Fisheries Research 142 (2013) 37–46

Contents lists available at SciVerse ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

A comparison of stock assessment uncertainty estimates using maximum likelihood and Bayesian methods implemented with the same model framework

Ian J. Stewart[a,*], Allan C. Hicks[a,b], Ian G. Taylor[a], James T. Thorson[b], Chantell Wetzel[a,b], Sven Kupschus[c]

[a] Fishery Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, 2725 Montlake Blvd East, Seattle, WA 98112, USA
[b] University of Washington, School of Aquatic and Fishery Sciences, Seattle, WA, USA
[c] Centre for Environment, Fisheries & Aquaculture Science, Pakefield Rd., Lowestoft, Suffolk NR33 0HT, UK

ARTICLE INFO

Article history:
Received 5 March 2012
Received in revised form 2 July 2012
Accepted 7 July 2012

Keywords:
Maximum-likelihood
Bayesian
Uncertainty
Stock assessment

ABSTRACT

Many fisheries stock assessment models are implemented specifically for likelihood-based estimation or for Bayesian inference (via full integration of the joint posterior distributions), but not all have appropriate structure for both statistical approaches. Bias correction of recruitment deviations, in particular, must be adjusted to achieve consistency in each case. Fisheries management often uses the two types of results similarly, setting future catch quotas based on expected values or posterior medians depending on which is available given time constraints. Using two recent examples from the U.S. west coast, Pacific hake and sablefish, both implemented in Stock Synthesis, we find that likelihood-based estimates of key management quantities, such as spawning biomass, corresponded well with posterior modes, but tend to be lower (on an absolute scale) than posterior median values and that the asymptotic approximation for uncertainty intervals based on the Hessian matrix tends to overestimate the likelihood of smaller stock sizes and underestimate that of larger stock sizes. This pattern may be caused by a basic asymmetry in most fisheries data-sets: the necessity of a minimum stock size to have generated the observed catch/time-series, but little information regarding the plausibility among much larger stock sizes. Where only one type of inference is available, this asymmetry may be important for management decision-making. Even if management takes explicit account of uncertainty, in some cases adding a precautionary buffer that scales with the relative uncertainty in point estimates, the differences in catch advice may turn out to be important and the relative reductions non-linear.

Published by Elsevier B.V.

1. Introduction

Stock assessments serve to inform fisheries management around the world. Specific stock assessment models and methods vary by region, data-availability, and management needs. The current state-of-the-art method for stock assessment includes integrated assessment models (Fournier and Archibald, 1982; Maunder, 2003; Methot and Wetzel, this issue; Maunder and Punt, this issue). These integrated models include a process model, which approximates relevant population dynamics, and an observation model, which approximates sampling processes. The process–model component may incorporate ecological detail including age, length, location, and gender. Parameters in integrated models are estimated by computing the deviance between predicted and observed data. Data are predicted by using the process model and parameters to calculate stock size and other population–dynamics variables, and the observation model calculates the data that would likely be observed given the parameters and resulting population dynamics. Sampling processes that differ by age, length, location, or gender (e.g. fishery selectivity) are generally included in the integrated assessment model when data are available regarding age-, length-, spatial, or sex-composition. Integrated models may also incorporate information from previous studies regarding difficult-to-estimate processes such as recruitment compensation (Hilborn and Liermann, 1998) through priors or likelihood penalties.

Many fisheries stock assessment models are implemented specifically for likelihood-based estimation (frequently penalized likelihood, including penalties as well as fit to observations; this method is broadly referred to here as “MLE”; it is also sometimes called “MPD-estimates”, and/or “delta-approximation based” in the literature) or for Bayesian integration, but not all have appropriate structure for both modes of inference. Bias correction of recruitment deviations (Methot and Taylor, 2011), implementation of prior probability distributions and/or likelihood equations may all need to be adjusted to achieve consistency in each case (Methot, 2011). Within Stock Synthesis (SS; a generalized integrated analysis platform programmed in AD Model Builder, ADMB;

* Corresponding author. Tel.: +1 206 302 2447; fax: +1 206 860 6792.
E-mail address: Ian.Stewart@noaa.gov (I.J. Stewart).

0165-7836/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.fishres.2012.07.003

Fournier et al., 2012; Methot and Wetzel, this issue), annual recruitment deviations from the stock–recruitment function are assumed to be generated from a log-normal distribution. Estimates of recruitment must therefore be bias-adjusted to correct the central tendency from the median to the mean, such that the central-tendency of model estimates is consistent with the long-term projected central tendency given variable recruitments. The degree of bias-correction required is a function of the variability in these deviations (σ_R, in SS), and the information in the data about recruitment. Typically the MLE estimated variability of recruitments from early model years (prior to the period which is well-informed by the composition data) is low, and thus requires a smaller adjustment (Methot and Taylor, 2011). However, the Bayesian posterior distribution for recruitments integrates over the full range of plausible values, and therefore requires a full bias-correction for all model years.

Uncertainty is increasingly important for interpretation of stock assessment results, and as a direct input to fisheries management (Patterson et al., 2001). Likelihood-based estimates provide the probability that repeating the experiment many times will result in a confidence interval that will contain the true parameter value with specific probability, not a specific probability statement about that parameter. Confidence intervals are commonly based on asymptotic theory, a normal error structure, and result in a symmetric distribution. Alternatively, Bayesian posteriors define the probability distribution (nonparametric) for a parameter value, and direct probabilistic statements can be made from that distribution. The Bayesian interpretation of uncertainty is appealing when specific probabilities are of interest (Punt and Hilborn, 1997). Fisheries management often treats these results similarly, despite the philosophical differences inherent in the two types of inference (e.g., Hilborn and Mangel, 1997; Carlin and Louis, 2000; Burnham and Anderson, 2002; Gelman et al., 2004).

The Pacific Fishery Management Council (PMFC), as mandated by the Magnuson-Stevens Reauthorization Act of 2006, reduces the target fishery catch from the estimate of the overfishing level (OFL; a level of harvest that if exceeded would constitute overfishing), This reduction is a function of the degree of scientific uncertainty around current stock estimates and a level of risk deemed acceptable for that stock (Pacific Fishery Management Council; www.pcouncil.org). The scientific uncertainty is based on either: (1) the retrospective level of uncertainty in spawning biomass (SB) derived via meta-analysis of many recent stock assessments (Ralston et al., 2011), or (2) the degree of uncertainty estimated from the current stock assessment result (by whatever method is available), whichever is larger. The use of the meta-analysis adjusts accordingly for cases where stock assessment uncertainty is underestimated; however, where greater uncertainty is estimated in the assessment it can still be included. This uncertainty is then used to create a probability distribution for the OFL, and the PMFC then selects a "P^*" (by law is <50%), which represents the estimated probability that the specified catch will be in excess of the OFL. Thus, the probability distribution for the OFL, particularly the lower quantiles, is very important in calculating the recommended catch target.

