SCRS/2012/154 Collect. Vol. Sci. Pap. ICCAT, 69(2): 913-937 (2013)

USING BAYES FACTORS TO EVALUATE THE CREDIBILITY OF STOCK-RECRUITMENT RELATIONSHIPS FOR WESTERN ATLANTIC BLUEFIN TUNA

Murdoch K. McAllister[1]

SUMMARY

This paper computes Bayesian posterior probabilities for different stock-recruit models for western Atlantic bluefin tuna on (a) whether regime shifts have affected recruitment and (b) the mathematical form of the stock-recruit function. The analysis uses stock-recruit data obtained from ICCAT's 2010 stock assessment of western Atlantic bluefin tuna. I present Bayes factors for the alternative hypotheses which reflect the ratio of Bayes' probability of the data for two "competing" hypothesis and identified improved formulations of prior probabilities removed their influence Bayes factors. The sensitivity of Bayes factors to different assumptions was evaluated, including settings for autocorrelation in stock-recruit deviates, priors, variance in recruitment deviates and the data to analyze. When applied to stock-recruit data from 1970 to 2006 and presuming a regime shift in 1977, Bayes factors were 4.8:1, favoring the Beverton-Holt no regime-shift model (i.e., the probability of the data was 4.8 times higher for the Beverton-Holt model compared to the two-line regime shift model). When different input settings were applied, the values obtained for Bayes factors for the two hypotheses ranged widely. They ranged, from about 1:2 against to 67:1 in favour of the Beverton-Holt model.

RÉSUMÉ

Ce document calcule les probabilités bayésiennes a posteriori pour différents modèles stock-recrutement pour les thons rouges de l'Atlantique Ouest sur la question de savoir si (a) des changements de régime ont affecté le recrutement et (b) la formule mathématique de la fonction stock-recrutement. L'analyse utilise les données de stock-recrutement obtenues de l'évaluation des stocks de thon rouge de l'Atlantique Ouest réalisée par l'ICCAT en 2010. Sont présentés des facteurs bayésiens pour les hypothèses alternatives qui reflètent le ratio de la probabilité bayésienne des données pour deux hypothèses "concurrentes" et sont identifiées des formulations améliorées de probabilités a priori qui ont rendu négligeable leur influence sur les facteurs bayésiens. La sensibilité des facteurs bayésiens à différents postulats a été évaluée, y compris les configurations pour l'auto-corrélation dans les déviations stock-recrutement, les priors, la variance dans les déviations du recrutement et les données à analyser. Lorsqu'ils sont appliqués aux données de stock-recrutement de 1970 à 2006 et postulant un changement de régime en 1977, les facteurs bayésiens s'élevaient à 4.8.1, favorisant le modèle sans changement de régime de Beverton-Holt (c.-à-d. la probabilité des données était 4,8 fois supérieure pour le modèle de Beverton-Holt que pour le modèle de changement de régime à deux lignes). Lorsque différentes configurations de données d'entrée étaient appliquées, les valeurs obtenues pour les facteurs bayésiens pour les deux hypothèses ont largement varié. Elles se sont situées à environ 1:2 par rapport à 67:1 en faveur du modèle de Beverton-Holt.

RESUMEN

En este documento se calculan las probabilidades bayesianas posteriores para diferentes modelos stock-reclutamiento para el atún rojo del Atlántico occidental (a) cuando los cambios de régimen afectan al reclutamiento y (b) en la forma matemática de la función stock-reclutamiento. En los análisis se utilizaron los datos stock-reclutamiento obtenidos de la evaluación de stock de ICCAT de 2010 de atún rojo del Atlántico oeste. Se presentan factores bayesianos para hipótesis alternativas que reflejan la ratio de probabilidad de Bayes de los datos para dos hipótesis "candidatas" y se identificaron formulaciones mejoradas de probabilidades previas que hicieron que su influencia en los factores bayesianos fuera inapreciables. Se evaluó la sensibilidad de los factores bayesianos a diferentes supuestos, lo que incluye especificaciones para la autocorrelación en desviaciones stock-reclutamiento, distribuciones previas, variación en desviaciones del reclutamiento y los datos que se tienen

[1] Fisheries Centre, AERL, 2202 Main Mall, University of British Columbia, Vancouver, B.C., Canada, V6T 1Z4, m.mcallister@fisheries.ubc.ca

que analizar. Cuando se aplican a los datos stock-reclutamiento de 1970 a 2006, y se parte del supuesto de un cambio de régimen en 1977, los factores Bayes fueron 4,8:1, favoreciendo el modelo sin cambio de régimen de Beverton-Holt (a saber, la probabilidad de los datos era 4,8 veces superior para el modelo Beverton-Holt que para el modelo de cambio de régimen de dos líneas). Cuando se aplicaron especificaciones diferentes para los valores de entrada, los valores obtenidos para los factores Bayes para las dos hipótesis oscilaron ampliamente. Oscilaron entre 1:2 y 67:1 a favor del modelo Beverton-Holt.

KEYWORDS

Stock-recruit function, stock assessment, regime shift, fisheries management, Bayes factors

INTRODUCTION

It is common in stock assessment for results to vary with input assumptions or hypotheses about the structural formulation of the stock assessment model considered. This may occur when different assumptions are made about recruitment to the exploited population, for example, about the form of the stock-recruit relationship, and there is uncertainty over which one is correct. This is often troublesome because different predictions about the consequences of alternative management actions can result from applying such different hypotheses in a stock assessment.

Traditionally, assessment scientists choose the most believable set of assumptions as a base case scenario, and then run models with other assumptions as sensitivity analyses. Decision-makers, when presented with a base case, and sensitivity analyses, tend to consider only the base case in choosing management policies. In the case when two or more hypotheses are considered believable, assessment scientists may present decision-makers with results for different sets of assumptions without choosing a preferred model. Managers are then presented with a variety of results in which the best action depends upon which state of nature is assumed to be true. Without any formal scientific guidance about how to weight the different hypotheses or assumptions, scientific considerations about their plausibility are often ignored in the decision making process.

This appears to have been the case in the stock assessment of western Atlantic bluefin tuna where some alternative stock recruitment relationships were hypothesized in the 1990s and two of these have since been applied in stock projections (ICCAT, 1999). ICCAT scientists have in their stock assessment reports indicated that the two alternative hypotheses on recruitment are equally plausible and that there’s been no scientific basis with which to evaluate the credibility of the two hypotheses (ICCAT 2002, 2008, 2010). Stock rebuilding measures that have been adopted for the western Atlantic bluefin tuna stock have been based primarily on projection results from only one of the two hypotheses, despite scientists’ assertion that both hypotheses have remained equally plausible. In contrast, McAllister *et al*. (2000a, b) suggested a Bayesian approach dealing with the uncertainty in these alternative recruitment hypotheses within the context of the provision of fisheries management advice. This paper takes up this issue by extending the approach and analysis of McAllister *et al*. (2000a, b).

Butterworth *et al*. (1996) addressed this problem in stock assessment where different models suggested different optimal policies and emphasized the importance of developing a scientific basis to weight the different models. They suggested a number of systematic approaches to constructing weightings for alternative hypotheses on model structure. Raftery and Richardson (1996), McAllister *et al.* (1999) and McAllister and Kirchner (2002) have suggested some Bayesian statistical approaches to constructing these weightings. McAllister *et al*. (2000a) explored one of these and presented Bayesian decision analysis as a formal approach to provide empirically based weightings for conflicting stock assessment results and provide scientific guidance to decision makers when they are faced with such conflicting results. The weightings come in the form of posterior probabilities which reflect the probability of a given model or hypothesis given the available data. McAllister *et al*. (2000a) proposed a statistical methodology to compute such probabilities for alternative models: the application of the sampling importance resampling (SIR) algorithm to the available data. This approach is amenable to be used in conjunction with existing ICCAT stock assessment methods because it can be used in concert with ADAPT VPA methods, catch-age methods, and age-structured and non-age-structured surplus production modeling. While it is useful to communicate about the credibility of a hypothesis by referring to posterior probabilities, these are often difficult to interpret since the probabilities by definition must sum to one and the probability values tend to get spread more thinly as the number of discrete alternative hypotheses increases.

In this paper I make use of Bayesian posterior probabilities and also Bayes factors for structurally different models (Kass and Raftery 1995). Bayes factors for alternative hypotheses reflect the ratio of the probability of the data given one particular model to the probability of the same data for a second model. They do not require the formulation and application of prior probabilities for different models, are directly proportional to posterior probabilities when the prior probabilities for the different hypotheses are equal, and based on recent experience (e.g., King *et al*. 2011) are easier to interpret than probability values.

McAllister *et al*. (2000a, b) examined two sets of hypotheses about western Atlantic bluefin tuna recruitment: whether the functional form of the stock-recruit relationship is Beverton-Holt or “2-line”, and whether there was or was not a regime shift (Hare and Francis 1995) that changed the stock-recruit relationship. This paper extends the data analysis to include western Atlantic bluefin tuna stock recruit data from ICCAT’s 2010 assessment of Atlantic bluefin tuna (ICCAT 2010) and the precise set of statistical assumptions applied in the 2010 assessment, for example, regarding autocorrelation in deviates from the stock-recruit functions considered. It also identifies prior probability distributions for stock-recruit model parameters that have negligible influence on Bayes factors for the different stock-recruit models and evaluates the sensitivity of Bayes factors to applying different years for the year of the regime shift, the final year of data to include and different assumptions about variance and autocorrelation in stock-recruit function deviates.

Methods

For descriptions of the general decision analytic approach and statistical protocols to dealing with structurally different stock assessment and stock-recruit models and computing marginal posterior probabilities for these using importance sampling see McAllister *et al*. (2000a). I outline below the statistical formulations of the stock-recruit models applied in McAllister *et al*. (2000b).

