

When can we reliably estimate the productivity of fish stocks?

Paul B. Conn, Erik H. Williams, and Kyle W. Shertzer

Abstract: In modern fishery stock assessments, the productivity of exploited stocks is frequently summarized by a scale-invariant “steepness” parameter. This parameter, which describes the slope of the spawner–recruit curve, determines resilience of a stock to exploitation and is highly influential when estimating maximum sustainable yield. In this study, we examined conditions under which steepness can be estimated reliably. We applied a statistical catch-age model to data that were simulated over a broad range of stock characteristics and exploitation patterns and found that steepness is often estimated at its upper bound regardless of underlying productivity. The ability to estimate steepness reliably was most dependent on the true value of steepness, the exploitation history of the stock, natural mortality, duration of the time series, and quality of an index of abundance; this ability was relatively unaffected by levels of stochasticity in recruitment and sampling intensity of age compositions. We further explored the method of inverse prediction to improve estimates of steepness and conclude that this approach holds promise. We illustrate the utility of simulation and inverse prediction methods with two fish stocks located off the southeastern United States, greater amberjack (*Seriola dumerili*) and gag grouper (*Mycteroperca microlepis*).

Résumé : Dans les évaluations modernes des stocks, la productivité du stock exploité est souvent résumée par un paramètre d’« inclinaison » qui est invariable à toutes les échelles. Ce paramètre, qui décrit la pente de la courbe reproducteurs–recrues, détermine la résilience d’un stock à l’exploitation et est très important dans la détermination du rendement maximum durable. Notre étude examine les conditions dans lesquelles l’inclinaison peut être estimée de façon fiable. Nous avons utilisé un modèle statistique des captures en fonction de l’âge avec des données simulées présentant un vaste éventail de caractéristiques des stocks et de patrons d’exploitation et avons découvert que l’inclinaison est souvent estimée à sa limite supérieure, quelle que soit la productivité sous-jacente. La possibilité d’estimer l’inclinaison de manière fiable dépend surtout de la valeur réelle de l’inclinaison, de l’histoire d’exploitation du stock, de la mortalité naturelle, de la durée de la série chronologique et de la qualité de l’indice d’abondance; cette possibilité est relativement peu affectée par les niveaux de stochasticité dans le recrutement et l’intensité de l’échantillonnage de la composition en âge. Nous avons éprouvé en plus la méthode de prédiction inverse pour améliorer les estimations de l’inclinaison et concluons que la méthode est prometteuse. Nous illustrons l’utilité de la simulation et de la prédiction inverse à l’aide de deux stocks de poissons du large du sud-est des États-Unis, ceux de la grande sériole (*Seriola dumerili*) et de la badèche baillou (*Mycteroperca microlepis*).

[Traduit par la Rédaction]

Introduction

Estimating the parameters of a spawner–recruit function is one of the most vexing challenges in fisheries stock assessment. Ultimately, these parameters control the productivity and resilience of modeled fish stocks and thus are fundamental for predicting a stock’s response to exploitation and for calculating management benchmarks. In practice, several factors act in unison to limit our knowledge of spawner–recruit relationships. This list includes process and observation errors, errors-in-variables, nonstationarity, time series bias, and — perhaps the most serious — lack of contrast in spawning biomass (Hilborn and Walters 1992; Quinn and Deriso 1999).

When virgin abundance is regulated by food or habitat (as opposed to cannibalism, for instance), the Beverton–Holt functional form satisfies desirable properties of spawner–recruit relationships (Ricker 1975). Assessments increasingly use a version of the Beverton–Holt model that is parameterized in terms of a “steepness” parameter h (Mace and Doonan 1988) that quantifies the degree of density dependence in recruitment. In this parameterization, steepness is bounded, $0.2 \leq h \leq 1.0$. Researchers often prefer this version because h is independent of scale (Shertzer et al. 2008) and thus affords comparison among stocks through, for example, meta-analysis, whereby data from similar stocks can be shared hierarchically (Liermann and Hilborn 1997; Myers et al. 1999; Michielsens and McAllister 2004).

Received 1 May 2009. Accepted 30 October 2009. Published on the NRC Research Press Web site at cjfas.nrc.ca on 15 February 2010. J21179

Paper handled by Associate Editor Ray Hilborn.

P.B. Conn,¹ E.H. Williams, and K.W. Shertzer. National Marine Fisheries Service, Southeast Fisheries Science Center, NOAA Center for Coastal Fisheries and Habitat Research, 101 Pivers Island Road, Beaufort, NC 28516, USA.

¹Corresponding author (e-mail: Paul.Conn@noaa.gov).

In this article, we address the question, “Under what conditions can we estimate steepness reliably?” We ask this both in a general sense and for specific populations. For the general case, our strategy is first to simulate data under a wide variety of life history strategies and management regimes and then to estimate h using statistical catch-age analysis (Fournier and Archibald 1982; Deriso et al. 1985; Methot 1990). We then relate variation in estimated steepness to simulation input variables. This approach is similar in spirit to previous simulation studies (e.g., Ianelli 2002; Yin and Sampson 2004; Magnusson and Hilborn 2007) but focuses explicitly on steepness.

Previous studies reached mixed conclusions about the impact of variables likely to affect h or management benchmarks derived from h . For instance, when addressing steepness in particular, Magnusson and Hilborn (2007) emphasized the importance of having age composition or relative abundance index data at low stock sizes. More recently, Haltuch et al. (2008) emphasized the length of catch time series and recruitment variability and found that variables related to sampling intensity (e.g., number of age compositions, precision of index values) were less important. Both of those studies had a substantial number of simulations in which h was estimated at its upper bound of 1.0. Such an estimate implies that recruitment is unrelated to spawning biomass, providing a highly optimistic picture of the resilience of the stock. Furthermore, high steepness shortens estimated rebuilding schedules of depleted stocks, because expected recruitment remains high even at the lowest stock sizes.

In addition to a large-scale simulation study, we also consider two case studies — greater amberjack (*Seriola dumerili*; GAJ) and gag grouper (*Mycteroperca microlepis*; GAG) off the southeastern United States. In recent stock assessments of these species, the point estimate of steepness was at its upper bound for gag (SouthEast Data, Assessment, and Review (SEDAR) 2006), but not for greater amberjack (SEDAR 2007). Both assessments used a bootstrapping procedure to approximate uncertainty around point estimates, and for both stocks, a large number of bootstrap iterations resulted in steepness estimated at its upper bound. This situation is clearly undesirable if the high steepness values reflect poor estimation rather than actual stock productivity. We try to assess the reliability of estimation using stock-specific simulations, and in particular, we explore the potential of inverse prediction (cf., Osborne 1991) to refine point estimates of steepness by calibrating them to known true values used in simulations.

This paper is structured as follows. First, we review relevant spawner–recruit theory. Second, we provide a brief summary of inverse prediction methods and describe how they can be useful in fisheries applications. We then describe the methods that we used for large-scale simulations, as well as methods for analyzing simulation results. Descriptions of the two case studies follow, including details of how inverse prediction methods were used to refine steepness estimates. After presenting results, we discuss implications of our work for stock assessment and benchmark determinations.

Materials and methods

Steepness

The Beverton–Holt spawner–recruit function (Beverton and Holt 1957) has a rich history in fish ecology for predicting the number of recruits to a population given a level of spawning biomass. In modern stock assessment, a common parameterization of this model is

$$R = \frac{0.8R_0hS}{0.2\Phi_0R_0(1-h) + (h-0.2)S}$$

where R is the number of new recruits, S is spawning biomass, h is steepness, R_0 is mean “virgin” recruitment (i.e., that expected prior to exploitation), and Φ_0 is unfished spawning biomass per recruit (Mace and Doonan 1988). The formal definition of h under this parameterization is “the proportion of R_0 produced by 20% of S_0 ,” where S_0 is the theoretical spawning biomass from an unfished stock (Shertzer et al. 2008). If natural mortality, growth, and maturity schedules are known, Φ_0 can be computed as

$$\Phi_0 = \sum_{a=1}^A W_a \mu_a \exp[-(a-1)M]$$

where A is the maximum age, W_a is weight at age a , μ_a is the proportion mature at age a , and M is the instantaneous natural mortality rate.

Given a time series of spawner–recruit data, the parameters R_0 and h may then be estimated using a method such as maximum likelihood. A lognormal error structure is often used to describe recruitment deviations (Chen 2004), and in modern stock assessment models, the time series of S and R are often estimated at the same time as the spawner–recruit parameters. If $h = 0.2$ (its lower bound), recruitment is directly proportional to spawning biomass. If $h = 1.0$ (its upper bound), recruitment is independent of spawning biomass. In the latter case, overfishing has no effect on recruitment, although it may reduce yield if fish are routinely being harvested before reaching an “optimal” size.

Inverse prediction

Fishery biologists increasingly use statistical catch-age (SCA) analysis to perform stock assessments. Model fitting requires optimizing a compound objective function that often includes penalties and weights to control various aspects of model fit. These penalties and weights make asymptotic variance estimates unreliable. Variance estimates also break down when bounded parameters lie on a boundary (e.g., 1.0 for steepness). We suggest that inverse prediction may be a viable approach in such cases.

