



## Can steepness of the stock–recruitment relationship be estimated in fishery stock assessment models?

Hui-Hua Lee<sup>a,c,\*</sup>, Mark N. Maunder<sup>b</sup>, Kevin R. Piner<sup>c</sup>, Richard D. Methot<sup>d</sup>

<sup>a</sup> Joint Institute for Marine and Atmospheric Research, University of Hawaii, 1000 Pope Street, Honolulu, HI 96822, USA

<sup>b</sup> Inter-American Tropical Tuna Commission, 8604 La Jolla Shores Drive, La Jolla, CA 92037, USA

<sup>c</sup> NOAA Fisheries, Pacific Islands Fisheries Science Center, 2570 Dole Street, Honolulu, HI 96822, USA

<sup>d</sup> NOAA Fisheries, Northwest Fisheries Science Center, 2725 Montlake Blvd. E., Seattle, WA 98112, USA

### ARTICLE INFO

#### Article history:

Received 6 August 2011

Received in revised form 3 March 2012

Accepted 4 March 2012

#### Keywords:

Pacific Coast groundfish stocks

Steepness

Stock assessment

Stock–recruitment

Stock synthesis

### ABSTRACT

Steepness of the stock–recruitment relationship is one of the most uncertain and critical quantities in fishery stock assessment and management. Steepness is defined as the fraction of recruitment from a virgin population obtained when the spawners are at 20% of the virgin level. Steepness directly relates to productivity and yield and is an important element in the calculation of many management reference points. Stock–recruitment relationships have traditionally been estimated from time series of recruitment and spawning biomass, but recently interest has arisen regarding the ability to estimate steepness inside fishery stock assessment models. We evaluated the ability to estimate steepness of the Beverton–Holt stock–recruitment relationship using simulation analyses for twelve US Pacific Coast fish stocks. A high proportion of steepness estimates from the simulated data and the original data occur at the bounds for steepness and the proportion decreased as the true steepness decreased. The simulation results indicate that, in most cases, steepness was estimated with moderate to low precision and moderate to high bias. The poorly estimated steepness indicates that often there is little information in the data about this quantity. However, reliable estimation is attainable with a good contrast of spawning stock biomass for relatively unproductive stocks when the model is correctly specified.

© 2012 Elsevier B.V. All rights reserved.

### 1. Introduction

Stock productivity is fundamental in fishery stock assessment and management. Productivity influences stock size, yield, and management reference points. When a stock is exploited to small population size, its regenerative capacity (resilience) becomes crucial to the recovery of the stock. A common measure of stock resilience is the steepness ( $h$ ) of the stock–recruitment relationship (Mace and Doonan, 1988; Francis, 1992). Steepness is defined as the fraction of recruitment from a virgin population obtained when the spawners are at 20% of the virgin level.

Unfortunately, steepness of the stock–recruitment relationship is one of the most uncertain population dynamic quantities, and determining its value is challenging. In practice, the form of a stock–recruitment relationship is usually unclear although its theory is well developed (Beverton and Holt, 1957/1993; Ricker, 1954; Shepherd and Cushing, 1980) and has been used to provide management advice. The inability to detect a relationship often results

from the lack of estimates of recruitment at either low spawning stock size or a lack of contrast in spawning stock size (Hilborn and Walters, 1992; Magnusson and Hilborn, 2007; Conn et al., 2010). Furthermore, the large amount of variability in recruitment caused by factors other than spawner abundance obscures the relationship.

In many stock assessments, steepness is not based on reliable data from the stock being studied, but is borrowed from other stocks or species (e.g., using meta-analyses; Myers et al., 1999, 2002; Dorn, 2002) or based on life history theory (He et al., 2006; Mangel et al., 2010). However, bias may occur in the chosen value because both recruitment values and the spawner values have error, selected stocks may not be representative, and environmental factors may have an influence on recruitment (Schirripa and Colbert, 2006).

There is increasing interest in estimating influential parameters such as natural mortality (Lee et al., 2011) and steepness (Magnusson and Hilborn, 2007) inside population dynamics models. The estimation of the influential parameters is facilitated by increases in computing power and available software (e.g., Fournier et al., 2012) allowing for the inclusion of a wider range of data types (e.g., tagging, environment data etc.) and more complex model processes (e.g., ageing error) (Maunder et al., 2009). In these integrated analyses, model processes are influenced by data sources that are

\* Corresponding author at: 2570 Dole Street, Honolulu, HI 96822, USA. Tel.: +1 808 983 5352; fax: +1 808 983 2902.

E-mail addresses: [Huihua.Lee@noaa.gov](mailto:Huihua.Lee@noaa.gov), [huihua.lee@gmail.com](mailto:huihua.lee@gmail.com) (H.-H. Lee).

not direct observations of that process via their connection to the underlying population dynamics. Therefore, parameters often fixed at assumed values because of a lack of direct observations, may in practice be estimable under certain conditions (Magnusson and Hilborn, 2007; Conn et al., 2010; Lee et al., 2011).

In this study, we build on previous work by Magnusson and Hilborn (2007) and Conn et al. (2010) by including a wider range of example species and more complex models to evaluate the reliability of stock–recruitment steepness estimates. We use simulation methods (Lee et al., 2011) in conjunction with example assessments to evaluate the influence of species life history, available data and historical exploitation patterns on the reliability of steepness estimates and overall model performance.

## 2. Methods

### 2.1. The assessment model

Stock assessment models (data and model configuration files) were obtained for twelve Pacific stocks (Table 1). Available stocks included pelagic (hake), flatfishes (arrowtooth flounder and english sole), rockfishes (black rockfish north, black rockfish south, blue rockfish, canary rockfish, chilipepper rockfish, darkblotched rockfish, shortbelly rockfish, yelloweye rockfish) and roundfishes (sablefish). All assessments were conducted using the integrated stock assessment model, stock synthesis (SS, Methot, 2000, 2006, 2009), and peer-reviewed (<http://www.pcouncil.org/groundfish/stock-assessments/>) in 2006–2008.

