OCEANS AND ATMOSPHERE
www.csiro.au

Reconsideration of OM structure and new data sources for 2017 reconditioning

Rich Hillary, Ann Preece, Campbell Davies

CCSBT-OMMP/1609/4

Prepared for the Seventh Operating Model and Management Procedure technical Meeting held in Kaohsiung, Taiwan 2th-4th of September 2016

CSIRO
Oceans & Atmosphere
Castray Esplanade, Battery Point, Hobart TAS 7000, Australia
Telephone : +61 3 6232 5222
Fax : +61 3 6232 5000

Copyright and disclaimer

© 2016 CSIRO To the extent permitted by law, all rights are reserved and no part of this publication covered by copyright may be reproduced or copied in any form or by any means except with the written permission of CSIRO.

Important disclaimer

CSIRO advises that the information contained in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law, CSIRO (including its employees and consultants) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

Contents

Abstract

This paper details potential structural changes to the current CCSBT OM required to deal with new data sources and scenarios for the 2017 reconditioning and for the next phase of MP testing. In terms of new data scenarios we will have non-member and attributable catch estimates; in terms of new data sources we will have both gene-tagging and half-sibling close-kin data. We also outline some potential changes to current projection dynamics, specifically around fixed future selectivity and growth relationships.

1 Background

The SBT operating models (OMs) will be reconditioned in 2017 for an updated assessment of stock status and to allow for testing (management strategy evaluation) of a new Management Procedure (MP). The 2016 OMMP and ESC therefore needs to consider structural changes that might be required, or new data sources that need to be incorporated.

In terms of new data sources and scenarios we have (i) non-member catch scenario estimates, inferred from non-member effort and member catch and effort properties (ii) member's attributable catch, which may be a mix of both scenarios and actual direct estimates (iii) gene-tagging data, and (iv) the next generation of close-kin data relating to half-sibling pairs. The first two data sources will not necessarily require technical adjustments to the current OM. They can be either assigned to an existing fleet (in terms of non-member catch), or a new fleet can be defined (and selectivity parameterisation) to deal with attributable catch that is not similar enough to existing fleets to be incorporated into them.

For the gene tagging and half-sibling close-kin data new likelihood functions are required for the OM. In the design study for a gene-tagging program, it was recommended that a Petersen model be used to calculate absolute abundance of a cohort of juveniles. This choice was considered more cost-effective compared with tagging multiple cohorts [1] which could provide mortality estimates via a Brownie estimation model, as was done in previous conventional tagging programs. A suggestion for the likelihood function to use in the OM was outlined and tested using the 1990s and 2000s tagging data. Here we also consider a flexible beta-binomial likelihood that has over-dispersion as a fundamental parameter in the likelihood itself (like the current Brownie likelihood for the 1990s tagging data). For the half-sibling data we first outline the key factors that influence the probability of two juveniles being half-siblings. A beta-binomial likelihood is also suggested for the half-sibling data.

A key feature of the current OM projection model is that both selectivity and growth are fixed at the most recent values, even though both change quite noticeably over time. It is quite likely that the factors that potentially caused both to change historically (i.e. stock abundance) may cause similar changes in the future, if the stock recovers under the next MP-driven rebuilding strategy. We outline the key population and monitoring data that these changes can effect, how they could alter our perception of the rebuilding of the stock, and how to incorporate such changes in the OM projection model.

2 Material & Methods

2.1 New data sources

2.1.1 Non-member catch

Paper CCSBT-ESC/1609/BGD 02 [2] details the estimates of non-member SBT catch, using two contrasting model approaches and assumptions relating to whether the effort data (and catchability) in question was "Taiwanese" or "Japanese" in nature. Irrespective of the relative weightings given to each modelling approach, or assumptions about how much of the effort relates to either of the assumed catchability scenarios, their inclusion should be fairly simple given the current OM structure. The LL_1 and LL_2 fleets represent the fleets for which Japanese and Taiwanese long-line catches belong to in the OM. If we are willing to assume a given amount of non-member effort can be assigned one of these catchability parameters, and the mean weight in those fleets are used to convert between weight and numbers in terms of catch [2], then assigning them the same selectivity function seems equally plausible. The estimate of non-member catch may then be included as additional historical catches in the relevant fleet (LL_1 or LL_2).

