Oceans & Atmosphere
www.csiro.au

Proposed new assessment structure for Macquarie Island toothfish using data upto and including 2018

R. Hillary & J.Day

13th April 2019

CSIRO Oceans & Atmosphere
Battery Point, Hobart 7000, Tasmania, Australia.

Copyright and disclaimer

Important disclaimer

Contents

1 Background

This paper details the structure of the proposed revised stock assessment model for Macquarie Island toothfish, the results of fitting the model to the available data. The first reason for moving the modelling framework from Stock Synthesis (SS) to a custom-designed model was future-proofing: the version of SS we currently use is itself an unsupported modification of a previous version of SS, adapted to accommodate time-varying reporting rates in the tagging data. The second reason was to systematically overhaul how the tagging data are included in the assessment itself. They are the major information contributor to historical and current abundance, migration, and fishing mortality. Tagging design work [1] has shown that a reformulation of the likelihood of the tagging model into the so-called spatial Brownie model would consistently improve estimates of migration rates while attaining the same precision in spatial abundance. Indeed, the only major change in the revised assessment is how the tagging data are modelled - the population model and spatial/fishery structure is the same, as are the length and conditional age-at-length data being used. The minor changes to the likelihood model for these latter data sets is also outlined below. The structural similarity for this first version of a proposed revised stock assessment model was deliberate, so as to make sure that the new model was as close as possible to the previous one. This way any differences between them will be - in principle - easier to diagnose. We do, however, outline later on some potential new directions to explore with the revised model, if it is agreed to be used going forward.

2 Data sets

A detailed summary of the three primary data sets can be found in [2].The four primary data inputs to the model are:

1. **Catch biomass**: in tonnes, per fishery
2. **Length frequency**: for each fishery, and using the number of hauls (not fish sampled) as the initial sample size
3. **Conditional age-at-length**: for each fishery and sex, we have the number of fish of a given age conditional on the length class samples came from
4. **Tagging data**: release events are now characterised by a length class and area of release, with repcature data being subsequent total recaptures (across all recapture lengths) in each of the spatial regions of the model.

The structure of the tagging data is the major difference between the previous assessment, and the proposed revised assessment. Before, as SS cannot deal with tag data releases-at-size, an approximation was used where releases are binned into a fixed number of release "age-classes" given the *expected* age-given-length, and subsequent recaptures treated as always belonging to those release cohorts. Realistically, given individual variability in age-given-length and recruitment variations, there are always likely to be more age classes present in the tag releases-at-size than assumed. Additionally, this requires the assumption of a single combined growth curve used to prepare the data for input into the assessment, yet in the assessment the growth function is sexually dimorphic and some parameters thereof are estimated within the assessment. These are the primary reasons for moving to a model for the tagging data where length-at-release, not approximate cohort-of-release, is a primary release covariate. Additionally, sex-at-release is not known and, given we assume selectivity that is driven by size, the recapture probabilities of the tags will be different depending on their sex-at-release. The current structure cannot deal with this additional uncertainty, but the proposed revised assessment deals with it directly.

3 Methods

The revised assessment framework uses the Template Model Builder (TMB) package in R [3]. This is, at present, the most efficient and flexible statistical modelling package available. It allows for highly complex statistical models (including the use of random effects) to be efficiently and robustly estimated. For the MCMC runs used generate the key probabilistic summaries of the assessment variables we use the tmbstan R package [4]. This links models written in TMB to the currently accepted most efficient MCMC sampler (the No U-turns or NUTS algorithm) and, for the models explored, runs in just over an hour as opposed to at least several days required for the SS MCMC algorithm.

3.1 Population and fishery models

The population model is essentially the same as that used in the current SS assessment. It is spatially structured, sex-structured, age and size based and with an assumed annual time-step. The initial population is assumed to be in an unfished equilibrium state. Movement is annual and considered at this stage to be independent of age or size, and is characterised via a time-independent matrix, Φ_{ij}: the probability that an animal in region i will migrate to region j from one year to the next (and we assume a closed population where $\sum_j \Phi_{ij} = 1$). Recruitment is assumed at age 1 and the mean value of total recruitment (across all areas) is governed by the following:

$$\tilde{R}_y = \frac{\alpha S_{♀,y-1}}{1 + \beta S_{♀,y-1}}$$

where $S_{♀,y-1}$ is the biomass of mature females (across all regions) in year $y - 1$ (calculated assuming 50% and 95% lengths-at-maturity of 139.6cm and 185.8cm, respectively). The parameters α and β are defined in terms of the steepness, h, unfished total female mature biomass, B_0, and the unfished total recruitment, R_0:

$$\alpha = \frac{4hR_0}{B_0(1-h)},$$
$$\beta = \frac{5h-1}{B_0(1-h)}.$$

The spatially and sexually disaggregated recruitment, $R_{y,s,r}$, is defined using the time-invariant spatial recruitment faction η_r ($\sum_r \eta_r = 1$), and the sex-ratio at birth, ζ_s ($\sum_s \zeta_s = 1$):

$$R_{y,s,r} = \tilde{R}_y \eta_r \zeta_s,$$

and if annual recruitment deviations, ϵ_y^R, are being estimated then there is an additional multiplier to $R_{y,s,r}$:

$$\begin{aligned} R_{y,s,r} &= \tilde{R}_y \eta_r \zeta_s \exp\left(\epsilon_y^R - \sigma_R^2/2\right), \\ \epsilon_y^R &\sim N\left(0, \sigma_R^2\right). \end{aligned}$$

For ages 2 to $A^+ - 1$ (where A^+ is the plus group) the abundance dynamics are

$$N_{y,a,s,r} = \sum_{r'} \Phi_{r',r} N_{y-1,a-1,s,r'} e^{-M} \left(1 - h_{y,a,s,r'}\right)$$

where M is the age-independent natural mortality rate, and $h_{y,a,s,r}$ is the fishery-aggregated harvest rate by sex and region. For the plus group we have that

$$\begin{aligned} N_{y,A^+,s,r} &= \sum_{r'} \Phi_{r',r} N_{y-1,A^+-1,s,r'} e^{-M} \left(1 - h_{y,A^+-1,s,r'}\right) \\ &+ \sum_{r'} \Phi_{r',r} N_{y-1,A^+,s,r'} e^{-M} \left(1 - h_{y,A^+,s,r'}\right) \end{aligned}$$

The harvest rates are calculated on a fishery-specific basis as follows:

$$\begin{aligned} X_{y,f} &= \sum_s \sum_a N_{y,a,s,r_f} e^{-\tau_{y,f} M} s_{a,s,f} w_{a,s}, \\ h_{y,a,s,f} &= \frac{C_{y,f}}{X_{y,f}} \times s_{a,s,f}. \end{aligned}$$

where:

- $s_{a,s,f}$ is the selectivity-at-age by sex and fishery (calculated from an estimated sexually aggregated selectivity-at-size and integrated across the distribution of size-at-age by sex)
- $w_{a,s}$ is the weight-at-age by sex (calculated from an estimated sexually aggregated length-weight relationship and integrated across the distribution of size-at-age by sex)
- $\tau_{y,f}$ is the fraction of the year at which the median fishing operations occurred
- r_f is the region in which fishery f occurs, and the regionally aggregated harvest rates, $h_{y,a,s,r}$ are just the sum (within year, age and sex) of all the individual fishery-specific harvest rates

3.1.1 Length related variables

All the key data series usesd in the assessment involve size-specific predicted quantities: length distributions in the catch, age-given-length, and length-specific recapture probabilities. The currency of the population and fishery model is primarily age-based, so we need to translate a number of age-based quantities into length:

- Predicted length frequency (aggregated across sexes) for each fishery
- Predicted distribution of age-given-length, accounting for ageing error, in each of the fisheries and for both sexes
- Predicted sex ratio-at-length for each region

- Predicted spatial recapture probability-at-length, derived from length-based harvest rates and the growth transition matrices for each sex

The predicted length frequency is actually derived from the “true” age-given-length distribution, so we begin with that first. The distribution of length-at-age is simply defined from the growth relationship. The mean length-at-age is defined via the Schnute parameterisation of the von Bertalanffy growth curve:

$$\mathbb{E}\left(l(a)\right) = l_1(a_1) + \left(l_2(a_2) - l_1(a_1)\right)\frac{1-\exp\left(-k\left(a-a_1\right)\right)}{1-\exp\left(-k\left(a_2-a_1\right)\right)},$$

where $l_1(a_1)$ and $l_2(a_2)$ are the lengths at reference ages a_1 and a_2 ($a_2 > a_l$), and k is the growth rate. To generate the distribution of length-at-age we assume a lognormal distribution (with a given standard deviation σ_l) around this mean length-at-age. This gives us a sex-specific distribution of length-at-age, $\pi_{l\,|\,a,s}$.

