ICES Journal of
Marine Science

ICES CIEM International Council for the Exploration of the Sea Conseil International pour l'Exploration de la Mer

ICES Journal of Marine Science (2018), 75(2), 585–595. doi:10.1093/icesjms/fsx175

Original Article

Uncertainty estimation and model selection in stock assessment models with non-parametric effects on fishing mortality

Leire Citores[1,2]*, Leire Ibaibarriaga[1], and Ernesto Jardim[3]

[1]*AZTI-Tecnalia, Marine Research Unit, Txatxarramendi Ugartea z/g, 48395 Sukarrieta, Basque Country, Spain*
[2]*BCAM Basque Center for Applied Mathematics, Applied Statistics, Mazarredo 14, E48009 Bilbao, Basque Country, Spain*
[3]*Sustainable Resources Directorate, Water and Marine Resources Unit, European Commission Joint Research Center (JRC), Via Enrico Fermi 2749, 21027 Ispra (VA), Italy*

**Corresponding author: tel: +34 667 174 393; fax: +34 946 572 555; e-mail: lcitores@azti.es*

Citores, L., Ibaibarriaga, L., and Jardim, E. 2017. Uncertainty estimation and model selection in stock assessment models with non-parametric effects on fishing mortality. – ICES Journal of Marine Science, 75: 585–595.

Received 10 March 2017; revised 8 July 2017; accepted 13 August 2017; advance access publication 11 October 2017.

Uncertainty coming from assessment models leads to risk in decision making and ignoring or misestimating it can result in an erroneous management action. Some parameters, such as selectivity or survey catchabilities, can present a wide range of shapes and the introduction of smooth functions, which up to now have not been widely used in assessment models, allows for more flexibility to capture underlying nonlinear structures. In this work a simulation study emulating a sardine population is carried out to compare three different methods for uncertainty estimation: multivariate normal distribution, bootstrap (without and with relative bias correction) and Markov chain Monte Carlo (MCMC). In order to study their performance depending on the model complexity, five different scenarios are defined depending on the shape of the smooth function of the fishing mortality. From 100 simulated datasets, performance is measured in terms of point estimation, coefficients of variation, bias, skewness, coverage probabilities, and correlation. In all approaches model fitting is carried out using the a4a framework. All three methods result in very similar performance. The main differences are found for observation variance parameters where the bootstrap and the multivariate normal approach result in underestimation of these parameters. In general, MCMC is considered to have better performance, being able to detect skewness, showing small relative bias and reaching expected coverage probabilities. It is also more efficient in terms of time consumption in comparison with bootstrapping.

Keywords: assessment model, bootstrap, delta method, Markov chain Monte Carlo, non-parametric smooth function, sardine.

Introduction

The current management system for marine fish and shellfish resources relies in stock assessments to set appropriate exploitation levels for commercial stocks. Stock assessments provide estimates of population and fleet dynamics, as well as past and current status of fish populations. This information constitutes the main product provided by science to policy makers and managers, in order for these to take evidence supported management decisions. Therefore, in this framework, stock assessment models should provide accurate estimates of the relevant parameters including their uncertainty. Such information is important to communicate the levels of confidence on scientific results and to allow the development of risk analysis (Francis and Shotton, 1997; Hilborn *et al.*, 2001).

Brooks and Deroba (2015) classified the uncertainty of stock assessment estimates into three broad categories: observation, structural and estimation uncertainty. Observation uncertainty is inherent to the input data whereas structural uncertainty is related to the model configuration describing the dynamics of the population. Estimation uncertainty is conditioned on the data and the model structure and can be quantified using different statistical tools such as the delta method, bootstrap, likelihood profiling or Markov chain Monte Carlo (MCMC). According to Patterson *et al.* (2001) these tools can be classified within the frequentist, likelihood or Bayesian paradigms. Previous works have compared the performance of different uncertainty methods using real or simulated datasets. Magnusson *et al.* (2012) compared the delta method, bootstrap and MCMC using simulated data and suggested that MCMC was the most reliable method given the dataset and the assessment method. Stewart *et al.* (2012) compared maximum likelihood (MLE) and Bayesian methods, concluding that MLE approximation under or overestimates the upper and lower portion of long tailed distributions. MacCall (2013) presented the delta method as a quick and practical solution for estimating precision of assessment quantities and in Elvarsson *et al.* (2014) bootstrap method was compared with Hessian-based approximations and proposed a comparison with MCMC methods as future work.

Stock assessment models have typically relied on parametric functions that can be easy to implement, but may lack enough flexibility to capture all the data features (Maunder and Harley, 2011; Hillary, 2012). Non-parametric models where the model structure is not fixed beforehand are well-established in regression models like generalized additive models (GAM's) (Härdle, 1990). Similar smooth functions are being progressively introduced into assessment models (Aarts and Poos, 2009; Thorson and Taylor, 2014) because they capture the underlying nonlinear structure of parameters like selectivity or catchability (Crone *et al.*, 2013; Maunder and Piner, 2014). However, little is known about how this might affect the uncertainty estimates. In general, the model fit improves when using smooth functions. However, the *CI*s of the smooth functions are wider at the data extremes where less observations are available (Marra and Wood, 2012). This is expected to affect especially highly complex functions and the final assessment years, where some cohorts have not been observed in full. Aarts and Poos (2009) found wider *CI*s at the beginning of the time series and they suggest that uncertainty in the last years of the assessment could be underestimated. In other studies, such as (Thorson and Taylor, 2014), uncertainty increased both at the beginning and at the end of the time series depending on the shape of the selectivity. In the first case *CI*s are derived from a multivariate distribution generated from MLE estimations and the corresponding hessian matrix using the percentile method and in the second case uncertainty is reported as 80% simulation intervals. This diverges from the usual approach in GAM's where the *CI*s provided for the smooths are computed using a Bayesian approach (Wahba, 1983; Wood, 2006b; Marra and Wood, 2012). So, it seems important not only to understand the uncertainty level when smooth functions are introduced into stock assessment models, but also to analyse if they differ depending on the estimation method.

The inclusion of smooth functions has been advocated as a way forward to avoid model miss-specification (Maunder and Harley, 2011), although it makes the model selection more difficult. Different criteria like AIC (Akaike, 1974) and BIC (Schwarz, 1978) are used to determine which parameters and which shapes are more appropriate in a specific assessment model. AIC tends to select models with a larger number of parameters while BIC tends to choose simpler models (Dziak *et al.*, 2012). Some examples in fisheries include (Wang and Liu, 2006) or (Butterworth and Rademeyer, 2008), but their use in models with smooth functions have not been compared.

The main objective of this work is to compare various uncertainty estimation methods and model selection statistics when nonlinear functions are included into the assessment model, specially, when fishing mortality is modelled by smooth functions depending on age and year.