### 2.2. Simulation analyses

Simulation analyses were conducted to evaluate the ability to estimate  $h$  within these assessment models. The underlying assumption behind this is that these assessment models reflect a “true” population with the stock’s life history and exploitation patterns described in the assessment. Therefore, the same model assumptions, catch, and parameter estimates used in the original assessments were used to generate simulated data (Table 2). We parametrically generated 500 data sets based on characteristics of the real data and the likelihood functions used in the model. In order to avoid potential bias from patterns in the recruitment residuals in the original assessment, recruitment deviates with the same variability ( $\sigma_R$ ) as assumed in the original assessment were randomly generated. This new random recruitment deviate vector and other parameters from the original assessment were used to simulate the new data sets that have same data components as the original assessment (Lee et al., 2011).

The model was then fit to the simulated data sets estimating  $h$  in the last phase of parameter estimation to avoid local minimum. The initial value of  $h$  was set to the true value. Priors were not used on any parameters. We defined model convergence when the eigen values of the Hessian matrix are positive definite. To determine if the estimation of  $h$  was the cause of non-convergence all simulated data sets were fit using a model with steepness fixed (not estimated) at the true value. If previously non-converged models converged after fixing  $h$ , we considered the estimation of  $h$  to be cause of the convergence issue. Non-converged model runs were then removed from further analysis. The resulting estimates of  $h$  from converged models were used to describe its distribution when data conformed to model assumptions. Finally,  $h$  was estimated inside the stock assessment model using the original data and compared with the results from the simulation analysis.

The performance of converged simulation model runs was evaluated by three indicators: (1) the proportion of estimates at the parameter boundary, (2) variability (coefficient of variation) of estimates, and (3) median of bias. The median of bias was used as statistics to summarize performance across simulated data sets.

$$\text{Median bias} = \text{median} \left( \frac{\hat{\theta} - \theta}{\theta} \right) \quad (1)$$

where  $\hat{\theta}$  is the estimated value of  $h$  from the simulated data sets and  $\theta$  is the true value from the original assessment model. The median value is used to make the performance indicator robust to outliers. The estimates are reliable when the proportion of estimates at the bounds is close to zero, median bias is close to zero, CV is small, and the estimate from the original data lies inside a distribution of estimates from the simulation analysis.

In order to illustrate if original data, in total or each likelihood component, was informative about  $h$ , a profile likelihood was computed over the entire range of  $h$  for each likelihood component as well as the total likelihood. If a log-likelihood profile shows a well-defined convex curvature with definite maximum, this component is informative about the true value of  $h$ . Conversely, the lack of a well-defined curvature would suggest a lack of information about  $h$ . The likelihood ratio statistic ( $LR$ ) was used to test if upper and lower bound of parameter have high enough likelihood to define curvature (Pawitan, 2001):

$$LR_i = 2 \ln \left( \frac{L(\hat{\theta})}{L(\theta_i)} \right) \sim \chi^2_1(0.95) \quad (2)$$

where  $\hat{\theta}$  is the maximum likelihood estimate (MLE) of  $h$ ,  $\theta_i$  is the specific value of interest for  $h$  ( $i=0.2$  and  $1.0$ ), and  $L(\hat{\theta})$  and  $L(\theta_i)$  are their likelihood functions. We reject the null hypothesis ( $H_0 : \hat{\theta} = \theta_i$ ) at  $\alpha=0.05$  if  $LR_i > \chi^2_1(0.95)$  where  $\chi^2_1(0.95)$  is the 0.95 quantile of a chi-squared distribution with one degree of freedom which is about 3.84. In other words, this value yields statistically significant different log-likelihood value. The well-defined curvature has definite maximum with rejection of  $H_0$  at both lower and upper bound of  $h$ .

To demonstrate how well steepness can be estimated within the assessment models when each stock experienced different exploited strength, two statistics were used to characterize exploitation history of the simulated stock; maximum depletion is defined as:

$$\text{Maximum depletion} = \min_{t=t_1, t_2, \dots, t_N} \left( \frac{SSB_t}{SSB_0} \right) \quad (3)$$

where  $SSB_t$  and  $SSB_0$  are the estimated spawning stock biomass in year  $t$  and estimated virgin spawning stock biomass. The lowest value of historical depletions (ratio of the estimated SSB in year  $t$  to the virgin SSB) was defined as maximum depletion level of the stock. Maximum population increase is used to describe stock increase after being maximally depleted in year  $t^*$ .

Maximum population increase

$$= \max_{t=t^*, \dots, t_N} \left( \frac{SSB_t}{SSB_0} \right) - \min_{t=t_1, t_2, \dots, t_N} \left( \frac{SSB_t}{SSB_0} \right) \quad (4)$$

Due to a different set of random recruitments used, each simulation run possessed a different degree of exploited strength. The medians of these two statistics were used to make general conclusion.

