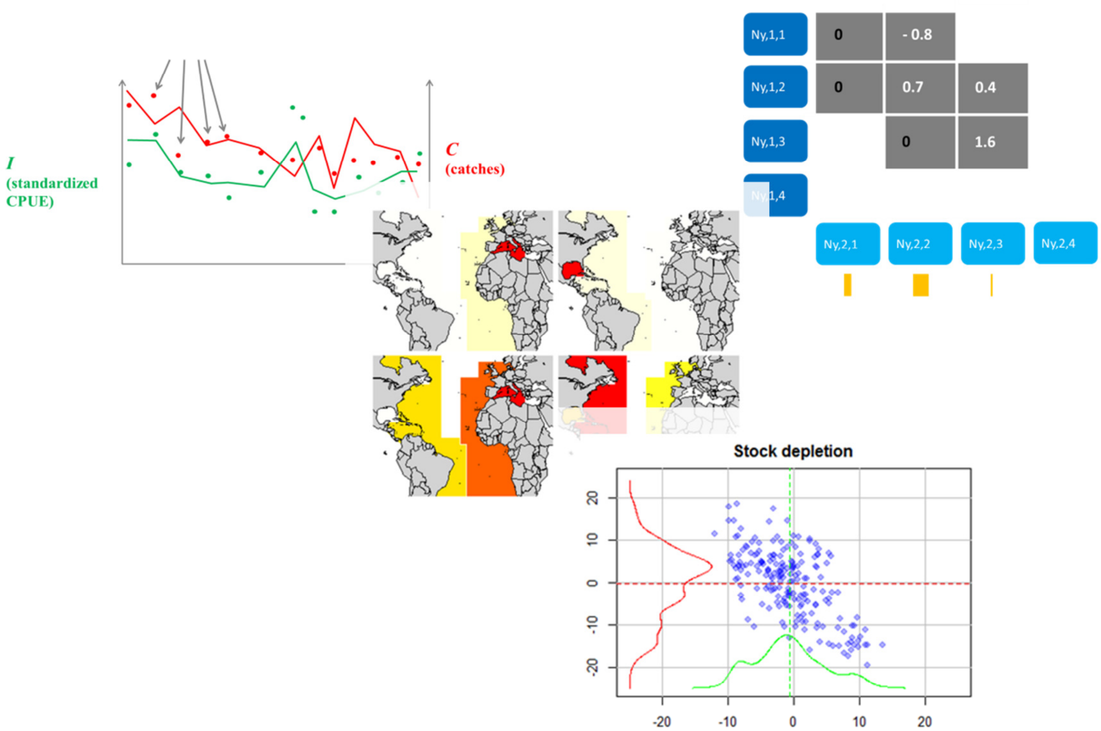
Evaluating Management Strategies for Atlantic Bluefin Tuna

Report 2: Operating model development and data requirements

September 2015

SHORT-TERM CONTRACT FOR MODELLING APPROACHES: SUPPORT TO BFT ASSESSMENT (GBYP 02/2015) OF THE ATLANTIC-WIDE RESEARCH PROGRAMME ON BLUEFIN TUNA (ICCAT-GBYP – Phase 5)



Tom Carruthers[1]
UBC

[1] t.carruthers@fisheries.ubc.ca 335 IOF, 2202 Main Mall, UBC V6T 1Z4 +1 604 822-6903

Executive Summary

The modifiable multi-stock model (M3) was developed as an operating model for characterizing credible hypotheses for Atlantic Bluefin Tuna (ABT) population and fishery dynamics. The M3 model may be fitted to a wide range of data including fishery catch and effort data, stock of origin data, catch size compositions, electronic tagging data and fishery independent surveys.

In this report I describe the structure and estimation framework of the M3 model and the various ways in which the available data may be used to inform the model. I also identify a range of possible hypotheses (alternative scenarios) for ABT relating to population dynamics, fishery dynamics and the interpretation of the various data.

A simulation evaluation was undertaken to demonstrate proof of concept (reliability of the M3 model structure and assumptions). These simulations indicate that the model can be expected to provide unbiased estimates of variables of management interest such as stock depletion, current stock size and spatial distribution. The strengths and weaknesses of the new operating model are discussed relative to previous mixing models.

In addition to model development this report summarizes the metadata that are available for fitting operating models for ABT. A metadata table is presented that categorizes data and describes their current availability to the GBYP MSE process.

In light of simulation testing, model development and the metadata summary, I highlight data collection priorities and also identify a possible strategy for proceeding with operating model development in the absence of complete data.

In order to meet contract deliverables it is necessary to obtain stock of origin data at the resolution of subyear and area. The second most important data priority is the provision of standardized indices for the Japanese and US longline fleets at the resolution of subyear and area.

Finally, I report on progress with respect to project deliverables and discuss wider possible benefits of the modelling work such as providing an alternative stock assessment that accounts for stock mixing.

Contents

1 Introduction

The Atlantic-Wide Research Programme on Bluefin Tuna (GBYP) aims to develop a new scientific management framework by improving data collection, knowledge of key biological and ecological processes, assessment models and management. A critical component of the GBYP is the construction of a robust advice framework consistent with the precautionary approach (GBYP 2014). A Management Strategy Evaluation (MSE, Cochrane 1998, Butterworth 1999, Kell et al. 2014, Punt et al. 2014) approach has been proposed to address this goal (Anon. 2014b). MSE establishes operating models that represent credible hypotheses for population and fishery dynamics which are used to quantify the efficacy of various management procedures. These management procedures may encompass a wide range of complexity from conventional stock assessments linked to harvest control rules (Hilborn 2003) through to simple empirical management procedures that calculate catch limits directly from resource monitoring data indices (Geromont and Butterworth 2014a;b, Kell et al. 2014).

MSE applications generally develop operating models from stock assessments that are fitted to data in order to ensure that model assumptions and estimated parameters are empirically credible (Punt et al. 2014, e.g. CCSBT 2011). In the case of Atlantic bluefin tuna, such a model requires enough complexity to capture the core uncertainties regarding Atlantic bluefin tuna dynamics (Fromentin et al. 2014, Leach et al. 2014). These include stock structure (Kell et al. 2012), stock mixing, migration (Fromentin and Lopuszanski 2014) and biases in observed data (e.g. annual catch data). Since operating models for Atlantic bluefin tuna must be able to represent hypotheses regarding spatio-temporal distribution and stock mixing (Kell et al. 2011, Arrizabalaga et al. 2014, Fromentin and Lopuszanski 2014) a suitable operating model must include spatial and seasonal structure.

In this report I document a novel spatial, multi-stock, statistical catch-at-length assessment model, referred to as the modifiable multi-stock model or M3 (v1.03)(referred to as 'the operating model' herein). These equations were also presented to the SCRS during the September 2015 bluefin tuna species group meeting (Carruthers et al. 2015a). The most notable multi-stock model previously applied to Atlantic bluefin tuna was MAST (Taylor et al. 2011). In the development of M3 I aimed to address a number of central weaknesses of MAST and produce a credible, robust and faster assessment model that can be applied to rapidly evaluate alternative hypotheses for Atlantic bluefin tuna.

The M3 model is able to accommodate the wide range of data that have been collected for Atlantic bluefin tuna including catch rate indices (Abid et al. 2015, Hanke et al. 2015, Kimoto et al. 2015a, Lauretta and Brown 2015, Santiago et al. 2015, and Walter 2015), aerial surveys (Bonhommeau et al. 2010), length composition data, larval surveys (Ingram et al. 2015), electronic tagging data (Block et al. 2005) and stock of origin data (Rooker et al. 2014). These data for Atlantic bluefin vary widely in their information content, may be interpreted in various ways (e.g. exploitation rate, growth, movement) and often provide contrasting information about fleet and population dynamics. For example, size composition data and catch rate indices may provide different inferences regarding stock depletion. A range of operating models may be developed based on the types of data used to fit the model and the way in which the model interprets these data.

A precursor to discussions about how data should be used by operating models is a metadata summary that outlines the types of data that have been collected for Atlantic bluefin tuna. In this report I summarize these data and describe their potential role in the development of operating models. We also describe an interim solution to operating model development in which a preliminary operating model is established rapidly from data that are freely available, after which the model may be refined as additional data become available.

The results of a simple simulation evaluation of the M3 model are also presented that demonstrates proof of model concept. These simulations serve as a basis for discussing possible model improvements and features.

2 Definition of spatial and seasonal structure

The 2015 bluefin tuna data preparatory meeting identified eight large marine areas for disaggregation of bluefin tuna data that may be used to model spatial dynamics (Figure 1a, Anon 2015a). Subsequently concerns have been expressed about the appropriateness of these spatial definitions for calculating standardized CPUE indices for the Japanese longline fleet (Kimoto et al. 2015b) and alternative spatial definitions have been proposed Figure 1b. A possible solution is including further spatial detail to encompass both of these spatial options (Figure1c).

Four subyear temporal definitions that may approximate the spatial distribution of bluefin tuna within model years have also been identified:

(subyear 1) 1st January – 15th March,
(subyear 2) 16th March – 15th May,
(subyear 3) 16th May – 15th July,
(subyear 4) 16th July – 31st December.

Concern has been raised however that size composition data may not be available at a temporal resolution finer than month (M. Lauretta personal communication). Regardless of the exact definition of spatial areas and subyear time steps, in the remainder of this document we refer to these as ‘area’ and ‘subyear’.

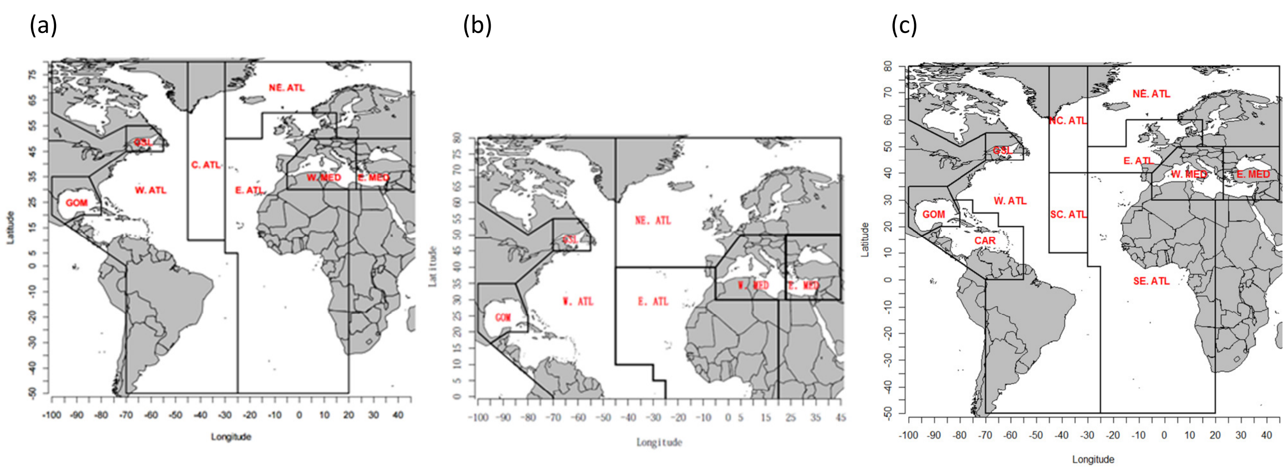


Figure 1. (a) Spatial definitions of the 2015 ICCAT bluefin tuna data preparatory meeting (Anon. 2015), (b) those proposed by Kimoto et al. (2015b) and (c) a spatial arrangement that can accommodate both spatial arrangements.

3 Structure and estimation framework of M3 operating model

3.1 Background

The most advanced multi-stock model previously applied to Atlantic bluefin tuna was MAST (Taylor et al. 2011). This model had a number of potential deficiencies which I aimed to address in the development of the M3 model.

MAST was fitted to age-composition data which is not advisable given that ages were derived directly from length data and there are numerous well established problems in the conversion of length to age (Allioud et al. 2014, Hilborn and Walters 1992, Kell and Kell 2011). To bypass these problems M3 is a statistical catch-at-length model that is fitted directly to length observations. The M3 model also requires relative abundance information by subyear and area. This places much higher demand on data-processing prior to model fitting but leads to large benefits in the estimation of spatial structure: if the model is fitted to indices disaggregated by time and space, only compatible movement dynamics may be estimated. In contrast, the MAST model could predict seasonal movements that were compatible with electronic tagging data but led to very little biomass in times/places where substantial catch rates were observed.

A principle limitation of the MAST approach is that it took a long time to run. This was related to two issues: (1) a large number of estimated parameters for exploitation rates and (2) the use of conventional tagging data to estimate exploitation rates. The first problem is resolved in the M3 model by using catch rate indices by time and area (only one catchability parameter is estimated per fleet). The second problem may be circumvented by avoiding calculation of conventional tag recapture probabilities that are in any case confounded by unknown and variable tag-reporting rates and tag shedding rates.