Methods

A comparison of three uncertainty estimation methods, approximate multivariate normal distribution based on the Hessian matrix, posterior probability intervals from MCMC and parametric bootstrap (without and with bias-correction), is done within a simulation testing framework (Deroba *et al.*, 2014) using the R package *Assessment for all* (a4a) (Jardim *et al.*, 2014). This is a new statistical framework for age-based fish stock assessment designed to be flexible in terms of model structure, using R's syntax for model building. For this study, smooth functions have been introduced for fishing mortality, which range from the classical separable age and year effect as factors to bi-dimensional smooths allowing the interaction between age and year.

Operating model and scenarios

In this study the operating model that represents the true population dynamics, is based on the a4a age-structured model fitted to sardine (*Sardina pilchardus*) in the Bay of Biscay (WGHANSA, 2015). Real data on this stock consists of 13 years (2002–2014) of landings and catch at age (WGHANSA, 2015), numbers at age and weights at age from the acoustic survey PELGAS (Masse *et al.*, in press) and an abundance index from the Daily Egg Production Method survey BIOMAN (Santos *et al.*, 2011, in press). There is observation data from surveys for every year in the model, being each survey an independent data source.

Let $N_{a,y}$ denote the number of individuals of age a at the beginning of year y. Then, according to a4a framework (Jardim *et al.*, 2014) the population dynamics are described by the usual equations:

$$N_{a,y} = N_{a-1,y-1} e^{-Z_{a-1,y-1}} \text{ for } y = 2, \ldots, Y \text{ and } a = 1, \ldots, A-1 \tag{1}$$

and

$$N_{A,y} = N_{A-1,y-1} e^{-Z_{A-1,y-1}} + N_{A,y-1} e^{-Z_{A,y-1}} \text{ for } y = 2, \ldots, Y, \tag{2}$$

where A represents the plus group with individuals aged A and older, which in this case is 6+. The total mortality of age a individuals during year y ($Z_{a,y}$) is decomposed as the sum of fishing mortality ($F_{a,y}$) and natural mortality ($M_{a,y}$):

$$Z_{a,y} = F_{a,y} + M_{a,y} \text{ for } y = 1, \ldots, Y \text{ and } a = 0, \ldots, A. \tag{3}$$

According to the Baranov equation, catch at age ($C_{a,y}$) are computed as the fraction of fishes dying each year due to fishing:

$$C_{a,y} = \frac{N_{a,y}(1 - e^{-Z_{a,y}}) F_{a,y}}{Z_{a,y}}. \tag{4}$$

The a4a model observations equations include observed catch-at-age, an age-structured abundance index and it might also include an aggregated abundance index. Catch at age, $\hat{C}_{a,y}$, are assumed to be log-normally (LN) distributed with variance σ_C:

$$\hat{C}_{a,y} \sim LN(C_{a,y}, \sigma_C) \tag{5}$$

The age-structured and aggregated abundance indices are also assumed to be LN distributed with variances σ_{I1} and σ_{I2}:

$$\hat{I}_{1\ \ a,y} \sim LN(q_1 N_{a,y}, \sigma_{I1}) \tag{7}$$

$$\hat{I}_{2\ \ y} \sim LN\left(q_2 SSB_y, \sigma_{I2}\right), \tag{9}$$

where q_1 and q_2 are survey catchabilities and SSB_y denotes spawning stock biomass in year y.

Natural mortality is assumed known, although it can have a different value for each year and age class. Therefore, the parameters to be estimated are related to recruitment ($R_y = N_{0,y}$), number of individuals at initial year ($N_{a,1}$), fishing mortality ($F_{a,y}$), survey catchabilities (q_1 and q_2) and observation variances (σ_C, σ_{I1}, and σ_{I2}) components. The a4a framework allows a variety of structures for each of these components: from the simplest constant value case to more complex and flexible structures represented by smooth functions depending on age, year or any additional covariate. Parameters are log transformed for estimation.

In this study, five different scenarios (Table 1) have been considered based on the fishing mortality (F) shape depending on age and year, where smoothing is introduced. In the first scenario ages and years are taken as categorical covariates while the rest of the scenarios include smoothing functions with different levels of complexity over ages and years. The first three scenarios represent a separable structure for F while the last ones are non-separable and allow interactions between age and year. Different levels of smoothing were tested by increasing the number of knots in the smoothers. Implementation of smooth functions in a4a is done through the “mgcv” package, (Wood, 2006a), in R. “mgcv” package is used to construct the structure of the selected smooth function with a given number of knots, i.e. the only thing imported from “mgcv” are the design matrices, which are created using “smooth.construct” type functions from this package for each particular smoothing option. a4a uses *s ()* and *te ()* methods which are used to set up the smoothers. *s ()*, given a fixed number of knots, sets up fixed degrees of freedom unpenalized thin plate spline for one dimension, while *te ()*, given also the number of knots, sets up an unpenalized tensor product of cubic splines for multi-dimensional cases. Thus, smoothness is fixed by the user for the generation of the design matrix through “mgcv” given that the parameter estimation method is done in ADMB and it has not been implemented as a method to estimate smoothing parameter numbers inside the assessment model.

For the sake of simplicity, constant catchabilities, q_1, q_2, and constant observations variances, σ_C, σ_{I1}, σ_{I2}, across ages and years were assumed. Recruitment and initial year numbers at age are estimated for each year and each age respectively.

Under these assumptions the models for each scenario were fitted to sardine data. F shapes derived from these structures for each scenario are shown in Figure 1. The estimated parameters were taken as the true values for the operating models in the simulation study.

Dataset simulation

True values were generated fitting real data to the defined operating model for each of the scenarios. Taking estimates of catch and indices (an acoustic index and a biomass index) from this initial run for each scenario as true values, new simulated datasets, consisting of catch-at-age and indices data, were generated multiplying a lognormal error. The coefficient of variation of these errors was set as 25% for catch data and 20% for indices data so that these values remain constant across all scenarios (Magnusson and Hilborn, 2007). For each scenario 100 simulated datasets were generated.

Implemented approaches

For each simulated dataset, the a4a assessment model was fitted using three different approaches for uncertainty estimation. The assessment settings were equal to the operating model used for data simulation, i.e. same parametrizations were used in the operating model and the assessment model of each scenario (Magnusson *et al.*, 2012).

The a4a stock assessment model is fitted with MLE and implemented in R/FLR/ADMB (Kell *et al.*, 2007; R Core Team, 2015). Automatic differentiation is used for likelihood maximization through ADMB (Fournier *et al.*, 2012).

Convergence from ADMB was checked obtaining that 100% of the fits had converged.

Multivariate normal approach

The first estimation method is based on the general MLE theory, according to which the MLE estimates are asymptotically unbiased, normally distributed with variance given by the inverse of the hessian matrix (Van Der Vaart, 1998). Thus, the empirical distribution of the parameters was approximated by obtaining a random sample of size 1000 from a multivariate normal distribution centred on the MLE point estimates with variance-covariance matrix given by the inverse of the hessian matrix. This variance-covariance matrix accounts for the correlation between the estimated parameters. Derived quantities, such as SSB, are computed for each sample, so that their empirical distribution can be derived.

