MARINE AND ATMOSPHERIC RESEARCH
www.csiro.au

Estimates of recruitment variability outside of the SBT operating model

Rich Hillary, Ann Preece, Campbell Davies

CCSBT-OMMP/1307/06
Prepared for the 4th CCSBT Operating Model and Management Procedure Technical Meeting 23-26 July 2013, Portland, Maine, US.

CSIRO Marine and Atmospheric Research
Wealth From Oceans National Research Flagship
Castray Esplanade, Batter Point, Hobart TAS 7000, Australia
Telephone : +61 3 6232 5222
Fax : +61 3 6232 5000

Contents

Abstract

Stochastic recruitment variance in the SBT operating model plays a crucial role in the uncertainty in stock projections, as well as being strongly influential on the estimates of historical recruitment and in particular steepness. It is also not directly estimable in the SBT operating model and an assumed value of 0.6 is currently used. This paper explores the use of a modified version of the population and estimation model used in the CCSBT Management Procedure to obtain an estimate of the recruitment variance outside of the SBT operating model.

1 Background

The variability in recruitment, often denoted by its lognormal standard deviation σ_r is a fundamental variable in almost all modern integrated stock assessments. Its primary role is as a way of defining the *a priori* variability in annual (or otherwise) deviations from the expected recruitment, but also in stochastic projections as a means of simulating future unknown recruitment deviations. In the CCSBT context the variable σ_r is central to both the historical conditioning of the OM and in the projections used to test the various MPs.

Another key factor that makes the value of σ_r important is that it is almost never estimated, with an assumed value being by far the most common approach, and with $\sigma_r = 0.6$ being the usual - if perhaps apocryphal in origin - assumption. To be able to estimate σ_r as an actual model parameter, one needs to treat the recruitment deviations as random effects. While there are a number of estimation methods and packages available that can do this, including in the `ADMB` framework [1] often used in fisheries models, it is done very rarely.

As detailed in [2] the recruitment deviation penalty, the magnitude of which is controlled directly by σ_r, also has a significant effect on the different levels of steepness resampled in the SBT OM. Indeed, since the preliminary inclusion of the close-kin data in [2] the recruitment penalty is the only real source of information on steepness, with the actual data having no overall preference at all. Whether or not this level of prior forcing is true information, by assuming a value for σ_r we are having an implicit effect on the steepness distribution so we should attempt to explore how good our assumed value of $\sigma_r = 0.6$ actually is.

Given the complexity of the SBT OM, and the non-separable nature of the recruitment deviations within the model, it was not possible to treat them as random effects and estimate σ_r within the current OM structure. One could, in theory, use a profile likelihood approach and look at the likelihood over a range of values of σ_r *but* this ignores the fact that σ_r changes the effective degrees of freedom (EDF) within the OM itself. Increasing σ_r will almost surely result in a better fit to the data, but this also entails an increase in the EDF and will likely result in less precise recruitment deviation estimates. By looking at only the first bit (better fits) but not accounting for the second and third points (which count against increasing σ_r) we will, almost surely, over-estimate σ_r.

In this paper we demonstrate and approach to estimate of σ_r that is obtained outside of the SBT OM structure. The idea is to use a modified form of the population and estimation model in the SBT MP. With some minor modifications, essentially turning the MP model into a full Bayesian hierarchical model, it can be configured to provide an estimate of the process variation in the *biomass* of juveniles aged 2 to 4. It should be stated that we do not propose an alternative MP in this paper; we are merely using a modification of the adopted MP structure for a very different purpose. We can then infer from this process variation what the likely variation in age 0 recruitment, σ_r, would be.

2 Material & Methods

This section is split into two sub-sections, the first of which specifies with the data sets used in the estimation procedure, and the second details the changes required to the formulation of the MP model to estimate σ_r.

2.1 Data sets

Three data sets are used in the proposed approach:

1. Standardised Japanese long-line CPUE used in the MP (1994-2012)
2. Scientific aerial survey [3] (1993-2000,2005-2013)
3. SAPUE biomass index [4] (2003-2013)

One clear difference with the MP model is that we include the SAPUE data, not just the CPUE and aerial survey data. The reason for this is that, when estimating variances, essentially the more data the better. The aerial survey is missing data from 2001 to 2004 and is the primary information source in the MP on recruitment trends. The recently revised SAPUE index begins in 2003 so, while not totally covering the gap in the aerial survey, does provide us with extra data on recruitment trends and thereby will hopefully assist in obtaining more precise and robust estimates of variance.