The M3 model is based on conventional age-structured accounting (e.g. Quinn and Deriso 1999, Chapter 8) which is common to stock assessment models such as Stock Synthesis 3 (Methot and Wetzel 2013), CASAL (Bull et al. 2012), Multifan-CL (Fournier et al. 1998) and iSCAM (Martell 2015). Similar to these assessment packages, M3 is developed using ADMB (Fournier et al. 2012) for its rapid and robust non-linear estimation performance for problems with relatively large numbers of parameters (i.e. more than 100 parameters). The more challenging aspects of developing a multi-stock spatial model relate to the modelling of movement and initializing the model.

3.2 Estimated parameters

The majority of parameters estimated by the model relate to movement probabilities and annual recruitment deviations (Table 1). The number of estimated parameters can be reduced substantially by limiting estimation to only those movements that have been recorded or are considered credible. For example, given a quarterly time step (e.g. Jan-Mar, April-Jun etc.) and the spatial definitions of the 2015 data preparatory meeting (Anon. 2015, Figure 1), an evaluation of conventional tagging for Atlantic bluefin tuna data reveals that less than 80 parameters of the 224 possible are required to characterize all of the possible movements recorded by these tagging data.

Table 1. The parameters estimated by the model. The example is for a possible bluefin tuna operating model of 8 areas (Figure 1), 4 subyears, 5 fleets, 65 years and 25 age classes.

Parameter	Number of parameters	Example	Symbol
Unfished recruitment	n_s	2	R_0
Length a modal selectivity	n_f	5	s_{mode}
Precision of selectivity	n_f	5	s_{prec}
Dome-shape of selectivity	n_f	5	s_{dome}
Recruitment deviations	$(n_{y+}n_a - 1) \cdot n_s$	178	r
Fleet catchability	n_f	5	q
Movement	Up to: $(n_r\text{-}1) \cdot (n_r) \cdot n_m$	224	ψ
Steepness (recruitment compensation)	n_s	2	h
Natural mortality rate modifier	n_s	2	M_{fac}
	Total	428	

3.3 Transition equations

The standard age-structured equations are complicated somewhat by the subyear temporal structure in which ageing and recruitment occur in a particular subyear. In this version of the model, spawning occurs for all stocks in a subyear *ms*, after subyear 1 (this is also likely to be the case in any final model fitted to bluefin tuna data since spawning in the Mediterranean and Gulf of Mexico is thought to occur after a period of movement early in the year).

Numbers of individuals *N*, for stock *s*, in a model year *y*, in the first subyear *m*=1, age class *a*, and area *r* are calculated from individuals that have moved $\vec{N}$, in the previous year, final subyear n_m, of the same age class subject to combined natural and fishing mortality rate *Z*:

1) $$N_{s,y,m=1,a,r} = \vec{N}_{s,y-1,n_m,a,r} \cdot e^{-Z_{s,y-1,n_m,a,r}}$$

where total mortality rate is calculated from annual natural mortality rate *M,* divided by the fraction of the year represented by the subyear t_m, and fishing mortality rate *F*, summed over all fleets *f*.

2) $$Z_{s,y,m,a,r} = \frac{M_{s,a}}{t_m} \sum_f F_{y,m,a,r,f}$$

Fishing mortality rate at age is derived from fishing mortality rate by length class *FL* and the conditional probability of fish being in length class *l,* given age *a* (an inverse age-length key, LAK).:

3) $$F_{y,m,a,r,f} = \sum_l FL_{y,m,l,r,f} \cdot LAK_{s,y,a,l}$$

The fishing mortality rate at length is calculated from an index of fishing mortality rate *I*, an estimated catchability coefficient *q* and a length selectivity ogive *s*, by fleet:

4) $$FL_{y,m,l,r,f} = q_f \cdot I_{y,f} \cdot s_{f,l}$$

Selectivity is calculated by the Thompson (1994) ogive and an estimate of mean length *L* of an age class *l*:

5) $$s_{f,l} = \frac{1}{1-s_{dome}} \cdot \left(\frac{1-s_{dome}}{s_{dome}}\right)^{s_{dome}} \cdot e^{s_{prec} \cdot s_{dome} \cdot (s_{mode}-L_l)} \cdot \frac{1}{1+e^{s_{prec} \cdot (s_{mode}-L_l)}}$$

In the spawning subyear *ms*, aging and recruitment occur:

6) $$N_{s,y,ms,a,r} = \vec{N}_{s,y,ms-1,a-1,r} \cdot e^{-Z_{s,y,ms-1,a-1,r}}$$

Recruitment is currently assumed to occur in user-specified spawning area for each stock *rs*. Recruitment is assumed to follow a Beverton-Holt function of spawning stock biomass *SSB* in the defined spawning areas *rs* relative to unfished spawning stock biomass *SSB0* and is subject to annual recruitment deviations *R*, for each stock.

7) $$N_{s,y,ms,1,rs} = R_{s,y} \cdot \frac{0.8 \cdot R0_s \cdot h_s \cdot SSB_{s,y}}{0.2 \cdot SSB0_{s,y} \cdot (1-h_s) + (h_s - 0.2) \cdot SSB_{s,y}}$$

where *h* is the steepness parameter (fraction of unfished recruitment at 1/5 unfished spawning stock biomass) and spawning stock biomass is calculated from moved stock numbers in the subyear prior to spawning subyear *ms*, in spawning area *rs*, weight of individuals at age *w*, and the fraction of individuals mature at age *mat*:

8) $$SSB_{s,y} = \sum_a \sum_r \vec{N}_{s,y,ms-1,a,rs} \cdot e^{-Z_{s,y,ms-1,a,rs}} \cdot w_{s,a} \cdot mat_{s,a}$$

where weight is calculated from length at age *l*:

9) $$w_{s,a} = \alpha_s \cdot l_{s,a}^{\beta_s}$$

and fraction mature at age is assumed to be a logistic function of age with parameters for the age at 50% maturity γ, and slope θ:

10) $$mat_{s,a} = 1/\left(1 + e^{(\gamma_s - a)/\vartheta_s}\right)$$

Stock numbers for subyears that are not the first subyear of the year and are not the spawning subyear are calculated:

11) $$N_{s,y,m,a,r} = \vec{N}_{s,y,m-1,a,r} \cdot e^{-Z_{s,y,m-1,a,r}}$$

In each subyear, after mortality and recruitment, fish are moved according to a Markov transition matrix *mov* that represents the probability of a fish moving from area *k* to area *r* at the end of the subyear *m*:

12) $$\vec{N}_{s,y,m,a,r} = \sum_k N_{s,y,m,a,k} \cdot mov_{s,m,k,r}$$

The movement matrix is calculated from a log-space matrix *lnmov* and a logit model to ensure each row *k*, sums to 1:

13) $$mov_{s,m,k,r} = e^{lnmov_{s,m,k,r}} / \sum_r e^{lnmov_{s,m,k,r}}$$

Movements from an area *k* to an area *r* that are considered not to be credible (e.g. from the Eastern Mediterranean to the Gulf of Mexico) are assigned a large negative number (essentially zero movement). For each area *k,* from which individuals can move, the first possible value is assigned a value of zero; subsequent possible movements are assigned an estimated parameter ψ (since rows must sum to 1 there is one less degree of freedom):

14) $$lnmov_{s,m,k,r} = \begin{cases} -1E10 & no\ movement\ from\ k\ to\ r \\ 0 & first\ possible\ movement\ from\ k\ to\ r \\ \Psi_{s,m,k,r} & other\ possible\ movements\ from\ k\ to\ r \end{cases}$$

This movement formulation limits estimation to only those movements that are possible given the data (e.g. consistent with observed tagging data).

3.4 Initializing the model

Compared with spatially aggregated models, initialization is more complex for spatial models, particularly those that may need to accommodate movement by age and include regional spawning and recruitment. The solution used here is to iterate the transition equations above (Equations 1, 6, 7, 11, 12) given zero fishing mortality until the spatial distribution of stock numbers converges for each of the subyears.

Prior to this iterative process an initial guess at the spatial and age structure of stock numbers $\widehat{N}$ is made using the estimated movement matrix and natural mortality rate at age *M*:

15) $$\widehat{N}_{s,m,a,r} = R0_s \cdot \mathrm{e}^{-\sum_1^a M_{s,a}} \cdot \sum_k \frac{1}{n_r} \cdot mov_{s,m,k,r}$$

It typically takes between 50 and 100 iteration years of unfished conditions for stock numbers to converge to within 1/10 of a percent of the previous iteration. To ensure stability of the estimation, a fixed number of iterations is defined by the user.

3.5 Predicting data

For each fleet *f*, total predicted catches in weight $\hat{C}$, are calculated from the Baranov equation:

16) $$\hat{C}_{y,m,r,f} = \sum_s \sum_a w_{s,a} \cdot N_{s,y,m,a,r} \cdot (1 - e^{-Z_{s,y,m,a,r}}) \cdot \left(\frac{F_{y,m,a,r,f}}{Z_{s,y,m,a,r}}\right)$$

Similarly predicted catches in numbers at age *CAA*, is given by:

17) $$\widehat{CAA}_{s,y,m,a,r,f} = N_{s,y,m,a,r} \cdot (1 - e^{-Z_{s,y,m,a,r}}) \cdot \left(\frac{F_{y,m,a,r,f}}{Z_{s,y,m,a,r}}\right)$$

This can be converted to a prediction of total catches in numbers by length class *CAL* using a stock specific inverse age-length key, *LAK*:

18) $$\widehat{CAL}_{y,m,l,r,f} = \sum_s \sum_a \widehat{CAA}_{s,y,m,a,r,f} \cdot LAK_{s,y,a,l}$$

The model predicts spawning stock biomass indices $\widehat{Issb}$, that are standardized to have a mean of 1 for each stock over the total number of years *ny*:

19) $$\widehat{Issb}_{s,y} = n_y \cdot SSB_{s,y} / \sum_y SSB_{s,y}$$

The model predicts vulnerable biomass indices $\hat{I}$, by fleet that are standardized to have a mean of 1 for each fleet:

20) $$\hat{I}_{y,m,r,f} = n_y \cdot n_m \cdot n_r \cdot V_{y,m,r,f} / \sum_y \sum_m \sum_r V_{y,m,r,f}$$

Where vulnerable biomass *V* is calculated:

21) $$V_{y,m,r,f} = \sum_l \left(s_{f,l} \cdot \sum_s \sum_a \left(N_{s,y,m,a,r,f} \cdot ALK_{s,y,a,l} \cdot w_{s,a}\right)\right)$$

The model predicts stock of origin composition of catches $\widehat{SOO}$, from predicted catch numbers at age:

22) $$\widehat{SOO}_{s,y,m,r,f} = \sum_a \widehat{CAA}_{s,y,m,a,r,f} / \sum_s \sum_a \widehat{CAA}_{s,y,m,a,r,f}$$

3.6 Likelihood functions, priors and the global objective function

The likelihood functions for the various data types are summarized in Table 2.

Table 2. Summary of the likelihood function for various data

Type of data	Disaggregation	Likelihood function
Total catches (weight)	year, subyear, area, fleet	Log-normal
Index of vulnerable biomass (e.g. a CPUE index)	year, subyear, area, fleet	Log-normal
Index of spawning stock biomass (e.g. a larval survey)	year, stock	Log-normal
Length composition	year, subyear, area	Multinomial
PSAT tag (known stock of origin)	stock, year, subyear, area	Multinomial
PSAT tag (unknown stock of origin)	year, subyear, area	Multinomial
Stock of origin	Year, subyear, area	Multinomial

A log-normal likelihood function was assumed for total catches by fleet. The log-likelihood was calculated:

23) $$OBJ_c = \sum_y \sum_m \sum_r \sum_f \frac{log(\sigma_{catch}) + \left(\log(\hat{C}_{y,m,r,f}) - \log(C_{y,m,r,f})\right)^2}{2 \cdot \sigma_{catch}^2}$$

Similarly the log-likelihood component for indices of vulnerable biomass and spawning stock biomass were calculated:

24) $$OBJ_i = \sum_y \sum_m \sum_r \sum_f \frac{log(\sigma_{index}) + \left(log(\hat{I}_{y,m,r,f}) - log(I_{y,m,r,f})\right)^2}{2 \cdot \sigma_{index}^2}$$

25) $$OBJ_{SSB} = \sum_s \sum_y \frac{log(\sigma_{SSB}) + \left(log(\widehat{Issb}_{s,y}) - log(Issb_{s,y})\right)^2}{2 \cdot \sigma_{SSB}^2}$$

The length composition data are assumed to be distributed multinomially. In traditional stock assessment settings catch composition data may often dominate the likelihood function due to the large number of observations. This is exacerbated by a failure to account for non-independence in size composition samples. There are two possible solutions: (1) manually specify the effective sample size (ESS) of length-composition samples or (2) use a multinomial likelihood function that includes the conditional maximum likelihood estimate of the ESS (perhaps even a freely estimated ESS, S. Martell personal communication). In this version of the code, ESS is user-specified.