Due to the technical, time, and reporting constraints stock assessment analyses may include either maximum likelihood, Bayesian, or both types of results. In the case of the PFMC, future catch targets have been based on both maximum likelihood estimates, applying a reduction in catch calculated from the uncertainty in log-SB (Ralston et al., 2011) or directly from the posterior medians for the OFL depending on which were available. Using two recent examples from the U.S. west coast, Pacific hake (Stewart et al., 2011a) and sablefish (Stewart et al., 2011b), both of which were implemented in SS, we compare the maximum likelihood estimates and Bayesian posterior distributions of key management quantities in an attempt to better understand how they differ and to what degree catch advice may vary depending on which method is applied to generate catch targets.

The primary objective of this analysis is to determine whether systematic differences occur in estimated quantities of management interest from integrated stock assessment based on maximum likelihood and Bayesian inference, and to identify which quantities are most susceptible, and how the probability distributions for the underlying model parameters lead to these differences. To illustrate the general difference between symmetric and asymmetric uncertainty, a simple hypothetical example is provided.

2. Methods

2.1. Statistical example

For the purposes of illustration, we compare two statistical distributions, each with identical modes. The first is log-normally distributed with median, m, and standard deviation in log space, σ^2; the mode of this distribution is: m/e^{σ^2}. The second is normally distributed, with the variance approximated by the negative inverse of the second derivative: $m^2\sigma^2/e^{2\sigma^2}$. Although this simple illustration is statistically unsurprising, it generally resembles the distributions observed for actual stock assessment quantities investigated below. Of particular importance is the degree and pattern of divergence in the percentiles of the lower tails, those values most relevant to management decision making. The differences between these distributions are presented for two modal values.

2.2. Stock assessments

We used the most recent stock assessments for sablefish and Pacific hake incorporated into management by the PFMC (Stewart et al., 2011a,b). Briefly, both are integrated stock assessments with similar data sources (Table 1). Fishery-dependent data consists of a time-series of catches assumed to be known without error and biological sampling (individual weights, lengths and ages) of those catches available for only a recent subset of years. Fishery-independent (survey) data consists of one or more indices of relative abundance covering a subset of recent years, as well as biological sampling.

These assessment models use SS as a population dynamics model that projects forward from initial conditions estimating all model parameters simultaneously. Parameters for both stock assessments include: initial age-structure deviations, recruitment deviations about a stock–recruit function for each model year, fishery and survey selectivity parameters, parameters describing somatic growth, and others (Tables 2 and 3).

The two assessments provide an interesting contrast: Pacific hake have a relatively short exploitation history and weak survey information, while sablefish have a much longer exploitation history and somewhat more informative survey information. In both cases, estimated uncertainty is large, due to relatively poor information regarding the absolute scale of the populations.

We first compare results from both types of inference for key parameters and management related quantities. For Pacific hake, the posterior distribution for the OFL was used by the PFMC for setting future catch.[1] For sablefish, the PMFC's default method to calculate a percent reduction from the MLE was used for setting

[1] The Pacific Fishery Management Council ultimately based management advice for 2011 on the model-averaged results from two assessment models reported together in Stewart et al., 2011a,b. For simplicity, in this paper we focus only on the results of the Stock Synthesis model.

Table 1
Primary data sources for each stock assessment.

Data source	Pacific hake	Sablefish
Fishery-dependent		
Fishery catch	1966–2010	1900–2010
Fishery biological	1975–2010	1978–1981, 1983, 1985–2010
Fishery-independent		
Abundance indices	1995, 1998, 2001, 2003, 2005, 2007, 2009	1980, 1983, 1986, 1989, 1992, 1995, 1997–2010
Survey biological	1995, 1998, 2001, 2003, 2005, 2007, 2009	1980, 1983, 1986, 1989, 1992, 1995, 1997–2010

Table 2
Summary of assessment model parameters for sablefish.

Parameter	Number estimated	Bounds (low, high)	Prior (mean, SD)
Stock, recruitment, and productivity			
Ln(R_0)	1	(8,12)	Uniform
Steepness (h)	–	NA	Fixed at 0.6
Recruitment SD (σ_r)	–	NA	Iterated to 1.15
Initial age deviations (ages 1–49 at age-0)	49	(−4,4)	Normal (0,σ_r)
Time-series recruitment deviations (1900–2010)	111	(−4,4)	Normal (0,σ_r)
Natural mortality (M, female)	1	(0.01,0.11)	Log(Normal) (−2.1791, 0.3384)
Natural mortality (M, male)	1	(0.01,0.11)	Log(Normal) (−2.0565,0.3375)
Survey catchability, selectivity and variability			
Trawl surveys			
Ln(catchability[Q])	5	Variable	Analytic solutions orLn(Q) uniform
Survey selectivity (double-normal)	13	Variable	Uniform
Extra additive SD for survey index	4	(0.001,1.3)	Uniform
Selectivity, retention, and discard mortality			
Fishery selectivity (cubic spline)	28	Variable	Uniform
Fishery retention (logistic)	7	Variable	Uniform
Fishery discard mortality	–	NA	Fixed at 20%, or 50% by fleet
Fishery size at first survival	–	NA	Fixed at 28 cm
Individual growth and maturity			
Females			
Length-weight coefficient (a)	–	NA	Fixed at 0.00000345
Length-weight coefficient (b)	–	NA	Fixed at 3.267
Length at 50% maturity	–	NA	Fixed at 58 cm
Logistic slope of maturity	–	NA	Fixed at −0.13
Length at age 0.5	1	(22,30)	Uniform
Length at age 30	1	(60,70)	Uniform
von Bertalanffy K	1	(0.15,0.35)	Uniform
CV of length at age 1	1	(0.03,0.15)	Uniform
CV of length at age 30	1	(0.03,0.15)	Uniform
Males			
Length-weight coefficient (a)	–	NA	Fixed at 0.00000367
Length-weight coefficient (b)	–	NA	Fixed at 3.251
Length at age 1 offset to females	–	NA	Fixed at 0.0
Length at age 30	1	(50,60)	Uniform
von Bertalanffy K	1	(0.2,0.45)	Uniform
CV of length at age 0.5	1	NA	Fixed at 0.0
CV of length at age 30	1	(0.03,0.15)	Uniform

Table 3
Summary of assessment model parameters for Pacific hake.