One quantity summarizing the weight of evidence in support of structurally different models is the marginal posterior probability for each model. This is given by:

$$P(m_j|\,\text{data}) = \frac{p(m_j)\int p(\theta_j)P(\text{data}|\theta_j)d\theta_j}{\sum_{l=1}^{N_m} p(m_l)\int p(\theta_l)P(\text{data}|\theta_l)d\theta_l} \qquad (1)$$

where $p(m_j)$ is the prior probability for model *j*, the relative plausibility for model *j* prior to evaluating such data, and $p(\theta_j)$ is the prior probability density function of parameter vector θ_j under model *j* and $P(\text{data}|\theta_j)$ is the probability density of the data obtained given the set of parameter values θ_j under model *j* (in Bayes theorem, often referred to as likelihood function of the data). The value $p(\theta_j)$ represents the probability for a given set of values for the parameters in model *j* prior to obtaining a set of data that can further our ability to discriminate among alternative parameter values. In absence of a consensus on which prior probability distribution to apply, a non-informative prior probability distribution may be applied as the base case prior. Such priors should allow the data to speak for themselves. For example, for discrete alternative hypotheses I would give each hypotheses the same prior probability: $p(m_j) = 1/N_m$ where N_m is the number of alternative hypotheses or models considered.

This paper considers the same stock-recruit data for western Atlantic bluefin tuna from ICCAT’s 2010 assessment. I’ve used the 2010 stock-recruit data for 1970-2008 (source Laurie Kell, ICCAT Secretariat, Madrid). These data together with the data applied in McAllister *et al*. (2000b) are shown in **Table 1**. A third dataset (Porch *et al*. 1960) for 1960-1998 was also considered in a sensitivity analysis since it extends further back when SSB and recruitment were higher. For most of the analyses, I have not included recruitment data after 2006 due to the decay in reliability of recruitment estimates in the latest recruits obtained from a VPA. However, since the 2010 assessment included the 2006 stock-recruit data to characterize stock-recruit functions, the reference case results in this paper include the 2006 data point. I’ve focused on the two alternative hypotheses on functional form of the stock-recruit function that have been considered in western Atlantic bluefin tuna stock assessments since 1997: (1) the Beverton-Holt model without a regime shift in the time series and (2) a 2-line model or hockey stock stock-recruit function where there has been a shift in the environmental regime affecting recruitment in the 1970s. The same steps for computing probabilities for structurally different models as outlined in McAllister *et al*. (2000a, b) are outlined below for the extended application to western Atlantic bluefin tuna.

Step 1: Identify alternative functional forms of the stock-recruit (SR) relationship

In the 1998 and subsequent assessments of western Atlantic bluefin tuna, plots and analyses of stock-recruit data and the fitted stock-recruit functions indicated that the empirical deviations from the stock-recruit function were autocorrelated (e.g., ICCAT 1999; ICCAT 2010). Thus, S-R deviates in projections have been modeled in these assessments to be autocorrelated with a 1-year time lag. As in McAllister *et al*. (2000b), I considered auto-correlation in S-R deviates and time-dependency in the magnitude of the variance for the deviates from the hypothesized stock-recruit relationship.

Beverton-Holt (BH) Stock-Recruit Model (after Francis 1992):

$$\hat{R}_{y+1} = \frac{S_y}{\alpha + \beta S_y} \exp\left(\varepsilon_y\right) \tag{2}$$

where:

$\hat{R}_{y+1}$, is the predicted number of recruits of age 1 in year y + 1.

S_y is the spawner biomass in year y.

$\alpha = \omega\,(1\text{-}h)\,/\,(4\,h)$

$\beta = \omega\,(5\,h - 1)\,/\,(4\,h\,B_0)$

ω is spawner biomass per recruit (fixed at 0.528 tons based on life history parameters for the western stock in ICCAT (2010)), *h* is steepness, the proportion of recruitment at virgin stock size, B_0, that results when $S = 0.2\,B_0$.

$$\varepsilon_y = \rho\varepsilon_{y-1} + X_y\sqrt{1-\rho^2} \tag{3}$$

ρ is the first order autocorrelation coefficient of ε_y where ε_y is normally distributed with mean 0 and a SD of σ_R. In some of the analyses, ρ was set at the value applied in the 2010 assessment, which applied 0.52 in both the Beverton-Holt and 2 line models. X_y is a normal random variable with mean 0 and SD = σ_R. Further below I describe how the parameters for this stock-recruit model and the other models were estimated.

<u>2 Line (2L) Stock-Recruit Model (ICCAT 1999)</u>

$$\hat{R}_{y+1} = \begin{cases} \overline{R} \times \exp\left(\varepsilon_y\right) \text{ if } S_y > S_{\inf} \\ \phi\, S_y \times \exp\left(\varepsilon_y\right) \text{ if } S_y \le S_{\inf} \end{cases} \tag{4}$$

where

$\overline{R}$ is the median recruitment when S_y is greater than the inflection point in the 2 line stock-recruit function, S_{inf},

$\phi = \overline{R}\,/\,S_{\text{inf}}$

ε_y is as described for the Beverton-Holt Stock-recruit function (Equation 3).

S_{inf} was determined according to the protocol in the 2010 assessment. This was to take the average of the six lowest values for S_y in the time series, the values from 1990-1995, i.e., 12,640 tons.

Step 2: Formulate a plausible hypothesis for there to have been a regime shift (RS) that altered considerably the form of the stock recruit function.

I considered the year, 1977 as the year in which the regime shift occurred because 1977 was the first year in the time series in which the estimated value for recruitment dropped below the long-term average, i.e., between 1976 and 1977 recruitment went from 1.17 to 0.97 of the long term average value between 1970 and 2010.

Step 3: Identify the alternative stock-recruit models.

As mentioned above, I focus mainly on the two alternative hypotheses on the structural form of the stock-recruit relationship for western bluefin tuna that have been considered in ICCAT's assessment of western Atlantic bluefin tuna: (i) Beverton-Holt and no regime shift; (2) 2 line and no regime shift. In the absence of any scientific arguments about the relative credibility of each structural alternative, the prior probabilities for each of the two alternative models, $p(m_j)$ was set at 0.5.

Step 4: Identify parameters to be estimated in the alternative stock-recruit models identified in steps 1-3.

Hypothesis 1: Beverton-Holt, no regime shift: The estimated parameters were

B_0 over the period 1970 and onwards,
h,
σ_R for ε_y before and after the year of the regime shift (YS) and
ε_y the annual recruitment deviates.

Hypothesis 2: 2 line, regime shift: The estimated parameters were

$\bar{R}_{<YS}$,
$\bar{R}_{\geq YS}$,
$\sigma_{R,1}$ for ε_y before YS,
$\sigma_{R,2}$ for ε_y after and including YS, and
ε_y the annual recruitment deviates.

For the year of the regime shift (YS) onwards, $S_{inf>YS}$ is assumed to be the average observed spawner biomass in years 1989-1993 (ICCAT 2010). The same slope, ϕ, is assumed below the inflection point for both regimes. Thus, for years before YS, $S_{\inf <YS} = \bar{R}_{<YS} / \phi$ and $\phi = \bar{R}_{\geq YS} / S_{\inf \geq YS}$.

Step 5: Define prior probability density (pdfs) functions for the parameters to be estimated in each of the alternative models.

Relatively non-informative priors were identified for each estimated parameter. The priors were chosen such that they were expected to have the lowest possible effect on the Bayes factors for different models and to allow the data to speak for themselves as much as possible. These are somewhat different from the priors applied in McAllister *et al*. (2000b) and the sensitivity of results to the application of different priors is evaluated further below.

The reference case priors for the two alternative models were as follows:

<u>B-H, no RS</u>

B_0 ~ Uniform(log(10,000t), log(5,000,000t))
h~Uniform(0.21, 0.99)
$\ln(\sigma_R)$ ~ Uniform(ln(0.1), ln(1))

<u>2-L RS</u>

$\ln(\sigma_{R,b})$ ~ Uniform(ln(0.1), ln(1)), b = 1, 2.
$\bar{R}$ ~ Uniform(log(10,000), log(5,000,000))

where b signifies the regime.

Step 6: Define the probability model of the data. A 1-year lag auto-regressive stock-recruit function implies that each observed value for recruitment, R_y, is lognormally distributed about the value for recruitment, $\hat{R}'_{j,y}$, predicted by the product of the recruitment given by the deterministic stock recruit model, *j*, in year y and the one-year-lag lognormal autoregressive term:

$$R_{j,y} \sim LogNormal\left(\hat{R}'_{j,y}, \left(1-\rho^2\right)\sigma_R^2\right) \quad (5)$$

where

$$\hat{R}'_{j,y} = \hat{R}^d_{j,y} \exp\left(\rho\varepsilon_{y-2}\right)$$

Note that the annual deviate, ε_{y-1}, that is applied to age 1 recruits, R_y, is referenced to the spawner biomass that produced that recruitment in the previous year (S_{y-1}). Thus, for a one year lagged autoregressive process, the deviate ε_{y-1} is modeled to be correlated to the deviate, ε_{y-2}.

and

$\hat{R}^d_{j,y}$ is the deterministic value for recruitment of age 1 fish in year y, predicted by S_{y-1} and the stock-recruit model parameters. For example, for the Beverton-Holt model (dropping the subscript j),

$$\hat{R}^d_y = \frac{S_{y-1}}{\alpha + \beta S_{y-1}}$$

and $\hat{R}'_{j,y}$ (dropping the "j" term) is given by:

$$\hat{R}'_y = \frac{S_{y-1}}{\alpha + \beta S_{y-1}} \exp\left(\rho\varepsilon_{y-2}\right) \quad (6)$$

For the first year in the time series with an age 1 recruitment observation, e.g., 1971, the value for ε_{y-2} (i.e., ε_{1969}) was fixed at 0. ε_{y-2} for recruitment "observations" in years from 1972 onwards was given by:

$$\varepsilon_{y-2} = \ln\left(R_{y-1}\right) - \ln\left(\hat{R}^d_{y-1}\right) \quad (7)$$

The likelihood function of the set of recruitment observations, $\underline{R}$, was formulated as follows:

$$L\left(\underline{R} \mid \theta_j\right) = \prod_{y=y_i}^{y_f} \frac{1}{R_y\sqrt{2\pi\sigma^2_{R,b}\left(1-\rho^2\right)}} \exp\left(-\frac{\left[\ln\left(R_y \big/ \hat{R}'_{j,y}\right)\right]^2}{2\sigma^2_{R,b}\left(1-\rho^2\right)}\right) \quad (8)$$

where y_i and y_f are the initial and final years of the stock-recruit data time series.