To get to the “true” distribution of age-given-length we use Bayes’ rule:

$$\tilde{\pi}_{a\,|\,y,l,s,f} = \frac{\pi_{l\,|\,a,s}\pi_{a\,|\,y,s,f}}{\pi_{l\,|\,y,s,f}},$$

where $\pi_{y\,|\,a,s,f}$ is the prior age distribution, and $\pi_{l\,|\,y,s,f}$ is the length distribution in the fishery:

$$\pi_{l\,|\,y,s,f} = \sum_a \pi_{l\,|\,a,s}\pi_{a\,|\,y,s,f},$$

and the prior age distribution is defined as follows:

$$\pi_{a\,|\,y,s,f} = \frac{N_{y,a,s,r_f}s_{a,s,f}}{\sum_i N_{y,i,s,r_f}s_{i,s,f}}.$$

For a given ageing error matrix, $A_{a,a'}$, where $\sum_a A_{a,a'} = 1$ and a' is the “true” age in this sense, the adjusted distribution of age-given-length (that we use to compare to the observations) is defined as

$$\pi_{a\,|\,y,l,s,f} = \sum_{a'} \tilde{\pi}_{a'\,|\,y,l,s,f}A_{a,a'}.$$

With fleet and sex-specific true and observed age-given-length distributions we can then derive the length-based harvest rates we will need later on for the tagging likelihood. The length-based harvest rate for each fleet is define to be

$$h_{y,l,s,f} = \sum_a h_{y,a,s,f}\tilde{\pi}_{a\,|\,y,l,s,f},$$

and the cumulative harvest rate in each area, $h_{y,l,s,r}$, is then just summed across the fisheries operating in a given area. The final required variable, required for calculating sex-aggregated length frequencies and the tagging likelihood, is the true sex-ratio by length and area, $\xi_{y,l,s,r}$:

$$\begin{aligned}
\pi_{a\,|\,y,s,r} &= \frac{N_{y,a,s,r}}{\sum_i N_{y,i,s,r}}, \\
\tilde{\pi}_{a\,|\,y,l,s,r} &= \frac{\pi_{l\,|\,a}\pi_{a\,|\,y,s,r}}{\sum_i \pi_{l\,|\,i}\pi_{i\,|\,y,s,r}}, \\
N_{y,l,s,r} &= \sum_a N_{y,a,s,r}\tilde{\pi}_{a\,|\,y,l,s,r}, \\
\xi_{y,l,s,r} &= \frac{N_{y,l,s,r}}{\sum_j N_{y,l,j,r}}.
\end{aligned}$$

For the tagging likelihood we need to calculate a sex-specific growth transition matrix given the length-based nature of this part of the model. This is done following the method outlined in [5] that deals with both the differing size of the length bins, and the stochastic uncertainty in the expected growth increments of the fish, given the growth curve. The transition matrix, $G_{l,l',s}$, is the probability that a fish in length bin l after a given time τ (taken to be one year here) will be in length bin l' (and $\sum_{l'} G_{l,l',s} = 1$).

3.1.2 Candidate selectivity functions

Selectivity is assumed to be inherently length-based and not sexually dimorphic, even though selectivity-at-age is given possibly different growth curves for males and females. We explored three potential selectivity functions:

1. **Double-logistic**: essentially a fully smooth function that encompasses the features of the double-normal and double-normal plateau functions
2. **Generalised gamma**: uses a modified gamma distribution-type kernel that is a reduced parameter dome-shaped distribution to avoid over-parameterisation and convergence issues of the double-logistic function when the plateau-type dynamics are absent
3. **Logistic**: usual logistic function that has no potential for dome-shaped dynamics

The main reason for not using the double-normal (or the plateau extension) is because of memory issues around taping for the AD part of the algorithm. True AD cannot deal with if/else type statements for parameters or variables derived from parameters. The two double-normal variables have exactly this structure and it can significantly increase the amount of memory used by AD algorithms, as they have to create exceptions in the taping procedure (doubling up of sorts every time you have an if/else condition). To maximise efficiency we therfore explored only truly smooth selectivity functions - noting that the ones we used have essentially the same flexibility as the piecewise smooth ones we have decided not to use.

The double-logistic function we use is defined as follows:

$$s_l^{\mathrm{dlog}} = \left(1 + 19^{-(l-\psi_{1,50})/\delta_{1,95}}\right)^{-1} \times \left(1 - \left(1 + 19^{-(l-\psi_{2,50})/\delta_{2,95}}\right)^{-1}\right),$$

and interpretation of parameters is as follows: $\psi_{1,50}$ is the length at which the selectivity first increases to 0.5; $\delta_{1,95}$ the additional “distance” to reaching 0.95; $\psi_{2,50}$ is the length at which selectivity has decreased to 0.5; and $\delta_{2,95}$ is the distance backwards from $\psi_{r,50}$ at which the selectivity is decreasing but at 0.95.

The generalised gamma selectivity is define as follows:

$$s_l^{\text{gamm}} = \frac{l^{a_g} \exp\left(-b_g l^{c_g}\right)}{\mathcal{G}},$$
$$\mathcal{G} = \left(\frac{a_g}{b_g c_g}\right)^{a_g/c_g} \exp\left(-a_g/c_g\right)),$$

and the function has a pre-defined maximum of 1 at $l = a_g/(b_g c_g)$, and the additional parameter c_g assists in the ability of the curve to reproduce strongly asymmetric dome-shaped dynamics.

The logistic function is just the standard robustly parameterised version:

$$s_l^{\log} = \left(1 + 19^{-(l-\psi_{50})/\delta_{95}}\right)^{-1},$$

where ψ_{50} is the length at 50% selection, and $\psi_{50} + \delta_{95}$ is the length at 95% selection.

3.2 Likelihood functions

We have now defined all the key population and fishery variables so we now move on to the likelihood functions of the three main observations used within the assessment.

3.2.1 Length frequency data

At this stage, we use sex-aggregated length frequency data (as is done in the current assessment) and the predicted length distribution is calculated as follows:

$$\pi_{l\,|\,y,f} = \frac{\pi_{l\,|\,y,s,f}\,\xi_{y,l,s,r_f}}{\sum_j \pi_{l\,|\,y,j,f}\,\xi_{y,l,j,r_f}}.$$

In the SS assessment model the data are currently modelled using the multinomial distribution, with the associated effective sample sizes “tuned” so that input variance assumptions match those coming out post-fitting. For the revised assessment model we propose an extension of the multinomial distribution called the Dirichlet-multinomial (D-M) distribution. The reason for doing this is so we have a flexible distribution where the correct weighting (effective sample size) for each fishery can be calculated on a robust statistical footing, as opposed to more *ad hoc* tuning algorithms. The D-M distribution has, at its core, the underlying multinomial distribution. The secondary part of this distribution assumes that the underlying probabilities that define the multinomial vary according to the Dirichlet distribution. This permits us to model additional variability (over-dispersion) in the length data in a coherent and statistically well-defined manner.

Let $n_{y,l,f}$ be the number of observations for which we have length data in bin l, in year y, for a given fishery f. The joint density of both the observations and the predicted length frequencies can be defined as follows:

$$p\left(\mathbf{n}\,|\,\mathbf{p},\boldsymbol{\alpha}\right) = p\left(\mathbf{n}\,|\,\mathbf{p}\right) p\left(\mathbf{p}\,|\,\boldsymbol{\alpha}\right),$$

where $\mathbf{n}$ is the vector of length data, $\mathbf{p}$ their underlying predicted multinomial probability, and $\boldsymbol{\alpha}$ the Dirichlet parameters. The first density is the multinomial, the second the Dirichlet in the above equation. What we really need though is the *marginal* distribution of $\mathbf{n}$ given $\boldsymbol{\alpha}$:

$$p\left(\mathbf{n}\,|\,\boldsymbol{\alpha}\right) = \int_{\mathbf{p}} p\left(\mathbf{n}\,|\,\mathbf{p}\right) p\left(\mathbf{p}\,|\,\boldsymbol{\alpha}\right) \mathrm{d}\mathbf{p}.$$

This integral has a closed-form solution and leads to the following marginal distribution for the length frequency data:

$$\Lambda^{l}_{y,f} = \frac{(n_{y,f}!)\Gamma(\omega_{y,f})}{\Gamma(n_{y,f}+\omega_{y,f})} \prod_{l} \frac{\Gamma(n_{y,l,f}+\omega_{y,f}\pi_{l\,|\,y,f})}{n_{y,l,f}!\Gamma(\omega_{y,f}\pi_{l\,|\,y,f})}$$

where $n_{y,f} = \sum_{l} n_{y,l,f}$, $\Gamma()$ is the gamma function, and the over-dispersion parameter, $\omega_{y,f}$, is defined as follows:

$$\omega_{y,f} = \frac{n_{y,f} - \varphi_f}{\varphi_f - 1},$$

and $\varphi_f > 1$ is the over-dispersion *factor*: the degree to which the multinomial variance is inflated due to correlation between the length classes. The point of going to the trouble of using the D-M formulation is that φ_f is an estimable parameter (as opposed to tuning to get the right value of $n_{y,f}$). The overall length frequency likelihood, Λ^{l} is obtained by multiplying over all years y and fisheries f.

