



CCAMLR

Commission for the Conservation of Antarctic Marine Living Resources  
Commission pour la conservation de la faune et la flore marines de l'Antarctique  
Комиссия по сохранению морских живых ресурсов Антарктики  
Comisión para la Conservación de los Recursos Vivos Marinos Antárticos

WG-SAM-2021/09

11 June 2021

Original: English

## Recruitment modelling for *Euphausia superba* stock assessments considering the recurrence of years with low recruitment

---

C. Pavez, S. Wotherspoon, D. Maschette, K. Reid and K. Swadling

WG-SAM



This paper is presented for consideration by CCAMLR and may contain unpublished data, analyses, and/or conclusions subject to change. Data in this paper shall not be cited or used for purposes other than the work of the CAMLR Commission, Scientific Committee or their subsidiary bodies without the permission of the originators and/or owners of the data.

# RECRUITMENT MODELLING FOR *EUPHAUSIA SUPERBA* STOCK ASSESSMENTS CONSIDERING THE RECURRENCE OF YEARS WITH LOW RECRUITMENT

Cassandra Pavez<sup>1</sup>, Simon Wotherspoon<sup>1,2</sup>, Dale Maschette<sup>1,2</sup>, Keith Reid<sup>3</sup>, Kerrie Swadling<sup>1</sup>

<sup>1</sup> Fisheries and Aquaculture Centre, Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, TAS 7001, Australia.

<sup>2</sup> Department of Agriculture, Water and the Environment, Australian Antarctic Division, 203 Channel Highway, Kingston, Tasmania, Australia and

<sup>3</sup> Commission for the Conservation of Antarctic Marine Living Resources, Hobart, 7000, Australia.

## RECOMMENDATIONS

For GYM assessments with proportional recruitment, we recommend to use:

- Formula-based methods for low variance in proportional recruitment;
- Simulation modelling for high variance in proportional recruitment to accurately reproduce mean and standard deviation;

as described in this paper and implemented in the Grym (open source package).

## ABSTRACT

Krill are a keystone species in the Southern Ocean food-web, and, as such, it is crucial to effectively manage the krill fishery to ensure its long-term sustainability. To assess the impacts of current harvesting pressures, evaluations rely on sampling and population modelling. Krill stock projections are developed with the Generalised Yield Model (GYM), which provides an assessment for stock status under current harvesting scenarios and various levels of uncertainties. One of the fundamental components of the GYM is the simulation of recruitment. De la Mare (1994) presents a proportional recruitment function for estimating krill recruitment based on estimates from field surveys. The De la Mare (1994) function uses estimates of the mean and variance of recruitment from survey data to determine the scaling of natural mortality and the distribution of random recruits that reproduce the observed mean and variance estimates. We evaluated De la Mare's (1994) proportional recruitment function and found that for large variations in recruitment the function does not reproduce the observed mean proportion of recruits and its variance accurately. We review the deficiencies within the model and provide two alternative methods, which can support a wider range of values and possible extreme scenarios, such as years of low recruitment.

## INTRODUCTION

The *Euphausia superba* (hereafter krill) fishery is, by tonnage, the largest fishery in the Southern Ocean (Siegel 2016). Krill are a keystone species in the Southern Ocean food-web, and, it is crucial to effectively manage the fisheries to assure long-term sustainability and conservation (Kawaguchi and Nicol 2007). In 1982 the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR) was established to conserve Antarctic marine life by, amongst other things, managing the Southern Ocean fisheries that currently target krill, Patagonian toothfish (*Dissostichus eleginoides*), Antarctic toothfish (*D. mawsoni*) and mackerel icefish (*Champsocephalus gunnari*) (Miller 2011, CCAMLR 2020). CCAMLR pioneered the use of the ecosystem-based management approach alongside precautionary catch limits to manage their fisheries. This approach works by setting catch limits that take account of the impact of harvesting activities on the ecosystem by ensuring enough biomass of the target species is left to support predators and safeguard healthy breeding populations (Miller 2011, CCAMLR 2020).

Understanding the dynamics and impacts of harvesting is essential for selecting appropriate catch limits. This optimization requires estimates of total abundance, age-specific mortality, and recruitment (Constable and de la Mare 2003, Siegel 2005). However, collecting these data from field surveys is currently limited in Southern Ocean species due to the harsh conditions of the Southern Ocean and the ensuing cost of surveying (Cox et al. 2011, Constable and Kawaguchi 2018, Krag et al. 2018). Uncertainties on population parameters for stock assessments led to the development of the Generalised Yield Model (GYM) (Constable and De la Mare 1996). The GYM was designed to simulate age-structured population projections based on estimated recruitment and mortality parameters from survey data. These projections are then used to explore population scenarios based on current harvesting levels and future harvesting strategies to establish sustainable catch limits for upcoming harvesting seasons (Constable and de la Mare 2003, CCAMLR 2020).

Recruitment and mortality are the key drivers of age-structured population projections. In the GYM, these parameters can be simulated using several methods, the most common being the proportional recruitment function (De la Mare 1994). Traditionally, fisheries rely on estimates of absolute biomass of recruits for stock projections rather than proportional recruitment. Since krill biomass at recruitment is difficult to estimate reliably, De la Mare's (1994) model uses krill surveys to determine the proportion of recruits seen in a particular age class, to then calculate the properties of the distribution of recruits. De la Mare's (1994) model aims to reproduce the observed values of the mean proportion of recruits and its variance from the survey by estimating the scaling of natural mortality and the distribution of random recruits (Constable and de la Mare 2003).