## 3. Results

Based on the original assessments, the stocks used in this study experienced various levels of historical depletion, ranging

**Table 1**  
Source of the twelve groundfish assessments used in this study.

Species	Assessment report
Arrowtooth	Stock assessment of the arrowtooth flounder ( <i>Atheresthes stomias</i> ) population off the West Coast of the United States in 2007 Isaac C. Kaplan and Thomas E. Helser
Black RF_N	Status of the black rockfish resource North of Cape Falcon, Oregon to the U.S.-Canadian border in 2006 Farron R. Wallace, Yuk Wing Cheng, and Tien-Shui Tsou
Black RF_S	The Status of Black rockfish off Oregon and California in 2007 David B. Sampson
Blue RF	The 2007 assessment of blue rockfish ( <i>Sebastes mystinus</i> ) in California Meisha Key, Alec D. MacCall, John Field, Debbie Aseltine-Neilson and Kirk Lynn
Canary	Status of the U.S. Canary rockfish resource in 2007 Ian J. Stewart
Chilipepper	Status of the chilipepper rockfish, <i>Sebastes goodei</i> , in 2007 John C. Field
Darkblotched	Status and future prospects for the darkblotched rockfish resource in waters off Washington, Oregon, and California as assessed in 2007 Owen S. Hamel
English sole	Updated U.S. English sole stock assessment: status of the resource in 2007 Ian J. Stewart
Hake	Stock assessment of Pacific hake, <i>Merluccius productus</i> (a.k.a Whiting) in U.S. and Canadian waters in 2008 Thomas E. Helser, Ian J. Stewart, and Owen S. Hamel
Sable	Status of the sablefish resource off the continental U.S. Pacific Coast in 2007 Michael J. Schirripa
Shortbelly	Stock assessment model for the shortbelly rockfish, <i>Sebastes jordani</i> , in the California current John C. Field, Edward J. Dick, and Alec D. MacCall
Yelloweye	Update to the status of yelloweye rockfish ( <i>Sebastes ruberrimus</i> ) off the U.S. West Coast in 2007 John R. Wallace

in population size from 10% to 60% of virgin spawning biomass and represented different levels of productivity (Figs. 1 and 2). These assessments had a wide range of fisheries characteristics (Table 2), ranging from one to multiple fisheries and surveys, and were fit to a variety of data types (Table 2), including common data types (total catch, indices/surveys, and length composition data) and specific data types (age composition, conditional age-at-length composition, mean length-at-age data or mean body-weight information).

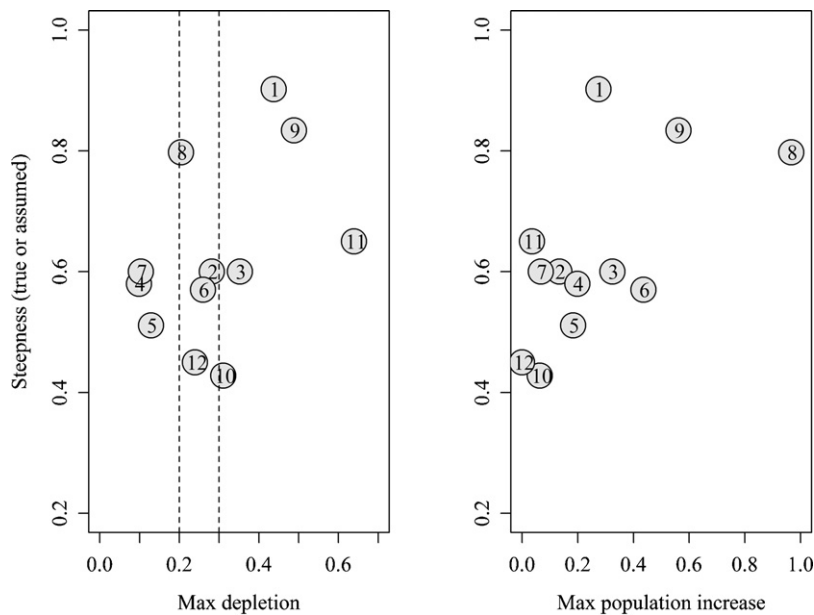
The proportion of converged models (Table 3) obtained for all simulated data sets ( $h$  estimated), ranged from 40% (chilipepper) to 100% (black rockfish south, blue rockfish, and sable fish). Subsequent model runs with  $h$  fixed at its “true” (assumed) value did not generally improve the proportion of models converging. The most improvement was noted in chilipepper rockfish and hake (Table 3). We also used the comparison of fixed and estimated  $h$  to determine that discarding those cases with non-converged models did not generate selection bias in results of the next set of analyses.

Across the simulations, 4 out of 12 assessments showed a large proportion of estimates at the bounds (i.e., more than 40% of estimates at either upper bound of 1.0 or lower bound of 0.2), particularly the upper bound that represents recruitment independent of stock size (Fig. 3). The percentage of model runs that estimated steepness at a bound decreased as the true steepness decreased. These bounded estimates of  $h$  resulted in ill-defined distributions of estimated  $h$  (Fig. 4). The performance of the simulation analyses was further evaluated from the unbounded estimates of  $h$ . The results showed that a moderate to low precision ( $CV \geq 17\%$ ) and moderate to high bias ( $\geq 10\%$ ) were obtained in the most cases (Table 3). In addition, there was no consistent direction to the bias among stocks.

Most assessments fitted to the original data estimated  $h$  at the parameter bounds, close to the bound or outside the distribution of the simulation results (Fig. 4). In the four cases for which the estimates of  $h$  from the original data were inside the distribution of the simulations and the estimates were close to “true” values, english sole, hake, and sablefish had the true value estimated in

**Table 2**  
The twelve assessment models in terms of data types used (denoted by check mark) and parameter estimated (denoted by check mark).

