

# Models of inter-moult period for Antarctic krill - Lack of progress and promulgation of unreliable models calibrated using indirect observations

---

## ABSTRACT

**Aims:** Two methods of calibrating regression models of inter-moult period (IMP) as a function of temperature exposure ( $T$ ) for crustaceans, in particular, Antarctic krill (*Euphausia superba*) are reviewed in terms of both theoretical and empirical properties in order to make recommendations on the application of the methods and/or the use of the resultant fitted models.

**Methodology:** The method and fitted model that used a meta-analysis of published results from laboratory-reared krill using means of directly observed IMP for a range of controlled, constant temperature regimes has valid theoretical and empirical support. The alternative used moult frequencies obtained as a “byproduct” of 5-d Instantaneous Growth Rate (IGR) experiments carried out at sea for which individual IMPs were not directly observed. Instead mean IMP given  $T$  and animal total length ( $L$ ) was predicted using the moult frequencies disaggregated to binary data of moulted versus not-moulted as dependent variable in the calibration of a logistic regression on  $T$  and  $L$ . The shape of the daily development rate,  $R$ , the inverse of IMP, versus  $T$  response curve fitted using direct observations is a classical monotonically increasing curve whereas for combinations of sex/maturity classes the curves fitted using indirect observations are parabola-like with sexually dimorphic concavities of either up or down. Four sources of bias in predictions of mean IMP using the indirect observations and estimation method are described. One source due to an unrepresentative sampling frame can lead to large positive bias in estimated mean IMP based on theory which has been absent until now and that applies a discrete uniform distribution for next moult date corresponding to ideal asynchrony. This bias and that due to IGR experimental measurement error in  $T$  cannot be remedied.

**Conclusion:** The indirect method and the corresponding fitted models are unreliable and

should not be used.

*Keywords: inter-moult period, laboratory experiments, at-sea IGR experiments, models, sampling methods*

## 1. INTRODUCTION

Inter-moult period (IMP) is as the name suggests the time between consecutive moults which involve exuviation of the exoskeleton for an individual of a species of crustacean. It is analogous to the development time for a consecutive stage in insect development such as that marked by the period between consecutive sheddings of the head capsule of larvae. Kawaguchi et al. (2006) [1] calibrated their model for adult Antarctic krill (*Euphausia superba*) (combined sex) for IMP as a function of temperature ( $T$ ) using a meta-analysis of published mean IMPs from studies that directly observed IMPs for individual, laboratory-reared, Antarctic krill using observed times between consecutive collections of exuviae for each animal maintained in its own experimental vessel under known constant temperature regimes which included ancillary light and feeding regimes. The latest of the studies used in the meta-analysis was published in 1991 [2]. The model of [1] for IMP considered only adults as did [3] who carried out more recent experiments in the Australian Antarctic Division's specialised krill aquarium using three temperature regimes of  $-1^{\circ}\text{C}$ ,  $1^{\circ}\text{C}$  and  $3^{\circ}\text{C}$  and observations of individual IMP and daily growth rate for males and females in maturity stages 3 to 5.

In Tarling et al. (2006) [4] the data used to calibrate their model of IMP as a joint function of  $T$  and animal total length ( $L$ ) were moult frequencies obtained as a "byproduct" of 5-d Instantaneous Growth Rate (IGR) experiments carried out at sea where individual IMPs were not directly observed and where moult frequencies were disaggregated to the individual animal level to give binary data of moulted versus not-moulted data recorded along with  $L$  and  $T$  where  $T$  was taken as the sea surface temperature at time of capture. These binary data with covariates of  $L$  and  $T$  were fitted using logistic regression.

Candy and Kawaguchi (2006) [5] used the model of [1] for IMP as a function of temperature combined with a model of seasonal average sea surface temperature (SST) and the models given by [1] for instantaneous growth rate (IGR) within an algorithm to predict average length at age for ages up to 6 years. Wiedenmann et al. (2008) [6] refined the predictive algorithm of [5] by making some very simple assumptions about movement through the water column using the proportion of the day that is spent in "deep" versus and "shallow" habitat using the monthly averages of temperature for these coarsely defined habitats in the regions around South Georgia. They also used the IGR model of [7] rather than that of [1] but used the model of [1] for IMP.

The models of IMP as a function of temperature and total length for separate sex and maturity classes of [4] were applied by [7] and more recently by [8] so that the models of [1] and [4] are clearly of an equally aged provenance. The first point to note is that these empirical models of IMP have not been updated or replaced since they were published in 2006. Therefore, it is reasonable to conclude that there has been a lack of progress since

2006 in the development of models of IMP which is hard to explain for such an ecologically and economically important species.

Wiedenmann et al. (2008) and [8] both noted the substantial difference in shape of the fitted temperature response curves between males and females for IMP in [4] and that both contradict the fundamentally different exponentially decreasing trend in IMP with increasing temperature in [1] for combined males and females. This is seen in Figure S3 of [8] where the curve shape in [4] for females is parabola-like and concave-up whereas Figure 8 of [4] shows this last trend along with a concave-down shape for males with the shapes for both sexes becoming more pronounced with increasing length. Wiedenmann et al. (2008) state that they explored both IMP models of [1] and [4] but chose the model of [1] for its “simplicity” but did not describe the outcome of their “exploration”. Bahlburg et al. (2023) note IMPs predicted by both models can be very different, especially at temperatures  $<1^{\circ}\text{C}$  or  $>3^{\circ}\text{C}$ .

Neither [6] nor [8] mention the fundamental difference in the respective datasets used for model calibration and its possible implications for growth projection whereas the two types of data were briefly mentioned by [3]. Bahlburg et al. (2023) only note the different coastal and oceanic sectors of Antarctica that the respective sample specimens studied were sourced for laboratory experiments ([1]) (i.e. East Antarctica) versus in situ studies (i.e. at sea) ([4]) (i.e. Atlantic sector of the Southern Ocean) and used for model calibration. However, this is not strictly correct since the meta-analysis of laboratory studies used by [1] were not exclusively based on observations from the Indian sector of the Southern Ocean. For example, the captive-reared krill experiments described in [2] were carried out “at the Polish Antarctic Station 'H. Arctowski' on King George Island, South Shetland Islands”.

Potential sources of these substantial differences in shape of the IMP response as a function of temperature for these two sets of models are investigated with respect to the fundamental difference in the respective datasets used for model calibration. Further, the theoretical underpinning of models calibrated using observed development times (analogous to IMP) is well known ([9],[10]) but such theory for the method of [4] of estimation of average IMP from moult frequencies without observation of actual sample IMPs was not given or referenced in [4] nor has been presented in subsequent literature. Four sources of bias in the predictions of mean IMP using the models of [4], that are avoided by using the models, estimation methods and type of data used by [1], are described. In investigating one of these sources a theory is presented that employs some specific assumptions on the distribution of next moult date that infer ideal asynchrony in moulting and derives a putative equivalence of probability of moulting obtained from IGR experiments with daily development rate based on directly observed IMPs.

Investigation of potential bias in the model predictions of [4] also further considers their empirical properties. To this end the development rate predictions across temperature values, given a fixed total length in the case of [4], are obtained from the published equations from [1] and [4] and graphically compared along with the meta-analytic IMP sample means used by [1].

