

Is the integrated assessment the best place for modelling the growth of Macquarie Island toothfish?

R. M. Hillary & J. Day
CSIRO
Oceans & Atmosphere,
Castray Esplanade,
Hobart,
TAS 7000,
Australia.

Abstract

Over the last few years the case for estimating growth inside of integrated stock assessment models, using conditional age-at-length data, has been increasingly expounded. The first argument is that, quite rightly, this type of approach can solve the issues around bias in traditional length-at-age data that arise from length-specific selectivity. The second is that it can propagate the uncertainty in the growth parameters to the variables of interest in the assessment itself. The stock assessment of the Macquarie Island Patagonian toothfish (*Dissostichus eleginoides*) fishery used to estimate some growth parameters internally, but had consistent issues with unrealistic parameter estimates that were inconsistent with estimates from the growth data only, and driven by other data sets. Using the extensive, sexually disaggregated multi-year ageing data for this fishery we develop a novel method to analyse these data in the length-conditional framework, thereby by-passing the necessity of having to estimate growth within the assessment if it is proving problematic. This work highlights one of the main risks with modelling growth inside of the assessment: the analyst has no control over which data informs the parameter of interest, and potentially spurious trends in unexpected data sets may combine with structural issues in the model to dominate clearly informative data sets, resulting in anomalous parameter estimates. For the example in question we suggest solutions that balance the issue of biased estimates with the second - and still very important - issue of variance propagation through the key assessment variables.

1 Introduction

Growth is a central driver of the individual (and population-level) dynamics of fish (and the stock they belong to) as well as those of the fishery itself. Given its importance, modelling growth has been a mainstay of fisheries science and stock assessment from the beginning. Many sources of data hold information on growth.:

Direct length-at-age, where age is determined from a hard part such as an otolith or, for elasmobranchs, the vertebrae. Mark-recapture data, where the growth increment between release and recapture is the information source, but there is no information on actual age. Length frequency data: in a closed population (within one recruitment cycle or across several) the progression of modes in length frequencies holds information on the growth rate of the population. As with mark-recapture data there is little direct information on age, save the number of age classes present if they are clear enough to be estimated well using mixture modelling.

The focus of this paper is for cases where the primary growth data are length and age measurements and, more specifically, modelling these data both inside and outside of an integrated stock assessment (Methot and Wetzel, 2013). For Macquarie Island toothfish, there is an extensive amount of sexually disaggregated length-at-age measurements from otoliths so there is ample information available for exploring appropriate growth models. The original stock assessment model was a modified version of the Stock Synthesis (Methot and Wetzel, 2013) modelling framework, and is sex-specific and spatially disaggregated, with fishery length-composition, conditional age-at-length, age-structured mark-recapture, and catch biomass data (Day *et al.*, 2015). The assessment is spatially structured, with two regions where age-independent movement is estimated within the assessment and the iterative reweighting approach outlined in (Francis, 2011) was used to weight the various data sets. The structure of the assessment has morphed over time as more and more data have become available, and early issues around the estimation of all the parameters of the female growth curve (Fay , 2010) led to asymptotic length becoming a fixed parameter in the assessment henceforth. The first indication of a potential problem appearing with the males was observed in 2014 (Day *et al.*, 2014), where the growth rate and maximum length parameters were suddenly estimated to be much smaller and larger, respectively, than both previous estimates and their female counterparts, at apparent odds with the age data that clearly showed males being consistently shorter than females (Day *et al.*, 2014). This problem worsened in 2015 with nonsensical estimates of both growth rate and asymptotic length and a notable deterioration in the actual fitting to the age data over the sample range (Day *et al.*, 2015). The main question was this: are the differences in the estimates both inside and outside the assessment driven by different assumptions and probability model structures, or something being driven

by the assessment itself and the other data therein? The solution that was explored was to develop a robust method for estimating growth externally, and design a custom-designed stock assessment model that could account for, if required, the parametric uncertainty of the growth relationship inside the stock assessment.

Estimating growth via paired length-at-age measurements has been the most common approach, historically speaking, presumably because of its relative ease of implementation. It assumes that length is random for a given age; age is known without error; and each length observation is independent of the others, given the model parameters. In almost all cases, and certainly for the specific example of Macquarie Island toothfish, we are fairly certain the selectivity is driven by length (driven by depth-specific ontogenic shifts as animals grow), we know that the age estimates are subject to error (Candy *et al.*, 2012), and the length measurements are almost certainly not independent.