The log-likelihood component for length composition data is calculated:

26) $$OBJ_{CAL} = -\sum_y \sum_m \sum_l \sum_r \sum_f CAL_{y,m,l,r,f} \cdot log\left(\hat{p}_{y,m,l,r,f}\right)/ESS_f$$

Where the model predicted fraction of catch numbers in each length class *p*, is calculated:

27) $$\hat{p}_{y,m,l,r,f} = \widehat{CAL}_{y,m,l,r,f} / \sum_l \widehat{CAL}_{y,m,l,r,f}$$

Similarly the log-likelihood component for PSAT tagging data of known stock of origin (SOO), released in year *y*, subyear *m*, area *r* and recaptured in year *y2,* subyear *m2*, and area *k* is calculated:

28) $$OBJ_{PSAT} = -\sum_s \sum_y \sum_m \sum_{y2} \sum_{m2} \sum_r \sum_k PSAT_{s,y,m,y2,m2,k} \cdot log\left(\hat{\theta}_{s,y,m,y2,m2,r,k}\right)$$

where recapture probabilities ϑ, are calculated by repeatedly multiplying a distribution vector *d*, by the movement probability matrix *mov*. For example for a tag released on a fish of stock 1 in year 2, subyear 3, and area 4, the probability of detecting the tag in year 3, subyear 2 for the various areas is calculated:

29) $$\hat{\theta}_{s=1,y=2,m=3,y2=3,m2=2,r=4,1:n_r} = \left(\left(d \cdot mov_{s,m=3}\right) \cdot mov_{s,m=4}\right) mov_{s,m=1}$$

where

30) $$d_k = \begin{cases} 0 & k \neq r \\ 1 & k = r \end{cases}$$

The log-likelihood component for PSAT tagging data of unknown stock of origin PSATu, is currently weighted according to the compound probability that a fish is of a particular stock given the track history for that tag. For example for a tag *t*, tracked in series of years y_i, subyears m_i, and regions r_i, the weight *w,* of that tag for a specific stock is calculated:

31) $$w_{t,s} = \frac{\prod_i [(\sum_a N_{si,yi,mi,ai,ri}) / (\sum_s \sum_a N_{si,yi,mi,ai,ri})]}{\prod_i [1 - (\sum_a N_{si,yi,mi,ai,ri}) / (\sum_s \sum_a N_{si,yi,mi,ai,ri})]}$$

This is simply the product of fractions of that stock in those time-area strata divided by the product of the fractions of other stocks in those time-area strata. An alternative approach would be to compare the relative probabilities of the observed movements among the stocks although it is unclear whether this circularity (PSAT data are a primary source of information regarding movement) could lead to estimation problems.

The weighted likelihood function is similar to that of the stocks of known origin but includes the appropriate weighting term for each tag

32) $$OBJ_{PSATu} = -\sum_t \sum_s \sum_y \sum_m \sum_{y2} \sum_{m2} \sum_r \sum_k PSATu_{t,s,y,m,y2,m2,k} \cdot log\left(\hat{\theta}_{s,y,m,y2,m2,r,k}\right) \cdot w_{t,s}$$

The log-likelihood component for stock of origin data *SOO* was also calculated assuming a multinomial distribution:

33) $$OBJ_{PSATu} = -\sum_{\boldsymbol{s}} \sum_y \sum_m \sum_r \sum_{\boldsymbol{f}} SOO_{s,y,\boldsymbol{m},\boldsymbol{r},\boldsymbol{f}} \cdot log\left(\widehat{SOO}_{s,y,m,r,f}\right)$$

In addition to these likelihood functions for observed data, priors may be placed on the steepness parameter *h*, of the stock recruitment relationship and a factor *Mfac*, multiplied by the user specified natural mortality rate-at-age schedule *Minit*.

34) $$M_{s,a} = Minit_{s,a} \cdot Mfac_s$$

The factor applied to the natural mortality rate-at-age schedule is assumed to be lognormally distributed according to user specified mean and standard deviation parameters.

35) $$OBJ_M = \sum_s \frac{log(\sigma M_s) + (Mfac_s - \mu M_s)^2}{2 \cdot \sigma M_s^2}$$

Steepness is parameterized by a logit model constrained between 0.2 and 1:

36) $$h_s = 0.2 + 0.8 \cdot e^{\hat{h}_s} / \left(1 + e^{\hat{h}_s}\right)$$

In the $logit^{-1}$ space, a normal prior is adopted for this transformed steepness $\hat{h}$, parameter that includes user specified mean $\widehat{\mu h}$, and standard deviation $\widehat{\sigma h}$, parameters. The corresponding log-likelihood component is:

37) $$OBJ_h = \sum_s \frac{log(\sigma h_s) + \left(\hat{h}_s - \mu h_s\right)^2}{2 \cdot \sigma h_s^2}$$

The global objective function OBJ_T, to be minimized is the summation of the weighted , likelihood components:

38) $$OBJ_T = \omega_c \cdot OBJ_c + \omega_i \cdot OBJ_i + \omega_{SSB} \cdot OBJ_{SSB} + \omega_{CAL} \cdot OBJ_{CAL} + \omega_{PSAT} \cdot OBJ_{PSAT} + \omega_{PSATu} \cdot OBJ_{PSATu} + \omega_M \cdot OBJ_M + \omega_h \cdot OBJ_h$$

4 Data to inform operating models

A coarse description of Metadata to support bluefin tuna operating model development is presented in Table 3.

4.1 CPUE indices

The operating model for Atlantic bluefin tuna is intended to be fitted to standardized catch rate data at the resolution of area and subyear. Providing relative abundance data at this resolution anchors the estimation of movement to only those scenarios that maintain a credible spatial distribution of vulnerable biomass. Models that include only annual indices have no such constraint and can predict movement and spatial distribution from tagging data that are not credible given fishery catch rate data (for example an absence of fish where observed seasonal catch rates are substantial).

It follows that when fitting spatial models, the value of relative abundance indices at the same resolution as estimated movements can be very high. Previous simulation evaluations have shown that reliable estimation of spatial distribution can be obtained from only spatial catch rate data and characterizing this reliably is a primary concern for the estimation of variables of management interest such as stock depletion and MSY reference points (Carruthers et al. 2011b). The same spatial catch rate data can also provide additional information about age- or size-specific movement if fleets have varying size selectivities. Since these indices constrain movement estimation to observed spatial distributions, the information content of electronic tagging can be used to explore additional characteristics of movement such as temporal variability or different movement of juvenile and mature fish. Since commercial catch rate data are generally reported at a sufficient temporal and spatial resolution to produce these indices at the scale of area and subyear, a relatively large quantity of information about spatial distribution is provided to the operating model with no additional data collection requirements.

This relatively strict data requirement also ensures that the model does not require the estimation of a very large number of fishing mortality rate parameters (e.g. up to 12,480 free F parameters estimated for an eight area model, with four subyears, 65 historical years of exploitation and 6 fleets). By including an index of fishing mortality rate (aka 'partial F'), just one catchability parameter q is estimated per fleet: $F=qI$ (e.g. 6 q parameters as opposed to 12,480 F parameters, Figure 2)

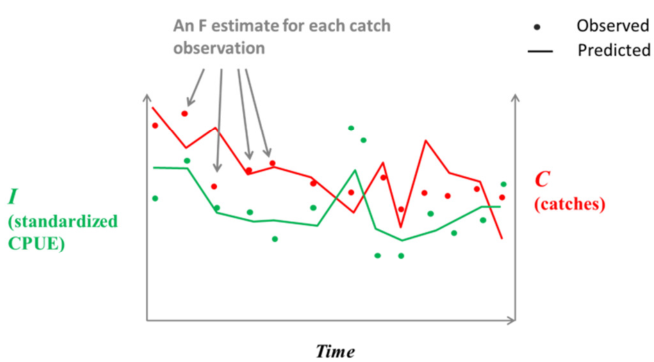


Figure 2. Rather than predicting each catch observation based on an individual estimated fishing mortality rate, catches may be divided by the relative abundance index to provide an index of fishing mortality rate and thus a single catchability parameter is required for each fleet over all areas, subyears and years.

The Japanese longline and US longline are important standardized indices for fitting operating models due to their relatively wide spatial and temporal coverage in mixing areas of the North Atlantic. A proposed US-Canadian combined longline index may provide additional continuity over a wider spatial range that covers the range of the population in the West Atlantic Ocean (Lauretta et al. 2015, SCRS-2015-171). In the Mediterranean, Moroccan, Portuguese, Spanish and Italian trap catch rate data may be used to provide indices for eastern and western areas. While these indices are a priority for operating model development, the calculation and testing of standardizing catch rates can be time intensive. We propose an interim solution in which freely available data, requested and organized by ICCAT, are used in preliminary operating models which are later updated with standardized indices. The most relevant data in this regard are the nominal (non-standardized) Task II catch and effort data, much of which is available at the temporal resolution of day and spatial resolution of 5x5 degree ocean square (ICCAT 2015b) consistent with the defined areas. These data do not include covariates that can be used to account for shifts in fishing practices that are typically included in catch rate standardization. Effects of spatial expansion and contraction may still be accounted for (e.g. Ahrens and Walters 2005; Carruthers et al. 2011a).

For a number of reasons (e.g. changes in spatial distribution, gear changes, species targeting), catch rate indices may not be linearly related to true changes in vulnerable biomass and may decline faster (hyperdepletion) or slower (hyperstability) than vulnerable biomass. Scenarios that incorporate hyperstability and hyperdepletion may be important for fitting operating models to historical relative abundance indices and also proposing observation models for MSE.

4.2 Larval indices

Recent papers have suggested that indices derived from larval surveys may be closely correlated with stock assessment predictions of spawning stock biomass (SCRS/2015/035). Larval surveys have been carried out in the Gulf of Mexico (1977-2013) and more recently the western Mediterranean (2001-2005, 2012-2013). It is possible that the operating model could be fitted to these indices under the assumption they are representative of trends in spawning stock biomass in spawning areas. If used in this way, these indices can be expected to strongly influence model predictions of trends in spawning stock biomass (since their interpretation is not contingent on estimates of selectivity and movement). Adding indices of spawning stock biomass may also indirectly inform age-based movement, maturity and selectivity of fishing in spawning areas.

4.3 Catches

The operating model predicts exploitation rate at the scale of subyear and area. It follows that catches are required at this scale and a fleet must be assigned to all catches (this may be an umbrella 'other' fleet of general size selectivity). In order to meet custom subyear definitions above it may be necessary to manually raise Task II catches for subyears and areas to sum to total reported landings.

Catch data are generally the only source of scale in stock assessments: if catches are reported in kilograms the estimates of MSY, catch recommendations and the estimate of current biomass will be a number 1000 times larger than the same catches reported in metric tonnes. It follows that consistent bias in catches across a time series (for example consistent 25% underreporting of catches) has no effect on estimates of stock depletion and fishing mortality rate and therefore is ignorable when specifying operating models (for which scale is not important unless a specific catch recommendation is to be considered). However temporal patterns in bias in catches can be significant (for example significant reductions in illegal, unreported and unregulated fishing of the eastern stock after 2008, ICCAT 2014b) and may provide contrasting view of historical stock trends and exploitation rate (e.g. Carruthers et al. 2015c). It follows that scenarios for catch underreporting (and related quantities such as dead discarding) could be used as alternative hypotheses that may be represented by operating models.

In terms of MSE results and selection of management procedures, a potentially more important issue is the level of bias in future observations of catches that is simulated in the management strategy evaluation. Carruthers et al. (2015d) found bias in catches to be amongst the most influential of observation processes determining the performance of management procedures. In their study, catch observation bias was often much stronger determinant of MP performance than current stock depletion or historical trend in exploitation rate (estimated by the fitted operating models potentially arising from historical biases in catches). It is therefore important to carefully consider plausible scenarios for biases in reported catches in the future.

4.4 Catch compositions

Currently the operating model is intended to be fitted to length sample data provided to ICCAT (Task II size). The majority of these data are reported at a sufficiently fine spatial and temporal resolution to be aggregated to the subyear and area definitions of the operating model. These data are the primary source of information regarding the size-selectivity of fishing. The representativeness of these data, their spatial-temporal coverage and the contribution to exploitation of the associated fleets is likely to determine how fleets are defined in the operating model.