## **2. METHODS**

### **2.1 Datasets and statistical methods used to calibrate models of IMP**

#### **2.1.1 Estimation using direct observations of IMP**

The nature of the data used by [4] compared to [1] to calibrate their models of IMP are fundamentally different. As noted earlier [1] calibrated their IMP model using a meta-analysis of published mean IMPs from directly observed IMPs for individual, laboratory-reared, krill using observed times between consecutive collections of exuviae for individual krill maintained each in its own experimental vessel under known constant temperature regimes which included ancillary light and feeding regimes. For example, the experiments of [11] considered three temperature regimes of 0.12°C, 0.97°C, and 4.48°C while [2] employed a single temperature regime of  $2.0 \pm 0.5^\circ\text{C}$ .

### **2.1.2 Estimation using indirect “observations” of IMP; moult counts or presence/absence of moults from at-sea IGR experiments**

In contrast the data used by [4] were from 51, 5-d IGR experiments carried out at sea where individual IMPs were not directly observed. The standard method [4] refer to (i.e. they denoted this the “1/MR” method) involves inferring from the frequency of moults in a 5-day incubation period the mean IMP using the formula  $\text{IMP} = Nd/m$  where  $N$  is the total sample of krill each incubated in individual 500-mL perforated polycarbonate jars ([4]),  $m$  is the number observed moulting during the incubation period of  $d$  days (i.e. standardised to 5-d) and this is applied to each IGR experiment. Note that  $m/N$  is the usual unbiased estimate of the probability,  $P(d)$ , of molting in the incubation period given the sum of  $N$  Bernoulli trials.

The regression method of [4] (they denote as the “IMP” method) uses the disaggregated binary or Bernoulli trial data (i.e. moult vs not-moult) for the complete set of IGR experiments fitted as a binomial/logistic model and so disaggregates the “1/MR” formula to individual krill with predictor variables of total length ( $L$ ) and temperature ( $T$ ) with the regression fitted separately for each sex/maturity class. Tarling et al. (2006) note problems with the “1/MR” method of obtaining IMP estimates of (i) empirical results giving an unreasonably wide range of estimated IMPs (ii) in theory, synchrony of molting could result in biased estimates of IMP, however, they claim that their “IMP” method overcomes these problems by aggregating across synchronous and asynchronous experimental cohorts.

It is important to note that [4] gave no theoretical justification for their “1/MR” method or its extension to their “IMP” method for estimating mean IMP from moult frequencies for a sample of captive krill using IGR-experimental incubations. In particular, the proportion of development per day for the hypothetical individual of average IMP, expressed as the inverse of that average IMP (within potential determinates such sex, maturity, and/or total length categories) is theoretically interpreted as a rate of development in units of proportion per day summing to unity from day 1 up to the day-equivalent to the average IMP for each temperature regime in constant temperature laboratory studies. This can be generalised to fluctuating temperatures in the field ([5],[9]). This is well supported theoretically ([10]) and empirically ([12]) for observed development times. Tarling et al. (2006), in their “1/MR” and “IMP” methods in effect equate a  $P^*$ , which is the conversion of  $P(d)$  to a daily value by dividing it by  $d$ , to the above daily rate but this equivalence has not been demonstrated in the literature either theoretically or empirically using simulation. This equivalence is established theoretically in the Appendix but only under some very restrictive assumptions combined with the mathematically correct calculation of  $P^*$  from  $P(d)$  as given later. Even if this equivalence is accepted more generally, it remains to also demonstrate whether the “1/MR” and “IMP” methods give either unbiased or biased estimates of mean IMP. Examination of such bias is given in Section 2.3.

## **2.2 Empirical Results: comparing development rate response curves**

To understand the difference between the IMP models of [1] and [4], Figure 1 shows the predictions of rate of development (i.e. reciprocal of IMP, see Equation 1 of [10]) as a function of temperature for adult krill, with the model of [4] predictions using a fixed total length ( $L$ ) of 55 mm. Figure 1 also overlays mean rates calculated using the meta-data used in [1], the reciprocal of each study mean IMP as extracted from Figure 5 of [1] using the PlotDigitizer software (<https://plotdigitizer.com/app>), and mean IMPs averaged over length and sex classes for the 1°C and 3°C regimes that were visually extracted from Figure 3 of [3] giving approximate values of 23 d and 18.5 d, respectively. Figure 1 also shows approximate 95% confidence bounds for the equivalent rate curve based on the corresponding approximate 95% confidence bounds of the model of IMP of [1] using digitized data of the upper bound shown in Figure 5 of [1] combined with calculating an approximate standard error of rate predictions using a first order Taylor series approximation to the variance of  $\log_{10}(\text{IMP})$  predictions.

## 2.3 Sources of bias in Tarling et al IMP model estimation

### 2.3.1 Mathematical error in converting probability of moult from a 5-d experimental incubation period to a 1-d period equivalent

For the [4] “1/MR” model/method it was noted earlier that for an IGR experiment the estimated mean IMP,  $\tilde{t}_m$ , is calculated as  $Nd/m$  [Equation (3) in [4]], where  $d$  is the experiment’s fixed incubation period, and express this as  $D=\tilde{t}_m=1/(P/d)$  where  $P=m/N$  (i.e. the estimate of the binomial probability from  $N$  Bernoulli trials with  $m$  “successes” or moults).

The extension of the “IMP” model/method of [4] gives the estimate of mean IMP,  $\tilde{t}_m(\mathbf{x})$ , given a vector of predictor variables  $\mathbf{x}$  and estimated regression parameters  $\hat{\boldsymbol{\alpha}}$  and incubation period  $d$  as  $\tilde{t}_m(\mathbf{x}) = d / P(\mathbf{x}'\hat{\boldsymbol{\alpha}}, d)$  [corresponding to Equation (6) of [4]] where the logistic regression model fitted to the binary data (i.e. moult vs non-moult during 5-day incubation period,  $d=5$ ) is given by

$$P(\mathbf{x}'\boldsymbol{\alpha}, d) = \frac{\exp\{\alpha_0 + \alpha_1 L + \alpha_2 T + \alpha_3 T^2\}}{1 + \exp\{\alpha_0 + \alpha_1 L + \alpha_2 T + \alpha_3 T^2\}}$$