Step 7: Carry out importance sampling.

See McAllister *et al*. (2000b) and McAllister and Ianelli (1997) for details. This is done to estimate marginal posterior probabilities for quantities of interest when the model of interest contains several estimated parameters. Draws are taken from a pre-specified density function called the importance function, which is constructed to be as similar as possible to the posterior density function of interest. The importance function applied for each estimation was the joint prior probability density function (pdf). Up to about 2,000,000 draws were taken from the importance function to obtain highly precise results (less than 10 minutes of computing time using Visual Basic 6.0 programming language and a 2.2 gigahertz notebook PC). A useful diagnostic to test whether enough importance sampling has been done to compute Bayes factor for each model is to monitor the coefficient of variation in the average importance weight:

$$CV\left(\overline{w}\left(\theta_j\right)\right)=\frac{\sqrt{\frac{1}{n_j}\sum_{k=1}^{n_j} w(\theta_{j,k})^2-\frac{1}{n_j^2}\left(\sum_{k=1}^{n_j} w(\theta_{j,k})\right)^2}}{n_j^{-0.5}\sum_{i=1}^{n_j} w(\theta_{j,i})} \tag{9}$$

where n_j is the number of draws taken for model *j* and

$$w\left(\theta_{j,k}\right)=\frac{P\left(data \mid \theta_{j,k}\right)p\left(\theta_{j,k}\right)}{h\left(\theta_{j,k}\right)} \tag{10}$$

where $h\left(\theta_{j,k}\right)$ is the value of the importance function (McAllister *et al*. 2000b) evaluated at the values for the parameters in parameter vector $\theta_{j,k}$ for model j.

This provides an approximation of the expected coefficient of variation (CV) in the marginal posterior probabilities computed for each alternative model. I applied, as a rule of thumb, the following stopping rule for importance sampling: the CV in the average importance weight for a given model should be less than 0.05 before stopping the importance sampling.

Step 8: Compute Bayes factor for each alternative model representing each joint hypothesis was given by (derived from McAllister et al. 2000a):

$$BF(m_j, m_{ref}) = \frac{\frac{1}{n_j}\sum_{k=1}^{n_j} w(\theta_{k,j})}{\frac{1}{n_{ref}}\sum_{i=1}^{n_{ref}} w(\theta_{ref,i})} \tag{10}$$

where m_j is a fully specified stock-recruit model including the assumptions about variance, autocorrelation, the presence of a regime shift and the value for the lag 1 autocorrelation term, m_{ref} is a model to which m_j is being compared, and n_j and n_{ref} are the number of draws taken from the importance density function used in importance sampling for model *j* and the reference model, respectively.

Step 9. Evaluate the sensitivity of results to different statistical assumptions. While it is common for a fixed pair of alternative models to be compared, it is common for there to be numerous somewhat arbitrary features to the model that could potentially influence the weightings obtained. I have made several comparisons between models with different assumptions to evaluate the robustness of results to the making of different statistical assumptions within the stock-recruit models compared. These are detailed in **Table 2**.

Reference case model settings

A chief aim of the analysis is to illustrate how to evaluate the credibility of the two alternative stock-recruit functions given the data used in ICCAT stock assessments (ICCAT 2010). The stock-recruit data included the number of age 1 recruits and spawner biomass from 1970-2006. The presumed year for the regime shift was set to the one considered in the latest stock assessment (i.e., the 1976 cohort) (ICCAT 2010). Autocorrelation in lag 1 autoregressive deviates from both stock recruit functions was applied with ρ set at 0.52 for both models as in ICCAT (2010). For the regime shift model, the variance in stock-recruit function deviates was assumed to be different between the two regimes. The priors for the location parameters for average unfished recruitment, i.e.,

R_{bar} in the 2 line model, and R_0 in the Beverton-Holt model, were made uniform on the natural logarithm of values for these parameters. This was to make the priors as similar as possible between the two models. The prior for σ_R was made uniform on the natural logarithm of σ_R, with the range of plausible values between 0.1 and 1, so that when a second σ_R parameter was estimated for the 2^{nd} regime, the effect of adding one additional parameter on Bayes factors for the two models was minimized, except in so far as it permitted a better fit to the data.

Evaluation of the sensitivity of Bayes factors to different model settings

The autocorrelation coefficient, ρ, computed at lag 1 for the stock-recruit deviates from the fit of the B-H model to the 1970-2006 data was 0.17. This was not significantly different from zero. In contrast the value of 0.52 was applied in ICCAT (2010). I thus evaluated the credibility of the B-H with the autocorrelation coefficient set at 0 and compared this model with the 2-L model with ρ also set to zero, and a few additional variants.

It is of interest to compare results obtained in the previous analyses (McAllister *et al*. 2000a; McAllister *et al*. 2000b) using S-R data for years included in those analyses. I've thus included two sets of comparisons. One uses the S–R up to 1993 from the 2010 stock assessment. The other uses the stock-recruit data up to 1993 that were used in McAllister *et al*. (2000b).

It is of interest to evaluate the effects of how successive data points impact the computation of Bayes factors. The effects of excluding the very large cohort in 2003, excluding 2006, and including the 2007 and 2008 datapoints were thus evaluated.

Other stock-recruit datasets have been formulated for western Atlantic bluefin tuna using catch-age methods (Porch *et al*. 2001; Taylor *et al*. 2011). Since a stock recruit-function was not included in Porch *et al*. (2001) catch-age estimation, the estimates of recruitment and spawning stock biomass are not influenced by a pre-existing stock-recruit function as they may be in Taylor *et al*. (2011). The alternative stock-recruit dataset for the years 1960-1998 in Porch *et al*. (2001) was thus analyzed.

The two alternative stock-recruit models in ICCAT assessments of western Atlantic bluefin tuna are a small subset of potentially plausible stock-recruit models. To meet the requirements of a minimal two factorial evaluation, I evaluated two additional model variants. Marginal posteriors were thus computed also for the Beverton-Holt, regime shift and 2 Line no regime shift model variants.

To minimize the effects of priors on Bayes factors, I'd applied in this paper uniform on log R_0 and uniform on log R_{bar} priors and a maximum value for σ_R of 1 in the uniform on log prior for σ_R. To evaluate the potential effects on Bayes factors of different formulations of priors for model parameters, alternative priors for R_{bar}, and R_0 in the 2 line and B-H models and a wider prior for the prior on σ_R were evaluated.

RESULTS

For those not familiar with probability definitions, the following are provided. A joint probability is the probability that a set of two or more non-mutually exclusive hypotheses is correct (i.e., provides an accurate representation of nature or what has actually happened). An example is the probability that both the Beverton-Holt stock-recruit function and no regime shift hypothesis are correct. A marginal probability is the probability that one particular hypothesis is correct, accounting for uncertainty across all of the other mutually exclusive hypotheses of interest, for example, the probability that the Beverton-Holt model is correct, integrated across the regime shift and non-regime shift hypotheses. A conditional probability is the probability that one hypothesis or set of hypotheses is correct given that some particular condition or other set of hypotheses is correct. An example is the probability that the Beverton-Holt model is correct given that there has been a regime shift. Posterior probabilities are shown below where there are two or more different models being compared.

Bayes factor for a given model represents the ratio of the total probability of the data for that model to that for some reference model. This conveys the same information as the marginal posterior probability for a given model given the data, except that it does not include the prior probabilities for the different models. Bayes factors are shown in some instances below to make it easier to make comparisons between two or more alternative models with regard to their credibility based on the available data.

In the first set of results, I have kept the models compared the same as those compared in McAllister *et al*. (2001b) but used priors that have reduced effects on Bayes factors, a regime shift year of 1977 instead of 1981, and the data for 1970-2006 from the 2010 ICCAT assessment (rather than only 1970-1993). The joint posteriors and Bayes factors for the four recruitment hypotheses in **Table 3a** indicate that the 2-line recruitment model hypotheses are four or more times less credible than the B-H models. According to the marginal posteriors (**Table 3b**), the Beverton-Holt functional form is more probable than 2-line (0.9 vs. 0.1). The no regime shift hypothesis is equally probable as the regime shift hypothesis when the Beverton-Holt and 2 line models are both considered together (0.49 for the regime shift and 0.51 for no regime shift). According to the conditional posterior probabilities (**Table 3c**), no regime shift is slightly more probable than regime shift under the Beverton-Holt model but much less probable under the 2 line model (**Figure 1**). Given these results, the two-line regime shift hypothesis, though less likely, cannot be discarded since its Bayes factor when compared to the Beverton-Holt no regime shift model is not very small, e.g., i.e., not less than 0.01. However, the data tend to support the Beverton-Holt function over the 2-line functional form, even with autocorrelation included and residual variance estimated separately for years before and after 1977. The large recruitments in the early part of the time series are not fitted well by any of the four models (**Figure 1**).

