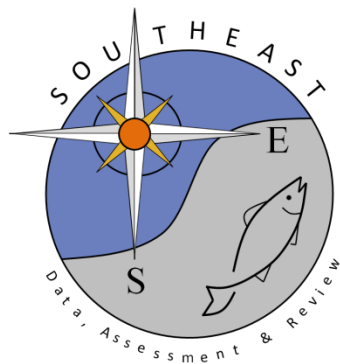JABBA: Just Another Bayesian Biomass Assessment

Henning Winker, Felipe Carvalho, Maia Kapur

SEDAR87-RD-12

November 2024

JABBA: Just Another Bayesian Biomass Assessment

Henning Winker[a,b,*], Felipe Carvalho[c], Maia Kapur[c,d]

[a] Department of Agriculture, Forestry and Fisheries, Private Bag X2, Vlaeberg, 8018, South Africa
[b] Centre for Statistics in Ecology, Environment and Conservation (SEEC), Department of Statistical Sciences, University of Cape Town, South Africa
[c] NOAA Pacific Islands Fisheries Science Center, 1845 Wasp Boulevard, Building 176, Honolulu, HI, 96818, United States
[d] Joint Institute for Marine and Atmospheric Research, University of Hawaii, 1845 Wasp Boulevard, Building 176, Honolulu, HI, 96818, United States

ARTICLE INFO

ABSTRACT

This study presents a new, open-source modelling software entitled 'Just Another Bayesian Biomass Assessment' (JABBA). JABBA can be used for biomass dynamic stock assessment applications, and has emerged from the development of a Bayesian State-Space Surplus Production Model framework, already applied in stock assessments of sharks, tuna, and billfishes around the world. JABBA presents a unifying, flexible framework for biomass dynamic modelling, runs quickly, and generates reproducible stock status estimates and diagnostic tools. Specific emphasis has been placed on flexibility for specifying alternative scenarios, achieving high stability and improved convergence rates. Default JABBA features include: 1) an integrated state-space tool for averaging and automatically fitting multiple catch per unit effort (CPUE) time series; 2) data-weighting through estimation of additional observation variance for individual or grouped CPUE; 3) selection of Fox, Schaefer, or Pella-Tomlinson production functions; 4) options to fix or estimate process and observation variance components; 5) model diagnostic tools; 6) future projections for alternative catch regimes; and 7) a suite of inbuilt graphics illustrating model fit diagnostics and stock status results. As a case study, JABBA is applied to the 2017 assessment input data for South Atlantic swordfish (*Xiphias gladius*). We envision that JABBA will become a widely used, open-source stock assessment tool, readily improved and modified by the global scientific community.

1. Introduction

Surplus Production Models (SPMs) are among the least parameter and data demanding population models that can produce estimates of Maximum Sustainable Yield (MSY) and associated fisheries reference points. Despite a number of limitations (Maunder, 2003; Punt and Szuwalski, 2012; Wang et al., 2014), SPMs remain an integral tool for data-limited to –moderate stock assessments (Dichmont et al., 2016; Punt et al., 2015) and meta-analyses of global fisheries (Froese et al., 2016; Rosenberg et al., 2017; Worm et al., 2009). SPMs approximate changes in biomass as a function of the biomass of the preceding year, the surplus production of biomass, and the removal by the fishery in the form of catch and are not differentiated by age and/or size. In SPMs, somatic growth, reproduction, natural mortality, and associated density-dependent processes are inseparably captured in the interplay of the two major parameters: the intrinsic rate of population increase r and carrying capacity K. The model requires an index of abundance, catch records, and an estimate of initial biomass (Prager, 1994).

A major criticism of SPMs is that they ignore the stock's size/age structure and therefore fail to account for dynamics in gear selectivity

(Wang et al., 2014) and lagged effects of recruitment and mortality (Aalto et al., 2015; Punt and Szuwalski, 2012), which can both lead to biased assessment results. Although these issues remain a limitation of SPMs, there has been considerable progress in optimizing the fitting procedures of SPMs, in consideration that they are continuously implemented by Regional Fishery Management Organizations (RFMOs) around the world. Such improvements include: Bayesian methods with improved prior formulations (McAllister et al., 2001); the development of estimation frameworks that allow incorporating both observation and process errors using mixed-effects (Punt, 2003; Thorson and Minto, 2015); and Bayesian state-space modelling approaches (Meyer and Millar, 1999; Millar and Meyer, 2000; Thorson et al., 2014). The Bayesian framework can reduce uncertainties about estimates of stock size, productivity, and biomass to carrying capacity ratios by using reasonably informative priors that incorporate information available from meta-analyses and published literature on historical stock levels and population demographics (McAllister et al., 2001; Punt and Hilborn, 1997). State-space models are regarded as powerful tools for modelling time-varying abundance indices because they simultaneously account for both process and observation errors (Buckland et al., 2004;

de Valpine, 2002; Meyer and Millar, 1999). Process error can account for model structural uncertainty as well as natural variability of stock biomass due to stochasticity in recruitment, natural mortality, growth, and maturation, whereas observation error determines the uncertainty in the observed abundance index due to measurement error, reporting error and other unaccounted variations in catchability (Francis et al., 2003; Meyer and Millar, 1999; Ono et al., 2012). More recently, Carvalho et al. (2014) investigated incorporating time blocks in SPMs to allow substantial variability in catchability over time caused by dominant changes in fishery strategies.

Due to their low data requirements, SPMs persist as a predominant assessment tool for large pelagic tuna, shark, and billfish assessments (Dichmont et al., 2016; Punt et al., 2015). The majority of SPMs implemented in RFMO stock assessments are based on third-party software, and among the most commonly used are 'A Stock Production Model Incorporating Covariates' (ASPIC; Prager, 2002, 1994) and 'Bayesian Surplus Production 2' (BSP2; McAllister, 2014).

ASPIC fits a generalized Pella and Tomlinson (1969) production function using several indices of abundance and conditions the stock population dynamics on either catch or effort. Uncertainty in ASPIC is evaluated with sensitivity tests, and a separate projections module is available in R. The greatest shortcoming of ASPIC is its inability to estimate process error, thus providing reference point estimates wholly reliant on the trends in data inputs alone. Regardless, ASPIC has been developed for and applied to several stock assessments that have been conducted by the International Commission for the Conservation of Atlantic Tunas (ICCAT). These include yellowfin tuna (*Thunnus albacares*), bigeye tuna (*Thunnus obesus*), skipjack tuna (*Katsuwonus pelamis*), albacore (*Thunnus alalunga*), swordfish (*Xiphias gladius*), and billfishes (Dichmont et al., 2016; Neilson et al., 2013; Punt et al., 2015). The software is provided on the National Oceanic and Atmospheric Administration's web-based toolbox as a standalone download (http://nft.nefsc.noaa.gov/ASPIC.html). Similarly, integrated statistical catch-at-age frameworks, such as Stock Synthesis (Methot and Wetzel, 2013), are capable of executing SPMs in a sophisticated modelling environment, though state-space formulation is not a possibility therein.

BSP2 (McAllister, 2014) is an updated version of ICCAT's deterministic Bayesian surplus production model (BSP) software documented in McAllister and Babcock (2006). The essential improvement of BSP2 over BSP is the implementation of a state-space approach, so that the user may include process error. Like ASPIC, BSP2 uses a Pella-Tomlinson production function, but the shape parameter is constrained to low values when biomass is less than biomass at sustainable yield (B_{MSY}) to avoid unrealistically high estimates of population increase at low population size (McAllister et al., 2000). A projection function for BSP2 was developed in 2012. The most current version of BSP2 has been developed through implementations in assessments of numerous Canadian fish stocks, as well as pelagic fish stocks in the Atlantic and Pacific oceans including North Pacific blue shark, Atlantic yellowfin, bigeye, and bluefin tuna, and billfishes. Historically, BSP2 was available on the ICCAT Stock Assessment Software Catalog.

Third-party software presents advantages by allowing analysts to explore varied assessment configurations and facilitating the peer-review of assessments. However, changes to such programs may be slow as development relies on few developers, and the record of issues addressed may be unclear. Sometimes, for independent software, access and collaboration with the original developer may be limited by website issues and it may be unclear to whom problems should be directed. Some fisheries scientists construct their own models from scratch, typically vetted internally by those involved with the singular assessment at hand. Code written independently for standalone assessments carries the risk of version corruption and likely presents redundancies or crucial discrepancies between similar models. All such assessments would benefit greatly from sourcing a standardized, central protocol from which changes could be suggested, implemented, and archived for the fishery in question.

The ubiquity of SPM use across RFMOs presents a clear need for a unified approach that is reproducible, well-documented, and easily implemented for a variety of fisheries. The open-source platform GitHub provides a means for fisheries scientists to share, document, and improve assessment procedures in a standardized manner, greatly reducing time spent constructing redundant models, and democratizing modelling approaches across nations. Hosting such tools in a globally-accessible repository also increases transparency in the assessment workflow; enables rapid, continuous modification of the code not limited to a single developer (via "forks"), and acts as an archive of model improvements over time (via the "issues" tracker).

This study presents a new, open-source modelling framework called 'Just Another Bayesian Biomass Assessment' (JABBA). The name is a reference to JAGS (Just Another Gibbs Sampler, Plummer, 2003), which is the language in which the Bayesian algorithm is executed. JABBA is a generalized Bayesian State-Space Surplus Production Model and represents an innovative approach to biomass dynamic modelling. The motivation for developing JABBA was to provide a user-friendly R (R Development Core Team, 2013) to JAGS (Plummer, 2003) interface for fitting generalized Bayesian State-Space SPMs to generate reproducible stock status estimates and diagnostics for a wide variety of fisheries. Specific emphasis has been placed on flexibility for specifying alternative model scenarios and achieving high stability and convergence rates throughout the development process. The open source R/JAGS interface provides a means to rapidly modify and run standard assessment scenarios, while still enabling the experienced user to customize the easily accessible R/JAGS source code for specific purposes. In this manuscript, several core features of JABBA are illustrated using data from the 2017 South Atlantic swordfish stock assessment conducted through ICCAT.

2. Materials and methods

JABBA emerged from the development of improved optimization procedures in Bayesian state-space modelling approaches (Meyer and Millar, 1999; Millar and Meyer, 2000), which were subsequently applied and tested for assessments of South Atlantic blue shark (ICCAT, 2016), North Pacific blue shark (ISC, 2017), Mediterranean albacore (ICCAT, 2017a), North and South Atlantic shortfin mako shark (ICCAT, 2017b), and South Atlantic swordfish (ICCAT, 2017c). Features of JABBA include: 1) an integrated state-space tool for averaging and automatically fitting multiple catch-per-unit-effort (CPUE) time series; 2) data-weighting through an estimation of additional observation variance for individual or grouped CPUE; 3) selection between Fox, Schaefer, or Pella-Tomlinson production functions, with the option to estimate the function as B_{MSY}/K; 4) options to fix or estimate the process and observation variance components; 5) model diagnostics tools; 6) future projections for alternative catch regimes; and 7) a suite of inbuilt graphics illustrating model fits and diagnostics, surplus production estimates, historical stock status trajectories, Kobe plot, and future projections.