3.2.2 Conditional age-at-length data

The data are the actual number of fish-at-age for a given fishery, length-class, sex and year: $n_{y,a,l,s,f}$. We assume a multinomial distribution for this likelihood as the default, primarily because we assume size dictates selectivity, so we would then expect that the distribution of age *within a given length class* would be random (i.e. multinomial in this case). So, the likelihood of the age-given-length data is as follows:

$$\Lambda^{a|l}_{y,l,s,f} = \prod_{a} \left(\pi_{a\,|\,y,l,s,f}\right)^{n_{y,a,l,s,f}}$$

3.2.3 Tagging data

So far the likelihoods are either slight modifications (such as the D-M for the length frequency data) or essentially the same (for the age-given-length data) as in the current SS assessment. For the tagging data we use a fundamentally different tag recapture model and associated likelihood. To set the scene for why we include the elements we do we first outline the processes that we have to account for:

- The sex-at-release is (almost always) not recorded, and the sexes will almost certainly have differing mortality rates and, therefore, different recapture probabilities
- The length-at-release is the key covariate, not age, and the fish grow probabilistically (via the size-transition matrix $\mathbf{G}$) over time
- It is more natural to follow the correlated recapture history of a given recapture event (pooled via year, length and region of release), not to have a kind of pooled tagged population and calculate the likelihood of recapture (both in total numbers and in terms of the spatial distribution) of these pooled tags year after year

The tag recapture model we derive fits within what would be considered a *multi-state mark-recapture model*. By this we mean there are a number of probabilistic states a tagged fish can inhabit over the recapture period of a given release event: which length class it is within, what spatial region it is in, what sex it is, and whether it has been recaptured or not. The release covariates are year, length class and region; the recapture covariates are year and region of recapture. So we will integrate across size at recapture and sex-at-release (we don't use the sexed tag recapture information) within the tagging model.

The number of tag releases is denoted by $T_{\tilde{y},\tilde{l},\tilde{r}}$, and the number of associated recaptures $R_{t,r}$ for $t = \tilde{y}+1,...,\tilde{y}+t_{\max}$, and $t_{\max}$ is either the maximum number of possible recapture events (and we have a pre-specified absolute maximum). As with the SS assessment, we do not use any within-season recapture events in the recapture data. To construct the overall tag recapture probability, we will first construct the sex-specific tag survival probabilities, and in the following $\tilde{\mathbf{z}} = \{\tilde{y},\tilde{l},\tilde{r}\}$ i.e. the vector of release covariates. To construct the tag survival probabilities we first need to track the size and spatial distribution of tag releases, $\gamma_{t,l,s,r}$ over the recapture period. Consider a single tag in given release event: $m_{\tilde{y},\tilde{l},s,\tilde{r}} = \pi^{\text{mort}}$, and zero elsewhere (and π^{mort} is the tag mortality probability). The dynamics of this single tag are as follows: for $t > \tilde{y}$:

$$m_{t,l,s,r} = \pi_t^{\text{tag}} e^{-M} \sum_{r'} \sum_{l'} m_{t-1,l',s,r'} G_{s,l',l} \Phi_{r',r} \left(1 - h_{t-1,l',s,r'}\right),$$

and π_t^{tag} is the probability of a double-tagged fish retaining *at least* one tag from time $t-1$ to time t. The equation itself is complex, but it combines tag loss/mortality, growth (via $\mathbf{G}$), migration (via $\mathbf{\Phi}$), natural mortality (via M) and fishing (via the harvest rates $h_{t,l,s,r}$).

The relative distribution of tags from a given release event (across both size and space) is defined as follows:

$$\gamma_{t,l,s,r} = \frac{m_{t,l,s,r}}{\sum_{l'} \sum_{r'} m_{t,l',s,r'}}.$$

The next variable we need is tag survival, π_t^s, marginalised across length and region. Obviously $\pi_{\tilde{y}}^s = \pi^{\text{mort}}$, and for $t > \tilde{y}$

$$\pi_t^s = \pi_{t-1}^s \pi_t^{\text{tag}} e^{-M} \sum_r \sum_l \gamma_{t-1,l,s,r}(1 - h_{t-1,l,s,r}),$$

which finally leads to the tag recapture probability (marginalised across length):

$$\pi_{t,s,r}^{\text{rec}} = \pi_t^s \pi_t^{\text{rep}} \sum_l \gamma_{t,l,s,r} h_{y,l,s,r},$$

where π_t^{rep} is the reporting rate at time t. While mathematically daunting, in plain words this is basically the product of three things:

1. The probability that the tagged fish survives with *at least* one tag attached to time t
2. The probability that the tag would be reported if recaptured at time t
3. The (length-averaged) probability of being caught at time t

We could go even more complex and condition on the length-at-recapture as well as time and region. In the positive column, this has far more direct information on growth (if estimating growth parameters within the model); in the negative column it massively increases the complexity of the tagging likelihood and the memory use required for the AD part of the optimisation algorithm. We are not - at this stage - exploring size-specific movement estimation, so the information on abundance and migration is not really going to benefit from having size as an additional recapture covariate. It is, however, something to consider for future possible configurations. The final step is to average out over the sex of the tagged animals (given we don't know this at release time) to get the overall recapture probability:

$$\pi^{\text{rec}}_{t,r} = \sum_s \pi^{\text{rec}}_{t,s,r} \xi_{\tilde{y},\tilde{l},s,\tilde{r}}$$

The base likelihood for the tagging data in this format is essentially the multinomial distribution, which is known loosely as the Brownie (size and spatially structured in this case) model [6]. This follows the recapture history of a given release event and, as we outlined in the background section, has been shown to be more informative on both abundance and migration, relative to the current SS two-stage likelihood. Tagging data are, however, well known to be often over-dispersed (i.e. more variable than the underlying base distribution would predict). To accommodate this process we again use the Dirichlet multinomial (D-M) distribution to model the likelihood of a given tagging event's recapture history. The D-M likelihood for the tagging data is

$$\Lambda^{\text{tag}}_{\tilde{\mathbf{z}}} = \frac{T_{\tilde{\mathbf{z}}}!\Gamma(\vartheta_{\tilde{\mathbf{z}}})}{\Gamma(T_{\tilde{\mathbf{z}}} + \vartheta_{\tilde{\mathbf{z}}})} \left(\prod_t \prod_r \frac{\Gamma(R_{t,r} + \vartheta_{\tilde{\mathbf{z}}} \pi^{\text{rec}}_{t,r})}{R_{t,r}!\Gamma(\vartheta_{\tilde{\mathbf{z}}} \pi^{\text{rec}}_{t,r})} \right) \frac{\Gamma(T_{\tilde{\mathbf{z}}} - \mathcal{R} + \vartheta_{\tilde{\mathbf{z}}}(1 - \tilde{\pi}^{\text{rec}}))}{(T_{\tilde{\mathbf{z}}} - \mathcal{R})!\Gamma(\vartheta_{\tilde{\mathbf{z}}}(1 - \tilde{\pi}^{\text{rec}}))}, \tag{3.1}$$

where

$$\vartheta_{\tilde{\mathbf{z}}} = \frac{T_{\tilde{\mathbf{z}}} - \varphi^{\text{tag}}}{\varphi^{\text{tag}} - 1},$$

and φ^{tag} is the tagging over-dispersion factor, and

$$\mathcal{R} = \sum_t \sum_r R_{t,r},$$

and

$$\tilde{\pi}^{\text{rec}} = \sum_t \sum_r \pi^{\text{rec}}_{t,r},$$

is the overall probability of recapturing a single tag from that release event. The probability of *never* recapturing a tag is the final term in Eq. (3.1) and is required to ensure that our recapture history probabilities sum to 1. The likelihood of all the tag data is just the product of the likelihoods of all the unique release events. As with the length frequency likelihood, the over-dispersion factor can now be estimated directly without needing to be tuned in some *ad hoc* manner.

