Models for exploring the information content of the RTTP-IO tagging data

R.M. Hillary[1]

Division of Biology,

Imperial College London, SW7 2AZ, UK.

October 2008

1 Introduction

This paper is intended to be methods a paper detailing a range of potential population and tagging models that can be applied to Indian Ocean tuna tagging and catch information outside of the integrated assessment framework. The models detailed in this paper are not explicitly spatial and are both age and length-based so as to be able to be applicable to all three key species. The models are custom-designed to try and explore an accommodate Indian Ocean tuna-specific issues but build on the many such external tagging models developed over the year.

Integrated assessments are an extremely useful tool not just for performing stock assessments but also for exploring the relative influence of a variety of data sets on our perceptions about the status of the stock. However, their generality can sometimes be their restriction - one cannot allow all permutations of all potential models so some issues will invariably be outside of the range of even the most flexible of assessment packages. Also, in such complex models parameters such as natural mortality are very difficult to estimate with confidence as they are confounded with so many other processes and parameters. In this regard more simplistic and focused models such as those proposed in this paper are quite useful as they relate to using tagging and catch information only and over a much more restricted age/length and time frame thereby permitting the potential estimation of key parameters such as natural mortality, as well as exploitation rates and

---

[1]E-mail: `r.hillary@imperial.ac.uk`

abundance.

2 The population models

We propose two conceptually very similar population and fishery models to be used - the first is age-based and the second is a length-based model, designed specifically to be used for skipjack. Each model is a yearly-quarterly non-spatial population model and the models for the tagged population are based upon the actual population model but with some modifications, designed to be able to mimic the dispersal of released tags into the various Indian Ocean gear types without having to design and explicitly spatial model.

2.1 The age-based population model

Here we model the dynamics of the numbers-at-age, $N_{y,q,a}$ in years $y = y_{\min}, ..., y_{\max}$, quarters $q = 1, ..., 4$ and ages $a = a_{\min}, ..., a_{\max}$. For quarters 2 to 4 and for $a > a_{\max}$:

$$N_{y,q,a} = N_{y,q-1.a} \exp\left(-M_{q-1,a}\right) \left(1 - \sum_{g=1}^{G} h_{y,g,q-1,a}\right) \tag{1}$$

where h is the harvest/exploitation rate by gear-type g and is calculated as follows:

$$h_{y,g,q,a} = \frac{C_{y,g,q,a}}{N_{y,q,a} \exp\left(-M_{q,a}/2\right)} \tag{2}$$

and $C_{y,g,q,a}$ is the gear-specific catch. For quarter one in all but the first year and for all ages greater than $a_{\min}$ we have that

$$N_{y,1,a} = N_{y-1,4,a-1} \exp\left(-M_{4,a-1}\right) \left(1 - \sum_{g=1}^{G} h_{y-1,g,4,a-1}\right) \tag{3}$$

In this framework, the initial numbers $N_{y,1,a=a_{\min}+1,...,a_{\max}}$ and the recruitment numbers-by-quarter $N_{y,q,a_{\min}}$ must be estimated in some way and this leads to the issue of how to model recruitment in this framework. The simplest option is that we have a single

recruitment event in each year, R_y, such that $N_{y,1,a_{\min}} = R_y$ and it follows for quarters 2 to 4 that

$$N_{y,q,a_{\min}} = N_{y,q-1,a_{\min}} \exp\left(-M_{q-1,a_{\min}}\right)\left(1 - \sum_{g=1}^{G} h_{y,g,q-1,a_{\min}}\right) \tag{4}$$

However, given tropical tuna biology this may not be an accurate representation of the dynamics. Another option is to apportion this recruitment value over the seasons using given recruitment probabilities $\pi^{\mathrm{rec}}_{y,q=1-3}$ and $\pi^{\mathrm{rec}}_{y,4} = 1 - \left(\pi^{\mathrm{rec}}_{y,1} + \pi^{\mathrm{rec}}_{y,2} + \pi^{\mathrm{rec}}_{y,3}\right)$ so that $N_{y,1,a_{\min}} = \pi^{\mathrm{rec}}_{y,1}R_y$ and that for quarters 2 to 4

$$N_{y,q,a_{\min}} = N_{y,q-1,a_{\min}} \exp\left(-M_{q-1,a_{\min}}\right)\left(1 - \sum_{g=1}^{G} h_{y,g,q-1,a_{\min}}\right) + \pi^{\mathrm{rec}}_{q,y}R_y \tag{5}$$