A JABBA assessment model is comprised of user-selected model processes, data, and statistical methods for comparing data to model predictions. A crucial development in the JABBA framework is the model diagnostic step, which aids in checking for parameter and model-structure misspecification and aims to avoid convergence errors. Systematic misfit to data or conflict between abundance indices within an assessment model should be considered a diagnostic of model misspecification. Unacceptable model fits (i.e., model estimates which do not match the data) can be detected by either the magnitude of the residuals being larger than implied by the observation error, or trends in residuals indicating systematic misfit (Carvalho et al., 2017). Data conflicts occur when different data series, given the model structure, provide conflicting information about important aspects of the dynamics. Unacceptable model misfit or data conflict can be dealt with by either data weighting or changing the model structure (Maunder and
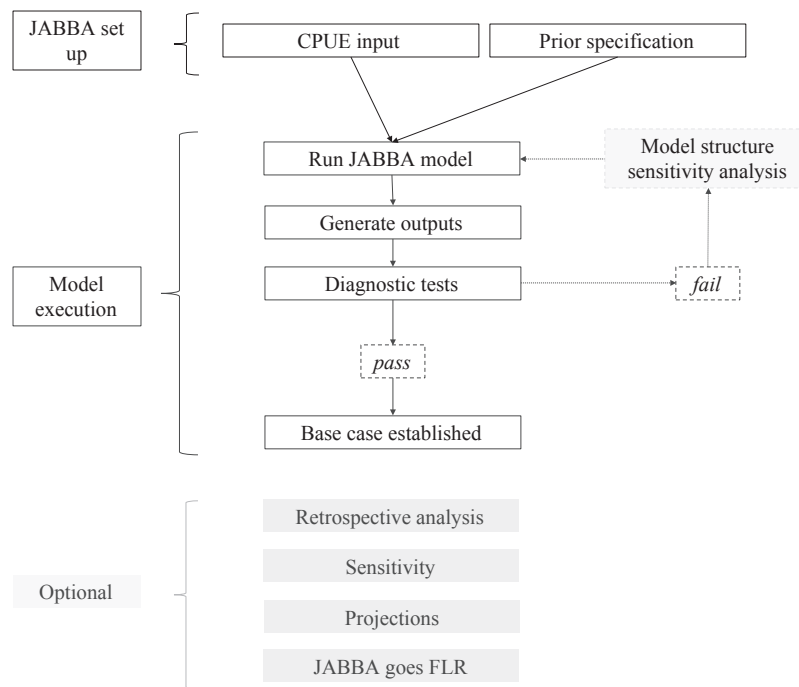
Fig. 1. JABBA workflow. In addition to the modelling framework itself, JABBA presents a new paradigm in model structure, diagnosis, and execution.

Piner, 2015). The ultimate goal of the diagnostic process is to create a dynamic model using all data inputs that fit the observations and are internally consistent. Internal consistency implies that all data are fit well, and their observational errors and residual trends are minimized. Fig. 1 shows a schematic of the modelling approach and internal structure of the JABBA model workflow.

2.1. JABBA model formulation

With JABBA, we seek to provide a generalized Bayesian state-space estimation framework for surplus production models (SPMs) by building on previous formulations by Pella and Tomlinson (1969), Gilbert (1992); (Wang et al., 2014) and Fletcher (1978); (Thorson et al., 2012). The surplus production function is formulated with the generalized three parameter SPM by Pella and Tomlinson (1969) of the form:

$$SP_t = \frac{r}{m-1}B_t\left(1 - \left(\frac{B_t}{K}\right)^{m-1}\right),$$ (1)

where r is the intrinsic rate of population increase at time t, K is the carrying capacity, B is stock biomass at time t, and m is a shape parameter that determines at which B/K ratio maximum surplus production is attained. If the shape parameter $m = 2$, the model reduces to the Schaefer form, with the surplus production (SP) attaining MSY at exactly $K/2$. If $0 < m < 2$, SP attains MSY at biomass levels smaller than $K/2$; the converse applies for values of m greater than 2. The Pella-Tomlinson model reduces to a Fox model (Fox, 1970) if m approaches one, resulting in maximum surplus production at $\sim 0.37 K$, but there is no solution for the exact Fox surplus production with $m = 1$. The shape parameter m can be directly translated into the biomass level where MSY is achieved, B_{MSY}, via the ratio B_{MSY}/K:

$$\frac{B_{MSY}}{K} = m^{\left(-\frac{1}{m-1}\right)}.$$ (2)

It follows that B_{MSY} is given by:

$$B_{MSY} = Km^{\frac{-1}{m-1}},$$ (3)

and the corresponding fishing mortality at MSY (F_{MSY}) is:

$$F_{MSY} = \frac{r}{m-1}\left(1 - \frac{1}{m}\right),$$ (4)

where the fishing mortality is an annual rate defined here as the ratio of:

$$F = \frac{C}{B},$$ (5)

where C denotes catch. Correspondingly, MSY can be expressed by:

$$MSY = F_{MSY}B_{MSY}.$$ (6)

We note that the Pella-Tomlinson formulation provides an approximate link to an age-structured model. Combining and re-arranging Eqs. (3), (4), and (6), it follows that r in Eq. (1) can be expressed as:

$$r = \frac{MSY}{B_{MSY}}\frac{m-1}{1-m^{-1}}.$$ (7)

Eq. (2) together with the re-rearranged Eq. (7) emphasizes the potential of translating estimates of MSY/B_{MSY} and B_{MSY}/K into r and m, respectively (Maunder, 2003; Thorson et al., 2012; Wang et al., 2014). This presents a crucial bridge for parameters derived from age-structured equilibrium models (e.g., a per-recruit analysis with integrated Beverton-Holt recruitment functions) to be implemented in an SPM.

In contrast to the alternative formulation of the Pella-Tomlinson used in several previous studies (Brodziak and Ishimura, 2012; Chaloupka and Balazs, 2007; Chang et al., 2014), this parametrization has the property that for any fixed input values of r and K, SP increases as m decreases (Fig. 2) due to the inclusion of m-1 as the denominator of r (Eq. (1)). However, as with the Fletcher parametrization, if $m \leq 1$, it has the biologically undesirable property of surplus production per unit biomass approaching infinity even as biomass approaches zero (Quinn and Deriso, 1999). This behaviour can be visually inferred from the close-to-vertical increase in surplus production near the curve's origin, in particular for the extremely low value of m = 0.1 (Fig. 2).

To address this anomaly, JABBA provides the additional option of combining the surplus production with a generic "hockey stick" recruitment function developed and tested by Froese et al. (2016). The hockey-stick, as proposed by Barrowman and Myers (2000), assumes that recruitment potential becomes increasingly impaired below a given
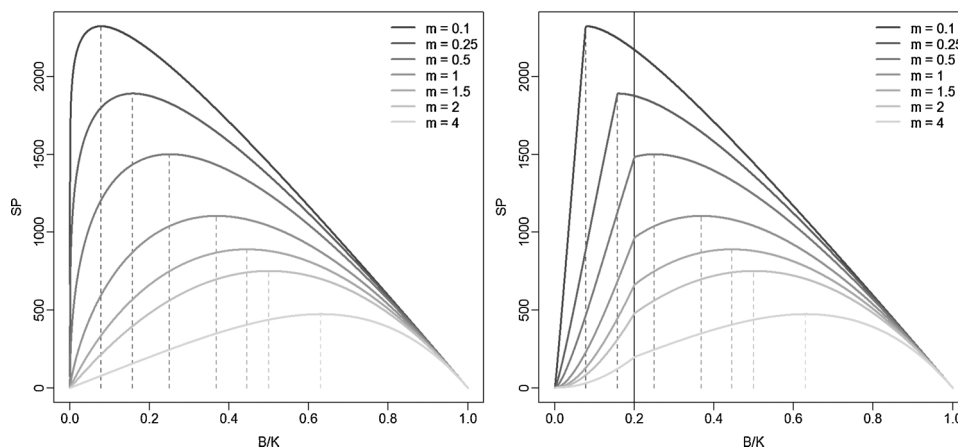
Fig. 2. Surplus production as a function of biomass for different values of the shape parameter $m = 0.1$–4 (from left to right) using the default Pella–Tomlinson model (left panel) and the optional Pella-Tomlinson hockey-stick composite model formulation for the case $P_{lim} = B_{lim}/K = 0.2$ (right panel).

biomass ratio level ($P_{lim} = B_{lim}/K$), with P_{lim} ranges of 0.2–0.25 having been widely adopted as thresholds for recruitment overfishing (Beddington and Cooke, 1983; Myers et al., 1994; Punt et al., 2013). The linear decrease of the underlying hockey-stick between 1 and 0 is implemented by introducing the multiplier to the surplus production function, so that for values of $B/K < P_{lim}$:

$$SP_t = \frac{r}{(m-1)} \frac{B_t}{(P_{lim}K)} B_t \left(1 - \left(\frac{B_t}{K}\right)^{m-1}\right) \quad \text{if} \frac{B_t}{K} < P_{lim} \tag{8}$$

This composite model formulation reduces to the Pella-Tomlinson model as P_{lim} approaches 0. The functional form of the Pella-Tomlinson hockey-stick formulation is illustrated for $P_{lim} = 0.2$ in Fig. 2. A notable property is that for $P_{lim} > B_{MSY}/K$, Eqs. (4) and (5) do not hold because maximum surplus production is attained at P_{lim} under such conditions (Fig. 2). For the condition $P_{lim} > B_{MSY}/K$, JABBA therefore automatically sets P_{lim} equal to the B_{MSY}/K inflection point.

2.2. State-space implementation

JABBA is formulated on the Bayesian state-space estimation framework proposed by Meyer and Millar (1999). The biomass B_y in year y is expressed as a proportion of K (i.e. $P_y = B_y/K$) to improve the efficiency of the estimation algorithm. The model is formulated to accommodate multiple CPUE series i. The initial biomass in the first year of the time series is scaled by introducing model parameter φ to estimate the ratio of the spawning biomass in the first year to K (Carvalho et al., 2014). The stochastic form of the process equation is given by:

$$P_y = \begin{cases} \varphi\, e^{\eta_y} & \text{for} y = 1 \\ \left(P_{y-1} + \frac{r}{(m-1)}P_{y-1}(1 - P_{y-1}^{m-1}) - \frac{\sum_f C_{f,y-1}}{K}\right)e^{\eta_y} & \text{for} P_{y-1} \geq P_{lim} \ \& \ y = 2,3,...,n \\ \left(P_{y-1} + \frac{r}{(m-1)}P_{y-1}(1 - P_{y-1}^{m-1})\frac{P_{y-1}}{P_{lim}} - \frac{\sum_f C_{f,y-1}}{K}\right)e^{\eta_y} & \text{for} P_{y-1} < P_{lim} \ \& \ y = 2,3,...,n \end{cases} \tag{9}$$

where η_y is the process error, with $\eta_y \sim N(0, \sigma_\eta^2)$; process variance ($\sigma_\eta^2$) can be either fixed or estimated. If estimated, the process variance prior is implemented using an inverse-gamma distribution (see next section);

$C_{f,y}$ is the catch in year y by fishery f.