Parameter	Number estimated	Bounds (low, high)	Prior (Mean, SD)
Stock, recruitment, and productivity			
Ln(R_0)	1	(13,18)	Uniform
Steepness (h)	1	(0.2,1.0)	~Beta(0.777,0.113)
Recruitment SD (σ_r)	–	NA	1.30
Initial age deviations (ages 1–19 at age-0)	19	(−6, 6)	~Ln(N(0, σ_r))
Time-series recruitment deviations (1966–2010)	46	(−4,4)	Normal (0,σ_r)
Natural mortality (M)	1	(0.05,0.4)	~Ln(N(0.2,0.1))
Survey catchability, selectivity and variability			
Acoustic survey			
Ln(catchability[Q])	1	NA	Analytic solutions
Age-based selectivity (non-parametric; ages 3–5)	3	(−5,9)	Uniform in scaled logistic space
Extra additive SD for survey index	4	(0.001,1.0)	Uniform
Selectivity, retention, and discard mortality			
Age-based selectivity (non-parametric; ages 2–5)	4	(−5,9)	Uniform in scaled logistic space
Individual growth and maturity			
Empirically derived weight and maturity matrix by age and year.			

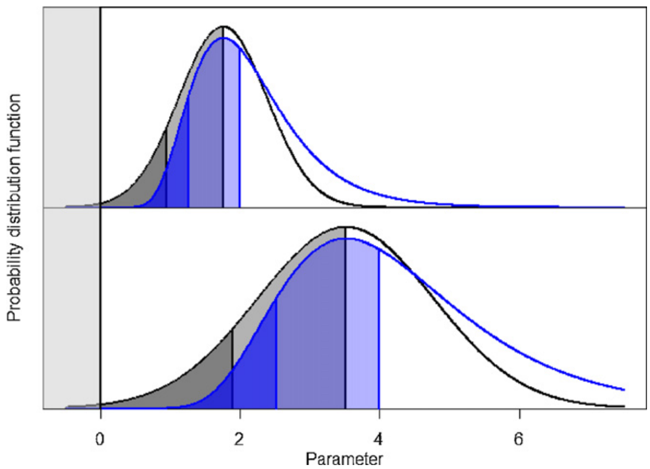

Fig. 1. Example lognormal (blue) and normal distributions (black). The standard deviations in log space are held constant, but the central tendencies are smaller (upper panel) or larger (lower panel). Shaded areas represent the lower 50% (light) and 10% (dark) of each distribution. Values less than zero are shaded on the left of the plot in light gray. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

future catches[2] (using a pre-specified log-SB sigma, 0.36; Ralston et al., 2011). We compare P^*-based catch targets from MLE estimates with Bayesian posterior distributions for the OFL.

Standard convergence diagnostics were applied to Bayesian results. These included running preliminary pilot Markov-Chain Monte-Carlo (MCMC) simulations to estimate the duration and thinning interval needed to generate stationary posterior parameter distributions, and multiple intermediate length MCMC simulations to compare parameter percentiles. Geweke tests for stationarity (Geweke, 1992; Plummer et al., 2006), Heidelberger and Welch (Heidelberger and Welch, 1983; Plummer et al., 2006) tests for sufficient burn-in, monitoring of maximum within-parameter autocorrelation (R Development Core Team, 2011) and effective sample size corrected for autocorrelation (Plummer et al., 2006) were all used to ensure sufficiently independent samples had been generated.

3. Results

3.1. Statistical example

The example lognormal distribution and normal approximation are shown in Fig. 1. Several obvious statistical patterns are visible: (1) the mode of the normal approximation is always smaller than the median of the lognormal distribution, (2) the range of normal approximation was always narrower than the lognormal distribution, and (3) the percentiles of the normal distribution were always more extreme in the lower tail than for the lognormal regardless of the mode. Importantly, the relative difference between the two percentiles increases as the percentile gets smaller; i.e., the 10th percentile of the normal approximation shows a bigger relative reduction from the lognormal 10th percentile than occurs at the 50th percentile (the median). In some cases, the normal approximation also included density for values less than zero (impossible for the lognormal). This artifact of the approximation is especially relevant here, because such values would be considered biologically implausible for most stock assessment parameters and derived quantities.

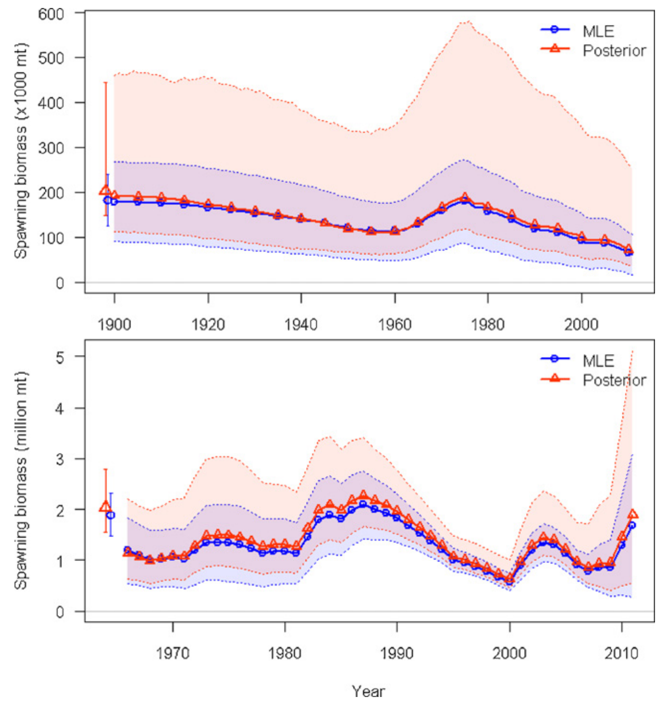

Fig. 2. Time-series of SB (with ~95% intervals) from the sablefish (upper panel) and Pacific hake stock (lower panel) stock assessments.

3.2. Stock assessments

Posterior distributions for the Pacific hake assessment were integrated via a 10,000,000 iteration MCMC chain, removing the initial chain value, and saving every 5000th iteration, resulting

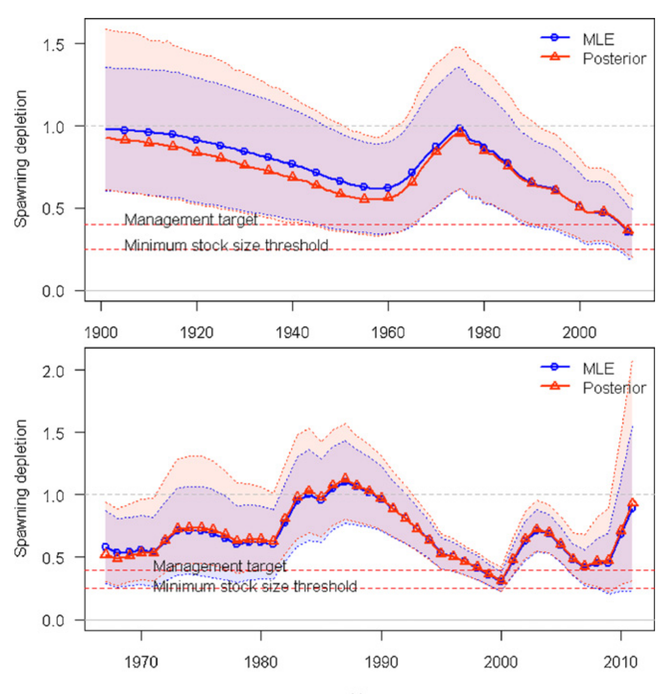

Fig. 3. Time-series of relative spawning depletion (with ~95% intervals) from the sablefish (upper panel) and Pacific hake (lower panel) stock assessments.