Table 3a shows that the normalized (priors * Likelihood) evaluated at the posterior mode for each of four alternative models are very similar. This number indicates how well the model with the best point estimate of the parameters fits the data, discounted by the log prior density at the posterior model. If a more complex model provides only a marginally better fit to the data, the statistic is not expected to change very much as shown. Normalizing the product of prior and likelihood at the mode shows relatively little difference between these naïve posterior probabilities. The statistic however ignores the variation in goodness of fit as a result of parameter uncertainty within each model. In contrast, the marginal posterior probability for each model (**Table 3a**) integrates the posterior distribution across all parameter values. The marginal posterior probability and Bayes factor are more appropriate than model selection criteria that are based on the maximum likelihood estimate or mode (e.g., AIC or BIC) because they account for parameter uncertainty in each model (Kass and Raftery 1995) whereas AIC and BIC do not. They will also tend to favor the most parsimonious models (models with fewer parameters) which fit the data well.

The final step in the provision of stock assessment results according to the Bayesian approach is to present tables from model projections for each alternative model/hypothesis and policy evaluated along with the probability values for each alternative model/hypothesis. To keep the illustration simple I have provided tables with indications of stock rebuilding potential based on results from the fitted stock-recruit models. Posterior modal values are provided for recruits per ton of spawner biomass (R/S) for each model at spawner biomass (S) equal to 0 and S equal to the inflection points of the two-line models (S_{inf}), i.e., the average observed spawner biomass for 1990-1995 for the second regime and the value given by the same slope and estimate of R_{bar} for the first regime (**Table 4**). Under the regime shift hypothesis, R/S at S=0 under 2 line model was about a third of that under the Beverton-Holt model. Under the no regime shift hypothesis, R/S at S=0 was 22% higher under the 2 line model than that under the Beverton-Holt model. Under the no regime shift hypothesis, R/S at S_{inf} under the 2 line model was 41% higher than that under the Beverton-Holt model. Under the regime shift model, R/S at S_{inf} was 0.5% lower under the 2 Line model than that under the Beverton-Holt model in the second regime. The value for steepness was considerably lower under the no regime shift model (0.47 (0.11)) compared to that under the regime shift model (0.72 (0.19)) and considerably more precise under the no regime shift hypothesis (posterior CVs shown in parenthesis after the posterior modal values).

A key comparison is between the ratio of recruits per spawner (R/S) for the 2 line regime shift model and the Beverton-Holt no regime shift model since these are the two alternatives considered in the provision of management advice. The R/S at S_{inf} under the 2 line regime shift model is slightly higher (7%) than that under the Beverton-Holt no regime shift model (**Table 4**). These results indicate that the resilience to exploitation as indexed by R/S predicted under each alternative model depends strongly on whether a regime shift has been hypothesized and on the form of the stock-recruit function.

A second key component of rebuilding potential is the apparent amount of rebuilding that is required given that stock biomass is considered to be low and in need of rebuilding as it is in recent assessments of western Atlantic bluefin tuna (ICCAT 2010). It is clear that with the mean maximum recruitment under the regime shift 2 line model (78,000) is far less than the average unfished recruitment expected under the Beverton-Holt model and no regime shift (399,000) (**Table 4**). The corresponding values for B_0 reflect this large variation between the different models (**Table 4**). These recruitment and B_0 reference points could be expected to offer a rough approximation of the relative levels of stock biomass at the maximum sustainable yield (B_{msy}) under these two alternative recruitment hypotheses, since B_{msy} is commonly in the range of about 0.2 to 0.4 B_0 for most exploited fish stocks.

To check the assumption of autocorrelated residuals, the marginal posterior probabilities for the different models under no autocorrelation versus the reference case level of auto correlation were computed. The marginal posterior for autocorrelation integrated across the four joint hypotheses on recruitment was 0.34, compared to 0.66 for no autocorrelation (**Table 5**). The favoring of autocorrelation was strong only for the 2 line no regime shift model which has much more strongly pronounced patterns in autocorrelation of stock-recruit deviates than the other models (**Figure 1**). Not surprisingly, if no autocorrelation was assumed, then the regime shift hypothesis obtained slightly higher marginal probability with 0.70, compared to 0.51 for no regime shift. The probability of the Beverton-Holt model over the 2 line model was not as high when no autocorrelation was applied, i.e., 0.62 compared to when an autocorrelation coefficient of 0.52 was assumed (**Table 5**).

To check the assumption that the variance in residuals was different for years before 1977, the marginal posterior probabilities were computed for the models under different versus same variance. Under separate variances, the standard deviation in log recruitment deviates was about twice as high in the first regime compared to the second regime (**Table 6a**). The marginal posterior for no difference in variance integrated across the four joint hypotheses on recruitment however was 0.46, compared to 0.54 for different variances (**Table 6**). The small number of data points for the first regime (only 7) likely prevented stronger support for there being a difference in the variance. The Bayes posteriors for the Beverton Holt and 2 Line models under regime shift and no regime shift when the same variance was assumed were very similar to those obtained when the variance was assumed to be different between the regimes. The favoring of the same variance assumption occurs mainly because of there being relatively few data points for the two separate time periods. The same variance assumption thus did not affect the other marginal and conditional posteriors regarding the 2 line and Beverton-Holt models (**Table 6**).

The stock-recruit data used in this analysis were different from those used in the initial analysis (McAllister *et al*. 2001) since then only data up to 1993 were available and different growth curve and age-slicing inputs were applied to produce the catch-age data. It is thus of interest to see whether results obtained with the updated analysis compare with those obtained by the previous analysis. In one analysis I used the same stock-recruit data, 1970-1993, as were applied in McAllister *et al*. (2000b). In a second analysis, I used the same years of stock-recruit data but instead those provided in the 2010 assessment (ICCAT 2010). When the same data were used as in McAllister *et al*. (2000b), considerably smaller differences in Bayes factors were obtained than those in McAllister *et al*. (2000b). With the updated analysis applied to the same data as in McAllister *et al*. (2000b), the Beverton-Holt no regime shift model was only about twice as likely as the 2 line regime shift model rather than about 70 times more likely in McAllister *et al*. (2000b) (**Table 7**). The much lower Bayes factor for the Beverton-Holt no regime shift model resulted mainly from the different priors that were applied for model parameters, i.e., the uniform on log priors applied to R_0 and R_{bar} in the Beverton-Holt and 2 line this time, rather than the uniform priors applied last time to B_0 and R_{bar} (see results below and Discussion section for more on this). The previous analysis applied a regime shift year of 1981 after the practice in earlier assessments; whereas this analysis applied the more recently applied year of 1977. Computations showed that this change in year of regime shift had very little effect on the Bayes factors. When the models were fitted to the same time segment of stock-recruit data from the 2010 assessment, the Bayes factors switch around to favor the 2 line regime shift model by a factor of 48. This difference can be attributed to the difference in the configuration of the stock-recruit data between those used in McAllister *et al*. (2001) and those obtained from ICCAT (2010). But when the full time series was used, i.e., 1970-2006, Bayes factor switched around in favor of the Beverton-Holt, no regime shift model by a factor of 4.8.

Table 8 shows how Bayes factors for the two alternative recruitment model hypotheses (BH-NRS vs. 2L RS) vary with different assumptions about autocorrelation, whether variance is estimated separately for the two regimes, the number of years of data included in the analysis, a different stock-recruit data set and with different specifications for the priors for the estimated parameters. With the autocorrelation coefficient set to zero, the 2L RS model obtained slightly higher Bayes factors when compared to a BH NRS model with the autocorrelation coefficient set to zero and either variance the same or different between the two regimes. With the variance set to be equal between the two regimes, the 2L RS model is about five times less credible than the BH model, indicating that the presumption of positive autocorrelation in the 2 L model makes it less credible against the BH NRS model (**Table 8**).

When the high 2003 recruitment point is removed from the analysis, the Bayes factor for the BH NRS model drops from 4.8 to 0.7, indicating that this high recruitment produced at low stock size contributes support for the BH NRS model. When applied to stock-recruit data from 1970 to 2005 and presuming a regime shift in 1977, the

Bayes factors were 1:1 and favored neither hypothesis. When the 2006, 2007 and 2008 recruitment data points were included, Bayes factors favoured the Beverton-Holt model by factors of 4.8, 4.1 and 6.1, respectively (e.g., the probability of the data was 4.1-6.1 times higher for the Beverton-Holt model compared to the two-line regime shift model). The posterior probabilities for the Beverton-Holt model computed using uninformative prior probabilities, for years 2000-2008 are shown in (Fig. 2). The posteriors switch from values close to zero up to 2003, to 30% in 2004, and to about 80% in 2006 and subsequently. The high sensitivity of the posterior to new data can potentially be dampened by square root or double square root transformations as shown in Fig. 2. When Bayes factors were computed for the 1960-1998 stock-recruit data set provided in SCRS 2001/52, Bayes factors supported the Beverton-Holt NRS model by a factor of 2.7.

As indicated above, the formulation of the priors for model parameters strongly impacted Bayes factors. When a uniform prior was placed on R_0 in the Beverton-Holt NRS model, Bayes factor increased from 4.8 to 55. When a uniform prior on $\log(B_0)$ was applied instead of the uniform on $\mathrm{Log}(R_0)$ prior, Bayes factor increased from 4.8 to 5.8. When the uniform prior for σ_R was extended from uniform on (0.1,1) to uniform on (0.1, 2) Bayes factor for the BH-NRS model increased from 4.8 to 5.6. When the uniform prior for σ_R was extended from uniform on (0.1,1) to uniform on (0.1, 2) and uniform on R_{bar} and uniform on B_0 priors were applied, Bayes factor for the BH-NRS model increased from 4.8 to 67. These results show that the most important formulation for priors on location parameters (e.g., R_0, R_{bar}, B_0) is to apply priors to the log transformed values of parameters, rather than to formulate prior density functions for the non-log transformed values. For parameters that are scale parameters such as σ_R, or that range between close to zero and close to 1, the formulation of the prior has very little impact on Bayes factors.