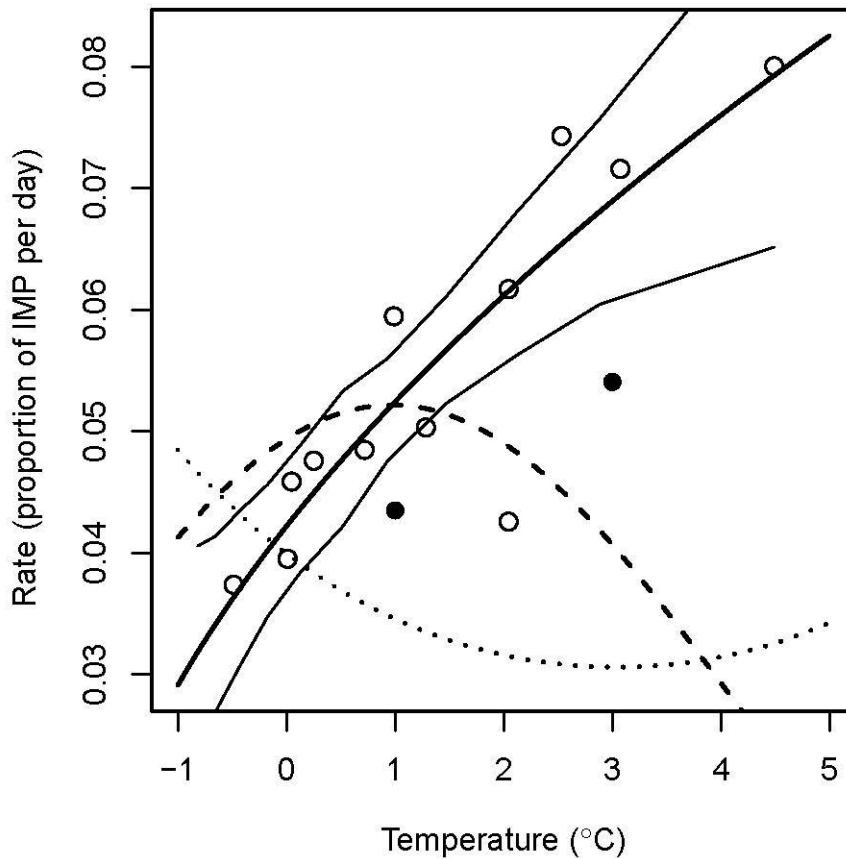
[i.e. Equation (7) of [4]] so that  $\mathbf{x}' = (1, L, T, T^2)$ .

In effect [4] in their Equation (6) take the prediction of  $P(\mathbf{x}'\boldsymbol{\alpha}, d)$  for the multi-day standard incubation period  $d$  and simply divide it by  $d$  to obtain a probability of moult for an individual day (i.e. a conversion of standardization from  $d$  days to 1-d). As described earlier, this 1-d standardized probability

$$P^*(\mathbf{x}'\boldsymbol{\alpha}) = P(\mathbf{x}'\boldsymbol{\alpha}, d) / d$$

is inferred by [4] to be a daily development rate fraction ( $R$ ) so that its reciprocal can be inferred as the development time  $D=1/R$  (i.e. IMP). However, this conversion by simply scaling by  $d$  is incorrect as seen when basic probability theory is applied. This conversion can be substantially in error and thus cause substantial positive bias in the estimate of IMP

with this bias increasing with increasing  $P^*(\mathbf{x}'\boldsymbol{\alpha})$ . This can be seen, after dropping dependence on  $\mathbf{x}$  and  $\mathbf{x}'\boldsymbol{\alpha}$  to simplify notation, as follows.



**Figure 1. Predicted rate curves from IMP.** Models of [1] (adults, heavy solid line; approximate 95% confidence bounds, fine solid lines) and [4] (adult female, dashed line; adult male, dotted line each for  $L$  of 55 mm) versus Temperature ( $T$ ). Study mean rates calculated using the meta-data used in [1] (open circles) were obtained as the reciprocal of each study mean IMP with these extracted from Figure 5 of [1] using the PlotDigitizer software (<https://plotdigitizer.com/app>) along with 95% bounds on the estimated rate curve and mean IMPs averaged over length and sex classes for the 1°C and 3°C regimes as visually extracted from Figure 3 of Brown et al. (2010) (filled circles).

For a period of  $d$ -days incubation, consider that no moult is observed for a random individual, the probability of this outcome is  $[1 - P(d)]$  and this corresponds to the probability that for each day in the period no moult occurs so that

$$[1 - P(d)] = (1 - P^*)^d .$$

Then it follows that the probability of observing a moult in the  $d$ -days period as given by  $P(d)$  converts to the 1-d period using the formula  $P^* = 1 - [1 - P(d)]^{1/d}$  and not the formula given by [4][their Equations (3) and (6)] corresponding to  $P^* = P(d)/d$ . The bias induced by the [4] method can be more clearly seen when we expand the correct formula using a  $d$  of 5-d as applied in the IGR experiments. This can be seen as

$$P(5) = 1 - (1 - P^*)^5 = P^* \left[ 1 + (1 - P^*) + (1 - P^*)^2 + (1 - P^*)^3 + (1 - P^*)^4 \right].$$

Therefore for  $P^*$  very small, say 0.01, the term in square brackets is close to 5 since each term after the "1" is close to 1 giving a total value of this term of 4.9, so that  $P^*_{\text{TO6}} = P(d)/d$  would be slightly less but close to the correct value  $P^* = 1 - [1 - P(d)]^{1/d}$ . However, given the inference of  $t_m = 1/P^*$  and a range of IMP of, for example 30-d to 10-d, the corresponding values for the above term are 4.678 and 4.095, respectively. The corresponding bias of the [4] formula for the estimate of IMP given true values,  $t_m$ , of 30-d and 10-d are calculated using the formula

$$\%bias = 100 \left[ d \left\{ 1 - [1 - P(d)]^{1/d} \right\} / P(d) - 1 \right] \text{ where } P(d) = 1 - (1 - P^*)^d$$

which give percent mathematical bias or error (as separate from bias due to transformation of a sample statistic, see below) values of 6.9% and 22.1% for  $d$  of 5 for the population mean IMP,  $t_m$ , given  $P^*$ , respectively. The above assumes, as in [4], that the IGR incubation period  $d$  is short enough that there is no chance for an individual animal to moult more than once in that period. Given that the standard value of  $d$  was 5-d this is a reasonable assumption.

### 2.3.2 Statistical bias due to inverting the estimate of $P^*$ to estimate mean IMP

The sample estimate of the proportion of moults out of a sample of  $N$  incubated krill in  $d$  days given by  $\hat{P}(d) = m/N$  has, from the above, an expected value

$P(d) = 1 - (1 - P^*)^d$  where under the discrete uniform distribution for next moult date for the population sampled by an IGR experiment has a probability of a randomly selected animal moulting on any given day between the distributional limits for next moult date of  $P^*$  (Appendix). The sample estimate of  $P^*$  is given by  $\hat{P}^* = 1 - [1 - \hat{P}(d)]^{1/d}$  while the (biased) estimate of [4] is given by  $\tilde{P}^* = \hat{P}(d) / d$ . An estimate of mean IMP for an IGR sample is given by either  $\hat{t}_m = 1 / \hat{P}^*$  or, using [4], as  $\tilde{t}_m = 1 / \tilde{P}^*$ . The respective transformations of  $\hat{P}(d)$  to an estimate for  $t_m$  each impart bias to their respective estimates of  $t_m$ . The expected value of the [4] estimate,  $\tilde{t}_m$ , is using a second-order Taylor series expansion approximately  $E(\tilde{t}_m) \{1 + \sigma_p^2 E(\tilde{t}_m)^2 / d^2\}$  where  $\sigma_p^2$  is the usual binomial variance formula,  $\sigma_p^2 = \hat{P}(d) \{1 - \hat{P}(d)\} / N$ , and  $E(\tilde{t}_m)$  is not equal to  $t_m$  given the results of the above section. Therefore, including bias due to using the incorrect formula for estimating  $P^*$ , if  $N = 100, d = 5, P(d) = 1/6, t_m = 30$  then the expected value of  $\tilde{t}_m$  is  $1.069 * 1.057 t_m$  corresponding to a combined 13% over-estimation of population mean  $t_m$ .

