fish Biol. Fish.* **17**(4): 501–516. doi:10.1007/s11160-007-9051-0.