3.2.4 Overall likelihood and objective function

The overall log-likelihood of the data is simply the sum of all three log-likelihoods of the data sources:

$$\ln \Lambda^{\text{tot}} = \ln \Lambda^{l} + \ln \Lambda^{a|l} + \ln \Lambda^{\text{tag}}$$

The full objective function to be maximised includes the recruitment prior and some additional penalties to stop harvest rates and tag recapture probabilities exceeding pre-specified maximum levels.

3.3 Estimated parameter options

The core set of estimated parameters are:

- Unfished total recruitment, R_0

- Selectivity parameters for each fishery
- Recruiment deviations for a pre-specified subset of years
- Spatial recruitment parameters, η_r
- Parameters of the migration matrix, $\mathbf{\Phi}$
- Over-dispersion parameters φ_f and φ^{tag}

The optionally estimable parameters are either all or some subset of the growth parameters, k, l_1, l_2 and σ_l.

4 Data & dimensions

This section deals with some high-level summaries of the input data, as well as the relevant dimensions of the model (years, ages, size classes etc.) and what specific choices are made about the different parameterisations for the various model processes.

4.1 Dimensions

The model runs from 1985 to 2018 (i.e. 10 years before fishing began), ages are from 1 to 52. Size-classes range from 0 to 190cm: 0 to 30 in 10cm bins, 30 to 140cm in 5cm bins, and from 140 to 190cm in 10cm bins. The model is run, as is the current assessment, as a two region model with a Northern and Southern region (with the same latitudinal separator for these regions as used in the current assessment). There are five fisheries:

1. Aurora trough trawl (ATT): assumed in region 2 (Southern region) and with an assumed time-invariant double-logistic selectivity
2. Northern valleys trawl (NVT): assumed in region 1 (Northern region) and with an assumed time-invariant generalised gamma selectivity
3. Aurora trough long-line (ATL): assumed in region 2 (Southern region) and with two possible selectivity options: generalised gamma or logistic
4. North Macquarie ridge longline (NMRL): assumed in region 1 (Northern region) and with two possible selectivity options: generalised gamma or logistic
5. South Macquarie ridge longline (SMRL): assumed in region 2 (Southern region) and with two possible selectivity options: generalised gamma or logistic

4.2 Data summaries

The major, more detailed summaries on the key date sets can be found in [2], in this paper we focus specifically on the summaries that relate directly to what we assume in the revised assessment model. For the length frequency data, we do not use the number of actually measured fish as the base sample size starting point. These often number in the thousands, and would imply a precision in the size data that is simply not either *a priori* credible, or born out by the fits to these data and what kind of *effective* sample sizes we actually estimate post-fitting. As the base unit of sample size, we actually employ the number of hauls, which range from less than 10 in some years to over 200 in the years where most catch is taken. This variable reflects relative changes in sampling effort fairly well - given the observer coverage more hauls clearly leads to more sampling effort - and is in the plausible range of values for what effective sample size might be. Given we estimate over-dispersion factors we can see *a posteriori* how well the number of hauls links to the overall effective sample size.

For the tagging data, we stipulate that only unique tagging events with at least five releases can be used. Not only does this minimise potentially uninformative tagging data simply adding to the memory required to run the model, but it ensures that the number of tag releases never falls below the over-dispersion factor - this is effectively an implicit zero sample size issue and causes the tagging likelihood to be undefined. We also stipulate there to be a user-defined maximum number of recapture events allowed for each tagging event. The reference value is set at 7: so after the year of release we allow a maximum of 7 recapture events. This includes over 90% of the actual recapture data. We do this to avoid recaptures at times-at-liberty for which we are unsure as to whether our assumptions about tag loss etc. are still holding true. We do, however, run a sensitivity with a maximum of 9 recapture events (including over 97% of the tag recaptures) to see what difference, if any, this assumption makes. For the tagging data we also explore a sensitivity test using the estimates of tag shedding from the actual tagging data themselves.

5 Results

This section summarises:

- Reference model configuration and fits to the various data sets
- Population dynamic summaries from the MCMC runs for the reference model
- Impact of the outlined sensitivity runs

5.1 Reference assessment model

The reference assessment model has the dimensions outlined in Section 4.1, and uses the data as outlined in Section 4.2. For the reference assessment model, we assume that the reference ages for the Schnute parameterisation of the von Bertalanffy growth function to be $a_1 = 5$ and $a_2 = 20$. This ensures that they are (a) are within the observed data range, and (b) are not too close or too far apart, relative to the data range. For the reference model we keep the growth parameters *fixed*, estimating them using the conditional age-at-length method outlined previously [7]. So, we are using these data to inform the model on population size and age structure (including recruitment), not growth - a sensitivity is explored estimating the growth parameters.

Variable	k	l_1	l_2	L_∞	t_0	σ_l
Female	0.055 (0.003)	0.49 (0.004)	1.15 (0.005)	1.67 (0.04)	-1.29 (0.18)	0.15 (0.015)
Male	0.071 (0.004)	0.48 (0.003)	1.01 (0.008)	1.29 (0.04)	-1.63 (0.21)	0.15 (0.02)

Table 5.1: *Maximum likelihood estimates (and approximate standard errors in brackets) of the growth parameters used in the reference model.*

A detailed summary of the estimation of the growth parameters can be found in [8] but Table 5.1 shows the estimate used as model inputs in the reference case. As seen in previous analyses, males seem to grow faster initially, but to a smaller asymptotic length; as a result, size-at-age (and weight) of females is greater than males from about age 5 onwards. The key mean length parameters (k, l_1, and l_2) are all very accurately estimated. Variability in mean length-at-age is very well estimated in both cases and the same for both sexes. The standard errors are informative in that it makes it fairly clear that uncertainty in growth is arguably the least of all the parameters used as inputs to the model, or estimated therein (see later).

5.2 Fitting summary for reference model

The fits to the length frequency data for the two trawl fleets are in Figure 5.1, and for the three longline fleets in Figure 5.2.

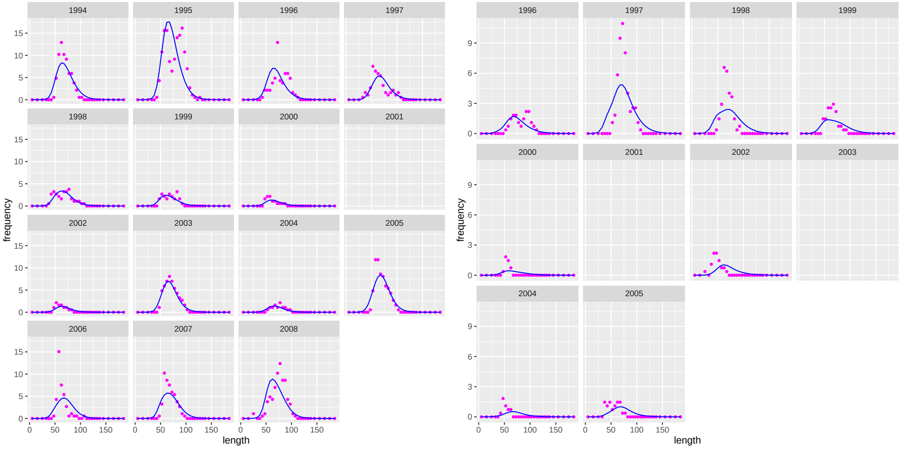

Figure 5.1: *Fits to the ATT (left) and NVT (right) trawl fisheries length data. Magenta circles are the observed data, and the blue lines the predictions.*

Figure 5.3 shows the fits to the female length-conditional age data for the two trawl fleets. and Figure 5.4 shows the same for the males. Figure 5.5 shows the fits to the female length-conditional age data for the three longline fisheries, and Figure 5.6 shows the same for the males.

The fits to the tagging data are summarised in four key ways:

1. Successive recaptures for each year of releases
2. Total recaptures for each year of release
3. Total recaptures for each year of recapture
4. Total recaptures for each year and region of recapture

All of these summaries aggregate across the size spectrum of releases and recaptures for visual brevity, and also because size-at-recapture is not an explicit part of the tagging likelihood.