	Arrowtooth	Black RF_N	Black RF_S	Blue RF	Canary	Chilipepper	Darkblotched	English sole	Hake	Sable	Shortbelly	Yelloweye
<b>Data</b>												
Number of fleets	4	3	6	3	12	4	1	2	2	3	1	7
Catch	X	X	X	X	X	X	X	X	X	X	X	X
Number of indices (surveys)	4	2	10	3	4	4	4	2	1	5	7	5
Discard							X	X		X		
Age data		X	X		X	X			X	X	X	X
Length data	X	X	X	X	X	X	X	X	X	X	X	X
Age-length	X			X	X		X	X				
Mean length-at-age		X	X							X		X
Mean body-weight			X					X		X		
<b>Parameter estimated</b>												
Unfished recruitment	X	X	X	X	X	X	X	X	X	X	X	X
Natural mortality					X				X			
Steepness								X	X	X		
Growth	X	X	X	X	X		X	X	X	X		X
Length CV					X		X	X	X	X		X
At least one catchability estimated	X	X	X	X	X	X	X	X	X		X	X
At least one asymptotic selectivity	X	X	X	X	X	X	X	X		X	X	X



**Fig. 1.** Stock resilience (steepness) under different levels of exploitation histories of twelve fish assessments. Maximum depletion is the lowest value of historical spawning stock biomass (SSB) relative to the virgin SSB (left panel) and maximum population increase describes the rate of stock increase after being maximally depleted (right panel). Vertical lines of left panel indicate 20% and 30% of the virgin level. Twelve stocks (numbers in the circles) are as follows: (1) arrowtooth flounder, (2) black rockfish north, (3) black rockfish south, (4) blue rockfish, (5) canary rockfish, (6) chilipepper rockfish, (7) darkblotched rockfish, (8) english sole, (9) hake, (10) sablefish, (11) shortbelly rockfish, and (12) yelloweye rockfish.

the original assessments. Although estimating  $h$  seems plausible for these assessments, the estimates of  $h$  from the original data for 3 out of 4 assessments were uncertain given the wide confidence intervals.

The results of the profile likelihoods suggested that, in most cases, neither individual data types nor all data taken together provided substantial information regarding the value of steepness. The exceptions were darkblotched rockfish, sablefish, and yelloweye which had likelihood profiles of some individual likelihood components as well as the total likelihood which indicated significant information on the magnitude of  $h$ . Table 4 summarizes comparison of profile likelihood among the twelve assessments.

The reliability of estimating steepness (smaller CV and less bias) tended to improve with increased depletion (Fig. 5). For the simulated stocks that had been depleted to low levels (dark circles in Fig. 6), increased population size after depletion tended to obtain less biased estimates of  $h$ . However, the relationship of bias and population increase was somewhat weaker due to small sample size of depleted stocks and there was no improvement in precision of the estimate.

#### 4. Discussion

Our results generally agreed with recent studies that  $h$  is poorly estimated except for certain conditions (Magnusson and Hilborn, 2007; Conn et al., 2010). Magnusson and Hilborn (2007) concluded that  $h$  can be estimated reliably from abundance-index or age-composition data when the data cover a period in which abundance varies substantially. Conn et al. (2010) further found that contrast in spawning biomass is important for estimating steepness and given equal contrast between stocks, two-way trips may result in better precision and accuracy than a one-way trip. Our study also found that contrast in the spawning biomass increased the precision and might reduce bias in estimates of  $h$ . However, unlike Conn et al. (2010) the two-way trip did not appear to improve the precision of the estimate. Unfortunately, we are unable to generalize the relationship because the twelve assessment examples are not a random sample from the population of all fish stocks and sample size is small. A more general conclusion may be obtained by including more assessments or simulating different depletion and recovery levels with different catch time

**Table 3**

True or assumed steepness, proportion of converged models, coefficient of variation (CV), and median bias of estimates from the simulation analyses where the CV and bias were based on the converged models and unbounded estimates. The median bias represents tendency of estimates towards the true value from the assessment models.

	True or assumed	Proportion converged ( $h$ estimated)	Proportion converged ( $h$ fixed)	CV	Median bias (in %)
Arrowtooth	0.902	0.94	0.97	0.19	−0.15
BlackRF.N	0.6	0.91	0.63	0.30	−0.20
BlackRF.S	0.6	1.00	1.00	0.32	−0.12
BlueRF	0.58	1.00	0.99	0.17	0.09
Canary	0.511	0.93	0.91	0.25	0.05
Chilipepper	0.57	0.40	0.54	0.20	0.13
Darkblotched	0.6	0.99	1.00	0.21	−0.02
English sole	0.797663	0.79	0.78	0.17	−0.04
Hake	0.834076	0.77	0.91	0.39	−0.41
Sable	0.427822	1.00	1.00	0.25	−0.13
Shortbelly	0.65	0.86	0.86	0.39	−0.56
Yelloweye	0.45	0.51	0.48	0.11	0.19

**Table 4**

Likelihood comparison from profile log-likelihood analysis across range of steepness (0.2–1.0) for the twelve assessment models. If a log-likelihood profile has a well-defined convex curve (see Section 2), likelihood component is informative about steepness (denoted by plus signs). Minus signs represent that likelihood component is uninformative and blank represent that data is not available.

Likelihood component	Arrowtooth	Black RF_N	Black RF_S	Blue RF	Canary	Chilipepper	Darkblotched	English sole	Hake	Sable	Shortbelly	Yelloweye
Total	–	–	–	–	–	–	+	–	–	+	–	+
Indices (survey)	–	–	–	–	–	–	+	–	–	–	–	+
Discard	–	–	–	–	–	–	–	–	–	–	–	–
Length comps	–	–	–	–	–	–	–	–	–	–	–	+
Age comps	–	–	–	–	–	–	+	–	–	+	–	–
Size-at-age	–	–	–	–	–	–	–	–	–	–	–	–
Mean body-weight	–	–	–	–	–	–	–	–	–	–	–	–

+, informative about steepness; –, uninformative about steepness.