2.1.2 Attributable catch

Attributable catch for various members might be less straightforward, in terms of OM inclusion. If the attributable catch comes from a fleet that is demonstrably similar to that for which reported catch is from, then the same approach as taken for the non-member catches would be appropriate: assign it the same selectivity and add the catch to the original reported catches. Where this approach will not work is when we cannot assume fleet-similarity for the attributable catch, relative to historically reported catches.

Focussing on the Australian example: recreational catches are not likely to have the same selectivity as the surface fishery. In fact, they are probably more biased to longer/older animals. If we were to include them as catch biomass in the OM, and assign them the surface selectivity, we would over-estimate the fishing mortality associated with those catches; the opposite would happen if we included them in terms of catch in numbers. For any member, attributable catch that cannot be assigned to an existing OM fleet must then be given an accompanying selectivity, or come with associated length/age frequency information from which selectivity can be estimated.

2.1.3 Gene Tagging

The driving idea behind the gene tagging is to be able to estimate the absolute abundance of age 2 fish using the well-established Petersen estimator [1]. Recall the basic abundance estimator as defined in the design study:

$$\widehat{N} = \frac{TS}{R}, \tag{2.1}$$

where T are the number of initially genotyped and released fish, S is the number of fish genotyped in the catch sample, and R the number of recaptures. The CV of this estimate of N will be approximately $R^{-1/2}$. The design study explored both the Brownie and Petersen models, but focussed on the Petersen design for an abundance estimate for juveniles, which required tagging a single cohort each year. The issue of bias from incorrect ageing was also explored. The likelihood proposed and tested in the gene-tagging design study was a gaussian model which could be corrected for ageing errors and for which over-dispersion could be incorporated after a number of years of data had been collected. The recommendation from the design study was to tag age 2 fish and take samples of the catch of age 3 fish, one year later, to obtain an estimate of the age 2 abundance in the original year of tagging.

Figure 2.1 shows the overlap in probable length distribution for age 2-4 fish. As noted in the design study, using a restricted length class of fish for tagging and catch sampling, and collecting otoliths at the time of catch sampling will help to understand the probable age distribution of both the tagged and scanned fish. Restricting the age-range of fish tagged to 2 and 3, and the catch sampled age-range to 3 and 4 would be achievable given what we know about current length-at-age in the GAB.

Some minor changes to the original estimator, while retaining the original general idea, are needed to incorporate age at length information. Let T_y, S_y, and R_y be the year-specific tagging, catch sample and recapture numbers. In terms of the sample length distributions we will have one each for the tagged and catch sampled fish: $p^{t,s}_{y,l}$. What we really need are the distributions of age-given-length so as to be able to probabilistically assign fish to specific cohorts when it comes to abundance estimation. There are a number of possible ways this could work, but let's proceed with one where we assume to already know the length-at-age distribution: $\pi_{l|a}$ (for our current application time-invariant but easily extendible). We need a prior age distribution for the sampled fish, $\pi^{t,s}_{y,a}$, to get the length-conditional age distribution we

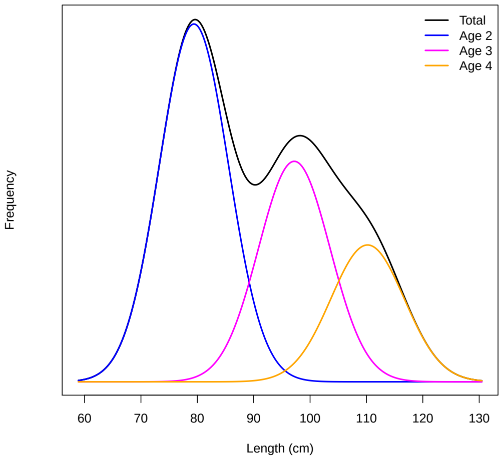

Figure 2.1: *Probably length distribution of ages 2, 3 and 4 year old fish in the GAB in summer given current mortality rates.*

really need:

$$\pi^{\bullet}_{y,a|l} = \frac{\pi_{l|a}\pi^{\bullet}_{y,a}}{\sum\limits_{a}\pi_{l|a}\pi^{\bullet}_{y,a}}, \tag{2.2}$$

and • denotes the specific sample i.e. tagged or catch sampled. For this set-up we will need to estimate $\pi^{\bullet}_{y,a}$ for each year and sampling stage. This can be done using the length distribution of the tagged and catch sampled fish, as the model-predicted length distribution will be $\sum\limits_{a}\pi_{l|a}\pi^{\bullet}_{y,a}$. We would then assume a multinomial likelihood for the length data and directly estimate the age prior, which is one of the approaches used to estimate the age composition of the surface fishery catch when length-at-age is assumed known. Age sampling of the catch sampled fish could make this a bit more involved, but the general idea is the same: obtain an understanding of the age-at-length distribution of all the fish in the study.