MCMC in ADMB

MCMC methods, used to sample from complex multi-dimensional distributions, are widely used in Bayesian inference. ADMB has an option for Bayesian analysis using MCMC (Monnahan *et al.*, 2014), which has been implemented in the a4a framework. It consists on sampling the MLE surface using MCMC, with a multivariate normal distribution as the proposal function. Although priors are not specified for this study, a flat prior based on parameters bounds is always assumed in ADMB (Millar, 2011). Note that ADMB developers suggest using MCMC results with caution given that there is no documentation available about this implementation. They recommend comparing results using another MCMC software and looking at standard diagnostics. Both recommendations have been followed in this work, looking at convergence diagnostics and obtaining equivalent results in JAGS (Plummer, 2003), verifying this way that obtained MCMC results are valid.

In order to determine the number of draws and the thinning for MCMC convergence was checked using original datasets for the fittings. Posterior traces and autocorrelations where checked and Geweke and Raftery and Lewis’s diagnostic tests were performed and passed. 100 000 draws were obtained with a thinning

Table 1. F structure definition in each scenario.

Scenario	Name	Notation in R	Description
S1	*age x year*	∼factor(age)+factor(year)	Classical age/year separable structure
S2	*high order splines*	∼s(age,k=6)+s(year,k=8)	Separable structure based on splines with 5 degrees of freedom for age and 7 for year.
S3	*low order splines*	∼s(age,k=3)+s(year,k=4)	Separable structure based on splines with 3 degrees of freedom for age and 4 for year.
S4	*high order tensor*	∼te(age,year,k=c(6,5)	Non-separable structure based on a tensor product with 29 degrees of freedom.
S5	low order tensor	∼te(age,year,k=c(3,3)	Non-separable structure based on a tensor product with 8 degrees of freedom.

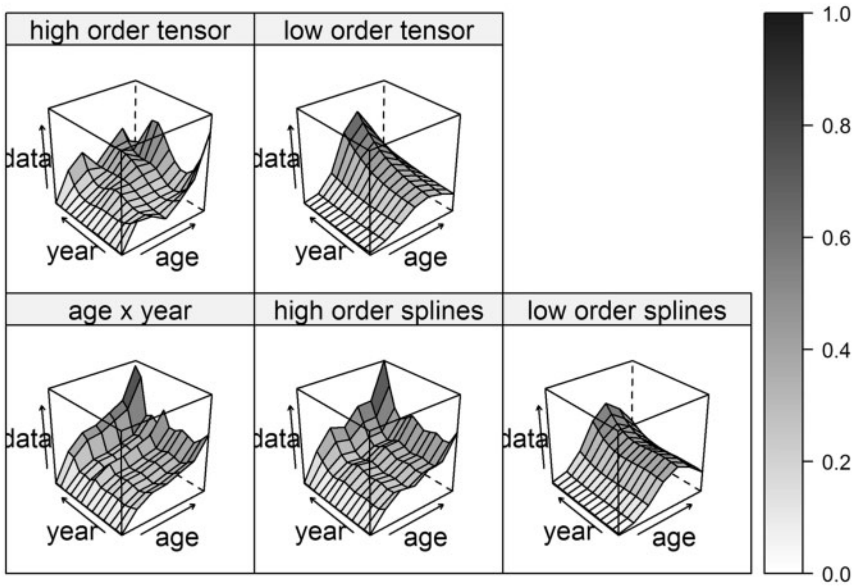

Figure 1. Selected fishing mortality shapes for each scenario as function of age and year.

of 100. Thus, a sample of size 1000 was saved for subsequent analysis.

Parametric bootstrapping

A parametric bootstrap (Efron and Tibshirani, 1994) was performed for each simulated dataset. From the a4a model estimates 1000 catch and indices bootstrap samples were generated according to the observation equations in (5–7). For each bootstrap sample the model is fitted obtaining 1000 estimates. A relative bias correction (BC) algorithm has been applied to these results. The applied algorithm is a relative BC and acceleration algorithm, setting the acceleration coefficient to zero. It adjusts for differences between the median of the bootstrap percentile density function and the estimate obtained with the original data sample (Efron and Tibshirani, 1994).

Performance evaluation

Performance of these methods was evaluated for the following estimated parameters and derived quantities: SSB_y (spawning stock biomass for each year y), R_y (recruitment for each year y), $F_{0,y}$, $F_{2,y}$, $F_{6+,y}$ (fishing mortality for ages 0, 2 and 6+ for each year y), Q_1, Q_2 (survey catchabilities) and σ_C, σ_{I1}, σ_{I2} (observations variances).

For each scenario and each simulated dataset (i), the uncertainty estimation of each parameter (in general denoted as θ) has been measured in terms of:

(1) Point estimate, $\hat{\theta}_i$, defined as the median of the 1000 iterations saved in simulation i.

(2) Coefficient of variation, $CV_{\hat{\theta}_i}$, defined as the cv (ratio of the *SD* to the mean) of the 1000 iterations saved in simulation i.

(3) Relative bias, $B_{\hat{\theta}_i}$, defined as $\frac{\hat{\theta}_i-\theta}{\theta}$, where θ are parameters' true values.

(4) Skewness, $SK_{\hat{\theta}_i}$, defined as $\frac{\mu_{3\theta_i}}{\sigma^3_{\theta_i}}$, where μ_3 is the third central moment and σ_{θ_i} is the *SD* of the 1000 iterations saved in simulation i.

(1) Coverage probability, $CP_{\hat{\theta}}$, is defined as $\Pr(\theta \in PI_{\theta_i})$: the proportion (out of the 100 simulations) of 90% *CI*s (computed using 5th and 95th percentiles of the 1000 iterations saved in simulation *i*) that contain the true value.

Point estimates, coefficient of variations, relative bias and skewness measures defined above have been analysed through their corresponding medians and 5th and 95th quantiles across simulation.

AIC and BIC accuracy evaluation

In order to evaluate the accuracy of AIC and BIC in choosing the correct model, all simulated datasets have been fitted using the five different models defined for each scenario and the AICs and BICs have been computed. In each case the model with the lowest value of either AIC or BIC was defined as the selected model,

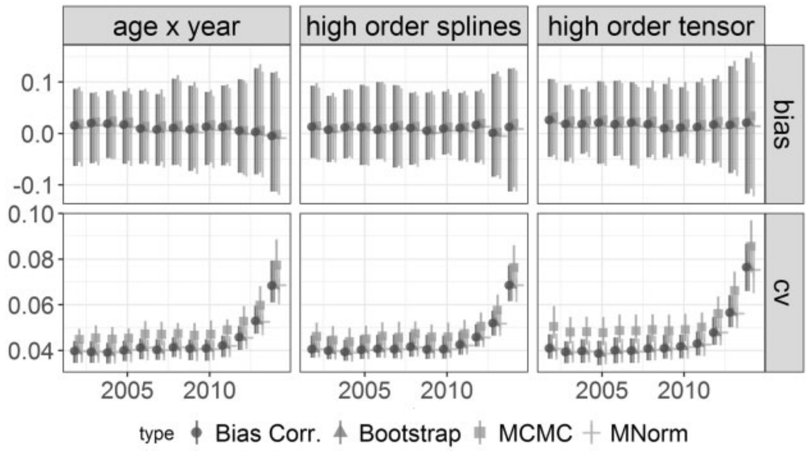

Figure 2. Medians (points) and 90% *CI*s (vertical bars) of the relative bias (top row) and the coefficient of variation (bottom row) of SSB for each year (x-axis). From left to right scenarios *age x year, high order splines,* and *high order tensor* (S1, S2, and S4) are represented in each column. Each point symbol and colour corresponds to an estimation approach.

computing afterwards the proportion of cases in which a correct selection was done for each scenario.