[2] The Pacific Fishery Management Council operates on a two-year management cycle for most groundfish species (excluding Pacific hake). For simplicity, in this paper we focus only on the distribution of the first year of projected OFL in order to provide a more analogous comparison with Pacific hake.

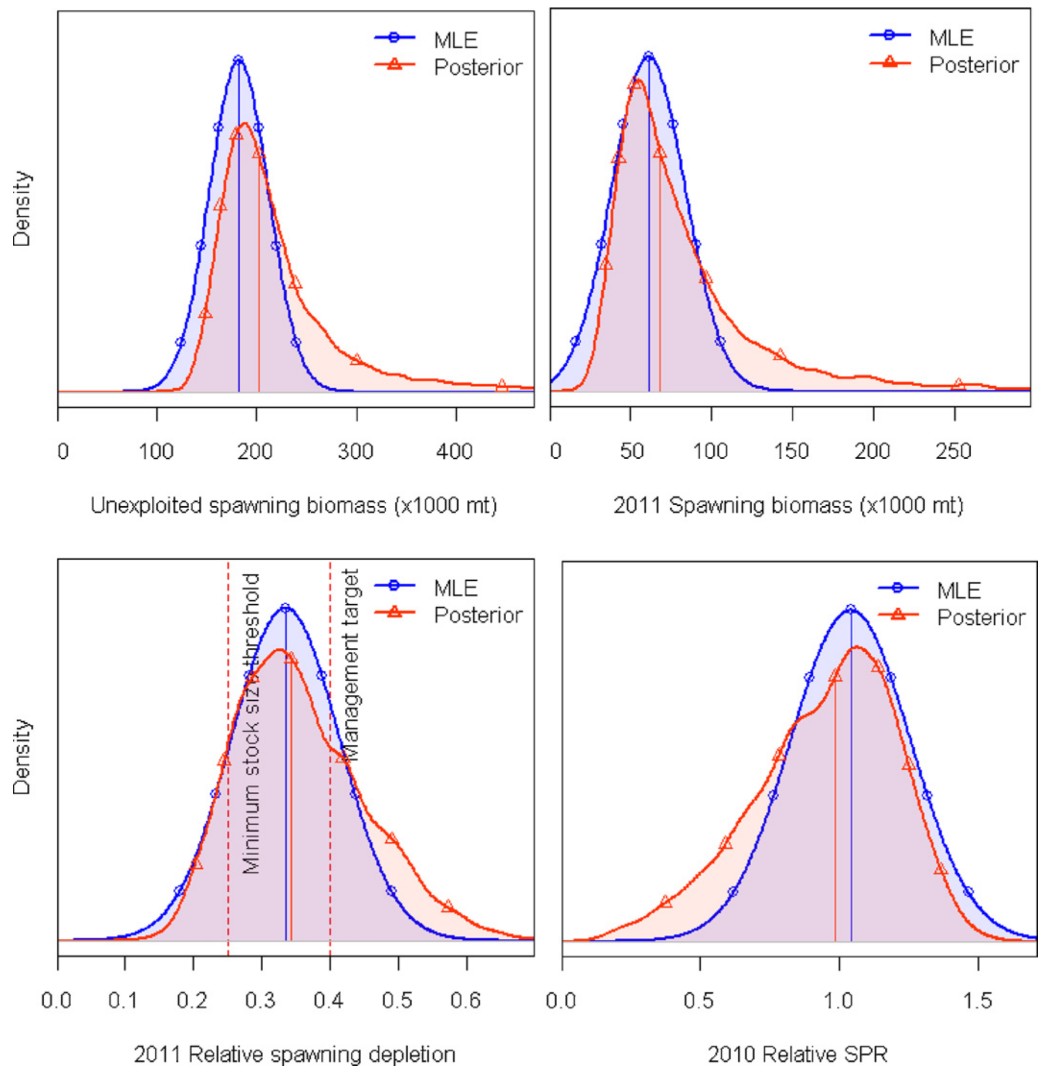

Fig. 4. Marginal distributions for quantities of management interest for sablefish. Vertical lines represent the maximum likelihood estimate and the posterior median.

in a minimum effective sample size over all estimated parameters of 894, and a maximum autocorrelation (at lag-1) of 0.11 (the estimated process error SD for the survey index of abundance, which was highly right-skewed; see Stewart et al. (2011a,b) for more description of this parameterization). Nine percent of model parameters had a Geweke statistic of absolute value >1.96 (~5% would be expected by random chance), and all passed the Heidelberger and Welch test for sufficient burn-in. Due to the skewness of the process error SD (an additive constant representing observation error not included in input estimates), and the symmetric approximation to the variance–covariance matrix used for the MCMC jump function, it is expected that convergence to an uncorrelated chain would be slowest for this parameter. These diagnostics uncovered no evidence of non-convergence and indicated that posterior distributions were unlikely to be appreciably changed by extended MCMC simulations and that the percentiles of the distributions for comparison in the analysis were likely to be reliably estimated.

Posterior distributions for the sablefish assessment were integrated via a 25,000,000 draw MCMC chain, removing the initial chain value, and saving every 50,000th iteration, resulting in a minimum effective sample size over all estimated parameters of 228 (an initial age-structure deviation, the next sparsest parameter, one of the selectivity at age values, had an effective sample size of 299), and a maximum autocorrelation (at lag-1) of 0.15 (an early recruitment deviation). Just under 5% of model parameters had a Geweke statistic of absolute value >1.96, and all passed the Heidelberger and Welch test for sufficient burn-in. The parameters that were slowest to converge were very poorly informed recruitment or initial age-structure deviations that were uncorrelated with key parameters and management-related quantities. Similar to the Pacific hake example, no evidence of non-convergence was identified.

SB and relative depletion (the ratio of the female SB in any year to the average unexploited equilibrium female spawning biomass [SB_0]) MLE time-series estimates were lower (on an absolute scale) than posterior median values (Figs. 2 and 3; Tables 4 and 5); however, the MLE and the mode of the Bayesian posterior for these quantities were generally similar (Figs. 4 and 5). This was due to the degree of right-skewness in the posterior distributions. Uncertainty, measured either by the breadth of the quartiles or the 95% intervals was much greater for the posterior distributions for all quantities (Tables 4 and 5). Further, the asymptotic approximation tends to overestimate the likelihood of smaller 2011 stock sizes and underestimate that of larger stock sizes (Figs. 4 and 5). The standard deviation of log-SB was 0.36 for sablefish (exactly equal to the PMFC's default level) and 0.41 for Pacific hake. Fishing intensity, here measured by relative spawning potential ratio (SPR), was

Table 4
Comparison of point estimates and distributions for sablefish.