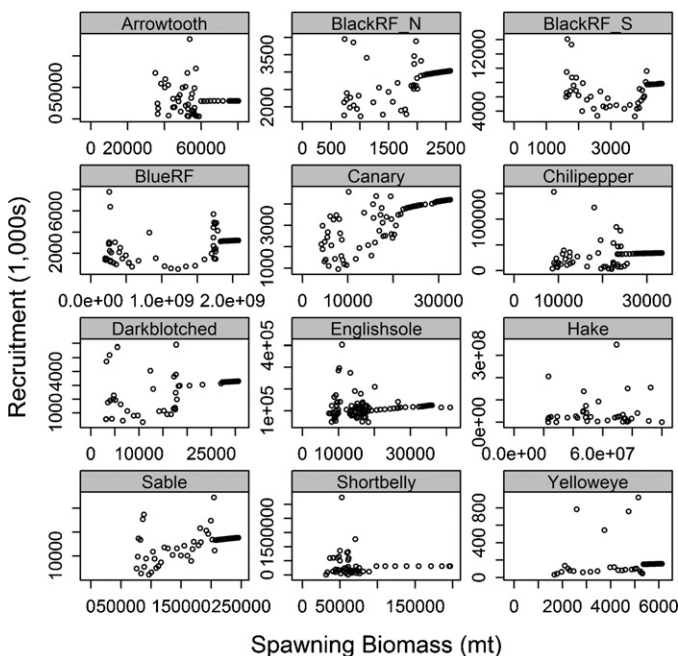
series for each stock. Notwithstanding its limitations, this study seems to reach similar conclusion to previous work.

A new result of this study is that the estimation of  $h$  did not in general lead to greater convergence issues than a fixed value. Our study included models that estimated both  $h$  and  $M$ , which are often thought to be too correlated to estimate simultaneously (Magnusson and Hilborn, 2007) resulting in convergence difficulties. For this work, we defined a model run with an invertible Hessian as a converged model run. Although this is standard practice, convergence would be more appropriately defined as a model that has found the global solution. This is often tested with random starting values or changes in phasing of parameters, but this search of the likelihood solution was beyond the scope of this work.

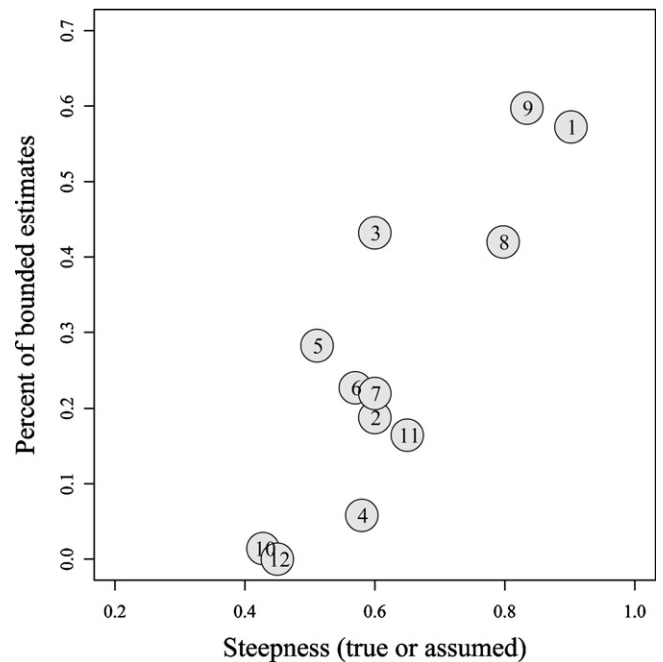
Our work represented a broad spectrum of life histories and human impacts. Stocks included in this work are from 4 different families of fishes with ecologies ranging from pelagic and migratory to demersal and sedentary. These stocks represented a range of productivity as described by the assumed levels of natural mortality and steepness, with productive stocks including the flatfishes and low productivity including the large bodied rockfishes. Although the twelve assessments may not cover full range of fish life histories, it is wider than often considered in other studies.

Furthermore, the selected assessments contain unevenness in the complexity of the models and the data as well as the differences in the quality and amounts of data which reflect the true variety encountered in actual stock assessments. This includes incorporating multiple fisheries/surveys, selectivity patterns and data types that is often too complex to create when using a single operating model. By increasing the complexity of the models beyond the previous analysis, we also included the complications due to information content among various data sources which were not applicable to previous studies.

Another difference between our study and others was the assumption made for the initial conditions. Conn et al. (2010) started their models at equilibrium fishing mortality equal to that in year 1. In contrast, the assumption made in our simulation studies was that almost all stocks were started at their virgin levels. As for the former, additional assumptions about harvest schedules need to be made to account for different exploitation history (Magnusson and Hilborn, 2007; Conn et al., 2010). As for the later, various deletion levels were compared between stocks. The depletion levels attained in each study differ. In Magnusson and Hilborn's study, one-way trip reached a lower depletion level than two-way trips;