Results

All methods (bootstrap, bias corrected bootstrap, MLE and MCMC) give nearly the same point estimates for SSB, recruitment, fishing mortality and catchability parameters, but different for the estimated variability for catch and indices. The median across the 100 simulations of the relative bias for SSB are slightly positive (around 0.01) but the 90% *CI*s do not show a trend toward positive or negative relative bias (Figure 2). The 90% CIs of the relative bias increase slightly for the last 2 years, related with higher coefficient of variations, being (−0.12, 0.12) in the last year. The empirical distributions of the CVs of SSB increase from less than 5% in the first years to nearly 10% in the last year (Figure 2). When comparing the different approaches the "multivariate normal" and the two types of bootstraps show nearly the same coefficients of variation while the MCMC option in ADMB results in nearly 15% higher CVs in most of the cases. When comparing scenarios, *High order tensor* scenario (S4) presents slightly higher values in the last 2 years and especially for the MCMC approach.

Similar results were obtained in terms of relative bias and CV's for recruitments (Figure 3). The 90% CIs of relative bias for recruitment, around (−0.1, 0.1), do not indicate systematic under or overestimation. The increase of the *CI*s of relative bias in the last year ranges from a ratio of 3.5 with respect to middle years in scenario S1 and S2 to nearly a ratio of 5 in scenario S4. For the CV's in recruitment estimates larger values than for biomass were found, especially in scenario S4 where median CVs in the last year are five times greater than those for SSB, reaching values from 45 to 60% depending on the method. For the rest of the years this difference is lower (around 2.5 times) taking CV values near 6%.

Medians across simulations for catchability parameters (q_1 and q_2) do not show significant relative bias in any of the scenarios and methods with 90% *CI*s around (−0.08, 0.07) for q_1 and (−0.1, 0.1) for q_2 (Figure 4). Observation variances are systematically underestimated in the MLE fit. As a result, generated samples for bootstrapping, where estimated variances are used, present lower variances, and when replicating the fitting it is again underestimated, obtaining even lower observation variance estimates than in the "multivariate normal" approach. The bias corrected version corrects this underestimation for these three parameters, with no significant effects in the rest of estimates. Figure 4 shows relative bias median values taken over the 100 simulations where negatives values are detected for bootstrap and multivariate normal approaches, indicating this underestimation. Most of the 90% CIs for these parameters include the zero value, except for σ_C and σ_{I1} using Bootstrap method. For multivariate normal approach median values are higher but still negative ranging from medians near −0.12 for σ_C to −0.05 for σ_{I2}. BC and MCMC approach seem to perform better in term of bias, showing nearly null median values. Except for scenario S4, where the two bootstrap approaches and the multivariate normal approach show lower median relative bias values for σ_C, similar results were obtained for the rest of scenarios. In all scenarios and with all methods σ_{I2} show wider relative bias *CI*s, around 2.5 times σ_{I1}'s intervals.

Regarding the CVs of the catchability and observation variances, q_2 presents slightly higher median CVs (around 6.5%) than q_1 (4.5%) with wider 90% CIs for all scenarios and methods (Figure 4). σ_C and σ_{I1} present similar CV values with a common pattern where BC method shows lower values with wider 90% CIs around (6.5–8.5)% in contrast with non-overlapping intervals for bootstrap approach. For σ_{I2} a similar pattern is observed with higher CVs around 23%.

Fishing mortality estimates do not show any trend in bias with 90% *CI*s narrower than (−0.3, 0.3) in most of the cases (Figure 5). A small increase of the intervals width can be seen for first and last ages as well as an increase for the last year (30–40% increase with respect to other ages or years), especially for *High order tensor* scenario (S4), where fishing mortality is modelled as a tensor product with high flexibility. In this last scenario estimates for the first age in the last year shows a more than three times wider *CI* in contrast to middle years, ranging from −0.5 to 1.5. Concerning CV medians for F estimates (Figure 6), different shapes are observed depending on the scenario, having similar structures across methods. Scenarios where F was modelled with a separable submodel present a

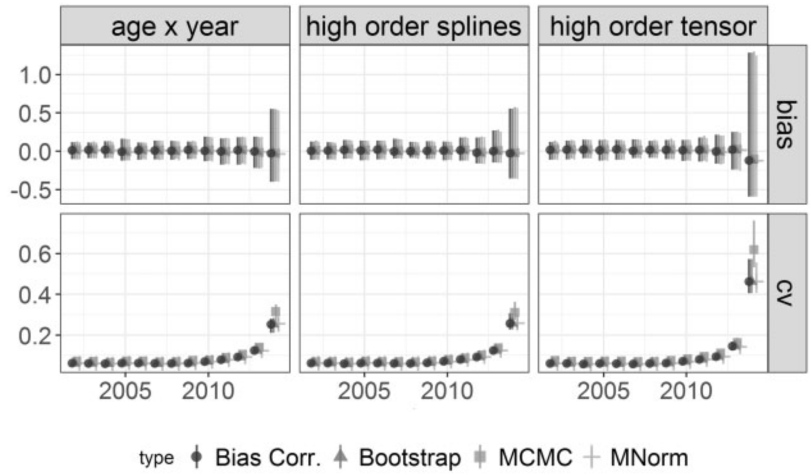

Figure 3. Medians (points) and 90% *CI*s (vertical bars) of the relative bias (top row) and the coefficient of variation (bottom row) of recruitment for each year (x-axis). From left to right scenarios *age x year*, *high order splines*, and *high order tensor* (S1, S2, and S4) are represented in each column. Each point symbol and colour corresponds to an estimation approach.