5.2.1 Relative data “weighting” estimates

The main difference with a tuning-type approach to data-weighting employed in the current SS assessment, and the one proposed in the revised assessment, is to move to actively estimating the key weighting parameters (so-called “right weighting” or, more accurately, empirical Bayesian inference for hierarchical models). Focussing on the trawl length data first: for the ATT and NVT fleets there is clear down-weighting of the haul data - more so for the NVT fleet. For the longline fleets, ATL and SMRL are down-weighted very little, but the NMRL fleet is clearly down-weighted. For the ATT data this looks like genuinely random variation; for the NVT data more some kind

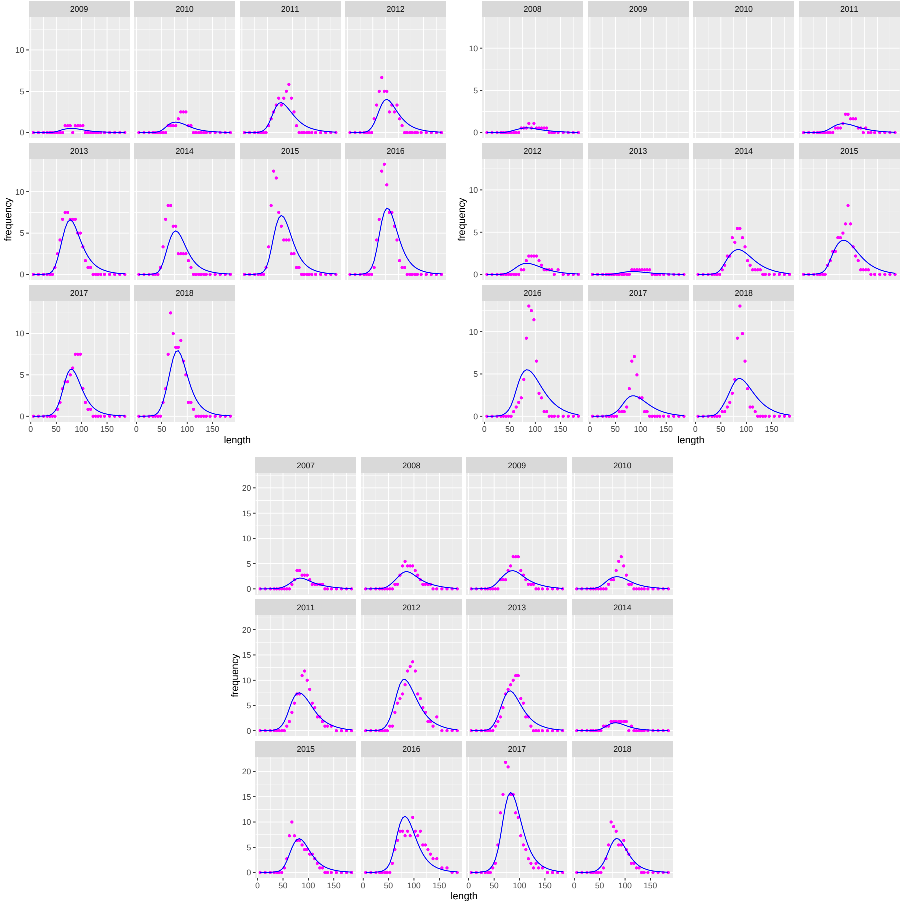

Figure 5.2: *Fits to the ATL (top left), NMRL (top right), and SMRL (bottom) longline fisheries length data. Magenta circles are the observed data, and the blue lines the predictions.*

of systemic lack of fit given the clear decrease in mean length over time (and the assumption of time-invariant selectivity). For the NMRL data by convention we assume logistic selectivity for this and the SMRL fleet to avoid the appearance of cryptic spawner biomass in the population. While logistic selectivity is actually the mode of choice for the ATL, and would be for SMRL if permitted the choice, it seems that we consistently over-estimate the right-hand limb of the length frequency curve in the last five years of data for the NMRL fleet.

Variable	φ_{ATT}	φ_{NVT}	φ_{ATL}	φ_{NMRL}	φ_{SMRL}	φ^{tag}
Estimate	1.86	2.74	1.2	1.84	1.1	1.84

Table 5.2: *Estimates of the over-dispersion factors for the size data for each fleet, φ_f, and the tagging data, φ^{tag}.*

For the tagging data we see that the estimate of $\varphi^{\text{tag}} = 1.84$ clearly suggests that the tagging data are over-dispersed, relative to the assumption of a straight multinomial recapture likelihood. Interestingly, it is close to the value of around 1.9 assumed (but not tuned) for the tagging data likelihood (for total recaptures) in the current SS assessment.

For the conditional age-at-length data we assumed a multinomial distribution, given the theory about size-selectivity versus age would suggest that age data from within a given length class

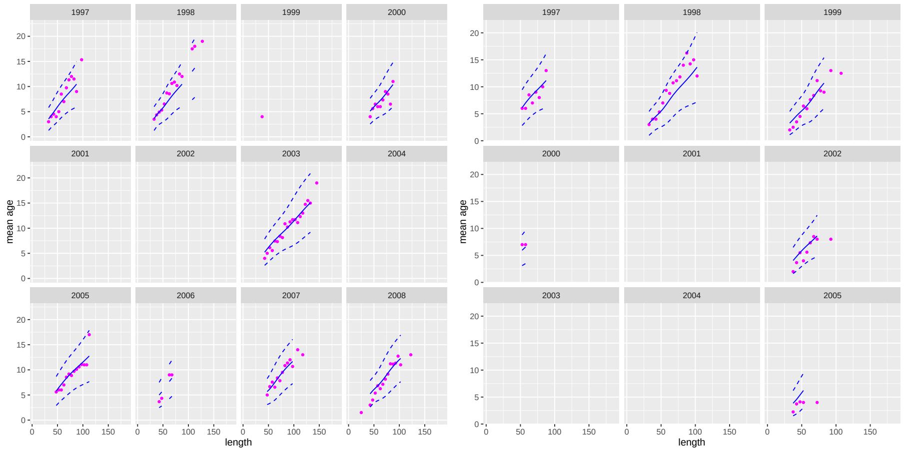

Figure 5.3: *Fits to the ATT (left) and NVT (right) trawl fisheries female age-given-length data. Magenta circles are the observed mean age, and the blue lines the predicted median and 95%ile.*

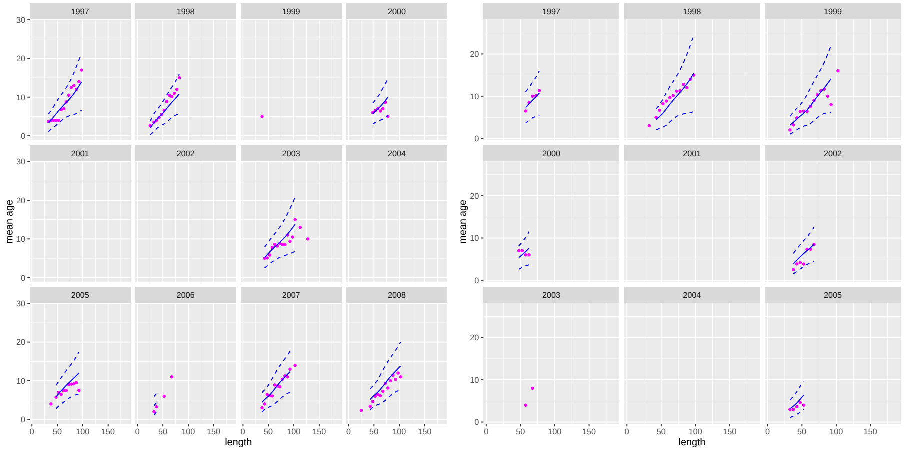

Figure 5.4: *Fits to the ATT (left) and NVT (right) trawl fisheries male age-given-length data. Magenta circles are the observed mean age, and the blue lines the predicted median and 95%ile.*