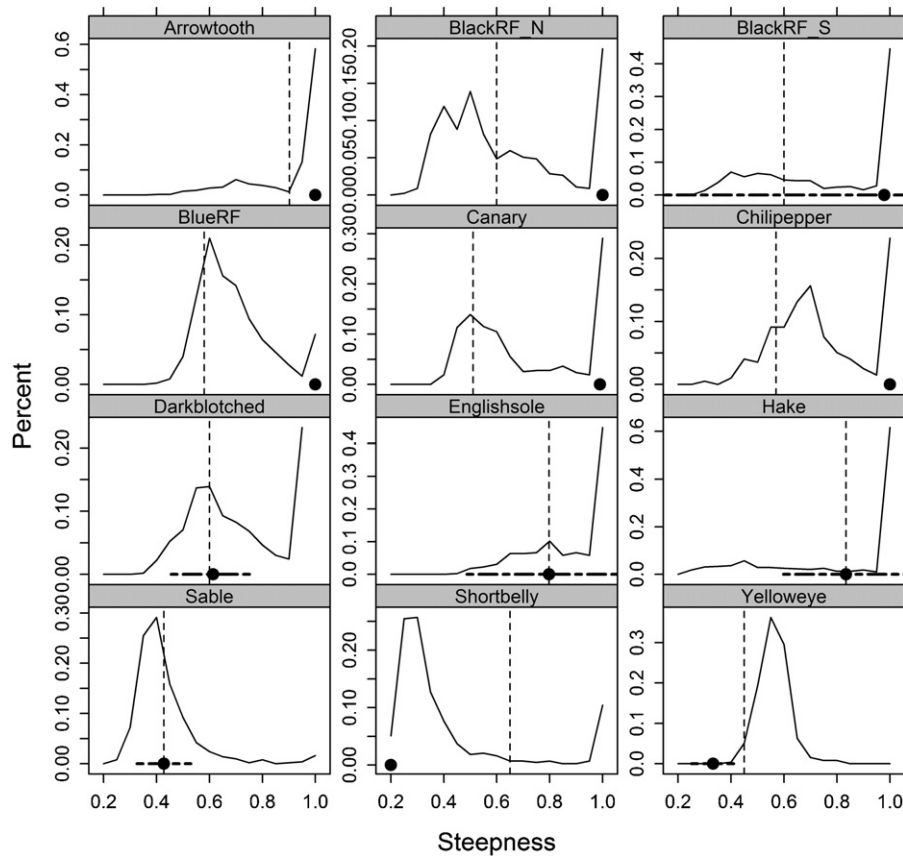


**Fig. 2.** Twelve examples of stock-recruitment estimates from groundfish stock assessments showing different levels of stock resilience. The lack of variability in recruitment at high biomass levels is a consequence of the stocks being modeled from an unexploited state and the lack of information (e.g., age or length composition data) about annual recruitment variability in the initial years. Therefore, the recruitment is the expectation of the stock-recruitment curve.



**Fig. 3.** The relationship between the true or assumed steepness and percent of bounded estimates of steepness. Twelve stocks (numbers in the circles) are as follows: (1) arrowtooth flounder, (2) black rockfish north, (3) black rockfish south, (4) blue rockfish, (5) canary rockfish, (6) chilipepper rockfish, (7) darkblotched rockfish, (8) english sole, (9) hake, (10) sablefish, (11) shortbelly rockfish, and (12) yelloweye rockfish.



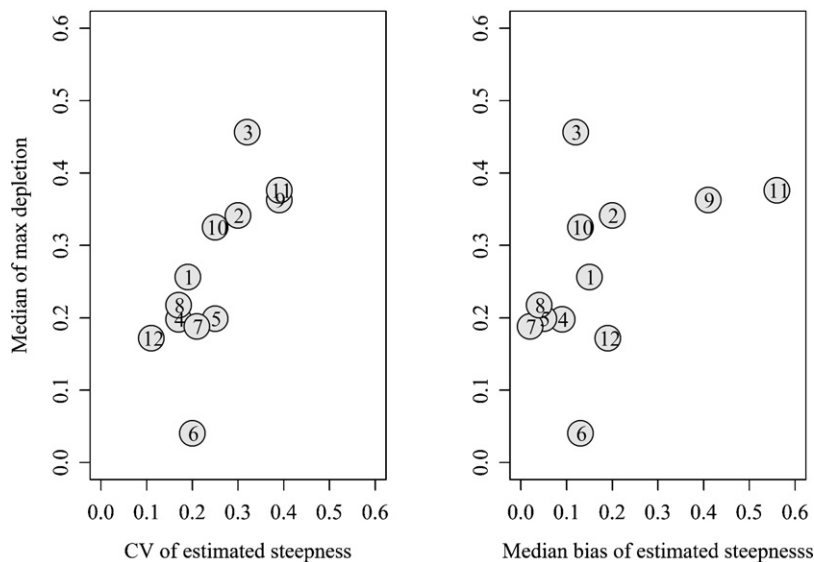


**Fig. 4.** Distribution of steepness estimated from the simulated data where converged simulations were selected based on the positive definite Hessian matrix. Dashed lines represent true or assumed values of steepness from original assessments. Dots represent estimated values of steepness from original data sets and bold dash lines represent confidence intervals around estimates.

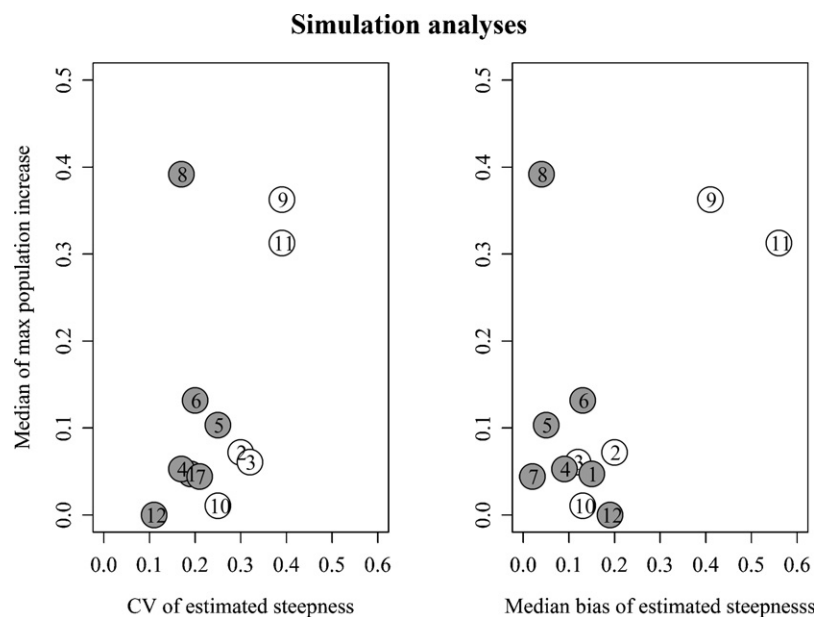
however, both one-way trip and two-way trips reached the same depletion level in Conn et al. study. In our study, trends of spawning biomass for simulated stocks generally followed those for true stocks and none of the stocks had identical exploitation patterns.