For the IMP model of [4] and using their corresponding estimate of mean IMP,  $\tilde{t}_m(\mathbf{x}) = d / P(\mathbf{x}'\hat{\boldsymbol{\alpha}}, d)$ , generalising the above the transformation-bias adjusted estimate of

$$\tilde{t}_m(\mathbf{x}) \text{ from the above gives } \tilde{t}'_m(\mathbf{x}) = \tilde{t}_m(\mathbf{x}) \left\{ 1 + \hat{\sigma}_{\tilde{t}_m(\mathbf{x})}^2 \tilde{t}_m^2(\mathbf{x}) / d^2 \right\}^{-1}$$

so that, to a close approximation, the estimate is positively biased as a percentage by the amount  $B = 100 \hat{\sigma}_{\tilde{t}_m(\mathbf{x})}^2 \tilde{t}_m^2(\mathbf{x}) / d^2$  excluding the bias induced by not using the correct formula for the estimate of mean IMP of  $\hat{t}_m(\mathbf{x}) = 1 - [1 - P(\mathbf{x}'\hat{\boldsymbol{\alpha}}, d)]^{1/d}$ . The expression for  $\hat{\sigma}_{\tilde{t}_m(\mathbf{x})}^2$  is given by

$$\hat{\sigma}_{\tilde{t}_m(\mathbf{x})}^2 = \mathbf{x}' \Sigma_{\hat{\boldsymbol{\alpha}}} \mathbf{x} \left[ \frac{dh^{-1}(\eta)}{d\eta} \right]^2 \text{ where } \Sigma_{\hat{\boldsymbol{\alpha}}} \text{ is the variance-covariance matrix of } \hat{\boldsymbol{\alpha}}, h(\cdot) \text{ is the}$$

logistic link function, and the linear predictor is  $\eta = \mathbf{x}'\boldsymbol{\alpha}$ . Although for reasonable sample sizes, if the "1/MR" method is used (i.e.  $\mathbf{x}' = 1, \boldsymbol{\alpha} = \alpha_0$ ) where for the above example of  $N$  taking a value of 100 this gives  $B$  as 6%, then the transformation bias is relatively small. However, this may not be the case for  $\tilde{t}_m(\mathbf{x})$  due to greater uncertainty for upper and lower ranges of combinations of  $L$  and  $T$  due to model prediction error, due to the term  $\mathbf{x}' \Sigma_{\hat{\boldsymbol{\alpha}}} \mathbf{x}$ , where this could be further exacerbated by subdividing the data into separate sex/maturity classes. It is not possible to investigate this bias without the estimate of  $\Sigma_{\hat{\boldsymbol{\alpha}}}$  which was not made available in [4]. Tarling et al. [4] failed to follow even standard reporting by not

presenting the standard errors of parameter estimates (i.e. square root of the diagonal elements of  $\hat{\Sigma}_{\alpha}$ ) from the fit of their model (7).

### 2.3.3 Unrepresentative sampling frame leading to positive bias in IMP estimates

In the Introduction it was noted that a theory for the method of [4] of estimation of average IMP from moult frequencies without observation of actual sample IMPs was not given or referenced in [4] nor has been presented in subsequent literature. A theory is presented in the Appendix that employs some specific assumptions on the distribution of next moult date that corresponds to ideal asynchrony in moulting and derives a putative equivalence of probability of moulting,  $P^*$ , obtained from IGR experiments with mean daily development rate based on directly observed IMPs.

Given the theory described in the Appendix the bias in estimates of mean IMP using the “1/MR” and “IMP” methods of [4] is elucidated. This bias occurs when any of the 5-consecutive day IGR sampling windows do not fall completely within the distributional limits of date of next moult for the population from which the particular IGR experiment’s sample is drawn. In practice, this can result in the artifact of a  $P$  of zero using the “1/MR” method, which gives an undefined estimate of average IMP, if the sampling window starts and ends before the onset of next moult, giving a structural zero for the IGR experiment as opposed to a sample-caused zero where moults in the population at large occur in the window but the sample did not include any of those moults by chance. Alternatively, the IGR sampling window could overlap either lower or upper distributional limits and this will bias low the estimated value of  $P^*$ . In both cases this bias in the estimate the population value of  $P^*$  is due to an unrepresentative sampling frame. The Appendix shows this using an assumed discrete uniform distribution for next moult dates and the partial overlap at the lower distributional limit that results in an “effective” value of  $d$ ,  $d'$ , that is less than  $d$  but greater than zero. The bias is due to a reduced (effective) value of  $d$  and if  $d'$  is zero this corresponds to a structural zero in moult frequency for those incubation days occurring before the onset of first moults. An example of the bias in the estimate of mean IMP ( $1/P^*$ ) is given in the Appendix where for a value of  $P^*$  of 0.05, a  $d$  of 5, and  $d'$  of 3, the percentage bias in the naïve estimate (i.e. assuming an effective value of  $d'$  of 5 when it is actually 3) of population mean IMP is very large at 65%.

Note that there is no compensatory mechanism which could offset the chances of  $d'$  being less than  $d$  such as an equal chance of  $d'$  being equally larger than  $d$  in some sort of symmetry since in fact  $d'$  can never be greater than  $d$  only equal to  $d$  or less than  $d$  but greater than or equal to zero. Therefore, the bias will always be either positive or zero, if in this last case, the IGR sampling window is always within the distributional limits so that  $d=d'$ . This distribution implies that within the sampled population next moult dates reflect perfect asynchrony which implies that the probability of a moult is equal for each day within the distribution limits and is not, for example bell-shaped, which would imply some level of synchrony (Appendix). The claim in [4] that mixing of individuals from the different cohorts sampled in their “IMP” method overcomes the bias caused by any synchrony by somehow

balancing out under-estimates with over-estimates, is a statement of hope given that they gave no theoretical or empirical support of this assertion. Tarling et al. [4] note that some of the “1/MR” estimates of IMP are unrealistically large and can span a range of up to 120 days, whereas from laboratory-rearing the range quoted by [2] was 10 to 30 d so the above potential source of positive bias could explain these extreme estimates.

It is demonstrated here that bias is possible and substantial even when there is perfect asynchrony (Appendix) and given  $d'$  is unknown there is no remedy for such an unrepresentative sampling frame. Unknown levels of synchrony in timing of moults greatly complicate any further investigation of the method of [4].

### 2.3.4 Measurement error in temperature exposure

An additional fundamental difference between the datasets used by [1] compared to [4] is that while the temperature the krill are exposed to over the complete IMP of each individual is known in the laboratory, for the moults of krill captured in at-sea IGR experiments only the sea-surface temperature at time of capture and that of the onboard holding vessels is known. Since the sampled animals have had their last moult in the wild there could be a substantial number of days before capture so that the temperature of their immediate environment during some or most of the period from previous moult that occurred prior to capture is unknown. Therefore, the temperature used in the Tarling et al. (2006) model calibration is a measurement with error that may not only be imprecise but also biased across IGR experiments and individual captures. For calibration of a model of IMP, the direct observation of individual IMP combined with a known temperature environment for the complete interval of the inter-moult period was available as used by Kawaguchi et al. (2006) so the above issue with measurement error of the temperature exposure variable is avoidable.