DISCUSSION

This paper offers an update in the analysis of the empirical credibility of the western Atlantic bluefin tuna stock recruit model hypotheses. It uses stock-recruit data for this stock updated from the 2010 assessment to further illustrate a methodology that could be applied in the current and future assessments of bluefin tuna and other ICCAT fish stock assessments to help deal with structural uncertainties in the provision of management advice. Several recent works have recognized the importance of providing a systematic and scientifically grounded approach to providing empirically based weights for structurally different stock assessment models (Butterworth *et al*. 1996; Punt and Hilborn 1997; McAllister and Kirkwood 1998; McAllister *et al*. 1999; Parma 2000; Punt *et al*. 2000; McAllister and Kirchner 2002). Importance sampling has been one of the statistical methods developed for this purpose (Kass and Raftery 1995; McAllister and Kirkwood 1998; McAllister *et al*. 2000; McAllister and Kirchner 2002). As mentioned above, the approach is compatible with the main stock assessment methodologies applied in ICCAT ADAPT VPA methods and surplus production modeling (McAllister *et al*. 2000). For example, the Bayesian methods can be applied to the stock and recruit data points produced by VPA or catch-age methods to indicate the weight of evidence in support of the different stock-recruit model assumptions. These weights could potentially be helpful to present when presenting the results of policy projections which were based on the different stock-recruit model assumptions.

This paper further demonstrates how importance sampling can be applied to compute empirical weightings for stock-recruit models with different assumptions about the recruitment of western Atlantic bluefin tuna. The paper extends the work in McAllister *et al*. (2000a, b) to further evaluate stock-recruit model assumptions. Autocorrelation in stock-recruit residuals was incorporated as in McAllister *et al*. (2000b). In contrast to McAllister *et al*. (2000b), Bayes factors were only mildly sensitive to autocorrelation in recruitment deviates except for the 2 line, no regime shift model. A difference in the variances of these deviates was also considered before and after 1977. Steepness in the Beverton Holt model was not estimated separately under the regime shift hypothesis for the two successive regimes mainly due to the paucity of data with which to estimate steepness for the 1st regime and the failure to find any support for the estimation of a second steepness parameter in McAllister *et al*. (2000b).

The findings of McAllister *et al*. (2000a,b) in favour of the Beverton-Holt no regime shift model were found in the current analysis to be too strong, due to the use of naïve priors for the location parameters, B_0 and R_{bar}. With adjustments to minimize the impact of the priors on Bayes factors, the stock-recruit data still tended to favour the Beverton-Holt formulations over the 2 line formulations (i.e., about 7 to 1). Bayes factor, i.e., 4.8, for the Beverton-Holt no regime shift model when the full time series, 1970-2006, was used was only mildly in favour of this model over the 2 line regime shift model. The 2 line formulations gave mostly fairly similar values this time for recruits per spawner compared the Beverton Holt ones over the current range of spawning stock sizes.

The lower empirical weights for the 2 line formulations would suggest that the 2 line formulations be treated with greater skepticism in future stock assessments.

Using the 1970-2006 data from the 2010 assessment, lead to Bayes factors ranging 2:1 against to 67:1 in favor of the Beverton-Holt no regime shift model hypothesis over the 2 line regime shift hypothesis when different assumptions about autocorrelation, the variance in recruitment deviates, and priors for the estimated parameters were considered. Under their reference case settings for the two models, McAllister *et al*. (2000b) obtained markedly higher Bayes factors in favor of the Beverton-Holt no regime shift hypothesis when compared to the 2 line no regime shift hypothesis. This is mainly a result of the different priors applied to R_{bar} and B_0 between McAllister *et al*. (2000b) and in this paper and the different configuration of the stock-recruit data, for years since 1977. In one of the sensitivity runs in this paper, we applied priors very similar to those in McAllister *et al*. (2000b) and obtained Bayes factors similarly as large for the Beverton-Holt no regime shift model as obtained in McAllister *et al*. (2000b), i.e., the new analysis gave 67 compared to about 70 in the previous analysis. The uniform on log B_0, uniform on log R_0, and uniform on log R_{bar} priors have less between-model influence on Bayes factors because these priors have $\log(R_{bar})$ and $\log(B_0)$ values in the denominator of the prior density function rather than just B_0 and R_{bar} values in the denominator of the uniform on B_0 and R_{bar} priors. Also over the range of support from the data, the values of R_{bar} under the 2 line models tend to be a few times higher than the B_0 values under the Beverton Holt models (Table 4). When a regime shift is considered, the straight uniform priors on B_0 or R_{bar} will thus tend to heavily down-weight the regime shift variants and the two line models. With log B_0, log R_0 or log R_{bar} in the denominator of the prior, the differential effects on Bayes factors for the different models are very low and differences in model fits to the data dominate the values obtained for Bayes factors. It was also an improvement to reparameterize the Beverton-Holt model from B_0 to R_0. The key location parameters in the Beverton-Holt and 2 line models were thus directly comparable, in the same units and having similar ranges of values. The uniform on log priors for these parameters thus minimized their influence on the values obtained for Bayes factors.

In addition, the range for the priors for the standard deviation in recruitment deviates (σ_R) was 1.9 in McAllister *et al*. (2000b) but on only 0.9 in this analysis. This prior with the wider interval will more strongly influence Bayes factors and down-weight the regime shift model with the estimation of the extra σ_R parameter in the regime shift model. The effect of wider priors on σ_R increased the Bayes factor in favor of the Beverton-Holt no regime shift model from 4.8 with the uniform priors on log R_{bar} and log R_0 and maximum on σ_R of 1 to a Bayes factor of 5.6. When both the wider priors on σ_R and uniform on R_{bar} and B_0 priors were applied, the Bayes factors in favour of the Beverton-Holt function became highly exaggerated at 67. This shows that inappropriate choice of priors for two or more model parameters can very strongly influence Bayes factors and priors should thus be chosen very judiciously to avoid having them influence Bayes factors. In this analysis, the use of uniform on log priors, reparameterization to make the location parameters more comparable in their range of plausible values, and use of priors that have narrower ranges thus tend to allow the data to speak for themselves more about the credibility of the different models.

The data used in the computation of Bayes factors in this analysis also strongly impacted the Bayes factor values obtained for the alternative stock-recruit models. When the initial set of years of data available for analysis were evaluated using the data from ICCAT (2010), i.e., 1970-1993, Bayes factors favoured the 2 line no regime shift model by a factor of 48. In contrast, the same set of years of data available from the 1998 stock assessment (ICCAT 1999), gave a Bayes factor of 2.1 in favour of the Beverton-Holt no regime shift model. When subsequent years of data were included in the analysis of the ICCAT (2010) assessment data, Bayes factor switched over in favour of the Beverton-Holt no regime shift model by a factor of about 5. The very high abundance of the 2002 cohort which was about two thirds of the average value for the large cohorts in the 1970s, strongly reduced the values of Bayes factors in favour of the two line, no regime shift model. The very low cohort in 2005, further shifted Bayes factors in favour of the Beverton-Holt model.

The swings in the values obtained for Bayes factors when the stock-recruit models were fitted to a growing time series of stock-recruit data suggest that when data sets are relatively short, e.g., less than 25 years, as was the case for data up to the mid-1990s, and the variation in spawning stock size has remained low for many years, as in the present analysis, Bayes factors obtained should be interpreted with caution. One-way trip data sets such as the ones analyzed in this paper in which abundance data start out high and progressively decrease have long been pointed out as problematic for parameter estimation in fisheries models (Hilborn and Walters 1992). One-way trip data can be expected to be relatively uninformative for the evaluation of the credibility of structurally different models (McAllister and Kirchner 2002), though in this case, it appears that Bayes factors quite strongly favoured the 2 line model with the models fitted to stock-recruit data 1970-mid-1990s to the early 2000s but then switched over to favour slightly the Beverton-Holt model with subsequent data. Strong swings in Bayes factors from one hypothesis to another suggest that used by itself Bayes factor may be overly sensitive to changes in the configurations of apparently informative data that is actually limited due to poor contrasts in the range of data obtained. Two alternative approaches to reduce the sensitivity of Bayes factors to slight changes in the configurations of datasets and to prevent Bayes factors from giving excessively high weights to one model when data are actually limited, are as follows. Double square root transform Bayes factors will reduce the potential for extremes in the values obtained and reduce the chance of false positives, i.e., placing too much weight on a given model prematurely when the data may be not have sufficient contrast to make strong conclusions. The precision in the likelihood function for the most recently obtained stock-recruit estimates could be reduced, since estimates of abundance for the most recent years are typically the most poorly determined. This would further reduce the sensitivity of Bayes factors to the most recently added data points and further reduce the potential for large changes in Bayes factors from the addition of the most recent years of stock-recruit data which tend to be the most imprecisely estimated points in the series and most prone to updating with updates in the stock assessment. Further research is needed to explore and develop this latter approach for practical application.

Posterior probabilities this time did not favor the autocorrelation structures, except for the 2 Line, no regime shift model. If no autocorrelation was assumed, the regime shift formulations became favored slightly over no regime shift ones (about 60:40). This also indicates that the regime shift and no regime shift hypotheses should not be ignored in the decision analyses of alternative stock rebuilding policies, despite the considerably more optimistic interpretations about the level of rebuilding effort required under the regime shift formulations.

Although the results did not strongly favor keeping the variances (and steepness in the Beverton-Holt model) separate for the earlier and later part of the time series, these assumptions were not strongly refuted by the data. Keeping the variances separate, especially to allow a larger variance for the earlier part of the time series is sensible, since these earlier data points may be less reliable than those for the latter part of the time series and a larger variance down-weights these earlier data in the estimation. While keeping the steepness values separate for the early and latter parts of the time series may seem sensible, as it is plausible that steepness can change depending on oceanographic regime, there are too few data in the early years to enable estimation of steepness. Model parsimony rationale would thus support assuming that either steepness or alpha remains the same for the first and second regime, should a regime shift hypothesis be considered for western Atlantic bluefin tuna.

Fitting the models to a stock-recruit data set derived from a catch-age methodology that did not apply within it a stock-recruit function and that went back further in time when recruitment was higher than in the recent past (Porch *et al*. 2001), provided Bayes factors that also favoured the Beverton-Holt no regime shift hypothesis. However, the support was only mildly pronounced. This may be due to the shortness and lack of contrast in the latter part of this time series.