The corresponding biomass for year y is:

$$B_y = P_y K, \tag{10}$$

The observation equation is given by:

$$I_{i,y} = q_i B_y e^{\varepsilon_{y,i}} \quad y = 1, 2, ..., n. \tag{11}$$

where q_i is the estimable catchability coefficient associated with the abundance index i, and $\varepsilon_{y,i}$ is the observation error, with $\varepsilon_{y,i} \sim N(0, \sigma_{\varepsilon,y,i}^2)$; $\sigma_{\varepsilon,y,i}^2$ is the observation variance in year y for index i (see further details, Section 2.3.2). The full Bayesian State-Space Surplus Production model projected over n years requires a joint probability distribution over all unobservable hyper-parameters $\theta = \{K, r, \varphi, \sigma_\eta^2 q_i, \sigma_{\varepsilon,y,i}^2\}$ and the n process errors relating to the vector of unobserved states $\eta = \{\eta_1, \eta_2, ...\eta_y\}$, together with all observable data in the form of the relative abundance indices i, $I_i = \{I_{i,1}, I_{i,2}, ..I_{i,y}\}$ (Meyer and Millar, 1999). According to Bayes' theorem, it follows that joint posterior distribution over all unobservable parameters, given the data and unknown states, can be formulated as:

$$p(\theta|\eta, \mathbf{I}) = p(K)p(r)p(\varphi)p(\sigma_\eta^2)p(q_i)p(\sigma_{\varepsilon,y,i}^2)$$
$$\times p\left(P_1 \middle| \varphi, K, \sigma_\eta^2\right) \prod_{y=1}^n p(P_y|P_{y-1}, K, r, \varphi, \sigma_\eta^2)$$
$$\times \prod_{y=1}^n p(I_{i,y}|P_t, q_i, \eta_t, \sigma_{\varepsilon,y,i}^2) \tag{12}$$

2.3. JABBA set up

2.3.1. Input fishery data

JABBA requires a minimum of two input comma-separated value files (.csv) in the form of catch and abundance indices. The 'Catch' input file contains the time series of year and catch by weight, aggregated across fleets for the entire fishery. Missing catch years or catch values are not allowed. JABBA is formulated to accommodate abundance indices from multiple sources (i.e., fleets) in a single 'CPUE' file, which contains all considered abundance indices. The first column of the CPUE input is year, which must match the range of years provided in the 'Catch' file. In contrast to the 'Catch' input, missing abundance index values are allowed, such that different abundance indices may correspond to smaller portions of the catch time series. Optionally, an additional 'se' input can be passed onto JABBA, containing standard error estimates associated with the abundance indices on a log scale. The 'se' input is a third file, structurally identical to the 'CPUE' input. Year- and series-specific standard errors are typically derived externally

(e.g., from a CPUE standardization model) (Maunder and Punt, 2004). Therefore, the user can assign as many unique 'se' values as there are unique fleets and years in the time series. Alternatively, this feature can be used to apply different weighting to individual abundance indices (Francis, 2011) by assigning varying coefficients of variation (CV) to each time series. This is optional, and JABBA provides diagnostic plots to aid the user in identifying whether modifying standard errors may be beneficial. If such weighting is implemented, it is advised that the CV chosen for each indexed year approximates the observed standard error on the log scale, such that the data weights are congruent with expectations as to how well the model should fit these data.

2.3.2. Prior specification

Most priors are stock-specific and therefore need to be specified by the user in the controlling R input file, the 'Prime' file herein. Key priors for any SPM assessment are typically the intrinsic rate of population increase *r*, the carrying capacity *K*, and the initial biomass depletion at the start of the available catch time series φ. JABBA provides two options to input priors for *r* and *K*; either as the mean and standard deviation (or CV in the case of *K*) of a lognormal distribution (Brodziak and Ishimura, 2012; Meyer and Millar, 1999) or as ranges of minimum and maximum of plausible values for a uniform distribution (Froese et al., 2016). Both options are regular choices in assessments. For example, mean and standard deviation for *r* are typically obtained from Monte-Carlo simulations of demographic equilibrium models (McAllister et al., 2001), whereas a uniform prior for *r* may range between lowest and highest estimates from different studies (ICCAT, 2017b) or approximate ranges based on resilience categories such as those provided in FishBase (Froese et al., 2016). Whether the corresponding lognormal prior is informative or "flat" depends on the width of the range. The larger the difference between lower and upper values, the less informative is the prior. A very wide range will result in a "flat" uninformative lognormal prior. Because lognormal priors on *r* and *K* are considered to have better convergence properties than uniform priors, the input ranges are converted into lognormal priors, such that:

$$\log(\overline{X}) = \frac{X_{\min} + X_{\max}}{2} \quad \text{and} \quad \sigma_r = \frac{\sqrt{(\log(\overline{X}) - \log(X_{\min}))^2}}{2}, \tag{13}$$

where *X* denotes the prior of interest, and σ_r is the approximated lognormal standard deviation for the assumed range of values for *X*.

Similarly, JABBA provides two prior distribution options, lognormal and beta, for the initial biomass depletion ratio φ. The beta distribution is provided as an alternative to the commonly used lognormal because of its desirable property of being bounded by 0 and 1 for depletion ratios. Here, both distributions can be specified by a mean and CV, which are then automatically converted into lognormal mean and standard deviation or beta distribution scale parameters, respectively.

Priors for process variance can be either fixed or estimated. If estimated, the process variance prior is implemented using an inverse-gamma distribution (Brodziak and Ishimura, 2012; Chaloupka and Balazs, 2007; Meyer and Millar, 1999). The JABBA default option for the process variance is $\sigma_\eta^2 \sim 1/gamma(4, 0.01)$, which corresponds to a process error mean of 0.059 with 95% confidence intervals at approximately 0.03 and 0.1 and a CV of 28% (Meyer and Millar, 1999; Millar and Meyer, 2000). This matches the level of process error where state-space SPMs are most likely to adequately perform (Ono et al., 2012; Thorson et al., 2014). Alternatively, the user may select a less informed inverse-gamma distribution by setting both scaling parameters of the inverse-gamma distribution to 0.001 (Brodziak and Ishimura, 2012; Carvalho et al., 2014; Chaloupka and Balazs, 2007). This choice of the inverse-gamma distribution implies that the parameters are approximately uniform on a log scale.

The user can choose to include process variance for all modelled years or only starting in the year when the first abundance index becomes available. However, it is important to note that catch data alone are unlikely to hold sufficient information to separate process

deviations from the deterministic population prediction. Therefore, estimating process error deviations in the absence of auxiliary information from abundance indices may pose the risk of over-fitting.

According to Francis et al. (2003), the observation variance comprises an external estimable observation error $\hat{\sigma}_{SE}$ and year-to-year variation in catchability. In JABBA, the total observation variance σ_ε^2 is separated into three components to further provide the option to distinguish between fixed input variance σ_{fix}^2 and estimable variance σ_{est}^2, where the prior for σ_{est}^2 assumes an uninformative inverse-gamma distribution with both gamma scaling parameters set to 0.001 (Chaloupka and Balazs, 2007). This value can be separately estimated for each abundance index *i*. All three variance components are additive in their squared form (Francis et al., 2003), with the total observation variance for abundance index *i* and year *y* is by given by:

$$\sigma_{\varepsilon,y,i}^2 = \hat{\sigma}_{SE,y,i}^2 + \sigma_{fix}^2 + \sigma_{est,i}^2. \tag{14}$$

Each variance component can be switched on or off in any combination to provide flexible data-weighting options to deal with data conflicts and model misspecifications in stock assessments (Carvalho et al., 2017; Francis, 2011; Maunder and Piner, 2015). This estimable observation variance σ_{est}^2 can be set to be the same value for all abundance indices, or estimated separately for each index. Adding a fixed observation error to externally estimated standard errors for abundance indices is common practice to account for additional sampling errors associated with abundance indices (Maunder and Piner, 2017), such as those caused by year-to-year variation in catchability (Francis et al., 2003). Total observation errors for abundance indices are typically assumed to range from 0.1–0.4 (Francis, 2011; Francis et al., 2003), while Francis (2003) suggested that the portion of the observation variance that accounts for the inter-annual variation in catchability can typically range from 0.15–0.2. Providing a fixed minimum plausible observation error will also inform the process variance of the biomass dynamics, as a portion of total variance is assigned a priori to observation variance. This can also help to increase model stability and convergence of state-space models.

2.4. Model execution

JABBA is executed in the R working environment and can be run on a personal computer. All customization and feature selection as described in Section 2.2 are selected via the Prime file. A simulation including projections and management plots with 2 chains of 10,000 MCMC iterations each for a single model type can be completed in under three minutes and can be monitored via a status bar with verbose updates. The user can specify the number of iterations, burn-in, and thinning rates as in JAGS to assist with MCMC convergence and chain-specific diagnostics. To ensure reproducibility, JABBA is distributed through the global open-source platform GitHub and is accessible free at https://github.com/JABBAmodel. The repository contains the current executable JABBA code, as well as descriptive vignettes and a working tutorial of the material presented in this manuscript.

2.4.1. Model diagnostics
2.4.1.1. CPUE fit.
The principal outputs from a default JABBA run are designed to give the user a cohesive visualization of their input data, model diagnostics, and CPUE fit. This enhances JABBA's utility as a full-fledged assessment model that can quickly provide material for the evaluation of a stock's status. To evaluate CPUE fits, JABBA provides two plots to illustrate several components of model performance. The first is called the "JABBA-residual plot" which includes: (1) colour coded lognormal residuals of observed versus predicted CPUE indices by fleet, (2) boxplots indicating the median and quantiles of all residuals available for any given year; the area of each box indicates the strength of the discrepancy between CPUE series (larger box means higher degree of conflicting information), and (3) a loess smoother
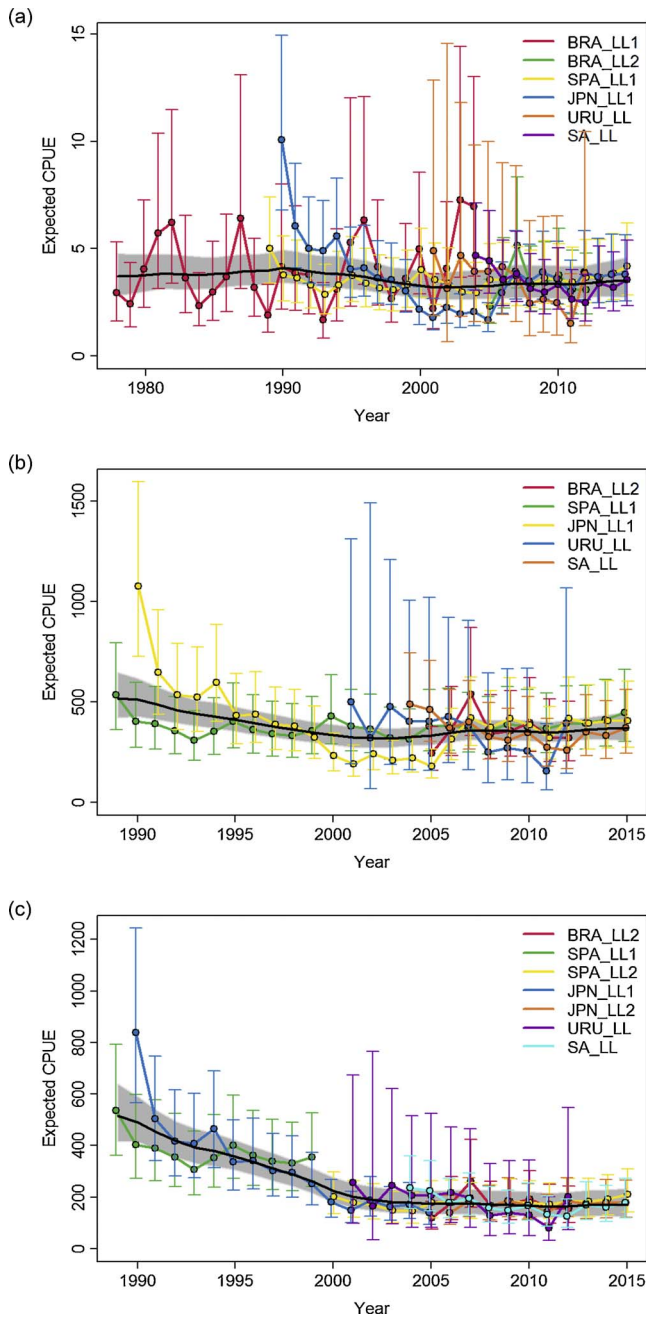
Fig. 3. Illustrations of trends in CPUE indices according to Scenario I (a), II (b), and III (c) for South Atlantic swordfish, which were produced using the state-space CPUE averaging tool implemented in JABBA. The underlying abundance trend is treated as an unobservable state variable that follows a log-linear Markovian process, so that the current mean relative abundance was assumed to be a function of the mean relative abundance in the previous year, an underlying mean population trend and lognormal process error term. The CPUE indices are aligned with the base index via estimable catchability scaling parameters.

through all residuals which highlights systematically auto-correlated residual patterns. The second plot is similar to those obtained from the Stock Synthesis output (Methot and Wetzel, 2013) using the program r4ss (Taylor et al., 2013), and it shows the observed and predicted CPUE values in log scale, as well as the 95% confidence interval (CI). JABBA also calculates the Root-Mean-Squared-Error (RMSE) to quantitatively evaluate the randomness of model residuals. In addition, the deviance information criterion (DIC) is provided for model selection purposes. JABBA will output summary results for CPUE residuals, RMSE, and DIC in a diagnostic results CSV file.