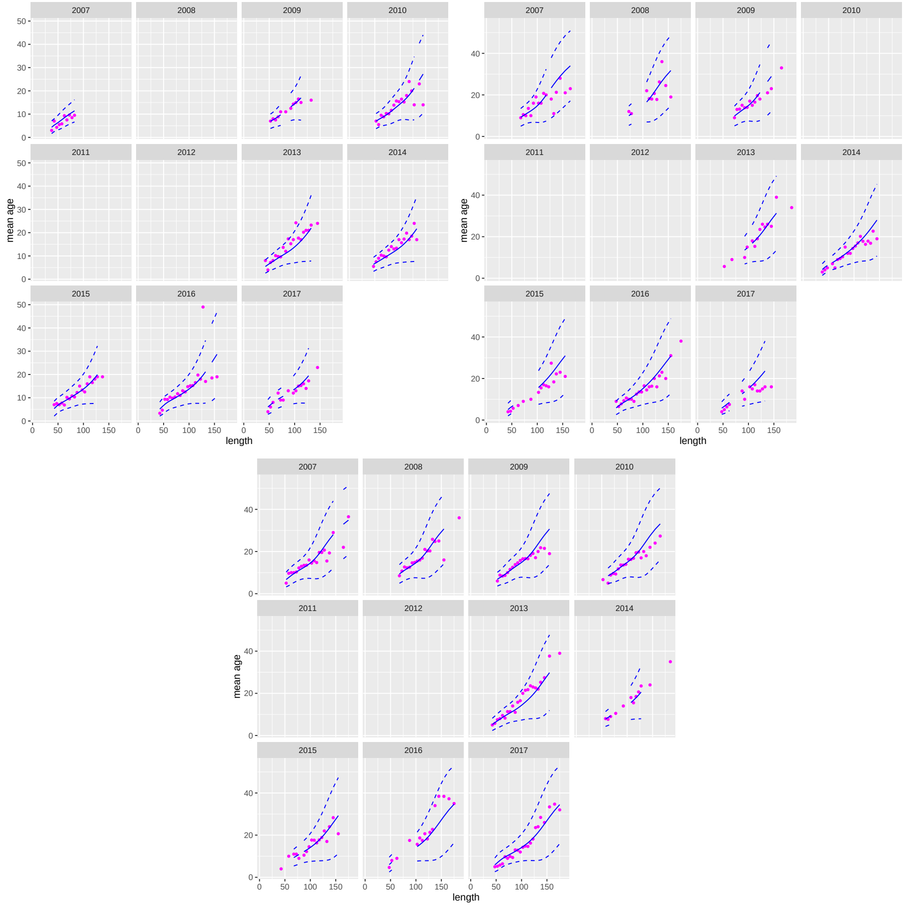

Figure 5.5: *Fits to the ATL (top left), NMRL (top right), and SMRL (bottom) longline fisheries female age-given-length data. Magenta circles are the observed mean age, and the blue lines the predicted median and 95%ile.*

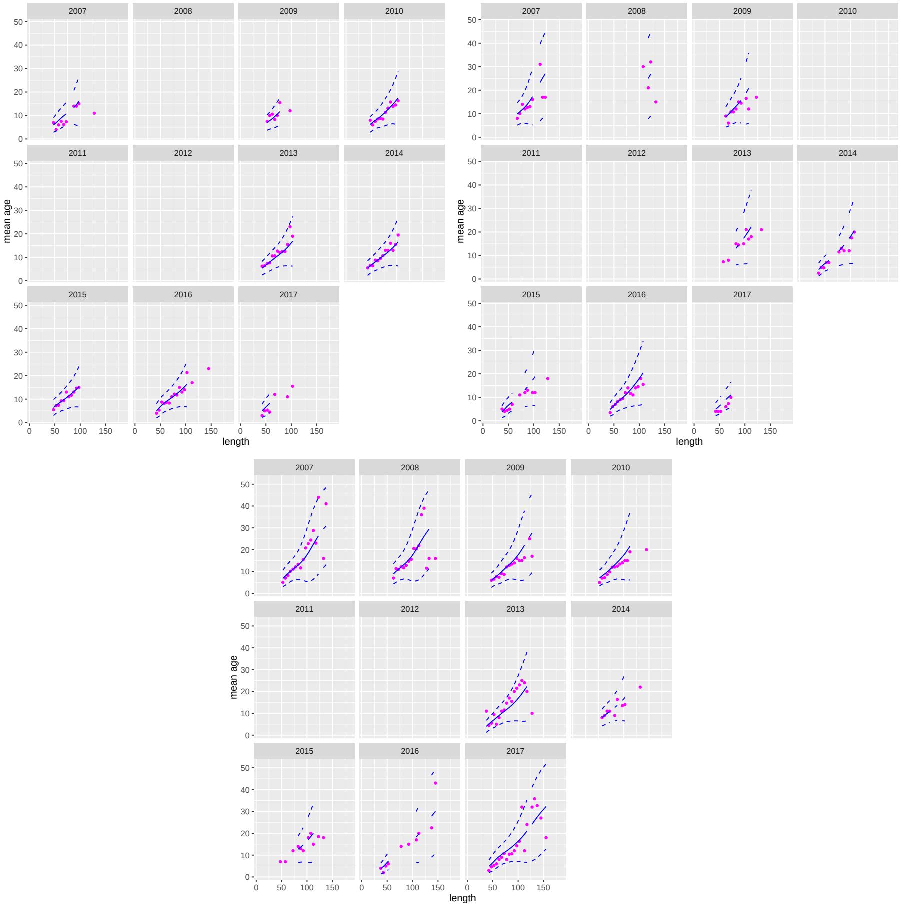

Figure 5.6: *Fits to the ATL (top left), NMRL (top right), and SMRL (bottom) longline fisheries male age-given-length data. Magenta circles are the observed mean age, and the blue lines the predicted median and 95%ile.*

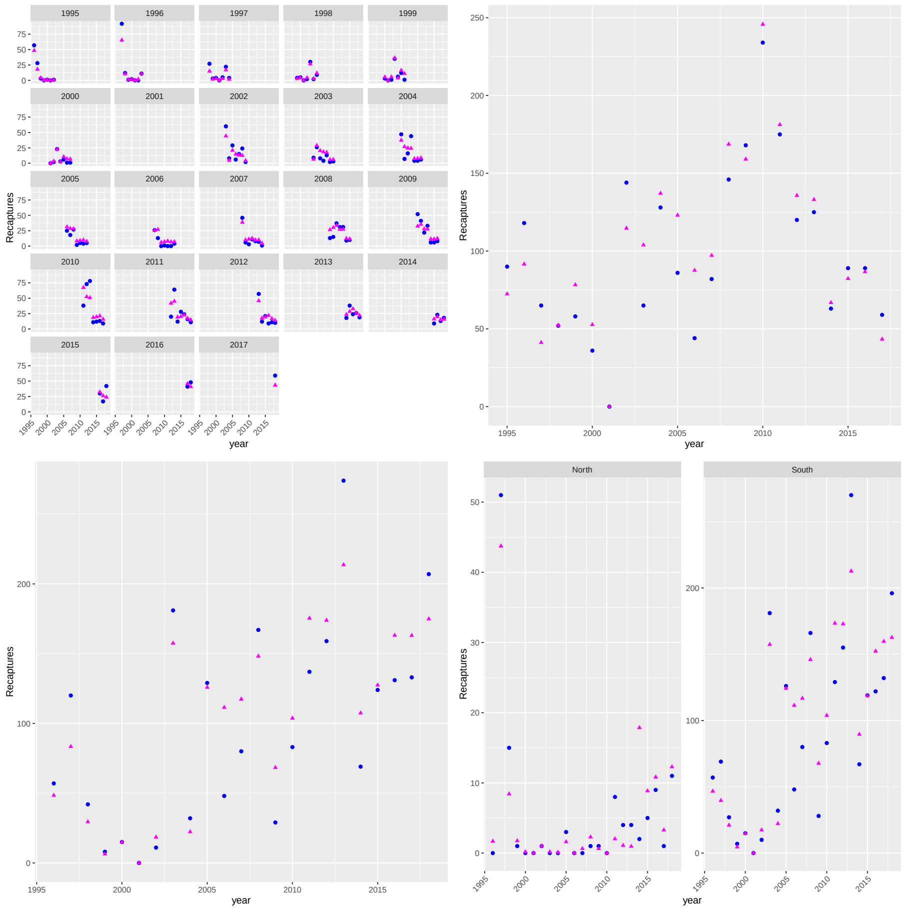

Figure 5.7: *Fits to the tagging data (blue circles, observed; magenta triangles, predicted) for recaptures following year of release (top left), total recaptures for each year of release (top right), total recaptures for year of recapture (bottom left), and recaptures for each year and region of recapture (bottom right).*

would be random (hence, the multinomial would be the right choice). The reality of whether this is true can only be gleaned once the model has been fitted to the data. When looking at all the fits to the data for each sex and fishery (Figs. 5.3–5.6) we see that, barring a few isolated examples, the observed mean length-at-age sits within the predicted 95% interval and doesn't systematically appear above or below the predicted mean. When one analyses the standardised residuals for over-dispersion (do they systematically appear greater than 1) there is no evidence that a move to the over-dispersion model (Dirichlet-multinomial) is required. This seems to suggest that:

- The multinomial distribution assumed for these data appears valid
- The model's predictions of age-given-length are clearly statistically consistent with the data and the assumed growth model
- At least for these data, the model has enough freedom to adequately explain the observations
- It would sort of back up the underlying assumption that size (not age) is the right underlying variable with which to parameterise selectivity

5.3 Population dynamic summaries from MCMC runs

For the reference assessment case, we used the `tmbstan` R-based MCMC package [4] to sample from the posterior distribution. The package uses the Hamiltonian MCMC algorithm, designed to solve a lot of the problems with the more traditional MCMC algorithms, when it comes to sampling from complex high-dimensional posterior surfaces. As a result, it is able to obtain a convergent MCMC sample from the posterior (1,000 iterations) in an about 80 mins (as opposed to days for the SS ADMB-driven MCMC algorithm). The key female SSB summaries can be found in Figure 5.8; total recruitment and the key spatial parameters (recruitment fraction to North, η_1, and migration rates between regions) can be found in Figure 5.9.

The current (*ca.* 2018) median estimate (and 95% credible interval) of female SSB depletion is 0.7 (0.65–0.76). As with previous assessments, the estimated overall level of female SSB (and depletion thereof) is consistently higher in the North, relative to the Southern region. Total recruitment has generally varied randomly around the mean level, with short periods of higher or lower recruitment, but not sustained periods of either (showing low levels of temporal autocorrelation).

The spatial recruitment fraction to the Northern region has a median (and 95% credible interval) of 0.15 (0.09–0.23) - this is quite different to previous estimates that have this value around the 0.45 level. Migration point estimates are similar (around 1% *per annum* from North to South, and 6% from South to North) albeit very slightly higher from South to North and with a marginal posterior less skewed to lower transfer rates in this direction. The reality is that one can obtain the same effective spatial distribution of animals by **either** depositing more or less recruits into a region, **or** having more or less fish move between regions. Additionally, a (comparatively) large change in the spatial recruitment parameter, can be offset by a much smaller proportional shift in a migration parameter. The spatial recruitment dynamic is a "one off" event; migration is the consistent movement of every age-class year upon year. It does not take much change in the latter to offset a change in the former (as is the case here).

Differences between the relative sizes of the Northern and Southern regions largely depend on

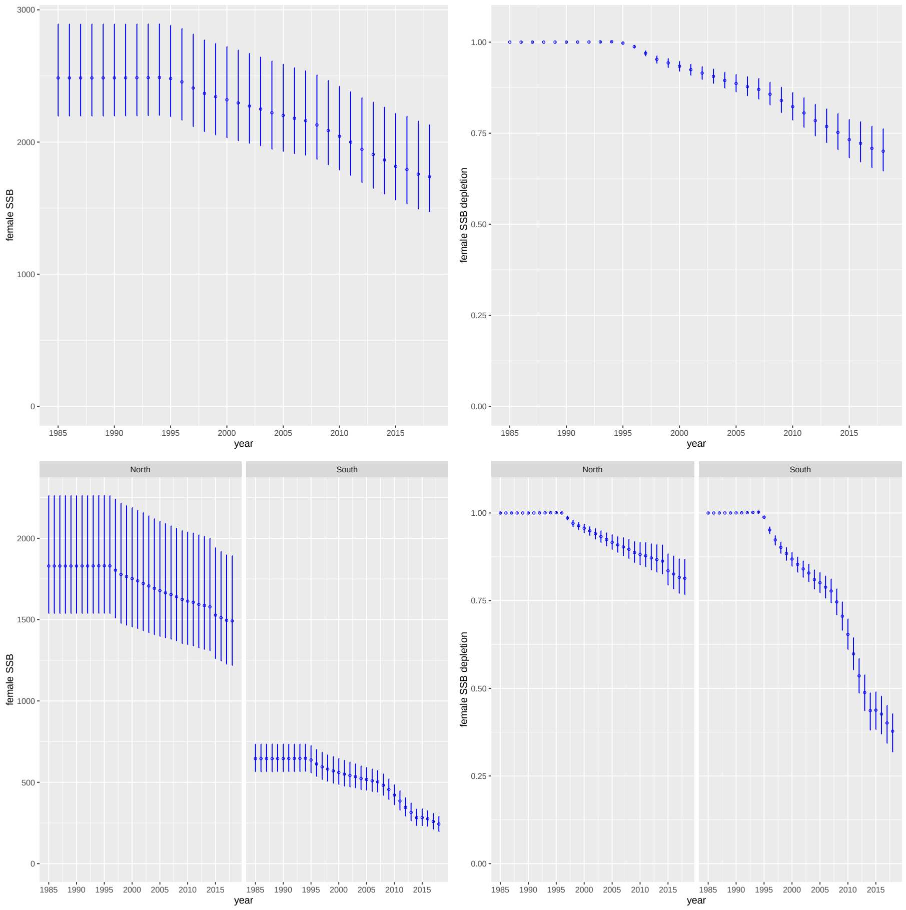

Figure 5.8: *Posterior median and 95% credible intervals for total female SSB (top left), female SSB relative depletion (top right), spatial female SSB (bottom left), and spatial female SSB relative depletion (bottom right).*

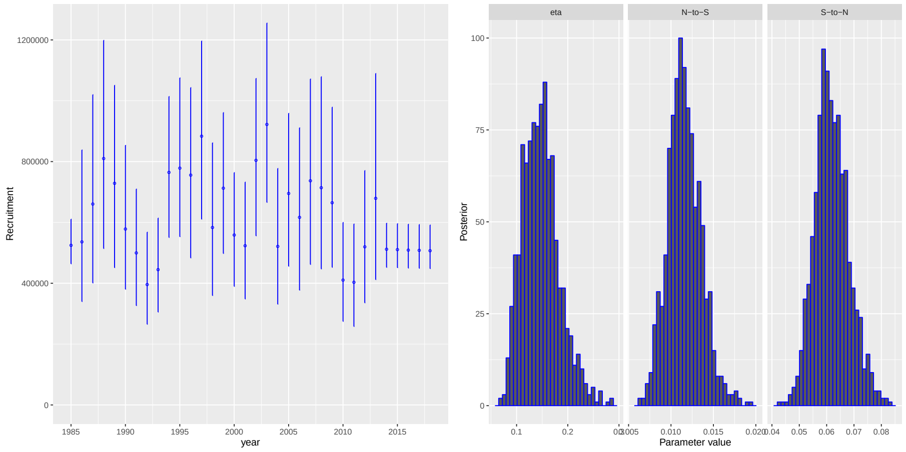

Figure 5.9: *Posterior median and 95% credible intervals for total recruitment (left), and the marginal posteriors for the three spatial parameters:* η_1, $\Phi_{1,2}$, *and* $\Phi_{2,1}$ *(right).*

the metric chosen. In terms of current female spawning biomass, clearly the model estimates more (almost six times more) in the North than in the South. If it is total numbers, there are in fact 50% *more* animals estimated to be in the Southern region - more recruits go here initially and these younger age-classes dominate the numbers. If our metric is, say, exploitable abundance currently accessible by the longline fleets then the estimated abundance in the North is around twice that in the South. If that metric is exploitable **biomass** there is around three times as much in the North, relative to the South. The point being the estimate of around 6–7 times predicted in the female SSB metric doesn't really reflect the estimated ratio for the more fishery or abundance focussed metrics.

5.4 Key sensitivity runs

Given this is the first time this assessment has been undertaken, we do not have a bridging-type analysis to show. We also don't - this time around at least - replicate a number of the year-on-year sensitivity tests done for the SS assessment (alternate values of M and steepness, dome-shaped selectivity for the long-line fleets). The reason is that the effect those tests have on the SS assessment will be qualitatively (and often quantitatively) very similar to the effect they have on the revised assessment.