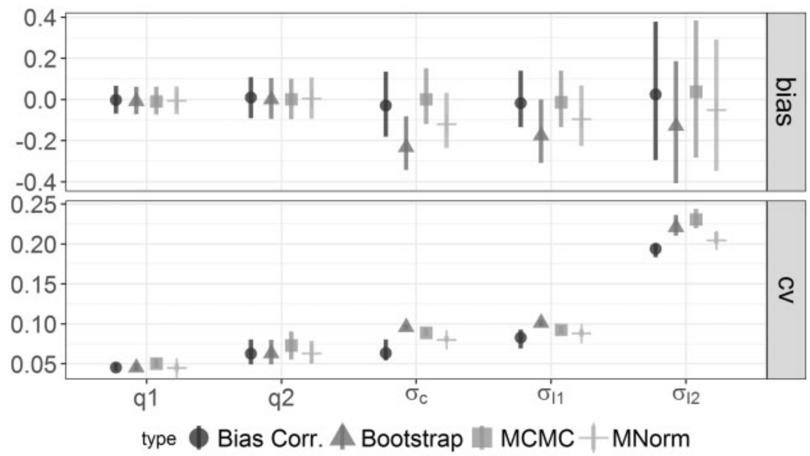

Figure 4. Medians (points) and 90% *CI*s (vertical bars) of the relative bias (top row) and the coefficient of variation (bottom row) of catchabilities (q_1 and q_2) and observation variances (σ_C, σ_{I1}, σ_{I2}) for scenario S2. Each point symbol and colour corresponds to an estimation approach.

flatter shape, with median CV values around 11% and a small increase for the last year, taking values up to 16%. Scenarios where F was modelled as non-separable the increase in median CVs was not only for last year but also for first years and first and last ages, ranging from 12% for middle years and ages to 50% for the most extreme value.

All selected parameter distributions seem to be symmetric as the skewness statistic range between −1 and 1 (Figure 7).

Coverage probabilities for parameters in the last year (Table 2) for the "multivariate normal" approach and bootstrapping are lower, taking values between 0.75 and 0.85 while MCMC method gives coverage probabilities above 0.9, as shown in Figure 8 for the whole time series of SSB, recruitment and fishing mortality parameters. Comparisons of coverage probabilities that are not shown in the table (for the rest of the years) are similar to the presented values. For biased estimations, such as observation variances, coverage probabilities are very low, mostly below 0.5, except for the MCMC approach where values near 0.9 are reached.

When comparing across methods, the MCMC shows a more stable performance in face of more complex F models. The results obtained were always close to 0.9, including when the F model complexity increased to a tensor product (S4). The other methods show a deterioration of the coverage in S4 when comparing with S1 and S2.

Concerning AIC and BIC accuracy evaluation for model selection, in the case of AIC selection criteria, the scenario with the smallest number of parameter show the lowest model selection accuracy, while the scenario with the highest number of parameters, low order splines scenario, has been selected correctly with an accuracy of 100% (Figure 9). The rest of scenarios present an accuracy of around 75–85%. For BIC selection criteria, all scenarios except the classical separable one, were selected correctly 100% of the times. For the classical separable scenario, the scenario with high order splines was selected 75% of the times. This scenario is the most similar one in terms of number of parameters and separability assumption. In miselection cases AIC and BIC values appear to be very close.

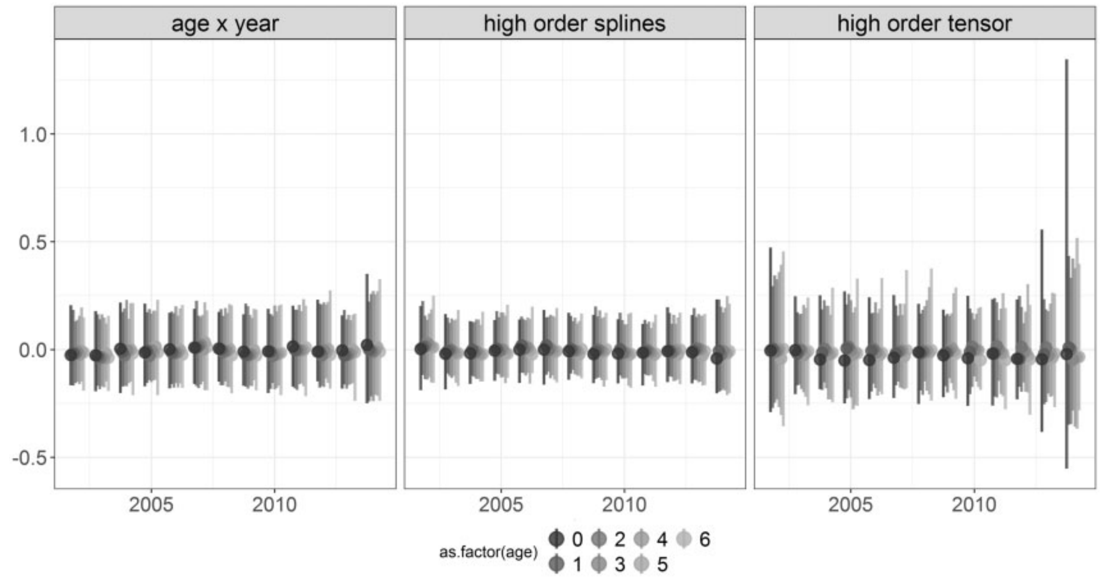

Figure 5. Medians (points) and 90% *CI*s (vertical bars) of the relative bias fishing mortality for each year (x-axis) for the MCMC approach. From left to right scenarios *age x year*, *high order splines*, and *high order tensor* (S1, S2, and S4) are represented in each column. Each colour corresponds to an age.

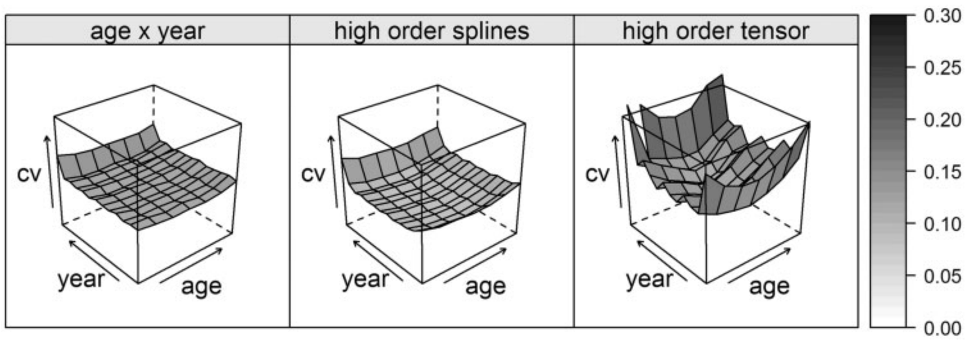

Figure 6. Median coefficients of variation of fishing mortality as a function of age and year for the MCMC approach. From left to right the panels correspond to scenarios *age x year*, *high order splines*, and *high order tensor* (S1, S2, and S4).