Now, instead of a simple single case as in the design study there would now be two plausible cases for the estimator, given the type of sampling regime: (i) where fish are sampled and recaptured the year after they were tagged; and (ii) where fish are sampled and recaptured two years after they were tagged (and this will only really be 2 year old releases). For the first case the modified estimator would be:

$$\widehat{N}_{y,a} = \frac{\left(\sum\limits_{l} T_{y,l}\pi^{t}_{y,a|l}\right)\left(\sum\limits_{l} S_{y+1,l}\pi^{s}_{y+1,a+1|l}\right)}{\sum\limits_{l} R_{y+1,l}\pi^{s}_{y+1,a+1|l}}, \tag{2.3}$$

whereas the second case would be

$$\widehat{N}_{y,a} = \frac{\left(\sum\limits_{l} T_{y,l}\pi^{t}_{y,a|l}\right)\left(\sum\limits_{l} S_{y+2,l}\pi^{s}_{y+2,a+2|l}\right)}{\sum\limits_{l} R_{y+2,l}\pi^{s}_{y+2,a+2|l}}. \tag{2.4}$$

In both cases, the inverse square-root of effective number of recaptures of that tagged age-class will be a good indicative estimate of the CV of the abundance estimate as before. A binomial likelihood is the most flexible option for the likelihood in the OM, where the probability of recapturing a fish released at age a in year y would be:

$$p_{y,a} = \frac{\sum_l T_{y,l}\pi^t_{y,a|l}}{N_{y,a}}, \tag{2.5}$$

and the associated sample size would be $\sum_l S_{y+r,l}\pi^t_{y+r,a+r|l}$, where, for the current sampling regime, $r = 1, 2$ depending on how many recapture events there are for that particular release event. The binomial is not only a natural choice for the base likelihood for the GT data, but can be easily extended to the beta-binomial model which can deal with over-dispersion (essentially process error). Some indicative over-dispersion scenarios were explored in the GT design study [1], and it was noted that it will take time to actually establish both the presence and - if true - extent of over-dispersion in the tagging data. As is the case with the 1990s tagging data, over-dispersion is calculated by a detailed analysis of the residual variance of the fitted recaptures [3]. A similar approach, requiring a number of years of release and recaptures, would be applicable for these data. The specifics of how to parameterise the beta-binomial distribution can be found in the Appendix.

2.1.4 Half-Sibling Pairs

The Parent-Offspring pair close-kin data (POPs) have already been incorporated into the CCSBT OM [3] and no proposed changes are suggested to the current likelihood for those data. The next generation sequencing methods that are replacing the previous micro-satellite approach used in the original CK data will permit the identification of half-sibling pairs (HSPs) among the juveniles [4]. While HSPs are clearly close-kin data, there are a number of important ways in which they are both different to - and potentially more informative than - POPs.

The actual identification of HSPs is much more involved than for POPs; POPs *have* to share at least one allele at a given locus on the genome, making their detection (relatively) more straightforward. We do not go into this more detailed procedure in this paper, and just assume that the identification of HSPs can be (and is) done. Interested readers way wish to read [4] for more details on this HSP identification procedure. For SBT, we will look for HSPs in the juveniles (already collected for the POP approach). So, we assume that we have juveniles i and j, and that they were born into cohorts c_i and c_j, respectively. Unlike the parent-offspring case, where overall spawner abundance and relative reproductive output (RRO) dictate the probability of finding a POP, with the half-sibling data there are more factors that influence their commonness, although these factors still ***all*** relate to adult, not juvenile, dynamics:

1. The expected total mortality accrued by an adult in the time between the two birth years c_i and c_j
2. The proportional increase (if any) in the relative reproductive output of an adult in the time between the two birth years c_i and c_j
3. The distribution of probable parents in the earliest cohort $c_{\min} = \min\{c_i, c_j\}$
4. The total spawning abundance in the latest cohort $c_{\max} = \max\{c_i, c_j\}$

Formalising this into a (sexually aggregated) probabilistic expression for whether the juvenile pair $\{i, j\}$ is an HSP:

$$p_{\text{hsp}} = \frac{4q_{\text{hsp}}}{S_{c_{\max}}}\left(\sum_a \left[\gamma_{c_{\min},a} \exp\left(-\sum_{y=c_{\min}}^{c_{\max}-1} Z_{y,a}\right)\varphi_{a+|c_i-c_j|}\right]\right), \tag{2.6}$$

where φ_a is the relative reproductive output as used in the POP probability [3], $S_y = \sum_a N_{y,a}\phi_a$ is the spawning abundance, and $\gamma_{y,a}$ is the age distribution of adults in year y:

$$\gamma_{y,a} = \frac{N_{y,a}\phi_a}{S_y}. \tag{2.7}$$

The HSP probability in (2.6) is quite complicated and it is worth discussing in detail - both to understand the factors involved and to begin to grasp the information content in the HSP data themselves. The first part of the equation is similar to the POP probability: the chance of finding the younger fish to be a half-sibling of the older fish is clearly, at some level, inversely proportional to the overall size of the spawning stock *in the younger fish's birth year*. The factor of 4, versus the factor of 2 one sees in the POPs, comes from the following observation: for a 50/50 sex-ratio the chance of finding a maternal or paternal HSP would be $1/(S/2)$ if $c_i \equiv c_j$, so adding these together (i.e. we don't distinguish between maternal and paternal cases) would yield a probability of $4/S$.

The terms inside the first bracket in (2.6) are where the HSP probability gets adjusted for three key factors: (i) we don't observe adults directly, so must integrate over all possible adult ages; (ii) for cross-cohort comparisons ($c_i \neq c_j$), all adults will have experienced some mortality between the two birth years in the comparison, making them less likely to be a future parent; and (iii) again for cross-cohort comparisons some adults will have increased their relative reproductive output, making them more likely to be a future parent.

The "catchability" factor, q_{hsp}, is included to account for the potential bias that might occur in estimates of adult abundance given systematic length-specific effects on relative reproductive ouptut of adults. This is not really an issue for POPs, unless such effects are somehow correlated with being captured. However, for HSPs because we are considering the output of the *same* individual at two points in time (i.e. the juvenile birth years), the relative reproductive potential of an adult that is lower/higher than average in the first birth-year could be similarly lower/higher than the average in the second birth-year. In future, a more detailed length-specific formulation of the HSP probability can be explored to deal with this bias potential directly, but for now this "catchability" approach will suffice - both for simulations and for the initial inclusion of the data in the OM.

In general, but especially for high-fecundity broadcast spawners like SBT, intra-cohort comparisons ($c_i = c_j$) are to be avoided. This is because of the potential for full/half-siblings from the same cohort to be far more probable than their cross-cohort counterparts - even in the absence of adult mortality and changes in reproductive output. Given the likely very high and very variable mortality rates of recently fertilised eggs, siblings fertilised at the same time, and in favourable conditions, are much more likely (relative to unrelated fish from the same cohort) to appear in the data. This would place a strongly downward bias on the abundance estimates if not accounted for. This effect can be dealt with, in principle, via the estimation of an intra-cohort inflation term for (2.6), but the simplest solution is not to make intra-cohort comparisons at all.

In terms of the likelihood for the HSP data, the binomial distribution is again a natural choice (as it was for the POPs) given the nature of the data (is it an HSP or not?). The extension to the beta-binomial distribution, to deal with over-dispersion if this is found to be present, can be done as outlined for the GT data in Section 2.1.3.

2.2 Structural changes for projections

This section focusses only on potential structural changes to the OM in projection mode that don't relate to new data generation, which are dealt with in [5], or UAM which is dealt with in [6]. Two substantial issues to consider are how to deal with time-varying selectivity and growth in projections. Currently, selectivity (for each fleet included in projections) and growth are fixed at their most recent values. At one level, this is a parsimonious approach: we do not have measurable (or easy to simulate) predictors that would be able to define what future levels of time-varying selectivity and growth might be. However, many of the reasons behind why both selectivity and growth have changed historically are likely to be linked to future stock dynamics (e.g. stock abundance).