It is noted that the stock-recruit data obtained from ICCAT's assessments remain possibly biased due to migrations of the much larger eastern Atlantic population into the western Atlantic. ICCAT stock assessments have assumed that all Atlantic bluefin tuna captured to the west of 45 degrees west, are of western origin. Recent modeling efforts to account for stock mixing in a spatially structured stock assessment model for eastern and western Atlantic bluefin tuna suggest that the fraction of eastern origin fish in the western Atlantic has varied substantially since the 1950s (Taylor *et al*. 2011). Some of the apparent variation in estimates of western Atlantic bluefin tuna spawning stock size and recruitment since 1970 seen in ICCAT assessments could thus be artifacts of variation in the migrations of the much larger eastern Atlantic bluefin tuna spawning population. It is thus recommended that further efforts be devoted to improving understanding of western Atlantic bluefin tuna recruitment using estimates of recruits and SSB that are not contaminated by eastern stock migrations.

These results demonstrate that, with suitable statistical methodology, stock-recruit data can themselves provide an empirical basis to weight structurally different stock recruit models. This can provide valuable scientific guidance for policy evaluation that has been previously left out of ICCAT stock assessments. Scientists can now provide the guidance that was previously lacking regarding empirical weightings for structurally different stock-recruit models. The particular data analysis could not readily distinguish whether a regime shift had occurred. The updated data analysis again cast doubt on the 2 line model which since the 1998 stock assessment has been given considerable attention by fishery scientists and managers. The data analyses instead tended to support the Beverton-Holt model no regime shift hypothesis which has all along been considerably less optimistic about stock rebuilding requirements than the two line regime shift hypothesis and has been consistently down-weighted by fishery managers, as they then lacked scientific guidance to suggest otherwise.

ACKNOWLEDGMENTS

This work was made possible by a grant from the Pew Charitable Trust. Laurie Kell is thanked for providing the stock-recruit data. Clay Porch, John Walter, Andy Cooper, Mark Maunder, Victor Restrepo, Gerald Scott, Juan Carlos Seijo, and Shana Miller are thanked for helpful comments on earlier versions of the analysis.

BIBLIOGRAPHY

BUTTERWORTH, D.S., Punt, A.E. and Smith, A.D.M. 1996. On plausible hypotheses and their weightings, with implications for selection between variants of the Revised Management Procedure. Rep. int. Whal. Commn 46: 481-491.

HARE, S.R. and Francis, R.C. 1995. Climate change and salmonid production in the northeast Pacific Ocean. p. 357-372. In R.J. Beamish [ed.] Climate change and northern fisher populations. Can. Spec. Publ. Fish. Aquat. Sci. 121.

HILBORN, R., and Walters, C. 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty. Chapman and Hall, New York.

KASS, R.E. and Raftery, A.E. 1995. Bayes factors. J. Am. Stat. Assoc. 90: 773-795.

KING, J.R., McAllister, M.K., Holt, K.R., and Starr, P.J. 2011. Lingcod (*Ophiodon elongates*) stock assessment and yield advice for outside stocks in British Columbia. Canadian Science Advisory Secretariat Research Document 2011/124. 177 p.

ICCAT. 1999. Collective Volume of Scientific Papers. XLIX (2).

ICCAT. 2010. REPORT OF THE 2010 ATLANTIC BLUEFIN TUNA STOCK ASSESSMENT SESSION. 132 p.

MCALLISTER, M.K., and Ianelli, J.N. 1997. Bayesian stock assessment using catch-age data and the Sampling/ Importance Resampling Algorithm. Can. J. of Fish. and Aquat. Sci. 54: 284-300.

MCALLISTER, M.K. and Kirkwood, G.P. 1998. Bayesian stock assessment: a review and example application using the logistic model. ICES J. Mar. Sci. 55: 1031-1060.

MCALLISTER, M.K., Babcock, E.A., and Pikitch, E.K. 2000a. Using Bayesian methods and decision analysis as a rational basis to dealing with stock assessment results while providing management advice. International Commission for the Conservation of Atlantic Tunas, Madrid. Document SCRS/00/34.

MCALLISTER, M.K., Babcock, E.A., and Pikitch, E.K. 2000b. Using Bayesian methods to provide an empirical basis to weight conflicting stock assessment results on stock rebuilding when there is uncertainty over processes affecting future recruitment. ICCAT Col. Vol. Sci. Papers SCRS/00/103.

MCALLISTER, M.K., Starr, P.J., Restrepo, V., and Kirkwood, G.P.. 1999. Formulating quantitative methods to evaluate fishery management systems: what fishery processes should be modeled and what trade-offs should be made? ICES Journal of Marine Science 56: 900-916.

MCALLISTER, M.K. and Kirchner, C.H. 2002. Accounting for structural uncertainty to facilitate precautionary fishery management: illustration with Namibian orange roughy. In "Targets, Thresholds, and the Burden of Proof in Fisheries Management" Mangel, M. ed. Bull. of Mar. Sci., 70(2). 499-540.

PARMA, A.M. 2000. Bayesian approaches to the analysis of uncertainty in the stock assessment of Pacific halibut. In J. M. Berkson, L.L. Kline and D.J. Orth (eds.) Incorporating uncertainty into fisheries models. .

PORCH, C.E., S.C. Turner, and G.P. Scott. 2001. Updated catch-at-age analyses of West Atlantic bluefin tuna 1960-1997. Col. Vol. Sci. Pap. ICCAT 52(3): 1052-1066.

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

PUNT, A.E., Pribac, R., Walker, T.I., Taylor, B.L. and Prince, J.D. 2000. Stock assessment of school shark, *Galeorhinus galeus*, based on a spatially explicit population dynamics model. Mar. Freshwater Res. 51: 205-20.

RAFTERY, A.E. and Richardson, S. 1996. Model selection via GLIB. *In* Bayesian Biostatistics. pp. 321-353. Ed. by D.A. Berry, and D.K. Dalene. Marcel and Dekker, Inc. 610 pp.

TAYLOR, N.G., McAllister, M.K., Lawson, G.L., Carruthers, T., Block, B.A. 2011. Atlantic Bluefin Tuna: A Novel Multistock Spatial Model for Assessing Population Biomass. PLoS ONE 6(12): 1-10.

Table 1. Stock –Recruit data for Western Atlantic Bluefin tuna from the 1998 assessment (digitized from Figure 38 of ICCAT (1999)) and from the 2010 assessment (source: Laurie Kell, ICCAT, Madrid).

Year	SSB from 1998 assessment	Age 1 recruits from 1998 assessment	SSB from 2010 assessment	Age 1 recruits from 2010 assessment
1970	49051	277190	51038	363536
1971	44356	244160	50782	321829
1972	44148	171620	51189	278125
1973	42590	499180	51460	150433
1974	44006	146140	46167	465130
1975	38152	138680	40953	163672
1976	37363	89715	36091	134434
1977	32162	56721	30952	111394
1978	29845	84324	27648	94041
1979	23919	69025	24448	98822
1980	23703	65117	22152	80273
1981	21390	63308	19019	79471
1982	20807	100600	17874	80865
1983	20016	71241	17109	102330
1984	16834	79292	16232	91014
1985	13724	89299	14612	96051
1986	12932	67782	14942	99517
1987	11483	89446	14310	86058
1988	10404	48340	14168	130704
1989	9244.6	69984	13714	112941
1990	8880.1	95498	13115	107223
1991	7728.9	50475	12784	85225
1992	6929.7	83861	12362	71585
1993	6935.8	32880	12468	66086
1994	NA	NA	12306	73885
1995	NA	NA	12756	97900
1996	NA	NA	13717	82474
1997	NA	NA	14535	69102
1998	NA	NA	14646	77109
1999	NA	NA	13949	72349
2000	NA	NA	13753	67294
2001	NA	NA	13131	75986
2002	NA	NA	12508	56892
2003	NA	NA	12016	60150
2004	NA	NA	12435	207191
2005	NA	NA	12871	76543
2006	NA	NA	12864	28708
2007	NA	NA	13751	42416
2008	NA	NA	14034	25297
2009	NA	NA	14072	25825

Table 2. List of comparisons between alternative stock-recruit (S-R) models with rationale provided for each comparison.