Quantity	2.5th percentile	25th percentile	Median or MLE	75th percentile	97.5th percentile
Unexploited average female SB (thousands mt)					
MLE	124	162	182	202	240
Posterior	149	179	202	238	445
2011 Female SB (thousands mt)					
MLE	16.4	45.6	61.0	76.3	105
Posterior	34.6	52.2	68.3	96.0	252
2011 Relative spawning depletion					
MLE	18%	28%	33%	39%	49%
Posterior	20%	29%	34%	42%	57%
2010 Relative SPR					
MLE	62%	89%	104%	119%	146%
Posterior	37%	78%	99%	114%	137%
Female natural mortality					
MLE	0.068	0.076	0.080	0.084	0.092
Posterior	0.068	0.075	0.080	0.085	0.094
Male natural mortality					
MLE	0.056	0.062	0.065	0.068	0.074
Posterior	0.055	0.061	0.065	0.068	0.075
2008 Recruitment deviation					
MLE	0.73	0.89	0.98	1.06	1.23
Posterior	0.77	0.94	1.03	1.13	1.31
Log(Unexploited average age-0 recruitment)					
MLE	9.58	9.86	10.01	10.16	10.44
Posterior	9.70	9.95	10.11	10.32	11.08

found to be left-skewed (Figs. 4 and 5) with the cumulative posterior distribution much greater at lower values than that based on maximum likelihood.

The time-series of relative depletion based on MLE and Bayesian inference were relatively more similar than estimates absolute stock size. Given the broad differences in the SB series', this is explained by the correlations among parameters. For sablefish, the SB_0 estimate was 90% correlated (based on the MLEs) with the 2011 SB estimate, for Pacific hake this value was 51%. By comparison, SB_0 was only 73% correlated with 2011 depletion for sablefish and 28% for Pacific hake. This pattern illustrates the greater uncertainty in absolute scale than in relative trend over time.

The posterior distributions for key model parameters were generally quite symmetric and did not differ considerably in central tendency from MLEs (Figs. 6 and 7). The exceptions to this were the additional variance components estimated for fishery-independent surveys which were highly right-skewed (e.g., Fig. 8), although largely uncorrelated with management related quantities.

Table 5
Comparison of point estimates and distributions for Pacific hake.

Quantity	2.5th percentile	25th percentile	Median or MLE	75th percentile	97.5th percentile
Unexploited average female SB (millions mt)					
MLE	1.467	1.746	1.893	2.040	2.319
Posterior	1.549	1.853	2.034	2.242	2.756
2011 Female SB (million mt)					
MLE	0.271	1.199	1.686	2.172	3.100
Posterior	0.631	1.334	1.874	2.646	5.140
2011 Relative spawning depletion					
MLE	22%	66%	89%	112%	156%
Posterior	35%	68%	91%	123%	203%
2010 Relative SPR					
MLE	36%	58%	70%	81%	103%
Posterior	30%	51%	64%	77%	105%
Steepness					
MLE	0.641	0.779	0.851	0.923	1.061
Posterior	0.570	0.730	0.810	0.875	0.958
Natural mortality					
MLE	0.176	0.201	0.214	0.227	0.252
Posterior	0.185	0.210	0.223	0.237	0.267
2008 Recruitment deviation					
MLE	1.782	2.330	2.617	2.904	3.453
Posterior	1.750	2.430	2.743	3.055	3.610
Log(Unexploited average age-0 recruitment)					
MLE	14.17	14.47	14.63	14.78	15.08
Posterior	14.31	14.60	14.77	14.95	15.36

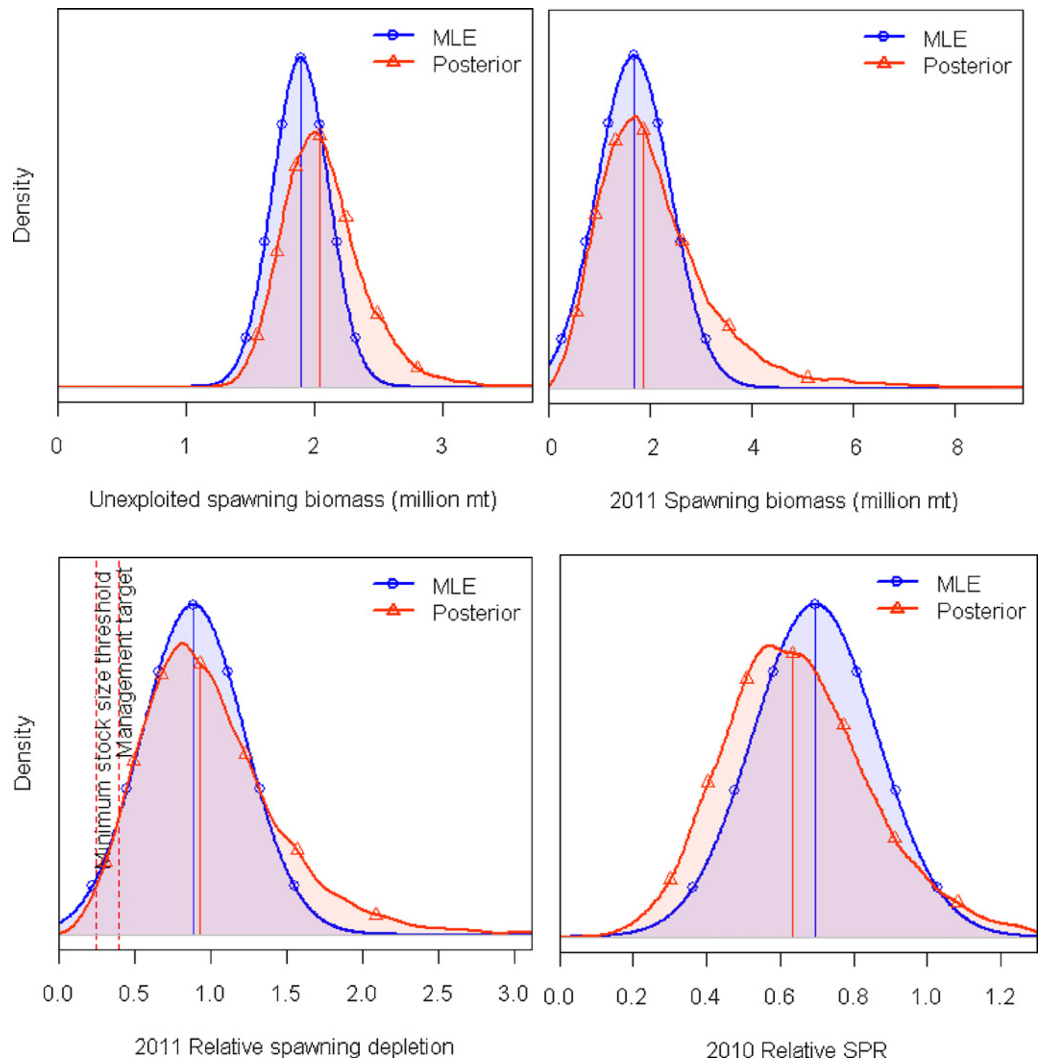

Fig. 5. Marginal distributions for quantities of management interest for Pacific hake. Vertical lines represent the maximum likelihood estimate and the posterior median respectively.