De la Mare's (1994) proportional recruitment model is a key element of the GYM when used for krill, and, it is important to evaluate its capacity and further development (SC-CAMLR 2019). Current environmental threats are enhancing the likelihood and severity of extreme events such as the recurrence of years with low recruitment (Reiss et al. 2008, Kawaguchi et al. 2009). Recruitment is generally accepted to be influenced by extrinsic and intrinsic factors, making it more episodic and not uniform as per the assumptions in models (Quetin and Ross 2003, Meyer et al. 2020). If periodicity breaks down, which one might expect under the impacts of expected climate change scenarios and increased harvesting pressures, unusual events with consecutive poor recruitment are more likely to occur (Siegel and Loeb 1995, Quetin and Ross 2003). The increase in the recurrence of years with low recruitment is likely to cause issues when using De la Mare's (1994) proportional recruitment model (see for example Kinzey et al. 2013).

In this study we explore the properties of the De la Mare (1994) proportional recruitment method and determine its capacity to reproduce the highly variable reproduction scenarios. We present a more direct derivation that facilitates the simulation of a broader range of scenarios which also allow for sporadic events of the reoccurrence of years with low recruitment. These methods are all implemented in the *Grym*: an open source package designed to reimplement the GYM and work within R (R Core Team, 2018) (Maschette et al. 2020, Wotherspoon and Maschette 2020).

## DEVELOPMENT OF THE MODEL

To model recruitment, De la Mare (1994) presents a method that aims to reproduce the mean ( $\hat{\mu}_{\mathcal{R}}$ ) and variance ( $\hat{\sigma}_{\mathcal{R}}^2$ ) of the proportion of recruits ( $\mathcal{R}$ ) observed in the survey data. In the model, De la Mare (1994) assumes that recruitment is independent of the size of the population and that krill in each age class have experienced constant mortality and have all come from a random distribution with constant mean and variance. De la Mare (1994) considers the proportion  $\mathcal{R}(r)$ , as the ratio of the number of individuals  $A_r$  in age class  $r$  to the number of individuals in that age class or above ( $A_k$ )

$$\mathcal{R}(r) = \frac{A_r}{\sum_{k=r} A_k}, \quad (1)$$

and proposes a model for recruitment in which the number of recruits to age class  $r$  is a stochastic multiple of the expected number of individuals in all older age classes

$$A_r = \frac{R}{1-R} E \left( \sum_{k=r+1} A_k \right), \quad (2)$$

where  $R$  is a random variable with range  $[0, 1)$ , and the mean and variance of  $R$  must be chosen in order to reproduce the mean and variance of  $\mathcal{R}(r)$  observed in survey data. The difficulty with this approach is that as the observed  $\mathcal{R}(r)$  is a non-linear function of the random variable  $R$ , it is necessary to apply multiple bias corrections when deriving the mean and variance of  $R$ , making the results difficult to generalize. Specifically, when presented with high variance cases, the model fails to reproduce the properties  $\mathcal{R}(r)$ . Here, we propose a more direct approach that facilitates generalisation of the recruitment model technique to a wide set of distributions.

In the proposed method, the number of recruits to the first age class is a stochastic multiple  $Q$  of the expected number of individuals in all older age classes

$$A_1 = Q E \left( \sum_{k=2} A_k \right). \quad (3)$$

In this formulation  $Q$  is a random variable with range  $[0, \infty)$ , and we have assumed the population recruits to the first age class. The mean  $\mu_Q$  and variance  $\sigma_Q^2$  of  $Q$  are again chosen to reproduce  $\hat{\mu}_{\mathcal{R}}$  and  $\hat{\sigma}_{\mathcal{R}}^2$  observed in the survey data. Eliminating the ratio  $R/(1-R)$  greatly simplifies the derivation of expressions for  $\mu_Q$  and  $\sigma_Q^2$ .

To derive the model, let  $A_k$  denote the number of individuals currently in age class  $k$ ,  $a_k$  the number of individuals originally recruited to that cohort, and  $S_k$  the survival from recruitment to the current year, so that for each age class  $A_k = a_k S_k$ , with  $A_1 = a_1$  and  $S_1 = 1$ . Further, we require that the survival  $S_k = S_k(M)$  are functions of a common instantaneous mortality  $M$ . For later convenience we define the sums as

$$s_0 = \sum_{k=2} S_k, \quad s_1 = S_r^{-1} \sum_{k=r+1} S_k, \quad s_2 = S_r^{-2} \sum_{k=r+1} S_k^2.$$

De la Mare (1994) explains that finding the proportion of recruits essentially translates to finding the rate of natural mortality. While in the De la Mare (1994) derivation  $M$  is constant across time and age classes, in the new formulation  $M$  is constant across the age classes but allowed to vary across time, so that younger age classes can experience higher mortality rates compared to older age classes as per natural conditions. Assuming the  $a_k$  are independently and identically distributed with mean  $\mu_a$  and variance  $\sigma_a^2$ . With these definitions Equation (3) can be written as  $a_1 = A_1 = Q \mu_a \sum_{k=2} S_k$ , and by using the sums  $a_1 = Q \mu_a s_0$ , Taking expectations of both sides shows that that  $\mu_Q$  and the sum of the  $S_k$  are inverses of each other, to signify that the rate at which individuals survive is balanced by the rate at which new individuals are recruited

$$\mu_Q = s_0^{-1}. \quad (4)$$

The model for the number of recruits can then be reduced to

$$a_1 = \frac{\mu_a}{\mu_Q} Q \quad (5)$$

where  $\mu_Q = s_0^{-1}$  and  $\mu_a$  is a free parameter.