A fleet in the operating model represents a distinct size selectivity of substantial exploitation rate. The purpose of including fleet heterogeneity in the operating model is to accurately characterize patterns in the exploitation of size classes in order to best estimate quantities of interest such as stock status and fishing mortality rate at maximum sustainable yield. Each fleet in the operating model must have paired observations of catch and an index of fishing mortality rate *I* (catch divided by a standardized index of abundance) by area and subyear. Each fleet must also have some size composition data. The catches of fleets that do not have both composition data and a standardized index of fishing mortality rate must be aggregated with catches of a fleet of similar size selectivity (that does have both composition data and an index of fishing mortality rate).

Due to the computational overhead, it is not possible to model the large number of flag and gear combinations recorded by ICCAT which have caught bluefin tuna historically. However it is relatively straightforward to aggregate fleets of similar size selectivity for model fitting and then predict expected individual catch rates *post-hoc*. This allows for parsimonious operating models (avoid including many fleets) while retaining predictive capacity for catch rates of individual fleets which may be a critical component of stakeholder utility.

A persistent issue in the fitting of integrated stock assessment models is the correct weighting of various sources of data (Candy 2008, Francis 2011, Maunder and Punt 2013). This is particularly applicable to size composition data that can often dominate the global objective function due to the number of observations and likelihood functions that are typically assumed for these data (e.g. a multinomial observation model). By fitting to length-composition data we bypass many of the issues associated with deriving age of fish from length data but do not avoid problems associated with overweighting composition data due to incorrectly accounting for non-independence among observations. For example bluefin tuna are observed to school in size mono-specific groups and hence multiple observations of length from a fishing set may in fact represent a single independent length observation. It may be the case that operating model results are sensitive to specification of this 'effective sample size'. It follows that various scenarios for fleet aggregation and effective sample size may be considered as alternative hypotheses to be represented by operating models in future MSEs.

4.5 Conventional tagging data

Conventional tagging data provide valuable information about the range of movements that are possible for Atlantic bluefin tuna and provide a basis for formulating alternative hypotheses about movement. Conventional tagging data have been used in previous bluefin tuna assessment models to estimate fishing mortality rates (and hence abundance) by quarter and area (MAST, Taylor et al. 2011). These approaches have assumed a prior for reporting rate that is constant over time, space and among fleets. There is however evidence that reporting rates vary in time, space and among fleets by factors as great as 500 (Carruthers et al. 2011, Carruthers and McAllister 2010, Hillary et al. 2008). Similarly large disparities between predicted catches and tag recapture probabilities can be expected if these variable reporting rates are not accounted for.

A fleet specific prior for reporting rate may be prescribed (e.g. Carruthers and McAllister 2010). The principal problem with this approach is that these reporting rate estimates are generally sensitive to alternative assumptions and can be as much as 1/3 or 3 times base-case estimates. While the difference between a reporting rate prior mean of 1/1000 or 3/1000 may not seem large, in relative terms this is highly uncertain and consistent with predictions of catches and abundance that differ by a factor of 3.

It is possible to estimate reporting rates by fleet inside the operating model but since these are confounded with recapture probability these data may only weakly inform abundance. This would not be a serious problem if the computational overhead associated with conventional tagging data was small. However it can be as large as 10 times the total number of other calculations slowing estimation speed by a similar margin (more so if reporting rates are estimated). An operating model of 4 subyears, 65 historical years, 8 areas and 6 fleets requires the calculation of approximately ~14M recapture probabilities in addition to the likelihood calculation for recaptured tags.

Many tags have been recaptured by observer programs for which it may be assumed that reporting rates are 100%. In order to retain information regarding recapture probability (and hence abundance) it is possible to fit to only observer tag recapture data and provide post-hoc estimates of reporting rates (also observed versus predicted recaptures) for the various commercial fleets.

The preliminary version of the operating model is designed to use conventional tagging data to identify the range of possible bluefin tuna movements and also provide information about possible shifts in growth rate of bluefin tuna. The latter is a novel use of conventional tagging data in the context of integrated stock assessment. It has been hypothesized that as exploitation rates increase there should be a shift in the size composition of the modelled population due to the attrition of faster growing individuals (Walters and Martell 2004). Traditional length-based assessments (e.g. MULTIFAN-CL, Fournier et al. 1998) have not accounted for this phenomenon and instead, renew the distribution of growth types in each cohort as it ages, irrespective of the level of exploitation. In constrast, a cohort may start with a normal distribution of maximum length among individuals which becomes strongly positively skewed as larger fish are preferentially removed from a cohort (perhaps at high exploitation rates, older fish would belong to the very lower tail of the unfished distribution of maximum length). Simulation studies indicate that estimates of fishing mortality rates from models that do not account for this phenomenon could be strongly positively biased (as much as 3 times that of true fishing mortality rate, C. Walters and A. Hordyk, personal communication).

The operating models under development include a dynamic inverse age-length key that accounts for attrition of faster/larger growth type groups in order to investigate this potentially important source of model mis-specification (SCRS/2015/179). Conventional tagging data could provide a valuable empirical validation of shifts in growth rate predicted by the model since each observation of a release and recapture provides an estimate of the growth parameters (e.g. Walters and Essington 2010). Given it may strongly impact estimates of stock status and productivity, growth type attrition may be considered an important alternative hypothesis to be captured in Atlantic bluefin tuna operating models.

4.6 Surgically implanted archival tags

Surgically implanted archival tags (SI tags, e.g. Walli et al. 2009, LPRC 2015) provide detailed tag track information and have been included in previous spatial population models to predict both movement and exploitation rates (Taylor et al. 2011). While reporting rates of archival tags are generally much higher than conventional tags (due to larger rewards for reporting archival tags), uncertainty over reporting rates and post-release mortality rate still complicates their use in the estimation of exploitation rate. SI tags could be used (1) to estimate exploitation rate and abundance with some assumption about reporting rate, (2) to estimate only recapture probabilities of tags reported by fleets of known reporting rate (e.g. observer fleets), (3) to estimate only movement assuming that releases and recaptures are independent of fishing (similarly to PSAT tags).

4.7 Pop-off Satellite Archival Tags (PSATs)

The primary source of information regarding movements from model areas to model areas among subyears comes from PSAT tags (Block et al. 2005, Lutcavage et al. 2012, Cermeno et al. 2015). The operating model includes these data formatted into separate subyear segments (i.e. has the fields 'year-from', 'area-from', 'area-to', 'subyear-from', 'age-from').

The majority of PSAT tags do not have a definite stock of origin (SOO) (they were not tagged in a spawning area or ocean area specific to a single stock). In such cases it is not clear how to weight these data such that information about movement can be correctly attributed to stock. Determining SOO (or SOO weights) may be undertaken prior to model fitting based on other SOO data. Alternatively this calculation can occur inside the model using the relative likelihood of SOO based on either (1) the model predicted composition of stocks in the areas of a tag track (as in MAST, Taylor et al. 2011) or (2) the predicted likelihood of tag track movements given estimated movement parameters of stocks. Both methods that occur inside the model can be expected to be unreliable since they are likely to lead to estimated movements that best divide the tags of unknown SOO (i.e. they can be expected to underestimate stock mixing simply due to the likelihood weighting). This potential source of bias should be simulation tested. The way in which SOO is assigned to PSAT tags may impact estimates of spatial distribution and movement and hence management reference points, and could be considered as alternative scenarios for operating models.

It may be possible to estimate spatial distribution of bluefin tuna in the absence of PSAT data using simplified spatial models that do not attempt to characterize the full range of movement among areas (e.g. the gravity models of Carruthers et al. (2011) and MAST Taylor et al. (2011)). However PSAT data contain vital information for evaluating alternative movement hypotheses such as age-based movement and temporally variable movement and their collection is a priority for operating model development.

4.8 Otolith microchemistry

It is generally not possible to estimate stock size and mixing using a multi-stock model without external data regarding stock of origin: without these data the magnitudes of the various stocks are confounded. Previous multi-stock models such as MAST have used otolith microchemistry data as the primary source of information regarding stock of origin (e.g. Rooker et al. 2008, Rooker et al. 2014). The challenge is processing microchemistry data at the resolution of subyear and area. While these data have been gathered in all of the areas of the operating model, they have not been processed at this resolution and only exist from 2005 onwards (i.e. Rooker et al. 2008)

Given that ICCAT data are already freely available that can be used to characterize trends in abundance, catch compositions and spatial distribution, data that provide information about stock of origin are arguably the most important for fitting a preliminary bluefin tuna operating model (after which other data sources can be refined).

4.9 Otolith shape analysis

Otolith shape analysis is an alternative approach for characterizing stock mixing which is both cheaper than micro-chemical analysis and also non-destructive. The method has been compared with traditional micro-chemical analysis and shows promise, achieving similar accuracy (SCRS/P/2015/004). Further detail on the approach is summarized in the report of the 2015 data preparatory meeting (ICCAT 2015a). Similar to other data that provide information regarding stock of origin, these data are particularly valuable in the early stage of operating model development.

4.10 Single nucleotide polymorphism (SNP)

SNP data provide another basis for assigning individuals to stock of origin and have been collected for both mixing areas in the Atlantic and spawning areas in the Gulf of Mexico and Mediterranean. The most recent SNP analyses (e.g. SCRS/2015/048) have made use of a larger number of genetic markers and have demonstrated an ability to accurately determine sub-stock structure such as spawning area (e.g. Gulf of Mexico, Balearic Sea, Strait of Sicily and Levantine Sea). It follows that SNP data could provide the necessary empirical basis for formalizing hypotheses that have been proposed for sub stock structure (e.g. Kell et al. 2011, Fromentin and Lopuszanski 2014), particularly for the Eastern stock that is thought to have spawning areas in western, central and eastern Mediterranean.

An ongoing challenge for operating model development is generating model datasets that can inform a model with more than one Mediterranean sub-stock. In principle this is possible if SNP data can be used to quantify fractions of individuals in each sub-stock in sub-areas of the Mediterranean and there are sufficient electronic tags (that can be assigned to each Mediterranean sub-stock) to estimate movement. Such data would still have to be provided at the resolution of subyear and area after which they may be too sparse to be considered a reliable characterization of sub-stock structure.

4.11 Other genetics data

Coupled with electronic tagging data, older genetic studies based on microsatellites (Carlsson et al. 2007) and mitochondrial DNA (Boustany et al. 2008) provided early confirmation of the broad east-west stock structure. It is not clear whether these data could be processed at the level of subyear and area in order to inform operating models. However this may be worth pursuing since these data broaden the temporal range of data on SOO.

Another potential for obtaining abundance information is close-kin analysis to estimate the spawning stock abundance (Bravington et al 2013), This provides the potential for genetic mark recapture experiments to estimate absolute abundance, mortality rates or migration, addressing directly some of the key uncertainties in BFT assessments. In particular the close-kin analysis could provide a fishery-independent estimate of spawning stock numbers, particularly for Western Atlantic Bluefin.

4.12 Fishery independent surveys

In addition to the fishery-independent larval surveys discussed above, aerial surveys have been suggested as potential source of index information. Aerial surveys (Bonhommeau et al. 2010) have been conducted in the Mediterranean since 2010 and may represent the longest running index of spawning biomass for the eastern stock. Similarly to larval indices, the aerial survey may strongly influence operating model estimates of recent trends in spawning biomass. Since the aerial surveys in the Mediterranean have covered areas that encapsulate eastern sub-stock structure, these data may also be used to inform alternative sub-stock structure hypotheses.

Similar aerial surveys have been proposed for the Western stock in addition to a western hydro acoustic survey and a hydro acoustic curtain in the Strait of Gibraltar. While such proposals may not necessarily support operating model development, these data may support management procedures in the future. It follows that it is important to characterize metadata for these types of future data collection programs.

4.13 Growth and aging

The operating model is fitted to catch-at-length data and predicts fishing mortality rate by length class for each fleet. The operating model therefore requires an inverse age-length key (probability of an individual being in a length class given it is of an age class) to convert estimates of fishing mortality rate at length to fishing mortality rate at age. The operating model either (1) applies a static, user-specified inverse age-length key or (2) attempts to generate a dynamic inverse age-length key from model predicted fishing mortality rate and user-specified variability in individual growths. In either case, growth must be characterized in order to develop these keys and also predict the weight of individuals of a given age class. Since the inverse age-length keys are a requirement of the operating model and are not yet available, their derivation is priority for developing a preliminary operating model. Accounting for attrition of growth types could be an important axis of uncertainty for alternative operating models.