The age-given-length framework (Morton & Bravington, 2008; Piner *et al.*, 2015) has consolidated its position as arguably the most robust method that can deal with the realities of paired age-length data (ageing error, size selectivity). In this approach the primary data source is the relative frequency of age *conditional* on length, not length-at-age. This assumes that age is sampled randomly within the specific length bin of interest. In the integrated stock assessment framework (Methot and Wetzel, 2013) the sample bias, attributed to length-specific selectivity, is dealt with directly as the selectivity is estimated along with the growth parameters via the inclusion of the length frequency data. Additionally, the ageing error can also be dealt with efficiently in this framework. The disadvantages of this approach are that it can be harder to visualise the data, relative to the length-at-age framework, and it is computationally more demanding. Also, in the stock assessment we are making the assumption that the age data are randomly sampled from the catch, and that the overall fleet selectivity can be used to construct the prior age distribution. This rarely happens in practice, and is not the case for the data in this paper where a wider size range is sought to obtain more informative data on longer older animals. A final disadvantage to implementing this approach in the integrated assessment framework, relevant to this paper, is parameter aliasing and correlation with other (misspecified) life-history parameters, as well as potentially spurious information from other data sources like abundance indices.

In this paper we explore a novel, highly applicable age-at-length approach that can be undertaken outside of the assessment framework. The point of the external age-at-length model is to serve as a bridge to the full assessment estimates. Length-at-age estimates have often differed from the age-at-length estimates in the assessment model - particularly for the males (Day *et al.*, 2014, 2015). By having an external estimation approach for growth that closely mimics how the data are modelled within the assessment, we are in then in a stronger

position to judge if it is another data set in the assessment causing the problem in relation to the growth estimates within the assessment. We also outline the approach taken to balancing unbiased parameter estimation with variance propagation in the stock assessment.

2 Data

For the Macquarie Island toothfish fishery the ageing of fish began at the start of the fishery in 1996, running through to the most recent data from 2019. Data are available by sex and are read by two experienced readers to get a robust estimate of age (and ageing error). Figure 1 outlines the overall size-at-age summary for both sexes, and Figure 2 details the samples-by-length for each year, for each sex. There are a total of 5,958 age estimates (39% male, 61% female). Overall sample size and average size of the aged fish has varied over time and both have increased as the fishery transitioned from a trawl dominated to a longline only fishery from around 2006–2008. Females, on average, grow at a slower rate but reach a larger expected asymptotic length than the males. An ageing error matrix was derived for a different stock of Patagonian toothfish, for the same otolith readers, in Candy *et al.* (2012) and we employ it in the analyses that follow.