The approach used in this paper is not conditioned on the original stock assessment having been correctly specified. We use simulated data sets that were conditioned on the original stock assessment to incorporate the complexity seen in real

### Simulation analyses



**Fig. 5.** The relationship between precision and bias of estimated steepness and maximum depletion from the simulation analyses. Left panel represents the coefficient of variance of estimated steepness and right panel represents the median bias of estimates (absolute values). Numbers in the circles represent twelve stocks as follows: (1) arrowtooth flounder, (2) black rockfish north, (3) black rockfish south, (4) blue rockfish, (5) canary rockfish, (6) chilipepper rockfish, (7) darkblotched rockfish, (8) english sole, (9) hake, (10) sablefish, (11) shortbelly rockfish, and (12) yelloweye rockfish.



**Fig. 6.** The relationship between the ability to estimate steepness and maximum population increase from the simulation analyses. Left panel represents the coefficient of variance of estimated steepness and right panel represents the median bias of estimates (absolute values). Stocks depleted below 30% of the virgin level were shown in dark circles and white circles indicate that stocks have not been depleted below 30% of the virgin level. Numbers in the circles represent twelve stocks as follows: (1) arrowtooth flounder, (2) black rockfish north, (3) black rockfish south, (4) blue rockfish, (5) canary rockfish, (6) chilipepper rockfish, (7) darkblotched rockfish, (8) english sole, (9) hake, (10) sablefish, (11) shortbelly rockfish, and (12) yelloweye rockfish.

world applications, but the simulation process we used controlled any potential model misspecification so the simulated data will generally perform better than applications which often have overdispersed data and misspecified models. The results therefore overestimate the ability to estimate steepness in normal fishery situations. For example, many fisheries models estimate recruitment trends due to data conflict within the model, which when associated with a trend in spawning biomass can result in very high or low steepness estimates. The simulation methods used here would generate data that would fit the model without conflict and therefore without producing recruitment trends, which may explain why the average steepness estimates from simulations tended to be much more central than estimates from the original data. For example, yelloweye rockfish has symmetric distribution of estimates of  $h$  from the simulation study with reasonable variability (11% CV) but estimates of steepness from the original data laid outside the distribution of simulations. This kind of model misspecification is an important factor for the estimation of steepness; however, evaluation of model misspecification is beyond the scope of this study.

Our results suggest that steepness is reliably estimable inside the stock assessment model only when the model is correctly specified for relatively low productive stocks with good contrast in spawning stock biomass. If the stock has never been highly depleted or has always been highly depleted, then recruitment data is not available from low or moderate/high spawning stock size, which is needed to inform estimates of steepness. For stocks with good contrast in spawning stock size, population increase from those low levels may decrease the bias in estimated  $h$ . And for stocks without good contrast, fixed values or informative priors may be still necessary. Dorn (2002) proposed a hierarchical Bayesian approach and Mangel et al. (2010) incorporated biological and ecological characteristic of species to construct a prior distribution for steepness. Despite our simulations showing that with the appropriate data collected at the right population sizes  $h$  may be estimable, common sense should be applied to any resulting estimates. Given the inability to estimate  $h$  reliably for many

stocks, management strategies robust to uncertainty in  $h$  should be adopted.

This study is based on the simulation method described in Lee et al. (2011), which was applied to natural mortality using the same stock assessment examples. Because both studies used the same original stock assessment models we can directly compare the reliability of estimation of  $M$  and  $h$ . Our results suggest that  $M$  may be more estimable than  $h$ , as  $M$  was reliably estimated in 9/12 cases (median bias <9% and CV <15%) and  $h$  in only 4/12 (median bias <9% and CV <25%). The estimates appear to degrade if both  $M$  and  $h$  are estimated simultaneously, but estimates of  $M$  are still more reliable than those for  $h$ . This result is counter to the more common, but still rare, practice of estimating  $h$  and specifying  $M$  at fixed values in applied stock assessment models.

Both studies tested the null hypothesis that controversial parameters can be reliably estimated inside integrated population dynamics models using best case approaches. We used the best case approach because failure to reliably estimate the parameters would provide the strongest evidence against the null. If the parameters could not be estimated under those ideal conditions, those results would strongly argue against estimation under routine conditions. However, as pointed out by Francis (2012) the failure to reject the null hypothesis is not equivalent to proving the null, especially given that routine stock assessment often contain unknown levels of model misspecification (Piner et al., 2011) that are far from best case.

Natural mortality and spawner-recruit steepness are both routinely assumed in stock assessments as these parameters are thought unreliably estimated as part of the dynamic model. Our results suggest that in routine applications this may be reasonable, particularly for steepness (e.g., Zhu et al., 2012), but under certain conditions, with appropriate data and well specified models, good estimates of these parameters may be possible (e.g., Maunder and Wong, 2011; Maunder, 2012). We do not propose that this simulation-based approach substitute for other approaches to estimate parameters (e.g., life history theory or empirical relationships) and traditional diagnostics (e.g., residual analysis or

retrospective analysis). We suggest that comparison of several approaches is necessary (e.g., Maunders and Wong, 2011; Piner et al., 2011; Maunders, 2012). Given that the other methods for estimating these parameters independent of the population dynamics model also suffer from bias, results of these “best case” approaches may lead to advancements in model development and data collection.