4.14 Maturity

There have been numerous studies focused on bluefin tuna reproductive biology with which to characterize maturity at age and fecundity in general (e.g. Corriero et al. 2005, Medina et al. 2007, Diaz 2010, Aranda et al. 2013, Knapp et al. 2014). However participants of the recent data-preparatory meeting (ICCAT 2015) concluded that the body of work on bluefin tuna reproduction should be reconsidered prior to the next assessment. It may be possible to identify a various hypotheses for bluefin tuna maturity and fecundity that are represented by operating models. Participants of the same meeting (ICCAT 2015) also reiterated an ongoing concern that the age-at-maturity estimated from fish on the spawning grounds may not be representative of reproductive contribution at age of the wider population. Again this source of uncertainty may be formalized in operating models.

Table 3. Simplified summary of data to inform operating models for Atlantic bluefin tuna. Spatial range either refers to the areas of Figure 1 or stock (Es = eastern stock, Ws = western stock).

Type of data (*Informs*)	Year range	Til	Spatial range	Can be by season?	Contact	Collab	Available to: TC	CMG	ICCAT	ALL	Used in OM?
1. CPUE indices ***(relative abundance, movement, performance at stakeholder level)***											
1.1. ICCAT task II CPUE	1950-2014	∞	All	Y	Carlos Palma (ICCAT)	Y	Y	Y	Y	Y	Yes
1.2. Japanese LL standardized spatial	1976-2013	∞	E, NE, W, C	Y	Ai Kimoto	Y	N	N	N	N	Not yet
	1990-2013	∞		Y		Y	N	N	N	N	Not yet
1.3. USA LL standardized spatial	1992-2014	∞	W	Y		Y	N	N	N	N	Not yet
	1992-2004	∞	GOM	Y		Y	N	N	N	N	Not yet
	2005-2014	∞	GOM	Y	Matt Lauretta (NOAA)	Y	N	N	N	N	Not yet
1.4. USA HL standardized spatial	1980-2014	∞	W	Y		Y	N	N	N	N	Not yet
1.5. USA RR standardized spatial	1992-2014	∞	W	Y							
1.6. USA-CAN LL standardized spatial	1992-2014	∞	W, C	Y	M. Lauretta (NOAA) /	Y	N	N	N	N	Not yet
1.7. USA-CAN HL standardized spatial	1993-2014	∞	W, C	Y	A. Hanke (DFO)	Y	N	N	N	N	Not yet
1.8. CAN LL standardized		∞	W, GSL	Y		Y	N	N	N	N	Not yet
1.9. CAN HL standardized	1981-2014	∞	GSL	Y	Alex Hanke (DFO)	Y	N	N	N	N	Not yet
	1988-2014	∞	W	Y		Y	N	N	N	N	Not yet
1.10. TWN LL standardized	1960-2004	2004	W, NE, E	Y	Julia Huang (NTOU)		N	N	N	N	Not yet
1.11. MOR TRAP standardized	1981-2014	∞	WM	Y	N. Abid		N	N	N	N	Not yet
1.12. POR TRAP standardized			W, WM	Y			N	N	N	N	Not yet
1.13. ESP TRAP standardized			W, WM	Y	Jose Miguel de la Serna		N	N	N	N	Not yet
1.14 ITA TRAP standardised			CM	Y	Pierantonio Addis	Y		N	N	N	Not yet
2. Larval indices ***(SSB, movement)***											
2.1. USA	1977-2013	∞	GOM	Y	Walter Ingram (NOAA)	Y	N	N	N	N	Not yet
2.2 ESP	01-'05 '12-'13	2018	W Med	Y	Franciso Alemany (IEO)	Y	N	N	N	N	Not yet
3. Catches ***(stock size, harvest rate)***											
3.1. ICCAT task I	1950-2015	∞	non-spatial	N	Carlos Palma (ICCAT)	Y	Y	Y	Y	Y	Yes
3.2. ICCAT task II			All	Y		Y	Y	Y	Y	Y	Yes
3.3 GBYP	1512-1950		E, M	Y	Carlos Palma (ICCAT)	Y	Y	Y	Y	Y	Not yet
4. Catch composition ***(selectivity, depletion)***											
4.1. ICCAT catch-at-size	1950-2015	∞	All	Y	Carlos Palma (ICCAT)	Y	Y	Y	Y	Y	Yes
4.2. Stereo video caging	2014	ended	WM, EM	Y	Mauricio Ortiz (ICCAT)		N	N	N	N	Not yet
4.3. Canadian fisheries				N	Alex Hanke (DFO)						
4.4 GBYP Historical catches	1910-1950	=	E, M	Y	Carlos Palma (ICCAT)	Y	N	Y	Y	Y	Not yet
5. Conventional tags **(*feasible movement, growth, GTG heterogeneity*)**											
5.1. ICCAT	1954-2014	2015	All	Y	Carlos Palma (ICCAT)	Y	Y	Y	Y	Y	Yes
6. SI archival tags **(*feasible movement*)**											
6.1. LPRC (n=4000)	2011-2015		W	Y	Molly Lutcavage		N	N	N	N	Not yet
7. PSAT tags **(*movement*)**											
7.1. LPRC (n=423)	2005-2009	ended	W	Y	Molly Lutcavage		N	N	N	N	Not yet
7.2. DFO (n=135)	2013-2015	∞	GSL,W,GOM	Y	Alex Hanke (DFO)	Y	N	N	N	N	Not yet
7.3. Stanford (n=1783)	1996-2010	∞	W	Y	Barbara Block		N	N	N	N	Not yet
7.4. GBYP (n = 103)	2012-2014	2015	E,MED	Y	Antonio Di Natale	Y	Y	N	N	N	Not yet
7.5. WWF (n = 100)				Y	Pablo Cermeno		N	N	N	N	Not yet
7.6. SEFSC (NOAA)	2011-2013		GOM,W,GSL	Y	Craig Brown		N	N	N	N	Not yet
7.7. Acadia (NS)			GSL	Y	Mike Stokesbury		N	N	N	N	Not yet
7.8. UCA	2011	ended	W, C, WM	Y	Antonio Medina	Y	Y	N	N	N	Not yet

Table 3 continued.

8. Otolith microchemistry (*stock of origin*)											
8.1. UMCES, TAMU	2012-2013			Y	David Secor	Y	N	N	N	N	Not yet
8.2. NOAA							N	N	N	N	Not yet
8.3. EU (AZTI)	2009-2011	ended	E	Y	Igaratza Fraile	Y	N	N	N	N	Not yet
8.4. DFO / UMCES	2011-2013	∞	W, GSL	Y	Alex Hanke (DFO)		N	N	N	N	Not yet
8.5 GBYP	2011-2015		All	Y	GBYP	Y		Y	Y		Not yet
9. Otolith shape analysis (*stock of origin*)											
9.1. GBYP GMIT (n=718)	2013	2015	E, W, C, WM	Y	Deirdre Brophy		N	N	N	N	Not yet
10. SNP (*population structure, genetic structure*)											
10.1. Med HCMR					Gianpaolo Zmpicinini		N	N	N	N	Not yet
10.2. GBYP UB	2011-2015		All		Gregory Puncher	Y	N	N	N	N	Not yet
10.3. AZTI (n=130)					Naiara Rodriguez Espeleta	Y	N	N	N	N	Not yet
10.4 NOAA/VIMS/CSIRO	2015		GOM/MED	N	John Walter		N	N	N	N	Not yet
10.5 GBYP Historical UB	200 BC - 1927		E, M	Y	Alessia Cariani	Y	N	N	N	N	Not yet
11. Other genetics on population structure (*population structure, genetic structure*)											
11.1. mtDNA					Barbara Block		N	N	N	N	Not yet
11.2. Micro Sat/ mtDNA (n=320 / 147)	2003	ended	GOM, WM	Y	Carlsson		N	N	N	N	Not yet
12. Fish. Ind. surveys (*relative abundance, movement*)											
12.1. ICCAT Aerial	2010-2015		M	Y	Antonio Di Natale	Y	N	N	N	N	Not yet
12.2. USA Aerial	2015-		W	Y	Molly Lutcavage		N	N	N	N	Not yet
12.3. USA Acoustic	2015-		W	Y	Molly Lutcavage		N	N	N	N	Not yet
12.4. SOG Hydro acoustic curtain (OTN)	proposed		W, WM	Y	Mike Stokesbury		N	N	N	N	Not yet
13. Growth, aging (*age-length keys, length-age keys*)											
13.1. Age-length keys (NOAA)				Y	John Walter		N	N	N	N	Not yet
13.2. Age-length keys (IEO)	2010-2012	ended	E, WM	Y	Enrique Rodriguez-Marin		N	N	N	N	Not yet
13.3. Age-length keys (DFO)	2010-2013	ended	GSL, W	Y	Alex Hanke (DFO)		N	N	N	N	Not yet
13.4. Derived from tagging	1963-2012	ended	Es, W s	Y	Lisa Allioud		N	N	N	N	Not yet
13.5 Age-length keys (GBYP)	2011-2015		E, M	Y	Antonio Di Natale	Y	N	Y	Y		Not yet
13.5 Ageing calibration (GBYP)	2014		E, M	Y	Antonio Di Natale	Y	N	Y	Y		Not yet
14. Maturity (*Spawning biomass*)											
14.1. Western (NOAA)	1975-1981	ended	GOM	Y	Guillermo Diaz (NOAA)		N	N	N	N	Not yet
14.2 Mediterranean		rew	M	Y	GBYP	Y	N	Y	Y		Not yet
15. Other ecological data (*spatial distribution, covariates for CPUE standardization, steepness, natural mortality rate, spawning locations etc.*)											
15.1. Larval ecology (IEO)		ended	WM	Y	Diego Alvarez Berastegui		N	N	N	N	Not yet
15.2. Habitat model				Y	Jean-Noel Druon		N	N	N	N	Not yet

5 Possible hypotheses/ data scenarios for ABFT

In this document we discuss various ways in which data can be interpreted by operating models. Along with traditional axes of uncertainty (Fromentin et al. 2014), these may be considered for alternative operating model hypotheses (Table 4).

Table 4. Possible hypotheses for alternative operating models.

Regarding:	Hypothesis 1	Hypothesis 2	Hypothesis 3
Population			
Mediterranean sub stock structure	1 stock mixed	2 sub populations east/west	3 sub populations east/central/west
Recruitment	Proportional to mature stock biomass in spawning area	Proportional to population-wide spawning biomass	
Recruitment	High recruitment scenario	Low recruitment scenario	
Recruitment	High resilience	Low resilience	
Recruitment	High variability	Low variability	
Mixing	MLE	Low	High
Natural mortality rate	High	Low	
Spatial definitions	E.g. Figure 1b	Figure 1c	
Growth	Renewable growth type	Growth type groups	
Data			
Larval indices	OM fitted	OM not fitted	
PSAT tags of unknown stock of origin	OM fitted	OM not fitted	
Assignment of PSATs of unknown stock of origin	Prior to fitting based on stock of origin data	Inside model according to stock distribution	Inside model according to estimated movement
Conventional tagging	OM not fitted	OM fitted to observer recaptures	OM fitted to all tags
Movement	Age-independent	Juvenile / mature	Continuous with age
Historical catch reporting	Unbiased	Negatively biased	More negatively biased
Future catch reporting	Unbiased	Negatively biased	More negatively biased
Catch-at-length weighting	Consistent with *apriori* estimates of effective sample size	More strongly weighted	Less strongly weighted
Reliability of relative abundance indices	Proportional	Hyper stable (where applicable)	
Fishery aggregation	By gear type	By flag / gear type	

6 Simulation evaluation of operating model

6.1 Design

The demonstration MSE framework previously presented to the GBYP Core Modelling Group (December 2014, Anon. 2014) was used to simulate data to determine whether the model could estimate quantities of interest such as stock depletion, current stock size and spatial distribution reliably. In this preliminary simulation evaluation, 200 datasets were simulated with varying stock depletion, exploitation history, gear selectivity, movement and spatial distribution. The simulation was kept relatively simple and included only two fleet types, 4 areas, 2 sub-years and 40 historical years of exploitation (Figure 3 illustrates the simplified mixing and spatial structure that borrows four areas from the spatial definitions of Anon. 2015, Figure 1). A summary of inputs and parameter ranges for the simulations is included in Table 3.