2.2.1 Selectivity

All fleets have time-varying non-parametric selectivity, and this level of freedom in the selectivity functions is warranted given the clear shifts in the length/age distributions in the various fisheries that seem independent of recruitment or mortality drivers. Selectivity is assigned in blocks of years for each fishery, with differing degrees of flexibility within a given time-block across the fisheries. Selectivity can change because of things the fishermen do, in terms of targeting and fishing practices, and because the size distribution of the population in the various regions changes over time (even if mortality rates and abundance at the population level is unchanged).

There has been a general shift in the major long-line fishery (LL_1) towards smaller, younger fish over the period in which the stock exhibited a strong decline in apparent abundance. Whether driven by economics, relative availability, or a combination of the two, it is reasonable to assume that, if the stock does indeed recover, the fleet will react to these changes and with attendant changes in future selectivity, relative to the present.

If these changes do occur, as it is reasonable to assume that they will, they impact on a number of important MP performance issues:

- Interpretation of CPUE as an abundance index. The Historical CPUE has a complicated, time-dependent and non-linear relationship to historical spawner abundance (see [5]). This relationship change over time is dominated by changes in selectivity, and accounting for them historically is vitally important to interpreting CPUE as an index of abundance. Any MP with CPUE as an index is assuming that CPUE links *in some way* to spawner abundance, as is done in the current MP. Assuming a fixed selectivity in the future, when it is quite possible it will change, will make CPUE seem like a better standalone index of sub-adult/adult abundance than it actually might be.
- Relative impact of catch levels from an MP when selectivity is changing, for a fixed relative catch allocation scheme. Fisheries in projections have specific agreed allocations. A notable change in the selection pattern of a given fishery, especially one who's data is to be used in a candidate MP, could yield quite different harvest rates across the exploited age-range, and with different outcomes (rebuilding performance, catch variability etc.) in an MP performance sense, for similar overall TAC levels. A well-tuned MP should be able to react to such changes, but they could occur none the less.

There is no reason to change the reference OM from the current assumption that selectivity remains fixed in the future, but we should consider robustness tests that do allow for future variations. A fairly simple example would be to link age at maximum selectivity for the long-line fleets (particularly LL_1) to increase with increasing abundance.

2.2.2 Growth

Figure 2.2 shows the change in mean length-at-age over time for SBT. Since informative growth data began being collected in the 1960s (tagging, then ageing data) mean length has increased notably at the youngest observed ages (1–4) and slightly decreased at the oldest ages (12 and above). The functional form of the growth relationship has changed also, with a much more pronounced two-stage growth pattern (slower–faster–slower) from age zero upwards, relative to the historically more von-Bertalanffy like dyanmics of gradually decreasing growth rate with age.

The most recent distribution of length-at-age is used in the projections, and it has long been a point of discussion in the ESC as to whether this may or may not have an effect on the projections. There are three probable (and, possibly, interacting) reasons why growth might have changed over time: (i)

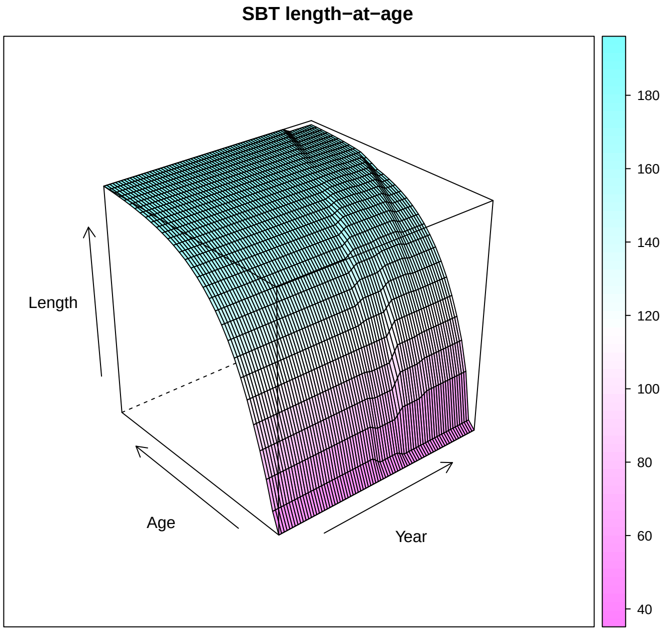

Figure 2.2: *Changes in mean length-at-age over time for SBT (ages 0–30 and years 1931–present).*

high historical exploitation removed slower growing individuals from the population (evolutionary effect); (ii) something fundamental in the ecosystem and/or surrounding environment changed the growth dynamics; and (iii) there are density-dependent growth dynamics present where competition for resources decreased, thereby increasing growth rates, as the overall stock abundance declined.