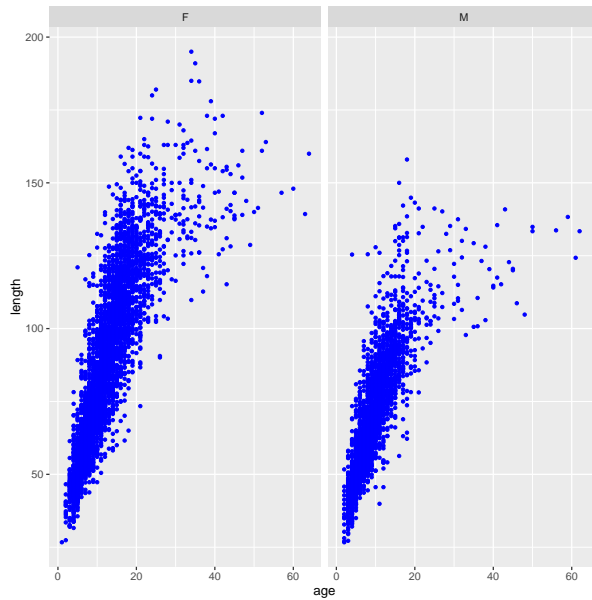


Figure 1: *Overall size-at-age for females (left) and males (right).*

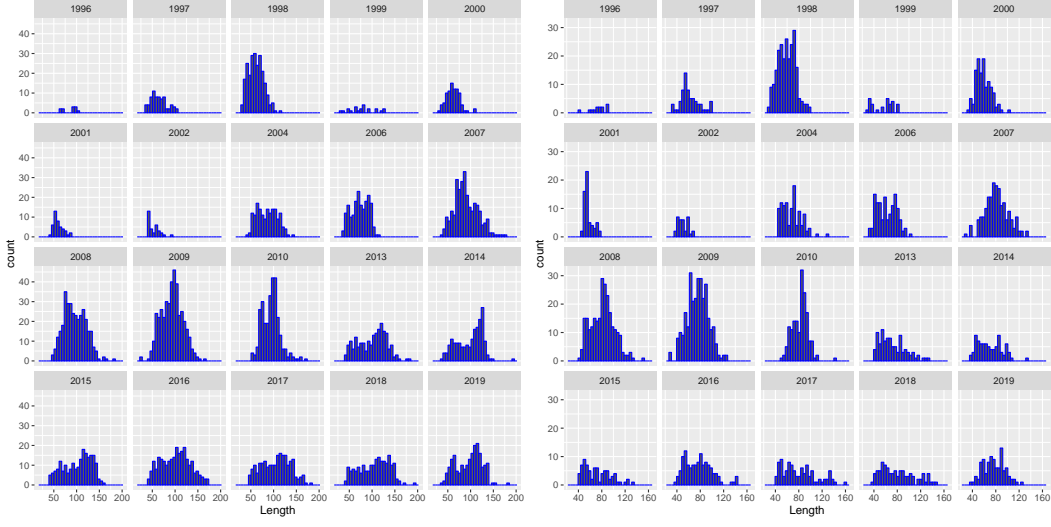


Figure 2: *Size frequency samples over time for females (left) and males (right).*

3 Methods

The mean length-at-age is defined via the Schnute parameterisation of the von Bertalanffy growth curve:

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

where $l_1(a_1)$ and $l_2(a_2)$ are the lengths at reference ages a_1 and a_2 ($a_2 > a_1$), and k is the growth rate.

To generate the distribution of length-at-age we assume a log-normal 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:

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

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

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

and the prior age distribution is defined as follows:

$$\pi_{a|y,s} \propto \text{LogN}(\mu_{y,s}, \sigma_{y,s}^2).$$

The above formulation works well for a largely unimodal distribution of length samples for any given year; if the empirical distribution is more complicated a more flexible model for the prior age distribution (e.g. smoothing splines) may be required. 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

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

For the length frequency data of the aged fish (again to be understood as being different to the length frequency data per fishery used in the assessment) we assume a Dirichlet-multinomial distribution:

$$\Lambda_{y,s}^l = \frac{(n_{y,s}!) \Gamma(\omega_{y,s})}{\Gamma(n_{y,s} + \omega_{y,s})} \prod_l \frac{\Gamma(n_{y,l,s} + \omega_{y,s} \pi_{l|y,s})}{n_{y,l,s}! \Gamma(\omega_{y,s} \pi_{l|y,s})}$$

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

$$\omega_{y,s} = \frac{n_{y,s} - \varphi_s}{\varphi_s - 1},$$

and $\varphi_s > 1$ 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_s is an estimable parameter (as opposed to tuning to get the right value of $n_{y,s}$).

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:

$$\Lambda_{y,l,s}^{a|l} = \prod_a (\pi_{a|y,l,s})^{n_{y,a,l,s}}$$

For the Schnute model reference ages we assume $a_1 = 5$ and $a_2 = 20$ as assumed in the revised assessment model. Length bins are in 10cm blocks from 20cm at the minimum to a maximum that ensures the largest length bin includes the largest animal observed in the data (for each sex). The parameters estimated in the full model (using both length and age-given-length data) are:

- Mean length-at-age parameters: l_1 , l_2 , and k
- Standard deviation in mean length-at-age: σ_l
- Prior mean μ_y and standard deviation σ_y of the prior age distribution

- Over-dispersion factor in the length data φ

The overall (sex-specific) joint log-likelihood is defined as follows:

$$\ln \Lambda_s^{\text{tot}} = \sum_y \left(\ln \Lambda_{y,s}^l + \sum_l \ln \Lambda_{y,l,s}^{a|l} \right).$$

We use the TMB package Kristensen *et al.* (2016) to find the parameters which maximise the joint likelihood of the length and age-given-length data, as well as give us approximate standard errors for each of the parameters and process variables.