Table 5. Simulation model specification

Parameter/variable	Symbol	Description	Value (range of simulated values)
Number of stocks	n_s		2
Number of fleets	n_f		2
Number of areas	n_r		4
Number of years	n_y	Historical years of exploitation	40
Number of subyears	n_m		2
Spawning subyear	*ms*		2
Unfished recruitment	*R0*		Stock 1: 225-450 Stock 2: 37 - 75
Natural mortality rate at age	*Minit*	User-specified natural mortality rate at age schedule	Stock 1: 0.49, 0.24, 0.24, 0.24, 0.24, 0.2, 0.175, 0.15, 0.125, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1. Stock 2: 0.1 (all ages)
Natural mortality rate factor	*Mfac*	Multiplier of the Natural mortality rate at age schedule	0.75 – 1.5
Steepness	*h*	Recruitment compensation	0.35 – 0.65
Inter-annual recruitment variation	σ_R	Log-normal standard deviation of recruitment deviations	0.1 - 0.3
Recruitment autocorrelation	Ψ_R		0.5 – 0.9
von Bertalanffy maximum growth rate parameter	κ		Stock 1: 0.087- 0.091 Stock 2: 0.091 - 0.095
Age at maturity	γ	Age when 50% of individuals are mature	Stock 1: 3.5 - 4.5 Stock 2: 8.5 - 9.5
Stock depletion	$SSB_{ny}/SSB0$	Spawning stock biomass relative to unfished	Stock 1: 0.035 - 0.2625 Stock 2: 0.3 - 0.4
Age at 100 selectivity	s_{mode}		5 – 8 (all fleets, all stocks)
Age at 5% selectivity		Ascending limb of sel. curve	2 – 3 (all fleets, all stocks
Sel. of oldest age			Fleet 1: 1 (all stocks) Fleet 2: 0.5 – 1 (all stocks)
Slope in recent exploitation rates			(-5) – 5 % per year (all fleets, all stocks)

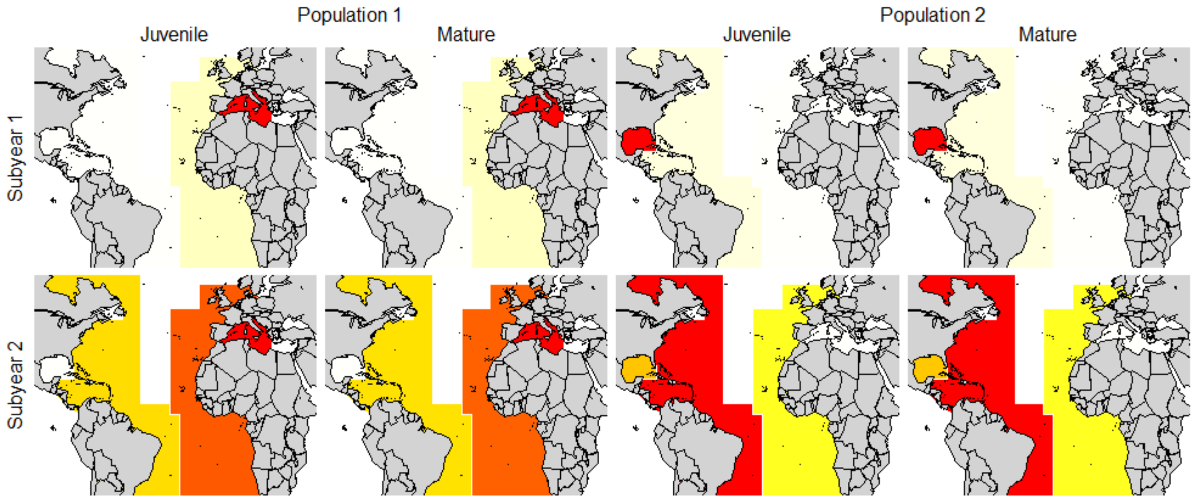


Figure 3. Spatio-temporal distribution of simulated population dynamics. In order to conduct a simplified test of estimation performance, only two subyears and four areas were simulated of the eight identified at the 2015 bluefin data preparatory meeting (ICCAT 2015, Figure 1).

This simulation evaluation was intended as a proof of model concept and consequently did not simulate biases in observed data (e.g. catch reporting, non-independence in length composition data) or evaluate model misspecification (e.g. incorrect aggregation of fleets, misspecification of selectivity, incorrect natural mortality rate at age, functional form of the stock recruitment relationship). The simulation results presented here are for a former version of the M3 (v1.02) model which does not include priors for steepness and natural mortality rate. In this simulation test these were assumed to be known perfectly without error.

6.2 Results

Simulation testing reveals that the model provides estimates of stock depletion, stock size and spatial distribution that are not strongly biased (Figure 4). For example, biases in estimates of stock depletion were on average within 2% of unbiased for both simulated stocks. The range of biases was also relatively low with standard deviations among simulations of 5.3% and 8.3% for stocks 1 and 2, respectively. Estimates of current stock size were somewhat negatively biased (around -5% for both stocks) but not strongly so. Among stocks, biases in estimates of current stock size were negatively correlated. This is to be expected since an overestimate of the size of stock 1 is likely to be paired with an underestimate of stock size 2, as the model aims to generate a similar total vulnerable biomass to that simulated. Among the simulations the biases in estimates of current stock size could be larger for stock 2 (a standard deviation of 10%) than stock 1 (a standard deviation of 6%) which should be anticipated since stock 2 is estmated to be smaller than stock 1 (unfished recruitment was 1/6 that of stock 1).

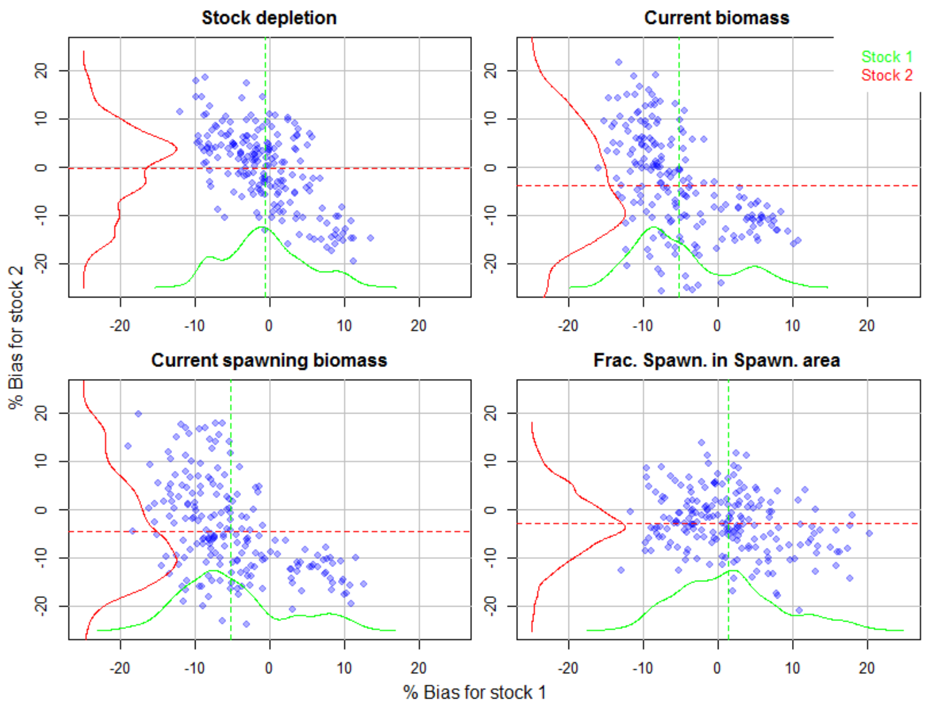


Figure 4. Bias in estimates of simulation model quantities ((estimated value – simulated value)/simulated value). Current refers to the final year of the simulation (ie most recent). Frac. Spawn. in Spawn. area is the fraction of the spawning stock biomass predicted in the spawning area for the final year of the simulation and is included here to examine the ability of the model to estimate spatial distribution reliably. The vertical and horizontal dashed lines represent the mean bias for stocks 1 and 2 respectively.

7 Discussion

7.1 Limitations

The first version of the M3 model was intended to initiate a dialogue with the GBYP core modelling group and wider SCRS regarding priorities for operating model development. In order to get a version of the model working and simulation tested, a number of features were omitted. One of the most important was the conventional tagging data which are not used to estimate exploitation rates in this preliminary M3 model.

It is relatively simple (although very computationally intensive) to include the model code to predict the dynamics of a tagged population of bluefin tuna and predict capture probabilities. However there are concerns that uncertain and variable reporting rates serve to contaminate these data which therefore could provide a misleading picture of movement and exploitation rate (and hence stock size). The confounding of reporting rate and fishing mortality rate estimates is a known problem in the use of conventional tagging data when no information exists to inform reporting rate estimates, as is the case for bluefin tuna. Previous spatial, multi-stock models of bluefin tuna such as MAST (Taylor et al. 2010) assumed similar reporting rates among fleets. There is however evidence that reporting rates may vary widely among fleets, for example between 1/1000 and 1/2 (Carruthers and McAllister 2010). Thus, the model may be confronted with observed recapture rates and observed catch rates that differ by a factor of 500 for some fleets. The likely result is a poorly defined objective function, model outputs that are sensitive to initial values and a failure to satisfy convergence criteria.

While the benefits of adding conventional tagging data are uncertain, the additional computational overhead may be greater than 500% for a model such as M3. It may be possible to include conventional tagging data and estimate the fleet-specific reporting rates but the benefit would be weak additional information regarding movements and exploitation rates at the cost of a much more computationally intensive model. Despite these potential drawbacks and the omission of these data in this first version of the model, conventional tags should still be considered as a source of information about exploitation rate and this should be a primary subject of discussion. There are a number of other ways in which conventional tagging data can be incorporated into the model such as characterising growth, informing estimates of fleet selectivity and vulnerability-at-age, and defining the full range of possible movements. Recaptures of conventional tags from observer-based programs provide one potential source of unbiased estimates as 100% reporting can be assumed. For example, the pelagic longline observer programs in the West Atlantic may provide unbiased estimates of recapture rates.

Currently the model does not attempt to model movement by age or length. Other approaches have aimed to estimate a separate set of movement parameters for juvenile and mature fish (e.g. MAST). This is a priority for model development. An alternative to estimating a separate movement model for certain ages or length classes would be to model continuous models of regional gravity or mixing rate with age (e.g. Carruthers et al. 2015). This has the benefit of adding the same number of parameters but having smooth transitions in movement among age classes rather than an abrupt shift to an alternative movement model at a particular life-stage.

Similar to other statistical catch-at-age and catch-at-length models which approximate historical fishing dynamics according to a finite number of fleets, the M3 model requires at least one selectivity curve to be either user-specified or assigned a 'flat-topped' (e.g. logistic) selectivity curve. The aggregation of fleets should also be considered carefully to avoid overly complex models with redundant fleets or those that have not contributed substantially to the exploitation of one or more of the stocks. Alternative selectivity assumptions, aggregations of fleets or temporal definitions of fleets may serve as alternative hypotheses to be taken into account in the operating models of a future MSE. Conventional tagging data may provide information on fleet-specific selectivity for some fishing fleets that operate in areas where a large number of tags have been released, e.g., the bait boat fisheries in the Bay of Biscay or the rod-and-reel fisheries off of North America.

7.2 Strengths and opportunities

Since all of the catch composition data for bluefin tuna are in the form of length samples, statistical catch-at-age models require these sampled lengths to be converted to ages. The problems associated with this practice are well established (Allioud et al. 2014, Hilborn and Walters 1992, Kell and Kell 2011). A central strength of the model proposed here is that it is fitted to these length samples directly. A statistical catch-at-length approach also provides avenues for accounting for varying growth rates among individuals which may be informed by conventional tagging data.

Once a range of plausible hypotheses have been identified for bluefin tuna and suitable operating models have been developed to represent these, there are potentially a large number of research questions that may also be addressed in addition to MSE. For example: to what extent is movement estimation biased by release location of PSAT tags? How should data be weighted in a spatial, integrated assessment model? What is a suitable experimental design of a genetic tagging program? When assessing populations with complex stock structure and highly migratory dynamics, what assessment model complexity is particularly important: space, age, both or neither? If explicit performance metrics are available, can suitable harvest control rules be derived?

An advantage of not integrating conventional tagging into the model is that it runs relatively quickly (e.g., in less than a few minutes). This may allow a simplified version of the approach to be included within a management procedure in future MSE analyses, broadening the range of complexity in management procedures. Rapid model fitting also allows for extensive simulation testing which confirms that the model has been programmed correctly and reveals potential areas for re-parameterization or simplification.

A number of model features are in development to allow for alternative hypotheses for bluefin tuna dynamics. Once such extension is an approximation to a Growth-Type Group model (GTG) via calculation of the inverse age-length key. A common oversight in fisheries stock assessment models is the inability to account for downward shifts in the expected length composition at age which occur as larger, faster growing individuals experience higher exploitation rates. Preliminary simulation evaluations using models that simulate many (300+) growth type groups, indicates that fishing mortality rate estimates can be highly inaccurate (as much as 200% biased) when this phenomenon is not taken into account. The conventional approach (also considered for the MAST model, Taylor et al. 2011) is to model discrete groups of individual of varying growth parameters. In Stock Synthesis these are referred to as 'platoons' (Methot and Wetzel 2013).