At least at the timescales the OM projections functions on, we can assume any evolutionary effect is quite likely to be effectively permanent. If an ecosystem/environmental shift occurred, and is very weakly dependent on SBT themselves, then the absence of information on the likely cause suggests the most sensible assumption is that it remains fixed in future. The density-dependent growth scenario is plausible (the correlation between mean length and overall abundance is approximately -0.8) and, if the stock recovers, is obviously incompatible with the assumption of fixed future growth dynamics. A major factor in deciding on robustness tests is their plausibility, relative to the reference OM assumptions and settings, and their potential for impacting the future dynamics so density-dependent growth is worth considering when viewed at in this manner.

Its potential impact is fairly simple to imagine. Current targets, in terms of stock rebuilding, mean trying to increase current spawner abundance to levels estimated in the late 1970s and early 1980s. Growth in this period was slower so, for the same natural mortality rates, the SSB-per-unit-recruit would be lower than it would be now, where growth is faster. The associated weight-at-age is also higher, so the same amount of catch biomass for a given selectivity-at-age will result in lower overall harvest rates. These two factors would, if density-dependent growth were true, tend to over-estimate how fast the stock could recover. As with the selectivity argument, a simple linkage between spawner depletion and mean length-at-age could be constructed to explore how much of an impact density-dependent growth might have, relative to our current fixed growth assumptions in the OM.

3 Discussion

This paper focussed on a number of issues relating to the inclusion of new data sources, and structural changes potentially required for the 2017 reconditioning and MP testing.

One of the key new information sources will be in relation to additional catch data - both from non-members and member-attributable. Estimates of non-member catch have focussed on assigning the relevant effort to be either "Japanese" or "Taiwanese" in nature [2], thereby assumed to have the same catchability and selectivity properties as those fisheries. This makes their historical inclusion in the OM fairly simple: the additional catch associated with the non-member effort is included in the reported catches for the relevant fleet (i.e. LL_1 or LL_2 in this case). For attributable catch, the situation is slightly more complex. If the catch can be demonstrated to be similar enough to an existing fleet it may be included there; if not, we have to either estimate (via composition data) - or agree on - an assumed selectivity for that attributable catch.

The gene tagging design study [1] focussed on releases and recaptures of age 2 and 3 fish, respectively, however we need to incorporate the complexities of probabilistically assigning age to tagged and catch sampled fish. Given what we know about the distribution of length-at-age of fish likely to be in the GAB in summer we can focus on tagging *mostly* age 2 fish and some age 3, and looking among *mostly* age 3 and some age 4 fish for subsequent recaptures. The specifics of how to adjust the estimator for this were outlined, as well as how information over multiple years that relates to the abundance of the same cohort can be included. A generic beta-binomial likelihood is defined for the gene tagging data that will be able to deal with over-dispersion directly if required.

The next-generation close-kin data focusses on identifying half-sibling pairs (HSPs) between juvenile-juvenile comparisons, as opposed to looking for parent-offspring pairs (POPs) between juvenile-adult comparisons. These data - independently of POPs - have the potential to be extremely informative on not just adult abundance, but also mortality rates [4]. The HSP likelihood is more complicated than for the POP case and is outlined in detail in this paper and - similar to the gene tagging case - we opt for a flexible beta-binomial likelihood for these data.

There are two key time-varying features in the current OM in terms of historical dynamics: selectivity and growth. Both are estimated to vary substantially over time, and accounting for this variation is vital to understanding both abundance trends in the data, and the estimation current stock status. One continuing issue for the OM in projection mode is that these factors remain fixed at their most recent values for all future years. Given that the reasons for the change in both these relationships over time is likely to be indelibly linked to changes in the population over time - most notably abundance - we explore hypotheses and options for how they might be simulated to vary in the future.