Inverse prediction has been used following controlled calibration experiments (see, e.g., Brown 1982; Osborne 1991; Benton et al. 2003). In typical applications, there is a variable of interest (say, X) that is difficult or expensive to measure, and another that is much easier to measure (call this variable Y). A calibration experiment is conducted to determine the relationship between X and Y , and this relationship is summarized in a model such as

$$Y = f(X, \theta) + \varepsilon$$

where $f(\cdot)$ is a function chosen by the investigator, θ is an unknown parameter vector, and ε is an error term. Once this relationship has been quantified, interest typically focuses on new observations Y_0 , where X_0 is unknown. That is, given a new response variable that is easy to measure (Y_0), what can one say about the unmeasured predictor variable?

One of the more flexible approaches to inverse prediction uses Bayesian inference to quantify uncertainty in X_0 given a new observation Y_0 . Following Racine-Poon (1988), inference about X_0 can be made with this approach using posterior simulation where

$$(1) \quad [X_0, \theta | Y_0, X, Y] \propto [Y_0 | X_0, \theta][X_0][X, Y | \theta][\theta]$$

Here A denotes the probability density function of a random variable A , $A|B$ denotes the conditional distribution of A given B , and X and Y are vectors of predictors and responses from the calibration experiment.

How is this framework useful for improving estimates of steepness from stock assessment? We propose that a simulation experiment be used to quantify the relationship between true and estimated steepness values, and that this relationship be used to infer a plausible distribution of steepness given a single value estimated from an SCA analysis on a given fish stock (cf., Fig. 1). This approach treats the SCA analysis as a “black box” in that a solution is reached but it is unnecessary for the resultant estimators to possess known theoretical properties (description of technical detail is deferred to Appendix A).

General simulation study

To investigate the reliability of steepness estimates produced by modern stock assessment models, we conducted a Monte Carlo simulation study. Our strategy was to consider variability in stock dynamics and data collection wherever we thought that such variability might influence estimation of steepness; in other cases, we fixed values across simulations or set them based on theory of life history invariants (e.g., Charnov 1993). Virtual stocks varied by natural mortality, productivity, exploitation pattern, and level of process error and were monitored for different lengths of time and with different levels of observation error (Table 1).

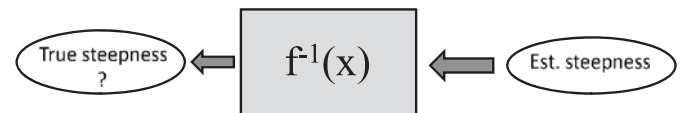
Four possible harvest regimes ($F_{\text{type}} \in 1, 2, 3, 4$) were considered, with fishing rates defined relative to F_{MSY} , the fishing rate that provides maximum sustainable yield (MSY). In regimes 1 and 2, mortality rates were constant over time at low ($F/F_{\text{MSY}} = 0.5$) or high ($F/F_{\text{MSY}} = 2.0$) levels, respectively. In regime 3, mortality rates increased linearly from $F/F_{\text{MSY}} = 0.5$ at the beginning of the study to $F/F_{\text{MSY}} = 2.0$ at the middle of the time series, before declining back to $F/F_{\text{MSY}} = 0.5$ at the end of the study. This scenario (“two-way trip”) mimics the situation in which exploitation increases until resource managers realize that overfishing is occurring and then act to decrease fishing pressure. Hilborn (1979) showed that rebuilding periods were important for estimating productivity with biomass-dynamic models, although their importance for modern statistical catch-age models is perhaps more equivocal (Magnusson and Hilborn 2007). Regime 4 consisted of a “one-way trip” whereby fishing mortality rates increased linearly from $F/F_{\text{MSY}} = 0.5$ at the beginning of the study to

Fig. 1. A visual depiction of inverse prediction as applied to steepness in stock assessment. The first step (a) consists of a calibration experiment in which simulated fisheries data are passed into a statistical catch-age analysis to estimate steepness. In the second step (b), it is assumed that the same statistical catch-age model from step a is used to estimate steepness for a given stock of interest. A functional relationship between true and estimated steepness from step a is then used to predict the true value of steepness given the SCA estimate. In practice, the second step can be accomplished by sampling eq. 1.

(a) Calibration experiment



(b) Inverse prediction



$F/F_{\text{MSY}} = 2.0$ at the end of the study. In all cases, initial stock status was set at equilibrium values associated with the fishing mortality rate in year one.

When simulating data, we assumed that stock dynamics were governed by a set of age-structured equations subject to lognormal process error in the Beverton–Holt recruitment relationship (Table 2). Selectivity and maturity schedules followed a logistic function, and harvesting and natural mortality processes were continuous with piecewise constant hazard rates (cf., Cox and Oakes 1984). The stock was exploited by a single fishery, which provided the assessment with three data sets for model fitting: total annual landings, samples of age compositions of landings, and an index of abundance derived from catch per unit effort. We assumed that observed landings and the index of abundance were subject to lognormal error and that observed age compositions were multinomially distributed (Table 2). All of these formulations are standard in models of fish dynamics (Quinn and Deriso 1999).

Estimation

We used two different estimation procedures on each data set. We started by fitting an SCA model to the data (Table 3), which was similar in structure to stock synthesis models (Methot 1989; 1990) and implemented in AD Model Builder software (Otter Research Ltd. 2004). Among other things, this procedure produced estimates of historical stock and recruit time series, as well as steepness. Spawner–recruit deviations were loosely constrained to follow a Beverton–Holt spawner–recruit model (Table 2). Because the constraint was loose, we found little difference between estimates of spawner and recruit time series using this approach and one with no defined spawner–recruit relationship. We denote estimates of steepness from AD Model Builder as \hat{h}^I to indicate that they were estimated internally in the stock

Table 1. Definitions of input variables with ranges used for general fish stock simulations. Each input variable was randomly sampled 10 000 times using a uniform distribution, with three replicate simulations performed at each such design point.

Parameter	Range	Definition
Y	(10, 45)	Years of data
M	(0.1, 0.4)	Instantaneous natural mortality rate
h	(0.21, 1.0)	Steepness
σ_R	(0.2, 0.8)	Standard deviation for multiplicative lognormal random errors applied to the spawner–recruit function
σ_I	(0.0, 0.5)	Standard deviation of abundance index
n	(50, 1000)	Number of fish aged annually
F_{type}	{1, 2, 3, 4}	An index describing which time series of fishing mortality rates were used
A	$1 + \sum_{a=1}^{39} I_{[\exp(-aM) > 0.02]}(a)$	Age when “plus” group begins
α	0.000001 (fixed)	Length–weight constant ($w = \alpha l^\beta$)
β	3.0 (fixed)	Length–weight exponent ($w = \alpha l^\beta$)
l_∞	1000 (fixed)	von Bertalanffy asymptotic size
t_0	−0.5 (fixed)	Age at size zero for von Bertalanffy growth model
K	$= M/1.65$	von Bertalanffy growth rate, parameterized using life history invariant of Charnov (1993)
a_{50}	$= \frac{1}{K} \log\left(\frac{3K+M}{M}\right)$	Age at 50% maturity and selectivity (Beverton 1992)
η	1.0 (fixed)	Slope parameter for logistic selectivity and maturity functions
R_0	100 000 (fixed)	Asymptotic recruitment when $F = 0$
q	0.000001 (fixed)	Catchability associated with abundance index

assessment model. Second, we used the predicted stock and recruit time series to estimate steepness externally via maximum likelihood. Following Dorn (2002), this approach involved maximizing a joint lognormal probability density function for recruitment with a bias correction:

$$\prod_y \frac{1}{R_y \sqrt{2\pi} \sigma_R} \exp \left[-\frac{1}{2\sigma_R^2} (\log(R_y) - \log(\tilde{R}_y) + \sigma_R^2/2)^2 \right]$$

where R_y denotes “observed” recruitment (i.e., output of the assessment model), and \tilde{R}_y denotes recruitment as predicted by the Beverton–Holt model. We denote estimates of steepness from this approach as \hat{h}^E to indicate that they were estimated externally. Estimates \hat{h}^I and \hat{h}^E were evaluated using median absolute relative error (MARE), a common metric for assessing estimator performance in fishery simulation studies (e.g., Prager 2002; Punt 2003; Wilberg and Bence 2006).

For simplicity, natural mortality was treated as a known constant in the assessment model and was set to its true value. Similarly, we assumed that the parameters describing growth and age–length conversions were known without error, and we supplied the assessment model with the correct standard deviations for the error in landings and the population index. Thus, the assessment model was provided a favorable scenario for estimating steepness.