Now consider the mean and variance of  $\mathcal{R}(r)$ . Although we have the population recruit to the first age class, this age class may not be observable in a survey and so we allow for  $r > 1$  in  $\mathcal{R}(r)$ . Second order approximations for the mean and variance of a ratio  $X/Y$  are (Stuart and Ord 1998)

$$E\left(\frac{X}{Y}\right) = \frac{E(X)}{E(Y)} - \frac{\text{COV}(X, Y)}{E(Y)^2} + \frac{\text{VAR}(Y) E(X)}{E(Y)^3}$$

$$\text{Var}\left(\frac{X}{Y}\right) = \frac{E(X)^2}{E(Y)^2} \left[ \frac{\text{Var}(X)}{E(X)^2} - 2 \frac{\text{Cov}(X, Y)}{E(X) E(Y)} + \frac{\text{Var}(Y)}{E(Y)^2} \right].$$

From which it follows that

$$\mu_{\mathcal{R}} = \frac{E(a_r S_r)}{E(\sum_{k=r} a_k S_k)} - \frac{\text{Cov}(a_r S_r, \sum_{k=r} a_k S_k)}{E(\sum_{k=r} a_k S_k)^2} + \frac{E(a_r S_r) \text{Var}(\sum_{k=r} a_k S_k)}{E(\sum_{k=r} a_k S_k)^3}$$

$$= \frac{1}{(1 + s_1)} + \frac{\sigma_a^2 (s_2 - s_1)}{\mu_a^2 (1 + s_1)^3},$$

and

$$\sigma_{\mathcal{R}}^2 = \frac{E(a_r S_r)^2}{E(\sum_{k=r} a_k S_k)^2} \left[ \frac{\text{Var}(a_r S_r)}{E(a_r S_r)^2} - 2 \frac{\text{Cov}(a_r S_r, \sum_{k=r} a_k S_k)}{E(a_r S_r) E(\sum_{k=r} a_k S_k)} + \frac{\text{Var}(\sum_{k=r} a_k S_k)}{E(\sum_{k=r} a_k S_k)^2} \right]$$

$$= \frac{\sigma_a^2 (s_1^2 + s_2)}{\mu_a^2 (1 + s_1)^4}.$$

From these expressions it follows that:

$$\sigma_a^2 = \frac{(1 + s_1)^4 \mu_a^2 \sigma_{\mathcal{R}}^2}{(s_1^2 + s_2)} \quad (6)$$

$$\mu_{\mathcal{R}} = \frac{1}{(1 + s_1)} + \frac{(1 + s_1) (s_2 - s_1) \sigma_{\mathcal{R}}^2}{s_1^2 + s_2}. \quad (7)$$

From Equation (4) it follows immediately that  $\sigma_Q^2 = \left(\frac{\mu_Q}{\mu_a}\right)^2 \sigma_a^2$ , and Equation (5) can be written as

$$\sigma_Q^2 = \frac{(1 + s_1)^4 \sigma_{\mathcal{R}}^2}{s_0^2 (s_1^2 + s_2)} . \quad (8)$$

Finally, as  $s_1$  and  $s_2$  are functions of natural mortality, given estimates of  $\mu_{\mathcal{R}}$  and  $\sigma_{\mathcal{R}}^2$  derived from surveys, Equation (6) can be solved for natural mortality  $M$  and hence  $s_0$ ,  $s_1$  and  $s_2$ . Then, in turn,  $\mu_Q$  and  $\sigma_Q^2$  can be estimated from Equations (4) and (8). Mean recruitment  $\mu_{\mathcal{R}}$  cannot be estimated from  $\mathcal{R}(r)$ , and must be estimated independently.

In the De la Mare (1994) formulation,  $R$  must lie in the interval  $[0, 1)$ , and since there are limited distributions for a bounded  $[0, 1)$  range, the formulation focuses on the Beta distribution. The great advantage of the new formulation is that it provides a direct relationship between  $\mu_Q$  and  $\sigma_Q^2$ , and  $\mu_{\mathcal{R}}$  and  $\sigma_{\mathcal{R}}^2$ , and since the  $Q$  element is bounded by a  $[0, \infty)$  range, any family of distributions with support on the positive half line can be used to generate  $Q$  parameters including the Gamma, Log Normal, truncated Normal and inverse Beta families. In particular, the “delta” or “zero inflated” family (Aitchison 1955) can also be added to simulate cases of occasional recruitment failure. In this case,  $Q$  is computed as

$$Q = \begin{cases} 0 & u < p \\ Q' & u \geq p \end{cases} , \quad (9)$$

where  $u$  is a uniform deviate,  $p$  is the probability of occurrence of low recruitment, and  $Q'$  is again a random deviate from a distribution with support on the positive half line. Expressions for  $\mu_Q$  and  $\sigma_Q^2$  in terms of the mean and variance of  $Q'$  are given by Aitchison (1955).

To demonstrate the validity of their original method, De la Mare (1994) presents a table of simulations. We reproduce these results using the Beta, Gamma, Log Normal and Inverse Beta distributions (Tab 1). The derivation presented here produces very similar results to the De la Mare (1994) values, due to the low CV values chosen for the simulation. The simulations were conducted for 10 000 trials for each distribution for consistency with the De la Mare (1994) results.