4 Acknowledgements

This work was funded by the Department of Agriculture & Water Resources and CSIRO Oceans & Atmosphere.

References

[1] Preece, A., Eveson, J. P., Davies, C. R., Grewe, P. G., Hillary, R. M., and Bravington, M .V. (2015) Report on gene-tagging design study. *CCSBT-ESC/1509/18*.

[2] Edwards, C. T. T., Williams, A., Hoyle, S. *et al.* (2016) Updated estimates of southern bluefin tuna catch by CCSBT non-member states. *CCSBT-ESC/1609/BGD 02*.

[3] Hillary, R. M., Preece, A., and Davies, C. R. (2014) Assessment of stock status of southern bluefin tuna in 2014 with reconditioned operating model. *CCSBT-ESC/1409/21*.

[4] Bravington, M. V., Skaug, H. J., and Anderson, E. C. (2016) Close-kin mark-recapture. *Stat. Sci.* (in press).

[5] Hillary, R. M., Preece, A., and Davies, C. R. (2016) Methods for data generation in projections. *CCSBT-OMMP/1609/7*.

[6] Preece, A., Davies, C. R., and Hillary, R. M. (2016) Advice on incorporating UAM in stock assessment and MP evaluation and implementation. *CCSBT-OMMP/1609/5*.

5 Appendix

In terms of the likelihood function required for the beta-binomial, we first begin by assuming that the *true* sampling probability, $\tilde{p}$, is described by a beta distribution: $\tilde{p} \sim B(\alpha, \beta)$. If there is no additional variability in the true sampling distribution, $\tilde{p}$, relative to the model-predicted sampling distribution, $\widehat{p}$, then with sample size n the mean and variance of the distribution (of recaptures in this case) are $n\widehat{p}$ and $n\widehat{p}(1-\widehat{p})$, respectively. To deal with additional variance we can introduce a new parameter, $\omega = (\alpha+\beta)^{-1} > 0$, so that while the mean of the distribution remains the same, the variance of R can be expressed as follows:

$$\mathbb{V}(R) = \frac{n\widehat{p}(1-\widehat{p})(n\omega+1)}{(1+\omega)}, \tag{5.1}$$

and so the inflation factor in the variance, the over-dispersion factor ϕ, is

$$\phi = \frac{(n\omega+1)}{(1+\omega)}. \tag{5.2}$$

It is ϕ that is usually estimated from the residuals in some way [3] and ω is easily solved for:

$$\omega = \frac{(\phi-1)}{(n-\phi)}. \tag{5.3}$$

The actual values of α and β for the distribution of $\tilde{p}$ are, after a little algebra, given by the following:

$$\alpha = \frac{(n-\phi)\widehat{p}}{(1-\widehat{p})(\widehat{p}+(1-\widehat{p})(\phi-1))} \tag{5.4}$$

and

$$\beta = \frac{n-\phi}{\widehat{p}+(1-\widehat{p})(\phi-1)}. \tag{5.5}$$

In terms of the final likelihood, we would like to integrate over the random variable $\tilde{p}$, and obtain the likelihood of observing the given recaptures, R, in terms of n and $\widehat{p}$. Fortunately, the marginal likelihood we are interested in (i.e. when integrating over $\tilde{p}$ in the joint beta-binomial likelihood) is of a known form:

$$\ell\left(R\,|\,\widehat{p}\right) \propto \frac{\Gamma(R+\alpha)\Gamma(n-R+\beta)\Gamma(\alpha+\beta)}{\Gamma(n+\alpha+\beta)\Gamma(\alpha)\Gamma(\beta)}, \tag{5.6}$$

where $\Gamma(\bullet)$ is the gamma function, and α and β are as defined in (5.4) and (5.5), respectively.

CONTACT US

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

YOUR CSIRO
Australia is founding its future on science and innovation. Its national science agency, CSIRO, is a powerhouse of ideas, technologies and skills for building prosperity, growth, health and sustainability. It serves governments, industries, business and communities across the nation.

FOR FURTHER INFORMATION

CSIRO Oceans and Atmosphere
Rich Hillary
t +61 3 6232 5452
e Rich.Hillary@csiro.au
w