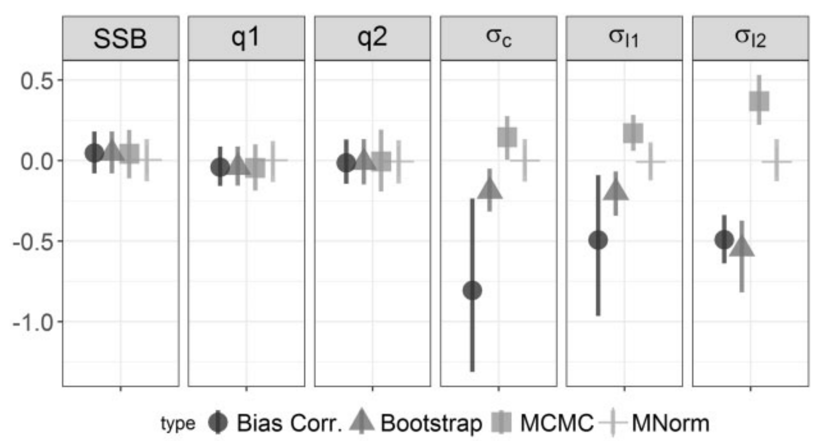

Figure 7. Medians (points) and 90% *CI*s (vertical bars) of the skewness SSB in the last year, catchabilities (q_1 and q_2) and observation variances (σ_C, σ_{I1}, σ_{I2}) for scenario S2. Each point symbol and colour corresponds to an estimation approach.

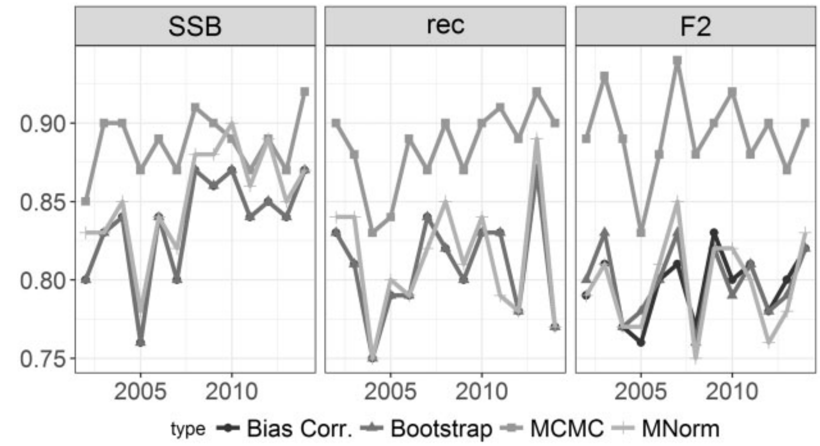

Figure 8. Coverage probabilities for each year and each method for SSB, recruitment (rec) and fishing mortality at age 2 (F2) parameters in scenario S4.

Table 2. Coverage probabilities for each scenario and each estimation method for catchabilities q_1 and q_2, observation variances σ_C, σ_{I1}, σ_{I2} and for SSB, R, fishing mortality at ages 0, 2, and 6 in the last year.

Scenario	Method	ssb	rec	f0	f2	f6	q_1	q_2	σ_C	σ_{I1}	σ_{I2}
S1	**Relative bias Corr.**	0.89	0.8	0.77	0.83	0.85	0.86	0.86	0.24	0.86	0.79
S1	**Bootstrap**	0.89	0.8	0.78	0.82	0.83	0.87	0.86	0	0.54	0.75
S1	**MCMC**	0.91	0.89	0.88	0.89	0.86	0.92	0.91	0.92	0.89	0.83
S1	**MNorm**	0.91	0.8	0.78	0.82	0.83	0.87	0.88	0.35	0.75	0.78
S2	**Relative bias Corr.**	0.86	0.85	0.9	0.88	0.91	0.92	0.9	0.54	0.79	0.83
S2	**Bootstrap**	0.86	0.85	0.9	0.89	0.92	0.92	0.93	0.14	0.4	0.75
S2	**MCMC**	0.92	0.9	0.97	0.92	0.94	0.93	0.94	0.91	0.93	0.9
S2	**Mnorm**	0.87	0.85	0.92	0.88	0.92	0.92	0.91	0.5	0.65	0.82
S3	**Relative bias Corr.**	0.85	0.84	0.88	0.89	0.88	0.91	0.92	0.89	0.73	0.85
S3	**Bootstrap**	0.85	0.84	0.88	0.89	0.86	0.91	0.92	0.37	0.38	0.74
S3	**MCMC**	0.89	0.88	0.91	0.9	0.9	0.93	0.95	0.91	0.93	0.89
S3	**MNorm**	0.85	0.85	0.88	0.89	0.89	0.91	0.93	0.74	0.6	0.82
S4	**Relative bias Corr.**	0.87	0.77	0.74	0.82	0.9	0.88	0.89	0.05	0.76	0.84
S4	**Bootstrap**	0.87	0.77	0.73	0.82	0.89	0.89	0.9	0	0.38	0.74
S4	**MCMC**	0.92	0.9	0.88	0.9	0.94	0.9	0.9	0.84	0.93	0.88
S4	**MNorm**	0.87	0.77	0.75	0.83	0.9	0.87	0.9	0.18	0.64	0.8
S5	**Relative bias Corr.**	0.9	0.87	0.86	0.86	0.81	0.87	0.93	0.78	0.76	0.85
S5	**Bootstrap**	0.9	0.87	0.86	0.86	0.83	0.87	0.93	0.29	0.35	0.7
S5	**MCMC**	0.92	0.9	0.89	0.89	0.88	0.92	0.94	0.9	0.94	0.89
S5	**MNorm**	0.9	0.85	0.86	0.87	0.8	0.88	0.93	0.62	0.61	0.83

Discussion

The interest in incorporating uncertainty in management processes is increasing given that the lack of consideration of different sources of uncertainty has caused many failures in fisheries management (Hilborn and Peterman, 1996). Thus, it becomes crucial to quantify correctly the uncertainty in fisheries assessments. In this work different methodologies were compared, focusing on estimation uncertainty. The work is focused on the introduction of non-parametric functions to model fishing mortality in stock assessment models. In addition, the accuracy of AIC and BIC in model selection has been analysed.

General results from the comparison of methods agreed with previous studies: (Magnusson *et al.*, 2012) identified MCMC method as the most reliable one and Stewart *et al.* (2012) suggest that MLEs generally misestimate skewed distributions in the tails which does not happen when using MCMC methods. As in Magnusson *et al.* (2012) the BC algorithm for bootstrapping resulted in an improvement of results. However in comparison to Magnusson *et al.* (2012) results, in this study it was not detected a systematic underestimation of coverage probabilities for MCMC method.

Parametric models are sometimes considered too restrictive to capture the complex dynamics of fish populations and the fleets exploiting them (Hillary, 2012). In Fronczyk *et al.* (2011) a Bayesian non-parametric approach based on a mixture model for the joint distribution of log-reproductive success and stock biomass is proposed and compared with simpler parametric and semi-parametric models for North Atlantic cod data. They conclude that the non-parametric model outperforms the simpler ones. In Hillary (2012) a Ricker model was compared with a non-parametric alternative and found the latter had better performance.