2.4.1.2. MCMC convergence. JABBA is run in JAGS (Plummer, 2003) to estimate the Bayesian posterior distributions of all quantities of interest by means of a Markov Chains Monte Carlo (MCMC) simulation. The JAGS model is executed from R using the wrapper function jags() from the library r2jags (Su and Yajima, 2012), which depends on rjags. Other packages and functions, such as 'jagsUI' or jags.parallel(), were tested during the development process. However, we found that jags() without parallel processing provided the most desirable MCMC mixing properties, thereby achieving MCMC convergence with a fairly small number of iterations. A critical issue when using MCMC methods is how to determine if random draws have converged to the posterior distribution. In JABBA, convergence of the MCMC samples to the posterior distribution is monitored via visual inspection of the trace, and Heidelberger and Welch (Heidelberger and Welch, 1992) and (Geweke, 1992) and Gelman and Rubin (1992) diagnostics as implemented in the coda package. To further improve the convergence properties of JABBA, we developed an optional "soft" boundary penalty on biomass within the JAGS code, which is enforced in cases where the biomass to K ratio, P_y, decreases below or is either equal to or greater than user-defined minimum and maximum values, respectively (default: min = 0.02, max = 1). The idea is that the likelihood is increasingly penalized the further that P_y diverges from the soft boundaries, thereby improving mixing behaviour of MCMC chains. The implementation appears to particularly increase the efficiency of process variance posterior convergence.

2.4.2. JABBA management plots

Stock status estimates can be visually classified based on color-coded biplots that project the biomass (B_y) and harvest rate ($F_y = C_y/B_y$) at year y as relative to their MSY-based reference points (B_y/B_{MSY} and F_y/F_{MSY}). As a default, JABBA produces a Kobe phase plot, which is widely used in tuna RFMOs (de Bruyn et al., 2013).The Kobe phase plot represents the status of the stock in terms of B_y/B_{MSY} on the x-axis and F_y/F_{MSY} on the y-axis. The plot is divided into four quadrants, defined for the stock biomass and fishing mortality relative to B_{MSY} and F_{MSY}, respectively. Use of these biological reference points assumes that managers have chosen B_{MSY} and F_{MSY} as reference points of interest. The color-coding is green if $B_y/B_{MSY} > 1$ and $F_y/F_{MSY} < 1$ and red if $B_y/B_{MSY} < 1$ and $F/F_{MSY} > 1$. However, the color-coding for the other two quadrants may differ among RFMOs. ICCAT, for instance, does not separate between the recovering biomass quadrant if $F_y/F_{MSY} < 1$ and $B_y/B_{MSY} < 1$ and the quadrant of higher fishing mortality if $F_y/F_{MSY} > 1$ and $B_y/B_{MSY} < 1$. Both are illustrated in yellow, whereas the Indian Ocean Tuna Commission (IOTC), for example, visually separates the recovering biomass quadrant from the higher fishing mortality quadrant by using yellow and orange, respectively. JABBA provides the user options to produce either version. In addition, an alternative 'post-modern' three-phase biplot (red, yellow, green) can be produced, which allows separating between target and limit reference points by closely following the principles put forward in Quinn and Collie (2005). In all biplot types, two-dimensional kernel density estimates are used to illustrate the confidence regions corresponding to the approximate 50%, 80%, and 95% credibility intervals for the F/F_{MSY} and B/B_{MSY} posteriors for the terminal year. Estimated probability that the stock is in the respective quadrant of the Kobe plot is provided in the legend of the plot.

In conjunction with the Kobe plot, JABBA produces the 'SP-phase-plot', which plots the surplus production (SP) curve along with the catch trajectory (y-axis) over the biomass range between 0 and K (x-axis). Conceptually, if current catch falls in the area below the *SP* curve, biomass is predicted to increase given that $SP > $ Catch. The maximum *SP* is equivalent to MSY, which corresponds to B_{MSY} on the x-axis. The inflection point at MSY is highlighted together with a shaded area denoting its 95% credibility region. The plot background follows the color scheme of the Kobe phase plot to facilitate interpretation. Additionally, it superimposes plot regions where biomass can recover under a
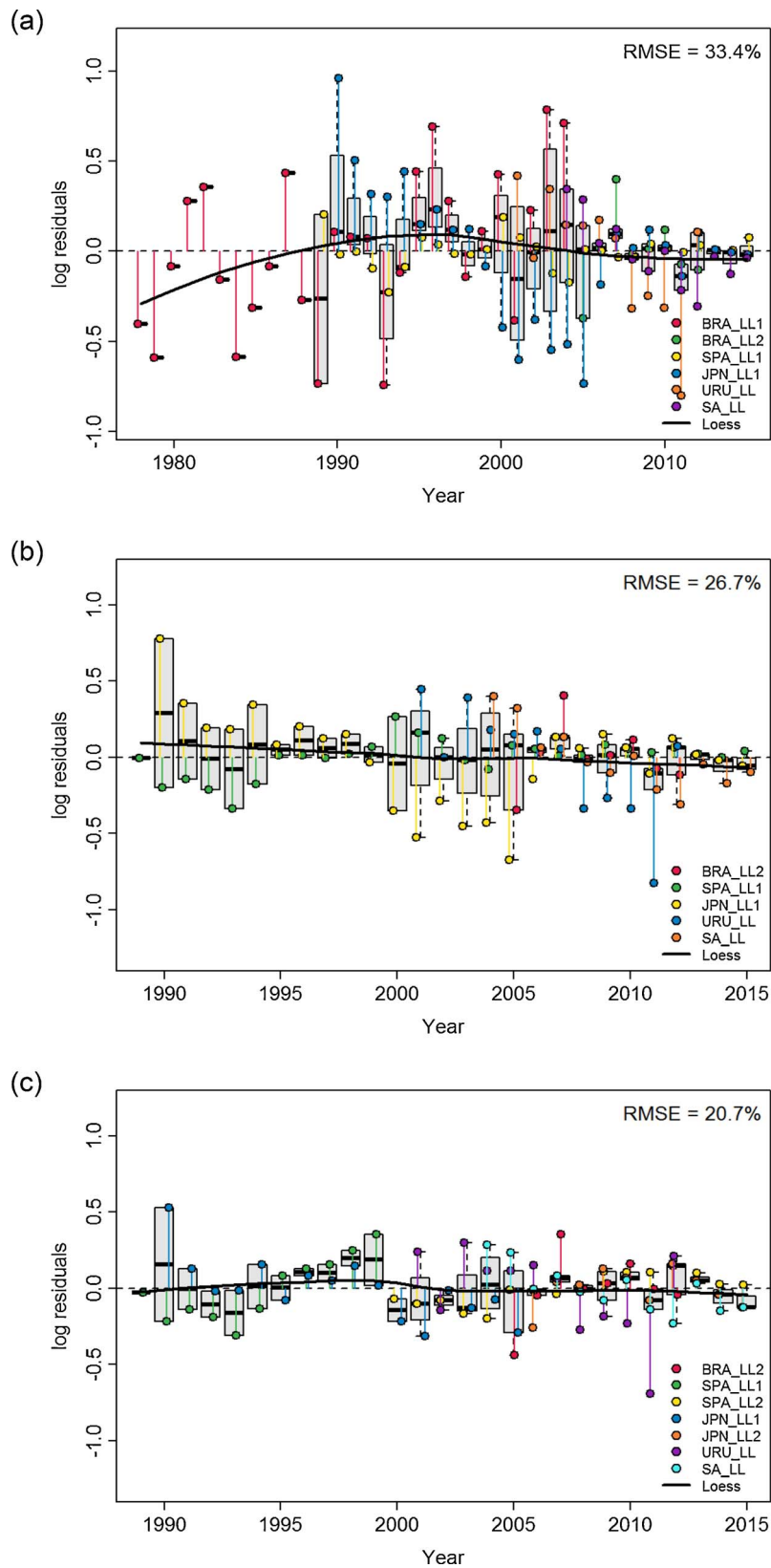
Fig. 4. JABBA residual diagnostic plots for various South Atlantic swordfish CPUE series under the three examined for Scenario I (a), II (b), and III (c). Boxplots indicate the median and quantiles of all residuals available for any given year, and solid black lines indicate a loess smoother through all residuals.

constant quota while in the red overfished state ($B < B_{MSY}$, $F > F_{MSY}$, but $SP >$ Catch). We also note that a constant quota would lead to overfishing if catch is above MSY despite the stock currently being in the green 'sustainable' ($F < F_{MSY}$, $B > B_{MSY}$, but MSY < Catch).