Now we have described the population model we need to define the model for the tagged population and this will be very similar though necessarily different in parts to the population model. The initial number of tags in a given release event is trivially the actual number of releases. Tagged fish are prone to both natural and fishing mortality as the untagged fish are but tagged animals also experience tag shedding as well as differential vulnerability to the gear types over time (a pseudo-spatial effect given the dispersal times of tagged fish across the Indian Ocean). To account for this we describe a compound tag survival probability, π^{TS}:

$$\pi^{TS}_{y,q,a} = \exp\left(-M_{q,a}\right)\pi^{R}\left(\tau\right)\left(1 - \sum_{g} h_{y,g,q,a}\kappa\left(g,\tau,a\right)\right). \tag{6}$$

The τ parameter denotes time-at-liberty; $\pi^{R}\left(\tau\right)$ is the probability of retaining a tag up to the given time-at-liberty, τ; and $\kappa\left(g,\tau,a\right)$ is a function ranging from zero to one that denotes the vulnerability of the tagged animals to a given gear type, at a given age and for a given time-at-liberty - this is the pseudo-spatial nature of the model whereby we allow for differential mixing times of the tagged animals into the different gear types given the spatially different locations of the fishing operations, relative to the point(s) of release of the tags. We can now define the dynamics of the tagged animals for quarters 2

to 4 (after release) as follows:

$$T_{y,q,a} = T_{y,q-1,a}\pi^{TS}_{y,q,a} \tag{7}$$

and for quarter one in all years after the year of release:

$$T_{y,1,a} = T_{y-1,4,a-1}\pi^{TS}_{y-1,4,a-1} \tag{8}$$

This fully defines the age-based population model so next we define the length-based population model.

2.2 The length-based population model

The basic structure of the length-based model is conceptually very similar to the age-based model - indeed the only real difference is that we have to model how the fish grow through the length-classes as time passes, instead of the simpler concept of age transitions. The main external information we need to do this is growth information - ideally a growth curve and the error around this curve. In general we need a growth increment function that gives us the predicted length of an animal, l', after time t and given it was originally of length l:

$$l' = \Gamma(l, t, \boldsymbol{\theta}), \tag{9}$$

where $\Gamma(\cdot, \cdot, \boldsymbol{\theta})$ is the growth increment function for a given parameter set $\boldsymbol{\theta}$. After fitting this growth model to data we could obtain a probability density function for the predicted length, given the initial length and the time increment:

$$p(l' \mid l, t, \boldsymbol{\theta}, \boldsymbol{\vartheta}), \tag{10}$$

where $\boldsymbol{\vartheta}$ are the process error parameters. Usually, we would integrate over the growth parameters to obtain a function predicting the distribution around a mean predicted length and conditional only on the process error parameters:

$$p_\theta(l' \mid l, t, \boldsymbol{\vartheta}) = \int p(l' \mid l, t, \boldsymbol{\theta}, \boldsymbol{\vartheta}) \, d\boldsymbol{\theta}. \tag{11}$$

Depending on the probability model for the data this density is usually normal, log-normal or gamma - the main point being that it is usually of a known form. For a given length interval $\ell = (l_{\min}, l_{\max})$ we can compute the probability of an animal of length l moving into length-class ℓ, after time t, as follows:

$$p\left(l \in \ell \,|\, l, t\right) = \int_{l_{\min}}^{l_{\max}} p_\theta\left(k \,|\, l, t, \boldsymbol{\vartheta}\right) dk. \tag{12}$$

Given a growth curve and the relevant process error parameters and distributions it is fairly simple to then compute a length-transition matrix, $\pi_{q,i,j}^T$, which is the probability of an animal in quarter $q-1$ growing from length-class i to length-class j in the following quarter.

To define the population dynamics we first need to define a length partition (similar to an age range):

$$\Lambda = \{l_1, ..., l_P\}. \tag{13}$$

For quarters 2 and 4 and for all length classes greater than l_1 we have the following dynamics:

$$N_{y,q,l} = \sum_{j \in \Lambda} N_{y,q-1,j} \exp\left(-M_{q-1,j}\right) \left(1 - \sum_g h_{y,g,q-1,j}\right) \pi_{q,j,l}^T, \tag{14}$$

and $h_{y,g,q,l}$ is calculated exactly as it was for the age-based model:

$$h_{y,g,q,l} = \frac{C_{y,g,q,l}}{N_{y,q,l} \exp\left(-M_{q,l}/2\right)}. \tag{15}$$

For quarter one for all years but the first we have that

$$N_{y,1,l} = \sum_{j \in \Lambda} N_{y-1,4,j} \exp\left(-M_{4,j}\right) \left(1 - \sum_g h_{y-1,g,4,j}\right) \pi_{4,j,l}^T. \tag{16}$$

In essence, in a length-based model all that happens is that to predict the numbers in a given length-class in a given quarter, all we do is take the proportion of animals from

all the other length class that will grow into this length class the quarter before and take away those that are lost to natural and fishing mortality and them sum them.