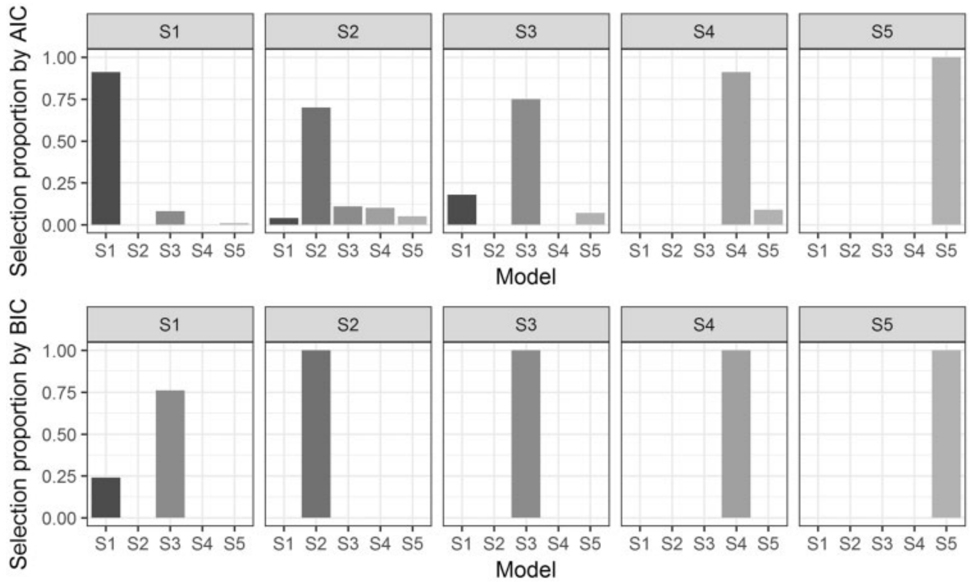

Figure 9. Selections proportion when using AIC (top) and when using BIC (bottom). Each panel correspond to a scenario from which data was generated and x axis represents the range of eligible models.

Most recent works have focused on introducing more flexibility in fishing mortality estimates using different approaches. In Fernandez *et al.* (2010) autocorrelation processes are used to model a more flexible selectivity and in Nielsen and Berg (2014) the incorporation of a correlation parameter to capture the temporal smooth development in selectivity is proposed. In these approximations a single parameter is included and is estimated from the data instead of pre-specifying the smoothness degree. Other works, such as Thorson and Taylor (2014) or Martell and Stewart (2013) have studied the usage of non-parametric functions to model selectivity in assessment models, finding in the first case, that non-parametric models have less relative bias and greater precision when the parametric function is misspecified. In the second case the authors suggested that when there is no precise knowledge about the fishery or catch data, adopting a flexible selectivity, such as an age-based selectivity interpolated over age and year using a bicubic spline, may be more appropriate than assuming constant selectivity.

Some available software for stock assessment include options for non-parametric modelling. Stock Synthesis (Methot and Wetzel, 2013) includes a non-parametric submodel for size selectivity, using waypoints and a set of linear segments while MULTIFAN-CL (Fournier *et al.*, 1998) uses cubic splines to model selectivity.

As mentioned earlier, non-parametric functions, splines and their tensor products in this study, present advantages in terms of the ability to capture flexible shapes in comparison with parametric options. It has been seen that more flexible shapes result in greater uncertainties, finding greater CVs and relative bias at terminal years and at terminal ages for most complex scenarios, probably due to a larger number of estimated parameters. Coverage probabilities were also found to deteriorate for the most complex model, although not for MCMC method. As in our case study, the extra freedom in non-parametric models used in Hillary (2012) resulted in higher uncertainty level. The differences in uncertainty quantification could have an effect in management advice and risk assessment when evaluating the consequences of different management actions under uncertainty.

The changing level of uncertainty estimated for last periods related to the degree of flexibility in submodels could be handled using model selection. In this work the most commonly used criteria (AIC and BIC) have been evaluated in terms of accuracy in selecting the correct model resulting in a better accuracy for BIC. Thorson *et al.* (2013) propose a stepwise model selection using AIC to select the degree of smoothness for time-varying parameters and Maunder and Harley (2011) state the need for an alternative to AIC and BIC criteria and perform cross validation model selection to determine non-parametric selectivity curves. An alternative to selecting a single model is performing model averaging, which eliminates the need for selecting a ‘best’ model and rejecting all alternative assumptions as proposed in Millar *et al.* (2014) and references therein. Model selection or model averaging are not very common practices in stock assessment yet, although the incorporation of these kind of approaches is an issue with an increasing interest (Anderson *et al.*, 2017).

Funding

This research was partly funded by the Basque Government through the IM16PELAGI project and the BERC 2014–2017 programme and by the Spanish Ministry of Economy and Competitiveness (MINECO) through the BCAM Severo Ochoa excellence accreditation SEV-2013-0323. LC is supported by a doctoral grant from AZTI Foundation.

Acknowledgements

We thank Andres Uriarte for general comments and recommendations on earlier versions of the article. We also thank two anonymous reviewers and the editor for their thoughtful reviews and constructive comments. This article is contribution No. 826 from AZTI (Marine Research).

References

Aarts, G., and Poos, J. J. 2009. Comprehensive discard reconstruction and abundance estimation using flexible selectivity functions. ICES Journal of Marine Science, 66: 763–771.

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19: 716–723.

Anderson, S. C., Cooper, A. B., Jensen, O. P., Minto, C., Thorson, J. T., Walsh, J. C., Afflerbach, J. *et al*. 2017. Improving estimates of population status and trend with superensemble models. Fish and Fisheries, 18: 732–741.

Brooks, E. N., and Deroba, J. J. 2015. When "data" are not data: the pitfalls of post hoc analyses that use stock assessment model output. Canadian Journal of Fisheries and Aquatic Sciences, 72: 634–641.

Butterworth, D. S., and Rademeyer, R. A. 2008. Statistical catch-at-age analysis vs. ADAPT-VPA: the case of Gulf of Maine cod. ICES Journal of Marine Science, 65: 1717–1732.

Crone, P., Maunder, M., Valero, J., McDaniel, J., and Semmens, B. 2013. Selectivity: theory, estimation, and application in fisheru stock assessment models. Center for the advancement of Populations assessment methodology (CAPAM), Workshop Series, Report, 1: 1–46.

Deroba, J. J., Butterworth, D. S., Methot, R. D., De Oliveira, J. A. A., Fernandez, C., Nielsen, A., Cadrin, S. X. *et al*. 2014. Simulation testing the robustness of stock assessment models to error: some results from the ICES strategic initiative on stock assessment methods. ICES Journal of Marine Science, 72: 19–30.

Dziak, J. J., Coffman, D. L., Lanza, S. T., and., and Li, R. 2012. Sensitivity and Specificity of Information Criteria. The Methodology Center and Department of Statistics, Penn State, The Pennsylvania State University, pp. 1–10.

Efron, B., and. Tibshirani, R. J. 1994. An Introduction to the Bootstrap, Monographs in Statistics and Applied Probability 57, Chapman and Hall/CRC.

Elvarsson, B., Taylor, L., Trenkel, V. M., Kupca, V., and Stefansson, G. 2014. A bootstrap method for estimating bias and variance in statistical fisheries modelling frameworks using highly disparate datasets. African Journal of Marine Science, 36: 99–110.