2.2 Modified MP population & estimation model

First, it makes sense to revisit the specifics of the MP population model: recruitment (R_y) and adult (B_y) biomass are related as follows:

$$B_{y+1} = R_y + g_y B_y, \tag{2.1}$$

where g_y is the adult biomass net growth effect (encompassing natural mortality, surplus production and exploitation effects). For the recruitment process the following model is assumed:

$$R_y = \exp\left(\mu_R + \epsilon_y^R\right), \tag{2.2}$$

with $\epsilon_y^R \sim N\left(-\sigma_R^2/2, \sigma_R^2\right)$. For the g_y a conceptually similar model is assumed and

$$g_y = \exp\left(\mu_g + \epsilon_y^g\right), \tag{2.3}$$

with $\epsilon_y^g \sim N\left(-\sigma_g^2/2, \sigma_g^2\right)$. For the aerial survey data I_y^{AS} a log-scale multivariate normal relationship to the recruiting biomass is assumed but with a one-year delay: $\ln \mathbf{I}^{AS} \sim MVN\left(q^R \mathcal{S}(\mathbf{R}), \Sigma_{AS}\right)$, where $\mathcal{S}(\circ)$ is just the right time-shift operator and Σ_{AS} is the aerial survey covariance matrix. The reason for this delay is because we assume that the aerial survey covers ages 2 to 4 and that the CPUE covers ages 4 to 12/18. To make sure that we are more likely to detect the movement of a signal in the aerial survey appearing in the CPUE data this delay is assumed as R_y represents the recruitment biomass contribution to the adult biomass (assumed covered by the CPUE). The situation is simpler for the CPUE likelihood and these data are assumed log-normally distributed about the adult biomass: $I_y^B \sim LN\left(q^B B_y, \sigma_B^2\right)$. The model is unidentifiable without additional information on the catchability ratio q_R/q_B and the details of how this is dealt with can be found in [5].

The approach to including the SAPUE index, I_y^S, in the MP model structure was first explored in [5]. As with the aerial survey we assume that $I_y^S \sim LN\left(q^S R_{y+1}, \sigma_{y,S}^2\right)$, given they are observing largely the same population in a similar geographical area. In this model formulation $\ln q^S$ is assigned a normal prior mean μ_{q^s} and variance $\sigma_{q^S}^2$ which results in the following normal conditional posterior:

$$p\left(\ln q^s\right) \sim N\left(\left(\frac{\mu_{q^s}}{\sigma_{q^S}^2} + \sum \frac{I_y^S}{R_{y+1}\sigma_{y,S}^2}\right) \times \left(\sigma_{q^S}^{-2} + \sum \sigma_{y,S}^{-2}\right)^{-1}, \left(\sigma_{q^S}^{-2} + \sum \sigma_{y,S}^{-2}\right)^{-1}\right). \tag{2.4}$$

When discussing the potential inclusion of the SAPUE data in the SBT OM [6] one issue that was raised as a problem was the potential correlation between the aerial survey and SAPUE indices, in terms of both observation and process error. A simple approach was used in this work to try and deal with this issue. For the aerial survey if σ_y^{AS} are the square roots of the diagonal elements of the aerial survey observation covariance matrix then those diagonal elements are replaced with the following:

$$\{\Sigma^{AS}\}_{yy} = \sigma_y^{AS}\left(\sigma_y^{AS} + \rho\sigma_{y,S}^{S}\right), \tag{2.5}$$

where ρ is the correlation between the survey and SAPUE indices. This adjustment does not alter the positive definite nature of the covariance matrix, it merely inflates the year-specific variance given the correlation coefficient. For the SAPUE index a similar adjustment is made to the year-specific observation variance. In principle, the correlation coefficient could also be estimated but such model structures were not explored in this paper, so an indicative value of $\rho = 0.25$ was assumed.

In terms of estimated parameters (and priors) we have:

- μ_R and μ_g with uniform priors
- ϵ_y^R and ϵ_y^g with normal priors
- B_{init} the initial relative adult biomass (disperse lognormal)
- $\ln q^S$ with a (disperse) normal prior
- σ_R and σ_g with log-uniform priors (i.e. $p(\sigma_\bullet) \propto \sigma_\bullet^{-1}$)

The fundamental changes to the model as it is implemented in the CCSBT MP are:

- The full aerial survey covariance matrix is used in the likelihood, as are the year-specific standard deviations in the SAPUE index
- Correlation between the two recruitment indices is (simply) accounted for
- The additional penalty forcing the year-averaged means of the recruitment and biomass growth random effects to be zero is removed
- The biomass of the initial state is now estimated
- The catchability (relative to the aerial survey and CPUE) of the SAPUE series, q^S, is estimated
- The recruitment and biomass growth random effect variances are now estimated, making this a Bayesian hierarchical model

In the MP the model acts as a biologically plausible smoother, with a number of constraints in place to make it stable but flexible enough to fit the CPUE and aerial survey data [5]. For this work, the emphasis is on estimating the variance parameters of the recruitment and biomass growth random effects as well - especially for recruitment. To do this robustly we have turned the MP model back into its original form of a relative abundance stock assessment, removing the additional penalties, estimating random effect variances as well as uncertainty in the initial states.