4 Results

Fits to the female and male size data can be seen in Figure 3, and the summary of the mean age-given-length can be found in Figure 4. Table 1 summarises the key parameter estimates. 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.

| Variable | k | l_1 | l_2 | L_∞ | t_0 | σ_l | φ |
|----------|---------------|---------------|--------------|-------------|--------------|---------------|------------|
| Female | 0.055 (0.003) | 0.494 (0.003) | 1.16 (0.004) | 1.68 (0.03) | -1.3 (0.15) | 0.15 (0.012) | 1.05* (NA) |
| Male | 0.067 (0.003) | 0.488 (0.002) | 1.02 (0.007) | 1.33 (0.03) | -1.86 (0.18) | 0.144 (0.016) | 1.05* (NA) |

Table 1: *Maximum likelihood estimates (and approximate standard errors in brackets) of key estimated parameters and process variables for each sex. The * for each of the over-dispersion coefficients indicate that the estimates hit the lower bound and, as such, we cannot produce sensible standard errors.*

When summarising the fits the length data, for both sexes the fits are generally fairly good, with no apparent systematic issues over time. For both sexes, the estimates of the over-dispersion factor were at the lower bound of 1.05 (we cannot have $\varphi = 1$ so 1.05 is a sensible lower bound), strongly indicating an apparent lack of over-dispersion in the size data *of aged animals* and, hence, the logical corollary that a multinomial distribution would in fact be as appropriate. Looking at the fits to the mean age-given-length data, we see good fits for both sexes and across years. Importantly, practically all the estimates sit within the approximate 95% CI. Also, analyses of the standardised residuals for these data show that the variance clusters around about 0.8-1.1 for both sexes - specifically they do not appear consistently over 1 and so the multinomial assumption also seems fine in this case.

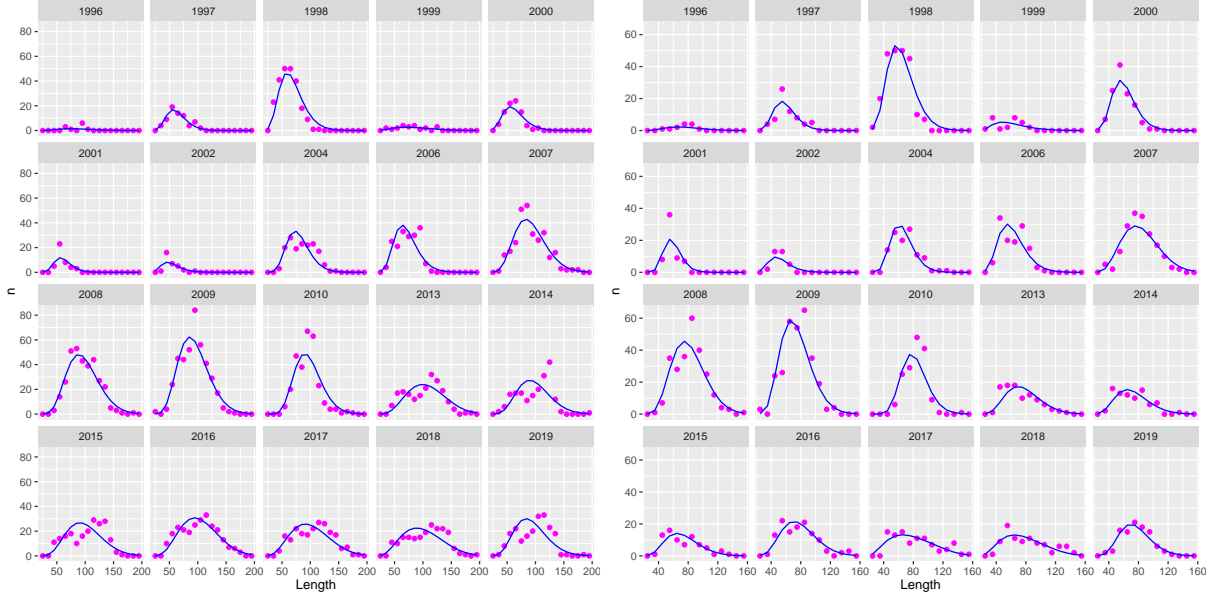


Figure 3: *Observed (magenta circles) and predicted (blue lines) length frequency summary for the female (left) and male (right) aged animals.*