Further details and interpretation of the 'SP-phase-plot' are presented in the results Section 4.
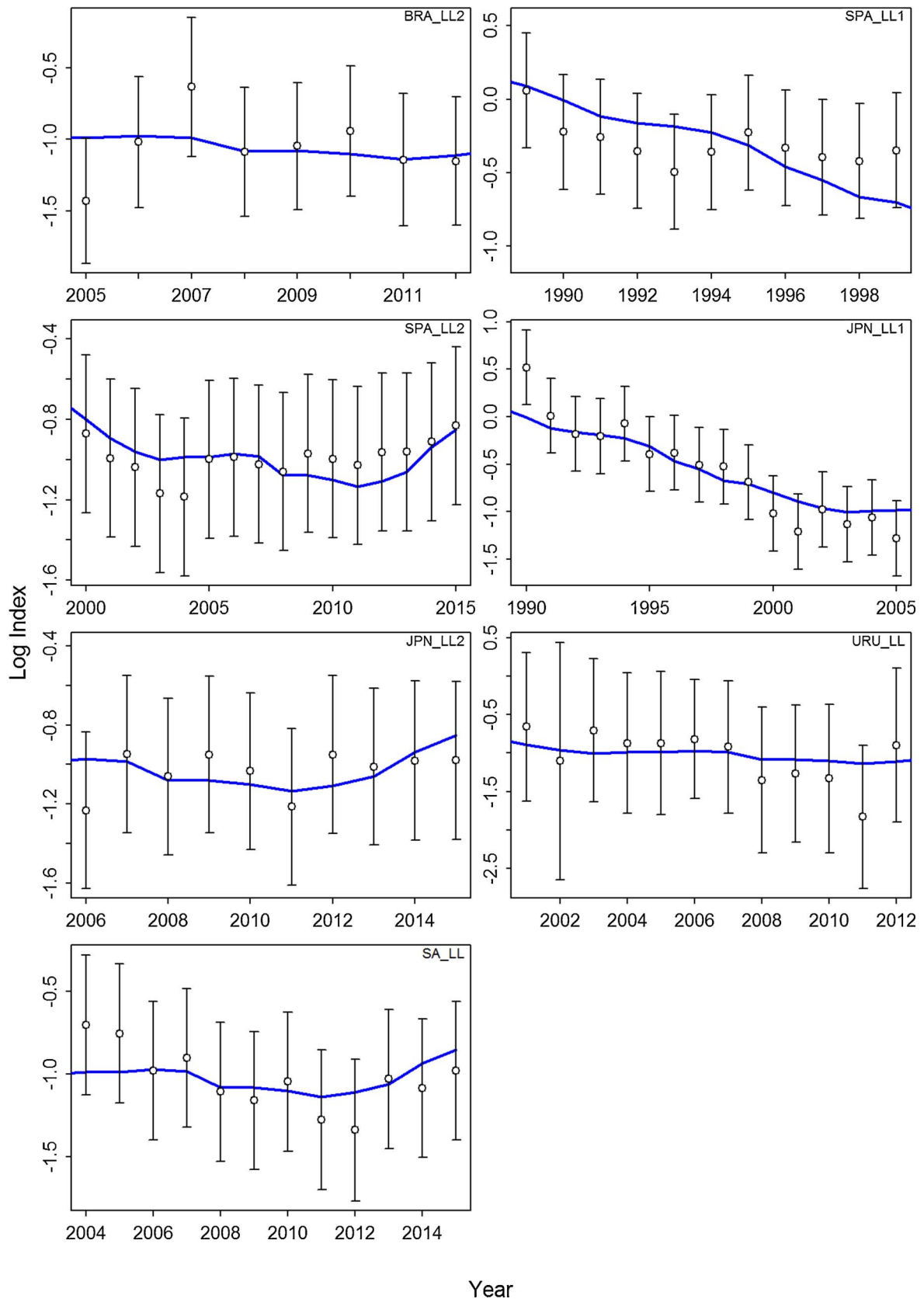
Fig. 5. JABBA fits to the standardized catch-per-unit-effort (CPUE) (in log scale) data sets from different fisheries for the JABBA base case scenario. The solid line is the model predicted value and the circles are observed data values. Vertical lines represent the estimated 95% confidence intervals around the CPUE values.

Table 1
Summary of posterior estimates (medians) and 95% Bayesian Credibility Intervals (C.I.s) of parameters from the South Atlantic swordfish JABBA base case scenario.

	JABBA estimates		
	Median	2.5%	97.5%
K (t)	111468.8	78494.5	163251.8
r	0.372	0.239	0.558
$\varphi = B_{1950}/K$	0.871	0.618	1.052
σ_η	0.055	0.032	0.095
F_{MSY}	0.313	0.201	0.469
B_{MSY} (t)	44585.5	31396.4	65297.7
MSY (t)	13978.3	12616.1	15370.5
B_{2015}/B_{MSY}	0.747	0.545	1.018
F_{2015}/F_{MSY}	0.985	0.723	1.313

2.5. Optional features

2.5.1. State-Space model for averaging of abundance indices

An integrated state-space tool provides a means to average multiple relative abundance indices. The average relative abundance is treated as an unobservable state variable that follows a log-linear Markovian process. The current mean relative abundance μ_y was assumed to be a function of the mean relative abundance in the previous year μ_{y-1}, an underlying mean population trend β and a lognormal process error term η. The process equation is formulated as:

$$\log(\mu_y) = \log(\mu_{y=1}) + \beta + \eta \quad \eta \sim N(0, \sigma_\eta^2),$$ (15)

and the corresponding observation equation is:

$$\log(I_{i,y}) = q_i + \log(\mu_y) + \varepsilon_{i,y} \quad \varepsilon_{i,y} \sim N(0, \sigma_{\varepsilon,i,y}^2),$$ (16)

where $I_{y,i}$ is the relative abundance value for year y and index i, σ_η^2 is the process error variance, $\varepsilon_{i,y}$ is the lognormal observation error term, $\sigma_{\varepsilon,i,y}^2$ is the total observation error variance for year y and index i as given by Eq. (10), and q_i is the catchability scaling parameter for index i. The abundance index with the earliest record (in order of occurrence) is taken as a reference index by fixing $q_1 = 1$, and the others are scaled to this reference index, respectively. The trend β has a flat normal prior, the scaling catchability coefficients follow a non-informative inverse-gamma distribution $q_i \sim 1/dgamma(0.001,0.001)$ and process and observation variance estimation is implemented as for the full JABBA model. This tool allows aligning multiple abundance indices at comparable scales for an initial inspection of an overall mean trend and potentially conflicting trends among time series.

2.5.2. Retrospective analysis

Retrospective analysis is another diagnostic approach widely used in stock assessments to evaluate the reliability of parameter and reference point estimates (Cadigan and Farrell, 2005; Hurtado-Ferro et al., 2014; Mohn, 1999). As an additional model performance diagnostic, we enabled the retrospective analysis procedure, which involves fitting a stock assessment model to the full dataset. The same model is then fitted to truncated datasets where the data for the most recent years are sequentially removed. Retrospective analysis is designed to reveal systematic bias in the estimation of B or F which could, for example, result from model mis-specification. In stock assessment, the ρ statistic proposed by Mohn (1999) is commonly used to evaluate the severity of retrospective patterns (Deroba, 2014). In JABBA, Mohn's ρ is calculated for biomass using the formulation proposed by Hurtado-Ferro et al. (Hurtado-Ferro et al., 2014)

$$\rho = \left(\overline{\frac{X_{Y-y,p} - X_{Y-y,\text{ref}}}{X_{Y-y,\text{ref}}}} \right),$$ (17)

where X is the quantity for which Mohn's ρ is being calculated, Y the final year of the assessment period, y the last year of a given "peel" p,

and ref the reference peel, which is the most recent assessment.

2.5.3. Projections

Users may elect to perform model projections to explore various, user-defined total future catch management scenarios across a given number of simulation years. Joint posteriors of B, F, and the ratios B/K, B/B_{MSY} and F/F_{MSY} are projected by passing on future catch trajectories to JAGS. The resulting 4-dimensional posterior matrix can be saved as an R object for further processing, for example, using the Fisheries library FLR (Kell et al., 2007). The inbuilt trajectory plot illustrates trajectories of the biomass depletion B/K and associated uncertainty over the range of specified catch values.

3. Case study: South Atlantic swordfish

The broadbill swordfish is the most widely distributed billfish species. In the Atlantic Ocean, fisheries catching swordfish are managed by ICCAT (see Neilson et al., 2013 for a review). In 2017, ICCAT carried out the latest stock assessment for the Atlantic swordfish stocks using four different assessment models: Stock Synthesis, BSP2, ASPIC, and JABBA (ICCAT, 2017c). Only the BSP2 and JABBA were applied to the South Atlantic stock (hereinafter referred as SA SWO) because of data availability. Both models produced very similar stock status results. Given considerations that JABBA is written in the open-source R/JAGS interface with more capabilities for future evolutions, it was recommended that management advice, including stock status and projections, should be based on the JABBA model (ICCAT, 2017c). In this case study, we focused specifically on illustrating the JABBA framework applied for this stock assessment. The results presented below are not intended to describe the full results, modelling and parameter choices, or management recommendations for the SA SWO stock assessment, but instead to illustrate the application and implementation of JABBA to an actual stock. For detailed information on the SA SWO stock assessment, please see ICCAT (2017c).

3.1. JABBA set-up for S. Atlantic swordfish

3.1.1. Input fishery data for S. Atlantic swordfish

In addition to a total catch time series, the following CPUE time series were made available for the SA SWO stock assessment (see Fig. 3): Brazil 1 (1978–2004), Brazil 2 (2005–2012), EU-Spain (1989–2015), Japan (1990–2015), Uruguay (2001–2012), and South Africa (2004–2015). Models were evaluated in detail assuming a Pella-Tomlinson production function with the B_{MSY}/K fixed at 0.4.

3.1.2. Prior formulations for S. Atlantic swordfish

Preliminary analysis conducted during the 2013 Atlantic SWO stock assessment showed that abundance indices were uninformative with respect to carrying capacity for SA SWO (ICCAT, 2014). A null hypothesis was formulated in which the carrying capacity per unit sea surface area in the range of SA SWO was presumed to be the same as that for North Atlantic SWO. The prior for K for SA SWO was thus obtained using the posterior for K from the North Atlantic SWO stock assessment model (ICCAT, 2014). This approach was also used in the 2017 SA SWO stock assessment, which resulted in a vaguely informative lognormal prior for K at 200,000 metric tons with a CV of 200% (ICCAT, 2017c). For r, a lognormal prior (mean = log(0.42), sd = 0.37) was assumed which was originally developed by McAllister (2014) using Leslie matrix population simulations. The initial biomass depletion prior ($\varphi = B_{1950}/K$) was input in the form of a lognormal prior, which assumes that the SA SWO stock was at K in 1950 with a CV = 0.25. All catchability parameters were formulated as uninformative uniform priors, while the process variance and observation variance priors were implemented by assuming the following inverse-gamma distributions:
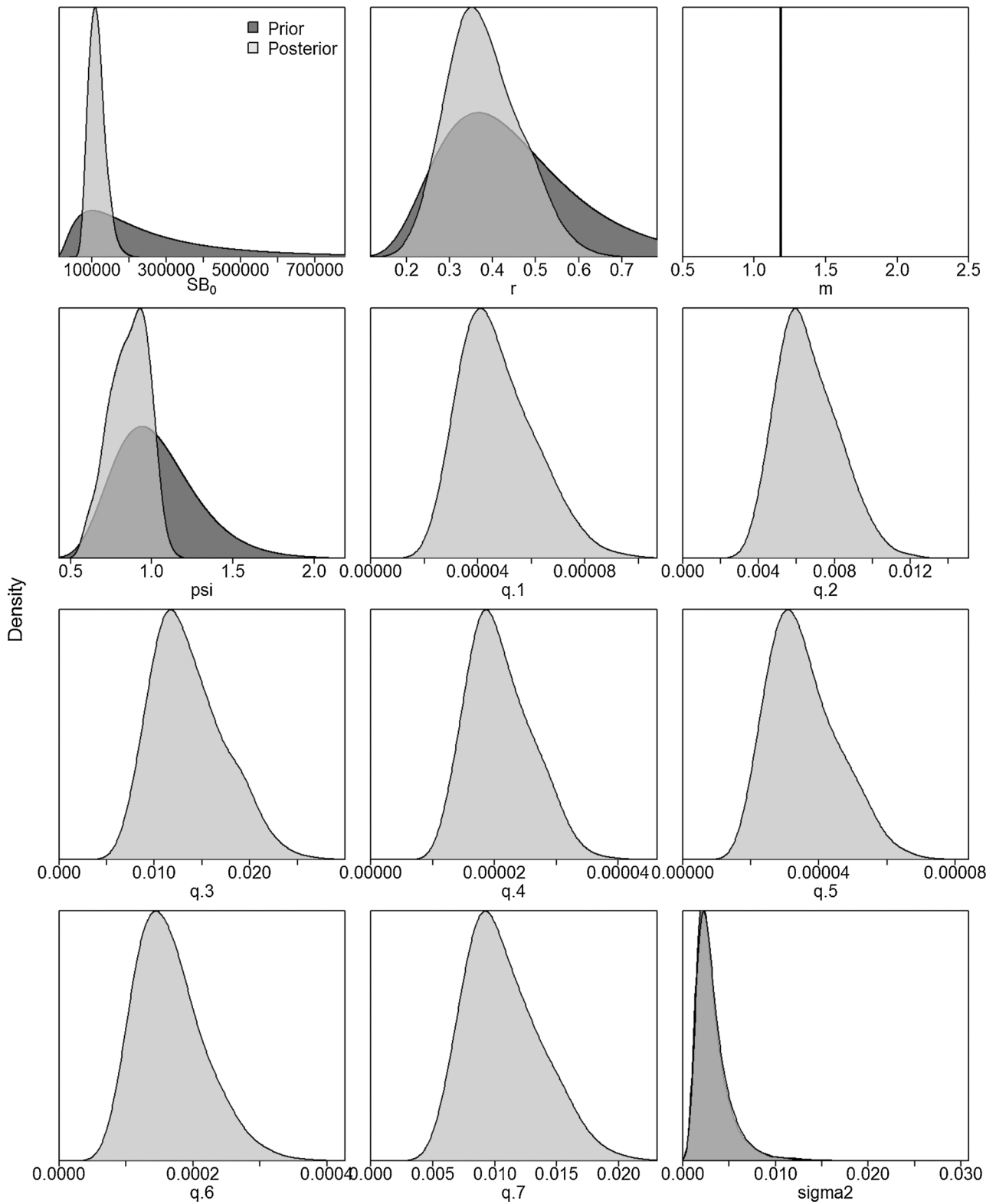
Fig. 6. Prior and posterior distributions of key model parameters for the South Atlantic swordfish JABBA base case scenario. Posteriors distributions are plotted using generic kernel densities.