To efficiently obtain a representative sample from the joint posterior of the parameters and hyperparameters a Metropolis-within-Gibbs MCMC routine was written in C++. A burn-in level of 1,000 iterations was used, with 1,000 being retained with a thinning factor of 100 employed to reduce auto-correlation in the Markov chains. Non-convergence of the chains was explored using regular diagnostic methods [7]

3 Results

3.1 Parameter summaries & data fits

Table 3.1 summarises the estimates of the time-independent parameters (μ_R, μ_g, B_{init}, and q^S) and hyper-parameters (σ_R and σ_g). Figure 3.1 details the summary of the key population dynamic variables

Parameter	*Summary*
μ_R	-1.5 (-1.8; -1.2)
μ_g	-0.35 (-0.5; -0.15)
B_{init}	0.91 (0.66; 1.34)
q^S	5.6 (4.8; 6.4)
Hyper-parameter	
σ_R	0.5 (0.35; 0.73)
σ_g	0.3 (0.09; 0.55)

Table 3.1: Summaries of time-independent parameters in terms of posterior median and lower and upper limits of the 95% credible interval.

R_y, B_y and g_y. As in previous years [5] the data are informative on the parameters in the MP but clearly also for the additional parameters estimated in this form of the model.

Figure 3.2 details the estimation performance summary of the model, in terms of fits to the data and posterior predictive performance - basically how well does the probability model predict the data post-estimation. The data are fitted well in general - similar to previous years [5] - with one difference: the survey index is higher than the SAPUE index in 2013 so the model predicts a recruitment biomass level in between the two (see Figure 3.2). In previous years, the agreement between the SAPUE and aerial survey indices in common years has been closer. This effect can also be detected in the posterior predictive performance summaries.

For posterior predictive analyses the Bayesian p-value is the probability with which the predicted discrepancy statistic (”closeness” of the simulated data to the deterministic prediction) is greater than the observed one (”closeness” of the actual data to the deterministic prediction). Ideally, one would like Bayesian p-values as close to 0.5 as possible, with values outside the range of 0.05-0.95 suggesting something systemically wrong with the model. For the CPUE the p-value is very close to 0.5 with a regular cloud of discrepancy statistics around the diagonal, so no issues there. For the aerial survey and SAPUE indices, while the cloud of points is nice and regular, the p-values are both around 0.35 suggesting the observed data are more variable than the predicted data. In previous years this has not been the case [5] and is driven by the divergence in trend in the aerial survey and SAPUE indices in 2013. The model splits the difference basically between the two series and is, therefore, adjudged unable to predict the variation in both series as well as before which is why we see smaller p-values.

3.2 Inferred estimates of σ_r

From the augmented version of the MP model used in this paper, the estimates of stochastic variation in recruitment biomass (as observed in the aerial survey and SAPUE indices so assumed to be ages 2 to 4) were around a level of 0.5 with a posterior CV of around 0.15. The final step in obtaining some kind of estimate of the variation in age 0 mean recruitment is to use this estimate of variance σ_R^2 to back-calculate what likely levels of σ_r^2 might be.

It first helps to think about what the process variance due to recruitment variability at age 0 might look like in the age-classes covered by the survey. In the most simple case, given that we assume that the survey and SAPUE indices cover ages 2 to 4 the noise in the indices will be constituted of the noise relating to 3 distinct cohorts (i.e. a moving average of previous recruitment deviations). Assuming that $\epsilon_y^r \sim N(0, \sigma_r)$ represents a simplistic view of the deviations in age 0 recruitment then the variation in age 2 to 4 abundance due to these variations will be