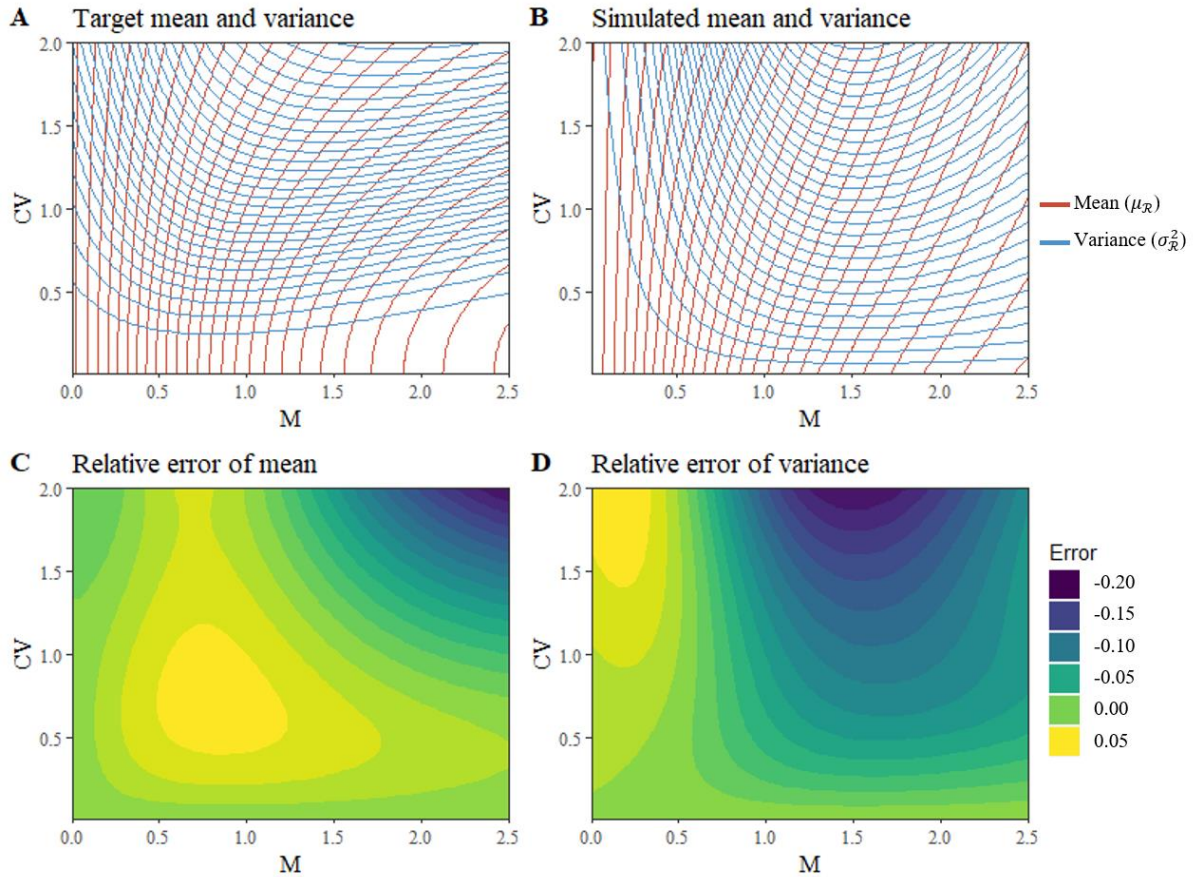
**Tab. 1** Exact outcomes of the 10 000 trials of each distribution for the four target parameters presented in De la Mare (1994).  $\hat{\mu}_{\mathcal{R}}$ ,  $\hat{\sigma}_{\mathcal{R}}^2$  and  $\widehat{CV}_{\mathcal{R}}$  represent the target values of  $\mathcal{R}$  parameters (highlighted in grey), and  $\mu_{\mathcal{R}}$ ,  $\sigma_{\mathcal{R}}^2$  and  $CV_{\mathcal{R}}$  represent the simulated values generated with the new formulation. The Beta distribution results are the same as the ones presented in De la Mare (1994), while the Gamma, Log Normal and Inverse Beta results are derived using the new formulation presented in this study. In this procedure, success is measured by the ability of the model to replicate the target  $\hat{\mu}_{\mathcal{R}}$ ,  $\hat{\sigma}_{\mathcal{R}}^2$  and  $\widehat{CV}_{\mathcal{R}}$  values

Distribution	$\hat{\mu}_{\mathcal{R}}$	$\mu_{\mathcal{R}}$	$\hat{\sigma}_{\mathcal{R}}^2$	$\sigma_{\mathcal{R}}^2$	$\widehat{CV}_{\mathcal{R}}$	$CV_{\mathcal{R}}$
Beta (de la Mare 1994)	0.3	0.300	0.00998	0.011	0.333	0.338
Gamma	0.3	0.300	0.00998	0.009	0.333	0.323
Log Normal	0.3	0.300	0.00998	0.009	0.333	0.310
Inverse Beta	0.3	0.300	0.00998	0.009	0.333	0.312
Beta (de la Mare 1994)	0.5	0.499	0.01	0.011	0.2	0.207
Gamma	0.5	0.499	0.01	0.010	0.2	0.195
Log Normal	0.5	0.500	0.01	0.009	0.2	0.188
Inverse Beta	0.5	0.500	0.01	0.009	0.2	0.186
Beta (de la Mare 1994)	0.5	0.496	0.02002	0.022	0.283	0.299
Gamma	0.5	0.500	0.02002	0.019	0.283	0.275
Log Normal	0.5	0.500	0.02002	0.016	0.283	0.249
Inverse Beta	0.5	0.500	0.02002	0.015	0.283	0.246
Beta (de la Mare 1994)	0.7	0.690	0.01002	0.012	0.143	0.158
Gamma	0.7	0.700	0.01002	0.011	0.143	0.149
Log Normal	0.7	0.700	0.01002	0.009	0.143	0.138
Inverse Beta	0.7	0.700	0.01002	0.009	0.143	0.132

These simulations show that when variance is low, any of these distributions are suitable for producing accurate results. However, both the method presented above and the original De la Mare (1994) formulation become increasingly inaccurate as  $\sigma_{\mathcal{R}}^2$  grows. Both methods rely on Taylor series approximations to the mean and variance of a ratio. But these approximations are only reliable for smaller variances, and so as  $\sigma_{\mathcal{R}}^2$  grows the expressions for  $\mu_Q$  and  $\sigma_Q^2$  become increasingly inaccurate



and the methods fail to reproduce the target  $\hat{\mu}_{\mathcal{R}}$  and  $\hat{\sigma}_{\mathcal{R}}^2$ . The nature of the errors in the approximations of the new formulation are presented in Figure 1. The relative error plots for  $\mu_{\mathcal{R}}$  and  $\sigma_{\mathcal{R}}^2$  show there are distinctive differences in the target and simulated mean and variance values in high CV and M zones; reflective of high variance scenarios.