$$\sigma_\eta^2 \sim \frac{1}{gamma\,(4,\,0.01)} \tag{18}$$

$$\sigma_{est,i}^2 \sim \frac{1}{gamma\,(0.001,\ 0.001)} \tag{19}$$

where $\sigma_{est,i}^2$ was estimated separately for each abundance index *i*.

The process variance prior corresponds to a mean process error of $\sigma_\eta = 0.056$ (CV = 0.65). Because most of the indices provided were considered over-precise with CVs < 0.1, an additional observation error variance of $\sigma_{fix}^2 = 0.2^2$ was added *a priori* to SEs of all time series using the fixed observation error option in JABBA, so that the total observation variance is given by:
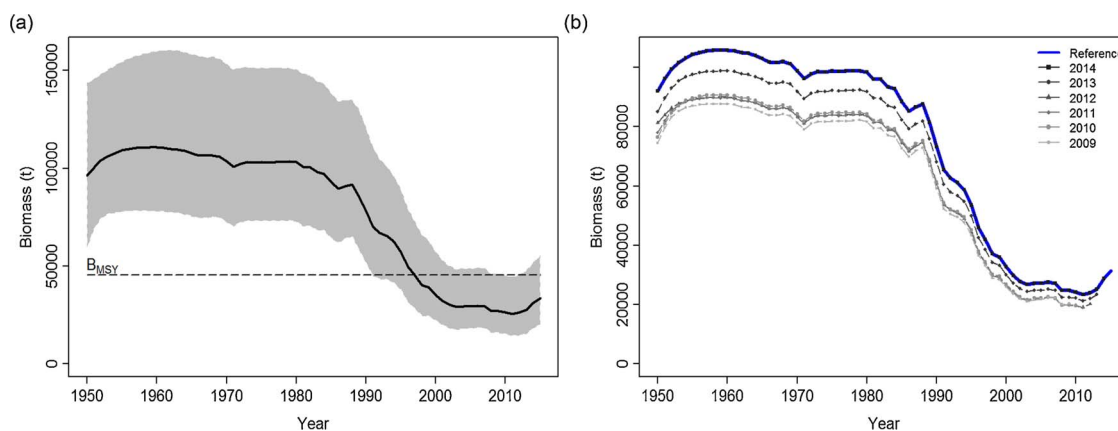
(a)   (b)

Fig. 7. Estimated trajectories (a) for stock biomass B_y and retrospective analysis (b) for the South Atlantic swordfish JABBA base case scenario. Grey-shaded areas denote 95% confidence intervals. The label "Reference" indicates the base case model fitted to the entire time series 1950–2015. The numeric year label indicates the retrospective results from the retrospective peel that includes data through to the referenced year. The mean value Mohn's ρ for Biomass (2009–2015) is −1.49.

$$\sigma^2_{\varepsilon,y,i} = \hat{\sigma}^2_{SE,y,i} + 0.2^2 + \sigma^2_{est,i}. \tag{20}$$

3.1.3. Optional features: retrospective analysis and future projections

A retrospective analysis was conducted for the last 6 years of the assessment time horizon to evaluate whether there were any strong changes in model results based on data availability. The selected period for the retrospective analysis was intended to cover two previous stock assessment cycles for the SA SWO. In addition, future projections for years 2015–2025 were developed. The simulations varied among a sequence of future catches from 4000 to 16,000 tons. The initial catch for the years 2016–2017 was set to the 2016 preliminary total catch reports of 10,056 tons.

4. Results

We have structured the description of results from the SA SWO case study to mirror the "Model Execution" section of the workflow schematic (Fig. 1). The goal is to provide a guide for the reader on how to interpret and customize stock assessments within the JABBA framework.

The first JABBA model for the SA SWO (Scenario I) included all CPUE indices available and (Fig. 3a) no additional model structure. The initial fit to Scenario I provided evidence that Brazil 1 CPUE, which was characterized by high variation, disguised the abundance signals of the other CPUE indices (Fig. 4a). This resulted in an overall poor fit associated with a high root mean squared error (RMSE) of 33.4%, which in this case was considered to *fail* the model diagnostic. Excluding Brazil 1 (Fig. 3b), CPUE (Scenario II) slightly improved the fit (RMSE = 26.7%), but revealed notable data conflicts between the standardized CPUE from Japan (1990–2015) and EU-Spain (1989–2015); therefore, this model was also considered to *fail* the diagnostics (Fig. 4b). The third model (Scenario III) was built after scientific experts suspected that, for Scenario II, the introduction of "American-style" gear had caused changes in swordfish catchability in the Spanish and Japanese fleets, and that those changes were not adequately captured in the standardization of those indices. Therefore, in Scenario III, additional time-blocks in catchability were implemented for EU-Spain and Japan (Fig. 3c). These changes in model structure implemented in Scenario III (Fig. 4c) substantially improved the residual pattern (RMSE = 20.7%). All three scenarios were able to converge adequately as judged by the Geweke (1992) and Heidelberger and Welch (1983) diagnostic tests, and via visually-evaluated stationary behaviour of the MCMC chains (Fig. A1).

Given the evidence of improved RMSE values alone, Scenario III was selected as the most plausible base case scenario (see discussion section for elaboration on the nature of model selection under JABBA's

diagnostic framework). The fits to the abundance indices were varied, but generally within the 95% CI. Overall, trends in observed and predicted CPUE were notably consistent for most indices included in the base case model. However, there was a distinct lack of fit for the EU-Spain index (1989–1999), which resulted in relatively wide confidence intervals due to larger model-internal estimates of $\sigma_{est,i}$ (Fig. 5). Model parameter and stock status estimates are presented for the base-case scenario in Table 1. The routine JABBA output plot, showing the posteriors and the assumed prior distribution for the key model parameters provided no evidence of severe prior misspecification (Fig. 6). The retrospective analysis of predicted trend in stock biomass B_y (Fig. 7a) showed a moderate retrospective pattern for the recent most recent assessment period 2009–2015 (Fig. 7b), with a mean value of Mohn's ρ = −1.49.

Degrees of stock depletion and overfishing in the base case model were illustrated using the "Kobe plot" (Fig. 8a). The current biomass (B_{2015}) is 27% below B_{MSY}, and the current fishing mortality (F_{2015}) is 0.01% below F_{MSY}. The confidence regions of the 2015 stock status posterior (50%, 80%, and 95% CI) straddle the boundary between the red and yellow regions with almost equal probability. The cumulative probability of the red and yellow regions suggests that current biomass levels are below B_{MSY} with a 98.5% probability. Yet, on inspection of the SP-phase plot (Fig. 8b), biomass is expected to increase if catch were maintained at current levels, considering that surplus production is notably larger than recent catches. This also helps explain the quick recovery suggested by projections under various future catch scenarios (Fig. 8c), which predict that stock levels around B_{MSY} would be already attained by 2018 under current catch levels of ∼ 10,000 tons (Fig. 8c). Even a substantial increase to a future catch of 14,000 tons from 2018 onwards is predicted to attain median biomass level at B_{MSY} over the 10-year projection period, albeit associated with considerably higher risk given the current uncertainty about the stock status. In summary, all three plots in conjunction prove useful for interpreting the stock status, with the newly designed SP-phase plot presenting a valuable additional perspective to the Kobe plot for evaluating stock status developments in relation to catch trajectories.

5. Discussion

The Bayesian state-space surplus production modelling tool JABBA presents a unifying yet flexible open-source stock assessment framework for biomass dynamics models. Several core JABBA features are illustrated using the recent 2017 ICCAT South Atlantic swordfish assessment as a case study, which is one of a few assessments already recommended for management for which JABBA was a candidate model (ICCAT, 2017c). The assessment input data comprised multiple,
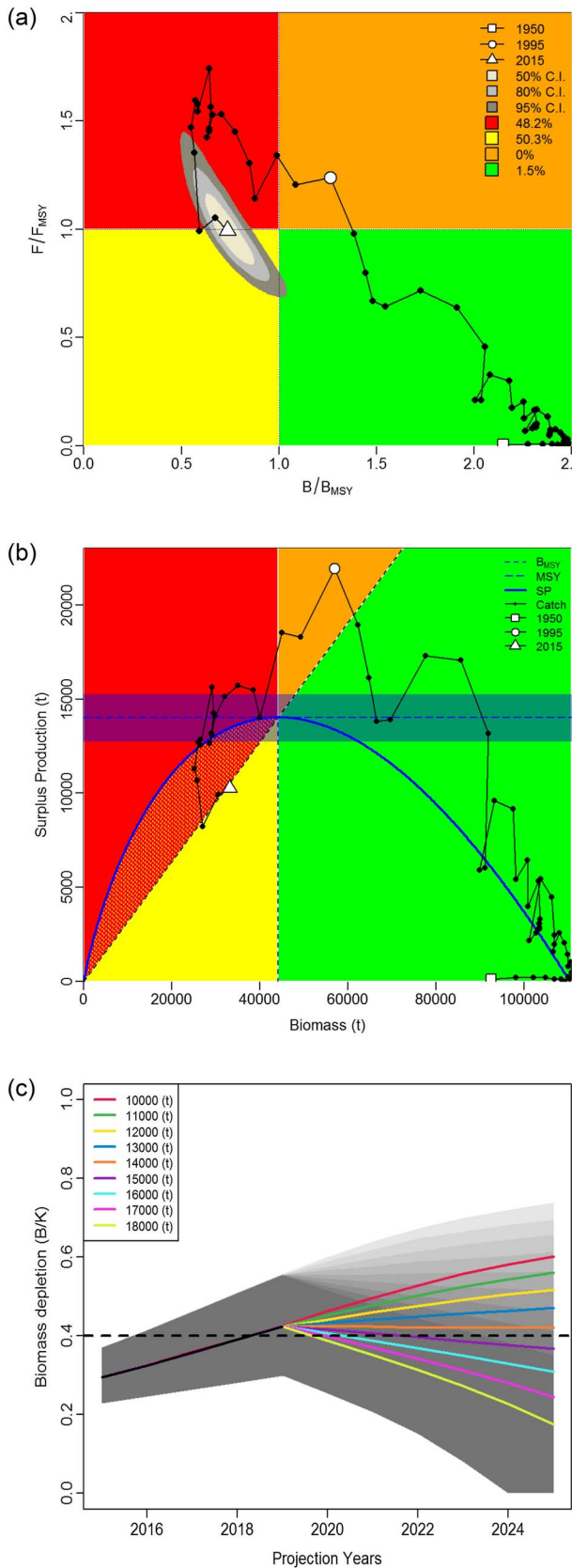
Fig. 8. a) Kobe phase plot for the JABBA base case scenario showing the estimated trajectories (1950–2015) of B/B_{MSY} and F/F_{MSY}. Different grey shaded areas denote the 50%, 80%, and 95% credibility interval for the terminal assessment year. The probability of terminal year points falling within each quadrant is indicated in the figure legend. b) JABBA SP-phase plot showing estimated surplus production curves and catch/biomass trajectories as a function of biomass shown for the base case scenario. MSY estimates are illustrated with 95% C.I.s (grey shaded area). c) Projections (2016–2025) based on the JABBA base case scenario over a sequence of future catches from 10,000–18,000 tons. The initial catch for the years 2016–2017 was set to the 2016 preliminary total catch reports of 10,056 tons. The dashed line denotes B_{MSY}.