Analysis of simulation performance data

Preliminary analysis suggested a bimodal distribution for estimated steepness (see, e.g., Fig. 2). In particular, steepness was estimated either at its upper bound of 1.0 or imprecisely but centered on or near the true value. We thus conceptualized \hat{h} as being drawn from a one-inflated and truncated normal distribution. Written as a finite mixture

distribution and taking into consideration that $\hat{h} \in (0.2, 1.0)$, the probability density function (pdf) for \hat{h} is then

$$(2) \quad f(\hat{h}|\mu_h, \sigma_h, \pi) = \pi + (1 - \pi) \frac{\mathcal{N}(\hat{h}|\mu_h, \sigma_h^2)}{\int_{0.2}^{1.0} \mathcal{N}(\hat{h}|\mu_h, \sigma_h^2) d\hat{h}}$$

Here, $\mathcal{N}(\mu_h, \sigma_h^2)$ denotes the normal pdf, and π gives the probability that h is estimated at its upper bound. Our interest focused on trying to identify important predictors of π (i.e., what factors influence steepness being estimated at its upper bound?) and σ_h (i.e., given that steepness is not estimated at 1.0, what factors influence precision of \hat{h} ?).

We initially found that estimating all parameters simultaneously was computationally prohibitive. Instead, we fit the model using a two-step process. First, π was estimated using a generalized additive model (GAM; Hastie and Tibshirani 1999) where the binary response variable was whether \hat{h} was estimated to be greater than 0.998. In this case, we used the function “gam” in the R programming language (R Development Core Team 2007), specifying cubic smoothing splines for all continuous variables. Second, conditional on $\hat{h} \leq 0.998$, we estimated σ_h using generalized linear models (GLMs; McCullagh and Nelder 1989) (a GAM was not used because the truncated normal distribution is not a standard error distribution). In this case, we assumed that steepness was being estimated without bias (that is, $\mu_h = h$), which indeed seemed to be the case when exploring results visually. We used the function “nlminb” for maximum likelihood inference on the constrained normal distribution and used a logit link function on σ to constrain its values to reasonable levels and to promote numerical stability.

Table 2. Formulation of population dynamics model and description of how data were generated.

Quantity	Symbol	Description or definition
Modeled quantities		
Weight at age	w_a	$w_a = al_a^b$
Maturity	m_a	$m_a = \frac{1}{1 + \exp[-\eta(a - a_{50})]}$
Fishery selectivity*	s_a	Data were generated with $s_a = m_a$, but logistic selectivity parameters were estimated within the SCA analysis
Fishery mortality rate*	F_y	Annual fishing mortality
Fishing mortality rate at age	$F_{a,y}$	$F_{a,y} = s_{a,y}F_y$, where F_y is the fully selected fishing mortality rate
Total mortality rate at age	$Z_{a,y}$	$Z_{a,y} = M + F_{a,y}$
Catchability coefficient*	q	Set to 0.00001 when generating data; estimated in the SCA analysis
Abundance at age*	$N_{a,y}$	$N_{1,1} = \frac{R_0(4h\phi_F - \phi_0(1-h))}{(5h-1)\phi_F} \exp(\varepsilon_0^R)$ $N_{a+1,1} = N_{a,1} \exp(-Z_{a,1}) \quad \forall a \in (1, \dots, A-1)$ $N_{A,1} = N_{A-1,1} \frac{\exp(-Z_{A-1,1})}{1 - \exp(-Z_{A,1})}$ $N_{0,y+1} = \frac{0.8R_0hS_y}{0.2\phi_0R_0(1-h) + (h-0.2)S_y} \exp(\varepsilon_y^R)$ $N_{a+1,y+1} = N_{a,y} \exp(-Z_{a,y}) \quad \forall a \in (1, \dots, A-1)$ $N_{A,y} = N_{A-1,y-1} \frac{\exp(-Z_{A-1,y-1})}{1 - \exp(-Z_{A,y-1})}$, where $\varepsilon_y^R \sim \text{Normal}(0, \sigma_R)$, ϕ_0 gives unfished spawning biomass per recruit, and ϕ_F gives spawning biomass per recruit at the initial F_y level
Mature biomass	S_y	$S_y = \sum_a N_{a,y}w_a m_a$, also referred to as spawning biomass
Population biomass	B_y	$B_y = \sum_a N_{a,y}w_a$
Landed catch at age	$C_{a,y}$	$C_{a,y} = \frac{F_{a,y}}{Z_{a,y}} N_{a,y} [1 - \exp(-Z_{a,y})]$
Landed age compositions	$p_{a,y}$	$p_{a,y} = \frac{C_{a,y}}{\sum_a C_{a,y}}$, proportional contribution of age class a in year y to the fishery
Data		
Observed landings	L_y	$L_y = \exp(\varepsilon_y^L) \sum_a C_{a,y}w_a$, where $\varepsilon_y^L \sim \text{Normal}(0, 0.05^2)$
Observed catch at age	$C'_{a,y}$	$C'_{a,y} \sim \text{Multinomial}(n, \mathbf{p}_y)$, where \mathbf{C}'_y and \mathbf{p}_y denote vectors of age-specific quantities in year y
Observed abundance index	U_y	$U_y = \varepsilon_y^U q \sum_a N_{a,y}s_a$, where $\varepsilon_y^U \sim \text{Normal}(0, \sigma_I^2)$

*These quantities estimated within SCA analyses.

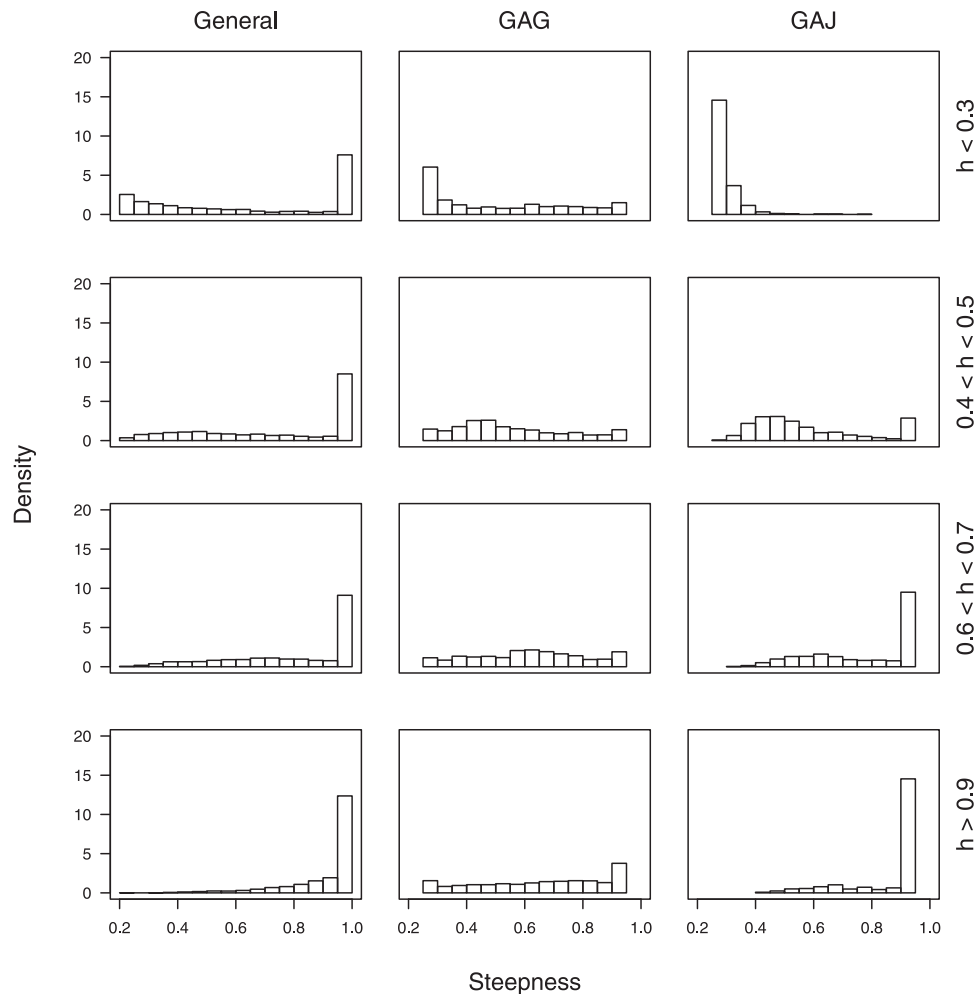
Table 3. Description of log-likelihood components and penalties used in the assessment model for estimation. Predicted quantities are indicated with a tilde and follow the same functional form as in Table 2 but use estimated parameters instead of true values. Unlike the generating model, separate parameters are estimated for selectivity and maturity. A small constant ($k = 0.00001$) is added in several places to prevent numerical errors.