$$\epsilon_y^R = \sum_{i=y-4}^{y-2} \omega_i \epsilon_i^r, \tag{3.1}$$

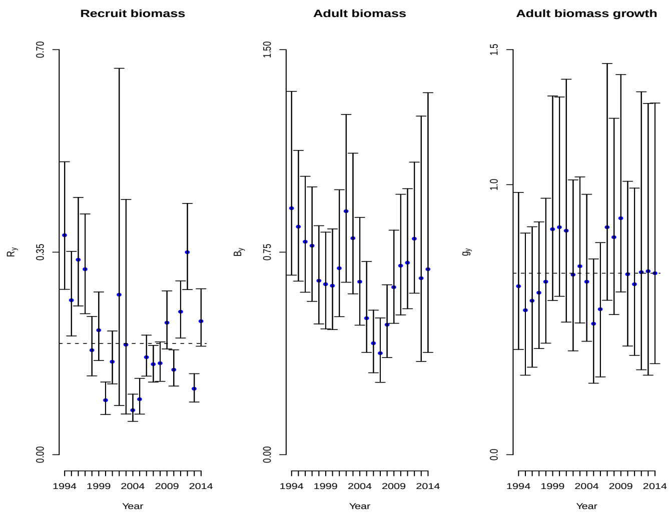

Figure 3.1: Median (blue circles) and 95% credible interval for the recruitment biomass (left), adult biomass (middle) and biomass net growth (right).

and ignoring variation in relative abundance-at-age and weight-at-age (i.e. $\omega \equiv 1$) ϵ_y^R will be an auto-correlated process with correlation coefficient $2/3$ and standard deviation $\sigma_r/\sqrt{3}$. The more cohorts you average over in the abundance index, the less variability due to recruitment but also the more correlated the variation becomes across time. Any autocorrelation in the age 0 recruitment deviates would further inflate the overall correlation expected in the recruitment biomass observed in the survey. This would need to be dealt with in the back-calculation of σ_r as not doing do would bias the results.

In reality, our variations are not additive but multiplicative in terms of how they affect the relative abundance of each cohort, and even at population equilibrium the relative changes in cohort abundance and weight-at-age mean we cannot ignore the weighting terms either. However, the simple system above *does* demonstrate that there is a clear relationship between age 0 variation and the random variation in the recruitment indices we have that are assumed to be covering ages 2 to 4.

There are no simple closed-form solutions for the real world system we are considering for SBT in terms of σ_r in terms of σ_R but we can numerically back-calculate the likely distribution of σ_r from our estimated distribution for σ_R nonetheless. This inferred distribution for σ_r is given in Figure 3.3 and the median and 95% credible interval is 0.64 (0.51-0.83). So this estimate is actually close to the value of $\sigma_r = 0.6$ assumed in the SBT OM **but** it is likely to be an over-estimate for this reason: it ignores process error in the form of temporal changes in the age distribution expected to be observed in the aerial survey and SAPUE indices (due to migration, mortality changes etc.). The levels of autocorrelation in the recruitment biomass random effects did not appear to be above the level expected for very low or effectively zero autocorrelation in the age 0 recruitment deviations, so we do not expect there to be any biases that need adjusting for in this regard.

4 Discussion

In this paper we have demonstrated how we can convert the population and estimation model that underpins the CCSBT MP into a form that can be used to sequentially obtain an estimate of the recruitment variance, σ_r^2, which is an influential but assumed variable in the SBT operating model. Some of the stabilising penalties included in the MP model are removed and the initial adult biomass state is now estimated not fixed, thereby permitting the estimation of the recruitment biomass and biomass net growth random effect variances. It is the recruitment biomass variance that can be linked back to the age zero recruitment

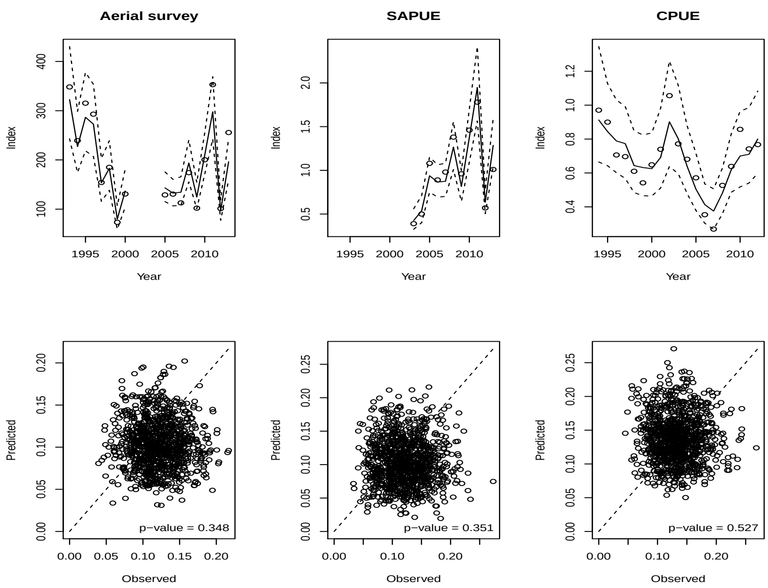

Figure 3.2: Top row summarises fits to aerial survey (left), SAPUE (middle) and CPUE (right) indices (observed, circles; predicted median (full) and 95% credible interval (dotted lines)). Bottom row summarises the posterior predictive performance of the model (including the p-values).

variance of interest.