diagnostic plots can be used to improve model fit, specifically via the introduction of time-blocked changes in catchability. This feature enabled us to identify and resolve data conflicts and settle upon a base case scenario, which was then used to infer the current stock status and make future projections under varying catch quotas.

JABBA is generalized in the sense that the production function can take on various forms, including conventional Fox and Schaefer production functions, and can be fit using a variety of error assumptions. JABBA is especially easy to use due to the modulated coding structure, which enables the user to simply turn features on or off from the Prime file, as described in Section 2.3.2. For example, by simply changing the settings in the Prime file, it is possible to compare observation error and process error models against a full state-space (total error) implementation, where both variance components are estimable (Ono et al., 2012; Punt, 2003). This structure resembles that of Stock Synthesis (Methot and Wetzel, 2013). JABBA also provides the option to save posteriors of stock trajectories (e.g. B/B_{MSY}, F/F_{MSY}, or B/K) in the form of multi-dimensional array Rdata objects, rendering JABBA outputs directly transferable to other platforms for further processing, such as the Fisheries library FLR (see example in Fig. A2).

A central focus during JABBA development was the workflow paradigm, illustrated in Fig. 1, which emphasizes the user's reliance upon diagnostic tools in the model fitting process. To this end, JABBA routinely produces the newly designed JABBA-residual plot type to aid in identifying poor fits and potential data conflicts among multiple abundance indices. Supporting goodness-of-fit statistics are provided in the form of the residual-mean-squared errors (RMSE). RMSE is a good measure of how precisely the model predicts the response and is the most important criterion in evaluating model fit if the purpose of the model is prediction. However, the best measure of model fit ultimately depends on the analyst's objectives. There is no guarantee, neither by JABBA nor by any stock assessment model, that a model with a great goodness-of-fit score adequately reflects the population dynamics of the stock.

Aside from technical features, JABBA is purposefully designed to run on a personal computer, which adds a revolutionary ability to quickly manipulate and test models even during weeklong RFMO assessment meetings. Most formal reviews of stock assessments take place during week-long meetings, for which the majority of model development and documentation is completed *a priori*. However, it is common for reviewer feedback to be incorporated into the assessment run during the meeting itself, with the intention of evaluating the impacts of alternative model structure and data on stock assessment results. For the SA SWO stock assessment with the MCMC specifications presented here, a single user is able to generate model outputs, including plots and diagnostics, for all three scenarios and projections for the base case in under 10 min.

Considering the rapid uptake of JABBA by multiple RFMOs, the expectation that subsequent assessments follow suit plus the continued need for transparency and reproducibility in assessment science, we chose to host JABBA on the open-source platform GitHub. The repository will retain fully reproducible tutorials of the assessment included in this case study. This will promote the development of additional modules, as users are free to replicate and contribute to the JABBA code in real time.

partially conflicting, fisheries-dependent abundance indices over varying time spans, as is commonly encountered in assessments of pelagic fishes. Our case study demonstrated how iterative examination of

An example of an optional extension has been implemented in JABBA in the form of the 'surplus production and recruitment' composite function (known as the hockey-stick function), which has been proposed and applied by Froese et al. (2016) to prevent the problem of overoptimistic stock status estimates and recovery potential for severely depleted stocks. The incorporation of the generic hockey-stick function through a simple multiplier has the desirable property of preventing surplus production per unit biomass approaching infinity at very low abundance (Fig. 2). As such, it closely resembles the underlying principle for the implementation of the Fletcher-Schaefer composite models in BSP2 (McAllister, 2014; McAllister et al., 2000) and a derivation thereof used in Depletion-Based Stock Reduction Analysis (DB-SRA; Dick and MacCall, 2011) to closely approximate the properties of an underlying Beverton and Holt stock recruitment function. Future JABBA extensions could include options for hierarchical (i.e. "hyper") prior specifications (Jiao et al., 2009), time-varying catchability (Wilberg et al., 2009) or productivity (Chang et al., 2014) estimation approaches, and testing non-proportionality in abundance indices (Davies and Jonsen, 2011). JABBA is also based on the assumption that catch observations are error free, which will usually not be true. It is possible to address this issue by modelling harvest rates as a separate and unobserved process (Pedersen and Berg, 2017). In such a formulation harvest rates can be estimated at any time even when catch information is not available. Further developments of JABBA should allow annual estimates of uncertainty in catch to be included in the assessment model fitting, which is then reflected in the uncertainty of estimated model parameters and management quantities.

JABBA is implemented as a flexible, user-friendly open-source tool to promote reproducibility and provide a platform for future research. The fairly fast convergence properties should encourage more extensive use of simulation-testing evaluations of Bayesian state-space surplus production models, which is currently limited to a few studies (Chang et al., 2014; Dichmont et al., 2016; Ono et al., 2012). An interesting research perspective is to potentially improve the comparability between age-structured models and JABBA by linking the parameterization of the production and shape parameters to dynamic pool models with integrated spawner-recruitment relationship (Maunder, 2003). To this end, we envision that JABBA will become an integral part of stock assessment tool boxes around the world, with potential applications for simulation testing and meta-analyses beyond routine assessment applications.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.fishres.2018.03.010.

References

Aalto, E.A., Dick, E.J., Maccall, A.D., 2015. Separating recruitment and mortality time lags for a delay-difference production model. Can. J. Fish. Aquat. Sci. 165, 161–165.

Barrowman, N.J., Myers, R.A., 2000. Still more spawner–recruitment curves: the hockey stick and its generalizations. Can. J. Fish. Aquat. Sci. 57, 665–676.

Beddington, J.R., Cooke, J., 1983. The potential yield of previously unexploited stocks. FAO Fish. Tech. Pap. 242.

Brodziak, J., Ishimura, G., 2012. Development of Bayesian production models for assessing the North Pacific swordfish population. Fish. Sci. 77, 23–34. http://dx.doi.org/10.1007/s12562-010-0300-0.

Buckland, S.T., Newman, K.B., Thomas, L., Koesters, N.B., 2004. State-space models for the dynamics of wild animal populations. Ecol. Model. 171, 157–175. http://dx.doi.org/10.1016/j.ecolmodel.2003.08.002.

Cadigan, N.G., Farrell, P.J., 2005. Local influence diagnostics for the retrospective problem in sequential population analysis. ICES J. Mar. Sci. 62, 256–265. http://dx.doi.org/10.1016/j.icesjms.2004.11.015.

Carvalho, F., Ahrens, R., Murie, D., Ponciano, J.M., Aires-da-silva, A., Maunder, M.N., Hazin, F., 2014. Incorporating specific change points in catchability in fisheries stock assessment models: an alternative approach applied to the blue shark (Prionace glauca) stock in the south Atlantic ocean. Fish. Res. 154, 135–146. http://dx.doi.org/10.1016/j.fishres.2014.01.022.

Carvalho, F., Punt, A.E., Chang, Y.J., Maunder, M.N., Piner, K.R., 2017. Can diagnostic tests help identify model misspecification in integrated stock assessments? Fish. Res. 192, 28–40. http://dx.doi.org/10.1016/j.fishres.2016.09.018.

Chaloupka, M., Balazs, G., 2007. Using Bayesian state-space modelling to assess the recovery and harvest potential of the Hawaiian green sea turtle stock. Ecol. Model. 205, 93–109. http://dx.doi.org/10.1016/j.ecolmodel.2007.02.010.

Chang, Y.-J., Brodziak, J., O'Malley, J., Lee, H.-H., DiNardo, G., Sun, C.-L., 2014. Model selection and multi-model inference for Bayesian surplus production models: a case study for Pacific blue and striped marlin. Fish. Res. 166, 129–139. http://dx.doi.org/10.1016/j.fishres.2014.08.023.

Davies, T.D., Jonsen, I.D., 2011. Identifying nonproportionality of fishery-independent survey data to estimate population trends and assess recovery potential for cusk (*Brosme brosme*). Can. J. Fish. Aquat. Sci. 68, 413–425. http://dx.doi.org/10.1139/f10-165.

de Bruyn, P., Murua, H., Aranda, M., 2013. The precautionary approach to fisheries management: how this is taken into account by Tuna regional fisheries management organisations (RFMOs). Mar. Policy 38, 397–406. http://dx.doi.org/10.1016/j.marpol.2012.06.019.

de Valpine, P., 2002. Review of methods for fitting time-series models with process and observation error and likelihood calculations for nonlinear, non-gaussian state-space models. Bull. Mar. Sci. 70, 455–471.

Deroba, J.J., 2014. Evaluating the consequences of adjusting fish stock assessment estimates of biomass for retrospective patterns using Mohn's Rho. North Am. J. Fish. Manag. 34, 380–390. http://dx.doi.org/10.1080/02755947.2014.882452.

Dichmont, C.M., Deng, R.A., Punt, A.E., Brodziak, J., Chang, Y.J., Cope, J.M., Ianelli, J.N., Legault, C.M., Methot, R.D., Porch, C.E., Prager, M.H., Shertzer, K.W., 2016. A review of stock assessment packages in the United States. Fish. Res. 183, 447–460. http://dx.doi.org/10.1016/j.fishres.2016.07.001.

Fletcher, R.I., 1978. On the restructuring of the Pella–Tomlinson system. Fish. Bull. 76, 512–515.

Fox, W.W., 1970. An exponential surplus-yield model for optimizing exploited fish populations. Trans. Am. Fish. Soc. 99, 80–88. http://dx.doi.org/10.1577/1548-8659(1970)99<80:AESMFO>2.0.CO;2.

Francis, R.I.C.C., Hurst, R.J., Renwick, J.A., 2003. Quantifying annual variation in catchability for commercial and research fishing. Fish. Bull. 101, 293–304.