**Fig. 1** Nature of the errors produced with the new recruitment derivation. A) Plot of the  $\mu_{\mathcal{R}}$  (red contours) and  $\sigma_{\mathcal{R}}^2$  (blue contours) of the target CV and M of recruitment. B) Plot of the simulated  $\mu_{\mathcal{R}}$  (red contours) and  $\sigma_{\mathcal{R}}^2$  (blue contours) using the recruitment derivation method. The relative errors are presented for the  $\mu_{\mathcal{R}}$  in plot C, and for the  $\sigma_{\mathcal{R}}^2$  in plot D. These high error regions (blue) are representative of areas where the simulated  $\mu_{\mathcal{R}}$  and  $\sigma_{\mathcal{R}}^2$  are not accurate reproductions of the target  $\mu_{\mathcal{R}}$  and  $\sigma_{\mathcal{R}}^2$ .

For large  $\sigma_{\mathcal{R}}^2$ , the required  $\mu_{\mathcal{Q}}$  and  $\sigma_{\mathcal{Q}}^2$  can be reliably determined by tabulation techniques. First note that natural mortality  $M$  determines the survivals  $S_k = S_k(M)$ , and in turn  $\mu_{\mathcal{Q}}$  through Equation (4); so, given  $M$  and  $\sigma_{\mathcal{Q}}^2$ , the corresponding  $\mu_{\mathcal{R}}$  and  $\sigma_{\mathcal{R}}^2$  can be determined to arbitrary accuracy by simulation. This process is used to tabulate the  $\mu_{\mathcal{Q}}$ ,  $\mu_{\mathcal{R}}$  and  $\sigma_{\mathcal{R}}^2$  corresponding to a range of  $M$  and  $\sigma_{\mathcal{Q}}^2$  parameters. Given a desired  $\mu_{\mathcal{R}}$  and  $\sigma_{\mathcal{R}}^2$ , the corresponding  $M$  (and hence  $\mu_{\mathcal{Q}}$ ) and  $\sigma_{\mathcal{Q}}^2$  are determined from the tabulated values by nearest neighbour interpolation. This tabulation method is straightforward,

(although computationally expensive) and will not be considered further due to the errors created in large variance scenarios.

De la Mare (1994) notes that the  $\hat{\mu}_{\mathcal{R}}$  and  $\hat{\sigma}_{\mathcal{R}}^2$  are estimates derived from a finite number of surveys and proposes resampling  $\hat{\mu}_{\mathcal{R}}$  from a Normal distribution and  $\hat{\sigma}_{\mathcal{R}}^2$  from a  $\chi^2$  distribution to account for sampling variation. However, as  $\mathcal{R}(r)$  is a proportion, its variance is bounded, and resampling and  $\hat{\sigma}_{\mathcal{R}}^2$  from a  $\chi^2$  distribution can lead to variances that are unrealizable. This is likely to be the cause of the failures reported by Kinzey et al. (2013) when simulating with variances drawn from empirical time-series that were larger than those used by De la Mare (1994). Maschette et al. (2020) proposed resampling  $\hat{\mu}_{\mathcal{R}}$  and  $\hat{\sigma}_{\mathcal{R}}^2$  by parametric bootstrap. Given observed  $\hat{\mu}_{\mathcal{R}}$  and  $\hat{\sigma}_{\mathcal{R}}^2$  derived from  $n$  field surveys, the methods described above are used to simulate  $n$  new surveys and proportions  $\mathcal{R}(r)$ , from which a new  $\hat{\mu}_{\mathcal{R}}$  and  $\hat{\sigma}_{\mathcal{R}}^2$  are estimated. The resampled  $\hat{\mu}_{\mathcal{R}}$  and  $\hat{\sigma}_{\mathcal{R}}^2$  produced in this way are guaranteed to be realizable.

As an alternative to the parametric bootstrap, we propose a constrained simulation method that resamples  $\mu_Q$  and  $\sigma_Q^2$  directly and overcomes the need for tabulation techniques when  $\sigma_Q^2$  is large. Similar to the tabulation method presented above, given  $M$  and  $\sigma_Q^2$  the method simulates  $n$  new surveys and proportions  $\mathcal{R}(r)$ , and computes the mean  $\tilde{\mu}_{\mathcal{R}}$  and variance  $\tilde{\sigma}_{\mathcal{R}}^2$  of the simulated  $\mathcal{R}(r)$ . But the sequence of stochastic fractions  $\{Q_i\}$  used in the simulation are generated by inverse probability transform; a sequence  $\{u_i\}$  of random uniform deviates  $u_i \sim U(0,1)$  is generated, and the sequence  $\{Q_i\}$  is calculated as

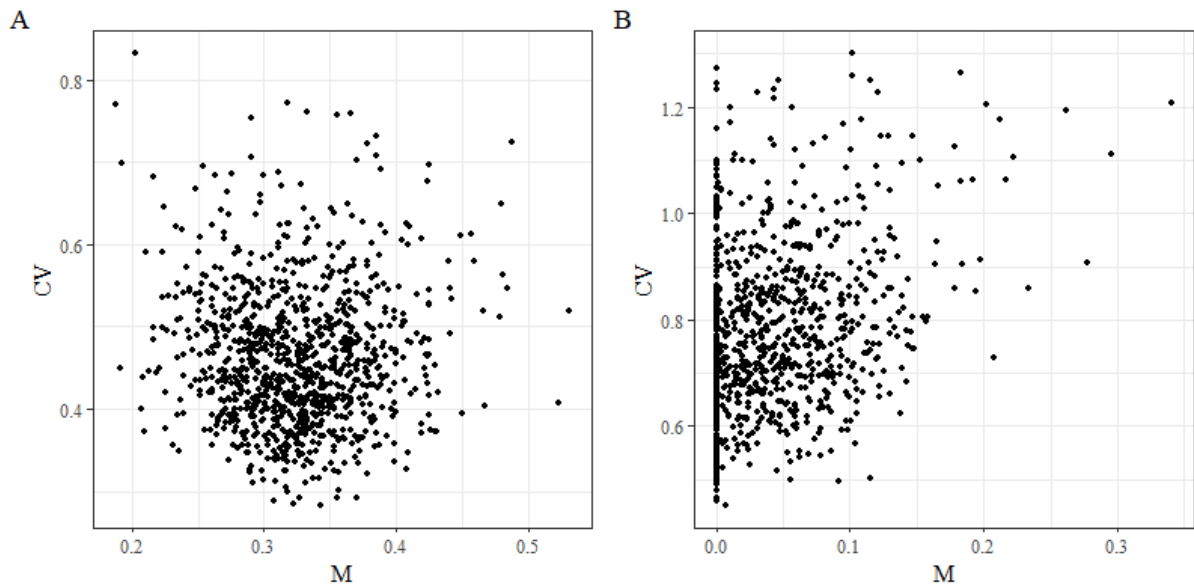
$$Q_i = F^{-1}(u_i; \mu_Q, \sigma_Q^2), \quad (10)$$