Instead we focus on two key sensitivity tests:

1. Estimating growth within the assessment itself
2. Using the estimates of tag shedding rates instead of the previous assumption of effectively zero tag loss over time

For the growth estimation sensitivity test, we actually don't try and estimate all the growth parameters - only k, l_1, and σ_l. The reason for this is that the assessment in previous years has shown a propensity to try and estimate *very* high L_∞ parameters, or very low k parameters, when given full freedom. So we fix the l_2 parameters at the values estimated in [8], and estimate the remaining parameters (while still assuming that $a_2 = 20$). For l_1 the assessment estimates (and approximate SDs) are 0.52 (0.004) and 0.51m (0.005) for females and males, respectively; for k they are 0.021 (0.003) and 0.08 (0.004); and for σ_l they are 0.14 (0.02) and 0.13 (0.02). Estimates of unfished female spawning biomass are higher (*ca.* 3,500t) but current estimates of depletion are *slightly* lower (0.66 with a CV of 0.1 as before). For the growth estimates, basically everything is the same as estimated in [8] apart from female k - estimated to be a little less than half of the reference case estimates. This is basically the previous effect (pushing for a low k-high L_∞ female combination) already outlined. Note, we fixed l_2 **not** L_∞ - that would require essentially defining $a_2 = \infty$ - so we did not further explore nuanced growth estimation within the assessment. What is clear is that, as with the SS assessment, it is the other data (the tags in fact) pushing this effect **not** the age-length data. These data give well-estimated parameters that do not have this effect [8] when considered in isolation. What is important, for the assessment, is that we have consistent estimates of the distribution of size-at-age for the exploited age-classes (which it appears we do). Where it does become a problem though is for estimating SSB_0 - the effect is clear here in that it has to push it up given the slower growth rates of females when growth is estimated. All the other key parameters (selectivity, recruitment, and migration) are essentially the same as for the fixed growth reference case.

For the tag shedding sensitivity test, we assumed what is effectively the worst-case scenario: where the tag shedding is defined as in [9] and this defines π_t^{tag}; as a result, we are basically then at the expected lower-bound of tag retention (for the purposes of detection post-capture). The estimate of current depletion is 0.69 for this case - as opposed to 0.7 for the reference case - and with basically the same CV. Any level of tag shedding will act so as to push down the estimates of abundance from the tagging data. What we see here is that what we might call a worst-case scenario makes the depletion just over 1% lower than the reference case.

6 Discussion

In this paper we detail a proposed new assessment model for the Patagonian tootfish fishery around Macquarie Island. The underlying population model is basically the same as used in the current Stock Synthesis assessment [2] - spatially, sexually and age-based with underlying size structure for all the key variables. The model uses the same data as the current assessment: catch biomass, size frequency data, length-conditional age data, and the tagging data. The size data (sexually aggregated as before) are now modelled using the flexible Dirichlet-multinomial likelihood with the over-dispersion factor now estimated (not tuned). The age-given-length data, by fishery and sex, are modelled using a multinomial distribution as before. The major difference between the two models is how they model the tagging data.

In the current SS assessment model, the tagging data are grouped into a number of representative "age" classes; they are then modelled in a two-stage likelihood for first total tag recaptures, then spatial distribution of recaptures second. In the proposed assessment model we develop a sex and length structured spatial Brownie [6] tagging model that contains the abundance and migratory information in one likelihood function. The key covariates of release are sex (assumed unknown), size and spatial region; the key recapture covariates are year of recapture and spatial region. The Dirichlet-multinomial distribution is used for the likelihood with the tagging over-dispersion factor also estimated (as it is for the size data for each fishery).

Another key difference is how the data weighting is done within the revised assessment. Instead of a tuning-type approach we use a more flexible likelihood function (for the length and tagging data) - specifically the Dirichlet-multinomial distribution - where the "weighting" is done via an estimated parameter (the over-dispersion factor), not manual tuning of sample sizes or control parameters. There are over-dispersion factors for each fishery's length data likelihood, and for the tagging data likelihood. For the age-length data we assumed a multinomial likelihood (as in the current assessment) and, in the post-fitting diagnostic analyses, we explored the appropriateness of the assumed likelihood for these data.

The estimation framework utilised is primarily the Template Model Builder (TMB) R-based package [3], with the MCMC analyses done in the extension package `tmbstan` [4]. Using this statistical estimation package both significantly extends the capabilities of ADMB (which underpins Stock Synthesis) in terms of random-effect models and distributional possibilities, but also drastically reduces estimation time and MCMC runtime. For the revised TMB-based assessment, MCMC runs (required for RBC scenarios) take just over 1 hour to achieve a convergent Markov chain of 1,000 iterations. For the current SS model this can take several days. One of the main aims of replacing the current assessment was to make the most of the near 20 years worth of software and statistical development that has occurred since ADMBs features were established and embedded within Stock Synthesis. It is now possible to do far more, in terms of model flexibility and statistical development, and do it more efficiently as this document outlines.

The results are, in terms of the key management variable depletion, essentially the same as the current updated SS assessment [2]: around 0.7 relative to the current unfished state with an 95% credible interval of 0.65–0.76. They do differ in the absolute, with the revised assessment's estimate of female SSB_0 of around 2,500t being lower than the level the current assessment has varied around. That being said, we do use a different growth relationship (for both females and males) that is fixed using the analysis in [8] - any changes in the mean length-at-age will alter the absolute estimates, even when the relative levels stay the same. The other clear difference is spatial recruitment fraction: the revised assessment estimates more recruitment (just over 80%) going to the Southern region than the current assessment does (around 55%). To some degree this is offset by marginally higher rates of South-to-North migration as they both still estimate far more female mature biomass in the North, relative to the South. This has a knock on effect to recruitment trends also, with the current SS assessment still showing a smoother lower-then-higher dynamic, relative to the more variable up-and-down recruitment estimated in the revised model.

In summary, the revised assessment still maintains a significant portion of the structure - both in terms of population model and likelihood functions - already well established in the current SS assessment model. We have altered the treatment of the length frequency data to effectively remove the need for tuning the effective sample sizes and directly estimate over-dispersion factors. The age-given-length data are modelled in the same way, and diagnostic analyses confirmed that the assumed multinomial likelihood was clearly able to explain the data and the variability therein. The tagging data model and associated likelihood was the largest change: sex structured (at release), length-based in terms of dynamics and a fully spatial Brownie model (no two-stage likelihood in terms of total recaptures and spatial recapture distribution). The weighting for these data is, like the length frequency data, done statistically via the estimation of a tagging over-dispersion factor, as opposed to fixing this as before. That being said, the estimate from this revised model was actually close to the assumed level in [2]. The tag data clearly represent a consistent and informative data set irrespective of the analysis method. The move to the spatial Brownie model appears to give more accurate estimates of spatial recruitment and movement parameters as we hoped it would [1], and we feel it more accurately reflects the realities of the tagging program (and the dynamics of the tags), so was worth the development.

7 Acknowledgements

Many thanks to AFMA, CSIRO and industry for funding an assistance, and to Tim Lamb for all his help in getting the data together.

References

[1] Hillary, R. M. and Day, J. (2017) Impact of spatial tagging rates for key estimates coming from the Macquarie Island toothfish assessment. *SARAG 56*.

[2] Day, J. and Hillary, R. M. (2019) Stock Assessment of the Macquarie Island fishery for Patagonian toothfish (*Dissostichus eleginoides*) using data up to and including August 2018. *SARAG 59*.

[3] Kristensen, K. *et al.* (2016) TMB: Automatic Differentiation and Laplace Approximation. *J. Stat. Soft.* **70**(5): 1–21.

[4] Monnahan, C. C. and Kristensen, K. (2018) No-U-turn sampling for fast Bayesian inference in ADMB and TMB: Introducing the adnuts and tmbstan R packages. *PLoS ONE*. **13**(5): e0197964.

[5] Hillary, R. M. (2011) A new method for estimating growth transition matrices. *Biometrics*. **67**: 76–85.

[6] Hillary, R. M. and Eveson, J. P. (2015) Length-based Brownie mark-recapture models: Derivation and application to Indian Ocean skipjack tuna. *Fish. Res.* **163**: 141–151.

[7] Hillary, R. M., Day, J. and Haddon, M. (2015) Age-at-length or length-at-age for modelling the growth of Macquarie Island toothfish? *SARAG, Feb 2015*.

[8] Hillary, R. M. (2019) Revised growth parameters for Macquarie Island toothfish using data up to and including 2017. *SARAG 59*.

[9] Hillary, R. M. (2019) Estimates of tag shedding in the Macquarie Island toothfish tagging program. *SARAG 59*.

CONTACT US

t 1300 363 400
+61 3 9545 2176
e csiroenquiries@csiro.au
w www.csiro.au