The stock assessment is now a custom-designed Bayesian sex-age-length-spatially structured assessment written in Template Model Builder (TMB) Hillary & Day (2021). Maximum posterior density estimation is used to configure the various model structures and Markov chain Monte Carlo (MCMC) is used (via the `tmbstan` package in R) to extract a suitable sample from the joint posterior distribution of the parameters. The current Harvest Control Rule (HCR) for this fishery is what is known as the CCAMLR rule (REF) so projections are done using the MCMC output to estimate the constant catch that meets the requirements of the CCAMLR HCR. To propagate parametric uncertainty in the growth parameters through the assessment variables we explored a variation of the Bayesian Model Averaging (BMA) approach: a representative discrete grid (e.g. balanced factorial design) of growth parameters are selected and the assessment model is fitted to each grid element (full MCMC but with an adaptive sample size per growth combination). Assuming uninformative priors for the growth parameters we use likelihood weighting of the growth scenarios as a proxy for the model posterior of each growth “model”. Ideas from BMA (see Appendix for details of how to do this very efficiently) are then applied to get a posterior sample from the joint growth and assessment parameter posterior and these are used in the HCR calculations. It is this two-stage process that was proposed to solve the issue with the Macquarie Island toothfish assessment: sensible growth estimates could not be obtained via internal estimation, but propagation of growth uncertainty in the assessment was considered an issue of generic

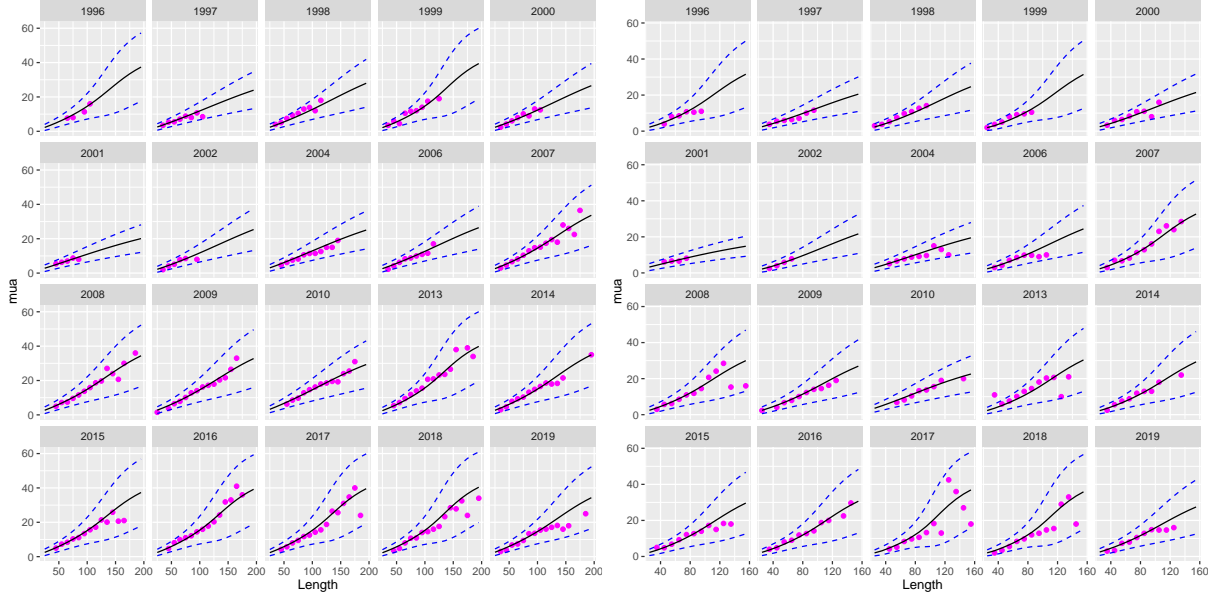


Figure 4: *Observed (magenta circles) and predicted median (full blue line) and 95% CI (dotted blue line) mean age-given-length summary for the female (left) and male (right) aged animals.*

importance. In practical terms this is not actually done for the current assessment Hillary & Day (2021) for the following reasons: the estimates of all the growth parameters are very precise (see Table 1) and adding in the additional complexity and computational burden of the BMA step makes no meaningful difference to either the stock status or the management advice outcomes. However, in other examples where the growth data are not as numerous and informative it might be very important to be able to propagate the uncertainty in the mean length-at-age, not just the uncertainty around the mean length-at-age. In those cases this approach can be a workable alternative that does not require more of an assessment than the ability to apply some kind of MCMC algorithm to obtain posterior samples for the parameters conditional on a suite of growth parameter scenarios.