Quantity	Symbol	Description or definition
Multinomial age compositions	Λ_1	$\Lambda_1 = n \sum_y \left\{ \sum_a (C'_{a,y} + k) \log[(p_{a,y} + k)] \right\}$
Lognormal landings	Λ_2	$\Lambda_2 = \sum_y \frac{\{\log[(L_y + k)/(\tilde{L}_y + k)]\}^2}{2(0.05)^2}$
Lognormal abundance index	Λ_3	$\Lambda_3 = \sum_y \frac{\{\log[(U_y + k)/(\tilde{U}_y + k)]\}^2}{2(\sigma_I^2)^2}$
Constraint on recruitment deviations	Λ_4	$\Lambda_4 = \sum_y R_y^2$, where R_y is the log of the recruitment deviation from the Beverton–Holt model in year y
Total likelihood	Λ	$\Lambda = \sum_{i=1}^4 \Lambda_i$, objective function minimized by the assessment model

The primary motivation of these simulations was to understand which stock characteristics were the most important for precise and accurate estimation of steepness. As omniscient simulators, we were able to use known parameter values and stock trajectory statistics as candidate predictor

variables for \hat{h} . To this end, we expressed eq. 2 as a function of simulation inputs and stock characteristics. Models were developed for $\text{logit}(\sigma_h)$ and $\text{logit}(\pi)$ that were written in terms of $h, Y, M, \sigma_R, \sigma_I, n, F_{\text{type}}$. In particular, we considered models ranging from an intercept-only model to one

Fig. 2. Histograms of estimated steepness, given different ranges of steepness used to generate data (ranges displayed on the right axis) for the general simulation study (left panels), gag grouper (*Mycteroperca microlepis* (GAG); center panels), and greater amberjack (*Seriola dumerili* (GAJ); right panels).



with an additive effect of all predictor variables and an interaction between σ_R and F_{type} (where present, F_{type} was included as a factor variable). We excluded more interactions because they make parameter estimates difficult to interpret. A stepwise AIC model selection procedure with backward selection (Venables and Ripley 2002) was used to select submodels for $\text{logit}(\sigma_h)$ and $\text{logit}(\pi)$ that were of reasonable structure. Once a final model had been selected, we plotted predicted values over the range of each input parameter to examine the influence of a particular input on estimates of steepness.

Case studies

We considered two case studies to investigate estimates of steepness in real applications, reliability of these estimates, and the influence of the estimates on management benchmarks. For illustration, we selected greater amberjack (GAJ) and gag grouper (GAG), both located off the southeastern United States and recently assessed using SCA models (SEDAR 2006; 2007). For GAJ, steepness was estimated at 0.74 in the base model run, but several sensitivity runs estimated steepness at its upper bound, which was set to 0.95 in both assessments. For GAG, steepness was estimated at the

upper bound in the base model run. In the original assessments, uncertainty around the benchmarks was quantified via a bootstrap procedure in which recruitment deviations were resampled a large number of times and parameters of the spawner–recruit curve were repeatedly re-estimated to generate distributions of MSY benchmarks. Many of the bootstrap iterations produced values of steepness at the upper bound, leading to bimodal distributions of benchmarks.

These two case studies highlight questions that arise in stock assessment whenever steepness is estimated. For instance, when steepness is estimated at the upper bound (as with GAG), does this provide some indication that the stock really is highly productive? Or, in such cases, are we better off adopting a proxy-based benchmark (see, e.g., Mace 1994; Williams and Shertzer 2003)? Can we be any more confident in our estimates when steepness is not estimated at the upper bound (as with GAJ)?

To answer these questions, we conducted simulations based on the exact number of years, sample sizes, gear types, and error structures used in each assessment. A complete description of model structure and data sources is described elsewhere (SEDAR 2006; 2007), but we provide a

few relevant details here. Both assessment models had structure similar to that provided in Table 2, but with multiple sectors (commercial and recreational), multiple gears (e.g., handline, spears), and discard mortalities. Sampling of length and age compositions of landings began for most fisheries in the late 1970s or 1980s. Fishery-dependent indices of abundance were also available beginning in the 1970s. The GAJ assessment model spanned 1946–2006, and the GAG model spanned 1962–2004.

For each stock, we simulated data with process and observation error. All simulation parameters were fixed to estimates from the assessments, with the exception of steepness, which we sampled from a Uniform(0.21, 1.0) distribution. Note that both stock assessment models imposed boundaries on steepness at 0.2 and 0.95 so that the simulated and estimated ranges differed slightly for the case studies. The same lognormal probability distributions estimated in the assessments were used to generate spawner–recruit deviations. That is, the sampled value of steepness was used to determine the underlying spawner–recruit curve, and random lognormal recruitment deviations were generated using the value of σ_R estimated in the original stock assessment. Each simulated data set was then analyzed with the assessment model, and this process was repeated 10 000 times.

As with the general simulation study, estimates of steepness (\hat{h}) in the case studies were assumed to be related to true steepness (h) via a normal mixture model. For both case studies, we assumed that the probability of steepness being estimated at its upper bound was a linear function of steepness on the logit scale:

$$\text{logit}(\pi_i) = \beta_1 + \beta_2 h_i^{\text{true}}$$

However, in contrast to the general simulation study, preliminary investigation of the GAJ data suggested that steepness estimates not occurring at the upper bound were negatively biased and that the magnitude of bias depended on the true steepness value. As such, we based inference on eq. 2, but considered the model

$$\text{logit}\left(\frac{\mu_h - 0.2}{0.75}\right) = \beta_3 + \beta_4 h_i^{\text{true}} + \beta_5 (h_i^{\text{true}})^2$$

This transformation constrains the predicted mean steepness value to fall within the (0.2, 0.95) range imposed by the SCA analyses in the stock assessments and allows bias in estimates of steepness to be a nonlinear function of the true steepness value. We also assumed a linear relationship between the standard deviation of the continuous part of the mixture distribution and true steepness:

$$\text{logit}(\sigma_i) = \beta_6 + \beta_7 h_i^{\text{true}}$$

We applied Markov chain Monte Carlo (MCMC) to sample each stock's posterior steepness distribution, using data from the simulation experiments for calibration (see Appendix A for details). This distribution was then used to calculate a posterior distribution of F_{MSY} . We compared these distributions with those obtained from 1000 bootstrap samples, where estimated spawner–recruit deviations were re-sampled and steepness was re-estimated (externally) at each

iteration. The latter approach was used to characterize uncertainty in the peer-reviewed stock assessments of GAJ and GAG. By conditioning on an F -weighted selectivity vector and on assumed growth, maturity, and natural mortality schedules, one could then solve for the corresponding distribution of F_{MSY} (SEDAR 2006).

Results

General simulation study

Overall, 56.5% of simulations had h^I estimated at its upper bound, and 36.9% had h^E estimated at its upper bound. MARE was also larger for internal estimates (0.227) than for external ones (0.193), suggesting that external estimates may be preferable (at least for the configuration used in this section). Subsequent simulation analyses thus use \hat{h}^E as the dependent variable.

Inspection of histograms of \hat{h}^E over different ranges of true steepness values suggested a bimodal distribution for estimated steepness (Fig. 2). The stepwise model selection procedure for π provided a high degree of support for all predictor variables (removing the least influential predictor resulted in an increase of 105 AIC points). However, predictive plots of the effect of each input variable clearly showed that some predictors were more important than others in determining whether steepness was estimated at its upper bound (Fig. 3). In particular, $\hat{\pi}$ increased substantially as the true value of steepness increased and as the coefficient of variation associated with the index of abundance increased. In contrast, the number of years of the study, natural mortality, age composition sample size, and recruitment variability appeared to be less important. The exploitation pattern (a proxy for biomass contrast) also appeared to be quite important, with two-way trips substantially decreasing (but not eliminating) the expected probability of having steepness estimated at its upper bound.

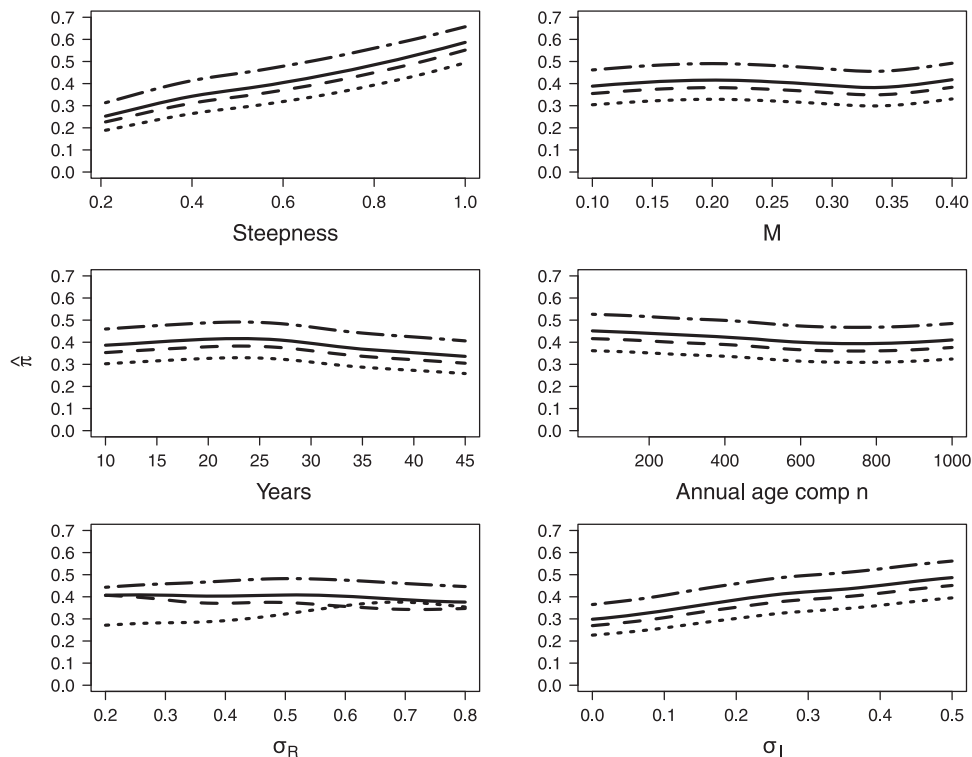
For the GLM analysis of factors affecting σ_h , the model with all predictor variables but n (the annual age composition sample size) had the lowest AIC score and was selected for effect comparisons. Here, the precision of \hat{h} tended to increase whenever true steepness, natural mortality, or the number of years of the study increased, or when the index of abundance had smaller standard error (Fig. 4). Precision was the best for two-way trips and for high but constant levels of fishing mortality. One-way trips resulted in slightly less precision, with constant but low levels of fishing mortality providing substantially less precise estimates. The effect of recruitment variability on σ_h differed depending on exploitation profile: higher recruitment variability resulted in more precise estimates when exploitation was constant but less precise estimates when fishing mortality rates varied from year to year (as a function of either one- or two-way trips).