Modelling recruitment is a little trickier in a length-based model as we have more flexibility in terms of potential processes. Again, assuming that we have a fixed number of recruits to the population in any given year, R_y, then how do we sensibly introduce these recruits into the population? The first option would be to again use the $\pi_{q,y}^{\text{rec}}$ probabilities to apportion recruitment to the initial length-class - it might be that fish recruit to a selection of length-classes given the growth dynamics and variation but it is much easier to assume this more simplistic recruitment dynamic. For further simplicity we assume that $N_{y,q,l_1} = \pi_{q,y}^{\text{rec}} R_y$ and that $l_1 = l_2$ - this makes the recruitment length bin (l_1, l_2) almost a dummy bin in that recruiting fish enter it in a given quarter and all leave it in the following quarter as it is of zero size.

Modelling the tagged population follows exactly the same format as in the age-based model in that we define a tag survival probability:

$$\pi_{y,q,l}^{TS} = \exp\left(-M_{q,l}\right) \pi^R\left(\tau\right) \left(1 - \sum_g h_{y,g,q,l} \kappa\left(g, \tau, l\right)\right), \tag{17}$$

so that the dynamics of tagged animals in quarters 2 to 4 after release are given by

$$T_{y,q,l} = \sum_{j \in \Lambda} T_{y,q-1,j} \pi_{y,q,j}^{TS} \pi_{q,j,l}^{T}, \tag{18}$$

and for quarter 1 for all years after release:

$$T_{y,1,l} = \sum_{j \in \Lambda} T_{y-1,4,j} \pi_{y-1,4,j}^{TS} \pi_{4,j,l}^{T}. \tag{19}$$

This concludes the definition of the length-based population model so now we can define our probability (likelihood) model for the tag recapture process.

3   Tag recapture model

There are many potential ways to model the recapture of tagged animals in an exploited population. All of them involve being able to create a probability of recapture (usually a composite of exploitation rate, catch information, reporting rate etc.) and then assuming some statistical model that relates observed and predicted returns of tagged animals.

For the case of the Indian Ocean tuna fisheries and tagging information we have a specific issue to overcome: we know that tagged animals are taken by all of the gear types but we have reliable reporting rate information for a single unloading port - Victoria in the Seychelles - which itself forms the destination of catch from a (large) subset of the purse-seine gear fishery. This gives us a number of difficulties that need to be solved:

1. We need to be able to model the loss of tags to the other gear types for which we do not have reporting rate information

2. To do this we need to combine catch and abundance information to obtain gear-specific exploitation rates to model the removal of tags by these gear types

3. To predict the tagging recaptures for which we have reporting rate information we have to (a) define a Seychelles catch-specific exploitation rate (as a subset of the total PS exploitation rate) or (b) use the catch associated with the tag recapture events (henceforth denoted reference catch) to define a probability of recover for these tags

For this work we chose to employ reference catch (associated with the tag return data) to create a probability of recapture, π^r, as follows:

$$\pi^r_{y,q,i} = \frac{C^{\text{ref}}_{y,q,i}}{N_{y,q,i} \exp\left(-M_{q,i}/2\right)} \times \pi^{\text{det}}_{y,q,i}, \tag{20}$$

where i relates to either age or length and $\pi^{\text{det}}_{\cdot,\cdot,\cdot}$ is the probability of detecting a tag (derived from the reporting rate information). The model-predicted number of tag returns is now

simply defined to be

$$\widehat{R}_{y,q,i} = \pi^r_{y,q,i} T_{y,q,i}.$$ (21)

The final choice is that the probability model we assume to compare model-predicted to observed recaptures, $R_{y,q,i}$. For this work we chose to use a Poisson model (with the possible inclusion of over-dispersion but not as a negative binomial/Poisson-gamma model) and we shall give the reasons for this assumption later on.

3.1 Pure Poisson model

If we assumed a pure Poisson process then the probability/likelihood of each recapture event (for a given release event t) is given by the following equation:

$$p\left(R_{y,q,i,t_r} \mid \widehat{R}_{y,q,i,t_r}\right) = \frac{\left(\widehat{R}_{y,q,i,t_r}\right)^{R_{y,q,i,t_r}} \exp\left(-\widehat{R}_{y,q,i,t_r}\right)}{R_{y,q,i,t_r}!}.$$ (22)

In terms of the likelihood it is important to note that we treat each release and the subsequent recapture events in isolation. Tag numbers and recaptures are not pooled together and this is to try and preserve the information content - particularly on natural mortality - that the tagging data might possess.