5 Discussion

The length-conditional modelling of age data Morton & Bravington (2008); Piner *et al.* (2015) is generally accepted as the best way to deal with dependence between size classes (via size-based selectivity) and ageing error. In an integrated stock assessment estimating growth internally is accepted to be best practice to ensure variance propagation in the key assessment variables, and length-conditional age data are likely to be the most informative (especially in relation to age-dependent growth parameters) given the issues with these data can be dealt

with in the assessment Methot and Wetzel (2013). One issue with internal estimation is when the estimates of the growth parameters from the assessment do not agree with the estimates one would obtain by correctly estimating the parameters externally using the age data, which one would assume to be the most informative and reliable.

This particular problem arose in the integrated assessment of the stock of Patagonian toothfish at Macquarie Island. The original stock assessment was a modified version of Stock Synthesis Methot and Wetzel (2013), and in successive assessments Fay (2010); Day *et al.* (2014, 2015) issues arose with female asymptotic length and then a similar issue arose with the same parameter for the males. The estimates were being driven by the mark-recapture data, and did not agree with external estimates of the growth parameters. This was a problem of two parts: (i) how to obtain the most robust estimates of growth to use in the assessment; and (ii) what, if any, structural issues in the model are interacting with the tagging data. In this paper we outline a novel approach to solving the first problem: a length-conditional age approach to estimating growth using length and age-given-length data to jointly estimate the parameters of the length-at-age distribution and the prior age distribution by year in the aged length data.

The method is simple to apply to annual age data where the size distribution of the aged animals changes markedly from year to year. It is especially useful if the size distribution of the aged animals is different from the size distribution of any particular fishery in the integrated assessment. For Macquarie Island toothfish we obtained estimates of slower growth but consistently higher size-at-age for the females relative to the males, but very similar variance in the mean length-at-age. The fits to both the length and age-given-length data are good and with no apparent over-dispersion in either data set.

The method is able to essentially replicate the key features of how we model these data within an assessment - account for the prior age distribution given size distribution, include ageing error. The obvious drawback is that, by estimating growth externally, we do not get automatic variance propagation of length-at-age uncertainty to the key assessment variables. When using MCMC in the assessment model, we outline an augmented Bayesian Model Averaging approach to efficiently constructing an MCMC algorithm to obtain samples from an approximation to the joint posterior of the assessment and growth parameters - more importantly this includes the key assessment variables such as biomass depletion and recruitment. This allows us to account for the variability in the size-at-age distribution jointly with the assessment variables, solving the variance propagation problem.

Estimating growth within an integrated stock assessment is rightly considered best practice. However, there will be cases - such as the case study in this paper - where this does not produce sensible growth estimates. One might expect for example age data to be the dominant

information source on growth, but the analyst cannot control which data set will exercise the most influence over the estimates. In cases such as this we need a robust method that can externally estimate growth and propagate the uncertainty in these external estimates through the key assessment variables and this paper suggests a method for doing this that only requires that the assessment method has some MCMC capability.

References

- Candy, S.G., Nowara, G. B., Welsford, D. C., and McKinlay, J. P. 2012. Estimating an ageing error matrix for Patagonian toothfish (*Dissostichus eleginoides*) otoliths using between-reader integer errors, readability scores, and continuation ratio models. *Fish. Res.* **115**: 14–23
- Day, J., Haddon, M., and Hillary, R.M. 2014. Stock Assessment of the Macquarie Island fishery for Patagonian toothfish (*Dissostichus eleginoides*) using data up to and including August 2014. SARAG.
- Day, J., Haddon, M., and Hillary, R.M. 2015. Stock Assessment of the Macquarie Island fishery for Patagonian toothfish (*Dissostichus eleginoides*) using data up to and including August 2015. SARAG.
- Fay, G. 2011. Stock assessment of the Macquarie Island fishery for Patagonian toothfish (*Dissostichus eleginoides*) using data up to and including June 2010.
- Francis, R.I.C.C. 2011. Data weighting in stock assessment models. *Can. J. Fish. Aquat. Sci.* **68**: 1124–1138.
- Hillary, R. M., Wayte, S., Day, J., and Haddon, M. 2014. Appropriate growth models for Macquarie Island toothfish. SARAG.
- Hillary, R. M., and Day, J. 2021. Integrated stock assessment for Macquarie Island toothfish using data upto and including 2020. SARAG.
- Kristensen, K. *et al.* (2016) TMB: Automatic Differentiation and Laplace Approximation. *J. Stat. Soft.* **70**(5): 1–21.
- 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.
- Morton, R., and Bravington, M. V. 2008. Comparison of methods for estimating age composition with application to Southern Bluefin Tuna (*Thunnus maccoyii*). *Fish. Res.* **93**: 22–28.
- Piner, K. R., Lee, H.-H., and Maunder, M. N. 2016. Evaluation of using random-at-length observations and an equilibrium approximation of the population age structure in fitting the von Bertalanffy growth function. *Fish. Res.* **180**: 128–137.