Case studies

Greater amberjack

Greater amberjack simulations resulted in estimates of steepness that were more precise than those estimated in the general simulation study. However, steepness was frequently estimated at the upper bound when true steepness was

Fig. 3. Predicted probability of steepness being estimated at 1.0 ($\hat{\pi}$) as a function of simulation inputs. Predictions are generated by varying x -axis values while fixing all other input variables to their mean. Input variables include the true value of steepness (steepness), natural mortality (M), the number of years for which data are available (years), the annual sample size for age composition data (n), the standard deviation of spawner–recruit errors (σ_R), and the assumed lognormal standard deviation for relative abundance data (σ_I). Solid lines give predictions for constant fishing mortality with $F/F_{MSY} = 0.5$ ($F_{type} = 1$); broken lines for constant F at $F/F_{MSY} = 2.0$ ($F_{type} = 2$); dotted lines for a two-way trip ($F_{type} = 3$); and dashed–dotted lines for strictly increasing F ($F_{type} = 4$).



greater than about 0.5 (Fig. 2). Unlike the general simulation study, estimates hit the upper bound more frequently when steepness was estimated externally than internally. Accordingly, MARE was lower when steepness was estimated internally (0.091) than when estimated externally (0.125). Also, when steepness was not estimated at the upper bound, it tended to be biased for certain values of true, underlying steepness. For instance, the mode of steepness estimates was positively biased at lower steepness values (Fig. 2).

Application of inverse prediction produced a posterior distribution of steepness that was noticeably smoother than a distribution generated by bootstrapping spawner–recruit residuals (Fig. 5). The bootstrap procedure resulted in a bimodal distribution of steepness estimates, with the dominant mode at the upper bound and the secondary mode near 0.55. The inverse prediction procedure resulted in a unimodal distribution, with the majority of estimates greater than 0.55. The posterior mean and median of the inverse prediction density were 0.66 and 0.64, respectively, and thus closer to that estimated by the SCA analysis. Comparisons of F_{MSY} for the two approaches mirrored steepness results, with the bootstrap procedure resulting in a bimodal distribution and the inverse prediction method resulting in a unimodal posterior distribution (Fig. 5).

Gag grouper

For gag grouper, simulations revealed little bias in estimates of steepness, although the estimator was imprecise

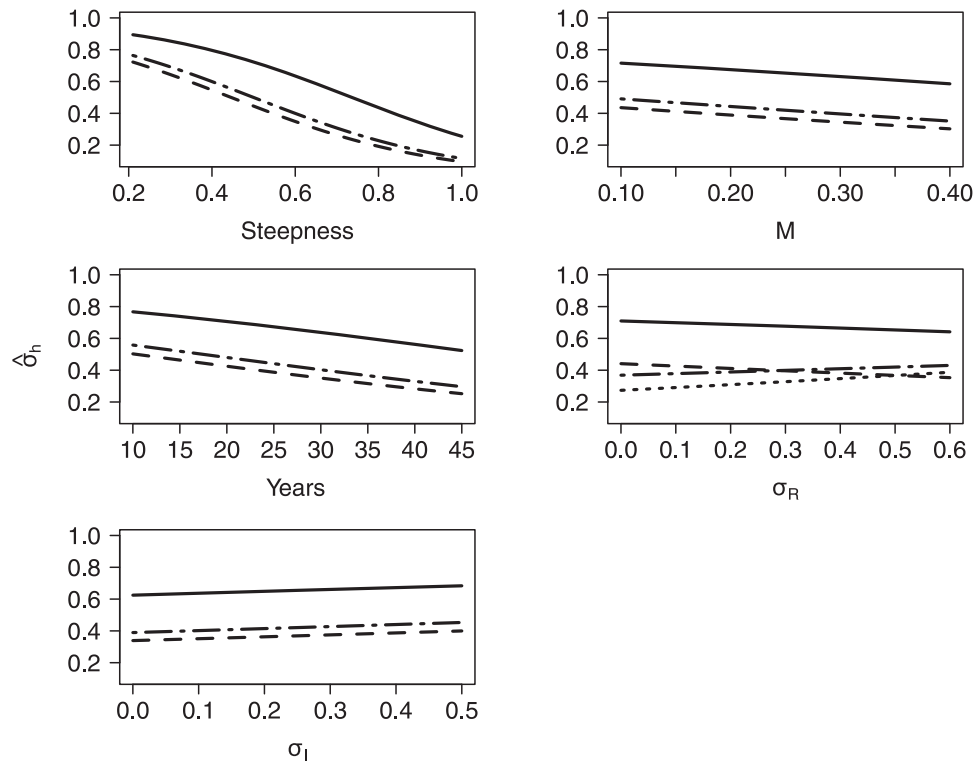
(Fig. 2). Steepness was estimated at the upper bound less often than in general or GAG simulations. In this case, the MARE for steepness was marginally lower when steepness was estimated externally (0.147) than internally (0.152).

Both inverse prediction and bootstrap procedures indicated that the stock was reasonably productive. However, inverse prediction once again outperformed the bootstrap approach, as it produced a posterior distribution of steepness that was smoother and with much less probability mass at the upper bound (Fig. 6). The posterior mean and median were 0.65 and 0.68, respectively, which were both lower than the SCA estimate of 0.95 from the GAG assessment. Paralleling the pattern in steepness, the bootstrap procedure resulted in substantial mass at $F_{MSY} = 0.24$, and the inverse prediction model admitted considerably more uncertainty (Fig. 6). For the inverse prediction approach, the posterior mean and median of F_{MSY} were both 0.14, suggesting that the value produced in the GAG assessment (0.24) may be too high.

Discussion

It is not uncommon in stock assessment for steepness to be estimated at its upper bound (e.g., GAG; SEDAR 2006). In this paper, we examined conditions in which steepness can and cannot be estimated reliably, as well as the factors that are important in discriminating between the two cases. In general, we found that the true steepness value and the

Fig. 4. Predicted standard deviation for estimated steepness as a function of simulation inputs for cases in which steepness was not estimated at its upper bound. Solid lines give predictions for constant fishing mortality with $F/F_{MSY} = 0.5$ ($F_{type} = 1$); broken lines for constant F at $F/F_{MSY} = 2.0$ ($F_{type} = 2$); dotted lines for a two-way trip ($F_{type} = 3$); and dashed-dotted lines for strictly increasing F ($F_{type} = 4$). The dotted line is indistinguishable from the broken line in most of the graphs. Simulation inputs include the true value of steepness (steepness), natural mortality (M), the number of years for which data are available (years), the standard deviation of spawner–recruit errors (σ_R), and the assumed lognormal standard deviation for relative abundance data (σ_I). The sample size of annual age compositions was not selected by AIC, so parameter effects are not presented here. Where not otherwise indicated, predictions are generated by evaluating the estimated GLM at the mean of other input variables.



exploitation history of the stock are important determinants of (i) whether steepness is estimated at its upper bound, and (ii) the precision of the steepness estimator. Two-way trips afforded the highest probability of estimating steepness precisely and accurately; in general, precision and accuracy increased as the number of years of data, quality of data (e.g., index CV), and natural mortality increased. Although the probability of estimating steepness at its upper bound increased as true steepness increased, it was still possible to arrive at such an estimate when productivity was low.

Our results differed somewhat from those of other studies that used SCA analysis to investigate the sensitivity of management benchmarks and model parameters (including h) to simulation inputs. For instance, Magnusson and Hilborn (2007) found that the reliability of estimates was much more dependent on having a good contrast in spawning stock biomass than on exploitation pattern (e.g., a one-way or two-way trip). In contrast, we found that two-way trips were quite important, improving both the accuracy and precision of estimates. However, the depletion levels attained in each study likely differ, and one-way trips tended to reach a lower depletion level than two-way trips in Magnusson and Hilborn's (2007) study.