The principal problem with these approaches is that it can take many extra growth type groups (300+) or several platoons to generate suitably smooth predicted length compositions. This is a major problem for models such as M3 that are already computationally intensive as each GTG or platoon is an additional dimension, and hence the number of calculations is increased by a factor equal to the number of GTGs or platoons. A solution under investigation here is an approximation that takes percentiles of the distribution of GTGs and uses linear interpolation to predict the shift in the length structure given historical fishing mortality rates. This is a much more tractable approach as it adds no additional dimensionality to the transition calculations: additional calculation is limited to the construction of the inverse age-at-length key. Furthermore, conventional tagging data may be used as an empirical source of growth data for fitting the inverse age-length key. This additional feature would add around 10% more computation time per iteration.

The operating model has been subject to simulation testing (SCRS/2015/179) to evaluate estimation performance. These simulations can be broadened to establish which data are most critical in determining model predictions of stock status and productivity. This type of value of information analysis may provide a more rigorous basis for prioritising data gathering to support operating model development.

An added benefit of developing an M3 model as part of this contract is that it may be presented alongside spatially aggregated, stock specific assessments and hence offer an alternative viewpoint of stock status and projections. A possible objective moving forward would be to present this alternative assessment results in 2016 alongside the conventional assessment approaches (assuming that the required data are provided, specifically stock of origin and standardized relative abundance indices at the resolution of subyear and area).

7.3 Future priorities

The M3 model differs from previous multi-stock model such as MAST in that it requires indices of abundance (e.g. standardized CPUE indices) for fleets by time-area strata (e.g. for a given ocean area and subyear). The core advantage of this is that the movement estimation is constrained to combinations of parameters that are consistent with other spatial data. Previous spatial modelling has demonstrated that these data alone are sufficient to estimate spatial heterogeneity reliably. This is often more important than estimating specific movement transitions. For example to estimate MSY-related reference points reliably, it is more important to know the spatial distribution of the stock at a given time of year than to know exactly what fraction of individuals moved to and from the various areas (Carruthers et al. 2011b). This means that the development of fleet-specific indices at the resolution of subyear and spatial area is a central priority for operating model development.

Arguably more important still is the provision of stock of origin data at the resolution of subyear and spatial areas. While non standardized indices may be developed from catch rate data reported to ICCAT (Task II) there is no substitute for stock of origin data and without these data, spatial multi-stock operating models cannot be developed.

In this paper, simulation testing was cursory and should be much more thorough for future releases of the model. For example future tests should examine various types of model misspecification such as ignoring differences in juvenile/mature movement and misspecification of maturity. Problematic observation processes may also be

simulated: for example persistent biases in annual catches, spatially biased length sampling and relative abundance indices that are non-linearly related to abundance. Model diagnostics in future simulation tests could also include other checks such as of the biases in MSY-related reference points.
In this paper we reference the spatial definitions of the 2015 data preparatory meeting (Anon. 2015, Figure 1). Kimoto et al. (2015b) recommend that the ICCAT Bluefin Working Group should carefully examine spatial stratifications particularly those of the northeast Atlantic, a region that is currently a main fishing area for Japanese longline vessels. Historical abundance indices for the Japanese longline fleet are perhaps the most important data for the stock assessment of Atlantic bluefin tuna and are likely to be pivotal in fitting operating models. Figure 1 originates from the bluefin mixing workshop in 2001 (Anon. 2002) which apart from the separation of the Gulf of Mexico and the Mediterranean, was not fully agreed upon by the group. Furthermore the reasoning behind the boundaries identified are not fully supported by knowledge accumulated over the last 15 years. It was recommended to revise the area stratification with new and updated information collected, in the light of both biological and fisheries aspects.

Version 1.03 of the M3 model includes user-specified priors for steepness (recruitment compensation) and the natural mortality rate factor *Mfac* (a multiplier of the user-specified natural mortality rate at age schedule). It is important to account for uncertainty in these parameters in operating models since these are among the least well known and most influential in the estimation of Atlantic bluefin tuna status and productivity. It follows that moving to a Bayesian version of the model using the Metropolis Hastings algorithm of ADMB is a future priority.

7.4 Strategy for operating model development and data priorities

In settings where MSE has been successful it has been an iterative process (Punt et al. 2014). It is generally recognised that stakeholders require prior exposure to MSE concepts in order to identify operating models and performance metrics. In recognition of this, it is a priority to develop a working MSE framework based on operating models fitted to data in order to advance the GBYP MSE process. A fitted operating model also allows data providers to see the benefit of their contributions and allows further refinement of the broader MSE framework.

A possible strategy for operating model development is to fit the model to ICCAT data that are freely available and admit more rigorous data to the model as they become available (e.g. standardized indices, PSAT tagging data). Catch data, relative abundance indices and length composition data are currently available as are the parameters to produce preliminary inverse age-length keys and maturity schedules. It follows that the most significant data gaps for developing a preliminary operating model are data regarding stock of origin (e.g. otolith microchemistry, SNP) and movement (e.g. PSAT data). If simplified gravity movement models are assumed tagging data are not necessary to estimate seasonal spatial distribution. As more detailed PSAT data become available alternative movement models can be considered such as age-based movement.

8 Progress relative to deliverables

8.0 The deliverables include documented, object-oriented SC4 Classes and C++ source code for the Operating Model (OM), including the Observation Error Model, (OEM) that can be used by third parties to develop and evaluate their own Management Procedures (MPs) consistent with the recommendations of the Modelling Coordinator, ICCAT population dynamics specialist and Modelling Steering Group. All code shall be available at the https://github.com/iccat-mse and https://github.com/generic-mse github repositories.

[90%] Developing the M3 model in ADMB has the relative merit that it is trivial to create more generic C++ libraries for use in other modelling frameworks such as template model builder (Kristensen 2015). Once the M3 model has been thoroughly simulation tested and fitted to real bluefin tuna data, the model will be uploaded onto the relevant websites. In terms of the coding this deliverable is essentially complete.

8.1 Based on the review of population hypotheses and stock structure, provide OM classes that can be used to conduct sensitivity analyses and then to implement hypotheses in the OM in order to evaluate alternative OEMs and MPs. *Specific outcome:* Provide examples for the review paper on population hypotheses and stock assumptions (see SCRS, 2013).

[20%] As identified here a range of alternative operating models may be proposed that interpret data and account for recruitment, growth and stock mixing in various ways. This deliverable is contingent on the provision of suitable data to inform the operating model and group decisions about which hypotheses to focus on (e.g. at a proposed working group meeting in Monterey January 2016). It follows that this will likely be one of the last deliverables to be completed prior to the February deadline. <u>To achieve this deliverable, stock of origin data at the resolution of subyear and area must be made available.</u>

8.2 Develop the Observation Error Model (OEM) that can be used to evaluate different data collection regimes e.g. aerial surveys, tagging programmes, catch and catch per unit effort (CPUE) and size to age conversions. *Specific outcome:* Use the OEM to conduct an analysis to show how improving data and knowledge can be used to reduce uncertainty and write up as an SCRS paper or peer review manuscript.

[100%] This deliverable has already been completed and was a requirement of the simulation testing described in this report.

8.3 Use the OM to evaluate alternative MPs developed and proposed by third parties. This requires to have a least one example of an MSE that evaluates a MP proposed by members of SCRS (e.g. Cooke, 2012). *Specific outcome:* Participate in manuscript, i.e. an SCRS or peer review paper, that documents the example(s).

[50%] Once the M3 has been fitted to data a range of possible MPs may be performance tested including current assessment methods and the MPs investigated by Carruthers et al. 2015d. In any case a range of MPs are already coded and compatible with the MSE framework.

8.4 Develop interactive tools in collaboration with other RFMOs for use with stakeholders based on Shiny (e.g. http://shiny.iphc.int/sample-apps/shiny/). *Specific outcome:* Published the interactive tool at http://rscloud.iccat.int:3838/bft-mse/

[0%] The development of shiny tools first requires an operating model fitted to data to represent various hypotheses and a closed-loop test of a range of MPs. It follows that this is likely to the last deliverable to be completed.

Acknowledgments

Many thanks to my coauthors of the two SCRS papers (Carruthers et al. 2015a;b) on which this report is based: Ai Kimoto, Matt Lauretta, Toshihide Kitakado, Doug Butterworth, Joe Powers, Laurie Kell, and Antonio Di Natale. This work was carried out under the provision of the ICCAT Atlantic-Wide Research Programme for Bluefin Tuna (GBYP), funded by the European Union, by several ICCAT CPCs, the ICCAT Secretariat and by other entities (see:http://www.iccat.int/GBYP/en/Budget.htm). The content of this paper does not necessarily reflect the point of view of ICCAT or of the other funders, which have no responsibility for it, neither does it necessarily reflect the views of the funders and in no way anticipates the Commission's future policy in this area.

9 References

Anon. 2015a. Report of the 2015 ICCAT bluefin data preparatory meeting (Madrid, Spain - March 2-6, 2015).

Abid, N., Faraj, A., De la Serna, J.M., Macías, D., Saber, S., Ortiz de Urbina J. 2015. Updated standardized joint CPUE index for bluefin tuna (*Thunnus thynnus*) caught by Moroccan and Spanish traps for the period 1981-2013. ICCAT, Col. Vol. Sci. Pap., 71(3): 1308-1312.

Ahrens, R., Walters, C. 2005. Why are there still large pelagic predators in the oceans? Evidence of severe hyper-depletion in longline catch-per-effort. WCPFC-SC1 ME-WP-3. Available at: https://www.wcpfc.int/system/files/SC1_ME_WP_3.pdf [accessed September 2015].

Ailloud, L.E., Smith, M.W., Then, A.Y, Omori, K.L., Ralph, G.M., Hoenig, J.M. 2014. Properties of age compositions and mortality estimates derived from cohort slicing of length data. ICES. J. Mar. Sci. 72(1): 1-6.

Aranda, G., Medina, A., Santos, A., Abascal, F.J., Galaz, T. 2013. Evaluation of Atlantic bluefin tuna reproductive potential from ovarian histology and its use in stock assessments. Journal of Sea Research. 76: 154-160.

Arrizabalaga, H., Fromentin, J.M., Kell, L.T., Kerr, L.A., Kimoto, A., Porch, C., Secor, D.H. 2014. Population hypotheses and stock assumptions for North Atlantic Bluefin tuna. SCRS/2014/xxx.

Block, B., A., Teo, S.L.H., Walli, A., Boustany, A., Stokesbury, M.J.W., Farwell, C.J., Weng, K.C., Heidi, D., Williams, T.D. 2005. Electronic tagging and population structure of Atlantic bluefin tuna. Nature. 434: 1121-1127.

Bonhommeau, S., Farrugio, H., Poisson, F, Fromentin, J-M. 2010. Aerial surveys of bluefin tuna in the western Mediterranean sea: retrospective, prospective, perspective. ICCAT, Col. Vol. Sci. Pap., 65(3): 801-811

Boustany, A. M., Reeb C. A., Block B. A. 2008. Mitochondrial DNA and electronic tracking reveal population structure of Atlantic bluefin tuna (*Thunnus thynnus*). Marine Biology 156:13–24.

Bravington, M. V., Grewe, P.G., Davies, C.R. 2013. Fishery-independent estimate of spawning biomass of Southern Bluefin Tuna through identification of close-kin using genetic markers. FRDC Report 2007/034. CSIRO, Australia.

Bull, B., Francis, R.I.C.C, Dunn, A., McKensie, A., Gilbert, D.J., Smith, M.H., Bain, R., Fu, D. 2012. CASAL (C++ algorithmic stock assessment library): CASAL user manual. V2.30-2012/03/21. NIWA Technical Report. 135. 280pp.

Butterworth, D.S., Punt, A.E., 1999. Experiences in the evaluation and implementation of management procedures. ICES J. Mar. Sci. 56, 985-998.

Candy, S.G. 2008. Estimation of effective sample size for catch-at-age and catch-at-length data using simulated data from the Dirichlet-multinomial distribution. CCAMLR Science. 15: 115-138.

Carruthers, T.R., McAllister, M. 2010. Estimating reporting rates using coincidental tag recaptures. Journal of Aquatic Living Resources. 23: 343-352.

Carruthers, T.R., Ahrens, R.N.M., McAllister, M.K., Walters, C. 2011a. Integrating imputation and standardization of catch rate data in the calculation of relative abundance indices. Fish. Res. 109(1): 157-167.