The bias in the estimate of mean IMP via the bias in the estimate of  $P^* \{P(\mathbf{x}'\hat{\boldsymbol{\alpha}}, d)\}$  where from Section 2.3.1

$$P^* \{P(\mathbf{x}'\hat{\boldsymbol{\alpha}}, d)\} = 1 - [1 - P(\mathbf{x}'\hat{\boldsymbol{\alpha}}, d)]^{1/d}$$

using the “IMP” method incurred due to measurement error in one of the covariates, that is  $T$  with its true and unknown value replaced by SST at time of capture,  $T^{(SST)}$ , is not possible to quantify. This is because there is no way to relate  $T^{(SST)}$  for an individual IGR experiment to the true temperature variable that is predictive of IMP. This true variable is potentially average temperature over the entire temperature exposure profile,  $\bar{T}^{(true)}$ , for some relevant cohort for an individual IGR experiment. Ideally, the correct relationship between IMP and temperature involves the empirical **integral** over time of development rate as a function of exposure temperature for each small time unit over the period between the unobserved last moult at-sea and observed on-board moult (i.e. the IMP) [see eqn (2) of [10]]. Adjusting for bias in binary response regression due to measurement error can be carried out using a calibration dataset ([13]) for which, in this case both  $\bar{T}^{(true)}$  and  $T^{(SST)}$

are available across a number of IGR experiments. Alternatively, an assumed known distribution for the error,  $\varepsilon_i = T_i^{(SST)} - \bar{T}_i^{(true)}$ , in the  $i^{\text{th}}$  IGR experiment can be applied where it is required that the expected value of  $\varepsilon$  is zero and a prior estimate of variance is available. This combined with an empirical distribution for  $\bar{T}^{(true)}$  and applied via Monte Carlo simulation was used by [14] to estimate bias in parameter estimates for a **binomial** response regression with a complementary log-log link function but the principle also applies to the empirically similar logistic link function. However, no such calibration dataset or data that could inform assumptions on the distributions of  $\bar{T}^{(true)}$  and  $\varepsilon$  across IGR cohorts are available from such experiments. Therefore, the bias in estimation of  $P(\mathbf{x}'\hat{\boldsymbol{\alpha}}, d)$  and thus mean IMP,  $\tilde{i}_m(\mathbf{x}'\hat{\boldsymbol{\alpha}}, d) = 1/P^* \{P(\mathbf{x}'\hat{\boldsymbol{\alpha}}, d)\}$  due to measurement error  $\varepsilon$ , that may have non-zero expectation, exists and could be substantial but is unknown.

Apart from the above model calibration issue, in applying such a model to predict IMP as part of an algorithm to predict the average length-at-age trajectory that was first developed by [5] [and not [6], since they fail to cite the former with respect to growth increment and not just prediction of IMP given that [6] simply express “step 3” of the algorithm of [5] with a simple mathematical equation] some assumptions on variable temperature exposure are required. In this respect, Candy and Kawaguchi (2006) use average sea-surface temperature (SST) that varies only with season and do not incorporate variability due to diurnal movements of krill through the water column and how that affects temperature exposure relative to the necessarily simplistic method of applying average SST. Wiedenmann et al. (2008) refine the predictive algorithm of [5] by making some very simple assumptions about movement through the water column using the proportion of the day  $t$ ,  $\tau(t)$ , that is spent in “deep” versus and “shallow”,  $1 - \tau(t)$ , habitat **and use** the monthly averages of temperature of such habitats in the regions around South Georgia. Although these are crude approximations to the actual unobserved individual temperature exposure profiles or, more simply, their average across a cohort, they did allow some investigation of the sensitivity of a range of values assumed for  $\tau(t)$  on model outputs with the result that these outputs were found to have low sensitivity. However, it should be noted that **the** sensitivity analysis is contingent on the above simplistic investigation using just two broad depth strata, their monthly temperature averages, and very general assumptions on the diurnal movement pattern. The unobserved temperature profiles are a limitation for IMP model application but given the above there is an alternative for model calibration that is the data obtained from laboratory-based investigations. It should be noted that there are also limitations or caveats that apply to laboratory-based studies that derive from the unnatural conditions in which krill are kept which could directly affect observed IMPs. There are a number of these factors that are important, including potentially **sub-optimal** feeding regimes, **sub-optimal** light regimes, the effects of isolation, and limited movement in individual-level experimental vessels but the one of key interest here is accounting for exposure to water column differences in temperature given laboratory rearing uses constant temperature regimes.

It may be possible to simulate such temperature variations due to diurnal movement in krill research installations such as that of the Australian Antarctic Division [16]). This would allow predictions of IMP using models such as that of [5] to be compared to observed IMP under

the equivalent average temperatures for a range of variable temperature regimes. Such studies have been carried out on insect species in which parameters of a development rate function of temperature have been estimated given rearing in both laboratory experimental regimes (see eqn 7 of [10]) and using the abovementioned integration under ambient screen temperatures in terrestrial field stations (see eqn 6.5 of [15]). However, it is obviously not possible to record the temperature exposure profile for krill in their natural marine environment.

### 3. DISCUSSION

#### 3.1 Empirical results

What is clear from Figure 1 is that the meta-data means based on observed IMPs for individual krill cover the full range of temperatures from  $-0.5^{\circ}\text{C}$  to  $5^{\circ}\text{C}$  and mean development rates show no indication of a decline above  $1^{\circ}\text{C}$  while there is a quite dramatic decline in rate above this temperature for adult females for  $L$  of 55 mm for the IMP model of [4]. This is also the case for smaller values of  $L$ , as can be seen in terms of minimum IMP corresponding to maximum rate (see Fig S3 of [8]). The other obvious characteristic of the rate curves obtained from the IMP models in [4] is the completely different shape for adult males compared to adult females. The adult male rate curve declines monotonically up to an upper temperature of approximately  $3^{\circ}\text{C}$  which it should be noted covers most of the range of temperatures that Antarctic krill are exposed to across their range from sub-Antarctic to Antarctic (including oceanic and shelf zones, see Figure 2 of [17]). It is safe to say that such a declining rate versus temperature response curve for ectotherms, including the most studied order of Arthropoda of Insecta ([12]), is unique if it is in fact an accurate characterisation. Ectotherms invariably show a declining development time (i.e. analogous to IMP) and a corresponding increasing development rate with increasing temperature as long as temperatures do not exceed tolerable levels for the species. If the upper range of experimental temperatures negatively affects subject animals then some decline in rates will be observed ([12]). This is clearly not the case for the meta-data used by [1] as seen in Figure 1. In one of the studies contributing to the metadata used in [1], Polek and Denys (1982) state that "the length of intermolt periods (IMP's) was inversely proportional to temperature (20.10 d, SD=1.60, at  $0.12^{\circ}\text{C}$ ; 16.87 d, SD=1.68, at  $0.97^{\circ}\text{C}$ ; and 12.48 d, SD=0.90, at  $4.48^{\circ}\text{C}$ )". This agrees with the IMP model of [1].