Descriptor	Variant	Run information
Ref. Reference case runs	Ref.B-H.NRS	Beverton-Holt, no regime shift, 1970-2006 data, auto-correlation coefficient (ρ) for stock-recruit function deviates set at 0.52
	Ref.2-L.RS	Two-line, 1977 regime shift, 1970-2006 data, ρ set at 0.52, variance in stock-recruit deviates also changed in 1977 (ICCAT 2010)
A. Auto-correlation and S-R variance assumptions	A.B-H.NRS.1	Beverton-Holt, 1970-2006, no regime shift, lag-1 auto correlation set at zero
	A.2-L.1	Two line, regime shift, auto-correlation set at zero, variance in S-R deviates changes in 1977, 1970-2006
	A.2-L.RS.2	Two line, regime shift, ρ set at zero, no change in variance in S-R deviates, 1970-2006
	A.2-L.RS.3	Two line, regime shift, ρ set at 0.52, variance in S-R deviates does not change in 1977, 1970-2006
B. Comparisons using SCRS 2000/103 inputs.	B.B-H.NRS.1	Beverton-Holt, no regime shift, S-R data (ICCAT 2010) for 1970-1993 only, ρ = 0.52
	B.2-L.RS.1	Two line, regime shift, S-R data (ICCAT 2010) for 1970-1993 only, ρ = 0.52, variance in S-R deviates changes in 1977
	B.B-H.NRS.2	Beverton-Holt, no regime shift, 1970-1993 data (ICCAT 1998), ρ = 0.52
	B.2-L.RS.2	Two line, regime shift, 1970-1993 data (ICCAT 1998), ρ = 0.52, variance in S-R deviates changes in 1977
	B.2-L.RS.3	Beverton-Holt, no regime shift, 1970-1993 data (ICCAT 1998), ρ = 0.52, uniform on B_0 and max σ_R at 2
	B.2-L.RS.3	Two line, regime shift, 1970-1993 data (ICCAT 1998), ρ = 0.52, variance changes 1977, uniform on B_0 and max σ_R at 2
C. The effects of adding or subtracting data points	C. B-H.NRS.1	Beverton-Holt NRS 1970-2006 but exclude 2003 cohort,
	C.2-L RS.1	2-Line Regime shift cohort, 1970-2006 but exclude 2003
	C. B-H.NRS.2	Beverton-Holt NRS, fitted to 1970-2005 series.
	C.2-L.RS.2	2-Line with regime shift in 1977 fitted to 1970-2005 series.
	C. B-H.NRS.3	Beverton-Holt NRS fitted to 1970-2007 series
	C.2-L RS.3	2-Line Regime shift fitted to 1970- 2007 series
	C. B-H.NRS.4	Beverton-Holt NRS fitted to 1970-2008 series
	C.2-L.RS.4	2-Line Regime shift in1977 fitted to 1970- 2008 series
D. Alternative stock-recruit data	D. B-H.NRS.1	Beverton-Holt NRS fitted to 1960-1998 series
	D.2-L.RS.1	2-Line Regime shift in1977 fitted to 1960- 1998 series
E. B-H regime shift and 2-L no regime shift variants	E.2-L.NRS.1	Two-line, no regime shift, 1970-2006 data, ρ set at 0.52 (ICCAT 2010)
	E.B-H.RS.1	Beverton-Holt, regime shift 1970-2006 data, variance in deviates changed in 1977; ρ set at 0.52 (ICCAT 2010)
F. Alternative priors for R_{bar}, B_0 and σ_R in the 2-L and B-H models	F.B-H.NRS.1	Beverton-Holt as reference case except for the prior for R_0 is uniform on R_0.
	F.2-L.RS.1	2 Line as reference case except for the prior on R_{bar} being uniform on R_{bar}
	F.B-H.NRS.2	Beverton-Holt as reference case except that model is reparameterized with a uniform on $\log(B_0)$ prior.
	Ref.2-L.RS.1	2 Line as reference case with the prior on R_{bar} being uniform on $\log(R_{bar})$
	F.B-H.NRS.3	Beverton-Holt as reference case except for the maximum σ_R is 2 rather than 1
	F.2-L.RS.3	2 Line as reference case except for the maximum σ_R is 2 rather than 1
	F.B-H.NRS.4	Beverton-Holt as reference case except for the prior for R_0 is uniform on R_0 and the maximum σ_R is 2 rather than 1
	F.2-L.RS.4	2 Line as reference case except for the the prior on R_{bar} being uniform on R_{bar} and the maximum σ_R is 2 rather than 1

Table 3. Bayesian posteriors and other statistics for alternative hypotheses on the form of the stock-recruit function and on whether a regime shift has occurred, for western Atlantic bluefin tuna.

a. Joint posteriors, Bayes factors, and log prior * likelihood for the four alternative hypotheses. A statistic (the coefficient of variation in the average of weights from importance sampling) showing numerical stability in the Bayes factors, i.e., CV< 0.05.

	H1: Beverton-Holt, no regime shift	H2: Beverton-Holt, regime shift	H3: 2-line, no regime shift	H4: 2-line, regime shift
Number of estimated parameters	3	5	2	4
Joint Posterior Probability	0.48	0.41	0.0005	0.10
Bayes factors (ref. to H1)	1	0.86	0.001	0.21
ln(prior*LH) at posterior mode	166.175	165.447	166.556	166.216
Normalized(prior*LH)	0.25	0.12	0.37	0.26
CV(average of weights)	0.022	0.022	0.011	0.025

b. Marginal posteriors

P(Regime Shift)	P(no Regime Shift)
0.51	0.49

P(Beverton Holt)	P(2 Line)
0.90	0.10

c. Conditional posteriors

P(Regime shift | BH)	P(no Regime shift | BH)
0.46	0.54

P(Regime shift | 2L)	P(no Regime shift | 2L)
0.995	0.005

Table 4. Median stock rebuilding potential as indexed by recruits per spawner under each of the four hypotheses on recruitment and occurring at S = 0 and $S=S_{inf}$ (12,640t). Proxies for reference case stock levels are indicated by the posterior modal values for average unfished recruitment (R0). The values for steepness are the posterior modal values. The number in parentheses shows the posterior coefficient of variation for steepness. Recruits per spawner is in recruits per ton of spawners.

	Regime shift		No regime shift	
Stock-Recruit Function	Beverton-Holt	2 Line	Beverton-Holt	2 Line
Bayes factors	0.86	0.21	1	0.001
R/S at S=0	19.1	6.2	6.8	8.3
R/S at S=Sinf, 1st regime	5.1	6.2	5.8	8.3
R/S at S=Sinf, 2nd regime	6.3	6.2	NA	NA
B-H mean steepness	0.72 (0.19)	NA	0.47 (0.11)	NA
Mean unfished recruitment (R_0, R_{bar}) 1st regime	292,000	288,000	399,000	104,000
Mean unfished recruitment (R_0, R_{bar}) 2nd regime	106,000	79,000	NA	NA
S_{inf} 1st regime	NA	46,150 t	NA	12,640 t
S_{inf} 2nd regime	NA	12,640 t	NA	NA
B_0 first regime	154,000 t	152,000 t	211,000 t	55,000 t
B_0 second regime	56,000 t	42,000 t	NA	NA
σ_r first regime	0.59	0.54	0.43	0.52
σ_r second regime	0.37	0.39	NA	NA

Table 5. Bayesian marginal posterior probabilities for alternative hypotheses on a. model versions with and without lag one autoregressive autocorrelation in recruitment deviates with the coefficient (ρ) set at the value in the 2010 assessment (0.52), b. whether autocorrelation is present in the stock recruit deviates integrating across the different models, c. the different recruitment hypotheses when no autocorrelation is assumed, d. regime shift versus no regime shift and Beverton-Holt versus 2 line assuming no autocorrelation, and e. the regime shift conditional on the recruitment model.

a. Marginal posteriors

	Autocorrelation	No autocorrelation
BH NRS	0.45	0.55
2L RS	0.12	0.88
BH RS	0.40	0.60
2L NRS	0.9995	0.0005

b. Marginal posteriors

P(autocorrelation)	P(no autocorrelation)
0.34	0.66

c. Joint posteriors for the four alternative hypotheses assuming no autocorrelation (but separate variances for before and after 1977)

	H1: Beverton-Holt, no regime shift	H2: Beverton-Holt, regime shift	H3: 2-line, no regime shift	H4: 2-line, regime shift
Joint Posteriors	0.30	0.31	0	0.39
Bayes factors	1	1.04	0	1.3

d. Marginal posteriors assuming no autocorrelation (but separate variances for before and after 1977)

P(Regime Shift)	P(no Regime Shift)	P(Beverton Holt)	P(2 Line)
0.70	0.30	0.62	0.38

e. Conditional posteriors assuming no autocorrelation (but separate variances for before and after 1977)

P(Regime shift | BH)	P(no Regime shift | BH)	P(Regime shift | 2L)	P(no Regime shift | 2L)
0.51	0.49	0.9999997	0.0000003

Table 6. Bayesian posteriors for alternative hypotheses on the form of the stock-recruit function and on whether a regime shift has occurred, for western Atlantic bluefin tuna assuming the same residual variance over the entire time series.

a. Marginal posteriors and posterior modal values for the SD in model residuals.

	P(different variance)		P(same variance)
	0.54		0.46
	σ≤76	σ>76	
BH, RS	0.59	0.37	0.43
2 L, RS	0.54	0.39	0.52

b. Joint posteriors for the four alternative hypotheses assuming the same variances for before and after 1977 (including autocorrelation)

	H1: Beverton-Holt, no regime shift	H2: Beverton-Holt, regime shift	H3: 2-line, no regime shift	H4: 2-line, regime shift
Joint Posteriors	0.53	0.36	<0.00001	0.11

c. Marginal posteriors assuming the same variances for before and after 1977 (including autocorrelation)

P(Regime Shift)	P(no Regime Shift)
0.47	0.53

P(Beverton Holt)	P(2 Line)
0.88	0.12

d. Conditional posteriors assuming the same variances for before and after 1977 (including autocorrelation)

P(Regime shift | BH)	P(no Regime shift | BH)
0.40	0.60

P(Regime shift | 2L)	P(no Regime shift | 2L)
0.995	0.005

Table 7. Bayes factors for the two main alternative recruitment hypotheses in assessments of western Atlantic bluefin tuna since 1998 considering data available for McAllister *et al*. (2000b). The first data set is that used in their analysis. The second dataset is that applied in this updated analysis but using data only up to 1993 as in the previous analysis.

Source	Years	Beverton-Holt, no regime shift	Two-line regime shift
ICCAT 2010	1970-2006	1	0.21
ICCAT 2010	1970-1993	1	48
ICCAT 1999	1970-1993	1	0.47

Table 8. Bayes factors for the Beverton Holt no regime shift and 2 line 1977 regime shift models under different assumptions about variance and autocorrelation in the recruitment deviates.