Like Haltuch et al. (2008), we found that the length of the catch time series was an important predictor of the precision of steepness; however, it did not appear as important a de-

terminant for h being estimated at its upper bound. Haltuch et al. (2008) found that the variance of recruitment deviations was also important, with larger recruitment variances leading to less accurate steepness estimates. Our results were more equivocal, indicating that this finding varies as a function of exploitation regime. If exploitation patterns produce good contrast in biomass, increasing recruitment variation tends to decrease our ability to estimate steepness reliably. In contrast, if exploitation rates are held constant, it is only through high recruitment variance that one can hope to get contrast in stock biomass; in this case, increasing spawner–recruit variance can actually improve our ability to estimate steepness. Another difference between our study and that of Haltuch et al. (2008) concerned the importance of sampling intensity. Like Haltuch et al. (2008), we found that the sample size of age compositions was relatively unimportant. However, unlike their study, we found index variability to be an important predictor of estimator performance. Although we were unable to pinpoint the reason for this difference, we suspect that it is linked to the depletion levels achieved during the two simulation studies. In our general simulation study, populations were initialized assuming a constant, historical level of fishing that predated monitoring. In contrast, the simulations of Haltuch et al. (2008) assumed that stocks were started at virgin levels and were reduced by 65%–90% by the end of the study.

Fig. 5. Distributions of steepness (\hat{h}^I) and F_{MSY} for greater amberjack (*Seriola dumerili*). These distributions were generated in two ways: through bootstrap resampling (left panels) of spawner–recruit residuals and by using inverse prediction (right panels) to estimate a posterior distribution. Vertical broken lines represent values from the SCA analysis.

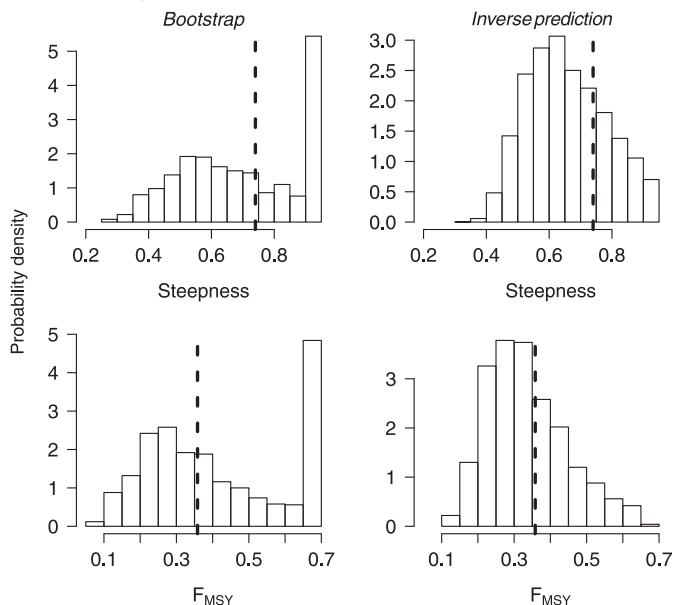
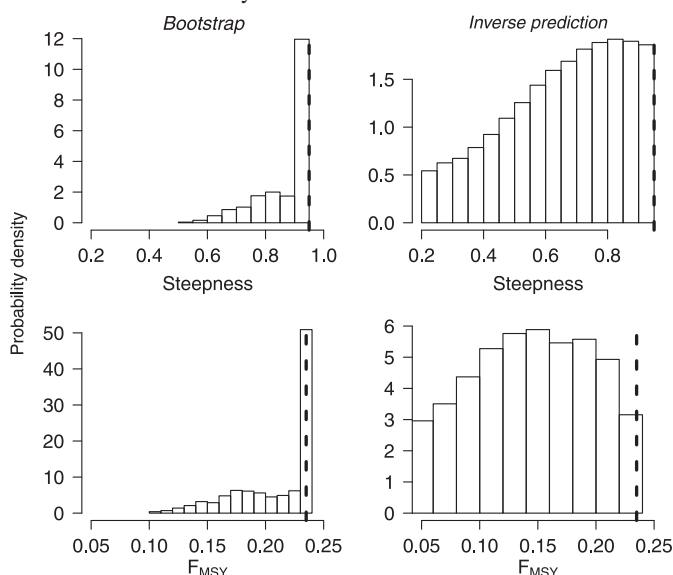


Fig. 6. Distributions of steepness (\hat{h}^I) and F_{MSY} for gag grouper (*Mycteroperca microlepis*). These distributions were generated in two ways: through bootstrap resampling of spawner–recruit residuals (left panels) and by using inverse prediction (right panels) to estimate a posterior distribution. Vertical broken lines represent values from the SCA analysis.



It is somewhat difficult to come to any definite conclusions based on comparisons of previous work and the study in this paper, in part because of a lack of adequate controls. For instance, each study differed in simulation design, types of data considered, model formulation, and objective functions. However, speculation is not entirely idle in this case. Clearly, contrast in spawning biomass is important for esti-

imating steepness; given equal contrast between stocks, our results suggest that two-way trips may result in better performance than one-way trips. Also, length of the catch time series appears to be a consistently important factor among studies, with longer time series leading to better performance of the steepness estimator. Similarly, age composition sample sizes do not appear to matter as much (although accuracy of age compositions may still be important). What is also clear from these studies is that although these factors can improve estimator performance, they do not guarantee that we will arrive at reliable estimates. For instance, we have shown here that simulations with good spawning biomass contrast and reasonable catch time series duration would still frequently result in steepness estimated at the boundary.

The general simulation study provided information on accuracy and precision of steepness in a relatively simple fishery (relatively short time series, a single index of abundance, a single gear, etc.). In contrast, case study simulations for GAJ and GAG were more detailed, mirroring the data sources and model structure used in the recent stock assessments. In these case studies, the accuracy and precision of steepness estimates increased dramatically, particularly for GAJ. In this case, it was rare that steepness was estimated at its upper bound when true steepness was less than 0.5. This finding suggests that estimates at the upper bound can actually be informative in some cases, at least for indicating which values of steepness are unlikely. When not estimated at its upper bound, steepness appears to be negatively biased when stocks are moderately to highly productive, a finding that agrees with Ianelli (2002), who found a similar pattern when examining the reliability of productivity estimates for ocean perch on the US west coast. Unfortunately, it appears necessary to perform fishery-specific simulations to examine bias and precision of the steepness estimator for individual stocks. The same holds true for examining whether internally or externally estimated values of steepness are more reliable.

We found inverse prediction methods to be extremely useful for characterizing posterior distributions of steepness that factored in species-specific simulation results. For GAJ, the posterior mean and median of true underlying steepness were almost identical to those obtained from the actual assessment and used for management, providing reinforcement that this steepness value was reasonable. For GAG, the inverse prediction approach suggested that the stock was likely highly productive, but that low productivity could not be discounted. Bayesian statisticians often prefer the posterior mean and median because they minimize squared and absolute error loss, respectively (Bernardo and Juárez 2003); for GAG, these values were 0.65 and 0.68, respectively. From this perspective, it appears that the estimate of 0.95 from the original assessment may be overly optimistic. As with steepness, we believe that the implied posterior distributions for F_{MSY} are more useful than those based on bootstrapping.

The inverse prediction approach is somewhat analogous to constructing a profile confidence interval for steepness. The latter approach is applicable when stock assessment models have been written strictly in terms of likelihood components (i.e., no ad hoc penalties, likelihood weightings, etc.). For

profile confidence intervals to be interpretable, one often needs to estimate a variance inflation factor to account for model misspecification or lack of fit (Burnham and Anderson 2002), in addition to estimating variance components internally or via the EM algorithm (Cotter and Buckland 2004; Deriso et al. 2007). Nevertheless, when interpretable, the profile confidence interval is preferable to the inverse prediction approach because the other parameters in the model are re-estimated when conditioning on particular steepness values. For instance, when comparing likelihood values with steepness fixed to a low value, we might expect a higher estimate of virgin recruitment (R_0) and lower estimates of fishing mortality. In contrast, the inverse prediction approach described in this paper conditions on virgin recruitment and fishing mortality rates that are originally estimated by the SCA analysis. We thus believe that the inverse prediction approach as described will tend to underestimate uncertainty. A possible solution is to vary more simulation inputs (e.g., R_0 and a fishing mortality scalar, b) when conducting a multivariate calibration experiment (Brown 1982; Benton et al. 2003).

We echo the caveat provided by Yin and Sampson (2004) and others that the simulation framework used here is a drastic simplification of how data are “generated” and collected in the real world. Despite tailoring some simulations to specific data features of individual stocks (for GAJ and GAG), data were analyzed using the same probability model, natural mortality rate, and CVs used to generate the data. As a general rule, in the real world, we would expect concurrent model and parameter misspecification. In addition, factors such as ageing error and recruitment autocorrelation will also tend to obscure the spawner–recruit relationship. If anything, realized estimates of steepness in the real world should be accorded more uncertainty than suggested here.