Appendix

To propagate the overall uncertainty in the length-at-age distribution through the stock assessment variables we proposed a Bayesian Model Averaging (BMA) approach (REF). To demonstrate how this works, we define the data and parameters of the growth relationship as D_g and $\boldsymbol{\vartheta}$, and those of the assessment model as D and $\boldsymbol{\theta}$. The two-step process we propose factorises the joint distribution of the growth and assessment parameters as follows:

$$\pi(\boldsymbol{\vartheta}, \boldsymbol{\theta} \mid D_g, D) = \pi(\boldsymbol{\theta} \mid D, \boldsymbol{\vartheta}) \pi(\boldsymbol{\vartheta} \mid D_g)$$

that is the assessment parameters are conditional on the growth parameters. To efficiently undertake this process, given we are using the `tmbstan` package to sample from the assessment parameter's posterior distribution, we construct an approximating discrete distribution for the growth parameters, $\boldsymbol{\vartheta}$. We use a representative suite of growth parameter combinations to define a discrete set of values for the growth parameters for each sex, $\boldsymbol{\vartheta}_{i,s}$ and $i \in 1, \dots, N$. Assuming essentially uniform quasi-uninformative priors for the growth parameters we use the full likelihood of that particular growth parameter vector so we have that

$$\pi(\boldsymbol{\vartheta}_{i,s} \mid D_g) \propto \Lambda_s^{\text{tot}}(D_g \mid \boldsymbol{\vartheta}_{i,s})$$

At this point, to make use of BMA techniques, we interpret the discrete posterior for each of the growth parameters-by-sex to be an individual model, M_j and $j = 1, \dots, \tilde{N}$, the probability of which is defined by the product of the discrete posterior of the male and female growth parameter vectors for that particular combination defined by the index j . For a naive combination of male and female growth parameter combinations $\tilde{N} \equiv N^2$ but we can apply efficient factorial design principles here also to reduce the number of overall discrete parameter combinations so that $\tilde{N} < N^2$.

From an inference point of view, we are interested in obtaining samples of the key assessment variables (SSB, recruitment *etc.*) from the *joint* posterior distribution of the assessment *and* growth parameters. The most straightforward but inefficient way to do this is define a fixed assessment parameter sample size n_1 , run the hybrid MCMC algorithm for the stock assessment parameters for each element of the growth parameter grid, j , then sample n_2 iterations from the discrete distribution of growth parameter combinations. For each one of these growth parameter combinations, we sample without replacement from the assessment parameter samples conditional on that growth parameter combination. Once complete this would give us n_2 samples from the joint assessment and growth parameter posterior distribution.

This would be inefficient because we are not going to need n_1 samples from the conditional posterior of the assessment variables for every growth parameter combination. If n_2 is the

overall sample size, we first sample n_2 growth parameter combinations; the relative frequency with which each growth parameter combination appears in the overall sample then tells us how many samples we need to take from each assessment parameter growth-conditional posterior. We can optimise the burn-in and thinning factor required for the hybrid MCMC runs quickly before running the main full sampling algorithm. All of this means that, while the most naive approach to the algorithm would be a problem that basically scaled computationally with $n_2 * N^2$ (which could be very large), a more efficient approach to the overall sampling algorithm can be made much much closer to n_2 in terms of the overall computational burden. Additionally, this makes the whole process simple to fit into a parallel computing paradigm if one considers each MCMC chain, corresponding to a particular growth parameter combination, something that can be sent to an individual core on a multi-core laptop or cloud resource. This makes it relatively straightforward to construct a sampling algorithm that is very close to the computational requirements for sampling from an assessment parameter posterior conditional on a single growth parameter combination.