where  $F^{-1}(x; \mu_Q, \sigma_Q^2)$  is the inverse distribution function for a suitable distribution with mean  $\mu_Q$  and variance  $\sigma_Q^2$ . In this way, the discrepancy

$$D(M, \sigma_Q^2, \{u_i\}) = (\tilde{\mu}_{\mathcal{R}} - \hat{\mu}_{\mathcal{R}})^2 + (\tilde{\sigma}_{\mathcal{R}} - \hat{\sigma}_{\mathcal{R}})^2 \quad (11)$$

between the observed and simulated  $\mu_{\mathcal{R}}$  and  $\sigma_{\mathcal{R}}^2$  can be viewed as a continuous function of  $M$ ,  $\sigma_Q^2$ , and the sequence of uniform deviates  $\{u_i\}$ . For each resample, the sequence  $\{u_i\}$  is held fixed, and  $M$  and  $\sigma_Q^2$  are estimated by minimizing  $D(M, \sigma_Q^2, \{u_i\})$ . This determines an  $M$  (and hence  $\mu_Q$ ) and  $\sigma_Q^2$  for a given sequence of random deviates  $\{u_i\}$  reproduces the observed  $\hat{\mu}_{\mathcal{R}}$  and  $\hat{\sigma}_{\mathcal{R}}^2$  in simulation. Successive  $M$  and  $\sigma_Q^2$  are drawn by repeating the procedure with different sequences  $\{u_i\}$ . For some pathological sequences  $\{u_i\}$  it will not be possible to reduce the discrepancy  $D(M, \sigma_Q^2, \{u_i\})$  to an acceptable level; this will lead to a large number of values in the sequence  $\{u_i\}$  to produce  $M = 0$ . Such sequences should be discarded, and the process repeated for a new sequence. Figure 2 shows the difference between the sequences which should be used and the ones which should be discarded.

One advantage of this process is that correlated recruitment sequences can be generated by simulating correlated sequences  $\{u_i\}$  of uniform deviates with a Gaussian copula approach (Chung and Zhong 2001, Xiao 2016). The sequence  $\{u_i\}$  is generated by first generating a sequence  $\{z_i\}$  of correlated



**Fig. 2**  $M$  and  $CV$  of target  $\hat{\mu}_{\mathcal{R}}$  and  $\hat{\sigma}_{\mathcal{R}}^2$  values generated from a random sequence. A) Plot derived from parameters used in De la Mare (1994) tests for the Gamma distribution assuming target  $\hat{\mu}_{\mathcal{R}}$  are estimated from 17 surveys. B) Plot showing the discrepancies generated with parameter from a sequence derived from low  $\hat{\mu}_{\mathcal{R}}$  values. Failed sequences (B) should be discarded and a new one generated in their place

Normal deviates  $z_i \sim N(0,1)$  with the desired correlation structure. A probability integral transformation is then used to transform this sequence of correlated Normal deviates to a sequence  $\{u_i\}$  of correlated uniform deviates

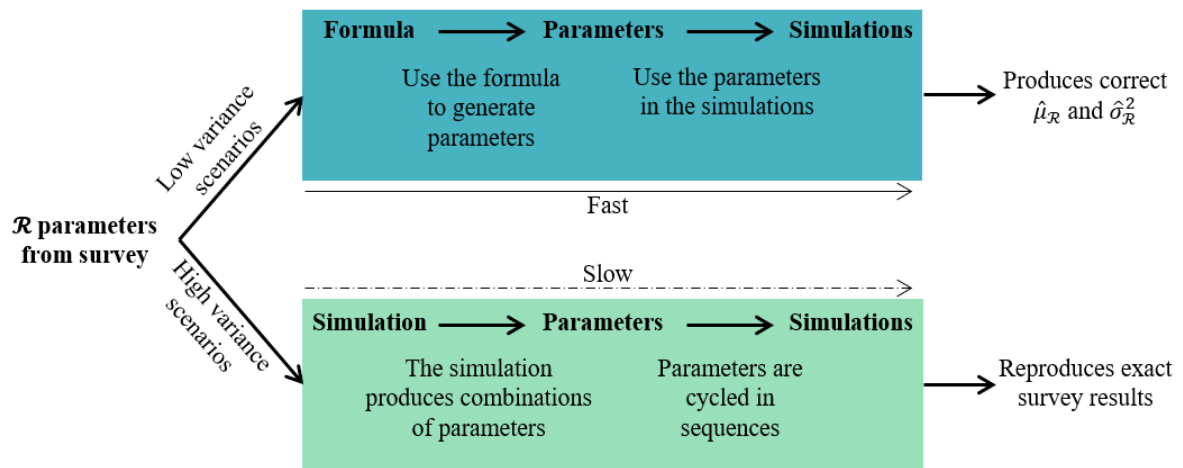
$$u_i = F(z_i), \quad (12)$$

where  $F$  is the distribution function of the standard Normal distribution. The  $\{u_i\}$  and the resulting recruitment sequence will have a similar correlation structure to the original  $\{z_i\}$ .