Polek and Denys (1982) also found IMP increased proportionally with total length ( $L$ ) at the two lowest temperature regimes while somewhat similarly in Figure 3 of [3] mean IMP decreased with increasing temperature and there was a significant positive slope between IMP as regressed on  $L$  for a restricted range of  $L$  from 30 to 40 mm but only for the lowest temperature regime of  $-1^{\circ}\text{C}$  (thus the reason for excluding this regime of [3] from Figure 1). However, [4] model predictions of IMP increasing with  $L$  as shown in their Figure 7 for  $L$  up to 60 mm but the increase in IMP with  $L$  is greatest at the highest temperatures which is not consistent with [11] and [3] though in this last case the range of  $L$  was considerably restricted. Even so, the results from these last two studies are based on observed IMP as response variable and not on indirect estimates based on moult frequencies.

Note that a linearly increasing rate-temperature response model which has a zero ordinate value for a given temperature (i.e. an estimated temperature threshold below which development cannot occur) is the basis of commonly used degree-day mediated predictions of insect emergence dates ([9]).

Additionally, to support this conclusion that the estimated male response curve of [4] must be viewed as extremely unusual if not an artifact of serious bias in parameter estimates, the obvious question is what possible physiological mechanism would result in the sexual dimorphism where adult males show a fundamental different temperature response to that of adult females? Figure 3 of [3] shows predominately a very similar average response for males and females. Also, [8] fail to comment on the shape of the IMP model of [4] for immature females which like mature females has a concave-up shape with temperature for given  $L$  but the “up-swing” for  $T$  above  $2^{\circ}\text{C}$  is much greater for mature females for an  $L$  of 50 mm (Figure 8 of [4]) corresponding to the strong “down-swing” seen in Figure 1 here. The slight “up-swing” for immature females is likely to have been statistically indistinguishable from a monotonic decline as is the case for the model of [1] which could have been a consideration if [4] had presented 95% confidence bounds for their predicted regression curves as was provided by [1] for their IMP model. Further, the trends for males under 50 mm and juveniles are weak (Figure 8 of [4]) and if 95% confidence bounds had been presented about these trends they might have enclosed the no-trend line set at mean IMP for these particular sex and maturity classes. The question then arises as to why these sex/maturity classes do not express a strong response of IMP with temperature as is the case in [1] for adults and for ectotherms in general.

The predicted trend of increasing in IMP with increasing temperature for the models of [4] for males, juveniles up to  $3^{\circ}\text{C}$ , and above  $2^{\circ}\text{C}$  (i.e. the up-swing) for mature females given the contradictory evidence of the monotonically decreasing trend model of [1] obtained using directly observed individual IMPs, and the general weight of empirical evidence on the nature this relationship for ectotherms in general, as mentioned earlier, should give rise to suspicions that the above models of [4] give biased predictions of mean IMP given temperature and  $L$ . Therefore, there are serious concerns with the IMP models of Tarling et al. (2006) and the predictions they produce (see Figure 1) are likely to be spurious given the above considerations.

### 3.2 Theoretical considerations of bias

Tarling et al. (2006) justify that their “IMP” method overcomes the problems they note with the “1/MR” method with the following; “This had the effect of mixing the krill; individuals from experiments with molt-synchrony were combined with others without synchrony; situations in which molt-synchrony may have increased the probability of molting were combined with others in which the probability was decreased.” Firstly, it is important to note that the “IMP” method is equivalent to the “1/MR” method when aggregating the binary moult frequencies across animals with the same values of predictors  $L$  and  $T$  to obtain the number of moulted,  $m$ , versus not-moulted,  $N-m$ , for each value of  $L$  and  $T$  and then using a generalised “1/MR” method to regress the binomial  $M=m$ , conditional on  $N$ , on  $T$  and  $L$  using logistic regression. This is because this last-mentioned regression gives identical maximum likelihood estimates of regression parameters to the binary (disaggregated) data as long as the modelled

probability from the logistic regression remains constant within aggregations (i.e. this is easily shown using products of the Bernoulli likelihoods). Therefore, contrary to the above hope-filled assurances by [4], the “IMP” suffers the same shortcomings as the “1/MR” method and this is the basis of the flaws described in the “1/MR” method being equally applicable to the “IMP” method. Secondly, this assumption of positive biases in estimates of IMP **effectively** balancing negative biases due to an unknown level of mixing of animals that developed synchronously with those that developed asynchronously without any theoretical or empirical support was shown in theory in Section 2.3.3 and the Appendix to be invalid. This is a due, given this theory, to the effective incubation period,  $d'$ , never being greater than an IGR experiment's actual incubation period,  $d$ .

The directions of bias for the “1/MR” and “IMP” methods of [4] were shown to be positive for all sources but the source of bias described in Section 2.3.4 for which it is unknown due to there being no way of estimating the temperature exposure between last (at-sea) moult and the observed moult under experimental incubation at either the individual animal level or on average at the cohort level. For the positive biases it was also shown that the magnitude of the bias depends upon a number of factors as described. For the bias due to the mathematical error (Section 2.3.1) and due to transformation bias (Section 2.3.2) these could be removed by adjustments to the “1/MR” and “IMP” methods based on the formulae given. However, for the biases that depend on unknown quantities given in Sections 2.3 and 2.4 there is no remedy possible.

Further, in terms of combining models of IMP and IGR in body length, [8] in their section “Towards more general krill growth models” state; “Another example of uncertain mechanisms are the two IMP-based models that predict growth as an increment in body length at the event of moulting. For the model dynamics, **more frequent moulting implies higher growth rates whereas fewer moulting events correspond to less growth, when growth rates are positive**” (bolding added). This ignores a very important result obtained by [3] that for the first time, their study was able to confirm, by fitting a bivariate Linear Mixed Model, that compensation mechanisms exist between IMP and IGR and to quantify the degree of compensation at both the between- and within-animal levels. Specifically, “compensation mechanisms” are a natural feedback mechanism that prevents a short IMP combining with a large IGR from contributing to an unrealistically large growth increment and thus accumulated to give unrealistically fast progression of the predicted length-at-age curve to its asymptote or conversely prevents long IMP combining with small IGR to give unrealistically small growth increments. Brown et al. (2010) note that taking such compensation into account would allow modelling of growth trajectories in the wild to quantify the distribution of length given age. However, this would require growth modelling to be carried out at the individual animal level in order to take into account this compensation mechanism.

#### 4. CONCLUSION

The empirical evidence is that the temperature response curves estimated by **the** “IMP” method of [4] **using** binary moult frequency data show (i) completely atypical response curves with temperature that are **of a** fundamental different form **to** the curves derived from observed IMPs, (ii) are highly heterogeneous across the sexes, and (iii) have weak to close to non-

existent trends for smaller sizes (e.g. 40 mm) and juveniles. Combined with this are the four sources of potentially serious bias in both the “1/MR” and “IMP” methods that were described where for two of these there is no possibly remedy.

Given the above, caution is warranted against further uncritical promulgation of the flawed models of [4] for predictive purposes and further fundamental research on IMP and growth is an urgent priority, especially given excellent and accessible research aquariums such as that at the Australian Antarctic Division at Kingston, Tasmania (Kawaguchi et al., 2010). Potentially the model of [1] could be updated with the data from [3] on adult males and females and any other similar more recently published and future studies. Also, the feasibility of experimental studies of IMP under variable temperature regimes that could potentially simulate diurnal movement within the water column is well worth investigating in order to validate any affect of such variation on predictions from models such as that of [1] calibrated from observations of IMP under constant temperature rearing.