3.2 Accounting for dispersion

We might assume the tag recapture process to be a Poisson process but it does have to be - in a Poisson model the mean and the variance are the same so a pure Poisson process has a **dispersion** value (σ^2/μ) of 1. However, some processes can vary less about the mean than a Poisson model would predict (under-dispersion) or more (over-dispersion). To be able to account for this type of process without having to move to a more complex statistical model (negative binomial or Poisson-gamma) we propose the following approach:

We assume *a priori* a Poisson recapture model - this implies that the variance and the mean of the recapture numbers are the same. One possible way to detect over-

dispersion is to look at the standardised residuals to detect if they differ from what we would expect from a Poisson process. In our case, and indeed in most cases, this is likely to be inaccurate as we do not have enough recapture events from which to sensibly estimate the (standardised) residual variance. For our example we have tried something different and we utilise the fact that we use Bayesian techniques and we have distributions of our predicted recaptures. The argument is as follows: we obtain a distribution of values for $\widehat{R}$ and this is our prediction of the actual recapture numbers so an approximate dispersion value might be

$$\varphi = \frac{\text{Var}(\widehat{R})}{E(\widehat{R})}. \tag{23}$$

This would in principle give us a dispersion value for each recapture event specific to each release event. For the sake of simplicity let us simply use an average dispersion value for each release (and subsequent) recapture event so that our log-likelihood then takes the following form:

$$\log \mathcal{L} = \sum_{t=1}^{T} \varphi_t^{-1} \sum_{tr=1}^{TR_t} \sum_{i} p\left(R_{t,tr,i} \mid \widehat{R}_{t,tr,i}\right), \tag{24}$$

where T is the number of release events, TR_t is the number of tag recapture events, for tagging event t and i again denotes length or age. In essence the dispersion acts as a likelihood weighting between release events and is recursively calculated in that we first run the model with $\varphi \equiv 1$ and then calculate the updated φ_t values, re-run the model the recalculate until they stabilise. This process should not take more than 2-3 runs to stabilise.

4   Priors & penalties

As mentioned in the previous section we intend to use Bayesian modelling and MCMC techniques and this means we need to define priors for the estimated parameters. Also, with such a model it is helpful to place some sensible constraints on areas of parameter space that are simply not plausible and try to dissuade the estimator from going into

these regions. This is normally done using penalties - negative contributions to the log-likelihood incurred when the estimator attempts to go to these nonsensical regions of the parameter space.

4.1 Prior distributions

We have a basic subset of parameters that have to be estimated for the model to work and these are:

- The recruitment values: R_y;

- The initial numbers: $N_{y_{\min},1,\cdot}$.

For these parameters we assume improper uniform priors by which we mean that there is absolutely no contribution to the objective function for any value of these parameters. Secondary estimable parameters would be natural mortality and the recruitment probabilities, $\pi_{y,q}^{\text{rec}}$. For natural mortality it would not be sensible to estimate the values of $M_{q,i}$ - in reality we would need to constrain these parameters in some way and the most sensible and simplistic solution is to simply estimate a single natural mortality multiplier, M, such that $M_{q,i} = M\chi_{q,i}$ where $\chi_{q,i}$ is a fixed ogive to determine the change in natural mortality by length and possibly by quarter. We might then simply apply a normal prior to M which can be informative or not depending on what other information we might wish to include on M from other analyses.

For the recruitment parameters we might be pushing the information limits of our data a little for in terms of expecting to estimate these values so informative prior/penalty constraints might be advisable if we choose to estimate these parameters.

4.2 Penalties

The final requirement is to define some safeguards within the objective function that simply act to keep the estimator out of areas of parameter space that are nonsensical. We

define two penalties for this population and estimation model framework:

- **Catch penalty**: for each instance where the total harvest rate (summed over all gears) is greater than or equal to one (i.e. there is more catch than fish) -100 is added to the log-likelihood

- **Tag penalty**: for each instance where there are not enough fish to tag in any given release event -100 is added to the log-likelihood

These two penalties would act so as to strongly discourage the estimator from going into areas of parameter space that violate these two principles. The log-posterior distribution is now simply the sum of the log-likelihood, log-prior and the log-penalties and a Metropolis-within-Gibb's sampler is employed to sample from the posterior distribution.