In light of our results, we suggest that stock assessment analysts exercise extreme caution in interpreting steepness estimates at the upper bound. In many cases, this result gives little indication of the true productivity of a stock, and overestimates of steepness would lead to overly optimistic management advice, particularly with regard to recovery projections for depleted stocks. Analysts should be equally cautious of steepness estimates at the lower bound. Although this phenomenon was rare in our SCA model configurations, it has been observed in other studies (e.g., in management strategy evaluations of hake; M. Haltuch, National Marine Fisheries Service, Northwest Fisheries Science Center, 2725 Montlake Blvd. East, Seattle, WA 98112–2097, personal communication, 2009) and would likely lead to overly pessimistic management advice. When steepness is estimated at either bound, application of proxy-based benchmarks or inverse prediction approaches may prove particularly useful. Likewise, knowledge of life history may prove useful for setting likely ranges on steepness (e.g., Rose et al. 2001; Goodwin et al. 2006; Brooks et al. 2009).

Acknowledgments

We thank R. Cheshire, P. Mace, M. Prager, and one anonymous referee for helpful comments on an earlier version of this manuscript. Opinions expressed are those of the authors

and do not necessarily reflect policies of any government agency.

References

- Benton, D., Krishnamoorthy, K., and Mathew, T. 2003. Inferences in multivariate–univariate calibration problems. *Statistician*, **52**: 15–39.
- Bernardo, J.M., and Juárez, M.A. 2003. Intrinsic estimation. *In Bayesian statistics. Edited by J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, and M. West.* Oxford University Press, Oxford, UK. pp. 465–476.
- Beverton, R.J.H. 1992. Patterns of reproductive strategy parameters in some marine teleost fishes. *J. Fish Biol.* **41**(Suppl. B): 137–160.
- Beverton, R.J.H., and Holt, S.J. 1957. On the dynamics of exploited fish populations. *Fisheries Investment Series 2*. Vol. 19. UK Ministry of Agriculture and Fisheries, London, UK. pp. 1–533.
- Brooks, E.N., Powers, J.E., and Cortés, E. 2009. Analytic reference points for age-structured models: application to data-poor fisheries. *ICES J. Mar. Sci.* **67**: 165–175.
- Brown, P.J. 1982. Multivariate calibration. *J. Roy. Stat. Soc. Ser. B (Meth.)*, **44**: 287–321.
- Burnham, K.P., and Anderson, D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. Springer-Verlag, New York.
- Charnov, E.L. 1993. Life history invariants: some explorations of symmetry in evolutionary ecology. Oxford University Press, Oxford, UK.
- Chen, D.G. 2004. Bias and bias correction in fish recruitment prediction. *N. Am. J. Fish. Manage.* **24**(2): 724–730. doi:10.1577/M03-053.1.
- Cotter, A.J., and Buckland, S.T. 2004. Using the EM algorithm to weight data sets of unknown precision when modelling fish stocks. *Math. Biosci.* **190**(1): 1–7. doi:10.1016/j.mbs.2004.03.001. PMID:15172800.
- Cox, D.R., and Oakes, D. 1984. Analysis of survival data. Chapman and Hall, London.
- Deriso, R.B., Quinn, T.J.I., II, and Neal, P.R. 1985. Catch-age analysis with auxiliary information. *Can. J. Fish. Aquat. Sci.* **42**(4): 815–824. doi:10.1139/f85-104.
- Deriso, R.B., Maunder, M.N., and Skalski, J.R. 2007. Variance estimation in integrated assessment models and its importance for hypothesis testing. *Can. J. Fish. Aquat. Sci.* **64**(2): 187–197. doi:10.1139/F06-178.
- Dorn, M.W. 2002. Advice on west coast rockfish harvest rates from Bayesian meta-analysis of stock–recruit relationships. *N. Am. J. Fish. Manage.* **22**(1): 280–300. doi:10.1577/1548-8675(2002)022<0280:AOWCRH>2.0.CO;2.
- Fournier, D., and Archibald, C.P. 1982. A general theory for analyzing catch at age data. *Can. J. Fish. Aquat. Sci.* **39**(8): 1195–1207. doi:10.1139/f82-157.
- Goodwin, N.B., Grant, A., Perry, A.L., Dulvy, N.K., and Reynolds, J.D. 2006. Life history correlates of density-dependent recruitment in marine fishes. *Can. J. Fish. Aquat. Sci.* **63**(3): 494–509. doi:10.1139/f05-234.
- Haltuch, M.A., Punt, A.E., and Dorn, M.W. 2008. Evaluating alternative estimators of fishery management reference points. *Fish. Res.* **94**(3): 290–303. doi:10.1016/j.fishres.2008.01.008.
- Hastie, T.J., and Tibshirani, R.J. 1999. Generalized additive models. Chapman and Hall/CRC Press, Boca Raton, Florida.
- Hilborn, R. 1979. Comparison of fisheries control systems that uti-

- lize catch and effort data. *J. Fish. Res. Board Can.* **36**: 1477–1489.
- Hilborn, R., and Walters, C.J. 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Chapman and Hall, New York.
- Ianelli, J.N. 2002. Simulation analyses testing the robustness of productivity determinations from west coast Pacific ocean perch stock assessment data. *N. Am. J. Fish. Manage.* **22**(1): 301–310. doi:10.1577/1548-8675(2002)022<0301:SATTRO>2.0.CO;2.
- Liermann, M., and Hilborn, R. 1997. Depensation in fish stocks: a hierarchical Bayesian meta-analysis. *Can. J. Fish. Aquat. Sci.* **54**(9): 1976–1984. doi:10.1139/cjfas-54-9-1976.
- Mace, P.M. 1994. Relationships between common biological reference points used as thresholds and targets of fisheries management strategies. *Can. J. Fish. Aquat. Sci.* **51**(1): 110–122. doi:10.1139/f94-013.
- Mace, P.M., and Doonan, I.J. 1988. A generalized bioeconomic simulation model for fish population dynamics. N.Z. Fisheries Association Resource Document No. 88/4.
- Magnusson, A., and Hilborn, R. 2007. What makes fisheries data informative? *Fish. Fish.* **8**: 337–358.
- McCullagh, P., and Nelder, J.A. 1989. Generalized linear models. Chapman and Hall, New York.
- Methot, R.D. 1989. Synthetic estimates of historical abundance and mortality for northern anchovy. In *Mathematical analysis for fish stock dynamics: reviews, evaluations and current applications*. Edited by E.F. Edwards and B.A. Megrey. Am. Fish. Soc. Symp. No. 6. pp. 66–82.
- Methot, R.D. 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. *Int. N. Pac. Fish. Comm. Bull.* **50**: 259–277.
- Michielsens, C.G.J., and McAllister, M.K. 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.
- Myers, R.A., Bowen, K.G., and Barrowman, N.J. 1999. Maximum reproductive rate of fish at low population sizes. *Can. J. Fish. Aquat. Sci.* **56**(12): 2404–2419. doi:10.1139/cjfas-56-12-2404.
- Osborne, C. 1991. Statistical calibration — a review. *Int. Stat. Rev.* **59**(3): 309–336. doi:10.2307/1403690.
- Otter Research Ltd. 2004. An introduction to AD Model Builder version 7.1.1. Otter Research Ltd., Box 2040, Sidney, British Columbia.
- Prager, M.H. 2002. Comparison of logistic and generalized surplus-production models applied to swordfish, *Xiphias gladius*, in the North Atlantic Ocean. *Fish. Res.* **58**(1): 41–57. doi:10.1016/S0165-7836(01)00358-7.
- Punt, A.E. 2003. The performance of a size-structured stock assessment method in the face of spatial heterogeneity in growth. *Fish. Res.* **65**(1–3): 391–409. doi:10.1016/j.fishres.2003.09.028.
- Quinn, T.J.I., and Deriso, R.B. 1999. Quantitative fish dynamics. Oxford University Press, New York.
- R Development Core Team. 2007. R: a language and environment for statistical computing [online]. Available at <http://www.R-project.org> [accessed 31 March 2009].
- Racine-Poon, A. 1988. A Bayesian approach to nonlinear calibration problems. *J. Am. Stat. Assoc.* **83**(403): 650–656. doi:10.2307/2289287.
- Ricker, W.E. 1975. Computation and interpretation of biological statistics of fish populations. *Bull. Fish. Res. Board Can.* No. 191.
- Rose, K.A., Cowan, J.H., Jr., Winemiller, K.O., Myers, R.A., and Hilborn, R. 2001. Compensatory density dependence in fish populations: importance, controversy, understanding and prognosis. *Fish. Fish.* **2**: 293–327.
- Shertzer, K.W., Prager, M.H., Vaughan, D.S., and Williams, E.H. 2008. Fishery models. In *Population dynamics*. Encyclopedia of Ecology. Vol. 2 of 5. Edited by S.E. Jorgensen and B. Fath. Elsevier, Oxford, pp. 1582–1593.
- SouthEast Data, Assessment, and Review. 2006. SEDAR 10 Assessment Workshop Report [online]. Available at <http://www.sefsc.noaa.gov/sedar/> [accessed 6 March 2009].
- SouthEast Data, Assessment, and Review. 2007. SEDAR 15 Assessment Workshop Report [online]. Available at <http://www.sefsc.noaa.gov/sedar/> [accessed 6 March 2009].
- Venables, W.N., and Ripley, B.D. 2002. Modern applied statistics with S. 4th ed. Springer, New York.
- Wilberg, M.J., and Bence, J.R. 2006. Performance of time-varying catchability estimators in statistical catch-at-age analysis. *Can. J. Fish. Aquat. Sci.* **63**(10): 2275–2285. doi:10.1139/F06-111.
- 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**(6): 710–720. doi:10.1139/f03-059.
- Yin, Y., and Sampson, D.B. 2004. Bias and precision of estimates from an age-structured stock assessment program in relation to stock and data characteristics. *N. Am. J. Fish. Manage.* **24**(3): 865–879. doi:10.1577/M03-107.1.