Future OFLs were more closely related to absolute stock size than relative depletion estimates. In general, wider Bayesian intervals (greater uncertainty), could offset the higher medians relative to MLE values. However, the broader uncertainty in the posterior distributions occurred largely in the upper tail of the distribution, so the use of a percentile below the median for the probability of overfishing did not offset the difference in point estimates. Differences in direct catch advice from the two types of inference were substantial across a range of potential P^* values (Table 6). The PMFC's default method, using uncertainty in log-SB to calculate a percent reduction from the OFL (Ralston et al., 2011) performed more like the Bayesian inference, however the method was still based on the MLE and therefore consistently underestimated the OFL for a given P^*. The exception to this pattern was observed at the smallest P^* value considered (0.1), where the method's relative performance was not consistent among the two examples. This was likely due to differences between the lower tail of the posterior and the strict log-normal assumption in the default method. In both examples,

Table 6
Example of catch target calculations for sablefish and Pacific hake as a function of the distributions for OFL and SB (mt), the probability of overfishing ($^*P$), and the perceived scientific uncertainty. Values in parentheses indicate percent of the posterior median for the OFL.

Probability of overfishing (P^*)	10%	25%	35%	45%	OFL (50%)
Sablefish					
MLE	3513 (49%)	4930 (69%)	5604 (78%)	6208 (86%)	6502 (90%)
MLE Log-normal from SB	4096	5104	5657	6209	6502 (90%)
Ralston et al. (2011)	(57%)	(71%)	(79%)	(86%)	
Posterior	4626 (64%)	5570 (78%)	6203 (86%)	6819 (95%)	7188 (100%)
Pacific hake					
MLE	290,229 (34%)	493,342 (58%)	589,887 (70%)	676,410 (80%)	718,502 (85%)
MLE Log-normal from SB	424,635	544,624	612,882	681,858	718,502
Ralston et al. (2011)	(50%)	(65%)	(73%)	(81%)	(85%)
Posterior	394,254 (47%)	580,041 (69%)	683,036 (81%)	791,223 (94%)	844,069 (100%)

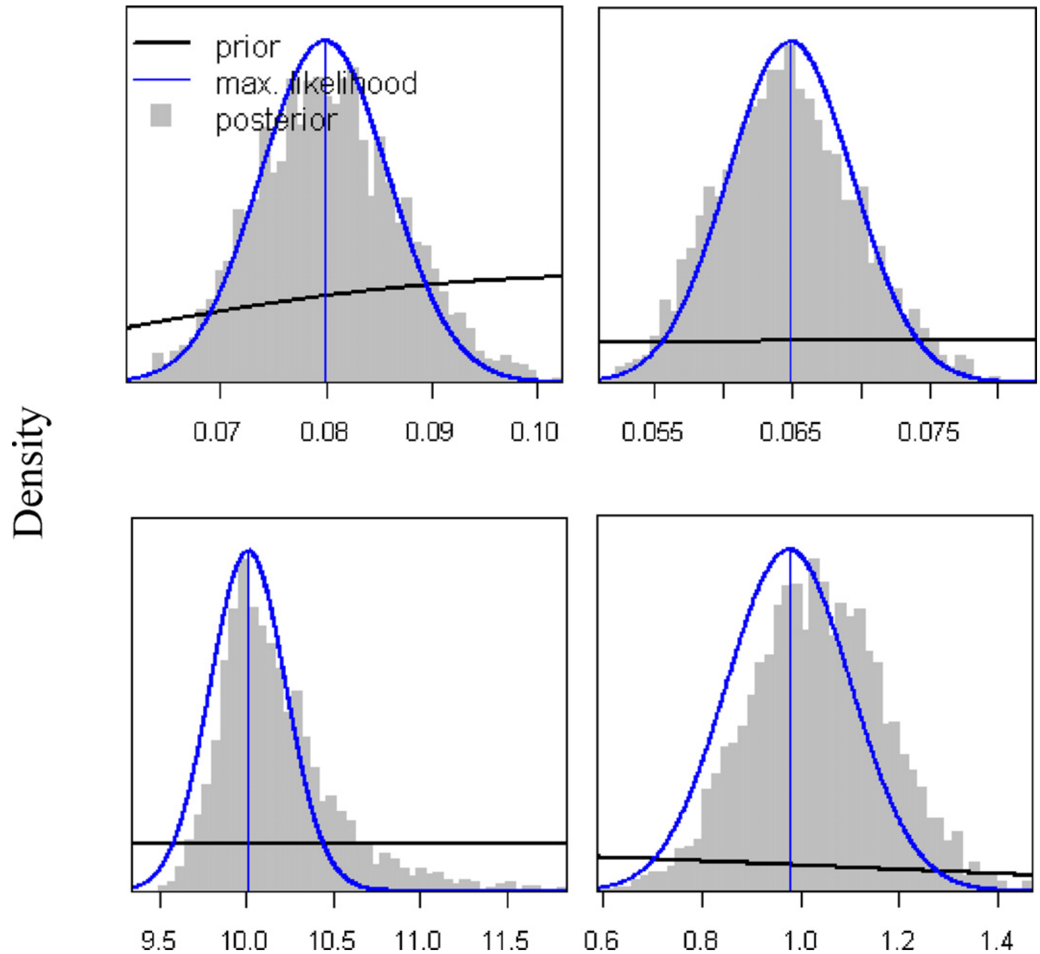

Fig. 6. Marginal distributions for select parameters for sablefish, including: female natural mortality (upper left), male natural mortality (upper right), log(unexploited average age-0 recruitment), and the 2008 log(recruitment deviation). Vertical lines represent the maximum likelihood estimates.

the likelihood-based approach was the most pessimistic regarding future OFL catches.

4. Discussion

Exploiting fisheries resources sustainably has been the aim of fisheries management for a long time, but recent advances in stock assessment methodologies have paved the way for a more formal inclusion of the uncertainty component in fisheries management. This paper compares two methods for estimating uncertainty that are frequently available using SS and other modeling applications built with ADMB. As illustrated with the theoretical example and observed in the actual stock assessment results, the upper portion of a distribution with a long right tail is systematically underestimated with a symmetric MLE approximation. This is consistent with the coverage of Bayesian intervals recently documented via simulation analysis (Magnusson et al., 2012). The spread of the lower portion of the distribution is systematically overestimated with the MLE approximation, with the degree depending on the percentile evaluated. Any truncation near zero will make this pattern even more pronounced. In fact this issue is clearly identified in the AD Model Builder manual (AD Model Builder. ADMB Project, 2009) yet seems to be frequently lost in the rush to produce and use assessment results. Fisheries stock assessment results that are associated with minimum values, such as the probability that SB is below some threshold, will likely be biased due to these differences. It is likely that the probability that a value is less than some critical value is overestimated when using MLEs, since the assessment parameters frequently have little density below some minimum value that is implausible given the catch history of the stock. At the extreme, the MLE approximation may even include an appreciable amount of probability less than zero, which is impossible.