## ACKNOWLEDGEMENTS

This research was funded by Scandy Statistical Modelling Pty Ltd.

## COMPETING INTERESTS

The Author has declared that no competing interests exist.

## COMPETING INTERESTS DISCLAIMER:

Authors have declared that they have no known competing financial interests OR non-financial interests OR personal relationships that could have appeared to influence the work reported in this paper.

## REFERENCES

- [1] Kawaguchi S, Candy S, King R, Naganobu M, Nicol S. Modelling growth of Antarctic krill. I. Growth trends with sex, length, season, and region. *Marine Ecology Progress Series* 2006;306:1–15. <https://doi.org/10.3354/meps306001>
- [2] Buchholz F. Moulting cycle and growth of Antarctic krill *Euphausia superba* in the laboratory. *Marine Ecology Progress Series*. 1991;69:217–229.

- [3] Brown M, Kawaguchi S, Candy S, Virtue P. Temperature effects on the growth and maturation of Antarctic krill (*Euphausia superba*). *Deep Sea Research Part II: Topical Studies in Oceanography*. 2010 57:672–682. <https://doi.org/10.1016/j.dsr2.2009.10.016>
- [4] Tarling GA, Shreeve RS, Hirst AG, Atkinson A, Pond DW, Murphy EJ, et al. Natural growth rates in Antarctic krill (*Euphausia superba*): I. Improving methodology and predicting intermolt period. *Limnology and Oceanography* 2006;51:959–972. <https://doi.org/10.4319/lo.2006.51.2.0959>
- [5] Candy S, Kawaguchi S. Modelling growth of Antarctic krill. II. Novel approach to describing the growth trajectory. *Marine Ecology Progress Series*. 2006;306:17–30. <https://doi.org/10.3354/meps306017>
- [6] Wiedenmann J, Cresswell K, Mangel M. Temperature-dependent growth of Antarctic krill: predictions for a changing climate from a cohort model. *Marine Ecology Progress Series*. 2008;358:191–202. <https://doi.org/10.3354/meps07350>
- [7] Atkinson A, Shreeve RS, Hirst AG, Rothery P, Tarling GA, Pond DW, et al. Natural growth rates in Antarctic krill (*Euphausia superba*): II. Predictive models based on food, temperature, body length, sex, and maturity stage. *Limnology and Oceanography*. 2006;51:973–987. <https://doi.org/10.4319/lo.2006.51.2.0973>
- [8] Bahlburg D, Thorpe SE, Meyer B, Berger U, Murphy EJ. An intercomparison of models predicting growth of Antarctic krill (*Euphausia superba*): The importance of recognizing model specificity. *PLoS ONE*. 2023;18(7): e0286036. <https://doi.org/10.1371/journal.pone.0286036>
- [9] Stinner RE, Gutierrez AP, Butler Jr GD. An algorithm for temperature-dependent growth rate simulation. *Canadian Entomologist*. 1974; 106:519–524
- [10] Régnière J, Powell J, Bentz B, Nealis V. Effects of temperature on development, survival and reproduction of insects: experimental design, data analysis and modeling. *Journal of insect physiology*. 2012; 58: 634-647. doi:10.1016/j.jinsphys.2012.01.010
- [11] Poleck TP, Denys CJ. Effect of temperature on the moulting, growth and maturation of the Antarctic krill *Euphausia superba* (Crustacea: Euphausiacea) under laboratory conditions. *Marine Biology*. 1982;70:255–265.
- [12] Rebaudo F, Rabhi VB. Modeling temperature-dependent development rate and phenology in insects: review of major developments, challenges, and future directions. *Entomologia Experimentalis et Applicata*. 2018; 166: 607-617. DOI: 10.1111/eea.12693
- [13] Carroll RJ, Wand MP. Semiparametric Estimation in Logistic Measurement Error Models. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1991;53: 573-585.
- [14] Candy SG. Empirical Binomial Sampling Plans: model calibration and testing using William's method III for generalised linear models with overdispersion. *Journal of Agricultural, Biological and Environmental Statistics*. 2002;7: 373-388.
- [15] Candy SG. Predictive models for integrated pest management of the leaf beetle *Chrysophtharta bimaculata* in *Eucalyptus nitens* plantations in Tasmania. 2000; University of Tasmania. Thesis. <https://doi.org/10.25959/23214194.v1>
- [16] Kawaguchi S, King R, Meijers R, Osborn JE, Swadling KM, Ritz DA, Nicol S. An experimental aquarium for observing the schooling behaviour of Antarctic krill (*Euphausia superba*). *Deep Sea Research Part II: Topical Studies in Oceanography*. 2010; 57: 683-692. doi:10.1016/j.dsr2.2009.10.017
- [17] Candy SG. Long-term trend in mean density of Antarctic krill (*Euphausia superba*) uncertain. *Annual Research & Review in Biology*. 2021;36: 27-43

## APPENDIX

### A distribution theory, structural zeros in sample moult frequency and effective incubation period

#### A distribution theory

Consider a single IGR experiment and its sample of a wild population of adult Antarctic krill with a sample,  $s(N)$ , of size of  $N$  from a total population,  $\square$ , possibly a krill swarm of unknown size in terms of total number. Let  $H_0$  be the date and time the sample is taken, which can be quantified as the number of days from an earlier fixed and arbitrary date and time, and each sample animal is incubated in individual jars for the next  $d$  consecutive days (for the following whole or integer values for number of days in a period will be considered so that fractions of a day are not accounted for though this does not limit the generality of results).

Let the date of the earliest next moult for the population relative to  $H_0$  be  $H_0 + t_0^{(\square)}$  and the corresponding date for the sample be  $H_0 + t_0^{(s)}$  where  $H_0 + t_0^{(s)} \geq H_0 + t_0^{(\square)}$  where  $t_0^{(\square)}$  and  $t_0^{(s)}$  are measured in days. Let the date of the latest next moult for the population be  $H_0 + t_e^{(\square)}$  and the corresponding date for the sample be  $H_0 + t_e^{(s)}$  where  $H_0 + t_e^{(s)} \leq H_0 + t_e^{(\square)}$ . The period of the IGR incubations is assumed to be common and between dates  $H_0$  and  $H_0 + d$ . The above assumes all date parameters refer to the start of the day while the IGR incubation period,  $d$ , refers to completed days. Since for the following only periods representing number of days between dates are considered the value of  $H_0$  can be considered zero without loss of generality. The distribution of the date,  $H$ , of next moult for a random animal from the population is assumed to have a discrete uniform distribution with range  $(t_0^{(\square)}, t_e^{(\square)})$  so that