Carruthers, T.R., McAllister, M.K. Taylor, N.G. 2011b. Spatial surplus production modelling of Atlantic Tunas and Billfish. Ecological Applications. 21: 2734-2755.

Carruthers, T.R., Punt, A.E., Walters, C.J., MacCall, A., McAllister, M.K., Dick, E.J., Cope, J. 2014a. Evaluating methods for setting catch limits in data-limited fisheries. Fish. Res. 153: 48-68.

Carruthers, T.R., Kimoto, A., Powers, J.E., Kell, L., Butterworth, D.S., Lauretta, M.V., Kitakado, T. 2015a. Structure and estimation framework for Atlantic bluefin tuna operating models. SCRS/2015/179.

Carruthers, T.R., Powers, J.E., Lauretta, M.V., Di Natale, A., Kell, L. 2015b. A summary of data to inform operating models in management strategy evaluation of Atlantic bluefin tuna. SCRS/2015/180.

Carruthers, T.R., Walters, J.F., McAllister, M.K., Bryan, M.D. 2015c. Modelling age-dependent movement: an application to red and gag groupers in the Gulf of Mexico. 72(8): 1159-1176. doi: 10.1139/cjfas-2014-0471

Carruthers, T.R., Kell, L., Maunder, M., Geromont, H., Walters, C., McAllister, M.K., Hillary, R., Kitakado, T., Davies, C. 2015d. Performance Review of Simple Management Procedures. [Drafted as part of this contract]

CCSBT. 2011. Report of the Sixteenth Meeting of the Scientific Committee. Bali, Indonesia. Commission for the Conservation of Southern Bluefin Tuna. Available at: http://www.ccsbt.org/userfiles/file/docs_english/meetings/meeting_reports/ccsbt_18/report_of_SC16.pdf

Cermeno, P., Quilez-Badia, G., Ospina-Alvarez, A., Sainz-Trapaga, S., Boustany, A., Seitz, A.C., Tudela, S., Block, B.A. 2015. Electronic tagging of Atlantic bluefin tuna (*Thunnus thynnus,* L.) reveals habitat use and behaviors in the Mediterranean Sea. PLoS ONE 10 (2): e0116638. doi: 0.1371/journal.pone.0116638

Cochrane, K L., Butterworth, D.S., De Oliveira, J.A.A., Roel, B.A., 1998. Management procedures in a fishery based on highly variable stocks and with conflicting objectives: experiences in the South African pelagic fishery. Rev. Fish. Biol. Fisher. 8, 177-214.

Corriero, A., Karakulak, S., Santamaria, N., Deflorio, M., Spedicato, D., Addis, P., Desantis, S., Cirillo, F., FenechFarrugia, A., Vassallo-Agius, R., de la Serna, J. M., Oray, Y., Cau, A., Megalofonou, P., Metrio, G., 2005. Size and age at sexual maturity of female bluefin tuna (*Thunnus thynnus* L. 1758) from the Mediterranean Sea. Journal of Applied Ichthyology 21, 483–486.

Diaz, G. 2010. A revision of western Atlantic bluefin tuna age of maturity derived from size samples collected by the Japanese longline fleet in the Gulf of Mexico (1975-1980). ICCAT, Collect. Vol. Sci. Pap. 66 (3): 1216-1226.

Fournier, D.A., Hampton, J., Sibert, J.R. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. Can. J. Fish. Aqua. Sci. 55: 2105-2116.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M., Nielsen, A. and Sibert, J. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Meth. Soft. 27 (2): 233-249.

Francis, R,I.C,C. 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aqua. Sci. 68(6): 1124-1138.

Fromentin, J-M., Lopuszanski, D. 2014. Migration, residency and homing of Bluefin tuna in the western Mediterranean Sea. ICES J. Mar. Sci. 71(3): 510-518.

Fromentin, J-M., Bonhommeau, S., Arrizabalaga, H., Kell, L.T. 2014. The spectre of uncertainty in management of exploited fish stocks: the illustrative case of Atlantic bluefin tuna. Mar. Pol. 47:8-14.

GBYP. 2014. ICCAT Atlantic wide research programme for Bluefin Tuna. Available online at: http://www.iccat.int/GBYP/en/index.htm [accessed October 2014]

Geromont, H.F., Butterworth, D.S. 2014a. Generic management procedures for data-poor fisheries; forecasting with few data. ICES J. Mar. Sci. doi:10.1093/icesjms/fst232.

Geromont, H.F., Butterworth, D.S. 2014b. Complex assessments or simple management procedures for efficient fisheries management: a comparative study. ICES J. Mar. Sci. doi:10.1093/icesjms/fsu017.

Hanke, A.R., Andrushchenko, I., Whelan, C. 2015. Indices of stock status from the Canadian bluefin tuna fishery:1981 to 2013. ICCAT, Col. Vol. Sci. Pap., 71(2): 983-1017.

Hilborn, R. 2003. The state of the art in stock assessment: where we are and where we are going. Scientia Marina 67 (supplement 1): 15-20.

Hilborn, R. & Walters, C.J. 1992. Quantitative fisheries stock assessment: Choice, dynamics and uncertainty. Chapman and Hall. 570 pp.

Hillary, R.M., Million, J., Anganuzzi, A., Areso, J.J. 2008. Tag shedding and reporting rate estimates for Indian Ocean tuna using double-tagging and tag-seeding experiments. Working paper of the Indian Ocean Tuna Commission. IOTC-2008-WPTDA-04. Available at: http://www.iotc.org/files/proceedings/2008/wptda/IOTC-2008-WPTDA-04.pdf [accessed September 2015]

ICCAT 2014a. Report of the 1st Meeting of ICCAT GBYP Core Modelling Group. Available at: http://www.iccat.int/GBYP/Documents/MODELLING/PHASE%204/BFTMSE_CMG1_Report.pdf [accessed September 2015]

ICCAT 2014b. Executive summary 2014, status of Eastern stock of Atlantic bluefin tuna. Available at: https://www.iccat.int/Documents/SCRS/ExecSum/BFT_EN.pdf [accessed September 2015]

ICCAT 2015a. Report of the 2015 ICCAT bluefin data preparatory meeting. Available at: https://www.iccat.int/Documents/Meetings/Docs/BFT_DATA_PREP_2015_eng.pdf [accessed September 2015).

ICCAT 2015b. Information published on TaskII – Catch and Effort (T2CE). Available at: https://www.iccat.int/Data/t2ce-ENG.pdf [accessed September 2015]

ICCAT 2015c. CATDIS readme file. Available at: http://www.iccat.int/Data/Catdis/Readme_catdis.docx [accessed September 2015]

Ingram, G.W., Jr., D. Alvarez-Berastegui, P. Reglero, R. Balbín, A. García5, and F. Alemany. Indices of larval bluefin tuna (Thunnus thynnus) in the Western Mediterranean Sea (2001-2013). SCRS/2015/035.Kell, L.T., Kell, A.A. 2011. A comparison of age slicing and statistical age estimation for Mediterranean swordfish (*Xiphias gladius).* ICCAT, Col. Vol. Sci. Pap., 66(4): 1522-1534.

Kell, L.T., Kell, A. 2011. A comparison of age slicing and statistical age estimation for Mediterranean swordfish (*Xiphias gladius*). Col. Vol. Sci. Pap. 66(4): 1522-1534.

Kell, L.T. T., Fromentin, J.-M., Bonhommeau, S. 2011. An Evaluation of the Implications of Population Structure on the Current Bluefin Tuna Advice Framework. SCRS/2011/110.

Kell, L.T. T., Fromentin, J.-M., Bonhommeau, S. 2012. An Evaluation of the Implications of Population Structure on the Current Bluefin Tuna Advice Framework. ICCAT, Col. Vol. Sci. Pap., 68(1): 370-378

Kell, L.T., Hillary, R., Fromentin, J.-M., Bonhommeau, S. 2014. An example management strategy evaluation of a model free harvest control rule. ICCAT, Col. Vol. Sci. Pap., 71(6): 2790-2797.

Kimoto, A., Takeuchi, Y., Itoh, T. 2015a. Updated standardized bluefin CPUE from the Japanese longline fishery in the Atlantic to 2014 fishing year. ICCAT, Col. Vol. Sci. Pap., 71(2): 1059-1097.

Kimoto, A., Takeuchi, T., Itoh, T. 2015b. Discussion on the area stratification in the North Atlantic for bluefin tuna mixing model. SCRS/2015/172.

Knapp, J.M, Aranda, G., Medina, A., Lutcavage, M. 2014. Comparative assessment of the reproductive status of female Atlantic bluefin tuna from the Gulf of Mexico and the Mediterranean Sea. PLOS One. doi: 10.1371/journal.pone.0098233.

Kristensen, K. 2015. Template Model Builder. Available at: http://www.admb-project.org/developers/tmb/tmb_r.pdf/at_download/file [accessed September 2015]

Lauretta, M.V., Brown, C.A. 2015. Standardized catch rates of bluefin tuna (*Thunnus thynnus*) from the rod and reel/handline fishery off the northeast United States during 1993-2013. ICCAT, Col. Vol. Sci. Pap., 71(3): 1223-1237.

Leach, A.W., Levontin, P., Holt, J., Kell, L.T., Mumford, J.D. 2014. Identification and prioritization of uncertainties for management of Eastern Atlantic bluefin tuna (*Thunnus thynnus*).

LPRC. 2015. Large pelagics research centre. Available at: http://www.tunalab.org/ [accessed September 2015]

Lutcavage, M.E., Galuardi, B., Lam, T.C.H. 2012. Predicting potential Atlantic spawning grounds of western Atlantic bluefin tuna based on electronic tagging results, 2002-2011. Col. Vol. Sci. Pap. ICCAT. 69(2): 955-961.

Martell, S. 2015. The iSCAM project. Available at: https://code.google.com/p/iscam-project/ [accessed 9/9/2015]

Maunder M.N., Punt, A.E. 2013. A review of integrated analysis in fisheries stock assessment. Fisheries Research. 142: 61-74.

Medina, A., Abascal, F.J., Aragon, L, Mourente G, Aranda, G., et al. 2007. Influence of sampling gear in assessment of reproductive parameters for bluefin tuna in the western Mediterranean. Mar. Ecol. Prog. Ser. 337: 221-230.

Methot, R.D. and Wetzel, C.R. 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fish. Res. 142: 86-99.

Punt, A.E., Butterworth, D.S., de Moore, C. L., De Oliveira, J. A. A., Haddon, M. 2014. Management strategy evaluation: best practices. Fish and Fisheries. doi: 10.1111/faf.12104

Quinn, T.J., Deriso, R.B. 1999. Quantitative fish dynamics. Oxford University Press, New York.

Rooker, J. R., Arrizabalaga, H., Fraile, I., Secor, D.H., Dettman, D.L., Abid, N., Addis, P., Deguara, S., Saadet Karakulak, F., Kimoto, A., Sakai, O., Macias, D., Santos, M.N. 2014. Crossing the line: migratory and homing behaviors of Atlantic bluefin tuna. Mar. Ecol. Prog. Ser. 504: 265-276.

Rooker JR, Secor DH, De Metrio G, Schloesser R, Block BA, et al. 2008. Natal homing and connectivity in Atlantic bluefin tuna populations. Science 322: 742–744.

Santiago, J., Arrizabalaga, H., Ortiz, M., Goñi, N. 2015. Updated standardized bluefin tuna CPUE index of the Bay of Biscay baitboat fishery (1952-2013). SCRS/2014/054.

Taylor, N.G., McAllister, M.K., Lawson, G. L. Carruthers, T.R., Block, B.A. 2011. Atlantic bluefin tuna: a novel multistock spatial model for assessment of population biomass. PLoS ONE. 6(12):1:10.

Walli A, Teo SLH, Boustany A, Farwell CJ, Williams T, Dewar, H., Prince, E., Block, B.A. 2009. Seasonal movements, aggregations and diving behavior of Atlantic bluefin tuna (Thunnus thynnus) revealed with archival tags. PloS ONE 4: e6151. doi: 10.1371/journal.pone.0006151

Walter J. 2015. Update of standardized catch rates of large bluefin tuna (*Thunnus thynnus*) from the U.S. pelagic longline fishery in the Gulf of Mexico 1987-2013 with correction for weak hook effects. ICCAT, Col. Vol. Sci. Pap., 71(3): 1260-1278.

Walters, C., Essington, T. 2010. Recovery of Bioenergetics Parameters from Information on Growth: Overview of an Approach Based on Statistical Analysis of Tagging and Size-At-Age Data. The Open Fish Science Journal, 2010, 3, 52-68