## Acknowledgements

Our sincere thanks to the stock assessment scientists at the Northwest Fisheries Science Center and Stacey Miller for providing copies of their stock synthesis files, which greatly facilitated this study. We are also grateful to the Southwest Fisheries Science Center for supporting this work. We want to thank Vidar Weststad from the American Fisherman's Research Foundation and Rick Deriso from the Inter-American Tropical Tuna Commission, Jim Ianelli from the Alaska Science Center and Andre Punt from the University of Washington for comments on an early draft of the paper. We also thank Kenneth Rose, Chris Legault and one anonymous reviewer for comments on the manuscript.

## References

- Beverton, R.J.H., Holt, S.J., 1957/1993. On the Dynamics of Exploited Fish Populations. Chapman and Hall, London.
- Conn, P.B., Williams, E.H., Shertzer, K.W., 2010. When can we reliably estimate the productivity of fish stocks? *Can. J. Fish. Aquat. Sci.* 67, 511–523.
- 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, 280–300.
- Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunders, M.N., Nielsen, A., Sibert, J., 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optim. Methods Softw.* 27, 233–249.
- Francis, R.I.C.C., 1992. Use of risk analysis to assess fishery management strategies: a case study using orange roughy (*Hoplostethus atlanticus*) on the Chatham Rise New Zealand. *Can. J. Fish. Aquat. Sci.* 49, 922–930.
- Francis, R.I.C.C., 2012. The reliability of estimates of natural mortality from stock assessment models. *Fish. Res.* 119–120, 133–134.
- He, X., Mangel, M., MacCall, A., 2006. A prior for steepness in stock–recruitment relationships, based on an evolutionary persistence principle. *Fish. Bull.* 104, 428–433.
- Hilborn, R., Walters, C.J., 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty. Chapman and Hall, New York.
- Lee, H.H., Maunders, M.N., Piner, K.R., Methot, R.D., 2011. Estimating natural mortality within a fisheries stock assessment model: an evaluation using simulation analysis based on twelve stock assessments. *Fish. Res.* 109, 89–94.
- Mace, P.M., Doonan, I.J., 1988. A generalized bio-economic simulation model for fish population dynamics. *N. Z. Fish. Assess. Res., Doc.* 88/4.
- Magnusson, A., Hilborn, R., 2007. What makes fisheries data informative? *Fish. Res.* 8, 337–358.
- Mangel, M., Brodziak, J., DiNardo, G., 2010. Reproductive ecology and scientific inference of steepness: a fundamental metric of population dynamics and strategic fisheries management. *Fish. Res.* 11, 89–104.
- Maunders, M.N., 2012. Evaluating the stock–recruitment relationship and management reference points: application to summer flounder (*Paralichthys dentatus*) in the U.S. mid-Atlantic. *Fish. Res.* 125–126, 20–26.
- Maunders, M.N., Schnute, J.T., Ianelli, J., 2009. Computers in fisheries population dynamics. In: Megrey, B.A., Moksness, E. (Eds.), *Computers in Fisheries Research*. Springer, pp. 337–372.
- Maunders, M.N., Wong, R.A., 2011. Approaches for estimating natural mortality: application to summer flounder (*Paralichthys dentatus*) in the U.S. mid-Atlantic. *Fish. Res.* 111, 92–99.
- Methot, R.D., 2000. Technical Description of the Stock Synthesis Assessment Program. U.S. Dept. Commer, NOAA Tech. Memo. NMFS-NWFSC-43.
- Methot, R.D., 2006. User Manual for the Integrated Analysis Program Stock Synthesis 2 (SS2). P.54. Model Version 1.23a. Northwest Fisheries Science Center.
- Methot, R.D., 2009. Stock assessment: operational models in support of fisheries management. In: Beamish, R.J., Rothschild, B.J. (Eds.), *The Future of Fisheries Science in North America*. Fish and Fisheries Series, 31, pp. 137–165.
- Myers, R.A., Bowen, K.G., Barrowman, N.J., 1999. Maximum reproductive rate of fish at low population sizes. *Can. J. Fish. Aquat. Sci.* 56, 2404–2419.
- Myers, R.A., Barrowman, N.J., Hilborn, R., Kehler, D.G., 2002. Inferring Bayesian priors with limited direct data: application of risk analysis. *N. Am. J. Fish. Manage.* 22, 351–364.
- Pawitan, Y., 2001. In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford University Press, Oxford.
- Piner, K.R., Lee, H.H., Maunders, M.N., Methot, R.D., 2011. A simulation-based method to determine model misspecification: examples using natural mortality and population dynamics models. *Mar. Coast Fish.* 3, 336–343.
- Ricker, W.E., 1954. Stock and recruitment. *J. Fish. Res. Bd. Can.* 11, 559–623.
- Shepherd, J.G., Cushing, D.H., 1980. A mechanism for density-dependent survival of larval fish as the basis of a stock–recruitment relationship. *J. Conserv. Int. Explor. Mar.* 39, 160–167.
- Schirripa, M.J., Colbert, J.J., 2006. Interannual changes in sablefish (*Anoplopoma fimbria*) recruitment in relation to oceanographic conditions within the California Current System. *Fish. Oceanogr.* 15, 25–36.
- Zhu, J.F., Chen, Y., Dai, X.J., Harley, S.J., Hoyle, S.D., Maunders, M.N., Aires-da-Silva, A., 2012. Implications of uncertainty in the spawner–recruitment relationship for fisheries management: an illustration using bigeye tuna (*Thunnus obesus*) in the eastern Pacific Ocean. *Fish. Res.* 119–120, 89–93.