$$\Pr(H = k) = P^*$$

where  $k \in \{t_0^{(\square)}, t_0^{(\square)} + 1, \dots, t_e^{(\square)}\}$ . Further if we specify  $t_e^{(\square)} = t_0^{(\square)} + t_m - 1$  where the mean IMP for  $\square$  is  $t_m$  and an integer then  $P^* = 1/t_m$ . Note also that there are  $t_m = t_e^{(\square)} + 1 - t_0^{(\square)}$  days in the interval including starting day  $t_0^{(\square)}$  so that summing the constant daily value  $P^*$  across  $t_m$  days gives a value of unity. So given these distributional properties the parallel of  $P^*$  with a constant daily development rate (i.e. proportion), as in Figure 1 for the case of observed IMPs and their corresponding means across the range of constant temperature regimes as used by [1] to estimate their model of IMP, is clear. This extends to the accumulation of the seasonally temperature dependent development rates (i.e.  $1/t_m$ ) across the austral summer in order to achieve cumulative integer values (i.e. an accumulation of a set of consecutive sums each to unity) that specify IMPs for the consecutive sequence of moults as given in step (1) of the algorithm of [5]. However, this putative theoretical

equivalence between  $P^*$  and constant daily development rate requires very restrictive assumptions on the distribution of  $H$ . The above discrete uniform distribution implies asynchrony in moulting since if for a subset of consecutive days  $k' \in \{t_0^{(\square)} + r, \dots, t_0^{(\square)} + r'\}$  where  $r < r' < (t_e^{(\square)} - t_0^{(\square)})$  and where the complement is  $k''$  so that  $k \in \{k', k''\}$ , then if  $\Pr(H = k') > \Pr(H = k'')$ , so that  $H$  no longer follows a discrete uniform distribution, then this would indicate a degree of synchrony of moulting. This follows since given the above more animals moult within a limited time period than expected under asynchrony.

In Section 2.3.1 it was shown that  $P^* = 1 - [1 - P(d)]^{1/d}$  and the only situation where the formula of [4],  $P_{T06}^* = P(d)/d$ , is correct is if the incubation period is restricted to a single day, so that  $d$  is 1.

Also note that the above theory requiring asynchrony and therefore a constant  $P^*$  for a given IGR experiment also assumes a constant temperature. This assumption is implemented in [4] by using a value of  $T$  of "sea surface temperature at time of capture" and as noted in Section 2.3.4 this is not the temperature the animals are exposed to over the whole inter-moult period prior to observed moults.

### Structural zeros in sample moult frequency and effective incubation period

The IGR sampling window is, given the above, from date of capture,  $H_0$ , the period between dates  $H_0$  and  $H_0 + d$  of  $d$  days. Again after setting  $H_0$  to zero without loss of generality, for  $d < t_0^{(\square)}$  then the moult frequency for the IGR sample is  $m \equiv 0$  (i.e. number of observed krill moulting in the  $d$ -day incubation period is zero) out of the total sample of  $N$ . This gives an undefined estimate of  $t_m$ ,  $\hat{t}_m = 1/\hat{P}^*$  (Section 2.3.2), since  $\hat{P}(d) = \hat{P}^* = P^* \equiv 0$ . These zeros are structural zeros since no matter how many IGR samples are taken or how large  $N$  is, the value of  $m$  is always zero since  $\Pr(m=0) = (1 - P^*)^d \equiv 1$ . For the case  $t_0^{(\square)} \leq d < t_0^{(s)}$  then for the realised sample  $m$  is zero but for other unrealised samples the case could be  $t_0^{(\square)} \leq t_0^{(s)} \leq d$  so that  $m$  can be greater than zero and in fact for  $d \geq t_0^{(\square)}$  then  $\Pr(m=0) = (1 - P^*)^d < 1$  since  $0 < P^* < 1$ . **Therefore**, zero values of  $m$  in this last case are sampling zeros and not structural zeros. If it is always the case that  $t_0^{(\square)} \leq 0$  and  $t_e^{(\square)} > d$  so that the IGR sample "window" falls within the distribution limits of  $H$ , then given the above distributional theory

$$E(m/N | N, d) = P(d) = 1 - (1 - P^*)^d = 1 - (1 - 1/t_m)^d$$

and given appropriate adjustment for bias induced by the inverse transformation and after applying the correct formula above (see Section 2.3.1) the "1/MR" method gives an unbiased

estimate of IMP. This conclusion also applies to estimates of mean IMP as a function of  $T$  and  $L$  obtained using the “IMP” method of [4].

The above cases with respect to the IGR sample window of either completely outside the date of next moult distributional limits,  $d \leq t_0^{(\square)}$ , and completely within those limits,  $t_0^{(\square)} \leq 0$  and  $t_e^{(\square)} > d$ , have been considered above. However, there is a third possible case of  $0 \leq d' < d$  where  $d' = d - t_0^{(\square)}$  and  $0 < t_0^{(\square)} < d$  where the IGR sampling window straddles  $t_0^{(\square)}$ . In this case the “effective” incubation period for which it is possible for a moult to occur,  $d'$ , is shorter than the nominal period,  $d$ , by  $d - d'$  days. For these days prior to  $t_0^{(\square)}$  then moulting frequency summed over these days is a structural zero. The probability of moult for the IGR  $d$ -day incubation is  $P(d) = 1 - (1 - P'^*)^d$  where  $P'^*$  is not the correct population value and  $d$  is not the true effective incubation period where a moult is possible for all days in the IGR experimental period  $d$ . The true population probability,  $P^*$ , is given by  $P(d, d') = 1 - (1 - P^*)^{d'}$ . Then, since  $P(d, d') = P(d) = E(m/N)$ , we can equate the above formulas to give

$$1 - (1 - P^*)^{d'} = 1 - (1 - P'^*)^d.$$

This equality can be expressed in terms of the true expected value of  $t_m$  and the naïve and incorrect value  $t'_m$  as

$$1 - (1 - 1/t_m)^{d'} = 1 - (1 - 1/t'_m)^d$$

giving

$$t'_m = \left[ 1 - (1 - 1/t_m)^{d'/d} \right]^{-1}.$$

The percentage bias of the naïvely calculated value of population mean IMP,  $t'_m$ , is given by

$$B = 100(t'_m - t_m) / t_m.$$

For example, for a true mean IMP of 20 d giving  $P^* = 0.05$ , and  $d$  of 5, and for a  $d'$  of 3 then  $B$  is 64.98%. For  $d'$  of 4 then  $B$  is 24.36%.

The key point is that occurrence of zero moult frequencies caused by sampling does not bias estimates of mean IMP, given the above assumption of a specific discrete uniform distribution and the assumption that for all IGR experiments the IGR sampling window falls completely within distributional limits,  $t_0^{(\square)} \leq 0$  and  $t_e^{(\square)} > d$ , if sampling is random since

$d' = d$ . In the cases described above where the sampling frame is not representative due to effective incubation period,  $d'$ , being less than the nominal incubation period,  $d$ , (where structural zeros correspond to  $d' = 0$ ) then a large positive bias is incurred. Note also that, given the above caveats, if  $d' = d$  then the bias is zero.

Given that there is no compensatory mechanism for the above biases (i.e. values of  $d'$  cannot be greater than  $d$ ) and the sampling frame cannot be informed due to unknown values of  $t_0^{(\square)}$  and  $t_e^{(\square)}$  and thus  $d'$  combined with the very restrictive assumption of asynchrony imposed via the above discrete uniform distribution the conclusion is that the "1/MR" and "IMP" methods should not be used.

UNDER PEER REVIEW