Francis, R.I.C.C., 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68, 1124–1138. http://dx.doi.org/10.1139/f2011-025.

Froese, R., Demirel, N., Coro, G., Kleisner, K.M., Winker, H., 2016. Estimating fisheries reference points from catch and resilience. Fish Fish. 18, 506–526. http://dx.doi.org/10.1111/faf.12190.

Gelman, A., Rubin, D.B., 1992. Lnference from iterative simulation using multiple sequences. Stat. Sci. 7, 457–472. http://dx.doi.org/10.2307/2246093.

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

Gilbert, 1992. A stock production modelling technique for fitting catch historiesto stock index data. New Zealand Fisheries Assessment Research Document 92/15.

Heidelberger, P., Welch, P.D., 1992. Simulation run length control in the presence of an initial transient. Oper. Res. 31, 1109–1144. http://dx.doi.org/10.1287/opre.31.6.1109.

Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L., Ono, K., Vert-Pre, K.A., Whitten, A.R., Punt, A.E., 2014. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES J. Mar. Sci. 99–110. http://dx.doi.org/10.1093/icesjms/fsu198.

ICCAT, 2014. Report of the 2013 Atlantic swordfish stock assessment session. Collect. Vol. Sci. Pap. ICCAT 70, 1484–1678.

ICCAT, 2016. Report of the 2015 ICCAT blue shark stock assessment session. Collect. Vol. Sci. Pap. ICCAT 72, 866–1019.

ICCAT, 2017a. Report of the 2017 ICCAT albacore species group intersessional meeting

(including assessment of Mediterranean albacore). Collect. Vol. Sci. Pap. ICCAT 74 (45).

ICCAT, 2017b. Report of the 2017 ICCAT shortfin mako assessment meeting. Collect. Vol. Sci. Pap. ICCAT 74, 1465–1561.

ICCAT, 2017c. Report of the 2017 ICCAT Atlantic swordfish stock assessment session. Collect. Vol. Sci. Pap. ICCAT 74, 841–967.

ISC, 2017. Stock assessment and future projections of blue shark in the North Pacific Ocean through 2015. pp. 154, (http://isc.fra.go.jp/pdf/ISC17/ISC17_Annex13-Stock_Assessment_and_Future_Projections_of_Blue_Shark.pdf).

Jiao, Y., Hayes, C., Corte, E., 2009. Hierarchical Bayesian approach for population dynamics modelling of fish complexes without species-specific data. ICES J. Mar. Sci. 66, 367–377. http://dx.doi.org/10.1093/icesjms/fsn162.

Kell, L.T., Mosqueira, I., Grosjean, P., Fromentin, J., Garcia, D., Hillary, R., Jardim, E., Mardle, S., Pastoors, M.A., Poos, J.J., Scott, F., Scott, R.D., 2007. FLR: an open-source framework for the evaluation and development of management strategies. ICES J. Mar. Sci. 64, 640–646.

Maunder, M.N., Piner, K.R., 2017. Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets. Fish. Res. 192, 16–27. http://dx.doi.org/10.1016/j.fishres.2016.04.022.

Maunder, M.N., 2003. Is it time to discard the Schaefer model from the stock assessment scientist's toolbox? Fish. Res. 61, 145–149. http://dx.doi.org/10.1016/S0165-7836(02)00273-4.

McAllister, M., Babcock, E., 2006. Bayesian surplus production model with the Sampling Importance Resampling algorithm (BSP): a user's guide.

McAllister, M.K., Babcock, E.A., Pikitch, E.K., Prager, M.H., 2000. Application of a non-equilibrium generalized production model to South and North Atlantic swordfish: combining Bayesian and demographic methods for parameter estimation. Collect. Vol. Sci. Pap. ICCAT 51, 1523–1550.

McAllister, M.K., Pikitch, E.K., Babcock, E.A., 2001. Using demographic methods to construct Bayesian priors for the intrinsic rate of increase in the Schaefer model and implications for stock rebuilding. Can. J. Fish. Aquat. Sci. 58, 1871–1890. http://dx.doi.org/10.1139/cjfas-58-9-1871.

McAllister, M.K., 2014. A generalized Bayesian surplus production stock assessment software (BSP2). Collect. Vol. Sci. Pap. ICCAT 70, 1725–1757.

Methot, R.D., Wetzel, C.R., 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fish. Res. 142, 86–99. http://dx.doi.org/10.1016/j.fishres.2012.10.012.

Meyer, R., Millar, R.B., 1999. BUGS in Bayesian stock assessments. Can. J. Fish. Aquat. Sci. 56, 1078–1086. http://dx.doi.org/10.1139/cjfas-56-6-1078.

Millar, R.B., Meyer, R., 2000. Non-linear state space modelling of fisheries biomass dynamics by using Metropolis-Hastings within-Gibbs sampling. J. R. Stat. Soc. Ser. C (Applied Stat.) 49, 327–342. http://dx.doi.org/10.1111/1467-9876.00195.

Mohn, R., 1999. The retrospective problem in sequential population analysis: an investigation using cod fishery and simulated data. ICES J. Mar. Sci. 56, 473–488. http://dx.doi.org/10.1006/jmsc.1999.0481.

Myers, R.A., Rosenberg, A.A., Mace, P.M., Barrowman, N., Restrepo, V.R., 1994. In search for thresholds of recruitment overfishing. Can. J. Fish. Aquat. Sci. 51, 191–205.

Neilson, J., Arocha, F., Cass-Calay, S., Mejuto, J., Ortiz, M., Scott, G., Smith, C., Travassos, P., Tserpes, G., Andrushchenko, I., 2013. The recovery of atlantic swordfish: the comparative roles of the regional fisheries management organization and species biology. Rev. Fish. Sci. 21, 59–97. http://dx.doi.org/10.1080/10641262.2012.754842.

Ono, K., Punt, A.E., Rivot, E., 2012. Model performance analysis for Bayesian biomass dynamics models using bias, precision and reliability metrics. Fish. Res. 125, 173–183. http://dx.doi.org/10.1016/j.fishres.2012.02.022.

Pedersen, M.W., Berg, C.W., 2017. A stochastic surplus production model in continuous time. Fish Fish. 18, 226–243. http://dx.doi.org/10.1111/faf.12174.

Pella, J.J., Tomlinson, P.K., 1969. A generalized stock production model. Inter-Am. Trop. Tuna Comm. Bull. 13, 421–458.

Plummer, M., 2003. JAGS: a program for analysis of bayesian graphical models using

gibbs sampling. In: 3rd International Workshop on Distributed Statistical Computing (DSC 2003). Vienna, Austria.

Prager, M.H., 1994. A suite of extensions to a nonequilibrium surplus-production model. Fish. Bull. 92, 374–389.

Prager, M.H., 2002. Comparison of logistic and generalized surplus-production models applied to swordfish *Xiphias gladius*, in the north Atlantic Ocean. Fish. Res. 58, 41–57. http://dx.doi.org/10.1016/S0165-7836(01)00358-7.

Punt, A.E., Hilborn, R., 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. Rev. Fish Biol. Fish. 7, 35–63. http://dx.doi.org/10.1023/a:1018419207494.

Punt, A.E., Szuwalski, C., 2012. How well can FMSY and BMSY be estimated using empirical measures of surplus production? Fish. Res. 134–136, 113–124. http://dx.doi.org/10.1016/j.fishres.2012.08.014.

Punt, A.E., Smith, A.D.M., Smith, D.C., Tuck, G.N., Klaer, N.L., 2013. Selecting relative abundance proxies for BMSY and BMEY. ICES J. Mar. Sci. 71, 469–483. http://dx.doi.org/10.1093/icesjms/fst162.

Punt, A.E., Su, N.-J., Sun, C.-L., 2015. Assessing billfish stocks: a review of current methods and some future directions. Fish. Res. 166, 103–118. http://dx.doi.org/10.1016/j.fishres.2014.07.016.

Punt, A.E., 2003. Extending production models to include process error in the population dynamics. Can. J. Fish. Aquat. Sci. 60, 1217–1228.

Quinn, T.J., Collie, J.S., 2005. Sustainability in single-species population models. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360, 147–162. http://dx.doi.org/10.1098/rstb.2004.1577.

Quinn, T.J., Deriso, R.B., 1999. Quantitative Fish Dynamics. Oxford University Press, New York.

R Development Core Team, 2013. R: A language and environment for statistical computing.

Rosenberg, A.A., Kleisner, K.M., Afflerbach, J., Anderson, S.C., Dickey-Collas, M., Cooper, A.B., Fogarty, M.J., Fulton, E.A., Gutiérrez, N.L., Hyde, K.J.W., Jardim, E., Jensen, O.P., Kristiansen, T., Longo, C., Minte-Vera, C.V., Minto, C., Mosqueira, I., Osio, G.C., Ovando, D., Selig, E.R., Thorson, J.T., Walsh, J.C., Ye, Y., 2017. Applying a new ensemble approach to estimating stock status of marine fisheries around the world. Conserv. Lett. 0, 1–9. http://dx.doi.org/10.1111/conl.12363.

Su, Y.-S., Yajima, M., 2012. R2jags: A Package for Running jags from R.

Taylor, I.G., Stewart, I.J., Hicks, A.C., Garrison, T.M., Punt, A.E., Wallace, J.R., Wetzel, C.R., Thorson, J.T., Takeuchi, Y., Ono, K., Monnahan, C.C., Stawitz, C.C., Teresa, Z.A., Whitten, A.R., Johnson, K.F., Emmet, R.L., Anderson, S.C., Iantaylornoaagov, M.I.T., 2013. Package r4ss: R Code for Stock Synthesis.

Thorson, J.T., Minto, C.C., 2015. Mixed effects: a unifying framework for statistical modelling in fisheries biology. ICES J. Mar. Sci. 72http://dx.doi.org/10.1093/icesjms/fsu213. (10.1093/icesjms/fst048).

Thorson, J.T., Cope, J.M., Branch, T.A., Jensen, O.P., Walters, C.J., 2012. Spawning biomass reference points for exploited marine fishes, incorporating taxonomic and body size information. Can. J. Fish. Aquat. Sci. 69, 1556–1568. http://dx.doi.org/10.1139/f2012-077.

Thorson, J.T., Ono, K., Munch, S.B., 2014. A Bayesian approach to identifying and compensating for model misspecification in population models. Ecology 95, 329–341.

Wang, S.-P., Maunder, M.N., Aires-da-Silva, A., 2014. Selectivity's distortion of the production function and its influence on management advice from surplus production models. Fish. Res. 158, 181–193. http://dx.doi.org/10.1016/j.fishres.2014.01.017.

Wilberg, M.J., Thorson, J.T., Linton, B.C., Berkson, J., 2009. Incorporating time-varying catchability into population dynamic stock assessment models. Rev. Fish. Sci. 18, 7–24. http://dx.doi.org/10.1080/10641260903294647.

Worm, B., Hilborn, R., Baum, J.K., Branch, T.A., Collie, J.S., Costello, C., Fogarty, M.J., Fulton, E.A., Hutchings, J.A., Jennings, S., Jensen, O.P., Lotze, H.K., Mace, P.M., McClanahan, T.R., Minto, C., Palumbi, S.R., Parma, A.M., Ricard, D., Rosenberg, A.a., Watson, R., Zeller, D., 2009. Rebuilding global fisheries. Science 325, 578–585. http://dx.doi.org/10.1126/science.1173146.