For the stocks examined in this paper, the right skew of management measures based on MLE estimates of SB suggest that the setting of the OFL would be precautionary rather than risk-neutral relative to the full posterior distribution. This is remains true for the use of a catch reduction based log-SB (Ralston et al., 2011), although that method could be improved by correcting for the mode vs. median of the assumed distribution. Given the correlation between fishing intensity and SB, the left skewed distribution for fishing intensity is intuitive and the probability that fishing intensity is below some threshold is likely underestimated. For management systems that manage on fishing mortality thresholds not estimated within the stock assessment model (e.g., International Council for the Exploration of the Sea [ICES]), this effect could result in additional precaution relative to target and limit reference points.

The patterns observed in this analysis may be related to a basic asymmetry in the information content of most fisheries data-sets: there must be a minimum stock size which was capable of having generated the observed catch time-series, but there may be little information in the data to rule out much larger stock sizes. Because uncertainty about MLEs is generally assumed to be symmetric, a right-skewed distribution will always be underestimated in the upper tail, and overestimated in the lower tail – precisely where management advice is most sensitive. In the converse, posterior distributions may contain appreciable probability over very large

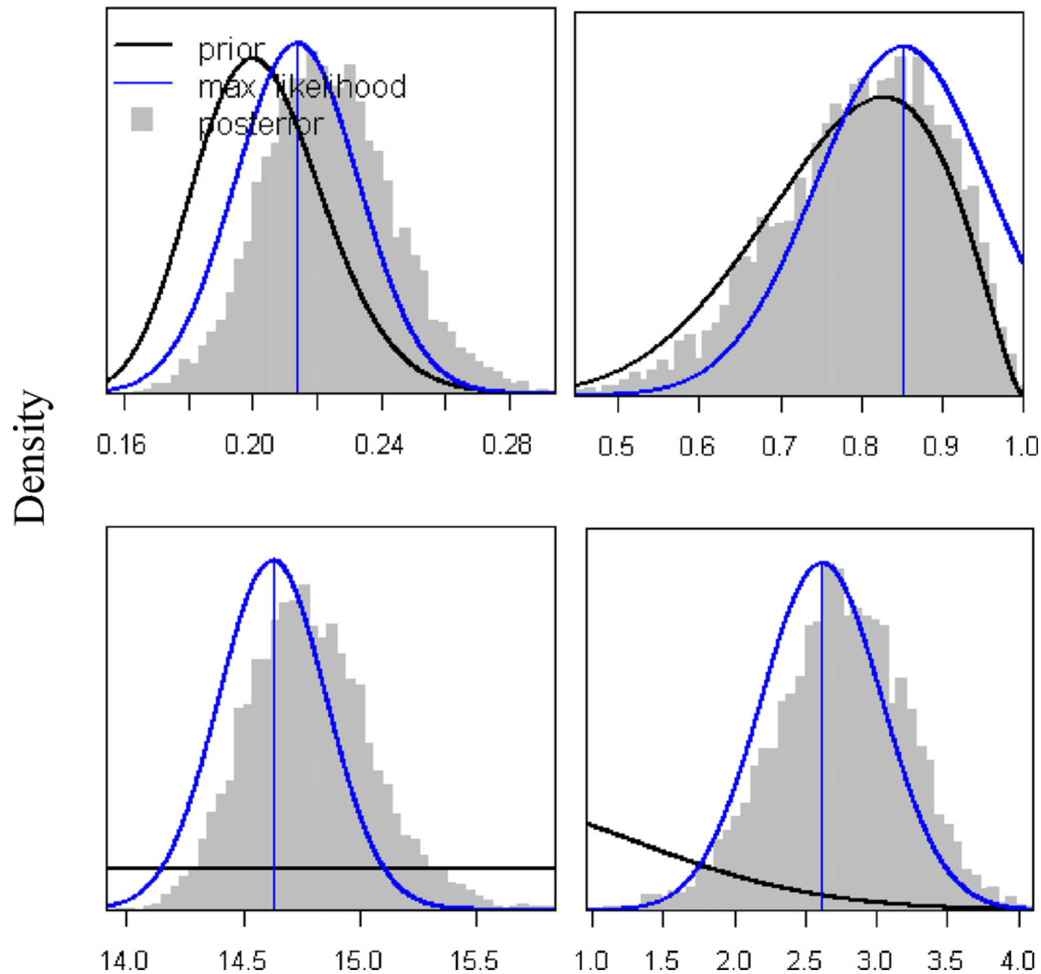

Fig. 7. Marginal distributions for select parameters for Pacific hake, including: natural mortality (upper left), steepness, log(unexploited average age-0 recruitment), and the 2008 log(recruitment deviation). Vertical lines represent the maximum likelihood estimates.

biomass levels. Whether this poses an issue to their direct use in calculating probability distributions for the application of P^*, and how best to incorporate additional information about 'plausible' stock sizes, remains an open question.

The two methods compared here differ in philosophy, and although we do not wish to feed the ongoing debate between frequentists and Bayesians (see Dennis, 1996; Carlin and Louis, 2000, Chapter 1) there is a fundamental difference between these two approaches in relation to how the PFMC sets catch levels.

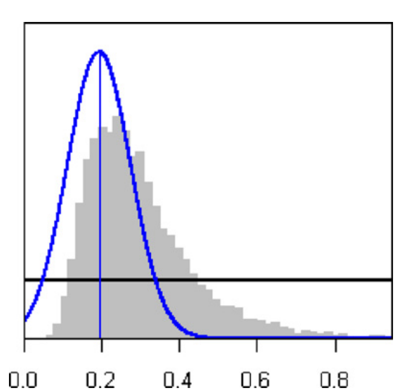

Fig. 8. Marginal distributions for the additive extra standard deviation component to observation error for the acoustic survey for Pacific hake. Vertical lines represent the maximum likelihood estimates.

Bayesian results provide the ability to make probabilistic statements about a parameter of interest, which is what US fishery managers attempt to do when applying P^* and buffering a catch level. The P^* approach requires calculation of the probability that the chosen catch level is less than or equal to the OFL. Although MLEs, and proxies derived from them, may serve as an interim solution, interpreting an interval which will contain the true parameter as a probability is awkward compared to directly specifying the probability distribution with a posterior distribution.