## DISCUSSION

Current environmental threats and future harvesting pressures are demanding flexibility from the GYM in terms of recruitment variability. Here we have developed a new flexible set of formulae that allow a greater range of distributions that facilitates diversity and capacity to model low recruitment events. We have improved upon the De la Mare (1994) model for cases with small variances and present a new broader procedure that operates with large variances.

For small variance, the testing of the new derivation showed success at reproducing  $\mu_R$  and  $\sigma_R^2$  values



**Fig. 3** Modelling processes of both methods presented in this study. The CV of  $\mathcal{R}(r)$  from the survey data defines which method should be used. For low variance scenarios, the formula method (presented in blue) should be used, as it gives fast and precise results. For high variance scenarios, the simulation method (presented in green) should be used. While the simulation method is very time costly, it is the only modelling process that accurately reproduces the survey parameters in high variance scenarios

within a limited range of values. Furthermore, the new derivation gives expressions for inverse Beta, Log Normal and Gamma distributions. At low variability, all three distributions produce broadly the same results, but the exact range depends on the higher order moments of the distributions. This difference in expressions means that subtle differences in the distributions will dictate the range of possible  $\mu_R$  and  $\sigma_R^2$  combinations, giving the new derivation a marginally wider coverage area. However, for extreme variances, derived from low recruitment events, Taylor series approximations fail using both the De la Mare (1994) recruitment model and the new derivation. For somewhat larger variances, the new derivation can be paired with the delta expressions with varying zero proportions to extend the coverage zones into higher parameter ranges. However, as the variance continues to increase, the

delta expressions will also fail. This occurs as Taylor delta methods rely on linear approximations to a non-linear relationship, and so simulations remain limited to certain parameters. High variance combinations will continue to fall outside of the plausible range and the simulations will continue to fail. De la Mare's (1994) recruitment model and the new derivation with the delta expressions fail to reproduce the desired parameters at high variance and are therefore not suitable for extreme scenarios.

The purpose of the GYM is to project possible future population scenarios so that sustainable harvesting levels can be established. If models are not capable of efficiently reproducing real life scenarios such as the recurrence of low recruitment events, then they are ineffective for fishery management purposes. Recruitment modelling needs to be more flexible as well as time efficient so it can efficiently deal with as broad a range of scenarios as possible, including extreme events that produce large CVs.

The new constrained simulation method resamples  $\mu_Q$  and  $\sigma_Q^2$  given  $M$  and  $\sigma_Q^2$ , and generates a random sequence of  $n$  new surveys and proportions  $\mathcal{R}(r)$ , so that the mean  $\tilde{\mu}_{\mathcal{R}}$  and variance  $\tilde{\sigma}_{\mathcal{R}}^2$  of the simulated  $\mathcal{R}(r)$  are an exact replica of the parameters extracted from survey data. Like the bootstrap method proposed by Maschette et al. (2020), whenever you run a constrained simulation, the sequence of random deviates must be maintained. It must be emphasized that when the  $M$  is set too low, the simulation will generate some pathological sequences  $\{u_i\}$  that do not accurately represent the desired  $M$  and  $CV$ , we suggest discarding such sequences and repeating the process with a new sequence. If the process is repeated and the new sequence  $\{u_i\}$  produces the same discrepancies, it could be indicative of an error in the original  $\mathcal{R}(r)$  from the surveys. Such situations could arise from the unlikely event that all surveys used to generate  $\hat{\mu}_{\mathcal{R}}$  and  $\hat{\sigma}_{\mathcal{R}}^2$  parameters sampled a total of zero recruits. While this scenario is highly unlikely, it should be emphasized that failed recruitment in the surveys is a general indication of the scarcity of recruits in the samples rather than a complete extinction of an age group in the population. There are still a number of uncertainties about krill recruitment failure and its periodicity for modelling. And, while population stability is regulated by the balance of krill survival ( $M$ ) and the rate at which new individuals are recruited, when recruitment is low, such as for events of failed recruitment,  $M$  is reflective of this, otherwise the population would decrease considerably in the models.

## CONCLUSION

Management of the Southern Ocean fisheries, especially the krill fishery, depend on the GYM and its flexibility to model recruitment considering extreme scenarios such as years of failed recruitment. The results of this study present a new derivation of De la Mare's (1994) recruitment function and an alternative derivation that supports recruitment failure. For low  $CV$  values, the new derivation is a fast

and precise tool that only requires simple calculations to estimate parameters of  $R$ . For high CV values, we developed a constrained simulation method which can generate results beyond the range of possible  $\mu_R$  and  $\sigma_R^2$  combinations in both De la Mare's (1994) recruitment model and the new derivation. Overall, the new derivation as well as the constrained simulation method can be used in the GYM to produce more reliable stock assessments. These adjustments to the GYM could have applications for large scale fisheries such as the krill fishery, as well as smaller scale fisheries with large variations in recruitment.