The SAPUE data are also included in the estimation scheme as they possess data which cover the period missing in the aerial survey index, and the potential correlation between the two indices is dealt with in a simple but somewhat *ad hoc* manner. The MP model (using a penalised likelihood approach) is now converted to a fully Bayesian hierarchical state-space model estimating recruitment and adult biomass and adult biomass net growth from 1994 to 2014 using the aerial survey, SAPUE and standardised CPUE abundance indices.

The data are informative for all the parameters in the augmented model - including the variances of principal interest - and the model explains the data well, albeit with the decrease in consistency between the aerial survey and SAPUE indices in 2013 causing a minor decrease in predictive performance relative to previous years [5]. Estimates of recruitment biomass variance, σ_R, had a mean of 0.5 with a posterior CV of around 0.15. Assuming that the process variance in the cohorts observed in the aerial survey and SAPUE indices is dominated by cohort strength variability, inferred estimates of σ_r had a median (and 95% credible interval) of 0.64 (0.51-0.83).

While this inferred estimate of σ_r is actually quite close to that assumed in the OM ($\sigma_r = 0.6$), it is most likely better interpreted as an upper bound at this time. In this model structure we are not able to account for the contribution of additional process variance linked to the age structure observed in the GAB by both the recruitment indices. So while we might expect recruitment variability to dominate the process variance in an index assumed to be comprised of 2 to 4 year old fish, we cannot discount this potential contribution to the estimates of σ_R and, by implication, σ_r. What we can say is that we do not appear to be overly-restricting the OM recruitment deviations *a priori*, nor do we seem to be under-estimating the level of future recruitment variability assumed in the OM projections used for MP evaluation.

5 Acknowledgements

This work was funded by AFMA and CSIRO's Wealth from Oceans National Research Flagship.

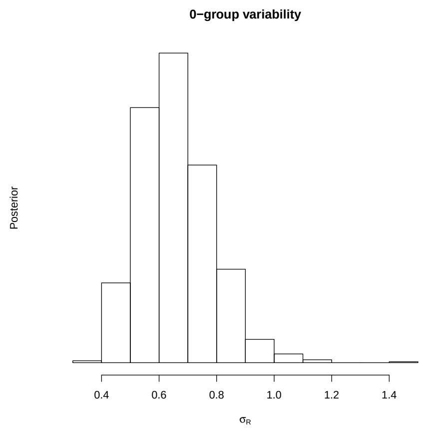

Figure 3.3: Distribution of σ_r given our estimation distribution of σ_R.

References

[1] Fournier, D.A., Skaug, H.J, Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A.,and J. Sibert. (2012) AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optim. Methods Softw.* 27, 233–249.

[2] Hillary, R. M., Preece, A., Davies, C. R., Bravington, M. V., Eveson, J. P., and Basson, M. (2012) Initial exploration of options for inclusion of the close-kin data into the SBT operating model. *CCSBT-ESC/1208/21*.

[3] Eveson, J. P., Farley, J. and Bravington, M. V. (2013) The aerial survey index of abundance: updated analysis methods and results for the 2012/13 fishing season. *CCSBT-ESC/1309/**.

[4] Farley, J. and Basson, M. (2013) Commercial spotting in the Australian surface fishery, updated to include the 2012/2013 fishing season. *CCSBT-ESC/1309/**.

[5] Hillary, R. M. and Preece. A. (2011) Updated technical specifications and performance analyses for MP1. *CCSBT-ESC/1107/12*.

[6] Hillary, R. M., Eveson, J. P., Preece, A., and Basson, M. (2012) Potential inclusion of SAPUE index in the SBT operating model. *CCSBT-ESC/1208/20*.

[7] Brooks, S. P., and Roberts, G. O. (1997) Assessing Convergence of Markov Chain Monte Carlo Algorithms. *Stat. & Comput.* 8, 319–335.

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 Marine and Atmospheric Research
Rich Hillary
t +61 3 6232 5452
e Rich.Hillary@csiro.au
w