Several alternate methods are available for generating asymmetric distributions about point estimates. Log-transformation of model parameters such as spawning biomass could be used to approximate the skewed posterior distributions observed in these examples. The performance of this and other potential transformations could be explored via simulation experiments. Implementation would require additional coding, but probably not substantially increased computational demands. Likelihood profiles are also relatively straightforward for directly estimated parameters, however they require a penalty function for derived quantities (including most outputs of management interest), which can be technically problematic. Such penalty functions are available in the underlying ADMB language (Fournier et al., 2012) and could be easily implemented. Bootstrapping (Efron and Tibshirani, 1994) is another method that which can provide asymmetric distributions for derived parameters. It is a time-consuming process, since "new" data must be simulated (either via independent code, or within SS), and the estimation conducted for each realization.

The analysis described here represents a comparison of uncertainty within the specified stock assessment model, but does not include uncertainty due to model misspecification, functional forms, fixed parameters, and choice or weighting of data. These types of uncertainty are often investigated through sensitivity analysis, which generally entails perturbing model input data, structure or fixed parameter values and comparing the results. Although useful for generating decision tables with alternate 'states of nature', sensitivity results are rarely amenable to the calculation of quantitative probabilities and therefore cannot be easily incorporated into P^*-type calculations. No single uncertainty distribution from a stock assessment model should be considered 'correct', since they are merely approximations of reality. The utility of these approximations must be interpreted with a careful inspection of nonrandom patterns in residuals to model fit, violation of error distribution assumptions, and other sources of unidentified bias present in data sources and the estimates of model parameters. These issues can be corrected, through re-parameterizing model structure, applying alternate error distributions and similar approaches, but only if identified. One such diagnostic is the relative corroboration of the posterior mode and maximum likelihood estimate. Where these differ considerably, model behavior should be investigated further before strict quantitative interpretation is made of either the point estimates or the uncertainty in those estimates.

Subjective decisions during all stages of data processing and model development are frequently discussed, but their effects are rarely understood in nearly all stock assessment analyses. Therefore it would be reasonable to think of the amount of uncertainty that is measured within an assessment model, for a single realization of the observed data and process of analysis as a minimum estimate. Despite this fact, fisheries managers must still make decisions and set catch levels. We suggest that the information provided in this analysis may be useful to spark further research (both meta-analysis and simulation studies) to investigate better application of currently available estimators and perhaps derive new methods to improve the performance of P^*-style approaches.

In the meantime, where only one type of inference is available for fisheries managers, the asymmetry apparent in these two examples may be important for management decision-making and may lead to lower projected catch levels when based on MLE results alone. Where management takes explicit account of uncertainty, in some cases adding a precautionary buffer that scales with the relative uncertainty in point estimates, the differences in catch advice may turn out to be important and the reductions from point estimates non-linear.

Acknowledgements

Many people contributed to the data, analyses and review of the stock assessments used for this comparison; these contributors are acknowledged directly in the stock assessment documents. We particularly recognize the contributions of Robyn Forrest and Chris Grandin to the 2011 Pacific hake assessment, and to the ideas leading to this analysis. We also thank Owen Hamel, Jim Hastie, and two anonymous reviewers for very helpful comments on preliminary versions of the paper.

References

AD Model Builder. ADMB Project, 2009. AD Model Builder: automatic differentiation model builder. Developed by David Fournier and freely. Available from admb-project.org.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd edition. Springer-Verlag, New York, 488 p.

Carlin, B.J., Louis, T.A., 2000. Bayes and Empirical Bayes Methods for Data Analysis, 2nd edition. Chapman and Hall/CRC, 440 p.

Dennis, B., 1996. Discussion: should ecologists become Bayesians? Ecol. Appl. 6, 1095–1103.

Efron, B., Tibshirani, R.J., 1994. An Introduction to the Bootstrap. Chapman and Hall/CRC, 456 p.

Fournier, D.A., Archibald, C.P., 1982. A general theory for analyzing catch at age data. Can. J. Fisheries Aquat. Sci. 39, 1195–1207.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., Sibert, J., 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods Software 27, 233–249.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis. Chapman and Hall/CRC, 668 p.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo, J.M., Berger, J., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics 4. Oxford University Press, Oxford, pp. 169–193.

Heidelberger, P., Welch, P.D., 1983. Simulation run length control in the presence of an initial transient. Oper. Res. 31, 1109–1144.

Hilborn, R., Liermann, M., 1998. Standing on the shoulders of giants: learning from experience in fisheries. Rev. Fish Biol. Fisheries 8, 273–283.

Hilborn, R., Mangel, M., 1997. The Ecological Detective: Confronting Models with Data. Princeton University Press, Princeton, N.J.

Magnusson, A., Punt, A., Hilborn, R., 2012. Measuring uncertainty in fisheries stock assessment: the delta method, bootstrap, and MCMC. Fish Fisheries Early View, 1–18.

Maunder, M.N., 2003. Paradigm shifts in fisheries stock assessment: from integrated analysis to Bayesian analysis and back again. Nat. Resource Model. 16, 465–475.

Methot, R.D., 2011. User manual for Stock Synthesis. Model version 3.21d. NOAA-NWFSC, Seattle, WA, 165 p.

Methot, R.D., Taylor, I.G., 2011. Adjusting for bias due to variability in estimated recruitments in fishery assessment models. Can. J. Fisheries Aquat. Sci. 68, 1744–1760.

Methot, R.D., Wetzel, C. Fisheries Research, this issue.

Patterson, K., Cook, R., Darby, C., Gavaris, S., Kell, L., Lewy, P., Mesnil, B., Punt, A.E., Restrepo, V., Skagen, D.W., Stefansson, G., 2001. Estimating uncertainty in fish stock assessment and forecasting. Fish Fisheries 2, 125–157.

Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: convergence diagnosis and output analysis for MCMC. R News 6, 7–11.

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

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, http://www.R-project.org/.

Ralston, S., Punt, A.E., Hamel, O.S., DeVore, J.D., Conser, R.J., 2011. A meta-analytic approach to quantifying scientific uncertainty in stock assessments. Fishery Bull. 109, 217–231.

Stewart, I.J., Forrest, R.E., Grandin, C.J., Hamel, O.S., Hicks, A.C., Martell, S.J.D., Taylor, I.G., 2011. Status of the Pacific hake (whiting) stock in U.S. and Canadian waters in 2011. Status of the Pacific Coast Groundfish Fishery through 2011. Stock Assessment and Fishery Evaluation: Stock Assessments, STAR Panel Reports, and rebuilding analyses. Pacific Fishery Management Council, Portland, Oregon, 217 p.

Stewart, I.J., Thorson, J.T., Wetzel, C., 2011. Status of the U.S. sablefish resource in 2011. Status of the Pacific Coast Groundfish Fishery through 2011. Stock Assessment and Fishery Evaluation: Stock Assessments, STAR Panel Reports, and rebuilding analyses. Pacific Fishery Management Council, Portland, Oregon, 442 p.