## REFERENCES

- Aitchison, J. 1955. On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of the American Statistical Association* **50**: 901-908.
- CCAMLR. 2020. Krill fisheries: Catch history. The Commission for the Conservation of Antarctic Marine Living Resources.
- Chung, K. L., and K. Zhong. 2001. *A course in probability theory*. Academic press.
- Constable, A., and W. De la Mare. 1996. A generalised model for evaluating yield and the long-term status of fish stocks under conditions of uncertainty. *CCAMLR Science* **3**: 31-54.
- Constable, A. J., and W. de la Mare. 2003. Generalised Yield Model, version 5.01b. Australian Antarctic Division, Kingston, Australia.
- Constable, A. J., and S. Kawaguchi. 2018. Modelling growth and reproduction of Antarctic krill, *Euphausia superba*, based on temperature, food and resource allocation amongst life history functions. *ICES Journal of Marine Science* **75**: 738-750.
- Cox, M. J., D. L. Borchers, D. A. Demer, G. R. Cutter, and A. S. Brierley. 2011. Estimating the density of Antarctic krill (*Euphausia superba*) from multi-beam echo-sounder observations using distance sampling methods. *Journal of the Royal Statistical Society Series C: Applied Statistics* **60**: 301-316.
- De la Mare, W. 1994. Modelling krill recruitment. *CCAMLR Science* **1**: 49-54.
- Kawaguchi, S., and S. Nicol. 2007. Learning about Antarctic krill from the fishery. *Antarctic Science* **19**: 219-230.
- Kawaguchi, S., S. Nicol, and A. J. Press. 2009. Direct effects of climate change on the Antarctic krill fishery. *Fisheries Management and Ecology* **16**: 424-427.
- Kinzey, D., G. Watters, and C. S. Reiss. 2013. Effects of recruitment variability and natural mortality on generalised yield model projections and the CCAMLR decision rules for Antarctic krill. *CCAMLR Science* **20**: 81-96.
- Krag, L. A., B. A. Krafft, A. Engås, and B. Herrmann. 2018. Collecting size-selectivity data for Antarctic krill (*Euphausia superba*) with a trawl independent towing rig. *PLoS ONE* **13**: e0202027.

Maschette, D., S. Wotherspoon, C. Pavez, P. Ziegler, S. Thanassekos, K. Reid, S. Kawaguchi, D. Welsford, and A. Constable. 2020. Generalised R Yield Model (Grym). Document SC-CAMLR-39/BG/19. CCAMLR, Hobart, Australia.

Meyer, B., A. Atkinson, K. S. Bernard, A. S. Brierley, R. Driscoll, S. L. Hill, E. Marschoff, D. Maschette, F. A. Perry, C. S. Reiss, E. Rombolá, G. A. Tarling, S. E. Thorpe, P. N. Trathan, G. Zhu, and S. Kawaguchi. 2020. Successful ecosystem-based management of Antarctic krill should address uncertainties in krill recruitment, behaviour and ecological adaptation. *Communications Earth & Environment* **1**: 1-12.

Miller, D. 2011. Sustainable Management in the Southern Ocean: CCAMLR Science. Smithsonian Inst Scholarly Press, Washington.

Quetin, L. B., and R. M. Ross. 2003. Episodic recruitment in Antarctic krill *Euphausia superba* in the Palmer LTER study region. *Marine Ecology Progress Series* **259**: 185-200.

Reiss, C. S., A. M. Cossio, V. Loeb, and D. A. Demer. 2008. Variations in the biomass of Antarctic krill (*Euphausia superba*) around the South Shetland Islands, 1996-2006. *ICES Journal of Marine Science* **65**: 497-508.

SC-CAMLR. 2019. Report of the thirty-eighth meeting of the Scientific Committee (SC-CAMLR-38). CCAMLR, Hobart, Australia.

Siegel, V. 2005. Distribution and population dynamics of *Euphausia superba*: Summary of recent findings. *Polar Biology* **29**: 1-22.

Siegel, V. 2016. *Biology and Ecology of Antarctic Krill*. Springer, Cham, Switzerland.

Siegel, V., and V. Loeb. 1995. Recruitment of Antarctic krill *Euphausia superba* and possible causes for its variability. *Marine Ecology Progress Series* **123**: 45-56.

Stuart, A., and K. Ord. 1998. *Kendall's advanced theory of statistics, distribution theory*. 6th edition. Oxford University Press, New York.

Wotherspoon, S., and D. Maschette. 2020. Grym: R implementation of the generalized yield model. R package version 0.1.0.9000. [github.com/AustralianAntarcticDivision/Grym](https://github.com/AustralianAntarcticDivision/Grym).

Xiao, Q. 2016. Calculating correlation coefficient for Gaussian copula. arXiv e-prints:arXiv: 1608.00738.

## SUPPLEMENTARY MATERIAL

Tab. 1 List of equation terms

Model parameters	Description
$\mathcal{R}$	Proportion of recruits observed in survey data
$\hat{\mu}_{\mathcal{R}}$	Mean of the proportion of recruits observed in survey data
$\hat{\sigma}_{\mathcal{R}}^2$	Variance of the proportion of recruits observed in survey data
$\widehat{CV}_{\mathcal{R}}$	Coefficient of variation observed in survey data
$\mu_{\mathcal{R}}$	Simulated mean of the proportion of recruits
$\sigma_{\mathcal{R}}^2$	Simulated variance of the proportion of recruits
$CV_{\mathcal{R}}$	Simulated coefficient of variation
r	The first age class observed in the survey data
$A$	The number of individuals in a given age class observed in the survey data
$A_k$	The number of individuals in every age class observed in the survey data
$R$	Random variable used in the stochastic fraction
$A_1$	Number of recruits to the first age class
$Q$	Random variable that replaces the stochastic fraction
$\mu_Q$	Mean of $Q$
$\sigma_Q^2$	Variance of $Q$
$a$	The number of individuals originally recruited to an age class
$\mu_a$	Mean of $a$
$\sigma_a^2$	Variance of $a$
$S_k$	Survival from recruitment to the observed age class in the survey data
M	Instantaneous mortality



$\boldsymbol{Q}'$	A random variable that supports the Delta expression
$\boldsymbol{u}$	Uniform deviate used in the Delta expression
$\boldsymbol{p}$	Probability of occurrence of low recruitment used in the Delta expression
$\tilde{\boldsymbol{\mu}}_{\mathcal{R}}$	Simulated mean of the simulated $\mathcal{R}$ in the bootstrap method
$\tilde{\boldsymbol{\sigma}}_{\mathcal{R}}^2$	Simulated variance of the simulated $\mathcal{R}$ in the bootstrap method
$\boldsymbol{u}_i$	Sequence of uniform deviates for the constrained simulation method
$\boldsymbol{z}_i$	Sequence of correlated normal deviates for the constrained simulation method