Fernandez, C., Cervino, S., Perez, N., and Jardim, E. 2010. Stock assessment and projections incorporating discard estimates in some years: an application to the hake stock in ICES Divisions VIIIc and IXa. ICES Journal of Marine Science, 67: 1185–1197.

Fournier, D. A., Hampton, J., and Sibert, J. R. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, Thunnus alalunga. Canadian Journal of Fisheries and Aquatic Sciences, 55: 2105–2116.

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

Francis, R., and Shotton, R. 1997. "Risk" in fisheries management: a review. Canadian Journal of Fisheries and Aquatic Science, 54: 1699–1715.

Fronczyk, K., Kottas, A., and Munch, S. 2011. Flexible modeling for stock-recruitment relationships using Bayesian nonparametric mixtures. Environmental and Ecological Statistics, 19: 183–204.

Härdle, W. 1990. Applied Nonparametric Regression, Cambridge University Press, Econometric Society monographs, Cambridge.

Hilborn, R., Maguire, J.-J., Parma, A. M., and Rosenberg, A. A. 2001. The Precautionary Approach and risk management: can they increase the probability of successes in fishery management? Canadian Journal of Fisheries and Aquatic Sciences, 58: 99–107.

Hilborn, R., and Peterman, R. M. 1996. The development of scientific advice with incomplete information in the context of the precautionary approach. FAO Fisheries Technical Paper, 350: 77–102.

Hillary, R. M. 2012. Practical uses of non-parametric methods in fisheries assessment modelling. Marine and Freshwater Research, 63: 606.

Jardim, E., Millar, C. P., Mosqueira, I., Scott, F., Osio, G. C., Ferretti, M., Alzorriz, N. *et al*. 2014. What if stock assessment is as simple as a linear model? The a4a initiative. ICES Journal of Marine Science, 72: 232–236.

Kell, L. T., Mosqueira, I., Grosjean, P., Fromentin, J.-M., Garcia, D., Hillary, R. M., Jardim, E., *et al*. 2007. FLR: an open-source framework for the evaluation and development of management strategies. ICES Journal of Marine Science, 64: 640–646.

MacCall, A. D. 2013. Use of the delta method to evaluate the precision of assessments that fix parameter values. Fisheries Research, 142: 56–60.

Magnusson, A., and Hilborn, R. 2007. What makes fisheries data informative? Fish and Fisheries, 8: 337–358.

Magnusson, A., Punt, A. E., and Hilborn, R. 2012. Measuring uncertainty in fisheries stock assessment: the delta method, bootstrap, and MCMC. Fish and Fisheries, 14: 325–342.

Marra, G., and Wood, S. N. 2012. Coverage properties of confidence intervals for generalized additive model components. Scandinavian Journal of Statistics, 39: 53–74.

Martell, S., and Stewart, I. 2013. Towards defining good practices for modeling time-varying selectivity. Fisheries Research, 158: 84–95.

Masse, J., Duhamel, E., Petitgas, P., Doray, M., and Huret, M. In press. Spring Acoustic Surveys: Pelgas survey. In Pelagic Surveys series for sardine and anchovy in ICES Areas VIII and IX (WGACEGG) - Towards an ecosystem approach. ICES Cooperative Research Report (CRR) No 332.

Maunder, M. N., and Harley, S. J. 2011. Using cross validation model selection to determine the shape of nonparametric selectivity curves in fisheries stock assessment models. Fisheries Research, 110: 283–288.

Maunder, M. N., and Piner, K. R. 2014. Contemporary fisheries stock assessment: many issues still remain. ICES Journal of Marine Science, 72: 7–18.

Methot, R. D., and Wetzel, C. R. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research, 142: 86–99.

Millar, C. P., Jardim, E., Scott, F., Osio, G. C., Mosqueira, I., and Alzorriz, N. 2014. Model averaging to streamline the stock assessment process. ICES Journal of Marine Science, 72: 93–98.

Millar, R. B. 2011. Maximum Likelihood Estimation and Inference: With Examples in R. SAS and ADMB, John Wiley & Sons, Chichester, UK.

Monnahan, C. C., Muradian, M. L., and Kuriyama, P. T. 2014. A Guide for Bayesian Analysis in AD Model Builder, doi: 10.13140/RG.2.1.2700.1043.

Nielsen, A., and Berg, C. W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fisheries Research, 158: 96–101.

Patterson, K., Cook, R., Darby, C., Gavaris, S., Kell, L., Lewy, P., Mesnil, B. *et al*. 2001. Estimating uncertainty in fish stock assessment and forecasting. Fish and Fisheries, 2: 125–157.

Plummer, M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *In* Proceedings of the 3rd International Workshop on Distributed Statistical Computing, p. 125. Vienna.

R Core Team. 2015. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Santos, M., Uriarte, A., and Boyra, G., and Ibaibarriaga, L. In press. Anchovy DEPM surveys 2003 - 2012 in the Bay of Biscay (subarea VIII) BIOMAN. *In* Pelagic Surveys series for sardine and anchovy in ICES Areas VIII and IX (WGACEGG) - Towards an ecosystem approach. ICES Cooperative Research Report (CRR) No 332.

Santos, M., Uriarte, A., and Ibaibarriaga, L. 2011. Spawning Stock Biomass estimates of the Bay of Biscay anchovy (Engraulis encrasicolus, L.) in 2010 applying the Daily Egg Production Method. Revista de Investigación Marina, 18: 76–91.

Schwarz, G. 1978. Estimating the dimension of a model. The Annals of Statistics, 6: 461–464.

Stewart, I. J., Hicks, A. C., Taylor, I. G., Thorson, J. T., Wetzel, C., and Kupschus, S. 2012. A comparison of stock assessment uncertainty estimates using maximum likelihood and Bayesian methods implemented with the same model framework. Fisheries Research, 142: 37–46.

Thorson, J. T., and Taylor, I. G. 2014. A comparison of parametric, semi-parametric, and non-parametric approaches to selectivity in age-structured assessment models. Fisheries Research, 158: 74–83.

Thorson, J. T., Zhou, S., E., Punt, A., and Smith, A. D. M. 2013. Astepwise-selected spline approximation to time-varying parameters,with application to occupancy modelling. Methods in Ecology and Evolution, 4: 123–132.

Van Der Vaart, A. W. 1998. Asymptotic Statistics, Cambridge University Press.

Wahba, G. 1983. Bayesian "confidence intervals" for the cross-validated smoothing spline. Journal of the Royal Statistical Society. Series B (Methodological), 45: 133–150.

Wang, Y., and Liu, Q. 2006. Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock–recruitment relationships. Fisheries Research 77: 220–225.

WGHANSA. 2015. Report of the Working Group on Southern Horse Mackerel, Anchovy and Sardine, Lisbon.

Wood, S. 2006a. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC.

Wood, S. N. 2006b. On confidence intervals for generalized additive models based on penalized regression splines. Australian and New Zealand Journal of Statistics, 48: 445–464.

Handling editor: Jan Jaap Poos