Appendix A. Details on MCMC sampling for steepness under inverse prediction

We assume that a simulation experiment is used to relate n estimates of steepness from an SCA analysis, $\mathbf{h}^{\text{est}} = \{h_1^{\text{est}}, h_2^{\text{est}}, \dots, h_n^{\text{est}}\}$, to true values of steepness used to generate the data ($\mathbf{h}^{\text{true}} = \{h_1^{\text{true}}, h_2^{\text{true}}, \dots, h_n^{\text{true}}\}$). The two are related through a function of the form $h_i^{\text{est}} = f(h_i^{\text{true}}, \theta_i) + \varepsilon_i$, which is supplied by the user (for example, by specification of a linear model). Here θ_i denotes a collection of unknown parameters, and ε_i is a random error term. For convenience, let this relationship be formalized through a joint probability density function given by

$$[\mathbf{h}^{\text{est}} | \mathbf{h}^{\text{true}}, \theta]$$

where $[X|Y]$ denotes the conditional distribution of X given Y .

We suppose that the investigator uses an SCA analysis to provide an estimate of steepness, h_0^{est} , for a single stock of interest. Understanding that h_0^{est} could have arisen in any number of ways, we propose to estimate a posterior distribution of true steepness, h_0^{true} . Following Racine-Poon (1988), this can be done by sampling from

$$(A1) \quad [h_0^{\text{true}}, \theta | \mathbf{h}^{\text{est}}, \mathbf{h}^{\text{true}}, h_0^{\text{est}}] \propto [h_0^{\text{est}} | h_0^{\text{true}}, \theta] [\mathbf{h}^{\text{est}} | \mathbf{h}^{\text{true}}, \theta] [\theta] [h_0^{\text{true}}]$$

Here, $[h_0^{\text{true}}]$ is a prior distribution of the true steepness of the unobserved stock, and $[\theta]$ specifies a joint prior distribution of remaining unknown parameters (i.e., those describing functional relationships in the calibration experiment). For purposes of this study, we set $[h_0^{\text{true}}]$ to be uniformly distributed on (0.2, 0.95) to correspond to the configuration of the assessment model (we note that this range is different from

the (0.21, 1.0) range used in the general simulation study). If data were available, another possibility would be to use meta-analysis to specify an informative prior distribution (cf., Myers et al. 1999; Dorn 2002). An improper but diffuse prior was specified for $[\theta]$ by setting it proportional to a fixed constant.

For both case studies, we assumed that true and estimated steepness were related by a normal mixture model (i.e., eq. 2). Define an indicator variable z_i for the i th simulation, which takes on the value of one if $h_i^{\text{obs}} = u$ and zero otherwise. The normal mixture model may then be rewritten as

$$(A2) \quad [\mathbf{h}^{\text{est}} | \mathbf{h}^{\text{true}}, \theta] = \prod_{i=1}^n \pi_i^{z_i} (1 - \pi_i)^{1-z_i} \left[\frac{\mathcal{N}(h_i^{\text{est}}; \mu_i, \sigma_i)}{\int_l^u \mathcal{N}(h_i^{\text{est}}; \mu_i, \sigma_i) d\hat{h}_i^{\text{est}}} \right]^{1-z_i}$$

where the upper (u) and lower (l) bounds of integration are set to the limits used in the SCA analysis. The model for observed steepness of the stock of interest (h_0^{obs}) is quite similar. In this case, define $z_0 = 1$ if $h_0^{\text{obs}} = u$ and $z_0 = 0$ otherwise. We then have

$$(A3) \quad [h_0^{\text{est}} | h_0^{\text{true}}, \theta] = \pi_0^{z_0} (1 - \pi_0)^{1-z_0} \left[\frac{\mathcal{N}(h_0^{\text{est}}; \mu_0, \sigma_0)}{\int_l^u \mathcal{N}(h_0^{\text{est}}; \mu_0, \sigma_0) d\hat{h}_0^{\text{est}}} \right]^{1-z_0}$$

In practice, there are two possible cases related to whether or not $z_0 = 1$. If $z_0 = 1$ (as was the case for GAG), $[h_0^{\text{est}} | h_0^{\text{true}}, \theta] = \pi_0^{z_0}$. Thus, when $z_0 = 1$, knowledge of σ and μ does not provide information about h_0^{true} (that is, σ and μ are ancillary for h_0^{true} in a formal statistical sense). If h_0^{true} is the only parameter of interest, this implies that the bracketed quantity in eq. A2 may be omitted during analysis. In contrast, if $z_0 = 0$, inference must simultaneously be drawn on all parameters.

For GAJ, we specified eq. A1 in terms of eight parameters: $h_0^{\text{true}}, \beta_1, \beta_2, \dots, \beta_7$. We assumed the following relationships:

$$(A4) \quad \begin{aligned} \text{logit}(\pi_i) &= \beta_1 + \beta_2 h_i^{\text{true}} \\ \text{logit}(\mu_i) &= \beta_3 + \beta_4 h_i^{\text{true}} + \beta_5 (h_i^{\text{true}})^2 \\ \text{logit}(\sigma_i) &= \beta_6 + \beta_7 h_i^{\text{true}} \end{aligned}$$

Three parameters were estimated for the GAG analysis: $h_0^{\text{true}}, \beta_1$, and β_2 . The definition for β_1 and β_2 was the same as for GAJ (i.e., as specified in eq. A4).

For each stock, parameters were sampled cyclicly using Metropolis–Hastings updates on their respective full conditional distributions (Gelman et al. 2004). We implemented a “random walk” chain (Givens and Hoeting 2005), whereby proposals (i.e., candidate parameter values) were generated as random normal variates with means centered at the parameter value at the previous iteration. Several sets of parameters were subjected to a high degree of correlation and were updated together in blocks. For instance, proposals for β_1 and β_2 were generated as multivariate normal random deviates with mean set to the previous iteration’s values and covariance specified to generate a correlation of -0.99 . This modification substantially improved mixing. The parameters β_3, β_4 , and β_5 were also updated together, with covariances chosen to achieve the correlations $\rho_{\beta_3, \beta_4} = -0.9$, $\rho_{\beta_3, \beta_5} = 0.9$, and $\rho_{\beta_4, \beta_5} = -0.9$. This modification improved mixing slightly, but autocorrelation was still high. The remaining parameters were updated independently. Standard deviations for Metropolis–Hastings proposals were specified so as to result in a 30%–40% acceptance rate, as suggested by Gelman et al. (2004). Convergence was determined by examining trace plots of individual parameters, and by computing Gelman–Rubin statistics (Gelman et al. 2004).

For GAJ, the MCMC sampling procedure was run twice, resulting in two chains of length 100 000. Gelman–Rubin statistics indicated convergence of all parameters after about 40 000 iterations; thus the last halves of each chain were combined to yield a posterior sample of 100 000. For GAG, mixing was much better; in this case, two chains of length 10 000 were run, and Gelman–Rubin statistics indicated convergence after about 2000 iterations. The second halves of the two chains were combined to yield a posterior sample of 10 000 from which posterior moments were calculated. All analyses were performed in the statistical programming language R (R Development Core Team 2007).

References

- Dorn, M.W. 2002. Advice on west coast rockfish harvest rates from Bayesian meta-analysis of stock–recruit relationships. *N. Am. J. Fish. Manage.* **22**(1): 280–300. doi:10.1577/1548-8675(2002)022<0280:AOWCRH>2.0.CO;2.
- Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. 2004. *Bayesian data analysis*. 2nd ed. Chapman and Hall, Boca Raton, Florida.
- Givens, G.H., and Hoeting, J.A. 2005. *Computational statistics*. John Wiley and Sons, Hoboken, New Jersey.
- Myers, R.A., Bowen, K.G., and Barrowman, N.J. 1999. Maximum reproductive rate of fish at low population sizes. *Can. J. Fish. Aquat. Sci.* **56**(12): 2404–2419. doi:10.1139/cjfas-56-12-2404.
- R Development Core Team. 2007. *R: a language and environment for statistical computing* [online]. Available at <http://www.R-project.org> [accessed 31 March 2009].
- Racine-Poon, A. 1988. A Bayesian approach to nonlinear calibration problems. *J. Am. Stat. Assoc.* **83**(403): 650–656. doi:10.2307/2289287.