Descriptor	Model Variant	Run information	Bayes factor	1/ Bayes factor
Ref. Reference case	Ref.B-H.NRS	Beverton-Holt, no regime shift with S-R data 1970-2006, auto-correlation coefficient (ρ) for stock-recruit function deviates set at 0.52 (ICCAT 2010)	1	1
	Ref.2-L.RS	Two-line, regime shift with S-R data 1970-2006, ρ set at 0.52, variance in S-R deviates also changed in 1977 (ICCAT 2010)	0.21	4.8
A. Auto-correlation and S-R variance assumptions	A.B-H.NRS.1	B-H, no regime shift, lag-1 auto correlation set at zero	1	1
	A.2-L.1	Two line, regime shift auto-correlation set at zero, variance in S-R deviates changes in 1977	1.27	0.8
	A.2-L.RS.2	Two line, regime shift, ρ set at zero, no change in variance in S-R deviates	2.05	0.5
	A.2-L.RS.3	Two line, regime shift, ρ set at 0.52, variance in S-R deviates does not change in 1977	0.18	5.7
C. The effects of adding or subtracting data points	C. B-H.NRS.1	B-H NRS 1970-2006 but exclude 2003 cohort	0.72	1.4
	C.2-L RS.1	2-L Regime shift, 1970-2006 but exclude 2003	1	1
	C. B-H.NRS.2	B-H NRS, fitted to 1970-2005 series.	1	1
	C.2-L.RS.2	2-Line with regime shift fitted to 1970-2005 series.	0.98	1.02
	C. B-H.NRS.3	B-H NRS fitted to 1970-2007 series	1	1
	C.2-L RS.3	2-L Regime shift fitted to 1970- 2007 series	0.23	4.4
	C. B-H.NRS.4	B-H NRS fitted to 1970-2008 series	1	1
	C.2-L.RS.4	2-L Regime shift in 1977 fitted to 1970- 2008 series	0.16	6.1
D. Alternative stock-recruit data	D. B-H.NRS.1	B-H NRS fitted to 1960-1998 series	1	1
	D.2-L.RS.1	2-L Regime shift fitted to 1960- 1998 series	0.37	2.7
F. Alternative priors for R_{bar}, B_0 and σ_R in the 2-L and B-H models	F.B-H.NRS.1	Beverton-Holt as reference case except for the prior for R_0 is uniform on R_0.	1	1
	F.2-L.RS.1	2 Line as reference case except for the prior on R_{bar} being uniform on R_{bar}	0.018	55
	F.B-H.NRS.2	Beverton-Holt as reference case except that model is reparameterized with a uniform on $\log(B_0)$ prior.	1	1
	Ref.2-L.RS.1	2 Line as reference case with the prior on R_{bar} being uniform on $\log(R_{bar})$	0.17	5.8
	F.B-H.NRS.3	Beverton-Holt as reference case except for the maximum σ_R is 2 rather than 1	1	1
	F.2-L.RS.3	2 Line as reference case except for the maximum σ_R is 2 rather than 1	0.18	5.6
	F.B-H.NRS.4	Beverton-Holt as reference case except for the prior for R_0 is uniform on R_0 and the maximum σ_R is 2 rather than 1	1	1
	F.2-L.RS.4	2 Line as reference case except for the the prior on R_{bar} being uniform on R_{bar} and the maximum σ_R is 2 rather than 1	0.015	67

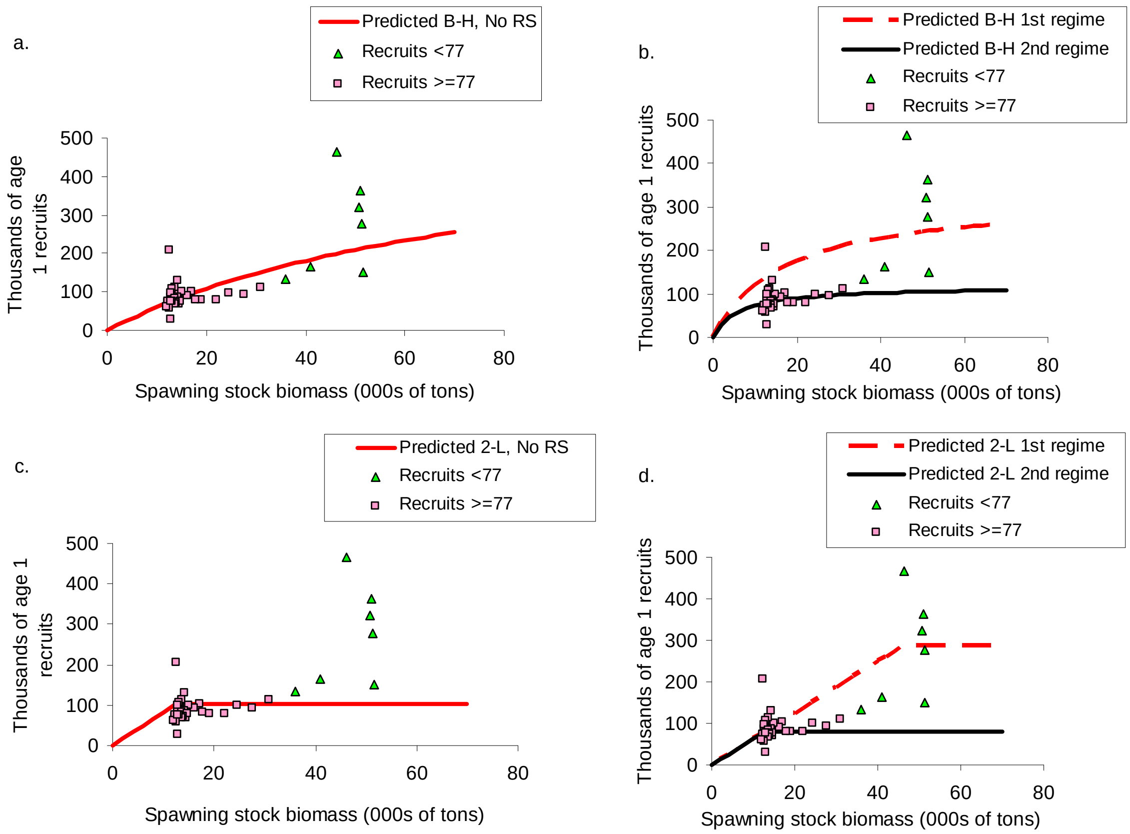

Figure 1. Posterior modal stock-recruit relationships for each of four alternative hypotheses on recruitment for western Atlantic bluefin tuna. a. Beverton-Holt function, no regime shift. b. Beverton-Holt function, regime shift in 1977. c. 2 Line, no regime shift. d. 2 Line regime shift in 1977.

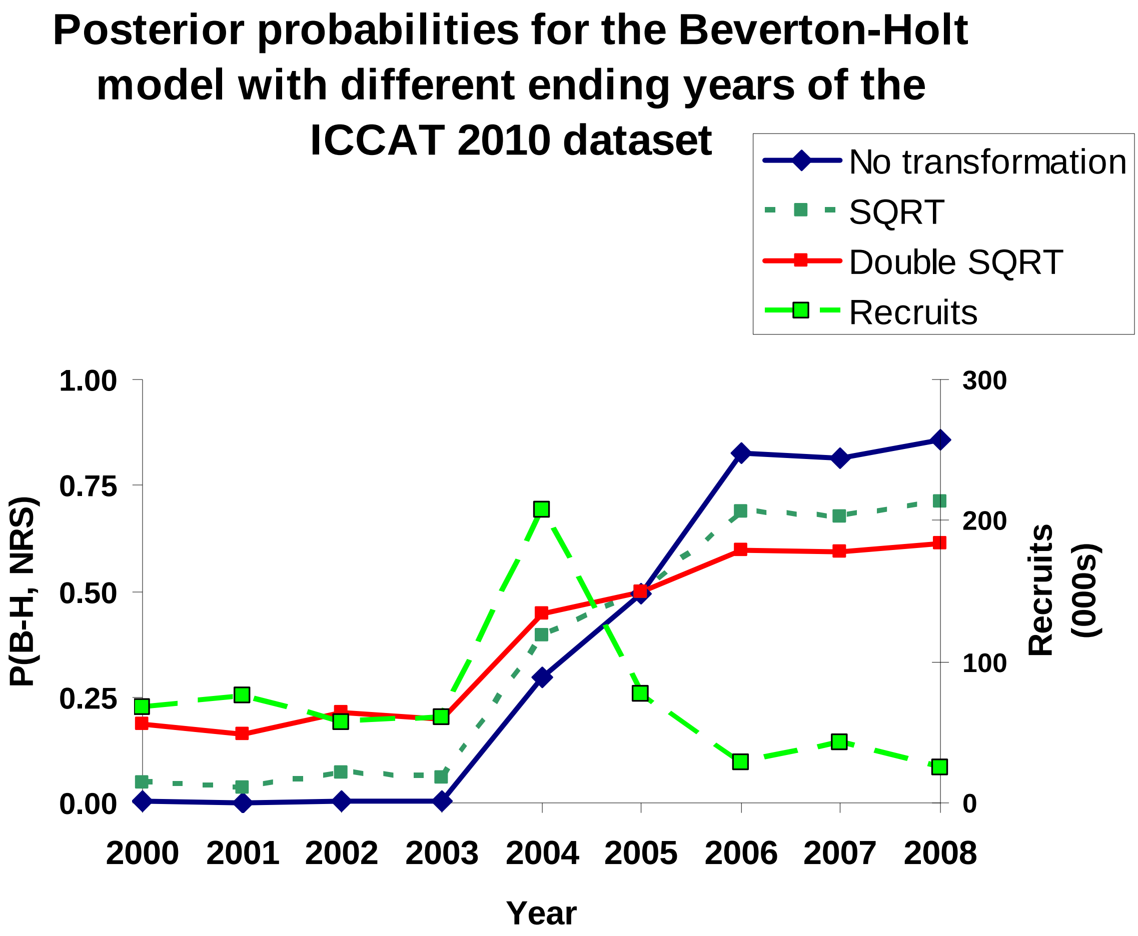

Figure 2. Bayesian posterior probabilities for the Beverton-Holt stock-recruit function when compared to the 2 Line regime shift model when compared to the 2010 estimates of age 1 recruits, and to square root and double square root transformations of the Bayes factors used to compute them. The latter was done to dampen the large interannual changes in the posterior probabilities with new data.